Extracting factors from heteroskedastic asset returns[☆]

Christopher S. Jones*

The William E. Simon Graduate School of Business Administration, University of Rochester, Rochester, NY 14627, USA

Abstract

This paper proposes an alternative to the asymptotic principal components procedure of Connor and Korajczyk (J. Financial Econom. 15 (1986) 373) that is robust to time series heteroskedasticity in the factor model residuals. The new method is simple to use and requires no assumptions stronger than those made by Connor and Korajczyk. It is demonstrated through simulations and analysis of actual stock market data that allowing heteroskedasticity sometimes improves the quality of the extracted factors quite dramatically. Over the period from 1989 to 1993, for example, a single factor extracted using the Connor and Korajczyk method explains only 8.2% of the variation of the CRSP value-weighted index, while the factor extracted allowing heteroskedasticity explains 57.3%. Accounting for heteroskedasticity is also important for tests of the APT, with p-values sometimes depending strongly on the factor extraction method used. © 2001 Elsevier Science S.A. All rights reserved.

JEL classification: G12; C13

Keywords: Principal components; Factor analysis; Heteroskedasticity; Arbitrage pricing theory

Nomenclature

T	number of sample time periods
K	number of factors
N	number of firms (in sample)
n	number of firms (asymptotic index variable)
d_t	average idiosyncratic variance at time t
$\bar{d}$	average d_t
f_t^k	time t realization of the kth rotated factor (the (k, t) element of F)
h_t	$K \times 1$ vector of time t unrotated factor (the tth column of H)
r_t^n	$n \times 1$ vector of time t excess returns
ε_t^n	$n \times 1$ vector of time t idiosyncratic errors
γ_t	$K \times 1$ vector of time t factor risk premia
B	$N \times K$ matrix of factor betas
$\hat{B}$	$N \times K$ estimate of B
B^n	$n \times K$ matrix of factor betas
D	$T \times T$ diagonal matrix of average idiosyncratic variances
$\hat{D}$	$T \times T$ estimate of D
E	$N \times T$ matrix of idiosyncratic errors
E^n	$n \times T$ matrix of idiosyncratic errors
F	$K \times T$ matrix of rotated realizations of factors ($M^{-1/2}H$)
$\tilde{F}$	$K \times T$ estimate of F
$\hat{F}$	$K \times T$ orthonormalized transformation of $\tilde{F}$
H	$K \times T$ matrix of realizations of (unrotated) factors
M	probability limit of $(1/n)B^{n'}B^n$
R	$N \times T$ matrix of excess returns
R^n	$n \times T$ matrix of excess returns
R^m	$m \times T$ matrix of excess returns, where some values are missing
$\mathscr{I}^m$	$m \times T$ matrix indicating what values of R^m are non-missing
$\overline{R^2}$	the average of K R-squares
$\underline{R^2}$	the minimum of K R-squares
X^n	$(1/n)H'B^{n'}B^nH$
Y^n	$(1/n)H'B^{n'}E^n$
Z^n	$(1/n)E^{n'}E^n$
Ω	$N \times N$ diagonal matrix of unconditional idiosyncratic variances
$\hat{\Omega}$	$N \times N$ estimate of Ω
Σ_t^n	$n \times n$ conditional covariance matrix of time t residuals

1. Introduction

The appeal of Ross's (1976) Arbitrage Pricing Theory (APT) rests in large part on the minimal assumptions necessary to derive it: that there are many

assets, that trading is costless, and that a factor model drives the returns generating process. To analyze the model empirically, however, one must impose additional structure. First, as Shanken (1982) emphasizes, obtaining an exact rather than approximate factor pricing relation requires an assumption about market equilibrium. One example is Chamberlain's (1983) assumption that the tangent portfolio is well-diversified. Second, some assumptions that ensure statistical identification are necessary. Possibilities include assuming the constancy of factor betas or selecting a particular set of factors, as was done in the well-known paper by Chen et al. (1986).

While prespecifying factors is a natural first step for investigating the sources of priced risks, it not an obvious starting point for tests of market efficiency. Because the APT is silent on what the factors driving returns are, prespecification of the factors leads to a joint test of market efficiency and the particular set of factors chosen.

Fortunately, numerous papers have de demonstrated that prespecifying the factors is an unnecessary step in testing the APT. One strategy is to assume that returns are Gaussian, that their covariances are constant, and that all comovement in asset returns can be attributed to factor movements. Given these restrictions, it is possible to use maximum likelihood factor analysis to estimate factor loadings. Roll and Ross (1980) used these loadings to test exact APT pricing with constant factor risk premia using simple cross-sectional regression tests, while Lehman and Modest (1988) use a more sophisticated factor decomposition algorithm to consider much larger cross-sections of returns under the same assumptions.

The normality of returns and the diagonal covariance matrix assumed in these papers is, however, neither empirically likely nor a theoretical requirement of the APT. Extending the results of Chamberlain and Rothschild (1983), Connor and Korajczyk (1986) introduced a novel method for factor extraction, which they called asymptotic principal components. The procedure avoids the undesirable diagonality assumption and is not likelihood-based, so it allows for non-Gaussian returns. In addition, the Connor and Korajczyk approach lends itself naturally to the consideration of time-varying factor risk premia.

The Connor and Korajczyk method remains popular in empirical finance and economics. Recent applications of the method include papers by Brennan et al. (1998), Brown et al. (1998), Chan et al. (1998, 1999), Elton (1999), Elton et al. (1998), Ferson and Korajczyk (1995), Jagannathan and Ma (2001), Levine and Zervos (1998), McCulloch and Rossi (1990, 1991), and Pastor and Stambaugh (1999).

Although significantly less restrictive than maximum likelihood factor analysis, the Connor and Korajczyk method maintains the assumption that the covariance matrix of the factor model residuals is constant through time. Of course, evidence against homoskedasticity in aggregate equity returns is

immense. French et al. (1987) observe that market volatility makes large and persistent deviations from its long-run mean, while Schwert (1990) notes that short-term movements in market volatility can be even more severe. Although time series heteroskedasticity in the idiosyncratic component of the returns of less diversified portfolios or individual firms is not as extensively documented, several studies suggest this is an empirical regularity as well. Schwert and Seguin (1990) find evidence that the volatilities of size-ranked portfolios exhibit movements that are independent of aggregate stock market volatility, while in a firm-level analysis Campbell et al. (2001) find that the volatilities of individual firms have market, industry, and idiosyncratic components, and that all three exhibit substantial time variation. Furthermore, firm level volatility dynamics are correlated, so that the average idiosyncratic variance across firms can change considerably from one month to the next.

Even without the results of Campbell et al. (2001), it would be compelling to think that at least some changes in aggregate volatility are associated with shifts in average residual volatility as well. With their results, however, it seems more clear that residuals are not as well-behaved as has been assumed. If this is the case, then Connor and Korajczyk's method is no longer valid, and we should expect the resulting factor estimates to explain neither the time variation in realized returns nor the cross-sectional variation in expected returns.

This paper demonstrates that accounting for residual heteroskedasticity is both important and straightforward. The method of Connor and Korajczyk (1986, hereafter CK), while adequate in many instances, often appears to severely misestimate the true latent factors, particularly in data from the last 10 years. The alternative proposed here, which I will call heteroskedastic factor analysis (HFA), is almost as easy to implement, requires no additional assumptions to those in CK, and is shown to generally outperform the CK method.

Section 2 of the paper will describe the method in detail and compare it with that of CK. Relative performance of the heteroskedastic factor analysis and CK methods is evaluated in Section 3 through a simple simulation study. The performance of the two methods in explaining actual stock market data is assessed in Section 4. Section 5 examines some basic implications for tests of the APT, while Section 6 concludes.

2. Heteroskedastic factor analysis

The central convergence result of CK states that given a large enough set of assets returns whose residuals are sufficiently uncorrelated, the realizations, over a fixed time period, of the unobserved factors (up to a nonsingular translation) may be recovered to any desired precision. This enables CK to test

the APT as though the true factors were known. In this section I review this result and show that a similar convergence result may be obtained in the case of residual heteroskedasticity.

2.1. Extracting the factors

Following CK, I assume a countably infinite set of assets. Under the equilibrium version of the APT (see Connor, 1984), the APT pricing relation holds exactly, so that for the first n assets we have

$$r_t^n = B^n h_t + \varepsilon_t^n, \tag{1}$$

$$\mathrm{E}[h_t] = \gamma_t, \tag{2}$$

$$\mathrm{E}[\varepsilon_t^n | h_t] = 0, \tag{3}$$

$$\mathrm{E}[\varepsilon_t^n \varepsilon_t^{n'}] = \Sigma_t^n, \tag{4}$$

where r_t^n is the $n \times 1$ vector of excess returns at time t, B^n is the $n \times K$ matrix of factor loadings, h_t is the $K \times 1$ vector of time t factor realizations, γ_t is the $K \times 1$ vector of time t factor risk premia, and ε_t^n is the time t $n \times 1$ vector of serially uncorrelated residuals.

Following Connor and Korajczyk, I will stack T columns together and rewrite Eq. (1) in matrix form,

$$R^n = B^n H + E^n, \tag{5}$$

in order to analyze the convergence of the $T \times T$ matrix $(1/n)R^{n'}R^n$, which is equal to

$$\frac{1}{n}H'B^{n'}B^nH + \frac{1}{n}H'B^{n'}E^n + \frac{1}{n}E^{n'}B^nH + \frac{1}{n}E^{n'}E^n$$
$$\equiv X^n + Y^n + Y^{n'} + Z^n. \tag{6}$$

As in Connor and Korajczyk (1986) I assume that the matrix $(1/n)B^{n'}B^n$ has a probability limit M, implying that X^n converges to $H'MH$. I also assume that M is full rank. It is notationally more economical to write this limit instead as $F'F$, where $F \equiv M^{1/2}H$ is the set of factors that would yield an equivalent realization of returns were the factor betas equal to $B^nM^{-1/2}$. Since factors and betas are rotationally indeterminate, changing the object of our estimation from H to F is inconsequential, so the remainder of the paper focuses on the estimation of F.

Assumption 7 of Connor and Korajczyk (1986), along with the assumed independence of the factors and residuals, allows the law of large numbers to be invoked for the Y^n and $Y^{n'}$ terms. Because the residual vector ε_t^n is mean zero, these two terms have probability limits of zero.

Their assumption 10 states that there exists an average residual variance and that it is constant through time. I relax this assumption in a minor but critical way. I still require that there exists an average idiosyncratic variance in each period, but I allow this quantity to change freely from one period to the next. Since residuals are still not serially correlated, the strong large law of numbers can again be invoked to show that the probability limits of the off-diagonal elements of Z^n are zero. Formally,

$$\plim_{n \to \infty} \frac{1}{n} \varepsilon_s^n \varepsilon_t^n = 0 \tag{7}$$

for $s \neq t$.

