When Will Mean-Variance Efficient Portfolios be Well Diversified?
Author(s): Richard C. Green and Burton Hollifield
Source: *The Journal of Finance*, Vol. 47, No. 5 (Dec., 1992), pp. 1785-1809
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2328996
Accessed: 25/11/2009 12:00

THE JOURNAL OF FINANCE • VOL. XLVII, NO. 5 • DECEMBER 1992

When Will Mean-Variance Efficient Portfolios Be Well Diversified?

RICHARD C. GREEN and BURTON HOLLIFIELD*

ABSTRACT

We characterize the conditions under which efficient portfolios put small weights on individual assets. These conditions bound mean returns with measures of average absolute covariability between assets. The bounds clarify the relationship between linear asset pricing models and well-diversified efficient portfolios. We argue that the extreme weightings in sample efficient portfolios are due to the dominance of a single factor in equity returns. This makes it easy to diversify on subsets to reduce residual risk, while weighting the subsets to reduce factor risk simultaneously. The latter involves taking extreme positions. This behavior seems unlikely to be attributable to sampling error.

THE MEAN-VARIANCE EFFICIENT frontier plays an important role in pedagogy and applications in finance. The properties of efficient portfolios are also central in both static and dynamic asset pricing.[1] For all its simplicity and intuitive appeal, however, practical implementation of mean-variance analysis has proved problematic. Portfolios constructed using sample moments of returns often involve very extreme positions. As the number of assets grow, the weights on individual assets do not approach zero as quickly as suggested by naive notions of "diversification."

Admittedly, the theoretical links between "diversification," in the sense of putting small weight on any asset, and variance minimization are tenuous. The Arbitrage Pricing Theory (APT) implies that efficient portfolios are well diversified, but the mean-variance Capital Asset Pricing Model (CAPM) requires this only insofar as the value weighted market portfolio is diversified. In applications, however, the connection between mean-variance efficiency and diversification is central. Black and Litterman (1990), for example, point to the inconsistency between naive diversification and the weights generated by asset allocation models as a major obstacle to their implementation. Practitioners are suspicious of portfolios that are not naively

* Green is from Carnegie Mellon University and Hollifield is from the University of British Columbia. We wish to thank seminar participants at Carnegie Mellon University, Rutgers University, Queens University, the University of Colorado at Boulder, the Stockholm School of Economics, and the 1990 Western Finance Association meetings for helpful comments on this work. The comments and guidance of the editor, René Stulz, and of two anonymous referees significantly improved the paper. Hollifield gratefully acknowledges the support of the Social Sciences and Humanities Research Council of Canada.

[1] See, for some recent examples, discussions in Chamberlain and Rothschild (1983), Hansen and Richard (1987), and Huberman, Kandel, and Stambaugh (1987).

diversified, so mean-variance methods are generally implemented with extensive sets of constraints that enforce such diversification. Numerous academic researchers have also documented problems in employing the weights that result from calculating efficient frontiers using sample moments.[2]

This paper studies the question of when mean-variance efficiency and "well diversified" coincide, and when they do not. First, we provide a theoretical answer to this question. Then, we use more specialized examples to advance possible explanations for the "poorly diversified" weights found in efficient portfolios calculated using sample return data.

Through duality theory, we provide simple and intuitive characterizations of conditions under which minimum-variance portfolios will be well diversified. These characterizations involve bounds on the means of portfolio returns in terms of their average absolute covariance with the individual assets. We use these results to study the equivalence between diversified frontier portfolios and exact APT pricing, first demonstrated in Chamberlain (1983). We show that an approximate factor structure is a stronger condition than needed for diversification to eliminate residual risk and ensure APT pricing. Intuitively, the eigenvalues of the covariance matrix can grow, if they grow more slowly than the rate at which the weights in the minimum-variance portfolios go to zero.

We then develop some simple cases and examples to better understand why efficient portfolios computed from sample moments tend to be so poorly diversified. If the extreme behavior of the weights is due simply to measurement error in the sample first and second moments, then one should employ Bayesian procedures, or shrinkage estimators, or simply rely on nonstatistical methods of estimating these quantities. Examples of these approaches include work by Frost and Savarino (1986, 1988) and Black and Litterman (1990). If, on the other hand, the behavior of the weights is due to characteristics of the population moments, then perhaps practitioners should question more critically the value of naive diversification.

Our analysis suggests the following explanation. The extreme weights in efficient portfolios are due to the dominance of a single factor in the covariance structure of returns, and the consequent high correlation between naively diversified portfolios. With small amounts of cross-sectional diversity in asset betas, well-diversified portfolios can be constructed on subsets of the assets with very little residual risk and different betas. A portfolio of these diversified portfolios can then be created that has a zero beta, thus eliminating the factor risk as well as the residual risk. Obviously, the greater the degree of domination of the first factor, the easier it is to implement this strategy on a relatively small asset universe.

As the number of assets grows to infinity, the weights that result from such a strategy will eventually grow small. The rate at which they grow small,

[2] See, for example, Kallberg and Ziemba (1983), Kroll, Levy, and Markowitz (1984), and Pulley (1981). Similarly, Best and Grauer (1991) document the extreme sensitivity of portfolio weights to small perturbations in means.

however, will depend on how similar the betas of the underlying assets are. If they are relatively constant in the cross section, portfolios diversified on subsets will have similar betas, and extreme weighting of these will be required to reduce factor risk. Minimizing variance may involve taking very large short positions in a diversified portfolio of assets and using the proceeds to finance very large long positions in another diversified portfolio.

These problems can become even more evident when the "assets" used in variance minimization are themselves well-diversified portfolios of a large underlying set of assets, as is often the case in applications. Strategic and tactical asset allocation procedures often use variance minimization to distribute funds across passively diversified portfolios such as country, industry, or size-based portfolios. Our results may shed some light on the difficulties encountered in these applications.[3]

We illustrate this behavior through example, and argue that the resulting weights correspond to those observed empirically. On the basis of this analysis, we argue that it is difficult to dismiss the observed extreme weightings as simply a consequence of sampling error. Given the dominance of a single factor, an extreme degree of similarity in the betas, or loadings, on that factor would be required to ensure that naive diversification corresponds to mean-variance efficiency for finite sets of assets. This is equally true of the population moments as it is of the sample moments. Using standard multivariate methods we provide statistical evidence that sufficient diversity in the betas is discernible even in the presence of sampling error.

The paper is organized as follows. Section I introduces notation, defines "diversification," and derives the dual to the system of inequalities that must have a solution if efficient portfolios are well diversified. Section II studies the relationship between diversification and APT pricing. Section III documents the behavior of efficient portfolios formed using sample moments and considers the extent to which the first principal component of the sample covariance matrix is the source of the extreme weights on sample mean-variance frontiers. Section IV briefly concludes.

I. Duality and Diversification

We consider the first two moments of the returns of n different securities. Let R denote the random vector of returns, with typical element R_i. Let V be the variance-covariance matrix for the returns, assumed to be strictly positive definite, with variances σ_i^2 on the diagonal and covariances σ_{ij} off diagonal. We denote the n-vector of mean returns as M, with typical element M_i. Let q denote an n-vector of ones. Then a portfolio is a vector of weights for the securities, w_p, satisfying $w_p'q = 1$, so that the weights sum to one. It has return $R_p = w_p'R$, expected return $E(R_p) = w_p'M$, variance $\sigma_p^2 = w_p'Vw_p$, and covariance with any other portfolio, s, $\sigma_{ps} = w_p'Vw_s$. Alternatively, we

[3] For example Black and Litterman (1990), Frost and Savarino (1986, 1988), Kallberg and Ziemba (1983), Kroll, Levy, and Markowitz (1984), and Pulley (1981).

can construct hedge portfolios that are zero-cost positions with "weights" that sum to zero. Let z_0 be such a position. Then $z_0'q = 0$, and we denote the payoff on such a position as $\Pi_0 = z_0'R$.

