ANNALS OF ECONOMICS AND FINANCE **6**, 185–207 (2005)

Dichotomous Asset Pricing Model

Liang Zou*

University of Amsterdam, The Netherlands
E-mail: L.Zou@uva.nl

Cross-asset derivative securities are studied and a dichotomous asset pricing model (DAPM) is derived that significantly enriches the Sharpe-Lintner-Black capital asset pricing model. An assets beta is shown to be observable ex ante through the price of its cross-market call or put, and the DAPM separately predicts the assets' expected return - beta relations under the upper-market and lower-market conditions. A sufficient condition for the DAPM to hold is that assets return distributions satisfy Ross' (1978) two-fund separation property, which implies that any well-diversified portfolio is both mean-variance and gain-loss efficient.

Key Words: Mean-variance; Gain-loss; Upper-market beta; Lower-market beta; Cross-asset derivative security; Dichotomous asset pricing.
JEL Classification Number: G12.

1. INTRODUCTION

The mean-variance[1] and the gain-loss[2] analyses are two simple approaches to portfolio selection and asset pricing. Both approaches assume that investment decisions involve the optimal trade-off between risk and expected

* I wish to thank several participants at the 2003 meetings of the Econometric Society and the European Economic Association in Stockholm, as well as colleagues and seminar participants at the University of Amsterdam, Hong Kong Baptist University, University of Hong Kong, Tinbergen Institute for helpful discussions, comments, and suggestions. Part of the paper was written during my visit at the HKBU in early 2003. I am particularly grateful to Prof. Yiu-Kwan Fan and Prof. Kin Lam for their hospitality, and to the Wing Lung Bank for financial support.

[1] 1E.g., Markowitz (1952), Tobin (1958), Sharpe (1964), Lintner (1965), and Black (1972). See also Cochrane and Saa-Requejo (2000).

[2] 2E.g., Domar and Musgrave (1944), Bawa (1975, 1976), Bawa and Lindenberg (1977), Fishburn (1977). See also Bernardo and Ledoit (2000). Unlike Bawa and Lindenberg (1977) who consider a more general class of mean-lower partial moment models, however, the aim of the present paper is to enrich the CAPM with stronger predictions. Thus we focus on expected-loss in the class of downside risk measures.

1529-7373/2005

return. Both imply two-fund separation in that any two efficient funds span the entire efficient frontier. When the risk-free asset exists, the portfolio problem boils down to finding a common risky portfolio that has the highest Sharpe ratio (risk premium over standard deviation) or the highest gain-loss ratio (expected gain over expected loss). The two approaches differ only in their "risk" measures. Since standard deviation and expected loss do not always rank risks in the same way, in general the mean-variance (MV) and the gain-loss (GL) approaches represent two different paradigms and they may yield different portfolio advice as well as asset pricing relations.

A quite common assumption made in portfolio theory as well as in the empirical asset pricing literature, however, is that the asset returns satisfy the multivariate normal distributions. As shown by Bawa and Lindenberg (1977), this assumption implies that the GL (or equivalently, mean-loss) criterion yields the same Sharpe-Lintner-Black capital asset pricing model (CAPM). Bernardo and Ledoit (2000) also show a one-to-one relation between the Sharpe ratio and the GL ratio under normally distributed and risk-neutral benchmark distributions. In other words, the multivariate normality assumption ensures that the optimal risky portfolio is both MV and GL efficient in having the highest Sharpe ratio and GL ratio among all assets and portfolios. Multivariate normality, on the other hand, is only a special case of the class of separating distributions delineated by Ross (1978), who shows that two-fund separation holds for all investor preferences[3] if and only if the asset returns satisfy a set of two-fund separability conditions (henceforth two-fund separability). One of the contributions of this paper is to show that two-fund separability also implies the existence of an optimal risky portfolio that is both MV and GL efficient.

Implications of MV and GL efficiency for asset-pricing relations have not been fully explored to date. This paper reports a number of new results that are of both theoretical and practical interest. One of these results is that a benchmark portfolio is MV and GL efficient if and only if a dichotomous asset pricing model (DAPM) holds. Denote by r_i the gross return on asset i and by r_0 the gross risk-free interest rate. We say that the DAPM holds if for all asset i their upper-market expected gains $\overline{x}_i$ ($= r_i - r_0$ if $r_m > r_0$ and $= 0$ otherwise) and lower-market expected losses $\underline{x}_i$ ($= r_0 - r_i$ if $r_m \leq r_0$ and $= 0$ otherwise) satisfy the following equations

[3]That is, for all expected utility preferences where the utility functions are monotone increasing and concave on the real line. An unverified conjecture is that Ross' findings might be extended to some nonexpected utility preferences as well. Although the MV and GL efficiency condition is motivated in this paper by way of no-approximate arbitrage, the condition can also be derived from MV and GL preferences of investors. The reader is referred to Zou (2003) for a more general class of non-linear preferences which includes the MV and GL preferences as special cases.

in relation to the benchmark portfolio m:

$$E(\overline{x}_i) = \beta_i^+ E(\overline{x}_m), \quad \beta_i^+ = \frac{E(\overline{x}_m \overline{x}_i}{E(\overline{x}_m^2)} \tag{1}$$

$$E(\underline{x}_i) = \beta_i^- E(\underline{x}_m), \quad \beta_i^- = \frac{E(\underline{x}_m \underline{x}_i}{E(\underline{x}_m^2)}, \tag{2}$$

$$\beta_i^+ = \beta_i^- = \beta_i^B = \frac{E(x_m x_i)}{E(x_m^2)}. \tag{3}$$

where $E(\cdot)$ denotes mathematical expectation. The β_i^+ and β_i^- will be called the asset's upper- and lower-market betas, and β_i^B the asset's "best beta" for its role in minimizing the potential pricing errors in a sense to be made clear. Equations in (1)-(2) represent two security market lines as depicted in Figure 1.

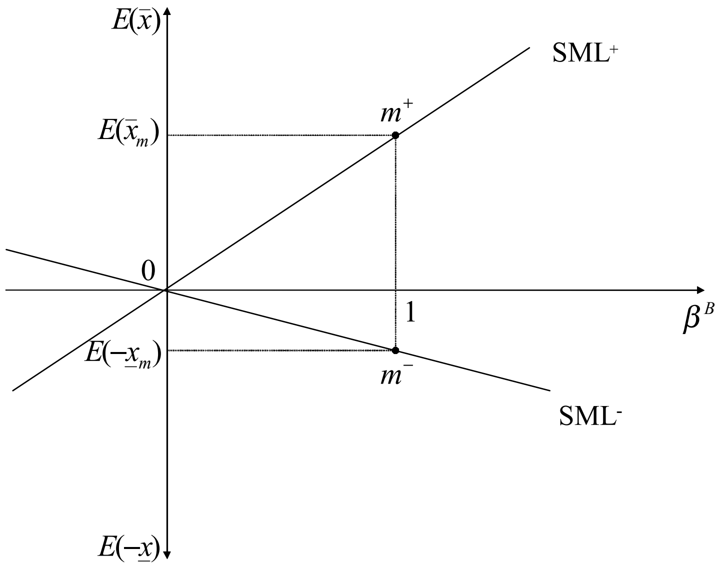

FIG. 1. The dichotomous asset pricing model (DAPM). The DAPM predicts two security market lines, the upper-market SML^+ and the lower-market SML^- where m is any benchmark portfolio that is both mean-variance and gain-loss efficient. The expected upper-market gain and lower-market loss on every asset are proportional to the benchmark portfolio's expected gain $E(x_m^+)$ $(= E(\overline{x}_m))$ and expected loss $E(x_m^-)$ $(= E(-\underline{x}_m))$ respectively. If the DAPM holds exactly, then the asset's "best beta" β^B is equal to the CAPM beta β.

