Review of Finance (2005) 9: 415–435
DOI 10.1007/s10679-005-2265-x

The Generalized Treynor Ratio [*]

GEORGES HÜBNER
HEC Management School – University of Liège and Maastricht University

Abstract. This paper extends the Treynor performance ratio for a single index to the case of multiple indexes. The new measure, called the Generalized Treynor Ratio, preserves the same key geometric and analytical properties of the original Treynor Ratio. The Generalized Treynor Ratio is defined as the abnormal return of a portfolio per unit of premium-weighted average systematic risk, normalized by the premium-weighted average systematic risk of the benchmark. Numerical simulations reveal that the portfolio rankings produced with this measure are more precise and more stable than the ones provided by Jensen's alpha and the Information Ratio.

1. Introduction

The Capital Asset Pricing Model developed by Treynor (1961), Sharpe (1964) and Lintner (1965) has provided the framework for a number of performance measures for managed portfolios, five of which have been broadly adopted in the financial literature. The Sharpe (1966) Ratio and the Treynor and Black (1973) 'Appraisal Ratio'[1] are derived from the Capital Market Line, with the level of risk being measured by the standard deviation of portfolio returns. Jensen's (1968) alpha, defined as the portfolio's excess return over the required average return, the Treynor (1966) ratio, defined as the alpha divided by the portfolio beta, and the Information Ratio, defined as the alpha divided by the standard deviation of the portfolio residual returns, are directly linked to the beta through the Security Market Line.

Beyond these five measures, the supply of new and widely accepted quantitative measures for the performance assessment of single index models has been scarce. Fama (1972) proposes a decomposition of performance between timing and selection abilities, while Treynor and Mazuy (1966) and Henriksson and Merton (1981) design performance measures aimed at measuring market timing abilities.

[*] I wish to thank Wayne Ferson, Pascal François, Eric Jacquier, Lars Tyge Nielsen, Nicolas Papageorgiou, Peter Schotman, Marco Pagano (the Editor) and three anonymous referees, as well as participants at the 2003 International AFFI Conference (Lyon), the 2003 NFA Conference (Québec) and seminars at Maastricht University and CORE (Louvain-la-Neuve) for helpful comments. The author acknowledges financial support from Deloitte Luxemburg and a research grant of the Belgian National Funds for Scientific Research (FNRS). Part of this research was completed while I was visiting HEC Montréal. All errors remaining are mine.

[1] We refer to the original performance measure developed by Treynor and Black (1973) and not to the one frequently quoted in the literature, defined later as the Information Ratio.

In a more recent paper, Modigliani and Modigliani (1997) propose an alternative measure of risk that refines the Sharpe Ratio by adjusting performance for portfolio leverage, using the volatility of returns within the CAPM framework.

Certain theoretical extensions to these performance measures exist for multi-index models. In the context of the Ross (1976) Arbitrage Pricing Theory, Jobson and Korkie (1984) and Connor and Korajczyk (1986) develop multi-factor counterparts of the Jensen (1968) and Treynor and Black (1973) measures, while Sharpe (1992, 1994) provides conditions under which the Sharpe Ratio can be applied in a multi-index model. Ferson and Schadt (1996) introduce and test conditional versions of the measures of Jensen (1968), Treynor and Mazuy (1966) and Henriksson and Merton (1981). There exists to date no way of generalizing the Treynor performance measure into a multi-index framework.

In this paper, we propose a generalization of the Treynor Ratio that holds for any unconditional multi-index linear asset pricing model. The measure proposed in this paper, entitled the Generalized Treynor Ratio, is defined as the abnormal return of a portfolio per unit of premium-weighted average systematic risk, normalized by the premium-weighted average systematic risk of the benchmark. This ratio ensures that the key economic and mathematical properties of the original Treynor Ratio are preserved. In particular, we obtain the same intuitive geometric interpretation as with the Treynor Ratio, namely the "portfolio performance per unit of systematic risk", i.e., the ratio of abnormal excess return (Jensen's alpha) to systematic risk exposure (the beta).

Our new ratio potentially addresses some well-documented issues raised with the application of the alpha. Currently, in the absence of a proper generalization of the Treynor Ratio, empirical performance studies on style-based portfolios with multi-index asset pricing models use Jensen's (1968) alpha. Being a scalar, the alpha can be measured and interpreted for any level of systematic risk exposure, including non-directional strategies. Thus, the alpha is a suitable performance metric when the systematic risk level of the funds is close to zero. In contrast, being a ratio, the Treynor measure is very sensitive to the denominator and provides unstable and imprecise performance measures for market neutral funds because of the risk of measurement error. For non-directional hedge funds, for instance, it is not likely to provide reliable performance values or stable portfolio rankings. For funds with negative betas, the Treynor Ratio is, in effect, inapplicable as it attributes a negative performance to funds with positive abnormal returns. As a matter of fact, in their comprehensive study, Kothari and Warner (2001) consider solely the alpha in their empirical comparison of multi-index asset pricing models for mutual fund returns. However, the alpha is likely to exhibit a major drawback when confronted with real data on directional portfolio strategies. While the portfolio risk exposures are controlled for in its derivation, the alpha itself is not invariant to a change in portfolio leverage. This phenomenon of "leverage bias", originally identified by Modigliani and Pogue (1974), diminishes the ability of Jensen's alpha to properly rank leveraged portfolios. The Generalized Treynor Ratio is insensitive to portfolio leverage. On this front, our new measure provides a more reliable assessment of performance than the alpha.

From a financial practitioners perspective, the unsystematic risk is of foremost importance. Hence, practitioners have gradually adopted the Information Ratio as a relevant performance measure for actively managed portfolios (see, e.g., Grinold and Kahn, 1992, 1995, 2000; Goodwin, 1998). Underlying this choice, the Information Ratio of a portfolio claims to provide, ex ante, a proxy for its return potential, and thus represents, ex post, a consistent tool for assessing the manager's ability to realize abnormal returns. Nonetheless, the Information Ratio is saddled with several serious drawbacks. Ledoit and Wolf (2003) show that measurement issues lead to strong overestimation of the residual variance, leading to a downward bias of the Information Ratio (Ledoit and Wolf, 2004). The biases due to measurement errors may alter the relative portfolio rankings based on this measure. The Generalized Treynor Ratio, whose denominator only depends on the portfolio betas, is not exposed to these biases.

The second part of this article shows, in an experimental setup, that if the difference in abnormal performance of multiple portfolios has been recorded at the expense of additional systematic risk, then the Generalized Treynor Ratio has a superior capacity to rank these portfolios than Jensen's alpha or the Information Ratio. We assess the quality of the ranking scheme produced by a performance measure by its precision to reproduce the true ranking and by the stability of the rankings it produces under alternative asset pricing models. For both criteria, the Generalized Treynor Ratio provides very reliable rankings, which lends support for its adoption as a useful performance metric for managed portfolios

The rest of the paper is organized as follows. Section 2 discusses the classical performance measures developed along with the CAPM. Section 3 proposes the generalization of the Treynor Ratio to a multi-index setup. The next section presents the results of numerical simulations on market-timing and stock-picking strategies. Section 5 concludes.

