Journal of Financial Economics 90 (2008) 252–271

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

How common are common return factors across the NYSE and Nasdaq?[☆]

Amit Goyal[a,*], Christophe Pérignon[b], Christophe Villa[c]

[a] Swiss Finance Institute, University of Lausanne, Lausanne 1015, Switzerland
[b] HEC Paris, 78351 Jouy-en-Josas, France
[c] Audencia Nantes School of Management and CREST-LSM, 44312 Nantes, France

ARTICLE INFO

Article history:
Received 27 September 2007
Received in revised form 27 December 2007
Accepted 31 January 2008
Available online 27 September 2008

JEL classification:
C63
G12
G29

Keywords:
Risk factors
Factor analysis
Asset pricing
Arbitrage pricing theory
Common subspace

ABSTRACT

We entertain the possibility of pervasive factors that are not common across two (or more) groups of securities. We propose and implement a general procedure to estimate the space spanned by common and group-specific pervasive factors. In our empirical analysis, we study the factor structure of excess returns on stocks traded on the NYSE and Nasdaq using our methodology. We find that there are only two common pervasive factors that govern the returns for both NYSE and Nasdaq. At the same time, the NYSE and Nasdaq each have one more group-specific factor that is not the same across the two exchanges. Our results point to the absence of complete similarity between the factors driving the returns on these exchanges.

1. Introduction

Multifactor models of asset returns have played a central role in finance. The arbitrage pricing theory (APT) of Ross (1976) was the first theoretically grounded multifactor model in asset pricing. A key assumption is that the random return of each security is a linear combination of a small number of common factors plus an asset-specific random variable, i.e., the idiosyncratic return. Idiosyncratic returns are supposed to be uncorrelated across firms, which is equivalent to saying that asset returns conform to an exact (or strict) factor structure. However, Chamberlain and Rothschild (1983) show that the diagonality of the residual covariance matrix is not necessary for the proof of the APT and derive a version of APT under a more general approximate factor structure.[1]

While the approximate factor structure allows idiosyncratic returns to be cross-correlated, it is still usually assumed that all factors are pervasive, in the sense that they influence a large number of assets. In an economy partitioned into several groups, such as different sectors, markets, or countries, the assumption that all pervasive factors are common to all groups is too strong. In other

[☆] We thank Michael Brennan, Robert Korajczyk, Richard Roll, Jay Shanken, Sylvie Viguier-Pla, seminars participants at ESSEC, HEC Montreal, Simon Fraser University, and Tilburg University, and especially an anonymous referee for their comments. Part of this research was done while Pérignon was a post-doctoral fellow at the UCLA Anderson School of Management.

* Corresponding author. Tel.: +41 21 692 3676; fax: +41 21 692 3435.
E-mail address: amit.goyal@unil.ch (A. Goyal).

[1] Ross (1976) notes that uncorrelated idiosyncratic returns are not required for APT but does not formally prove it.

0304-405X/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jfineco.2008.01.004

words, the existence of *common pervasive factors* (that affect returns of securities in all groups) and *group-specific pervasive factors* (that affect returns of securities only in some groups) is, a priori, not ruled out by APT. Indeed, Connor and Korajczyk (1993, p. 1264) argue that "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 particularly relevant instance of a natural group structure in financial markets is provided by the two main stock exchanges in the United States, the NYSE and Nasdaq. While the NYSE and Nasdaq provide the same service, their underlying structures, rules, and governing principles are very different. For instance, the NYSE is a specialist-based auction system and Nasdaq is a computer-based dealer market. There is also a lot of evidence that suggests that (the return structures of) the securities that trade on these two exchanges are very different. Nasdaq has been shown to be less integrated with the NYSE and the Amex than the latter two are with each other (Naranjo and Protopapadakis, 1997). Fama and French (2004) report that Nasdaq accounts for most new listings after 1972: the number of NYSE new listings is approximatively equal to 10% of Nasdaq new listings. Before the October 1987 crash, Nasdaq stock volatility was comparable with that observed in other U.S. markets, but Nasdaq has since become unusually volatile (Schwert, 2002). Although Campbell, Lettau, Malkiel, and Xu (2001) attest that the level of idiosyncratic risk has globally trended up in the United States, Malkiel and Xu (2003) report that the rise has been much stronger for Nasdaq stocks than for NYSE stocks. Baruch and Saar (2006) explain that the listing decision between NYSE and Nasdaq is not due to differences in the structure or rules of the markets, but rather to different return patterns of the securities that are traded on these markets.

In spite of this body of evidence, there has been very little work done on pinpointing the exact nature of the difference(s) between factors across these two groups. A central difficulty in this endeavor, namely identifying common pervasive factors and group-specific pervasive factors, has been the absence of a formal procedure in extracting these factors. Cho, Eun, and Senbet (1986) use a variant of factor analysis called inter-battery factor analysis to estimate the factor sensitivity matrix for factors common across pairs of countries. Unfortunately, inter-battery factor analysis can estimate neither common factors for more than two groups nor country-specific factors, which are not ruled out by the international APT. As the first contribution in this paper, we propose a general procedure for estimating common pervasive risk factors in the presence of several groups of assets under the framework of a large number of assets in each group.

Our methodology takes advantage of a multigroup version of the standard principal component analysis (PCA). The key to the analysis is that the space spanned by the latent factors can be consistently estimated in each group. Our procedure then estimates a single common-factor subspace that resembles all group-factor subspaces as closely as possible. Since we study large cross-sections with limited time series, our methodology generalizes the pioneering approach of Connor and Korajczyk (1986, 1988) by accounting for the group structure of the returns. We conduct several simulation experiments to test the usefulness of our method. We find that the traditional PCA falls short in estimating the space spanned by the common pervasive factors even in large samples. In contrast, our method outperforms the standard PCA for estimating the space spanned by common pervasive factors, as well as the space spanned by all (common and group-specific) pervasive factors, especially in the presence of cross-correlated residuals.

Our simulations show that, in some settings, traditional techniques have comparable performance to our approach in identifying the space of all pervasive factors. In these situations, the efficacy of our procedure comes through in separating these factors into common pervasive and group-specific pervasive factors. We demonstrate, however, that it can be important to identify the exact nature of the split in factors across groups (and hence to rely on our approach) in some applications (for example, in the estimation of the price of risk of factors).

We apply our methodology to the return structure of NYSE/Amex and Nasdaq securities by studying excess returns on stocks. For ease of notation, we frequently refer to the group consisting of stocks traded on NYSE and Amex as the "NYSE group" while the second group is referred to as the "Nasdaq group." We consider a 25-year sample covering the 1978–2002 period, which we divide into five 60-month subperiods. For each of the subperiods, we extract common pervasive factors and group-specific pervasive factors. We find that, while there are three factors in each group, there are at most two common pervasive factors. In other words, we find that there are two common factors that govern the returns for both the NYSE and Nasdaq. At the same time, the NYSE and Nasdaq each have one more group-specific factor that is not the same across the two exchanges. Therefore, our empirical results point to the absence of complete correspondence in the return structures of stocks on these two exchanges.

A comparison of our technique versus traditional PCA, wherein the data are pooled, reveals that traditional techniques have difficulties in uncovering the common factors across the two groups (even though, as noted before, their performance is similar to that of our procedure in uncovering *all* pervasive factors in some sample periods). This is comforting since it shows that the efficacy of our procedure, which we document using simulation experiments, carries over to real data. We also make attempts to "identify" our common pervasive factors. Because of the well-known rotational indeterminacy problem of PCA, we can only compare canonical correlations of our factors with other factors. We find that our common pervasive factors are highly related to the standard benchmark factors, such as the three Fama and French (1993) factors. Although the standard factors explain less of the common variation in returns in recent years than in the early part of the sample, our results still point to the usefulness of the Fama-French model because the canonical correlation of these factors with our common pervasive factors is in excess of 0.95 for most of the sample period.

Beyond the theoretical contribution mentioned earlier, our paper makes at least two practical contributions. First, our methodology can be applied to compare the price of risk of factors in different groups, such as various countries. Second, we add to the voluminous literature on multifactor models, which is motivated in part to describe the covariance structure of asset returns in a parsimonious manner. We find that there are at most three pervasive factors in each group, two of which are common. Even though we estimate these factors ex post, the consistency of the result across a long time horizon is comforting.

Our paper adds to the small literature on the analysis of group structure superimposed on classical factor analysis. In addition to the study by Cho, Eun, and Senbet (1986) mentioned earlier, Bekaert, Hodrick, and Zhang (2005) and Korajczyk and Sadka (2007) have recently proposed a two-step procedure to identify common and group-specific factors. However, we show in this paper that the first step of their procedure, which aims to estimate common factors, does not work well when, for instance, some pervasive factors are common to a subset of groups only and other pervasive factors are group-specific, or if the sample size or the factor sensitivities (betas) vary across groups. In contrast, our methodology is general and is able to estimate common pervasive factors as well as group-specific pervasive factors. As will become clear later, one can also view our method as a multigroup version of the heteroskedastic factor analysis (HFA) of Jones (2001). Finally, our work is related to the studies on differences between the NYSE and Nasdaq exchanges. The closest companion from this strand of literature to our paper is Baruch and Saar (2006) who show that a stock is more liquid when it is listed on a market where "similar" securities are traded. Our study differs from theirs in focus (we investigate sources of common and group-specific covariation in returns), approach (we introduce a new methodology), and sample period (we study a 25-year period versus their three-year sample period).

The outline of the rest of the paper is as follows. Section 2 introduces the group structure factor model. Section 3 details our methodology and provides simulation results to discuss the usefulness of our procedure relative to other existing methods. In Section 4, we apply our methodology to NYSE and Nasdaq stocks. Section 5 concludes and provides directions for future research.

2. Background

In the APT of Ross (1976), arbitrage arguments are used to show that the expected return on a specific stock is a linear function of sensitivities to a finite number of pricing factors. The return-generating process is assumed to be driven by a K-factor model given by

$$R = BF^K + \varepsilon \tag{1}$$

where R is a $(N \times T)$ matrix of excess returns, B is a $(N \times K)$ matrix of factor sensitivities, F^K is a $(K \times T)$ matrix of K pricing factor returns, ε is a $(N \times T)$ matrix of idiosyncratic returns, N denotes the number of assets, and T denotes the number of observations. The key assumptions are $\mathrm{E}(\varepsilon) = 0$, $\mathrm{E}(F^K F^{K\prime}) = I_K$, $\mathrm{E}(F^K \varepsilon') = 0$, and $\mathrm{E}(\varepsilon\varepsilon') = D$ (Assumptions 1–4) where I_K is a $(K \times K)$ identity matrix and D is a diagonal matrix. Consequently, one can express the $(N \times N)$ excess return covariance matrix Σ in terms of the parameters of the model as $\Sigma = BB' + D$. Model (1), together with Assumptions 1–4, is known as the exact factor model. When N is relatively small, and much smaller than T, estimation of the B and D matrices can be obtained by classical factor analysis. Although not essential, the joint-normality of the stock returns and residuals is often assumed and maximum likelihood estimation is employed. The main limitation of this approach is that as the number of parameters increases with N, computational difficulties make it necessary either to consider a small cross-section or to combine the numerous available series into portfolios.

Chamberlain and Rothschild (1983) propose a more general setting, called the approximate factor model, by allowing cross-sectionally correlated residuals, where the residual variance matrix is not required to be diagonal. This model requires the following assumptions: $\mathrm{E}(\varepsilon\varepsilon') = V$, $\mathrm{plim}_{N\to\infty}\,\lambda_1(v) < \infty$ and $\mathrm{plim}_{N\to\infty}\,\lambda_K(BB') = \infty$ (Assumptions 5–7). Here $\lambda_k(\cdot)$ denotes the k^{th} eigenvalue of a given covariance matrix. Assumption 6 states that V is a matrix with bounded eigenvalues. Assumption 7 implies that none of the K latent factors can be neglected and that all factors are pervasive.

In an approximate factor model, Chamberlain and Rothschild (1983) show that the factor structure estimated using PCA converges to a rotation of the true pervasive factors when the covariance matrix of returns is known and the number of assets N diverges. Connor and Korajczyk (1986, 1988) show that principal components have good properties also in samples having a finite number of observations T ($T \ll N$). They show that the first K eigenvectors of the $(T \times T)$ cross-product matrix $\Omega = (1/N)R'R$ converge to a rotation of the true pervasive factor returns under the assumptions of no serial correlation and homoskedasticity in the residuals. Jones (2001) generalizes the Connor and Korajczyk's method (hereafter APCA) to the heteroskedastic case by letting the limit of the cross-product matrix of residuals be diagonal. Bai and Ng (2002) and Bai (2003) show that PCA remains consistent in the presence of both serial and cross-sectional correlation in the residuals, as well as heteroskedasticity over time and across residuals, under the framework of large N and large T. Indeed, the estimated factor structure can be obtained by running a PCA either from the $(N \times N)$ sample covariance matrix when $N < T$, or by running an APCA from the $(T \times T)$ cross-product matrix when $T < N$.

