Investigaciones Económicas. Volumen XVII (1) - Enero 1993, pp. 3-41

PERFORMANCE MEASUREMENT OF MANAGED PORTFOLIOS: A SURVEY

Gonzalo RUBIO*
Universidad del País Vasco

Este trabajo presenta un panorama de las técnicas de evaluación de la gestión de carteras. El artículo enfatiza los aspectos críticos de las técnicas tradicionales y analiza nuevas alternativas. Se discuten las dificultades conceptuales de la separación entre la calidad de la información relacionada con la selección de activos individuales y con la sincronización del mercado.

Asimismo, se discuten brevemente los resultados empíricos obtenidos a partir de las técnicas presentadas anteriormente. Dicha evidencia parece indicar que los resultados son robustos a cambios en las carteras de referencia más razonables, así como a las técnicas empleadas.

1. Introduction

Performance evaluation of managed portfolios is a key issue of financial economics. Of course, this should not be surprising. Taking into account the important volume of resources that is being allocated through investment companies in the major stock exchanges in the world, it becomes natural to investigate whether an important component of these companies —the mutual fund— is being professionally managed and is efficiently using the resources available. On the other hand, it seems rather obvious that a prerequisite to any systematic evaluation consists of having an understanding of the techniques available to measure the performance of managed portfolios, and the critical issues involved in the evaluation. The objective of this paper is to present a technical survey of these measurements, their potential problems in empirical applications, and their economic justification.

From the point of view of individual clients, the ability to distinguish the performance of competitive funds should be the main factor in altering the

* Most of the work reported in this paper was prepared whilst the author was Visiting Professor at the University of California, Berkeley. I appreciate helpful comments from two anonymous referees and from Greg Connor, David Modest and Rafael Repullo. Gil Mehrez provided excellent research assistance. I am also grateful to Eudal Canadell and Verónica Grandón for helping with the data. Finally, I appreciate the financial support provided by the Comisión Nacional del Mercado de Valores and the Berkeley Program in Finance. The usual disclaimer applies.

amount of resources allocated to different portfolio managers. In addition, the manager can benefit from continuous evaluation by identifying the sources of strength and weakness in his or her management process.

Evaluation of performance is also interesting from the efficient market hypothesis point of view. It is well known that, according to the efficient market hypothesis, prices should reflect all available information. The so called «strong form» of the efficient hypothesis suggests that portfolio managers should gain just enough return to be compensated for bearing risk and for the costs associated with their activities.

Lastly, the tasks assigned to the regulatory agencies of stock markets in terms of supervision and control of market activities require an understanding of the evaluation process of investment companies. Given the acceptance of a framework for continuous evaluation, the regulatory and legal system for investment companies may more efficiently accomplish its role of providing the fundamental competitive safeguards that prevent abuse of individual investors.

All these reasons justify a survey about the performance measurement techniques available in the literature.

The common methodology employed in most previous studies is mainly based on estimating the average returns of the funds during some sample period. These returns are generally adjusted for the behavior of the market during that period, and by some measure of risk. Hence, performance is and should be evaluated on a relative basis, and not on absolute basis. This is to say, performance is usually measured relative to some definition of what is meant by the market. Moreover, it is generally the case that we are interested in determining not only the existence of superior performance, but also in distinguishing what part is due to the manager's capacity in selecting appropriate assets, and what portion is due to his or her skill in changing the portfolio holdings through time, so that the risk level of the overall portfolio is adjusted to market conditions at appropriate times. Hence, ideally we would like to have a measure of selectivity and a measure of market timing.

Unfortunately, we encounter two main difficulties in interpreting the results obtained when we work along the lines presented above. Firstly, it must be pointed out that the traditional measures of risk and return are not necessarily correct in a framework of asymmetric information between the manager and the evaluator. It is interesting to note that these measures developed primarily to evaluate mutual funds cannot properly address the issue of asymmetric information, which is a major consideration when we try to justify the existence of these companies. This is to say, it is certainly reasonable to identify superior performance with superior information. However, the traditional measures of risk and return may present important biases if the possibility of superior information is recognized. In this context, the composition of the portfolios evaluated will surely change over time. It is not clear, therefore, that a single estimable parameter of risk is the relevant or appropriate measure of the true risk known only to the manager.

Secondly, when we establish a ranking of funds based on the traditional measures of risk and return, it is not clear whether this ranking is an indication of true performance or whether it just reflects the position of the bechmark employed in the study within the well known mean-standard deviation diagram. The mean variance efficiency of the benchmark used in the empirical application becomes a relevant issue in order to interpret unambiguously the results.

This discussion implies basically that to know the actual composition of the portfolios period by period facilitates enormously the task of performance evaluation. In theory, it becomes the most important data in performance evaluation. Unfortunately, it is extremely difficult to obtain series of monthly holdings long enough to provide a reasonable basis for statistical tests.

Finally, an additional difficulty is related to the ability of distinguishing between skill and luck. It is not always clear that the availability of techniques in order to evaluate managed portfolios would provide a clear picture of the reasons of either success or failure. An the end, the question of persistence of performance becomes an empirical issue which should be explicitly debated in any research related to performance evaluation.

This paper is organized as follows: the next section reviews the traditional measures of performance. Section three describes the problems associated with them. Some recent alternatives developed to overcome some of these problems are discussed. The fourth section contains a brief presentation of the empirical results obtained in the Spanish mutual fund industry when the previously discussed techniques are implemented. Finally, we conclude with a summary and discussion of the main ideas developed throughout the paper.

2. Risk-Adjusted Measures of Performance

2.1. Asset pricing models

Financial equilibrium in the context of modern asset pricing theory requires that the expected return of risky assets be linearly related to the appropriate measures of non-diversifiable risk and risk premia on one or more relevant state variables. Thus, we find asset pricing models with returns related to one measure of systematic risk or, alternatively, models in which asset pricing is dominated by multiple measures of exposure to systematic risk.

By assuming that the rates of return on risky assets are generated by a multiple factor model, we are able to derive the Arbitrage Pricing Theory (APT) of Ross (1976) and Connor (1984). On the other hand, by imposing more structure on the preferences of investors and/or the distributional characteristics of returns, one can obtain the traditional Capital Asset Pricing Model (CAPM) of Sharpe (1964)-Lintner (1965)-Mossin (1966)-Black (1972) and the intertemporal models of Merton (1973), Rubinstein (1976), Breeden (1979) and Cox, Ingersoll and Ross (1985).

These models represent a fundamental step towards the implementation of performance measures. In particular, these models are able to offer sharp predictions about the nature of risk premiums for risky assets. In fact, the differences among them are mainly due to the way in which the risk premium is obtained.

In the context of the CAPM, market equilibrium requires that the true value-weighted market portfolio be an efficient portfolio. This indicates that the mathematical conditions to any efficient portfolio necessarily apply to the market portfolio. Therefore, there is a linear relationship between the expected return of an asset and the covariance of its return with the return of the market portfolio [1].

$$E(r_i) = E(r_z) + [E(r_m) - E(r_z)]\beta_i \quad ; \quad i = 1, \ldots, N \tag{1}$$

where,

$E(r_i)$ is the expected excess return on asset i with respect to the riskless lending rate;

$E(r_m)$ is the expected excess return on the market portfolio;

$E(r_z)$ is the expected excess return on the frontier portfolio with zero correlation with respect to the market portfolio;

β_i is $\text{cov}(r_i, r_m) / \text{var}(r_m)$; and

N is the number of assets.

When riskless lending and borrowing opportunities are assumed, the traditional version of the CAPM predicts that:

$$E(r_z) = 0 \quad \text{and} \quad E(r_m) > 0$$

and the risk premium on risky assets are proportional to their market betas.

On the other hand, when a riskless lending and borrowing rate is not assumed, the constrained version of the CAPM predicts:

$$E(r_z) \geq 0 \quad \text{and} \quad [E(r_m) - E(r_z)] > 0$$

The model says that there is (in equilibrium) a linear and positive relationship between the expected return on any risky asset and nondiversifiable risk or beta risk. This is the only risk priced by the market. Idiosyncratic or unique risk can be diversified away. For this reason, beta risk is also known as syste-

[1] The presentation of equations [1] through [4] could have been stated in terms of conditional expectations relative to the information set available to the representative consumer in the optimization problem, Φ_{t-1}. By the law of iterated expectations, it is possible to take the expectation with respect to any information set subset of Φ_{t-1}. This implies that we can take the expectation with respect to the subset Φ_0, so that the equations are written in terms of unconditional expectations.

matic risk. It is the only variable needed to explain cross-sectional differences in expected stock returns. Note that beta is the sensitivity relative to a single common factor, the market return.

On the other hand, Ross (1976) argued that systematic risk need not be adequately represented by a single common factor such as the return on the market. On the contrary, the APT assumes that returns are generated by a linear K-factor process:

$$R_i = E(R_i) + \sum_{k=1}^{K} b_{ik} f_k + \varepsilon_i \qquad [2]$$

where,

$E(R_i)$ is the expected return on asset i;

f_k is the common factor affecting all individual assets;

b_{ik} is the sensitivity (or the factor loading) of asset i to the movements in common factor k; and

ε_i is the zero-mean nonsystematic (or idiosyncratic) return component of asset i with finite variance[2].

It should be noted that equation [2] is based upon a very natural assumption. It says that the return we observe ex-post has two components. The first one is its expected value while the second is due to the unanticipated component of the return. This surprise (innovation) also has two components. The first one is related to some common factors affecting all individual securities in different degrees, whilst the second one is specifically related to each individual security. The sensitivities to the common factors represent the degree in which different individual securities are affected by them.

If the nonsystematic component were zero for each asset, $\varepsilon_i = 0$, and given that the returns are linearly related by equation [2], it can be shown that nonarbitrage arguments guarantee that the expected returns also have a linear relationship:

$$E(R_i) = \lambda_0 + b_{i1}\lambda_1 + \ldots + b_{iK}\lambda_K \qquad [3]$$

In this case, expected returns are exactly and linearly related to multiple measures of exposure of systematic risk, where λ_0 is a constant and $\lambda_1, \ldots, \lambda_k$ are the factor risk premia.

Unfortunately, once we recognize that the idiosyncratic component of returns are nonzero, pure arbitrage pricing theory does not give an exact pricing

[2] When we assume that the idiosyncratic returns are uncorrelated across securities, we have the so-called «strict factor model», to be distinguished from the «approximate factor model» where correlations among residuals are allowed.

model. This implies that the following «approximately equal» relationship holds:

$$E(R_i) \approx \lambda_0 + b_{i1}\lambda_1 + \ldots + b_{iK}\lambda_K \quad [4]$$

This relation should price most assets with negligible error but, unfortunately, need not price all securities arbitrarily well. What the original APT model says is that the sum of squared deviations from an exact relationship does not go to infinity with the number of assets. In other words, in large N economies deviations are bounded for most assets, but can be very large for any particular asset. It is the case that a small number of securities could be priced arbitrarily badly.

It must be said that equilibrium requirements are crucial in distinguishing both the CAPM and APT perspectives presented above. Interestingly, equilibrium versions of the APT have also been developed in the literature. Once again, they attempt to provide a linear relationship between expected returns and the factor betas. Both Dybvig (1983) and Grinblatt and Titman (1983) are able to develop a utility based equilibrium model that results in an approximate linear relation. It is interesting to note that they are able to provide with economic content the deviations from the exact linear relationship. Finally, Connor (1984) obtains an exact equilibrium APT by assuming that the equilibrium allocations of individuals do not incorporate any idiosyncratic risk. It turns out that this assumption implies that in finite economies there is only one optimal well-diversified portfolio held by all individuals. Of course, this is the market portfolio, so that this condition becomes very similar to the ideas behind the CAPM. In other words, individuals' marginal utilities are not influenced by idiosyncratic risk. To conclude, an exact linear relationship exists whenever the efficient frontier contains at least one portfolio that is fully diversified. According to the exact equilibrium K-factor APT, some weighted average of K portfolios playing the role of the K factors intersects the global efficient frontier. Given that the weighted average of those portfolios contains no idiosyncratis risk, the intersection point with the global efficient frontier has just systematic risk. Exactness must be achieved. Note that in the pure arbitrage APT, there was no economic reason leading to such an intersection [3].

