Chapter 3: An Empirical Test for Two-Fund Separation[1]

We develop an empirical test for two-fund separation (TFS) at the market level. The test is useful for analyzing representative investor models of capital market equilibrium. Combined with a test for stochastic dominance efficiency, our test can determine if (i) all well-behaved representative investor models can be accepted, (ii) all these models have to be rejected or (iii) only some models can be accepted and hence the conclusions are not robust with respect to model specification error.

TWO-FUND SEPARATION (TFS) implies that all well-behaved investors mix the market portfolio of risky assets with the riskless asset. Tobin (1958), Cass and Stiglitz (1970) and Ross (1978), among others, provide a rigorous analysis of this concept. Separating return distributions such as elliptical distributions (including the normal distribution) are frequently used to justify the use of representative investor models of capital market equilibrium. For example, without a separating distribution, there hardly is a case for the classic mean-variance CAPM. Of course, the mean-variance model applies also if all investors have quadratic utility. However, quadratic utility is generally seen as highly implausible because it allows for decreasing utility (satiation) and does not allow for decreasing absolute risk aversion (DARA). A similar case can be made against representative investor models that rely on other types of utility functions; economic theory simply does not forward a specific functional form for investor utility.

Given the central role of TFS in capital market models, there have been surprisingly few empirical tests of TFS. One way to test the TFS is by examining the actual investment portfolios of investors (for instance, Blume and Friend (1975)) or participants in controlled laboratory experiments (for example, Kroll, Levy and Rapoport (1988)). However, the observed behavior of individuals may not represent the behavior of the aggregated capital market. Unfortunately, as far as we know, the literature currently does not forward empirical tests for TFS at the market level. This study tries to fill the void by developing such a test.

As a primer on our test, Figure 1 below shows the empirical mean-beta plot for the ten CRSP size decile stock portfolios, using the CRSP all-share index as the market portfolio. Clearly, the observed mean-beta combinations come very close to the predicted security market line, although the small cap, high-beta portfolios appear undervalued. At first glance,

[1] This chapter is an adaptation of Post-Versijp (2005). In order to obtain a clearer exposition for this thesis, parts of this article are moved to other chapters, for example, an empirical illustration can be found in Chapter 4.

these results seem to support the mean-variance CAPM. However, if the mean-variance model is motivated by a separating return distribution, we would expect similar results for all other representative investor models.

One alternative model is the mean-lower partial moment (LPM) model of Bawa and Lindenberg (1977), which replaces the variance and regular beta with the general lower partial moment and LPM beta.[2] Figure 1 includes the results for this model if we select the second-order LPM with a target rate of return of -/-10% — a model of "mean-tail risk investors". The evidence in favor of the this model seems weaker than for the mean-variance model. The tail betas of the small cap portfolios are substantially smaller than the regular betas. In fact, the tail betas of the ten portfolios all are very close to unity. Presumably, this reflects the increased correlation between stocks during the more severe market downturns. Since the tail betas converge, investors who trade off mean return against tail risk will not hold the market portfolio but rather will underweight large cap stocks and overweight small cap stocks to capture the mean spread between the two types of stocks.

Since the market portfolio doesn't seem to be the optimal portfolio of risky assets for mean-tail risk investors, TFS appears not to apply in this case. Paradoxically, this also weakens the support for the mean-variance model, which becomes implausible if TFS is violated.

The above example considers only the mean-LPM model as an alternative to the mean-variance model and the fit may worsen even further for other alternative models. Also, the example relies on visual interpretation rather than statistical analysis and does not account for the variance-covariance structure of the portfolios. A proper empirical test for TFS would consider all well-behaved representative investor models and would do so in a statistically meaningful way. This study aims to develop such a test.

[2] Lower partial moments are generally considered more meaningful measures of risk than variance, because variance treats upside and downside deviations in a symmetric way, which is inconsistent with decreasing absolute risk aversion. By contrast, LPMs focus on the downside exclusively, and may use another reference point than the mean to measure deviations. For a thorough treatment of LPMs and their relation with the SD rules, see for instance Bawa (1975).