It can be verified that if m is interpreted as the market portfolio, then (1)-(3) imply the CAPM prediction

$$E(x_i) = \beta_i E(x_m), \quad \beta_i = \frac{Cov(x_m, x_i)}{Var(x_m)} \tag{4}$$

where $Cov(\cdot;\cdot)$ and $Var(\cdot)$ denote covariance and variance. The reverse, however, is not necessarily true. Equations (1)-(3) separately predict the assets' upper- and lower-market expected return - beta relations and thus are strictly stronger than the CAPM prediction in (4). In relation to the traditional normal-distribution assumption made to justify the CAPM, the stronger predictions in (1)-(3) are also more significant in that they are derived from weaker distributional assumptions, e.g., two-fund separability.

Another new result is related to the cross-market (CM) derivative securities that are defined and analyzed in this paper. Although these derivative securities are not yet popularly traded in todays marketplace, they are easy to construct. Call c_i the uppermarket call and p_i the lower-market put of asset i whose payoffs are defined by

$$c_i : \begin{cases} r_i - r_0 & \text{if } r_m > r_0 \\ 0 & \text{if } r_m \leq r_0 \end{cases}, \quad p_i : \begin{cases} 0 & \text{if } r_m > r_0 \\ r_0 - r_i & \text{if } r_m \leq r_0 \end{cases}. \tag{5}$$

Thus the payoffs of c_i $(= \overline{x}_i)$ and p_i $(= \underline{x}_i)$ depend on the joint performance of m as well as the underlying asset i. In particular, c_m and p_m are the call and put options on one dollar of the benchmark portfolio with strike price equal to the gross risk-free rate r_0. In general, however, since for some assets the payoffs defined in (5) can be negative, c_i and p_i are not options in the traditional sense. Yet, to prevent arbitrage the prices of c_i and p_i must be equal for all i. This property follows from a general cross-asset put-call parity that will be derived in the next section.

From the DAPM and the cross-asset put-call parity it is easy to show that[4]

$$c_i = \beta_i^+ c_m, \quad p_i = \beta_i^- p_m \text{ for all } i \tag{6}$$

Note an important difference between (6) and (1)-(3): the prices c_i and p_i are directly observable by trading these derivative securities, while the expectations $E(\overline{x}_i)$ and $E(\underline{x}_i)$ are not. Consequently, if CM derivatives are tradable and hence priced, then akin to the implied volatility of the Black-Scholes (1973) option-pricing model, the assets' implied betas, β_i^+ and β_i^- , will be observable from the CM derivative prices according to (6). No-arbitrage will ensure that $c_i = p_i$, thus if the DAPM holds then for all

[4]Without ambiguity c_i and p_i denote both the CM derivatives as well as the prices of such derivatives.

asset i,[5]

$$\beta_i^+ = \beta_i^- = \beta_i = \beta_i^B = c_i/c_m = p_i/p_m \tag{7}$$

In light of the empirical difficulties in estimating the CAPM betas accurately (e.g., Fama and French, 1997; Pástor and Stambaugh, 1999), the theoretical relations established in (7) are significant. A potential market for the CM-type of derivative securities could be anticipated. For instance, such derivatives could be traded like futures contracts and be settled daily or marked to market. The implied betas from the observed CM derivatives' prices would then reveal the investors' aggregate beliefs about assets' betas and help reduce the beta uncertainty.

Indeed, the DAPM offers a theoretical foundation for separately investigating the average returns conditional on the market being up or down. Empirical tests of the DAPM predictions in (1)-(3) requires one to partition or dichotomize the space of assets' excess returns into the upper-market ($x_m > 0$) and lower-market ($x_m \leq 0$) subspaces. Such dichotomization could lead to new insights in the cross-sectional differences in average returns and help explain the empirical regularities that seem to violate the CAPM. The focus of the present paper is on the theoretical development of the DAPM, empirical applications of the model are investigated in a subsequent paper (Zou, 2004).

The next section introduces the cross-asset derivative securities and proves a general put-call parity relation for these securities. Section 3 is devoted to the derivation and analysis of the DAPM. Section 4 summarizes the main findings of the paper. The Appendix contains the proofs of the lemmas and theorems.

2. THE CROSS-ASSET DERIVATIVE SECURITIES

Let an investment opportunity set Λ_t be given that includes $n(\geq 2)$ primary assets $i = 1, \ldots, n$ with prices $S_{i,t}$ at time t. Assume that the capital market is perfectly competitive, there is no tax and transaction costs, and that the market allows free portfolio formation in that if $p, q \in \Lambda_t$ then the portfolio $\theta p + (1 - \theta)q \in \Lambda_t$ for all $\theta \in \mathbb{R}$. Over any period

[5] "Upside" and "downside" betas have been defined differently in various empirical papers. For instance, Ang and Chen (2002) define upside (downside) beta as the ratio of covariance over variance conditional on that both the asset and the benchmark portfolio make higher (lower) returns than their means. They reject the hypothesis that their downside beta, and cannot reject that their upside beta, is equal to what would be implied by jointly normal distributions. An important feature of the upper- and lower-market betas in this paper is that they are theoretically derived, rather than empirically motivated. Furthermore, the market here is partitioned according to the benchmark returns only, and with reference to the risk-free rate. No assumption is made throughout the paper that asset return distributions are jointly normal.

$[t, t+\tau], \tau > 0$, let $r_{i,t+\tau}$ denote the gross return (including dividends) on assets i. Let $r_{m,t+\tau}$ denote the gross return on some benchmark portfolio $m \in \Lambda_t$ (e.g., an equal- or value-weighted average of all primary assets in Λ_t used as a market proxy). Let $r_{0,t+\tau}$ denote the gross risk-free interest rate (if exists) or the return on a "zero-beta" portfolio whose return is stochastically uncorrelated with $r_{m,t+\tau}$. Assume that $E_t(r_{m,t+\tau} - r_{0,t+\tau}) > 0$ where $E_t(\cdot)$ is the expectation operator conditional on the information at time t. To ease exposition, assume that all dividends are reinvested in the same assets so that $r_{i,t+\tau} = S_{i,t+\tau}/S_{i,t}$.

Let $X_{t+\tau}$ denote the set of (random) payoffs of all zero-price, or self-financed, portfolios that can be formed with assets in Λ_t. For example, $r_{i,t+\tau} - r_{j,t+\tau} \in X_{t+\tau}$ is the payoff of a portfolio formed by a one-dollar long position in asset i and one-dollar short position in asset j. In particular, define $x_{i,t+\tau} = r_{i,t+\tau} - r_{0,t+\tau} \in X_{t+\tau}$ to be the excess return, and $E_t(x_{i,t+\tau})$ the risk premium, on asset i. As a technical condition, assume that the returns $r_{i,t+\tau}, i = 1, \ldots, n$, are continuously distributed on a subset of $\mathbb{R}^n$ (even though their observed returns are discrete). This helps avoid possible corner solutions that may not be characterized by the first-order condition in portfolio optimization. Define $x^+ = \max(x, 0)$ and $x^- = \min(x, 0)$. Throughout the paper, assume that the first two moments of all returns exist and that the no-arbitrage principle holds in that for all t, τ, there exists no zero-price payoff $x_{t+\tau} \in X_{t+\tau}$ such that $E_t(x^+_{t+\tau}) > 0$ and $x^-_{t+\tau} \equiv 0$.

For all $x_{t+\tau} \in X_{t+\tau}$ define

$$\overline{x}_{t+\tau} = \begin{cases} x_{t+\tau} & \text{if } x_{m,t+\tau} > 0 \\ 0 & \text{if } x_{m,t+\tau} \leq 0 \end{cases} ; \quad \underline{x}_{t+\tau} = \begin{cases} 0 & \text{if } x_{m,t+\tau} > 0 \\ -x_{t+\tau} & \text{if } x_{m,t+\tau} \leq 0 \end{cases} . \tag{8}$$

Note that $x_{t+\tau} = \overline{x}_{t+\tau} - \underline{x}_{t+\tau}$. Call $\overline{x}_{t+\tau}$ the upper-market gain and $\underline{x}_{t+\tau}$ the lower-market loss (neither need be positive) of $x_{t+\tau}$ with respect to the given benchmark portfolio m. For notational convenience, I often drop the subscripts t and/or $t + \tau$ unless they are needed for clarity.