2. Performance Measures in the CAPM

This section discusses three classical performance measures developed within the context of the CAPM and analyzes them in relation to the set of properties displayed by the Treynor Ratio.

2.1. JENSEN'S ALPHA, THE TREYNOR RATIO AND THE INFORMATION RATIO

We consider performance measures that use the Security Market Line derived by Treynor (1961), Sharpe (1964) and Lintner (1965) as our framework. This line represents the expected total return of every security or portfolio i as a linear function of the return of the market portfolio m:

$$\mathrm{E}_i = R_f + \beta_i \left[\mathrm{E}_m - R_f\right] \tag{1}$$

where $E_i \equiv E[R_i]$ denotes the unconditional continuous expected return of security i, R_f denotes the continuous return on the risk-free security and $\beta_i = \frac{\mathrm{cov}(R_i, R_m)}{\sigma^2(R_m)}$ is the beta of security i.

This equilibrium relation corresponds to the market model $r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it}$, where $r_i = R_i - R_f$ denotes the excess return on security i. The ex post (realized) equation is:

$$\overline{r}_i = \widehat{\alpha}_i + \widehat{\beta}_i \overline{r}_m \tag{2}$$

where $\widehat{\beta}_i = \frac{\widehat{\mathrm{cov}}(r_i, r_m)}{\widehat{\sigma}^2(r_m)}$ and $\widehat{\alpha}_i = \overline{r}_i - \widehat{\beta}_i \overline{r}_m$ are the estimators of β_i and α_i, respectively.

If the CAPM holds and if markets are efficient, α_i should not be statistically different from 0.

Equation (2) constitutes the source of three major performance measures of financial portfolios: Jensen's alpha (1968), the Treynor Ratio (1966) and the Information Ratio.

Jensen's alpha is measured by $\widehat{\alpha}_i$ in Equation (2). It is the percentage excess return earned in addition to the required average return over the period. This is an absolute performance measure – it is measured in the same units as the return itself – after controlling for risk.

The Treynor Ratio (*TR*) is the ratio of Jensen's alpha over the stock beta:

$$TR_i = \frac{\widehat{\alpha}_i}{\widehat{\beta}_i} \tag{3}$$

Thus, it is interpreted as the percentage abnormal return earned per unit of systematic risk of the portfolio.[2]

Finally, the Information Ratio (*IR*) is a measure of Jensen's alpha per unit of portfolio specific risk, measured as the standard deviation of the market model residuals:

$$IR_i = \frac{\widehat{\alpha}_i}{\widehat{\sigma}(\varepsilon_i)} \tag{4}$$

In this simple classical setup, both the *IR* the *TR* appear to provide additional and complementary information with respect to Jensen's alpha: two securities with different systematic or specific risk levels that provide the same excess returns over the same period will have the same alpha but will differ with respect to the other two ratios.

[2] Even though financial practice has been persistently defining the Treynor Ratio as the ratio of total return over the beta, the formulation of Equation (3) corresponds to the original measure developed by Treynor (1966). Thus, this paper will proceed from now on with the original measure.

2.2. PROPERTIES OF PERFORMANCE MEASURES

The only pure axiomatic analysis of performance measures for financial portfolios is proposed by Chen and Knez (1996). Their paper focuses on the set of properties that so-called "admissible" performance measures should respect to sustain the law of one price in securities markets. The authors come up with four conditions for the admissibility of a performance measure: it should (i) assign zero performance to every reference portfolio (i.e., any portfolio that would be held at equilibrium using public information), (ii) be linear – any rescaling or mix of portfolios should not improve the manager's performance ranking, (iii) be continuous – portfolios that produce indistinguishable returns are ranked the same, and (iv) be nontrivial – if some traded security produces nonzero abnormal returns, it is also assigned nonzero performance. Among admissible performance measures, Chen and Knez (1996) identify Jensen's alpha, but also the APT-based performance measure of Connor and Korajczyk (1986), the period weighting measures of Grinblatt and Titman (1989) and the intertemporal marginal rates of substitution of Glosten and Jagannathan (1994). However, they show that the Sharpe (1966) performance measure is not admissible as it fulfills neither condition (i) nor (ii), because a ratio of the linear combination of its elements is not equal to the linear combination of the ratios. For the same reason, the Treynor Ratio and the Information Ratio violate condition (ii) as well.

Yet, Chen and Knez (1996) acknowledge that their axiomatic approach relates to the no-arbitrage and pricing properties of financial markets, but that it does not yield normative results about other fundamental issues such as the ranking of portfolios or the consistency with utility maximization. Specifically, they recognize that the Sharpe Ratio may be adequate for these purposes. This measure is independent of the parametrization of asset pricing models and consistent with maximization of expected utility whenever the mean-variance framework is applicable for investors. Such properties are not met by the alpha. In particular, condition (ii) set forth by Chen and Knez (1996) allows for the leverage bias identified by Modigliani and Pogue (1974), thus not warranting that admissible performance measures provide reasonable portfolio rankings.

Instead of establishing a normative framework similar to Chen and Knez (1996), this article adopts a more pragmatic view for practical and empirical purposes. We identify the key properties that are desired by researchers or analysts when selecting performance measures associated with a particular asset pricing specification. Specifically, the measure of risk-adjusted portfolio performance must be monotonically decreasing with respect to its associated risk measure – whether systematic, specific or total – and monotonically increasing with respect to the return measure. When applied to portfolios, the performance measure must also exhibit an averaging property: this implies that a portfolio manager should not be able to manipulate her performance by acquiring a convex combination of several funds whose performance measure would be greater than the one of the best performing fund in the portfolio. The performance measure should also be leverage-invariant, i.e., be homogenous of degree zero with respect to the scale of the return, which rules out Jensen's alpha. Finally, as for admissible performance measures, there must exist a reference portfolio in the set of zero-performance funds.

