MANAGEMENT SCIENCE
Vol. 56, No. 6, June 2010, pp. 905–923
ISSN 0025-1909 | EISSN 1526-5501 | 10 | 5606 | 0905
DOI 10.1287/mnsc.1100.1157

Impossible Frontiers

Thomas J. Brennan
School of Law, Northwestern University, Chicago, Illinois 60611, t-brennan@law.northwestern.edu

Andrew W. Lo
Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; and AlphaSimplex Group, Cambridge, Massachusetts 02142, alo@mit.edu

A key result of the capital asset pricing model (CAPM) is that the market portfolio—the portfolio of all assets in which each asset's weight is proportional to its total market capitalization—lies on the mean-variance-efficient frontier, the set of portfolios having mean-variance characteristics that cannot be improved upon. Therefore, the CAPM cannot be consistent with efficient frontiers for which every frontier portfolio has at least one negative weight or short position. We call such efficient frontiers "impossible," and show that impossible frontiers are difficult to avoid. In particular, as the number of assets, n, grows, we prove that the probability that a generically chosen frontier is impossible tends to one at a geometric rate. In fact, for one natural class of distributions, nearly one-eighth of all assets on a frontier is expected to have negative weights for *every* portfolio on the frontier. We also show that the expected minimum amount of short selling across frontier portfolios grows linearly with n, and even when short sales are constrained to some finite level, an impossible frontier remains impossible. Using daily and monthly U.S. stock returns, we document the impossibility of efficient frontiers in the data.

Key words: short selling; long/short; portfolio optimization; mean-variance analysis; CAPM
History: Received September 2, 2009; accepted January 13, 2010, by Wei Xiong, finance. Published online in *Articles in Advance* April 9, 2010.

1. Introduction

A cornerstone of modern portfolio management is the "efficient frontier" of mean-variance analysis: the set of portfolios for which the lowest variance possible is attained for a given level of expected return, or the highest possible expected return is attained for a given level of variance. The main thrust of the capital asset pricing model (CAPM) is that the market portfolio—the portfolio of all assets in which each asset's weight is proportional to its total market capitalization—must lie somewhere on the efficient frontier. Because, by definition, every component of the market portfolio has a nonnegative weight (since its market capitalization must be nonnegative), we would expect at least one portfolio on the efficient frontier to have this property. If, for a given a set of asset-return parameters (means, variances, and covariances), the corresponding efficient frontier does not have any such portfolio, we call this an "impossible frontier" for obvious reasons.

In this paper, we show that as the number of assets grows large, nearly all efficient frontiers are impossible.

Specifically, for any arbitrary set of expected returns and for a randomly chosen covariance matrix, we show that the probability that the resulting frontier is impossible approaches one as the number of assets increases without bound. This result depends, of course, on the specific distribution from which we draw the covariance matrix; we consider two classes: the uniform distribution (Haar measure), and distributions centered around linear factor models such as the CAPM and Ross's (1976) arbitrage pricing theory (APT). For both classes of distributions, mean-variance-efficient frontiers are almost surely impossible.

This remarkable result is not an artifact of pathological parameters, except in the two-asset case, but is a generic property of mean-variance-efficient portfolios. For typical parameter values, *every* portfolio on the efficient frontier will contain at least one short position, i.e., a negative weight. This implies that such an efficient frontier cannot be consistent with a CAPM equilibrium in which every investor holds the tangency portfolio, for such an equilibrium requires all weights to be positive for that portfolio. Alternatively, our impossibility result implies that the set of expected-return vectors and covariance matrices, $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, that are consistent with a CAPM equilibrium is extremely small—in fact, measure zero in the limit—hence we should not expect typical empirical estimates of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to yield plausible portfolios from the CAPM perspective unless the CAPM is literally true and estimation error is negligible.

Our results provide one explanation for the skepticism that most long-only portfolio managers have for standard mean-variance optimization. From their perspective, an impossible frontier is truly impossible for them to implement. Moreover, it is well known that the output of standard portfolio optimizer yields weights that must be constrained, but until now, the nonnegativity restriction that has become second nature to practitioners was thought to be a consequence of estimation error. The results in our paper show that even in the ideal case in which the means and covariance matrix of asset returns are known with perfect certainty, the efficient frontier will almost always contain negative weights. To the extent that estimation error generates means and covariances that deviate from the CAPM, such sampling variation will only exacerbate the problem, making it more likely that the sample efficient frontier is impossible. Our impossibility results may also provide a partial explanation for the recent popularity of so-called "active extension" strategies such as "130/30" portfolios in which a limited amount of short selling is permitted.

We begin in §2 with a brief review of the literature, and in §3 we derive analytical results for the two- and three-asset cases to build intuition and motivate our more general results. The main results of this paper are contained in §4, where we propose two classes of probability measures over the space of all possible covariance matrices and show that under both of these classes of measures, impossible frontiers become the rule, not the exception, as the number of assets increases without bound. We also show that the expected minimum amount of short selling across frontier portfolios grows linearly with n, and even when short sales are constrained to some finite level, an impossible frontier remains impossible. Given the importance of the CAPM, in §5 we examine the linear one-factor return-generating model in more detail, and show how to construct a covariance matrix that does not yield an impossible tangency portfolio. In §6, we provide an empirical illustration of our theoretical findings using daily and monthly returns for a subset of S&P 500 stocks, and show that the usual sample estimators of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ do yield impossible frontiers. We conclude in §7 with a discussion of the theoretical and practical significance of our results.

2. Literature Review

Any review of the mean-variance portfolio selection literature must begin with Markowitz (1952), who first introduced this powerful framework to the economics literature. Building on the Markowitz mean-variance framework, Tobin (1958), Sharpe (1964), and Lintner (1965) derived the equilibrium implications under the assumption that all investors held mean-variance-optimal or "efficient" portfolios, culminating in the "capital market line," the line connecting the risk-free rate on the expected-return axis with the tangency portfolio on the efficient frontier in mean–standard deviation space.

The role of short sales in mean-variance analysis has also been considered by several authors. In fact, Markowitz (1959, p. 132) recognized the importance of implementing constraints on portfolio weights, one of which was a nonnegativity or short-sales constraint. However, Lintner (1965) was perhaps the first to study the impact of short sales on capital market equilibrium, deriving alternative equilibria under short-sales prohibitions as well as short-sales constraints. Lintner concluded that investors would not engage in short sales in equilibrium because of the Tobin separation theorem, i.e., all investors are indifferent between holding portfolios of all assets and portfolios of just two funds: the riskless asset and the tangency portfolio.

The importance of the mean-variance efficiency of the market portfolio was recognized early on by many authors, who initiated a series of debates on the testable implications of the CAPM. In particular, Roll (1977), Rudd (1977), Roll and Ross (1977), Green (1986), Green and Hollifield (1992), and Best and Grauer (1985, 1992) all acknowledged the potential incompatibility of an arbitrary set of means and covariances with positive weights for frontier portfolios. Roll (1977) and Roll and Ross (1977) present qualitative arguments for frontier portfolios to have positive weights, and Rudd (1977) corrects and quantifies some of those arguments. Using the dual of the standard mean-variance quadratic optimization problem, Green (1986) derives a fascinating necessary and sufficient condition for the efficient frontier to be impossible—the existence of a nontrivial zero-expected-return arbitrage portfolio (a portfolio with weights that sum to zero) that has nonzero correlation with all assets. Green and Hollifield (1992) derive conditions under which frontier portfolios will be well diversified, meaning they contain no extreme weights, and argue that extreme weights are likely in the presence of a single dominant factor in asset returns. Jagannathan and Ma (2003) show that, despite this fact, portfolios with nonnegativity constraints often perform better than unconstrained counterparts because the constraints reduce the impact of estimation error. And given a positive vector of market weights and a covariance matrix, Best and Grauer (1985) derive the restrictions on the vector of expected returns that imply that the given vector of market weights is an efficient portfolio.

But the most relevant paper for our purposes is Best and Grauer (1992). In addition to providing conditions for all frontier portfolios to have positive weights, they derive a beautiful result: if a frontier contains portfolios with positive weights, they all lie on a continuous segment of the frontier! Using this fact, they show that as the number of assets increases without bound, and assuming that the CAPM holds (so that the vector of expected returns and the covariance matrix is consistent with the mean-variance efficiency of the vector of market weights), the line segment of frontier portfolios with positive weights converges to a single point under certain conditions. Given the implausibility of these conditions, and the sensitivity of frontier portfolios to small perturbations in the vector of expected returns (documented in Best and Grauer 1991a, b), they argue that frontiers with all positive weights are highly unlikely in practice. Our results confirm this intuition, but without any restrictions whatsoever on the vector of expected returns.

More recently, Markowitz (2005, p. 17) argued that empirical deviations from the CAPM are not surprising in light of the counterfactual assumptions on which the CAPM is based. In particular, he observes that "[w]hen one clearly unrealistic assumption of the capital asset pricing model is replaced by a real-world version, some of the dramatic CAPM conclusions no longer follow." An example is the fact that unlimited borrowing and lending at identical yields is not possible in practice, and this limitation implies that the market portfolio need not be mean-variance efficient in equilibrium.

Markowitz's (2005) caveats are well taken, but the results of our paper are considerably stronger—we show that even if all the assumptions of the CAPM are true, the market portfolio need not be mean-variance efficient. Specifically, Markowitz (2005) states the following assumptions:

(A1) Transaction costs and other illiquidities can be ignored.

(A2) All investors hold mean-variance-efficient portfolios.

(A3) All investors hold the same (correct) beliefs about means, variances, and covariances of securities.

(A4) Every investor can lend all she or he has or can borrow all she or he wants at the risk-free rate.

Markowitz (2005) then concludes that:

(C1) The market portfolio is a mean-variance-efficient portfolio.

The results of §§3–5 show that there exist certain combinations of means, variances, and covariances for which *every* mean-variance-efficient portfolio contains short positions, implying that none can be the market portfolio. And as the number of assets grows without bound, the likelihood of coming across a set of parameter values with this characteristic approaches certainty.

3. Some Examples of Impossible Frontiers

We begin with some notation. Let $\boldsymbol{\mu}$ be the vector of expected returns for n assets, and let $\boldsymbol{\Sigma}$ be the covariance matrix of those returns.[1] For a given level of expected return, μ_o, the corresponding portfolio on the efficient frontier is the vector $\boldsymbol{\omega}$, which minimizes the value of

$$\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} \quad \text{subject to} \quad \boldsymbol{\omega}'\boldsymbol{\iota} = 1, \quad \boldsymbol{\omega}'\boldsymbol{\mu} = \mu_o, \tag{1}$$

where $\boldsymbol{\iota}$ is a column vector of ones of the appropriate length. The set of optimal $\boldsymbol{\omega}$ can be found using the method of Lagrange multipliers (see, for example, Merton 1972):

$$\mathcal{F} = \left\{\boldsymbol{\omega}\colon \boldsymbol{\omega} = \frac{BC}{D}\left(\mu_o - \frac{B}{C}\right)(\boldsymbol{\omega}_\mu - \boldsymbol{\omega}_g) + \boldsymbol{\omega}_g \quad \text{for } \mu_o \geq \frac{B}{C}\right\}, \tag{2}$$

where

$$\begin{aligned} A &\equiv \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \quad & B &\equiv \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}, \\ C &\equiv \boldsymbol{\iota}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}, \quad & D &\equiv AC - B^2, \end{aligned} \tag{3}$$

and

$$\boldsymbol{\omega}_g \equiv \boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}/C, \quad \boldsymbol{\omega}_\mu \equiv \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/B. \tag{4}$$

Note that $\boldsymbol{\omega}_g$ is the global minimum-variance portfolio, and $\boldsymbol{\omega}_\mu$ is the vector that maximizes the Sharpe ratio relative to the risk-free rate of zero, i.e., $\boldsymbol{\omega}_\mu$ maximizes the function $\boldsymbol{\mu}'\boldsymbol{\omega}/\sqrt{\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}}$.

The frontier starts at the expected return level $\mu_o = B/C$. In fact, we can compute minimum-variance portfolios for values of μ_o less than B/C, but these portfolios would lie on the "inefficient" branch of the portfolio frontier, i.e., the portion of the frontier for which the expected return is not maximized for a given level of risk.

We call a frontier "impossible" with respect to the ith component if the weight of the ith component at each point on the frontier is negative. Clearly, a sufficient condition for a frontier to be impossible is that it be impossible for the ith asset, $1 \leq i \leq n$. From (2), we see that every point on an efficient frontier can be written in the form

$$\boldsymbol{\omega} = \frac{C}{D}\left(\mu_o - \frac{B}{C}\right)\boldsymbol{\omega}_P + \boldsymbol{\omega}_g, \tag{5}$$

where $\boldsymbol{\omega}_P \equiv B\boldsymbol{\omega}_\mu - B\boldsymbol{\omega}_g$. The values of C and D are nonnegative by the Cauchy-Schwartz inequality, so

[1] Throughout this paper, we maintain the following notational conventions: (1) all vectors are column vectors unless otherwise indicated; (2) matrix transposes are indicated by primes, hence $\boldsymbol{\omega}'$ is the transpose of $\boldsymbol{\omega}$; and (3) vectors and matrices are always typeset in boldface, i.e., X and μ are scalars and $\mathbf{X}$ and $\boldsymbol{\mu}$ are vectors or matrices.

a frontier will be impossible with respect to the ith asset exactly when $\boldsymbol{\omega}_g$ and $\boldsymbol{\omega}_P$ both have negative ith components.

Our technique for proving that an efficient frontier is impossible is to show that the ith elements of both $\boldsymbol{\omega}_g$ and $\boldsymbol{\omega}_P$ are negative for some i. Using this method, we can calculate a lower bound for the probability that a generically chosen efficient frontier is impossible, as well as lower bounds for the expected number of negative weights of portfolios on an impossible frontier and lower bounds on the expected amount of total short sales at each point on the frontier.

In §3.1, we investigate the special case of $n = 2$ and find that certain frontiers are impossible, but only under some rather unnatural conditions. However, in §3.2, we show that when $n = 3$, a variety of frontiers become impossible without any unnatural conditions.

3.1. The Two-Asset Case

For the case of $n = 2$ assets, we can characterize all situations in which a frontier will be impossible (proofs are included in the appendix).