To summarize, the APT explicitly recognizes the possibility of families of portfolios with the same CAPM beta, but having at the same time different sensitivities to other possible sources of systematic risk. In this sense, it becomes a multi-beta model of asset valuation. Unfortunately, we are not able to theoretically identify these common factors. Economic intuition suggests, however, that macroeconomic variables such as inflation, productivity, term structure or default risk may be reasonable candidates for the unidentified common factors.

Both the CAPM and the APT models of asset valuation play an important role in the performance measurement techniques to be discussed in this paper.

[3] See Grinblatt and Titman (1987).

2.2. *Traditional measures of performance*

a) THE JENSENS'S ALPHA

Accepting that we are able to measure the periodic returns for a portfolio during an evaluation interval, we still need the portfolio's risk level. Assuming that the client is a well diversified investor, the appropriate measure of risk is the beta coefficient or non-diversifiable risk presented in the previous section.

In order to implement empirically the CAPM, several difficulties must be overcome. Equation [1] is expressed in terms of expectations. The ex-post version of the CAPM involves the assumption of rational expectations; i. e., the realized returns for a particular time period are drawings from the ex-ante probability distributions of returns on those assets. Therefore, it can be written:

$$R_{it} = R_{ft} + (R_{mt} - R_{ft})\beta_i + \varepsilon_{it} \qquad [5]$$

which is the ex-post version of the CAPM, and where R_f is the rate of return of the risk-free asset and ε_i is the zero-mean idiosyncratic return. It should be pointed out that a similar ex-post version of the APT can be obtained.

If we regress the excess return (over the risk-free rate) of mutual fund p, on an intercept and the excess return of the market, we have the following expression[4].

$$R_{pt} - R_{ft} = \alpha_p + \beta_p(R_{mt} - R_{ft}) + \varepsilon_{pt} \qquad [6]$$

where a portfolio's alpha is the excess return that would be expected on the mutual fund if the excess return on the market portfolio were zero. Note that under this condition, on average, and according to the CAPM, this mutual fund should have gained the risk free-rate. This implies that intuitively a positive alpha would be desirable and a negative alpha undesirable.

The characteristic line has two components. The first one, $\alpha_p + \varepsilon_{pt}$, is the nonmarket component of mutual fund p's excess return, whilst the second one, given by $\beta_p(R_{mt} - R_{ft})$, is the market component of mutual fund p's excess return. It should be clear that α_p is the expected value of the nonmarket component, and ε_{pt} is the deviation of the observed nonmarket component from its expected value. It turns out that the standard deviation of ε_p is the appropriate measure for the idiosyncratic risk of portfolio p.

This discussion may be summarized by noting that α_p is the difference between the observed return of the fund and the return this fund should have obtained for its level of beta risk. Hence, on average, a positive (negative) α_p

[4] It is known as the «characteristic line».

seems to imply superior (inferior) performance. This coefficient is known as the Jensen's alpha[5].

It is interesting to note that the alpha of a portfolio in each moment of time is just a weighted average of the individual alphas[6].

$$\alpha_p = \sum_{i=1}^{N} w_i \alpha_i \qquad [7]$$

where w_i represents the fraction of the portfolio held in asset i, and N is the number of assets in the portfolio. Equation [7] suggests that superior (inferior) selectivity of individual assets by the manager would imply superior (inferior) portfolio performance.

b) THE TREYNOR RATIO[7]

It measures the mutual fund's average excess return per unit of beta risk. Once again, nondiversifiable risk becomes the appropriate measure of risk. It can be expressed as:

$$IT_p = \frac{\bar{R}_p - R_f}{\hat{\beta}_p} \qquad [8]$$

where $\bar{R}_p$ is the average return of the fund over the sample period and $\hat{\beta}_p$ is the estimate of the systematic risk of portfolio p given by [6]. Over the sample period, the CAPM suggests that, on average, the following should hold:

$$\bar{R}_p = R_f + (\bar{R}_m - R_f)\,\hat{\beta}_p \qquad [9]$$

where $\bar{R}_m$ is the sample average of the market portfolio return. Equation [9] implies that,

$$IT_p = \frac{\bar{R}_p - R_f}{\hat{\beta}_p} = \bar{R}_m - R_f \qquad [10]$$

[5] See Jensen (1968). Connor and Korajczyk (1986) developed a similar measure under the APT framework. They showed that the Jensen coefficient is theoretically compatible with the exact competitive equilibrium version of the APT.

[6] Note that the framework employed in this section is static. This implies that the decomposition of Jensen's alpha in equation [7] only makes sense when the portfolio weights are constant. When applying these equations to an intertemporal evaluation of mutual funds, we will encounter serious difficulties in interpreting the performance of dynamic strategies followed by most mutual funds. More will be said later about this issue.

[7] See Treynor (1965). It is also known as the «reward to volatility ratio». Given that the interest of the techniques presented in the next two subsections is basically empirical, it was decided to discuss them in terms of sample moments rather than population moments.

Over any given interval, the fund's average excess return may deviate from the market's excess return. This implies that equation [9] becomes,

$$\bar{R}_p = R_f + (\bar{R}_m - R_f)\hat{\beta}_p + d_p \qquad [11]$$

where d_p is the fund's deviation for the period. If $d_p > 0$, the investment will be successful since $(\bar{R}_p - R_f) / \hat{\beta}_p > (\bar{R}_m - R_f)$.

It should be clear that both the Jensen's alpha and the Treynor ratio are very closely related. We can estimate alpha directly from [9]:

$$\hat{\alpha}_p = (\bar{R}_p - R_f) - \hat{\beta}_p(\bar{R}_m - R_f) \qquad [12]$$

dividing both sides of [12] by $\hat{\beta}_p$ and rearraging,

$$\frac{\bar{R}_p - R_f}{\hat{\beta}_p} = \frac{\hat{\alpha}_p}{\hat{\beta}_p} + (\bar{R}_m - R_f) \qquad [13]$$

The left hand side of [13] is the Treynor ratio. Since superior performance implies that $(\bar{R}_p - R_f) / \hat{\beta}_p > (\bar{R}_m - R_f)$, $\hat{\alpha}_p/\hat{\beta}_p$ must also be positive. Generally, the estimate of beta is nonnegative, which indicates that superior performance by Treynor implies superior performance by Jensen's alpha and vice versa. However, it might be possible to find inconsistencies between both measures when ranking of mutual funds is pursued.

c) THE SHARPE RATIO [8]

It measures the mutual fund's average excess return per unit of total risk. In this case, we allow for the possibility of measuring performance by considering that the fund is not perfectly well diversified. As we have already discussed, both systematic and idiosyncratic risk are relevant if the final holding is not a well diversified portfolio. It can be expressed as:

$$IS_p = \frac{\bar{R}_p - R_f}{\sigma_p} \qquad [14]$$

where σ_p is the standard deviation (total risk) of the portfolio's return. In order to define superior performance, equation [14] is compared to the equivalent ratio using the average market return and the market standard deviation. It is important to note that if the Treynor ratio (or alpha) indicates that the porfolio outperformed the market, it is possible for the Sharpe ratio to indicate that the portfolio did not perform as well as the market. The reason should be clear. The mutual fund may have a relatively large amount of idiosyncratic risk.

[8] See Sharpe (1966). It is also known as the «reward to variability ratio». See also Fama (1972) for a related treatment of performance.

d) MARKET TIMING

— The Quadratic Approach to Timing

The basic idea behind the most popular measure of performance to date, the Jensen's alpha, is that portfolio managers must select securities that outperform other of comparable nondiversifiable risk. However, there is an alternative way to achieve superior performance. The manager may switch from risk class at appropriate times. More specifically, the manager should hold a high beta portfolio prior to market rises and a low beta portfolio prior to market declines. This behavior is known as market timing. It turns out that if the manager has the necessary ability to time the market, his or her success should eventually be reflected in a positive (ex-post) performance[9].

This strategy might be followed by either switching among stocks with different betas or by changing the proportions invested in stocks and bonds. In any case, overall performance is usually separated between security selection and market timing. The possibility of distinguishing both components has been an important issue in modern financial economics. As was pointed out by Admati, Bhattacharya, Pfleiderer, and Ross (1986) (ABPR), this should not be surprising. There are sound economic reasons behind the importance of the distinction between the two sources of performance. First of all, by differentiating selectivity and timing, it is possible to achieve a much better idea of the service offered by mutual funds. Secondly, the resulting distribution of asset returns may be different depending on whether private information about market aggregates dominates private information about individual firms.

In order to understand the traditional techniques employed in estimating timing ability, we should realize that the actual observations around the fitted regression line given by [6] will have a curved (convex) shape. If the manager has timing ability, when the market increases substantially, the fund will tend to do better that it would have otherwise done. The observations will tend to be concentrated above the ex-post characteristic line. When the market declines, the fund will tend to decline less than it would have otherwise declined. Once again, most of the observations will be above the fitted regression line.

In order to capture this curved shape in the observations, Treynor and Mazuy (1966), Pfleiderer and Bhattacharya (1983), and ABPR (1986) have proposed the following quadratic regression:

$$r_{pt} = \alpha_p + \eta_{1p} r_{mt} + \eta_{2p}(r_{mt})^2 + \omega_{pt} \qquad [15]$$

where, $r_{pt} = R_{pt} - R_{ft}$, and $r_{mt} = R_{mt} - R_{ft}$.

[9] This is a very delicate statement. In principle, it seems clear that good timing should imply positive performance. However, it does not necessarily imply that positive performance would be captured through the particular specification of the model employed to measure total performance. As we will see later, for a given empirical application (for example using Jensen's alpha), issues related to asymmetric information between the manager and the evaluator may invalidate the statement in the text.

The basic argument behind [15] suggests that the addition of the squared term (which results in a curved shape) will improve the fit of the least squares regression. If market timing is successful, the coefficient, η_{2p}, should be positive and statistically different than zero.

Unfortunately, there are some subtle, yet serious difficulties with some of the arguments presented above. The main idea is that superior performance should be identified with superior information. Intuitively, this idea implies that we need to measure the quality of private information possessed by the manager. The precision of the signal received by the manager becomes the main issue. By simply running the regression suggested above, it is possible to erroneously identify aggressiveness with quality of information [10]. It can be shown that by trading aggressively, a portfolio manager without significantly good information may be able to replicate some of the performance achieved by a manager with better information.

For both selectivity and timing, it is important to recognize that performance should be evaluated on the basis of the quality of information. The value added by a manager crucially depends on the precision of his or her information [11].

It turns out that, in principle, we are able to estimate the quality of timing information without knowing the quality of selectivity information. At the same time, however, it is only possible to infer whether there exists selectivity information. Following both Pfleiderer and Bhattacharya (1983) and ABPR (1986), it becomes necessary to specify how managers react to information. It is assumed that the response is linear which is consistent with the maximization of expected utility by a constant absolute risk averse individual. This assumption is very common in the literature on asymmetric information. Appendix A presents a detailed and technical analysis of the issues involved.

