Journal of Business Finance & Accounting, 33(9) & (10), 1484–1507, November/December 2006, 0306-686X
doi: 10.1111/j.1468-5957.2006.00636.x

Seasonality, Market Timing and Performance Amongst Benchmarks and Mutual Fund Evaluation

Juan C. Matallín-Sáez*

Abstract: Mutual fund performance is normally measured by comparing results of active management with those obtained by one or several benchmarks that should represent the fund's investment. In this context, this paper examines the effect on mutual fund assessment if a relevant benchmark is omitted. This effect is analysed in three elements of active management: stock selection, market timing, and seasonality. The latter is defined as fund management at specific moments of time with the objective of achieving positive abnormal returns to improve performance. For a sample of Spanish mutual funds, we find that the omission of style benchmarks, particularly that corresponding to small-cap stocks, leads to greater evidence of negative market timing and positive seasonality at year beginning. However, the positive abnormal returns of the seasonality at year end, month end and especially at the beginning of July hold regardless of benchmark omission. The paper therefore also analyses the relation between performance and seasonality, finding that positive seasonality at year beginning and at July beginning improves performance; however, at other moments it implies a possible window dressing strategy in mutual fund management.

Keywords: mutual funds, omitted benchmarks, performance, market timing, seasonality

1. INTRODUCTION

The growth in the mutual funds industry has led to the development of a large body of literature on portfolio management with a particular focus on mutual fund performance. The assessment of mutual fund management is a subject of great

* The author is from the Finance and Accounting Department, Universitat Jaume I. He wishes to express his gratitude to an anonymous referee for providing helpful comments and suggestions, to María Paz Jordá of the University of Valencia for her support in obtaining Spanish market data from *Intertell* and to Carlos Forner of the University of Alacant for his comments about the momentum factor in the Spanish stock market. This study is part of research project P1 1A2004-08, supported by Fundació Bancaixa-Castelló. The author also wishes thank the Public Relations department of the CNMV (Spanish Securities and Markets National Commission), AFI (International Financial Analysts) and MSCI (Morgan Stanley Capital International) for making their data available. (Paper received September 2003, revised version accepted April 2006. Online publication August 2006)

Address for correspondence: Juan C. Matallín-Sáez, Dpto. de Finanzas y Contabilidad, Universitat Jaume I, E-12080 Castellón, Spain.
e-mail: matallin@cofin.uji.es

interest to investors, managers, pensions plans and funds of funds among others. In this context, three elements of active portfolio management may be considered to assess mutual fund. The first of these is performance, strictly defined as managers' skills in improving the results of passive management. The second element is market timing, or managers' ability to anticipate market movements and adjust the level of mutual fund risk accordingly. Most mutual fund studies divide performance into two components: market timing and stock selection, or macroforecasting and microforecasting skills respectively in Holmes and Faff (2004). Stock selection involves managers' ability to include (remove) undervalued (overvalued) stocks in (from) the portfolio. The third element of active management is seasonality, which we define as management at specific moments of time, also with the objective of improving mutual fund performance.

If an empirical analysis of mutual funds assessment is carried out, the evaluator requires information about mutual fund management. An internal evaluator may have access to detailed information on managers' decisions and portfolio holdings. For an external evaluator, however, this information is not readily available. Thus, the vast majority of mutual fund papers use returns data, adjusting them using a linear model that includes a set of risk factors or benchmarks and also other regressors that assess active mutual fund management. Thus, for instance, the constant term or the mean tracking error of this model is the abnormal return that usually measures performance. But, is this a true picture? This paper aims to show how one must be cautious when assessing non-observable active management and linking abnormal returns with it. Specifically, we analyse the effect of omitting relevant benchmarks when adjusting mutual funds returns. This is a crucial issue, as the model estimates could be misinterpreted.

For example, let us suppose a mutual fund with a style that invests 50% in large-cap stocks and the remaining 50% in small-cap stocks. Suppose that this mutual fund is assessed against a single benchmark, such as a weighted stock market index that, by size weighting, is closer to large-cap stocks. This mutual fund can therefore show performance, market timing ability and seasonality that are not the result of active management, but rather a consequence of the fact that the small-cap stocks have behaved this way in performance, market timing and seasonality compared with this single benchmark. If, for instance, over the sample period the small-cap stocks have achieved better (worse) results that large-cap, the constant term of the linear model might be positive (negative), in which case one could infer that managers had done a good (poor) job, even though in fact they did nothing. This is the effect of omitting benchmarks relevant to the mutual fund evaluation. To solve this problem, a specific benchmark must be included, in this case for small-cap stocks.

In this vein, Sharpe (1992) and Elton et al. (1993) demonstrate the convenience of incorporating benchmarks that represent all classes of assets making up the mutual fund portfolio. More specifically, Pástor and Stambaugh (2002) analyse the effect on performance of omitting these benchmarks. These studies provide evidence that standard performance measures can be improved by including a greater number of benchmarks. In view of this evidence, the present paper contributes by extending the analysis of the effect of omitting benchmarks to the study of market timing ability and seasonality. We show algebraically how the estimates of the measurement of performance, timing, and seasonality could be biased if the mutual fund invests in stocks not represented by the benchmarks. In addition, in an attempt to assess the economic impact of this effect, an empirical study is undertaken. Hence, the paper shows that the more micro the assessment of the mutual fund management is, the more relevant the effect of omitting benchmarks will be. In this sense, the analysis of timing and seasonality is especially sensitive. Thus, part of the previous evidence of timing and seasonality is not a result of active management, but an implicit result of the stocks in the portfolio.

These results show the weakness of the link between abnormal returns and active management. This contribution is especially relevant to the financial literature since it can be extended to other research areas that focus on the abnormal returns or the tracking errors of linear models, such as event studies, among others. In fact, over the past decade, much of the financial literature on asset pricing has been developed by adding new regressors with power to explain the error term of previous linear models, the first of these being the market model.

On the other hand, the existence of relations of performance, market timing and seasonality within a set of benchmarks is also of interest in mutual fund asset allocation. If persistence exists in these relations, managers could use this information to improve mutual fund evaluation through asset allocation. So for instance, let us suppose two benchmarks with a positive linear relation. Thus, the presence of 'market timing' between these benchmarks is in fact an effect of the asymmetry between the returns on these classes of assets, since by definition, the benchmarks are passive portfolios. In this way, the evidence of, for example, perverse market timing would indicate negative asymmetry. This means that in the presence of down (up) markets, the relation between the benchmarks increases (diminishes), and in consequence the benefits of diversification through asset allocation diminish.

In the empirical work of the paper, the effect of benchmark omission is analysed for a sample of Spanish equity mutual funds. The Spanish mutual funds industry does not differ greatly from the industries of other markets in Europe or the US. However, as its development essentially took place during the second half of the 1990s, the Spanish industry has not yet reached the same maturity as other countries. This fact has conditioned the period sample length for empirical studies on Spanish mutual funds. However, the results of these studies are not substantially different from findings in other markets. Most empirical studies on mutual fund performance evaluation conclude that mutual funds, on average, underperform the appropriate benchmark return. For the Spanish market, Rubio (1995), Matallín and Fernández (2000), Menéndez and Álvarez (2000) and Martínez (2003) have confirmed this result. Furthermore, in line with the results of other markets, Rubio (1995) and Matallín (2002) find evidence of negative market timing when the quadratic market model is applied. The present study also aims to supplement the existing literature on Spanish mutual funds. On this issue, we are not aware of any previous papers that analyse seasonality, and more specifically, the effect of omitting benchmarks.

The following section outlines the methodology employed to assess the three elements of active management: performance, market timing ability and seasonality, and also analyses the effect of benchmark omission. This is followed by a description of the characteristics of the sample of mutual funds used in the empirical work. The final sections present the results and conclusions.

2. PORTFOLIO EVALUATION AND THE EFFECT OF BENCHMARK OMISSION

The literature on mutual fund performance is extensive and includes a large number of theoretical models and empirical applications. A common thread in this literature is the comparison of the results obtained from the active management of a mutual fund with those obtained by a passive portfolio that replicates its risk or style. As Sharpe (1991) argues, if a mutual fund aims to improve the performance of a benchmark, it has to differentiate from this benchmark. Thus, the tracking error between the return of a mutual fund and the benchmark would have to be, on average, positive and exceed the costs of active management.

