

NBER WORKING PAPER SERIES

DIVERSIFICATION AND THE OPTIMAL
CONSTRUCTION OF BASIS PORTFOLIOS

Bruce N. Lehmann
David M. Modest

Working Paper 9461
<http://www.nber.org/papers/w9461>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2003

We are grateful to the Columbia Business School for providing us with the computational resources to carry out this analysis, and to Arthur Warga for helpful comments. The research reported here is part of the NBER's research program in Asset Pricing. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research or Morgan Stanley. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

©2003 by Bruce N. Lehmann and David M. Modest. All rights reserved. Short sections of text not to exceed two paragraphs, may be quoted without explicit permission provided that full credit including notice, is given to the source.

Diversification and the Optimal Construction of Basis Portfolios
Bruce N. Lehmann and David M. Modest
NBER Working Paper No. 9461
January 2003
JEL No. G1

ABSTRACT

Nontrivial diversification possibilities arise when a factor model describes security returns. In this paper, we provide a comprehensive examination of the merits of various strategies for constructing basis portfolios that are, in principle, highly correlated with the common factors underlying security returns. Three main conclusions emerge from our study. First, increasing the number of securities included in the analysis dramatically improves basis portfolio performance. Our results indicate that factor models involving 750 securities provide markedly superior performance to those involving 30 or 250 securities. Second, comparatively efficient estimation procedures such as maximum likelihood and restricted maximum likelihood factor analysis (which imposes the APT mean restriction) significantly outperform the less efficient instrumental variables and principal components procedures that have been proposed in the literature. Third, a variant of the usual Fama-MacBeth portfolio formation procedure, which we call the minimum idiosyncratic risk portfolio formation procedure, outperformed the Fama-MacBeth procedure and proved equal to or better than more expensive quadratic programming procedures.

Bruce N. Lehmann
University of California, San Diego
IR/PS
9500 Gilman Drive
La Jolla, CA 92093-0519
and NBER
blehmann@ucsd.edu

David M. Modest
Morgan Stanley

One of the basic tenets of investment management is that diversification pays: risk can be reduced by spreading investments across a large number of imperfectly correlated assets. The merits of diversification are particularly clear in linear factor models for equity returns. These models assume returns are generated by a small number of common factors plus an additional random component that can be diversified away in large portfolios. Linear factor models form the basis of the arbitrage pricing theory (APT) of Ross (1976, 1977). They are also widely used in quantitative investment management for a variety of purposes including risk management, alpha detection, and performance evaluation.

One agnostic and prevalent approach to implementing factor models is to treat the underlying common factors as unobserved random variables, and to infer them from the covariance structure of returns. In this approach, the factor model parameters of individual securities are estimated and then used to form *basis* or *reference* portfolios to mimic the common factors. Cross-sectional regression based on factor sensitivity and idiosyncratic risk estimates remains the most widely used portfolio formation procedure. In theory, these *basis portfolio* returns will be both highly correlated with the common factors and relatively free of unsystematic risk. In practice, there is no guarantee that a basis portfolio formation procedure based on a finite sample will mimic the factors sufficiently well. Performance differences might be expected across portfolio construction and factor model estimation procedures and might well vary with the number of securities included and/or the number of factors being considered. It is clearly important to know which methods do a good job of mimicking the common factors and which do not.

Which strategy is best? This is an empirical question that has not been previously addressed. In this paper, we remedy this omission by providing a comprehensive examination of

different basis portfolio formation strategies. In particular, we provide a detailed analysis of the performance of variants of all of the portfolio formation procedures and estimation methods that have been proposed in the literature, as well as an examination of the performance gains in cross-sections of up to 750 securities.

The paper is organized as follows. The next section briefly reviews multifactor models of asset returns and their implications for the behavior of diversified portfolios. The third section delineates the different portfolio formation methods considered while the fourth describes the estimation methods used to generate the inputs to these procedures. The fifth section analyzes the problem of basis portfolio performance evaluation. The penultimate section empirically compares and contrasts different estimation methods, different portfolio formation procedures, and different numbers of securities. The final section provides some concluding remarks.

1. Diversification in Multifactor Models of Asset Returns

A return generating process within which it is easy to characterize diversifiable risk is given by the following linear factor structure:

$$\tilde{R}_t = E + B\tilde{\delta}_t + \tilde{\varepsilon}_t; E[\tilde{\delta}_t] = E[\tilde{\varepsilon}_t\tilde{\delta}_t'] = 0; E[\tilde{\delta}_t\tilde{\delta}_t'] = I_K; E[\tilde{\varepsilon}_t\tilde{\varepsilon}_t'] = \Omega \quad (1)$$

where \tilde{R}_t and E are N -vectors of returns and expected returns, respectively, B is an $N \times K$ matrix of factor loadings, $\tilde{\delta}_t$ is a K -vector of mean zero common factors normalized to have a $K \times K$ identity covariance matrix I_K , and $\tilde{\varepsilon}_t$ is an N -vector of mean zero disturbances with positive definite covariance matrix Ω . The assumption that Ω is positive definite implies that no asset contains only factor risk. No substantive assumptions are made in (1) except for the existence of the relevant first and second moments.¹ In particular, no assumption has yet been made regarding the potential to diversify away idiosyncratic risk.

Additional structure can be placed on (1) by assuming that the common factors are the dominant source of covariation among returns or, equivalently, that the residuals represent diversifiable risk. The residuals will be diversifiable if a weak law of large numbers applies to them. Chamberlain and Rothschild (1983) provide a particularly convenient sufficient condition for such a weak law of large numbers to hold: that the largest eigenvalue of Ω remains finite as $N \rightarrow \infty$.² Chamberlain and Rothschild (1983) termed (1) coupled with this constraint on Ω an *approximate factor structure*.

What diversification opportunities arise in a world with an *approximate factor structure*?

An arbitrary portfolio with weights w_p has first and second moments of returns given by:

$$\begin{aligned}\tilde{R}_{pt} &= w_p' \tilde{R}_t = w_p' E + w_p' B \tilde{\delta}_t + w_p' \tilde{\epsilon}_t; \text{Var}[\tilde{R}_{pt}] = b_p' b_p + \sigma_{\tilde{\epsilon}_p}^2 \\ &= E_p + b_p' \tilde{\delta}_t + \tilde{\epsilon}_{pt}; \sigma_{\tilde{\epsilon}_p}^2 = w_p' \Omega w_p\end{aligned}\tag{2}$$

where E_p is the portfolio's expected return, b_p is a $K \times 1$ vector of the portfolio's factor loadings, and $\tilde{\epsilon}_{pt}$ is the portfolio's residual return at time t . Portfolio p is a *well-diversified* portfolio if its weights are of the order $1/N$ and the sum of squared weights converges to zero (i.e., $w_p' w_p \rightarrow 0$). The idiosyncratic risk of a *well-diversified portfolio* vanishes in the limit when returns have an *approximate factor structure* because $\sigma_{\tilde{\epsilon}_p}^2 = w_p' \Omega w_p \leq w_p' w_p \xi_{\max}(\Omega) \rightarrow 0$ where $\xi_{\max}(\Omega)$ is the largest eigenvalue of Ω .

A *well-diversified* portfolio with no exposure to factor risk (i.e., $b_p = 0$) generally exists unless a vector of ones approximately lies in the column span of B . Such a *factor-neutral portfolio* cannot be formed from a finite asset menu if the factors can be normalized so that $B\iota = \mathbf{1}$, where ι is a suitably conformable vector of ones. A convenient sufficient condition for the

¹ At this stage, these moments can be either conditional or unconditional.

existence of such a portfolio in the infinite dimensional case is that the covariance matrix of the rows of B is nonsingular in the limit.³ That is: $Var[B] = \frac{1}{N} \sum_{j=1}^N (\beta_j - \bar{\beta})(\beta_j - \bar{\beta})'$ is positive definite where $\bar{\beta} = \frac{1}{N} \sum_{j=1}^N \beta_j$ and β_j is the j^{th} row of B. Note that the return of such a portfolio (if it exists) must converge to the riskfree rate as $N \rightarrow \infty$ or else there is an arbitrage opportunity. The possibility of creating a riskless portfolio from a set of risky assets is one of the main differences between mean-variance analysis in finite and infinite-dimensional asset markets.

The main restriction on the limiting behavior of *well-diversified* portfolios involves their expected returns. Since the returns of any $K+1$ well-diversified portfolios are perfectly correlated in the limit under these assumptions, their limiting expected returns must be linearly dependent as well. Accordingly, there exists a K -vector λ and a constant λ_0 such that $E_p \rightarrow \lambda_0 + b_p' \lambda$ for all *well-diversified* portfolios p . Note that this mean restriction is *not* the APT, a theory that converts this *exact* expression for *well-diversified* portfolio expected returns into an *approximate* expression for *individual asset* expected returns. Hence, this analysis is free of the controversy surrounding the APT.

2. Basis Portfolio Formation

A statistician seeking to estimate $\tilde{\delta}_t$ in (2), given *a priori* knowledge of $\tilde{R}_t, B, E,$ and $\Omega,$ would naturally use generalized least squares. This is the minimum variance linear unbiased estimator:

² In what follows, all limits will be taken as $N \rightarrow \infty,$ with N left implicit to avoid notational clutter.

³ Set $B\iota = \iota + \eta$ and consider a unit cost portfolio $w_p.$ The sum of its betas is given by $w_p' B\iota \equiv b_p' \iota = w_p' \iota + w_p' \eta.$ Hence, this sum will generically be one if p is well diversified and η can be eliminated via diversification. A column vector of ones will approximately lie in the span of B if, for example, one of the common factors is unanticipated inflation and all but a finite number of securities returns are equally affected by this factor.

$$\hat{\delta}_t^{GLS} = (B'\Omega^{-1}B)^{-1}B'\Omega^{-1}[\tilde{R}_t - E] = \delta_t + (B'\Omega^{-1}B)^{-1}B'\Omega^{-1}\tilde{\varepsilon}_t; \text{Var}[\hat{\delta}_t^{GLS} - \tilde{\delta}_t] = (B'\Omega^{-1}B)^{-1} \quad (3)$$

where $(B'\Omega^{-1}B)^{-1}B'\Omega^{-1}\tilde{\varepsilon}_t$ is the error in tracking δ_t . The estimator $\hat{\delta}_t^{GLS}$ is consistent if the smallest eigenvalue of $B'B$ grows without bound as $N \rightarrow \infty$, the usual OLS consistency condition. The rate at which $\hat{\delta}_t^{GLS}$ converges hinges on two factors: (a) the structure of Ω and (b) the dispersion across the rows of B . *Ceteris paribus*, convergence is faster when the securities used in estimation have smaller idiosyncratic variances, lower correlations among their idiosyncratic risks, and greater differences among their risk exposures. The last factor is well understood in a regression setting: precise estimates of the covariance between the dependent and independent variables are hard to obtain in small samples if the independent variables vary little across observations. Similarly, large cross-sections are needed to obtain precise factor estimates if securities typically possess similar risk exposures. For example, suppose two of the factors are unexpected changes in expected inflation and unanticipated inflation but only a small fraction of the assets respond differently to them (i.e., inflation has a neutral impact on most security returns). In this case, accurate estimation of both common factors would require a much larger cross-section than would typically be needed to merely eliminate idiosyncratic risk.