The tth diagonal element of Z^n, given by $(1/n)\varepsilon_t^n \varepsilon_t^n$, converges to the average idiosyncratic variances in period t, so we define

$$d_t = \plim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \Sigma_t^n(i, i). \tag{8}$$

The probability limit of Z^n can therefore be written as a diagonal matrix D, where $D(t, t) = d_t$. Under the original Connor and Korajczyk assumptions, $d_s = d_t$, so D could be written as dI. So while $(1/n)R^{n'}R^n$ converged to the matrix $F'F + dI$ under the CK assumptions it converges instead to the somewhat less restricted $F'F + D$ under my modified assumptions.

It should be noted that some forms of time series heteroskedasticity in return residuals are not incompatible with the Connor and Korajczyk assumptions. In particular, if idiosyncratic volatilities evolve independently across firms, then the average variance across a large sample of firms will still remain constant. It is only when the dynamics of idiosyncratic variances have a common component (as Campbell et al., 2001, find empirically) that their average will exhibit time variation.

The cost of generalizing the assumption of constant variances is that the matrix F can no longer be recovered by simply computing the eigenvectors of the cross-product matrix, $F'F + D$, as it could be in the Connor and Korajczyk setup. Fortunately, other methods for decomposing the more general for into its components F and D are well known. The estimation procedure of Joreskog (1967), for instance, can be used to compute the maximum likelihood estimate of the matrix B from the sample covariance matrix of R_t. While this estimator cannot be used when n is large relative to T or when the residual covariance matrix is nondiagonal, the computational algorithm proposed by Joreskog will be used here.

In the maximum likelihood factor analysis of Joreskog, the residuals ε_t^n are assumed independent and identically distributed through time, and the factor risk premia are assumed constant. The covariance matrix of returns then has a probability limit of $B'B + \Omega$, where Ω is now the assumed *diagonal* residual covariance matrix and where, without loss of generality, $\mathrm{Cov}(f_t, f_t') = I$.

Joreskog's iterative algorithm recovers the maximum likelihood estimates of B and Ω. The consistency of MLE implies that as the estimated covariance matrix approaches its probability limit, $B'B + \Omega$, the resulting estimates of B must converge to the true B matrix.

Because the probability limit of the cross-product matrix (which does allow for heteroskedasticity, residual correlations, and time-varying risk premia) is $F'F + D$, which has the same for as Joreskog's covariance matrix, the same iterative procedure that Joreskog used to extract B can be used here to extract F. This procedure follows several steps:

1. Compute the cross-product matrix $(1/n)R^{n'}R^n$ and denote it as C.
2. Guess an initial estimate, $\hat{D}$, of the diagonal matrix D.[1]
3. Collect the K eigenvectors corresponding to the K largest eigenvalues of the matrix $\hat{D}^{-1/2}C\hat{D}^{-1/2}$.[2] Let Λ be the diagonal matrix with the K largest eigenvalues on the diagonal, arranged in descending order. Denote the matrix of eigenvectors by V, where the ith column of V is the eigenvector corresponding to the ith diagonal element of Λ.
4. Compute an estimate of the factor matrix $\tilde{F}$ as $\hat{D}^{1/2}V(\Lambda-I)^{1/2}$.
5. Compute a new estimate $\hat{D}$ as the diagonal of $C - \tilde{F}'\tilde{F}$.
6. Return to 3 to iterate again or terminate the process if estimates have converged.[3]

Finally, to facilitate comparison with the Connor and Korajczyk factors, which are naturally written as an orthonormal collection, I transform the extracted factors $\tilde{F}$ into an orthonormal set $\hat{F}$. The first column of the transformed factor matrix $\hat{F}$ is set equal to the normalized first column of $\tilde{F}$. For every $1 < k \leqslant K$, I regress $\tilde{F}_k$ on $\hat{F}_1$ through $\hat{F}_{k-1}$ (without an intercept) and set $\hat{F}_k$ equal to the normalized residual of this regression. This produces a $T \times K$ matrix $\hat{F}$ for which $\hat{F}'\hat{F} = I$.

2.2. Firms with missing observations

Although the procedure described can potentially make use of extremely large cross sections of firm-level returns data, it is limited by the requirement that each firm in the sample have no missing observations. Connor and Korajczyk (1987) show, however, that it is fairly straightforward to incorporate firms that are only observed for a part of the sample, and their

---

[1] In this paper I use the 0.5 times the diagonal of C, implying that half the return variance comes from the idiosyncratic component.

[2] Since eigenvectors will be computed only from symmetric matrices, we may assume without loss of generality that all eigenvectors extracted from such matrices for an orthonormal set.

[3] As discussed in Knez et al. (1994) it is possible for the algorithm to result in an estimate $\hat{D}$ that contains some negative elements. In the applications in this paper, this problem does not occur.

results are generally supportive of the value of this modification of their original procedure. The same approach to missing data can be adopted when residuals are assumed heteroskedastic.

Following their exposition, let m denote the total number of firms with some returns data observed within the sample period. Let R^m denote the $m \times T$ matrix of excess returns, where all missing values are replaced by zeros, and $\mathscr{I}^m$ the $m \times T$ indicator matrix containing a one wherever the corresponding element of R^m is nonmissing and zero otherwise. If all of the returns in R^m are generated according to Eq. (1), the limit of $(R^{m'}R^m) \div (\mathscr{I}^{m'}\mathscr{I}^m)$ is identical to that of $(1/n)R^{n'}R^n$, where $\div$ denotes element-by-element division.[4]

When there are no systematic differences between the firms with missing observations and those with none, then we would expect the elements of $(R^{m'}R^m) \div (\mathscr{I}^{m'}\mathscr{I}^m)$ to be computed with lower variance because of the larger data sample used. As Connor and Korajczyk (1987) discuss, however, this is not necessarily the case, especially since the firms with missing observations are more likely to be small. Because small firm returns are often highly volatile, incorporating more of them could possibly increase the variance of the resulting estimator.

It is also possible that changing the composition of firms mid-sample could induce average idiosyncratic variances to change even though individual firm residuals are homoskedastic. Evidence presented in Safdar (2000) suggests that this is likely. Maintaining a constant sample of firms should therefore lessen the relevance of the method introduced in this paper. Since most results will be presented for data samples without missing observations, the importance of accounting for residual heteroskedasticity may sometimes be understated.

2.3. Data

Data used in the paper consist of excess returns on stocks traded on the NYSE, Amex, and Nasdaq exchanges as recorded in the CRSP monthly stock file. Following Connor and Korajczyk (1988), I consider a number of five-year intervals, with the first interval (1979–1983) corresponding to the last one considered by Connor and Korajczyk.

Five-year intervals were chosen for several reasons. First, different time periods display very different patterns of market and firm-level volatility, as shown by Campbell et al. (2001). To the extent that changes in volatility regimes influence the efficacy of one extraction method over another, using four separate intervals should provide some evidence for the robustness of each method. Second, the relatively short five-year intervals should minimize violations of the assumption that the betas are constant and should reduce the

---

[4] Observe that if all returns are observed, then $\mathscr{I}^m$ is an $m \times T$ matrix of ones, and that $\mathscr{I}^{m'}\mathscr{I}^m$ will be a $T \times T$ matrix where each element is equal to m.

severity of the survival bias caused by including only firms with no missing data.

In addition, I will also consider the 20-year sample that merges the smaller intervals. Because the vast majority of firms have an incomplete time series of returns over this period, the data set of firms with no missing observations is relatively small. I therefore attempt to improve the factor extraction over this period by also using a sample that also contains returns of firms with missing observations.

Table 1 shows that the five-year intervals differ in several ways. Panel A shows that the number of firms traded in these markets has generally increased over the past 20 years, with the greatest increases coming from Nasdaq. In the 1979–1983 sample, Nasdaq stocks comprised 35% of the sample, growing to 48% by 1994–1998.

In the 20-year samples, requiring a full time series of returns restricts the sample to just 1281 firms. If missing observations are allowed, then the sample grows dramatically, with over 19,000 firms contributing some returns data to the analysis. Because they have missing values, the number of observations increases more moderately as these firms are added, but the expanded data set still averages roughly five times the number of returns per month as the sample with no missing observations. The table also shows that most of these additional firms are traded on Nasdaq, making the composition of the larger sample much different from the sample without missing observations, which is comprised primarily of much larger firms.

Panel B of Table 1 shows that these changes in the composition of the sample have coincided with an increase in the level of residual heteroskedasticity. Assuming either one, three, six, or 12 factors, the factor extraction procedure described above was used to produce an estimate of D. The square roots of the 60 or 240 months of estimated average idiosyncratic variances (the diagonal elements of D) were then used to compute the standard deviation of $\sqrt{d_t}$. This measure of residual heteroskedasticity is about 2.5 times higher over the 1989–1993 interval than it is for 1979–1983 or 1984–1988, suggesting that violations of the homoskedasticity assumption have become more relevant since the Connor and Korajczyk (1986, 1988) papers were written.

Interestingly, heteroskedasticity is lowest for the relatively select 20-year sample of firms with no missing observations, but it is relatively large over the same period when firms with missing observations are included. This suggests that changes in average residual variance are primarily due to the addition of new firms to the sample, a finding consistent with Safdar (2000).

Finally, Panel C reports the autocorrelations of $\sqrt{d_t}$. As in other studies, volatility is persistent, with autocorrelations ranging from 0.21 to 0.5. Since the asymptotic standard error given a null of zero autocorrelation is at most 0.13, most of these correlations are statistically significant at 5% levels.

Table 1
Descriptive statistics

The primary data set is comprised of firms from NYSE, Amex, and Nasdaq over the 20-year period from 1979 to 1998. Four five-year subsamples are considered in which a firm is included as long as it has a complete set of returns over that interval. In addition, a 20-year sample is formed of firms that have 20 complete years of returns data. Finally, a 20-year sample of all firms, including those with missing observations, is denoted by the m superscript. Panel A describes the number of firms in each one of the samples. In order to measure of the degree of residual heteroskedasticity in each sample, the heteroskedastic factor extraction algorithm of Section 2.1 was applied to each sample for various values of K (the number of factors). The time series d_t, which is an output of that algorithm, measures the average residual variance across firms in each period. The standard deviation of $\sqrt{d_t}$, reported in Panel B, provides a summary statistic of the variability of average residual variance. Panel C, which reports the autocorrelation of $\sqrt{d_t}$, summarizes the peristence of average residual variance.

