Extracting Factors with Maximum Explanatory Power*

Yexiao Xu

School of Management
The University of Texas at Dallas

All comments welcomed

Version: February 2007

Abstract

Security returns are heteroscedastic both cross-sectionally and over time, which affects the accuracy of standard factor extraction methods. In order to reduce the impact of such heterogeneity and to preserve the true factor structure, this paper studies the performance of a factor extracting method based on maximizing the explanatory power of the extracted factors. The implementation of the methodology is largely based on the principal components analysis on a correlation structure of asset returns. However, such a simple extension allows us to improve the finite sample performance over other popular approaches when returns are heteroscedastic both across individual assets and over time. Moreover, the out-of-sample study suggests that the extracted factors are not only stable across different sample groups, but also more pervasive in explaining the out-of-sample individual stock returns than other methods. These factors even have better out-of-sample explanatory power than the Fama and French factors. In addition, we shed light on the issue of choosing the correct number of factors.

*I am grateful to Stephen J. Brown, Lewis Chan, Nai-fu Chen, Ted Day, Yacine Hammami, Chris Jones, Raymond Kan, Bob Korajczyk, Craig MacKinlay, Jianping Mei, Larry Merville, John Wei, Steven Wei, Chu Zhang, Guofu Zhou, and seminar participants at SMU, Hong Kong University of Science and Technology, Hong Kong Polytech University, 2002 Western Finance Association Conference, and 2002 European Financial Management Association Conference for helpful comments. Address correspondence to: Yexiao Xu, School of Management, The University of Texas at Dallas, PO Box 688, Richardson, Texas 75080, USA; Email: yexiaoxu@utdallas.edu
Extracting Factors with Maximum Explanatory Power

Abstract

Security returns are heteroscedastic both cross-sectionally and over time, which affects the accuracy of standard factor extraction methods. In order to reduce the impact of such heterogeneity and to preserve the true factor structure, this paper studies the performance of a factor extracting method based on maximizing the explanatory power of the extracted factors. The implementation of the methodology is largely based on the principal components analysis on a correlation structure of asset returns. However, such a simple extension allows us to improve the finite sample performance over other popular approaches when returns are heteroscedastic both across individual assets and over time. Moreover, the out-of-sample study suggests that the extracted factors are not only stable across different sample groups, but also more pervasive in explaining the out-of-sample individual stock returns than other methods. These factors even have better out-of-sample explanatory power than the Fama and French factors. In addition, we shed light on the issue of choosing the correct number of factors.

The empirical failure of the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965) is amply documented and has motivated many researchers toward the use of either multifactor models in the spirit of Merton (1973) or the Arbitrage Pricing Theory (APT) of Ross’s (1976). Different from the CAPM model where asset returns are determined by a single endogenous market risk factor, the APT model uses a set of $K$ unspecified exogenous factors to price an asset. Since the APT theory does not provide any guidance with respect to the number and the nature of the underlying pricing factors, in application, researchers typically either rely on prespecified macroeconomic variables (Chen, Roll, and Ross, 1986) and fundamental variables (Fama and French, 1993), or used a “reverse engine” approach based on statistical methods, such as factor analysis or principal components analysis (for example, Roll and Ross, 1980; and Connor and Korajczyk, 1988, 1993).

If returns of individual assets are indeed determined linearly by $k$ factors, a “reverse engine” approach should be able to uncover the true underlying factors when there are enough assets. This rationale is theoretically sounded, and factors extracted using statistical methods usually offer tight in-sample fit to actual data. However, the out-of-sample performance seems to be questionable, which hinders the wide application of the APT model. In recent years, we have experienced a declining trend in the explanatory power of the fundamental factors, including the Fama and French factors, due to an increasing trend in the idiosyncratic volatility documented by Campbell, Lettau, Malkiel, and Xu (2001). For example, the average $R^2$ from fitting a market model to individual stocks’ monthly returns has decreased from 30% in the 1960s to 10% in late 1990s. Therefore, it is useful to revisit statistical methods if we can overcome the out-of-sample issue.

Two commonly used statistical approaches in extracting factors are principal components analysis and factor analysis. In finance, researchers seem to prefer principal components analysis over factor analysis since principal components can be uniquely defined up to a scalar trans-
formation without prespecifying the number of factors. Moreover, the order of the principal
compound factors is also unique. Chamberlain and Rothschild (1983) have even shown that
the approach can be applied to extract factors when asset returns have an approximate factor
structure. In an empirical study, Shukla and Trzcinka (1990) have explicitly compared principal
components analysis to factor analysis and have concluded that principal components analysis
is at least as good as factor analysis.

However, principal components analysis has one serious drawback. By design, the principal
components analysis maximizes the total variance of an extracted component and pays no atten-
tion to individual assets’ idiosyncratic volatilities. In finite samples, therefore, these extracted
factors tend to be severely contaminated by idiosyncratic returns. In the worst case scenario,
a particular asset itself can serve as a principal component when the asset returns have a huge
total volatility. This case won’t be a concern if the total volatility is largely systematic. Un-
fortunately, for individual stock returns, especially for daily returns, the total return volatilities
are dominated by idiosyncratic volatilities. As Malkiel and Xu (1997) have shown that the id-
iosyncratic volatility and the firm capitalization are negatively correlated, principal component
factors may weigh heavily on small stocks. More important, the poor out-of-sample performance
of the extracted factors can be a result of overweighting on idiosyncratic returns that change
unexpectedly over time.

In this paper, we modify the standard principal components (PC) analysis to deal with
the issue of cross-sectional heteroscedasticity of individual assets. This approach allows us to
choose each factor with maximum explanatory power across assets. When there is a pervasive
asset pricing factor, it should be highly correlated with most individual assets, which suggests
that the explanatory power of such a factor should be high. Thus, by maximizing the aggre-
gate explanatory power across all assets, the modified approach greatly reduces the influence of
heteroscedasticity in extracts factors and provides evenly distributed explanatory power across
individual assets. For convenience, we call this approach the “maximum explanatory compo-
component” (MEC) analysis. By focusing on the correlations structure, the MEC factors can also preserve the out-of-sample explanatory power.

The MEC approach is largely based on eigenvalue from a correlation matrix of asset returns. In this sense, we are not the first one to propose the methodology. However, this study contributes in three aspects. First, we provide a theoretical interpretation for the PC factors extracted based on a correlation matrix instead of a covariance matrix. We also show that the MEC factors have good large sample properties. In fact, the MEC factors converge to the true factor structure when the covariance structure of asset returns is known (also see Chamberlain and Rothschild, 1983). Drawing from Bai and Ng’s (2002) analysis, we demonstrate that the extracted factors will still converge to the true underlying factors in probability even when the covariance structure has to be estimated. Second, we perform an in depth analysis of the MEC factors empirically with respect to its ability to deal with the heterogeneous return structure. More important, we have demonstrated that the MEC factors have very good out-of-sample consistence properties and explanatory power using the actual NYSE/AMEX stock returns. Third, we extend the approach to individual stocks where the number of stocks far exceeds the number of time observations using adjusted cross-sectional dispersions. As a by-product, the explanatory power of the extracted factors measured by the average coefficients of determination are obtained directly without running regressions for individual stocks. This property leads to an efficient algorithm in computing individual $t$ statistics of factor loadings indirectly.

An equally important issue in the application of the APT model is to determine the number of pricing factors. We apply an information criterion similar to that of Bai and Ng’s (2002). This approach not only estimates the number of factors consistently but also can be conveniently implemented when using MEC factors. In addition, we provide three ad hoc criteria—the naive criterion, the eigenvalue criterion, and the $t$-ratio criterion to determine the number of factors. Since the true number of pricing factors is unobservable, the power of different approaches is studied through simulations.

---

2In fact, many statistics books have mentioned applying PC analysis to a correlation matrix as an alternative.
Related to this paper is a recent study by Jones (2001), who has combined the Connor and Korajczyk’s (1988) approach with the factor analysis in order to explicitly estimate the time-varying idiosyncratic volatilities. His methodology is designed to deal with heteroscedasticity over time under the assumption of constant idiosyncratic volatilities across individual assets. In contrast, the MEC approach is designed to account for heteroscedasticity across individual assets. As shown from simulation, the MEC approach is also capable of accommodating time-varying volatilities. Connor and Korajczyk (1988) have also proposed an iterative approach to deal with the heteroscedasticity across assets based on scaling individual asset returns by their idiosyncratic volatilities.3

The paper is organized as follows. In Section 1, we provide a theoretical analysis of the MEC approach. The issue of selecting the correct number of factors is then studied in Section 2 using the four different criteria. In Section 3 we analyze the finite sample properties of the MEC factors for both individual stocks and portfolios through simulation. Finally, we study the out-of-sample performance of the MEC factors using NYSE/AMEX stocks in Section 4 under variety measures including cross-validation and explanatory power. Concluding comments are provided in Section 5.

3Their approach would have been a more efficient one than ours if the true idiosyncratic volatilities were known. But the approach will result in choosing individual securities themselves as factors since the idiosyncratic volatilities have to be estimated simultaneously. See the discussion of a numerical example in the next section.
1 The Methodology

The Arbitrage Pricing Theory of Ross (1976) describes a linear relationship between the expected returns of individual securities and the expected returns of the underlying pricing factors.\(^4\) Since, the APT theory is silent about both the number of and the economic nature of the underlying pricing factors, a factor analysis technique can be used in extracting these factors. However, factor analysis approach requires a pre-specification of the number of factors and involves some kind of iterative procedure to estimate the factors. Chamberlain and Rothschild (1983) have proposed to use the principal components (PC) analysis instead to extract factors. In fact, they have shown that, under the approximate \(K\)-factor structure, PC factors converge to a rotation of the true underlying factors. Without assuming the knowledge of the true variance-covariance matrix of asset returns, Connor and Korajczyk (1986) have extended the PC analysis to the case of individual stocks. They have also provided a convergence result for the extracted factors.

\(PC\) analysis chooses a factor component by maximizing its total variance. The resulting components will be best in representing the true factor structure if each security has the same return idiosyncratic volatility. For a finite sample, however, residual returns are heteroscedastic and are correlated to some degree across individual assets. In fact, Malkiel and Xu (2001) and Trzcinka (1986) have shown that idiosyncratic return volatilities not only are very different across individual securities but also account for large portions of the total return volatility. Therefore, \(PC\) based factors may explain too much firm specific variation when applied to security returns. In other words, stocks with huge total volatility but less covariance with each other may be overweighted. This deficiency could also hamper the out-of-sample performance.

A simple solution to overcome the drawback of the standard PC analysis is to maximize the explanatory power of the extracted factors instead. Intuitively, if a factor is “pervasive,” it ought to correlate with majority asset returns. In other words, collectively, the true factors

\(^4\) Such a pricing relationship holds under both an exact factor structure, where the idiosyncratic returns are orthogonal to each other (see Ross, 1976; Dybvig, 1983; and Connor, 1984) and an approximate factor structure (Chamberlain and Rothschild, 1983), where the idiosyncratic returns can be correlated to some degree.
should have the largest average explanatory power across individual assets. Focusing on the explanatory power reduces the influence of idiosyncratic volatilities. Factors extracted under the new objective function are thus call “Maximum Explanatory Components,” or MEC for short. In this section, we first illustrate issues facing the PC approach using a simple numerical example.

1.1 A Simple Example

As discussed above, our focus is on idiosyncratic volatility. For illustration purpose, we assume there are two stocks that are determined by a common factor $r_F$.

\[
\begin{align*}
\text{Stock 1:} & \quad r_1 = r_F + \epsilon_1, \quad \epsilon_1 \sim N(0, \sigma^2), \\
\text{Stock 2:} & \quad r_2 = r_F + \epsilon_2, \quad \epsilon_2 \sim N(0, 4\sigma^2),
\end{align*}
\]

where idiosyncratic returns $\epsilon_1$ and $\epsilon_2$ are independent. We also assume that the common factor $r_F$ is distributed as $N(0, \sigma^2)$.

The PC Approach:

Applying the PC approach, we obtain the following two component portfolios,

\[
\begin{align*}
\hat{r}_{p1} = 0.23r_1 + 0.77r_2 = 1.00r_F + 0.23\epsilon_1 + 0.77\epsilon_2 = 1.00r_F + \epsilon_{p1}, \quad \epsilon_{p1} \sim N(0, 2.43\sigma^2), \\
\hat{r}_{p2} = 0.77r_1 - 0.23r_2 = 0.54r_F + 0.77\epsilon_1 - 0.23\epsilon_2 = 0.54r_F + \epsilon_{p2}, \quad \epsilon_{p2} \sim N(0, 0.81\sigma^2),
\end{align*}
\]

which leads to the following observations.

Observation 1: The first component portfolio does not purely represent the underlying factor $r_F$. It is severely contaminated by idiosyncratic returns due to a large weight on the second stock that has a large idiosyncratic return.