For the purpose of ranking portfolio managers, the previous discussion provides relevant tools to analyze the performance measures. In relation with those dimensions, four key features of the univariate Treynor Ratio – denoted (U1) to (U4) – can be emphasized, provided that systematic risk is positive in all cases:

- (U1) *Ratio of Euclidean distances:* $TR(\widehat{\alpha}_i, \widehat{\beta}_i, \overline{r}_m) \equiv TR_i = \frac{D(0,\widehat{\alpha}_i)}{D(0,\widehat{\beta}_i)}$, where D(.,.) denotes the Euclidean distance operator.
- (U2) *Homogeneity*
  - (U2a) *with respect to abnormal return (degree 1):* $TR(k\widehat{\alpha}_i, \widehat{\beta}_i, \overline{r}_m) = kTR(\widehat{\alpha}_i, \widehat{\beta}_i, \overline{r}_m)$ for $k > 0$
  - (U2b) *with respect to risk (degree −1):* $TR(\widehat{\alpha}_i, k\widehat{\beta}_i, \overline{r}_m) = k^{-1}TR(\widehat{\alpha}_i, \widehat{\beta}_i, \overline{r}_m)$ for $k > 0$
  - (U2c) *with respect to risk premium (degree 0):* $TR(\widehat{\alpha}_i, \widehat{\beta}_i, k\overline{r}_m) = TR(\widehat{\alpha}_i, \widehat{\beta}_i, \overline{r}_m)$ for $k > 0$
- (U3) *Proportionality*
  - (U3a) *to portfolio abnormal return:* $TR(\widehat{\alpha}_p, \widehat{\beta}_p, \overline{r}_m) = (\lambda + (1-\lambda)k)TR(\widehat{\alpha}_i, \widehat{\beta}_i, \overline{r}_m)$ for $\overline{r}_p = \lambda\overline{r}_i + (1-\lambda)\overline{r}_{i'}$, $\widehat{\alpha}_{i'} = k\widehat{\alpha}_i$, $\widehat{\beta}_{i'} = \widehat{\beta}_i$.
  - (U3b) *inverse to portfolio required return:* $TR(\widehat{\alpha}_p, \widehat{\beta}_p, \overline{r}_m) = (\lambda + (1-\lambda)k)^{-1}TR(\widehat{\alpha}_i, \widehat{\beta}_i, \overline{r}_m)$ for $\overline{r}_p = \lambda\overline{r}_i + (1-\lambda)\overline{r}_{i'}$, $\widehat{\alpha}_{i'} = \widehat{\alpha}_i$, $\widehat{\beta}_{i'} = k\widehat{\beta}_i$ and $(\lambda + (1-\lambda)k) > 0$
- (U4) *Reference portfolio*: $TR(\widehat{\alpha}_i, \widehat{\beta}_m, \overline{r}_m) = \widehat{\alpha}_i$ for some ex post reference portfolio m

The Treynor Ratio, Jensen's alpha and the Information Ratio diverge as to whether they satisfy this set of properties. Jensen's alpha violates properties (U1), (U2) and (U3), while the Information Ratio only violates property (U3).

Unlike the alpha, the Treynor Ratio and the Information Ratio are both slope measures, i.e., ratios of a Euclidean norm in the returns space to a Euclidean norm in the space of risk. Property (U1) provides an economic justification related to the nature of the returns and risk measures. The homogeneity properties (U2) reflect the basic improvement of the Treynor Ratio and the Information Ratio with respect to Jensen's alpha, the linear sensitivity to risk. This implies that the Treynor Ratio and the Information Ratio are leverage-invariant. Property (U3) means that the performance of a portfolio of two assets that have the same beta or the same alpha is directly or inversely proportional, respectively, to the Treynor Ratios of its components. This property is indeed specific to the Treynor Ratio, because the beta of a portfolio is linearly related to the betas of its components, unlike the *IR* that relies on a measure of standard deviation. As the required return does not appear in the formula for Jensen's alpha, it cannot behave proportionally to the level of required return of a portfolio. Finally, the identification of a reference portfolio enables the comparability of all funds using the performance measure. This is the only property shared by all three measures.

To summarize, the Treynor Ratio is, in the context of the CAPM, the single measure that uses a slope (ratio of distances) measure, ensures homogeneity with respect to risk and return, presents the averaging property for portfolios and is leverage invariant. The alpha is the only admissible measure in the sense of Chen and Knez (1996), but meets none of these requirements. The Information Ratio does not present the averaging property and so is subject to manipulation.

3. The Generalization of the Treynor Ratio

Since the publication of Ross's (1976) Arbitrage Pricing Theory (APT), it has been widely recognized that the use of a single index in a market model is probably insufficient to keep track of the systematic sources of portfolio returns in excess of the risk free rate. Consequently, a variety of unconditional asset pricing models are developed to account for this under-specification for empirical purposes.[3] They have mostly taken four directions. The family of empirical extensions of the classical CAPM increments the basic model by adding additional risk premiums related to the "attributes" of the assets, like in Fama and French (1993) and Carhart (1997). APT-based approaches use a multi-factor specification, like in Chen, Roll and Ross (1986). Thirdly, the stream of asset-class models initiated by Sharpe (1992) relates observed security returns to several classes of assets like equity, bonds or currencies. Finally, multi-moment asset pricing models originated by the seminal paper of Kraus and Litzenberger (1976) build on investors' preferences for higher moments than the variance.

In spite of their differences in conception and statistical assumptions, all these approaches share a very general affine model specification that can be summarized as follows, still considering the ex-post multidimensional equation corresponding to (2):

$$\overline{r}_i = \widehat{\alpha}_i + \sum_{j=1}^{K} \widehat{\beta}_{ij} \overline{r}_j = \widehat{\alpha}_i + \widehat{\mathbf{B}}_i \overline{\mathbf{R}} \tag{5}$$

where $j = 1, \ldots, K$ denotes the number of distinct risk premiums, the line vector $\widehat{\mathbf{B}}_i = (\widehat{\beta}_{i1}, \ldots, \widehat{\beta}_{iK})$ and the column vector $\overline{\mathbf{R}} = (\overline{r}_1, \ldots, \overline{r}_K)^\top$ represent risk loadings and average returns for the premiums, respectively. For Equation (5) to make sense in the context of the measurement of risk-adjusted performance, we

[3] We do not encompass here the stream of theoretical extensions of the CAPM to general equilibrium or exchange economies setups, which is beyond the scope of this paper. Similarly, we shall not discuss the class of conditional asset pricing models, bearing in mind that the results obtained in this paper could be extended to conditional performance measures along the methodology developed by Ferson and Schadt (1996).

only consider cases where $\widehat{\mathbf{B}}_i\overline{\mathbf{R}} \geq 0$. The alpha remains a scalar whereas the systematic risk measures of the portfolio is made of a column vector of loadings of the individual risk premiums whose sum can be equal to zero.

In the portfolio management area, the sole use of Jensen's alpha could result in unfair judgements over the performance of portfolios that have been invested in very different classes of risks. The failure to account for the portfolio risk exposure leads to unduly favoring well-performing risky portfolios (high positive alphas) while condemning bad performers in the same population (low negative alphas).

The most likely explanation for the current shortage of risk-adjusted, relative performance measure such as the Treynor measure is the lack of a natural reduction method, mapping from the K-dimensional risk space to a one-dimensional measure. The most intuitive reduction rule that would account for all sources of risk, namely $\frac{\widehat{\alpha}_i}{\sum_{j=1}^{K}\widehat{\beta}_{ij}}$ is unsatisfactory, as it does not respect the scale independence (homogeneity of degree 0 with respect to scale) property. The denominator of this ratio cannot be considered as a Euclidean distance measure either: to see this, it suffices to note that the sum of betas can be negative, and thus not represent a distance, without altering the risk-return relationship. These basic obstacles appeared to be sufficient to prevent theoretical research from digging any deeper in this area.