When the available information takes the form of portfolio holdings, active management can be directly compared with passive management. When the data available are returns, we have to propose a model that allows us to compare mutual fund returns with passive benchmark returns. From Sharpe's ratio (1966), measures have been developed such as those by Ackermann et al. (1999) and Graham and Harvey (2001), which compare the mutual fund with a passive portfolio with the same total risk. The evolution of asset pricing models has led to the development of linear models which adjust: to systematic risk in Jensen (1968); to evaluation through options in Glosten and Jagannathan (1994); to public information in the conditional approach of Ferson and Schadt (1996); and to various risk and style factors in Sharpe (1992), Elton et al. (1993) and Carhart et al. (2002). In this paper, we show the effect of benchmark omission in linear models. To this effect, we propose the broad model (1) that jointly measures performance, market timing and seasonality. From this, we will also apply shorter models, selecting some exogenous variables, to analyse the effect of their introduction and of omitting the benchmarks:

$$r_{p,t} = \alpha_p + \sum_{j}^{J} \beta_{pj} r_{j,t} + \sum_{j}^{J} \gamma_{pj} r_{j,t}^2 + \sum_{s,S} s_p S_t + \varepsilon_{p,t}. \tag{1}$$

In model (1) $r_{p,t}$ and $r_{j,t}$ are respectively, the return on t of portfolio p and of benchmark j, both in excess of the risk-free return. The constant term α_p measures the fund's performance once the return has been adjusted to the J factors. When the stock market is the sole factor, this constant is Jensen's alpha. The second sum of the model (1) generalises the Treynor and Mazuy (1966) market timing measurement for J benchmarks.

Market timing is one of the elements of active management defined as the ability of managers to anticipate market movements and modify the risk position of the portfolio accordingly. Various measures can be applied to evaluate market timing ability.[1] If the available information consists of returns, decisions on market timing must be incorporated into the return generating process. When the stock market is the most relevant factor, Treynor and Mazuy (1966) propose a model with a dynamic beta as the sum of a constant as a long term average beta and a slope γ_{pm} which measures the market timing ability, following a strategy which increases (decreases) the beta when the excess market return is positive (negative). When the dynamic beta is established in the classic model we obtain the Treynor and Mazuy quadratic regression,[2] where α_p

1 The application of a market timing measure depends on the information available to the evaluator: the ideal would be to analyse managers' timing decisions (e.g. Chance and Hemler, 2001). However, in practice these decisions cannot be observed. Information from portfolio holdings can then be used (e.g. Grinblatt and Titman, 1989b). However, this information is less accessible than return data.

2 Another classic timing measure is that by Henriksson and Merton (1981), which establishes two different risk levels, and thus, a dichotomous asymmetric effect. However, in the face of the empirical evidence, the Treynor and Mazuy measure is more appropriate since it incorporates a gradual asymmetric effect.

measures the asset selection. To date, many extensions and applications of this measure have been conducted (e.g. Bollen and Busse, 2001).

The third sum of model (1) measures another possible element of active management, i.e. seasonality, defined as particular management in certain moments of time in order to achieve specific targets. One of these targets might be the improvement in the ranking of a mutual fund within its category, for instance by means of temporal changes in the risk position (e.g. Brown et al., 1996; and Koski and Pontiff, 1999). Within this line, Haugen and Lakonishok (1988) investigate the relation between mutual fund manager behaviour and the January effect. Lee et al. (1998) analyse two hypotheses for this behaviour: window dressing and performance hedging. Hedging suggests that managers habitually track a benchmark, but at the beginning of January, change their strategy and buy, in particular, small-cap stocks in order to beat the benchmark and to improve performance. Window dressing, as defined by Lee et al. (1998), implies that at the end of the year, managers present portfolio holdings that essentially include well-known securities, but when the new year begins, these are changed to include less well-known securities and small-cap stocks, also with the aim of beating the benchmark. Their results show that manager behaviour is the result of hedging and not of window dressing. Carhart et al. (2002) define a window dressing strategy as one that aims to artificially increase the mutual fund net asset value. By using information on daily returns, they find evidence of seasonality. Holmes and Faff (2004) analyse monthly seasonality in a sample of Australian mutual funds. Their results confirm anomalous returns for the months of January and July, especially in the latter case.

Following some of the methods proposed by Carhart et al. (2002), in model (1) we use dummies, S_t, that take the value of 1 on the days indicated and zero on the remaining days. Holmes and Faff (2004) also use dummy variables to measure anomalous results to analyse seasonal market effects. By defining dummies, we also attempt to capture the existence of anomalous returns at certain moments of time. The third sum of the model (1) expands the next six dummies, S_t. First we define the 'year beginning' variable YB_t and 'year end' variable YE_t which take the value of 1 if t corresponds respectively to the last and first day of the year the stock market is open. The 'half year beginning' variable HB_t and 'half year end' HE_t are similarly defined. The 'month beginning' MB_t and 'month end' ME_t variables are also established if t is respectively the end or the beginning of the month, whenever they are neither the end nor the beginning respectively, of neither the year nor the half year.

To analyse the effect of benchmark omission on performance we put forward a specific case, which can be generalised, in which the excess return on fund p is explained by $J + K$ benchmarks. The works of Sharpe (1992) and Elton et al. (1993) establish the convenience of carrying out evaluations that include benchmarks for all classes of assets in which the fund invests. Thus, the inclusion of K benchmarks is proposed in the fourth sum of the model (2). If a mutual fund invests effectively in $J + K$ asset classes and on evaluating the fund, (1) is proposed instead of (2), benchmarks are omitted and therefore, the fund assessment in (1) will be attributable in part to the relations between benchmarks and not to managers:

$$r_{p,t} = \alpha^*_p + \sum_j^J \beta^*_{pj} r_{j,t} + \sum_j^J \gamma^*_{pj} r^2_{j,t} + \sum_{s,S} s^*_p S_t + \sum_k^K \beta^*_{pk} r_{k,t} + \varepsilon^*_{p,t}. \tag{2}$$

Pástor and Stambaugh (2002) assume a linear relation between the benchmarks and evaluate the omission effect in performance. We extend their analysis to include the effect on market timing and seasonality evaluation. Expression (3) shows the linear relation between the omitted benchmark k and the benchmarks considered initially in (1):

$$r_{k,t} = \alpha_k + \sum_j^J \beta_{kj} r_{j,t} + \sum_j^J \gamma_{kj} r_{j,t}^2 + \sum_{s,S} s_k S_t + u_{k,t}. \tag{3}$$

When (3) is substituted into (2), we obtain (4) and on arranging terms (5). On comparing (5) with (1), expressions (6)–(10) are put forward showing the impact of benchmark omission on portfolio assessment:

$$\begin{aligned} r_{p,t} = {} & \alpha_p^* + \sum_j^J \beta_{pj}^* r_{j,t} + \sum_j^J \gamma_{pj}^* r_{j,t}^2 + \sum_{s,S} s_p^* S_t + \varepsilon_{p,t}^* \\ & + \sum_k^K \beta_{pk}^* \alpha_k + \sum_k^K \beta_{pk}^* \sum_j^J \beta_{kj} r_{j,t} + \sum_k^K \beta_{pk}^* \sum_j^J \gamma_{kj} r_{j,t}^2 + \sum_k^K \beta_{pk}^* \sum_{s,S} s_k S_t + \sum_k^K \beta_{pk}^* u_{k,t} \end{aligned} \tag{4}$$

$$\begin{aligned} r_{p,t} = {} & \alpha_p^* + \sum_k^K \beta_{pk}^* \alpha_k + \sum_j^J r_{j,t} \left(\beta_{pj}^* + \sum_k^K \beta_{pk}^* \beta_{kj} \right) \\ & + \sum_j^J r_{j,t}^2 \left(\gamma_{pj}^* + \sum_k^K \beta_{pk}^* \gamma_{kj} \right) + \sum_{s,S} S_t \left(s_p^* + \sum_k^K \beta_{pk}^* s_k \right) + \varepsilon_{p,t}^* + \sum_k^K \beta_{pk}^* u_{k,t} \end{aligned} \tag{5}$$

$$r_{p,t} = \alpha_p + \sum_j^J \beta_{pj} r_{j,t} + \sum_j^J \gamma_{pj} r_{j,t}^2 + \sum_{s,S} s_p S_t + \varepsilon_{p,t} \tag{1}$$

$$\alpha_p = \alpha_p^* + \sum_k^K \beta_{pk}^* \alpha_k \tag{6}$$

$$\beta_{pj} = \beta_{pj}^* + \sum_k^K \beta_{pk}^* \beta_{kj} \tag{7}$$

$$\gamma_{pj} = \gamma_{pj}^* + \sum_k^K \beta_{pk}^* \gamma_{kj} \tag{8}$$

$$s_p = s_p^* + \sum_k^K \beta_{pk}^* s_k \tag{9}$$

$$\varepsilon_{p,t} = \varepsilon_{p,t}^* + \sum_k^K \beta_{pk}^* u_{k,t}. \tag{10}$$

As we would expect, the greater β_{pk}^* is for the portfolio evaluated, the greater the effect of the omission will be in (6)–(10). The bias in the estimation of the parameters and in the variance of $\varepsilon_{p,t}$ is also the result of transferring to (1) the whole cross-sectional and temporal structure of variances and covariances between all the $u_{k,t}$ and $\varepsilon_{p,t}^*$. If we focus on the effect on performance, the sign of the intercept relating the benchmarks implies a bias of the same sign in the performance measurement in (6). In other words, the estimated alpha in (1) is the sum of the true alpha, α_p^*, and the bias resulting from the sum of the products between β_{pk}^* and the performance, α_k, of the omitted benchmark k with regard to the J benchmarks solely considered in (1).