Observation 2: The second component portfolio does not purely represent the idiosyncratic returns, which makes the identification of the factor structure difficult.
The MEC Approach:

In contrast, the MEC factors can be constructed as,

\[ r_{p1} = 0.5r_1 + 0.5r_2 = 1.00r_F + 0.5\varepsilon_1 + 0.5\varepsilon_2 = r_F + \varepsilon_{p1}, \quad \varepsilon_{p1} \sim N(0, 1.25\sigma^2), \]
\[ r_{p2} = 0.5r_1 - 0.5r_2 = 0.00r_F + 0.5\varepsilon_1 - 0.5\varepsilon_2 = \varepsilon_{p2}, \quad \varepsilon_{p2} \sim N(0, 1.25\sigma^2), \]

Comparing with the PC factors, we see that,

**Observation 3:** The first constructed factor is less sensitive to idiosyncratic returns, and is much closer to the true factor than the first PC factor.

We can also compare the beta estimates and \( R^2 \)s for each stock using factors estimated using different approaches:

<table>
<thead>
<tr>
<th></th>
<th>The First Stock</th>
<th>The Second Stock</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\beta}<em>{r</em>{p1}} )</td>
<td>( R^2_{Stk1} )</td>
<td>( \hat{\beta}<em>{r</em>{p1}} )</td>
</tr>
<tr>
<td>True</td>
<td>1.00</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>PC</td>
<td>0.36</td>
<td>0.22</td>
<td>1.19</td>
</tr>
<tr>
<td>MEC</td>
<td>0.67</td>
<td>0.80</td>
<td>1.33</td>
</tr>
<tr>
<td>eCK</td>
<td>0.20</td>
<td>0.10</td>
<td>1.00</td>
</tr>
</tbody>
</table>

These estimates further reveal that,

**Observation 4:** When the PC factor is used, the beta estimate of the first stock is severely biased downward. The bias is much smaller when the MEC factor is used.

This observation implies that the factor premium estimated from cross-sectional regressions using PC factor loadings will also suffer a large bias.

**Observation 5:** Although the average \( R^2 \)s from different factor estimates are close, individual \( R^2 \)s are more evenly distributed across stocks when using the MEC factor than using the PC factor. In fact, the PC factor tends to penalize stocks with smaller variances.
Malkiel and Xu (1997) have shown that the idiosyncratic volatility and the market capitalization of a firm are negatively correlated. Observation 5 may suggest that the PC factors tend to put too much weight on small stocks. Relative to idiosyncratic returns, the first stock has a larger systematic component than that of the second stock. Thus, the extracted factors should better explain returns of the first stock than the second stock. This is exactly the case when the MEC factor is used, while it is just the opposite when using the PC factor.

Connor and Korajczyk (1988) have also proposed an extension (denoted as “eCK” above) to deal with idiosyncratic returns. This method requires a scaling of each asset returns by its idiosyncratic volatility first, followed by a standard PC analysis. Since idiosyncratic volatility is unobservable, an iterative approach is used in practice. Surprisingly, the extracted factor is just the second stock itself in our numerical example. This is because the idea behind the eCK approach is to extract factors that capture the maximum variance for the idiosyncratic volatility scaled returns. Since idiosyncratic volatility estimate depends on the estimate of the factor, the closer is the estimated factor to the return of a stock with the largest variance (i.e., the second stock), the smaller the idiosyncratic return for that stock will be. In this way, the variance of the scaled returns of the second stock will be huge. This iterative approach will converge to a point where the factor weight is just one on the second stock. In other words, the, the algorithm automatically chooses the second stock as a factor.\textsuperscript{5}

Since MEC factors tend to equalize the explanatory power across different assets, one concern might be that these factors could give less weight to stocks with large systematic components but with the same idiosyncratic volatilities. By design, PC factors should be the best choice in this case. Our numerical results show that, however, the MEC factors are not very different from the PC factors in this unrealistic case in terms of the beta estimates, but can dramatically increase $R^2$s of stocks with low betas.\textsuperscript{6} We discuss next how to directly construct the

\textsuperscript{5}This phenomenon seems to mimic the “Haywood” case in factor analysis. The difference is that “Haywood” case might occur in factor analysis depending on the nature of a sample, while the phenomenon for the eCK approach is guaranteed to occur independent of the characteristics of a sample.

\textsuperscript{6}Results are available from the author.
MEC factors from the perspective of maximizing their explanatory power, and the asymptotic properties.

1.2 The Maximum Explanatory Components (MEC) Analysis

The APT model starts with the following linear structure for the $N$ asset returns,

$$
R = \alpha + \beta F + \tilde{\epsilon} \quad (1)
$$

where $R, \tilde{\epsilon} \in \mathbb{R}^{N \times T}$ are matrices of total returns and idiosyncratic returns, respectively, $\beta \in \mathbb{R}^{N \times K}$ is a matrix of factor loadings with respect to the fundamental factor return matrix, $\tilde{F} \in \mathbb{R}^{K \times T}$, and $\alpha \in \mathbb{R}^{N \times 1}$ is a vector of individual assets’ expected returns that satisfies the usual pricing conditions. In order to allow for an approximate factor structure, we also assume that the largest eigenvalue of matrix $D$ is bounded. Since a factor should be pervasive (i.e., be important for most of the assets), we assume that the smallest eigenvalue of matrix $\beta'\beta$ is unbounded when $N$ goes to infinity. In analyzing the asymptotic properties, we will apply the main result from Chamberlain and Rothschild (1983). Thus, their other assumptions including the mixing condition and the bounded higher moment conditions are also adopted here. In addition, we assume that the diagonal elements of the variance-covariance matrix $\Sigma = \text{Cov}(R, R')$ is bounded and nonzero.

In order to reduce the influence of idiosyncratic volatilities, we extract factors by maximizing the average explanatory power of the extracted factors across individual assets. For ease of exposition, denote $\tilde{R}$ as the demeaned asset returns, $\Sigma = \text{Cov}(\tilde{R}, \tilde{R}')$ and $\Omega = \text{Corr}(\tilde{R}, \tilde{R}')$ as the variance-covariance matrix and the correlation matrix of asset returns, respectively, and $V = [\text{diag}(\Sigma)]^{1/2}$ as a diagonal matrix of standard deviations. Given the correlation matrix $\Omega$ can be written as $\Omega = V^{-1}\text{Cov}(\tilde{R}, \tilde{R}')V^{-1} = \text{Cov}(V^{-1}\tilde{R}, \tilde{R}'V^{-1})$, we can interpret it as the “covariance” matrix of the transformed return matrix $R^*(= V^{-1}\tilde{R})$. Following the convention,
we assume that the eigenvectors \( \mathbf{a}_i \) \( (i = 1, 2, \cdots, N) \) of matrix \( \Omega \) are orthogonal to one another with a unit norm, and eigenvalues \( \lambda_1, \lambda_2, \cdots, \geq \lambda_N \geq 0 \) are sorted in a descending order. Under these notations, the following main proposition describes the procedure to construct MEC factors.

**Proposition 1** The \( i \)-th MEC factor, \( \mathbf{f}_i^* \), that maximizes the average coefficient of determination from regressing individual asset returns on this factor conditioned on the first \( (i - 1) \) factors is given by,
\[
\mathbf{f}_i^* = \frac{1}{\sqrt{\lambda_i}} \mathbf{R}^{*'} \mathbf{a}_i = \frac{1}{\sqrt{\lambda_i}} \tilde{\mathbf{R}}' \tilde{\mathbf{V}}^{-1} \mathbf{a}_i.
\]

The corresponding average \( \bar{R}_i^2 \) of the \( i \)-th factor can be directly obtained as,
\[
\bar{R}_i^2 = \frac{1}{N} \lambda_i.
\]

**Proof:** See Appendix A.

This proposition says that the first MEC factor is the best in explaining the \( N \) asset returns. Similarly, the second factor is the best linear factor for explaining the residual returns after controlling for the first factor, and so forth. Although some researchers in statistics have applied the standard PC analysis on a correlation matrix and have lead to the same factor structure as that of equation (2), Proposition 1 provides a rationale for such practice and an interpretation of the eigenvalues from the procedure. This simple, but important extension to the standard PC methodology, allows us to overcome its weakness of over-weighting on assets with large idiosyncratic volatilities while providing the maximum explanatory power. In fact, MEC factors can retain common variations as much as possible even when volatilities vary greatly across individual stocks.

It is also true that \( R^2 \)'s from regressing individual stock returns on MEC factors tend to vary less than those on PC factors.\(^7\) In other words, the MEC factors will have more evenly distributed explanatory power across individual stocks. However, this fact does not necessarily mean that the average \( R^2 \) using the MEC factors will be very different from the average \( R^2 \)'s using other factors.

\(^7\)Mathematically, this is due to the fact that the objective function of equation (22) in Appendix A is concave. Thus, an optimum is reached when individual values are not too far away from the mean.
In implementation, the MEC approach requires scaling individual asset returns by their total volatilities. This procedure will not under weigh those assets with large systematic returns but small idiosyncratic returns as shown in our numerical example. This is because those assets will also have large covariances with the common factors. In other words, the relative importance of the systematic component to total returns is more important than the absolute importance.\footnote{The absolute importance may be dampened to some degree when the idiosyncratic return component is very small. This is unlikely when dealing with individual stocks.}

Since the constructed factors are orthogonal to one another, it is easy to establish the following result from Proposition 1,

**Corollary 1** The average $R^2$ from regressing individual asset returns on the first $K$ MEC factors is: $\frac{1}{N}\sum_{i=1}^{K} \lambda_i$.

This corollary is useful in designing different approaches to determine the number of factors.

For computational efficiency, we restate equation (2) in a matrix notation as,

$$F^* = \Lambda_{A_K}^{-1/2} A'_K R^*,$$

where $A_K$ and $\Lambda_{A_K}$ are the matrix of the first $K$ eigenvectors and the eigenvalue matrix, respectively. We can also obtain the corresponding factor loading matrix as,

$$B^* = \text{Cov}(\tilde{R}, F^*) [\text{Var}(F^*)]^{-1} = \text{VCov} \left( R^*, R^* A_K \Lambda_{A_K}^{-1/2} \right) = V \Omega A_K \Lambda_{A_K}^{-1/2} = V A_K \Lambda_{A_K}^{-1/2}. \ (5)$$

It is useful to know the statistical significance of the loading estimates. Although one can run individual regressions for each stocks, it is tedious when there are several thousand stocks. Fortunately, due to the special structure of the MEC factors, all the $t$-ratios can be computed together in one step.

**Corollary 2** The matrix of $R^2$ with the $(i, j)$-th element being the coefficient of determination from regressing the $i$-th stock return on the $j$-th MEC factor, and the matrix of $\tau^2$ with the $(i, j)$-th element being the squared $t$-ratio of the $j$-th MEC factor loading for stock $i$, can be computed directly as,

$$R^2 = (V^{-1}B^*) \odot (V^{-1}B^*), \quad \tau^2 = \frac{R^2 \odot (11' - R^2 11')}{}, \quad (6)$$

$$R^2 = (V^{-1}B^*) \odot (V^{-1}B^*), \quad \tau^2 = \frac{R^2 \odot (11' - R^2 11')}{}, \quad (7)$$
where \(\circ\) and \(\oslash\) are element-by-element multiplication and division, respectively.

**Proof:** See Appendix B.

### 1.3 Estimation of the MEC Factors

The first step in MEC analysis is to estimate the variance-covariance matrix, \(\Sigma\). This is usually done using the maximum likelihood estimator of, \(\hat{\Sigma} = \frac{1}{T-1}\hat{R}\hat{R}'\). \(\hat{\Sigma}\) is invertible as long as \(T > N\). In many applications, however, the number of securities far exceeds the number of return observations, especially for individual stocks. Moreover, it might be more accurate to estimate the factor structure over a relatively short sub-sample period since both the expected returns and the variance-covariance structure change over time as indicated in many empirical studies. For these reasons, we extend the MEC approach to the case of \(T < N\).