The intuition leading to the Generalized Treynor Ratio is straightforward. The main problem we are looking to resolve is that the actual pricing framework involves sources of risk that can be interdependent and of unequal importance. If one restates the asset pricing model with totally independent sources of risk that correspond to identical risk premiums, the problem is solved: all portfolios with identical sum of the risk loadings (betas) have comparable global exposures to sources of systematic risk. As a generalization of the single index Treynor Ratio, the proposed performance measure must satisfy a similar set of basic properties as the ones listed above. It should be interpreted as a ratio of distances, respect the same kinds of heterogeneity and proportionality conditions, and yield the alpha as the performance of the benchmark. Furthermore, to properly accomplish its role, this performance measure ought to be as parsimonious as possible, i.e., to display the lowest number and least repetition of parameters as possible. The solution is provided in the following proposition.

PROPOSITION 1. If the ex-post asset pricing equation of every portfolio i corresponds to a K-factor model that can be represented by Equation (5) where $\sum_{j=1}^{K}\hat{\beta}_{ij}\overline{r}_j > 0$, the Generalized Treynor Ratio corresponding to the following equation:

$$GTR_i = \hat{\alpha}_i \frac{\sum_{j=1}^{K}\hat{\beta}_{mj}\overline{r}_j}{\sum_{j=1}^{K}\hat{\beta}_{ij}\overline{r}_j} \tag{6}$$

$$= \hat{\alpha}_i \frac{\hat{\mathbf{B}}_m\overline{\mathbf{R}}}{\hat{\mathbf{B}}_i\overline{\mathbf{R}}} \tag{7}$$

where m is the benchmark portfolio with identical leverage as portfolio i, is the most parsimonious performance measure that respects the same properties as the Treynor ratio, that is independent of the model specification and that reduces to the Treynor ratio for $K = 1$.

Proof. See Appendix.

The key to the geometric argument underlying this Proposition is to reduce the risk dimension to orthogonal, normalized axes and to pick a portfolio that has the same coordinates on every axis. The output of this procedure allows us to use the ingredients of the initial asset pricing model, even if risk premiums are not independent.

This expression for *GTR* has an interpretation that directly relates to the original Treynor Ratio: it provides the abnormal return of portfolio i per unit of premium-weighted average systematic risk, normalized by the market premium-weighted average systematic risk or, in short, by some proper unit of systematic risk. To see this, note that:

$$GTR_i = \frac{\widehat{\alpha}_i}{\left(\dfrac{\sum_{j=1}^{K} w_j \widehat{\beta}_{ij}}{\sum_{j=1}^{K} w_j \widehat{\beta}_{mj}}\right)} \tag{8}$$

where $w_j = \frac{\overline{r}_j}{\sum_{j=1}^{K} \overline{r}_j}$. Notice that the required excess return is constrained to be positive for both the portfolio under study and the benchmark.[4]

A crucial property of this ratio is the perfect independence of the measure with the choice of model specification. Consider two alternative specifications that use different risk premiums, but that provide exactly the same required returns for all portfolios. As both the numerator and the denominator of the ratio in Equation (6) use the total required returns of portfolios as inputs, the *GTR* will be the same for all portfolios under each specification.

Alternatively, this property can be viewed in light of required returns: it states that two portfolios that exhibit different risk loadings but whose required returns are similar may eventually be ranked according to their alphas. This is the translation of the rather natural requirement that portfolios that bear the same risk in aggregate are perfectly equivalent on the risk side and should only be compared through their abnormal performance. However, this should not be confused with the outcome of a simple releveraging of the portfolio, as the *GTR* shows that the performance per unit of systematic risk is comparable for any vector of factor loadings, including but not limited to proportional values of the betas: two portfolios that only differ with respect to their leverage deliver the same *GTR* as their benchmark portfolios display the same leverage as well.

[4] The absence of a benchmark portfolio could still lead to an interpretable measure, but its usefulness could only be sustained within a particular version of the asset pricing model. Then, the dependence of *GTR* on the scale of the risk premiums would forbid any cross-model comparison.

4. Numerical Simulations

We now assess the quality of the performance measures within a controlled experiment. We compare the *GTR* to the alpha and the *IR*, as the latter two are the most popular performance measures. The purpose of this exercise is to determine whether, in the presence of realized portfolio performance at the expense of an increase in systematic risk exposure, the *GTR* provides more reliable rankings than the alpha and the *IR*, or at least equally good ones. Unlike Kothari and Warner (2001), who use the alphas as instruments to assess the ability of asset pricing specifications to detect abnormal performance, our objective is exactly the converse. The performance measures are under study, and the pricing models are used as instruments.

The quality of a performance measure is assessed along two dimensions: (i) the ability of a criterion to rank funds in a way that is as close as possible to the theoretical ranking induced by the simulation procedure; (ii) the stability of these rankings when confronted with alternative asset pricing models, whether over-specified or under-specified. The first dimension represents the precision with which a performance measure reproduces the perfect ranking of portfolio managers. The second dimension corresponds to the reliability of the ranking obtained with a performance measure when uncertainty prevails vis-à-vis the correct asset pricing specification. The outcome of this analysis yields information about the stability of the rankings based on a given measure.

4.1. CONSTRUCTION OF SIMULATED PORTFOLIOS

Starting from a known return generating process, we inject positive or negative abnormal returns stemming from the stock picking or market timing abilities of managers. The injection procedure guarantees that the use of the alpha, the *GTR* or the *IR* would produce rankings that are identical to the true rankings of the managers. Then, we introduce a pure noise component in the returns of each portfolio. This noise component corresponds to the sampling error. With these series of returns, we assess the quality of the performance measures by confronting them with generated data.

We construct two sets of 50 simulated portfolios whose monthly returns are measured from January 1994 to December 2003 (120 months). Each of these portfolios shares a common risk profile called the basis portfolio b. Its monthly return denoted r_{bt} is a weighted average of the excess return on the Weighted US Market index over the three-months T-Bill rate, the SMB and the HML factors, obtained from Kenneth French's website. To simplify the procedure, the weighting of each risk premium is such that the total monthly expected excess return $\overline{r}_b = 1\%$ and the contribution of each risk premium to the monthly return is 0.33%. As the average monthly risk premiums were 53.2 b.p., 11.7 b.p. and 4.8 b.p. respectively during this period, the resulting betas are set to 0.63, 2.84 and 6.92.[5]

Next, we assume that the managers differ either in their market timing for the first set of portfolios, or in their stock picking abilities for the second set. Market timing abilities are introduced in the following manner: each month, we observe whether the total return obtained for the basis portfolio is greater or lower than the 1% average. A manager with good market timing abilities will increase the beta of each component by a factor w^+ whenever the observed return at time t is above average, but will reduce her exposure by a fraction w^- otherwise. A bad market timer will systematically underweight good months and overweight bad ones. The weights are chosen so that the *GTR* of the portfolio is equal to a given performance level k and that, at the same time, the average required return of this portfolio is increased by the same factor k. Specifically, this imposes the following set of restrictions on the weights:

$$\sum_{t=1}^{120} w_t = k \text{ where } w_t = \begin{cases} w^+ \text{ if } r_{bt} \geq \overline{r}_b \\ w^- \text{ otherwise} \end{cases} \tag{9}$$

$$w^+ \sum_{t=1}^{120} \max\left(r_{bt} - \overline{r}_b, 0\right) - w^- \sum_{t=1}^{120} \max\left(\overline{r}_b - r_{bt}, 0\right) = k(1+k) \tag{10}$$

where the first condition amounts to increasing the average beta by a factor k and the second one to setting $\widehat{\alpha} = k(1+k)$, so that the *GTR* of the portfolio is equal to $k(1+k)/(1+k) = k$.