In practice, the choice among basis portfolio construction methods is further complicated by the need to estimate B and Ω . The estimation of Ω requires the placement of constraints on its functional form. Popular choices are the statistical factor analysis model, in which Ω is diagonal, and the principal components model, in which Ω is a diagonal matrix with equal variances. Measurement error in the estimates of B and Ω can impair the small sample properties of basis portfolio estimators and, as a consequence, large cross-sections may be needed to produce reliable basis portfolio estimates.

In what follows, we consider four procedures for constructing basis portfolios. Two of the methods involve biased and unbiased versions of the generalized least squares estimator discussed above. Unbiasedness is not a particular virtue, particularly in the presence of measurement error and in the absence of a natural scale for the latent factors. Hence, it may be desirable to seek a biased estimator with potentially lower variance. The other two methods use mathematical programming procedures that constrain the reference portfolios to be well-diversified, which might mitigate some of the harmful effects of both sampling error in and the imposition of constraints on Ω .

Before considering alternative portfolio formation procedures, it is useful to translate this statistical formulation into the language of optimal portfolio formation after the fashion of Litzenberger and Ramaswamy (1979) and Rosenberg and Marathe (1979). The generalized least squares estimator provides what we usually refer to as Fama-MacBeth portfolios, after rescaling the portfolios to cost a dollar.⁴ In particular, we choose the N portfolio weights w_j to mimic the j^{th} factor so that they:

$$\min_{w_j} w_j' D w_j \text{ subject to } w_j' b_j = 1; w_j' b_k = 0 \quad \forall j \neq k \quad (4)$$

where b_k is the k^{th} column of the *sample* factor loading matrix B and D is the diagonal matrix consisting of the *sample* idiosyncratic variances.⁵ The solution is the *unbiased minimum idiosyncratic risk portfolio* that mimics the j^{th} unobservable common factor. We rescale the weights w_j so that the portfolio costs a dollar (i.e., $w_j' \mathbf{1} = 1$) to maintain comparability with other basis portfolio formation procedures.

⁴ Of course, Fama and MacBeth (1973) used the ordinary least squares estimator. Our usage is, however, common.

An alternative method that produces what we term *minimum idiosyncratic risk portfolios* involves choosing portfolio weights w_j so that:

$$\min_{w_j} w_j' D w_j \text{ subject to } w_j' \mathbf{1} = 1; w_j' b_k = 0 \quad \forall j \neq k \quad (5)$$

In principle, this procedure should produce minimum idiosyncratic risk portfolios whose fluctuations are proportional to the common factor. In contrast to the unbiased OLS estimator, the proportionality factor need not equal one.⁶ This is easily seen by comparing (4) and (5).

It is easy to distinguish these minimum idiosyncratic risk portfolios from the more familiar Fama-MacBeth portfolios in the one factor case. Assume for simplicity that the idiosyncratic variances are identical (i.e., $D = \sigma_\epsilon^2 I_N$). In this case, the Fama-MacBeth portfolio solves the programming problem:

$$\min_w w' w \text{ subject to } w' b = 1 \quad (6)$$

with solution $w = (b'b)^{-1} b$ prior to rescaling so that the portfolio weights sum to one where b is the vector of *sample* factor loadings. Similarly, the *minimum idiosyncratic risk portfolio* satisfies:

$$\min_w w' w \text{ subject to } w' \mathbf{1} = 1 \quad (7)$$

which is solved by the equally weighted portfolio $w = \frac{1}{N} \mathbf{1}$. Thus Fama-MacBeth portfolio weights are proportional to the individual security sample factor loadings (i.e., the betas) and, thus, take advantage of the differing information content of individual securities regarding the fluctuations in the common factor. Minimum idiosyncratic risk portfolios, however, are merely well-

⁵ Note that we are now ignoring off-diagonal elements of Ω such as industry effects. As a consequence, our procedures actually are better characterized as weighted least squares or diagonal generalized least squares.

⁶ Let $B = (b_1 b_2 \dots b_K)$ and suppose we are interested in mimicking the j^{th} factor. The minimum idiosyncratic risk estimator is $D^{-1} B^* [B^* D^{-1} B^*]^{-1}$ where $B^* = (b_1 b_2 \dots \mathbf{1} \dots b_K)$ and $\mathbf{1}$ is a vector of ones in the j^{th} column.

diversified and do not exploit such information. Note that the factor loading of the minimum idiosyncratic portfolio is the average beta of the securities (i.e., $\bar{b} = \frac{1}{N}b'\mathbf{1}$) while that of the Fama-MacBeth portfolio is unity prior to rescaling.

The diversification properties of Fama-MacBeth portfolios depend on the normalization of the common factors. If the factors are normalized so that factor loadings are typically close to one,⁷ both procedures yield portfolio weights of order $1/N$. However, Fama-MacBeth portfolios have very large weights in finite cross-sections when average factor loading estimates are on the order of 0.001 to 0.0001 as is typical in daily data under the normalization $Var(\tilde{\delta}_t) = I_K$. This is not a problem when the factor model parameters are measured without error but it is a potentially serious source of difficulty when large factor loadings can reflect measurement error as well as responsiveness to common factors.

The comparative merits of minimum idiosyncratic risk and Fama-MacBeth portfolios in the presence of measurement error can be investigated more fully by again considering the one factor model:

$$\tilde{R}_t = E + \beta\tilde{\delta}_t + \tilde{\varepsilon}_t; E[\tilde{\varepsilon}_t\tilde{\varepsilon}_t'] = \sigma_\varepsilon^2 I_N; \sigma_\beta^2 > 0 \quad (8)$$

where $\beta \neq \bar{\beta}\mathbf{1}$ is the $N \times 1$ vector of the (not identical) population factor loadings and the idiosyncratic disturbances are assumed, for simplicity, to have zero means, common (known) variances σ_ε^2 , and to be distributed independently of one another. Suppose that we measure the factor loadings β with error:

$$b = \beta + v; E[v] = 0; E[vv'] = \Sigma_v \quad (9)$$

⁷ We are not aware of any study in which the factors are normalized to ensure that the typical loading is unity. The following transformation yields typical loadings of one. Transform B so

where \boldsymbol{v} is an N vector of the deviations of the sample factor loadings from their population counterparts with \boldsymbol{v} assumed to be independent of $\tilde{\delta}_t$ and $\tilde{\epsilon}_t \forall t$.⁸ We also normalize the factor loading estimates such that $\mathbf{b}'\mathbf{b} = \mathbf{b}'\mathbf{1}$. This normalization implies that $0 < \bar{b} < 1$ to ensure that the sample variance of the b_i 's is positive. The Fama-MacBeth procedure yields a basis portfolio with a sample loading of unity and with weights that sum to one. Finally, we assume that the cross-section is large enough to ensure that $\bar{b} \approx \bar{\beta} = \frac{1}{N} \sum_{j=1}^N \beta_j$ [i.e., $\frac{1}{N} \sum_{j=1}^N v_j \approx 0$] to simplify the arithmetic.

In this setting it is easy to characterize the behavior of the Fama-MacBeth portfolio and the relations between its returns and the common factor. The portfolio returns are:

$$\begin{aligned} \tilde{R}_t^{FM} &= (\mathbf{b}'\mathbf{b})^{-1} \mathbf{b}'[E + \beta\tilde{\delta}_t + \tilde{\epsilon}_t] = (\mathbf{b}'\mathbf{b})^{-1}[\beta'E + \beta'\beta\tilde{\delta}_t + \beta'\tilde{\epsilon}_t + \mathbf{v}'E + \mathbf{v}'\beta\tilde{\delta}_t + \mathbf{v}'\tilde{\epsilon}_t] \\ &\approx \frac{1}{N\bar{\beta}}[\beta'E + \beta'\beta\tilde{\delta}_t + \beta'\tilde{\epsilon}_t + \mathbf{v}'E + \mathbf{v}'\beta\tilde{\delta}_t + \mathbf{v}'\tilde{\epsilon}_t] \end{aligned} \quad (10)$$

where the approximation $\bar{b} \approx \bar{\beta}$ implies that $N\bar{\beta} \approx \mathbf{b}'\mathbf{b}$. The variance of Fama-MacBeth portfolio returns as well as their squared correlation with the common factor are (approximately) given by:

$$\begin{aligned} \text{Var}[\tilde{R}_t^{FM}] &\approx \frac{1}{N^2\bar{\beta}^2}[(\beta'\beta^2 + \beta'\Sigma_v\beta)\sigma_\delta^2 + \beta'\beta\sigma_\epsilon^2 + E'\Sigma_v E + \sigma_\epsilon^2 \text{tr}(\Sigma_v)] \\ \text{Corr}[\tilde{R}_t^{FM}, \tilde{\delta}_t]^2 &\approx \frac{\beta'\beta^2\sigma_\delta^2}{(\beta'\beta^2 + \beta'\Sigma_v\beta)\sigma_\delta^2 + \beta'\beta\sigma_\epsilon^2 + E'\Sigma_v E + \sigma_\epsilon^2 \text{tr}(\Sigma_v)} \end{aligned} \quad (11)$$

that $\mathbf{B}'\mathbf{D}^{-1}\mathbf{B}$ is diagonal and denote the j^{th} diagonal element as γ_j . Then $\mathbf{B}^* = \mathbf{B}\mathbf{A}$ where \mathbf{A} is a diagonal matrix with ζ_j/γ_j along the diagonal where $\zeta = \mathbf{B}'\mathbf{D}^{-1}\mathbf{1}$ has a typical loading of one.

⁸ We assume that our estimates are unbiased for simplicity. The assumption that \boldsymbol{v} is independent of $\tilde{\delta}_t$ and $\tilde{\epsilon}_t$ is less innocuous since we typically estimate factor loadings and form basis portfolios on the same sample. Accounting for such problems would complicate the analysis considerably without altering the basic point. For example, if \mathbf{b} were estimated from an ordinary least squares regression, then $\text{cov}[\mathbf{v}_i, \tilde{\epsilon}_t] = \frac{\bar{\delta}\sigma_\epsilon^2}{\sum_{t=1}^T \delta_t^2} = \frac{\bar{\delta}\sigma_\epsilon^2}{T[\sigma_\delta^2 + \bar{\delta}^2]}$ would typically be quite small in moderately large samples.

where σ_δ^2 is the variance of the common factor.