Consider now a general class of cross-asset call C_{ij} and cross-asset put P_{ij} whose payoffs are defined by the asset prices $S_{i,t+\tau}, S_{j,t+\tau}$, as well as some "threshold prices" K_i, K_j, as follows:[6]

$$C_{ij} : \begin{cases} S_{i,t+\tau} - K_i & \text{if } S_{j,t+\tau} > K_j \\ 0 & \text{if } S_{j,t+\tau} \leq K_j \end{cases} , \quad P_{ij} : \begin{cases} 0 & \text{if } S_{j,t+\tau} > K_j \\ K_i - S_{i,t+\tau} & \text{if } S_{j,t+\tau} \leq K_j \end{cases} . \tag{9}$$

When $i = j$ and $K_i = K_j$, C_{ij} and P_{ij} reduce to the usual call and put options. For a more general example, suppose i is IBM and j is Microsoft.

[6]These cross-asset derivative contracts resemble forward contracts more than options, because their payoffs can be negative and the holder of such contracts has not only the right to receive but also the obligation to pay the contingent cash difference.

Then the cross-asset call C_{ij} gives the security holder the right and obligation to receive, at time $t+\tau$, the difference (which may be negative) between IBM's share price $S_{i,t+\tau}$ and the threshold price K_i if Microsoft's share price $S_{j,t+\tau} > K_j$; otherwise the security holder receives nothing. The no-arbitrage principle then implies a cross-asset put-call parity relation as stated in the following lemma.

LEMMA 1. *Assume that the risk-free asset exists and* $C_{ij}, P_{ij} \in \Lambda_t$. *Then no-arbitrage implies*

$$C_{ij} + \frac{K_i}{r_{0,t+\tau}} = P_{ij} + S_i \tag{10}$$

Obviously, (9) embeds the usual put-call parity as a special case. Unaware of any earlier work on this simple yet important cross-asset put-call parity, I offer a proof here by no-arbitrage. Consider the cash flows of a portfolio formed at time t with a long call C_{ij}, a short put P_{ij}, a short stock S_i and cash deposit $K_i/r_{0,t+\tau}$:

Position	time t	time $t+\tau$	
		$(S_{j,t+\tau} > K_j)$	$(S_{j,t+\tau} \leq K_j)$
Long call	$-C_{ij}$	$S_{i,t+\tau} - K_i$	0
Short put	P_{ij}	0	$-(K_i - S_{i,t+\tau})$
Short stock	S_i	$-S_{i,t+\tau}$	$-S_{i,t+\tau}$
Cash deposit	$-K_i/r_{0,t+\tau}$	K_i	K_i
Total Cash Flow	$P_{ij} + S_i - C_{ij} - K_i/r_{0,t+\tau}$	0	0

Since the payooff this portfolio at time $t+\tau$ is always zero, it must have a price of zero at time t to avoid arbitrage. Consequently, the parity relation in (9) must hold. It is easily seen that as a special case where $j = m, S_i = S_j = 1$, and $K_i = K_j = r_{0,t+\tau}$, (9) reduces to

$$c_i = p_i \quad \text{for all } i \tag{11}$$

where c_i and p_i are the cross-market call and put defined in (5). Since $c_m = p_m > 0$, equation (10) implies that for all i

$$\frac{c_i}{c_m} = \frac{p_i}{p_m} = \varphi_i \quad \text{for some } \varphi_i$$

Consequently, no-arbitrage implies that for all $c_{i,t}$ and $p_{i,t} \in \Lambda_t$, and all t, there exists a $\varphi_{i,t} \in \mathbb{R}$ such that

$$c_{i,t} = \varphi_{i,t} c_{m,t} \text{ and } p_{i,t} = \varphi_{i,t} p_{m,t} \tag{12}$$

where the subscript t is added back for clarity.

I do not assume that all cross-asset derivative securities are tradable. Instead, I only assume that Λ_t includes the primary assets and their cross-market calls and/or puts defined in (5), as well as all portfolios that can be formed by these securities. From the parity relation in (9) it is clear that as long as $c_{i,t}$ is tradable, so is the payoff of $p_{i,t}$ C and vice versa. The cross-market derivative securities and their price relations in (11) will play an important role in motivating the derivation of the DAPM.

3. THE DICHOTOMOUS ASSET PRICING MODEL

Following the modern approach to asset pricing I begin with a general pricing relation whereby the prices of all assets $S_t \in \Lambda_t$ at time t are expressed as the expectation of their stochastically discounted time $t + \tau$ prices, $S_{t+\tau}$ (including dividends). The stochastic discount factor (SDF) $\delta_{t+\tau}$, common to all assets, is strictly greater than zero if and only if Λ_t permits no-arbitrage:

$$S_t = E_t(\delta_{t+\tau} S_{t+\tau}) \tag{13}$$

Recall that $x_{t+\tau}$ denotes the payoff of a zero-price portfolio ($S_t = 0$), thus (12) is equivalent to

$$\begin{aligned} 1 &= E_t(\delta_{t+\tau} r_{t+\tau}) \text{ for all } S_t \neq 0 \\ 0 &= E_t(\delta_{t+\tau} x_{t+\tau}) \text{ for all } S_t = 0 \end{aligned}$$

When the market is incomplete, as is assumed in this paper, it is well known that the SDF approach to asset pricing is generally valid but the set of admissible SDFs that satisfy (12) is typically large (see, e.g., Hansen and Jagannathan, 1997). Among the admissible SDFs, we naturally prefer the simple ones provided that they make good economic sense.

3.1. Mean-Variance Efficiency

By MV efficiency I broadly mean that there is a uniform upper (and lower) bound for the Sharpe ratio $Sh(x_\pi)$ $(= E(x_\pi)/\sqrt{Var(x_\pi)})$ of all portfolios $\pi \in \Lambda$. An extremely high Sharpe ratio indicates a "good deal" in the terminology of Cochrane and Saa-Requejo (2000). Provided that there are sufficiently many investors who prefer such good deals, it is conceivable that prices would adjust whenever a good deal is available and that the dynamic price processes would maintain a MV efficient investment opportunity set.

There is an important difference between MV efficiency and the "no good deal" approach of Cochrane and Saa-Requejo (2000), however. Here, the Sharpe ratio is assumed to be bounded endogenously so that we can talk about an unconstrained MV ecient portfolio. In contrast, Cochrane and Saa-Requejo study the effect of bounds on the Sharpe ratio (or on the second moment of the SDF) that are imposed exogenously by the modeler. By comparison, MV efficiency implies a unique SDF with respect to the benchmark portfolio while the no-good-deal bounds produce a class of admissible SDFs.

More generally, the Sharpe ratio can be defined for any zero-price payoffs $x \in X$. I introduce a "modified Sharpe ratio" and show its relation to the traditional MV efficiency.

Definition 3.1. Λ_t is MV efficient (over period $[t, t+\tau]$) iff there exists $m \in \Lambda_t$ such that $E_t(x_{m,t+\tau}) = E_t(r_{m,t+\tau} - r_{0,t+\tau}) > 0$ and that for all $x_{t+\tau} \in X_{t+\tau}$.

$$\eta_t^2(x_{t+\tau}) \leq \eta_t^2(x_{m,t+\tau}) < 1, \text{ where } \eta_t(x) = \frac{E_t(x)}{\sqrt{E_t(x^2)}}. \tag{14}$$

The portfolio m satisfying (13) is called a mean-variance efficient portfolio (over period $[t, t+\tau]$), and $\eta_t(x_{t+\tau})$ is called the time-t "modified Sharpe ratio" of the zero-price payoff $x_{t+\tau}$.