In (8), the true timing ability is measured by γ_{pj}^* and the result of the sum of $\beta_{pk}^*\gamma_{kj}$ is the bias when benchmarks are omitted. It should be taken into account that, given that the benchmarks k and j are by definition passive, γ_{kj} does not measure the timing ability, but rather an asymmetric effect between these benchmarks. If γ_{kj} is significantly positive, it indicates that the slope between the benchmarks is greater (less) when $r_{j,t}$ is positive (negative). However, if this parameter is negative, the result will be the opposite. If for example, we suppose that the true timing is zero and that the result of the sum of $\beta_{pk}^*\gamma_{kj}$ is positive (negative), it would imply that the relation between the benchmarks is increasing (decreasing), and therefore γ_{pj} would present the same sign, from which it would be spuriously deduced that the fund evaluated has carried out correct (perverse) market timing with respect to benchmark j.

Finally, if the existence of seasonality between benchmarks is considered, expression (9) shows the effect of benchmark omission on mutual fund seasonality measurement. In the same way as for performance and timing measurement, if seasonality exists between the omitted benchmarks and those used to assess the mutual fund, evidence of seasonality may be found, not as a result of management, but rather as a spurious result of this omission.

3. SAMPLE SELECTION AND DATA

The Spanish mutual funds sector has expanded considerably in the last decade, placing Spain third in Europe and sixth in the world in terms of the number of mutual funds managed.[3] From this sector, the empirical analysis of this paper is carried out on the mutual funds that mainly invest in the Spanish stock market. So, a sample of 220 Spanish mutual funds FIM (*Fondos de Inversión Mobiliaria*) from July 1998 to September 2004 was used in the empirical analysis. This sample was made up of all the mutual funds for which complete data during this period are available.[4] This could create a survivorship bias caused by the mutual funds not included in this sample, i.e. both missing funds and new funds during this period.[5] To assess this bias we created two further samples: the first with 37 non-survivor funds, missing in the second half of the sample period, and the second with 55 new funds that were launched in the first half. The empirical analysis is also applied to these two samples, because even though the length of these time series is only three complete years, this will report more about mutual fund management than if they were totally omitted.

The funds included in the sample, like nearly all Spanish mutual funds, are accumulation funds which do not pay dividends. Thus, the daily return was calculated

3 According to the Mutual Fund Fact Book (46th Edition, 2006) published by the Investment Company Institute.

4 The Spanish mutual fund industry developed primarily in the second half of the 1990s. For this reason, there is an inverse relationship between the length of the time series and the number of funds with complete data. If we had selected an earlier start date, the number of funds in the sample would have been drastically reduced. Thus, the use of daily data has enabled us to analyse a large volume of information for all the funds existing on 30 June, 1998.

5 Literature reviews conducted by Elton, Gruber and Blake (1996) and Holmes and Faff (2004) provide evidence of the interest in survivorship bias and its effect in portfolio performance and persistence.

as the variation relative to the daily net asset value of the mutual fund. This value, which is taken as the price for investors' purchases and refunds, is computed net of fees and trading costs and therefore returns are also net to investors. The daily data were provided by the Spanish Securities and Markets National Commission (CNMV) and are free from asynchronous biases.[6] In accordance with the CNMV classification, four categories ranging from the highest to the lowest percentage of investment in stocks as opposed to bonds were differentiated: variable income (VI), mixed variable income (MVI), mixed fixed income (MFI) and fixed income (FI). The funds studied were predominantly risky investments in the domestic market, in the variable income and mixed variable income categories. The use of daily data avoids the presence of biases derived from greater frequencies when market timing is measured, as shown by Goetzmann et al. (2000) and Bollen and Busse (2001). Moreover, for the seasonality analysis we propose, the use of daily data is important when, as in this case, returns are the only information that is normally available with this frequency.

We used the following benchmarks to represent the classes of assets held by the mutual funds in the sample. The first, $r_{m,t}$, is the excess return of the Ibex 35 index, proxy for the Spanish stock market, which is a value-weighted index with the 35 most-traded stocks. This index is taken as the underlying reference asset in the Spanish derivatives market and is commonly used as a benchmark for Spanish equity mutual funds. The second, $r_{d,t}$, is the excess return of the AFI Government Debt fixed income index, a portfolio made up of medium and long term maturity Treasury bonds. To enlarge the number of benchmarks, we used the Morgan Stanley Capital International (MSCI) style indexes for the Spanish market: the small-cap index which covers stocks with lower stock market size ($r_{s,t}$); the growth index which includes stocks with higher price-to-book ratios ($r_{g,t}$); and the value index with equities of a low value for this ratio ($r_{v,t}$). These benchmarks have also been used by other previous studies on mutual fund performance and they are not adjusted by dividends. The return implicit in T-bills repos was used as the risk free return. Data sources used were Internet information providers from the *Sociedad de Bolsas*, the *Analistas Financieros Internacionales* (AFI) and the MSCI. We have not found a momentum benchmark available for the Spanish market, although researchers such as Forner and Marhuenda (2003) have constructed a monthly momentum factor to analyse asset pricing in this market. Thus, following Carhart (1997) and Forner and Marhuenda (2003), we construct a zero-investment portfolio as the equal-weight average of firms, in the Spanish stock market, with the highest 30 percent eleven-month returns lagged one month minus the respective lowest 30 percent. The momentum portfolio ($r_{n,t}$) is re-balanced monthly, but the returns are calculated daily. The stock prices data was provide by *Intertell*. From this data, we can specify the expression (1) to (11), to be applied in the empirical analysis:

$$\begin{aligned} r_{p,t} &= \alpha_p + \beta_{pm} r_{m,t} + \beta_{pd} r_{d,t} + \beta_{ps} r_{s,t} + \beta_{pg} r_{g,t} + \beta_{pv} r_{v,t} + \beta_{pn} r_{n,t} + \gamma_{pm} r^2_{m,t} \\ &\quad + yb_p \mathrm{YB}_t + ye_p \mathrm{YE}_t + hb_p \mathrm{HB}_t + he_p \mathrm{HE}_t + mb_p \mathrm{MB}_t + me_p \mathrm{ME}_t + \varepsilon_{p,t}. \quad (11) \end{aligned}$$

6 Basarrate and Rubio (1999) found that net asset values for some mutual funds are not correctly matched with their respective dates. This problem, which they term nonsimultaneous prices, was evidenced using the Madrid Stock Exchange Bulletin database. The problem occurs in this database because the net asset data is provided periodically by the fund management company and different criteria could be used when reporting this data. However, we use a different database, from the CNMV, which is quarterly and collected directly from the funds' accounting books and therefore this problem of asynchronousity does not arise.

Table 1 shows some descriptive statistics on the mutual funds sample and benchmarks. To carry out an aggregate analysis, two portfolios were also formed from the returns on the individual mutual funds. Thus, EFVI and EFMVI are equally weighted funds made up of variable income and mixed income mutual funds respectively. A non-survivor fund sample is made up for missing funds after August 2001 and new funds created after this date. Then, aggregated and equally weighted portfolios are formed, EFVI-9801 and EFVI-0104, respectively for 16 missing and 14 new VI type funds, and EFMVI-9801 and EFMVI-0104 for 21 missing and 41 new MVI mutual funds.

4. RESULTS

(i) Analysis of the Relation Between Benchmarks

We first examine the relations between the benchmarks in order to establish the possible bias arising should any one of them be omitted. Given that the mutual funds analysed invest mainly in stocks, the bias that would occur if the stock market index, m, were omitted is obvious. We therefore analyse the relation of other, *a priori*, less important benchmarks, which are consequently more likely to be omitted. To do this, we propose successive regressions where the endogenous variables are these benchmarks and where performance, market timing and seasonality are analysed.

Table 2 shows the results obtained. The excess return of the stock market, $r_{m,t}$, has hardly any explanatory power on $r_{d,t}$, the excess return of the bonds market. However, when the endogenous variable is $r_{s,t}$, $r_{g,t}$ or $r_{v,t}$, i.e., the excess return of the small-cap, growth and value benchmarks, $r_{m,t}$ does have a major explanatory power. These results give an idea of the effect of omitting these benchmarks, under the hypothesis that the fund evaluated actually invests in these classes of assets. The effect will be greater the higher the mutual fund investment in these assets and the stronger the relation between them and $r_{m,t}$.