Denote \(\hat{V} = [diag(\hat{\Sigma})]^{1/2}\) as a diagonal matrix with diagonal elements equal to those of \(\hat{\Sigma}\), and \(\hat{\Omega} = \frac{1}{T-1}\hat{R}^*\hat{R}^*\) as an estimate of the correlation matrix, where \(\hat{R}^* = \hat{V}^{-1}\hat{R}\). For the case of \(T < < N\), the first \(T\) eigenvalues and eigenvectors of the cross-time pseudo correlation matrix \(\hat{Q} = \frac{1}{N-1}\hat{R}^*\hat{R}^*\) is equivalent to those of \(\hat{\Omega}\). Let \(\hat{A}\) and \(\hat{A}_T\) be the eigenmatrix and eigenvalue matrix of \(\hat{\Omega}\), respectively, and \(\hat{H}\) and \(\hat{\Lambda}_H\) be the eigenmatrix and eigenvalue matrix of \(\hat{Q}\), respectively. The following analysis demonstrates the equivalence of obtaining the MEC factors from matrix \(\hat{Q}\).

\[
\left(\frac{1}{N-1}\hat{R}^*\hat{R}^*\right)\hat{H} = \hat{H}\hat{\Lambda}_H; \quad \hat{A}_T = \begin{bmatrix} \hat{\Lambda}_H & 0 \\ 0 & 0 \end{bmatrix}
\]

\[
\Rightarrow \left(\frac{1}{N-1}\hat{R}^*\hat{R}^*\right)\hat{H}\hat{\Lambda}_H^{-1/2} = \hat{H}\hat{\Lambda}_H^{-1/2}\hat{\Lambda}_H
\]

\[
\Rightarrow \left(\frac{1}{T-1}\hat{R}^*\hat{R}^*\right)\left[\frac{1}{\sqrt{N-1}}\hat{R}^*\hat{H}\hat{\Lambda}_H^{-1/2}\right] = \left[\frac{1}{\sqrt{N-1}}\hat{R}^*\hat{H}\hat{\Lambda}_H^{-1/2}\right] \left(\frac{N-1}{T-1}\hat{\Lambda}_H\right).
\]

In other words, the first \(T\) eigenvectors and eigenvalues of the correlation matrix \(\hat{\Omega}\) can be written as \(\hat{A}_T = \frac{1}{\sqrt{N-1}}\hat{R}^*\hat{H}\hat{\Lambda}_H^{-1/2}\) and \(\hat{\Lambda}_T = \frac{N-1}{T-1}\hat{\Lambda}_H\). Therefore, the first \(K\) MEC factors
can be expressed according equation (4) as,
\[
\hat{F}^* = \hat{A}^{-1/2}_K \hat{A}'_K \hat{R}^* = \left( \sqrt{\frac{T-1}{N-1}} \hat{A}^{-1/2}_H \hat{H}_K \hat{A}'_K \right) \hat{R}^* = \sqrt{T-1} \hat{R}'_K. \tag{8}
\]
Without running individual regressions for each stock, we can immediately obtain the corresponding factor loadings as,
\[
\hat{B}^* = \frac{\hat{R} \hat{F}^*}{T-1} \left( \frac{\hat{F}^* \hat{F}^*'}{T-1} \right)^{-1} = \hat{R} \left( \sqrt{T-1} \hat{H}_K \right) \left[ (T-1) \hat{H}'_K \hat{H}_K \right]^{-1} = \frac{1}{\sqrt{T-1}} \hat{R} \hat{H}_K. \tag{9}
\]
Equation (8) suggests that the scaled \( i \)-th eigenvector of \( \hat{Q} \) is already the estimator of the \( i \)-th MEC factor. Moreover, according to Proposition 1, the average \( R^2 \) from regressing individual asset returns on the \( i \)-th MEC is the scaled \( i \)-th eigenvalue of matrix \( \hat{Q} \).

We summarize the steps in applying the MEC procedure under different scenarios:

**Case 1:** For portfolios, where \( N < T \), the \( i \)-th factor returns can be estimated from \( \hat{f}_i^* = \frac{1}{\sqrt{\hat{\lambda}_i}} \hat{R}'_i \hat{a}_i \) where \( \hat{\lambda}_i \) and \( \hat{a}_i \) are the \( i \)-th eigenvalue and eigenvector of the correlation matrix \( \hat{\Omega} = \frac{1}{T-1} \hat{R}' \hat{R} \), respectively. \( \frac{\hat{\lambda}_i}{N} \) is also the average \( R^2 \) from regressing individual portfolio returns on the \( i \)-th extracted factor alone. Moreover, loadings can be computed according to equation (5), with \( t \)-ratios obtained directly from equation (7), where,
\[
R^2 = (\hat{A}_K \hat{A}^{1/2}_K) \odot (\hat{A}_K \hat{A}^{1/2}_K).
\]

**Case 2:** For individual stocks, where \( N > T \), the \( i \)-th factor returns can be estimated from \( \hat{f}_i^* = \sqrt{T-1} \hat{h}_i \) and \( \hat{\lambda}_i = \frac{N-1}{T-1} \hat{\lambda}_{H,i} \) where \( \hat{\lambda}_{H,i} \) and \( \hat{h}_i \) are the \( i \)-th eigenvalue and eigenvector of the matrix \( \hat{Q} = \frac{1}{N-1} \hat{R}' \hat{R} \), respectively. Again, \( \frac{\hat{\lambda}_i}{N} \) is also the average \( R^2 \) from regressing individual stock returns on the \( i \)-th extracted factor alone. Loadings are obtained directly from equation (9) with \( t \)-ratios computed directly from equation (7), where,
\[
R^2 = \frac{1}{T-1} (\hat{R}' \hat{H}_K) \odot (\hat{R}' \hat{H}_K).
\]
Before we proceed to empirical study, it is necessary to take a detailed look at the issue of what these constructed factors might represent.
1.4 Asymptotic Properties for the MEC Factors

Similar to the asymptotic properties of PC factors, MEC factors also converge to a rotation of the true underlying factors that drive asset returns when the correlation structure is known (see Appendix C). In this sense, both the PC and the MEC approaches are equivalent asymptotically despite very different finite sample properties. When the covariance structure $\Sigma$ has to be estimated, Connor and Korajczyk (1986) have proposed an asymptotic principal components approach. Additional asymptotic results are provided by Bai and Ng (2002).

Under mild conditions that allow for both heteroscedasticity over time and serial dependence in the residuals, Bai and Ng (2002) have shown in their Theorem 1 that the standard PC factors converge to a rotation of the true underlying factors with a convergence rate of $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. Following the same logic, we can extend this convergence result to our case with an additional assumption of $V^{-1} < MI_N$, where $M$ is a positive constant. In particular, from the definition of eigenvectors for matrix $\hat{Q}$, we can rewrite the factor estimates as $\hat{F}^* = \frac{1}{N-1} \hat{\Lambda}_H^{-1} \hat{F}^* (\hat{R}' \hat{\beta})$. Therefore, the following result can be obtained,

$$
C_{NT}^2 \left( \frac{1}{T} \sum_{t=1}^T \| \hat{F}^{*k}_t - L^K \hat{F}_t \|^2 \right) = O_p(1),
$$

where $L^K = \frac{1}{N-1} \hat{\Lambda}_H^{-1} \hat{F}^{*k} (\hat{F}' \hat{\beta} V^{-2} \hat{\beta})$ and $k (k \leq K)$ is the chosen number of factors. That is, the $k$ factors ($\hat{F}^{*k}$) estimated using the MEC approach converge in probability to a rotation of the true factor $L^K \hat{F}$ at a speed of $C_{NT}$.

In general, the MEC approach not only allows us to choose factors that maximize the total explanatory power for individual asset returns, but also the can exclude the possibility of putting too much weight on assets with huge idiosyncratic volatilities in a finite sample. In addition, we offer a clear interpretation of the eigenvalues from the analysis and provide a simple procedure to compute the loadings and the corresponding $t$-ratios without running individual regressions. Therefore, given the convergence results, it is better to focus on the correlation structure instead of the covariance structure.
2 Factor Selection

It is equally challenging to determine the correct number of pricing factors from the observed stock returns. Commonly used approaches are based on the idea of testing residual returns being independent after controlling for $K$ factors. Alternatively, one can treat this problem as a classical model selection problem as in Bai and Ng (2002). We use a similar information criterion in conjunction with the MEC factors. Since the consistency of these approaches rely on asymptotic properties, some ad hoc approaches may be useful for finite samples.

2.1 The Information Criterion

When factors $\hat{F}$ are estimated from minimizing an objective function, $V(k)$, for a given number of factors $k$, the loss function $L(k)$ in model selection is usually defined as,

$$L(k) = V(k) + kg(N, T) = \min_{\hat{F}^k} V(k, \hat{F}^k) + kg(N, T),$$

where $g(N, T)$ satisfies the condition of (i) $g(N, T) \to 0$ and (ii) $C_{NT} \cdot g(N, T) \to \infty$ as $N, T \to \infty$. Although the objective function $V(k)$ is decreasing in $k$, its incremental reduction will not offset the increase in the efficiency loss when $k$ exceeds the true number of factors. In other words, from a model selection perspective, the number of factors can be determined by choosing $k$ that minimizing the loss function. For example, since PC analysis is equivalent to solving the following minimization problem,

$$V(k) = \min_{\beta^k, \hat{F}^k} V(k, \hat{F}^k) = \min_{\beta^k, \hat{F}^k} \frac{1}{N} \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} (\hat{R}_{i,t} - \hat{\beta}_i^k \hat{f}_t^k)^2 = \frac{1}{N} \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} (\hat{R}_{i,t} - \hat{\beta}_i^k \hat{f}_t^k)^2,$$

Bai and Ng (2002) have shown in their Theorem 2 that the true number of factors can be consistently estimated from $\arg \min_k L(k)$ given the principal component factors. In security analysis applications, we often face the case of $N >> T$. From their simulation, Bai and Ng have suggested that the most effective choice of loss function (denoted as $C_p$) is,

$$C_p(k) = V(k) + k\hat{\sigma}^2 \frac{N + T}{NT} lnT,$$
where \( \hat{\sigma}^2 = \frac{1}{N} \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} (\hat{R}_{i,t} - \hat{\beta}_i^{k_{\text{max}}} \hat{f}_t^{k_{\text{max}}})^2 \) and \( k_{\text{max}} \) is the maximum number of factors considered. In order to implement equation (13), the residual variance of each stock has to be estimated for all possible values of \( k \).

We propose a similar, but easy to implement, loss function with respect to the MEC factors. Since maximizing \( R^2 \) is equivalent to minimizing \( (1 - R^2) \), MEC factors can be equivalently estimated from imposing the following objective function based on the OLS regression theory,

\[
\rho(k) = \min_{\beta^*, F^k} \rho(k, F^k) = \min_{\beta^*, F^k} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} (\check{R}_{i,t} - \beta_i^k \check{f}_t^k)^2 = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} (\check{R}_{i,t} - \hat{\beta}_i^{*k} \hat{f}_t^{*k})^2,
\]

where \( \hat{F}^{*k} \) is a vector of the \( k \) MEC factors, and \( \sigma_i^2 \) is asset \( i \)'s total return variance. From corollary 1, the computation of the objective function \( \rho(k) \) can be greatly simplified to,

\[
\rho(k) = 1 - \frac{1}{N} \sum_{i=1}^{k} \lambda_i,
\]

where \( \lambda_i \) is the \( i \)-th eigenvalue. Using \( \rho(k) \) as the objective function in equation (11), and choosing the same \( g(N, T) \) function as in (13), we can estimate the number of factors by minimizing the following information criterion,

\[
C_p^{MEC}(k) = 1 - \frac{1}{N} \sum_{i=1}^{k} \lambda_i + k \hat{\lambda} \frac{N + T}{NT} \ln T \tag{14}
\]

where \( \hat{\lambda} = 1 - \frac{1}{N} \sum_{i=1}^{k_{\text{max}}} \lambda_i \). The estimated number of factors \( k \) from equation (14) is thus consistent.\(^9\) This simple extension is very easy to implement and does not require individual regressions or additional computations for each choice of \( k \).

### 2.2 The Ad Hoc Approaches

The consistency of the \( C_p \) approach relies on asymptotic properties. For a finite sample, some ad hoc approaches may work well. In fact, the unique properties of the MEC factors allow us to propose three additional criteria in estimating the number of factors.

\(^9\)This is because the MEC factor estimates converge in probability to a rotation of the true factors as discussed in the last section, which satisfies the conditions for consistency in Corollary 2 of Bai and Ng (2002).
Brown (1989) has shown that the standard PC analysis produces a pattern in the eigenvalues even if the data are generated from $k$ equally important factors. In particular, the first eigenvalue will be extremely large followed by $(k - 1)$ moderate and slowly decreasing eigenvalues. Eigenvalues after the $k$-th eigenvalue will be very small. In other words, there will be a noticeable drop in the level of eigenvalues from the $k$-th to the $(k + 1)$-th eigenvalues. A similar pattern occurs when applying the MEC approach. In fact, it is much easy to detect such a pattern since the scaled eigenvalues can be directly interpreted as $R^2$'s. This suggests the following criterion,\(^\text{10}\)

**Naive Criterion:** A total of $k$ ($k > 1$) factors should be considered when the average incremental $R^2$ from including the $(k + 1)$-th factor drops significantly from that of the $k$-th factors, and the incremental $R^2$'s decrease steadily afterwards.

When an additional factor only captures the comovement of a few stocks, the incremental gain in the average $R^2$ by including that factor will be relatively small, say less than $\delta\%$. One caveat is that a moderate incremental average $R^2$ could be simply due to a large $R^2$s for a few stocks in a finite sample. As discussed above, this is unlikely since MEC factors tend to have evenly distributed explanatory power across stocks. Due to the direct relationship between the incremental $R^2$s and the eigenvalues from Proposition 1, we can establish the following criterion.

**Eigenvalue Criterion:** The first $k$ components should be chosen as factors if the $(k + 1)$-th eigenvalue drops below $N \times \delta\%$ for the first time.

The conventional “eigenvalue one rule” is a special case here, where the $\delta$ is set to $\frac{1}{N}\%$.\(^\text{11}\) In applying the criterion, we need to choose the appropriate $\delta$.

Since the MEC factors are orthogonal to one another, we have the following relationship

\(^{10}\)Once can consider this approach as a modified ‘scree-plot’ (See Bartholomew, Steele, Moustaki, and Galbraith, 2002).