In the remainder of the paper I assume that the investment opportunity set Λ_t is MV efficient for all t. Although the modified Sharpe ratio differs from the Sharpe ratio in penalizing deviations from zero rather than from the mean, the squares of the two ratios give the same ranking for all zero-price payoffs. This follows from the fact that η and Sh are related by

$$\eta^2(x) = \frac{[E(x)]^2}{[E(x)]^2 + Var(x)} = \frac{[E(x)]^2/Var(x)}{[E(x)]^2/Var(x) + 1} = \frac{[Sh(x)]^2}{[Sh(x)]^2 + 1} \tag{15}$$

so that $\eta_2(x) < 1$ if and only if $[Sh(x)]^2 < \infty$, and that for any two zero-price payoffs x_1 and x_2

$$\eta^2(x_1) \geq \eta_2^2(x_2) \text{ if and only if } [Sh(x_1)]^2 \geq [Sh(x_2)]^2.$$

An advantage of using the modified Sharpe ratio, however, is to derive a modified CAPM that relates the MV efficiency with the GL efficiency in a more transparent manner.

Theorem 1. *Over period $[t, t+\tau]$, portfolio $m \in \Lambda_t$ is MV efficient iff for all $x_{t+\tau} \in X_{t+\tau}$ (including the special case $x_{i,t+\tau} = r_{i,t+\tau} - r_{0,t+\tau}$)*

$$E_t(x_{t+\tau}) = \beta_t^B(x_{t+\tau}) E_t(x_{m,t+\tau}) \tag{16}$$

where $\beta_t^B(x_{t+\tau})$ *is called the "best beta" of* $x_{t+\tau}$ *given by*

$$\beta_t^B(x_{t+\tau}) = \frac{E_t(x_{t+\tau}x_{m,t+\tau})}{E_t(x_{m,t+\tau}^2)} \tag{17}$$

The modified CAPM (15)-(16) is equivalent to the CAPM in (4) if either model holds exactly. To see this, note that

$$\begin{aligned}
\beta_i^B E(x_m) &= \frac{E(x_m x_i)}{E(x_m^2)} E(x_m) \\
&= \frac{Var(x_m)}{E(x_m^2)} \times \frac{[E(x_m)]^2 E(x_i) + Cov(x_m, x_i)E(x_m)}{Var(x_m)} \\
&= \frac{[E(x_m)]^2}{E(x_m^2)} E(x_i) + \frac{Var(x_m)}{E(x_m^2)} \beta_i E(x_m) \\
&= \eta^2 E(x_i) + (1-\eta^2)[\beta_i E(x_m)].
\end{aligned}$$

Therefore

$$E(x_i) = \beta_i^B E(x_m) \iff E(x_i) = \beta_i E(x_m)$$

The reason for β_i^B to be called the "best beta", however, comes from the fact that it minimizes the potential pricing errors of the model. More specifically, note that there is no loss of generality to write

$$x_i = b_i x_m + \epsilon_i \text{ for all } i$$

Define the pricing error by

$$E(\epsilon_i^2) = E(x_i - b_i x_m)^2$$

Then minimizing the pricing error yields

$$\min_{b_i} E(x_i - b_i x_m)^2 \Rightarrow b_i = \beta_i^B$$

The next theorem shows an equivalent expression of the modified CAPM in terms of the SDF.

THEOREM 2. *Over period* $[t, t+\tau]$, *portfolio* $m \in \Lambda_t$ *is MV efficient iff there exists an SDF,* $\delta_{t+\tau}$, *satisfying*

$$\delta_{t+\tau} = A_{m,t}(1 - \lambda_{m,t} x_{m,t+\tau}) \tag{18}$$

where $\lambda_{m,t}$ *and* $A_{m,t}$ *are time-t constants given by*

$$\lambda_{m,t} = \frac{E(x_{m,t+\tau})}{E(x_{m,t+\tau}^2)}, \quad A_{m,t} = \frac{1}{r_{0,t+\tau}(1-\eta_{m,t}^2)} \tag{19}$$

In particular, the cross-market calls and puts are priced under MV efficiency by

$$c_i = A_m[E(\overline{x}_i) - \lambda_m E(\overline{x}_i x_m)] = A_m[E(\overline{x}_i) - \lambda_m E(\overline{x}_i \overline{x}_m)] \quad (20)$$

$$p_i = A_m[E(\underline{x}_i) - \lambda_m E(\underline{x}_i x_m)] = A_m[E(\underline{x}_i) + \lambda_m E(\underline{x}_i \underline{x}_m)] \quad (21)$$

A limitation of the SDF in (17)-(18), as Dybvig and Ingersoll (1982) first recognized, is that it may be negative for large $x_{m,t+\tau}$ and therefore may permit arbitrage under complete markets. On the other hand, Ingersoll (1987, p.99) shows that if the asset returns are jointly normal then there exists no arbitrage opportunity under the CAPM (hence under the modified CAPM). The MV pricing model in (17)-(18) can also be defended by restricting its scope to the pricing of primary assets only (Dybvig and Ingersoll, 1982).

3.2. Gain-Loss Efficiency

The definition of GL efficiency is similar to MV efficiency except that the expected loss is penalized instead of the second moment. The same symbol m is used for notational convenience; it does not suggest that m is MV efficient in this subsection.

DEFINITION 3.2. Λ_t is GL efficient (over period $[t, t+\tau]$) iff there exists $m \in \Lambda_t$ such that $E_t(x_{m,t+\tau}) = E_t(r_{m,t+\tau} - r_{0,t+\tau}) > 0$ and that for all $x_{t+\tau} \in X_{t+\tau}$, either $E_t(x^+_{t+\tau}) = E_t(x^-_{t+\tau}) = 0$ or

$$Z_t(x_{t+\tau}) \leq Z_t(x_{m,t+\tau}) < \infty, \text{ where } Z_t(x) = \frac{E_t(x^+)}{E_t(-x^-)} \quad (22)$$

The portfolio m satisfying (21) is called a gain-loss efficient portfolio (over period $[t, t+\tau]$), and $Z_t(x_{t+\tau})$ is called the time-t "gain-loss ratio" of the zero-price payoff $x_{t+\tau}$.

In the remainder of the paper I assume that the investment opportunity set Λ_t is also GL efficient for all t. It is easily seen that if Λ is gain-loss efficient then the investment opportunity set permits no arbitrage. Should there exist a zero-price portfolio with $x \in X$ such that $E(x^+) > 0$ and $x^- \equiv 0$, then $Z(x) = \infty$. But since the reverse is not necessarily true (e.g., $E(x^+) > 0$ and $E(x^-) = 0$ imply $Z(x) = \infty$ but do not imply $x^- \equiv 0$), GL efficiency is a somewhat stronger condition than no-arbitrage. Using the term of Bernardo and Ledoit (2000), we may say that violation of GL efficiency means the presence of some sort of "approximate arbitrage" in that there are investment strategies that could generate arbitrarily high expected gains with arbitrarily low expected losses. Therefore, provided that there are sufficiently many investors who are willing to allocate (even just a fraction of) their capital to higher gain-loss ratio assets it is reasonable that the investment opportunity set would be GL efficient.

The gain-loss ratio Z is related to the "mean-loss" ratio $E(x)/E(-x^-)$ (e.g., Bawa and Lindenberg, 1977) by $Z(x) - 1 = E(x)/E(-x^-)$. Thus the two ratios rank zero-price payoffs in exactly the same way. A small advantage of the GL ratio, although inessential, is that its sign is always positive whereas the mean-loss ratio may change signs. Note also a symmetric property of the GL ratio: $Z(x) = 1/Z(-x)$, which implies that the set of zero price payoffs are symmetric around the 45° line passing through the origin in a gain-loss diagram. Since the gain-loss and mean-loss ratios are one-to-one, however, in discussing the implications of the model I shall refer to the one that is more convenient.

The notion of GL efficiency is more related to Bawa and Lindenberg's (1977), while diering from Bernardo and Ledoit's (2000) gain-loss approach in two respects. First, Bernardo and Ledoit define gain and loss by taking expectation under the benchmark risk-adjusted probability measures; whereas these are defined here in (21) by way of the true (either objective or subjective) probability measures. Second, Bernardo and Ledoit do not assume GL efficiency in the sense of Definition 2; instead, similar to Cochrane and Saa-Requejo (2000) they study effects of exogenously imposed bounds on the GL ratios of zero-price payoffs under the benchmark risk-adjusted probability measures. These bounds are shown to restrict the deviations of admissible SDFs from the benchmark SDF in a similar fashion. Here I focus on the unique SDF that characterizes a GL efficient benchmark portfolio. With such a benchmark, the risk-adjusted GL ratio of all zero-price payoffs must be equal to one. The following theorem provides an enriched version of Bawa and Lindenberg's (1977) pricing model under mean-loss efficiency.