It should also be mentioned that the possibility of observing the quality of timing information is embedded in the traditional definitions of timing and selectivity. As it can be observed from either [6] or]15], selectivity is statistically independent of the market return or, more generally, the timing portfolio return. Once again, this holds just by definition. In this way the problem of distinguishing both concepts is clearly simplified.

For practical purposes, however, it may be difficult to accept that both issues are not related to each other. In fact, ABPR (1986) formally showed that selectivity and timing cannot be conceptually different.

The interesting implication of this discussion is that precisely because of the inherent problems of this approach, it becomes plausible to detect the existence of selectivity and the quality of timing information.

[10] Aggressiveness measures the degree of the reaction of managers to new information, without specifying the quality of that information. It may be captured through the risk aversion coefficient. See Appendix 1.

[11] See Pfleiderer and Bhattacharya (1983), Admati and Ross (1985), and ABPR (1986).

The conceptually correct approach involves using some of the ideas behind the APT. It is not necessary, however, to use a particular asset pricing model to identify the quality of information. By doing so, we avoid addressing the issue of aggregation. Under this approach, timing information is about the realizations of factors that influence the returns of many assets, without becoming necessary to prespecify a timing portfolio. On the other hand, selectivity information allows for information to be anything specific about individual asset returns without restriction. Unfortunately, because of the large number of interactions between information signals and asset returns, this approach becomes unfeasible from the estimation point of view. In other words, although the second approach has conceptual advantages over the traditional approach, testing is possible only imposing independence between selectivity and timing [12].

As a summary of the subtile points discussed above, it should be first emphasized that even if we accept the usual distinction between selectivity and timing, and the model represented by the quadratic regression, the coefficient η_{2p} in [15] does not provide a quantitative measure of the quality of timing information. On the other hand, however, and as shown by Appendix A, this measure may be retrieved from a further regression based on the same quadratic specification. Secondly, the usual distinction between selectivity and timing is conceptually inappropriate, as they should not be statistically independent [13].

— Market Timing and Options

There is a second way of dealing with market timing. The papers by Merton (1981), Henriksson and Merton (1981), and Henriksson (1984) presented an approach based on the ideas behind option pricing.

In the basic model, the manager forecasts when stocks will outperform riskless assets and when riskless assets will outperform stocks. However, the manager does not predict the magnitude of the relative returns. In fact, up markets are defined as periods for which the market return is greater than or equal to the risk-free rate, and down markets as periods where the market return is lower than the riskless rate [14].

In order to be a successful timer, a manager should have a higher beta position in the up market than in the down market. In other words, the slope of

[12] The traditional approach and the «correct» approach are known as the portfolio and factor approach respectively.

[13] I would like to thank one the referees for providing a clean summary of the ideas of this subsection. Moreover, even if we are not able to directly estimate a quantitative measure of the quality of timing information, testing whether $\eta_{2p} = 0$ is still useful. If the null hypothesis cannot be rejected, the implication would be consistent with no timing ability at all.

[14] This is less general than the model presented in Appendix A to motivate the quadratic regression.

the ex-post characteristic line for up markets should be greater than the slope for down markets. Obviously, two linear regressions could be run to see whether the slope coefficients are different using data for up markets in the first regression and data for down markets in the second regressions. It is interesting to observe that successful timing would imply a nonlinear payoff when both fitted regressions lines are graphed together. Of course, nonlinear payoffs, are the type of payoffs produced by options.

It turns out that two separate regressions are not needed. These ideas can be captured through one single regression with a dummy variable, D_t, which is equal to zero if $R_{mt} \geq R_{ft}$, and it takes a value of -1 if $R_{mt} < R_{ft}$:

$$R_{pt} - R_{ft} = \alpha_p + \gamma_{1p}(R_{mt} - R_{ft}) + \gamma_{2p}[D_t(R_{mt} - R_{ft})] + \omega_{pt} \qquad [16]$$

If $D_t = -1$, the equation [16] becomes:

$$\begin{aligned} R_{pt} - R_{ft} &= \alpha_p + \gamma_{1p}(R_{mt} - R_{ft}) - \gamma_{2p}(R_{mt} - R_{ft}) + \omega_{pt} \\ &= \alpha_p + (\gamma_{1p} - \gamma_{2p})\,(R_{mt} - R_{ft}) + \omega_{pt} \end{aligned} \qquad [17]$$

This is like the usual regression with $\beta = \gamma_{1p} - \gamma_{2p}$. In fact, the difference between both coefficients is the portfolio's down-market beta.

If, on the other hand, $D_t = 0$, equation [16] becomes:

$$R_{pt} - R_{ft} = \alpha_p + \gamma_{1p}(R_{mt} - R_{ft}) + \omega_p \qquad [18]$$

In this case, $\beta = \gamma_{1p}$ is the portfolio's up-market beta.

For timing to be successful, it is required that $\gamma_{1p} > \gamma_{1p} - \gamma_{2p}$ which implies that γ_{2p} is the difference between the up-market beta and the down-market beta. If the manager is able to successfully time the market, the coefficient, γ_{2p}, in [16] must be positive.

Given this discussion, it can be easily verified that the regression [16] may alternatively be written as:

$$R_{pt} - R_{ft} = \alpha_p + \gamma_{1p}(R_{mt} - R_{ft}) + \gamma_{2p}\max(0, R_{ft} - R_{mt}) + \omega_{pt} \qquad [19]$$

The second explanatory variable, $\max(0, R_{ft} - R_{mt})$, is exactly the payoff of a European put option on the market with exercise price equal to the risk-free rate.

More formally, Merton (1981) established an isomorphism between the market timing strategy and a strategy consisting of a passive mix of stocks and risk-free assets, together with a number of options to exchange one asset for another. What gives an option to exchange assets considerable interest is that it is equivalent to the option to obtain the greater of the appreciation of two assets. Of course, these two assets could be the market and the risk-free asset.

In particular, consider the following case. Suppose you can invest in a mutual fund which can infallibly predict whether the market porfolio will have a rate of return over the next period which will be greater or less than the interest rate. If the fund predicts that the market will appreciate faster than the interest rate, it secretly invest all of its assets in the market; otherwise, it secretly invests everything in Treasury Bills. Let F_0 and F_T be the current value and the end-of-period value of the fund respectively.

It turns out that the value of this perfect forecast is equal to a riskless investment of F_0 plus a purchased call option on the market, with the exercise price equal to $(1+R_f)F_0$, or alternatively equal to an investment of F_0 in the market plus a purchased put option on the market with exercise price of $(1+R_f)F_0$. To see this, it should be noted that the mutual fund is promising max $[F_T, (1+R_f)F_0]$ which can be decomposed in two ways:

$$\max[F_T, (1+R_f)F_0] = (1+R_f)F_0 + \max[0, F_T - (1+R_f)F_0]$$
$$\max[F_T, (1+R_f)F_0] = F_T + \max[0, (1+R_f)F_0 - F_T] \quad [20]$$

It should be mentioned that, in principle, it is also possible to use the Henriksson-Merton (1981) approach with a multiple factor APT bechmark. It would give rise to an estimable relation between the excess return of each fund and a put option payoff on each of the factors[15].

Both approaches to market timing have been extensively employed in different empirical applications[16]. It might be relevant to point out that the coefficients from least squares estimation of [19] and [15] produce consistent estimates of both the selectivity and timing parameters of performance. The model suggests that α_p captures selectivity information, while γ_{2p}(and η_{2p}) reflects timing information. However, as Henriksson and Merton (1981) point out, the structure of the theoretical model implies that beta is not stationary. In particular, as the authors show, this makes the standard deviation of ω_{pt} to be an increasing function of $|r_{mt}|$. A correction for heteroskedasticity to improve the reliability of the statistical inference becomes necessary. The same type of reasoning may be applied to the quadratic approach to market timing[17].

3. Problems of Risk-Adjusted Performance Measures and New Alternatives

3.1. Asymmetric information

In the last section, it was argued that the usual methodology in evaluating performance depends heavily upon the use of the following regression:

$$r_{pt} = \alpha_p + \beta_p r_{mt} + R_{pt} \quad [21]$$

[15] A similar argument applies to the quadratic approach to timing.

[16] See Rubio (1992). Moreover, the conceptual limitations of the usual distinction between selectivity and timing information discussed before also apply to the Henriksson-Merton approach.

[17] In his empirical application of these techniques, Rubio (1992) employs White's (1980) consistent covariance matrix estimator.

where r_{pt} and r_{mt} are the excess returns (over the risk-free rate) of the fund and the market respectively.

In a very influential article, Dybvig and Ross (1985a) showed that a well informed manager who makes the adequate timing decisions between the market and the riskless asset can appear to have inferior performance in terms of the expression [21]. In other words, a successful market timer can have a negative Jensen's alpha.

It must be recalled that the asset pricing model behind [21] (the CAPM) assumes homogeneous beliefs. It is not surprising that it becomes difficult to reconcile the model with the information acquisition technology that gives rise to the existence of superior information[18].

In more practical terms, we confront a problem of asymmetric information. Suppose that a regulatory agency wants to evaluate a portfolio manager. In general, the outsider cannot observe the information set available to the manager. It turns out that this problem is reflected in the traditional measure of risk, the beta coefficient. In particular, timing ability produces an upward bias in beta. It must be pointed out that even if beta is able to explain the risk premia of individual assets, it exhibits serious biases when trying to measure the risk of dynamic investment strategies. Naturally, these are the usual types of investments followed by mutual funds[19]. Of course, the problem may be more or less serious depending upon the prior objective of the fund. It is certainly the case that funds holding high percentages of fixed income securities over time may be relatively easier to evaluate.

It has been argued above that the upward bias in beta might be so severe that the traditional Jensen's alpha may erroneously designate successful portfolio managers as inferior. To elaborate this point, imagine that the manager receives two signals, H and L, for high and low market respectively. The manager is constrained to select either a high beta portfolio or a low beta portfolio depending on the signal. Suppose that the benchmark is mean-variance efficient, so that the two characteristic lines (for the up market and for the down market) will go through the origin. It should be clear that the only difference between both lines is their slope. The uninformed evaluator, running the usual regression given by [21], will capture the average risk between both positions which is given by the slope connecting the high beta and the low beta strategies. Unfortunately, it can be shown that the slope of this line will be steeper than either of the two previous characteristic lines. An upward bias in beta has artificially been introduced. Moreover, the intercept of the regression line can even be negative suggesting inferior performance. Appendix 2 has a brief technical presentation of how Jensen's alpha may be negative if there is timing information.

[18] Connor and Korajczyk (1986) imposed some restrictions on the particular information held by informed investors. In particular, they have superior information about asset-specific events only. They do not have timing information. Under these conditions, the APT alpha is a valid measure of performance.

[19] For an excellent presentation of these issues, see Grinblatt and Titman (1989b).

In summary, the problem of performance is essentially different when the managers change their portfolio composition upon obtaining private information than when the composition of the portfolio analyzed is known to be constant. Unfortunately, the most common case must deal with frequent changes in the portfolio's risk level through changes in portfolio holdings. The impossibility of observing the impact of these changes on the portfolio's risk level makes the performance evaluation of mutual funds a very difficult and complex problem.

It should be emphasized that the real difficulty lies behind the asset pricing models employed. It is not just a question of solving an econometric problem. We would need a model involving heterogeneous beliefs, asymmetric information, and rational behavior of agents in a multiperiod framework. This is an extremely difficult task. Admati and Ross (1985) proposed a model along the lines suggested above. As they pointed out, however, the crucial feature of multiple time periods cannot be addressed by their model[20].