\(^{11}\)The shortcoming of the “eigenvalue one rule” is apparent. Given the same structure of asset returns, one would include more factors when the number of assets increases.
between the \((k + 1)\)-th factor loading and the incremental \(R^2\) of asset \(i\).

\[
R^2_{i,\hat{f}_{k+1}^*} = \frac{1}{T - 1} t^2_{i,\hat{f}_{k+1}^*} (1 - \sum_{l=1}^{k+1} R^2_{i,\hat{f}_l^*}).
\]  

(15)

If all the loadings for the \((k + 1)\)-th factor are at least significant at the \(\nu\)% level, i.e., \(t_{i,\hat{f}_{k+1}^*} \geq t_\nu\), we can sum both sides of equation (15) over individual assets as,

\[
\sum_{i=1}^{N} R^2_{i,\hat{f}_{k+1}^*} = \lambda_{i,k+1} \geq \frac{1}{T - 1} t^2_{\nu} \sum_{i=1}^{N} (1 - \sum_{l=1}^{k+1} R^2_{i,\hat{f}_l^*}) = \frac{t^2_{\nu}}{T - 1} (N - \sum_{l=1}^{k+1} \lambda_l).
\]  

(16)

This result suggests that we can set,

\[
\delta = \frac{t^2_{\nu}}{N(T - 1)} (N - \sum_{l=1}^{k+1} \lambda_l).
\]  

(17)

Certainly, this is a necessary condition for choosing \(k\). At the same time, we do not require a factor to be significant at the conventional level for all individual stocks. \(\nu\) can be set at a lower significance level, for example \(\nu = 10\%\).

In an approximate factor structure of Chamberlain and Rothschild (1983), idiosyncratic returns of individual stocks can be weakly correlated with one another. In other words, the \((k + 1)\)-th eigenvector and the eigenvectors thereafter will not be exactly zero, but could be ignored if they only capture the comovement of a limited number of assets. Therefore, we can define a pervasive factor as one whose corresponding factor loading estimates are statistically significant for sufficient number of assets, that is

\textit{t-Ratio Criterion} : Only the first \(k\) factors should be included if there are less than \(\gamma\)% of the \((k + 1)\)-th factor loadings of individual assets being statistically significant at a conventional level.

Although the choice of \(\gamma\) is subjective, we believe \(\gamma = 30\%\) is reasonable for factor loadings to be significant at a 5% level based on our simulation results in the next section. This criterion is also easy to implement without the need to run individual regressions for each asset against each factor. A computational procedure is given in corollary 2.
3 Performance Comparison

Not only the statistical distributions for portfolio returns are very different from those of individual stocks, but the estimation procedure are different as well. Therefore, we will investigate the finite sample properties of MEC factors for both portfolios and individual stocks. For individual stocks, we will compare the MEC approach to both the Connor and Korajczyk’s (1986) approach (CK) and the heteroscedasticity consistent approach (HFA) of Jones (2001). In order to retain the complicated return structure of individual stocks, we will generate return data from the characteristics of actual returns. For portfolios, we compare MEC factors to standard PC factors since both CK and HFA factors are designed for individual stocks.

3.1 Simulation Results for Portfolios

$R^2$s from fitting the Fama and French’s (1993) three-factor model to the 100 size-beta sorted portfolios are relatively large ranging from 35% to 85%. Thus, we simulate portfolio returns with an average $R^2$ of 60% using three equally important factors in order not to give advantages to any factor extracting methods. In particular, we assume that the $i$-th portfolio return is generated from,

$$R_i = \beta_{1,i}f_1 + \beta_{2,i}f_2 + \beta_{3,i}f_3 + \epsilon_i,$$

where $f_1$, $f_2$, and $f_3$ are the three independent factors with $i.i.d$ normal distribution of unit variance. To allow for sufficient variations in factor loadings, $\beta_{1,i}$, $\beta_{2,i}$, $\beta_{3,i}$ are all generated according to a normal distribution of $N(1,0.25)$. We also assume that the idiosyncratic return $\epsilon_i$ is $i.i.d.$ of $N(0,\sigma_i^2\kappa_i^2)$ in order to produce sufficient variations in idiosyncratic volatilities across portfolios.\(^{12}\) Evidence from mutual fund data suggests that portfolio idiosyncratic volatilities and betas are related. Parameter $\kappa$ is used to capture such correlation. In addition, we define an aggregate beta as, $\beta_i^2 = (\beta_{1,i}^2 + \beta_{2,i}^2 + \beta_{3,i}^2)$ and explore the following three cases by controlling $\kappa$.

\(^{12}\)We do not consider heteroscedasticity over time for portfolios since none of the approaches studied here designed to correct for this effect.
Random $R^2$: Set $\kappa_i = 1$, and $\sigma_i \sim \text{uniform}(0.3, 3.0)$. The $R^2$s of individual assets are completely random with an average $R^2$ of 60.6%.

Increasing $R^2$: Set $\kappa_i = \beta_i$ and $\sigma_i = \gamma(k)$ if the corresponding $\beta_i$ is at the $k$-th rank from sorting $\beta_i$ in ascending order, where $\gamma_i$ is linearly decreasing from 1.53 to 0.17. In this case, $R^2$s not only strictly increase with $\beta_i$s but also have an average value of 60.6%.

Decreasing $R^2$: Set $\kappa_i = \beta_i$, and $\sigma_i = \gamma(k)$ if the corresponding $\beta_i$ is at the $k$-th rank from sorting $\beta_i$ in ascending order, where $\gamma_i$ is linearly increasing from 0.17 to 1.53. The average $R^2$ is again equal to 60.6%. The individual $R^2$s decrease with their $\beta_i$s.

The number of months is set at $T = 75 \times 12$, which resembles that of monthly observations on the CRSP tape. In addition, we study three portfolio sizes of $N = 20$, $N = 50$, and $N = 100$.

As Brown (1989) has shown that, even when starting from a structure with equally important factors, the PC analysis will rotate the underlying factors so that the first extracted factor captures most of the variations. This feature makes it difficult to compare the performance of different methods directly. However, since both the PC and the MEC factors reflect weighted average of individual asset returns, we can indirectly compute the loading estimates condition on the true factor returns.

$$
R_i = \hat{\beta}_{1,i}f_1 + \hat{\beta}_{2,i}f_2 + \hat{\beta}_{3,i}f_3 + e_i
= \sum_{l=1}^{3} \hat{\beta}_{l,i} \sum_{j=1}^{N} w_{l,j} (\beta_{1,j}f_1 + \beta_{2,j}f_2 + \beta_{3,j}f_3 + e_j) + e_i
= \sum_{k=1}^{N} \beta_{k,j} \sum_{l=1}^{3} w_{l,j} \hat{\beta}_{l,i})f_k + \sum_{j=1}^{N} \sum_{l=1}^{3} w_{l,j} \hat{\beta}_{l,i} e_j + e_i,
$$

where $w_{l,j}$s are the weights for the constructed factors using different methods. Therefore, we can define a pseudo estimator, $\tilde{\beta}_{k,i} = \sum_{j=1}^{N} \beta_{k,j} \sum_{l=1}^{3} w_{l,j} \hat{\beta}_{l,i}$, for the $i$-th factor loading. In this way, we are able to compute the root mean square errors (RMSE) for different methods.

We first investigate the case of “random $R^2$,” where no method inherits any advantages by design. As shown in Table 1, there are significant difference between the two approaches. For
example, the RMSEs of $\tilde{\beta}_1$ using the MEC factors are 35%, 30%, and 16% lower than those using the PC factors for the case of 20, 50, and 100 portfolios, respectively.\textsuperscript{13} Furthermore, there are about 80% chance for the PC loading estimates and 90% chance for the MEC loading estimates to be less than 10% away from the true value (not shown in the table). We also see that the $R^2$s from regressing individual portfolio returns on the three extracted factors are a little higher for the MEC factors than for the PC factors.

[ Insert Table 1 here ]

It is also interesting to see how estimation errors are related to the true factor loadings. In Figure 1, we have plotted the estimation errors of the first factor against the corresponding true betas. When the PC factors are used, the estimation errors are negatively correlated with the true betas as shown in Panel A of Figure 1. This evidence suggests that the PC factors systematically underestimate the beta loadings for high risk portfolios and overestimate the beta loadings for low risk portfolios. If these estimated loadings are used to test the APT model cross-sectionally, we tend to reject the model more often due to smaller variations in the loading estimates. Such a relationship between estimation errors and the true betas does not exist when the MEC factors are used as shown in panel B of Figure 1.

[ Insert Figure 1 here ]

In order to assess the accuracy of the factor estimates themselves, we define a precision measure $\eta$, as the percentage of the total variance of the extracted factor explained by the variance of the true factors.\textsuperscript{14} For the same case of “random $R^2$,” not only the precision of each factor estimate increases with the number of portfolios, but also the first factor can be precisely estimated with the $\eta$ measures above 95% in general. Significant differences occur in the second

\textsuperscript{13}Since we use the pseudo estimator $\tilde{\beta}$ and we have assumed equal importance of each factor, the beta coefficient estimates for other two factors are very similar.

\textsuperscript{14}This definition is motivated by the fact that an estimated factor represents a rotation of the true underlying factors contaminated by idiosyncratic returns of individual assets.
and the third extracted factors. For example, when the number of portfolios is small, the second and the third \textit{PC} factors only contain 42\% and 20\% of the true factors, respectively, while the second and the third \textit{MEC} factors count 66\% and 41\% of the true factors, respectively. Moreover, the fourth \textit{PC} factor which is supposed to be purely idiosyncratic still contains 9\% of the true factors. In contrast, the fourth \textit{MEC} factor contains virtually no true factors. When the number of portfolios exceeds 100, the precision measures for the second and third \textit{MEC} factors are close to 90\%, and those of the \textit{PC} factors are about 85\%.

For the structure of “increasing $R^2$,” this is the “worst” case scenario for \textit{MEC} factors since \textit{PC} factors have an edge by design. Although the \textit{PC} approach outperforms the \textit{MEC} approach, the actual difference is minimum. For example, under the sample size of 20, the RMSEs are 0.109 and 0.115 for \textit{PC} and \textit{MEC} factor loadings, respectively. There is virtually no difference when the number of portfolios increases to 100. In general, both approaches work well.$^{15}$

In reality, less (more) risky stocks or portfolios tend to have large (small) $R^2$s. This phenomenon can be simulated using the structure of “decreasing $R^2$” in Table 1. Clearly, there are large differences in performance between the two approaches even when the number of portfolios exceeds 100. For example, the RMSE of 0.15 for the \textit{PC} factors is twice as large as that of the \textit{MEC} factors even under the portfolio size of 100. The precision measures for the second and the third \textit{MEC} factors are about 94\%, while those for the \textit{PC} factors are only about 67\%. When the number of portfolios is small, the differences are much larger. Moreover, it is difficult to distinguish high order \textit{PC} factor from idiosyncratic returns. Therefore, the \textit{MEC} approach not only has advantages in the efficiency of the beta estimates and the precision of the extracted factors themselves, but also is robust in different scenarios.

$^{15}$In our simulation, $R^2$s of individual portfolios have a perfect relationship with the aggregate beta measure. When such a perfect relationship is broken by introducing a small noise, the \textit{MEC} approach again outperforms the \textit{PC} approach.
3.2 Simulating Individual Stock Returns

Simulation methods not only allow us to assess the finite sample properties, but also help us to establish a common ground to compare different approaches. Compared to portfolios, it is difficult to model individual stock returns. Even the Fama and French’s (1993) three-factor model can only explain less than 20% of the return variations in recent years. In order to design our simulation as close to reality as possible, we utilize the characteristics of individual stocks covered by the CRSP tape. Similar to the setting of most factor structure studies, we focus on monthly frequency over the sample period from 1963 to 1999. Due to the number of stocks available and possible changes in the factor structure over time, we divide the whole sample period into the following six subsample periods of six years each: 1964-1969, 1970-1975, 1976-1981, 1982-1987, 1988-1993, and 1994-1999. For consistency, we have further restricted our sample to NYSE/AMEX stocks since NASDAQ stocks have been added to CRSP tape only after 1973, and NASDAQ stocks tend to be in favor of the MEC approach.

3.2.1 Simulation Design

In order not to give advantage to any approaches, we assume that individual stock returns are generated by factors that have similar characteristics as those of the Fama and French’s (1993) three factors. In particular, asset returns used in our simulations are generated according to,

$$ \hat{R}_{i,t} = \sum_{k=1}^{3} \beta_{i,k} f_{k,t} + \sigma_{i,t} e_{i,t}, \quad i = 1, \ldots, N_p, \quad \text{and} \quad t = 1, \ldots, 72 $$

(19)

where $\beta_{i,k}$ is the $k$-th factor loading estimated from regressing stock $i$’s returns on the orthogonalized Fama and French’s three factors. The reason for using the actual loading estimates is to retain the distributional structure of observed stock returns. The means and the standard deviations of these loading estimates for each subsample period shows sufficient variations (see

---

16 Recent studies suggest that the factor structure itself might change over time (see Merville and Xu, 2001). This is also supported by the empirical evidence on the time varying risk premium.