THEOREM 3. *Over period $[t, t+\tau]$ portfolio $m \in \Lambda_t$ is GL efficient iff for all $x_{t+\tau} \in X_{t+\tau}$*

$$E_t(\overline{x}_{t+\tau}) \;=\; \varphi_t(x_{t+\tau})E_t(\overline{x}_{m,t+\tau}) \qquad (23)$$

$$E_t(\underline{x}_{t+\tau}) = \varphi_t(x_{t+\tau})E_t(\underline{x}_{m,t+\tau}) \qquad (24)$$

for some $\varphi_t(x_{t+\tau}) \in \mathbb{R}$. In particular, $\varphi_t(x_{i,t+\tau}) = \varphi_{i,t}$ for all $i \in \Lambda_t$, as given in (11). An equivalent expression of (22)-(23) is that either $E_t(\overline{x}_{t+\tau}) = E_t(\underline{x}_{t+\tau}) = 0$ or

$$Z_t^R(x_{t+\tau}) = \frac{E_t(\overline{x}_{t+\tau})}{E_t(\underline{x}_{t+\tau})} = \frac{E_t(\bar{(x)}_{m,t+\tau})}{E_t(\underline{x}_{m,t+\tau})} = Z_{m,t} \qquad (25)$$

where $Z_t^R(x_{t+\tau})$ is called the "relative gain-loss ratio" of $x_{t+\tau}$ (i.e., relative to the benchmark m).

The next theorem presents the SDF under GL efficiency.[7]

THEOREM 4. *Over period $[t, t+\tau]$ portfolio $m \in \Lambda_t$ is GL efficient iff there exists an SDF satisfying*

$$\delta_{t+\tau} = \delta_{t+\tau}^+ + \delta_{t+\tau}^- \tag{26}$$

where

$$\delta_{t+\tau}^+ = \begin{cases} B_{m,t} & \text{if } x_{m,t+\tau} > 0 \\ 0 & \text{if } x_{m,t+\tau} \leq 0 \end{cases}, \quad \delta_{t+\tau}^- = \begin{cases} 0 & \text{if } x_{m,t+\tau} > 0 \\ B_{m,t}Z_{m,t} & \text{if } x_{m,t+\tau} \leq 0 \end{cases} \tag{27}$$

and $B_{m,t}$ is a time-t constant given by

$$B_{m,t} = \frac{1}{r_{0,t+\tau}[Pr(x_{m,t+\tau} > 0) + Z_{m,t}Pr(x_{m,t+\tau} \leq 0)]} \tag{28}$$

Comparing the pricing model (25)-(27) with the general formula in (12), we find that under GL efficiency the SDF $\delta_{t+\tau}$ can be dichotomized into two "conditionally constant discount factors" an upper-market discount factor $\delta_{t+\tau}^+$ and a lower-market discount factor $\delta_{t+\tau}^-$. These two discount factors have their own meaning as well: $E_t(\delta_{t+\tau}^+)$ is the price of a "binary call option" that pays 1 dollar if $x_{m,t+\tau} > 0$ and 0 otherwise, and $E_t(\delta_{t+\tau}^-)$ is the price of a "binary put option" that pays 1 dollar if $x_{m,t+\tau} \leq 0$ and 0 otherwise. In the derivation of the upper- and lower-market discount factors no assumption is made that such binary options are tradable, however. For the record, the time-t price of a zerocoupon risk-free bond that pays 1 dollar at time $t+\tau$ is the sum of the binary call and put, and is correctly priced at

$$E_t(\delta_{t+\tau}^+) + E_t(\delta_{t+\tau}^-) = \frac{1}{r_{0,t+\tau}} \tag{29}$$

Under GL efficiency the cross-market options are priced by

$$c_{i,t} = E_t(\delta_{t+\tau}^+ \overline{x}_{i,t+\tau}) = B_{m,t}E_t(\overline{x}_{i,t+\tau}) \tag{30}$$

$$p_{i,t} = E_t(\delta_{t+\tau}^- \underline{x}_{i,t+\tau}) = B_{m,t}Z_{m,t}E_t(\underline{x}_{i,t+\tau}). \tag{31}$$

[7]The SDF under GL efficiency has a number of attractive features that are explored in Zou (2000).

3.3. MV and GL efficiency

A thrust of the CAPM is its prediction that the market portfolio is MV efficient. It is well-known that this prediction follows either from the assumption that the investors have MV preferences, or from the assumption that asset returns are normally distributed. The normal distributions of asset returns would guarantee that the market portfolio is also GL efficient. In general, however, asset return distributions need not be normal (e.g., Longin and Solnik, 2001, Ang and Chen, 2002) and there is no guarantee that a MV efficient portfolio will be also GL efficient. We wish to know under what conditions this is the case.

Ross (1978) shows that the multivariate normal distributions belong to a more general class of separating distributions. Instead of restricting the investor preferences, separating distributions ensure portfolio separation for all risk averse investors who maximize (von Neumann-Morgenstern) expected utility. Ross' two-fund separation turns out to imply that there exists a portfolio (e.g., the market portfolio in equilibrium) that is both MV and GL efficient, as depicted in Figure 2.

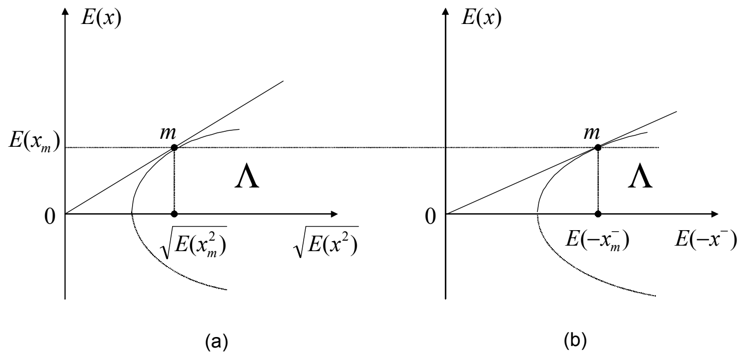

FIG. 2. Two-fund separation implies that the market portfolio is both mean-variance and gain-loss efficient. Measuring risk (a) by standard deviation (equivalently, by $\sqrt{E(x^2)}$) or (b) by expected loss $E(-x^-)$ yields the same optimal risky portfolio m.

DEFINITION 3.3. Over period $[t, t+\tau]$ the primary asset returns $r_{i,t+\tau}, i = 1, \ldots, n$, exhibit two-fund separability under Λ_t iff there exists two mutual funds $m_1, m_2 \in \Lambda_t$ such that for any portfolio $\pi \in \Lambda_t$, any invested capital w, and any increasing and concave utility function U, there exists a portfolio $\theta m_1 + (1-\theta)m_2$ with

$$E_t U(w\theta r_{m_1,t+\tau} + (1 - theta)wr_{m_2,t+\tau}) \geq E_t U(wr_{\pi,t+\tau})$$

(if the expectations exist).

The above definition is the weak form of two-fund separability adapted from Ross (1978).

LEMMA 2. *The primary asset returns $r_{i,t+\tau}, i = 1, \ldots, n$, exhibit two-fund separability under Λ_t (over period $[t, t+\tau]$) if and only if there exist random variables $y_{1,t+\tau}$ and $y_{2,t+\tau}$, and two mutual funds $m_1, m_2 \in \Lambda_t$ such that for all i*

$$\begin{aligned} r_{i,t+\tau} &= y_{1,t+\tau} + b_{i,t} y_{2,t+\tau} + \varepsilon_{i,t+\tau} \\ E_t(r_{i,t+\tau}) &= E_t(y_{1,t+\tau}) + b_{i,t} E_t(y_{2,t+\tau}) \\ 0 &= E(\varepsilon_{i,t+\tau} | y_{1,t+\tau} + \xi y_{2,t+\tau}) a.e. \textit{for all } \xi \in A \\ 0 &\equiv \varepsilon_{m_1,t+\tau} \equiv \varepsilon_{m_2,t+\tau} \end{aligned}$$

where A is some properly define interval (for a proof see Ross, 1978 Theorem 2).