3.2. Mean-variance efficiency

One of the most important ideas of modern financial economics is that mean-variance efficiency of the true market portfolio implies the CAPM[21]. In fact, a linear relationship between expected return and beta must hold when beta is measured relative to any mean-variance efficient portfolio. This result is just a consequence of the mathematics behind efficient portfolios. Moreover, it is also true that an exact linear function of beta must always hold with respect to ex-post data if the proxy chosen as the market portfolio is ex-post efficient. The immediate implication of this result is that all securities and portfolios must have zero alphas. In other words, there is no theoretical ability to discriminate between investments whenever the market proxy employed is ex-post efficient[22].

On the other hand, if the benchamark portfolio lies inside the efficient frontier; i.e. the proxy is not mean-variance efficient, securities or portfolios may have either positive or negative alphas. In this case, in principle, we are able to rank portfolios. Unfortunately, it is not clear whether this ranking indicates true performance or just the position of the benchmark in the mean-standard deviation space[23].

[20] The work of Wang (1990), promises interesting future applications along these lines.

[21] A mean-variance efficient portfolio is obtained when, for a given standard deviation, the portfolio gives the highest possible expected return and, for a given expected return, the portfolio provides the smallest standard deviation. See Roll (1977) for an excellent discussion about the pricing implications of mean-variance efficient portfolios.

[22] It can be shown that similar conclusions can be reached using the APT framework. See Grinblatt and Titman (1987).

[23] See Roll (1978) and Dybvig and Ross (1985b).

Despite the fact that the comments above represent an important challenge to the evaluation of managed portfolios, there is a theoretical argument under which a manager with superior information may, on average, obtain a positive alpha, whilst a manager who lacks superior information may, on average, have zero alpha. The main idea is that the benchmark portfolio is efficient from the perspective of managers without superior information. It was pointed out above that they are not able (and they should not be able) to distinguish between passive strategies or even between active strategies based on public information. However, the same benchmark falls inside the efficient frontier of informed managers. This implies that active investment strategies based on superior private information will and should be above the ex-post or empirically implemented CAPM. In other words, the slope connecting the risk-free return to a benchmark portfolio located on the efficient frontier of managers who lack superior information will be less than the slope of the efficient frontier of the informed managers. The sign of alpha will, on average, be positive for managers with high quality information[24].

These arguments suggest that the first step in the evaluation process should try to identify mean-variance efficient benchmarks. To understand why this is crucial, it should be noted that efficiency of a benchmark will be rejected if a different portfolio with constant weights obtains a positive and significant performance measure relative to the benchmark. It must follow that if a mutual fund realizes superior performance relative to a benchmark for which efficiency cannot be rejected, the positive result must be explained by changes in portfolio holdings (weights) due to superior information. This would not necessarily be true if efficiency of the benchmark can be rejected[25].

3.3. Option related issues

In the previous discussion of this section, we have presented the two relevant problems of performance evaluation. The issues of asymmetric information between the manager and the outsider and the efficiency of the benchmark employed in the evaluation process must be fully addressed in the analysis of mutual funds.

On the other hand, it was also argued that, in principle, it should be possible to separate the questions of selectivity information and timing information. Of course, it should be clear that throughout that discussion the problems of asymmetric information and efficiency of the benchmark were ignored.

Unfortunately, it turns out that even under ideal conditions, both the quadratic and the Henriksson-Merton approach to timing might be misspecified.

The first important aspect to be noted is that both approaches are in fact option motivated. The quadratic regression indicates that the portfolio's beta fluctuates over many values, depending on the size of the market's excess

[24] See Dybvig and Ross (1985b) and Grinblatt (1986).

[25] See Grinblatt and Titman (1989b) for a detailed analysis.

returns. In other words, the slope is continually increasing from left (bear market) to right (bull market)[26].

The Henriksson-Merton approach indicates that the portfolio's beta fluctuates just between two values, depending upon whether the market return is larger or lower than the risk-free rate.

As we can see, both ideas rest on non-linear payoff structures. In fact, the quadratic regression may exaggerate the option-like characteristics of the Henriksson-Merton approach.

In order to understand the potential misspecification of these models, it should be pointed out that the market timing ability considered by them, can be viewed as a free option[27]. This is to say, if we invest in regular options we would obtain the same structure of payoffs as when using market timing. However, it is clear that options have a non-negative cost. This implies that if mutual funds have options in their portfolio composition (or option-like securities), the reduction in return produced by the cost of options will appear in the alpha coefficient. Given that a positive timing coefficient in equation [19] is equivalent to buying a market put without paying the price of the option, the decrease in return from the cash paid by purchasing the option will show up as a negative alpha coefficent. At the same time, a negative timing coefficient is equivalent to selling a market put without receiving the price for the option. Now, the increase in return from the cash obtained by selling the option will show up as a positive alpha coefficient. Hence, we should expect a negative correlation between measures of timing and selectivity if mutual funds incorporate options in their portfolio holdings.

It should be noted that besides holding actual options in their portfolios, there are two ways in which funds could include option payoffs in their returns. First of all, it is well known that options can be replicated using dynamic strategies between cash and stocks. If mutual funds trade more frequently than the measurement of returns, this may create false evidence of timing. In the limit, this argument would imply continuous trading and, therefore, perfect option replication without any special timing information.

More importantly, it is also well known that equity in a leveraged firm is just a call option on the assets of the firm with a striking price equal to the face value of debt. Moreover, debt can be viewed as holding a risk-free zero coupon bond and selling a put option on the value of the firm with the striking price being the face value of debt. Whenever mutual funds buy debt and equity of leveraged firms, their payoff will be affected by these implicit options.

To conclude, as long as mutual funds hold firms with debt in their capital structure, the coefficients α_p and γ_{2p} in regression [19] will be negatively correlated. The same is true for the quadratic regression. The available models

[26] See Lehmann and Modest (1987) for a detailed analysis of the quadratic regression.

[27] See Jagannathan and Korajczyk (1986).

using returns to capture market timing are misspecified because of option-like payoff structures[28].

3.4. *Risk-adjusted performance measures: new alternatives*

a) THE POSITIVE PERIOD WEIGHTING

The problem of asymmetric information discussed previously mainly implies that the traditional measures of risk are biased due to unobservable changes in the portfolio holdings of mutual funds. Unfortunately, data on portfolio holdings are very costly and difficult to obtain. We would like to measure performance by observing returns and, at the same time, to be able to correctly interpret the results.

In a very important step towards the clarification of the issues involved in performance evaluation, Grinblatt and Titman (1989b) developed a performance measure which does not assign negative performance to market timers. The problem of overestimating risk because of market timing ability is solved by their proposal.

The positive period weighting (PPW) finds a time series of weights that sum to one with the two following properties:

1. They are non-negative ($w_t > 0$)

$$2)\ \sum_{t=1}^{T} w_t (R_{mt} - R_{ft}) = 0 \qquad [22]$$

where the benchmark, m, is mean-variance efficient.

Once these weights have been estimated, the PPW is given by:

$$\mathrm{PPW} = \sum_{t=1}^{T} w_t (R_{pt} - R_{ft}) \qquad [23]$$

where R_{pt} is the return of the mutual fund at time t. PPW greater (lower) than zero indicates superior (inferior) performance.

This measure of performance is intuitively simple. The estimated weights force the weighted average excess return of the benchmark to be zero. If, with the same weights, the fund's weighted average excess return is positive, it must be an indication of superior performance relative to the benchmark.

The non-negative constraint imposed on the weights solves the problem of assigning negative performance to market timers. Otherwise, when the market excess return is large, the weights in the second condition of [22] will be negative. This may imply a negative performance measure given by [23] preci-

[28] All empirical applications of both techniques with the exception of Lehmann and Modest (1987) obtain results consistent with the type of misspecifications discussed.

sely when we observe a bull market. If this market behavior were successfully captured by the manager, by allowing negative weights, we would incorrectly assign negative performance to skilled portfolio managers. Hence, the non-negative constraint becomes a crucial component of the new performance measure.

To see what is behind the PPW, let the initial wealth of the representative investor be 1. When his or her wealth is divided between the market portfolio and the risk-free rate, future wealth, for any given time period t, is given by:

$$W = (1-w)\,(1+R_f) + w(1+R_m) = 1 + (1-w)R_f + w\,R_m \qquad [24]$$

where w is the fraction of his initial wealth invested in the market portfolio.

The problem of the representative investor is to maximize the expected utility of his future wealth by choosing the optimal fraction w:

$$\max E[U(1 + (1-w)\,R_f + w\,R_m)] \qquad [25]$$

The optimality condition is given by[29]:

$$E[U'(W)\,(R_m - R_f)] = 0 \qquad [26]$$

where $U'(W)$ is the marginal utility of future wealth. At this point, it becomes necessary to impose more structure on the problem, so that we can provide $U'(W)$ with empirical content. In particular, we are going to impose a very popular utility function in the literature of financial economics[30].

$$U(W) = W^{1-b}/(1-b) \qquad [27]$$

where b is the coefficient of relative risk aversion.

With T observations during the sample period and imposing [27], the optimality condition [26] becomes:

$$\sum_{t=1}^{T} \{1 + wR_{mt} + (1-w)R_{ft}\}^{-b}\,(R_{mt} - R_{ft}) = 0 \qquad [28]$$

This expression can now be used to solve for the optimal weight, w^*, during the whole sample period. Therefore, for each month (or any other interval), it becomes possible to estimate:

[29] By assuming that the utility function is concave, the first order necessary conditions are also sufficient.

[30] This function is known as the power utility function. It presents decreasing absolute risk aversion and constant relative risk aversion. These two properties not only make intuitive sense, but they are also convenient from a theoretical point of view. At the same time, this implies that the empirical implementation of the PPW depends upon a particular coefficient of the aggregate relative risk aversion. Fortunately, for the Spanish capital market, the work by Alonso, Rubio and Tusell (1990) contains the evidence on the magnitude of the b coefficient in Spain over the past twenty years.

$$\{1 + w^* R_{mt} + (1 - w^*) R_{ft}\}^{-b} \quad [29]$$

It is crucial to point out that [28] has the same structure as the second condition of [22]. This is to say, the term in brackets given by [29] represents the time series of weights that we are looking for. It turns out that once these weights are scaled to sum to one, they satisfy the conditions given by [22]. In other words, the weights are just given by the marginal utility of the representative investor's future wealth. In the last step, these weights are applied to the portfolio excess returns to obtain an estimate of the PPW.

Unfortunately, the PPW contains some drawbacks which may mitigate its empirical validity. Without imposing a constant relative risk aversion utility function, its implementation certainly becomes extremely difficult. This suggests that correct knowledge of the relative risk aversion coefficient is required. It seems clear that performance measurement could be misleading if these conditions are inappropriate. Morover, its application involves maintaining the optimal weights for the market and the risk free asset constant throughout the sample period. It should be noted, however, that the final weights given by [29] change in every period. Finally, as we will see below, the PPW is very similar to Jensen's alpha. In fact, the same general static flavor is behind both measures of performance. This should help to understand the results reported later in the empirical section of the paper.

As pointed out by Grinblatt and Titman (1988), the PPW measure, like the Jensen's alpha, is a linear weighting of returns. The implication is that the standard t-statistic can be used to test whether it is significantly different from zero, when conditioned on the excess returns of the portfolios in the alternative benchmarks. To appreciate this point, it should be noted that under the null hypothesis that the manager does not have superior selectivity or timing information, the PPW may be written as:

$$\text{PPW} = \sum_{t=1}^{T} w_t(\beta_p r_{mt} + \varepsilon_{pt}) = \beta_p \sum_{t=1}^{T} w_t r_{mt} + \sum_{t=1}^{T} w_t \varepsilon_{pt} = \sum_{t=1}^{T} w_t \varepsilon_{pt}$$

Assuming that the weights are distributed independently of ε_{pt} and that ε_{pt} is homoscedastic, the test statistic is given by:

$$\text{PPW}\,(\sigma_\varepsilon^2 \sum_{t=1}^{T} \text{weights}_t^2)^{-1/2} \quad [30]$$

where σ_ε^2 can be computed using the residuals from the regression used to compute Jensen's alpha[31].