Panel A	Number of firms			
	Total	NYSE	Amex	Nasdaq
1979–1983	3340	1647	524	1169
1984–1988	3619	1510	445	1664
1989–1993	4364	1765	531	2068
1994–1998	4792	2026	472	2294
1979–1998	1281	889	124	268
1979–1998^m	19,144	4736	2101	12,307

Panel B	Standard deviation of $\sqrt{d_t}$			
	$K = 1$	$K = 3$	$K = 6$	$K = 12$
1979–1983	1.88	1.78	1.80	1.79
1984–1988	1.90	1.88	1.81	1.70
1989–1993	5.20	5.02	4.97	4.49
1994–1998	2.66	2.68	2.64	2.57
1979–1998	1.79	1.74	1.69	1.56
1979–1998^m	3.87	3.87	3.82	3.72

Panel C	Autocorrelation of $\sqrt{d_t}$			
	$K = 1$	$K = 3$	$K = 6$	$K = 12$
1979–1983	0.48	0.49	0.47	0.50
1984–1988	0.38	0.32	0.33	0.36
1989–1993	0.25	0.22	0.21	0.22
1994–1998	0.43	0.43	0.41	0.42
1979–1998	0.44	0.42	0.43	0.43
1979–1998^m	0.49	0.49	0.49	0.50

2.4. Measuring the equivalence of two sets of factors

Many of the results in this paper concern the equivalence of two sets of factors. In some cases, the two sets consist of a set of true factors and a set of

estimates, while in other cases both sets are estimates. I will be principally interested in whether the two sets of factors have the same span, or whether the two sets explain the same variation in stock returns, and I will measure this by regressing one set on the other and calculating the regression R-squares.

There are several ways one might do this. When comparing two sets of K factors, for example, one option would be to regress the first factor from the first set on the first factor of the second set, then regress the second factor from the first set on the second factor of the second set, and so on. Of course, if the two sets of factors were identical but arranged in different orders this procedure could return R-squares of zero, implying zero equivalence when in fact the spanning of the two sets is identical. A better measure of equivalence would therefore be calculated by regressing each factor in the first set on *all* the factors of the second set. If all the R-squares are close to one then the equivalence of the two sets of factors would be considered very high.

Because the comparison of two sets of factors involves running K regressions, I will reduce the quantity of numerical results by reporting only the average ($\overline{R^2}$) and the minimum ($\underline{R^2}$) of the K R-squares. The average is intended to summarize the overall equivalence of the two sets, while the minimum focuses more on the "marginal factor" that is hardest to detect in the data. Reporting the minimum is intended to highlight whether, as the number of factors is increased, the accuracy with which the additional factors are estimated is too low to make them of any benefit.

When measuring the equivalence of a set of extracted factors and a set of "true" factors, I will always place the true factors on the left hand of the regression, thereby focusing on the ability of the extracted factors to span the true ones. From an econometric perspective, this choice is counterintuitive, since it is the extracted factors that are measured with error. From the standpoint of asset pricing theory, however, this choice reflects the result that if the factor proxies span the true factors, then a linear expected return relation will hold using either the proxies or the true factors. The reverse, however, is not true.

3. A simulation study

Extraction of the true factors is at best only approximate, since two key results are valid only asymptotically. First, as in Connor and Korajczyk, the cross-product matrix only converges to its probability limit as the number of assets N grows large. Second, even given the true cross-product matrix, the decomposition procedure of Joreskog is iterative, with any feasible solution being an approximation of the iterative limit.

In this section I address, through the use of simulations, the degrees with which these approximations are likely to be valid in the data sets commonly

used by researchers in empirical asset pricing. I seek to assess the performance of the current method and that of Connor and Korajczyk under alternative data generating processes by altering assumptions about residual hetero-skedasticity and by changing the properties of the generated returns to reflect differences between the data samples described in Table 1.

Because a realistic model of how firms are added and deleted from the sample is not obvious, I restrict attention in this section to the cases in which firms have no missing observations. Since Table 1 indicates that the addition of firms with missing observations may contribute to heteroskedasticity, the experiments below may lessen the relevance of the HFA approach.

3.1. The simulation algorithm

The data generating processes (DGP) used in the simulations are designed to mimic the actual data as closely as possible. Rather than simulating factors under some arbitrary assumptions, bootstrap samples of factor estimates extracted from the actual data are used as the true factors in the simulations. Given estimates of the $T \times K$ matrix $\hat{F}$ of factor realizations and the $T \times T$ diagonal matrix $\hat{D}$ of average residual variances, I sample (with replacement) T rows of $\hat{F}$ and the corresponding diagonal elements of $\hat{D}$ to use as the true factors and average residual variances in the simulations. Let F_i denote the ith bootstrap draw of the factor matrix and D_i the corresponding draw of the D matrix.[5]

All alphas are assumed to be zero in the simulations. The factor betas assumed in the DGP will be bootstrap samples of the least squares estimates of the betas from the actual data, and are assumed constant over time. If $\hat{B}$ is the $N \times K$ matrix of OLS estimates of the factor betas from real data, I draw with replacement N rows of the $\hat{B}$ matrix to use as the true betas in the simulations. In addition, I draw the corresponding elements of the $N \times N$ diagonal matrix $\hat{\Omega}$, whose (n, n) element is the unconditional sample variance of the residual of stock n. Let B_i denote the ith bootstrap draw of the beta matrix and Ω_i the corresponding draw of Ω.

Two scenarios are considered in the simulations. In the first, I assume that residuals are homoskedastic, and I apply the CK method to extract factor estimates ($\hat{F}^{CK}$) from each of the samples described in Table 1. The beta estimates ($\hat{B}^{CK}$) and residual variances ($\hat{\Omega}^{CK}$) used in the simulations come from the OLS regression of excess stock returns on these factor estimates. Using the bootstrap draws (F_i, B_i, and Ω_i), from these estimates, the $N \times T$ matrix of simulated excess returns R_i will then be generated by the equation

$$R_i = B_i F_i + \Omega_i^{1/2} E_i, \tag{9}$$

---

[5] If the factors are constructed under the assumption of homoskedastic errors, no $\hat{D}$ matrix is produced, so I skip this step.

where $\Omega_i^{1/2}$ is the Cholesky factor of Ω_i and E_i is an $N \times T$ matrix of independent standard normals.[6]

In the second scenario, residuals are assumed to display heteroskedasticity, and so the factor estimates ($\hat{F}^{\text{HFA}}$) and average idiosyncratic variance matrix ($\hat{D}$) are extracted using the heteroskedastic factor analysis procedure with 100 Joreskog iterations. Corresponding beta estimates ($\hat{B}^{\text{HFA}}$) and residual variances ($\hat{\Omega}^{\text{HFA}}$) come from the OLS regression of returns on the factor estimates. Now using the bootstrap draws ($F_i, D_i, B_i,$ and Ω_i) from these estimates, excess returns are generated by

$$R_i = B_i F_i + \Omega_i^{1/2} E_i^*, \tag{10}$$

where the matrix E_i^* is an $N \times T$ matrix of independent, zero mean, but now heteroskedastic normal random variables. Specifically, if $d_{i,t}$ is the (t, t) element of D_i and $\bar{d}_i = (1/T)\Sigma_{t=1}^{T} d_{i,t}$, then the tth column of E_i^* has variance $d_{i,t}/\bar{d}_i$. On average, therefore, a stock's residual variance will be given by the appropriate element of Ω_i, but this variance will change over time as suggested by the draw D_i.

3.2. Convergence of the Joreskog iteration

The first experiment attempts to assess the convergence of the Joreskog iteration. Since the iterative method is only necessary when residuals are heteroskedastic, I generate five thousand artificial return data sets under the second scenario, in which the true factors and betas are constructed using data from 1994 to 1998 and the number of factors is assumed to be either one, three, six, or 12. Each simulated data sample therefore consists of 60 monthly observations of returns on 4792 firms.

I assume in all cases that the true number of factors is known, so from each set of artificial returns that is generated, the correct number of factors is extracted using the HFA method. These extracted factors are then used as regressors, along with an intercept, in a set of regressions with the true factors as the dependent variables. Assume that the true number of factors is K, and let $f_{i,t}^k$ denote the kth true factor at time t, or the (k, t) element of F_i. Let $\hat{f}_{i,t}^k(iter)$ denote the time-t estimate of the kth factor extracted with *iter* iterations.

To assess convergence of the estimates as the number of iterations is increased, I run K time series regressions for each bootstrap sample i:

$$f_{i,t}^j = \delta_0 + \sum_{K=1}^{K} \delta_k \hat{f}_{i,t}^k(iter) + \xi_t \tag{11}$$

---

[6] Recall that Ω is a diagonal matrix, so the (n, n) element of the Cholesky factor is just the estimated residual standard deviation of the nth firm in the sample.

for $j = 1, ..., K$. For each set of regressions, the average $(\overline{R^2}_i)$ and minimum $(\underline{R^2}_i)$ of the K R-squares are calculated.

  The means of these statistics calculated across all bootstrap draws are tabulated in Table 2 as a function of the number of Joreskog iterations. The standard errors of all mean estimates are below 0.001. Because of sampling error, both statistics should have averages below unity, but they are fairly close to unity in all cases. When a single factor is extracted, a single iteration of the algorithm is adequate, but for $K > 1$ both average R-squares $(\overline{R^2})$ and minimum R-squares $(\underline{R^2})$ increase noticeably as the number of iterations is increased from one to three. Increasing the number of iterations beyond five

Table 2
Convergence of the Joreskog iteration

 Five thousand return data sets are simulated under the assumption that residuals are heteroskedastic and the number of factors, K, is either one, three, six, or 12. Each simulated sample consists of 60 monthly observations of returns on 4792 firms, consistent with the size of the 1994–1998 subsample. Factors are then extracted from these returns using the heteroskedastic factor extraction method of Section 2.1, where the number of iterations (*iter*) of the algorithm is varied from one to 100. The extracted factors, $\hat{f}^k_{i,t}(iter)$, are then used as regressors in the following set of regressions in which the true factors, $\hat{f}^k_{i,t}$, are the dependent variables:

$$f^j_{i,t} = \delta_0 + \sum_{k=1}^{K} \delta_k \hat{f}^k_{i,t}(iter) + \xi_t$$

for $j = 1, ..., K$. For each of the simulations, the average $(\overline{R^2}_i)$ and minimum $(\underline{R^2}_i)$ of the K R-squares are calculated. Higher values of $\overline{R^2}_i$ and $\underline{R^2}_i$ indicate better spanning of the true factors by the extracted ones. The average values of these statistics are reported in Panels A and B, respectively.

	$K = 1$	$K = 3$	$K = 6$	$K = 12$
Panel A		Mean of $\overline{R^2}_i$		
1 Iteration	0.992	0.947	0.956	0.963
3 Iterations	0.993	0.949	0.964	0.974
5 Iterations	0.993	0.949	0.964	0.974
10 Iterations	0.993	0.949	0.965	0.974
25 Iterations	0.993	0.949	0.965	0.975
100 Iterations	0.993	0.949	0.965	0.975
Panel B		Mean of $\underline{R^2}_i$		
1 Iteration	0.992	0.888	0.903	0.892
3 Iterations	0.993	0.895	0.925	0.932
5 Iterations	0.993	0.895	0.925	0.934
10 Iterations	0.993	0.895	0.925	0.935
25 Iterations	0.993	0.895	0.925	0.936
100 Iterations	0.993	0.895	0.925	0.937

yields extremely small benefits, suggesting that allowing too few iterations is unlikely to be a source of substantial approximation error.