DEFINITION 3.4. Assume that two-fund separability holds for $r_{i,t+\tau}, i = 1, \ldots, n$, under Λ_t (over period $[t, t+\tau]$). Then a portfolio $m \in \Lambda_t$ is "well diversified" if and only if $\varepsilon_{m,t+\tau} \equiv 0$, i.e., m has no specific or ε risk.

The next lemma shows an important implication of two-fund separability.

LEMMA 3. *Assume that the primary asset returns $r_{i,t+\tau}, i = 1, \ldots, n$, exhibit two-fund separability under Λ_t (over period $[t, t+\tau]$), then any well-diversified portfolio $m \in \Lambda_t$ with $E_t(x_{m,t+\tau}) > 0$ is both MV and GL efficient.*

The next theorem shows the main result of the paper.

THEOREM 5. *Portfolio $m \in \Lambda_t$ satisfying $E_t(x_{m,t+\tau}) > 0$ is both MV and GL efficient over period $[t, t+\tau]$ iff the dichotomous asset pricing model (DAPM) holds with respect to m. That is, for all asset $i \in \Lambda_t$,*

$$E_t(\overline{x}_{i,t+\tau}) = \beta^+_{i,t} E_t(\overline{x}_{m,t+\tau}) \tag{32}$$

$$E_t(\underline{x}_{i,t+\tau}) = \beta^-_{i,t} E_t(\underline{x}_{m,t+\tau}) \tag{33}$$

where

$$\beta^+_{i,t} = \frac{E_t(\overline{x}_{m,t+\tau}\overline{x}_{i,t+\tau})}{E_t(\overline{x}^2_{m,t+\tau})}, \beta^-_{i,t} = \frac{E_t(\underline{x}_{m,t+\tau}\underline{x}_{i,t+\tau})}{E_t(\underline{x}^2_{m,t+\tau})}, \textit{and } \beta^+_{i,t} = \beta^-_{i,t} = \beta^B_{i,t} = \varphi_{i,t} \tag{34}$$

By Lemma 3, the DAPM holds whenever the primary asset returns satisfy twofund separability and the benchmark portfolio is well diversified. It is worth remarking, however, that two-fund separability is only a sufficient, and not necessary, condition for the DAPM. Ross' two-fund separability is of particular interest, however, because it has least restrictions on investor preferences.

Remarks:

a) The standard textbook interpretation of the CAPM is "high (systematic) risk high (expected) return", where systematic risk is measured by beta. The DAPM could be interpreted as saying "high beta high (expected) gain when the market goes up and high (expected) loss when the market goes down." In this interpretation, however, beta takes on a somewhat different meaning. It seems more appropriate in the DAPM to interpret the upper-market beta as a measure of the asset's upper-market potential, and the lowermarket beta the asset's lower-market risk. The model predicts that these two betas are equal if and only if a MV and GL efficient benchmark portfolio is correctly specified.

b) It is important to note that Theorem 5 is pure deductive theory. Under certain assumptions, the theorem shows the equivalence between two formal statements: (i) a benchmark portfolio m is both MV and GL ecient within a given opportunity set Λ and (ii) the prices and excess returns of all assets in Λ are related to m according to the DAPM. Unlike the CAPM which predicts that the unobservable market portfolio is MV efficient (e.g., Roll's critique, 1977), a test of the DAPM can be confined to any given opportunity set. The test would mainly concern the hypothesis that the specified benchmark portfolio is both MV and GL efficient within the opportunity set.

c) Since in general investors prefer greater upper-market potential and smaller lowermarket risk, the DAPM could be applied to performance evaluation with more fine-tuned measures. For instance, the upper-to-lower-market beta ratio (β^+/β^-) could be used as a (or an additional) measure of market-timing ability. The benchmark case where all assets are fairly priced according to the DAPM is that this ratio equals 1 for all assets and portfolios. If a fund manager persistently achieves a ratio β^+/β^- that is greater than one, then the fund is likely to be attractive. Other measures such as the gain-loss ratio (Z) and the relative gain-loss ratio (Z^R) could be useful as well.

4. SUMMARY AND CONCLUSION

This paper derives a number of new and potentially useful results that are summarized as follows.

• If the asset-return distributions satisfy two-fund separability, which includes the multivariate normality as a special case, then any well-diversified portfolio is both mean-variance (MV) and gain-loss (GL) efficient.

• A general cross-asset put-call parity is derived. In particular, it is shown that the cross-market (CM) call and put derivative securities (interpreting the benchmark portfolio as a market proxy) must have the same price in order to prevent arbitrage. The price of the CM call or put is directly related to the asset's beta and the price of the (at the forward-money) call or put option on the benchmark portfolio.

• If the cross-market calls or puts belong to the investment opportunity set, then the necessary and sufficient condition for a benchmark portfolio to be MV and GL efficient is that the dichotomous asset pricing model (DAPM) holds. According to the DAPM, the expected excess return - beta relations hold separately in the upperand lower-market regimes and can be characterized by two security market lines as depicted in Figure 1. This observation enriches the single security market line prediction of the CAPM.

• Provided that the DAPM holds, the price of a CM call (or put) is the product of the underlying asset's beta and the call (or put) option on one-dollar of the benchmark portfolio with strike price equal to the gross risk-free rate. As a result, akin to the implied volatility of the Black-Scholes (1973) option-pricing model, trading of CM derivatives can reveal the assets' implied betas and help reduce the beta uncertainty.

The extent to which the dichotomous approach may help improve our current understanding about the cross-section of average returns is an important issue that remains to be investigated. Since the theoretical model (DAPM) derives from economically sensible and meaningful assumptions (e.g., MV and GL efficiency or no-approximate arbitrage), and since the DAPM is also an implication of two-fund separability which is frequently assumed to motivate the CAPM, it can be concluded that any future empirical research adopting the dichotomous approach presented in this paper is theoretically justified.

APPENDIX

Proof of Theorem 1:

(The necessity part of the proof is standard and the sufficiency part is somewhat new.) Let any benchmark portfolio m be given with $E(x_m) > 0$ and consider any zero-price payoff $x \in X$. Let π denote a portfolio that is formed by one dollar of m and θ amount of exposure in x in that the return and excess return on π are

$$r_\pi = r_m + \theta x \text{ and } x_\pi = x_m + \theta x$$

Maximizing the square of the modified Sharpe ratio $\eta_\pi^2 = [E(x_\pi)]^2/E(x_\pi^2)$ implies a first-order condition:

$$\max_\theta \eta_\pi^2 \tag{A.1}$$

$$\Rightarrow \frac{\partial \eta_\pi^2}{\partial \theta} = \frac{2E(x_\pi)}{[E(x_\pi^2)]^2}[E(x)E(x_\pi^2) - E(x_\pi)E(x_\pi x)] = 0. \tag{A.2}$$

Define

$$\Psi(\theta) = E(x)E(x_\pi^2) - E(x_\pi)E(x_\pi x)$$

For x_π to have the highest modified Sharpe ratio we must have $E(x_\pi) > 0$, thus

$$\frac{\partial \eta_\pi^2}{\partial \theta} = 0 \iff \Psi(\theta) = 0$$

We also need to consider the second-order condition

$$\frac{\partial^2 \eta_\pi^2}{\partial \theta^2} = \frac{\partial}{\partial \theta}\left[\frac{2E(x_\pi)}{[E(x_\pi^2)]^2}\right]\Psi(\theta) + \frac{2E(x_\pi)}{[E(x_\pi^2)]^2}\Psi'(\theta) \le 0 \tag{A.3}$$

where

$$\begin{aligned} \Psi'(\theta) &= 2E(x)E(x_\pi x) - E(x)E(x_\pi x) - E(x_\pi)E(x^2) \\ &= E(x)E(x_\pi x) - E(x_\pi)E(x^2) \end{aligned} \tag{A.4}$$

With the above preparation we are now ready to prove the theorem.