While (A)PCA appears to be a general method to estimate the space spanned by the true pervasive factors, this method says nothing about the space spanned by common pervasive factors and group-specific pervasive factors. APT is consistent with the existence of common pervasive factors that affect the returns of all securities in all groups, *and* group-specific pervasive factors that affect the returns of all securities only in some (but not all) groups. We therefore generalize the above framework by considering an economy characterized by the return-generating process given in Eq. (1) and Assumptions 1–3. This economy is divided into a finite number G of sectors or countries, each of them containing $N_g = (N/G) + \gamma_g$ securities with $\Sigma_{g=1}^{G}\gamma_g = 0, (T < N_g)$. The number of securities in each sector goes to infinity with N since G and γ_g are constants. In our model, some of the common factors $F^K = (F_1, F_2, \ldots, F_K)'$ in Eq. (1) are possibly not pervasive in all groups. We assume that the number of pervasive factors in group g is $K_g \leqslant K$. Moreover, we denote by K_C the number of common pervasive factors. The return generating process in group g is

$$R^g = B^g F^{K_g} + \varepsilon^g \quad (2)$$

where R^g is a $(N_g \times T)$ matrix of excess returns in group g, B^g is a $(N_g \times K_g)$ matrix of factor sensitivities, ε^g is a $(N_g \times T)$ matrix of idiosyncratic returns, and F^{K_g} is a $(K_g \times T)$ matrix of K_g pricing factor returns. The group-specific factors F^{K_g} are a subset of F^K ($F^K : F^K = \bigcup_{g=1}^{G} F^{K_g}$). Also, the K_C common pervasive factors are given by $\bigcap_{g=1}^{G} F^{K_g}$.

The key assumptions are $\mathrm{E}(\varepsilon^g \varepsilon^{g\prime}) = V^g$, $\mathrm{plim}_{N_g \to \infty} \lambda_1 (V^g) < \infty$, and $\mathrm{plim}_{N_g \to \infty} \lambda_{K_g}(B^g B^{g\prime}) = \infty$ (Assumptions 5b–7b). In our model, for $i \neq j$, $i,j = 1, \ldots, K$, it is possible to encounter the following situation: in group g_1, $\mathrm{plim}_{N_{g_1} \to \infty} \lambda_i (B^{g_1} B^{g_1\prime}) = \infty$ and $\mathrm{plim}_{N_{g_1} \to \infty} \lambda_j (V^{g_1}) < \infty$ and in an other group g_2, $\mathrm{plim}_{N_{g_2} \to \infty} \lambda_i (V^{g_2}) < \infty$ and $\mathrm{plim}_{N_{g_2} \to \infty} \lambda_j (B^{g_2} B^{g_2\prime}) = \infty$. Therefore, the i^{th} (j^{th}) factor in group g_1 (g_2) is pervasive while it is not pervasive in group g_2 (g_1). Security returns are thus determined by either *common pervasive* risk factors that influence almost all stocks in all groups, or *group-specific pervasive* risk factors that only affect stocks in some groups. We emphasize that model (2) is a generalization of the traditional APT models only in the sense of additional group structure superimposed on the underlying securities. The expected return relation, either approximate as in Ross (1976) and Chamberlain and Rothschild (1983) or exact as in Connor (1984), continues to hold in this model. However, the group structure imposes additional difficulties in estimation, which we next explore.

3. Estimating common pervasive factors in an approximate factor model

In order to estimate a model similar to that in Eq. (2), Bekaert, Hodrick, and Zhang (2005) and Korajczyk and Sadka (2007) propose the following two-step procedure. In the first step, they use the first principal components estimated from the pooled data across all groups as an estimate of the space spanned by the common factors. In a second step, the specific factors are estimated using data within each group (the first principal components in each group), and then the specific factors are orthogonalized with respect to the common ones. Although this approach works extremely well when all pervasive factors are pervasive in all groups (i.e., $K_C = K = K_g$), it is not perfectly suited for many other cases encountered in practice. For instance, the first principal components estimated from pooled data might not estimate the space spanned by the true common pervasive factors if some factors are common to a subset of groups only, or if the sample sizes or betas vary across groups (see our simulation results in Tables 1–3 for details). Furthermore, data from the G groups are needed to perform the first step of the procedure. As shown by Boivin and Ng (2006), the quality of the estimation with an APCA of the true pervasive factors can deteriorate since more data are not always better for estimating an approximate factor model.[2]

A first attempt to generalize principal components to a G-group method was made by Krzanowski (1979), whose approach consists of a descriptive technique for comparing the subspaces spanned by the first K components in each group. This is achieved by approximating the G K-dimensional subspaces by a single K-dimensional subspace. His idea was to find the minimum angle between the axis of the subspaces spanned by the first K eigenvectors of several covariance matrices. Krzanowski (1979, Theorem 3) proves that the solution is given by the first K eigenvectors of $P = \sum_{g=1}^{G} P_g$ where $P_g = J_g^K J_g^{K\prime}$ is the g^{th} group's eigenprojection corresponding to the first K eigenvalues and J_g^K are the first K eigenvectors of the covariance matrix Σ_g.

Another multigroup method is the common principal component analysis (hereafter CPCA) of Flury (1984). Flury considers the situation in which the G groups have a common subspace for all g, i.e., $K_g = K$. In CPCA, all group-specific covariance matrices Σ_g, $g = 1, \ldots, G$, are simultaneously diagonalizable by the same orthogonal matrix J. The CPCA transformation can be viewed as a rotation yielding variables that are as uncorrelated as possible simultaneously in G groups. However, this model has one major drawback. Under CPCA, the assumption that the same orthogonal matrix diagonalizes all Σ_g simultaneously is often too restrictive in practice. One generalization of CPCA suggested by Flury (1987) is to assume that K eigenvectors of each matrix span the same subspace. Flury's common subspaces can correspond to any principal components, not necessarily those with the largest eigenvalues.[3]

More recently, Schott (1999) relaxes the assumptions of Krzanowski (1979) and Flury (1987) by considering the PCA of G groups when, for each group, the first K_g for $g = 1, \ldots, G$ principal components account for most of the total variability of observations in each group. The set of these $\sum_{g=1}^{G} K_g$ principal component vectors spans a space of dimension K, where K is less than $\min(\sum_{g=1}^{G} K_g, T)$. This subspace is estimated as the first K eigenvectors of $P = \sum_{g=1}^{G} P_g$ where P_g is the g^{th} group's eigenprojection $P_g = J_g^{K_g} J_g^{K_g\prime}$ corresponding to the first K_g eigenvalues. The

[2] The results of Boivin and Ng (2006) hold only in finite samples, since the estimators that they consider are consistent.

[3] Pérignon and Villa (2006) provide an application of Flury's techniques to model interest rates.

Table 1
Simulation results

We consider three groups of data ($G = 3$), in which the data generating process for excess returns, R_{it}, is assumed to be

$$R_{it} = \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i3}F_{3t} + \varepsilon_{it}, \quad \forall i \in [1, N_1]$$
$$R_{it} = \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i4}F_{4t} + \varepsilon_{it}, \quad \forall i \in [N_1 + 1, N_1 + N_2]$$
$$R_{it} = \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i5}F_{5t} + \varepsilon_{it}, \quad \forall i \in [N_1 + N_2 + 1, N_1 + N_2 + N_3]$$

where the factors $F_{j\cdot}, j = 1, \ldots, 5$ are orthogonal and N(0, 1) variables. The factor loadings $\beta_{\cdot j}, j = 1, \ldots, 5$ are N(1, 1) variates and $\varepsilon_{it} = \sigma u_{it}$ with $\sigma = \sqrt{3}$ and $u_{i\cdot}$ is N(0, 1). Furthermore, we use $T = 60$ and $N_g = 1,000$. Factors are estimated by applying alternatively APCA (on the pooled data) and MGFA. The span of estimated factors, $\hat{F}$, is compared to that of actual factors, F, by the statistic

$$S = \frac{\text{trace}(F'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'F)}{\text{trace}(F'F)}$$

We report the mean and standard deviation (across 1,000 simulations) of S separately for the two common pervasive factors (first two columns) and for all pervasive factors (last two columns). Panel A is for the setting described above while the other panels explore variations on this basic theme. In each panel, boldface figures denote the best-performing method.

The number of common factors (K_C) is determined to be the number of eigenvalues of the matrix P that are equal to G. The average of these eigenvalues is also reported in the line following the panel description. For example, the number 2.997 for the first eigenvalue in Panel A means that on average the first factor is pervasive in 2.997 groups. The standard deviations of these averages are typically very small and are not reported.

	Common pervasive factors		All pervasive factors	
	APCA	MGFA	APCA	MGFA
Panel A: Basic setting				
(Average eigenvalues of $P = [2.997, 2.994, 1.014, 1.000, 0.986]$)				
Mean	0.9510	**0.9992**	**0.9987**	**0.9987**
Std. Dev.	0.0330	0.0001	0.0001	0.0001
Panel B: No group-specific pervasive factors, $\beta_{\cdot 3} = \beta_{\cdot 4} = \beta_{\cdot 5} = 0$				
(Average eigenvalues of $P = [2.999, 2.996]$)				
Mean	**0.9995**	**0.9995**	**0.9995**	**0.9995**
Std. Dev.	0.0001	0.0001	0.0001	0.0001
Panel C: No common pervasive factors, $\beta_{\cdot 1} = \beta_{\cdot 2} = 0$				
(Average eigenvalues of $P = [1.007, 1.000, 0.994]$)				
Mean	Failure	**1.0000**	**0.9995**	**0.9995**
Std. Dev.	Failure	0.0000	0.0001	0.0001
Panel D: Common pervasive factors with higher sensitivity, $\beta_{\cdot 1}$ and $\beta_{\cdot 2}$ are N(2, 1)				
(Average eigenvalues of $P = [2.999, 2.994, 1.015, 1.000, 0.985]$)				
Mean	0.9801	**0.9994**	0.9986	**0.9987**
Std. Dev.	0.0022	0.0001	0.0002	0.0001
Panel E: Group-specific pervasive factors with higher sensitivity, $\beta_{\cdot 3}$, $\beta_{\cdot 4}$, and $\beta_{\cdot 5}$ are N(2, 1)				
(Average eigenvalues of $P = [2.996, 2.994, 1.010, 1.000, 0.990]$)				
Mean	0.3447	**0.9991**	**0.9992**	**0.9992**
Std. Dev.	0.0035	0.0001	0.0001	0.0001
Panel F: Different group sample sizes, $N_1 = 2,000$, $N_2 = 1,000$, and $N_3 = 500$				
(Average eigenvalues of $P = [2.997, 2.993, 1.015, 1.000, 0.985]$)				
Mean	0.9180	**0.9991**	**0.9985**	**0.9985**
Std. Dev.	0.0205	0.0001	0.0002	0.0002
Panel G: Same group-specific pervasive factors in two groups, $F_3 = F_4$ and $K = 4$				
(Average eigenvalues of $P = [2.997, 2.994, 1.998, 1.000]$)				
Mean	0.6881	**0.9992**	0.9989	**0.9990**
Std. Dev.	0.1619	0.0001	0.0001	0.0001

matrix P will have rank equal to K, and the space spanned by the eigenvectors corresponding to the positive eigenvalues of P will be the partial common principal component subspace as named by Schott. Moreover, the first K_C eigenvectors of P corresponding to the first K_C eigenvalues give an estimation of the common principal component subspace. In theory, the first K_C eigenvalues are all equal to G.