Let w_μ denote the portfolio that for a given mean μ, has minimum variance. It minimizes $w'Vw$ subject to $w'q = 1$ and $w'M = \mu$. The first order condition for this problem, which is both necessary and sufficient, is the existence of scalars λ_μ, γ_μ such that the following equation holds:

$$Vw_\mu = \lambda_\mu q + \gamma_\mu M. \tag{1}$$

Merton (1972) gives the following expressions for the two multipliers in this equation:

$$\lambda_\mu = \frac{(M'V^{-1}M) - (q'V^{-1}M)\mu}{(M'V^{-1}M)(q'V^{-1}q) - (q'V^{-1}M)^2} \tag{2}$$

$$\gamma_\mu = \frac{(q'V^{-1}q)\mu - (q'V^{-1}M)}{(M'V^{-1}M)(q'V^{-1}q) - (q'V^{-1}M)^2} \tag{3}$$

At the global minimum-variance portfolio, where the constraint on the means is not binding, γ_μ will be zero. Along the efficient part of the minimum-variance frontier $1/\gamma_\mu$ will be positive, and it will be negative along the lower boundary. For any mean other than that of the global minimum-variance portfolio, let R_μ be the return on the efficient portfolio with weights w_μ and let w_z denote the weights on the minimum-variance portfolio that is uncorrelated with R_μ. Premultiplying (1) by these weights and setting the result equal to zero gives as the "zero-beta" rate

$$E(R_z) = -\frac{\lambda_\mu}{\gamma_\mu}. \tag{4}$$

Similarly, premultiplying (1) by w_μ gives

$$\frac{1}{\gamma_\mu} = \frac{E(R_\mu) - E(R_z)}{\sigma_\mu^2} \tag{5}$$

Equation (5) says that $1/\gamma_\mu$ measures the excess return to variance ratio for efficient portfolios.

We take the term "well diversified" to mean that weight in the portfolio is spread reasonably evenly across many assets, and thus the weight on any one asset gets small as the number of assets grow. This can be given formal content by imposing a bound on the norm of the weights. The norm that proves tractable for our purposes here is the sup-norm. We simply require $|w_i| \leq K_n$. As an alternative, Chamberlain (1983) identifies portfolio diversification with the sum of the squared weights, or the l_2-norm. The differences between these approaches are discussed in Section III. The bound we impose on the absolute value of the weights depends on the size of the cross section. With diversification, we typically intend that the weights on individual assets get small as the universe of available assets expands.

We can now pose formally the question raised in the Introduction. We say the mean-variance frontier is *well diversified* at mean μ if there exists a portfolio that satisfies:

P1

$$Vw_\mu = \lambda_\mu q + \gamma_\mu M \tag{6}$$

$$-K_n \leq w_i \leq K_n \quad \text{for } i = 1,\cdots, n \tag{7}$$

The next lemma provides the dual to diversified frontier portfolios.

Lemma 1: *There exists a solution to* P1 *if and only if there is no n-vector y satisfying*

D1

$$\left|\lambda_\mu(y'q) + \gamma_\mu(y'M)\right| > K_n \sum_{i=1}^{n} |\text{Cov}(y'R, R_i)| \tag{8}$$

Proof: (If) Suppose the weights, w_μ, are poorly diversified. We will construct a solution to the dual. Assume $|w_{\mu j}| > K_n$ and let e_j denote the jth column vector of the n-by-n identity matrix. It has zeros in all components but the jth, where it has a one. Note that

$$w_{\mu j} = e_j' w_\mu = \lambda_\mu\left(e_j' V^{-1} q\right) + \gamma_\mu\left(e_j' V^{-1} M\right) \tag{9}$$

Consider the portfolio $y_j \equiv V^{-1} e_j$. Since

$$|\lambda_\mu(y_j' q) + \gamma_\mu(y_j' M)| = |\lambda_\mu\left(e_j' V^{-1} q\right) + \gamma_\mu\left(e_j' V^{-1} M\right)| = |w_{\mu j}| \tag{10}$$

the first two terms in (8), evaluated at y_j, have absolute value greater than K_n. Since $Vy_j = e_j$, $\sum_i |\text{Cov}(R'y_j, R_i)| = 1$, so the right side of (8) is simply K_n. Thus, (8) is dual to the system P1.

(Only if) Suppose that the efficient portfolio with mean μ is well diversified, so $|w_{\mu i}| \leq K_n$. Let y be any n-vector. Premultiply (6) by y' to obtain

$$y'Vw_\mu = \sum_i \text{Cov}(R'y, R_i) w_{\mu i} = \lambda_\mu(y'q) + \gamma_\mu(y'M). \tag{11}$$

Therefore,

$$\begin{aligned} \left|\lambda_\mu(y'q) + \gamma_\mu(y'M)\right| &= \left|\sum_i \text{Cov}(R'y, R_i) w_{\mu i}\right| \\ &\leq \sum_i |\text{Cov}(R'y, R_i)||w_{\mu i}| \\ &\leq K_n \sum_i |\text{Cov}(R'y, R_i)| \end{aligned} \tag{12}$$

Thus, if the frontier is well diversified at μ, no n-vector that satisfies (8) exists. ■

Proceeding from the dual inequality (8) to the conclusion of the following theorem is simply a matter of keeping track of the various ways in which the vector y can be normalized. We use ν to denote the mean return on the global minimum-variance portfolio and σ_ν^2 its variance.

Theorem 1: *The efficient portfolio with mean $\mu \neq \nu$ is well diversified if and only if the return, R^*, on every portfolio with weights that sum to one, satisfies*

$$|E(R^*) - E(R_z)| \leq \left|\frac{K_n}{\gamma_\mu}\right| \sum_{i=1}^{n} |\mathrm{Cov}(R^*, R_i)| \tag{13}$$

and the payoff, Π^*, *on every hedge position, with weights that sum to zero, satisfies*:

$$|E(\Pi^*)| \leq \left|\frac{K_n}{\gamma_\mu}\right| \sum_{i=1}^{n} |\mathrm{Cov}(\Pi^*, R_i)| \tag{14}$$

Proof: (If) Suppose the frontier is not well diversified at $\mu \neq \nu$. Then, by Lemma 1, there exists y such that (8) is satisfied. Suppose $y'q \neq 0$. Define $R^* = y'R/y'q$. The portfolio with this return has weights that sum to one. By (8):

$$\left|\left[(\lambda_\mu/\gamma_\mu) + (y'M)/(y'q)\right]\gamma_\mu(y'q)\right| > K_n \sum_{i=1}^{n} |\mathrm{Cov}(y'R, R_i)|, \tag{15}$$

divide both sides by $|\gamma_\mu(y'q)|$ to get:

$$|-E(R_z) + E(R^*)| > |K_n/\gamma_\mu| \sum_{i=1}^{n} |\mathrm{Cov}(R^*, R_i)|, \tag{16}$$

violating (13).

Suppose $y'q = 0$. Define $\Pi^* = y'R$. By (8):

$$|(y'M)\gamma_\mu| > K_n \sum_{i=1}^{n} |\mathrm{Cov}(y'R, R_i)|. \tag{17}$$

Divide both sides by $|\gamma_\mu|$ to get:

$$|E(\Pi^*)| > |K_n/\gamma_\mu| \sum_{i=1}^{n} |\mathrm{Cov}(\Pi^*, R_i)|, \tag{18}$$

which violates (14).

(Only if) Suppose either (13) or (14) are violated at $\mu \neq \nu$, where $\gamma_\mu \neq 0$. We will construct a solution to the dual in these cases, implying the frontier is not well diversified. First, suppose there is a portfolio, with return R^* and weights that sum to one, violating (13). Let y be the weights on this portfolio. Thus, $y'q = 1$, and using (4) we will write the violation of (13) as:

$$|y'M + \lambda_\mu(y'q)/\gamma_\mu| > \left|\frac{K_n}{\gamma_\mu}\right| \sum_{i=1}^{n} |\mathrm{Cov}(y'R, R_i)|. \tag{19}$$

Multiply the above by $|\gamma_\mu|$ to get:

$$|(y'M)\gamma_\mu + \lambda_\mu(y'q)| > K_n \sum_{i=1}^{n} |\mathrm{Cov}(y'R, R_i)|, \tag{20}$$

which shows y satisfies the dual, (8), at μ.