The first five models in Table 2 show the performance between benchmarks. For $r_{s,t}$, $r_{g,t}$ and $r_{n,t}$ (momentum factor) in models T2, T3 and T5, the constant α_k is not significant. In model T1 and T4, when the endogenous variable is $r_{d,t}$, and $r_{v,t}$ respectively, the constant takes on a positive and borderline significant value. The bias in the performance will therefore have a positive sign, particularly for funds with a greater investment in bonds and value stocks.

In the next five models in Table 2 measure the 'market timing' ability between benchmarks. Since by definition, the benchmarks are passive portfolios, the 'market timing ability' measure becomes an analysis of the asymmetry in the relation between these benchmarks. First, models T6, T8 and T9 respectively show the relation between $r_{d,t}$, $r_{g,t}$ and $r_{v,t}$ with $r^2_{m,t}$. In these cases, the coefficient γ_{km} of the quadratic term is positive but not significant. However, for the $r_{s,t}$ and $r_{n,t}$ benchmarks in models T7 and T10, the γ_{km} coefficient is negative, which implies a negative asymmetry between the returns of these benchmarks and m, the stock market index. Particularly for small-cap stocks, this asymmetry is markedly both negative and significant. This implies that the relation between the small-cap stocks and the stock market is greater in down markets than in up markets, which leads to a decrease in the possible benefits of diversification. Thus, if we assess the market timing ability of a mutual fund that invests in small-cap assets and omit this benchmark, the measure of market timing ability will be negatively

Table 1

Mutual Funds Sample and Descriptive Statistics of Daily Excess Returns

	Number of Funds	*Max.*	*Min.*	*Mean*	*Median*	*Standard Deviation*	*Skewness*	*Kurtosis*
Survivor mutual funds sample:								
EFVI	79	0.0526	−0.0584	−0.0001	0.0006	0.0129	−0.1007	4.6391
EFMVI	144	0.0263	−0.0294	−0.0002	0.0002	0.0067	−0.0984	4.4624
Non-survivor mutual funds sample:								
EFVI-9801	16	0.0490	−0.0585	−0.0002	0.0007	0.0124	−0.3350	5.2883
EFVI-0104	14	0.0426	−0.0407	−0.0001	0.0003	0.0115	0.1050	4.2706
EFMVI-9801	21	0.0228	−0.0234	−0.0001	0.0000	0.0056	−0.2318	4.2777
EFMVI-0104	41	0.0264	−0.0251	−0.0002	0.0000	0.0066	0.1109	4.5001
Benchmarks:								
$r_{m,t}$ (Ibex 35)		0.0652	−0.0709	−0.0002	0.0006	0.0159	−0.0335	4.7860
$r_{d,t}$ (Bond index)		0.0089	−0.0118	0.0001	0.0002	0.0026	−0.2683	3.7738
$r_{s,t}$ (Small)		0.0470	−0.0528	0.0001	0.0003	0.0093	−0.5175	7.2870
$r_{g,t}$ (Growth)		0.0849	−0.0879	−0.0003	0.0002	0.0201	0.0737	5.1529
$r_{v,t}$ (Value)		0.0810	−0.0673	0.0001	0.0004	0.0146	0.1517	5.4416
$r_{n,t}$ (Momentum)		0.0168	−0.0218	0.0000	0.0000	0.0043	−0.2159	4.8803

Notes:
The table shows information on the mutual funds sample and some descriptive statistics of daily excess returns for the mutual funds and benchmarks. The mutual funds sample is made up of equity mutual funds (VI or variable income funds) and balanced equity-bond mutual funds (MVI or mixed variable income). The sample period for survivor mutual funds sample runs from July 1, 1998 to September 30, 2004. Two aggregated and equally weighted portfolios were formed, EFVI and EFMVI for VI and MVI funds, respectively. A non-survivor funds sample is made up for missing funds after August 2001 and new funds created after this date. Then, aggregated and equally weighted portfolios are formed, EFVI-9801 and EFVI-0104, respectively for VI type missing and new funds, and EFMVI-9801 and EFMVI-0104 for MVI mutual funds. The value of the maximum, minimum, average, median and standard deviations are presented by multiplying by 100.

Table 2

Analysis of the Relation Between Benchmarks

Model	Endog. Variable k	*Exogenous Variable j* Constant	$r_{m,t}$	$r_{d,t}$	$r^2_{m,t}$	YB_t	YE_t	HB_t	HE_t	MB_t	ME_t	Adj. R^2
T1	$r_{d,t}$	0.0108	−3.6827									5.002
		(0.098)	(0.000)									
T2	$r_{s,t}$	0.0146	38.4628	−9.0267								44.476
		(0.476)	(0.000)	(0.247)								
T3	$r_{g,t}$	−0.0102	118.9410	−4.5194								89.428
		(0.544)	(0.000)	(0.470)								
T4	$r_{v,t}$	0.0268	84.4928	−1.2193								84.436
		(0.074)	(0.000)	(0.831)								
T5	$r_{n,t}$	−0.0037	−8.2019	13.0370								10.750
		(0.742)	(0.000)	(0.007)								
T6	$r_{d,t}$	0.0052	−3.6642		22.0868							5.115
		(0.473)	(0.000)		(0.102)							
T7	$r_{s,t}$	0.0722	38.3414	−7.1296	−228.4106							45.915
		(0.000)	(0.000)	(0.360)	(0.000)							
T8	$r_{g,t}$	−0.0116	118.9438	−4.5633	5.2897							89.421
		(0.510)	(0.000)	(0.466)	(0.904)							
T9	$r_{v,t}$	0.0192	84.5088	−1.4691	30.0777							84.436
		(0.231)	(0.000)	(0.796)	(0.502)							
T10	$r_{n,t}$	0.0084	−8.2276	13.4376	−48.2271							10.996
		(0.451)	(0.000)	(0.005)	(0.070)							
T11	$r_{d,t}$	0.0089	−3.7214			0.0965	−0.1395	−0.0655	−0.0334	0.0162	0.0539	4.997
		(0.175)	(0.000)			(0.584)	(0.181)	(0.284)	(0.697)	(0.700)	(0.083)	
T12	$r_{s,t}$	−0.0021	38.0199	−9.6783		1.2552	0.5373	0.2929	0.1010	0.0370	0.1847	45.265
		(0.920)	(0.000)	(0.205)		(0.000)	(0.045)	(0.026)	(0.718)	(0.724)	(0.010)	
T13	$r_{g,t}$	−0.0120	118.9510	−4.1540		−0.2294	−0.0259	0.6349	−0.0199	−0.0198	0.0179	89.438
		(0.512)	(0.000)	(0.510)		(0.363)	(0.892)	(0.003)	(0.940)	(0.797)	(0.803)	
T14	$r_{v,t}$	0.0254	84.4490	−1.2364		0.2892	0.2743	−0.4498	0.0979	0.0792	−0.0566	84.464
		(0.104)	(0.000)	(0.830)		(0.425)	(0.229)	(0.001)	(0.678)	(0.357)	(0.372)	
T15	$r_{n,t}$	−0.0037	−8.0796	13.3958		−0.4412	0.0732	0.0346	0.1157	0.0274	−0.0090	10.871
		(0.755)	(0.000)	(0.005)		(0.327)	(0.650)	(0.822)	(0.194)	(0.618)	(0.840)	

Notes:
The table shows the estimate of the parameters for different combinations of the regressors of the expression (11), where the portfolio assessed is in this case a benchmark. The value of the parameters is presented by multiplying the estimated coefficient by 100. The p-value of each parameter is given in parenthesis. The heteroskedasticity and autocorrelation consistent covariance estimator is by Newey and West (1987).

biased, in such a way that spurious evidence of perverse market timing ability might even be found.

The analysis of benchmark seasonality is presented in models T11 to T15 in Table 2. The first point of interest is that most of the coefficients measuring the seasonality take on a positive value. Thus, the value of α_k for $r_{d,t}$, $r_{s,t}$, $r_{g,t}$ and $r_{v,t,}$ is lower compared with that obtained from models T1 to T4. This is particularly marked in the case of small-cap stocks, where the constant term is 1.14 times more negative. If we look at the row of model T12 for the small-cap benchmark, the positive return at the year beginning is specially important and significant. For small-cap, value and growth benchmarks the parameter for the dummy at the beginning of July is also positive and significant. If we assess a mutual fund that invests in these types of assets and omit these benchmarks the evidence of seasonality could be biased. This might lead us to infer that managers have followed a window dressing or hedging strategy, when in fact such specific management has not occurred.

(ii) Evaluation of Mutual Fund Management

(a) Aggregate Results

Firstly, we carry out an aggregate analysis of the portfolios EFVI and EFMVI which are formed as equally weighted funds made up of variable income (VI) and mixed variable income (MVI) mutual funds respectively.