Proposition 1. *For $n = 2$, let the assets be ordered so that $\mu_1 < \mu_2$, let σ_i denote the risk of the ith asset, and let ρ denote the correlation between the assets. The efficient frontier is impossible if and only if*

$$\frac{\sigma_2}{\sigma_1} < \rho.$$

Because $\rho \leq 1$, the proposition implies that a necessary condition for a frontier with two assets to be impossible is that $\sigma_2 < \sigma_1$. Also, because the volatilities are both nonnegative, it is also necessary that $\rho > 0$. Thus, for a frontier to be impossible, the asset with higher expected return must also have lower risk, and the two assets must be positively correlated. In such a circumstance, it will be optimal to have a short position in the low-return/high-risk asset at every point on the efficient frontier.

This condition is unnatural because the lower expected-return asset is strictly dominated by the higher expected-return asset, given that the latter is less risky than the former. Therefore, on purely economic grounds, it is possible to rule out impossible frontiers in the two-asset case. In the next section, however, we show that with just one more asset, there is no natural way to avoid impossible frontiers.

3.2. The Three-Asset Case

For $n = 3$ assets, we describe a class of mean returns, variances, and covariances indexed by a parameter $\varepsilon > 0$ such that for sufficiently small values of ε, the efficient frontier corresponding to the specified parameters is impossible, and that the short-sale amount throughout the frontier becomes arbitrarily large as ε approaches 0. Unlike the case of the condition in Proposition 1, no dominance relation among the assets is necessary; in fact, all three assets have the same ratio of expected return to standard deviation. After deriving the formal results for this class of parameters, we provide a specific numerical example and also indicate the range of values of ε for which frontiers in this class are impossible.

For each $0 < \varepsilon < 1$, let $d = 1 - \varepsilon$, and define the following set of asset-return parameters:

$$\boldsymbol{\mu} = \mu_2 \begin{pmatrix} d \\ 1 \\ 1/d \end{pmatrix}, \quad \boldsymbol{\sigma} = \sigma_2 \begin{pmatrix} d \\ 1 \\ 1/d \end{pmatrix}, \quad \text{and}$$
$$\mathscr{C} = \begin{pmatrix} 1 & d & d^4 \\ d & 1 & d \\ d^4 & d & 1 \end{pmatrix}, \tag{6}$$

where $\boldsymbol{\mu}$ is the vector of expected returns, $\boldsymbol{\sigma}$ is the vector of standard deviations, and $\mathscr{C}$ is the correlation matrix. We assume that $\mu_2 > 0$ and $\sigma_2 > 0$. The ratio of expected return to standard deviation for each asset is thus μ_2/σ_2. Also, the correlation between the first and the second asset is positive and identical to that between the second and third asset, whereas the correlation between the first and third asset is positive but somewhat smaller because $d^4 < d$. We prove below that the efficient frontiers corresponding to the class of parameters described by (6) are impossible for sufficiently small values of ε. Before deriving this result, however, we must confirm that the correlation matrix specified in (6) is positive definite, which is accomplished by the following lemma (see the appendix for the proof):

Lemma 1. *The correlation matrix $\mathscr{C}$ specified in* (6) *is positive definite for all $0 < \varepsilon < 1$.*

With this lemma in hand, we now prove that frontiers described in (6) are impossible when ε is sufficiently small.

Proposition 2. *For sufficiently small values of ε, the frontier corresponding to the parameters described in* (6) *is impossible. Each frontier portfolio can be expressed as*

$$\boldsymbol{\omega} = \alpha\tilde{\boldsymbol{\omega}}_P + \boldsymbol{\omega}_g,$$

for $\alpha \geq 0$, where $\tilde{\boldsymbol{\omega}}_P = (12/B)\boldsymbol{\omega}_P$. There are values $0 < \alpha_1 < \alpha_2$ such that the weight of the second asset is negative for each portfolio in the range $0 \leq \alpha \leq \alpha_2$ and the first asset is negative for each portfolio in the range $\alpha > \alpha_1$. Therefore, there is always at least one asset with a negative weight, and for values of α between α_1 and α_2, two assets have negative weights. Moreover, each portfolio on the frontier has a negative weight that is at least as large as $1/(3\varepsilon) + O(1)$.

This proposition demonstrates that impossible frontiers can arise easily in the three-asset case, and without imposing any unnatural conditions on the asset-return parameters. In practice, numerical computations show that frontiers corresponding to the parameters in (6) become impossible whenever $\varepsilon \leq 0.317$.

For concreteness, consider a numerical example with the following parameter values: $\varepsilon = 0.2$, $\mu_2 = 12\%$, and $\sigma_2 = 20\%$. In this case, we have

$$\boldsymbol{\mu} = \begin{pmatrix} 0.096 \\ 0.120 \\ 0.150 \end{pmatrix}, \quad \boldsymbol{\sigma} = \begin{pmatrix} 0.160 \\ 0.200 \\ 0.250 \end{pmatrix}, \quad \text{and}$$

$$\mathscr{C} = \begin{pmatrix} 1.0000 & 0.8000 & 0.4096 \\ 0.8000 & 1.0000 & 0.8000 \\ 0.4096 & 0.8000 & 1.0000 \end{pmatrix}.$$

The parametrization of the frontier described in Proposition 2 takes the form

$$\boldsymbol{\omega} = \alpha \begin{pmatrix} -2.6021 \\ 0.5204 \\ 2.0817 \end{pmatrix} + \begin{pmatrix} 1.3553 \\ -0.9104 \\ 0.5551 \end{pmatrix}.$$

The value of α for which the first coordinate of $\boldsymbol{\omega}$ is zero is $\alpha_1 = 0.5208$, and the value of α for which the second coordinate of $\boldsymbol{\omega}$ is zero is $\alpha_2 = 1.7493$. For values of α between 0 and α_1, the weight of the second asset is negative, and the weights of the other two assets are positive. For values of α between α_1 and α_2, the weights of both the first and second assets are negative, and the weight of the third asset is positive. Finally, for values of α greater than α_2, the weight of the first asset is negative, and the weights of the other two assets are positive. Therefore, the efficient frontier is impossible. Moreover, each portfolio on the frontier has a total short-sale amount of at least -63.9% of the total asset value.

The circumstances in which impossible frontiers can arise in the case of three assets is not limited to just those parameters described by (6). In fact, the efficient frontier will continue to be impossible if the values of the risk, return, and covariance parameters are allowed to vary within a small neighborhood of (6). In addition, many other three-asset examples of impossible frontiers with empirically plausible parameters can be constructed. By increasing the number of assets from two to three, the set of impossible frontiers has grown significantly. In §4, we show that this is no coincidence, and that as n increases without bound, an arbitrarily chosen frontier is almost surely impossible.

4. The General Case

In this section, we consider the general case of an arbitrary number of n assets. Unfortunately, simple analytical results like those for the two- and three-asset cases of §3 are not available for an arbitrary number of assets (but see Green 1986; Best and Grauer 1985, 1992 for other useful characterizations of impossibility, including computationally explicit methods for identifying impossible frontiers). Instead, we propose to conduct the following thought experiment: for a given vector $\boldsymbol{\mu}$ of expected returns and a randomly selected covariance matrix $\boldsymbol{\Sigma}$, what is the likelihood that the resulting frontier is impossible? To compute such a probability, we must, of course, propose a probability distribution for a covariance matrix, which is not a straightforward exercise. Although distributions of covariance matrices have been developed in the statistics literature, e.g., the Wishart distribution, they are sampling distributions of covariance-matrix *estimators* applied to independently and identically distributed multivariate normal data (see Anderson 1984, Chap. 7). Such distributions are highly parametric—if multivariate normality does not hold, then neither does the Wishart—and also do not necessarily capture the randomness that we seek, i.e., the random drawing of an arbitrary population covariance matrix from the space of all possible covariance matrices. In particular, Wishart distributions are typically "centered" at the estimated sample covariance matrix with multivariate tails that decline exponentially fast. This may be a reasonable model of the randomness associated with sampling variation, but seems less compelling as a mechanism for drawing an arbitrary covariance matrix at random.

Instead, we seek a more general distribution, such as a uniform distribution over the space of all possible covariance matrices, i.e., the space of all $(n \times n)$ symmetric positive-definite matrices with real elements. However, because this space is not compact, a uniform distribution over this space will have infinite mass. Nevertheless, in the same way that an "improper prior" can be specified in Bayesian inference (see, e.g., Jeffreys 1961, pp. 180–181; Box and Tiao 1973, p. 426), we can construct an "uninformative" distribution as a proxy for the uniform. We provide such a distribution for covariance matrices in §4.1, which will allow us to gauge the probability that a randomly selected covariance matrix gives rise to an impossible frontier, yielding the conclusion that impossible frontiers are almost certain to arise as the number of assets increases without bound.

However, it may be argued that an uninformative distribution of covariance matrices will not yield economically relevant draws because the resulting covariance matrices lack the factor structure hypothesized in the most popular asset-pricing models. We address this concern in §4.2 by introducing another class of probability distributions centered around the covariance matrices generated by linear factor-pricing models such as the CAPM and APT, and derive lower bounds on the probability that a frontier is impossible if it is chosen randomly with respect to one of the distributions in this class. We show that this lower bound also approaches unity as n grows without bound.

In §4.3, we calculate lower bounds on the expected number of assets with respect to which a frontier will be impossible, as well as estimates for the expected minimum size of short positions across frontier portfolios. We also find that an impossible frontier will remain impossible even if constraints are placed on the total amount of short selling allowed in any portfolio.

4.1. Impossibility with a Uniform Prior

Without any prior information for what a covariance matrix "should" look like, a plausible starting point for generating a "randomly selected" covariance matrix is to apply a uniform distribution over the space of all real symmetric positive definite $(n \times n)$ matrices. This problem is closely related to the *invariance principle* in Bayesian statistical decision theory, which states that "if two decision problems have the same formal structure..., then the same decision rule should be used in each problem" (Berger 1985, p. 390). This principle depends critically on the absence of prior information which, in turn, requires that the prior distribution also be uninformative and "invariant," i.e., scale independent. To formalize this notion of invariance, Berger (1985, Chap. 6.6, Definition 2) defines a group structure on the set of transformations of the data, and then defines a parametric family of probability measures to be invariant under this group if the parameters are invariant across all actions of the group. For example, under the group of scale transformations

$$Y = g_c(X) \equiv cX, \tag{7}$$

the exponential density function

$$f(x \mid \theta) = \theta^{-1} \exp(-x/\theta)$$

is invariant because the density of Y is given by (see Berger 1985, p. 394)

$$c^{-1} f(y/c \mid \theta) = (c\theta)^{-1} \exp(-y/(c\theta)) = f(y \mid c\theta). \tag{8}$$

Of course, the principle of invariance is considerably more general, and can be applied to arbitrary groups defined on arbitrary collections of random variables. Berger (1985, Chap. 6.6) shows that the prior distribution function associated with invariant decision rules is given by Haar measure, an extension of the uniform distribution to more general settings. We apply this same approach to define an uninformative prior on the space of all $(n \times n)$ covariance matrices, which we denote by $\mathcal{P}_n$.

For $\mathcal{P}_n$, the natural group of transformations to consider is the general linear group $\mathbf{GL}_n$, the group of all invertible linear transformations on $\mathbf{R}^n$ (or, equivalently, the group defined by all invertible $(n \times n)$ matrices). This is the multivariate analog of the multiplicative scale transformation (7), where we wish to define invariance with respect to the matrix multiplication of a vector of n random asset returns $\mathbf{r}$ by an arbitrary invertible matrix $\mathbf{A}$, yielding $\mathbf{Ar}$. If the covariance matrix of $\mathbf{r}$ is $\boldsymbol{\Sigma}$, then the covariance matrix of $\mathbf{Ar}$ is $\mathbf{A\Sigma A}'$; hence, we seek to construct an uninformative prior on $\mathcal{P}_n$ that is invariant between $\boldsymbol{\Sigma}$ and $\mathbf{A\Sigma A}'$ for all invertible $\mathbf{A}$. It turns out that Haar measure is the only measure satisfying this property.

More formally, Haar measure is the unique measure (up to a constant) that is invariant under the natural action of the group $\mathbf{GL}_n$ of invertible linear transformations on $\mathbf{R}^n$ on the space of covariance matrices. For $\mathbf{G} \in \mathbf{GL}_n$, this action is defined by $\boldsymbol{\Sigma} \mapsto \mathbf{G\Sigma G}'$ for $\boldsymbol{\Sigma} \in \mathcal{P}_n$, and any covariance matrix $\boldsymbol{\Sigma}$ can be mapped to any other covariance matrix under some such action. Thus, any such action takes a neighborhood around a specified covariance matrix to a corresponding neighborhood around any other covariance matrix, and Haar measure assigns the same volume to every such image of the original neighborhood. In this sense, Haar measure behaves uniformly on all of $\mathcal{P}_n$ and represents an uninformative prior distribution over all possible $(n \times n)$ covariance matrices. The following definition summarizes Haar measure on $\mathbf{GL}_n$ (see Jorgenson and Lang 2005 for further discussion).

Definition 1. Haar measure on $\mathcal{P}_n$ is the measure, ν_n, that is invariant under transformations of the form $\boldsymbol{\Sigma} \mapsto \mathbf{G\Sigma G}'$, for $\mathbf{G} \in \mathbf{GL}_n$. Thus, for any region $\mathbf{S} \subseteq \mathcal{P}_n$, Haar measure has the property that

$$\nu_n(\mathbf{S}) = \nu_n(\mathbf{GSG}') \tag{9}$$

for all $\mathbf{G} \in \mathbf{GL}_n$. This measure is unique up to multiplication by a positive constant, and in terms of the elements of the matrix $\boldsymbol{\Sigma} = [\Sigma_{i,j}]$, we have

$$d\nu_n(\boldsymbol{\Sigma}) = \frac{1}{(\det(\boldsymbol{\Sigma}))^{(n+1)/2}} \prod_{i \le j} d\Sigma_{i,j}, \tag{10}$$

where $d\Sigma_{i,j}$ is the element of Euclidean measure.

Under Haar measure, the entire space $\mathcal{P}_n$ has infinite volume so we cannot scale by a constant to transform Haar measure into a proper probability density. Instead, we calculate the probability that a selected frontier is impossible on cross sections of $\mathcal{P}_n$ using the probability density induced by Haar measure on those cross sections. We first need to introduce a useful system of coordinates on $\mathcal{P}_n$ with respect to which we can easily define our cross sections.