We use Schott's (1999) results to extend Connor and Korajczyk (1986, 1988) to group factors. This hybrid approach gives us a straightforward way to estimate the spaces spanned by both common pervasive factors and all

Table 2
Simulation results—cross-correlated residuals

We consider three groups of data ($G = 3$), in which the data-generating process for excess returns, R_{it}, is assumed to be

$$R_{it} = \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i3}F_{3t} + \sum_{j=1}^{n}\gamma_{ij}L_{jt} + \varepsilon_{it}, \quad \forall i \in [1, N_1]$$

$$\gamma_{ij} \sim N(1,1), \quad j = 1,\ldots,n, \text{ if } i \in [1+(j-1)N_1/n, jN_1/n] \text{ and 0 otherwise}$$

$$R_{it} = \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i4}F_{4t} + \sum_{j=1}^{n}\gamma_{ij}L_{jt} + \varepsilon_{it}, \quad \forall i \in [N_1+1, N_1+N_2]$$

$$\gamma_{ij} \sim N(1,1), j = 1,\ldots,n, \text{ if } i \in [N_1+1+(j-1)N_2/n,\ N_1+jN_2/n] \text{ and 0 otherwise}$$

$$R_{it} = \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i5}F_{5t} + \sum_{j=1}^{n}\gamma_{ij}L_{jt} + \varepsilon_{it}, \quad \forall i \in [N_1+N_2+1, N_1+N_2+N_3]$$

$$\gamma_{ij} \sim N(1,1), \ j = 1,\ldots,n, \text{ if } i \in [N_1+N_2+1+(j-1)N_3/n, N_1+N_2+jN_3/n] \text{ and 0 otherwise}$$

where the factors $F_{j\cdot}, j = 1,\ldots,5$ are orthogonal and N(0,1) variables. The factor loadings $\beta_{\cdot j}, j = 1,\ldots,5$ are N(1,1) variates and $\varepsilon_{it} = \sigma u_{it}$ with $\sigma = \sqrt{3}$ and $u_{i\cdot}$ is N(0,1). We divide each group into $n = 10$ subgroups (with ten percent of the data in each subgroup) and suppose that, in each subgroup, asset returns are also driven by an additional non-pervasive factor. Furthermore, we use $T = 60$, $N_g = 1,000$, and we perform in each case 1,000 simulations. Factors are estimated by applying alternatively APCA and MGFA. The span of estimated factors, $\hat{F}$, is compared to that of actual factors, F, by the statistic

$$S = \frac{\text{trace}(F'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'F)}{\text{trace}(F'F)}$$

We report the mean and standard deviation (across 1,000 simulations) of S separately for the two common pervasive factors (first two columns) and for all pervasive factors (last two columns). Panel A is for the setting described above while the other panels explore variations on this basic theme. In each panel, boldface figures denote the best-performing method.

The number of common factors (K_C) is determined to be the number of eigenvalues of the matrix P that are equal to G. The average of these eigenvalues is also reported in the line following the panel description. The standard deviations of these averages are typically very small and are not reported.

	Common pervasive factors		All pervasive factors	
	APCA	MGFA	APCA	MGFA
Panel A: Basic setting				
(Average eigenvalues of $P = [2.995, 2.987, 1.018, 0.999, 0.984]$)				
Mean	0.9401	**0.9916**	0.9880	**0.9913**
Std. Dev.	0.0032	0.0009	0.0013	0.0008
Panel B: No group-specific pervasive factors, $\beta_{\cdot 3} = \beta_{\cdot 4} = \beta_{\cdot 5} = 0$				
(Average eigenvalues of $P = [2.998, 2.989]$)				
Mean	**0.9872**	**0.9872**	**0.9872**	**0.9872**
Std. Dev.	0.0007	0.0007	0.0007	0.0007
Panel C: No common pervasive factors, $\beta_{\cdot 1} = \beta_{\cdot 2} = 0$				
(Average eigenvalues of $P = [1.057, 0.976, 0.967]$)				
Mean	Failure	**1.0000**	0.9626	**0.9719**
Std. Dev.	Failure	0.0000	0.0022	0.0016
Panel D: Common pervasive factors with higher sensitivity, $\beta_{\cdot 1}$ and $\beta_{\cdot 2}$ are N(2,1)				
(Average eigenvalues of $P = [2.997, 2.987, 1.015, 1.000, 0.985]$)				
Mean	0.9740	**0.9946**	0.9927	**0.9954**
Std. Dev.	0.0026	0.0004	0.0014	0.0005
Panel E: Group-specific pervasive factors with higher sensitivity, $\beta_{\cdot 3}$, $\beta_{\cdot 4}$, and $\beta_{\cdot 5}$ are N(2,1)				
(Average eigenvalues of $P = [2.992, 2.987, 1.018, 0.997, 0.985]$)				
Mean	0.3370	**0.9958**	0.9916	**0.9929**
Std. Dev.	0.0037	0.0006	0.0007	0.0005
Panel F: Different group sample sizes, $N_1 = 2,000$, $N_2 = 1,000$, and $N_3 = 500$				
(Average eigenvalues of $P = [2.994, 2.985, 1.019, 0.999, 0.982]$)				
Mean	0.9085	**0.9914**	0.9762	**0.9912**
Std. Dev.	0.0179	0.0008	0.0091	0.0007
Panel G: Same group-specific pervasive factors in two groups, $F_3 = F_4$ and $K = 4$				
(Average eigenvalues of $P = [2.995, 2.987, 1.996, 1.000]$)				
Mean	0.6677	**0.9915**	0.9902	**0.9917**
Std. Dev.	0.1542	0.0008	0.0008	0.0006

Table 2 (*continued*)

	Common pervasive factors		All pervasive factors	
	APCA	MGFA	APCA	MGFA
Panel H: Number of non-pervasive factors ($n = 5$)				
(Average eigenvalues of $P = [2.994, 2.982, 1.030, 0.995, 0.976]$)				
Mean	0.9274	**0.9822**	0.9046	**0.9853**
Std. Dev.	0.0038	0.0022	0.0505	0.0014

Table 3

Simulation results—missing data

We consider three groups of data ($G = 3$), in which the data-generating process for excess returns, R_{it}, is assumed to be

$$R_{it} = \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i3}F_{3t} + \varepsilon_{it}, \quad \forall i \in [1, N_1]$$
$$R_{it} = \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i4}F_{4t} + \varepsilon_{it}, \quad \forall i \in [N_1 + 1, N_1 + N_2]$$
$$R_{it} = \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i5}F_{5t} + \varepsilon_{it}, \quad \forall i \in [N_1 + N_2 + 1, N_1 + N_2 + N_3]$$

where the factors $F_{j\cdot}, j = 1, \ldots, 5$ are orthogonal and $N(0, 1)$ variables. The factor loadings $\beta_{\cdot j}, j = 1, \ldots, 5$ are $N(1, 1)$ variates and $\varepsilon_{it} = \sigma u_{it}$ with $\sigma = \sqrt{3}$ and $u_{i\cdot}$ is $N(0, 1)$. Furthermore, we use $T = 60$ and $N_g = 1,000$. Half the firms are chosen to have missing observations. We set 15 months (25% of the total number of observations) to have missing values. The firms and months with missing values are chosen at random. Missing values are replaced by zero. Factors are estimated by applying alternatively APCA (on the pooled data) and MGFA. The span of estimated factors, $\hat{F}$, is compared to that of actual factors, F, by the statistic

$$S = \frac{\text{trace}(F'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'F)}{\text{trace}(F'F)}$$

We report the mean and standard deviation (across 1,000 simulations) of S separately for the two common pervasive factors (first two columns) and for all pervasive factors (last two columns). Panel A is for the setting described above while the other panels explore variations on this basic theme. In each panel, boldface figures denote the best-performing method.

The number of common factors (K_C) is determined to be the number of eigenvalues of the matrix P that are equal to G. The average of these eigenvalues is also reported in the line following the panel description. The standard deviations of these averages are typically very small and are not reported.

	Common pervasive factors		All pervasive factors	
	APCA	MGFA	APCA	MGFA
Panel A: Basic setting				
(Average eigenvalues of $P = [2.996, 2.991, 1.017, 1.000, 0.983]$)				
Mean	0.9510	**0.9988**	**0.9978**	**0.9978**
Std. Dev.	0.0036	0.0002	0.0002	0.0002
Panel B: No group-specific pervasive factors, $\beta_{\cdot 3} = \beta_{\cdot 4} = \beta_{\cdot 5} = 0$				
(Average eigenvalues of $P = [2.998, 2.994]$)				
Mean	**0.9993**	**0.9993**	**0.9993**	**0.9993**
Std. Dev.	0.0001	0.0001	0.0001	0.0001
Panel C: No common pervasive factors, $\beta_{\cdot 1} = \beta_{\cdot 2} = 0$				
(Average eigenvalues of $P = [1.008, 1.0000, 0.992]$)				
Mean	Failure	**1.0000**	**0.9992**	**0.9992**
Std. Dev.	Failure	0.0000	0.0001	0.0001
Panel D: Common pervasive factors with higher sensitivity, $\beta_{\cdot 1}$ and $\beta_{\cdot 2}$ are $N(2, 1)$				
(Average eigenvalues of $P = [2.998, 2.990, 1.022, 1.000, 0.979]$)				
Mean	0.9785	**0.9986**	0.9956	**0.9965**
Std. Dev.	0.0031	0.0003	0.0014	0.0005
Panel E: Group-specific pervasive factors with higher sensitivity, $\beta_{\cdot 3}$, $\beta_{\cdot 4}$, and $\beta_{\cdot 5}$ are $N(2, 1)$				
(Average eigenvalues of $P = [2.994, 2.990, 1.013, 1.000, 0.987]$)				
Mean	0.3452	**0.9985**	**0.9986**	0.9982
Std. Dev.	0.0041	0.0002	0.0002	0.0002
Panel F: Different group sample sizes, $N_1 = 2,000$, $N_2 = 1,000$, and $N_3 = 500$				
(Average eigenvalues of $P = [2.995, 2.990, 1.018, 1.000, 0.982]$)				
Mean	0.9140	**0.9986**	**0.9974**	**0.9974**
Std. Dev.	0.0322	0.0002	0.0004	0.0003
Panel G: Same group-specific pervasive factors in two groups, $F_3 = F_4$ and $K = 4$				
(Average eigenvalues of $P = [2.996, 2.991, 1.997, 1.000]$)				
Mean	0.6803	**0.9988**	**0.9983**	**0.9983**
Std. Dev.	0.1570	0.0002	0.0002	0.0002

pervasive factors. Our procedure, which we refer to as multigroup factor analysis (hereafter MGFA), can be presented as follows:

1. Perform an APCA in each group.
   (a) Compute the cross-product matrix $\hat{\Omega}_g = (1/N_g)R_g'R_g$ of returns.
   (b) Compute the first K_g eigenvectors $\hat{J}_g^{K_g}$ of $\hat{\Omega}_g$, corresponding to the first K_g eigenvalues.
   (c) Compute $\hat{P}_g = \hat{J}_g^{K_g}\hat{J}_g^{K_g\prime}$, the g[th] group's eigenprojection corresponding to the first K_g eigenvalues.
2. Compute the eigenvectors $\hat{J}_0$ of $\hat{P} = \sum_{g=1}^{G}\hat{P}_g$.
3. The common pervasive space spanned by the true common pervasive factors is estimated as the first K_C columns of $\hat{J}_0$, i.e., the first K_C eigenvectors of $\hat{P}$ corresponding to the first K_C eigenvalues.
4. The pervasive space spanned by all true pervasive factors is estimated as the first K columns of $\hat{J}_0$, i.e., the first K eigenvectors of $\hat{P}$ corresponding to the first K eigenvalues.

The key to the analysis is that the space spanned by the latent factors can be consistently estimated with APCA in each group and the reduced subspace retains most of the variability *within* all groups, which is not the case with an APCA to the pooled data. Moreover, if (following Jones, 2001) the space spanned by the true factors in each group is estimated by HFA in step one of the above procedure, our approach can also be viewed as a multigroup version of HFA.

3.1. Number of factors

The above procedure implicitly assumes knowledge of the number of factors. Of course, the estimation of the number of factors is an integral part of factor analysis. Under the assumption of a large number of assets in each group, the number of pervasive factors in a group (K_g) or in the economy (K) can be estimated using the information criteria (PC_p and IC_p) developed by Bai and Ng (2002). The PC statistics are similar to the conventional panel data criteria and the IC statistics are similar to the conventional time-series criteria. The important difference is that the penalty is a function of both the number of cross-sections and the length of the time series. These criteria estimate the number of factors even under heteroskedasticity in both the time and cross-sectional dimensions and also under weak serial dependence and cross-sectional dependence. Further details on these criteria are provided in Appendix A.

In the special case of two groups, K_C is given by the number of canonical correlations among the pervasive factors in each group that are equal to one (see Bai and Ng, 2006). To the best of our knowledge, there are no established criteria to estimate the dimension (K_C) of the common pervasive subspace in the presence of more than two groups. However, we can rely on the results of Dauxois and Pousse (1975) to solve this problem. Dauxois and Pousse point out that the spectral analysis of $P = P_1 + P_2$ is a canonical analysis of two subspaces. Therefore, with two groups, K_C equals the number of eigenvalues from the spectral analysis of P that are equal to two. Although canonical correlation and the eigenvalue-based method provide the same results with two groups, the latter technique can be used with any number of groups. We thus set K_C as the number of eigenvalues of $P = \sum_{g=1}^{G} P_g$ that are equal to G. As shown in the simulations in the next subsection, this approach leads to very accurate estimates for the number of common factors.