17 Both the size and the book-to-market variables are orthogonalized to the market variable first. The size variable is then orthogonalized to the book-to-market variable.
Panels B and C of Table 2). $\tilde{f}_{k,t}$ is the $k$-th factor return generated from a normal distribution with the variance set to that of the corresponding Fama and French factor. $N_p$ is the number of individual stocks in the $p$-th subsample period reported in the first row of Table 2, and $e_{i,t} \sim N(0,1)$ is the idiosyncratic return.

[ Insert Table 2 here ]

The volatilities of individual stock returns vary greatly in each of the subsample periods as shown in Panel D of Table 2. Although the median monthly return volatility is about 8.6% in the late 1990s, the 90-th percentile is as high as 16%. In order to capture such a large variation in volatility, we vary idiosyncratic volatility using the structure of $\sigma_{i,t} = \sigma_i \times \sigma_t$, where $\sigma_i$ is estimated directly from the Fama and French three-factor model. For the case of homoscedastic returns across stocks, we set $\sigma_i$ to the mean of the estimated individual idiosyncratic volatilities, $\bar{\sigma}$. We can also allow for heteroscedasticity over time as used in Jones (2001). In this case, we can let $\sigma_t$ increase from 0.8 to 1.2 (or 50%) evenly over the six-year sample period so that the average level of $\sigma_i$ is not affected.\footnote{Xu and Malkiel (2003) have shown that the average volatility of the 20 most volatile stocks has gone up three times from 1963 to 1998, or 33% over a six-year period. Therefore, our specification here is unrealistic. However, the heteroscedasticity consistent method of Jones (2001) works best when there are significant differences in idiosyncratic volatilities over time.}

In summary, we study the following three cases: 1) “Homoscedastic Return Across Stocks”, where $\sigma_i = \bar{\sigma}$ and $\sigma_t = 0.8, \ldots, 1.2$; 2) “Homoscedastic Return Over Time”, where $\sigma_i$ is estimated directly from the Fama and French three-factor model and $\sigma_t = 1$; and 3) “Heteroscedastic Return”, where $\sigma_i$ is estimated and $\sigma_t = 0.8, \ldots, 1.2$. Our simulation is carried out with 1000 replications.

Table 2 also reports other characteristics, such as the average explanatory power ($R^2$). Over the six subsample periods, $R^2$ was highest in the early 1970s for the first factor (more than 29%), but gradually decreased to 13% in the late 1990s. This is consistent with Campbell, Lettau, Malkiel, and Xu’s (2001) finding of deceasing explanatory power in the recent decade. It is also true that the second or the third factors only have less than 3% of the explanatory power in some
of the subsample periods. Thus, finding factors with greater explanatory power is important.

### 3.2.2 Comparing Different Approaches

We compare the relative performance of the MEC approach to both the Connor and KORajczyk’s (1986) approach (CK) and the heteroscedasticity consistent estimator (HFA) proposed by Jones (2001) through simulation. The CK approach is a PC analysis on the matrix \( \frac{1}{N-1} \tilde{R}' \tilde{R} \); while the HFA approach modifies the CK approach using factor analysis. Unlike the CK approach which assumes homoscedastic idiosyncratic returns over time, HFA estimates idiosyncratic volatilities at each point in time as the corresponding elements in the \( T \times T \) diagonal matrix \( D \) from the matrix \( \frac{1}{N-1} \tilde{R}' \tilde{R} \) using an iterative maximum likelihood approach. In particular, we follow Jones’ (2001) iterative procedure to estimate the matrix \( D \). Once \( \hat{D} \) is obtained, factors are re-estimated as the principal components from matrix \( \frac{1}{N-1} \tilde{R}' \tilde{R} - \hat{D} \) in order for the factors to have a comparable rotation as those of other methods.\(^{19}\)

We first examine the overall explanatory power of different sets of extracted factors. In each subsample period, we run GLS regressions of generated stock returns on various sets of extracted factors for each stock. The corresponding average \( R^2 \)s are reported in Table 3. Under each return structure, all three approaches produce very similar \( R^2 \)s with the MEC approach being the best. The \( R^2 \)s using MEC factors are even slightly larger than those using true factors since MEC factors also explain some of the idiosyncratic returns in a finite sample. Therefore, similar to portfolios, the differences in performance should not be measured by average \( R^2 \)s for individual stocks.

Perhaps, it is more important to evaluate the accuracy of factor estimates themselves from each approach. However, the difficulty lies in the fact that all extracted factors represent a rotation of the true factors. Therefore, we reconstruct a comparable estimator condition on the

\(^{19}\)Since maximum likelihood approach is used in HFA to extract \( D \) from \( C = F'F + D \), it has a similar structure to a standard factor analysis. One way to estimate \( F \) in the factor analysis is the principal components approach performed on the matrix \( (C - D) \) when the true \( D \) is known. Therefore, the approach used here is justified as long as \( \frac{1}{N-1} \tilde{R}' \tilde{R} \) approaches to \( C \) in a large sample.
true factor loadings,

\[
\tilde{R}_i = \tilde{B}_i \tilde{F} + e_i = \tilde{R}_i \hat{H}_K \hat{H}_K' + e_i
\]

\[
= (B_i F + \epsilon_i) \hat{H}_K \hat{H}_K' + e_i = B_i (F \hat{H}_K \hat{H}_K') + (e_i + \epsilon_i \hat{H}_K \hat{H}_K')
\]

(20)

where \( \hat{H}_K \) is a matrix that consists of the first \( K \) eigenvectors in the implementation of either of the approaches. The second equality in equation (20) holds because of equations (8) and (9) for MEC factors. Similar results to those of equations (8) and (9) can be obtained for both the CK approach and the HFA approach when eigenvectors are estimated from the matrix \( (\frac{1}{N-1}\tilde{R}'\tilde{R}) \) and the matrix \( (\frac{1}{N-1}\tilde{R}'\tilde{R} - \hat{D}) \), respectively. Therefore, we can define \( \tilde{F} = F \hat{H}_K \hat{H}_K' \) as a pseudo estimator of the un-rotated factors for any of the approaches. In this way, we are still able to compare the root-mean-squared-errors (RMSEs) for different methods. These results are shown in Table 3.

[ Insert Table 3 here ]

When residual returns are heteroscedastic across individual assets only, both HFA factors and CK factors are equivalent by construction. Indeed, the pseudo CK factors virtually yield the same RMSEs as those of the pseudo HFA factors.\(^{20}\) In contrast, all three pseudo MEC factors outperform the corresponding pseudo CK or HFA factors by 15% to 43% in RMSEs. The largest differences occur in the most recent two subsample periods. Relatively speaking, the RMSEs of the first pseudo factor is greater those of the second or third under any approaches. This is because the first Fama and French factor (market factor) plays a much important role than the other two factors in time-series regression. Since the extracted factors are rotations of these pseudo factors, the estimated MEC factors are much more accurate in representing the true underlying factors than the other two approaches.

For the case of increasing idiosyncratic volatilities over a six-year period by 50% but homoscedastic across individual stocks, the pseudo HFA factors should be the best among the

\(^{20}\)These tiny differences between the CK approach and the HFA approach in RMSEs are due to the fact that we are using the estimated \( D \) matrix not the theoretical one.
three approaches by construction. However, the differences among the first pseudo factors using different approaches are minimal (at most 3%). There are some advantages of the second and the third pseudo $HFA$ factors over others in terms of RMSEs. Although this case is highly unlikely, one can consider it as a worst case scenario for the $MEC$ factors. In fact, there is virtually no difference among the three approaches if idiosyncratic volatilities only increase by 25%.

A more practical case is that idiosyncratic volatilities are both increasing over time and varying across individual stocks. As shown in Table 3, the general pattern in the RMSEs is very similar to that of the first case above, namely, the pseudo $MEC$ factors have a significant lead over both the pseudo $HFA$ factors and the pseudo $CK$ factors. The additional heteroscedasticity over time seems to have much more impact on the RMSEs of the pseudo $CK$ factors than to those of the $MEC$ factors. This suggests some concern for the $CK$ approach in extracting factors from individual stock returns when heteroscedasticity exists in both dimensions. In contrast, the $MEC$ approach is most effective in handling heteroscedasticity of individual stock returns.

### 3.2.3 Estimating the Number of Factors from Simulated Returns

The APT theory provides no guidance on the determination of the number of underlying pricing factors. It is left as an empirical issue to a large extent. Recently, Bai and Ng (2002) have proposed using an information criterion to determine the number of factors. In this section, we study the empirical performance of various factor selecting criteria discussed in Section 2 under the simulation design of the last subsection. In particular, we first apply the $C_p$ criterion of equation (13) to $CK$ factors and $HFA$ factors. We then compare them to the $C_p^{MEC}$ criterion of equation (14), the eigenvalue criterion, and the $t$-Ratio criterion. We study a single factor structure, a two-factor structure, and a three-factor structure.

21 An incomplete list includes, Connor and Korajczyk (1993), Lehman and Modest (1988), MacKinlay and Richardson (1991), Mei (1993), Shanken (1982), and so on.
When idiosyncratic volatilities are only different across individual stocks, the $C_p$ criterion using either CK factors or HFA factors (denoted as $C_{p}^{CK}$ and $C_{p}^{HFA}$ respectively) accurately estimates the true number of factors in the first three subsample periods for either a one-factor model or a two-factor model. However, as shown in Table 4, these methods have substantially underestimated the number of factors for the recent three subsample periods. In contrast, the MEC factor based $C_{p}^{MEC}$ criterion has at most underestimated the number of factors by 10% for the two-factor model. For a three-factor model, both the CK and the HFA factor based approaches performed much worse than the MEC factor based approach, although the MEC approach also underestimate the number of factors by a half to one factor over different subsample periods.

For the case of increasing idiosyncratic volatilities over time only, the MEC factor-based approach correctly estimates the number of factors for all three models during the first two subsample periods, while other two approaches underestimate the number of factors. For the rest of the subsample periods, the results are mixed. Although this unrealistic case is ideal for the HFA factor-based approach, it does not have a clear advantage over the MEC factor-based approach. When returns exhibit both cross-sectional and time series heteroscedasticity, the results are very similar to those in the first case. Therefore, for a realistic return structure, the MEC factor-based $C_{p}^{MEC}$ criterion can estimate the number factor more consistently than both the $C_{p}^{CK}$ and the $C_{p}^{HFA}$ criteria. In general, it is difficult to distinguish high order factors when they offer less than 3% of additional explanatory power in terms of $R^2$.

It is also interesting to note that when the overall idiosyncratic volatility decreases by 30%, the $C_p$ criterion correctly estimates the true number of factors no matter which factor estimates are used. This is equivalent to an increase in the average explanatory power of individual assets from 25% to 47.5%. In other words, both the $C_{p}^{CK}$ and the $C_{p}^{HFA}$ criteria can work well in determining the number of factors for portfolios rather than individual stocks, while the $C_{p}^{MEC}$ criterion is effective for both portfolios and individual stocks. Different from the implementation
of the other two approaches that requires running regressions for individual stocks, the $C_p^{MEC}$ criterion depends only on eigenvalues with no additional computation.

The Eigenvalue Criterion (A1) and the $t$–Ratio Criterion (A2) are also applied to determine the number of factors. As discussed before, we choose $t^2_{\nu} = 2.8$ in the eigenvalue criterion and $\gamma = 30\%$ in the $t$–ratio criterion. Results from Table 4 suggest that both criteria are surprisingly effective and truthfully estimate the number of underlying factors no matter which return structure is assumed. Therefore, these ad hoc approaches are useful and can be used in conjunction with the information criterion.
4 Out-of-Sample Performance for Actual Stock Returns

In theory, each approach is designed to be optimal in-sample under its own assumptions. One way to assess the relative performance of different approaches is the simulation approach implemented under the same plausible scenarios. However, simulation approaches do not address the stability issue of the extracted factors when used out-of-sample. When the extracted factors are heavily influenced by idiosyncratic volatilities, the out-of-sample performance is expected to be poor since idiosyncratic volatilities change frequently. In this section, we investigate the significance of controlling for idiosyncratic volatilities in factor extraction from an out-of-sample perspective. In particular, we compare different approaches in terms of consistency across different samples of stocks and the explanatory power in the subsequent sample periods.

For the same reasons discussed in Section 3.2, we divided the whole sample period from 1964 to 1999 into six six-year subsample periods. During any subsample period, NYSE/AMEX stocks on the CRSP tape are selected into our sample if it has continuous monthly return records over the six-year period. Since most empirical studies have suggested that there are less than six factors, we only extract six factors using the HFA, the CK, and the MEC approaches.

4.1 Cross-validation Analysis

If a particular method can estimate the true underlying factors accurately, the extracted factors using one set of stocks should be consistent with those using a different set of stocks over the same time period. This is the idea of cross-validation used in Jones (2001). In particular, we randomly divide the whole sample of stocks into two groups with equal numbers. For a specific approach, we then regress each of the six extracted factor returns from the first group of stocks on all the six factor returns extracted from the second group of stocks. In order to avoid selection bias, this random grouping exercise is repeated 100 times. The average $R^2$s from such regressions

22 For the same reason stated in section 2.2 of Jones (2001), we ignore the case of missing observations.
are reported in each panel (that corresponds to each subsample period) of Figure 2.  