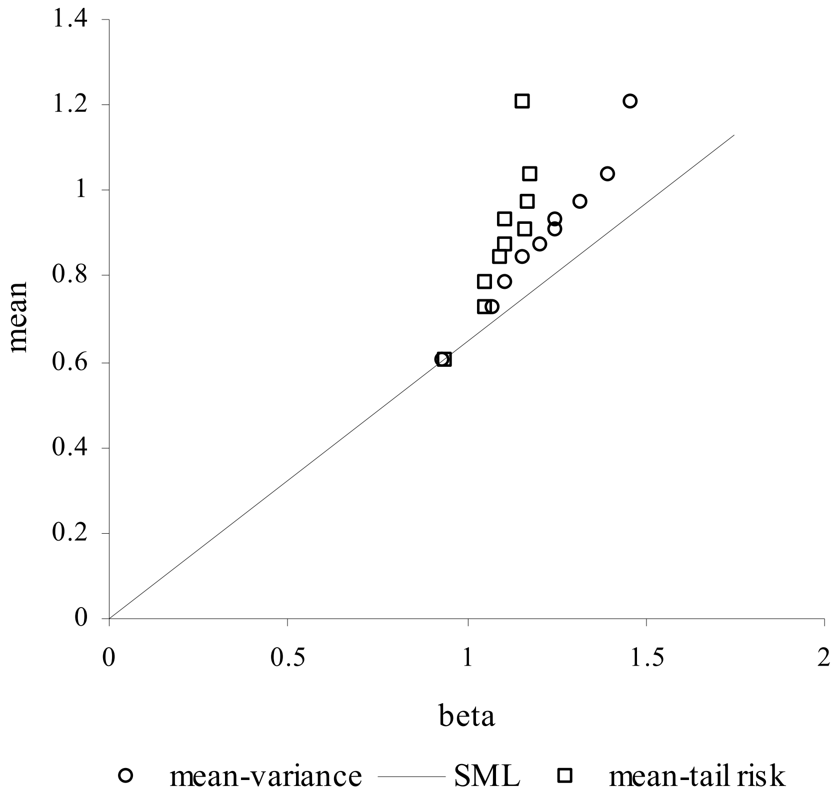

Figure 1

Empirical Fit of Mean-Variance and Mean-tail Risk Model

This figure shows the mean-beta relation for the mean-variance model (circles) and the mean-tail risk model (squares) for ten CRSP size decile portfolios of US stocks, using the CRSP all-share index as the market portfolio. The results are based on monthly excess returns from July 1926 to December 2003 (T=930 months). Tail risk is measured by the second-order lower partial moment with a target rate of return of -10%. The straight line through the origin represents the security market line (SML) with the slope set equal to the sample equity premium. The raw return data are taken from the on-line data library of Kenneth French: http://mba.tuck.dartmouth.edu/ pages/faculty/ken.french/data_library.html. Excess returns are computed by subtracting the one-month US Treasury bill (from Ibbotson Associates) from the raw returns.

Our TFS test is related to the recently developed tests for stochastic dominance efficiency of the market portfolio of Post (2003) and Post and Versijp (2007). Those tests ssentially ask if there exists *some* well-behaved utility function for which the market portfolio is optimal. Our TFS test basically asks if the market portfolio is the optimal portfolio of risky assets for *all* well-behaved utility functions. Combined with a test for stochastic dominance efficiency, the TFS test effectively covers all representative investor models of capital market equilibrium.

The combined tests are useful as first-stage screening devices in empirical asset pricing. If stochastic dominance efficiency is rejected, all representative investor models have to be rejected, and if TFS is not rejected, no representative investor model can be rejected. In both cases, there is no need for parameterization of the model, as the result of every (well-behaved) model is known in advance. Similarly, the tests are useful for robustness analysis. If the stochastic dominance efficiency test and the TFS test yield the same result (accept/reject), the results of every representative investor model are robust with respect to the model specification. By contrast, if stochastic dominance efficiency is not rejected but TFS is rejected, the results are not robust, which is a cause for concern given the ambiguity surrounding the model specification.

Computational burden is an important consideration in our analysis. Specifically, the typical approach to testing efficiency is to *minimize* a (weighted) sum of squared errors across all admissible utility functions. This problem can be solved with straightforward convex quadratic programming (CQP) techniques. Unfortunately, the same approach is more difficult for *maximizing* the sum of squares, which requires non-convex quadratic programming techniques. To obtain a computationally tractable test, we focus on maximizing the maximum (standardized) error rather than the sum of squared errors. The associated optimization problem can again be solved with CQP techniques.

The remainder of this text is structured as follows. Section 3.1 introduces preliminary notation, assumptions and definitions. Since the TFS test is especially useful in combination with a SSD efficiency test, we adhere to the framework of Post and Versijp (2007; also see Chapter 2). Section 3.2 discusses our novel TFS test, including the practical issues of computation, statistical inference and statistical size and power. Finally, Section 3.3 gives concluding remarks and suggestions for further research. The Appendix gives formal proofs of our theorems.

3.1 Preliminaries

We consider a single-period, portfolio-based, representative-investor model of the capital market that satisfies the following assumptions:

ASSUMPTION 1 (Investor preferences) The representative investor is nonsatiable and risk-averse. The representative investor chooses an investment portfolio to maximize the expected utility associated with the portfolio's return. Throughout the text, we will denote utility functions by $u:\Re \to P$, $u \in U_2$, with U_2 for the set of twice continuously differentiable, strictly increasing and concave, von Neumann-Morgenstern utility functions, and P for a nonempty, closed, and convex subset of $\Re$.[3]

ASSUMPTION 2 (Portfolio possibilities) The investment universe consists of N risky assets and a riskless asset. Throughout the text, we will use the index set $\mathrm{I} \equiv \{1,\cdots,N\}$ to denote the different risky assets. Investors may diversify between the assets, and we will use $\boldsymbol{\lambda} \in \Re^N$ for a vector of portfolio weights. The market portfolio will be denoted by $\boldsymbol{\tau} \in \Re^N$.

ASSUMPTION 3 (Return distribution) The excess return vectors $\boldsymbol{x} \in \Re^N$ are random variables with a continuous joint cumulative distribution function (CDF) $G:\Re^N \to [0,1]$. To exclude a dominance relationship between the market portfolio and the riskless asset, we assume $\Pr[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau} < 0] > 0$ and $E[\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau}] > 0$. Further, to derive the asymptotic sampling distribution of our test statistics, we assume that the first-moment vector $\boldsymbol{\mu}(u) \equiv E[u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}] = \int u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}dG(\boldsymbol{x})$ is finite and that the second-moment matrix $\boldsymbol{\Omega}(u) \equiv E[u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})^2 \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}]$ is finite and positive-definite for all $u \in U_2$.

Under these assumptions, the representative investor's optimization problem can be summarized as $\max_{\boldsymbol{\lambda} \in \Re^N} E[u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\lambda})]$. The market portfolio $\boldsymbol{\tau} \in \Re^N$ is optimal for a given utility function $u \in U_2$ if and only if the Euler equation is satisfied:

---

[3] Throughout the text, we will use $\Re^N$ for an N-dimensional Euclidean space, and $\Re^N_+$ denotes the positive orthant. Further, to distinguish between vectors and scalars, we use a bold font for vectors and a regular font for scalars. Finally, all vectors are column vectors and we use $\boldsymbol{x}^{\mathrm{T}}$ for the transpose of $\boldsymbol{x}$.

$$E[u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}] = \mathbf{0}_N \qquad (1)$$

Using the terminology that is common in the asset pricing literature, the marginal utility function $u'(x)$ represents the *pricing kernel* and $E[u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}]$ represents the vector of *pricing errors*. If $E[u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})x_i] > 0$, asset $i \in \mathrm{I}$ is undervalued and its weight in the portfolio should be increased relative to τ_i. Similarly, if $E[u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})x_i] < 0$, asset $i \in \mathrm{I}$ is overvalued and its weight in the portfolio should be decreased.