These criteria for determining the number of common pervasive factors are also very intuitive. As an illustration, consider two groups wherein stock returns in each group are driven by three factors among which two are common and one is group-specific. In this economy, the total number of pervasive factors is four (i.e., two common across two groups and two group-specific factors, one for each group). The sequence of eigenvalues of $P = P_1 + P_2$ is [2, 2, 1, 1, 0, 0], which suggests that two factors affect both groups ($K_C = 2$) and two factors affect one group only. Alternatively, the first three canonical correlations between the pervasive factors in each group are [1, 1, 0].

3.2. Simulations

We perform a variety of simulations to compare the ability of our MGFA procedure and APCA methods to estimate pervasive factors. We carry out this analysis for uncorrelated and cross-correlated residuals.

3.2.1. Basic setting

Our basic setting consists of three groups of data, $G = 3$, in which the excess returns R_{it} are driven by five pervasive factors, $K = 5$. There are two common pervasive factors ($K_C = 2$) and three group-specific factors. Each group has three pervasive factors $K_g = 3, \forall g$. In other words, out of the three pervasive factors for each group, two are common across groups (i.e., they affect asset returns in all groups) and one is group-specific. The exact return-generating process is assumed to be

$$\begin{aligned} R_{it} &= \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i3}F_{3t} + \varepsilon_{it}, \quad \forall i \in [1, N_1] \\ R_{it} &= \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i4}F_{4t} + \varepsilon_{it}, \quad \forall i \in [N_1+1, N_1+N_2] \\ R_{it} &= \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i5}F_{5t} + \varepsilon_{it}, \quad \forall i \in [N_1+N_2+1, N_1+N_2+N_3] \end{aligned} \tag{3}$$

where the factors F_{jt}, $j = 1, \ldots, 5$, are orthogonal and N(0,1) variables. The factor loadings β_{ij}, $j = 1, \ldots, 5$ are N(1,1) variates and $\varepsilon_{it} = \sigma u_{it}$ with $\sigma = \sqrt{3}$ and $u_{i.}$ is N(0,1). We use $T = 60$ and $N_g = 1{,}000$, and in each case we perform 1,000 simulations. Factors are estimated by alternatively applying APCA (to pooled data) and MGFA. Since both methods only identify the space spanned by the factors (e.g., the estimated factors need not coincide with the true factors), we compare the span of the estimated factors, $\hat{F}$, with the span of the true factors, F. Following Boivin and Ng (2006), we calculate for each extraction method the statistic

$$S = \frac{\text{trace}(F'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'F)}{\text{trace}(F'F)} \tag{4}$$

A large S statistic indicates a small discrepancy between the spaces spanned by the actual and estimated factors. The S statistic ranges from zero (totally unpredictable) to one (perfectly predictable). In the first set of simulations, we also assume that the errors are mutually uncorrelated within and across groups. We report the mean and standard-deviation (across 1,000 simulations) of the S statistic separately for the two common pervasive factors (first two columns) and for all pervasive factors (last two columns).

We consider the setting described above in Panel A and explore variations on the basic theme in subsequent panels of Table 1. In Panel B, we assume that the excess returns are driven by common factors only ($\beta_{.3} = \beta_{.4} = \beta_{.5} = 0$). In Panel C, we change the basic setting by removing the effect of all common factors ($\beta_{.1} = \beta_{.2} = 0$). We study the effect of a higher sensitivity to all common pervasive factors in Panel D ($\beta_{.1}$ and $\beta_{.2}$ are N(2, 1)), whereas we study the effect of higher sensitivity to group-specific factors in Panel E ($\beta_{.3}$, $\beta_{.4}$, and $\beta_{.5}$ are N(2, 1)). In Panel F, we investigate the effect of different group sample sizes by considering $N_1 = 2,000$, $N_2 = 1,000$, and $N_3 = 500$. Finally, we study in Panel G the effect of the presence of a factor that is common in only two groups ($F_3 = F_4$ and $K = 4$). We boldface figures denoting the best-performing method in each panel.

In all simulations, we estimate the number of factors (both K_g and K_C) for MGFA. We determine K_g using the PC_p and IC_p criteria of Bai and Ng (2002), and set K_C to be the number of eigenvalues of the matrix P that are equal to G. However, for APCA, we assume that the first two factors are common pervasive factors, even though APCA itself provides no guidance on this subject. We report the average of the eigenvalues of P in the line following the panel description. For example, the number 2.997 for the first eigenvalue in Panel A means that on average the first factor is pervasive in 2.997 groups. The standard deviations of these averages are typically very small and are not reported. In all panels, we can see that the number of common factors is estimated quite precisely. We are able to determine that there are two common pervasive factors in all groups in all panels (except Panel C, where we correctly deduce that there are no common pervasive factors). The criterion is also useful for detecting that there is an additional factor common to only two groups in Panel G.

The results in Table 1 suggest that MGFA systematically outperforms APCA in estimating the true common pervasive factors. The presence of group-specific pervasive factors seriously reduces the ability of the APCA method in estimating the common factor structure. Obviously, the absence of group-specific pervasive factors (as in Panel B) makes our method the same as standard APCA and there is no difference in the S statistic for this case. Unlike APCA, our procedure is able to identify that there are no common pervasive factors, i.e., the intersection of the first common eigenvectors in each group is empty, when we fix $\beta_{.1} = \beta_{.2} = 0$ (i.e., no common pervasive factors in Panel C). Group-specific factors with a relatively high sensitivity have an especially deleterious impact on the performance of APCA—the S statistic is 0.345 and 0.999 for APCA and MGFA, respectively, in Panel E. It is also interesting to note that, even under the basic setting, different sample sizes reduce the ability of standard APCA to identify common pervasive factors (the S statistic for APCA is 0.951 and 0.918 in Panels A and F, respectively) while there is no deterioration in the performance of MGFA.

In the last two columns of Table 1, we compare the ability of the APCA and MGFA to estimate all K pervasive factors. We see in all panels that APCA and MGFA provide similarly accurate results in estimating the space spanned by the K pervasive factors: the average S statistic is systematically in excess of 0.998 and the standard errors are close to zero.

3.2.2. *Cross-correlation*

In a second series of simulations, we study the effect of cross-correlation within and across groups. To this end, we suppose that there exist ten *non-pervasive* factors in each group. More precisely, we divide each group into $n = 10$ subgroups (with ten percent of the data in each subgroup) and suppose that, in each subgroup, asset returns are also driven by an additional factor. This extra feature introduces the effect of a factor affecting the returns of all firms in a given industry, regardless of their group. With $L_{j\cdot}$, $j = 1, \ldots, n$ orthogonal and N(0, 1) variables, we modify the return-generating processes as follows:

$$
\begin{aligned}
R_{it} &= \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i3}F_{3t} + \sum_{j=1}^{n} \gamma_{ij}L_{jt} + \varepsilon_{it}, \forall i \in [1, N_1] \\
&\quad \gamma_{ij} \sim \mathrm{N}(1,1), \quad j = 1, \ldots, n, \\
&\quad \text{if } i \in [1 + (j-1)N_1/n, jN_1/n] \text{ and 0 otherwise} \\
R_{it} &= \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i4}F_{4t} + \sum_{j=1}^{n} \gamma_{ij}L_{jt} + \varepsilon_{it}, \forall i \in [N_1 + 1, N_1 + N_2] \\
&\quad \gamma_{ij} \sim \mathrm{N}(1,1), \quad j = 1, \ldots, n, \\
&\quad \text{if } i \in [N_1 + 1 + (j-1)N_2/n,\ N_1 + jN_2/n] \\
&\quad \text{and 0 otherwise} \\
R_{it} &= \beta_{i1}F_{1t} + \beta_{i2}F_{2t} + \beta_{i5}F_{5t} + \sum_{j=1}^{n} \gamma_{ij}L_{jt} + \varepsilon_{it}, \\
&\quad \forall i \in [N_1 + N_2 + 1, N_1 + N_2 + N_3] \\
&\quad \gamma_{ij} \sim \mathrm{N}(1,1), \quad j = 1, \ldots, n, \\
&\quad \text{if } i \in [N_1 + N_2 + 1 + (j-1)N_3/n, N_1 + N_2 + jN_3/n] \\
&\quad \text{and 0 otherwise}
\end{aligned} \tag{5}
$$

The results of this set of experiments are reported in Table 2 (with the same format as that of Table 1). As far as common pervasive factors are concerned, the quality of the estimated space is reduced slightly for both estimation methods—the S statistic is lower in Table 2 than the corresponding values in Table 1. However, the deterioration in performance is greater for APCA than it is for MGFA. Therefore, the span of the estimated factors is still better with MGFA than with APCA. Most notably, the standard errors with APCA are larger than those with our procedure. As in Table 1, we contrast in the last two columns of Table 2 the span of the estimated K pervasive factors given by the two competing procedures. With the exceptions of Panels B and C, the reduction in the value of the S statistic is more severe with APCA than with our

MGFA procedure. As a result, our novel method dominates the standard factor extraction method in panels A, D, E, F, G, and H in estimating the space spanned by all true pervasive factors. Moreover, if we combine panels F and H, the performance of asymptotic PCA drops (average $S = 0.7804$). At the same time, MGFA yields an average S statistic equal to 0.9849 with a standard error three times smaller.

Overall, we conclude that APCA falls short in estimating the space spanned by the common pervasive factors even in large samples. Furthermore, in the presence of cross-correlated residuals, MGFA outperforms the standard APCA for estimating the space spanned by *all* (common and group-specific) pervasive factors in finite samples. In unreported results, we also perform another set of simulations in which we introduce heteroskedasticity in the residuals. For these experiments, we utilize the HFA as recommended by Jones (2001) instead of standard APCA as well in the first step of MGFA. We find that while HFA offers an improvement over standard APCA (as in Jones), our method, which incorporates the group structure, is still able to outperform the HFA procedure.

3.3. Missing data

The first step of our MGFA is performing an APCA on each group. Specifically, we calculate the eigenvectors of the cross-product matrix $\hat{\Omega}_g = (1/N_g)R_g'R_g$ of returns in each group. This calculation requires a balanced sample of complete data history for each firm. However, unbalanced samples (where not all firms have complete return history) and/or missing data are ubiquitous features of real data. Does MGFA work under this setting? Since MGFA relies on APCA as a preliminary step, this question can be rephrased as whether APCA is robust under missing data. Connor and Korajczyk (1987) answer this question affirmatively. They show that the $\hat{\Omega}$ matrix approaches a transformation of the factors as the number of assets grows large, even when the sample contains firms with missing observations (replaced with zeros). The intuition for this result is that $\hat{\Omega}$ computed with the full sample is merely an average over a greater number of observations than the corresponding $\hat{\Omega}$ computed with the restricted sample. As long as the assets with missing observations are not "too" different from continuously traded assets, the approximation error in estimating the factors is, in fact, smaller with samples including firms with missing observations than those without.[4]

We check in simulations whether these claims are indeed justified. The return-generating process is assumed to be the same as in the basic setting of Section 3.2.1. In addition, we consider a case where half of the firms in all groups have missing data. The proportion of missing data is 25% of the total—in other words, for each firm with missing data, 15 months out of 60 months are set to have a missing value. The firms and months chosen to have missing values are, of course, selected at random in each simulation. The missing values are replaced by zero and MGFA and APCA are performed in each simulation. The results of this set of simulation experiments are reported in Table 3 (again with the same format as that of Table 1). Confirming the results of Connor and Korajczyk (1987), we find no noticeable deterioration in the performance of either APCA or MGFA, as the S statistics are quite close to each other in Tables 1 and 3. Therefore, echoing the earlier results, MGFA is again better at estimating the common pervasive factors than APCA.

Note that our only claim is that data samples with missing data will not lead to greater deterioration in performance with MGFA than with APCA. As mentioned earlier, the key requirement for the validity of this analysis is that the non-missing returns of the firms with missing data still follow the same data-generating process as for the rest of the firms. Since we ensure this by design in our simulation experiments, we cannot guarantee at this stage that MGFA procedure works equally well with real data samples with missing data. We will check the veracity of our claims for real data in Section 4.