Definition 2. Each matrix $\mathbf{M} \in \mathcal{P}_n$ can be uniquely expressed in terms of (partial) Iwasawa coordinates as $(\mathbf{X}, \mathbf{W}, \mathbf{V})$, where $\mathbf{W} \in \mathcal{P}_2$, $\mathbf{V} \in \mathcal{P}_{n-2}$, and $\mathbf{X} \equiv [\mathbf{x}_1, \mathbf{x}_2]$, with $\mathbf{x}_1, \mathbf{x}_2 \in \mathbf{R}^{n-2}$. The relationship between $\mathbf{M}$ and $(\mathbf{X}, \mathbf{W}, \mathbf{V})$ is defined by the formula

$$\mathbf{M} = \begin{pmatrix} \mathbf{I}_2 & \mathbf{X}' \\ \mathbf{0} & \mathbf{I}_{n-2} \end{pmatrix} \begin{pmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{pmatrix} \begin{pmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{X} & \mathbf{I}_{n-2} \end{pmatrix} = \begin{pmatrix} \mathbf{W} + \mathbf{X}'\mathbf{V}\mathbf{X} & \mathbf{X}'\mathbf{V} \\ \mathbf{V}\mathbf{X} & \mathbf{V} \end{pmatrix}. \tag{11}$$

Moreover, each matrix $\mathbf{W}$ can be uniquely expressed in terms of Iwasawa coordinates as (y, u, v), where $u, v \in \mathbf{R}_+$ and $y \in \mathbf{R}$, according to the relationship

$$\mathbf{W} = \begin{pmatrix} 1 & y \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u & 0 \\ 0 & v \end{pmatrix} \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix} = \begin{pmatrix} u + y^2 v & yv \\ yv & v \end{pmatrix}. \tag{12}$$

Finally, we can also express each matrix $\mathbf{X}$ in terms of polar coordinates $(r_1, \ldots, r_{n-2}, \theta_1, \ldots, \theta_{n-2})$, where $r_i \in \mathbf{R}_+$ and $\theta_i \in S^1 = [0, 2\pi]$, using the relationships $x_{1,i} = r_i \cos\theta_i$ and $x_{2,i} = r_i \sin\theta_i$. Therefore, each $\mathbf{M} \in \mathcal{P}_n$ can be written in terms of coordinates

$$\mathbf{M} = (r_1, \ldots, r_{n-2}, \theta_1, \ldots, \theta_{n-2}, y, u, v, \mathbf{V}) \tag{13}$$

so that the space $\mathcal{P}_n$ can be viewed as the product

$$\mathcal{P}_n = (\mathbf{R}_+)^{n-2} \times (S^1)^{n-2} \times \mathbf{R} \times \mathbf{R}_+ \times \mathbf{R}_+ \times \mathcal{P}_{n-2}. \tag{14}$$

Using the coordinate system of Definition 2, we consider cross sections of $\mathcal{P}_n$ that have fixed values for all coordinates except the θ_i. We write $Z = Z(r_1, \ldots, r_{n-2}, y, u, v, \mathbf{V})$ for such a cross section with specified fixed values of the coordinates $r_1, \ldots, r_{n-2}$, y, u, v, and $\mathbf{V}$. This cross section is thus a product of $(n-2)$ copies of S^1, and the measure on this cross section induced by Haar measure on $\mathcal{P}_n$ is

$$d\nu_Z = \frac{1}{(2\pi)^{n-2}} d\theta_1 \cdots d\theta_{n-2}. \tag{15}$$

The measure ν_Z is therefore a proper probability distribution on the cross-sectional space Z; although probabilities cannot be computed with respect to Haar measure on all of $\mathcal{P}_n$, they *can* be computed with respect to ν_Z on each cross section Z of $\mathcal{P}_n$.

To calculate the probability that a covariance matrix $\mathbf{\Sigma}$ gives rise to an impossible frontier, it is convenient first to change variables from $\mathbf{\Sigma}$ to $\mathbf{M}$ using the correspondence

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{M}\mathbf{A}', \tag{16}$$

where $\mathbf{A} = \mathbf{A}(c_1, \ldots, c_n)$ is the unique matrix in $\mathbf{GL}_n$ with columns defined by

$$\mathbf{A}\mathbf{e}_1 = c_1 \boldsymbol{\iota}, \quad \mathbf{A}\mathbf{e}_2 = c_2 \boldsymbol{\mu}, \quad \text{and} \quad \mathbf{A}\mathbf{e}_j = c_j \mathbf{e}_j \quad \text{for } 3 \le j \le n \tag{17}$$

for specified values of $c_i > 0$, where $\mathbf{e}_j$ is an $(n \times 1)$-vector with 0s in all entries except for a 1 in the jth entry, $j = 1, \ldots, n$. Haar measure is invariant under this change of variables, so we can replace $\mathbf{\Sigma}$ with $\mathbf{M}$ and use Haar measure on $\mathbf{M}$ as the basis for our probability calculations. We calculate the probability that a matrix $\mathbf{\Sigma} = \mathbf{A}\mathbf{M}\mathbf{A}'$ gives rise to an impossible frontier for a matrix $\mathbf{M}$ in a cross section Z, where the probability is calculated with respect to the distribution ν_Z. In Theorem 1, we obtain a lower bound for this probability of impossibility that depends only on the parameter y in the specification of Z. However, first we need a lemma that provides a useful test for impossibility.

Lemma 2. *For a frontier to be impossible with respect to the ith coordinate, it is necessary and sufficient that*

$$\mathbf{e}_i'\mathbf{\Sigma}^{-1}\boldsymbol{\iota} < 0 \quad \text{and} \quad \mathbf{e}_i'\mathbf{\Sigma}^{-1}\boldsymbol{\mu} - \left(\frac{\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\mathbf{\Sigma}^{-1}\boldsymbol{\iota}}\right)\mathbf{e}_i'\mathbf{\Sigma}^{-1}\boldsymbol{\iota} < 0. \tag{18}$$

If $\mathbf{\Sigma} = \mathbf{A}\mathbf{M}\mathbf{A}'$ and $i > 2$, the conditions in (18) *are equivalent to*

$$\cos\theta_{i-2} > y \sin\theta_{i-2} \quad \text{and} \quad \sin\theta_{i-2} > 0, \tag{19}$$

where $\mathbf{M}$ has coordinates as in Definition 2.

Theorem 1. *Let $\mathbf{M} \in Z = Z(r_1, \ldots, r_{n-2}, y, u, v, \mathbf{V})$ be chosen randomly with respect to the distribution ν_Z. The probability p_Z that the covariance matrix $\mathbf{\Sigma} = \mathbf{A}\mathbf{M}\mathbf{A}'$ gives rise to an impossible frontier is bounded below as*

$$p_Z \ge 1 - \left(1 - \frac{1}{2\pi}\cot^{-1} y\right)^{n-2} \ge 1 - \left(1 - \frac{1}{2\pi(1 + \max(0, y))}\right)^{n-2}. \tag{20}$$

This theorem shows that for any fixed value of y, the probability that a covariance matrix in a cross section Z gives rise to an impossible frontier tends to 1 geometrically as n grows.[2] Moreover, if y is bounded

[2] The intuition for this result is similar to the intuition for the fact that the fraction of an n-dimensional unit sphere not lying in the positive orthant tends to 1 geometrically with n. The sphere is defined by n Euclidean variables subject to the condition that their squares sum to 1. Under the uniform Haar measure on the sphere, each of these n variables is independent of all of the others, and each is equally likely to be positive or negative. Thus, the likelihood that at least one is negative, so that a point is not in the positive orthant, is $1 - 2^{-n}$. For the case of the probability distribution in Theorem 1, each of the $(n-2)$ variables θ_i is independent of all the others, and there is a likelihood p_y, dependent only on y, that the ith of these variables gives rise to a frontier that is impossible with respect to the $(i+2)$th coordinate. As a result, the probability that the frontier is impossible with respect to at least one coordinate is at least as great as $1 - (1 - p_y)^{n-2}$.

above by y_+, the probability for any cross section Z with such a y coordinate tends to 1 at least as quickly as

$$p_Z \geq 1 - \left(1 - \frac{1}{2\pi(1+\max(0, y_+))}\right)^{n-2}.$$

The following corollary extends the previous results to yield a lower bound on the probability of impossibility for probability densities on the entire space $\mathcal{P}_n$.

COROLLARY 1. *Let φ be any probability density on $\mathcal{P}_n$ that factors into a product of densities*

$$\varphi = \left(\prod_{i=1}^{n-2} \varphi_{r_i}\right) \times \left(\prod_{i=1}^{n-2} \varphi_{\theta_i}\right) \times \varphi_y \times \varphi_u \times \varphi_v \times \varphi_{\mathbf{V}}, \quad (21)$$

where the φ_{θ_i} are uniform probability densities on S^1, and the other distributions are aribtrary distributions on the respective spaces $r_i \in \mathbf{R}_+$, $y \in \mathbf{R}$, $u \in \mathbf{R}_+$, $v \in \mathbf{R}_+$, and $\mathbf{V} \in \mathcal{P}_{n-2}$. Let $\boldsymbol{\Sigma} = \mathbf{AMA}'$ be an arbitrary covariance matrix, with $\mathbf{A}$ as defined in (17), *and with $\mathbf{M}$ chosen randomly in accordance with the distribution φ. The probability p that $\boldsymbol{\Sigma}$ gives rise to a frontier that is impossible is bounded below by*

$$p \geq 1 - \int_{\mathbf{R}} \left(1 - \frac{1}{2\pi(1+\max(0, y))}\right)^{n-2} \varphi_y(y). \quad (22)$$

4.2. Impossibility with Linear-Factor Model Priors

Although the generality of Haar measure in representing the selection of an arbitrary covariance matrix is compelling, some may consider it too general because it does not differentiate among outcomes according to their economic plausibility. In particular, Haar measure places the same probabilistic weight on covariance matrices arising from quantum mechanics as it does on those from economic models—there is nothing intrinsic to Haar measure that incorporates economic structure. Accordingly, one could argue that Haar measure places too much weight on financially irrelevant covariance matrices. This argument is debatable, not in the least because we do not usually develop economic theories to yield specific implications for covariance matrices, and hence it is not clear what "financially relevant" covariance matrices look like.

However, there does exist an important class of financial models that places restrictions on asset-return covariance matrices, and that is the set of linear factor models such as the CAPM and APT. If a linear factor-pricing model holds, then a "typical" covariance matrix drawn randomly from this economy will have a different distribution than a Haar measure. Accordingly, financial economists may prefer a more informed prior when it comes to covariance matrices.

In this section, we introduce a class of probability distributions based on covariance matrices motivated by linear factor models such as the CAPM and APT, and we calculate probabilities of impossibility with respect to distributions in this class. The construction of this class uses many of the techniques and notations developed in connection with our analysis of Haar measure in §4.1, so our exposition will be less detailed.

We start with $\mathbf{T}_0 = \mathbf{T}_0(\boldsymbol{\mu}, \mu_m, \sigma_m, r_f)$, the covariance matrix implied by a linear one-factor model for a chosen value of the excess expected-return vector $\boldsymbol{\mu}$ and for arbitrarily specified values of the excess expected return on the market, μ_m, and the market volatility, σ_m. The excess expected return values are defined as the excess of the expected return values over the risk-free rate, r_f. Also, we assume for the moment that there are no idiosyncratic components to asset returns. The matrix $\mathbf{T}_0$ can be written

$$\mathbf{T}_0 = \sigma_m^2 \boldsymbol{\beta}\boldsymbol{\beta}',$$

where $\boldsymbol{\beta}$ is the vector of "beta" values, $\boldsymbol{\beta} = \boldsymbol{\mu}/\mu_m$ (recall that we have temporarily assumed no idiosyncratic shocks).

To incorporate independent idiosyncratic risks,[3] nonnegative amounts can be added to diagonal elements of $\mathbf{T}_0$, and the elements of the matrix may be additionally adjusted to reflect deviations from the CAPM. We define a family of such matrices,

$$\begin{aligned}\mathcal{T} &= \mathcal{T}(\boldsymbol{\mu}, \mu_m, \sigma_m, r_f)\\ &= \{\mathbf{T} = \sigma_m^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \mathbf{diag}(\boldsymbol{\delta}) + \varepsilon \boldsymbol{\iota}\boldsymbol{\iota}' \colon \varepsilon \geq 0, \delta_i \geq 0\},\end{aligned}$$

and we define the subfamily $\mathcal{T}_2$ to be those matrices in $\mathcal{T}$ with $\delta_1 = 0$, $\delta_2 = 0$, and $\delta_i > 0$ for $3 \leq i \leq n$. The notation $\mathbf{diag}(\boldsymbol{\delta})$ represents the diagonal matrix with diagonal elements equal to the elements of the vector $\boldsymbol{\delta}$.

For any $\mathbf{T} \in \mathcal{T}_2$, we can write $\mathbf{T} = \mathbf{AA}'$, where $\mathbf{A} = \mathbf{A}(c_1, \ldots, c_n)$ is defined in (17), with $c_1 = \varepsilon^{1/2}$, $c_2 = \sigma_m/\mu_m$, and $c_i = \delta_i^{1/2}$ for $3 \leq i \leq n$. We write covariance matrices $\boldsymbol{\Sigma}$ in the form $\boldsymbol{\Sigma} = \mathbf{AMA}'$ for some $\mathbf{M} \in \mathcal{P}_n$, and we consider probability distributions on $\boldsymbol{\Sigma}$ defined in terms of probability distributions on $\mathbf{M}$. Because every $\boldsymbol{\Sigma}$ corresponds to a unique $\mathbf{M}$ under this relationship, every probability distribution for $\boldsymbol{\Sigma}$ can be realized in this way. Also, when $\mathbf{M} = \mathbf{I}_n$, we

[3] Of course, if the CAPM were literally true, then the vector of idiosyncratic risks $\boldsymbol{\epsilon}$ would have to satisfy the linear restriction $\boldsymbol{\omega}_m'\boldsymbol{\epsilon} \equiv 0$, where $\boldsymbol{\omega}_m$ is the vector of market weights (in other words, the random idiosyncratic shocks must sum to zero in each and every realization, implying that their n-dimensional joint distribution is, in fact, degenerate, and lies in an $(n-1)$-dimensional subspace). However, we are not assuming that the CAPM is true; otherwise, by definition the efficient frontier cannot be impossible. Instead, we are using the CAPM's one-factor model as motivation for constructing an alternative to Haar measure for the express purpose of generating a "randomly chosen" covariance matrix that may be more relevant for financial applications.

have $\mathbf{\Sigma} = \mathbf{T}$, so distributions for $\mathbf{\Sigma}$ are "centered" on the CAPM-based matrix $\mathbf{T}$ to the same extent that distributions for $\mathbf{M}$ are centered on $\mathbf{I}_n$. We can now define a broad class of probability distributions for $\mathbf{\Sigma}$ that are "centered" on CAPM-based matrices $\mathbf{T} \in \mathcal{T}_2$.