Necessity: Assume that mis MV efficient so that x_m has the highest modified Sharpe ratio η_m. Then $\theta = 0$ is a solution for the program in (A.1). It follows that for all $x \in X$

$$\begin{aligned} \Psi(0) &= E(x)E(x_m^2) - E(x_m)E(x_m x) = 0 \\ \iff & E(x) = \frac{E(xx_m)}{E(x_m^2)}E(x_m) = \beta^B E(x_m) \end{aligned} \tag{A.5}$$

Sufficiency: Assume that m satisfies the first order condition (A.5) for all $x \in X$. Note from (A.3) that $\partial^2\eta_\pi^2/\partial\theta^2$ and Ψ' have the same sign when $\Psi(0) = 0$. Therefor substituting $\theta = 0$ in (A.4) and using (A.5) we obtain

$$\begin{aligned} \Psi'(0) &= E(x)E(x_m x) - E(x_m)E(x^2) \\ &= \left[\frac{[E(x)]^2}{E(x^2)} - \frac{[E(x_m)]^2}{E(x_m^2)}\right]\frac{E(x^2)}{E(x_m)} \le 0 \iff \frac{[E(x)]^2}{E(x^2)} \le \frac{[E(x_m)]^2}{E(x_m^2)} \end{aligned}$$

In other words, all local extrema with strictly positive risk premiums are the local maxima. To show that the local maximum is also the global maximum, assume that some other x is also a local extremum satisfying (A.2) with $E(x) > 0$. Then exchanging places between x and x_m yields a first-order condition

$$E(x_m)E(x^2) - E(x)E(xx_m) = 0 \tag{A.6}$$

It follows by comparing (A.6) with (A.5) that

$$\eta(x) = \frac{E(x)}{\sqrt{E(x^2)}} = \frac{E(x_m)}{\sqrt{E(x_m^2)}} = \eta_m$$

▌

Proof of Theorem 2: By Theorem 1, for $x_i = r_i - r_0$ we have

$$0 = E(x_i)E(x_m^2) - E(x_m)E(x_m x_i) \tag{A.7}$$

Defining $\lambda_m = E(x_m)/E(x_m^2)$ yields an equivalent expression

$$\begin{aligned} E(r_i - r_0) &= E(x_m(r_i - r_0))\frac{E(x_m)}{E(x_m^2)} = E(x_m(r_i - r_0))\lambda_m \\ &\iff E(r_i) - E(x_m r_i)\lambda_m = r_0(1 - E(x_m)\lambda_m) \\ &\iff E(r_i(1 - x_m\lambda_m)) = r_0(1 - E(x_m)\lambda_m) \end{aligned}$$

Note that

$$E(x_m)\lambda_m = \frac{[E(x_m)]^2}{E(x_m^2)} = \frac{[E(x_m)]^2}{[E(x_m)]^2 + Var(x_m)} < 1$$

Thus $(1 - E(x_m)\lambda_m) = 1 - \eta_m^2 > 0$. Define

$$\delta = A_m(1 - \lambda_m x_m), \quad A_m = \frac{1}{r_0(1 - \eta_m^2)}$$

It is easily seen that δ is a SDF: for all i,

$$1 = E_t(\delta r_i), \quad 0 = E_t(\delta x_i) \tag{A.8}$$

The equivalence between (A.7) and (A.8) is thus established. ▌

Proof of Theorem 3: (Again the necessity part of the proof is standard and the sufficiency part is somewhat new.) Let a benchmark portfolio m be given such that $E(x_m) > 0$. Consider any zero-price portfolio with payoff $x \in X$ (including the special case $x_i = r_i - r_0$). Let π denote a portfolio that is formed by one dollar of m and θ amount of exposure in x so that the upper-market gain and lower-market loss of π are

$$\overline{x}_\pi = \overline{x}_m + \theta\overline{x},\ \underline{x}_m + \theta\underline{x}$$

The GL ratio of π is $Z_\pi = E(x_\pi^+)/E(-x_\pi^-)$.

Necessity: Assume that m is GL efficient. If $E(\overline{x}) = E(\underline{x}) = 0$, then $Z_\pi = Z_m$ for all θ. Otherwise (e.g., Bawa and Lindenberg, 1977), under our assumption that the asset returns are continuously distributed, the maximum Z_π necessarily satisfies the first-order condition evaluated at $\theta = 0$ (so that $x_\pi = x_m$):[1]

$$\begin{aligned}
\max_\theta Z_\pi \quad &\Rightarrow \quad \frac{\partial Z_\pi}{\partial \theta}|_{\theta=0} = 0 \\
&\Rightarrow \quad E(\overline{x})E(\underline{x}_m) - E(\overline{x}_m)E(\underline{x}) = 0 \\
&\Longleftrightarrow \quad \frac{E(\overline{x})}{E(\overline{x}_m)} = \frac{E(\underline{x})}{E(\underline{x}_m)} \equiv \varphi \text{ for some } \varphi \text{ hence (22)-(23)} \\
&\Longleftrightarrow \quad Z^R \equiv \frac{E(\overline{x})}{E(\underline{x})} = Z_m \text{ hence (24)}
\end{aligned}$$

Sufficiency: Assume that (24) holds for m. Define $E(\cdot, x_m \leq 0) = E(\cdot|x_m \leq 0)Pr(x_m \leq 0)$ where $E(\cdot|x_m \leq 0)$ is the expectation conditional on event $x_m \leq 0$ (likewise, with event $x_m > 0$). Suppose on the contrary that m is not GL efficient. Then there exists $\hat{m}$ with $E(x_{\hat{m}}) > 0$ such that

$$Z_m - 1 \equiv \frac{E(x_m)}{E(-x_m; x_m \leq 0)} < Z_{\hat{m}} - 1 = \frac{E(x_{\hat{m}})}{E(-x_{\hat{m}}; x_{\hat{m}} \leq 0)} \tag{A.9}$$

However, noting $E(x_{\hat{m}; x_{\hat{m}} \leq 0)}) \leq E(x_{\hat{m}}; x_m \leq 0)$ and from (24) we have

$$Z_{\hat{m}} - 1 = \frac{E(x_{\hat{m}})}{E(-x_{\hat{m}}; x_{\hat{m}} \leq 0)} \leq \frac{E(x_{\hat{m}})}{E(-x_{\hat{m}}; x_m \leq 0)} = Z_m - 1$$

which contradicts (A.9). The contradiction shows that m is GL efficient. ▌

Proof of Theorem 4: From Theorem 3, m is GL efficient if and only if for all $i \in \Lambda$

$$E(r_i - r_0; x_m > 0) = Z_m E(r_0 - r_i; x_m \leq 0) \tag{A.10}$$

or equivalently

$$E(r_i; x_m > 0) + Z_m E(r_i; x_m \leq 0) = r_0[Pr(x_m > 0) + Z_m Pr(x_m \leq 0)]$$

[1]If returns are discretely distributed, a GL efficient portfolio may be found, e.g., using Bernardo and Ledoit's (2000) technique. The first-order condition, however, may not be satisfied if the GL efficient portfolio m exhibits $Pr(x_m = 0) > 0$. In such cases, more assumptions would be needed to ensure the characterizations in (22)-(24). This line of technical generalization is not pursued here.

It is easy to verify that for the SDF δ defined in (25)-(27), $E(\delta r_i) = 1$ and $E(\delta x_i) = 0$ if and only if (A.10) hold for all i. ▮

Proof of Lemma 3: Assume the two-fund separability holds. Let r_0 denote the risk-free rate or return on any "zero-beta" portfolio that is uncorrelated with m. By definition of two-fund separability we can derive $r_i - r_0 = b_i y + \varepsilon_i$ with $E(\varepsilon_i|y) = 0$ for all i. Suppose m is a well diversified portfolio so that $x_m = b_m y$ and $E(x_m) = b_m E(y) > 0$. Consider any arbitrary portfolio π with $x_\pi = b_\pi y + \varepsilon_\pi$ and we are ready to prove the lemma.