It is easier to evaluate the corresponding quantities for the minimum idiosyncratic risk portfolio because it ignores the information in the betas. Hence, its returns and their corresponding moments are:

$$\begin{aligned}\tilde{R}_t^{MIRP} &= \frac{1}{N} t[E + \beta \tilde{\delta}_t + \tilde{\varepsilon}_t] = \bar{E} + \bar{\beta} \tilde{\delta}_t + \bar{\varepsilon}_t; \quad Var[\tilde{R}_t^{MIRP}] \approx \bar{\beta}^2 \sigma_\delta^2 + \frac{\sigma_\varepsilon^2}{N} \\ Corr[\tilde{R}_t^{MIRP}, \tilde{\delta}_t]^2 &\approx \frac{\bar{\beta}^2 \sigma_\delta^2}{\bar{\beta}^2 \sigma_\delta^2 + \sigma_\varepsilon^2 / N}\end{aligned}\quad (12)$$

What are we to make of this tedious arithmetic? The latent nature of the factors makes their scale arbitrary, making high correlation the appropriate objective as opposed to low tracking error. When β is measured without error [i.e., $v = 0$], the Fama-MacBeth portfolio return is more highly correlated with the common factor than that of the minimum idiosyncratic risk portfolio since:

$$Corr[\tilde{R}_t^{FM}, \tilde{\delta}_t]^2 \approx \frac{\beta' \beta \sigma_\delta^2}{\beta' \beta \sigma_\delta^2 + \sigma_\varepsilon^2} = \frac{(\sigma_\beta^2 + \bar{\beta}^2) \sigma_\delta^2}{(\sigma_\beta^2 + \bar{\beta}^2) \sigma_\delta^2 + \sigma_\varepsilon^2 / N} \geq \frac{\bar{\beta}^2 \sigma_\delta^2}{\bar{\beta}^2 \sigma_\delta^2 + \sigma_\varepsilon^2 / N} \approx Corr[\tilde{R}_t^{MIRP}, \tilde{\delta}_t]^2 \quad (13)$$

where $\sigma_\beta^2 = \frac{1}{N} \sum_{j=1}^N (\beta_j - \bar{\beta})^2 > 0$ is the sample cross-sectional variance of the population betas.

The same cannot be said when there is sampling error in the betas: the squared correlation of minimum idiosyncratic risk portfolio returns with the common factor is unaffected by measurement error in β while that of the Fama-MacBeth portfolio falls in the presence of sampling error.

For example, consider the special case in which β is estimated by ordinary least squares regression of $\tilde{R}_t - E$ on $\tilde{\delta}_t$, in which case the error covariance matrix is (approximately):

$$\Sigma_v \approx \frac{\sigma_\varepsilon^2}{T \sigma_\delta^2} I_N \quad (14)$$

where T is the sample size and the approximation arises because we replaced the sample mean and variance of $\tilde{\delta}_t$ with their population values for ease of exposition. The squared correlation of the Fama-MacBeth portfolio with the common factor is then:

$$\begin{aligned} \text{Corr}[\tilde{R}_t^{FM}, \tilde{\delta}_t]^2 &\approx \frac{\beta' \beta^2 \sigma_\delta^2}{\beta' \beta^2 \sigma_\delta^2 + \beta' \Sigma_v \beta \sigma_\delta^2 + \beta' \beta \sigma_\varepsilon^2 + E' \Sigma_v E + \sigma_\varepsilon^2 \text{tr}(\Sigma_v)} \\ &\approx \frac{1}{1 + \frac{\sigma_\varepsilon^2}{N} \left[\frac{T+1}{T(\sigma_\beta^2 + \bar{\beta}^2) \sigma_\delta^2} + \frac{(\sigma_E^2 + \bar{E}^2 + \sigma_\varepsilon^2)}{T(\sigma_\beta^2 + \bar{\beta}^2)^2 \sigma_\delta^4} \right]}; \quad \sigma_E^2 = \frac{1}{N} \sum_{j=1}^N (E_j - \bar{E})^2 \end{aligned} \quad (15)$$

which will be smaller than that of the minimum idiosyncratic risk portfolio when:

$$T(\sigma_\beta^2 + \bar{\beta}^2)^2 \sigma_\delta^2 < (T+1) \bar{\beta}^2 (\sigma_\beta^2 + \bar{\beta}^2) \sigma_\delta^2 + \bar{\beta}^2 (\sigma_E^2 + \bar{E}^2 + \sigma_\varepsilon^2) \quad (16)$$

which can easily obtain when the sampling variation in the betas is small or, equivalently (given the normalization $\frac{1}{N} b' b = \bar{b}$ and the approximation $\bar{b} \approx \bar{\beta}$), when $\bar{\beta}$ is close to one. Note that the variances of sample betas computed with respect to the usual market proxies are often quite small.

While both of these procedures produce well-diversified basis portfolios in the limit, they may not do so in finite asset menus. This possibility led Chen (1983) to employ (proprietary) mathematical programming methods to produce well-diversified basis portfolios with a finite number of securities. Such portfolios might suffer only marginally from errors in estimating the factor loadings and idiosyncratic variances due to the diversification constraint. Accordingly, our other two portfolio formation procedures involve quadratic programming subject to fixed upper and lower bounds as in:

$$\min_{w_j} w_j' D w_j \quad \text{subject to } w_j' \mathbf{1} = 1; \quad w_j' b_k = 0 \quad \forall j \neq k; \quad \ell_j \leq w_j \leq u_j \quad (17)$$

where ℓ_j and u_j are the fixed lower and upper lower bounds. We examined two choices for these bounds. Following Chen (1983), we produced portfolios with non-negative weights (i.e., $\ell_j = 0$)

that could take on maximum values of one to two per cent. We also examined portfolios that were merely constrained to be well-diversified with weights taking on maximum and minimum values of plus and minus one or two per cent. We experimented with constraints as large as five per cent in absolute value and the results did not differ materially from the ones presented below and, hence, are not reported.

Finally, we formed sample factor neutral portfolios for each portfolio formation method. These portfolios are used to construct the excess return basis portfolios analyzed below. The differences among excess return portfolios are also best illustrated in the one factor case with identical and uncorrelated idiosyncratic risks. In this case, the required orthogonal portfolio solves the programming problem:

$$\min_{w_{R_f}} w'_{R_f} w_{R_f} \text{ subject to } w'_{R_f} \mathbf{1} = 1; w'_{R_f} \boldsymbol{\beta} = 0 \Rightarrow w_{R_f} = \frac{1}{N} \left[\frac{\sigma_{\beta}^2 + \bar{\beta}^2}{\sigma_{\beta}^2} \mathbf{1} - \frac{\bar{\beta}}{\sigma_{\beta}^2} \boldsymbol{\beta} \right] \quad (18)$$

The corresponding excess return portfolio weights are obtained by subtracting these weights from $w = \frac{1}{N} \mathbf{1}$, the minimum idiosyncratic risk portfolio in this case as noted in (7) above. This yields:

$$w_{MIRP} - w_{R_f} = \frac{1}{N} \mathbf{1} - \frac{1}{N} \left[\frac{\sigma_{\beta}^2 + \bar{\beta}^2}{\sigma_{\beta}^2} \mathbf{1} - \frac{\bar{\beta}}{\sigma_{\beta}^2} \boldsymbol{\beta} \right] = \frac{\bar{\beta}}{N\sigma_{\beta}^2} (\boldsymbol{\beta} - \mathbf{1}\bar{\beta}) \quad (19)$$

The Fama-MacBeth excess return portfolio solves the programming problem:

$$\min_{w_{FM}^{ex}} w_{FM}^{ex}{}' w_{FM}^{ex} \text{ subject to } w_{FM}^{ex}{}' \mathbf{1} = 0; w_{FM}^{ex}{}' \boldsymbol{\beta} = 1 \Rightarrow w_{FM}^{ex} = \frac{1}{N\sigma_{\beta}^2} (\boldsymbol{\beta} - \mathbf{1}\bar{\beta}) \quad (20)$$

These manipulations yield one nontrivial insight — the minimum idiosyncratic risk procedure produces weights for excess return portfolios that are proportional to the Fama-MacBeth excess return portfolios. Not surprisingly, the factor of proportionality is the average factor loading $\bar{\beta}$. Once again, when the average factor loading is much less than one, the Fama-

MacBeth procedure will produce very large positive and negative portfolio weights. In contrast, the minimum idiosyncratic procedure yields a well-diversified excess return portfolio. A similar result arises in the multiple factor case.

The choice of basis portfolio formation procedure should be a second order consideration since well-diversified portfolios possess no idiosyncratic risk in the limit and so any K imperfectly correlated well-diversified portfolios should suffice. In this section, we have argued that the construction of reliable basis portfolios in practice requires a large enough asset menu so as to virtually eliminate idiosyncratic risk and an asset menu with sufficiently different factor sensitivities so as to allow statistical discrimination among the factors. Further, we have shown that asymptotically efficient basis portfolio formation procedures that make use of sample differences in factor sensitivities may have inferior small sample properties to those methods that simply emphasize diversification and orthogonality.

3. Estimation Methods

The choice among estimation methods involves different tradeoffs than the choice among basis portfolio formation procedures. The tradeoff here is between statistically efficient but computationally costly methods such as factor analysis, and less efficient but less costly methods such as instrumental variables or principal components. The question is whether comparatively inefficient methods provide performance comparable to that of computationally burdensome efficient estimation methods.

When the factors are not observed, estimates of B and Ω may be obtained from the covariance structure of returns, $\Sigma = BB' + \Omega$. When returns are normally and independently distributed — that is, when $\tilde{\delta}_i$ and $\tilde{\epsilon}_i$ are jointly normal and serially independent — the higher moments of returns contain no information about these parameters. In these circumstances, the

log likelihood function is given by:

$$\begin{aligned}\mathcal{L}(\Sigma | S) &= -\frac{NT}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma| - \frac{1}{2} \sum_{t=1}^T (\tilde{R}_t - \bar{R})' \Sigma^{-1} (\tilde{R}_t - \bar{R}) \\ &= -\frac{NT}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma| - \frac{T}{2} \text{tr}(S \Sigma^{-1})\end{aligned}\tag{21}$$

since the sample covariance matrix $S = \frac{1}{T} \sum_{t=1}^T (\tilde{R}_t - \bar{R})(\tilde{R}_t - \bar{R})'$ follows a Wishart distribution.

Unfortunately, the estimation of B and Ω when security returns possess an *approximate factor structure* requires the imposition of further constraints on Ω . The two main alternatives are factor analysis and principal components: Ω is a diagonal matrix D in the former while $\Omega = \sigma_\varepsilon^2 I_N$ in the latter. This observation highlights the intuitive distinction between factor analysis and principal components — factor analysis implicitly involves weighted least squares estimates of the factors and factor loadings (where the weights are the estimated idiosyncratic variances) while principal components provides the corresponding ordinary least squares estimates. Hence, factor analysis should tend to perform better than principal components when there is nontrivial cross-sectional variation in idiosyncratic variances.

We used two relatively inexpensive techniques for estimating the statistical factor analysis model. The first is a modified version of the EM algorithm as applied to factor analysis in Lehmann and Modest (1988), which used an iterative multivariate regression procedure. The first two steps of the EM algorithm in this paper are unchanged but the third step is new:

1. Given B and D , compute the minimum variance *biased* estimator of the factors

$$\text{via } \hat{\delta}_t^{BGLS} = (I + B' \Omega^{-1} B)^{-1} B' \Omega^{-1} [\tilde{R}_t - E];^9$$

⁹ $\hat{\delta}_t^{BGLS}$ has lower mean squared error than (3) since it uses the information in the normalization $\text{Var}(\tilde{\delta}_t) = I_K$.