Different models, from T16 to T23, were proposed, made up by means of a combination of the regressors from model (11). Table 3 presents the results for the EFVI portfolio, where firstly models T16, T18, T20 and T22 analyse the performance, timing and seasonality with the stock market as the only benchmark. Thus, in T16, the performance measured by α_p is negative but not significant, coinciding with most of the literature on mutual funds performance. In T18, market timing is considered, and it is negative and significant with a *p*-value of 0.086. This evidence is common in the financial literature: Treynor and Mazuy (1966), Henriksson (1984), Grinblatt and Titman (1989a), Coggin et al. (1993), Wenchi-Kao et al. (1998), Volkman (1999), Edelen (1999), Jiang (2003) and Holmes and Faff (2004) amongst others, do not find significant evidence of market timing in general, but in some cases they do however, find a large number of mutual funds with perverse timing. The presence of seasonality is evaluated in model T20 with the parameters from yb_p to $me_{p,}$ all of which are positive and also significant, except for the seasonality for the half year end and month beginning.

On the other hand, if we compare the results of models T16, T18, T20 and T22, we can analyse the interaction between performance, timing and seasonality. Thus, when we compare the performance in model T18, which measures timing, with T16, it continues to be non-significant but takes on a higher positive value. The same occurs for the small-cap benchmark in Table 2, so when timing is measured the intercept becomes more positive and significant. The inverse relation between asset selection ability, measured by α_p, and market timing ability is widely reported in the financial literature. However, some authors (e.g., Henriksson, 1984; Jagannathan and Korajczyk, 1986; and Coggin et al., 1993) have explained how this result can be produced spuriously. If we compare model T20, which measures seasonality, with T16, the performance in T20 is lower. This result shows how part of the positive return of the mutual funds was concentrated in specific moments of time. Finally, model T22 jointly

Table 3

Mutual Fund Management Evaluation

$p = EFVI$	Model *T16*	*T17*	*T18*	*T19*	*T20*	*T21*	*T22*	*T23*
α_P	−0.0010	−0.0010	0.0067	0.0025	−0.0069	−0.0060	0.0013	−0.0021
	(0.815)	(0.777)	(0.161)	(0.564)	(0.114)	(0.115)	(0.785)	(0.626)
β_{pm}	80.710	81.740	80.684	81.686	80.598	82.119	80.566	82.057
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_{pd}		−7.322		−7.205		−7.331		−7.211
		(0.000)		(0.000)		(0.000)		(0.000)
β_{ps}		7.083		6.915		6.738		6.545
		(0.000)		(0.000)		(0.000)		(0.000)
β_{pg}		−2.433		−2.390		−2.638		−2.587
		(0.285)		(0.289)		(0.253)		(0.257)
β_{pv}		−1.715		−1.654		−1.817		−1.750
		(0.445)		(0.460)		(0.424)		(0.439)
β_{pn}		−3.452		−3.562		−3.452		−3.559
		(0.011)		(0.008)		(0.012)		(0.009)
γ_{pm}			−30.371	−13.893			−31.829	−15.165
			(0.086)	(0.416)			(0.069)	(0.371)
yb_p					0.242	0.149	0.265	0.162
					(0.023)	(0.197)	(0.023)	(0.180)
ye_p					0.200	0.159	0.193	0.156
					(0.022)	(0.028)	(0.028)	(0.032)
hb_p					0.281	0.266	0.280	0.266
					(0.000)	(0.000)	(0.000)	(0.000)
he_p					0.038	0.034	0.038	0.034
					(0.324)	(0.242)	(0.327)	(0.232)
mb_p					0.020	0.021	0.018	0.020
					(0.385)	(0.341)	(0.425)	(0.357)
me_p					0.055	0.047	0.053	0.046
					(0.005)	(0.018)	(0.006)	(0.002)
Adj. R^2	98.576	98.761	98.588	98.763	98.620	98.790	98.634	98.793

Notes:
Results for the EFVI portfolio, an equi-weighted portfolio made up of variable income mutual funds. Models are run from different combinations of the regressors of the expression (11). Performance is measured by α_p, market timing is evaluated by γ_{pm} and seasonality by yb_p up to me_p. The value of the parameters is presented by multiplying the estimated coefficient by 100. The p-value of each parameter is given in parenthesis. The heteroskedasticity and autocorrelation consistent covariance estimator is by Newey and West (1987).

measures timing and seasonality, and the results are consistent with models T18 and T20.

Below, we analyse the effect of omitting benchmarks by comparing the results in Table 3 of the previous models, T16, T18, T20 and T22 with the corresponding models with style benchmarks, T17, T19, T21 and T23. First, comparing models T16 and T17, the performance measured by α_p is observed to be very similar, and it is not significant in any case. Here, the omission of style benchmarks does not imply an outstanding effect in the aggregated performance of mutual funds. In this respect, we should not forget that in this sample period, the performance between benchmarks, in models T1 to T5, is not significant either, except for the value benchmark, but at a 10% significance level. Moreover, expression (6) shows how this effect is a result of compensation of the biases caused by the omission of these benchmarks, and these were positive for the small-cap and value benchmarks, but negative for growth.

Second, we analyse the effect of benchmark omission on the evaluation of market timing ability, comparing T18 and T19. Thus, when style benchmarks are introduced in T19 the γ_{pm} parameter that measures timing is less negative and is no longer significant. This effect could be due fundamentally to the introduction of the small-cap benchmark, because it presents an important and negative asymmetry or 'perverse market timing' in model T7 in Table 2. If mutual funds hold small-cap stocks and this benchmark is omitted, we could mistakenly infer evidence of perverse market timing solely because these stocks have shown this asymmetry.

Third, models T20 and T21 show the effect on seasonality measurement. Thus, the seasonality at the beginning of year yb_p is positive and significant in T20, but is not significant in T21 when the style benchmarks were not omitted. This result is consistent with the evidence for seasonality at the beginning of the year in model T12 in Table 2 for the small-cap benchmark. Therefore, if we omit this benchmark, we may infer that managers had followed a window dressing or hedging strategy as defined by Lee et al. (1998) at the beginning of January. However, this result is linked with the small stocks held by the mutual fund, since they show an important January effect (e.g. Keim, 1983, 1989 and 1999). Year-end seasonality ye_p is positive and significant in the models, even after style benchmarks have been included. Although year-end seasonality in T12 in Table 2 is significant for the small cap benchmark, this evidence holds when this benchmark is included. Consistently, we could infer that the evidence of positive year-end seasonality is linked to mutual fund management. Furthermore, the significance of seasonality at the beginning of the second half year hb_p is positive and clearly significant in all models. As can be observed, there are practically no variations in the value of hb_p nor in its significance, which indicates that the abnormal return at the beginning of July cannot be explained by asset allocation as in the case of yb_p. Also, Holmes and Faff (2004), using monthly data, find a substantially higher performance of Australian mutual funds during the month of July. The seasonality corresponding to the end of the half year he_p is not significant in any of the models proposed, and the same holds for that corresponding to month beginning mb_p. Analysis of quarterly seasonality was also carried out, which was not significant either at the beginning or the end of each period. On the other hand, month-end seasonality is significant and positive in all models.

Lastly, from T22 and T23 the joint effect of timing and seasonality measurement and omitting benchmarks is analysed. It should be noted that on introducing the seasonality measure in T22, the evidence of negative perverse market timing is greater than in model T18. It can also be observed that the seasonality at the beginning of year is particularly significant, which could be interpreted as a moment of positive timing. The dummy captures the positive abnormal return at this moment and thus the evidence of negative timing is greater. However when the style benchmarks are introduced in T23, the evidence of negative timing and year beginning abnormal returns disappear, consistent with the results pointed out above.

Table 4 shows the results of the EFMVI portfolio, made up of mutual funds from the mixed variable income group. As was to be expected, the risk in this fund was lower than that presented for the EFVI. With regard to performance, the constant takes on more negative values than in the previous cases, and is significant in most of the models, especially when seasonality is measured and style benchmarks are considered. The parameter measuring market timing takes on negative values but they are not significant. As with the EFVI portfolio, when the style benchmarks are included, evidence of perverse market timing is reduced. In addition, seasonality at year-end and the beginning of the second half year is significant and positive.

(b) Non-parametric Approach to Seasonality Evaluation

We also carried out an analysis of robustness to study the effect of omitted benchmarks in the seasonality evaluation. This analysis uses the Kruskal-Wallis non-parametric test applied to the residuals of models T18 and T19 obtained by adjusting the aggregate portfolios, EFVI in Table 3 and EFMVI in Table 4. Whereas dummies capture an anomalous return on average, with the K-W statistic we contrast the null hypothesis that on days that define seasonality the residuals have an identical distribution to the residuals from the remaining days. This analysis is more robust with the presence of outliers and is especially pertinent considering the limited number of days that define seasonality.