Definition 3. For $c > 0$, a distribution φ on $\mathbf{\Sigma} \in \mathcal{P}_n$ is in the class $\mathcal{D}(\mathcal{T}_2; c)$ if the corresponding distribution $\varphi_{\mathbf{M}}$ on $\mathbf{M} \in \mathcal{P}_n$ can be factored into a product of distributions

$$\varphi = \left(\prod_{i=1}^{n-2} \varphi_{r_i}\right) \times \left(\prod_{i=1}^{n-2} \varphi_{\theta_i}\right) \times \varphi_y \times \varphi_u \times \varphi_v \times \varphi_{\mathbf{V}}, \qquad (23)$$

where the φ_{θ_i} are uniform probability densities on S^1, where φ_y is bounded above by $c\,e^{-y^2}$ for $y \geq 0$, and where the other distributions are arbitrary distributions on the respective spaces $r_i \in \mathbf{R}_+$, $u \in \mathbf{R}_+$, $v \in \mathbf{R}_+$, and $\mathbf{V} \in \mathcal{P}_{n-2}$. Here we use the notation of Definition 2 for the coordinates for $\mathbf{M}$, and we use the correspondence $\mathbf{\Sigma} = \mathbf{AMA}'$ for the relationship between $\mathbf{\Sigma}$ and $\mathbf{M}$.

We now turn to the central result of this section: a lower bound for the probability of impossibility that is uniform across all distributions in the class $\mathcal{D}(\mathcal{T}_2; c)$.

Theorem 2. *For any given excess expected-return vector $\boldsymbol{\mu}$, excess expected return on the market μ_m, market volatility σ_m, and risk-free rate r_f relative to which excess returns are defined, let φ be a probability distribution in $\mathcal{D}(\mathcal{T}_2; c) = \mathcal{D}(\mathcal{T}_2(\boldsymbol{\mu}, \mu_m, \sigma_m, r_f); c)$ for a specified $c > 0$. With respect to this distribution, the probability that a random choice of $\mathbf{\Sigma}$ gives rise to an efficient frontier that is impossible is bounded below by*

$$P_I \geq 1 - \left(\frac{6}{7}\right)^{n-2} - 4c\exp\left(-\left(\frac{n-2}{3\pi}\right)^{2/3}\right). \qquad (24)$$

This lower bound holds uniformly across all $\varphi \in \mathcal{D}(\mathcal{T}_2; c)$, as well as across all choices of $\boldsymbol{\mu}$, μ_m, σ_m, and r_f. As n increases without bound, the probability that a generically chosen frontier is impossible tends to unity.

The remarkable generality of Theorem 2 raises the question of how tight the lower bound can be, especially given the fact that we have placed no restrictions on the excess expected-return vector $\boldsymbol{\mu}$. Table 1 shows that even for the 50-asset case—a relatively small number of assets for most financial applications—the likelihood of an impossible frontier is nearly certain.

It should come as no surprise that Theorem 2 can easily be extended to the case in which returns satisfy any linear k-factor model, $k \ll n$. In this case, the factor $(n-2)$ in (24) is replaced by $(n-k-2)$, and some of the constants are slightly different, but the asymptotic implications of the bound are identical. As n increases without bound, the probability of an impossible frontier approaches unity.

Table 1 Lower Bound for Probability of Impossible Frontier with *n* Assets

n	Lower bound
25	0.9059
50	0.9787
75	0.9920
100	0.9966

Note. This table shows the lower bound for the probability that a randomly chosen n-asset covariance matrix yields an impossible frontier under any measure $\varphi \in \mathcal{D}(\mathcal{T}_2; 0.1)$ over the space of all $(n \times n)$ symmetric positive-definite covariance matrices with real elements.

4.3. Characterizing the Short Positions

In this section, we derive several additional results about impossible frontiers. We determine the expected number of the total number of assets n with respect to which a generic frontier will be impossible, and derive a lower bound for the expected sizes of short positions across a generic frontier. We also generalize Theorem 2 to the case in which a constraint is placed on the total size of short positions at each point on the frontier.

Theorem 3. *For any given excess expected-return vector $\boldsymbol{\mu}$, excess expected return on the market μ_m, market volatility σ_m, and risk-free rate r_f relative to which excess returns are defined, let φ be an arbitrary probability distribution in $\mathcal{D}(\mathcal{T}_2; c) = \mathcal{D}(\mathcal{T}_2(\boldsymbol{\mu}, \mu_m, \sigma_m, r_f); c)$. Under this probability distribution, the expected number of the total number of assets n with respect to which the frontier corresponding to a random choice of $\mathbf{\Sigma}$ is impossible is bounded below by*

$$E_n \geq c'(n-2) \qquad (25)$$

for a positive constant c' defined as

$$c' \equiv \int_{\mathbf{R}} \left(\frac{1}{2\pi(1+\max(0, y))}\right) \varphi_y(y), \qquad (26)$$

which depends only on the factor φ_y of the probability distribution φ. If φ_y is a normal distribution with unit variance, a numerical lower bound for E_n is $(n-2)/8$.

This result follows from an estimate of the integral defining the expected value (see the appendix), and shows that the number of assets requiring short positions on a typical frontier grows linearly with the number of assets.

We can also determine lower bounds for the aggregate size of the short positions among efficient-frontier portfolios. The following definition makes this notion precise:

Definition 4. For $1 \leq i \leq n$, let S_i denote the infimum of the short position in the ith asset, measured as a fraction of the portfolio's net asset value, where the infimum is taken over all points on a given efficient frontier. Let S denote the infimum of the aggregate amount of short selling, where the infimum is also taken over all portfolios on a given efficient frontier.

With this definition, we are able to derive a lower bound on the magnitude of shorting among efficient-frontier portfolios:

Theorem 4. *For any given excess expected-return vector* $\boldsymbol{\mu}$, *excess expected return on the market* μ_m, *market volatility* σ_m, *and risk-free rate* r_f *relative to which excess returns are defined, let* φ *be an arbitrary probability distribution in* $\mathcal{D}(\mathcal{T}_2; c) = \mathcal{D}(\mathcal{T}_2(\boldsymbol{\mu}, \mu_m, \sigma_m, r_f); c)$. *Under* φ, *for* $3 \le i \le n$, *the expected value of* S_i *satisfies*

$$E[S_i] \ge c_i, \tag{27}$$

where

$$c_i \equiv \frac{1}{2\pi}\left(\int_0^{\infty} r_{i-2}\,\varphi_{r_{i-2}}\right)\left(\int_{-\infty}^{0}(1-y)\,\varphi_y\right), \tag{28}$$

and $\varphi_{r_{i-2}}$ *and* φ_y *are as in Definition* 3. *Note that if all the functions* $\varphi_{r_{i-2}}$ *are identical, so that all the* c_i *have a common value* c_*, *then the expected value of* S *has the following lower bound*:

$$E[S] \ge c_*(n-2). \tag{29}$$

Finally, we consider the effect of imposing short-sales constraints by first defining the concept of a constrained efficient frontier:

Definition 5. For $b \ge 0$, a constrained efficient frontier, $\mathcal{F}_b$, is the set of portfolio weight vectors that provide maximum returns for given levels of volatility, subject to the condition that the total size of the short positions in such weight vectors be no more than a fraction b of the portfolio's net asset value. Such a constrained frontier is an impossible frontier if every point on $\mathcal{F}_b$ has a negative weight for at least one asset.

Remarkably, imposing short-sales constraints does not decrease the probability that a frontier is impossible, as the next result shows:

Theorem 5. *Let* $\mathcal{F}$ *be an unconstrained efficient frontier, and let* $\mathcal{F}_b$ *be the corresponding constrained efficient frontier for some* $b > 0$. *If* $\mathcal{F}$ *is an impossible frontier, then* $\mathcal{F}_b$ *is an impossible frontier as well.*

Therefore, the probability that a constrained efficient frontier, with $b > 0$, is impossible is at least as large as the probability that an unconstrained efficient frontier is impossible.

5. The One-Factor Model

Given the overwhelming importance of the CAPM to financial theory and practice, we consider the special case of the linear one-factor model that underlies the CAPM. In particular, let the $(n \times 1)$-vector of returns of n assets be given by the following linear one-factor model:

$$\mathbf{r} = \boldsymbol{\iota} r_f + \boldsymbol{\beta}(r_m - r_f) + \boldsymbol{\epsilon}, \tag{30}$$

where r_m is the stochastic market return, $\boldsymbol{\beta}$ is an $(n \times 1)$ constant vector, and $\boldsymbol{\epsilon}$ is an $(n \times 1)$ stochastic vector of idiosyncratic shocks. We assume that the expected value of $\boldsymbol{\epsilon}$ is zero, and we write $\boldsymbol{\Omega}$ for its covariance matrix.

Let μ_m and σ_m denote the expected return and standard deviation of r_m, respectively. According to the CAPM relation (30), the mean vector and covariance matrix for asset returns, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, can be written in terms of μ_m, σ_m, $\boldsymbol{\beta}$, r_f, and $\boldsymbol{\Omega}$ as

$$\boldsymbol{\mu} = \boldsymbol{\iota} r_f + \boldsymbol{\beta}(\mu_m - r_f) \quad \text{and} \quad \boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\beta}'\sigma_m^2 + \boldsymbol{\Omega}. \tag{31}$$

The tangency portfolio implied by the CAPM is $\boldsymbol{\omega}_\mu$, defined in (4) as $\boldsymbol{\omega}_\mu \equiv \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}/B$. And under the assumption that $\boldsymbol{\Omega}$ is diagonal and $\boldsymbol{\mu}$ contains all positive elements, it can be shown that the tangency portfolio is, in fact, not impossible; i.e., it contains no negative weights and may, therefore, be consistent with capital market equilibrium in which the weights are proportional to the market capitalizations of the securities. In this section, we explore the impossibility of the tangency portfolio for more general residual covariance matrices $\boldsymbol{\Omega}$ and with no constraints on $\boldsymbol{\mu}$, and find that, as before, impossibility is the rule, not the exception as n increases without bound.

In §5.1, we introduce the techniques needed to characterize impossible tangency portfolios, and in §5.2, we derive a lower bound on the probability that a randomly selected tangency portfolio is impossible. In §5.3, we show how to construct the unique covariance matrix that is consistent with a given vector of means $\boldsymbol{\mu}$, the risk-free rate r_f, a set of market-capitalization weights $\boldsymbol{\omega}_m$, and the CAPM equilibrium (i.e., where those market weights correspond to those of the tangency portfolio), and which is as "close" as possible to a given covariance matrix $\boldsymbol{\Sigma}$. In other words, we derive the covariance matrix that is as close as possible to $\boldsymbol{\Sigma}$, but that is also consistent with the CAPM.

5.1. Characterizing Impossible Tangency Portfolios

As in §4, the key to characterizing impossible tangency portfolios is the choice of coordinates in which to express the covariance matrix, which will allow us to focus on the portion of the matrix that is relevant for impossibility. Any covariance matrix $\boldsymbol{\Sigma}$ can be written in the form $\boldsymbol{\Sigma} = \mathbf{AMA}'$, where $\mathbf{M}$ is a positive-definite symmetric matrix, and where $\mathbf{A}$ is the unique matrix that takes $\mathbf{e}_1$ to $\boldsymbol{\mu}$ and $\mathbf{e}_i$ to $\mathbf{e}_i$ for $2 \le i \le n$.[4]

[4] Note that the definitions of $\mathbf{A}$ and $\mathbf{M}$ are slightly different here than in §4.1, but we keep the same notation because these matrices play the same role as before.

Also, using techniques similar to those in §4.1, $\mathbf{M}$ can be expressed in terms of partial Iwasawa coordinates as

$$\mathbf{M} = \begin{pmatrix} w & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{pmatrix}\left[\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{x} & \mathbf{I}_{n-1} \end{pmatrix}\right] = \begin{pmatrix} w+\mathbf{x}'\mathbf{V}\mathbf{x} & \mathbf{x}'\mathbf{V} \\ \mathbf{V}\mathbf{x} & \mathbf{V} \end{pmatrix}, \tag{32}$$

where $w \in \mathbf{R}_+$, $\mathbf{x} \in \mathbf{R}^{n-1}$, and $\mathbf{V}$ is a covariance matrix of dimension $(n-1)\times(n-1)$. Here we have used the notation $\mathbf{G}[\mathbf{H}]$ for $\mathbf{H}'\mathbf{G}\mathbf{H}$.

In these coordinates, the portfolio $\boldsymbol{\omega}_\mu$ can be expressed simply as

$$\boldsymbol{\omega}_{\mu,i} = \begin{cases} (1+\mu_2 x_1+\cdots+\mu_n x_{n-1})/d & \text{for } i=1, \\ -\mu_1 x_{i-1}/d & \text{for } 2\le i\le n. \end{cases} \tag{33}$$
$$d \equiv 1+(\mu_2-\mu_1)x_1+\cdots+(\mu_n-\mu_1)x_{n-1}.$$

Therefore, $\boldsymbol{\omega}_\mu$ is completely determined by $\mathbf{x}$ and $\boldsymbol{\mu}$. This allows us to characterize the impossibility of the tangency portfolio via the following proposition:

Proposition 3. *The tangency portfolio, $\boldsymbol{\omega}_\mu$, implied by the CAPM is impossible if and only if any one of the following three conditions holds*: (i) *two elements of* $\mathbf{x}$ *have different signs*; (ii) *all elements of* $\mathbf{x}$ *have the same sign as* μ_1/d; *or* (iii) *the quantity* $(1+\mu_2x_1+\cdots+\mu_nx_{n-1})/d$ *is negative, where* $d\equiv 1+(\mu_2-\mu_1)x_1+\cdots+(\mu_n-\mu_1)x_{n-1}$.

We will make the most use out of the first condition for impossibility in Proposition 3 because it describes the bulk of the cases in which the tangency portfolio is impossible.