Our TFS test is best understood in combination with the Post and Versijp (2007) test for second-order stochastic dominance (SSD) efficiency of the market portfolio. That test uses the following definition:

DEFINITION 1 (SSD Efficiency) The market portfolio $\boldsymbol{\tau} \in \Re^N$ is SSD efficient if and only if it is optimal for some well-behaved utility function, that is, $\exists u \in U_2 : E[u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}] = \mathbf{0}_N$.

The definition of TFS is a variation to this theme:

DEFINITION 2 (Two-Fund Separation) Two-fund separation applies if and only if the optimal portfolio for every well-behaved utility function $u \in U_2$ consists of a positive fraction invested in the market portfolio and the remainder in the riskless asset, that is,

$$E[u'(\kappa\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}] = \mathbf{0}_N,\ \kappa > 0 \quad \forall u \in U_2 \qquad (2)$$

If $\kappa > 1$, the investor borrows (takes a short position in the riskless asset) and takes a levered position in the market portfolio; if $\kappa < 1$, the investor lends (takes long position in the riskless asset); and if $\kappa = 1$, the investor holds the market portfolio without lending or borrowing. We may reformulate Definition 2 in a way that does not involve lending or borrowing:

THEOREM 1 (Reformulation TFS) *Two-fund separation applies if and only if the market portfolio is the optimal portfolio for all well-behaved investors who do not lend or borrow, that is,*

$$E[u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}] = \mathbf{0}_N \ \forall u \in U_2 : E[u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau}] = 0 \tag{3}$$

Combined, the concepts of SSD efficiency and TFS yield three scenarios (summarized in the table below). First, if the market portfolio is SSD inefficient, it is not optimal for any utility function and hence no representative investor model applies. In this case, we need a more complex model to describe capital market equilibrium, for instance an intertemporal consumption-oriented model or a heterogeneous investor model. Of course, in case of SSD inefficiency, TFS does not apply.

Second, if TFS applies, the market portfolio is the optimal portfolio of risky assets for all well-behaved utility functions and hence all representative investor models apply. This form of robustness for the model specification is comforting given that there exist few prior arguments for selecting a specific model (for instance the mean-variance model or the mean-tail risk model). Of course, if TFS applies, then the market portfolio must also be SSD efficient.

Third, when TFS does not apply and the market portfolio is SSD efficient, the market portfolio is optimal for only some utility functions and hence only some representative investor models apply. This situation is uncomforting given the uncertainty surrounding the correct specification.

	SSD efficiency	SSD inefficiency
TFS	All representative investor models apply	
No TFS	Some representative investor models apply	No representative investor model applies

To test the null hypothesis of TFS, we need full information on the CDF $G(\boldsymbol{x})$. In practical applications, $G(\boldsymbol{x})$ generally is not known and information is limited to a discrete set of T time-series observations.

ASSUMPTION 4 (Data set) The time-series observations are serially independently and identically distributed (IID) random draws from the CDF. Throughout the text, we will represent the observations by the matrix $\mathbf{X} \equiv (\boldsymbol{x}_1 \cdots \boldsymbol{x}_T)$, with $\boldsymbol{x}_t \equiv (x_{1t} \cdots x_{Nt})^{\mathrm{T}}$. Since the timing of the draws is inconsequential, we are free to label the observations by their ranking with respect to the evaluated portfolio, that is, $\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{\tau} < \boldsymbol{x}_2^{\mathrm{T}}\boldsymbol{\tau} < \cdots < \boldsymbol{x}_T^{\mathrm{T}}\boldsymbol{\tau}$.

Using the data set and the utility gradient vector $\nabla u \equiv (u'(\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{\tau})\cdots u'(\boldsymbol{x}_T^{\mathrm{T}}\boldsymbol{\tau}))^{\mathrm{T}}$, we can construct the following sample counterparts of $\boldsymbol{\mu}(u)$ and $\boldsymbol{\Omega}(u)$:[4]

$$\hat{\boldsymbol{\mu}}(\nabla u) = T^{-1}\mathbf{X}\nabla u \tag{4}$$

$$\hat{\boldsymbol{\Omega}}(\nabla u) = T^{-1}(\mathbf{X} \circ \nabla u^{\mathrm{T}})(\nabla u \circ \mathbf{X}^{\mathrm{T}}) \tag{5}$$

Since the observations are assumed to be serially IID, $\hat{\boldsymbol{\mu}}(\nabla u)$ and $\hat{\boldsymbol{\Omega}}(\nabla u)$ converge in probability to $\boldsymbol{\mu}(u)$ and $\boldsymbol{\Omega}(u)$ respectively. These properties are useful for statistical inference about TFS, as will be shown in the next section.

3.2 An Empirical Fest for Two-Fund Separation

A. THE TEST STATISTIC

Our test statistic for two-fund separation is best understood by first considering the test statistic for SSD efficiency of Post and Versijp (2007):

$$\xi(\mathrm{B}_2) \equiv \min_{\boldsymbol{\beta}\in \mathrm{B}_2}\{T\hat{\boldsymbol{\mu}}(\boldsymbol{\beta})^{\mathrm{T}}\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1}\hat{\boldsymbol{\mu}}(\boldsymbol{\beta})\}$$

with

$$\mathrm{B}_2 \equiv \left\{\boldsymbol{\beta} \in \Re_+^T : T^{-1}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{1}_T = 1; \beta_t \geq \beta_{t-1} \quad t = 2,\cdots,T\right\}$$

In the spirit of GMM, this test statistic minimizes a weighted sum of squares of the empirical pricing errors $\hat{\boldsymbol{\mu}}(\boldsymbol{\beta})$, with the elements of the inverted sample second-moment matrix $\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1}$ as weights. The polyhedron B_2 represents the properties of the gradient vector ∇u of well-behaved utility functions $u \in U_2$, standardized such that the mean equals unity. This standardization is typically used (explicitly or implicitly) in asset pricing tests, including the

[4] Slightly departing from the conventional notation, the Hadamard operator $\circ$ is used here for element-by-element multiplication of the elements of every column (row) of a matrix with a given column (row) vector.

well-known Gibbons, Ross and Shanken (GRS; 1989) test for mean-variance efficiency. Since utility functions are unique up to the level of a positive linear transformation, the standardization does not affect the efficiency classification or the p-values and critical values.