MV efficiency of m: We have

$$\begin{aligned} E(x_\pi) &= b_\pi E(y), E(x_\pi^2) = b_\pi^2 E(y^2) + E(\varepsilon_\pi^2) \\ \eta_\pi^2 &= \frac{b_\pi^2 [E(y)]^2}{b_\pi^2 E(y^2) + E(\varepsilon_\pi^2)} \le \frac{[E(y)]^2}{E(y^2)} = \eta_m^2 \end{aligned}$$

GL efficiency of m: Assume $b_\pi \ge 0$ (for $b_\pi \le 0$ the proof is analogous). Since $\min(b_\pi y + \varepsilon_\pi, 0)$ is a concave function of ε_π conditional on all y, by Jensen's inequality

$$E(\min(b_\pi y + \varepsilon_\pi, 0)|y) \le \min(E(b_\pi y + \varepsilon_\pi|y), 0) = \min(b_\pi y, 0)$$

Taking expectation over y yields

$$E(\min(b_\pi y + \varepsilon_\pi, 0)) \le E(\min(b_\pi y, 0))$$

It follows that for all portfolio $\pi = b_\pi y + \varepsilon_\pi$

$$\begin{aligned} -E(\min(x_\pi, 0)) &= -E(\min(b_\pi y + \varepsilon_\pi, 0)) \ge -E(\min(b_\pi y, 0)) \\ &\Rightarrow Z_\pi - 1 = \frac{E(x_\pi)}{-E(x_\pi^-)} \le \frac{E(b_\pi y)}{-E(b_\pi y^-)} = \frac{E(x_m)}{-E(x_m^-)} = Z_m - 1 \end{aligned}$$

▮

Proof of Theorem 5: Necessity: Suppose m is both MV and GL efficient so that (19) and (29) hold. By comparison we have

$$\begin{aligned} c_i &= A_m[E(\overline{x}_i) - \lambda_m E(\overline{x}_i \overline{x}_m)] = B_m E(\overline{x}_i) \\ &\Rightarrow (A_m - B_m) E(\overline{x}_i) = \lambda_m E(\overline{x}_i \overline{x}_m) && \text{(A.11)} \\ &\Rightarrow (A_m - B_m) E(\overline{x}_m) = \lambda_m E(\overline{x}_m^2) && \text{(A.12)} \end{aligned}$$

Since $\lambda_m > 0$, from (A.12) $A_m - B_m > 0$. Thus dividing (A.11) by (A.12) yields (31):

$$E(\overline{x}_i) = \frac{E(\overline{x}_i \overline{x}_m)}{E(\overline{x}_m^2)} E(\overline{x}_m) = \beta_i^+ E(\overline{x}_m)$$

For the downside, (20) and (30) imply

$$\begin{aligned} p_i &= A_m[E(\underline{x}_i) + \lambda_m E(\underline{x}_i \underline{x}_m)] = B_m Z_m E(\underline{x}_i) \\ &\Rightarrow (B_m Z_m - A_m) E(\underline{x}_i) = \lambda_m E(\underline{x}_i \underline{x}_m) \qquad \text{(A.13)} \\ &\Rightarrow (B_m Z_m - A_m) E(\underline{x}_m) = \lambda_m E(\underline{x}_m^2) \qquad \text{(A.14)} \end{aligned}$$

From (A.14) and $\lambda_m > 0, B_m Z_m - A_m > 0$. Thus (32):

$$E(\underline{x}_i) = \frac{E(\underline{x}_i \underline{x}_m)}{E(\underline{x}_m^2)} E(\underline{x}_m) = \beta_i^- E(\underline{x}_m)$$

Finally, (22) and (23) imply that $\varphi_i = \beta_i^+ = \beta_i^- = \beta_i^B$, where the last equation derives from the mathematical property that for real numbers a, b, c, d, if $a/b = c/d$ and $b, d > 0$ then $a/b = c/d = (a+c)/(b+d)$.

Sufficiency: Assume that (31)-(33) hold for m. Then subtracting (32) from (31) and using (33) yields the modified CAPM conditions (15)-(16); hence by Theorem 1 m is MV efficient. Now dividing (31) by (32) and using (33) yields (24), which is equivalent to conditions (22)-(23). By Theorem 3) m is GL efficient. ∎

REFERENCES

Ang, A. and J., Chen, 2002, Asymmetric correlations of equity portfolios. *Journal of Financial Economics* **63**, 443-494.

Bawa, Vijay S., 1975, Optimal rules for ordering uncertain prospects. *Journal of Financial Economics* **2**, 95-121.

Bawa, Vijay S., 1976, Admissible portfolios for all individuals. *The Journal of Finance* **XXXI**, 1169-1183.

Bawa, V.S. and E.B., Lindenberg, 1977, Capital market equilibrium in a mean-Lower partial moment framework. *Journal of Financial Economics* **5**, 189-200.

Bernardo, A.E. and O., Ledoit, 2000, Gain, loss, and asset pricing. *Journal of Political Economy* **vol. 108, no. 1**, 144-172.

Black, F., 1972, Capital market equilibrium with restricted borrowing. *Journal of Business* **45**, 444-455.

Black, F. and M. Scholes, 1973, The valuation of options and corporate liabilities. *Journal of Political Economy* **81**, 637-654.

Cochrane, J. H. and J., Saa-Requejo, 2000, Beyond arbitrage: Good-deal asset price bounds in incomplete markets. *Journal of Political Economy* **108**, 79-119.

Cochrane, J. H., 2001, *Asset Pricing.* Princeton University Press, Princeton, New Jersey.

Domar, E.V. and R.A. Musgrave., 1944, Proportional income taxation and risk-taking. *Quarterly Journal of Economics* **58**, 389-422.

Dybvig, P. H. and J.E. Ingersoll, 1982, Mean-variance theory in complete markets. *Journal of Business* **55**, 233-251.

Fama, E. F. and K.R. French, 1997, The industry costs of equity. *Journal of Financial Economics* **43**, 153-193.

Fishburn, P.C., 1977, Mean-risk analysis with risk associated with below-target returns. *The American Economic Review* **67**, 116-126.

Hansen, L. P. and R. Jagannathan, 1997, Assessing specification errors in stochastic discount factor models. *Journal of Finance* **52**, 557-590.

Ingersoll, J.E. Jr., 1987, *Theory of Financial Decision Making.* Rowman & Littlefield Publishers, Inc., Maryland.

Lintner, J., 1965, The valuation of risk assets and the selection of risky investment in stock portfolios and capital budgets. *Review of Economics and Statistics* **47**, 13-37.

Longin F. and B. Solnik, 2001, Extreme correlation of international equity markets. *Journal of Finance* **LVI**, 649-676.

Markowitz, H. M., 1952, Portfolio selection. *Journal of Finance* **7**, 77-91.

Pástor, L. and R. F. Stambaugh, 1999, Costs of equity capital and model mispricing. *Journal of Finance* **54**, 67-121.

Roll, R., 1977, A critique of the asset pricing theory's tests: Part I. *Journal of Financial Economics* **4**, 129-176.

Roll, R. and S. A. Ross, 1994, On the cross-sectional relation between expected returns and betas. *Journal of Finance* **49**, 101-121.

Ross, S.A., 1978, Mutual fund separation in financial theory: The separating distributions. *Journal of Economic Theory* **17**, 254-286.

Sharpe, W. F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* **19**, 423-439.

Tobin, J., 1958, Liquidity preference as behavior toward risk. *Review of Economic Studies* **25**, 65-85.

Zou, L., 2000, Inherent efficiency, security markets, and the pricing of investment strategies. *Tinbergen Institute Discussion Paper.* 00-108/2. http://www.tinbergen.nl/home.html (Publications/Discussion Papers).

Zou, L., 2003, Compound utility theory. *Working Paper.* Finance Group, University of Amsterdam, submitted to the Review of Economic Studies.

Zou, L., 2004, Dichotomous analysis of the Fama-French portfolios. *Working Paper.* Finance Group, University of Amsterdam.