5.2. The Probability of Impossible Tangency Portfolios

To determine the probability that the CAPM tangency portfolio is impossible, we need to choose a probability distribution on the underlying variables $\boldsymbol{\beta}$, μ_m, σ_m, r_f, and $\boldsymbol{\Omega}$. Once these variables are determined, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are determined as well, and Proposition 3 will allow us to assess the impossibility of the corresponding tangency portfolio.

For our probability distribution, we allow $\boldsymbol{\beta}$, μ_m, σ_m, and r_f to be specified arbitrarily—our results will hold uniformly across any choice of these variables. With respect to $\boldsymbol{\Omega}$, we decompose the matrix into components and allow all but one of those components to be specified arbitrarily. Specifically, we write $\boldsymbol{\Omega}$ as

$$\boldsymbol{\Omega} = \begin{pmatrix} \Omega_{11} & \mathbf{0} \\ \mathbf{0} & \widetilde{\boldsymbol{\Omega}} \end{pmatrix}\left[\begin{pmatrix} 1 & \boldsymbol{\gamma}' \\ \mathbf{0} & \mathbf{I}_{n-1} \end{pmatrix}\right] = \begin{pmatrix} \Omega_{11} & \Omega_{11}\boldsymbol{\gamma}' \\ \Omega_{11}\boldsymbol{\gamma} & \widetilde{\boldsymbol{\Omega}}+\Omega_{11}\boldsymbol{\gamma}\boldsymbol{\gamma}' \end{pmatrix}, \tag{34}$$

where $\boldsymbol{\gamma}\in\mathbf{R}^{n-1}$, $\widetilde{\boldsymbol{\Omega}}$ is an $(n-1)\times(n-1)$ positive-definite matrix, and $\Omega_{11}>0$. The values of $\widetilde{\boldsymbol{\Omega}}$ and Ω_{11} can be specified arbitrarily. With respect to $\boldsymbol{\gamma}$, we impose a probability distribution $\varphi_\gamma(\boldsymbol{\gamma})$ on $\mathbf{R}^{n-1}$, and in our probability calculations we consider several possible choices of φ_γ. Thus, for our probability distribution on the underlying variables, we allow completely arbitrary specification of all terms except $\boldsymbol{\gamma}$, and with respect to $\boldsymbol{\gamma}$ we focus on a number of different choices of probability distributions on $\mathbf{R}^{n-1}$.

The characterization of impossibility in Proposition 3 relies on expressing $\boldsymbol{\Sigma}=\mathbf{A}\mathbf{M}\mathbf{A}'$ in terms of coordinates $\mathbf{x}$, w, and $\mathbf{V}$ for $\mathbf{M}$. Our probability distribution, however, is expressed in terms of another set of coordinates for $\boldsymbol{\Sigma}$, namely, $\boldsymbol{\gamma}$, Ω_{11}, $\widetilde{\boldsymbol{\Omega}}$, σ_m, and $\boldsymbol{\beta}$. Thus, we need to calculate the relationship between these choices of coordinates to apply the characterization of impossibility to draws from our distribution. The relationship of primary importance will be the expression of $\mathbf{x}$ in terms of the coordinates for the probability distribution, so we now turn to this calculation.

Multiplying on the right by $(\mathbf{A}')^{-1}$ and on the left by $\mathbf{A}^{-1}$ in the expression for $\boldsymbol{\Sigma}$ in (31), and using the definition of $\mathbf{A}$, we see that

$$\mathbf{M} = \mathbf{e}_1\mathbf{e}_1'(\sigma_m/\mu_m)^2 + \begin{pmatrix} \Omega_{11}/\mu_1^2 & (\Omega_{11}/\mu_1)\mathbf{z}' \\ (\Omega_{11}/\mu_1)\mathbf{z} & \widetilde{\boldsymbol{\Omega}}+\Omega_{11}\mathbf{z}\mathbf{z}' \end{pmatrix} = \begin{pmatrix} \Omega_{11}/\mu_1^2+(\sigma_m/\mu_m)^2 & (\Omega_{11}/\mu_1)\mathbf{z}' \\ (\Omega_{11}/\mu_1)\mathbf{z} & \widetilde{\boldsymbol{\Omega}}+\Omega_{11}\mathbf{z}\mathbf{z}' \end{pmatrix}, \tag{35}$$

where $\mathbf{z}=\boldsymbol{\gamma}-\tilde{\boldsymbol{\mu}}/\mu_1$, with $\tilde{\boldsymbol{\mu}}=(\mu_2,\ldots,\mu_n)'$. In light of the expression for $\mathbf{M}$ in (32), we have

$$\mathbf{x}=\mathbf{V}^{-1}(\mathbf{V}\mathbf{x})=(\Omega_{11}/\mu_1)(\widetilde{\boldsymbol{\Omega}}+\Omega_{11}\mathbf{z}\mathbf{z}')^{-1}\mathbf{z}. \tag{36}$$

Because $\boldsymbol{\mu}=\boldsymbol{\iota}r_f+\boldsymbol{\beta}(\mu_m-r_f)$ and $\mathbf{z}$ is determined by $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$, we see that (36) expresses $\mathbf{x}$ in terms of the coordinates for our probability distribution, as desired. After some algebraic manipulation, we can also write this expression for $\mathbf{x}$ as

$$\mathbf{x}=\left(\frac{\Omega_{11}/\mu_1}{1+\Omega_{11}\left\|\widetilde{\boldsymbol{\Omega}}^{-1/2}\mathbf{z}\right\|^2}\right)\widetilde{\boldsymbol{\Omega}}^{-1}\mathbf{z}. \tag{37}$$

This is a more useful formula for $\mathbf{x}$ because we are interested in the signs of the elements of $\mathbf{x}$, and this expression shows these are the same as the signs of the elements of $\widetilde{\boldsymbol{\Omega}}^{-1}\mathbf{z}/\mu_1$ (because the remaining multiplicative factor is always positive).

Theorem 6. *Let p be the probability that the tangency portfolio implied by the CAPM is impossible when the probability distribution on the term $\boldsymbol{\gamma}$ underlying the covariance matrix $\boldsymbol{\Omega}$ has a distribution given by φ_γ. A lower bound for p is*

$$p\ge\det(\widetilde{\boldsymbol{\Omega}})\int_{\mathbf{R}^{n-1}}F(\boldsymbol{\gamma})\,\varphi_\gamma(\widetilde{\boldsymbol{\Omega}}\boldsymbol{\gamma}+\tilde{\boldsymbol{\mu}}/\mu_1), \tag{38}$$

where F is equal to 1 *whenever* $\boldsymbol{\gamma}$ *has two elements with different signs*, *and equal to* 0 *otherwise*.

We now make the result more concrete by applying the theorem to specific choices for the distribution φ_γ.

Corollary 2. *If* φ_γ *has an* $(n-1)$*-dimensional multivariate normal distribution with mean* $\tilde{\boldsymbol{\mu}}/\mu_1$ *and covariance matrix* $s\widetilde{\boldsymbol{\Omega}}^2$ *for some* $s>0$, *then the probability that* $\boldsymbol{\omega}_\mu$ *is impossible satisfies*

$$p \geq 1 - 2^{2-n}, \tag{39}$$

and this result is independent of the choice of s.

Note that choices of φ_γ not centered at $\tilde{\boldsymbol{\mu}}/\mu_1$ will generally have a lower probability of impossibility. However, for choices of φ_γ that are close to the uniform distribution, e.g., those with large variance, the probability of impossibility will have a lower bound similar to that in Corollary 2.

5.3. A Nonimpossible Covariance Matrix

Given the simple structure of the linear one-factor model (30), it should be possible to find some covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ "close" to $\boldsymbol{\Sigma}$ in some sense that yields a "nonimpossible" tangency portfolio, i.e., a tangency portfolio that has strictly positive market-capitalization weights $\boldsymbol{\omega}_m$ and is consistent with $\boldsymbol{\mu}$, $\boldsymbol{\beta}$, and r_f. Using the techniques developed in §5.1, we construct such a nonimpossible covariance matrix in this section and show how it is related to Black and Litterman's (1992) approach to asset allocation with prior information.

Suppose that a mean return vector, $\boldsymbol{\mu}$, and a market-capitalization weight vector, $\boldsymbol{\omega}_m$, are given, and consider a covariance matrix $\boldsymbol{\Sigma}$ that is derived either empirically or from prior information, but that is not necessarily compatible with $\boldsymbol{\mu}$ and $\boldsymbol{\omega}_m$ in the sense that $\boldsymbol{\omega}_m \neq \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, as required by the CAPM. The matrix most compatible with the observed $\boldsymbol{\Sigma}$ but still conforming to the known values of $\boldsymbol{\mu}$ and $\boldsymbol{\omega}_m$ can be determined in the following manner. Write $\boldsymbol{\Sigma} = \mathbf{AMA}'$, and write $\mathbf{M}$ in terms of w, $\mathbf{x}$, and $\mathbf{V}$, as in (32). Replace $\mathbf{x}$ by $\tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}}$ is defined by

$$\tilde{x}_i = \frac{-\omega_{m,\,i+1}}{\mu_1 + (\mu_2-\mu_1)\omega_{m,\,2} + \cdots + (\mu_n - \mu_1)\omega_{m,\,n}} \tag{40}$$

for $1 \leq i \leq n-1$. The formula in (40) inverts the relationship between $\boldsymbol{\omega}_m$ and $\mathbf{x}$ from (33), so the value of $\tilde{\mathbf{x}}$ is the unique value compatible with the market weight vector $\boldsymbol{\omega}_m$ and the expected-return vector $\boldsymbol{\mu}$.

The change from $\mathbf{x}$ to $\tilde{\mathbf{x}}$ described in the last paragraph corresponds to a change in the overall covariance matrix. Replace $\boldsymbol{\Sigma}$ by $\widetilde{\boldsymbol{\Sigma}}$ where

$$\widetilde{\boldsymbol{\Sigma}} \equiv \mathbf{A}\widetilde{\mathbf{M}}\mathbf{A}', \quad \widetilde{\mathbf{M}} \equiv \begin{pmatrix} w + \tilde{\mathbf{x}}'\mathbf{V}\tilde{\mathbf{x}} & \tilde{\mathbf{x}}'\mathbf{V} \\ \mathbf{V}\tilde{\mathbf{x}} & \mathbf{V} \end{pmatrix}. \tag{41}$$

This new covariance matrix, $\widetilde{\boldsymbol{\Sigma}}$, is then compatible with $\boldsymbol{\omega}_m$ and $\boldsymbol{\mu}$ in that $\boldsymbol{\omega}_m$ is the tangency portfolio resulting from this mean and covariance. In addition, $\widetilde{\boldsymbol{\Sigma}}$ is the covariance matrix most compatible with the specified values of $\boldsymbol{\mu}$ and $\boldsymbol{\omega}_m$ and the observed value of $\boldsymbol{\Sigma}$ in that it requires precisely the amount of alteration to $\boldsymbol{\Sigma}$ needed to make the three sets of parameters compatible.

Therefore, for those who have strong conviction that the CAPM must hold and that $\boldsymbol{\mu}$ and $\boldsymbol{\omega}_m$ are, in fact, the correct expected returns and market weights, and $\boldsymbol{\Sigma}$ is their best estimate of the covariance matrix, the covariance matrix they should adopt is $\widetilde{\boldsymbol{\Sigma}}$ given in (41).

6. Empirical Analysis

To gauge the empirical relevance of our impossibility results, we use daily and monthly returns for stocks in the S&P 500 index to estimate portfolio parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and show that the realizations of impossible frontiers in the historical record are nontrivial. These results update and confirm similar empirical examples in Green (1986) and Best and Grauer (1992).

6.1. The Data

The monthly data consist of returns for stocks listed on the S&P 500 in December of 1995 for which monthly return data were available for the period from January 1980 through December 2005. The daily data consist of returns for stocks listed on the S&P 500 in December of 1995 for which daily return data was available for the period from January 1, 1996, through December 31, 2005. There are a total of 271 stocks in the monthly data set and 326 stocks in the daily data set.

6.2. A 100-Stock Empirical Efficient Frontier

For concreteness, we construct the efficient frontier for the first 100 assets for both daily and monthly returns using standard estimators for the means and covariance matrices. The two frontiers are plotted in Figure 1, and we find that both are impossible. The thin lines indicate the unconstrained frontiers, and the thick lines indicate the frontiers constrained to allow only 50% short selling. Figure 2 shows the amount of short selling for points on both of these frontiers. Clearly the short-sales constraints do not eliminate the problem of impossible frontiers and have a significant impact on the characteristics of the constrained optimal portfolio.