We propose the following test statistic for two-fund separation:

$$\theta(\mathrm{B}_2^+) \equiv \max_{\boldsymbol{\beta}\in\mathrm{B}_2^+}\{\max_{i\in\mathrm{I}}\{T^{1/2}(\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta}))_i^{-1/2}\hat{\boldsymbol{\mu}}(\boldsymbol{\beta})\}\} \tag{6}$$

with $(\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta}))_i^{-1/2}$ for the i-th row of $\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1/2}$ and

$$\mathrm{B}_2^+ \equiv \left\{\boldsymbol{\beta}\in\mathrm{B}_2 : T^{-1}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\boldsymbol{\tau}=0\right\} \tag{7}$$

Compared with $\xi(\mathrm{B}_2)$, $\theta(\mathrm{B}_2^+)$ has three distinguishing features. First, while $\xi(\mathrm{B}_2)$ employs the utility function that gives the *best* possible fit, $\theta(\mathrm{B}_2^+)$ uses the utility function with the *worst* fit possible; the minimization operator is replaced with a maximization operator.

Second, the sum of squared (standardized) errors $\hat{\boldsymbol{\mu}}(\boldsymbol{\beta})^{\mathrm{T}}\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta})^{-1}\hat{\boldsymbol{\mu}}(\boldsymbol{\beta})$ is replaced with the maximum (standardized) error $\max_{i\in\mathrm{I}}\{T^{1/2}(\hat{\boldsymbol{\Omega}}(\boldsymbol{\beta}))_i^{-1/2}\hat{\boldsymbol{\mu}}(\boldsymbol{\beta})\}$. This replacement is made to obtain a computationally tractable test (see Section II).

Third, while $\xi(\mathrm{B}_2)$ employs the admissible set B_2, $\theta(\mathrm{B}_2^+)$ uses B_2^+, which adds the restriction that the sample pricing error of the market portfolio should equal zero. This is the empirical equivalent of the restriction $E[u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau}]=0$, which stems from the need to exclude lending and borrowing; see Theorem 1. The GRS test for mean-variance efficiency also uses this restriction. Still, for efficiency tests, the restriction is not very important, as minimizing the individual pricing errors typically yields a near-zero pricing error for the market portfolio anyway. Indeed, the efficiency statistic $\xi(\mathrm{B}_2)$ does not impose this restriction and imposing it generally has a negligible effect for that statistic. By contrast, excluding this restriction in the TFS test (which maximizes the errors) generally leads to a large increase in the pricing errors, as the test will identify a utility function for which a high degree of borrowing or lending is optimal.

B. Computational Issues

Computational burden is an important consideration for using statistic (6). The SSD efficiency statistic $\xi(\mathrm{B}_2)$ minimizes a weighted sum of squared errors across all admissible utility functions, with the weights depending on the variance-covariance structure of the errors. For a given covariance matrix $\hat{\mathbf{\Omega}}(\gamma), \gamma \in \mathrm{B}_2^+$, this problem generally boils down to minimizing a convex quadratic function over a polyhedron—a relatively simple convex quadratic programming (CQP) problem; we only need to check if a potential solution is a Karush-Kuhn-Tucker point. The entire problem—including the determination of the optimal covariance matrix—can be solved by iteratively solving the CQP problem, where every iteration uses the covariance matrix based on the optimal gradient vector $\boldsymbol{\beta}^*$ from the previous iteration.

Unfortunately, the same approach is more difficult for maximizing the sum of squares; in this case, we maximize a convex function over a polyhedron and non-convex quadratic programming techniques are required. These techniques generally are computationally very demanding for high-dimension problems, such as ours (the problem dimension equals the length of the time-series (T)). By contrast, our TFS test statistic—based on the maximum absolute error rather than the sum of squared errors—can closely approximated by solving a small series of CQP problems.

The first step to reduce the computational burden is to proxy $\hat{\mathbf{\Omega}}(\boldsymbol{\beta})$ using $\phi(\boldsymbol{\beta})\hat{\mathbf{\Omega}}(\mathbf{1}_T)$, with $\hat{\mathbf{\Omega}}(\mathbf{1}_T) = (T^{-1}\mathbf{X}\mathbf{X}^{\mathrm{T}})$ and

$$\phi(\boldsymbol{\beta}) \equiv \left[\frac{T^{-1}(\boldsymbol{\beta} \circ \mathbf{X}^{\mathrm{T}}\boldsymbol{\tau})^{\mathrm{T}}(\boldsymbol{\beta} \circ \mathbf{X}^{\mathrm{T}}\boldsymbol{\tau})}{T^{-1}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\tau})^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\tau})} \right] \tag{8}$$

Thus, we assume that the effect of changing the gradient vector on every error covariance term is proportional to the effect on the error variance of the market portfolio. In typical asset pricing applications, the correlation between asset returns is fairly high and stable across different return levels, making this a plausible assumption. For example, the assumption does not materially affect the test for TFS for the CRSP size decile portfolios in Chapter 4; weighting with $\phi(\boldsymbol{\beta})\hat{\mathbf{\Omega}}(\mathbf{1}_T)$ produces very similar p-values as weighting with $\hat{\mathbf{\Omega}}(\boldsymbol{\beta})$, even for "extreme" kernels.