Lastly, the importance of non-negative weights in [22] should be recalled. In fact, this is the difference between Jensen's alpha and the PPW. In this regard, it can be shown that Jensen's alpha can be written as:

[31] Given the explicit recognition of market timing, the estimate of the standard deviation of the residuals might be adjusted for heterokedasticity. See Cumby and Glen (1990).

$$\alpha = \sum_{t=1}^{T} w_t (R_{pt} - R_{ft}) \tag{31}$$

To see this, it should be pointed out that Jensen's alpha is obtained by letting the weights be given by:

$$w_t = \frac{\sigma_m^2 - (r_{mt} - \bar{r}_m)\,\bar{r}_m}{T\,\sigma_m^2} \tag{32}$$

with the important difference, relative to the definition of the PPW, that the weights are allowed to be negative. For this reasons, we might observe negative alphas when there is timing ability. In order to understand why we allow negative weights in the traditional Jensen's alpha, it should be pointed out that the CAPM holds with either normal returns or quadratic preferences. It is well known that the quadratic utility function may present negative marginal utility. Given that the weights are just marginal utilities over time, it implies the possibility of erroneously designating superior portfolio managers as inferior.

b) THE CONNOR-KORAJCZYK MEASURE OF TOTAL PERFORMANCE

It has been mentioned before that it is extremely difficult to measure timing ability by using returns. The theoretical framework developed by Henriksson-Merton (1981) and Pfleiderer and Bhattacharya (1983) is misspecified because of the implicitly option-like payoffs whenever a mutual fund acquires a leveraged firm.

In order to avoid this problem, we must recognize explicitly that mutual funds buy or short sell implicit costly options. This is the insight of the performance measure proposed by Connor and Korajczyk (1990) (CK).

Unfortunately, their measure does not separate selectivity information and timing information. It is just a measure of total performance. Its advantage is that, once again, we have a measure of performance which is not contaminated by market timing ability. In this sense, it provides the evaluator with the same advantage as the PPW.

CK assume that the mutual fund invests the proceeds (positive or negative) from the put transaction in Treasury Bills. This implicit cost may be captured through the following expression:

$$\text{Nput} = \max(0, R_{ft} - R_{mt}) - (1+R_{ft})P \tag{33}$$

where P is the price of a European put option against a non-dividend paying market index, and Nput means net put. The Black Scholes valuation model might be used in order to price this implicit put option.

The framework developed before to estimate market timing was based on the following regression:

$$r_{pt} = \alpha_p + \gamma_1 r_{mt} + \gamma_2 \text{put} + \omega_{pt} \quad [34]$$

On the other hand, by using [33], we can write:

$$r_{pt} = \alpha'_p + \gamma_1 r_{mt} + \gamma_2[\text{put} - (1+R_{ft})\, P] + \omega_{pt} \quad [35]$$

By combining [34] and [35] we obtain (on average):

$$\alpha_p = \alpha'_p - \gamma_2\,(1+R_f)P \quad [36]$$

The model given by [35] predicts $\alpha'_p = 0$. This becomes the CK's performance measure. It should be noted that when $\alpha'_p = 0$, the Jensen's alpha is $-\gamma_2(1+R_f)P$. It is possible now to explain the negative correlation coefficient between α_p and γ_2 suggested by the misspecification of the Henriksson-Merton model. In other words, by using the traditional timing model we would predict $\alpha_p = 0$, whereas the CK model predicts $\alpha_p = -\gamma_2(1+R_f)P$. By making the adjustment of [35], we can correctly predict $\alpha'_p = 0$.

As before, with the PPW, this technique is not free of conceptual and empirical limitations. In particular, this model is based on the Henriksson-Merton framework. It must be recalled that their regression rests on the arbitrary assumption that the manager's timing information relates only to the sign of the excess return of the benchmark, and not at all to its level. Moreover, it seems difficult to really assess the relative importance of option-like securities with non-linear payoffs in the portfolio of mutual funds. Finally, its empirical implementation requires an option valuation model. Taking into account the peculiar characteristics of this model, and the sensitivity of option pricing models to the estimation of the standard deviation of returns, this issue does not seem to be trivial.

c) Using Portfolio Holdings Data

The last performance measurement to be discussed is related to the separation of selectivity information and timing information. This topic is one of the most promising areas of research in performance evaluation, and as discussed below is closely related to the availability of data on portfolio holdings.

The importance of recognizing explicitly dynamic strategies in order to be able to separate selectivity and timing has been discussed by Cornell (1979), Grinblatt and Titman (1989b), and Elton and Gruber (1991). In particular, the last two papers emphasize the role of the covariance between the dynamic beta and the excess return of the fund as the appropriate measure of timing ability. Unfortunately, these ideas assume the observability of portfolio holdings over time.

It is well known that Jensen's alpha can be written as:

$$J_\alpha = \bar{r}_p - \hat{\beta}_p\, \bar{r}_m \quad [37]$$

where, $\bar{r}_p$, $\hat{\beta}_p$ and $\bar{r}_m$ are the average excess return of the fund, the estimate of the beta of the fund, and the average excess return of the reference portfolio respectively.

On the other hand, the dynamic beta of the fund can be defined as:

$$\beta_{pt} = \sum_{i=1}^{N} w_{it}\, \hat{\beta}_i \qquad [38]$$

where, w_{it} is the percentage of each security relative to the value of all assets in the fund in each month t, and $\hat{\beta}_i$ is the estimate of beta for each security over a given sample period[32]. Of course, the whole point is that the weights, w_{it}, will change over time.

Using the dynamic beta given by [38], we can write:

$$r_{pt} = \hat{\beta}_{pt}\, r_{mt} + \varepsilon_{pt} \qquad [39]$$

Taking the average of [39], and after a simple manipulation:

$$\bar{r}_p = \bar{\beta}_p\, \bar{r}_m + \frac{1}{T} \sum_{t=1}^{T} \hat{\beta}_{pt}(r_{mt} - \bar{r}_m) + \bar{\varepsilon}_p \qquad [40]$$

where $\bar{\beta}_p$ is the average of the dynamic beta over the sample period[33]. Using [40] in the expression [37], Jensen's alpha can be written as:

$$J_\alpha = (\bar{\beta}_p - \hat{\beta}_p)\bar{r}_m + \frac{1}{T} \sum_{t=1}^{T} \beta_{pt}(r_{mt} - \bar{r}_m) + \bar{\varepsilon}_p \qquad [41]$$

which contains the three components of total performance as measured by Jensen's alpha. The first term is the bias of the estimate of beta, the second represents a measure of timing information, and the last term corresponds to the selectivity component. The last two terms can be written as:

$$TM_\alpha = \mathrm{cov}(\beta_{pt}, r_{pt}) \qquad [42]$$

which can be estimated to obtain a measure of timing ability, and

$$S_\alpha = \bar{r}_p - \frac{1}{T} \sum_{t=1}^{T} \beta_{pt}\, r_{mt} \qquad [43]$$

[32] It should be recognized that the empirical validity of this approach rests on the betas for individual assets remaining constant over the sample period. This limitation does not imply that this approach is not robust to the objections raised in section 3.1. However, it is based on models which do not properly value options or option-like assets with non-linear payoffs.

[33] It is important to note that this beta may be different than the beta estimated by the usual regression methodology. In fact, the difference between these two types of betas is one of the components of Jensen's alpha.

which may be employed to estimate the selectivity component of total performance.

4. Some Empirical Results Using Spanish Data

4.1. The data

Data necessary to compute monthly returns of individual securities for a period of twenty-nine years are available. The twenty-nine year period starts in December 1962 and ends in December 1991. This means that a total of 348 monthly returns are available. These returns are adjusted for cash dividends and changes in the capital structure. The sample is composed of 152 individual stocks.

It was decided to use the returns on all securities in the sample to compute an estimate of the monthly return on the market portfolio. Two market returns were calculated. The first one assumes an unweighted market portfolio (EW); the monthly return is the arithmetic mean of the returns on all available securities for that month. The second one assumes a value-weighted market portfolio (VW), the weights being the market values of each security at the end of the preceding year.

Morover, APT factors were estimated by the methodology known as asymptotic principal components[34]. Finally, one reasonable alternative to traditional benchmarks such as the stock market indexes or the APT factors, consists of forming a benchmark on the basis of stocks' characteristics. The idea would be to construct a benchmark such that passive strategies based on beta, size, or dividend yield could not generate statistically significant alphas. This basically would imply having (by construction) a mean-variance efficient benchmark[35].

Spain did not have short-term goverment securities during the beginning of the period covered by the results reported in this paper. During these months, the riskless rate was based on lending rates offered by financial institutions. Although these rates were not insured, the assumption seems to be a reasona-

[34] See Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986).

[35] In order to generate this benchmark, ten monthly return portfolios based on beta, size and dividend yield were formed. From January 1966 to December 1979, regressions of the excess returns of these portfolios on a constant and our four bencmarks were run. The intercepts from these regressions were then used as a reference point for selecting the portfolios to be included in our benchmark. This is referred to as the MVDY benchmark in the empirical illustration below. It is a five-portfolio benchmark consisting of three size-based portfolios, and two dividend yield-based portfolios. The smallest decile of firms comprised the first size-based portfolio; the average of the second through sixth smallest size portfolios comprised the second portfolio; and the average of the three largest size portfolios comprised the third portfolio. Finally, the average of the seventh and eighth, and the average of the two highest dividend yield portfolios were included in the MVDY benchmark. Given that the results are reported for the eighties, it is interesting to note that the tests were run «out of sample».

ble approximation. Treasury Bills sold by auction and priced at a discount were traded for the first time in 1982. The monthly equivalent of the one-year Treasury Bill rate obtained from secondary market operations was employed as the risk-free rate for the rest of the period. The one-month interbank deposit rate was also collected.

Data necessary to compute monthly returns of 50 mutual funds were collected from December 1970 to June 1990. The returns were obtained from the net asset values at the end of each month. Of course, data are not fully available for all mutual funds. As usual, it is important to realize that the net asset values and therefore the returns, are obtained after subtracting expenses and management fees. On the other hand, these returns do not necessarily indicate the return earned by the participants in the fund, since there may be a load charge and/or a redemption fee, and they are not included in the estimation of net asset values. In any case, in Spain, the load charge is relatively uncommon, and the redemption fee only applies for holding periods of less than two years.

The sample of mutual funds is divided into two subsamples. The first one — 33 funds— includes mutual funds with risky assets. In order to be classified in this subsample, mutual funds must have at least 25 percent of their portfolios in risky assets throughout the whole period for which data are available. The second subsample incorporates funds with mostly fixed income securities in their holdings. The risk level of the first subsample measured by beta was, on average, clearly higher than the risk level of the second group.

It must be mentioned that our effort was centered on collecting information about funds holding at least 25 percent of their portfolios in risky securities. Throughout the whole period, 1971-1990, our sample represents 95 percent of the total value of assets for this sector of the industry.

Finally, data necessary to compute the dynamic beta given by equation [38] was collected for 14 mutual funds holding risky securities. This implies that we have monthly portfolio holdings for these funds between January 1986 and June 1990[36]. Given the serious degree of concentration in the industry, this sample represents 90 percent of the value of assets of all existing funds with holdings of risky securities. Over the 54 months of the sample period, these mutual funds had, on average, 66 percent of their holdings in risky assets.