Because there is virtually no improvement in performance beyond five iterations, this value will be used in all subsequent calculations.

3.3. Do extracted factors span the true factors?

The second simulation exercise expands on the issue of how well the estimated factors would explain the true factors, were they observed. Five thousand sets of artificial returns will be simulated both assuming homo-skedasticity (scenario 1) and assuming heteroskedasticity (scenario 2). In addition, because the degree of heteroskedasticity has changed over time (see Table 1), the DGP is further varied by using the four different five-year subperiods as well as the single 20-year period to construct the true factors, factor loadings, and residual variances. As before, I consider the cases of one, two, six, and 12 factors.

For each simulated data set, both the CK and HFA methods will be used to extract factors. The simulations will therefore reveal how well the CK procedure performs when residuals are heteroskedastic, as well as how the HFA procedure performs when they are not.

As in the previous exercise, I report statistics on the average and minimum R-squares of the K regressions of the true factors on the extracted factors,

$$f_{i,t}^{j} = \delta_0 + \sum_{K=1}^{K} \delta_k \hat{f}_{i,t}^{k} + \xi_t \tag{12}$$

for $j = 1, \ldots, K$, where $\hat{f}_{i,t}^{k}$ may come out of either the CK or HFA procedure. High average and minimum R-squares indicate that the true factors of the DGP are spanned by the factor estimates.

Table 3 presents the results for five different time periods, 1979–1983, 1984–1988, 1989–1993, 1994–1998, and 1979–1998. Panel A explores the homo-skedastic first scenario, in which returns are generated according to Eq. (9), while Panel B reports on the heteroskedastic DGP given by Eq. (10). If the mean $\overline{R^2}$ or $\underline{R^2}$ for one method is higher than that of the other method by 0.01 or more, the mean of the superior method appears in bold face type. Standard errors in this table are extremely small, and all bold faced means are at least 25 standard errors higher than the mean of the competing method.

Panel A shows in almost every case that the performance of the two methods is very similar when residuals are homoskedastic. In three out of the four five-year samples and for the 20-year sample, this performance is very good, while the 1989–1993 period presents somewhat of a problem for the HFA method. While the HFA procedure is considerably worse in extracting the first factor in this period, for higher factors the performance of the two methods is identical.

Table 3
Simulation performance of extracted factors in spanning true factors

Five thousand return data sets are simulated with either one, three, six, or 12 factors under various assumptions about residual heteroskedasticity. Each simulated sample consists of a set of monthly returns that has the same number of months and number of firms as the actual sample of data observed over the time interval listed. Factors are then extracted from these returns using the Connor and Korajczyk (1986) procedure (CK) and the heteroskedastic factor extraction method of Section 2.1 (HFA). Each set of extracted factors, $\hat{f}_{i,t}^{k}$, is then used as the set of regressors in the following set of regressions in which the true factors, $f_{i,t}^{k}$, are the dependent variables:

$$f_{i,t}^{j} = \delta_0 + \sum_{k=1}^{K} \delta_k \hat{f}_{i,t}^{k} + \xi_t$$

for $j = 1, \ldots, K$. For each of the simulations, the average $(\overline{R^2}_i)$ and minimum $(\underline{R^2}_i)$ of the K R-squares are calculated. Higher values of $\overline{R^2}_i$ and $\underline{R^2}_i$ indicate better spanning of the true factors by the extracted ones. The average values of these statistics are reported in the table. Panel A uses returns simulated under Scenario 1, in which residuals are homoskedastic, while Panels B and C consider Scenario 2, where residuals are heteroskedastic. Panel C differs from Panel B in its use of the block bootstrap to capture serial dependence in average residual variance. Bold face type indicates a value that is higher than the corresponding value of the competing method by 0.01 or more.

	Extracted using CK				Extracted using HFA			
	$K=1$	$K=3$	$K=6$	$K=12$	$K=1$	$K=3$	$K=6$	$K=12$
Panel A	Scenario 1:							
	Returns simulated with homoskedastic residual							
				Mean of $\overline{R^2}_i$				
1979–1983	0.998	0.962	0.972	0.979	0.998	0.961	0.972	0.978
1984–1988	0.997	0.939	0.961	0.975	0.997	0.938	0.961	0.974
1989–1993	**0.766**	0.959	0.964	0.975	0.641	0.958	0.963	0.975
1994–1998	0.992	0.957	0.970	0.980	0.992	0.957	0.970	0.979
1979–1998	0.995	0.835	0.846	0.869	0.995	0.834	0.846	0.868
				Mean of $\underline{R^2}_i$				
1979–1983	0.998	0.933	0.946	0.953	0.998	0.932	0.945	0.951
1984–1988	0.997	0.893	0.923	0.942	0.997	0.892	0.921	0.941
1989–1993	**0.766**	0.909	0.933	0.944	0.641	0.908	0.932	0.942
1994–1998	0.992	0.925	0.941	0.954	0.992	0.924	0.940	0.951
1979–1998	0.995	0.671	0.741	0.780	0.995	0.670	0.740	0.779
Panel B	Scenario 2:							
	Returns simulated with heteroskedastic residuals							
				Mean of $\overline{R^2}_i$				
1979–1983	0.997	0.869	0.883	0.879	0.998	**0.965**	**0.973**	**0.969**
1984–1988	0.996	0.903	0.935	0.934	0.997	**0.951**	**0.968**	**0.976**
1989–1993	0.709	0.691	0.751	0.930	**0.815**	**0.838**	**0.959**	**0.972**
1994–1998	0.978	0.919	0.926	0.940	**0.993**	**0.949**	**0.964**	**0.974**
1979–1998	0.995	0.805	0.816	0.833	0.995	**0.831**	**0.849**	**0.871**

Table 3 (*continued*)

	Extracted using CK				Extracted using HFA			
	$K = 1$	$K = 3$	$K = 6$	$K = 12$	$K = 1$	$K = 3$	$K = 6$	$K = 12$
				Mean of $\underline{R}^2_i$				
1979–1983	0.997	0.736	0.682	0.518	0.998	**0.937**	**0.947**	**0.861**
1984–1988	0.996	0.816	0.837	0.769	0.997	**0.913**	0.935	**0.940**
1989–1993	0.709	0.469	0.480	0.725	**0.815**	0.719	**0.918**	**0.904**
1994–1998	0.978	0.846	0.816	0.789	**0.993**	**0.895**	**0.925**	**0.934**
1979–1998	0.995	0.641	0.669	0.646	0.995	**0.698**	**0.749**	**0.774**

Panel C	Scenario 2:							
	Returns simulated with heteroskedastic residuals, Block bootstrap							
				Mean of $\overline{R^2}_i$				
1979–1983	0.997	0.923	0.897	0.934	0.998	**0.961**	**0.964**	**0.973**
1984–1988	0.996	0.913	0.942	0.946	0.997	**0.951**	**0.968**	**0.973**
1989–1993	0.695	0.732	0.644	0.849	**0.801**	**0.943**	**0.968**	**0.975**
1994–1998	0.985	0.739	0.812	0.899	0.993	**0.954**	**0.970**	**0.957**
1979–1998	0.995	0.775	0.839	0.846	0.995	**0.810**	**0.863**	**0.878**
				Mean of $\underline{R}^2_i$				
1979–1983	0.997	0.866	0.660	0.792	0.998	**0.935**	**0.903**	**0.917**
1984–1988	0.996	0.836	0.868	0.827	0.997	**0.915**	**0.938**	**0.921**
1989–1993	0.695	0.517	0.181	0.475	**0.801**	**0.890**	**0.921**	**0.921**
1994–1998	0.985	0.451	0.500	0.427	0.993	**0.916**	**0.944**	**0.734**
1979–1998	0.995	0.598	0.622	0.672	0.995	**0.674**	**0.700**	**0.769**

It is somewhat puzzling in this case that the HFA method extracts the higher order factors more accurately than it does the first factor.

In the Panel B, in which residuals are heteroskedastic, the HFA procedure is clearly superior, with mean $\overline{R^2}$ and $\underline{R}^2$ that are universally higher, many by a wide margin, particularly for $K = 6$ and 12. In addition, many of the lowest R-squares occur in the two more recent five-year samples and in the 20-year sample, more evidence that residual heteroskedasticity may be more of a problem today than it was when the Connor and Korajczyk papers were written.

A weakness of the bootstrap experiments thus far is that they have ignored serial dependence in factors and residuals, and therefore may generate unrealistic sampling distributions of the two goodness of fit measures.[7] Volatility persistence, as reported in Table 1, is a primary example of such

---

[7] I am grateful to Bob Korajczyk for this observation.

dependence, and is clearly not captured in the previous experiments due to the independence of the 60 or 240 draws of d_t. Panel C of Table 3 therefore uses a block bootstrap algorithm to capture some of the serial dependence in the data. Rather than using serially independent draws of F_t and the corresponding d_t, I now draw five or 20 complete calendar years of F_t and d_t, again with replacement. The results using this block bootstrap procedure are qualitatively unchanged, as the HFA method consistently outperforms CK.

Performance over the 20-year sample is generally worse than the average performance across the four subperiods, except that in all cases there are no problems with either method extracting the first factor. Reflecting the result from Table 1 that the degree of heteroskedasticity for this sample is low, the differences between the CK and HFA methods are never particularly large. Because the number of firms in the 20-year sample is low, however, the extracted factors from both methods are generally mediocre, suggesting that it might be preferable to use shorter sample periods or to make use of firms with missing observations.

4. Explaining actual returns data

While the results of the previous section demonstrated that the HFA method could be useful in situations where the CK method performed poorly, the results relied on somewhat stylized data generating processes. Among other limitations, these DGPs assumed the independence of residuals across firms, a simplification that is no doubt at odds with the data.

Rather than present more simulation evidence, this section seeks to assess the performance of each method in explaining actual returns data. First, I perform a cross validation exercise in which factors extracted from half of the firms in the sample are used to explain factors extracted from the other half. Next, I show how well factors extracted using each method explain the returns of some commonly-used benchmark portfolios.

4.1. A cross validation exercise

If the factors extracted from one set of returns are not spurious, then they should have power to explain factors extracted from the returns of a different set of securities. I investigate this property in a very simple cross validation exercise. In each of the six data samples, firms are split randomly into two equally-sized groups.[8] For both the CK and HFA procedures, the returns of

---

[8] For the experiments using the 20-year sample with missing observations, firms vary in the number of returns they contribute to their groups. Although the groups are equally-sized in the number of firms, there will be some differences in the number of return observations.

each half are used to construct two separate sets of factor estimates. The process is repeated five thousand times, with the only difference between iterations being the assignment of firms into the two halves.