Our approximation is closely related to the approach by Hansen and Jagannathan (HJ, 1997) of replacing $\hat{\mathbf{\Omega}}(\boldsymbol{\beta})$ by $\hat{\mathbf{\Omega}}(\mathbf{1}_N)$.[5] However, rather than using $\hat{\mathbf{\Omega}}(\mathbf{1}_N)$ directly, we first multiply this matrix with the factor $\phi(\boldsymbol{\beta})$, so as to correct for the volatility introduced by the shape of the kernel. Without this correction, we would systematically underestimate the

[5] The matrix $\hat{\mathbf{\Omega}}(\mathbf{1}_N)^{-1}$ is commonly referred to as "the HJ weighting matrix".

elements of $\hat{\mathbf{\Omega}}(\boldsymbol{\beta})$ for "high-volatility kernels", which would adversely affect the statistical size of our test (that is, TFS would be rejected too often).[6]

Replacing $\hat{\mathbf{\Omega}}(\boldsymbol{\beta})$ with $\phi(\boldsymbol{\beta})\hat{\mathbf{\Omega}}(\mathbf{1}_T)$ in (6) yields the following approximation for the TFS test statistic:

$$\tilde{\theta}(\mathrm{B}_2^+) \equiv \max_{\boldsymbol{\beta}\in\mathrm{B}_2^+}\{\max_{i\in\mathrm{I}}\{T^{1/2}\phi(\boldsymbol{\beta})^{-1/2}(\hat{\mathbf{\Omega}}(\mathbf{1}_T))_i^{-1/2}\hat{\boldsymbol{\mu}}(\boldsymbol{\beta})\}\} \tag{9}$$

The second step is to express (9) in terms of a series CQP problems. We may without harm change the ordering of the two maximization operators, that is, $\tilde{\theta}(\mathrm{B}_2^+) = \max_{i\in\mathrm{I}}\{\max_{\boldsymbol{\beta}\in\mathrm{B}_2^+}\{T^{1/2}\phi(\boldsymbol{\beta})^{-1/2}(\hat{\mathbf{\Omega}}(\mathbf{1}_T))_i^{-1/2}\hat{\boldsymbol{\mu}}(\boldsymbol{\beta})\}\}$. Also, we may deal with the embedded optimization problem $\max_{\boldsymbol{\beta}\in\mathrm{B}_2^+}\{T^{1/2}\phi(\boldsymbol{\beta})^{-1/2}(\hat{\mathbf{\Omega}}(\mathbf{1}_T))_i^{-1/2}\hat{\boldsymbol{\mu}}(\boldsymbol{\beta})\}$, $i\in\mathrm{I}$, by changing the standardization $T^{-1}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{1}_T = 1$ used in the admissible set B_2^+. The purpose of this standardization is to ensure that $\boldsymbol{\beta} < \infty$, so as to exclude infinitely large positive pricing errors. We may without harm replace this standardization with $\phi(\boldsymbol{\beta}) \leq 1$, or, equivalently, $(\boldsymbol{\beta}\circ\mathbf{X}^{\mathrm{T}}\boldsymbol{\tau})^{\mathrm{T}}(\boldsymbol{\beta}\circ\mathbf{X}^{\mathrm{T}}\boldsymbol{\tau}) \leq (\mathbf{X}^{\mathrm{T}}\boldsymbol{\tau})^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\tau})$. Obviously, the inequality will be binding, or $\phi(\boldsymbol{\beta}) = 1$, for all nested problems $i\in\mathrm{I}$ with a positive solution. Hence, the term $\phi(\boldsymbol{\beta})^{-1/2}$ in (9) can be dropped if we change the standardization.

Changing the ordering of the maximization operators and using the new standardization, we can rewrite (9) in the following way:

$$\tilde{\theta}(\mathrm{B}_2^+) = \max_{i\in\mathrm{I}}\{\max_{\boldsymbol{\beta}\in\tilde{\mathrm{B}}_2^+}\{(\mathbf{X}\mathbf{X}^{\mathrm{T}})_i^{-1/2}(\mathbf{X}\boldsymbol{\beta})\}\} \tag{10}$$

with

$$\tilde{\mathrm{B}}_2^+ \equiv \Big\{\boldsymbol{\beta}\in\Re_+^T : (\boldsymbol{\beta}\circ\mathbf{X}^{\mathrm{T}}\boldsymbol{\tau})^{\mathrm{T}}(\boldsymbol{\beta}\circ\mathbf{X}^{\mathrm{T}}\boldsymbol{\tau}) \leq (\mathbf{X}^{\mathrm{T}}\boldsymbol{\tau})^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\tau});$$
$$T^{-1}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\boldsymbol{\tau} = 0; \beta_t \geq \beta_{t-1} \quad t = 2,\cdots,T\Big\} \tag{11}$$

[6] Indeed, if we use $\hat{\mathbf{\Omega}}(\mathbf{1}_N)$ in the application in Section 3.3, the test rejects TFS based on the large pricing errors that occur for mean-tail risk models (see also Figure 1). However, this approach ignores that it is difficult to estimate tail risk in a reliable way and that the mean-tail risk pricing errors are highly uncertain. If we correct for this additional uncertainty, we can no longer reject mean-tail risk models.

The embedded problem $\max_{\boldsymbol{\beta}\in\tilde{B}_2^+}\{(\mathbf{XX}^T)_i^{-1/2}(\mathbf{X}\boldsymbol{\beta})\}$, $i \in \mathrm{I}$, involves maximizing a linear objective function under linear equality and inequality constraints and a convex quadratic inequality constraint. This is a standard CQP problem. Thus, the entire problem can be solved by enumerating a series of *N* standard CQP problems (one CQP problem for every asset).

The change of the standarisation does not affect the test statistic because the resulting change of the pricing errors is compensated by the effect on the weighting matrix. Still, when analyzing the pricing kernel and the pricing errors rather than the test statistic, it can be useful (for comparison with other studies and methodologies) to return to the original standardization (the sample mean of the gradient vector equals unity) by using $\boldsymbol{\beta}' = \boldsymbol{\beta}(T^{-1}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{1}_T)^{-1}$.