Now suppose there exists a payoff Π^*, with weights that sum to zero, that violates (14). Let y be the weights in this position. Since (14) is violated and $y'q = 0$:

$$\left| y'M + \lambda_\mu (y'q)/\gamma_\mu \right| > \left| \frac{K_n}{\gamma_\mu} \right| \sum_{i=1}^{n} |\mathrm{Cov}(y'R, R_i)|. \tag{21}$$

Multiply the above by $|\gamma_\mu|$ to get:

$$\left| (y'M)\gamma_\mu + \lambda_\mu (y'q) \right| > K_n \sum_{i=1}^{n} |\mathrm{Cov}(y'R, R_i)|, \tag{22}$$

and y satisfies the dual, (8) at μ. ■

A corollary follows from consideration of the global minimum-variance portfolio. At its mean, ν, the constraint on the mean return is non-binding, so $\gamma_\nu = 0$. Premultiplying the first-order condition, (6), by the weights in the portfolio then gives the variance of the minimum-variance portfolio as the multiplier for the constraint that the weights sum to unity. That is, $\sigma_\nu^2 = \lambda_\nu$. Obviously, the conditions ensuring that the weights in this portfolio are well diversified involve only the characteristics of the covariance matrix, and not those of the means.

Corollary 1: *The portfolio with the minimum feasible variance is well diversified if and only if the return, R^*, on every portfolio with weights that sum to one, satisfies*

$$\sigma_\nu^2 \le K_n \sum_{i=1}^{n} |\mathrm{Cov}(R^*, R_i)|. \tag{23}$$

Proof: Consider inequality (8), which characterizes solutions of the dual at any mean return, with $\gamma_\mu = 0$, and $\lambda_\mu = \sigma_\nu^2$. If y solves the dual at this point,

$$|\sigma_\nu^2 (y'q)| > K_n \sum_{i=1}^{n} |\mathrm{Cov}(y'R, R)|. \tag{24}$$

Normalizing by $|y'q|$ gives

$$\sigma_\nu^2 > K_n \sum_{i=1}^{n} |\mathrm{Cov}(R^*, R_i)|, \tag{25}$$

where R^* is the return on a portfolio with weights that sum to one. It thus violates (23). Now suppose there is a portfolio violating (23). Such a portfolio obviously satisfies (8), since $\lambda_\mu = \sigma_\nu^2$ and $\gamma_\nu = 0$. ■

The results thus far characterize the first two moments of asset returns when the efficient portfolio with a particular mean is well diversified. We now pose a broader question: When will there be no portfolios anywhere on the mean-variance frontier that are well diversified? We show that this will be the case if and only if there is a single set of portfolio weights that solve the dual for all possible target mean returns.

Theorem 2: *There are no well-diversified portfolios anywhere on the mean-variance frontier if and only if there is a portfolio y^* with return $R^* =$*

$R'y^*$ that has the same mean as the global minimum-variance portfolio ($E(R^*) = \nu$), and that satisfies

$$\sigma_\nu^2 > K_n \sum_{i=1}^n |\mathrm{Cov}(R^*, R_i)|. \tag{26}$$

Proof: (If) Since $q'y^* = 1$, we can use the expressions for the multipliers, (2) and (3) to write the left-hand side of (8) as:

$$\left|\lambda_\mu(q'y^*) + \gamma_\mu(M'y^*)\right| = \left|\frac{1}{D}(E(R^*)(q'V^{-1}q) - (q'V^{-1}M))\mu - \frac{1}{D}(E(R^*)(q'V^{-1}M) - (M'V^{-1}M))\right|, \tag{27}$$

where $D = (M'V^{-1}M)(q'V^{-1}q) - (q'V^{-1}M)^2$. Since $\gamma_\nu = 0$, $\nu = (q'V^{-1}M)/(q'V^{-1}q)$ and the coefficient multiplying μ on the right-hand side is zero at $E(R^*) = \nu$. The second term, at $E(R^*) = \nu$, is simply $-\lambda_\nu = -\sigma_\nu^2$. Thus, at y^*, the left-hand side of (8) is simply σ_ν^2, which clearly satisfies the dual if (26) is satisfied, and y^* solves the dual at all mean returns.

(Only if) Suppose the mean-variance frontier is nowhere well diversified. The set $S = \{w_\mu: -\infty < \mu < \infty\}$ is, by the two-fund theorem, a convex set. By hypothesis, it is disjoint from the set $I = \{w: |w_i| \le K_n\}$, which is also convex. The two may therefore be separated by a hyperplane, and there exists a vector b such that $w_\mu' b > w'b$ for all μ and all $w \in I$. In particular,

$$w_\mu' b = \lambda_\mu(q'V^{-1}b) + \gamma_\mu(M'V^{-1}b) > \sup_{w \in I} w'b = K_n \sum_i |b_i|, \tag{28}$$

where the last step follows since $w'b$ is maximized by choosing $w_i = K_n$ or $w_i = -K_n$ depending on the sign of b_i.[4] Define $\hat{y} = V^{-1}b$. Then, $q'\hat{y} = q'V^{-1}b$, $M'\hat{y} = M'V^{-1}b$, and $|\mathrm{Cov}(R_i, R'\hat{y})| = |b_i|$. It follows then from (28) that $\hat{y}$ satisfies the dual inequality, (8), at every mean return, μ. At $\mu = \nu$, $\gamma_\nu = 0$, and by (28) $\lambda_\nu q'\hat{y} > K_n \sum_i |b_i| > 0$. This implies $q'\hat{y} \neq 0$, so we can define $y^* \equiv \hat{y}/\hat{y}'q$ and $R^* = R'y^*$. Then for all μ,

$$\lambda_\mu + \gamma_\mu E(R^*) > K_n \sum_i |\mathrm{Cov}(R_i, R^*)|, \tag{29}$$

and in particular (26) is satisfied. The right-hand side of (29) does not depend on μ. Expressing the left-hand side as in (27), we find it is linear in μ. The only way, then, (29) can hold for all μ is for the coefficient on μ to be zero. As noted in connection with (27), this implies $E(R^*) = \nu$. ■

Theorem 2 shows that when the frontier is nowhere well diversified, the portfolio that solves the dual at every mean is one that solves it at the minimum-variance point. In addition, it has the same mean as the global minimum-variance portfolio. Thus, verifying the existence of just one such

[4] Note that it is in dealing with inequality (28) that the choice of the sup-norm provides tractability. In the definition of the set I, we do not impose $w'q = 1$ so that we can set $\sup_{w \in I} w'b = K_n \sum_i |b_i|$.

portfolio obviates the need to check for solutions to the dual problems at other means. The intuition for this result is that the dual problem, like the primal, has a "two-fund" structure. Any hyperplane that separates the well-diversified portfolios from the efficient frontier, which is a ray in Euclidean n-space, must meet only two criteria. It must separate the well-diversified portfolio from the minimum-variance portfolio, and it must be parallel to the ray along which all the other efficient portfolios lie.

II. Factors and Diversification

In this section we consider the duality results developed above when returns have a factor structure. The issues we study here have been considered within a Hilbert space setting in Chamberlain and Rothschild (1983) and Chamberlain (1983). The finite dimensional duality theory developed above provides a very straightforward intuition for these relationships. It also provides a very simple means of proving Chamberlain's result that, given an approximate factor structure, well-diversified portfolios on the efficient frontier imply that the mean of each asset converges to the APT expected return. This result is stronger than the convergence in mean-square in Ross (1976). Finally, our approach allows us to generalize the result to situations where the covariance matrix of residuals has eigenvalues that are unbounded but grow at a relatively slow rate.