In this setup, managers will simply be ranked by their associated value of k. We create portfolios in decreasing order of performance from a monthly *GTR* of + 49 b.p. to −49 b.p. with a tick of 2 b.p. between each portfolio. All portfolios are directional, i.e., have a positive risk exposure, as their required monthly excess returns range from 51 to 149 b.p. Stock picking abilities are created in the same manner. Good stock pickers will attribute, each month, a higher weight to the risk premiums achieving a higher return than the mean and a lower weight to the risk premium that provides a below average return. Again, the sum of additional weights is equal to k. The conditions then become:

$$\sum_{\substack{i=Mkt,SMB,\\ HML}} \sum_{t=1}^{120} w_{it} = k \text{ where } w_{it} = \begin{cases} w^+ \text{ if } r_{it} \geq \overline{r}_{bt} \\ w^- \text{ otherwise} \end{cases} \tag{11}$$

$$w^+ \sum_{\substack{i=Mkt,SMB,\\ HML}} \sum_{t=1}^{120} \max\left(r_{it} - \overline{r}_{bt}, 0\right) - w^- \sum_{\substack{i=Mkt,SMB,\\ HML}} \sum_{t=1}^{120} \max\left(\overline{r}_{bt} - r_{it}, 0\right) = k(1+k) \tag{12}$$

[5] These values are chosen such that, for each premium, the product $\beta_j \overline{r}_j = 0.33\%$.

As for market timing ranked portfolios, managers are ranked by *GTR* values ranging from + 49 b.p. to −49 b.p. with a tick of 2 b.p. between each portfolio.

Given that investors only access the time series of portfolio returns, they are not able to retrieve the market timing or stock picking abilities of portfolio managers. Thus, to rank managers and compare their qualities, investors use an unconditional asset pricing specification.

In this context, by construction, the series behave in such a way that the *GTR* as well as the alpha of each portfolio, derived from the calibration of the original Fama and French (1993) three-factor model, is an exact measure of performance. As a result, the rankings produced by the use of alphas must be identical to the ones produced by the *GTR*.

In order to introduce idiosyncratic risk in each portfolio, we then add a random residual component to the monthly return. This component follows a uniform distribution in the interval [−5%, +5%].[6] As the source of this risk is not systematic, the rankings produced by the *IR* should also provide the same ranking as the *GTR* and the alphas, as the numerator of the *IR* is the alpha (which is monotonically increasing with k) and the denominator is the standard deviation of the residual risk, which is a monthly 2.89%. Table I reports the descriptive statistics of the performance values.

For both simulated data sets, the descriptive statistics show that the main differences between the distributions are to be found in the location and scaling of performance values. Proportionally to their ranges (max – min value) , the standard deviations of alpha and the *IR* are greater than the one of *GTR* for both series, except for the alpha series for market timing portfolios (ratio st. dev/range = 0.237 against 0.241 for *GTR*). Thus, within its sample bounds, the Generalized Treynor Ratio exhibits smaller (or at most equal) relative standard deviations than its competitors. This disadvantage serves our purposes of testing the precision and stability of the *GTR* as the finding of superior performance should not raise suspicion over manipulations in the sample construction.

4.2. NUMERICAL PRECISION AND STABILITY OF PERFORMANCE MEASURES

As the return generating process for our numerical simulations is strictly controlled, we can construct adequate tools to measure the precision and stability of performance measures. The procedures and results are reported in the next subsections.

4.2.1. *Precision*

By construction, the required return of the ith portfolio in our series must be $(1+k_i)$ where $k_i = -0.49 + 0.02(i-1)$, and the abnormal return of this portfolio must be

[6] The results obtained for the simulations are qualitatively extremely robust when the residuals follow a normal distribution or when the support of the uniform distribution changes.

Table I. Descriptive statistics for simulated portfolios

This table reports the summary statistics of performance measures associated with simulated portfolio monthly returns for the 1994–2003 period. The first (respectively second) part of the table reports results for a sample of 50 portfolios that exhibit different market timing (respectively stock picking) abilities of portfolio managers. We use the Fama and French (1993) three-factor model for the computation of alpha, *GTR* and *IR*. All figures are in percentages except the skewness and kurtosis.

	Mean	Median	Min	Max	S.D.	Skew.	Kurt.
Timing							
$\overline{r}$	1.06	1.00	0.10	2.34	0.62	0.19	2.07
$\widehat{\alpha}$	0.06	0.00	−0.58	0.87	0.34	0.12	2.64
GTR	−0.02	0.00	−0.83	0.59	0.34	−0.07	2.72
IR	0.04	0.00	−0.41	0.53	0.23	0.06	2.55
Picking							
$\overline{r}$	1.10	0.96	0.14	2.47	0.65	0.26	2.02
$\widehat{\alpha}$	0.11	−0.01	−0.42	1.01	0.39	0.18	2.50
GTR	0.02	−0.02	−0.75	0.68	0.36	0.02	2.23
IR	0.07	−0.01	−0.28	0.61	0.26	0.08	2.30

$k_i(1+k_i)$. With each performance measure, the corresponding theoretical ordering is known in advance: portfolios 1 to 50 are ranked from the worst to the best.

Therefore, any deviation in the ordering of the portfolios when the alpha, *GTR* or *IR* criterion is used must be due to the incapacity of the measure to extract the performance from the interference introduced by the noise component. The precision of a performance measure represents its ability to reproduce the "true" classification of portfolios when it is known, as is the case here. A simple way to measure this precision is to use the correlation between the series generated by each measure and the theoretical one. The full account for the correlation structure of our measures is reported in Table II.

We report the parametric Pearson product-moment correlation and the Spearman rank correlation coefficient. Both methods yield the same kind of qualitative results. The results for the alpha and the *IR* are very similar. The precision of the *GTR* is systematically higher than its competitors. The gap, although small in magnitude, is higher with the rank correlation than with the Pearson coefficient. Given the difference in skewness coefficients of the series of measures, the assessment of the precision of each measure is probably more reliable under the nonparametric method. Hence, in the scope of our simulations, the *GTR* appears to be more precise than the other performance measures when the abnormal performance of a portfolio is obtained at the expense of higher systematic risk.