2. Estimate B and D via multivariate regression of \tilde{R}_i on $\hat{\delta}_i^{BGLS}$ and
3. Compute the MLE of B given D from step 2 via $B = D^{1/2}\Gamma(D,S^*)\Lambda^{-1/2}$ (D,S*) where $\Lambda(D,S^*)$ is a diagonal matrix comprised of the K largest eigenvalues of $S^* = D^{-1/2}SD^{-1/2}$ and $\Gamma(D,S^*)$ are the associated eigenvectors. Return to step 1 if the algorithm has not converged.

This revised EM algorithm usually converges much more rapidly than the conventional EM algorithm.¹⁰ We refer to this below as unrestricted maximum likelihood.

The second technique, which we refer to as restricted maximum likelihood, makes use of the information on B implicit in the vector of sample mean security returns when the APT is true. In this case, expected security returns are linear combinations of the product of their factor loadings and a set of factor risk premiums. In order to exploit this information, we also performed maximum likelihood factor analysis subject to this constraint by maximizing:

$$\mathcal{L}(B, D, \lambda_0, \lambda | S) = -\frac{NT}{2} \ln 2\pi - \frac{T}{2} \ln |\Sigma| - \frac{T}{2} \text{tr}(S\Sigma^{-1}) - \frac{T}{2} (\bar{R} - \iota\lambda_0 - B\lambda)' \Sigma^{-1} (\bar{R} - \iota\lambda_0 - B\lambda) \quad (22)$$

where λ is the vector of factor risk premiums and $\lambda_0 = 0$ if a column of ones lies in the column span of B. We used the unrestricted estimates of B and D from maximum likelihood factor analysis as starting values and then obtained these restricted maximum likelihood estimates using a variant of the EM algorithm.

Principal components has been advocated by Chamberlain and Rothschild (1983) and

¹⁰ We found it necessary to create this modification because the conventional EM algorithm spent a great deal of time slightly refining the factor loading estimates in the neighborhood of the maximum of the likelihood function. Jones (2001) uses a very similar iterative principal components procedure that differs in how the D estimate is updated. He shows that the resulting factor estimates are consistent as $N \rightarrow \infty$ when B is time invariant if the factors are serially correlated and the residuals are heteroskedastic and provides a method for dealing with missing-at-random data along the lines of Connor, Korajczyk, and Uhlauer (2002). See Lehmann (1992) for related results.

Connor and Korajczyk (1986,1988) as an inexpensive alternative to maximum likelihood factor analysis. This method requires the extraction of the K largest eigenvalues of S and their associated eigenvectors or, equivalently, the K largest singular values of the matrix of returns and their corresponding singular vectors. We extracted all of the eigenvalues and eigenvectors of S and scaled each eigenvector by the square root of the corresponding eigenvalue to set $Var(\tilde{\delta}_t) = I_K$. We then estimated D by substituting the transformed eigenvectors associated with the K largest eigenvalues for B into the model $\Sigma = BB' + D$.¹¹

Finally, instrumental variables is another inexpensive alternative to maximum likelihood factor analysis, a variant of which was employed in Chen (1983). The basic idea is quite simple: substitute consistent estimates of the factors for the factors themselves in (1) and estimate the factor loadings and idiosyncratic variances by ordinary least squares regression. Following Madansky (1964) and Hagglund (1982), we normalize the factors so that I_K is the factor loading matrix of the first K securities, making are the corresponding factors correlated, and decompose the vector of returns into three components:

$$\begin{aligned}\tilde{r}_{1t} &= \tilde{R}_{1t} - E_1 = \tilde{\delta}_t + \tilde{\epsilon}_{1t} \\ \tilde{r}_{2t} &= \tilde{R}_{2t} - E_2 = \gamma_2' \tilde{\delta}_t + \tilde{\epsilon}_{2t} \\ \tilde{r}_{3t} &= \tilde{R}_{3t} - E_3 = \Gamma_3' \tilde{\delta}_t + \tilde{\epsilon}_{3t}\end{aligned}\tag{23}$$

where \tilde{r}_{1t} , \tilde{r}_{2t} , and \tilde{r}_{3t} are the demeaned returns on the first K , the $K+1^{\text{st}}$, and the last $N-K-1$ assets, respectively, and γ_2 and Γ_3 are the corresponding factor loading vector and factor loading matrix of the $K+1^{\text{st}}$ and the last $N-K-1$ assets, respectively. Consider the regression of \tilde{r}_{2t} on \tilde{r}_{1t} :

¹¹ Connor, Korajczyk, and Uhlener (2002) avoid the comparatively expensive eigenvalue/eigenvector decomposition by using a two-pass cross-sectional regression method that is a close cousin of the EM algorithm. They also consider methods for dealing with data that is missing-at-random.

$$\tilde{r}_{2t} = \gamma_2' \tilde{r}_{1t} + \tilde{u}_{2t} = \gamma_2' \tilde{r}_{1t} + \tilde{\varepsilon}_{2t} - \gamma_2' \tilde{\varepsilon}_{1t} \quad (24)$$

Clearly application of ordinary least squares will lead to biased and inconsistent estimates of γ_2 since \tilde{r}_{1t} is correlated with its own idiosyncratic disturbance term. If instead we first regress \tilde{r}_{1t} on \tilde{r}_{3t} :

$$\tilde{r}_{1t} = \Pi \tilde{r}_{3t} + \tilde{u}_{3t} \quad (25)$$

and then replace \tilde{r}_{1t} in equation (24) with the fitted values from this regression, ordinary least squares can then be used to estimate γ_2 consistently. This estimate is consistent because the fitted values from (25) are purged of $\tilde{\varepsilon}_{1t}$ and the idiosyncratic disturbances $\tilde{\varepsilon}_{3t}$ are uncorrelated or sufficiently weakly correlated with $\tilde{\varepsilon}_{2t}$ by assumption. Repeated application of this procedure after swapping \tilde{r}_{2t} with each element of \tilde{r}_{3t} yields consistent estimates of the corresponding rows of Γ_3 . Finally, solution of the matrix equations:

$$\begin{aligned} \Gamma'(S - \Gamma\Phi\Gamma' - D)\Gamma &= 0 \\ \text{Diag}(S - \Gamma\Phi\Gamma' - D) &= 0 \end{aligned} \quad (26)$$

yields estimates of the factor covariance matrix Φ and the idiosyncratic variances D . The Φ estimate can be used to transform Γ into B by implicitly rescaling the factors to be uncorrelated with unit variances.

4. Basis Portfolio Comparison

It would be a simple matter to determine which combination of basis portfolio formation procedure and estimation method works best if the factors were observed. Of course, we would not need to construct basis portfolios if we observed the common factors. One heuristic approach to comparing basis portfolios is to examine the behavior of their weights and the sample means and variances of their returns in order to check whether the results appear to be reasonable. For

example, basis portfolios ought to be well-diversified in order to diminish idiosyncratic risk. The quadratic programming portfolios are well-diversified by construction but the same need not be true of Fama-MacBeth or minimum idiosyncratic risk portfolios in finite samples. Similarly, the sample moments of basis portfolio returns can reveal peculiarities in their performance. For example, we have examined basis portfolios with mean returns and standard deviations as high as 120 percent and 450 percent per month, respectively. Good basis portfolios probably do not exhibit such behavior. Unfortunately, searching for reasonable reference portfolios in this fashion is not likely to eliminate many candidates.

Fortunately, some plausible large sample approximations facilitate the comparison of *well-diversified* basis portfolios. The sample mean vector and covariance matrix of \tilde{R}_{pt} are given by:

$$\begin{aligned}\bar{R}_p &= \frac{1}{T} \sum_{t=1}^T \tilde{R}_{pt} = \iota \lambda_0 + B_p \bar{\lambda} + \bar{\varepsilon}_p; \quad \bar{\delta} = \frac{1}{T} \sum_{t=1}^T \tilde{\delta}_t; \quad \bar{\lambda} = (\lambda + \bar{\delta}); \quad \bar{\varepsilon}_p = \frac{1}{T} \sum_{t=1}^T \tilde{\varepsilon}_t \\ S_p &= \frac{1}{T} \sum_{t=1}^T (\tilde{R}_{pt} - \bar{R}_p)(\tilde{R}_{pt} - \bar{R}_p)' = B_p S_\delta B_p' + B_p S_{\delta \varepsilon_p} + S_{\delta \varepsilon_p}' B_p' + S_{\varepsilon_p} \\ S_\delta &= \frac{1}{T} \sum_{t=1}^T (\tilde{\delta}_t - \bar{\delta})(\tilde{\delta}_t - \bar{\delta})'; \quad S_{\delta \varepsilon_p} = \frac{1}{T} \sum_{t=1}^T (\tilde{\delta}_t - \bar{\delta}) \tilde{\varepsilon}_{pt}'; \quad S_{\varepsilon_p} = \frac{1}{T} \sum_{t=1}^T \tilde{\varepsilon}_{pt} \tilde{\varepsilon}_{pt}'\end{aligned}\tag{27}$$

where λ_0 is zero if a vector of ones (approximately) lies in the column span of B and is the riskless rate otherwise. $S_{\delta \varepsilon_p}$, the matrix comprised of the sample covariances between the actual factor realizations and the idiosyncratic returns of the basis portfolios, should be close to its probability limit of zero in large samples (i.e., $S_p \approx B_p S_\delta B_p' + S_{\varepsilon_p}$). Similarly, the sample mean vector $\bar{\varepsilon}_p$ will be close to zero as well if the basis portfolios are large, well-diversified, and constructed so that their weights are not systematically related to the $\tilde{\varepsilon}_{pt}$ realizations (i.e., $\bar{R}_p \approx \iota \lambda_0 + B_p \bar{\lambda}$). We do *not* assume that S_{ε_p} is close to zero since $\bar{\varepsilon}_p$ will converge to zero faster

than S_{ε_p} for well-diversified portfolios of large numbers of securities.

Now consider the usual χ^2 statistic for testing the hypothesis that the K mean excess returns of the basis portfolios over λ_0 are all zero:

$$\begin{aligned} T(\bar{R}_p - \lambda_0)' S_p^{-1} (\bar{R}_p - \lambda_0) &\approx T \bar{\lambda}' B_p' S_p^{-1} B_p \bar{\lambda}; S_p^{-1} = S_{\varepsilon_p}^{-1} - S_{\varepsilon_p}^{-1} B_p (S_{\delta}^{-1} + B_p' S_{\varepsilon_p}^{-1} B_p)^{-1} B_p' S_{\varepsilon_p}^{-1} \\ &\approx T \bar{\lambda}' [(B_p' S_{\varepsilon_p}^{-1} B_p)^{-1} + S_{\delta}]^{-1} \bar{\lambda}; B_p' S_p^{-1} B_p = [(B_p' S_{\varepsilon_p}^{-1} B_p)^{-1} + S_{\delta}]^{-1} \end{aligned} \quad (28)$$

The structure of S_p simplifies matters considerably and the χ^2 statistic is (approximately) bounded by the maximum attainable value of $T \bar{\lambda}' S_{\delta}^{-1} \bar{\lambda}$. When $B_p' S_{\varepsilon_p}^{-1} B_p$ is large (i.e., when $\xi_{\min}(B_p' S_{\varepsilon_p}^{-1} B_p)$ is large), $[(B_p' S_{\varepsilon_p}^{-1} B_p)^{-1} + S_{\delta}]^{-1}$ is small and the χ^2 statistic will be large as well. *Ceteris paribus*, reductions in the idiosyncratic variances and increases in the factor loadings of the basis portfolios will increase the percentage of basis portfolio return variance explained by $\tilde{\delta}_t$ and, thus, the χ^2 statistic. Hence, the usual χ^2 statistic for testing the joint significance of mean excess basis portfolio returns can rank different combinations of portfolio formation procedures and factor loading estimation methods.