If the factors are extracted correctly and precisely, then the two halves should produce equivalent sets of factor estimates. Measurement of this equivalence is the focus of the exercise. Using either the CK or HFA procedures, let $\hat{f}^{j}_{i,t}$ denote the estimate of the jth factor extracted from the first half on repetition i. Let $\check{f}^{j}_{i,t}$ denote the corresponding factor estimate extracted from the second half. I then run a set regressions similar to previous ones:

$$\hat{f}^{j}_{i,t} = \delta_0 + \sum_{k=1}^{K} \delta_k \check{f}^{k}_{i,t} + \xi_t \tag{13}$$

for $j = 1, \ldots, K$. Equivalence of the two sets of factors is measured by high average ($\overline{R^2}_i$) and minimum ($\underline{R^2}_i$) R-squares across the K regressions.

The means of $\overline{R^2}_i$ across the five thousand repetitions are plotted in Fig. 1. The means of ($\underline{R^2}_i$) are plotted in Fig. 2. The solid lines represent the means for the CK method, while the dashed lines represent the HFA method. If a method produces accurate estimates of the underlying factors, then the means should be close to unity. Lower R-squares are evidence that a method results in poorly-estimated or even spurious extracted factors.
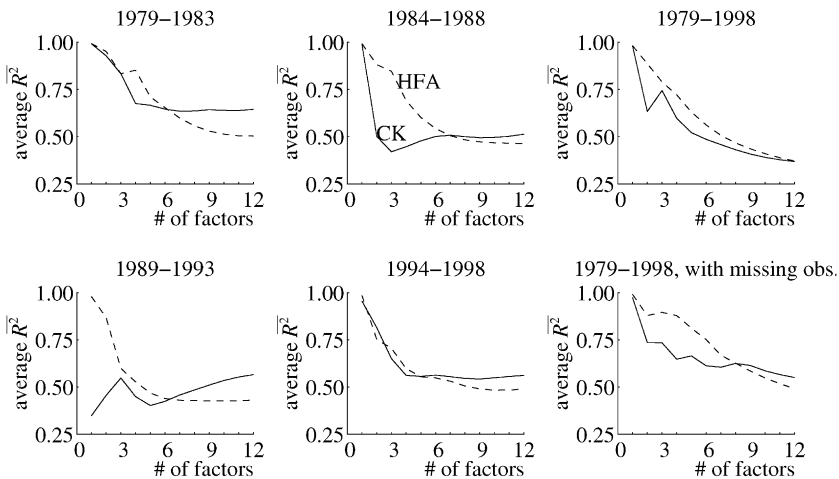

Fig. 1. Cross validation average $\overline{R^2}$. Random samples are formed by randomly dividing up firms into two groups of equal size. Each of the K factors extracted using returns from one group are regressed on all the factors extracted from the other group. For each sample, the average R-square, $\overline{R^2}$, of these K regressions is calculated. Higher values of $\overline{R^2}$ indicate greater consistency between the factors extracted from each half of the data. Solid and dashed lines represent average $\overline{R^2}$ across samples for the CK and HFA methods, respectively, as a function of the number of factors, K.
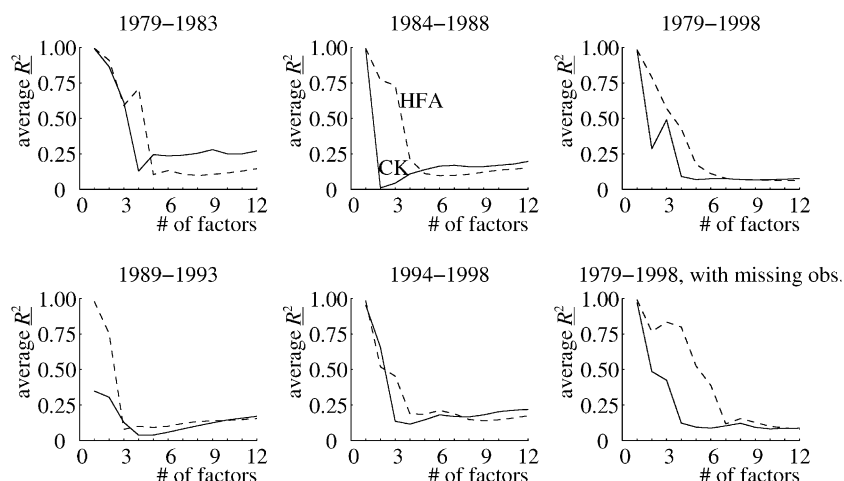
Fig. 2. Cross validation average $\underline{R}^2$. Random samples are formed by randomly dividing up firms into two groups of equal size. Each of the K factors extracted using returns from one group are regressed on all the factors extracted from the other group. For each sample, the minimum R-square, $\overline{R^2}$, of these K regressions is calculated. Higher values of $\underline{R}^2$ indicate greater consistency between the factors extracted from each half of the data. Solid and dashed lines represent average $\underline{R}^2$ across samples for the CK and HFA methods, respectively, as a function of the number of factors, K.

The top left panels of Figs. 1 and 2, for instance, show that the CK and HFA methods are roughly equivalent for $K = 1$–3 in the 1979–1984 period. For $K = 4$, however, the average $\overline{R^2}$ is around 0.65 for CK, while the HFA value is around 0.85. The corresponding average $\underline{R}^2$ is less than 0.2 for CK, compared with 0.7 for HFA, indicating that on average one of the first four factors is estimated with very low precision when using the CK method. When larger numbers of factors are extracted, however, the CK method seems to perform better, as the HFA factors produce average $\underline{R}^2$ that are fairly close to zero for $K > 4$.

In the 1984–1988 period, the CK method seems to fail again, this time for $K = 2$ and 3. The worst period for the CK method, however, is 1989–1993, when even the first factor appears to be estimated with extremely low precision. The first HFA factor, in contrast, seems to be estimated extremely accurately, as it is nearly unchanged no matter how the sample is split. In 1994–1998, the two methods perform more comparably, with HFA somewhat more consistent in extracting the third factor.

In the 20-year sample without missing observations, CK fares poorly in extracting the second factor. The R-squares generally increase when firms with missing observations are added, consistent with the larger sample size. The gain

is much larger for the HFA method, however, reflecting the fact that the larger sample has a greater level of residual heteroskedasticity.

Overall, it appears that the CK procedure typically has difficulty in the extraction of some relatively low-order factor, such as the second factor in 1984–1988 or the first factor in 1989–1993. One plausible explanation is that because the CK factor is spurious, additional higher-order factors prove to be beneficial. Low-order factors for the HFA method are comparatively well-estimated, so little common variation remains for higher-order factors to explain. Because higher order factors are not called for, extracting them amounts to an overfitting of the data, so they are inconsistent with data not used in their construction.

To summarize, low order factors from the Connor and Korajczyk often appear to be spurious, while it is the high order HFA factors that are imprecisely estimated. One implication is that more CK factors will appear to be necessary to explain covariation in returns, so that estimates of the number of factors in returns may be biased upward when using the CK extraction procedure.

4.2. Explaining common benchmarks

Although it is not necessary, many investigations of the APT choose to test prespecified factors, and a number of recent papers have identified benchmark portfolios that appear to fare well as APT factors. In this section I look briefly at the explanatory power that the CK and HFA factor estimates have over these benchmark returns.

The portfolios I consider are the value-weighted and equal-weighted market indexes, Fama and French's (1993) size and book-to-market factors, Carhart's (1997) momentum factor, and a 30-year fixed term bond index from CRSP. I regress each excess return series on between one and twelve CK or HFA factors and again look at regression adjusted R-squares, which are plotted in Figs. 3–6. As before, accurately estimated factors should result in higher adjusted R-squares, although we no longer expect the to equal unity.[9]

Over the 1979–1983 sample, whose results are depicted in Fig. 3, there are few differences between CK and HFA factors. Somewhat larger are the differences in the 1984–1988 sample (Fig. 4), with HFA outperforming CK most noticeably for the momentum factor. In both these samples, just one factor is sufficient to explain the returns of the equal-weighted market portfolio. Since Brown (1989) notes that the first principal component should approximate the equal-weighted portfolio even when there are multiple factors, this result is not unusual.

---

[9] Although raw R-squares are used throughout the remainder of the paper, adjusted R-squares are used here because the focus is directly on the change in fit as more factors are considered.
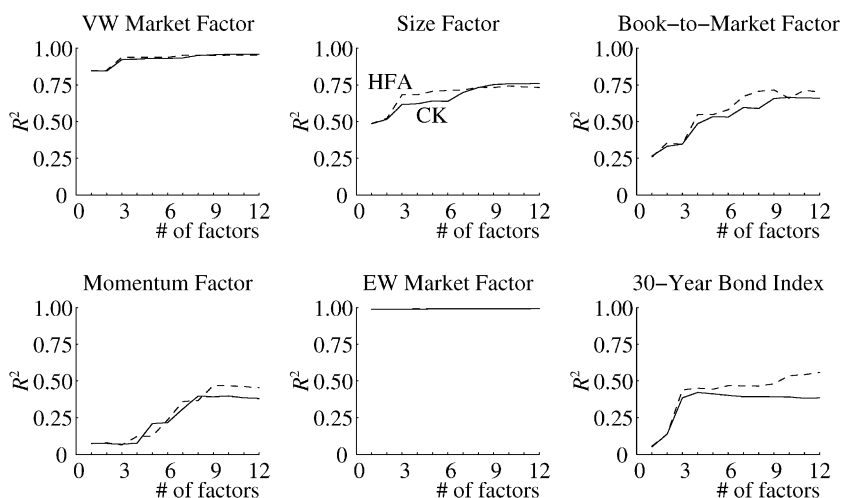
Fig. 3. Portfolio return adjusted *R*-squares for 1979–1983. Excess returns of six bench mark portfolios are regressed on one to 12 extracted factors using monthly data from 1979–1983. Solid and dashed lines denote the adjusted *R*-squares for the CK and HFA methods, respectively, as a function of the number of factors, *K*.

Fig. 4. Portfolio return adjusted *R*-squares for 1984–1988. Excess returns of six benchmark portfolios are regressed on one to 12 extracted factors using monthly data from 1984–1988. Solid and dashed lines denote the adjusted *R*-squares for the CK and HFA methods, respectively, as a function of the number of factors, *K*.