Table 5 reports the results of the K-W test for the seasonality analysis. For the EFVI portfolio and model T18, the distributions of the residuals on year beginning, year end, half year beginning and end month respectively, are not the same as for the days without seasonality. These results are consistent with the evidence for model T22 with dummies in Table 3. The results from the residuals of model T19 are also consistent with respect to model T23 with dummies in Table 3. According to the results obtained from passing from model T22 to T23, by introducing style benchmarks, the evidence of seasonality at the year beginning vanishes. In relation to the EFMVI portfolio and comparing the results in Table 5 with those in Table 4, the evidence of seasonality at year end coincides; however in Table 5 the residuals at half year beginning do not differ significantly from the remaining days.

(c) Individual Results

Models T16-T23 are applied individually to the mutual funds in the sample. The panels in Table 6 summarise the results obtained for VI mutual funds, which are consistent with the aggregate results shown above. In performance measurement, practically all the alphas are not significant. Only when market timing is evaluated do some positive and significant alphas appear, but as we pointed out for the aggregated results, this is an artificial result (Henriksson, 1984; Jagannathan and Korajczyk, 1986; and Coggin et al., 1993). In Panel A, the number of negative alphas is higher when the seasonality measurement in models T20 and T21 is incorporated than in T16 and T17 when it is left out. We could take this result as evidence of how seasonality management in mutual

Table 4

Mutual Fund Management Evaluation

	Model							
$p = EFMVI$	*T16*	*T17*	*T18*	*T19*	*T20*	*T21*	*T22*	*T23*
α_p	−0.0091	−0.0090	−0.0060	−0.0090	−0.0127	−0.0119	−0.0098	−0.0122
	(0.037)	(0.029)	(0.216)	(0.050)	(0.006)	(0.006)	(0.048)	(0.009)
β_{pm}	40.494	40.527	40.484	40.526	40.465	40.839	40.453	40.844
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
β_{pd}		−7.239		−7.239		−7.119		−7.129
		(0.000)		(0.000)		(0.000)		(0.000)
β_{ps}		5.032		5.031		4.926		4.942
		(0.000)		(0.000)		(0.000)		(0.000)
β_{pg}		−1.508		−1.508		−1.657		−1.661
		(0.475)		(0.477)		(0.438)		(0.439)
β_{pv}		−0.867		−0.866		−0.999		−1.004
		(0.699)		(0.701)		(0.658)		(0.659)
β_{pn}		−3.200		−3.200		−3.348		−3.340
		(0.015)		(0.016)		(0.001)		(0.011)
γ_{pm}			−12.129	−0.087			−11.343	1.230
			(0.444)	(0.996)			(0.478)	(0.939)
yb_p					0.034	−0.036	0.042	−0.037
					(0.756)	(0.748)	(0.709)	(0.742)
ye_p					0.201	0.168	0.199	0.168
					(0.000)	(0.000)	(0.000)	(0.000)
hb_p					0.094	0.081	0.093	0.081
					(0.009)	(0.041)	(0.009)	(0.041)
he_p					0.073	0.070	0.073	0.070
					(0.136)	(0.122)	(0.140)	(0.122)
mb_p					0.019	0.020	0.018	0.020
					(0.479)	(0.456)	(0.490)	(0.452)
me_p					0.034	0.029	0.033	0.029
					(0.126)	(0.198)	(0.135)	(0.199)
Adj. R^2	93.663	94.066	93.667	94.062	93.698	94.088	93.701	94.084

Notes:
Results for the EFMVI portfolio, an equi-weighted portfolio made up of mixed variable income mutual funds. Models are run from different combinations of the regressors of the expression (11). Performance is measured by α_p, market timing is evaluated by γ_{pm} and seasonality by yb_p up to me_p. The value of the parameters is presented by multiplying the estimated coefficient by 100. The p-value of each parameter is given in parenthesis. The heteroskedasticity and autocorrelation consistent covariance estimator is by Newey and West (1987).

Table 5
Seasonality and the Kruskal-Wallis Statistic

	p = EFVI		*p = EFMVI*	
	Model in Table 3		*Model in Table 4*	
	T18	*T19*	*T18*	*T19*
YB	4.407	1.927	0.006	0.533
	(0.036)	(0.165)	(0.938)	(0.465)
YE	3.828	3.511	8.752	7.489
	(0.050)	(0.061)	(0.003)	(0.006)
HB	17.565	17.144	2.906	2.040
	(0.000)	(0.000)	(0.088)	(0.153)
HE	0.708	0.662	1.577	1.689
	(0.400)	(0.416)	(0.209)	(0.194)
MB	0.939	1.587	0.445	0.584
	(0.333)	(0.208)	(0.505)	(0.445)
ME	6.068	4.961	2.522	1.877
	(0.014)	(0.026)	(0.112)	(0.171)

Notes:
In this table seasonality is analysed with the Kruskal-Wallis statistic. On the residuals of models T18 and T19, in Table 3 for the EFVI portfolio and in Table 4 for EFMVI, it tests the null hypothesis that the empirical distribution of residuals on seasonality days is the same for the residuals on the remaining days. Thus, YB represents the year beginning days, YE the year end days, HB the half year beginning days, HE for half year end, MB for month beginning and ME for month end (whenever they are neither the end nor the beginning respectively, of neither the year nor the half). The table shows the K-W statistic, and the *p*-value of each test is given in parenthesis.

funds may improve performance, in other words, to interpret it as a hedging strategy, as defined by Lee et al. (1998).

When market timing ability is measured, the number of mutual funds with negative values is higher than those with positive values, which could be taken as evidence of perverse timing. The number of mutual funds with negative and significant market timing decreases when the style benchmarks are included. These results are consistent with the evidence for the aggregate results.

For models T20-T23, Panels B, C and D in Table 6 show the results of the seasonality analysis. The parameters which measure the seasonality corresponding to year beginning, year end, second half year beginning and month end are positive in most of the funds and are significant practically only in this case. Comparing results obtained in T20 with T21 and those from T22 with T23, when style benchmarks are included, the evidence of positive seasonality decreases, particularly at the year beginning. This result has already been reported in the aggregate results in Tables 3 and 5. This implies that part of the evidence of seasonality is attributable to the seasonality between benchmarks and not to managers. The abnormal return corresponding to the beginning of July shows the highest number of significant cases in the seasonality analysis, and this evidence is less sensitive to benchmark omission.

Table 7 summarises the individual results obtained for MVI mutual funds. Panel A shows how the performance is preponderantly negative. In models T21-T23, when the style benchmarks are incorporated, the number of negative alphas is higher. We do not find evidence of market timing ability, although a slight number of significant

Table 6
Individual Results of Variable Income (VI) Mutual Funds

	Number of Mutual Funds							
Panel A	*Performance* α_p				*Market Timing* γ_{pm}			
			p-value < 0.05				*p-value* < 0.05	
Model	<0	>0	<0	>0	<0	>0	<0	>0
T16	49	30	2	5				
T17	50	29	3	6				
T18	30	49	0	20	60	19	18	0
T19	39	40	2	14	45	34	9	0
T20	64	15	10	0				
T21	64	15	15	2				
T22	43	36	5	12	59	20	18	0
T23	46	33	6	7	47	32	9	0
Panel B	yb_p				ye_p			
			p-value < 0.05				*p-value* < 0.05	
Model	<0	>0	<0	>0	<0	>0	<0	>0
T20	13	66	1	29	3	76	1	22
T21	23	56	1	21	7	72	1	18
T22	13	66	1	28	3	76	1	22
T23	25	54	1	22	7	72	2	17
Panel C	hb_p				he_p			
			p-value<0.05				*p-value*<0.05	
Model	<0	>0	<0	>0	<0	>0	<0	>0
T20	5	74	0	51	27	52	1	5
T21	7	72	0	51	28	51	0	4
T22	5	74	0	51	27	52	2	5
T23	7	72	0	51	28	51	1	5
Panel D	mb_p				me_p			
			p-value < 0.05				*p-value* < 0.05	
Model	<0	>0	<0	>0	<0	>0	<0	>0
T20	28	51	3	11	6	73	0	20
T21	32	47	3	10	13	66	1	18
T22	31	48	2	11	7	72	0	20
T23	33	46	3	10	11	68	1	15

Notes:
Performance measured by α_p, market timing evaluated by γ_{pm} and seasonality by yb_p up to me_p. Estimation of the parameters for models T16-T23 applied in Table 3. The heteroskedasticity and autocorrelation consistent covariance estimator is by Newey and West (1987).

parameters are negative, involving perverse market timing. With respect to seasonality, the number of mutual funds with anomalous returns at year end and beginning of July are relevant. These results are consistent with the evidence for the aggregate results in Table 4.