C. STATISTICAL INFERENCE

For the purpose of statistical inference, we need to characterize the sampling distribution of the TFS test statistic. Unfortunately, the restrictions imposed on the utility function make it difficult to characterize the exact sampling distribution. This problem is encountered in constrained parametric models as well, since the constraints may be binding or non-binding depending on the return distribution in an unknown manner. Nevertheless, it is possible to derive a bound on the asymptotic p-values and critical values. Specifically, using $\Phi(\cdot)$ for the standard normal CDF, we may derive the following result:

THEOREM 2 (Asymptotic sampling distribution) Asymptotically, $\Pr[\theta(\mathrm{B}_2^+) \geq y] \geq 1 - \Phi(y)^N$ for every $y \geq 0$.

This theorem can be used for statistical inference. Specifically, the asymptotic *p*-value associated with the observed value of the test statistic is always greater than or equal to $1 - \Phi(\theta(\mathrm{B}_2^+))^N$. Equivalently, the asymptotic critical value associated with a significance level of α is always greater than or equal to $\Phi^{-1}((1-\alpha)^{1/N})$. Thus, we cannot reject TFS at a significance level of α if $1 - \Phi(\theta(\mathrm{B}_2^+))^N \geq \alpha$, or if $\theta(\mathrm{B}_2^+) \leq \Phi^{-1}((1-\alpha)^{1/N})$. To give some feeling for the bound, the table below gives the asymptotic critical value $H_N^{-1}(\alpha)$ for several values for the significance level (α) and the number of assets (*N*).

$\Phi^{-1}((1-\alpha)^{1/N})$	α=10%	α=5%	α=1%
N=10	2.309	2.568	3.089
N=25	2.635	2.870	3.351
N=100	3.075	3.283	3.718

Similar to Theorem 2, the p-value for the SSD efficiency test statistic $\xi(\mathrm{B}_2)$ is bounded from above (rather than from below); see Post and Versijp (2004, Theorem 2). By considering the best possible well-behaved utility function, the efficiency test effectively maximizes the statistical size (probability of rejection when the null is correct). By contrast, the TFS test considers the worst possible utility function and thus maximizes the statistical power (probability of rejecting the null hypothesis when the null is incorrect).

As discussed in Section 3.1, the combined concepts of SSD efficiency and TFS yield three scenarios ("all models apply", "some models apply", "no model applies"). By using an upper bound for the p-value of the SSD efficiency test and a lower bound for the p-value of the TFS test, the empirical tests allows for a similar classification (summarized in the table below).

First, if SSD efficiency is rejected, all representative investor models can be rejected. Since the true p-value of the SSD test is smaller than the reported p-value, this conclusion applies for all significance levels smaller than or equal to α. Again, in this situation, we need to relax the assumptions behind single-period, consumption oriented, representative investor models in order to describe capital market equilibrium.

Second, if TFS is not rejected, no representative investor model can be rejected. Since the true p-value of the TFS test is larger than the reported p-value, this conclusion again applies for all significance levels smaller than or equal to α. To repeat, this form of robustness for the model specification is comforting given that there exist few prior arguments for selecting a specific model.

Third, if TFS is rejected but SSD efficiency is not rejected, we cannot draw an unambiguous conclusion; some representative investor model cannot be rejected but other ones can. Also, the rejection of TFS may reflect a lack of size of the TFS test (the true p-value of the TFS test is larger than the reported p-value) and the failure to reject SSD efficiency may reflect a lack of power of the SSD test (the true p-value of the SSD test is smaller than the reported p-value). As discussed before, this situation is very uncomforting given the uncertainty surrounding the appropriate model specification.

The table also includes the situation where SSD efficiency has to be rejected, but TFS cannot be rejected. If both tests would use the same aggregate error measure, this situation would never occur. However, the two tests use different error measures; while the SSD test uses the weighted sum of squared errors, the TFS test uses the maximum (standardized) error. Logically speaking, we therefore need to account for the possibility that the SSD p-value is smaller than the TFS p-value due to the maximum error being relatively small compared to the sum of squared errors. This situation is especially relevant if the error distribution has a thin right tail. However, for every utility function, the errors converge to a normal distribution (see the Proof to Theorem 2) and hence do not have thin tails. Also, the TFS test uses the worst possible utility function, thus inflating the errors, while the SSD efficiency test uses the best possible utility function, thus deflating the errors. For these reasons, the SSD p-value generally is greater than the TFS p-value and situations where SSD efficiency has to be rejected but TFS cannot be rejected are very rare.

	Accept SSD efficiency at significance level α	Reject SSD inefficiency at significance level α
Accept TFS at significance level α	Accept all representative investor models at significance level α	No logical conclusion can be drawn (a rare situation)
Reject TFS at significance level α	No robust conclusion can be drawn at significance level α	Reject all representative investor models at significance level α

D. SIMULATED SIZE AND POWER PROPERTIES

Post and Versijp (2004) analyze the statistical properties of the SSD and GRS efficiency tests (among other tests) by means of a simulation experiment. The simulation experiment builds on a data set of monthly excess returns of the CRSP beta decile stock portfolios during the period from January 1933 to December 2002 (840 months). Random samples were drawn from (1) a multivariate normal distribution with joint population moments equal to the first two sample moments of the portfolios—a case where the SSD test has no added value compared to the GRS test—and (2) the empirical, non-normal return distribution—a case where the SSD test becomes relevant. In this section, we will extend this simulation experiment to analyze the size and power properties of our TFS test.

We will first apply the test to the case of the normal distribution. In this case, we know that TFS applies and the tangency portfolio is the optimal portfolio of risky assets for all well-behaved investors. Hence, we may analyze the statistical size (rejection rate if the null applies) by the relative frequency of random samples in which TFS is wrongly rejected if we use the ex ante tangency portfolio as the market portfolio.