Table II. Numerical precision of performance measures

This table reports the product-moment (Pearson) and rank (Spearman) correlation coefficients of, respectively, the series of average monthly returns and portfolio rankings between the individual performance measures of simulated portfolios and their true expected value for the 1994–2003 period.

	Pearson		Spearman	
Measure	Timing	Picking	Timing	Picking
$\widehat{\alpha}$	0.866	0.888	0.878	0.868
GTR	0.869	0.906	0.887	0.902
IR	0.867	0.887	0.877	0.867

4.2.2. *Stability*

The stability of a performance measure relates to its robustness to a change in the returns generating process. The stability of the performance measure is an important feature for empirical as well as practical matters. If one has omitted a factor or has chosen risk premiums that provide too poor explanatory power, all subsequent analysis will be corrupt unless the performance measure is relatively insensitive to the specification issue. This stability property is especially important in the context of persistence analysis, as the number and types of risks to be priced may change over time.

There are three interesting instances when the issue of stability can arise using alternative linear specifications: one model can be perfectly specified, i.e., it exactly corresponds to the returns generating process, while the alternative can be under-specified, i.e., at least one source of systematic risk is not properly priced; one can be perfectly specified while the alternative is over-specified, i.e., at least one risk premium is redundant with the others; or one is over-specified and one is under-specified.

In the simulation procedure, the proper returns generating process is the Fama and French (1993) three-factor model. We remove one risk premium, namely the HML premium in our application, to obtain an under-specified model, or add a fourth premium, namely the momentum factor UMD, to yield over-specification. Again, stability is measured through the product-moment and the rank correlation coefficients between paired series of performance measures. The results are reported in Table III.

The correlation coefficients between the three-factor and two-factor (resp. four-factor) model accounts for the stability of the measure when one of the models is under-specified (resp. over-specified), while the correlations between the four-factor and the two-factor models represent the effect of switching from an over-specified to an under-specified model.

Table III. Numerical stability of performance measures

This table reports the product-moment (Pearson) and rank (Spearman) correlation coefficients of, respectively, the average monthly returns and portfolio rankings between the individual performance measures simulated for the 1994–2003 period under the correct (three-factor), overspecified (four-factor) and underspecified (two-factor) models. We use the Fama and French (1993) model for the correct specification. The HML variable is removed for the two-factor model. Carhart's (1997) UMD variable is added for the four-factor model.

Panel A: Pearson product-moment correlations

	$\widehat{\alpha}$		*GTR*		*IR*	
	2-f	3-f	2-f	3-f	2-f	3-f
Timing						
three-factor	0.951		0.996		0.966	
four-factor	0.942	0.984	0.977	0.978	0.937	0.982
Picking						
three-factor	0.964		0.988		0.973	
four-factor	0.951	0.991	0.978	0.987	0.955	0.991

Panel B: Spearman rank correlations

	$\widehat{\alpha}$		*GTR*		*IR*	
	2-f	3-f	2-f	3-f	2-f	3-f
Timing						
three-factor	0.933		0.993		0.995	
four-factor	0.928	0.964	0.961	0.962	0.959	0.963
Picking						
three-factor	0.915		0.978		0.988	
four-factor	0.908	0.990	0.970	0.985	0.977	0.989

The results obtained for the three measures using the product-moment or the rank correlation coefficients are quite similar. Their behavior is quite different depending on whether one of the models is over- or under-specified. In the first case, all three measures remain very stable, with a very small advantage (ranging from 0.2% to 0.6% depending on the series and correlation measure) for alphas over *GTR*s. Information Ratios perform similarly to alphas.

The pattern is very different when one model is under-specified. With the Pearson correlation coefficient, the stability of the *GTR* exceeds those of *IR* and alpha by 3% and 4.5% (Timing) and 1.5% and 2.4% (Picking) respectively. The same pattern is observable when the four-factor and two-factor results are compared, with differences between the *GTR* and *IR* and alphas, respectively, ranging from 2.3% to 4%. Although the relative performance of the *IR* picks up when the Spearman coefficient is used, the alpha is much less stable than the *GTR*, with a difference ranging from 3.3% to 6.3%.

This outcome suggests that the problem of model under-specification is potentially more harmful to the stability of a performance measure than over-specification. Among these measures, the alpha is unambiguously the worst performer, especially when portfolio rankings are compared instead of performance values. The *IR* does not necessarily prove to be inferior to the *GTR* for ranking purposes.

From these simulations, we can thus conclude that there is no indication of inferior rating or ranking abilities of the *GTR* versus the alpha or the *IR* from a controlled experiment. If any winner is to be extracted from this analysis of precision and stability of performance measures, the Generalized Treynor Ratio probably emerges.

5. Conclusion

This paper has developed a new performance measure applied to multi-index models. Our Generalized Treynor Ratio remains a simple measure of performance that makes sense intuitively while strictly obeying the same properties as its one-dimensional counterpart. We have shown numerically that this measure is precise, i.e., it measures what it is supposed to measure, as well as stable, i.e., its rankings are robust to changes in the asset pricing models, provided that performance is attributable to the portfolio manager's exposure to systematic risk.

Important additional work remains to be done with this new performance metric. Our numerical experiment calls for some additional work focusing on the empirical properties of performance measures applied to multi-index models when confronted with real data. Furthermore, the adaptation of this framework to conditional asset pricing specifications is not straightforward and involves the development of a conditional distance measure. On the applied side, the use of this measure to analyze performance persistence of mutual or directional hedge fund managers is also likely to prove very fruitful. This list of extensions is part of our ongoing research agenda.

Appendix: Proof of Proposition 1

Let the Generalized Treynor Ratio (*GTR*) be the risk-adjusted performance measure that would fill the same role as the Treynor Ratio in a multivariate setup. This particular role entails that it respects at least the same basic properties as its univariate counterpart. Considering that for every portfolio, the sum of risk premiums is positive ($\sum_{j=1}^{K} \widehat{\beta}_{ij}\overline{r}_j > 0$), one can adapt the four properties (U1) to (U4) listed in Section 2:

- (G1) *Ratio of Euclidean distances:* $GTR(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, \overline{\mathbf{R}}) \equiv GTR_i$ is the ratio of an Euclidean norm in the returns space to a Euclidean norm in the K-dimensional risk hyperspace.
- (G2) *Homogeneity*
  - (G2a) *with respect to abnormal return (degree 1):* $GTR(k\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, \overline{\mathbf{R}}) = kGTR(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, \overline{\mathbf{R}})$ for $k > 0$
  - (G2b) *with respect to the vector of risk loadings (degree* -1*):* $GTR(\widehat{\alpha}_i, k\widehat{\mathbf{B}}_i, \overline{\mathbf{R}}) = k^{-1}GTR(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, \overline{\mathbf{R}})$ for $k > 0$
  - (G2c) *with respect to the vector of risk premiums (degree 0):* $GTR(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, k\overline{\mathbf{R}}) = GTR(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, \overline{\mathbf{R}})$ for $k > 0$
- (G3) *Proportionality*
  - (G3a) *to portfolio abnormal return:* $GTR(\widehat{\alpha}_p, \widehat{\mathbf{B}}_p, \overline{\mathbf{R}}) = (\lambda+(1-\lambda)k)GTR(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, \overline{\mathbf{R}})$ for $\overline{r}_p = \lambda\overline{r}_i+(1-\lambda)\overline{r}_{i'}$, $\widehat{\alpha}_{i'} = k\widehat{\alpha}_i$, $\widehat{\mathbf{B}}_{i'}\overline{\mathbf{R}} = \widehat{\mathbf{B}}_i\overline{\mathbf{R}}$.
  - (G3b) *inverse to portfolio required return:* $GTR(\widehat{\alpha}_p, \widehat{\mathbf{B}}_p, \overline{\mathbf{R}}) = (\lambda + (1-\lambda)k)^{-1}GTR(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, \overline{\mathbf{R}})$ for $\overline{r}_p = \lambda\overline{r}_i + (1-\lambda)\overline{r}_{i'}$, $\widehat{\alpha}_{i'} = \widehat{\alpha}_i$, $\widehat{\mathbf{B}}_{i'}\overline{\mathbf{R}} = k\widehat{\mathbf{B}}_i\overline{\mathbf{R}}$ and $(\lambda + (1-\lambda)k) > 0$
- (G4) *Reference portfolio*: $GTR(\widehat{\alpha}_i, \widehat{\mathbf{B}}_m, \overline{\mathbf{R}}) = \widehat{\alpha}_i$ for some reference portfolio m

In addition, the particular issue of a multi-factor model involves that three additional properties must be met by the Generalized Treynor Ratio:

- (G5) *Model irrelevance:* $GTR(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i^*, \overline{\mathbf{R}}^*) = GTR(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, \overline{\mathbf{R}})$, where $\widehat{\mathbf{B}}_i^*\overline{\mathbf{R}}^* = \widehat{\mathbf{B}}_i\overline{\mathbf{R}}$
- (G6) *Reducibility:* $GTR(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, \overline{\mathbf{R}}) = TR(\widehat{\alpha}_i, \widehat{\beta}_i)$ if $K = 1$
- (G7) *Parsimony:* Among all measures that respect conditions (G1) to (G6), the Generalized Treynor Ratio requires the lowest number and least repetition of parameters.

To have property (G1), the *GTR* should be written as a ratio of Euclidean distances. Therefore, the general analytical expression should look like the following:

$$GTR_i = \frac{f\left(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, \widehat{\mathbf{B}}_m, \overline{\mathbf{R}}\right)}{g\left(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, \widehat{\mathbf{B}}_m, \overline{\mathbf{R}}\right)}$$

where f and g are measures of Euclidean norms in the one-dimensional returns space and the K-dimensional risk space, respectively.

In the returns space, the Euclidean distance $f\left(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, \widehat{\mathbf{B}}_m, \overline{\mathbf{R}}\right) = \mathrm{D}(0, \widehat{\alpha}_i) = \widehat{\alpha}_i$ as the distance between the origin and a point on an axis is the coordinate of this point. As it is assumed that the returns generating process for each security i is a factor model, then one can write:

$$r_{it} = \sum_{j=1}^{K} \widehat{\beta}_{ij} r_{jt} + \varepsilon_{it} = \sum_{j=1}^{K} \widehat{b}_{ij} \pi_{jt} + \eta_{it} \tag{13}$$

for some set of properly scaled orthogonal risk premiums $\pi_1, \ldots, \pi_K$ such that $\sum_{j=1}^{K} \overline{r}_j = \sum_{j=1}^{K} \overline{\pi}_j$ and $\overline{\pi}_j > 0$, where $\mathrm{E}(\varepsilon_{it}) = \mathrm{E}(\eta_{it}) = \mathrm{E}\big(\pi_{jt}\eta_{it}\big) = 0$ for all j.

Consider the reference portfolio m. Using the same factor model as in (12), one writes:

$$\begin{aligned} r_{mt} &= \sum_{j=1}^{K} \widehat{\beta}_{mj} r_{jt} + \varepsilon_{mt} = \sum_{j=1}^{K} \widehat{b}_{mj} \pi_{jt} + \eta_{mt} \\ &= \sum_{j=1}^{K} \pi^*_{jt} + \eta_{mt} \end{aligned} \tag{14}$$

where $\pi^*_{jt} = \widehat{b}_{mj} \pi_{jt}$. Equation (13) allows to rescale each risk premium so that the corresponding beta of the reference portfolio is normalized to 1.

The realized return of security i is given by:

$$\overline{r}_i = \widehat{\alpha}_i + \sum_{j=1}^{K} \widehat{b}_{ij} \overline{\pi}_j \tag{15}$$

$$= \widehat{\alpha}_i + \sum_{j=1}^{K} \widehat{b}^*_{ij} \overline{\pi}^*_j \tag{16}$$

where $\widehat{b}^*_{ij} = \frac{\widehat{b}_{ij}}{\widehat{b}_{mj}}$.

As the factors are orthogonal, so is the K-dimensional space of factor risk loadings. To obtain an orthonormal space, every risk premium should be normalized to a single unit. Define the weight of factor j by:

$$\omega_j = \frac{\sqrt{K}\,\overline{\pi}^*_j}{\sum_{j=1}^{K} \overline{\pi}^*_j} \tag{17}$$

Using (16), Equation (15) can be rewritten as:

$$\overline{r}_i = \widehat{\alpha}_i + \sum_{j=1}^{K} \widehat{b}^*_{ij} \omega_j \left(\frac{\sum_{j=1}^{K} \overline{\pi}^*_j}{\sqrt{K}} \right) = \widehat{\alpha}_i + \sum_{j=1ij}^{K} \widehat{b}^{**}_{ij} \overline{\Pi} = \alpha_i + \Psi_i \overline{\Pi} \tag{18}$$

where $\overline{\Pi} = K^{-1/2} \sum_{j=1}^{K} \overline{\pi}_j^*$, $\Psi_i = \sum_{j=1}^{K} \widehat{b}_{ij}^* \omega_j = \sum_{j=1}^{K} \widehat{b}_{ij}^{**}$. Risk factors corresponding to coefficients $\widehat{b}_{ij}^{**}$ are orthogonal to each other and have identical magnitudes.

To respect property (G6), the *GTR* of all portfolios requiring the same return should provide the same ranking as the one obtained with their alpha. In particular, combining this with property (G2), this yields $GTR_i >$ (resp.$<, =$) GTR_p for a portfolio p where $\widehat{\alpha}_p >$ (resp.$<, =$) $\widehat{\alpha}_i$, $\Psi_p = \Psi_i$ and $\widehat{b}_{p1}^{**} = \widehat{b}_{p2}^{**} = \cdots = \widehat{b}_{pK}^{**} = \frac{\Psi_i}{K}$.