Table 7
Individual Results of Mixed Variable Income (MVI) Mutual Funds

	Number of Mutual Funds							
Panel A	*Performance* α_p				*Market Timing* γ_{pm}			
			p-value < 0.05				*p-value < 0.05*	
Model	*<0*	*>0*	*<0*	*>0*	*<0*	*>0*	*<0*	*>0*
T16	121	20	31	1				
T17	120	21	33	1				
T18	93	48	39	11	76	65	18	0
T19	104	37	49	3	58	83	11	1
T20	131	10	53	1				
T21	129	12	57	1				
T22	107	34	48	5	72	69	20	2
T23	113	28	59	3	55	86	10	1
Panel B	yb_p				ye_p			
			p-value < 0.05				*p-value < 0.05*	
Model	*<0*	*>0*	*<0*	*>0*	*<0*	*>0*	*<0*	*>0*
T20	75	66	8	14	8	133	0	77
T21	87	54	11	12	12	129	1	65
T22	71	70	7	16	8	133	0	79
T23	85	56	10	13	13	128	1	69
Panel C	hb_p				he_p			
			p-value<0.05				*p-value<0.05*	
Model	*<0*	*>0*	*<0*	*>0*	*<0*	*>0*	*<0*	*>0*
T20	28	113	1	45	40	101	0	22
T21	30	111	1	41	43	98	0	18
T22	27	114	1	43	40	101	0	22
T23	30	111	1	41	43	98	0	18
Panel D	mb_p				me_p			
			p-value < 0.05				*p-value < 0.05*	
Model	*<0*	*>0*	*<0*	*>0*	*<0*	*>0*	*<0*	*>0*
T20	104	37	2	9	24	117	1	19
T21	104	37	2	11	27	114	1	17
T22	108	33	2	9	27	114	1	20
T23	108	33	2	9	28	113	1	17

Notes:
Performance measured by α_p, market timing evaluated by γ_{pm} and seasonality by yb_p up to me_p. Estimation of the parameters for models T16-T23 applied in Table 4. The heteroskedasticity and autocorrelation consistent covariance estimator is by Newey and West (1987).

(d) Non-survivor Mutual Funds Evaluation

To assess non-survivor bias, the four equi-weighted portfolios with missing and new mutual funds are analysed. EFVI-9801 is made up of 16 VI mutual funds missing in the second half of the sample period, and EFVI-0104 of 14 VI new funds created in the first half of the sample period. In the same way for MVI mutual funds: EFMVI-9801 and EFMVI-0104 are made up of 21 (41) missing (new) mutual funds. We should interpret the results of assessment of these portfolios with caution because the number of funds is small and the length of the time series is only three complete years.[7] However, these results could inform us about the bias of omitting these mutual funds in the analysis in Tables 3 to 7, only with survivor mutual funds.

When the results for VI mutual funds are compared, the alphas are not significant, but new funds show better performance than missing funds. The perverse timing evidence for the missing funds is greater than the evidence for the survivor funds. Market timing is better for new funds, but not significant. For both groups, and as in Table 3, when style benchmarks are considered, the evidence of perverse timing is reduced. Moreover, the most significant seasonality occurs at the half year beginning. There are differences for the seasonality at the end and beginning of year, but these are not significant. For the missing and new MVI mutual funds, the alphas are negative as in Table 4, but with different significance. Market timing is also not significant and the evidence of seasonality coincides partially with the evidence found for the survival funds.

Therefore, the results for non-survivor mutual funds are generally consistent with those for survivor mutual funds, especially for the performance, timing and seasonality at the July beginning. However, for the seasonality at beginning and end of the year the results for non-survival funds show some differences with respect to survival funds. This latter result could be due to the short length of the time series for non-survival mutual funds.

(e) Seasonality and Performance

It is of interest to analyse the economic relevance of seasonality. To do this, we analyse the relationship between the evidence of seasonality in model T23 and the evidence of performance in model T19, before measuring the seasonality but after controlling for the style benchmarks. By means of regressions, for each group of mutual funds, we analyse the cross-sectional relation between the seasonality in model T23 and the alpha of model T19.

The left panel in Table 8 summarises the results for the group of VI mutual funds. The analysis shows a direct and significant relation between yb_p in T23 and performance from model T19. The anomalous return of the year beginning may explain the 27% of cross-sectional performance in these mutual funds. As Lee et al. (1998) point out, these results could be due to a hedging strategy. This strategy could be implemented either through specific management in the market or through a specific allocation in stocks that show good performance at the beginning of January. Moreover, the Spanish financial newspapers also publish rankings from the year beginning, and a good start therefore brings a high position in the ranking from this moment on. We also found evidence of a significant and positive relation between performance and the seasonality at July beginning. This latter result is also evidenced for the group of MVI mutual funds in the right panel of Table 8. In contrast, the seasonality at other moments does not imply this positive relation, and in some cases it is even significantly

7 To reduce the extent of the paper, the tables with the results for non-survivor mutual funds are not reported.

Table 8
Analysis of the Economic Relevance of Seasonality

	VI Mutual Funds			*MVI Mutual Funds*		
	Constant	*Slope*	R^2	*Constant*	*Slope*	R^2
YB	−0.0020	2.735	26.519	−0.0087	0.689	1.839
	(0.290)	(0.000)		(0.000)	(0.263)	
YE	0.0052	−1.742	3.727	−0.0050	−2.352	4.276
	(0.043)	(0.266)		(0.019)	(0.036)	
HB	−0.0052	2.883	11.450	−0.0128	4.668	12.608
	(0.142)	(0.002)		(0.000)	(0.000)	
HE	0.0021	1.135	0.650	−0.0053	−5.276	14.204
	(0.248)	(0.693)		(0.001)	(0.000)	
MB	0.0020	2.540	0.909	−0.0077	−6.660	2.592
	(0.295)	(0.472)		(0.000)	(0.037)	
ME	0.0001	5.102	2.005	−0.0085	−1.574	0.187
	(0.960)	(0.360)		(0.000)	(0.736)	

Notes:
By means of regressions, for each group of mutual funds we analyse the cross-sectional relation between the evidence of seasonality in model T23 and the alpha or performance, before measuring the seasonality but after controlling for the style benchmarks, from model T19. YB represents the year beginning days, YE the year end days, HB the half year beginning days, HE for half year end, MB for month beginning and ME for month end (whenever they are neither the end nor the beginning respectively, of neither the year nor the half) The *p*-value of each parameter is given in parenthesis. The heteroskedasticity consistent covariance estimator is by White (1980).

negative, particularly for MVI funds. This latter evidence could be explained by the Carhart et al. (2002) hypothesis if mutual funds carry out a window dressing strategy without improving performance, or even quite the reverse.

5. CONCLUSIONS

The use of linear models to evaluate portfolio management has been extensively applied in the performance literature when the information available takes the form of returns data. In this context the current paper has shown the bias produced when we omit benchmarks that represent asset classes in which the assessed mutual fund actually invests. The assessment of three elements of active management, performance, market timing and seasonality, may be biased when relevant benchmarks are omitted and they reveal relations of performance, 'market timing' or asymmetry in returns, and seasonality.

Firstly, we analyse the relation between a set of benchmarks in Spanish markets and find general evidence of an alpha or performance not different from zero. Especially for small-cap stocks, the evidence implies negative or perverse market timing, positive seasonality at the year end and year beginning, at month end and at the July beginning.

We analyse a sample of Spanish mutual funds using a model with only one benchmark for the broad stock market. In aggregate terms, the results obtained coincide with evidence from previous studies on mutual funds. Performance is not significant in many cases, but the presence of negative values is greater. In general market timing ability does not exist, but there is greater evidence of perverse or negative timing. The presence of seasonality is evident in a large number of mutual funds, with positive abnormal returns, and is particularly significant at the turn of the year, the beginning of July and at the month end.

However, these results will vary depending on the models used, essentially according to the benchmarks considered. When style benchmarks are introduced, the market timing is less negative and significant. This result will be a consequence of the asymmetrical relation between the small stocks benchmark and the broad market benchmark. Thus, some of the evidence of market timing is not attributable to the portfolio managers, but rather to the exogenous behaviour of the small stocks in the portfolio. Likewise, when style benchmarks are considered, the evidence of seasonality is reduced. However, while the seasonality at year beginning is more sensitive to style benchmarks, the evidence of seasonality at year end, at July beginning and at month end is less sensitive and holds through all the models. In both individual and aggregate results, the abnormal return at July beginning is the most significant. On the other hand, given that in general seasonality produces positive returns, when this analysis is not omitted, the mutual funds performance is poorer.

We also analyse the economic relevance of seasonality. The seasonality at year beginning shows a relevant and positive relation with the cross-sectional performance of the most risky mutual funds. For the whole sample of mutual funds, we also found evidence of a significant and positive relation between performance and the seasonality at July beginning. These results could be due to a hedging strategy. In contrast, the seasonality at other moments does not imply this positive relation, and in some cases is even significantly negative, which reveals a possible window dressing strategy.