[ Insert Figure 2 ]

Several interesting observations can be made from Figure 2. First of all, no matter which approach is used, $R^2$s for the first extracted factor from cross-validation regressions are very close to one for all three approaches. Thus, the first factor is very stable independent of the methods. For higher order factors, however, the MEC approach consistently outperforms the other two approaches. For example, the solid line is almost always above the other two lines by a large margin. In fact, the $R^2$s for the second MEC factor are all above 95% except for the first subsample period. In contrast, the $R^2$s for the second HFA factor are between 55% and 92%, while those from the second CK factor are between 80% and 95%.

Similarly, the $R^2$s for the third MEC factor are above 90% for most subsample periods. The $R^2$s remain high even for the fourth MEC factor. These numbers are much larger than those from the other two approaches. When comparing the HFA factor to the CK factor, the third (or higher order) HFA factor slightly outperforms the CK factor in most of the subsample periods. It is also important to note that the largest differences between MEC factors and HFA (or CK) factors shown in Figure 2 occurred in the two recent subsample periods. Such large differences could be due to the fact that idiosyncratic volatilities have not only increased but also been very different across individual stocks in the recent decade.

4.2 Out-of-sample Performance–A Cross-sectional Perspective

In a time series analysis, a holdout sample is usually used to study the out-of-sample performance. For a cross-sectional comparison, we use a similar idea by dividing the whole sample randomly into two equal numbered subgroups, and use one group as the holdout sample. In particular, during each subsample period, we run OLS regressions for the second group of stocks. Similarly, we can regress factors extracted from the second group of stocks on that of the first group of stocks. The results are very similar to those reported in Figure 2, which means a good random sampling.
against six factors extracted from the first group of stocks only. The average $R^2$s from the regressions and the percentage of stocks with significant beta loading estimates at a 5% level are reported in Table 5.

Among the three approaches, $MEC$ factors produce the highest $R^2$ in any subsample periods with $CK$ and $HFA$ factors being similar. This result is not by design since we are examining the out-of-sample performance. On average, the $R^2$s from using $MEC$ factors are 1.75% and 4% more than those from the other two alternatives in the first four and the last two subsample periods, respectively. Such incremental performance is especially important during periods with decreasing overall explanatory power for the same number of factors. For example, the six $MEC$ factors still offer a 35% explanatory power in the late 1990s, while either $CK$ or $HFA$ factors only provide a 30% explanatory power over the same period.

When a factor is pervasive and the method of extracting such a factor is efficient, the corresponding factor loading estimates should be statistically significant for most individual stocks. As an alternative, we report the percentages of significant loading estimates under different methods in Table 5 over different subsample periods. During the first four subsample periods, the percentages of the significant first factor loading estimates at a 5% level are very close to each other (more than 97%) among the three approaches. However, there are significant differences over the last two subsample periods. For example, the average percentages are 85%, 74%, and 71% for the $MEC$, the $HFA$, and the $CK$ factors, respectively. This evidence further indicates that the $MEC$ approach is especially useful during periods with large idiosyncratic volatilities.

For the second factor, half of the time the $MEC$ approach is the best and half of the time the $CK$ approach is the best. There is no clear dominance among the three methods for the third factor. From the fourth factor on, the $MEC$ factor again dominates, especially for the last three
sub-sample periods. Therefore, \textit{MEC} factors not only have a high out-of-sample explanatory power, but also are more pervasive than factors extracted using alternative approaches.

4.3 Out-of-sample Performance–A Time Series Perspective

We can also directly compare different statistical factors to the Fama and French factors. In this case, we hold out a three-year period in each of the subsample period. In order to be comparable to the Fama and French model, we only extract three factors. During each of the six-year subsample period, we use the first three-year monthly returns to construct the three-factor weights. These weights are then used to compute the corresponding factor returns used in the second three-year period. We then regress individual stock returns on different set of constructed factors including the Fama and French factors in the second three-year period. Panel A of Table 6 reports the average $R^2$'s from these regressions.

[ Insert Table 6 ]

In all the six subsample periods, $R^2$'s for the \textit{MEC} factors not only are higher than those of the \textit{HFA} and the \textit{CK} factors, but also have exceeded those of the Fama and French factors by 0.4% in the early 1960s and by 4% in the late 1990s. Results are mixed for either the \textit{HFA} or the \textit{CK} factors when compared to the Fama and French factors. The \textit{HFA} factors performed even better in recent subsample periods when idiosyncratic volatilities have gone up.

If the factor structure is stable over time at least in short-run, we can reverse the order of above exercise by using the second three-year sample to estimate factor weights that are applied in the construction of factor returns in the first three-year period. Similar results are reported in Panel B of Table 6. Again, the \textit{MEC} factors have outperformed all other factors including the Fama and French factors except for the third subsample period where the results are very close to each other. Therefore, \textit{MEC} factors are more stable and have better out-of-sample properties than other popular factors including the Fama and French factors.
5 Concluding Comments

This paper focuses on the impact of idiosyncratic volatilities on the extracted factors using different approaches. With a simple extension to the standard principal components (PC) analysis, we extract factors by maximizing their explanatory power for individual assets. Although these MEC (Maximum Explanatory Component) factors are equivalent to the PC factors extracted based on a correlation structure of asset returns, they have many advantages over other approaches. The MEC factors not only converge to the true underlying factor structure, but also are better in handling heteroscedasticity both across asset returns and over time in finite samples. Through simulations for both portfolios and individual stocks, we have shown that the MEC approach improves the accuracy of both the factor estimates and the loading estimates over those using other popular approaches. The MEC approach also offers great computational advantages. We provide simple formulas to compute both the loading estimates and their t-ratios for both portfolios and individual stocks without running regressions.

It seems that the APT model has received limited attention in recent years. Meanwhile, the Fama and French’s three factor model has become the standard model in application. This is not because of the interpretation of the three Fama and French factors, rather it is due to their out-of-sample stability. In fact, as Cochrane (2001) has pointed out that both the size and the book-to-market variables can only be proxies for unknown risk factors. Although statistical factors also suffer the drawback of lacking interpretation, the biggest challenge is the out-of-sample performance. Since idiosyncratic volatilities account for majority portions of individual stocks’ total return volatilities, out-of-sample volatilities could be very different from those in-sample. MEC factors improve the out-of-sample performance by reducing the dependence on idiosyncratic volatilities. Empirical evidence from the cross-validation exercise suggests that MEC factors are consistent across randomly selected groups of stocks. Moreover, based on out-of-sample studies, we show that the MEC factors not only have larger explanatory power than the Fama and French factors, but also are more pervasive than other factors especially in
recent sample periods.

We also shed light on the issue of determining the number of pricing factors. The MEC approach allows us to design four different criteria, the information criterion, the naive criterion, the eigenvalue criterion, and the $t$–ratio criterion in estimating the number of factors. We show that these criteria offer good estimates in finite samples.
Appendix A: Proof of Proposition 1

In order to show that the first MEC factor, \( f_1^* = \frac{1}{\sqrt{\lambda_1}} \tilde{R}'V^{-1}a_1 \), maximize the average \( R^2 \), we first examine the correlation vector between the first factor and individual asset returns,

\[
Corr(\tilde{R}, f_1^*) = \sigma_{f_1^*}^{-1}\tilde{V}^{-1}Cov(\tilde{R}, f_1^*) = \frac{1}{\sqrt{\lambda_1}\sigma_{f_1^*}}Cov(V^{-1}\tilde{R}, \tilde{R}'V^{-1}a_1) = \frac{1}{\sqrt{\lambda_1}\sigma_{f_1^*}}\Omega a_1,
\]

(21)

where the factor variance can be written as, \( \sigma_{f_1^*}^2 = \frac{1}{\lambda_1} a_1' V^{-1}Cov(\tilde{R}, \tilde{R}')V^{-1}a_1 = \frac{1}{\lambda_1} a_1' \Omega a_1 \). Basic econometrics theory suggests that the \( R^2 \) from regressing a particular stock returns on the first factor is just the squared correlation between the two. Thus, the average coefficient of determination from similar regressions for all individual assets is,

\[
\bar{R}_1^2 = \frac{1}{N} Corr(\tilde{R}, f_1^*')Corr(\tilde{R}, f_1^*) = \frac{1}{N\lambda_1 \sigma_{f_1^*}^2} a_1' \Omega a_1 = \frac{1}{N} a_1' \Omega a_1.
\]

(22)

To verify that \( \bar{R}_1^2 \) has the maximum value with respect to all the linear factors, \( f_1 = \frac{1}{\sqrt{\lambda_1}} \tilde{R}'V^{-1}x \), we can obtain the average \( R^2 \) for \( f_1 \) in a similar way as,

\[
R^2(x) = \frac{1}{N} x'\Omega'y x.
\]

(23)

The first order condition from maximizing equation (23) with respect to \( x \) can be expressed as,

\[
\frac{\partial R^2(x)}{\partial x'} = \frac{2}{N} \left[ \Omega x - \frac{x'\Omega x}{x'\Omega x} \right] = 0.
\]

(24)

The last equality holds when \( x = a_1 \) is the first eigenvector of the correlation matrix \( \Omega \). By definition, \( \lambda_1 = \frac{a_1' \Omega a_1}{a_1' \Omega a_1} \) is its corresponding eigenvalue. This result and equation (22) suggest that the maximum for the average coefficient of determination, \( \bar{R}_1^2 \), is just \( \frac{1}{N} \lambda_1 \). By construction, the first eigenvalue is the largest. Equation (24) also holds for other eigenvectors. Therefore, the proposition obtains.

Appendix B: Proof of Corollary 2

Since the MEC factors are orthogonal to each other, \( R^2 \) from regressing asset returns on a factor equals the squared correlation between the return and the factor. In other words, we have,

\[
R^2_{i,f_k^*} = \frac{Var(f_{k,t})}{Var(f_{k,t})Var(R_{i,t})} = \left[ \frac{Cov(f_{k,t}, R_{i,t})}{Var(f_{k,t})} \right]^2 \times \frac{Var(f_{k,t})}{Var(R_{i,t})} = (\sigma_{f_{k,t}}^{-1} \beta_{i,k}) \times (\sigma_{f_{k,t}}^{-1} \beta_{i,k}).
\]

(25)

The last equality in equation (25) holds because \( Var(f_{k,t}) = 1 \) by construction. Therefore, equation (6) is just equation (25) in matrix notation. It is also true that the \( R^2 \) from regressing returns on several factors equals the sum of \( R^2 \)s from regressing on individual factors. Thus, we can compute the \( t \)-ratio for the \( k \)-th factor loading of asset \( i \) as,

\[
t^2_{i,f_k^*} = \left( \frac{Cov(f_{k,t}, R_{i,t})}{Var(f_{k,t})} \right)^2 \times \frac{\sigma_{resid,i}^2}{\sigma^2_{f_{k,t}}}.
\]

36
\[
\frac{\text{Cov}^2(f_{k,r}^* \tilde{R}_{i,t})}{\text{Var}(f_{k,r}^*)} \times \frac{T - K}{(1 - R_{i,f_1}^2, \ldots, R_{i,f_{K}}^2) \text{Var}(\tilde{R}_{i,t})} = \frac{\text{Cov}^2(f_{k,r}^* \tilde{R}_{i,t})}{\text{Var}(f_{k,r}^*) \text{Var}(\tilde{R}_{i,t})} \times \frac{T - K}{1 - \sum_{l=1}^{k+1} R_{i,f_l}^2} = (T - K) \frac{R_{i,f_{K+1}}^2}{1 - \sum_{l=1}^{k+1} R_{i,f_l}^2},
\]

Using matrix notation, equation (26) can be rewritten as equation (7).

**Appendix C: Convergence of MEC Factors When the Correlation Structure Is Known**

To demonstrate that the MEC factors also converge to a rotation of the actual (or fundamental) factors that drive asset returns, we write the variance-covariance structure for the true return generating process of equation (1) as,

\[
\Sigma_N = \beta_N \beta_N' + D_N,
\]

where \( N \) denotes the number of assets. Equation (27) holds because of the assumption of \( \text{Cov}(\tilde{F}, \tilde{F}') = I_K \).

Similarly, the true correlation structure can be written as,

\[
\Omega_N = V_N^{-1} \Sigma_N V_N^{-1} = V_N^{-1} \beta_N \beta_N' V_N^{-1} + V_N^{-1} D_N V_N^{-1} = B_N B_N' + G_N,
\]

where \( B_N = V_N^{-1} \beta_N \) and \( G_N = V_N^{-1} D_N V_N^{-1} \). Since \( V_N = [\text{diag}(\Sigma_N)]^{1/2} \) is a positive and bounded diagonal matrix, \( \Omega_N \) is still positive definite. Moreover, the largest eigenvalue of \( G_N \) will be bounded if \( D_N \) is bounded.