The Euclidean distance between 0 and the coordinates of this portfolio on the K-dimensional orthonormal risk axis is:

$$\begin{aligned} \mathrm{D}(0, \widehat{b}_{p1}^{**}, \widehat{b}_{p2}^{**}, \ldots, \widehat{b}_{pK}^{**}) &= \left(\sum_{j=1}^{K} \widehat{b}_{pj}^{**\,2} \right)^{1/2} = \left(\sum_{j=1}^{K} \left(\frac{\Psi_i}{K} \right)^2 \right)^{1/2} \\ &= \frac{\sum_{j=1}^{K} \widehat{b}_{ij}^* \overline{\pi}_j^*}{\sum_{j=1}^{K} \overline{\pi}_j^*} \end{aligned} \tag{19}$$

The last line follows from noticing that $\Psi_i = \frac{\sum_{j=1}^{K} \widehat{b}_{ij}^* \overline{\pi}_j^* \sqrt{K}}{\sum_{j=1}^{K} \overline{\pi}_j^*} > 0$.

From equations (15) and (12), $\sum_{j=1}^{K} \widehat{b}_{ij}^* \overline{\pi}_j^* = \sum_{j=1}^{K} \widehat{\beta}_{ij} \overline{r}_j$. Similarly, applying equation (14) to ex-post values yields $\sum_{j=1}^{K} \overline{\pi}_j^* = \sum_{j=1}^{K} \widehat{b}_{mj} \overline{\pi}_j = \sum_{j=1}^{K} \widehat{\beta}_{mj} \overline{r}_j$.

Setting that $g\left(\widehat{\alpha}_i, \widehat{\mathbf{B}}_i, \widehat{\mathbf{B}}_m, \overline{\mathbf{R}}\right) = \mathrm{D}(0, \widehat{b}_{p1}^{**}, \widehat{b}_{p2}^{**}, \ldots, \widehat{b}_{pK}^{**})$ leads to equation (6) that fulfills properties (G1) to (G6).

The Generalized Treynor Ratio involves $2K$ different betas repeated once, K risk premiums repeated twice, and 1 measure of abnormal returns. Consider another performance measure that fulfills properties (G1) to (G6). To respect properties (G1), (G2) (or (G3)) and (G4), it must include at least one measure of abnormal return, the K betas of portfolio i, and the K betas of portfolio m, respectively. Therefore, it can only be more parsimonious than the *GTR* thanks to a lower number of risk premiums. In order to be homogenous of degree 0 with the vector of risk premiums, the sum of powers of any scalar multiplier of the vector must be equal to 0. If any risk premium appears once, it should bear an exponent of 0 and the expression would violate property (G5). Therefore, no other performance measure that fulfills properties (G1) to (G6) is more parsimonious than the *GTR*. Hence, the *GTR* fulfills condition (G7). □

References

Carhart, M. M. (1997) On persistence in mutual funds performance, *Journal of Finance* **52**, 57–82.

Carpenter, J N. and Lynch, A. W. (1999) Survivorship bias and attrition effects in measures of performance persistence, *Journal of Financial Economics* **54**, 337–374.

Chen, Z. and Knez, P. J. (1996) Portfolio performance measurement: Theory and applications, *Review of Financial Studies* **9**, 511–555.

Chen, N.-F., Roll, R., and Ross, S. A. (1986) Economic forces and the stock market, *Journal of Business* **59**, 383–403.

Connor, G. and Korajczyk, R. A. (1986) Performance measurement with the arbitrage pricing theory. A new framework for analysis, *Journal of Financial Economics* **15**, 373–394.

Fama, E. F. (1972) Components of investment performance, *Journal of Finance* **27**, 551–567.

Fama, E. F. and French, K. (1993) Common risk factors in the returns of stocks and bonds, *Journal of Financial Economics* **33**, 3–56.

Ferson, W. E. and Schadt, R. W. (1996) Measuring fund strategy and performance in changing economic conditions, *Journal of Finance* **51**, 425–461.

Glosten, L. and Jagannathan, R. (1994) A contingent claims approach to performance evaluation, *Journal of Empirical Finance* **1**, 133–166.

Goodwin, T. H. (1998) The information ratio, *Financial Analysts Journal* **54**(4), 34–43.

Grinblatt, M. and Titman, S. (1989) Mutual fund performance: an analysis of quarterly portfolio holdings, *Journal of Business* **62**, 393–416.

Grinold, R. C. and Kahn, R. N. (1992) Information analysis, *Journal of Portfolio Management* **18**(3), 14–21.

Grinold, R. C. and Kahn, R. N. (1995) *Active Portfolio Management*, Richard D. Irwin, Chicago, Illinois.

Grinold, R. C. and Kahn, R. N. (2000) The efficiency gains of long-short investing, *Financial Analysts Journal* **56**(6), 40–53.

Henriksson, R. D. and Merton, R. C. (1981) On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills, *Journal of Business* **54**, 513–533.

Jensen, M. J. (1968) The performance of mutual funds in the period 1945–1964, *Journal of Finance* **23**, 389–416.

Jobson, J. D. and Korkie, B. M. (1984) On the Jensen measure and marginal improvements in portfolio performance: A note, *Journal of Finance* **39**, 245–251.

Kothari, S. P. and Warner, J. B. (2001) Evaluating mutual fund performance, *Journal of Finance* **56**, 1985–2010.

Kraus, A. and Litzenberger, R. H. (1976) Skewness preference and the valuation of risk assets, *Journal of Finance* **31**, 1085–1099.

Ledoit, O. and Wolf, M. (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance* **10**, 603–621.

Ledoit, O. and Wolf, M. (2004) Honey, I shrunk the sample covariance matrix, *Journal of Portfolio Management* **30**(4), 110–119.

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

Modigliani, F. and Modigliani, L. (1997) Risk-adjusted performance, *Journal of Portfolio Management* **23**(2), 45–54.

Modigliani, F. and Pogue, G. A. (1974) An introduction to risk and return – II, *Financial Analysts Journal* **30**(3), 69–86.

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

Sharpe, W. F. (1964) Capital asset prices: A theory for market equilibrium under conditions of risk, *Journal of Finance* **19**, 425–442.

Sharpe, W. F. (1966) Mutual fund performance, *Journal of Business* **39**, 119–138.

Sharpe, W. F. (1992) Asset allocation: Management style and performance measurement, *Journal of Portfolio Management* **18**(2), 29–34.

Sharpe, W. F. (1994) The Sharpe ratio, *Journal of Portfolio Management* **21**(1), 49–58.

Treynor, J. L. (1961) Toward a theory of market value of risky assets, Mimeo, subsequently published in R. A. Korajczyk (ed.) (1999) *Asset Pricing and Portfolios Performance: Models, Strategy and Performance Metrics*, Risk Books, London.

Treynor, J. L. (1966) How to rate management investment funds, *Harvard Business Review* **43**, 63–75.

Treynor, J. L. and Black, F. (1973) How to use security analysis to improve portfolio selection, *Journal of Business* **46**, 66–86.

Treynor, J. L. and Mazuy, K. (1966) Can mutual funds outguess the market? *Harvard Business Review* **43**, 131–136.