3.4. Price of risk

Our simulations show that APCA falls short in estimating the space spanned by the common pervasive factors even in large samples. However, in some settings both APCA and MGFA have comparable performance in identifying the space of *all* pervasive factors. Even under these situations, it may be important to identify the exact nature of the split in factors across groups (and hence rely on MGFA). One example is the estimation of the price of risk of factors. The main implication of the APT is that expected returns on assets are approximatively linear in their sensitivities to the factors

$$E[r] = \iota\lambda_0 + B\lambda \tag{6}$$

where λ_0 is a constant, λ is a vector of factor risk premia, and ι is a vector of ones. One testable restriction of the model is that the implied risk premia are the same across subsets of assets (see Brown and Weinstein, 1983). That is, if we partition the return vector r into $r^1, r^2, \ldots, r^G$ with B^g representing the same partitioning of B, and investigate the subset pricing relations

$$E[r^g] = \iota\lambda_0^g + B^g\lambda^g, \quad g = 1, 2, \ldots, G \tag{7}$$

then $\lambda_0^g = \lambda_0$ and $\lambda^g = \lambda$ for all g. These tests can be viewed as tests of the law of one price—the price of risk is the same across subgroups, conditional on an estimated factor model.

There are a number of studies that use the APT to analyze asset returns across two or more countries since integration across national markets requires that common sources of risk be priced in a consistent manner across countries. Cho, Eun, and Senbet (1986) use inter-battery factor analysis (IBFA) to estimate the factor sensitivity matrix for factors common across pairs of countries and then test for consistent pricing across countries in a manner similar to that of Brown and Weinstein (1983). However, since IBFA picks out only common factors, the values of λ_0^g could differ across countries since they can incorporate the risk premia for factors specific to that

[4] Jones (2001) shows that HFA is also robust under missing data.

country which are still not globally diversifiable. Connor and Korajczyk (1995, footnote 25) point out that "Financial market integration does not imply or require that the countries be engaged in producing the same goods. Therefore, financially integrated countries might still have assets that are subject to country-specific productivity shocks. A country specific, but priced, factor could occur if the country in question is not small relative to the world economy."

In order to illustrate the possible errors in estimating λ's when the APCA or IBFA are used, we perform a simple simulation exercise. Our basic setting consists of two groups of data, $G=2$, in which the returns r_{it} are driven by two pervasive factors, $K=2$. There is one common pervasive factor ($K_C=1$) and one group-specific factor. The first group has two pervasive factors while the second one has only one common pervasive factor. The exact return-generating process is assumed to be

$$\begin{aligned} r_{it} &= E[r_{it}] + \beta_{i1} f_{1t} + \beta_{i2} f_{2t} + \varepsilon_{it}, \quad \forall i \in [1, N_1] \\ r_{it} &= E[r_{it}] + \beta_{i1} f_{1t} + \varepsilon_{it}, \quad \forall i \in [N_1+1, N_1+N_2] \end{aligned} \tag{8}$$

where the realization of the factors f_{jt}, $j=1,\ldots,2$, are orthogonal and $N(0,1)$ variables. The factor loadings β_{ij}, $j=1,\ldots,2$, are $N(1,1)$ variates and ε_{it} is $N(0,1)$. We use $T=60$, and in each case we perform 1,000 simulations. The true prices of risk are $\lambda_0=5\%$, $\lambda_1=20\%$, and $\lambda_2=10\%$. We perform the analysis in three steps. In the first step, we extract the factors. In the second step, we estimate betas by running a time-series regression for each stock. The third step involves running the following cross-sectional regression model

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t = \iota\lambda_0 + \hat{B}\lambda + \nu \tag{9}$$

where ν is a vector of pricing errors. When we extract the factors using APCA on the pooled data, the third-stage cross-sectional regression uses loadings on two factors for both groups (since APCA does not identify common pervasive factors). When we extract the factors using MGFA, the third-stage cross-sectional regression uses loadings on two factors for the first group but loadings on only the first factor for the second group. Finally, to replicate the spirit of IBFA, we also perform a third experiment in which we run the third-stage cross-sectional regression using the *true* betas on the first factor only for the two groups.

Table 4 reports the results of this simulation exercise. Panel A is for the same number of firms in both groups

Table 4
Simulation results—prices of risk

We consider two groups of data ($G=2$), in which the data-generating process for returns, r_{it}, is assumed to be

$$\begin{aligned} r_{it} &= E[r_{it}] + \beta_{i1} f_{1t} + \beta_{i2} f_{2t} + \varepsilon_{it}, \quad \forall i \in [1, N_1] \\ r_{it} &= E[r_{it}] + \beta_{i1} f_{1t} + \varepsilon_{it}, \quad \forall i \in [N_1+1, N_1+N_2] \end{aligned}$$

where the realization of the factors $f_{j\cdot}$, $j=1,\ldots,2$ are orthogonal and $N(0,1)$ variables. The factor loadings $\beta_{\cdot j}$, $j=1,\ldots,2$ are $N(1,1)$ variates and $\varepsilon_{i\cdot}$ is $N(0,1)$. Expected returns are given by

$$E[r] = \iota\lambda_0 + B\lambda$$

where $\lambda_0=5\%$, $\lambda_1=20\%$, and $\lambda_2=10\%$. We use $T=60$. We perform the analysis in three steps. In the first step, the factors are extracted. In the second step, we estimate betas by running a time-series regression for each stock. The third step involves running the following cross-sectional regression model:

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t = \iota\lambda_0 + \hat{B}\lambda + \nu$$

where ν is a vector of pricing errors. We report the averages of these prices of risks (with standard deviation across simulations in parenthesis). Panel A has the same number of firms in both groups ($N_1=N_2=1,000$) while Panel B has fewer firms in the second group ($N_2=100$).

Method	Group	$\widehat{\lambda_0}$	$\widehat{\lambda_1}$	$\widehat{\lambda_2}$
Panel A: $N_1=1,000$, $N_2=1,000$				
APCA	Group 1	0.0549 (0.0013)	0.2147 (0.0031)	0.0148 (0.0061)
	Group 2	0.0533 (0.0013)	0.1779 (0.0034)	0.0769 (0.0075)
MGFA	Group 1	0.0550 (0.0014)	0.1933 (0.0021)	0.0933 (0.0011)
	Group 2	0.0534 (0.0011)	0.1935 (0.0019)	–
Common true factor	Group 1	0.1528 (0.0039)	0.1965 (0.0029)	–
	Group 2	0.0533 (0.0013)	0.1967 (0.0009)	–
Panel B: $N_1=1,000$, $N_2=100$				
APCA	Group 1	0.0550 (0.0013)	0.2075 (0.0022)	0.0553 (0.0073)
	Group 2	0.0538 (0.0039)	0.1447 (0.0140)	0.1281 (0.0169)
MGFA	Group 1	0.0550 (0.0017)	0.1939 (0.0040)	0.0929 (0.0038)
	Group 2	0.0533 (0.0038)	0.1934 (0.0038)	–
Common true factor	Group 1	0.1532 (0.0047)	0.1971 (0.0032)	–
	Group 2	0.0537 (0.0039)	0.1964 (0.0026)	–

($N_1 = N_2 = 1,000$) while Panel B has fewer firms in group 2 ($N_2 = 100$). In both panels, we see that the estimates of $\widehat{\lambda}_0$ and $\widehat{\lambda}_1$ with MGFA are virtually equal in both groups. However, in the other two experiments, the estimates of the prices of risk are different across groups. Consequently, one sometimes rejects the hypothesis that the prices of risk are the same across groups even when the hypothesis is true in the population. For example, even though APCA often roughly estimates the span of the two pervasive factors correctly (the unreported S statistic is close to one), it fails because the estimates of $\widehat{\lambda}_1$ are quite different from each other across the two groups. The IBFA experiment estimates the $\widehat{\lambda}_1$ correctly but the estimates of the price of zero-beta risk, $\widehat{\lambda}_0$, vary dramatically between groups.

We conclude that, at least in the setting explored here, it is not enough to identify the space of all pervasive factors. The split between common and group-specific factors is also important and, therefore, one would prefer MGFA.

4. NYSE and Nasdaq

While the NYSE and Nasdaq provide the same service, their underlying structures, rules, and governing principles are very different. For instance, the NYSE is a specialist-based auction system and Nasdaq is a computer-based dealer market. There are many reasons why firms choose to list on one exchange versus the other. For instance, it is commonly assumed that firms in high-technology industries would want to list on Nasdaq. However, industry might not be the only criterion. Baruch and Saar (2006) show that the liquidity of a stock increases when it lists on an exchange where similar securities trade, which leads them to argue that listing decisions by managers reflect conscious choices to increase the liquidity of their firms' stocks. There are also differences in listing requirements. For example, the listing requirements at the NYSE in the early years could be met only by very large firms, while the Nasdaq market tried to attract small startups.

It therefore seems plausible that the underlying return structure of stocks listed on the NYSE can be very different from that of stocks listed on Nasdaq. However, there has been very little work done on pinpointing the exact nature of the difference(s) between the factors across these two groups.[5] We use our MGFA methodology to overcome the problem of identifying common pervasive factors and group-specific pervasive factors.

We analyze the congruence between factor structures for NYSE and Nasdaq stocks in several steps. The first obvious step is to extract factors in each group. As discussed previously, this step is accomplished by running an APCA (or HFA) separately in each group. The validity of this step, even in the presence of group structure, arises from the fact that the space spanned by the latent factors can be consistently estimated in each group and the reduced subspace retains most of the variability within all groups.

Armed with estimates of factors for each group, we proceed to compare the two groups of factors. The comparison done via canonical correlation analysis (CCA) provides a visual inspection of the potential number of common factors across the two groups. The next step is to formally estimate the number of common factors and then extract them. We accomplish this task by using MGFA and, for comparison, also estimate common factors via APCA (or HFA) on pooled data. A horserace (again via CCA) is then run between these different sets of estimated common factors to compare their ability to explain the common variation in returns in each of the two groups. The final steps in our analysis deal with the comparison of statistical factors estimated using MGFA and the set of standard risk factors (motivated by theory or empirical characteristics).

4.1. Data

We use excess returns on stocks traded on the NYSE, Amex, and Nasdaq as recorded in the Center for Research in Security Prices (CRSP) monthly stock file. We calculate excess returns by subtracting the interest rate on the one-month Treasury bill from the individual stock returns. We consider a 25-year sample covering the 1978–2002 period, which we divide into five 60-month subperiods. In each subperiod, we systematically exclude stocks (1) with missing returns data, (2) that are not traded either on the NYSE, Amex, or Nasdaq, (3) that change exchange during the subperiod, and (4) that are not common stocks (we exclude ADRs, REITs, closed-end funds, and other securities that do not have a CRSP share type code of 10 or 11). This filtering procedure yields 2,942 firms for 1978–1982, 3,068 for 1983–1987, 3,757 for 1988–1992, 3,928 for 1993–1997, and 4,023 for 1998–2002. We partition our original universe of stocks into two groups, the first containing stocks traded on the NYSE and Amex ("NYSE group") and the second containing stocks traded on Nasdaq ("Nasdaq group").

Table 5 presents the descriptive statistics on the excess returns. For each subperiod and each group, we compute these statistics from the time series of each stock and then report the cross-sectional average. The statistics can therefore be interpreted as those for a representative stock. Panels A and B give these statistics for the NYSE group and the Nasdaq group, respectively. The number of stocks in each group is not quite the same, with the NYSE group having more stocks in the early subperiods and the Nasdaq group having more stocks in the later subperiods. The volatility of stocks on Nasdaq is higher than that of stocks on the NYSE. The difference in volatility across the two groups has increased over the years, echoing the results in Malkiel and Xu (2003). It is also apparent from the table that Nasdaq stocks have higher skewness and kurtosis than their counterparts on the NYSE. While this evidence, per se, does not imply that the factor structure of returns is different across groups, it is suggestive of differences in NYSE returns and Nasdaq returns.

[5] Baruch and Saar (2006) show differences in loadings on principal components of returns across the two exchanges. They, however, estimate principal components separately for the different groups and, as such, do not undertake an analysis of extracting factors that are common across the two groups.

Table 5
Descriptive statistics on excess returns

We use excess monthly returns on stocks traded on the NYSE, Amex, and Nasdaq calculated by subtracting the interest rate on the one-month Treasury bill from the individual stock returns. We exclude stocks (1) with missing returns data, (2) that are not traded either on the NYSE, Amex, and Nasdaq, (3) that change exchange during the subperiod, and (4) that are not common stocks. We partition our original universe of stocks into two groups, the first containing stocks traded on the NYSE and Amex, and the second with stocks traded on Nasdaq. The sample period is 1978–2002, which we divide into five 60-month subperiods. For each group and each subperiod, we report the descriptive statistics on the excess returns. Each of these statistics is calculated using the time series of individual stocks and then cross-sectionally averaged across stocks. The statistics can, therefore, be interpreted as those for a representative stock.