We begin by considering a "strict factor structure." It is assumed that the n-vector of returns, R, has the following structure:

$$R = E(R) + \beta\delta + \varepsilon \tag{30}$$

The K^*-vector of "factors," δ, are taken to be mutually orthogonal with unit variance, and the n-vector of "residuals," ε, has covariance matrix Ω, which is diagonal. We denote the variance of the residual for asset i as ω_i^2. The covariance matrix of returns can then be decomposed as:

$$V = \beta\beta' + \Omega \tag{31}$$

where the matrix β consists of n rows and K^* columns of "factor loadings." The variance, $w'Vw$, of a given portfolio, w, consists of two terms. We will refer to the term $w'\beta\beta'w$ as the "factor risk" of the portfolio, and the term $w'\Omega w$ as its "idiosyncratic risk."

We say exact APT pricing holds if the mean returns are in the span of the betas and a constant vector.

$$E(R_j) = \left(1 - \sum_k \beta_{jk}\right)E(R_0^*) + \sum_k \beta_{jk} E(R_k^*) \tag{32}$$

The portfolios R_k^*, $k = 0,\cdots, K^*$, are "factor-mimicking" portfolios. We say that exact APT pricing holds in the limit if, as the number of assets, n, increases, there exist sequences of feasible factor-mimicking portfolios R_{nk}^*,

$k = 0,\cdots, K^*$, $n = 1,\cdots,\infty$ such that:

$$\lim_{n\to\infty} E(R_j) - \Big(\big(1 - \textstyle\sum_k \beta_{jk}\big)E(R^*_{n0}) + \sum_k \beta_{jk} E(R^*_{nk})\Big) = 0, \quad (33)$$

for any fixed j.

Consider, as an example, a single factor structure. In this case,

$$\mathrm{Cov}(R_i, R_j) = \begin{cases} \beta_j^2 + \omega_j^2 & \text{if } i = j \\ \beta_i\,\beta_j & \text{if } i \neq j. \end{cases} \quad (34)$$

Using this we can bound the assets' means. Assume the portfolio with mean μ is well diversified, and let y_j be a portfolio that puts weight one on asset j. Equation (13) from Theorem 1 implies:

$$\begin{aligned} |E(R_j) - E(R_z)| &\leq \left|\frac{K_n}{\gamma_\mu}\right| \textstyle\sum_{i=1}^{n} |\mathrm{Cov}(R_j, R_i)| \\ &\leq \left|\frac{K_n}{\gamma_\mu}\right| \Big(\textstyle\sum_{i=1}^{n} |\beta_j \beta_i| + \omega_j^2\Big) \\ &= \left|\frac{K_n}{\gamma_\mu}\right| |\beta_j| \Big(\textstyle\sum_{i=1}^{n} |\beta_i| + h_j\Big), \end{aligned} \quad (35)$$

where $h_j = \omega_j^2/|\beta_j|$. It follows that if the efficient portfolio with mean μ is well diversified then

$$\left|\frac{E(R_j) - E(R_z)}{\beta_j}\right| \leq \left|\frac{K_n}{\gamma_\mu}\right| \Big(\textstyle\sum_{i=1}^{n} |\beta_i| + h_j\Big), \quad (36)$$

for every $j = 1,\cdots, n$.

Recall from (5) that $1/\gamma_\mu$ measures the tradeoff between risk and return along the frontier. We assume that it is bounded away from zero and infinity as the cross section grows. If the factor is "pervasive", then we expect $\sum_{i=1}^{n} |\beta_i|$ to grow without bound as $n \to \infty$, while $K_n \to 0$. Otherwise there would be an infinitely large subset of assets with infinitesimal factor risk. Factor risk, in such a case, could be eliminated simply through diversification. The only term on the right-hand side of (36) which is specific to the asset is h_j, which does not change with the cross section. If K_n goes to zero at the same rate as the sum of the absolute betas grows, then the right-hand side of (36) will go to a bound which is independent of the asset being considered. Variation in the means outside of these bounds must be due to variation in the betas. For a fixed cross section of assets, as asset j's beta goes to zero, its expected return must also be approaching the zero-beta rate for the frontier to be well diversified. The larger the beta is in absolute value, the larger is the allowable variation in expected returns. With multiple factors similar relations can be derived that bound mean returns in terms of weighted averages of the betas across factors, as well as assets.[5]

The bound in (36) above will be very tight for assets with very little factor risk. By constructing such portfolios we can obtain much tighter bounds for all assets than those above, and for each asset these converge to the expected return predicted by the APT. Using a superscript n to denote the efficient set parameters when there are n assets, we make the following assumptions.

Assumption 1: *There is a strict factor structure.*

Assumption 2: *For any* $\mu \neq \lim_{n\to\infty} \nu^n$, γ_μ^n *is uniformly bounded away from zero.*

The first of these is self-explanatory. The second requires that the mean-variance frontier not become vertical in the limit, and hence that there remains a meaningful tradeoff between mean and variance. In Chamberlain and Rothschild (1983) this condition is ensured by "absence of arbitrage" in their limiting, Hilbert space economy. Thus, with the appropriate definitions, we could replace Assumption 2 with "there is no asymptotic arbitrage." The following theorem is analogous to Chamberlain's (1983, Theorem 1 and Corollary) result for the Hilbert space setting. The proof follows the standard strategy of constructing a return from the "factor-mimicking" portfolios with the same betas as the return on some particular asset.

Theorem 3: *Under Assumptions* 1 *and* 2 *a sufficient condition for exact APT pricing to hold in the limit is that, when* $K_n \to 0$ *as* $n \to \infty$, *every portfolio in the sequence of efficient portfolios for some mean return* $\mu \neq \lim_{n\to\infty} \nu^n$ *is well diversified.*

Proof: Let H_n be the set of demeaned portfolios.

$$H_n \equiv \left\{R - E(R) \colon R = \sum_{i=1}^{n} x_i R_i,\ \sum_{i=1}^{n} x_i = 1\right\}. \tag{37}$$

We construct the factor-mimicking portfolios by projecting factor k onto H_n, so that

$$\delta_k = R_{nk}^* - E(R_{nk}^*) + \zeta_{nk} \tag{38}$$

where $E(\zeta_{nk} R_j) = 0$ for $j = 1, \cdots, n$. Similarly, we construct R_{n0}^* by projecting the zero vector onto H_n and let ζ_{n0} be the projection error. (Note that R_{n0}^* will be the minimum-variance portfolio which has zero betas.) For asset j,

[5] Consider the 1-norm of the betas for factor k, $\|\beta_k\|_1 \equiv \sum_i |\beta_{ik}|$. We can use these as weights to compute a measure of asset j's responsiveness to factor movements

$$\hat{\beta}_j \equiv \sum_k |\beta_{jk}| \left(\frac{\|\beta_k\|_1}{\sum \|\beta\|}\right).$$

The Cauchy-Schwartz inequality, and steps similar to those above now leads to an inequality like (36) with $\hat{\beta}_j$ substituting for β_j and $\sum_k \|\beta_k\|_1$ substituting for $\Sigma_i |\beta_i|$ on the right-hand side.

consider the combination of factor-mimicking portfolios with the same factor risk:

$$R_{nj}^* \equiv \left(1 - \sum_k \beta_{jk}\right) R_{n0}^* + \sum_k \beta_{jk} R_{nk}^*. \tag{39}$$

Let

$$\begin{aligned}
\Pi_j^n &\equiv R_j - R_{nj}^* \\
&= c_j + \left(1 - \sum_k \beta_{jk}\right)[R_{n0}^* - E(R_{n0}^*)] + \sum_k \beta_{jk}(\delta_k - [R_{nk}^* - E(R_{nk}^*)]) + \varepsilon_j \\
&= c_j + \left(1 - \sum_k \beta_{jk}\right)\zeta_{n0} + \sum_k \beta_{jk}\zeta_{nk} + \varepsilon_j
\end{aligned} \tag{40}$$

where

$$c_j = E(R_j) - \left(\left(1 - \sum_k \beta_{jk}\right) E(R_{n0}^*) + \sum_k \beta_{jk} E(R_{nk}^*)\right). \tag{41}$$

If the efficient frontier contains a well-diversified portfolio, the duality conditions from Theorem 1 imply:

$$\left|E(\Pi_j^n)\right| \leq \left|\frac{K_n}{\gamma_\mu^n}\right| \sum_{i=1}^n \left|\mathrm{Cov}(\Pi_j^n, R_i)\right|. \tag{42}$$