Next, we will consider the empirical, non-normal distribution. In this case, TFS does not hold and different investors have different optimal portfolios of risky assets.[7] For example, mean-tail risk investors can be demonstrated to assign a lower weight to low-beta stocks and a higher weight to high-beta stocks than mean-variance investors do. We will analyze the statistical power (rejection rate if the null is violated) of the TFS test by its ability to reject TFS for an ex ante SSD efficient benchmark portfolio.[8] The benchmark portfolio is selected by minimizing the second-order lower partial moment with a target rate of return of -/-10%, subject to the mean return being equal to that of tangency portfolio. For a separating distribution, this LPM(-10) portfolio would coincide with the tangency portfolio and would be optimal both for the mean-tail risk investor and the mean-variance investor. However, for the empirical distribution, the two portfolios differ substantially and the LPM(-10) portfolio is no longer optimal for the mean-variance investor. Thus, we may gauge the power of the TFS test by the relative frequency of random samples in which TFS is rejected if we use the LPM(-10) portfolio as the market portfolio.

Table I summarizes the simulation results. For the sake of comparison, Panel A includes the results for the GRS test.[9] Panel B includes the results for the TFS test. The left-hand side of the table reports the statistical size for various combinations of the sample size (T) and the significance level (α) and the right-hand side reports the power properties.

---

[7] We do not use the size decile portfolios in this section because TFS cannot be rejected for the empirical distribution of these portfolios (see Chapter 4), rendering any analysis of the statistical power based on the empirical distribution impossible.

[8] The Post and Versijp (2004) simulations used the SSD inefficient equal-weighted portfolio as the market portfolio to gauge the power of the efficiency tests. By contrast, we use an SSD efficient portfolio in this study, because we wish to separate the effect of violations of TFS from the effect of inefficiency.

[9] The SSD test is not included, because the LPM(-10) portfolio is SSD efficient and hence can not be used to gauge the power of the SSD test. By contrast, the LPM(-10) portfolio is highly mean-variance inefficient and hence allows for establishing the power of the GRS test.

Table I

Statistical Properties of the GRS and TFS Tests

The table displays the simulated size and power properties of the two tests for various numbers of time-series observations (T) and for various levels of significance (α). Panel A represents the Gibbons, Ross and Shanken (1989) mean-variance efficiency test. Panel B gives the results for our two-fund separation test. The simulations build on a data set of monthly excess returns of the CRSP beta decile stock portfolios and the CRSP all-share index from January 1933 to December 2002. The size is measured as the relative frequency of rejections of TFS in 10,000 random samples from a multivariate normal distribution with joint moments equal to the sample moments of the original data set and using the ex ante LPM(-10) portfolio as the market. Power is measured as the rejection rate in 10,000 random samples from the empirical, non-normal distribution.

Panel A: GRS mean-variance efficiency test						
	Size			Power		
	α=5%	α=2.5%	α=1%	α=5%	α=2.5%	α=1%
T=50	0.052	0.027	0.010	0.095	0.049	0.021
T=100	0.057	0.029	0.012	0.163	0.100	0.055
T=200	0.053	0.029	0.013	0.284	0.200	0.120
T=500	0.054	0.031	0.013	0.641	0.544	0.429
T=1,000	0.059	0.030	0.012	0.931	0.891	0.827
T=2,000	0.056	0.029	0.012	0.999	0.998	0.996
Panel B: Two-fund separation test						
	Size			Power		
	α=5%	α=2.5%	α=1%	α=5%	α=2.5%	α=1%
T=50	0.486	0.444	0.391	0.660	0.620	0.573
T=100	0.439	0.391	0.335	0.702	0.649	0.590
T=200	0.424	0.383	0.308	0.788	0.736	0.671
T=500	0.397	0.350	0.297	0.922	0.882	0.858
T=1,000	0.378	0.317	0.280	0.992	0.987	0.970
T=2,000	0.360	0.295	0.272	1.000	1.000	1.000

The size properties reflect our approach to statistical inference. For the GRS test, the exact sampling distribution is known, and the size equals the nominal significance level (save some simulation error). By contrast, for the TFS test, the exact distribution is not known and an asymptotic lower bound is used (see Theorem 2). Hence the p-values exceed the nominal significance level, even for large samples. For example, for T=1,000 and α=2.5%, the size is 31.7%. Recall that the asymptotic lower bound minimizes Type II error (failure to detect violations of TFS) and a relatively high size is the price one pays for this orientation. Indeed, our results suggest that the TFS test has very favorable power properties, even for relatively small samples. For example, for T=200 and α=2.5%, the power of the TFS test is 73.9%, compared with only 20.8% for the GRS test.

In conclusion, the TFS test appears very powerful for typical asset pricing applications, substantially more powerful than GRS test, which seems not well-suited for detecting violations of TFS. Of course, we have to pay a price for this power; the TFS test compromises size, like the SSD test compromises power for a favorable size. However, the size remains reasonable, at least in large samples.

3.3 Concluding Remarks

Given the pivotal role of separating distributions in asset pricing theory, our test for two-fund separation (TFS) seems a useful addition to the research instruments for empirical asset pricing. The test seems especially useful when combined with a test for stochastic dominance efficiency of the market portfolio. Both types of tests can serve as screening devices prior to testing parameterized models such as the mean-variance model or the mean-tail risk model, and for testing the robustness of such models.

In Section 3.2, parts B-D, we have shown how to deal with the practicalities of computation and statistical inference. The test statistic can be closely approximated by solving a convex quadratic programming (CQP) problem for every asset. Further, the simulation study suggests that a test procedure that uses our bound to the asymptotic sampling distribution has favorable statistical properties in typical asset pricing applications.