4.2. Empirical results: the evidence from monthly returns

As mentioned in the previous section, testing for mean-variance efficiency of the benchmarks employed is a crucial issue in performance evaluation. Rubio (1992) performs a comprehensive analysis of the benchmarks suggested in

[36] The mutual funds are BBV Indice, BBV Rendimiento, Fonbancaya, Nuvofondo, Hispafondo, Fonbanif, Fondhispano, Inespa, Inrenta, Fontisa, Banserfond, Ahorrofondo, Eurovalor and Dineuro Bolsa.

subsection 4.1. It turns out that the performance of the VW index in terms of mean-variance efficiency is clearly inferior to the performance of the EW index or APT factors. Over all periods, the performance of both the APT benchmarks and the EW index is very similar. Unfortunately, however, and for both 1971-90 and 1980-90, it seems clear that passive portfolio strategies based on size and dividend yield realize abnormal performance relative to either CAPM or APT benchmarks. On the other hand, average absolute mispricing tends to be much smaller relative to our fourth benchmark. It should be recalled that this benchmark is built so that the well known anomalies previously reported tend to disappear. The results based on this benchmark provide an important reference point in order to interpret the general results without ambiguity.

The results for different benchmarks and performance measures are contained in Table 1 for individual mutual funds[37]. For most of the mutual funds, and independently of the benchmark and performance measurement employed, the results from 1980 to 1990 are clearly disappointing. The average annualized alpha varies from −1.8 percent for the VW index, to −6.9 percent for the EW benchmark. It is important to recall that the VW index is the most inefficient bechmark. It can potentially generate spurious conclusions about mutual fund performance. In this sense, the EW index or the APT factors are more reliable benchmarks. Moreover, it can be observed from Table 1 that the alphas obtained under the fourth benchmark closely follow the alphas under the EW index. Given this evidence, we may conclude that the poor performance of the Spanish mutual funds is not sensitive to reasonable benchmarks. Taking into account that the VW index is much more affected by the behavior of large firms, it seems possible to suggest that the average results for the sampling period reflect the negative influence of a very strong size effect. It must be pointed out that the Spanish mutual funds tend to be concentrated in large companies. This is just a consequence of a serious concern of investment companies about the liquidity of their investments[38].

[37] See Rubio (1992) for a detailed presentation of the empirical results concerning the behavior and performance of Spanish mutual funds. In Table 1, Fondonorte and Rentfondo can be considered to be fixed income funds. In interpreting the results regarding these two funds, one should take into account that they might not be as well diversified as risky funds (funds having at least 25 percent of their portfolios in risky assets throughout the sample period). If this were actually the case, beta risk would not be the appropriate measure of risk supported by these funds. Total risk, as measured by standard deviation, should be considered. The Sharpe ratio would be a more relevant performance measure.

[38] We have stressed on the importance of testing for benchmark efficiency throughout the paper. However, most of the results in Table 1 are based on the traditional benchmarks for which mean-variance efficiency seems to be rejected. Of course, one of the objectives of this exercise is to see whether the results are robust to alternative bechmarks. In this sense, it seems natural to include the results for all benchmarks. Moreover, given the evidence in Table 1, it would be possible to report the results in terms of any of the three last bechmarks employed. Finally, it should be pointed out that the efficiency of the EW index and the APT factors could not be rejected during the last five years of the sample period. This was clearly not the case for the VW index.

On the other hand, the remarkable similarity between different measures across benchmarks is striking. It can be concluded that the negative results obtained are not sensitive to the choice of the performance evaluation technique. If there is any market timing ability among managers, it does not seem to affect Jensen's alpha in a significant way. This is an interesting result. It suggests that a very straightforward methodology can be employed in the evaluation of mutual funds [39].

To further explore the reasons behind the similarities found in Table 1, we should first investigate the relevance of market timing. The results reported by Rubio (1992) suggest that the models proposed to distinguish between selectivity information and timing information just by employing returns data are not useful. As previously mentioned, there are reasons to suspect that these models are misspecified due to non-linear payoffs related to option-like securities. In any case, it is not possible to believe that portfolio managers are systematically engaged in wrong market timing strategies. Rubio (1992) employs the data described at the end of subsection 4.1 to implement the analytical framework developed in part C of subsection 3.4. In particular, both the timing component and the selectivity component are estimated using expressions [42] and [43]. The results indicate that timing contributes positively to total performance for both the VW and EW indexes. The negative performance of Table 1 seems to be related to the either the poor capacity of selecting individual assets and/or to the impact of turnover costs in the selection process [40].

The fact that timing is an important component of performance tend to suggest that the Jensen measure presented in Table 1 should be misleading. However, given the robustness of the results to alternative measures of performance, we may be tempted to conclude that the empirical relevance of imposing non-negative weights in the PPW methodology is basically insignificant. On the other hand, it is interesting to note that, in most cases, performance appears to be slightly worse using the PPW. It should be possible to argue that

[39] Regarding the implementation of the PPW, it must be said that the empirical evidence of Alonso, Rubio and Tusell (1990) suggests that a coefficient of relative risk aversion around 4 is a reasonable estimate. This reference point is employed in our estimations reported in Table 1. In any case the results were not sensitive to the choice of *b*. The weights found in the first step necessary to implement the PPW were 97 percent and 73 percent for the EW and VW index respectively. These weights clearly reflect the best performance of the EW index relative to the VW index. Using a similar algorithm to the one employed to obtain the PPW, an optimal ex-post combination of 5 APT factors was obtained. The optimal weight for the first factor was 84 percent. On the other hand, in order to implement the CK methodology, the Black-Scholes formula was employed to price the put option in each month during the sample period. The values of the exercise price, the underlying asset, and the time to expiration were normalized to be equal to one. Finally, the whole sample period was divided into four five-years subperiods. The unbiased standard deviation of the excess market return was estimated for each subperiod. It was then used as the input for the standard deviation of the underlying asset in the put option expression.

[40] The negative performance directly related to the selectivity component appears to be connected with the size effect.

TABLE 1
Measures of performance for alternative techniques and benchmarks: individual mutual funds, 1980-1990[1]

MUTUAL FUNDS	VW INDEX			EW INDEX			APT FACTORS[2]			MVDY[3]
	Alpha[4]	PPW[5]	CK[6]	Alpha	PPW	CK	Alpha	PPW	CK	Alpha
Ahorrofondo	0.298	0.300	0.293	0.061	0.050	0.063	0.157	0.137	0.162	0.105
	(1.43)	(1.44)	(1.42)	(0.28)	(0.23)	(0.30)	(0.74)	(0.97)	(0.77)	(0.45)
Banserfond	0.410	0.410	0.400	−0.014	−0.045	−0.013	0.228	0.164	0.241	0.232
	(1.34)	(1.34)	(1.44)	(−0.04)	(−0.14)	(−0.05)	(0.76)	(0.74)	(0.86)	(0.69)
BBVIndice	−0.406	−0.432	−0.401	−0.892	−0.988	−0.887	−0.677	−0.769	−0.666	−0.815
	(−1.43)	(−1.52)	(−1.44)	(−2.99)	(−3.27)	(−3.49)	(−2.49)	(−2.85)	(−2.84)	(−2.59)
BBV Rendimiento	−0.438	−0.458	−0.430	−0.849	−0.918	−0.846	−0.645	−0.713	−0.638	−0.684
	(−1.43)	(−1.49)	(−1.46)	(−2.60)	(−2.77)	(−2.92)	(−2.14)	(−2.67)	(−2.39)	(−1.96)
Eurovalor	−0.427	−0.508	−0.377	−0.727	−0.944	−0.711	−0.593	−0.772	−0.556	−0.736
	(−1.61)	(−1.90)	(−1.68)	(−2.64)	(−3.37)	(−3.20)	(−2.13)	(−2.76)	(−2.33)	(−2.49)
Fonbancaya	−0.152	−0.232	−0.119	−0.472	−0.689	−0.461	−0.323	−0.474	−0.298	−0.302
	(−0.87)	(−1.32)	(−0.68)	(−2.41)	(−3.46)	(−2.13)	(−1.71)	(−2.39)	(−1.37)	(−1.47)
Fondonorte	0.060	0.026	0.065	−0.030	−0.138	−0.024	0.082	0.013	0.095	0.082
	(0.21)	(0.09)	(0.23)	(−0.10)	(−0.46)	(−0.08)	(0.28)	(0.06)	(0.32)	(0.28)
Fontisa	0.534	0.501	0.537	−0.001	−0.103	0.002	0.217	0.128	0.226	0.123
	(1.83)	(1.72)	(1.99)	(−0.00)	(−0.35)	(0.00)	(0.74)	(0.68)	(0.83)	(0.40)
Hispafondo	−0.035	−0.048	−0.030	−0.556	−0.600	−0.553	−0.387	−0.461	−0.374	−0.490
	(−0.13)	(−0.18)	(−0.12)	(−2.21)	(−2.34)	(−2.26)	(−1.61)	(−2.01)	(−1.55)	(−1.82)
Inespa	−0.373	−0.383	−0.376	−0.959	−0.996	−0.956	−0.760	−0.824	−0.754	−0.859
	(−1.37)	(−1.40)	(−1.43)	(−3.76)	(−3.84)	(−4.04)	(−2.95)	(−3.70)	(−3.01)	(−3.10)
Inrenta	−0.095	−0.109	−0.093	−0.649	−0.693	−0.646	−0.457	−0.529	−0.446	−0.588
	(−0.35)	(−0.40)	(−0.39)	(−2.50)	(−2.63)	(−2.84)	(−1.74)	(−2.23)	(−1.93)	(−2.15)
Mediterráneo	−0.881	−0.932	−0.848	−1.026	−1.771	−1.010	−1.009	−1.109	−0.977	−1.053
	(−2.83)	(−2.99)	(−3.17)	(−3.10)	(−3.56)	(−4.13)	(−3.11)	(−3.57)	(−3.88)	(−2.97)
Nuvofondo	−0.258	−0.278	−0.257	−0.810	−0.863	−0.808	−0.614	−0.671	−0.611	−0.717
	(−1.07)	(−1.15)	(−1.16)	(−3.63)	(−3.81)	(−3.84)	(−2.66)	(−2.87)	(−2.80)	(−2.95)
Rentfondo	−0.056	−0.034	−0.069	−0.192	−0.150	−0.193	−0.144	−0.124	−0.145	−0.144
	(−0.30)	(−0.18)	(−0.40)	(−1.00)	(−0.77)	(−1.09)	(−0.77)	(−0.44)	(−0.85)	(−0.77)

[1] Until June 1990.
[2] The results are based on 5 APT factors.
[3] Benchmark on the basis of stocks' characteristics.
[4] The Jensen's alpha, in percentage terms.
[5] The Positive Period Weighting of Grinblatt and Titman.
[6] The measure of total performance of Connor and Korajczyk.

the difference between Jensen's alpha and the PPW is a manifestation of the positive contribution of timing to total performance. This would be the case if the PPW abstracts from the market timing impact.

This conjecture was investigated by running a generalized least square regression of the difference between both measures of performance for each individual fund on a constant and the timing measure estimated by [42]. The results show a positive statistically reliable relation between both variables. This may indicate that the lack of timing biases in the PPW measure is statistically relevant.

Despite the fact that there is some economic reasoning behind the apparently small differences between Jensen's alpha and the PPW, it should be recognized that the influence of these differences in the qualitative conclusions regarding the performance of mutual funds is negligible. A possible explanation could be related to the empirical evidence reported by Alonso, Rubio and Tusell (1990). They could not find important differences among asset pricing models under alternative specifications for the utility function of the representative investor.