Consider the covariances in the bounds above:

$$\begin{aligned}
\mathrm{Cov}(\Pi_j^n, R_i) &= \mathrm{Cov}\left(\left(1 - \sum_k \beta_{jk}\right)\zeta_{n0} + \sum_k \beta_{jk}\zeta_{nk} + \varepsilon_j, R_i\right) \\
&= \left(1 - \sum_k \beta_{jk}\right)\mathrm{Cov}(\zeta_{n0}, R_i) + \sum_k \beta_{jk}\mathrm{Cov}(\zeta_{nk}, R_i) \\
&\quad + \mathrm{Cov}(\varepsilon_j, R_i) \\
&= \begin{cases} 0 & \text{if } i \neq j \\ \omega_j^2 & \text{if } i = j \end{cases}
\end{aligned} \tag{43}$$

Therefore

$$\left|E(\Pi_j^n)\right| \leq \left|\frac{K_n}{\gamma_\mu}\right| \omega_j^2 \tag{44}$$

and since as $n \to \infty$, $K_n \to 0$, $\gamma_\mu^n \to \gamma_\mu > 0$, the right-hand side goes to zero as $n \to \infty$. ■

This theorem differs from Chamberlain's in the way "diversification" is formalized. Chamberlain requires that there be a portfolio on the efficient frontier that is well diversified in the sense of having weights that converge to zero in the sum of squares, or l_2-norm. Our result requires that there be some point on the frontier where portfolios are well diversified in the sense of the sup-norm. The sup-norm is a weaker notion of diversification. We are dealing with sequences of portfolio weights where the nth vector in the sequence is zero in all but the first n components. Thus, for every finite n, the norm-equivalence relation (see, e.g., Stewart (1973), p. 170) states

$$\|w\|_\infty \leq \|w\|_2 \leq n^{1/2}\|w\|_\infty. \tag{45}$$

The l_2-norm, which is what Chamberlain considers, will go to zero if K_n, which restricts the sup- or ∞-norm, goes to zero at rate $n^{-1/2}$. Clearly, however, K_n may go to zero in situations where the l_2-norm does not. Thus, portfolios that are well diversified in Chamberlain's (1983) sense will also be well diversified by our criteria, but not the reverse. In this respect, the theorem is a generalization of Chamberlain's (1983) result. Using a weaker notion of diversification we still find the relationship between efficient, diversified portfolios and APT pricing holds. On the other hand, as stated above, our result relies on a strict factor structure whereas Chamberlain's simply requires that the eigenvalues of the residual covariance matrix have a uniform bound, an "approximate factor structure."

In fact, though, the duality conditions expressed in terms of the sup-norm reveal that even an approximate factor structure is stronger than needed. Consideration of equations (42)–(44) makes clear that to ensure convergence to strict APT pricing when there are well-diversified efficient portfolios, all that is really needed is that the bound on the weights go to zero faster than the 1-norm of the residual covariance matrix. That is, we require

$$\lim_{n \to \infty} K_n \max_j \left\{ \sum_{i=1}^{n} |\text{Cov}(\varepsilon_j, \varepsilon_i)| \right\} = 0. \tag{46}$$

This can certainly happen in situations where the eigenvalues are not uniformly bounded. Intuitively, diversification still "works," in the sense that well-diversified portfolios have smaller and smaller residual variances, even if the eigenvalues of Ω_n, the residual covariance matrix, are not bounded; they just cannot grow too fast relative to K_n. The empirical evidence in, for example, Trzcinka (1986) suggests such situations may indeed be relevant. These studies document the existence of a dominant eigenvalue which clearly grows linearly with the number of assets. Other eigenvalues appear to be important, but they seem to grow at a much slower rate. The evidence on whether the additional factors are actually priced appears mixed (Shukla and Trzcinka (1990)). The following corollary to Theorem 3 gives an explicit expression for the speed at which they can grow, although we show by example below that for specific covariance matrices, faster rates of growth may be permissible.

Corollary 3: *Assume, in place of Assumption* 1, *that* $\lim_{n \to \infty} K_n \psi^n n^{1/2} = 0$, *where* ψ^n *denotes the largest eigenvalue of the n-asset residual covariance matrix* Ω^n. *Then exact APT pricing holds in the limit.*

Proof: The arguments proceed as in the Theorem up to equation (38), from which we have,

$$\text{Cov}(\Pi_j^n, R_i) = \text{Cov}(\varepsilon_j, \varepsilon_i). \tag{47}$$

Therefore

$$\begin{aligned} \left|E(\Pi_j^n)\right| &\le \left|\frac{K_n}{\gamma_\mu}\right| \sum_{i=1}^{n} \left|\mathrm{Cov}(\varepsilon_j, \varepsilon_i)\right| \\ &\le \left|\frac{K_n}{\gamma_\mu}\right| \max_j \sum_{i=1}^{n} \left|\mathrm{Cov}(\varepsilon_j, \varepsilon_i)\right| \\ &\le \left|\frac{K_n}{\gamma_\mu}\right| \psi^n n^{1/2} \xrightarrow[n \to \infty]{} 0. \end{aligned} \tag{48}$$

The last inequality follows from the equivalence relations between matrix norms (see, e.g., Stewart (1973), p. 183). ■

The following is an example of a residual covariance matrix that does not have any bounded eigenvalues, and so does not have an approximate factor structure. Nevertheless, diversification "works" in the sense that the equally weighted portfolio has zero limiting residual variance. It also illustrates that the sufficient conditions in Corollary 3, while weaker than bounded eigenvalues, are still not necessary for well-diversified efficient portfolios to imply APT pricing.

Example 1: Suppose, for $0 < \alpha < 1$,

$$\omega_{ij} = (1 + |i - j|)^{\alpha} - (|i - j|)^{\alpha}. \tag{49}$$

The row or column sums of this matrix "telescope," so that:

$$\sum_j \omega_{ij} = (i)^{\alpha} - 1 + (n - i - 1)^{\alpha}. \tag{50}$$

Note that both of the exponentiated terms on the right-hand side of (43) are less than or equal to n^{α}. Thus an upper bound on $\sum_j \omega_{ij}$ is $2n^{\alpha}$. Further, we have $i \ge 0$ and $n - i - 1 \ge 0$, so $i^{\alpha} + (n - i - 1)^{\alpha} \ge (n - i - 1 + i)^{\alpha}$, and $(n - 1)^{\alpha} - 1$ gives a lower bound on $\sum_j \omega_{ij}$. Thus we see that all of the row sums grow at rate n^{α}, and since all terms in this covariance matrix are positive, the row sums correspond to the quantities in the bounds on the means in, for example, equation (48). Condition (46) will then hold if $K_n n^{\alpha} \to 0$ or if K_n goes to zero faster than $n^{-\alpha}$. For this covariance matrix a well-diversified efficient frontier is equivalent to APT pricing as long as "well diversified" implies bounds on portfolio weights that go to zero faster than $n^{-\alpha}$. For example, with $\alpha = 1/2$ the row sums (and, as shown below, the eigenvalues) of this matrix grow at rate $n^{1/2}$. To have APT pricing we must have efficient portfolios with weights that go to zero as the universe of assets grows at least as fast as the square root of $1/n$. Thus, if $\alpha = 1/2$ the restriction that K_n is $o(n^{-\alpha})$ strengthens the l_2 notion of diversification. For $\alpha < 0.5$, however, portfolios that are well diversified in the sense that $n^{\alpha}K_n$ goes to zero need not be well diversified in the l_2 sense. For such a case we employ both a weaker definition of diversification and a weaker notion of "diversifiable risk" than Chamberlain employs, and still find the relationship between efficient, diversified portfolios and APT pricing holds.