What *is* surprising is that in the 1989–1993 Sample, shown in Fig. 5, this relation breaks down for the CK method, with the first factor explaining the equal-weighted portfolio with an *R*-squared of less than 0.6. The first HFA

Fig. 5. Portfolio return adjusted R-squares for 1989–1993. Excess returns of six benchmark portfolios are regressed on one to 12 extracted factors using monthly data from 1989–1993. Solid and dashed lines denote the adjusted R-squares for the CK and HFA methods, respectively, as a function of the number of factors, K.
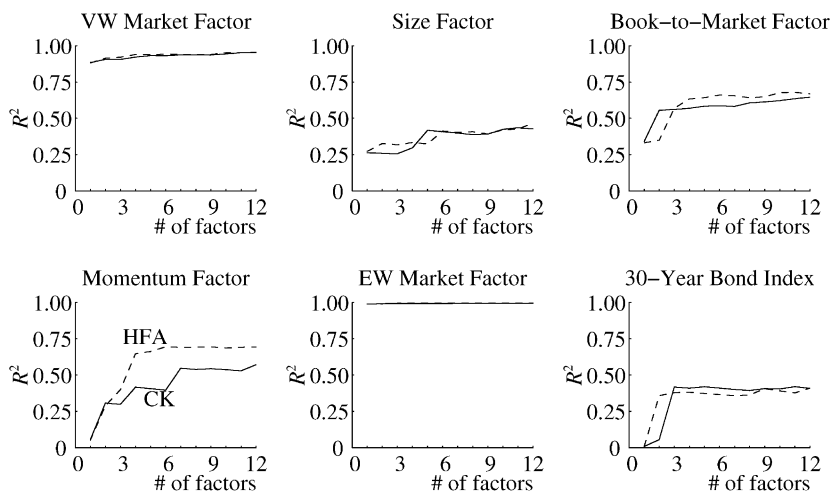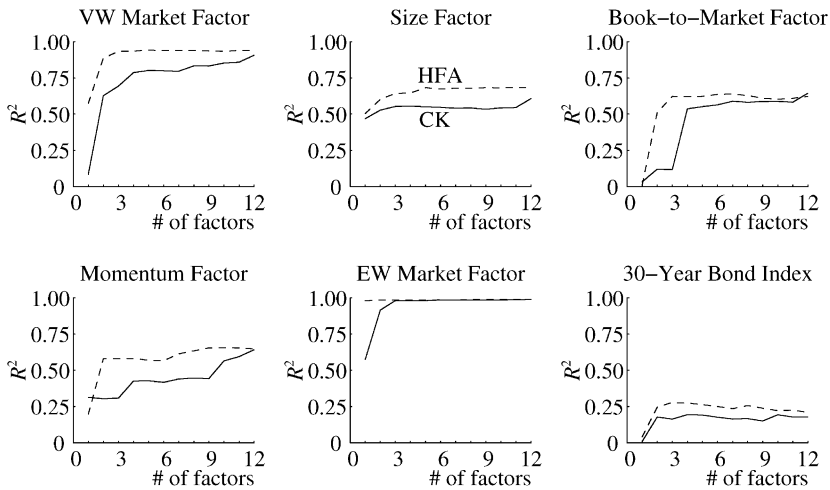

Fig. 6. Portfolio return adjusted R-squares for 1994–1998. Excess returns of six benchmark portfolios are regressed on one to 12 extracted factors using monthly data from 1994–1998. Solid and dashed lines denote the adjusted R-squares for the CK and HFA methods, respectively, as a function of the number of factors, K.
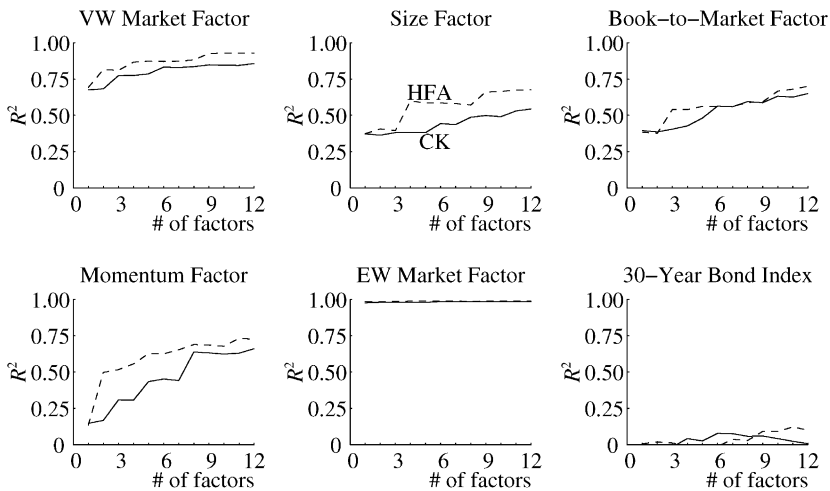
factor, meanwhile, still explains the equal-weighted market almost perfectly, suggesting that Brown's argument is correct as long as residual heteroskedasticity is taken into account.

The other portfolios over the 1989–1993 sample are also better explained by the first HFA factor. The value-weighted portfolio's R-squared is not even 0.1 when it is regressed on the CK factor, while the corresponding value is about 0.6 for the HFA factor. Even as higher numbers of factors are considered, the HFA factors outperform CK. Only once 12 factors are considered do the R-squares of the CK and HFA factors roughly converge. Given the cross validation results, however, it seems unlikely that more than six of these factors are estimated with any accuracy.

Over the 1994–1998 sample (Fig. 6), the HFA factors continue to generally outperform CK, although the differences are not as large as 1989–1993.

The 20-year sample with no missing observation is considered in Fig. 8, while results with missing observations are shown in Fig. 9. These can be compared informally to Fig. 7, which plots the average adjusted R-squares over the four subperiods.

Fig. 8 reveals few meaningful differences between the two extraction methods. When firms with missing observations are added (Fig. 9), however, the HFA method regains its advantage over CK. In comparing the results in Figs. 8 and 9, there appears to be no general pattern. While momentum returns are better explained by the larger data set, the opposite result is obtained for the 30-year bond and the value-weighted market index. This last finding is perhaps not too surprising, since the data set without missing values was

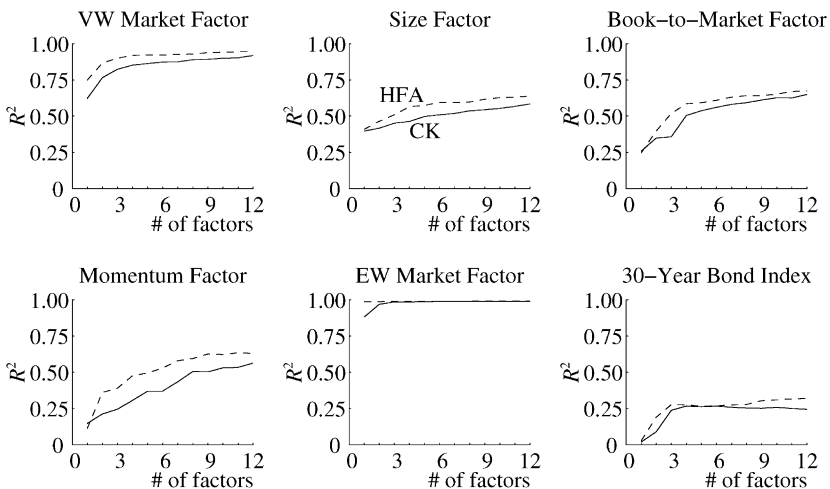

Fig. 7. Portfolio return adjusted R-squares averaged across subsamples. Excess returns of six benchmark portfolios are regressed on one to 12 extracted factors over four different five-year subperiods. Solid and dashed lines denote the average of the four adjusted R-squares for the CK and HFA methods, respectively, as a function of the number of factors, K.
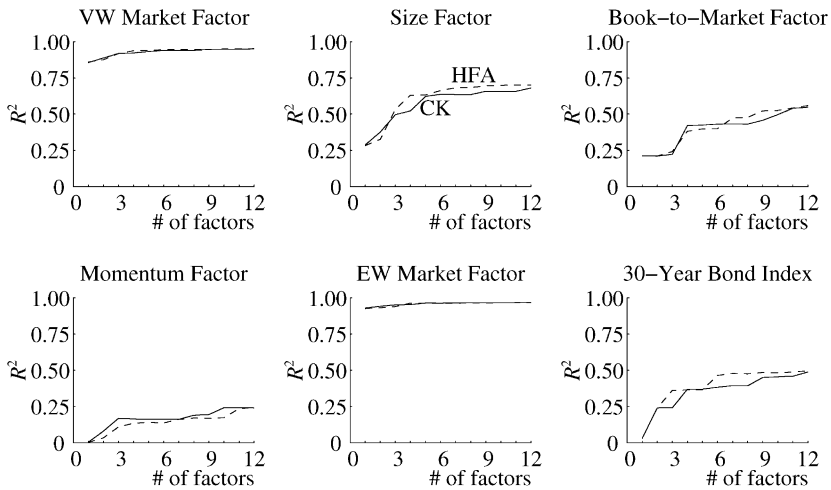
Fig. 8. Portfolio return adjusted R-squares for 1979–1998. Excess returns of six benchmark portfolios are regressed on one to 12 extracted factors using monthly data from 1979–1998. Solid and dashed lines denote the adjusted R-squares for the CK and HFA methods, respectively, as a function of the number of factors, K.

Fig. 9. Portfolio return adjusted R-squares for 1979–1998, with missing observations. Excess returns of six bench mark portfolios are regressed on one to 12 extracted factors using monthly data from 1979–1998, including data on firms with missing observations. Solid and dashed lines denote the adjusted R-squares for the CK and HFA methods, respectively, as a function of the number of factors, K.

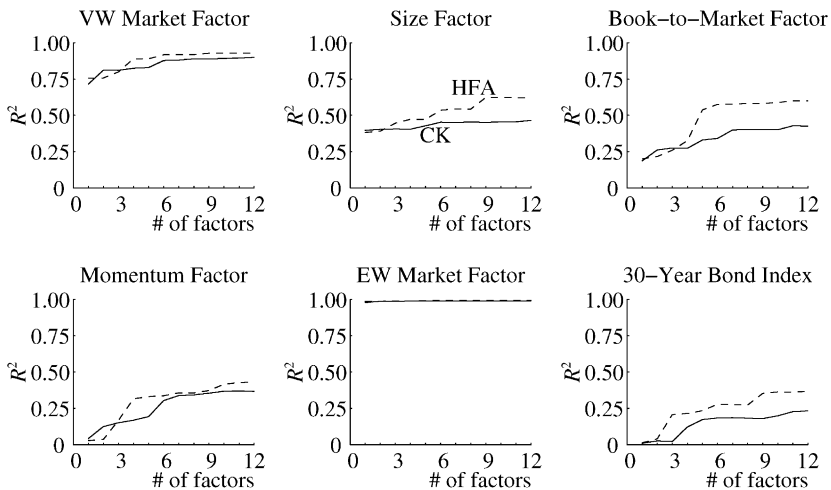mainly comprised of the large firms that make up that index. Consistent with this result, the equally-weighted index is better explained by the sample that includes firms with missing observations.