	1978–1982	1983–1987	1988–1992	1993–1997	1998–2002
Panel A: NYSE-Amex stocks					
Mean	0.0110	0.0056	0.0083	0.0101	0.0030
Std. Dev.	0.1110	0.1084	0.1108	0.0892	0.1265
Skewness	0.4268	0.0486	0.5258	0.3469	0.3624
Kurtosis	4.1554	4.7175	4.8058	4.2624	4.6601
N	1,690	1,500	1,547	1,763	1,760
Panel B: Nasdaq stocks					
Mean	0.0134	0.0047	0.0129	0.0137	0.0119
Std. Dev.	0.1319	0.1469	0.1583	0.1543	0.2082
Skewness	0.7026	0.6557	0.9808	0.8544	0.9271
Kurtosis	5.2443	5.8006	6.4174	5.5704	6.0987
N	1,252	1,568	2,210	2,165	2,263

Table 6
Number of pervasive factors in NYSE/Amex and Nasdaq

We run an asymptotic principal component analysis (APCA) on stock returns in each of two groups of stocks: group G1 consisting of stocks on NYSE/Amex and group G2 consisting of stocks on Nasdaq. This table presents the optimal number of factors obtained in each of these groups. The optimal number of common factors in each group is the one that minimizes the Bai and Ng (2002) information criteria, $IC_{pi}(k)$ and $PC_{pi}(k)$ (refer to Appendix A for equations). We calculate each information criterion for $k = 1, \ldots, 8$. The sample consists of all common stocks that have no missing data and the sample period 1978–2002 is divided into five subperiods of 60 months each.

	1978–1982		1983–1987		1988–1992		1993–1997		1998–2002	
	G1	G2	G1	G2	G1	G2	G1	G2	G1	G2
IC_{p1}	3	1	2	1	2	1	2	1	3	2
IC_{p2}	3	1	2	1	2	1	2	1	3	2
IC_{p3}	3	1	2	1	2	1	2	1	3	2
PC_{p1}	2	1	2	1	2	1	2	1	3	2
PC_{p2}	2	1	2	1	2	1	2	1	3	2
PC_{p3}	2	1	2	1	2	1	2	1	3	2

4.2. How similar are the pervasive factors across the two groups?

We first run a separate APCA on each of these two groups and obtain estimates of factors *within* the groups.[6] The number of factors is estimated using the information criteria (PC_p and IC_p) proposed by Bai and Ng (2002). We report the values of these statistics in Table 6. We see that, depending on the sample period, the PC_p and IC_p select between one and three factors in both the NYSE and Nasdaq groups. The variation in number of pervasive factors over time is puzzling from an economic standpoint. If some factors are pervasive in one sample period, they should be pervasive in *all* sample periods. We therefore rely more on economic intuition than on purely statistical criteria, and use three as the number of pervasive factors in each group in our empirical analysis (in other words, we set $K_1 = K_2 = 3$). Note that this is a conservative assumption in that setting the number of factors to be too low could potentially lead to our missing some important factors. This number also coincides with the number of factors used in the multifactor model of Fama and French (1993). Since the total number of pervasive factors in the economy is bounded by the sum of group-specific pervasive factors ($K \leqslant \sum_{g=1}^{G} K_g$), we also conclude that there are between three (i.e., all factors are common across the NYSE and Nasdaq) and six (i.e., no factors are common across the NYSE and Nasdaq) pervasive factors for both exchanges for all sample periods.

We proceed by comparing the estimated factors in each exchange group. We want to test whether the latent

[6] Note that based on our terminology, these are not group-specific factors. Rather, we refer to them as factors in a group (that include common pervasive *and* group-specific pervasive factors).

Table 7
Comparison of factors in NYSE/Amex and Nasdaq

We estimate pervasive factors in each group using APCA in Panel A and HFA in Panel B (the number of pervasive factors is assumed to be three in both groups). We present the first three canonical correlations between all the estimated pervasive factors in NYSE/Amex and Nasdaq. The numbers in parentheses underneath the canonical correlation are the 95% confidence intervals for the canonical correlation (refer to Appendix B for equations). We also present the first three eigenvalues of the matrix *P* which is the sum of the eigenprojection matrices for each group. The sample consists of all common stocks that have no missing data and the sample period 1978–2002 is divided into five subperiods of 60 months each.

		1978–1982	1983–1987	1988–1992	1993–1997	1998–2002
Panel A: APCA in each group						
Canonical correlations between factors	First	0.955	0.953	0.962	0.786	0.901
		(0.933, 0.977)	(0.930, 0.976)	(0.944, 0.981)	(0.691, 0.882)	(0.853, 0.948)
	Second	0.707	0.655	0.814	0.418	0.804
		(0.582, 0.832)	(0.514, 0.796)	(0.729, 0.899)	(0.228, 0.609)	(0.715, 0.893)
	Third	0.506	0.331	0.196	0.144	0.165
		(0.328, 0.684)	(0.136, 0.526)	(0.016, 0.377)	(0.000, 0.308)	(0.000, 0.336)
Eigenvalues of *P*	First	1.985	1.963	1.971	1.926	1.962
	Second	1.854	1.654	1.914	1.434	1.693
	Third	1.101	1.162	1.229	1.135	1.359
Panel B: HFA in each group						
Canonical correlations between factors	First	0.962	0.959	0.954	0.829	0.879
		(0.943, 0.981)	(0.939, 0.979)	(0.932, 0.977)	(0.750, 0.908)	(0.822, 0.937)
	Second	0.720	0.689	0.369	0.451	0.793
		(0.600, 0.840)	(0.559, 0.820)	(0.175, 0.563)	(0.264, 0.637)	(0.700, 0.886)
	Third	0.490	0.314	0.022	0.105	0.240
		(0.309, 0.671)	(0.119, 0.508)	(0.000, 0.095)	(0.000, 0.251)	(0.051, 0.428)
Eigenvalues of *P*	First	1.986	1.975	1.972	1.949	1.969
	Second	1.824	1.825	1.682	1.466	1.918
	Third	1.557	1.517	1.181	1.169	1.481

pervasive factors in the NYSE group correspond to the latent factors in the Nasdaq group. In other words, we want to gauge the general coherency between F^{NYSE} and F^{Nasdaq}. Since we are only able to estimate the subspace spanned by the true factors in each group, following common practice we compare the two sets of factors using canonical correlations.[7] We follow the methodology of Bai and Ng (2006), who provide statistics that indicate the extent to which the two sets of factors differ. This procedure compares the (individual or set of) observed variables with estimates of the unobserved factors. The point of departure with standard theory is that they work with large dimensional panels. While all nonzero population canonical correlations should be unity if the two sets of factors span the same space, the confidence interval for the smallest estimated nonzero canonical correlation provides a bound for the weakest correlation between the two sets of factors. More details on the CCA are provided in Appendix B. Here we just note that canonical correlations are a generalization of the univariate correlations (or regression R^2s as used by Jones, 2001).

We calculate the first three canonical correlations and present these correlations in Table 7. Panel A reports the results for standard APCA while Panel B reports the results when we use HFA on each group. Recall that high values of canonical correlations are indicative of the presence of common factors across the two groups. Only the first canonical correlation is typically in excess of 0.9 in both panels. This suggests only one well-defined relation between the two sets of factors (in other words, only one common factor). The second correlation is around 0.8 suggesting a weak remaining relation between the two sets of factors. The canonical correlations in the fourth period are intriguing—the first correlation is close to only 0.8 while the second correlation is close to 0.5. This suggests, based purely on statistical criteria, that there is possibly no common factor between the two sets of factors in this sample period. The third canonical correlation is very low and practically zero in the last three sample periods. This strongly suggests that all three pervasive factors in Nasdaq stocks cannot span the same factor space in NYSE stocks.

The results in Panel B of Table 7 from HFA are similar to those from APCA with the exception of the nonzero canonical correlations in periods three and four. In the third period, the value of the second nonzero canonical correlation is half as large with HFA (0.369) as with APCA (0.814). Furthermore, the values of the first nonzero canonical correlations in the fourth period differ more between the two methods than those during the first part of our sample period. As Jones (2001) suggests, the latter half of our sample is characterized by rising volatility of Nasdaq stocks and the technology boom of the 1990s. It is therefore not surprising that there is a larger difference between the two techniques in the latter half of the sample than in the first half of the sample. We use the results from HFA later in the paper as HFA better accounts

[7] Obviously, if $K_1 \leqslant K_2$, then K_1 latent factors in the NYSE group cannot span the space of the K_2 latent factors in Nasdaq. This, however, is ruled out by our assumption of $K_1 = K_2 = 3$.

for the heteroskedasticity prevalent in the sample than APCA. Thus, we conclude from Panel B that there could be either one or two well-defined relations in the factors governing the returns structure of NYSE and Nasdaq stocks.

4.3. Common pervasive factors

We estimate the number of common pervasive factors between the two groups following the methodology in Section 3.1. We report the first three eigenvalues of P also in Table 7 for each subsample. Recall that an eigenvalue of P equal to two (in the population) implies the existence of a common pervasive factor. Focusing on Panel B (HFA), this criterion indicates two common pervasive factors in the first, second, and fifth subsamples, and weaker evidence of two common pervasive factors in the third and fourth subsamples. If we relax the criteria somewhat and consider any eigenvalue greater than 1.5 to be suggestive of a common factor, we can conclude that there are three common factors in subsamples one and two, two common factors in subsamples three and five, and one common factor in subsample four. There is, of course, strong evidence of at least one common pervasive factor in all subsamples. This is not surprising because a common market factor would naturally arise in equilibrium versions of APT (Connor, 1984) and, frequently, the first estimated factor in traditional PCA is identified closely with the (equal-weighted) market factor.

We note that our conclusions regarding the number of factors are the same based on the eigenvalues of P or on the canonical correlations between factors, as discussed in the preceding section. However, as was the issue with the number of all pervasive factors, the variation in the number of common pervasive factors over time indicated by statistical criteria is troubling from an economic standpoint. We therefore rely again more on economic intuition than on purely statistical criteria and use two as the number of common pervasive factors in each group in the rest of our empirical analysis.

The stage is now set for us to estimate common pervasive factors. In addition to our MGFA methodology from Section 3, we also use APCA and HFA techniques in estimating common factors. The latter two methods are used for comparison purposes and are, of course, performed on pooled data. We assume that the number of common factors is two for all three techniques. The natural question at this stage is whether the factors estimated with the MGFA method are in fact the underlying unobserved true common factors in each group. We answer this question by comparing the span of common factors to that of all pervasive factors in each group. It is useful to recall that all pervasive factors for a group capture, by definition, the greatest degree of common variation in stock returns within a group. The idea behind our test (of comparing these factors to common pervasive factors) is that if the estimated common factors are indeed common, then they should have high canonical correlations with the factors for each group.

As in Jones (2001), we assume that the true pervasive factors in each group are well estimated by HFA. Following Bai and Ng (2006), we construct confidence intervals for

Table 8

Comparison of the common pervasive factors and all pervasive factors for NYSE/Amex and Nasdaq

This table presents the smallest nonzero (i.e., second) canonical correlation between common pervasive factors and all pervasive factors in each group (NYSE/Amex and Nasdaq). Pervasive factors for a group are calculated using HFA. The common pervasive factors are estimated using MGFA (Panel A), APCA (Panel B), and HFA (Panel C). Data for APCA and HFA are pooled across both groups to calculate common pervasive factors. The number of all pervasive factors is three in each group and the number of common pervasive factors is two. Numbers in parentheses underneath the canonical correlation are the 95% confidence intervals for the canonical correlation. The sample consists of all common stocks that have no missing data and the sample period 1978–2002 is divided into five subperiods of 60 months each.