Now, consider the variance of the equally weighted portfolio:

$$\begin{aligned}\frac{1}{n^2}(q'\Omega_n q) &= \frac{1}{n^2}\sum_j \left(\sum_i \omega_{ij}\right)\\ &\le \frac{n}{n^2}\max_j\left(\sum_i \omega_{ij}\right)\\ &\le \frac{1}{n}(2n^\alpha - 1) \xrightarrow[n\to\infty]{} 0.\end{aligned} \tag{51}$$

Thus, the equally weighted portfolio has zero residual variance in the limit. Nevertheless, the maximal eigenvalue is unbounded. In fact, it grows faster than the sufficient condition in the Corollary permits.

$$\begin{aligned}\psi^n &= \sup_x \frac{x'\Omega_n x}{x'x} \ge \frac{q'\Omega_n q}{q'q} = \frac{1}{n}\sum_j \left(\sum_i \omega_{ij}\right) \ge \frac{n}{n}\min_j\left(\sum_i \omega_{ij}\right)\\ &\ge (n-1)^\alpha - 1\end{aligned} \tag{52}$$

so ψ^n grows at least at rate n^α. The matrix in the example has the property that, as the number of assets increases, the 1-norm grows at the same rate as the 2-norm, or the maximal eigenvalue. In general, we can only be sure that the 1-norm grows at $n^{1/2}$ times the rate at which the maximal eigenvalue grows.

III. Frontiers Calculated from Sample Moments

In this section we analyze the behavior of portfolio weights in minimum-variance portfolios that are calculated using historical returns. We argue that the extreme weightings seen in efficient portfolios are largely due to the cross-sectional dispersion of the "loadings" associated with the first principal component, or "factor."[6] We show that there is reason to believe this is not simply a result of measurement error.[7] We provide examples that show behaviors similar to those encountered using sample data can be due to the ability to diversify on subsets to reduce idiosyncratic risks. These portfolios can then be combined to simultaneously eliminate factor risk. This tends to result in extreme portfolio weights.

In the first subsection we analyze conditions under which the efficient frontier fails to appear well diversified. In the second subsection we will argue that these conditions are an empirically plausible explanation for the poor diversification observed in sample efficient portfolios.

[6] We refer to an asset's beta and its weight in the first principal component interchangeably. Strictly speaking, they are only asymptotically equivalent as the number of assets grows. For the most part, however, our arguments concern behaviors that arise as n grows large.

[7] This is a different explanation than those made in Black and Litterman (1990), Best and Grauer (1991), and Frost and Savarino (1986), for example.

A. The Role of a Dominant Factor and Diversification on Subsets

Suppose returns have a single-factor structure. If the betas of the assets are constant across assets, ($\beta = cq$, where c is a scalar) then any portfolio with $w'q = 1$ must bear significant factor risk. Variance-minimizing portfolios will tend to weight assets equally. On the other hand, if there is sufficient diversity in the betas, portfolios can be created with very little residual risk and different betas. These, in turn, can be combined to eliminate factor risk, but the weightings involved in this second step may be very extreme. We argue in this section that such behavior is the source of the extreme weights observed in efficient portfolios calculated using historical estimates of the covariance matrix. The effects of diversification on subsets can be illustrated by example.

Example 2: Assume there is a strict factor structure governing returns with one factor that has unit variance. All assets have constant idiosyncratic variance ω^2. There are $2n$ assets, and we will be considering the consequences of letting n grow without bound. All the assets have betas of unity, except for two finite subsets. These consist of otherwise identical low-beta and high-beta assets.

$$\beta_i = \begin{cases} 1 - e & \text{for } i = 1,\cdots, N \\ 1 & \text{for } i = N + 1,\cdots, n \\ 1 + e & \text{for } i = n + 1,\cdots, n + N \\ 1 & \text{for } i = n + N + 1,\cdots, 2n \end{cases} \tag{53}$$

Now consider two equally weighted portfolios, of the first n and second n assets, respectively. They will each have idiosyncratic variance of ω^2/n. The first portfolio will have $\beta_L = 1 - e(N/n)$. The second, $\beta_H = 1 + e(N/n)$. The minimum-variance portfolio of these two equally weighted portfolios can be calculated directly. For the first portfolio this weight is:

$$w_L = \frac{1}{2}\left(1 - \frac{\dfrac{Ne}{n}}{\dfrac{\omega^2}{2n} + \left(\dfrac{Ne}{n}\right)^2}\right) = \frac{1}{2}\left(1 - \frac{Ne}{\dfrac{\omega^2}{2} + \dfrac{(Ne)^2}{n}}\right) \tag{54}$$

This weight does not become well diversified, that is close to one-half, as the number of assets increases. Rather, it approaches $(1/2) - (Ne/\omega^2)$, which, for reasonably large N, will be negative and large in absolute value. Note that this is true despite the fact that for n large relative to N, the two portfolios would appear to be virtually identical. Their betas both converge to unity. They approach this limit, however, at the same rate as the residual variance goes to zero. The role of the residual variance term is evident in the denominator of the middle expression in (54). Were ω^2 divided by a number that grew faster than n, such as n^2, the weights in the minimum-variance portfolio would explode with n. If it were divided by $n^{1/2}$ the weights would approach one-half. As it is, the need to take increasingly aggressive short positions in the low-beta portfolio, in order to eliminate factor risk as the betas converge, just offsets the exposure to residual variance this entails. Therefore, the weights approach a constant. This constant may be quite far from weights we would think of as well diversified, however. Behavior of this sort provides a possible explanation for the extreme weightings in efficient portfolios composed of other portfolios we document below.

The weightings on individual assets in this example will not, of course, be variance minimizing, but the intuition developed is still helpful in understanding the behavior of the individual weights on the frontier. For example, consider the simpler case, where we omit the assets with unit betas. The weights in the global minimum-variance portfolio can be computed directly for this case. For the low-beta and high-beta assets the weights are, respectively:

$$w_L = \frac{1}{2N} + \frac{e}{\omega^2 + 2Ne^2} \quad w_H = \frac{1}{2N} - \frac{e}{\omega^2 + 2Ne^2} \tag{55}$$

These weights do go to zero as N increases, but much more slowly if the difference between the betas (or e) is small. The second term, which summarizes the deviation from equal weighting, has e in the numerator, while N in the denominator is multiplied by e^2. The tendency to offset the two subsets through aggressive shorting of one will slow the speed with which the weights go to zero. For reasonable parameter values the weights in (55) are similar in magnitude to the average absolute values we report for sample frontiers below.[8]

To summarize, the analysis here provides conditions under which we might expect to see extreme weightings in the global minimum-variance portfolio, and, by continuity, in other portfolios on the efficient frontier. The source of the extreme weightings would be the tendency to offset positions that are themselves well diversified on subsets of the assets. These positions should be highly correlated with each other. Furthermore, the extreme weightings should disappear if the heterogeneity in betas is eliminated. The next section investigates these empirical predictions. Note that, as the conditions we provide seem just as plausible for population as for sample moments, our explanation contrasts with traditional ones that ascribe pathologies in the mean-variance frontier completely to sampling error.[9]

B. Characteristics of the Weights in Sample Minimum-Variance Portfolios

We first examine portfolios constructed using returns from the Center for Research in Security Prices (CRSP) monthly returns file, from the period January 1971 to December 1985. This fifteen-year time period was broken into three five-year subperiods. Within each subperiod we randomly selected

[8] For example, with $n = 50$, $e = 0.1$, and $\omega = 0.1$ the average absolute value of the weights is 9.5%. With e reduced to 0.05, the average absolute value of the weights for 50 assets is 19%.

[9] See, for example, Frost and Savarino (1986, 1988) and Black and Litterman (1990).

firms from the set of all CRSP firms, and then retained those with complete records over that particular five-year subperiod. This procedure gave us just over 900 sixty-month return series which were divided fairly evenly between the three subperiods. Note, however, that a particular firm could easily appear in more than one subperiod. Within each of the five-year subperiods, we constructed subsets of 10, 30, and 50 firms, and from these calculated sample means, variances, and covariances. The larger subsets simply regroup the smaller subsets. For example, the first 30-stock subset from 1/71 to 12/75 is simply the firms from the first three 10-stock subsets. For each subset we calculated efficient portfolio weights and other statistics. No difference in the behaviors of interest was evident across the three subperiods, so the descriptive results reported below are aggregated across subperiods as well as across subsets of firms within subperiods. Over all three subperiods we had available 90 subsets of 10 stocks each, 30 subsets of 30 stocks each, and 18 subsets of 50 stocks each.