While the performance in Figs. 8 and 9 is in some cases superior to the average performance shown in Fig. 7, the largest discrepancies suggest that the best overall performance comes from using the HFA method over five-year intervals.

5. Pricing the Fama-French factors

A principal use of the extracted factors is for testing the Arbitrage Pricing Theory of Ross (1976). In this section I briefly consider the implications of residual heteroskedasticity for tests of the APT and propose several strategies for dealing with it. The question I ask is simple: do factors extracted from the cross-section of stock returns price the three factors of Fama and French (1993)?

The portfolios identified by Fama and French as "factors" have received a great deal of attention, and it is not without debate that they represent factors in the sense intended by Ross. Nevertheless, they represent feasible zero-investment portfolios that could be maintained over time with a limited amount of portfolio rebalancing.[10] The portfolios, consisting of the value-weighted market index, the Small Minus Big size-related portfolio, and the High Minus Low book-to-market portfolio, are particularly interesting because the apparent dependence of expected returns on size and book-to-market remains one of the least understood features of equity markets.

I first consider a test of the APT based on the test statistic of Gibbons et al. (1989). Let R denote the $T \times N$ matrix of excess returns on N assets, and let X denote the $T \times (K + 1)$ matrix whose first column is vector of ones and whose remaining columns consist of the time series of K extracted factors. Finally, let $\hat{\alpha}$ denote the OLS estimate of factor model intercepts and $\hat{\Sigma}$ the standard unbiased estimate of the residual covariance matrix. Under the assumption of a constant residual covariance matrix,

$$\frac{T-N-K}{(T-1)N_C} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim F_{N,T-N-K},\tag{14}$$

where c is the (1,1) element of $(X'X)^{-1}$.

Since residual heteroskedasticity invalidates this statistic, an alternative test is desirable. I use the GMM test of MacKinlay and Richardson (1991), which

---

[10] The "momentum factor" of Carhart (1997) also represents the return on a portfolio of stocks, but the composition of this portfolio changes dramatically over time. Since the feasible return on this portfolio may be heavily affected by transactions costs, I do not considered it here.

has the added advantage of being robust to non-normality of returns. Letting x_t denote the tth row of X and $\hat{\varepsilon}_t$ the $1 \times N$ vector of time t OLS residuals, define

$$J = \frac{1}{T} \sum_{t=1}^{T} (I \otimes x_t' x_t), \tag{15}$$

$$L = \frac{1}{T} \sum_{t=1}^{T} (\hat{\varepsilon}_t' \hat{\varepsilon}_t \otimes x_t' x_t), \tag{16}$$

$$Q = I \otimes [0, 1]. \tag{17}$$

Then

$$T\hat{\alpha}' [Q(J'L^{-1}J)^{-1}Q']^{-1}\hat{\alpha} \sim X^2(N) \tag{18}$$

holds asymptotically.

Table 4 reports these test statistics and their p-values over the six samples for a variety of values of K. Factors are extracted both using the CK and HFA methods. In the cases where the use of different factor extraction methods leads to different accept/reject outcomes at the 5% significance level, all statistics appear in bold face type.

For three of the five-year samples, there is at least one value of K for which the Fama-French portfolios result in a rejection of the APT using one set of extracted factors but not the other. In general, there appear to be only a few patterns in Table 4. The HFA factors may lead to larger or smaller values of the test statistic, depending on both K and the sample period. In the first two subperiods, the two methods lead to similar test statistics, reflecting the unimportance of the extraction method in these samples.

Consistent with previous sections, the most dramatic differences are found in the 1989–1993 subperiod, in which the HFA factors lead to much stronger rejections of exact factor pricing. For $K = 1$, the p-value for the GRS test with CK factors is 0.303, but it declines to just 0.013 when HFA factors are used instead. This result is somewhat expected, since the greater explanatory ability of the HFA factors should increase the power of the test. Table 5, which reports results for univariate tests, confirms this intuition, as both the OLS and White (1980) standard errors of $\hat{\alpha}$ are consistently smaller for the HFA factors, particularly for the 1989–1993 subsample.

The ordering is reversed for the 1994–1998 sample, so that using CK factors instead of HFA factors now lowers the p-values of the GRS test, sometimes substantially. This is slightly surprising since the CK standard errors are still somewhat larger than those computed using the HFA factors. The corresponding panel of Table 5 implies that the strength of these rejections is generally the result of larger alphas for the CK method, suggesting that the

Table 4
Multivariate tests of APT pricing of the Fama-French factors

Factors extracted using the Connor and Korajczyk (1986) method (CK) and the heteroskedastic factor extraction method of Section 2.1 (HFA) are used in multivariate tests of the APT. Extracted factors are used in OLS regressions to estimate the degree of mispricing ($\hat{\alpha}$) in the Value Weighted index and the Small Minus Big and High Minus Low portfolios of Fama and French (1993). Tests statistics are computed using the tests of Gibbons et al. (1989) and MacKinlay and Richardson (1991) as reviewed in Section 5. P-values are reported in parentheses. Bold face type indicates cases in which the two factor extraction methods lead to different accept/reject outcomes at the 5% significance level.

	Gibbons et al. (1989) $F(3, T-K)$				MacKinlay and Richardson (1991) $\chi^2(3)$			
	$K = 1$	$K = 3$	$K = 6$	$K = 12$	$K = 1$	$K = 3$	$K = 6$	$K = 12$
				1979–1983 (T=60)				
CK	2.595	**3.713**	4.277	8.090	**8.860**	13.673	15.738	24.969
	(0.061)	**(0.017)**	(0.009)	(0.000)	**(0.031)**	(0.003)	(0.001)	(0.000)
HFA	2.347	**2.698**	4.346	10.397	**7.187**	8.941	15.384	36.118
	(0.082)	**(0.055)**	(0.008)	(0.000)	**(0.066)**	(0.030)	(0.002)	(0.000)
				1984–1988 (T=60)				
CK	2.624	6.208	6.077	6.687	8.157	18.978	20.648	27.727
	(0.059)	(0.001)	(0.001)	(0.001)	(0.043)	(0.000)	(0.000)	(0.000)
HFA	2.721	7.042	6.494	6.141	8.362	21.837	24.622	25.427
	(0.053)	(0.000)	(0.001)	(0.001)	(0.039)	(0.000)	(0.000)	(0.000)
				1989–1993 (T=60)				
CK	**1.244**	3.939	4.843	4.782	**3.628**	12.193	16.038	16.479
	(0.303)	(0.013)	(0.005)	(0.006)	**(0.305)**	(0.007)	(0.001)	(0.001)
HFA	**3.954**	9.672	8.804	11.941	**11.770**	30.942	31.121	43.987
	(0.013)	(0.000)	(0.000)	(0.000)	**(0.008)**	(0.000)	(0.000)	(0.000)
				1994–1998 (T=60)				
CK	3.504	**3.247**	6.237	6.537	11.422	10.909	20.410	22.697
	(0.021)	**(0.029)**	(0.001)	(0.001)	(0.010)	(0.012)	(0.000)	(0.000)
HFA	3.194	**2.370**	3.895	6.056	10.082	8.133	14.821	23.104
	(0.030)	**(0.081)**	(0.014)	(0.001)	(0.018)	(0.043)	(0.002)	(0.000)
				1979–1998 (T=240)				
CK	7.102	7.418	7.048	10.319	22.263	23.187	22.965	36.716
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HFA	6.966	6.582	6.464	10.314	21.621	20.068	21.253	34.686
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
				1979–1998, with missing observations (T=240)				
CK	5.370	6.501	10.154	14.658	16.857	20.953	34.602	46.779
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
HFA	5.302	5.052	14.150	17.177	16.321	16.876	40.465	49.805
	(0.001)	(0.002)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)

Table 5
Ordinary least squares alphas for the Fama-French factors

Factors extracted using the Connor and Korajczyk (1986) method (CK) and the heteroskedastic factor extraction method of Section 2.1 (HFA) are used in univariate tests of the APT. Extracted factors are used in OLS regressions to estimate the degree of mispricing ($\hat{\alpha}$) in the Value Weighted (VW) index and the Small Minus Big (SMB) and High Minus Low (HML) portfolios of Fama and French (1993). OLS standard errors are reported in parentheses, while White (1980) standard errors appear in square brackets.

	$K = 1$		$K = 3$		$K = 6$		$K = 12$	
	CK	HFA	CK	HFA	CK	HFA	CK	HFA
				1979–1983				
VW	−0.49	−0.45	−0.44	−0.39	−0.45	−0.39	−0.44	−0.45
	(0.25)	(0.24)	(0.17)	(0.16)	(0.16)	(0.16)	(0.12)	(0.13)
	[0.21]	[0.21]	[0.15]	[0.15]	[0.14]	[0.15]	[0.13]	[0.13]
SMB	0.36	0.38	0.30	0.31	0.31	0.24	0.27	0.23
	(0.24)	(0.24)	(0.21)	(0.19)	(0.19)	(0.18)	(0.15)	(0.16)
	[0.22]	[0.22]	[0.19]	[0.18]	[0.18]	[0.18]	[0.16]	[0.16]
HML	0.84	0.82	0.75	0.73	0.78	0.93	1.00	0.97
	(0.34)	(0.34)	(0.32)	(0.32)	(0.26)	(0.25)	(0.22)	(0.21)
	[0.37]	[0.38]	[0.35]	[0.35]	[0.30]	[0.26]	[0.26]	[0.22]
				1984–1988				
VW	0.33	0.38	0.17	0.08	0.01	−0.03	−0.05	0.01
	(0.23)	(0.23)	(0.21)	(0.19)	(0.18)	(0.17)	(0.14)	(0.14)
	[0.22]	[0.22]	[0.22]	[0.21]	[0.18]	[0.17]	[0.16]	[0.14]
SMB	−0.39	−0.38	−0.33	−0.23	−0.22	−0.30	−0.21	−0.35
	(0.24)	(0.24)	(0.24)	(0.23)	(0.21)	(0.23)	(0.21)	(0.21)
	[0.24]	[0.23]	[0.26]	[0.26]	[0.24]	[0.23]	[0.23]	[0.19]
HML	0.66	0.64	0.85	0.88	0.77	0.61	0.79	0.66
	(0.27)	(0.27)	(0.22)	(0.22)	(0.21)	(0.20)	(0.19)	(0.19)
	[0.27]	[0.27]	[0.22]	[0.22]	[0.22]	[0.19]	[0.19]	[0.18]
				1989–1993				
VW	0.54	0.02	−0.04	−0.12	−0.11	−0.07	−0.01	−0.06
	(0.46)	(0.32)	(0.27)	(0.12)	(0.21)	(0.12)	(0.14)	(0.11)
	[0.46]	[0.26]	[0.25]	[0.13]	[0.21]	[0.13]	[0.15]	[0.13]
SMB	−0.39	−0.52	−0.53	−0.48	−0.51	−0.54	−0.54	−0.60
	(0.24)	(0.24)	(0.23)	(0.20)	(0.22)	(0.19)	(0.20)	(0.18)
	[0.24]	[0.23]	[0.22]	[0.20]	[0.23]	[0.21]	[0.21]	[0.21]
HML	0.05	0.20	0.22	0.38	0.25	0.40	0.28	0.42
	(0.31)	(0.32)	(0.30)	(0.19)	(0.21)	(0.19)	(0.18)	(0.18)
	[0.30]	[0.28]	[0.28]	[0.20]	[0.22]	[0.20]	[0.19]	[0.19]