It is also evident from Table 1 that the CK measure of performance and the Jensen measure give similar results. It was already pointed out in section 3 that the framework of the traditional asset pricing models suffers from serious limitations when an outsider tries to identify superior information. The empirical evidence reported in this paper might be a consequence of the limited capabilities of the adjustments proposed to overcome the problems of asymmetric information. Once again, the development of equilibrium intertemporal asset pricing models under asymmetric information is the most promising step in improving the understanding of performance evaluation.

Finally, it is well known that to estimate alpha under the assumption of riskless lending and borrowing can be criticized because realistic borrowing rates are typically higher than Treasury Bill rates. It turns out that measures of portfolio performance using Treasury Bills discriminate in favor of portfolios with betas less than one, and against portfolios with betas higher than one. Although not reported in the results, we are dealing with conservative mutual funds. Their betas are always well below one. In order to be more confident about the negative results found in Table 1, the same regressions were run using the interbank rate. The average annualized interbank rate during the eighties was 13.7 percent. The average risk-free rate employed in the regressions above was 9.8 percent. Our arguments suggest that the overall performance should be worse using the interbank rate. As expected, alphas were even more negative than before across all benchmarks. Given the consistently low level of betas over different periods, the use of our risk-free estimate will tend to give conservative results in terms of negative performance.

4.3. Persistence of performance

An important part of any research regarding performance evaluation is related to the distinction between luck and ability. Despite the overall negative results briefly described thus far, there exists the possibility that a few funds may be skilled enough to consistently perform so that their management fees and turnover costs are compensated. On the other hand, it is important to study whether funds with negative results maintain their poor performance through time.

In order to analyze persistence among mutual funds, the period between 1976 and 1990 was divided into three five-year subperiods, 1976-1980, 1981-85 and 1986-90. Our persistence tests regress the Jensen measures calculated for each individual fund between 1976 to 1980 against the corresponding Jensen measures for each fund between 1981 to 1985. The same procedure was repeated using the 1981 to 1985 as the initial period, and 1986 to 1990 as the testing period. In particular, the following generalized least squares regressions were performed:

$$\alpha_{pt} = a + b\ \alpha_{pt+1} + \eta_{pt} \qquad [44]$$

where a positive and significant slope coefficient would indicate persistence of performance. The results reported in Table 2 seem to be consistent with persistence of performance across benchmarks[41]. Given the similarity observed among different performance measurements and alternative benchmarks, the results do not seem to be dependent upon the specific techniques employed in Table 2.

TABLE 2
Persistence of performance across sample periods

Generalized least squares cross-sectional regressions of Jensen's alpha for some period t on Jensen's alpha for the following period $t+1$. The results are reported for the VW market index, the EW market index, and 5 APT factors
$\alpha_{pt} = a + b\ \alpha_{pt+1} + \eta_{pt}; p = 1, ..., N$

PERIOD	1976/80-1981/85			1981/85-1986/90[1]		
BENCHMARK	Intercept	Slope Coeff.	R^2 Adj.	Intercept	Slope Coeff.	R^2 Adj.
VW Index	−0.589 (−6.11)[2]	0.372 (1.64)	0.12	0.262 (2.74)	0.780 (3.97)	0.55
EW Index	−0.154 (−1.75)	0.608 (3.24)	0.44	0.216 (0.99)	0.558 (2.42)	0.29
5APT Factors	−0.583 (−7.26)	0.441 (2.66)	0.34	0.362 (1.73)	0.774 (2.75)	0.35

[1] Until June 1990.
[2] t-statistics in parenthesis. They are computed using the unbiased cross-product of OLS time series residuals as the weighting matrix.

However, it has been argued by Brown, Goetzmann, Ibbotson, and Ross (1991) that the apparent persistence in abilities among fund managers might be due to survivorship bias. A manager who is willing to accept a large amount of risk will have a high probability of failure. However, if he or she is

[41] Other randomly selected subperiods were also employed. The results were not sensitive to the choice of the sample period. Persistence of performance has also been found in the U.S. with a variety of techniques by Grinblatt and Titman (1988), and Hendricks, Patel and Zeckhauser (1990).

able to survive, there is a high probability that this manager won his or her important bet. This implies that high returns tend to persist. Otherwise, we would not find the fund in our sample anymore. In their own words: «to the extent that survivorship depends on past returns, ranking managers who survive by realized returns may induce an apparent persistence in performance». They present some numerical evidence to show that this effect may bė partially responsible for persistence. On the other hand, Grinblatt and Titman (1989a) argue that the distortion caused by survivorship bias is very small. In particular, they report that this effect is reponsible for only 0.1 to 0.4 percent return per year measured on a risk adjusted basis before transaction costs and management fees.

From our point of view, it does not seem reasonable to believe that our sample suffers from survivorship bias. During the sample period used in the analysis, only one fund, Fondiberia, disappeared from the sample. The results presented in Table 2 do not contain this fund. When the same regressions were run with Fondiberia incorporated into the sample, the results were almost identically the same.

To further investigate the extent to which our findings are due to the best and worst performing funds, equally-weighted portfolios of two funds with the top and bottom Jensen measures were constructed. Positive (negative) performance for the funds ranked best (worst) would indicate persistence of performance. The results (not reported) show sign in favor of this hypothesis. It is interesting to point out that, between 1986 and 1990, all funds performed so poorly that even the funds previously ranked best had negative alphas.

It can be concluded that these results reflect differences in the performance ability of fund managers over time. Nevertheless, it must also be recognized that persistence tests might not be a fair tests of management skills. Skilled managers are probably willing to take a considerable amount of idiosyncratic risk. At the same time, managers who do not have good performance skills will tend to hold large well diversified passive portfolios increasing their probability of not showing inferior performance. Skilled managers are, in fact, taking more chances of finding themselves in the bottom side of the rankings. Fortunately, taking into account the large number of observations available, we should be confident of the results regarding persistence of management ability over time.

5. Conclusions

This paper has presented a survey of the techniques employed in order to evaluate the quality of portfolio management. The survey has been constructed around the basic problems contained in the traditional measures of performance. It was pointed out that the issue of asymmetric information between the outsider and the manager is the fundamental problem confronted by the evaluator. Moreover, the efficiency of the benchmark used in the evaluation should be properly addressed by the researcher.

Given these two ideas, the empirical results were reported on the basis of construct validity. Alternative ways of measuring performance were discussed. As we pointed out above, the specific problems regarding each technique were described. The same strategy was followed in terms of competing benchmarks. Empirical results were reported relative to each performance measurement and benchmark. In general, the results were not sensitive to the choice of the performance measurement procedure. On the other hand, it was found that the results involving the VW index were consistently different from the results obtained with other benchmarks. Taking into account that the VW index is the most inefficient benchmark, it was concluded that results about performance evaluation should always include the EW index, APT factors, or an alternative benchmark robusts to well known cross-sectional anomalies of asset pricing. Of course, the EW index is a rather simple way of constructing a benchmark. It should be clear that the important point about any benchmark is its mean-variance efficiency. Otherwise, there is always the incentive, from the manager's point of view, to take advantage of the deficiencies in the particular benchmark employed. This is even more serious if there is some type of direct relationship between rewards and performance evaluation.

Finally, it was mentioned that any evaluation of the quality of information should contain an attempt to distinguish between two main sources of performance, selectivity information and market timing information. It was pointed out that, in principle, this is attainable. Unfortunately, it must be concluded that, with monthly returns, it is not possible to measure the magnitudes of selectivity and timing. We are only able to measure the magnitude of total performance. If we want to be more precise about the reasons behind performance, portfolio holdings are absolutely necessary. The intuition should be clear. The gap between the information available to the evaluator and the information possessed by the manager is reduced by incorporating frequent revisions of portfolio holdings.

It seems possible to conclude that the future of performance evaluation will be centered around the availability of portfolio holdings. The inference validity of future research might crucially depend on this type of data. This data will help not only to be much more precise about the real services offered by mutual funds, but also to analyze the specific relation between performance and turnover costs. Moreover, intertemporal models of asset pricing with asymmetric information should be the complementary issue in the future of performance evaluation.

Finally, it should be mentioned that given the investment restrictions imposed on mutual funds, it might be necessary to construct an ex-ante benchmark that takes explicitly into account these legal constraints.

Appendix 1. The Quality of Timing Information[A1]

It is assumed throughout this appendix that the manager maximizes the expected utility of individual investors with constant absolute risk aversion[A2].

π_t is defined to be, $r_{mt} - E(r_m)$, where $E(r_m)$ is the unconditional expected market excess return[A3]. It is implicitly assumed that during the period for which the analysis is performed, there will be no portfolio revisions. Let π_t^* be the conditional expected value of π_t given the manager's information set at the beginning of the period:

$$\pi_t^* = E(\pi_t | \Phi_t) \qquad [A.1]$$

Under the assumptions that the conditional distribution of π_t is normal and constant absolute risk aversion, it can be shown that:

$$\beta_{pt} = \beta_{pT} + \theta\, \pi_t^* \qquad [A.2]$$

where β_{pT} is the target beta of the fund, and θ is the manager's linear response to information. It can be shown that $\theta = 1/a \text{ var } (\pi_T | \Phi_t)$, where the preferences of individuals with constant absolute risk aversion are parametrized by a, and β_{pt} is equal to $\theta\, E(R_m)$. The interpretation of θ is very natural. The larger the uncertainty of the conditional forecast and/or the larger the manager's risk aversion, the smaller will be the change in the portfolio's risk level.

Let us identify the manager's information set with a signal received at the beginning of the period given by $\pi_t + u_t$, where u_t is a mean-zero normal deviate which is independent of π_t.

The idea is to minimize the variance of the forecast error, so that the objective becomes:

$$\min E[\pi_t - E(\pi_t | \pi_t + u_t)]^2 \qquad [A.3]$$

Given our assumptions, $E(\pi_t | \pi_t + u_t)$ is linear on $(\pi_t + u_t)$. This indicates that we want to minimize:

$$\min_{\zeta} E[\pi_t - \zeta(\pi_t + u_t)]^2 \qquad [A.4]$$

It can be easily shown that the optimal forecast is given by:

$$\pi_t^* = \zeta(\pi_t + u_t) \qquad [A.5]$$
$$\text{with } \zeta = \sigma_\pi^2 / (\sigma_\pi^2 + \sigma_u^2)$$

where ζ may be understood to represent the manager's quality of information.

[A1] The appendix is drawn from Jensen (1972), and Pfleiderer and Bhattacharya (1983).

[A2] This assumption implies that the response to information is linear. Equivalently, this implies that the slope of the indifference curve in the traditional mean-variance space is constant.

[A3] Note that we assume that the expected market return does not change over time.