The upper panel of Table I provides some summary measures of the degree of diversification in portfolios constructed using sample moments, and how these measures vary as the size of the asset universe grows. These are the weights in the global minimum-variance portfolios. The first three columns concern the maximum absolute value of the weights on individual assets within each of the portfolios. This is a measure of how big the vectors of weights are in the sup-norm. In terms of the formulae in the paper, these numbers tell us how big K_n would have to be for the portfolios in question to be "well diversified." The second three columns report the same summary statistics for the average of the absolute values of the weights in the portfolios. This measures how large the portfolios are in the 1-norm.

By almost any criteria these portfolio weights seem large. On average across the different subsets of assets, between 40% and 50% of the value of the portfolio is invested in a single asset. The percentage of value invested in one asset does not seem to decrease as the number of assets in the portfolio grows.

Furthermore, this lack of diversification does not appear to be the result of simply "loading up" on one or two assets. In the fifty-asset case, where equal weighting puts 2% in each asset, the *average* absolute value of the weights varies between 8.5% and 24%. Variance minimization appears to be achieved by selling a substantial number of the stocks short and using the proceeds to finance ever larger long positions in a substantial number of other stocks.

Portfolios on the positively sloped portion of the frontier tend to have weights that are slightly more extreme, but of similar orders of magnitude and qualitatively similar, to those in the global minimum-variance portfolios.[10] Since the global minimum-variance portfolio is computed without reference to the mean returns, this suggests that estimation error in the

[10] This is documented in the working paper versions of this paper, available from the authors on request.

Table I

Characteristics of Weights in the Global Minimum-Variance Portfolio Using 60 Months of Data for Three Random Samples of CRSP Firms from 1/71–12/75, 1/76–12/80, and 1/81–12/85.

The body of the table reports measures of the size of the weights, and the average correlation between components of the portfolios for various asset universes of various sizes.

	Maximum Absolute Value of Weights			Average Absolute Value of Weights			Correlation Between Long and Short Positions		
Number of assets per subsets (n)	10	30	50	10	30	50	10	30	50
Total number of subsets	90	30	18	90	30	18	90	30	18
Using sample covariance matrix									
Max. over all subsets	0.829	0.789	1.219	0.207	0.113	0.220	0.896	0.974	1.000
Min. over all subsets	0.164	0.201	0.295	0.100	0.062	0.085	NA[a]	0.855	0.992
Mean over all subsets	0.462	0.406	0.588	0.136	0.088	0.128	NA[a]	0.944	0.997
First component replaced with constant vector									
Max. over all subsets	0.493	0.118	0.073	0.102	0.033	0.020	0.560	0.634	0.654
Min. over all subsets	0.128	0.062	0.039	0.100	0.033	0.020	NA[b]	NA[b]	NA[b]
Mean over all subsets	0.229	0.080	0.050	0.100	0.033	0.020	NA[b]	NA[b]	NA[b]

[a] 1 of 90 portfolios had no negative weights.
[b] All but one portfolio had no negative weights.

means is not the primary source of poor diversification in efficient portfolios. In addition, the ability to diversify on subsets, and short one subset to finance large positions in the other, boosts the mean at very little cost in variance. As a result, the tradeoff between mean and variance along the frontier, as measured by $1/\gamma_\mu$, increases dramatically as the number of assets increases.

The effects of diversification on subsets are clearly illustrated in the last three columns of the upper panel in Table I. Here we take the short and the long positions in the minimum-variance portfolio and renormalize each by the sum of the negative and the sum of the positive weights, respectively. Thus, we end up treating the long and short positions as separate portfolios. These portfolios appear to be very highly correlated, and the average correlation grows to over 90% as the number of assets increases to 50. This suggests that an important source of the extreme weights lies in how the assets are related to some single source of risk, or factor. This factor can, in turn, be associated with the "dominant" eigenvalue which various empirical studies have shown tends to grow much faster than any of the others (see, e.g., Trzcinka (1986)).[11]

As a check on the importance of these phenomena in determining the weights, we performed the experiment documented in the lower panel of Table I. Let ψ_k be the kth largest eigenvalue of the covariance matrix and let a_k denote the associated eigenvector. We can decompose the covariance matrix as follows (see, e.g., Morrison (1976)):

$$V = \sum_k \psi_k a_k a_k'. \tag{55}$$

We constructed new matrices, denoted V^*, by subtracting off the first principal component and replacing it with an equicorrelated component, a vector of constants with the same norm as the original:[12]

$$V^* = V - \psi_1\left(a_1 a_1' - \frac{1}{n^2} q q'\right) \tag{56}$$

Examination of the lower panel of Table I shows that with this one substitution the frontier tends to look well diversified. The minimum-variance portfolio has all positive weights in all but one case with 30 or 50

[11] Brown (1989) argues that dominance of a single eigenvalue could arise even in a model with multiple "equally important" factors. The "betas" in his model, however, are drawn from the same i.i.d. cross-sectional distribution, so in large cross sections the factors are, in fact, close to linearly dependent.

[12] This operation need not preserve the relative importance of the largest principal component. It may be that the other eigenvectors will not be close to orthogonal to q, if the original eigenvector, a_1, is sufficiently different from the constant vector. We calculated the percentage of trace due to the first eigenvalue both before and after the substitution described above, to check whether the relative influence of this component was affected appreciably. It does not appear to be.

assets. Over the 18 cases with 50 assets, the maximum amount ever placed in a single asset is 7.3%.[13]

We view these results as evidence that the "first factor," or "dominant principal component" is, to a surprising degree, the source of the extreme weightings in frontier portfolios. Substitution for the first principal component eliminated the high correlation between long and short positions in the minimum-variance portfolio. We emphasize that we are not "recommending" this substitution in applications. Given the dominance of the first factor, as long as the cross-sectional heterogeneity in the betas is real, and not simply due to sampling error, it can and should be exploited in minimizing variance.

The use of mean-variance analysis in "asset allocation" frequently involves constructing portfolios of assets that are themselves reasonably well diversified. The reduction in residual variance that is achieved by employing portfolios instead of individual assets should reduce sampling error in the parameter estimates. Table II suggests that it does not reduce the extreme magnitudes of the weights; it aggravates the phenomenon. Even if formation of portfolios reduces measurement errors in the first principal component, there appears to be enough cross-sectional variation in the loadings to construct portfolios that are uncorrelated with the first factor. There is very little other variance that is not eliminated simply through diversification.

The weights reported in the upper panel of this table were calculated using monthly data from twenty portfolios formed on the basis of size using all

Table II

Characteristics of Weights in the Global Minimum-Variance Portfolio

	Maximum Absolute Value	Average Absolute Value
20 Size-based portfolios		
1/71–12/80	1.160	0.539
1/71–12/75	1.622	0.887
1/76–12/80	1.678	0.531
20 Beta-based portfolios		
1/76–12/85	0.635	0.191
1/76–12/80	0.799	0.282
1/81–12/85	0.614	0.195

[13] The portfolios on the positively sloped portion of the frontier also have weights that are much less extreme with this substitution, as is shown in working paper versions of this paper, available from the authors on request. The greater degree of diversification in the weights with this substitution for the first principal component is similar to a result obtained by Frost and Savarino (1986). Their Bayesian procedure involves prior beliefs of a single factor model with constant betas, and they find in simulations that the resulting efficient portfolios perform much better out of sample than those which use the sample covariance matrix. In the simulation, however, the "true" moments are based on historical estimates of the means and the covariance matrix. Thus the Bayesian pulls the covariance estimates away from the "truth" covariance matrix and towards the covariances matrix employed in the lower panels of Table I. Whether this is appropriate depends on the priors that the betas are, in fact, homogeneous.

firms on the CRSP daily returns file.[14] The second set of portfolios was constructed from a random sample of CRSP firms with continuous histories on the monthly returns file from 1971–1985. For each of these firms the first sixty months of returns were regressed on the equally weighted CRSP market index. The firms were then ranked according to their "betas" and divided into twenty equally weighted portfolios of eleven firms each. The sample means and sample covariances of these portfolios over the subsequent ten years were used to compute the weights in the efficient portfolios, which are reported in the lower panel. In each case, we find that portfolios which are first diversified naively, and then combined to minimize variance, tend to be weighted in an extreme manner. The table also contains descriptive statistics for the weights in a portfolio on the positively sloped portion of the frontier. These are, if anything, slightly less extreme than those in the global minimum-variance portfolio, again indicating that the basic problem is not with the estimated mean returns.