Table 5 (*continued*)

	K = 1		K = 3		K = 6		K = 12	
	CK	HFA	CK	HFA	CK	HFA	CK	HFA
1994–1998								
VW	0.62	0.69	0.53	0.39	0.34	0.30	0.24	0.17
	(0.30)	(0.29)	(0.25)	(0.23)	(0.22)	(0.19)	(0.19)	(0.14)
	[0.31]	[0.29]	[0.26]	[0.24]	[0.22]	[0.20]	[0.20]	[0.15]
SMB	−0.89	−0.85	−0.85	−0.72	−0.87	−0.74	−0.80	−0.70
	(0.29)	(0.28)	(0.28)	(0.28)	(0.27)	(0.23)	(0.23)	(0.20)
	[0.27]	[0.27]	[0.27]	[0.28]	[0.27]	[0.24]	[0.25]	[0.22]
HML	0.33	0.28	0.36	0.21	0.55	0.33	0.56	0.61
	(0.26)	(0.26)	(0.25)	(0.22)	(0.22)	(0.22)	(0.19)	(0.18)
	[0.25]	[0.25]	[0.25]	[0.21]	[0.22]	[0.23]	[0.19]	[0.19]
1979–1998								
VW	−0.03	−0.03	−0.12	−0.18	−0.12	−0.15	−0.15	−0.15
	(0.11)	(0.11)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
	[0.10]	[0.10]	[0.08]	[0.09]	[0.07]	[0.07]	[0.07]	[0.07]
SMB	−0.31	−0.31	−0.23	−0.14	−0.23	−0.19	−0.25	−0.25
	(0.14)	(0.14)	(0.12)	(0.12)	(0.10)	(0.10)	(0.10)	(0.09)
	[0.13]	[0.13]	[0.12]	[0.12]	[0.10]	[0.10]	[0.10]	[0.09]
HML	0.54	0.54	0.55	0.49	0.42	0.43	0.53	0.52
	(0.15)	(0.15)	(0.15)	(0.15)	(0.13)	(0.13)	(0.12)	(0.12)
	[0.15]	[0.15]	[0.16]	[0.16]	[0.13]	[0.13]	[0.12]	[0.12]
1979–1998, with missing observations								
VW	0.17	0.16	0.12	0.02	−0.07	−0.17	−0.13	−0.12
	(0.15)	(0.14)	(0.13)	(0.13)	(0.10)	(0.08)	(0.09)	(0.08)
	[0.14]	[0.13]	[0.12]	[0.12]	[0.10]	[0.09]	[0.09]	[0.08]
SMB	−0.31	−0.30	−0.31	−0.25	−0.24	−0.16	−0.21	−0.26
	(0.13)	(0.13)	(0.13)	(0.12)	(0.12)	(0.11)	(0.12)	(0.10)
	[0.12]	[0.12]	[0.12]	[0.12]	[0.12]	[0.11]	[0.12]	[0.10]
HML	0.49	0.49	0.54	0.45	0.57	0.51	0.56	0.54
	(0.15)	(0.15)	(0.15)	(0.15)	(0.14)	(0.11)	(0.13)	(0.11)
	[0.15]	[0.15]	[0.15]	[0.15]	[0.15]	[0.12]	[0.14]	[0.11]

differences in the inferences drawn using each method are not solely the result of differences in power.

For the 20-year samples, all tests result in highly significant rejections, generally because of the extracted factors' inability to explain the large positive

returns of the HML portfolio. By looking at the individual alphas, we can see a few differences between the HFA and CK methods. The value-weighted index, for example, has a significant alpha under HFA but not CK when $K = 6$.

In some cases, it is difficult to attribute stronger rejections to either larger alphas or smaller standard errors. In the 1979–1983 sample with $K = 3$, for example, the CK factors lead to a much stronger rejection, even though their alphas and standard errors are comparable to those generated by the HFA factors. In these cases, the different outcomes of the CK and HFA methods must be attributed to different covariance estimates implied by the two methods.

The results for MacKinlay and Richardson's (1991) test are very similar, but generally result in lower p-values than GRS. One interpretation is motivated by Shanken's (1990) derivation of the MacKinlay–Richardson test as Hotelling's T^2 test with a White heteroskedasticity-consistent covariance matrix. In comparing the OLS standard errors with the White standard errors in Table 5, we see that the White standard errors are often smaller, and so tend to produce stronger rejections than the OLS standard errors that underlie the GRS test.

6. Conclusion

Heteroskedastic factor analysis (HFA) is a simple and natural extension of the method of Connor and Korajczyk (1986). Overall, there occasionally seems to be a large efficiency gain from allowing residual time series heteroskedasticity through the use of HFA instead of the CK method.

In the data considered, the 1989–1993 sample represents one period in which the CK procedure appears to seriously misestimate the true factors, with the HFA factors offering much greater explanatory power for common benchmark returns. While the two methods performed more similarly in earlier periods, the HFA method still generally outperformed CK in explaining returns. The treatment of residual heteroskedasticity was also shown to be important in tests of the APT.

Over a 20-year period, choosing firms with no missing observations led to a fairly small sample that did not generate extremely accurate results but that also did not generate the conspicuous mistakes of the 1989–1993 sample. By adding firms with missing observations, the sample size increased dramatically, but the introduction and subtraction of firms made the sample residuals much more heteroskedastic, and the HFA procedure again offered a significant performance advantage.

The proposed method may be particularly relevant in future research. Campbell et al. (2001) find that although overall market volatility did not increase over the period from 1962 to 1997, the idiosyncratic volatilities of individual firms increased substantially. As these residuals grow in importance,

it seems likely that accounting for their heteroskedasticity will become even more essential.

References

Brennan, M., Chordia, T., Subrahmanyam, A., 1998. Alternative factor specifications, security characteristics, and the cross-section of expected stock returns. Journal of Financial Economics 49, 345–373.

Brown, S., 1989. The number of factors in security returns. Journal of Finance 44, 1247–1262.

Brown, S., Goetzmann, W., Grinblatt, M., 1998. Positive portfolio factors. Unpublished working paper. NBER #6418.

Campbell, J., Lettau, M., Malkiel, B., Xu, Y., 2001. Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk. Journal of Finance 56, 1–43.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

Chamberlain, G., 1983. Funds, factors, and diversification in arbitrage pricing models. Econometrica 51, 1305–1323.

Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure, and mean-variance analysis of large asset markets. Econometrica 51, 1281–1304.

Chan, L., Karceski, J., Lakonishok, J., 1998. The risk and return from factors. Journal of Financial and Quantitative Analysis 33, 159–188.

Chan, L., Karceski, J., Lakonishok, J., 1999. On portfolio optimization: forecasting covariances and choosing the risk model. Review of Financial Studies 12, 937–974.

Chen, N., Roll, R., Ross, S., 1986. Economic forces and the stock market. Journal of Business 59, 383–403.

Connor, G., 1984. A unified beta pricing theory. Journal of Economic Theory 34, 12–31.

Connor, G., Korajczyk, R., 1986. Performance measurement with the arbitrage pricing theory. Journal of Financial Economics 15, 373–394.

Connor, G., Korajczyk, R., 1987. Estimating pervasive economic factors with missing observations. Unpublished working paper, Northwestern University, Evanston, IL.

Connor, G., Korajczyk, R., 1988. Risk and return in and equilibrium APT: application of a new test methodology. Journal of Financial Economics 21, 255–289.

Elton, E., 1999. Expected return, realized return, and asset pricing tests. Journal of Finance 54, 1199–1220.

Elton, E., Gruber, M., Busse, J., 1998. Do investors care about sentiment? Journal of Business 71, 477–500.

Fama, E., French, K., 1993. Common risk factors in the returns of bonds and stocks. Journal of Financial Economics 33, 3–53.

Ferson, W., Korajczyk, R., 1995. Do arbitrage pricing models explain the predictability of stock returns? Journal of Business 68, 309–349.

French, K., Schwert, G., Stambaugh, R., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Gibbons, M., Ross, S., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Jagannathan, R., Ma, T., 2001. Risk reduction in large portfolios: a role for portfolio weight constraints. Unpublished working paper, Northwestern University, Evanston, IL.

Joreskog, K., 1967. Some contributions to maximum likelihood factor analysis. Psychometrica 34, 183–202.

Knez, P., Litterman, R., Scheinkman, J., 1994. Explorations into factors explaining money market returns. Journal of Finance 49, 1861–1882.

Lehmann, B., Modest, D., 1988. The empirical foundations of the arbitrage pricing theory. Journal of Financial Economics 21, 213–254.

Levine, R., Zervos, S., 1998. Stock markets, banks, and economic growth. American Economic Review 88, 537–558.

MacKinlay, A., Richardson, M., 1991. Using generalized method of moments to test mean-variance efficiency. Journal of Finance 46, 511–527.

McCulloch, R., Rossi, P., 1990. Posterior, predictive and utility based approaches to testing arbitrage pricing theory. Journal of Financial Economics 28, 7–38.

McCulloch, R., Rossi, P., 1991. A Bayesian approach to testing the arbitrage pricing theory. Journal of Econometrics 49, 141–168.

Pastor, L., Stambaugh, R., 1999. Costs of equity capital and model mispricing. Journal of Finance 54, 67–121.

Roll, R., Ross, S., 1980. An empirical investigation of the arbitrage pricing theory. Journal of Finance 35, 1073–1103.

Ross, S., 1976. An arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360.

Safdar, I., 2000. Why has idiosyncratic volatility increased? Unpublished working paper, University of Rochester.

Schwert, G., 1990. Stock volatility and the crash of '87. Review of Financial Studies 3, 77–102.

Schwert, G., Seguin, P., 1990. Heteroskedasticity in stock returns. Journal of Finance 45, 1129–1155.

Shanken, J., 1982. The arbitrage pricing theory: is it testable? Journal of Finance 37, 1129–1140.

Shanken, J., 1990. Intertemporal asset pricing: an empirical investigation. Journal of Econometrics 45, 99–120.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.