For example, in the one factor case discussed earlier, the χ^2 statistics for the Fama-MacBeth and minimum idiosyncratic risk portfolios with no measurement error in β are (approximately):

$$\begin{aligned} \chi_{FM}^2 &\approx T \frac{\beta' \beta^2 \lambda^2}{\beta' \beta^2 \sigma_{\delta}^2 + \beta' \beta \sigma_{\varepsilon}^2} = T \frac{\lambda^2}{\sigma_{\delta}^2 + \frac{\sigma_{\varepsilon}^2}{N \bar{\beta}^2} (\sigma_{\beta}^2 + \bar{\beta}^2)} \\ \chi_{MIRP}^2 &\approx T \frac{\bar{\beta}^2 \lambda^2}{\bar{\beta}^2 \sigma_{\delta}^2 + \frac{\sigma_{\varepsilon}^2}{N}} = T \frac{\lambda^2}{\sigma_{\delta}^2 + \frac{\sigma_{\varepsilon}^2}{N \bar{\beta}^2}} \end{aligned} \quad (29)$$

That is, the Fama-MacBeth portfolio performs better as long as $\sigma_{\beta}^2 > 0$. However, the inequality

can be reversed when there is measurement error in b since:

$$\begin{aligned}\chi_{FM}^2 &\approx T \frac{\beta' \beta^2 \lambda^2}{\beta' \beta^2 \sigma_\delta^2 + (1 + \frac{1}{T}) \beta' \beta \sigma_\varepsilon^2 + \frac{\sigma_\varepsilon^2}{T \sigma_\delta^2} (\beta' \beta^2 \lambda^2 + N \sigma_\varepsilon^2)} \\ &\approx T \frac{\lambda^2}{\sigma_\delta^2 + \frac{\sigma_\varepsilon^2}{N \bar{\beta}^2} \frac{\bar{\beta}^2}{(\sigma_\beta^2 + \bar{\beta}^2)} \left\{ (1 + \frac{1}{T}) + \frac{N}{T \sigma_\delta^2} [1 + \lambda^2 (\sigma_\beta^2 + \bar{\beta}^2)] \right\}}\end{aligned}\quad (30)$$

This statistic will be smaller than χ_{MIRP}^2 if $\sigma_\beta^2 < \bar{\beta}^2 \left\{ \frac{1}{T} + \frac{N}{T \sigma_\delta^2} [1 + \lambda^2 (\sigma_\beta^2 + \bar{\beta}^2)] \right\}$, which will generically occur for N sufficiently large. Sampling error can indeed cause minimum idiosyncratic risk portfolios to outperform their Fama-MacBeth counterparts.

This reasoning points to a problem with testing the APT after using the χ^2 statistic to assess basis portfolio performance. Consider the fitted multivariate regression of $\tilde{R}_t - \iota \lambda_0$ on $\tilde{R}_{pt} - \iota \lambda_0$ and a constant:

$$\tilde{R}_t - \iota \lambda_0 = \hat{\alpha} + \hat{B}(\tilde{R}_{pt} - \iota \lambda_0) + \hat{\varepsilon}_t \quad (31)$$

where $\hat{\alpha}$ is the vector of estimated intercepts, \hat{B} is the estimated factor loading matrix, and $\hat{\varepsilon}_t$ is the fitted residual vector. The population value of α will be zero if the APT is true and the basis portfolios measure the common factors with negligible error. The usual χ^2 for testing this hypothesis is:

$$\begin{aligned}T \hat{\alpha}' \hat{\Sigma}_\varepsilon^{-1} \hat{\alpha} &= T [\bar{R} - \iota \lambda_0 - \hat{B}(\bar{R}_p - \iota \lambda_0)]' \hat{\Sigma}_\varepsilon^{-1} [\bar{R} - \iota \lambda_0 - \hat{B}(\bar{R}_p - \iota \lambda_0)] \\ &= T [(\bar{R} - \iota \lambda_0)' S^{-1} \hat{\Sigma}_\varepsilon \hat{\Sigma}_\varepsilon^{-1} \hat{\Sigma}_\varepsilon^{-1} S^{-1} (\bar{R} - \iota \lambda_0)] \\ &= T (\bar{R} - \iota \lambda_0)' S^{-1} (\bar{R} - \iota \lambda_0) - T (\bar{R}_p - \iota \lambda_0)' S_p^{-1} (\bar{R}_p - \iota \lambda_0)\end{aligned}\quad (32)$$

where $\hat{\Sigma}_\varepsilon$ is the sample residual covariance matrix of $\hat{\varepsilon}_t$. The last equality follows directly from three observations: (1) $\hat{\alpha} = \hat{\Sigma}_\varepsilon^{-1} S^{-1} (\bar{R} - \iota \lambda_0)$; (2) $\hat{B} = S \omega_p (\omega_p' S \omega_p)^{-1}$; and (3) $\hat{\Sigma}_\varepsilon = S - S \omega_p' (\omega_p' S \omega_p)^{-1} \omega_p S$

where ω_p is the $N \times K$ matrix of portfolio weights of the true basis portfolios. This representation makes economic sense: the difference in squared Sharpe ratios measures the performance improvement in mean-variance space arising from the addition of the individual securities to the basis portfolios, which will be zero, apart from sampling error, if $\alpha = 0$.¹²

Thus the χ^2 statistic for testing the null hypothesis $\alpha = 0$ is the difference between the χ^2 statistics for testing the joint significance of two sets of mean excess returns, those of individual securities and those of the basis portfolios. Choosing the basis portfolios with the largest χ^2 statistic minimizes the test statistic for $\alpha = 0$ and thus reduces its power, although the magnitude of the bias when the APT is false cannot be analyzed without further assumptions. Fortunately, this problem can be mitigated to a considerable extent by using known empirical anomalies such as those associated with market to book, firm size and dividend yield to increase the power of tests of the APT since basis portfolios tend to be well-diversified while the anomalies are not. That is, characteristics such as *small* market to book or firm size, *zero* dividend yield, or *high* dividend yield are clearly not distributed uniformly over securities.

Finally, it is worth emphasizing a major limitation of the analysis. It is not possible to test analytically whether the differences in χ^2 statistics are significant at conventional levels because they are dependent across portfolio formation and estimation procedures.¹³ Hence, the results that follow should be thought of as suggestive ones to be weighed along with other evidence such as the sensitivity of absolute and relative mutual fund performance measures to alternative basis portfolio construction methods as documented in Lehmann and Modest (1987).

5. Empirical Results

Evaluation of basis portfolio performance requires consideration of the frequency of

¹² This was first noted in Jobson and Korkie (1982).

observation, the sample period, the asset menu, and the number of postulated factors. The main virtue of daily data is the increased precision of second moment estimates while the biases in mean returns associated with bid-ask spreads and the biases in second moments arising from non-trading and thin trading constitute their principal defects. As in most investigations of the APT, we opted for the putative virtues of a large sample and used daily data to estimate the factor loadings and idiosyncratic variances.

We estimated factor models for four subperiods covered by the CRSP daily returns file: 1963 through 1967, 1968 through 1972, 1973 through 1977, and 1978 through 1982. In each period, we confined our attention to continuously listed firms in order to have the same number of observations for each security and ignored any potential selection bias associated with this choice. This yielded samples of 1001, 1350, 1350, and 1346 securities and 1259, 1234, 1263, and 1264 daily observations in these four subperiods, respectively. To guard against any biases induced by the natural progression of letters (General Dynamics, General Electric, etc.), we randomly reordered the securities in each subperiod.

We also made choices as to the number of securities and the number of factors included in the analysis. In order to study the impact of the size of the asset menu on reference portfolio performance, we estimated factor models for the first 30, 250, and 750 securities in our randomly sampled data files for each period after seeing little improvement in experiments with larger cross-sections. We chose to remain agnostic about the true number of factors and examined models containing five, ten, and fifteen factors, although for obvious reasons we did not estimate a fifteen factor model with only thirty securities. We also normalized the models so that the factors were uncorrelated with unit variances and that $B'D^{-1}B$ was diagonal. This is the standard

¹³ The bootstrap could be employed here under suitable regularity conditions.

normalization in the literature, although it is worth noting that Fama-MacBeth portfolios are not invariant with respect to normalization.

We first provide evidence on the diversification properties of alternative combinations of Fama-MacBeth and minimum idiosyncratic risk portfolios and estimation methods, omitting the quadratic programming portfolios since they are well-diversified by construction. There are several ways to quantify the diversification properties of reference portfolio weights. We confine our attention to one simple summary measure: the sum of squared portfolio weights, which converges to zero as the number of assets grows without bound when a portfolio is well-diversified. The minimum sum that can be attained with portfolio weights that sum to one is the inverse of the number of securities: $0.0333\bar{3}$ for thirty securities, 0.004 for 250 securities, and $0.00133\bar{3}$ for 750 securities.

Table 1 summarizes the diversification properties of Fama-MacBeth and minimum idiosyncratic risk portfolios across estimation methods, numbers of assets, and factors. For each combination, we report two numbers: the average sum of squared portfolio weights across both factors and sample periods and the sample standard deviation of the sum of squared portfolio weights. These quantities are given by:

$$\overline{w'_k w_k} = \frac{1}{T} \sum_{t=1}^T \frac{1}{K} \sum_{k=1}^K w'_{kt} w_{kt} = \frac{1}{TK} \sum_{t=1}^T \sum_{k=1}^K \sum_{i=1}^N w_{ikt}^2; \quad Var(\overline{w'_k w_k}) = \frac{1}{TK} \sum_{t=1}^T \sum_{k=1}^K (w'_{kt} w_{kt} - \overline{w'_k w_k})^2 \quad (33)$$

where k indexes factors, t refers to time periods (where T=4), and i indexes firms. Obviously, these measures are descriptive and are not appropriate for inference without further assumptions.

The overwhelming message of Table I is that Fama-MacBeth portfolios formed under the usual factor model normalization are extremely poorly diversified. In contrast, minimum idiosyncratic risk portfolios proved to be quite well-diversified with mean sums of squared weights only ten to twenty times the minimum attainable ones. The contrast is striking — Fama-

MacBeth portfolios yielded values of $\overline{w_k'w_k}$ between 40 and 50,000 times those of minimum idiosyncratic risk portfolios. Examination of the results for individual basis portfolios (not reported here) reveals that similar differences occur in the disaggregated data as well. In fact, every minimum idiosyncratic risk portfolio was better diversified than its Fama-MacBeth counterpart. The evidence is less clear on the comparative performance of different estimation methods. Minimum idiosyncratic risk portfolios formed from principal components estimates appear to be slightly better diversified than those computed from the other three estimation methods but the difference is slight and is reversed often when individual basis portfolios are compared.