6.3. More Impossible Frontiers

Applying the usual sample mean and covariance-matrix estimators to daily and monthly returns, we compute estimates $(\hat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})$ and construct efficient fron-

Figure 1 Unconstrained Efficient Frontier for 100 Stocks in the S&P 500 Index, as Well as the Frontier Constrained to Allow No More Than 50% Short Selling, Based on (a) Daily Returns from January 1, 1996, to December 31, 2005, and (b) Monthly Returns from January 1980 to December 2005

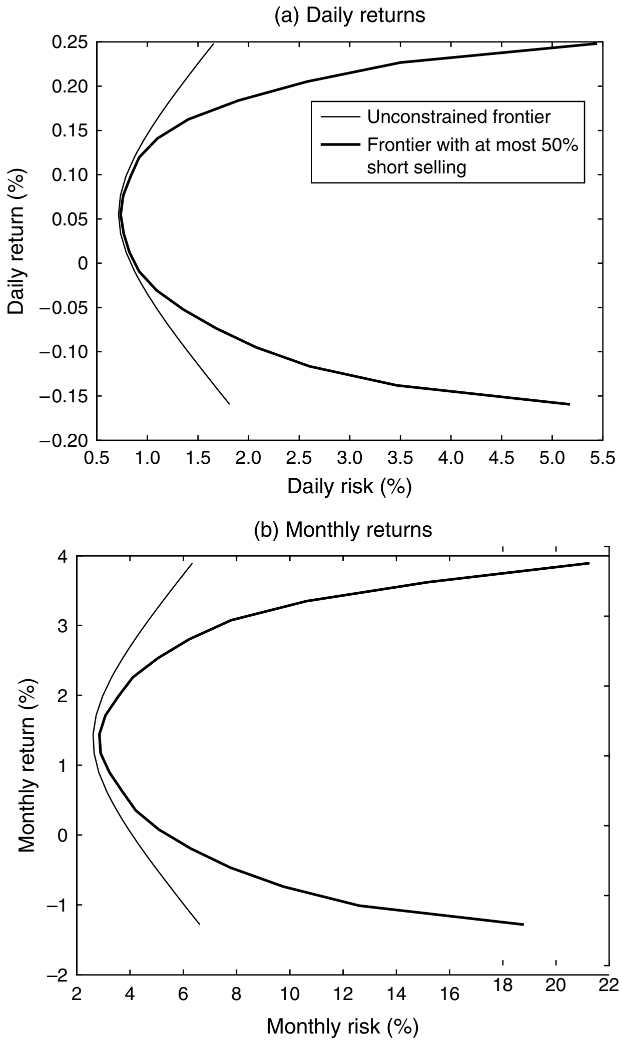
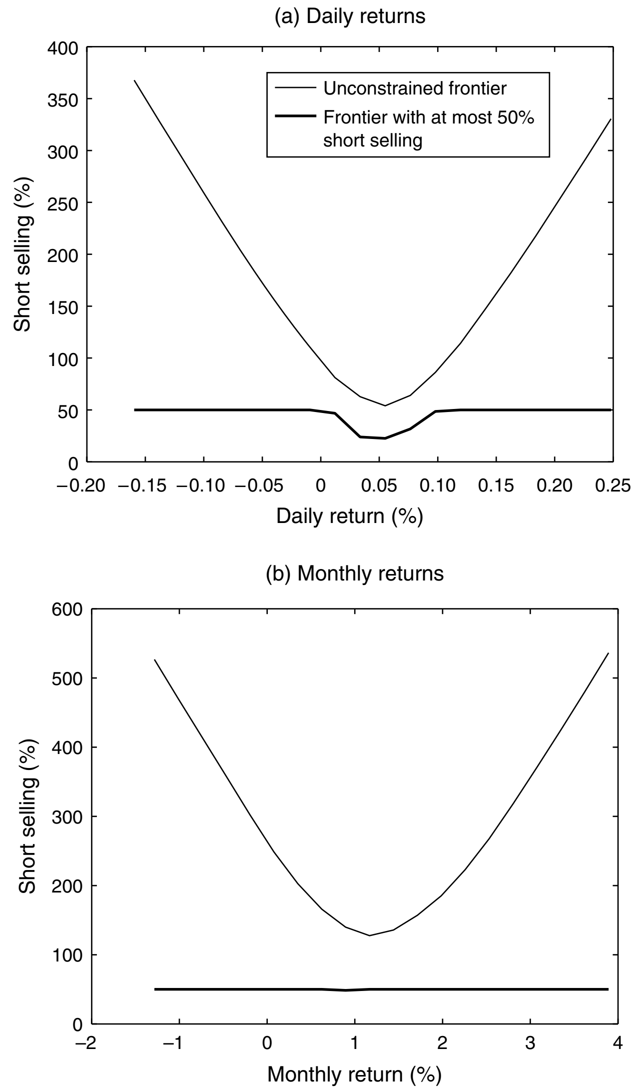


Figure 2 Magnitude of Short Positions for Points on the Unconstrained Efficient Frontier for 100 Stocks in the S&P 500 Index, as Well as the Frontier Constrained to Allow No More Than 50% Short Selling, Based on (a) Daily Returns from January 1, 1996, to December 31, 2005, and (b) Monthly Returns from January 1980 to December 2005



tiers for each of 2 through 326 assets for daily returns, and 2 through 271 assets for monthly returns. Figure 3 shows the fraction of assets with respect to which each frontier is impossible. Figure 4 shows the size of the short positions in the portfolios $\boldsymbol{\omega}_g$ and $\boldsymbol{\omega}_\mu$ for each of these frontiers. These results show that negative holdings are the rule rather than the exception for empirical efficient frontiers, and nonnegativity constraints are likely to have a major impact on the characteristics of mean-variance-optimized portfolios.

6.4. Estimation Error

One possible critique of our empirical analysis is that estimation error is likely to yield sample means and covariances that are inconsistent with the CAPM, so it is not surprising that we find impossible frontiers in the data. But this observation only underscores the ubiquity of impossible frontiers in practice. Because the population means and covariance matrix must always be estimated in financial applications, estimation error is an unavoidable aspect of practical portfolio management. Although a number of authors have explored the impact of estimation error on portfolio optimization (see, e.g., Brown 1976; Bawa et al. 1979; Frost and Savarino 1986; Jorion 1986; Jagannathan and Ma 2003; Tu and Zhou 2004, 2007, 2008; Wang 2005; Garlappi et al. 2007; Kan and Zhou 2007; DeMiguel et al. 2009), and alternatives such as Bayesian inference (Brown 1976), robust portfolio optimization (Fabozzi et al. 2007), and resampling

Figure 3 The Fraction of Assets with Respect to Which the Empirical Frontiers Are Impossible, as a Function of the Number of Assets Underlying the Frontiers, Based on a Subset of S&P 500 Stocks Using (a) Daily Returns from January 1, 1996, to December 31, 2005, with the Number of Stocks, n, Ranging from 2 to 326; and (b) Monthly Returns from January 1980 to December 2005, with the Number of Stocks, n, Ranging from 2 to 271

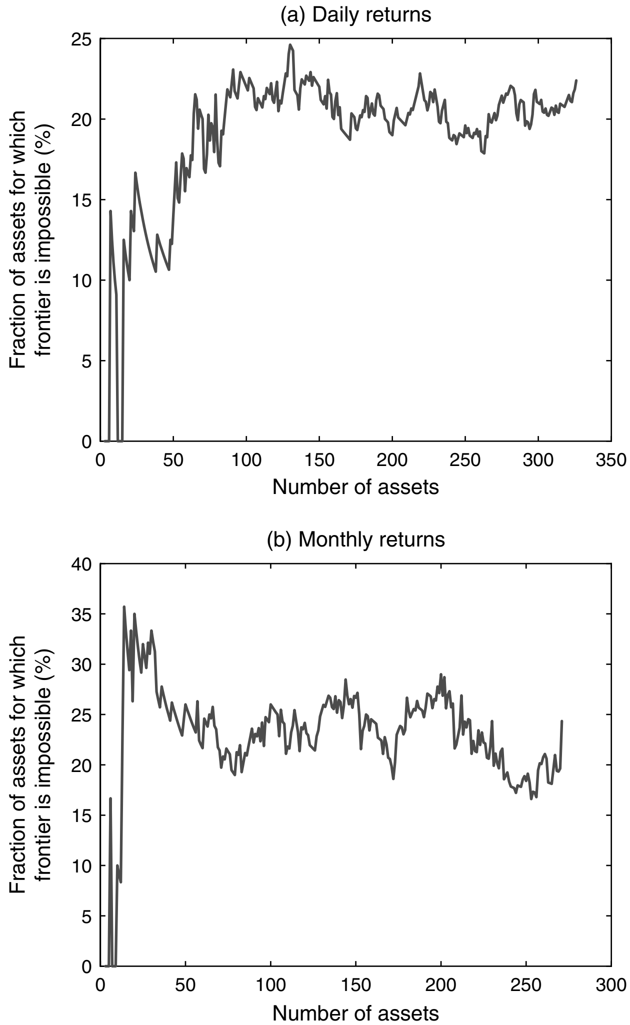
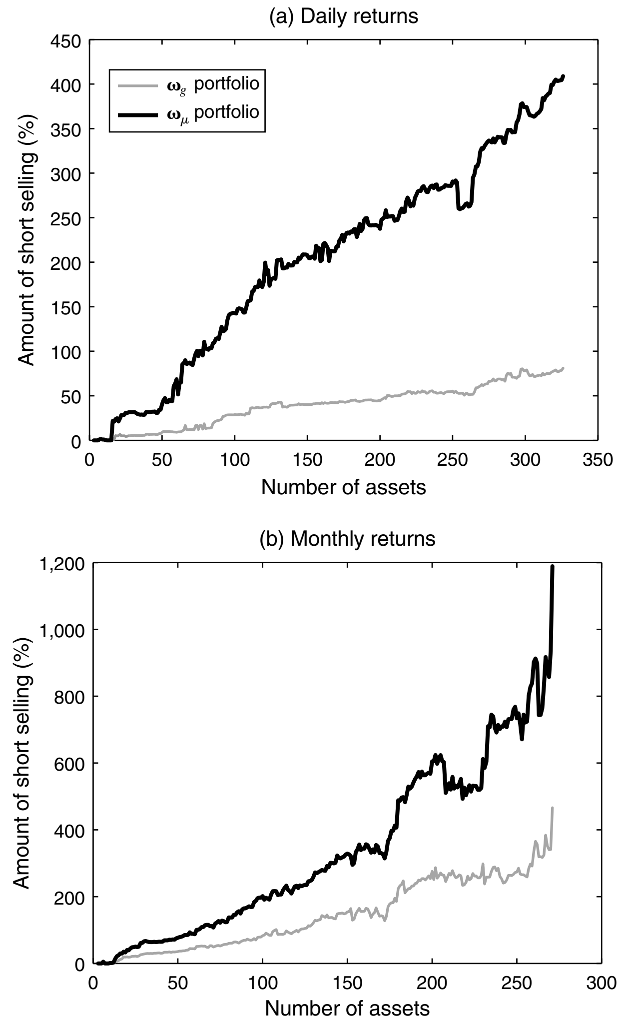


Figure 4 Magnitude of Short Positions in the Portfolios ω_g and ω_μ for the Empirical Frontiers, Based on a Subset of S&P 500 Stocks Using (a) Daily Returns from January 1, 1996, to December 31, 2005, with the Number of Stocks, n, Ranging from 2 to 326; and (b) Monthly Returns from January 1980 to December 2005, with the Number of Stocks, n, Ranging from 2 to 271



(Michaud 1998) have been developed in response, none of these methods addresses the impossibility of the *population* mean-variance-efficient frontier.

In particular, Theorems 1 and 2 show that impossible frontiers are almost certain to occur, even in the absence of estimation error. To the extent that estimation error can be viewed as random perturbations of population parameters (as opposed to perturbations that yield parameters closer to those satisfying a CAPM/APT relation), it is even more likely that estimated means and covariances will yield impossible frontiers. In other words, if a frontier is impossible for a set of population parameters, adding random noise to those parameters is unlikely to yield frontiers that are consistent with the CAPM.

7. Conclusion

In this paper, we have shown that mean-variance-efficient frontiers almost always contain short positions, implying a fundamental inconsistency between efficiency and economic equilibrium as described by the CAPM. This result is distinct from earlier concerns in the literature regarding the mean-variance efficiency of the market portfolio. Those concerns involved the observability of the total market portfolio, the existence of nontraded assets such as human capital, estimation errors in the sample means and covariance matrix, nonstationarities, asymmetric information, and other capital-market imperfections. Even in a frictionless world where all parameters are fixed and known, and where all of the other perfect-markets assumptions of the CAPM hold, mean-variance-efficient frontiers are almost always impossible.

This surprisingly general result provides a potential explanation for the near universal disdain with which long-only portfolio managers regard standard mean-variance optimization techniques. These investment professionals—who comprise the majority of end users of commercial portfolio construction software such as the BARRA Optimizer and the Northfield Portfolio Optimizer—have railed against mindless optimization for years, arguing that portfolio weights obtained in this manner are ill behaved and must be constrained or otherwise postprocessed. However, the typical rationale for these complaints is that the weights of frontier portfolios are too unstable and too sensitive to estimation error to be of practical value. We have identified a distinctly different rationale, which is the ubiquity of short positions in frontier portfolios, even in the absence of estimation error. An impossible frontier is, in fact, literally impossible for the long-only portfolio manager. The surging popularity of 130/30 and more general long/short strategies among such managers and their investors may well be a practical manifestation and an unintended consequence of the impossibility of mean-variance-optimal portfolios.

The virtual certainty of impossible frontiers also has implications for the interpretation of economic equilibrium. The converse of our impossibility theorem is that the set of parameters, $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, that are "possible," i.e., that are consistent with the mean-variance efficiency of the market portfolio, is a vanishingly small set as the number of assets grows without bound. In particular, in a CAPM equilibrium, covariances are also endogenously determined via supply and demand, despite the fact that most asset-pricing models focus exclusively on the properties of expected returns in equilibrium. Is it any wonder that the set of n means and $n(n+1)/2$ covariances that is consistent with capital-market equilibrium is apparently quite sparse?

To the disciples of general equilibrium theory, this may be heretical, but from a broader and more practical perspective, it should not be too surprising that the likelihood of simultaneous equality of supply and demand across a large number of markets is small, and increasingly less likely as the number of assets grows. With the techniques developed in this paper, we hope to be able to deduce other generic properties of financial market equilibria as well as their practical implications.

Acknowledgments

The views and opinions expressed in this paper are those of the authors only and do not necessarily represent the views and opinions of AlphaSimplex Group, MIT, Northwestern University, or any of their affiliates and employees. The authors make no representations or warranty, either expressed or implied, as to the accuracy or completeness of the information contained in this paper, nor are they recommending that this paper serve as the basis for any investment decision—this paper is for information purposes only. Research support from AlphaSimplex Group, the MIT Laboratory for Financial Engineering, and the Northwestern University School of Law Faculty Research Program is gratefully acknowledged. The authors thank Henry Cohn, Jayna Cummings, Sanjiv Das, Arnout Eikeboom, Leonid Kogan, Tri-Dung Nguyen, Tjeerd van Cappelle, the editor (Wei Xiong), the associate editor (Haitao Li), two referees, and participants of the JOIM (*Journal of Investment Management*) Fall 2008 Conference and the MIT Finance Lunch for helpful comments and discussions.

Appendix

In this appendix, we provide proofs for the main results of this paper.

Proof of Proposition 1. When there are only $n = 2$ assets, we may write the set of points on the frontier simply as

$$\mathcal{F} = \left\{ \boldsymbol{\omega} \colon \boldsymbol{\omega} = \left[\frac{\mu_o - \mu_1}{\mu_2 - \mu_1} \;\; \frac{\mu_2 - \mu_o}{\mu_2 - \mu_1} \right]' \text{ for } \mu_o \geq \frac{B}{C} \right\}.$$

Thus, for all expected returns μ_o with $\mu_1 < \mu_o < \mu_2$, points on the frontier have positive weight in both components, but for all values of μ_o outside this range, every point on the frontier has exactly one negative component. If the minimum value of μ_o, namely, $\mu_o = B/C$, is less than μ_2, then at least some point on the frontier has all positive weights, but if this value of μ_o is greater than μ_2, then all points on the frontier have at least one negative weight.