The analysis in Section III.*A* suggests that, at least in sample data, as the number of assets grows it becomes possible to construct portfolios that are orthogonal to the dominant principal component and yet diversify away most of the risk not associated with this factor. The following question arises: Is this simply due to measurement error in the "loadings" on the dominant factor, or is it a characteristic of the (unconditional) population moments? In applications, if one believes this behavior is the result of purely statistical error, one might try to "control" for it by constraining the weights. If it is characteristic of the population moments, one might try to exploit the ability to diversify on subsets and still reduce factor risk as a strategy to achieve low variance and high return.

At a minimum we can ask whether there is so much noise in the data that we cannot reliably attribute any of the heterogeneity we observe in the betas to the characteristics of the population moments. Under the assumption of multivariate normality, we can test the hypothesis that the first principal component is a constant vector, or an equicorrelated component. Such tests account explicitly for sampling error in the components of the eigenvector. We employ the test described in Morrison (1976, p. 295). Let a_i denote the ith eigenvector of the covariance matrix. Under the null hypothesis that $a_1 = qn^{-1/2}$, the following statistic has an asymptotic (as $T \to \infty$) chi-squared distribution with $n - 1$ degrees of freedom:

$$\chi^2 = T(\psi_1 n^{-1} q' S^{-1} q + n^{-1} \psi_1^{-1} q' S q - 2) \tag{59}$$

Table III shows the results of a number of tests using this statistic. We first computed this statistic for sets of 30 and 50 assets. The chi-squared statistics reported in Table III are the minimum values attained over these subsets. The associated probability values reject the hypothesis that the first principal component is the equally weighted portfolio at very high levels of confidence. The marginal probability that a given one of the chi-squares, across the

[14] We are very grateful to Eric Sirri for providing these data.

Table III

Tests for a Constant First Principal Component

The null hypothesis is that the first principal component is proportional to a constant vector.

	Number of Subsets	Minimum χ^{2}[a]	d.f.
Subsets of 30 assets (time periods aggregated)	30	160.82[b]	29
Subsets of 50 assets (time periods aggregated)	18	859.68[b]	49
20 Beta-based portfolios			
1/76–12/85	1	204.45	19
1/76–12/80	1	280.33	19
1/81–12/85	1	91.34	19
20 Size-based portfolios			
1/71–12/80	1	317.59	19
1/71–12/75	1	231.48	19
1/76–12/80	1	191.81	19
2 Beta-based portfolios			
1/76–12/85	1	112.98	1
1/76–12/80	1	101.61	1
1/81–12/85	1	26.04	1
2 Size-based portfolios			
1/71–12/80	1	101.10	1
1/71–12/75	1	61.66	1
1/76–12/80	1	31.66	1

[a]All the chi-squared statistics listed reject the null hypothesis of a constant first eigenvector at extremely high levels of confidence.

[b]The probability that all the χ^2's across the subsets exceed this value is bounded above by the marginal probability that any one exceeds this value. The marginal probability under the null is very close to zero.

subsets, would attain a value as large or greater than those in the table is virtually zero; and this probability provides an upper bound on the probability that, under the null hypothesis, all of the chi-squares would be above this critical value. These tests are not independent across the 30-asset and 50-asset groupings, of course. The same return data used to form the 30-asset subsets are simply regrouped to form the 50-asset subsets.

Performing these tests with individual assets fails, to some extent, to tell us what we want to know: can diversity in the betas be maintained while diversifying on subsets so as to eliminate residual variance? Accordingly, we also perform the tests using large portfolios of stocks. These results are reported in the lower panels of Table III. For sets of size-based portfolios, the null hypothesis that the first principal component is a constant vector is rejected at extremely high levels of confidence. Beta-based portfolios were also used to perform tests such as those in Table II. These portfolios are constructed, using out-of-sample data, to have cross-sectional variation in the betas on the equally weighted portfolio. The equally weighted portfolio is, in turn, highly correlated with the first principal component. In all these cases, the hypothesis that the first principal component is constant is rejected. This evidence suggests that the extreme weightings we see in variance-minimizing portfolios constructed with sample data are not just a result of measurement error.

With the possible exception of the tests just reported, an important limitation of our analysis is its descriptive nature, and the lack of formal tests of the hypothesis advanced regarding the source of the extreme weightings observed in sample efficient frontiers. Given the difficulties associated with posing this hypothesis in a manner amenable to formal inference, we have attempted instead to take a range of different approaches to documenting the phenomenon, and evaluating its robustness.

IV. Conclusions

We have shown that the existence of well-diversified portfolios on the frontier is equivalent to conditions that bound excess returns, over the zero-beta rate, with measures of the degree to which a portfolio covaries with the returns on all the individual assets. For there to be no well-diversified portfolios anywhere on the frontier, there must exist portfolios for which these measures are less than the variance of the minimum-variance portfolio. These bounds then provide simple means of illustrating the relationship between mean-variance efficiency, APT pricing, and diversification. We also investigate the sources of the empirical behaviors of efficient portfolios' weights. These appear to be attributable to variation across the elements of the eigenvector associated with the first principal component, and the ability to form well-diversified portfolios on subsets of stocks that have different loadings on this factor. This allows for reduction of both factor risk and idiosyncratic risk, but also tends to result in extreme portfolio weightings.

REFERENCES

Best, M. J. and R. R. Grauer, 1991, On the sensitivity of mean-variance efficient portfolios to changes in asset means: Some analytical and computational results, *Review of Financial Studies* 4, 315–342.

Black, F. and R. Litterman, 1990, Asset allocation: Combining investor views with market equilibrium, Technical report, Goldman Sachs Fixed Income Research.

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

Chamberlain, G., 1983, Funds, factors and diversification in arbitrage pricing models, *Econometrica* 51, 1305–1324.

——— and M. Rothschild, 1983, Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica* 51, 1281–1304.

Frost, P. A. and J. E. Savarino, 1986, An empirical Bayes approach to efficient portfolio selection, *Journal of Financial and Quantitative Analysis* 21, 293–306.

———, 1988, For better performance: Constrain portfolio weights, *Journal of Portfolio Management* 14, 29–34.

Hansen, L. P. and S. F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–613.

Huberman, G., S. Kandel, and R. F. Stambaugh, 1987, Mimicking portfolios and exact arbitrage pricing, *Journal of Finance* 42, 1–10.

Kallberg, J. G. and W. T. Ziemba, 1983, Comparison of alternative utility functions in portfolio selection problems, *Management Science* 29, 1257–1276.

Kroll, Y., H. Levy, and H. M. Markowitz, 1984, Mean-variance versus direct utility maximization, *Journal of Finance* 39, 47–61.

Merton, R. C., 1972, An analytic derivation of the efficient portfolio frontier, *Journal of Financial and Quantitative Analysis* 7, 1851–1872.

Morrison, D. G., 1976, *Multivariate Statistical Methods* (McGraw-Hill, New York).

Pulley, L. B., 1981, General mean-variance approximation to expected utility for short holding periods, *Journal of Financial and Quantitative Analysis* 16, 361–373.

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

Shukla, R. and C. Trzcinka, 1990, Sequential tests of the Arbitrage Pricing Theory: A comparison of principal components and maximum likelihood factors, *Journal of Finance* 45, 1541–1564.

Stewart, G. W., 1973, *Introduction to Matrix Computations* (Academic Press, New York).

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