Another promising route for further research is the inclusion of additional assumptions about investor preferences. Like the definition of SSD efficiency, the definition of two-fund separation relies only on nonsatiation and risk aversion. The power of the efficiency test can be improved by replacing the SSD criterion with higher-order stochastic dominance rules, including the third-order stochastic dominance (TSD) rule and the fourth-order stochastic dominance (4SD) rule. The TSD rule adds the condition that investors like positive skewness (a necessary condition for decreasing absolute risk aversion). The 4SD rule further adds the condition that investors dislike kurtosis. Arguably, any realistic utility function should exhibit these two properties. Post and Versijp (2007) show that tests for TSD efficiency and 4SD efficiency can be obtained by adding further linear restrictions to the utility gradient vector. A similar approach can be adopted in our TFS test to improve the statistical size of the test; we need to add the additional restrictions to the restrictions imposed on B_2^+. Further research could focus on the additional statistical size obtained through adding such restrictions. Still, we stress that such additional restrictions won't change our empirical findings. We cannot find a single utility function, not even an "exotic" one that violates skewness preference or kurtosis aversion, for which the market portfolio is not optimal in our application. Thus, adding further restrictions will not change our conclusions.

Appendix

Proof of Theorem 1 (Reformulation TFS) The necessary condition is straightforward; if all investors combine the market portfolio and the riskless asset, the investors who do not borrow or lend must hold the market portfolio.

The sufficient condition requires us to show that (3) implies that every investor will combine the market portfolio and the riskless asset. For this purpose, it is useful to introduce a transformed version of a given $u \in U_2$:

$$v_u(x) \equiv \kappa_u^{-1} u(\kappa_u x)\; E[u'(\kappa_u \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})]^{-1} \tag{i}$$

with marginal utility $v_u'(x) = u'(\kappa_u x)E[u'(\kappa_u \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})]^{-1}$ and with κ_u set such that the market portfolio has a zero pricing error (no lending or borrowing):

$$E[v_u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau}] = E[u'(\kappa_u \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau}]E[u'(\kappa_u \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})]^{-1} = 0 \tag{ii}$$

There exists a unique $\kappa_u > 0$, because $k(\kappa) \equiv E[u'(\kappa \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau}]E[u'(\kappa \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})]^{-1}$ is a continuous and strictly decreasing function; increasing κ increases the weight assigned to the lowest returns (due to decreasing marginal utility) and hence lowers $k(\kappa)$.

Since $u'(x) \geq 0 \Rightarrow v_u'(x) \geq 0$ and $u'(x) \geq u'(y) \Rightarrow v_u'(x) \geq v_u'(y)$, $u \in U_2$ implies

$$v_u \in U_2 \tag{iii}$$

Given (ii) and (iii), (3) implies

$$E[v_u'(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}] = \boldsymbol{0}_N \Rightarrow E[u'(\kappa_u \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})\boldsymbol{x}] = \boldsymbol{0}_N \tag{iv}$$

Thus, the optimal portfolio for every $u \in U_2$ consists of a fraction κ_u invested in the market portfolio and the remainder in the riskless asset. **Q.E.D.**

Proof of Theorem 2 (Asymptotic sampling distribution) Consider the utility functions v_u, $u \in U_2$, as defined in (i). We know that the utility functions are feasible (iii) and the market portfolio is optimal for these utility functions (iv). Also, by construction, $E[v'_u(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\tau})] = 1$. Thus, asymptotically, $\nabla v_u \in \mathrm{B}_2^+$ and

$$\theta(\mathrm{B}_2^+) \xrightarrow{p} \max_{u \in U_2} \zeta(\nabla v_u) \tag{v}$$

where

$$\zeta(\nabla v_u) \equiv \max_{i \in \mathrm{I}} \{T^{1/2}(\hat{\boldsymbol{\Omega}}(\nabla v_u))_i^{-1/2} \hat{\boldsymbol{\mu}}(\nabla v_u)\}$$

We can derive the asymptotic distribution of $\zeta(\nabla v_u)$ for any given $u \in U_2$ from known results. Since the observations are serially IID, it follows from the Levy-Lindenberg central limit theorem that $\hat{\boldsymbol{\mu}}(\nabla v_u)$ obeys an asymptotic multivariate normal distribution with mean $\mathbf{0}_N$ (using $\boldsymbol{\mu}(v_u) = \mathbf{0}_N$ from (iv)) and variance-covariance matrix $T^{-1}\boldsymbol{\Omega}(v_u)$, i.e., $\hat{\boldsymbol{\mu}}(\nabla v_u) \xrightarrow{d} N(\mathbf{0}_N, T^{-1}\boldsymbol{\Omega}(v_u))$. Using this result and $\hat{\boldsymbol{\Omega}}(\nabla v_u) \xrightarrow{p} \boldsymbol{\Omega}(v_u)$, we find that $T^{1/2}\hat{\boldsymbol{\mu}}(\nabla v_u)^{\mathrm{T}}\hat{\boldsymbol{\Omega}}(\nabla v_u)^{-1/2}$ obeys an asymptotic multivariate normal standard normal distribution, i.e., $T^{1/2}\hat{\boldsymbol{\mu}}(\nabla v_u)^{\mathrm{T}}\hat{\boldsymbol{\Omega}}(\nabla v_u)^{-1/2} \xrightarrow{d} N(\mathbf{0}_N, \mathbf{I}_N)$. This distribution allows us to characterize the asymptotic distribution of $\zeta(\nabla v_u)$.

$$\Pr[\zeta(\nabla v_u) \geq y] = 1 - \Pr[T^{1/2}(\hat{\boldsymbol{\Omega}}(\nabla v_u))_i^{-1/2} \hat{\boldsymbol{\mu}}(\nabla v_u) \leq y] \quad i = 1, \cdots, N$$

$$= 1 - \Phi(y)^N \tag{vi}$$

Since $\theta(\mathrm{B}_2^+)$ asymptotically maximizes across all v_u $u \in U_2$ (see (v)), the asymptotic sampling distribution of $\theta(\mathrm{B}_2^+)$ must be bounded from below by that of $\zeta(\nabla v_u)$ for any given $u \in U_2$ gradient vector. Hence, asymptotically,

$$\Pr[\theta(\mathrm{B}_2^+) \geq y] \geq 1 - \Phi(y)^N \tag{vii}$$

for every $y \geq 0$. **Q.E.D.**