What accounts for the sharp contrast in the diversification behavior of these two portfolio formation methods? As noted earlier, the answer is scaling: Fama-MacBeth portfolios are the minimum idiosyncratic risk portfolios with a loading of one on one factor and loadings of zero on the other factors prior to rescaling to cost a dollar. Factor loading estimates as usually normalized are typically much smaller than one (on the order of 0.001 to 0.0001 in daily data) and, hence, some weights must be very large and positive to insure a portfolio loading of one on the factor being mimicked while others must be large and negative in order to have loadings of zero on the other factors. This problem does not arise in the CAPM context since the natural beta on a market proxy is one. By contrast, minimum idiosyncratic risk portfolios need not have any particular loading on the factor being mimicked and they need only have small positive and negative weights to insure orthogonality to the other factors.

The remainder of this section is devoted to the evaluation of alternative methods using the χ^2 statistic (28). Tables 2 through 4 contain aggregate χ^2 statistics summed over the four subperiods, a valid procedure since sums of independent χ^2 statistics are distributed χ^2 as well

and the subperiod statistics are independent by assumption. Few nontrivial differences in performance are obscured by aggregation and we summarize any relevant ones in the text. Table 2 provides the fitted χ^2 statistics based on daily data and five, ten, and fifteen factor models for various combinations of portfolio formation procedures, estimation methods, and number of securities in the asset universe. Table 3 presents the same statistics based on daily returns from the subsequent five-year period to guard against any overfitting caused by using the same sample to estimate the factor models and to compute the χ^2 statistics.¹⁴ These statistics are based on a smaller number of securities since not all securities were continuously listed during both five year periods. Table 4 contains the same statistics computed from weekly returns in order to mitigate any bid-ask spread bias that might inflate mean basis portfolio returns.

Each table contains a plenitude of information. Panels A, B, and C report on 5, 10, and 15 factor models, respectively. The four columns listed at the top of each table correspond to the four portfolio formation methods under consideration: minimum idiosyncratic risk, Fama-MacBeth, positive weight quadratic programming, and well-diversified quadratic programming. Eight combinations of estimation methods and numbers of securities comprise the rows of each table: maximum likelihood and restricted maximum likelihood factor analysis with 30, 250, and 750 securities and principal components and instrumental variables with 750 securities. The χ^2 statistic for the null hypothesis that the raw and excess mean returns of the K basis portfolios are all zero along with the associated marginal significance level (in parentheses underneath it) is reported for each portfolio formation method, estimation method, and number of securities. The excess returns are the difference between raw reference portfolio returns and minimum idiosyncratic risk orthogonal portfolio returns. Note that the χ^2 statistics for the excess returns of

¹⁴ Hence, there are no results for the first five year period.

the minimum idiosyncratic risk and Fama-MacBeth portfolios are identical as noted earlier.¹⁵

The results strongly suggest the importance of using large cross-sections to construct basis portfolios. The 750 security basis portfolios always outperformed their 250 security counterparts which, in turn, always dominated 30 security portfolios. The typical χ^2 statistics for 250 and 750 securities were roughly twice and three times as large, respectively, as those for 30 securities. These rankings hold across estimation and portfolio formation procedures, time periods, and observation intervals for both raw return and excess return portfolios (not all of which are reported here). These differences increase with the number of factors in sample in the daily data, although this finding does not persist in the weekly data or out of sample. Note that this need not have occurred — we have found randomly selected 250 security portfolios with larger χ^2 statistics than those of 750 securities.

For example, the χ^2 statistics for the minimum idiosyncratic risk basis portfolio mean excess returns based on unrestricted maximum likelihood factor analysis were 35.20, 81.14, and 116.88 using 30, 250, and 750 securities, respectively. The mean returns of the 30 security raw return basis portfolios were insignificant at the ten per cent level for two out of four subperiods in daily data while the corresponding mean excess returns were insignificant in three out of four periods. In contrast, the 250 and 750 security raw return portfolios were highly significant at conventional levels in all subperiods while the excess return portfolios were highly significant in daily data for all but the five factor model in the third period.

The four estimation methods exhibited similarly striking contrasts. Restricted maximum

¹⁵ Recall that the Fama-MacBeth and minimum idiosyncratic risk excess return portfolio weights are proportional where the factor of proportionality is the inverse of the average factor loading in the one factor model. Fama-MacBeth portfolio weights tend to be much larger since the average loading is typically much smaller than one. However, the proportionality factor cancels out in the construction of the χ^2 statistic.

likelihood factor analysis systematically outperformed its unconstrained counterpart in daily and weekly data but this dominance did not persist out of sample, where they achieved almost identical performance. Both maximum likelihood procedures consistently outperformed the less efficient instrumental variables and principal components methods. Consider, for example, ten factor minimum idiosyncratic risk excess return portfolio results in daily data in sample: the χ^2 statistics based on restricted and unrestricted maximum likelihood factor analysis, principal components, and instrumental variables were 143.92, 116.88, 97.88, and 82.94, respectively. Perusal of the tables indicates that the differences in χ^2 statistics across estimation methods were typically much larger for ten and fifteen factor models than for five factor models. The superiority of the relatively efficient maximum likelihood procedures also persisted out of sample and in weekly data. The four estimation methods generated χ^2 statistics of 53.82, 53.50, 37.13 and 39.07 out of sample and 90.33, 73.83, 53.08, and 62.48 in weekly data, respectively.

An examination of the subperiod results mostly supports this conclusion, albeit with some interesting intertemporal variation in performance. Principal components performed well in-sample only in the first five year period while achieving performance comparable to the 250 security basis portfolios in the subsequent three periods. Instrumental variables provided more consistent performance than principal components in-sample but was consistently inferior to the more efficient estimation procedures. The out-of-sample results were broadly consistent with these findings with some idiosyncrasies in the individual subperiods. In five factor models, the efficient methods proved superior in the second subperiod while all methods yielded similar performance in the first and third subperiods. In ten factor models, all methods provided similar performance in the first and second subperiods except for the superior performance of instrumental variables in the first and its inferior performance in the second. The efficient

methods outperformed the inefficient methods by a wide margin in the final out-of-sample period. Similar inconsistencies emerged in the fifteen factor runs. The methods provided similar performance in the first out-of-sample period except for the inferior results provided by principal components. In the second subperiod, only instrumental variables provided inferior results to the similar performance of the other three methods. Again, the efficient methods outperformed the inefficient methods in the final subperiod.

The final comparisons are among the four portfolio formation methods. Minimum idiosyncratic risk portfolios provided almost identical performance to well-diversified quadratic programming portfolios using 750 securities and provided consistently superior performance with 250 securities. Other observations depend on the estimation method and choice of raw or excess returns.¹⁶ For the efficient estimation methods and raw returns, minimum idiosyncratic risk portfolios consistently dominated Fama-MacBeth portfolios in-sample in both daily and weekly data which, in turn, typically outperformed the positive weight quadratic programming portfolios. For instance, the daily in-sample χ^2 statistics for five factor minimum idiosyncratic risk and Fama-MacBeth raw return portfolios based on maximum likelihood factor analysis were 131.96 and 92.28, respectively. The corresponding numbers for ten factors were 201.21 and 144.94, respectively, and for fifteen factors were 227.77 and 168.24, respectively. However, Fama-MacBeth portfolios slightly outperformed minimum idiosyncratic risk portfolios out-of-sample while both continued to dominate the positive weight quadratic programming portfolios. The picture regarding the relative merits of the minimum idiosyncratic risk, Fama-MacBeth, and quadratic programming excess return basis portfolios was virtually identical for the efficient

¹⁶ As noted above, the minimum idiosyncratic risk and Fama-MacBeth portfolio formation procedures yield identical χ^2 statistics for excess return basis portfolios. However, the χ^2 statistics

estimation methods. The only difference is that the positive weight quadratic programming portfolios provided superior performance for five factor models in-sample, an improvement which did not persist out-of-sample.

The final question considered here is whether less efficient estimation methods or smaller cross-sections coupled with quadratic programming provides a good substitute for more computationally costly alternatives. With two exceptions, no such combination outperformed minimum idiosyncratic risk portfolios of 750 securities based on efficient estimates. Both exceptions involve five factor excess return portfolios in-sample in daily data. First, positive weight quadratic programming portfolios based on instrumental variables estimates yielded a χ^2 statistic of 77.64 compared to 64.40 obtained from minimum idiosyncratic risk portfolios based on unrestricted maximum likelihood estimates. Second, 250 security quadratic programming portfolios based on efficient estimates outperformed 750 security minimum idiosyncratic risk basis portfolios in-sample. Neither result obtained out-of-sample.

6. Conclusion

This paper has provided a comprehensive examination of the merits of different basis portfolio formation strategies. The analysis involved the main factor model estimation and portfolio formation methods proposed in the literature as well as some that have not been considered previously. In addition, this study provided a detailed evaluation of the impact of increases in the number of securities underlying the analysis. The result is a detailed set of data measuring the performance of excess and raw return basis portfolios both in and out of sample over a variety of time periods and observation frequencies.

Three conclusions emerge from the examination of the more than 2300 statistics reported

of the raw return basis portfolios are relevant if excess returns are computed with respect to a

on in this document. First, increasing the number of securities included in the analysis dramatically improves basis portfolio performance with 750 security factor models proving markedly superior to those involving 30 or 250 securities. Second, comparatively efficient estimation procedures such as unrestricted and restricted maximum likelihood factor analysis significantly outperform the less efficient instrumental variables and principal components procedures. In particular, less efficient estimation methods typically performed as well as maximum likelihood factor analysis with 250 securities. Third, the minimum idiosyncratic risk portfolio formation procedure proposed in the second section outperformed both the Fama-MacBeth and positive weight quadratic programming portfolios and proved equal to or better than the more expensive well-diversified quadratic programming procedure. The Fama-MacBeth procedure, the dominant method in the literature, yielded poorly diversified portfolios that provided inferior performance in this context, albeit performance that can be improved upon with more appropriate factor model normalizations. In sum, if an investigator had to choose one basis portfolio formation strategy from the formidable list considered here, the clear winner is minimum idiosyncratic risk portfolios coupled with maximum likelihood factor analysis of 750 securities.

given riskless rate such as the Treasury bill rate.