The convergence result is established if we can show that a MEC factor-based decomposition converges to equation (28). Since the matrix notation for the MEC factors of equation (2) can be expressed as \( \tilde{F}^* = \Lambda_K^{-1/2} \Lambda_K' R^* \), where \( \Lambda_K \) and \( \Lambda_K \) are the matrices consisting of the first \( K \) eigenvectors and eigenvalues of \( \Omega_N \), respectively, the transformed factor loading \( B_N^* \) can be derived in the following way,

\[
B_N^* = V_N^{-1} \beta_N = V_N^{-1} E[\tilde{R} \tilde{F}'] (E[\tilde{F} \tilde{F}'])^{-1} = E[R' F'] = E[R' R'] A_K A_K^{-1/2} = \Omega_N A_K A_K^{-1/2} = A_K A_K^{-1/2},
\]

The following decomposition of the correlation matrix thus follows,

\[
\Omega_N = B_N^* B_N^* + G_N,
\]

where, \( G_N = A_{N-K} A_{N-K}' A_{N-K} \), \( A_{N-K} \) is an eigenmatrix of \( \Omega_N \) containing the last \( (N - K) \) eigenvectors, and \( A_{N-K} \) is a diagonal matrix containing the last \( (N - K) \) eigenvalues of \( \Omega_N \) as its diagonal elements.

Under the approximate K-factor structure, Chamberlain and Rothschild (1983) have suggested a similar decomposition for the variance-covariance matrix \( \Sigma_N \) based on factors obtained using the PC analysis. Because of the analogy between equations (29) and the decomposition offered in Chamberlain and Rothschild (1983), their convergence Theorem 4 should continue to hold with respect to the decomposition of equation (29). In other words, \( B_N^* \) should converge to \( B_N \), i.e. \( V_N B_N^* \) converges to \( \beta_N \) as the number of assets goes to infinity. Therefore, the “Maximum Explanatory Component” approach allows us to recover the true factor structure up to an orthogonal rotation in the limit.
References


### Table 1: A Numerical Comparison between PC Factors and MEC Factors

This table reports simulation results for portfolios applying both PC factors and MEC factors. We assume that there are 75 * 12 time periods. Portfolio i’s return is generated from a three-factor model, 

\[ R_i = \beta_{1i}f_1 + \beta_{2i}f_2 + \beta_{3i}f_3 + \epsilon_i, \]

where \( f_1, f_2, f_3 \sim iidN(0, 1) \), \( \epsilon_i \sim iidN(0, \kappa_i^2 \sigma_i^2) \), and \( \beta_i^2 = (\beta_{1i}^2 + \beta_{2i}^2 + \beta_{3i}^2) \). The true factor loadings \( \beta_{1i}, \beta_{2i}, \) and \( \beta_{3i} \) are independent and are generated from the distribution of \( N(1, 0.25) \). Under “Random \( R^2 \),” we set \( \kappa_i = 1 \), and generate \( \sigma_i \) according to a uniform distribution of \( U(0.3, 3.0) \). Under “Increasing \( R^2 \),” we set \( \kappa_i = \beta_i \) and \( \sigma_i = \gamma_i \) if the corresponding \( \beta_i \) is at the \( k \)-th rank from sorting \( \beta_i \) in ascending order, where \( \gamma_i \) is linearly decreasing from 1.53 to 0.17. In this case, \( R^2 \) is strictly increasing with \( \beta_i \). Similarly, under “Decreasing \( R^2 \),” we set \( \kappa_i = \beta_i \) and \( \sigma_i = \gamma_i \), where \( \gamma_i \) is linearly increasing from 0.17 to 1.53. Thus, \( R^2 \) increases with \( \beta_i \). The pseudo \( \hat{\beta} \) is defined in equation (18). \( \hat{f} \) is the factor estimate. “RMSE” stands for root mean squared error. “\( \eta \) Ratio” is the percentage of the total variance of a constructed factor being captured by the variance of the true factors. It is reported in percentage.

<table>
<thead>
<tr>
<th># of</th>
<th>Method</th>
<th>RMSE</th>
<th>( \eta ) Ratio</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \beta_1 )</td>
<td>( \beta_2 )</td>
<td>( \beta_3 )</td>
</tr>
<tr>
<td>20</td>
<td>PC</td>
<td>0.229</td>
<td>0.227</td>
<td>0.230</td>
</tr>
<tr>
<td></td>
<td>MEC</td>
<td>0.148</td>
<td>0.146</td>
<td>0.147</td>
</tr>
<tr>
<td>50</td>
<td>PC</td>
<td>0.119</td>
<td>0.118</td>
<td>0.119</td>
</tr>
<tr>
<td></td>
<td>MEC</td>
<td>0.081</td>
<td>0.082</td>
<td>0.082</td>
</tr>
<tr>
<td>100</td>
<td>PC</td>
<td>0.080</td>
<td>0.081</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>MEC</td>
<td>0.067</td>
<td>0.068</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>PC</td>
<td>0.109</td>
<td>0.109</td>
<td>0.110</td>
</tr>
<tr>
<td></td>
<td>MEC</td>
<td>0.115</td>
<td>0.115</td>
<td>0.117</td>
</tr>
<tr>
<td>100</td>
<td>PC</td>
<td>0.065</td>
<td>0.065</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>MEC</td>
<td>0.067</td>
<td>0.067</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>PC</td>
<td>0.308</td>
<td>0.311</td>
<td>0.307</td>
</tr>
<tr>
<td></td>
<td>MEC</td>
<td>0.139</td>
<td>0.140</td>
<td>0.139</td>
</tr>
<tr>
<td>50</td>
<td>PC</td>
<td>0.260</td>
<td>0.255</td>
<td>0.256</td>
</tr>
<tr>
<td></td>
<td>MEC</td>
<td>0.087</td>
<td>0.087</td>
<td>0.088</td>
</tr>
<tr>
<td>100</td>
<td>PC</td>
<td>0.151</td>
<td>0.150</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>MEC</td>
<td>0.076</td>
<td>0.076</td>
<td>0.076</td>
</tr>
</tbody>
</table>
Table 2: **Subsample Summary Statistics**

This table reports the number of stocks, the coefficient of determination, the mean and standard deviation of the Fama and French’s factor loadings, and the volatility distribution of individual stocks for each of the six subsample periods from 1964 to 1999. Each NYSE/AMEX stock is selected into our sample if it survives over the corresponding subsample period.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td># of Stk</td>
<td>1430</td>
<td>1833</td>
<td>1612</td>
<td>1560</td>
<td>1700</td>
<td>2044</td>
</tr>
</tbody>
</table>

**Panel A: Coefficient of determination**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Factor</td>
<td>22.82</td>
<td>29.26</td>
<td>28.24</td>
<td>28.73</td>
<td>16.73</td>
<td>12.89</td>
</tr>
<tr>
<td>2 Factors</td>
<td>28.17</td>
<td>38.36</td>
<td>33.00</td>
<td>32.46</td>
<td>21.35</td>
<td>16.62</td>
</tr>
<tr>
<td>3 Factors</td>
<td>30.49</td>
<td>41.53</td>
<td>36.16</td>
<td>34.11</td>
<td>23.95</td>
<td>21.36</td>
</tr>
</tbody>
</table>

**Panel B: Mean of loadings**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Fct.</td>
<td>1.069</td>
<td>1.132</td>
<td>1.126</td>
<td>1.110</td>
<td>1.033</td>
<td>0.8795</td>
</tr>
<tr>
<td>2nd Fct.</td>
<td>0.8835</td>
<td>0.9551</td>
<td>0.6976</td>
<td>0.6378</td>
<td>0.5986</td>
<td>0.4158</td>
</tr>
<tr>
<td>3rd Fct.</td>
<td>0.3041</td>
<td>0.3635</td>
<td>0.3354</td>
<td>0.1044</td>
<td>0.2971</td>
<td>0.5241</td>
</tr>
</tbody>
</table>

**Panel C: Standard deviation of loadings**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Fct.</td>
<td>0.5177</td>
<td>0.4141</td>
<td>0.4845</td>
<td>0.4176</td>
<td>0.5955</td>
<td>0.5330</td>
</tr>
<tr>
<td>2nd Fct.</td>
<td>0.9854</td>
<td>0.8874</td>
<td>0.8853</td>
<td>0.8258</td>
<td>0.9453</td>
<td>0.6573</td>
</tr>
<tr>
<td>3rd Fct.</td>
<td>0.7320</td>
<td>0.6781</td>
<td>0.6281</td>
<td>0.6406</td>
<td>0.8088</td>
<td>0.7007</td>
</tr>
</tbody>
</table>

**Panel D: Distribution for volatilities of individual stocks**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>0.053</td>
<td>0.076</td>
<td>0.058</td>
<td>0.063</td>
<td>0.049</td>
<td>0.035</td>
</tr>
<tr>
<td>Median</td>
<td>0.092</td>
<td>0.120</td>
<td>0.098</td>
<td>0.101</td>
<td>0.089</td>
<td>0.086</td>
</tr>
<tr>
<td>90%</td>
<td>0.165</td>
<td>0.181</td>
<td>0.161</td>
<td>0.160</td>
<td>0.172</td>
<td>0.158</td>
</tr>
</tbody>
</table>
Table 3: Root Mean Squared Error for the Estimated Pseudo Factor Returns