Using the optimal forecast and the values of β_{pT} and π_t, the usual characteristic line,

$$r_{pt} = \alpha_p + \beta_p r_{mt} + \varepsilon_{pt} \qquad [A.6]$$

can be written as:

$$r_{pt} = \alpha_p + \theta \{E(r_m) + \zeta[r_{mt} - E(r_m) + u_t]\} r_{mt} + \varepsilon_{pt} \qquad [A.7]$$

rearranging,

$$r_{pt} = \alpha_p + \theta\, E(r_m)\, (1-\zeta) r_{mt} + \zeta\, \theta (r_{mt})^2 + \theta\, \zeta\, u_t\, r_{mt} + \varepsilon_{pt} \qquad [A.8]$$

grouping terms, we have:

$$r_{pt} = \eta_{0p} + \eta_{1p}\, r_{mt} + \eta_{2p}\, (r_{mt})^2 + \omega_{pt} \qquad [A.9]$$

The large sample coefficient estimates of the regression above will be:

$$\begin{aligned} \text{plim } \eta_{0p} &= \alpha_p \\ \text{plim } \eta_{1p} &= \theta\, E(r_m)\, (1-\zeta) \\ \text{plim } \eta_{2p} &= \zeta\, \theta \end{aligned} \qquad [A.10]$$

The existence of selectivity can be detected from α_p. It turns out that we can estimate the quality of timing information from the residual term of equation [A.9]:

$$\omega_{pt} = \theta\, \zeta\, u_t\, r_{mt} + \varepsilon_{pt}$$

By regressing $(\omega_{pt})^2$ on $(r_{mt})^2$, we have:

$$(\omega_{pt})^2 = \theta^2\, \zeta^2\, \sigma_u^2 (r_{mt})^2 + \delta_{pt} \qquad [A.11]$$

where,

$$\delta_{pt} = \theta^2\, \zeta^2 (r_{mt})^2\, (u_t^2 - \sigma_u^2) + (\varepsilon_{pt})^2 + 2\, \theta\, \zeta\, r_{mt}\, u_t\, \varepsilon_{pt}$$

Given that both u_t and ε_{pt} are independent of r_{mt}, δ_{pt} is uncorrelated with r_{mt}. This second regression gives a consistent estimate of $\theta^2\, \zeta^2\, \sigma_u^2$. From [A.10] we also have a consistent estimator of $\theta\zeta$, which implies that we can recover σ_u^2. Using an estimate of σ_π^2, it becomes possible to obtain ζ which is, as shown by Jensen (1972), equal to the square of the correlation coefficient between the prediction and the realization of the market innovation π_t[A4]. Naturally, this would be an ideal measure of the quality of timing information. Note that it is even plausible to estimate the coefficient of absolute risk aversion. In particular, it can be shown that $1/a = \theta\, \zeta\, \sigma_u^2$.

[A4] Using a generalized least square version, it is also possible to obtain efficient estimates.

Appendix 2. The Dybvig and Ross Model

Their work assumes a model with two assets: a riskless asset with return R_f and a risky asset or portfolio whose return is generated by the following process:

$$R = R_f + \gamma + S + \varepsilon \qquad [A.12]$$

where, γ is the risk premium, S is a signal observed by the manager, and ε is an unobserved noise. Both S and ε are independent mean zero normal variables and, without loss of generality,, R_f is assumed to zero. The risk premium is a positive constant, and the manager is assumed to have valuable information, that is, the variance of S, σ_S^2, is greater than zero, but his or her information is not complete, that is, the variance of ε, σ_ε^2, is also greater than zero.

The manager attempts to maximize the expected utility from an exponential utility function of the form $-e^{-aW}$ for some $a > 0$. Note that this function presents constant absolute risk aversion with respect to future wealth W.

Let $w(S)$ be the fraction of funds which are placed on the risky asset. The manager's problem can be written as:

$$\max_{w(S)} E[e^{-aw(S)(\gamma + S + \varepsilon)}] \qquad [A.13]$$

It can be shown that the first order conditions imply that the optimal fraction is given by:

$$w(S) = (\gamma + S) / (\sigma_\varepsilon^2) \qquad [A.14]$$

Given that we are imposing a utility function with constant absolute risk aversion, the similarity between [A.14] and θ in [A.2] is not surprising.

The Jensen's alpha is now given by:

$$\alpha = E[w(S)R] - \{\text{cov}[R, w(S)\, R]/\text{var}(R)\}\, E(\gamma + S + \varepsilon) \qquad [A.15]$$

They continued their analysis to show that [A.15] can be negative even if the manager possesses valuable information as long as $\gamma^2 > \sigma_\varepsilon^2 + \sigma_S^2$. The only case in which the correct sign of [A.15] can be guaranteed is when the manager does not have any information; i.e., $S = 0$. In this case the traditional performance evaluation measures will provide the outsider with the correct analysis. Once again, it should be clear that the main problem is that, in general, the outsider does not have the same information set as the manager. His or her evaluation must be generally performed without knowing S or $w(S)$.

References

Admati, A. and Ross, S. (1985): «Measuring Investment Performance in a Rational Expectations Model», *Journal of Business* 58, pp. 1-26.

Admati, A.; Bhattacharya, S.; Pfleiderer, P. and Ross, S. (1986): «On Timing and Selectivity», *Journal of Finance* 41, pp. 715-730.

Alonso, A.; Rubio, G. and Tusell, F. (1990): «Asset Pricing and Risk Aversion in the Spanish Stock Market», *Journal of Banking and Finance* 41, pp. 351-369.

Black, F. (1972): «Capital Market Equilibrium with Restricted Borrowing», *Journal of Business* 45, pp. 444-455.

Breeden, D. (1979): «An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities», *Journal of Financial Economics* 7, pp. 265-296.

Brown, S.; Goetzmen, W.; Ibbotson, R. and Ross, S . (1991): «Survivorship Bias in Performance Studies», Manuscript prepared for the Finance Seminar, University of California, Berkeley.

Chamberlain, G. and Rothschild, M. (1983): «Arbitrage and Mean-Variance Analysis on Large Markets», *Econometrica* 51, pp. 57-72.

Connor, G. (1984): «A Unified Beta Pricing Theory», *Journal of Economic Theory* 34, pp. 13-31.

Connor, G. and Korajczyk, R. (1986): «Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis», *Journal of Financial Economics* 15, pp. 373-394.

Connor, G. and Korajczyk, R. (1988): «Risk and Return in an Equilibrium APT», *Journal of Financial Economics* 21, pp. 255-289.

Connor, G. and Korajczyk, R. (1990): «The Attributes, Behavior and Performance of U.S. Mutual Funds», *Review of Quantitative Finance and Accounting* 1, pp. 1-22.

Cornell, B. (1979): «Asymmetric Information and Portfolio Performance Measurement», *Journal of Financial Economics* 7, pp. 381-391.

Cox, J.; Ingersoll, J. and Ross, S. (1985): «An Intertemporal General Equilibrium Model of Asset Prices», *Econometrica* 53, pp. 363-384.

Cumby, R. and Glen, J. (1990): «Evaluating the Performance of International Mutual Funds», *Journal of Finance* 45, pp. 497-521.

Dybvig, P. (1983): «An Explicit Bound on Individual Assets' Deviations from APT Pricing in a Finite Economy», *Journal of Financial Economics* 12, pp. 483-496.

Dybvig, P. and Ross, S. (1985a): «Differential Information and Performance Measurement Using a Security Market Line», *Journal of Finance* 40, pp. 383-399.

Dybvig, P. and Ross, S. (1985b): The Analytics of Performance Measurement Using a Security Market Line», *Journal of Finance* 40, pp. 401-416.

Elton, E. and Gruber, M. (1991): «Differential Information and Timing Ability», *Journal of Banking and Finance* 15, pp. 117-131.

Fama, E. (1972): «Components of Investment Performance», *Journal of Finance* 27, pp. 233-265.

Grinblatt, M. (1986): «How to Evaluate a Portfolio Manager», *Journal of Financial Markets and Portfolio Management* 1, pp. 9-20.

Grinblatt, M. and Titman, S. (1983): «Factor Pricing in a Finite Economy», *Journal of Financial Economics* 12, pp. 497-507.

Grinblatt, M. and Titman, S. (1987): «The Relation Between Mean-Variance Efficiency and Arbitrage Pricing», *Journal of Business* 60, pp. 97-112.

Grinblatt, M. and Titman, S. (1988): «The Evaluation of Mutual Fund Performance: An Analysis of Monthly Returns», Working Paper, University of California, Los Angeles.

Grinblatt, M. and Titman, S. (1989a): «Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings», *Journal of Business* 62, pp. 393-416.

Grinblatt, M. and Titman, S. (1989b): «Portfolio Performance Evaluation: Old Issues and New Insights», *Review of Financial Studies* 2, pp. 393-421.

Hendricks, D.; Patel, J. and Zeckhauser, R. (1990): «Hot Hands in Mutual Funds: The Persistence of Performance», Working Paper, The Kennedy School of Government, Harvard University.

Henriksson, R. (1984): «Market Timing and Mutual Fund Performance: An Empirical Investigation», *Journal of Business* 57, pp. 73-96.

Henriksson, R. and Merton, R. (1981): «On Market Timing and Investment Performance: Statistical Procedures for Evaluating Forecasting Skills», *Journal of Business* 54, pp. 513-533.

Jagannathan, R. and Korajczyk, R. (1986): «Assessing the Market Timing Performance of Managed Portfolios», *Journal of Business* 59, pp. 217-235.

Jensen, M. (1968): «The Performance of Mutual Funds in the Period 1945-1964», *Journal of Finance* 23, pp. 389-416.

Jensen, M. (1972): «Optimal Utilization of Market Forecasts and the Evaluation of Investment Portfolio Performance». In *Mathematical Methods in Investment and Finance,* Edited by G. Szego and K. Shell, North Holland.

Lehmann, B. and Modest, D. (1987): «Mutual Fund Performance Evaluation: A Comparison of Benchmarks and Benchmark Comparisons», *Journal of Finance* 42, pp. 233-265.

Lintner, J. (1965): «The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets», *Review of Economics and Statistics* 47, pp. 13-37.

Merton, R. (1973): «An Intertemporal Capital Asset Pricing Model», *Econometrica* 41, pp. 867-887.

Merton, R. (1981): «On Market Timing and Investment Performance: An Equilibrium Theory of Value for Market Forecasts», *Journal of Business* 54, pp. 363-406.

Mossin, J. (1966): «Equilibrium in a Capital Asset Market», *Econometrica* 35, pp. 768-783.

Pfleiderer, P. and Bhattacharya, S. (1983): «A Note on Performance Evaluation», Working Paper, Stanford University.

Roll, R. (1977): «A Critique of the Asset Pricing Theory's Tests: On Past and Potential Testability of the Theory», *Journal of Financial Economics* 4, pp. 129-176.

Roll, R. (1978): «Ambiguity when Performance is Measured by the Security Market Line», *Journal of Finance* 33, pp. 1051-1069.

Ross, S. (1976): «The Arbitrage Theory of Capital Asset Pricing», *Journal of Economic Theory* 13, pp. 341-360.

Rubinstein, M. (1976): «The Valuation of Uncertain Income Streams and the Pricing of Options», *Bell Journal of Economics* 7, pp. 407-425.

Rubio, G. (1992): «Further Evidence on Performance Evaluation: Portfolio Holdings, Recommendations and Turnover Costs». Institute of Business and Economic Research, Finance Working Series no. 222, Haas School of Business, University of California at Berkeley.

Sharpe, W. (1964): «Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk», *Journal of Finance* 19, pp. 425-442.

Sharpe, W. (1966): «Mutual Fund Performance», *Journal of Business* 39, pp. 119-138.

Treynor, J. (1965): «How to Rate Management of Investment Funds», *Harvard Business Review* 43, pp. 63-75.

Treynor, J. and Mazuy, F. (1966): «Can Mutual Funds Outguess the Market?», *Harvard Business Review* 44, pp. 347-368.

Wang, J. (1990): «A Model of Intertemporal Asset Prices Under Asymmetric Information», Working Paper, Sloan School of Management, MIT.

White, H. (1980): «A Heteroscedasticity-Consistent Covariance Estimator and a Direct Test for Heteroscedasticity», *Econometrica* 48, pp. 817-839.

Abstract

This paper presents a survey of the evaluation techniques of managed portfolios. The paper emphasizes the critical aspects of traditional methodologies, and analyze new alternatives. A discussion about the conceptual issues behind the separation of selectivity information and timing information is also presented.
An empirical application of the techniques is briefly reported. The evidence seems to indicate that the results are robust to reasonable benchmarks and measurement techniques.

Recepción del original, octubre de 1991
Versión final, diciembre de 1992