The condition that $B/C < \mu_2$ is the same as the condition that

$$\frac{\mu_1(\mathbf{e}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}) + \mu_2(\mathbf{e}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota})}{(\mathbf{e}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}) + (\mathbf{e}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota})} < \mu_2.$$

The denominator on the left-hand side is nonnegative, according to the Cauchy-Schwarz inequality, so we may cross-multiply and collect terms to see that the inequality holds exactly when $\mathbf{e}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota} > 0$. This, in turn, is the same as the inequality $\rho < \sigma_2/\sigma_1$, so we see that a frontier will be impossible just when $\sigma_2/\sigma_1 < \rho \leq 1$ and $\mu_1 < \mu_2$, which is the assertion of Proposition 1. $\square$

Proof of Lemma 1. We note that an $(n \times n)$ correlation matrix is positive definite if its elements are the cosines of angles between a pairs of vectors chosen from a set of n unit vectors that span an n-dimensional space. Thus, for a 3×3 matrix

$$\mathbf{M} = \begin{pmatrix} 1 & a & c \\ a & 1 & b \\ c & b & 1 \end{pmatrix},$$

positive definiteness holds if there are three vectors, $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$, such that a is the cosine of the angle between the first two vectors, b is the cosine of the angle between the last two vectors, and c is the cosine of the angle between the first and the last vector. Any values for a and b may be specified provided that $-1 < a, b < 1$, with strict inequality necessary to avoid linear dependence among the vectors. If such values of a and b are specified, then the value of c must follow the law of cosines so that

$$c = ab + \tilde{c}\sqrt{1-a^2}\sqrt{1-b^2} \tag{42}$$

for some $-1 < \tilde{c} < 1$. Strict inequality is again necessary to avoid linear dependence among the vectors. The definition of $\mathscr{C}$ in (6) specifies that $a = d$, $b = d$, and $c = d^4$. These values meet the requirements for positive definiteness, because c can be expressed in the form given in (42) with $\tilde{c} = d^2$. Thus $\mathscr{C}$ is positive definite and the lemma follows. □

Proof of Proposition 2. The idea of the proof is to compute $\boldsymbol{\omega}_\mu$, $\boldsymbol{\omega}_g$, and $\boldsymbol{\omega}_P$ directly. We obtain expressions for each of these portfolios in terms of ε, and we then analyze the results to show that, for small values of ε, all portfolios on the efficient frontier are impossible. We also analyze the results to obtain a lower bound for the minimum amount of short selling in any portfolio throughout the entire impossible frontier.

We turn first to the explicit computation of the three portfolios. To begin, we note that the inverse covariance matrix corresponding to the parameters in (6) is

$$\boldsymbol{\Sigma}^{-1} = \frac{\sigma_2^4(1-d^2)}{\det(\boldsymbol{\Sigma})}\begin{pmatrix} 1/d^2 & -(1+d^2) & d^2 \\ -(1+d^2) & (1+d^2)(1+d^4) & -d^2(1+d^2) \\ d^2 & -d^2(1+d^2) & d^2 \end{pmatrix}.$$

For $\boldsymbol{\mu}$ as defined in (6), we can multiply the above expression for $\boldsymbol{\Sigma}^{-1}$ by $\boldsymbol{\mu}$ and use the relationship $d = 1-\varepsilon$ to obtain

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \frac{\mu_2\sigma_2^4(1-d^4)}{\det(\boldsymbol{\Sigma})}\begin{pmatrix} \varepsilon+\varepsilon^2+\varepsilon^3+\varepsilon^4+O(\varepsilon^5) \\ -2\varepsilon+6\varepsilon^2-4\varepsilon^3+\varepsilon^4 \\ \varepsilon-\varepsilon^2 \end{pmatrix}.$$

Dividing the bracketed vector in this expression by the sum of its elements, we obtain

$$\boldsymbol{\omega}_\mu = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\boldsymbol{\iota}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}} = \frac{1}{72\varepsilon}\begin{pmatrix} 12+18\varepsilon+17\varepsilon^2+O(\varepsilon^3) \\ -24+60\varepsilon-10\varepsilon^2+O(\varepsilon^3) \\ 12-6\varepsilon-7\varepsilon^2+O(\varepsilon^3). \end{pmatrix},$$

which is the computation of $\boldsymbol{\omega}_\mu$ in terms of ε we desire. The computation of $\boldsymbol{\omega}_g$ proceeds similarly, with a direct computation of $\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}$ and a rescaling by the sum of vector components to obtain

$$\boldsymbol{\omega}_g = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}} = \frac{1}{72\varepsilon}\begin{pmatrix} 12+30\varepsilon+35\varepsilon^2+O(\varepsilon^3) \\ -24+60\varepsilon-22\varepsilon^2+O(\varepsilon^3) \\ 12-18\varepsilon-13\varepsilon^2+O(\varepsilon^3) \end{pmatrix}.$$

Finally, we compute $\boldsymbol{\omega}_P$ using the foregoing results to obtain

$$\boldsymbol{\omega}_P = B(\boldsymbol{\omega}_\mu - \boldsymbol{\omega}_g) = \frac{B}{12}\begin{pmatrix} -2-3\varepsilon+O(\varepsilon^2) \\ 2\varepsilon+O(\varepsilon^2) \\ 2+\varepsilon+O(\varepsilon^2) \end{pmatrix},$$

where

$$B = \boldsymbol{\iota}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \frac{\mu_2\sigma_2^4(1-d^4)}{\det(\boldsymbol{\Sigma})}(6\varepsilon^2-3\varepsilon^3+2\varepsilon^4) > 0.$$

The final inequality in the above expression holds because $\mu_2 > 0$, $\det(\boldsymbol{\Sigma}) > 0$, and $0 < \varepsilon < 1$.

We now use the foregoing computations to show that the efficient frontier is impossible for sufficiently small values of ε. Using (5), we see that every portfolio on the efficient frontier is of the form

$$\boldsymbol{\omega} = \alpha\tilde{\boldsymbol{\omega}}_P + \boldsymbol{\omega}_g,$$

where $\alpha \geq 0$ and $\tilde{\boldsymbol{\omega}}_P = (12/B)\boldsymbol{\omega}_P$. Inserting the above expressions for $\boldsymbol{\omega}_P$ and $\boldsymbol{\omega}_g$, we rewrite this expression for $\boldsymbol{\omega}$ as

$$\boldsymbol{\omega} = \alpha\begin{pmatrix} -2-3\varepsilon+O(\varepsilon^2) \\ 2\varepsilon+O(\varepsilon^2) \\ 2+\varepsilon+O(\varepsilon^2) \end{pmatrix} + \begin{pmatrix} 1/(6\varepsilon)+5/12+(35/72)\varepsilon+O(\varepsilon^2) \\ -1/(3\varepsilon)+5/6-(11/36)\varepsilon+O(\varepsilon^2) \\ 1/(6\varepsilon)-1/4-(13/72)\varepsilon+O(\varepsilon^2) \end{pmatrix}.$$

By solving for the value of α that makes the second coordinate zero, we see that the weight of the second asset is negative for portfolios corresponding to the range $0 \leq \alpha \leq \alpha_2$, where $\alpha_2 = 1/(6\varepsilon^2) + O(1/\varepsilon)$. Similarly, by solving for the value of α that makes the first coordinate zero, we see that the weight of the first asset is negative for portfolios corresponding to the range $\alpha \geq \alpha_1$, where $\alpha_1 = 1/(12\varepsilon) + O(1)$. For sufficiently small values of ε, we have $\alpha_2 \gg \alpha_1$, and so the efficient frontier is impossible. Specifically, the frontier has negative weights in the second asset if α is between 0 and α_2, and negative weights in the first asset if α is greater than α_1. Because α_1 is less than α_2, there is always at least one asset with a negative weight, and in the range between α_1 and α_2, both the first and second assets have negative weight.

Finally, we note that the portfolio with the minimum amount of short selling on the entire frontier occurs at the point α_1. To see this, we observe that for α values less than α_1, only the second asset has a negative weight, and this weight decreases as α increases. Also, for α values greater than α_1, the first asset has a negative weight, and it increases with α more quickly than the negative weight of the second asset decreases. Thus, the minimum amount of short selling occurs at $\alpha = \alpha_1$, and for small values of ε, this amount of short selling is $1/(3\varepsilon) + O(1)$. □

Proof of Lemma 2. The first inequality in (18) states that the ith component of $\boldsymbol{\omega}_g$ is negative, and the second inequality states that the ith component of $\boldsymbol{\omega}_P = B\boldsymbol{\omega}_\mu - B\boldsymbol{\omega}_g$ is also negative. Together, these inequalities imply that the ith component of each portfolio on the entire efficient frontier has a negative weight, because C and D are always positive, by the Cauchy-Schwartz inequality, and because frontier portfolios have the form described in (5). This demonstrates the sufficiency of the condition for impossibility in the ith asset. The necessity also follows readily because a negative ith component of each portfolio is only possible if there is a negative ith component in the minimum risk portfolio, $\boldsymbol{\omega}_g$, as well as in the high-risk portfolios that tend toward a positive multiple of $\boldsymbol{\omega}_\mu - \boldsymbol{\omega}_g$.

To deduce the equivalence between the conditions in (18) and (19), we note that

$$\boldsymbol{\Sigma}^{-1} = (\mathbf{A}^{-1})'\mathbf{M}^{-1}\mathbf{A}^{-1}.$$

As in Definition 2, we can express $\mathbf{M}$ in terms of coordinates as $(\mathbf{X}, \mathbf{W}, \mathbf{V})$, and we have

$$\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{I}_2 & \mathbf{0} \\ -\mathbf{X} & \mathbf{I}_{n-2} \end{pmatrix} \begin{pmatrix} \mathbf{W}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_2 & -\mathbf{X}' \\ \mathbf{0} & \mathbf{I}_{n-2} \end{pmatrix} = \begin{pmatrix} \mathbf{W}^{-1} & -\mathbf{W}^{-1}\mathbf{X}' \\ -\mathbf{X}\mathbf{W}^{-1} & \mathbf{V}^{-1} + \mathbf{X}\mathbf{W}^{-1}\mathbf{X}' \end{pmatrix}.$$

From the definition of $\mathbf{A} = \mathbf{A}(c_1, \ldots, c_n)$ in (17), we see that

$$\mathbf{A}^{-1}\boldsymbol{\iota} = \mathbf{e}_1/c_1, \quad \mathbf{A}^{-1}\boldsymbol{\mu} = \mathbf{e}_2/c_2, \quad \text{and}$$
$$\mathbf{A}^{-1}\mathbf{e}_j = \mathbf{e}_j/c_j, \quad \text{for } 3 \le j \le n.$$

We write $\mathbf{W} = [w_{ij}]$ so that

$$\mathbf{W}^{-1} = \frac{1}{\det(\mathbf{W})} \begin{pmatrix} w_{22} & -w_{12} \\ -w_{12} & w_{11} \end{pmatrix}.$$

After some algebraic rearrangements, we see that the conditions in (18) are equivalent to

$$x_{1,(i-2)} w_{22} - x_{2,(i-2)} w_{12} > 0 \quad \text{and} \quad x_{(i-2),2} > 0, \tag{43}$$

where we have used the facts that $\det(\mathbf{W}) > 0$ and $w_{22} > 0$, because $\mathbf{W}$ is positive definite.

With the notation from Definition 2, we write $w_{22} = v$, $w_{12} = yv$, $x_{1,i-2} = r_{i-2}\cos\theta_{i-2}$, and $x_{2,i-2} = r_{i-2}\sin\theta_{i-2}$. Equation (43) can be rewritten in terms of these new coordinates as

$$\cos\theta_{i-2} - y\sin\theta_{i-2} > 0 \quad \text{and} \quad \sin\theta_{i-2} > 0$$

because both $r_{i-2} > 0$ and $v > 0$, and this is the condition in (19). □

Proof of Theorem 1. From Equation (19) of Lemma 2 we see that the probability, p_i, that a frontier is impossible with respect to the ith coordinate, for $i > 2$, is just the probability that the conditions of (19) are fulfilled when θ_{i-2} is chosen from the uniform distribution on $S^1 = [0, 2\pi]$. The conditions are satisfied exactly when $\theta \in (0, \pi)$ and $y < \cot\theta_{i-2}$, and this corresponds to a probability of impossibility

$$p_i = \frac{1}{2\pi}\cot^{-1} y,$$

where $\cot^{-1}$ denotes the branch of the inverse cotangent with values between 0 and π.