TABLE 1: Average Sum of Squared Portfolio Weights
 [Standard deviation of sums in parentheses]

| Number of Securities | Estimation Method | Five Factors | | Ten Factors | | Fifteen Factors | |
|----------------------|-------------------------------|----------------------------|------------------------|----------------------------|------------------------|----------------------------|-----------------------|
| | | Minimum Idiosyncratic Risk | Fama-MacBeth | Minimum Idiosyncratic Risk | Fama-MacBeth | Minimum Idiosyncratic Risk | Fama-MacBeth |
| 30 | Maximum Likelihood | 0.397 (0.494) | 75.011 (459.206) | 1.276 (6.063) | 98.150 (846.203) | N/A | N/A |
| | Restricted Maximum Likelihood | 0.399 (0.496) | 2178.232 (16057.81) | 1.257 (5.941) | 273.641 (2781.711) | N/A | N/A |
| 250 | Maximum Likelihood | 0.048 (0.049) | 4.491 (28.928) | 0.063 (0.045) | 51.643 (551.863) | 0.071 (0.041) | 502.614 (4253.059) |
| | Restricted Maximum Likelihood | 0.048 (0.049) | 5.209 (34.645) | 0.063 (0.045) | 195.268 (2277.061) | 0.071 (0.041) | 81.613 (492.380) |
| 750 | Maximum Likelihood | 0.016 (0.016) | 0.659 (4.297) | 0.022 (0.015) | 72.689 (798.788) | 0.025 (0.014) | 255.234 (3267.704) |
| | Restricted Maximum Likelihood | 0.016 (0.016) | 0.659 (4.310) | 0.022 (0.015) | 22.570 (189.855) | 0.025 (0.014) | 46.992 (469.129) |
| | Principal Components | 0.013 (0.019) | 1.391 (5.784) | 0.017 (0.035) | 26.992 (251.929) | 0.019 (0.043) | 244.917 (2912.448) |
| | Instrumental Variables | 0.015 (0.018) | 77.021 (605.504) | 0.019 (0.013) | 1087.395 (9194.892) | 0.022 (0.012) | 93.953 (1310.661) |

TABLE 2: Basis Portfolio Comparisons: Aggregate Daily In-Sample χ^2 Statistics
[p-values in parentheses]

| | | Portfolio Formation Procedure | | | | | | | |
|--------------------------|-------------------------------|-------------------------------|-------------------|-------------------|-------------------|-----------------------|-------------------|-------------------|-------------------|
| | | Minimum Idiosyncratic Risk | | Fama-MacBeth | | Quadratic Programming | | | |
| | | Raw Returns | Excess Returns | Raw Returns | Excess Returns | Raw Returns | Excess Returns | Raw Returns | Excess Returns |
| Number of Securities | Estimation Method | Raw Returns | Excess Returns | Raw Returns | Excess Returns | Raw Returns | Excess Returns | Raw Returns | Excess Returns |
| Panel A: Five Factors | | | | | | | | | |
| 30 | Maximum Likelihood | 47.37 (0.001) | 26.30 (0.160) | 41.96 (0.003) | 26.30 (0.160) | 42.83 (0.002) | 33.11 (0.033) | N/A | N/A |
| | Restricted Maximum Likelihood | 50.22 (0.000) | 29.68 (0.075) | 44.44 (0.001) | 29.68 (0.075) | 42.06 (0.003) | 32.81 (0.035) | N/A | N/A |
| 250 | Maximum Likelihood | 102.10 (0.000) | 61.53 (0.000) | 83.76 (0.000) | 61.53 (0.000) | 81.38 (0.000) | 75.61 (0.000) | 86.40 (0.000) | 50.89 (0.000) |
| | Restricted Maximum Likelihood | 108.26 (0.000) | 70.09 (0.000) | 91.08 (0.000) | 70.09 (0.000) | 83.32 (0.000) | 80.03 (0.000) | 91.18 (0.000) | 58.29 (0.000) |
| 750 | Maximum Likelihood | 131.96 (0.000) | 64.49 (0.000) | 92.28 (0.000) | 64.49 (0.000) | 106.85 (0.000) | 76.32 (0.000) | 131.93 (0.000) | 64.32 (0.000) |
| | Restricted Maximum Likelihood | 138.26 (0.000) | 73.89 (0.000) | 99.66 (0.000) | 73.89 (0.000) | 118.77 (0.000) | 78.42 (0.000) | 138.13 (0.000) | 73.77 (0.000) |
| | Principal Components | 109.00 (0.000) | 57.34 (0.000) | 84.18 (0.000) | 57.34 (0.000) | 92.93 (0.000) | 37.57 (0.000) | 108.17 (0.000) | 57.08 (0.000) |
| | Instrumental Variables | 122.81 (0.000) | 54.53 (0.000) | 70.49 (0.000) | 54.53 (0.000) | 126.04 (0.000) | 77.64 (0.000) | 125.66 (0.000) | 55.39 (0.000) |
| Panel B: Ten Factors | | | | | | | | | |
| 30 | Maximum Likelihood | 54.59 (0.062) | 35.20 (0.690) | 50.17 (0.130) | 35.20 (0.690) | 53.63 (0.073) | 46.85 (0.210) | N/A | N/A |
| | Restricted Maximum Likelihood | 57.77 (0.034) | 37.70 (0.570) | 52.41 (0.090) | 37.70 (0.570) | 25.79 (0.690) | 19.66 (0.930) | N/A | N/A |
| 250 | Maximum Likelihood | 129.94 (0.000) | 81.14 (0.000) | 102.78 (0.000) | 81.14 (0.000) | 103.64 (0.000) | 87.96 (0.000) | 105.96 (0.000) | 63.49 (0.010) |
| | Restricted Maximum Likelihood | 154.72 (0.000) | 105.57 (0.000) | 124.81 (0.000) | 105.57 (0.000) | 113.10 (0.000) | 95.79 (0.000) | 124.08 (0.000) | 81.28 (0.000) |
| 750 | Maximum Likelihood | 201.21 (0.000) | 116.88 (0.000) | 144.94 (0.000) | 116.88 (0.000) | 133.84 (0.000) | 101.76 (0.000) | 201.68 (0.000) | 115.83 (0.000) |
| | Restricted Maximum Likelihood | 225.06 (0.000) | 143.92 (0.000) | 171.22 (0.000) | 143.92 (0.000) | 140.36 (0.000) | 112.74 (0.000) | 225.40 (0.000) | 142.58 (0.000) |
| | Principal Components | 143.62 (0.000) | 97.88 (0.000) | 124.14 (0.000) | 97.88 (0.000) | 123.07 (0.000) | 74.52 (0.000) | 143.64 (0.000) | 97.73 (0.000) |
| | Instrumental Variables | 160.28 (0.000) | 82.94 (0.000) | 102.36 (0.000) | 82.94 (0.000) | 139.14 (0.000) | 97.84 (0.000) | 162.13 (0.000) | 85.61 (0.000) |
| Panel C: Fifteen Factors | | | | | | | | | |
| 250 | Maximum Likelihood | 136.61 (0.000) | 90.57 (0.007) | 114.23 (0.000) | 90.57 (0.007) | 95.41 (0.003) | 78.08 (0.058) | 115.33 (0.000) | 73.55 (0.110) |
| | Restricted Maximum Likelihood | 162.99 (0.000) | 116.21 (0.000) | 138.54 (0.000) | 116.21 (0.000) | 97.66 (0.002) | 78.55 (0.054) | 134.88 (0.000) | 93.32 (0.004) |
| 750 | Maximum Likelihood | 227.77 (0.000) | 137.59 (0.000) | 168.24 (0.000) | 137.59 (0.000) | 142.48 (0.000) | 112.92 (0.000) | 228.59 (0.000) | 136.25 (0.000) |
| | Restricted Maximum Likelihood | 261.81 (0.000) | 174.62 (0.000) | 204.05 (0.000) | 174.62 (0.000) | 144.82 (0.000) | 116.15 (0.000) | 261.14 (0.000) | 172.99 (0.000) |
| | Principal Components | 169.59 (0.000) | 112.50 (0.000) | 139.73 (0.000) | 112.50 (0.000) | 139.94 (0.000) | 92.04 (0.000) | 168.96 (0.000) | 112.36 (0.000) |
| | Instrumental Variables | 168.29 (0.000) | 106.75 (0.000) | 135.14 (0.000) | 106.75 (0.000) | 149.52 (0.000) | 109.47 (0.000) | 174.90 (0.000) | 109.25 (0.000) |

TABLE 3: Basis Portfolio Comparisons: Aggregate Daily Out-of-Sample χ^2 Statistics
 [p-values in parentheses]

| | | Portfolio Formation Procedure | | | | | | | |
|--------------------------|-------------------------------|-------------------------------|------------------|-------------------|------------------|-----------------------|------------------|-------------------|------------------|
| | | Minimum Idiosyncratic Risk | | Fama-MacBeth | | Quadratic Programming | | | |
| | | Raw Returns | Excess Returns | Raw Returns | Excess Returns | Raw Returns | Excess Returns | Raw Returns | Excess Returns |
| Number of Securities | Estimation Method | Raw Returns | Excess Returns | Raw Returns | Excess Returns | Raw Returns | Excess Returns | Raw Returns | Excess Returns |
| Panel A: Five Factors | | | | | | | | | |
| 250 | Maximum Likelihood | 40.98 (0.000) | 17.95 (0.270) | 44.24 (0.000) | 17.95 (0.270) | 32.86 (0.005) | 19.04 (0.210) | 37.60 (0.001) | 14.11 (0.520) |
| | Restricted Maximum Likelihood | 41.52 (0.000) | 18.37 (0.240) | 44.61 (0.000) | 18.37 (0.240) | 31.82 (0.007) | 18.45 (0.240) | 37.81 (0.001) | 14.18 (0.510) |
| 750 | Maximum Likelihood | 60.29 (0.000) | 33.03 (0.005) | 63.32 (0.000) | 33.03 (0.005) | 35.66 (0.002) | 27.03 (0.028) | 61.55 (0.000) | 33.13 (0.005) |
| | Restricted Maximum Likelihood | 60.79 (0.000) | 33.27 (0.004) | 63.61 (0.000) | 33.27 (0.004) | 35.83 (0.002) | 27.14 (0.028) | 62.02 (0.000) | 33.36 (0.004) |
| | Principal Components | 57.70 (0.000) | 24.96 (0.050) | 41.82 (0.000) | 24.96 (0.050) | 36.24 (0.002) | 16.31 (0.360) | 57.69 (0.000) | 24.68 (0.054) |
| | Instrumental Variables | 49.62 (0.000) | 20.90 (0.140) | 50.26 (0.000) | 20.90 (0.140) | 34.93 (0.003) | 15.86 (0.390) | 49.77 (0.000) | 20.76 (0.140) |
| Panel B: Ten Factors | | | | | | | | | |
| 250 | Maximum Likelihood | 56.83 (0.002) | 31.11 (0.410) | 58.57 (0.001) | 31.11 (0.410) | 34.61 (0.260) | 33.23 (0.310) | 46.04 (0.031) | 21.24 (0.880) |
| | Restricted Maximum Likelihood | 58.27 (0.002) | 31.83 (0.380) | 59.07 (0.001) | 31.83 (0.380) | 34.23 (0.270) | 29.15 (0.510) | 46.08 (0.031) | 23.07 (0.810) |
| 750 | Maximum Likelihood | 80.67 (0.000) | 53.50 (0.005) | 85.09 (0.000) | 53.50 (0.005) | 45.83 (0.032) | 38.83 (0.130) | 81.49 (0.000) | 53.82 (0.005) |
| | Restricted Maximum Likelihood | 80.90 (0.000) | 53.82 (0.005) | 85.41 (0.000) | 53.82 (0.005) | 44.73 (0.041) | 37.90 (0.150) | 81.69 (0.000) | 54.14 (0.004) |
| | Principal Components | 70.36 (0.000) | 37.13 (0.170) | 60.88 (0.001) | 37.13 (0.170) | 43.33 (0.055) | 28.18 (0.560) | 71.22 (0.000) | 37.29 (0.170) |
| | Instrumental Variables | 65.83 (0.000) | 39.07 (0.120) | 69.75 (0.000) | 39.07 (0.120) | 45.44 (0.035) | 37.13 (0.170) | 72.15 (0.000) | 40.07 (0.100) |
| Panel C: Fifteen Factors | | | | | | | | | |
| 250 | Maximum Likelihood | 65.45 (0.025) | 40.44 (0.670) | 66.93 (0.019) | 40.44 (0.670) | 42.61 (0.570) | 38.51 (0.740) | 57.51 (0.100) | 31.25 (0.940) |
| | Restricted Maximum Likelihood | 64.59 (0.029) | 40.58 (0.660) | 67.18 (0.018) | 40.58 (0.660) | 40.95 (0.640) | 39.15 (0.720) | 56.84 (0.110) | 30.87 (0.950) |
| 750 | Maximum Likelihood | 101.63 (0.000) | 66.13 (0.022) | 99.86 (0.000) | 66.13 (0.022) | 57.96 (0.093) | 53.17 (0.190) | 99.26 (0.000) | 64.77 (0.028) |
| | Restricted Maximum Likelihood | 101.42 (0.000) | 66.62 (0.020) | 100.36 (0.000) | 66.62 (0.020) | 57.82 (0.095) | 54.54 (0.160) | 101.22 (0.000) | 65.29 (0.026) |
| | Principal Components | 79.33 (0.001) | 46.25 (0.420) | 73.68 (0.005) | 46.25 (0.420) | 47.08 (0.390) | 34.31 (0.880) | 80.49 (0.001) | 46.44 (0.410) |
| | Instrumental Variables | 91.62 (0.000) | 53.26 (0.190) | 88.85 (0.000) | 53.26 (0.190) | 40.06 (0.680) | 26.46 (0.990) | 90.45 (0.000) | 53.11 (0.190) |