	1978–1982	1983–1987	1988–1992	1993–1997	1998–2002
Panel A: MGFA					
NYSE	0.924	0.915	0.804	0.836	0.945
	(0.888, 0.961)	(0.873, 0.956)	(0.715, 0.893)	(0.760, 0.912)	(0.918, 0.972)
Nasdaq	0.924	0.915	0.804	0.836	0.945
	(0.887, 0.961)	(0.873, 0.956)	(0.715, 0.893)	(0.760, 0.912)	(0.918, 0.972)
Panel B: APCA					
NYSE	0.965	0.935	0.642	0.844	0.834
	(0.947, 0.982)	(0.904, 0.967)	(0.497, 0.78)	(0.771, 0.916)	(0.757, 0.911)
Nasdaq	0.831	0.815	0.883	0.649	0.845
	(0.753, 0.909)	(0.730, 0.899)	(0.827, 0.938)	(0.506, 0.792)	(0.772, 0.917)
Panel C: HFA					
NYSE	0.973	0.976	0.758	0.848	0.842
	(0.960, 0.987)	(0.964, 0.988)	(0.651, 0.864)	(0.777, 0.919)	(0.768, 0.915)
Nasdaq	0.827	0.772	0.785	0.609	0.907
	(0.748, 0.907)	(0.671, 0.874)	(0.688, 0.881)	(0.454, 0.763)	(0.862, 0.952)

the smallest nonzero (i.e., second) canonical correlations between these sets of factors. The results are in Table 8. Panel A shows that the canonical correlation between common pervasive factors estimated using MGFA and group factors has the same value for both the NYSE and Nasdaq. Moreover, this value is very close to one, indicating that our two common factors estimated using MGFA are able to capture the source of common variation across both groups of stocks. Comparison of Panels A and B shows that, in some subsamples and for some groups, the correlation of group pervasive factors is higher with common factors extracted using APCA than with factors extracted using MGFA. However, it is never the case that APCA common pervasive factors exhibit higher correlations with *both* groups' pervasive factors. For example, in the first subsample (1978–1982), the canonical correlation of MGFA (APCA) factors with the NYSE group is 0.924 (0.965), while that with the Nasdaq group is 0.924 (0.831). With MGFA, it is possible to simultaneously estimate a reduced subspace of dimension K for all groups while retaining most of the variability within all the groups. APCA, on the other hand, pools the data together and in some subsamples favors one group over the other. The APCA factors can thus have higher correlation with one of the groups' factors, but they are unable to simultaneously account for the commonality. The common factors estimated with APCA tend to favor NYSE over Nasdaq in some periods and vice versa in other periods. We find in Panel C that HFA gives markedly better results than APCA, in a confirmation of Jones (2001). Moreover, controlling for heteroskesdasticity removes the bias inherent in APCA of favoring high-volatility stocks/groups. More pertinent to our study is the observation that canonical correlations using HFA are not higher for both the groups than those using MGFA.

It is also interesting to calculate the canonical correlations between the set of common factors extracted from MGFA, APCA, and HFA. This comparison provides yet another perspective on the degree of commonality between the "common" factors extracted using various techniques. In unreported results, we find that the canonical correlations between the first common factor across the three techniques are fairly high (usually 0.98 or 0.99) for all subsamples. The second canonical correlation between MGFA and HFA factors is higher than that between MGFA and APCA. However, the most striking feature is that the canonical correlation between MGFA and APCA is very low in the last two subperiods (0.29 and 0.51, respectively, for 1993–1997 and 1998–2002). This shows that the common factors across the three techniques are not that common.

We conclude that common risk factors are best estimated using MGFA and represent the highest degree of variation in returns that is common across the two groups. In particular, our methodology yields the two common pervasive factors that explain the returns structure of both NYSE and Nasdaq stocks. In contrast, the traditional techniques of APCA and HFA applied to the pooled data are unable to uncover the covariation that is *common* across both exchanges.

4.3.1. *A robustness exercise*

Our analysis suggests that, even though there are three pervasive factors in each group, there are at most two common pervasive factors. One critique of our methodology can be a sampling bias problem. More specifically, it is possible that there are some outliers in the data that could potentially lead to us not being able to identify closer correspondence between NYSE and Nasdaq stocks. We address this issue in the following way. We randomly divide the stocks in each group into two non-overlapping, equal-sized groups and compare canonical correlations between these two subgroups. We estimate pervasive factors in each subgroup using HFA (the number of pervasive factors is assumed to be three in each subgroup). We repeat the estimation 1,000 times and present the average value of the three canonical correlations, with their standard deviations in parentheses, in Table 9. Panel A (B) presents the results for NYSE (Nasdaq) stocks. We find that the first two canonical correlations are well in excess of 0.9, and the third is around 0.8, for both the NYSE and Nasdaq across all time periods. This lends support to our earlier conclusion about the number of factors in each group being three. More important, it shows that the presence of a few outliers (say in the more volatile Nasdaq stocks) is not driving our earlier results. We also randomly combine 50% of the NYSE/Amex stocks and 50% of the Nasdaq stocks into two subgroups. Canonical correlations between the factors estimated from each of these two "random" groups provide further evidence of the degree of closeness between the factors in each group. Panel C shows that the average third canonical correlation between these two groups is around 0.5 (the first is in excess of 0.95 and the second one is around 0.8), very similar to the results in Table 7. The consistency of these randomization results with our earlier results shows that sampling bias does not plague our conclusions.

4.4. *Common pervasive factors and standard benchmark factors*

The next set of experiments is designed to explore the relation between our common pervasive statistical factors and the standard benchmark factors that are often used in asset pricing. The set of standard factors that we include are the three Fama and French (1993) factors and a momentum factor. While the well-known rotational indeterminacy obviously rules out exact correspondence between statistical factors and real economic variables/factors, our attempts are designed only to establish correlation (or lack thereof) between the two time series. We present canonical correlations between these two sets of factors in Table 10.

There are many interesting patterns revealed by these canonical correlations. For instance, the market factor typically has the highest correlation with the common pervasive factors. This is not too surprising since, as mentioned before, a market factor is commonly identified as the first principal component. What is more surprising is the fact that the importance of the market factor in

Table 9
A randomization experiment

In Panel A, NYSE/Amex stocks are randomly divided into two nonoverlapping, equal-sized groups and the first three canonical correlations are estimated between the estimated subgroup-specific pervasive factors. We repeat the analysis with Nasdaq stocks in Panel B. In Panel C, we randomly combine 50% of the NYSE/Amex stocks with 50% of the Nasdaq stocks into two subgroups and the first three canonical correlations are estimated between the estimated subgroup-specific pervasive factors. We estimate pervasive factors in each subgroup using HFA (the number of pervasive factors is assumed to be three in each subgroup). We repeat the estimation 1,000 times and present the average value of the three canonical correlations with their standard deviations in parentheses. The sample period 1978–2002 is divided into five subperiods of 60 months each.

	1978–1982	1983–1987	1988–1992	1993–1997	1998–2002
Panel A: NYSE/Amex					
First canonical correlation	0.9978	0.9975	0.9957	0.9929	0.9952
	(0.0004)	(0.0005)	(0.0009)	(0.0015)	(0.0011)
Second canonical correlation	0.9824	0.9804	0.9746	0.9783	0.9839
	(0.0044)	(0.0045)	(0.0057)	(0.0091)	(0.0038)
Third canonical correlation	0.9410	0.8716	0.7927	0.8424	0.9667
	(0.0293)	(0.0451)	(0.1095)	(0.1183)	(0.0101)
Panel B: Nasdaq					
First canonical correlation	0.9953	0.9954	0.9936	0.9883	0.9957
	(0.0010)	(0.0009)	(0.0013)	(0.0023)	(0.0009)
Second canonical correlation	0.9364	0.9136	0.9382	0.9380	0.9823
	(0.0136)	(0.0178)	(0.0129)	(0.0124)	(0.0038)
Third canonical correlation	0.7036	0.8357	0.4618	0.7784	0.9295
	(0.1418)	(0.0337)	(0.2100)	(0.0731)	(0.0329)
Panel C: NYSE/Amex and Nasdaq					
First canonical correlation	0.9783	0.9763	0.9725	0.8982	0.9351
	(0.0030)	(0.0029)	(0.0038)	(0.0116)	(0.0077)
Second canonical correlation	0.8312	0.8083	0.5853	0.6598	0.8773
	(0.0250)	(0.0264)	(0.0677)	(0.0356)	(0.0038)
Third canonical correlation	0.6184	0.5196	0.1550	0.2884	0.4827
	(0.1142)	(0.0686)	(0.0964)	(0.0819)	(0.0333)

explaining the returns has declined over time. The smallest nonzero correlation used to be 0.91 in 1978–1982 but is only 0.71 in the last subsample, 1998–2002. The other two Fama-French factors (SMB and HML) have lower correlations than the market factor. The importance of the SMB factor seems to have declined while the importance of the HML factor has increased over the years (although the patterns are not monotonic). The momentum factor has the lowest correlation with the common pervasive factors—the confidence interval for the last two subsamples even has a lower bound of zero. The above facts imply that the market factor is closely related to our common factors. The relation between the other two Fama-French factors and the common factors is weak, and there is no discernible relation between the momentum factor and the common factors.

The last two rows of Table 10 present the correlations between the common pervasive factors and the set of standard factors taken together. Mirroring the earlier results, we again find that the correlations of the standard factors with the common pervasive factors have declined over time. The addition of the momentum factor only marginally improves the explanatory power of the standard factors. The positive news from these results is that the standard Fama-French factors are able to do a good job of explaining the common variation in returns across securities listed both on the NYSE and Nasdaq—the canonical correlation with the pervasive factors is well in excess of 0.95 for most of the sample periods. Also note that there is no dichotomy between our finding that there are *two* common pervasive factors and the *three*-factor model of Fama and French (1993). We both say that three factors are required to model stock returns in each group.

4.5. Missing data

Our analysis thus far is based on a sample of stocks with complete returns history over each subsample. This obviously induces a survivorship bias. In this section, therefore, we want to explore whether MGFA can effectively handle missing data. Our results from Section 3.3 already provide some comfort in this regard. However, if firms with missing data are very different from firms with full time series, then it is not clear whether both APCA and MGFA can effectively extract factors.

Table 10
Comparison of common pervasive factors and standard benchmark factors

This table presents the smallest nonzero canonical correlation between the common pervasive factors and the standard benchmark factors. The standard benchmark factors are the three Fama and French (1993) factors, FF3 = (MKT, SMB, HML), and a momentum factor, MOM. The common pervasive factors are estimated using MGFA for the two groups of stocks (NYSE/Amex and Nasdaq). The number of common pervasive factors is two. Numbers in parentheses underneath the canonical correlation are the 95% confidence intervals for the canonical correlation. The sample consists of all common stocks that have no missing data and the sample period 1978–2002 is divided into five subperiods of 60 months each.

	1978–1982	1983–1987	1988–1992	1993–1997	1998–2002
MKT	0.913 (0.870, 0.955)	0.911 (0.869, 0.954)	0.941 (0.912, 0.970)	0.787 (0.692, 0.883)	0.714 (0.592, 0.836)
SMB	0.504 (0.325, 0.682)	0.392 (0.199, 0.584)	0.700 (0.573, 0.827)	0.366 (0.172, 0.560)	0.384 (0.190, 0.577)
HML	0.377 (0.184, 0.571)	0.421 (0.231, 0.611)	0.210 (0.027, 0.393)	0.549 (0.380, 0.718)	0.624 (0.474, 0.774)
MOM	0.421 (0.231, 0.612)	0.196 (0.016, 0.376)	0.211 (0.028, 0.395)	0.097 (0.000, 0.240)	0.137 (0.000, 0.299)
FF3	0.991 (0.987, 0.996)	0.987 (0.981, 0.994)	0.981 (0.971, 0.991)	0.955 (0.933, 0.978)	0.897 (0.848, 0.946)
FF3+MOM	0.992 (0.988, 0.996)	0.990 (0.985, 0.996)	0.985 (0.977, 0.993)	0.956 (0.934, 0.978)	0.936 (0.908, 0.969)

As before, we extract data from the CRSP monthly stock file for the sample period 1978–2002. We no longer impose the requirement of continuous trading and replace the missing observations with zeros. This leads to a significant increase in the size of the cross-section, with the number of stocks in some subsamples more than three times than in the restricted sample.[8] There are two additional considerations in this experiment. First, Connor and Korajczyk (1987) point out that the only requirement of the data to satisfy consistency requirements for APCA is that the non-missing returns follow the same process as the rest of the data. When missing observations are caused by suspension of trading, this assumption is violated. The reason is that CRSP includes the returns for the missing periods in the first non-missing period return (for example, the calculated return after a month of non-trading is a two-month return). We follow the lead of Connor and Korajczyk in eliminating the first non-missing observation. Second, a significant amount of non-synchronous trading leads to distortions in estimation of true covariance-factor structures (see Shanken, 1987). We assume that the extent of this problem is not severe for monthly data and ignore this issue.

Using this new sample, we follow the same steps as before. In particular, we estimate the factors in each group using APCA and HFA. Table 11 reports the canonical correlations between these sets of factors as well as the first three eigenvalues of matrix P. The numbers in this table are roughly similar to those in Table 7. Therefore, our earlier conclusions about the number of common factors remain valid. In particular, there is strong evidence of two common pervasive factors in all subsamples. The only noteworthy feature seems to be that the canonical correlations between the sets of factors extracted using APCA differ from those between the sets of factors extracted using HFA to a greater degree in Table 11 than in Table 7, in general.