Equation (19) of Lemma 2 also shows that for a fixed value of y, impossibility in the ith coordinate is independent of impossibility in the jth coordinate, for $i, j > 2$. Thus, the probability of impossibility in at least one of the coordinates $i > 2$ is bounded below as

$$p \ge 1 - \left(1 - \frac{1}{2\pi}\cot^{-1} y\right)^{n-2},$$

and this implies the first inequality of the theorem. The second inequality follows directly, because the inequality $\cot^{-1} y \ge 1/(1+\max(0, y))$ holds for all y. □

Proof of Theorem 2. From Corollary 1, we see that the probability in the theorem is bounded below as

$$\begin{aligned} P_I &\ge 1 - \int_{\mathbf{R}} \left(1 - \frac{1}{2\pi(1+\max(0, y))}\right)^{n-2} \varphi_y(y) \\ &\ge 1 - \int_{-\infty}^{0} \left(1 - \frac{1}{2\pi}\right)^{n-2} \varphi_y(y) \\ &\quad - c\int_0^{\infty} \left(1 - \frac{1}{2\pi(1+y)}\right)^{n-2} e^{-y^2}\,dy. \end{aligned}$$

The first integral in the last line is bounded above by $(1 - 1/(2\pi))^{n-2}$. The second integral is bounded above by the sum

$$c\int_0^2 \left(1 - \frac{1}{6\pi}\right)^{n-2} e^{-y^2}\,dy + c\int_2^{\infty} \left(1 - \frac{1}{3\pi y}\right)^{n-2} e^{-y^2}\,dy.$$

The first integral in this sum is bounded above by $c(1 - 1/(6\pi))^{n-2}$, and the second integral is bounded above by $2ce^{-((n-2)/(3\pi))^{2/3}}$. This last bound follows from the fact that $(1 - 1/(3\pi y))^{n-2}$ is bounded above by $e^{-((n-2)/(3\pi))^{2/3}}$ for $0 \le y \le ((n-2)/(3\pi))^{1/3}$, as well as the fact that

$$\int_{((n-2)/(3\pi))^{1/3}}^{\infty} e^{-y^2}\,dy < e^{-((n-2)/(3\pi))^{2/3}}.$$

Combining these results, we see that the probability is bounded below by

$$P_I \ge 1 - \left(\frac{6}{7}\right)^{n-2} - c\left(\frac{19}{20}\right)^{n-2} - 2c\exp\left(-\left(\frac{n-2}{3\pi}\right)^{2/3}\right),$$

where we have made use of the fact that $6/7 > 1 - 1/(2\pi)$ and the fact that $19/20 > 1 - 1/(6\pi)$. Finally, we note that

$$(19/20)^{n-2} \le 2\exp(-((n-2)/(3\pi))^{2/3}).$$

This inequality can be proven by noting that it holds if and only if the logarithm of the right-hand side minus the logarithm of the left-hand side is positive for $n \ge 2$. Computation of the value of this difference shows that it is positive for $n \le 27$, and computation of the derivative of the difference, taken with respect to n, shows that the derivative is positive for $n \ge 27$. As a result, the difference is positive for all n, and the inequality holds. Thus, we can bound P_I below as

$$P_I \ge 1 - \left(\frac{6}{7}\right)^{n-2} - 4c\exp\left(-\left(\frac{n-2}{3\pi}\right)^{2/3}\right). \quad \square$$

Proof of Theorem 3. The expected number of assets with respect to which an efficient frontier is impossible satisfies

$$E_n \ge \int_{\mathbf{R}} \left(\sum_{i=3}^{n} \frac{1}{2\pi(1+\max(0, y))}\right) \varphi_y(y).$$

This follows from the proof of Theorem 1, which shows that the ith summand in the integrand is a lower bound for the probability that a covariance matrix gives rise to an impossible frontier for a fixed value of y. We thus see that

$$E_n \ \ge\ (n-2)\int_{\mathbf{R}} \left(\frac{1}{2\pi(1+\max(0, y))}\right) \varphi_y(y),$$

and this last integral is the constant c' from the statement of the theorem. Also, in the case in which $\varphi_y(y)$ is a normal distribution with unit variance, we see from a numerical computation that

$$E_n \geq \frac{n-2}{8},$$

and this is the final claim of the theorem. $\square$

Proof of Theorem 4. If a frontier meets the necessary and sufficient conditions of Lemma 2 for the ith coordinate where $3 \leq i \leq n$, then it is an impossible frontier with respect to the ith asset. In this case, the ith components of both $\boldsymbol{\omega}_g$ and $\boldsymbol{\omega}_P$ are negative, so the total amount of short selling in the ith asset throughout the frontier is bounded below by the amount of short selling in the ith asset for the minimum-variance portfolio. Thus, we see that

$$S_i(\boldsymbol{\Sigma}) \geq -\mathbf{e}_i'\boldsymbol{\omega}_g = -\frac{\mathbf{e}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\iota}} = -(-r_{i-2}\cos\theta_{i-2} + y r_{i-2}\sin\theta_{i-2}),$$

where we have used the change of coordinates $\boldsymbol{\Sigma} = \mathbf{AMA}'$ and the coordinates for $\mathbf{M}$ from Definition 2 to establish the final equality. We thus see that the expected amount of short selling with respect to the ith asset satisfies

$$E[S_i] \geq \int_0^{\infty}\left(\int_0^{\pi}\int_{-\infty}^{\cot\theta_{i-2}}(r_{i-2}\cos\theta_{i-2} - y r_{i-2}\sin\theta_{i-2})\,\varphi_y\frac{d\theta_{i-2}}{2\pi}\right)\varphi_{r_{i-2}}, \quad (44)$$

where we have used the result from Lemma 2 in which a frontier is impossible with respect to the ith asset exactly when $\theta_{i-2} \in (0, \pi)$ and $y < \cot\theta_{i-2}$. We have also used the notation φ_y and $\varphi_{r_{i-2}}$ from Definition 3.

Because the integrand in (44) is positive throughout the region of integration, we can find a smaller lower bound by restricting the size of the region of integration. We calculate

$$E[S_i] \geq \left(\int_0^{\infty} r_{i-2}\varphi_{r_{i-2}}\right)\left(\int_0^{\pi/2}\int_{-\infty}^{0}(\cos\theta_{i-2} - y\sin\theta_{i-2})\,\varphi_y\frac{d\theta_{i-2}}{2\pi}\right)$$
$$= \frac{1}{2\pi}\left(\int_0^{\infty} r_{i-2}\varphi_{r_{i-2}}\right)\left(\int_{-\infty}^{0}(1-y)\,\varphi_y\right).$$

This is the lower bound in the theorem for $E[S_i]$. The lower bound for $E[S]$ follows immediately if the $\varphi_{r_{i-2}}$ are identical for $3 \leq i \leq n$. $\square$

Proof of Theorem 5. Let $\mathcal{F}_0$ be the frontier constrained to allow no short selling that corresponds to $\mathcal{F}$ and $\mathcal{F}_b$. Let σ_0 be the risk of the minimum risk portfolio on $\mathcal{F}_0$, and let μ_0 be the expected return of the maximum-expected-return portfolio on $\mathcal{F}_0$. Each portfolio on $\mathcal{F}_b$ with a lower risk than σ_0 must involve short selling, because σ_0 is the minimum possible risk without short selling. Similarly, each portfolio on $\mathcal{F}_b$ with a higher expected return than μ_0 must involve short selling, because μ_0 is the maximum possible expected return without short selling. Thus, we need only show that each portfolio on $\mathcal{F}_b$ with a risk greater than or equal to σ_0 and an expected return less than or equal to μ_0 must involve short selling.

Let $\boldsymbol{\omega}_b$ be a portfolio on $\mathcal{F}_b$ with risk and expected return σ_b and μ_b, respectively, such that $\sigma_b \geq \sigma_0$ and $\mu_b \leq \mu_0$. There are weight vectors $\boldsymbol{\omega}_0$ and $\boldsymbol{\omega}_U$ on $\mathcal{F}_0$ and $\mathcal{F}$, respectively, with the same expected return as $\boldsymbol{\omega}_b$. For $0 \leq \lambda \leq 1$, write

$$\boldsymbol{\omega}_\lambda = (1-\lambda)\boldsymbol{\omega}_0 + \lambda\boldsymbol{\omega}_U$$

so that each $\boldsymbol{\omega}_\lambda$ also has the same expected return μ_b. Let σ_λ denote the risk of $\boldsymbol{\omega}_\lambda$. Note that σ_λ is a decreasing function of λ, for $0 \leq \lambda \leq 1$, because σ_λ^2 is a quadratic function of λ, and because $\lambda = 1$ corresponds to the minimum risk portfolio for the level of expected return μ_b. We assume here that $\mathcal{F}$ is impossible so that $\boldsymbol{\omega}_U$ involves short selling and is therefore distinct from $\boldsymbol{\omega}_0$. We thus see that each $\boldsymbol{\omega}_\lambda$ with $\lambda > 0$ has lower risk than $\boldsymbol{\omega}_0$ but the same level of return μ_b. Also, the amount of short selling in $\boldsymbol{\omega}_\lambda$ is positive for all $\lambda > 0$ but goes to zero as $\lambda \to 0$. As a result, there is some $\lambda^* > 0$ such that the amount of short selling on $\boldsymbol{\omega}_{\lambda^*}$ is no more than b. The existence of this $\boldsymbol{\omega}_{\lambda^*}$ implies that σ_b must be no greater than $\boldsymbol{\omega}_{\lambda^*}$, and hence strictly less than the risk of $\boldsymbol{\omega}_0$. Because the risk of $\boldsymbol{\omega}_b$ must be strictly less than the risk of $\boldsymbol{\omega}_0$, it follows that $\boldsymbol{\omega}_b$ must involve short selling, as desired. $\square$

Proof of Theorem 6. From the condition for impossibility in Proposition 3 and from the expression for $\mathbf{x}$ in (37), we see that the probability is bounded below by

$$p \geq \int_{\mathbf{R}^{n-1}} F(\mathbf{x})\varphi_\gamma(\boldsymbol{\gamma})$$
$$= \int_{\mathbf{R}^{n-1}} F\left(\left(\frac{\Omega_{11}/\mu_1}{1+\Omega_{11}\|\widetilde{\boldsymbol{\Omega}}^{-1/2}\mathbf{z}\|^2}\right)\widetilde{\boldsymbol{\Omega}}^{-1}\mathbf{z}\right)\varphi_\gamma(\boldsymbol{\gamma}),$$

where $\mathbf{z} = \boldsymbol{\gamma} - \tilde{\boldsymbol{\mu}}/\mu_1$. Because F depends only on the signs of the elements of its argument, we have

$$p \geq \int_{\mathbf{R}^{n-1}} F(\widetilde{\boldsymbol{\Omega}}^{-1}\mathbf{z})\,\varphi_\gamma(\boldsymbol{\gamma}),$$

and after a change of variables, we see that

$$p \geq \det(\widetilde{\boldsymbol{\Omega}})\int_{\mathbf{R}^{n-1}} F(\boldsymbol{\gamma})\,\varphi_\gamma(\widetilde{\boldsymbol{\Omega}}\boldsymbol{\gamma} + \tilde{\boldsymbol{\mu}}/\mu_1)$$

as desired. $\square$

Proof of Corollary 2. Substitution of the specified choice of φ_γ into the result of Theorem 6 shows that

$$p \geq \frac{1}{(2\pi s)^{(n-1)/2}}\int_{\mathbf{R}^{n-1}} F(\boldsymbol{\gamma})\exp\left(-\frac{1}{2s}\boldsymbol{\gamma}'\boldsymbol{\gamma}\right)d\boldsymbol{\gamma}.$$

This integral is simply an expression for the fraction of the unit sphere in $\mathbf{R}^{n-1}$ that does not have either all negative or all positive coordinates, and this fraction is $1 - 2^{2-n}$, as desired. $\square$

References

Anderson, T. 1984. *An Introduction to Multivariate Statistical Analysis*, 2nd ed. John Wiley & Sons, New York.

Bawa, V., S. Brown, R. Klein. 1979. *Estimation Risk and Optimal Portfolio Choice*. North-Holland, Amsterdam.

Berger, J., 1985, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer-Verlag, New York.

Best, M. J., R. R. Grauer. 1985. Capital asset pricing compatible with observed market value weights. *J. Finance* **40** 85–103.

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

Best, M. J., R. R. Grauer. 1991b. Sensitivity analysis for mean-variance portfolio problems. *Management Sci.* **37** 980–989.

Best, M. J., R. R. Grauer. 1992. Positively weighted minimum-variance portfolios and the structure of asset expected returns. *J. Financial Quant. Anal.* **27** 513–537.

Black, F., R. Litterman. 1992. Global portfolio optimization. *Financial Analysts J.* **47** 28–43.

Box, G., G. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA.

Brown, S. 1976. Optimal portfolio choice under uncertainty: A Bayesian approach. Ph.D. dissertation, University of Chicago, Chicago.

DeMiguel, V., L. Garlappi, R. Uppal. 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Rev. Financial Stud.* **22** 1915–1953.

Fabozzi, F., P. Kolm, D. Pachamanova, S. Focardi. 2007. *Robust Portfolio Optimization and Management*. Wiley Finance, New York.

Frost, P., J. Savarino. 1986. An empirical bayes approach to efficient portfolio selection. *J. Financial Quant Anal.* **21** 293–305.

Garlappi, L., R. Uppal, T. Wang. 2007. Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Rev. Financial Stud.* **20** 41–81.

Green, R., B. Hollifield. 1992. When will mean-variance efficient portfolios be well diversified? *J. Finance* **47** 1785–1809.

Green, R. C. 1986. Positively weighted portfolios on the minimum-variance frontier. *J. Finance* **41** 1051–1068.

Jagannathan, R., T. Ma. 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *J. Finance* **58** 1651–1683.

Jeffreys, H. 1961. *Theory of Probability*, 3rd ed. Oxford University Press, Oxford, UK.

Jorgenson, J., S. Lang. 2005. *$Pos_n(R)$ and Eisenstein Series*. Springer-Verlag, Berlin.

Jorion, P. 1986. Bayes-Stein estimation for portfolio analysis. *J. Financial Quant. Anal.* **21** 279–292.

Kan, R., G. Zhou. 2007. Optimal portfolio choice with parameter uncertainty. *J. Financial Quant. Anal.* **42** 621–656.

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Rev. Econom. Statist.* **47** 13–37.

Markowitz, H. 1952. Portfolio selection. *J. Finance* **7** 77–91.

Markowitz, H. 1959. *Portfolio Selection*. Cowles Foundation, Yale University, New Haven, CT.

Markowitz, H. 2005. Market efficiency: A theoretical distinction and so what? *Financial Analysts J.* **60** 17–30.

Merton, R. 1972. An analytic derivation of the efficient portfolio frontier. *J. Financial Quant. Anal.* **7** 1851–1872.

Michaud, R. 1998. *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Harvard Business School Press, Boston.

Roll, R. 1977. Critique of the asset pricing theory's tests: Part I: On past and potential testability of the theory. *J. Financial Econom.* **4** 129–176.

Roll, R., S. A. Ross. 1977. Comments on qualitative results for investment proportions. *J. Financial Econom.* **5** 265–268.

Ross, S. 1976. The arbitrage theory of capital asset pricing. *J. Econom. Theory* **13** 341–360.

Rudd, A. 1977. Note on qualitative results for investment proportions. *J. Financial Econom.* **5** 259–263.

Sharpe, W. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *J. Finance* **19** 425–442.

Tobin, J. 1958. Liquidity preference as behavior towards risk. *Rev. Econom. Stud.* **67** 65–86.

Tu, J., G. Zhou. 2004. Data-generating process uncertainty: What difference does it make in portfolio decisions? *J. Financial Econom.* **72** 385–421.

Tu, J., G. Zhou. 2007. Incorporating economic objectives into Bayesian priors: Portfolio choice under parameter uncertainty. Working paper, Olin School of Business, Washington University, St. Louis.

Tu, J., G. Zhou. 2008. Being naive about naive diversification: Can investment theory beat the 1/N strategy? Working paper, Olin School of Business, Washington University, St. Louis.

Wang, Z. 2005. A shrinkage approach to model uncertainty and asset allocation. *Rev. Financial Stud.* **18** 673–705.