TABLE 4: Basis Portfolio Comparisons: Aggregate Weekly In-Sample χ^2 Statistics
[p-values in parentheses]

| | | Portfolio Formation Procedure | | | | | | | |
|--------------------------|-------------------------------|-------------------------------|-------------------|-------------------|-------------------|-----------------------|------------------|-------------------|-------------------|
| | | Minimum Idiosyncratic Risk | | Fama-MacBeth | | Quadratic Programming | | | |
| | | Raw Returns | Excess Returns | Raw Returns | Excess Returns | Raw Returns | Excess Returns | Raw Returns | Excess Returns |
| Number of Securities | Estimation Method | Raw Returns | Excess Returns | Raw Returns | Excess Returns | Raw Returns | Excess Returns | Raw Returns | Excess Returns |
| Panel A: Five Factors | | | | | | | | | |
| 30 | Maximum Likelihood | 34.95 (0.020) | 26.51 (0.150) | 31.47 (0.049) | 26.51 (0.150) | 31.55 (0.049) | 26.61 (0.015) | N/A | N/A |
| | Restricted Maximum Likelihood | 37.24 (0.011) | 29.62 (0.076) | 33.89 (0.027) | 29.62 (0.076) | 30.77 (0.058) | 26.35 (0.150) | N/A | N/A |
| 250 | Maximum Likelihood | 65.69 (0.000) | 42.12 (0.003) | 53.50 (0.000) | 42.12 (0.003) | 44.71 (0.001) | 37.02 (0.012) | 55.08 (0.000) | 36.01 (0.015) |
| | Restricted Maximum Likelihood | 69.89 (0.000) | 47.45 (0.001) | 58.02 (0.000) | 47.45 (0.001) | 47.85 (0.000) | 40.58 (0.004) | 58.73 (0.000) | 41.02 (0.004) |
| 750 | Maximum Likelihood | 77.69 (0.000) | 42.20 (0.003) | 55.48 (0.000) | 42.20 (0.0030) | 56.83 (0.000) | 55.80 (0.000) | 78.09 (0.000) | 42.31 (0.003) |
| | Restricted Maximum Likelihood | 82.27 (0.000) | 47.76 (0.000) | 59.97 (0.005) | 47.76 (0.000) | 57.64 (0.000) | 56.00 (0.000) | 82.79 (0.000) | 47.98 (0.000) |
| | Principal Components | 49.69 (0.000) | 28.09 (0.110) | 40.95 (0.000) | 28.09 (0.110) | 37.43 (0.100) | 20.03 (0.460) | 49.56 (0.000) | 28.00 (0.110) |
| | Instrumental Variables | 71.49 (0.000) | 37.87 (0.009) | 42.56 (0.009) | 37.87 (0.009) | 51.41 (0.000) | 45.35 (0.001) | 71.94 (0.000) | 37.93 (0.009) |
| Panel B: Ten Factors | | | | | | | | | |
| 30 | Maximum Likelihood | 43.59 (0.32) | 34.94 (0.700) | 39.72 (0.480) | 34.94 (0.700) | 43.65 (0.320) | 40.91 (0.430) | N/A | N/A |
| | Restricted Maximum Likelihood | 45.81 (0.240) | 37.32 (0.590) | 41.69 (0.400) | 37.32 (0.590) | 23.41 (0.800) | 19.84 (0.920) | N/A | N/A |
| 250 | Maximum Likelihood | 87.22 (0.000) | 58.69 (0.028) | 71.11 (0.002) | 58.69 (0.028) | 67.18 (0.005) | 61.19 (0.017) | 71.87 (0.002) | 50.68 (0.120) |
| | Restricted Maximum Likelihood | 105.24 (0.000) | 75.51 (0.001) | 87.32 (0.000) | 75.51 (0.001) | 70.63 (0.002) | 65.57 (0.007) | 84.44 (0.000) | 63.90 (0.010) |
| 750 | Maximum Likelihood | 121.29 (0.000) | 73.83 (0.001) | 88.69 (0.000) | 73.83 (0.001) | 84.07 (0.000) | 73.72 (0.001) | 122.63 (0.000) | 74.01 (0.001) |
| | Restricted Maximum Likelihood | 137.12 (0.000) | 90.33 (0.000) | 105.22 (0.000) | 90.33 (0.000) | 92.54 (0.000) | 84.43 (0.000) | 138.51 (0.000) | 90.39 (0.000) |
| | Principal Components | 76.37 (0.000) | 53.08 (0.081) | 66.00 (0.006) | 53.08 (0.081) | 61.55 (0.016) | 49.45 (0.150) | 76.40 (0.000) | 52.98 (0.082) |
| | Instrumental Variables | 104.51 (0.000) | 62.48 (0.013) | 72.03 (0.001) | 62.48 (0.013) | 85.23 (0.000) | 78.25 (0.000) | 104.88 (0.000) | 65.63 (0.007) |
| Panel C: Fifteen Factors | | | | | | | | | |
| 250 | Maximum Likelihood | 94.95 (0.003) | 68.77 (0.200) | 82.18 (0.000) | 68.77 (0.200) | 64.04 (0.340) | 54.28 (0.680) | 82.97 (0.026) | 59.71 (0.490) |
| | Restricted Maximum Likelihood | 114.98 (0.000) | 87.19 (0.012) | 100.20 (0.001) | 87.19 (0.012) | 62.89 (0.370) | 52.51 (0.740) | 101.14 (0.001) | 77.10 (0.068) |
| 750 | Maximum Likelihood | 149.60 (0.000) | 94.10 (0.003) | 109.57 (0.000) | 94.10 (0.003) | 94.55 (0.003) | 90.94 (0.006) | 151.71 (0.000) | 93.91 (0.003) |
| | Restricted Maximum Likelihood | 174.71 (0.000) | 119.72 (0.000) | 134.83 (0.000) | 119.72 (0.000) | 100.82 (0.001) | 98.09 (0.001) | 175.89 (0.000) | 119.34 (0.000) |
| | Principal Components | 94.73 (0.003) | 63.84 (0.340) | 79.67 (0.046) | 63.84 (0.340) | 80.73 (0.038) | 67.64 (0.230) | 94.68 (0.003) | 63.93 (0.340) |
| | Instrumental Variables | 112.48 (0.000) | 81.06 (0.036) | 97.52 (0.002) | 81.06 (0.036) | 98.26 (0.001) | 96.63 (0.002) | 119.52 (0.000) | 84.65 (0.020) |

References

- Chamberlain, Gary and Michael Rothschild, 1983. "Arbitrage and Mean Variance Analysis on Large Asset Markets," *Econometrica* 51, 1281-1304.
- Chen, Nai-fu, 1983. "Some Empirical Tests of the Theory of Arbitrage Pricing," *Journal of Finance* 38, 1393-1414.
- Connor, Gregory and Robert A. Korajczyk, 1986. "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis." *Journal of Financial Economics* 15, 373-394.
- Connor, Gregory and Robert A. Korajczyk, 1988. "Risk and Return in an Equilibrium APT: Application of a New Test Methodology." *Journal of Financial Economics* 21, 255-289.
- Connor, Gregory, Robert A. Korajczyk, and Robert T. Uhlaner 2002, "Sunspots, Iterative Two-Pass Cross-Sectional Regressions, and Asymptotic Principal Components," unpublished manuscript.
- Fama, Eugene F. and James D. MacBeth, 1973. "Risk, Return, and Equilibrium: Some Empirical Tests," *Journal of Political Economy* 81, pp. 607-636.
- Hagglund, Gösta, 1982. "Factor Analysis by Instrumental Variables Methods," *Psychometrika* 47, pp. 209-221.
- Jobson, J. D. and Robert M. Korkie, 1982. "Potential Performance and Tests of Portfolio Efficiency," *Journal of Financial Economics* 10, pp. 433-466.
- Jones, Christopher S., "Extracting Factors from Heteroskedastic Asset Returns." *Journal of Financial Economics* 62 (September 2001): 293-325.
- Lehmann, Bruce N., 1992. "Notes on Dynamic Factor Pricing Models." *Review of Quantitative Finance and Accounting* 2, 69-87.
- Lehmann, Bruce N. and David M. Modest, 1987. "Mutual Fund Performance Evaluation: A Comparison and Benchmarks and Benchmark Comparisons." *Journal of Finance* 42, 233-265.
- Lehmann, Bruce N. and David M. Modest, 1988. "The Empirical Foundations of the Arbitrage Pricing Theory." *Journal of Financial Economics* 21, 213-254.
- Litzenberger, Robert H. and Krishna Ramaswamy, 1979. "The Effects of Dividends on Common Stock Prices: Theory and Empirical Evidence," *Journal of Financial Economics* 7, pp. 163-195.
- Madansky, Albert, 1964. "Instrumental Variables in Factor Analysis." *Psychometrika* 29, 105-113.

Roll, Richard and Stephen A. Ross, 1980. "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance* 35, 1073-1103.

Rosenberg, Barr and Vinay Marathe, 1979. "Tests of Capital Asset Pricing Hypotheses" in *Research in Finance: A Research Annual*, Greenwich, Connecticut: JAI Press.

Ross, Stephen A., 1976. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13, 341-360.

———, 1977. "Return, Risk and Arbitrage" in Irwin Friend and James L. Bicksler (eds.) *Risk and Return in Finance*. Cambridge, MA: Ballinger.