The existence of relations of performance, asymmetry in returns or 'market timing' and seasonality between benchmarks is therefore of interest in portfolio management. The omission of these benchmarks in mutual fund evaluation makes it difficult to delimit managers' added value. With regard to the benefits of diversification and the use of asset allocation as a management tool, research is needed to analyse the persistence over time of the relations between benchmarks.

REFERENCES

Ackermann, C., R. Mcenally and D. Ravenscraft (1999), 'The Performance of Hedge Funds: Risk, Return and Incentives', *Journal of Finance*, Vol. 54, No. 3, pp. 833–73.

Basarrate, B. and G. Rubio (1999), 'Nonsimultaneous Prices and the Evaluation of Managed Portfolios in Spain', *Applied Financial Economics*, Vol. 9, No. 3, pp. 273–81.

Bollen, N. and J. Busse (2001), 'On the Timing Ability of Mutual Fund Managers', *Journal of Finance*, Vol. 56, No. 3, pp. 1075–94.

Brown, K., W. Harlow and L. Starks (1996), 'Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry', *Journal of Finance*, Vol. 51, No. 1, pp. 84–109.

Carhart, M. (1997), 'On Persistence in Mutual Fund Performance', *Journal of Finance*, Vol. 52, No. 1, pp. 57–82.

———, R. Kaniel, D. Musto and A. Reed (2002), 'Leaning for the Tape: Evidence of Gaming Behavior in Equity Mutual Funds', *Journal of Finance*, Vol. 57, No. 2, pp. 661–93.

Chance, M. and M. Hemler (2001), 'The Performance of Professional Market Timers: Daily Evidence from Executed Strategies', *Journal of Financial Economics*, Vol. 62, No. 2, pp. 377–411.

Coggin, D., F. Fabozzi and S. Rahman (1993), 'The Investment Performance of U.S. Equity Pension Fund Managers: An Empirical Investigation', *Journal of Finance*, Vol. 48, No. 3, pp. 1039–55.

Edelen, R. (1999), 'Investor Flows and the Assessed Performance of Open-end Mutual Funds', *Journal of Financial Economics*, Vol. 53, No. 3, pp. 439–66.

Elton, E., M. Gruber and C. Blake (1996), 'Survivorship Bias and Mutual Fund Performance', *Review of Financial Studies*, Vol. 9, No. 4, pp. 1097–120.

——— ——— and M. Hlavka (1993), 'Efficiency with Costly Information: A Reinterpretation of Evidence for Managed Portfolios', *Review of Financial Studies*, Vol. 6, No. 1, pp. 1–22.

Ferson, W. and R. Schadt (1996), 'Measuring Fund Strategy and Performance in Changing Economic Conditions', *Journal of Finance*, Vol. 51, No. 2, pp. 425–61.

Forner, C. and J. Marhuenda (2003), 'Contrarian and Momentum Strategies in the Spanish Stock Market', *European Financial Management*, Vol. 9, No. 1, pp. 67–88.

Glosten, L. and R. Jagannathan (1994), 'A Contingent Claim Approach to Performance Evaluation', *Journal of Empirical Finance*, Vol. 1, No. 2, pp. 133–60.

Goetzmann, W., J. Ingersoll and Z. Ivkovic (2000), 'Monthly Measurement of Daily Timers', *Journal of Financial and Quantitative Analysis*, Vol. 35, No. 3, pp. 257–90.

Graham, J. and C. Harvey (2001), 'Graham-Harvey Performance Metrics' (http://www.duke.edu/˜charvey/performance˙eval/lettab1.htm).

Grinblatt, M. and S. Titman (1989a), 'Portfolio Performance Evaluation: Old Issues and New Insights', *The Review of Financial Studies*, Vol. 2, No. 3, pp. 93–421.

——— ——— (1989b), 'Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings', *Journal of Business*, Vol. 62, No. 3, pp. 393–416.

Haugen, R. and J. Lakonishok (1988), *The Incredible January Effect* (Dow Jones-Irwin, Homewood, Illinois).

Henriksson, R. (1984), 'Market Timing and Mutual Fund Performance: An Empirical Investigation', *Journal of Business*, Vol. 57, No. 1, pp. 73–96.

——— and R. Merton (1981), 'On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills', *Journal of Business*, Vol. 54, No. 4, pp. 513–33.

Holmes, K. and R. Faff (2004), 'Stability, Asymmetry and Seasonality of Fund Performance: An Analysis of Australian Multisector Managed Funds', *Journal of Business Finance & Accounting*, Vol. 33, Nos. 3 & 4, pp. 539–78.

Jagannathan, R. and R. Korajczyk (1986), 'Assessing the Market Timing Performance of Managed Portfolios', *Journal of Business*, Vol. 59, No. 2, pp. 217–35.

Jensen, M. (1968), 'The Performance of Mutual Funds in the Period 1945–1964', *Journal of Finance*, Vol. 23, No. 2, pp. 389–415.

Jiang, W. (2003), 'A Nonparametric Test of Market Timing', *Journal of Empirical Finance*, Vol. 10, No. 4, pp. 399–425.

Keim, D. (1983), 'Size-related Anomalies and Stock Market Seasonality: Further Empirical Evidence', *Journal of Financial Economics*, Vol. 12, No. 1, pp. 13–32.

——— (1989), 'Trading Patterns, Bid-ask Spreads, and Estimated Security Returns: The Case of Common Stocks at Calendar Turning Points', *Journal of Financial Economics*, Vol. 25, No. 1, pp. 75–97.

——— (1999), 'An Analysis of Mutual Fund Design: The Case of Investing in Small-cap Stocks', *Journal of Financial Economics*, Vol. 51, No. 2, pp. 173–94.

Koski, J. and J. Pontiff (1999), 'How are Derivatives Used? Evidence from the Mutual Fund Industry', *Journal of Finance*, Vol. 54, No. 2, pp. 791–816.

Lee, C., D. Porter and D. Weaver (1998), 'Indirect Tests of the Haugen-Lakonishok Small-firm/January Effect Hypotheses: Window Dressing versus Performance Hedging', *Financial Review*, Vol. 33, No. 2, pp. 177–94.

Martínez, M. (2003), 'Legal Constraints, Transaction Costs, and the Evaluation of Mutual Funds', *European Journal of Finance*, Vol. 9, No. 3, pp. 199–218.

Matallín, J. (2002), 'Analysis of the Relation Between Size and Management of Mutual Funds', X Foro de Finanzas de la Asociación Española de Finanzas (Universidad Pablo de Olavide, Sevilla).

——— and M. Fernández (2000), 'Style Analysis and Performance Evaluation of Spanish Mutual Funds', *Journal of Asset Management*, Vol. 1, No. 2, pp. 151–71.

Menéndez, S. and S. Álvarez (2000), 'La rentabilidad y persistencia de los resultados de los fondos de inversión españoles de renta variable', *Revista Española de Financiación y Contabilidad*, Vol. 29, No. 103, pp. 15–36.

Newey, W. and D. West (1987), 'A Simple, Positive Semi-definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix', *Econometrica*, Vol. 55, No. 3, pp. 703–8.
Pástor, L. and R. Stambaugh (2002), 'Mutual Fund Performance and Seemingly Unrelated Assets', *Journal of Financial Economics*, Vol. 63, No. 3, pp. 315–49.
Rubio, G. (1995), 'Further Evidence on Performance Evaluation: Portfolio Holdings, Recommendations, and Turnover Costs', *Review of Quantitative Finance and Accounting*, Vol. 5, No. 2, pp. 127–53.
Sharpe, W. (1966), 'Mutual Fund Performance', *Journal of Business*, Vol. 39, No. 1, pp. 119–38.
——— (1991), 'The Arithmetic of Active Management', *Financial Analysts Journal*, Vol. 47, No. 1, pp. 7–9.
——— (1992), 'Asset Allocation: Management Style and Performance Measurement', *Journal of Portfolio Management*, Vol. 18, No. 1 (Winter), pp. 7–19.
Treynor, J. and M. Mazuy (1966), 'Can Mutual Funds Outguess the Market?', *Harvard Business Review*, Vol. 44, No. 4, pp. 131–36.
Volkman, D. (1999), 'Market Volatility and Perverse Timing Performance of Mutual Fund Managers', *Journal of Financial Research*, Vol. 22, No. 4, pp. 449–70.
Wenchi-Kao, G., L. Cheng and K. Chan (1998), 'International Mutual Fund Selectivity and Market Timing During Up and Down Market Conditions', *Financial Review*, Vol. 33, No. 2, pp. 127–44.
White, H. (1980), 'A Heteroscedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity', *Econometrica*, Vol. 48, No. 4, pp. 817–38.