This table reports root mean squared errors (RMSEs) of the extracted pseudo factor returns (see equation (20)) extracted from simulated individual stock returns. “HFA” denotes that factors are estimated using heteroscedasticity consistent estimator of Jones (2001); “CK” denotes that factors are estimated using the Connor and Korajczyk (1986) method; and Fct. 1 stands for the first factor, and so on. Individual stock returns used in simulation are generated according to Fama and French’s (1993) three-factor model. “Homo. Returns Across Stocks” means that the idiosyncratic volatility is assume to be constant across individual stocks and increases 50% gradually over time, “Homo. Returns Over Time” means that the idiosyncratic volatility is the residual variance of the three-factor model applied to actual returns, and “Heteroscedastic Returns” means that the idiosyncratic volatility is the residual variance of the three-factor model and increases 50% gradually over time.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fct.1</td>
<td>Fct.2</td>
<td>Fct.3</td>
</tr>
<tr>
<td>True</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1964-1969</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFA</td>
<td>0.5340</td>
<td>0.5223</td>
<td>0.6328</td>
</tr>
<tr>
<td>C.K.</td>
<td>0.5328</td>
<td>0.5210</td>
<td>0.6299</td>
</tr>
<tr>
<td>MEC</td>
<td>0.4492</td>
<td>0.4048</td>
<td>0.4333</td>
</tr>
<tr>
<td>1970-1975</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>40.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFA</td>
<td>0.7043</td>
<td>0.5680</td>
<td>0.6104</td>
</tr>
<tr>
<td>C.K.</td>
<td>0.7039</td>
<td>0.5672</td>
<td>0.6086</td>
</tr>
<tr>
<td>MEC</td>
<td>0.6687</td>
<td>0.5136</td>
<td>0.5004</td>
</tr>
<tr>
<td>1976-1981</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>35.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFA</td>
<td>0.6215</td>
<td>0.4873</td>
<td>0.6025</td>
</tr>
<tr>
<td>C.K.</td>
<td>0.6208</td>
<td>0.4860</td>
<td>0.6003</td>
</tr>
<tr>
<td>MEC</td>
<td>0.5819</td>
<td>0.4038</td>
<td>0.4693</td>
</tr>
<tr>
<td>1982-1987</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>37.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFA</td>
<td>0.7180</td>
<td>0.5908</td>
<td>0.6987</td>
</tr>
<tr>
<td>C.K.</td>
<td>0.7171</td>
<td>0.5884</td>
<td>0.6950</td>
</tr>
<tr>
<td>MEC</td>
<td>0.6608</td>
<td>0.4311</td>
<td>0.4834</td>
</tr>
<tr>
<td>1988-1993</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>27.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFA</td>
<td>0.5857</td>
<td>0.7439</td>
<td>0.8016</td>
</tr>
<tr>
<td>C.K.</td>
<td>0.5842</td>
<td>0.7415</td>
<td>0.7988</td>
</tr>
<tr>
<td>MEC</td>
<td>0.4414</td>
<td>0.4165</td>
<td>0.4522</td>
</tr>
<tr>
<td>1994-1999</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>27.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFA</td>
<td>0.6831</td>
<td>0.7744</td>
<td>0.6474</td>
</tr>
<tr>
<td>C.K.</td>
<td>0.6814</td>
<td>0.7713</td>
<td>0.6455</td>
</tr>
<tr>
<td>MEC</td>
<td>0.5669</td>
<td>0.5153</td>
<td>0.4391</td>
</tr>
</tbody>
</table>
This table reports the estimates of the number of factors when factors are estimated using heteroscedasticity consistent estimator of Jones (2001) (HFA), the Connor and Korajczyk (1986) (CK) approach and the MEC approach on simulated individual stock returns. These return data are generated according to Fama and French's (1993) three-factor model with different volatility structure. “Homo. Returns Across Stocks” means that the idiosyncratic volatility is assume to be constant across individual stocks and increases 50% gradually over time, “Homo. Ret. Over Time” means that the idiosyncratic volatility is the residual variance of the three-factor model applied to actual returns, and “Heteroscedastic Returns” means that the idiosyncratic volatility is the residual variance of the three-factor model and increases 50% gradually over time. “A1” and “A2” stand for the eigenvalue criterion and the t−Ratio criterion described in Section 3, respectively.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K=1  K=2  K=3</td>
<td>K=1  K=2  K=3</td>
<td>K=1  K=2  K=3</td>
</tr>
<tr>
<td>With HFA</td>
<td>1.000 1.982 2.040</td>
<td>1.578 1.847 2.473</td>
<td>1.000 1.983 2.044</td>
</tr>
<tr>
<td>With CK</td>
<td>1.000 1.982 2.040</td>
<td>1.578 1.847 2.473</td>
<td>1.000 1.988 2.055</td>
</tr>
<tr>
<td>With MEC</td>
<td>1.000 1.999 2.332</td>
<td>1.000 2.000 3.000</td>
<td>1.000 1.998 2.389</td>
</tr>
<tr>
<td>With A1</td>
<td>1.000 2.000 3.000</td>
<td>1.000 2.000 3.000</td>
<td>1.000 2.000 3.000</td>
</tr>
<tr>
<td>With A2</td>
<td>1.000 2.000 2.999</td>
<td>1.000 2.000 3.000</td>
<td>1.000 2.000 2.999</td>
</tr>
<tr>
<td>With HFA</td>
<td>1.000 2.000 2.408</td>
<td>1.558 2.131 2.922</td>
<td>1.000 1.999 2.424</td>
</tr>
<tr>
<td>With CK</td>
<td>1.000 2.000 2.408</td>
<td>1.558 2.131 2.922</td>
<td>1.000 2.000 2.464</td>
</tr>
<tr>
<td>With MEC</td>
<td>1.000 2.000 2.646</td>
<td>1.000 2.000 3.000</td>
<td>1.000 2.000 2.699</td>
</tr>
<tr>
<td>With A1</td>
<td>1.000 2.000 2.994</td>
<td>1.000 2.000 3.000</td>
<td>1.000 2.000 2.994</td>
</tr>
<tr>
<td>With A2</td>
<td>1.000 2.000 2.999</td>
<td>1.000 2.000 3.000</td>
<td>1.000 2.000 3.000</td>
</tr>
<tr>
<td>With HFA</td>
<td>1.000 1.940 2.476</td>
<td>1.000 1.970 2.639</td>
<td>1.000 1.944 2.498</td>
</tr>
<tr>
<td>With CK</td>
<td>1.000 1.940 2.476</td>
<td>1.000 1.974 2.678</td>
<td>1.000 1.949 2.530</td>
</tr>
<tr>
<td>With MEC</td>
<td>1.000 1.884 2.506</td>
<td>1.000 1.764 2.069</td>
<td>1.000 1.908 2.565</td>
</tr>
<tr>
<td>With A1</td>
<td>1.000 2.000 2.999</td>
<td>1.000 2.000 3.000</td>
<td>1.000 2.000 2.999</td>
</tr>
<tr>
<td>With A2</td>
<td>1.000 2.000 3.000</td>
<td>1.000 2.000 3.000</td>
<td>1.000 2.000 3.000</td>
</tr>
<tr>
<td>With HFA</td>
<td>1.000 1.356 1.487</td>
<td>1.000 1.250 1.352</td>
<td>1.000 1.382 1.510</td>
</tr>
<tr>
<td>With CK</td>
<td>1.000 1.356 1.487</td>
<td>1.000 1.301 1.404</td>
<td>1.000 1.409 1.549</td>
</tr>
<tr>
<td>With MEC</td>
<td>1.000 1.773 1.826</td>
<td>1.000 1.070 1.094</td>
<td>1.000 1.826 1.864</td>
</tr>
<tr>
<td>With A1</td>
<td>1.000 2.000 2.832</td>
<td>1.000 1.988 2.450</td>
<td>1.000 2.000 2.841</td>
</tr>
<tr>
<td>With A2</td>
<td>1.000 2.000 2.991</td>
<td>1.000 1.939 2.213</td>
<td>1.000 1.999 2.924</td>
</tr>
<tr>
<td>With HFA</td>
<td>1.000 1.652 1.941</td>
<td>1.000 1.992 2.506</td>
<td>1.000 1.672 1.956</td>
</tr>
<tr>
<td>With CK</td>
<td>1.000 1.652 1.941</td>
<td>1.000 1.995 2.553</td>
<td>1.000 1.692 1.966</td>
</tr>
<tr>
<td>With MEC</td>
<td>1.000 1.973 2.081</td>
<td>1.000 1.906 2.053</td>
<td>1.000 1.978 2.118</td>
</tr>
<tr>
<td>With A1</td>
<td>1.000 2.000 2.995</td>
<td>1.000 2.000 2.998</td>
<td>1.000 2.000 2.996</td>
</tr>
<tr>
<td>With A2</td>
<td>1.000 2.000 2.965</td>
<td>1.000 2.000 2.959</td>
<td>1.000 2.000 2.968</td>
</tr>
<tr>
<td>With HFA</td>
<td>1.000 1.308 1.769</td>
<td>1.000 1.575 2.149</td>
<td>1.000 1.344 1.793</td>
</tr>
<tr>
<td>With CK</td>
<td>1.000 1.308 1.769</td>
<td>1.000 1.627 2.219</td>
<td>1.000 1.362 1.830</td>
</tr>
<tr>
<td>With MEC</td>
<td>1.000 1.791 2.139</td>
<td>1.000 1.150 1.415</td>
<td>1.000 1.838 2.201</td>
</tr>
<tr>
<td>With A1</td>
<td>1.000 2.000 3.000</td>
<td>1.000 2.000 3.000</td>
<td>1.000 2.000 3.000</td>
</tr>
<tr>
<td>With A2</td>
<td>1.000 2.000 2.994</td>
<td>1.000 1.969 2.923</td>
<td>1.000 2.000 2.995</td>
</tr>
</tbody>
</table>
Table 5: Out-of-Sample Performance for the NYSE/AMEX Stocks
This table reports the out-of-sample performance of each method using the actual NYSE/AMEX stock returns for the six subsample periods from 1964-1999. We use the coefficient of determination and the percentage of significant loading estimates (at a 5% level) from regressing each stock returns in the first randomly selected group against factors extracted from stocks returns of the second randomly selected group. “HFA” denotes that factors are estimated using heteroscedasticity consistent estimator of Jones (2001), and “C.K.” denotes that the factors are estimated using the Connor and Korajczyk (1986) method. Individual stocks that are survived in a subsample period are included in the analysis.

<table>
<thead>
<tr>
<th>Approach</th>
<th>% of stocks with significant loading estimates</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$f_1$</td>
<td>$f_2$</td>
</tr>
<tr>
<td>1964-1969</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFA</td>
<td>97.65</td>
<td>23.56</td>
</tr>
<tr>
<td>C.K.</td>
<td>96.87</td>
<td>26.34</td>
</tr>
<tr>
<td>MEC</td>
<td><strong>97.70</strong></td>
<td>25.90</td>
</tr>
<tr>
<td>1970-1975</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFA</td>
<td>89.75</td>
<td>33.71</td>
</tr>
<tr>
<td>C.K.</td>
<td>98.92</td>
<td><strong>41.87</strong></td>
</tr>
<tr>
<td>MEC</td>
<td><strong>99.24</strong></td>
<td>41.30</td>
</tr>
<tr>
<td>1976-1981</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFA</td>
<td>99.33</td>
<td>28.60</td>
</tr>
<tr>
<td>C.K.</td>
<td>99.46</td>
<td>30.92</td>
</tr>
<tr>
<td>MEC</td>
<td><strong>99.51</strong></td>
<td>34.19</td>
</tr>
<tr>
<td>1982-1987</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFA</td>
<td>96.13</td>
<td>27.13</td>
</tr>
<tr>
<td>C.K.</td>
<td>96.95</td>
<td>33.71</td>
</tr>
<tr>
<td>MEC</td>
<td><strong>97.67</strong></td>
<td>33.81</td>
</tr>
<tr>
<td>1988-1993</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFA</td>
<td>73.72</td>
<td>33.57</td>
</tr>
<tr>
<td>C.K.</td>
<td>78.50</td>
<td><strong>39.62</strong></td>
</tr>
<tr>
<td>MEC</td>
<td><strong>83.18</strong></td>
<td>34.17</td>
</tr>
<tr>
<td>1994-1999</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFA</td>
<td>71.42</td>
<td>27.07</td>
</tr>
<tr>
<td>C.K.</td>
<td>73.99</td>
<td>24.41</td>
</tr>
<tr>
<td>MEC</td>
<td><strong>84.77</strong></td>
<td><strong>36.00</strong></td>
</tr>
</tbody>
</table>
Table 6: Out-of-Sample Comparison with the Fama and French Factors

This table reports the out-of-sample performance of different factors relative to the Fama and French’s (1993) factors using the actual NYSE/AMEX stock returns. In each of the six-year subsample period, we use the first three years of data to construct the three-factor weights using different factor extracting methods. These weights are then used to compute the factor returns used in the second three-year period. $R^2$s are obtained from running regressions of individual stock returns on different set of constructed factors including the Fama and French factors as a benchmark. These results are reported in Panel A. Panel B is similar to Panel A except that we use the second three-year returns to estimate the factor weights used to compute factor returns in the first three years, and compute the $R^2$s accordingly for each subsample period. “FF” denotes that the Fama and French’s (1993) three factors are used, “HFA” denotes that factors are estimated using heteroscedasticity consistent estimator of Jones (2001), and “C.K.” denotes that the factors are estimated using the Connor and Korajczyk (1986) method. $\Delta R^2$ is the difference in $R^2$s between the current three-factor model and the Fama and French’s three-factor model. Individual stocks that are survived in a subsample period are included in the analysis.

<table>
<thead>
<tr>
<th>Period</th>
<th>FF</th>
<th>HFA</th>
<th>CK</th>
<th>MEC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R^2$</td>
<td>$R^2$</td>
<td>$\Delta R^2$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>Period</td>
<td>$R^2$</td>
<td>$R^2$</td>
<td>$\Delta R^2$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>Panel A: Use the first 3-year data to construct factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1964-1966</td>
<td>38.55</td>
<td>38.40</td>
<td>-0.14</td>
<td>37.51</td>
</tr>
<tr>
<td>1970-1972</td>
<td>47.94</td>
<td>46.74</td>
<td>-1.21</td>
<td>46.70</td>
</tr>
<tr>
<td>1976-1978</td>
<td>37.81</td>
<td>39.22</td>
<td>1.41</td>
<td>37.04</td>
</tr>
<tr>
<td>1982-1984</td>
<td>43.37</td>
<td>44.90</td>
<td>1.53</td>
<td>44.86</td>
</tr>
<tr>
<td>1988-1990</td>
<td>24.84</td>
<td>23.34</td>
<td>-0.50</td>
<td>20.23</td>
</tr>
<tr>
<td>1994-1996</td>
<td>27.02</td>
<td>28.66</td>
<td>1.64</td>
<td>29.12</td>
</tr>
<tr>
<td>Average</td>
<td>36.25</td>
<td>36.88</td>
<td>0.62</td>
<td>35.91</td>
</tr>
<tr>
<td>Panel B: Use the second 3-year data to construct factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1973-1975</td>
<td>40.67</td>
<td>40.58</td>
<td>-0.10</td>
<td>40.79</td>
</tr>
<tr>
<td>1979-1981</td>
<td>42.96</td>
<td>41.05</td>
<td>-1.91</td>
<td>42.42</td>
</tr>
<tr>
<td>1985-1987</td>
<td>31.44</td>
<td>33.73</td>
<td>2.29</td>
<td>33.39</td>
</tr>
<tr>
<td>1991-1993</td>
<td>32.10</td>
<td>31.01</td>
<td>-1.09</td>
<td>30.56</td>
</tr>
<tr>
<td>1997-1999</td>
<td>20.88</td>
<td>21.28</td>
<td>0.40</td>
<td>20.98</td>
</tr>
<tr>
<td>Average</td>
<td>32.82</td>
<td>32.44</td>
<td>-0.38</td>
<td>32.50</td>
</tr>
</tbody>
</table>
Figure 1 Relationship Between Estimated Errors and True Betas
This graph shows the relationship between estimation errors of the first factor and the corresponding true betas. We assume that there are 20 portfolios with returns generated from a three factor model. Each factor returns are generated independently from $N(0, 1)$ with 900 observations. The three factor loadings are drawn independently from $N(1, 0.25)$. The idiosyncratic volatility of each portfolio is chosen randomly so that the average $R^2$ is equal to 50%. 

Panel A: The Relationship Between Estimation Errors Using PC Factors and True Betas

Panel B: The Relationship Between Estimation Errors Using MEC Factors and True Betas
Figure 2 Cross-Validation
This graph shows the average coefficients of determination of regressing each of the six factor returns extracted from the first randomly selected sample of stocks on all the six factor returns extracted from the second group of non-overlapping randomly selected stocks. This exercise is repeated 100 times for each subsample period using three different factor extraction methods.