We then proceed to extract the common pervasive factors using MGFA. In unreported results, we find that the canonical correlations of all pervasive factors are higher with the common pervasive factors extracted with MGFA than with factors extracted with APCA/HFA. We do find, similar to Table 8, that in some subsamples for some groups, the correlation of group pervasive factors is higher with common factors extracted using APCA/HFA than with factors extracted using MGFA. However, the correlations with both groups' pervasive factors are never higher for APCA/HFA common pervasive factors.

5. Conclusion

We propose a new procedure called MGFA to analyze pervasive factors in several groups. The key to the analysis is that the space spanned by the latent factors can be consistently estimated in each group of large dimension. Our approach is very general and can be viewed as a multigroup version of standard approaches like APCA (Connor and Korajczyk, 1986) and HFA (Jones, 2001). More precisely, our method estimates the subspace spanned by all pervasive factors and at the same time the subspace of common factors that are pervasive in all groups.

[8] The breakdown between NYSE and Nasdaq firms is as follows for the five subperiods: 2,384/3,689, 2,365/6,060, 2,303/5,748, 2,619/6,739, 2,789/6,163, respectively.

Table 11
Comparison of factors in NYSE/Amex and Nasdaq (including stocks with missing data)
We estimate pervasive factors in each group using APCA in Panel A and HFA in Panel B (the number of pervasive factors is assumed to be three in both groups). We present the first three canonical correlations between all the estimated pervasive factors in NYSE/Amex and Nasdaq. The numbers in parentheses underneath the canonical correlation are the 95% confidence intervals for the canonical correlation (refer to Appendix B for equations). We also present the first three eigenvalues of the matrix *P* which is the sum of the eigenprojection matrices for each group. The sample consists of all common stocks (including the stocks that have missing data) and the sample period 1978–2002 is divided into five subperiods of 60 months each.

		1978–1982	1983–1987	1988–1992	1993–1997	1998–2002
Panel A: APCA in each group						
Canonical correlations between factors	First	0.922	0.931	0.963	0.818	0.951
		(0.885, 0.960)	(0.897, 0.964)	(0.945, 0.982)	(0.734, 0.901)	(0.927, .975)
	Second	0.661	0.617	0.806	0.033	0.717
		(0.522, 0.801)	(0.465, 0.770)	(0.717, 0.894)	(0.000, 0.121)	(0.596, 0.838)
	Third	0.240	0.116	0.011	0.000	0.016
		(0.052, 0.429)	(0.000, 0.268)	(0.000, 0.065)	(0.000, 0.007)	(0.000, 0.080)
Eigenvalues of *P*	First	1.964	1.964	1.977	1.907	1.973
	Second	1.811	1.786	1.901	1.183	1.849
	Third	1.490	1.292	1.107	1.029	1.136
Panel B: HFA in each group						
Canonical correlations between factors	First	0.968	0.951	0.948	0.899	0.950
		(0.952, 0.984)	(0.927, 0.975)	(0.923, 0.974)	(0.851, 0.948)	(0.926, 0.975)
	Second	0.652	0.701	0.522	0.292	0.640
		(0.509, 0.794)	(0.574, 0.827)	(0.348, 0.697)	(0.099, 0.486)	(0.494, 0.786)
	Third	0.221	0.466	0.012	0.006	0.201
		(0.036, 0.406)	(0.281, 0.650)	(0.000, 0.066)	(0.000, 0.046)	(0.019, 0.382)
Eigenvalues of *P*	First	1.985	1.972	1.973	1.951	1.973
	Second	1.802	1.833	1.745	1.645	1.803
	Third	1.470	1.678	1.105	1.080	1.443

Application of our methodology to the return structure of NYSE and Nasdaq securities reveals that, while there are three factors in each group of stocks, there are at most two common pervasive factors. In other words, our empirical results uncover little evidence of close correspondence in the return structure of stocks on the two exchanges. At the same time, we emphasize that one should not interpret our results to imply absence of *any* integration between the two exchanges. To the contrary, we find that there are two common factors that drive returns of stocks on both the exchanges. Our results only point to the absence of complete similarity between the factors driving the returns on these exchanges. More importantly, we estimate these common and group-specific factors.

Future research could apply our methodology to identifying the sources of risk that drive returns in different countries or asset classes (e.g., equity vs. hedge funds). Another potential application would be to identify the common and group-specific driving forces underpinning macroeconomic variables and asset prices.

Appendix A. Number of factors

In earlier approaches to determining the number of factors, Trzcinka (1986) and Brown (1989) investigate the behavior of the eigenvalues, and mainly the number of eigenvalues whose values increase with the number of securities considered in the analysis. Alternative approaches are based on the incremental explanatory power of additional factors (see Connor and Korajczyk, 1993) or test whether the additional factors have some additional risk that is not already priced by the first factors (see Kandel and Stambaugh, 1989). Bai and Ng (2002) set up the determination of factors as a model-selection problem. In consequence, the proposed criteria depend on the usual tradeoff between good fit and parsimony. Since the problem is non-standard—not only because account needs to be taken of the sample size in both the cross-section and the time-series dimensions, but also because the factors are not observed—they develop a new theory that does not rely on sequential limit, nor does it impose any restrictions between N (number of assets) and T (length of the time series). The results hold under heteroskedasticity in both the time and the cross-section dimensions and also under weak serial dependence and cross-sectional dependence. Simulations show that the criteria have good finite-sample properties.

The criteria developed by Bai and Ng (2002) to determine the number of factors are PC_p and IC_p which are given by

$$PC_{p1}(k) = V(k, \widehat{F}^k) + k\widehat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right)$$
$$PC_{p2}(k) = V(k, \widehat{F}^k) + k\widehat{\sigma}^2\left(\frac{N+T}{NT}\right)\ln C_{NT}^2$$
$$PC_{p3}(k) = V(k, \widehat{F}^k) + k\widehat{\sigma}^2\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right) \quad \text{(A.1)}$$

and

$$IC_{p1}(k) = \ln V(k,\widehat{F}^k) + k\left(\frac{N+T}{NT}\right)\ln\left(\frac{NT}{N+T}\right)$$
$$IC_{p2}(k) = \ln V(k,\widehat{F}^k) + k\left(\frac{N+T}{NT}\right)\ln C_{NT}^2$$
$$IC_{p3}(k) = \ln V(k,\widehat{F}^k) + k\left(\frac{\ln C_{NT}^2}{C_{NT}^2}\right) \quad (A.2)$$

respectively. Here $V(k,\widehat{F}^k) = N^{-1}\sum_{i=1}^{N}\widehat{\sigma}_i^2$, $\widehat{\sigma}_i^2 = T^{-1}\widehat{e}_i'\widehat{e}_i$, $\widehat{\sigma}^2 = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{e}_{it}^2$, $C_{NT} = \min(\sqrt{N},\sqrt{T})$, and k is the number of estimated factors.

The *PC* statistics resemble the conventional panel data criteria and the *IC* statistics resemble the conventional time-series criteria. The important difference is that the penalty is a function of both *N* and *T*. There are three types of each criterion which are asymptotically equivalent but can have different finite-sample properties. Finally, as usual, the number of factors is to be chosen as k that maximizes $PC(k)$ or the $IC(k)$ criteria.

Appendix B. Canonical correlations

CCA seeks to identify and quantify the associations between two sets of variables. The goal is to find the maximal correlation between a chosen linear combination of the first set of variables and a chosen linear combination of the second set of variables. The pair of linear combinations are called the canonical variables and their correlations the canonical correlations. The canonical correlations thus measure the strength of the association between the two sets of variables. CCA is more general than several other methods. For instance, regression analysis is CCA with the first subset of variables consisting of a single numerical variable.

If F_t and G_t are two sets of factors with dimensions $(r\times 1)$ and $(m\times 1)$, respectively, then the canonical correlations, denoted by $\widehat{\rho}_k^2$, $k=1,\ldots,\min(m,r)$, are given by the largest eigenvalues of the $(r\times r)$ matrix

$$S_{FF}^{-1}S_{FG}S_{GG}^{-1}S_{GF} \quad (B.1)$$

where S is the sample covariance matrix. Bai and Ng (2006) show that having to estimate F and G has no effect on the sampling distribution of $\widehat{\rho}_k^2$. This allows them to construct $(1-\alpha)$ percent confidence intervals for the population canonical correlations as:

$$(\rho_k^{2-},\rho_k^{2+}) = \left(\widehat{\rho}_k^2 - 2\Phi_\alpha\frac{\widehat{\rho}_k(1-\widehat{\rho}_k^2)}{\sqrt{T}}, \widehat{\rho}_k^2 + 2\Phi_\alpha\frac{\widehat{\rho}_k(1-\widehat{\rho}_k^2)}{\sqrt{T}}\right) \quad (B.2)$$

where Φ is the cumulative density function for standard normal variables and T is the length of the time series.

References

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–171.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Bai, J., Ng, S., 2006. Evaluating latent and observed factors in macroeconomics and finance. Journal of Econometrics 131, 507–537.

Baruch, S., Saar, G., 2006. Asset returns and the listing choice of firms. Review of Financial Studies, forthcoming.

Bekaert, G., Hodrick, R.J., Zhang, X., 2005. International stock return comovements. Working Paper, Columbia University.

Boivin, J., Ng, S., 2006. Are more data always better for factor analysis? Journal of Econometrics 132, 169–194.

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

Brown, S.J., Weinstein, M.I., 1983. A new approach to asset pricing models: the bilinear paradigm. Journal of Finance 38, 711–743.

Campbell, J., Lettau, M., Malkiel, B.G., Xu, Y., 2001. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. Journal of Finance 56, 1–46.

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

Cho, D.C., Eun, C.S., Senbet, L.W., 1986. International arbitrage pricing theory: an empirical investigation. Journal of Finance 41, 313–329.

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

Connor, G., Korajczyk, R.A., 1986. Performance measurement with the arbitrage pricing theory: a new framework for analysis. Journal of Financial Economics 15, 373–394.

Connor, G., Korajczyk, R.A., 1987. Estimating pervasive economic factors with missing observations. Working Paper, Northwestern University.

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

Connor, G., Korajczyk, R.A., 1993. A test for the number of factors in an approximate factor model. Journal of Finance 48, 1263–1291.

Connor, G., Korajczyk, R.A., 1995. The arbitrage pricing theory and multifactor models of asset returns. In: Jarrow, R., et al. (Eds.), Handbooks in Operations Research and Management Science, vol. 9. North Holland, Amsterdam, pp. 87–144.

Dauxois, J., Pousse, A., 1975. Une extension de l'analyse canonique: quelques applications. Annales de l'Institut Henri Poincaré 11, 355–379.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E., French, K., 2004. New lists: fundamentals and survival rates. Journal of Financial Economics 73, 229–269.

Flury, B., 1984. Common principal components in k groups. Journal of the American Statistical Association 79, 892–898.

Flury, B., 1987. Two generalizations of the common principal component model. Biometrika 74, 59–69.

Jones, C.S., 2001. Extracting factors from heteroskedastic asset returns. Journal of Financial Economics 62, 293–325.

Kandel, S., Stambaugh, R.F., 1989. A mean-variance framework for tests of asset pricing models. Review of Financial Studies 2, 125–156.

Korajczyk, R.A., Sadka, R., 2007. Pricing the commonality across alternative measures of liquidity. Journal of Financial Economics, forthcoming.

Krzanowski, W.J., 1979. Between-groups comparison of principal components. Journal of the American Statistical Association 74, 703–707 (correction 76, 1022).

Malkiel, B.G., Xu, Y., 2003. Investigating the behavior of idiosyncratic volatility. Journal of Business 76, 613–644.

Naranjo, A., Protopapadakis, A., 1997. Financial market integration tests: an investigation using US equity markets. Journal of International Financial Markets, Institutions, and Money 7, 93–135.

Pérignon, C., Villa, C., 2006. Sources of time variation in the covariance matrix of interest rates. Journal of Business 79, 1535–1549.

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

Schott, J.R., 1999. Partial common principal component subspaces. Biometrika 86, 899–908.

Schwert, G.W., 2002. Stock volatility in the new millennium: how wacky is Nasdaq? Journal of Monetary Economics 49, 1–24.

Shanken, J., 1987. Nonsynchronous data and the covariance-factor structure of returns. Journal of Finance 42, 221–231.

Trzcinka, C., 1986. On the number of factors in the arbitrage pricing model. Journal of Finance 41, 347–368.