Selection, timing and total performance of equity mutual funds: On the relevance of model specification

Sebastian Krimm[+]

Hendrik Scholz[++]

Marco Wilkens[+++]

Working Paper

Catholic University of Eichstaett-Ingolstadt

This Version: November 30, 2008

[+] Sebastian Krimm, Ingolstadt School of Management, Catholic University of Eichstaett-Ingolstadt, Auf der Schanz 49, 85049 Ingolstadt, Germany, phone: +49 841 937 1865, fax +49 841 937 2865, e-mail: sebastian.krimm@ku-eichstaett.de

[++] Hendrik Scholz, Ingolstadt School of Management, Catholic University of Eichstaett-Ingolstadt, Auf der Schanz 49, 85049 Ingolstadt, Germany, phone: +49 841 937 1878, fax +49 841 937 2878, e-mail: hendrik.scholz@ku-eichstaett.de

[+++] Marco Wilkens, Ingolstadt School of Management, Catholic University of Eichstaett-Ingolstadt, Auf der Schanz 49, 85049 Ingolstadt, Germany, phone: +49 841 937 1883, fax +49 841 937 2883, e-mail: marco.wilkens@ku-eichstaett.de

Selection, timing and total performance of equity mutual funds: On the relevance of model specification

Abstract

This is the first paper to analytically and empirically compare the results of the Jensen, the Treynor-Mazuy and the Henriksson-Merton models. We analyze to which the extent the choice of performance models influence the selection, timing and total performance of funds. Our main findings are: i) Under realistic market conditions, the analytical results show that the Jensen alpha of a fund is very similar to its total performance based on the Treynor-Mazuy and Henriksson-Merton. ii) The empirical results confirm very small differences between these measures for a broad sample of US equity mutual funds. iii) In contrast, the contributions of selection and timing activities to the funds' total performance based on the Treynor-Mazuy and the Henriksson-Merton approaches clearly depend on the model choice. However, the influence of model choice on fund rankings based on selection or timing performance is comparatively small.

JEL classification: G11, G12

Keywords: Performance evaluation; total performance; selection performance; timing performance; model specification

1. Introduction

A multitude of approaches to evaluate the performance of actively managed portfolios are discussed today. In this context, many studies analyze the impact of the choice of market factors in one- and multifactor settings on performance evaluation: Roll (1978) theoretically discusses the relevance of the efficiency of the market portfolio for performance measurement.[1] Lehmann and Modest (1987), Grinblatt and Titman (1994) and Otten and Bams (2004) analyze the empirical effects of the choice of market factors, and they all find it to be relevant.[2] However, the question of how the modeling of timing activities of portfolio managers affects the measured abnormal returns has reached little attention so far.[3] Therefore this paper is the first to theoretically and empirically analyze the relations of three commonly used models to measure fund performance – the ones proposed by Jensen (1968), Treynor and Mazuy (1966) and Henriksson and Merton (1981). These models differ in respect to the consideration of timing activities: In the case of Jensen alpha, the exposure to market risk is assumed to be constant, while the model of Treynor-Mazuy (Henriksson-Merton) allows for continuous (discrete) variations of market risk depending on the market risk premium realized within the next period. Thus, these two models explicitly account for respective timing activities of portfolio managers.

The impact of market timing activities on the measured performance of funds has been long discussed: Jensen (1972) shows that the Jensen alpha (henceforth JA) is biased downwards when applied to portfolios which successfully time the market. Grant (1977) shows that the bias of the JA due to timing activities depends on the mean and the variance of the market

1 Grinblatt and Titman (1989) present a solution by dissociating performance measurement from capital market models like the CAPM.
2 Other studies that apply different market models or factors are Elton et al. (1996), Carhart (1997), Fletcher and Forbes (2002).
3 Coles et al. (2006) deal with this question on the basis of a bootstrap analysis which we will discuss later.

return and can be negative as well. Grinblatt and Titman (1989) decompose the JA and attribute its market timing bias to an under- or overestimation of the expected exposure to market risk. They also show that the direction of this bias depends on the mean and the variance of the market return. We extend the literature by analytically deriving the differences in total performance – defined as the abnormal return due to selection and due to timing activities – between the models. We further determine the economic significance of these differences, first by assuming identically and independently (i.i.d), normally distributed market returns, and by considering more realistic distributions of market returns based on a bootstrap approach.

Jagannathan and Korajczyk (1986) theoretically discuss the problem of choosing between the models of Treynor and Mazuy (henceforth TM) and Henriksson and Merton (henceforth HM) when measuring timing performance. A solution is given by Glosten and Jagannathan (1994), who develop a contingent claim approach for performance measurement and show that the JA, the TM and the HM models can be interpreted as different methods of approximating the form of this contingent claim. Pesaran and Timmermann (1994) develop a generalization of the non-parametric test of Henriksson and Merton (1981), assuming that the portfolio manager categorizes his forecast of the market return into a number of classes and adapts the market risk of his portfolio accordingly.[4] We do not employ such a generalized approach as we are specifically interested in the impact of the consideration of timing activities on the measured performance.

The abnormal returns as measured by the models discussed here have been compared rarely: To the best of our knowledge, Coles et al. (2006) is the only study dealing with the relevance of model specification in which the JA, the TM and the HM model are applied. Conducting a

[4] As the number of classes increases, this timing strategy increasingly resembles that of the TM model.

bootstrap analysis for 327 funds, Coles et al. find that the choice of model does not affect inferences about total abnormal performance of funds, but that the selection and the timing performance are strongly influenced by it. Due to the nature of their bootstrap approach, they do not show analytically results. Besides a theoretical analysis, we present empirical results for a large sample of US equity mutual funds which strongly confirm our analytical results: In almost all cases, the three models yield nearly identical inferences about the total performance of funds. We also find that the selection and timing performance in absolute terms systematically differ between the Treynor-Mazuy and the Henriksson-Merton model. However, ranking funds according to selection or timing performance produces nearly identical rank orders for both models.

The remainder of this paper is organized as follows: Section 2 discusses the models and the interpretation of their resuls. Section 3 shows the analytical relations between selection, timing and total performance as measured by the Jensen, the Treynor-Mazuy and the Henriksson-Merton models. Subsequently, a bootstrap analysis shows the magnitudes of the differences between these measures. In Section 4 we interpret our empirical analysis of US mutual equity funds. Section 5 concludes.

2. Performance analysis with the Jensen, the Treynor-Mazuy und the Henriksson-Merton model

2.1. Detection of selection and timing activities

Jensen (1968) measures the performance of a fund *i* as the constant of a regression of its returns in excess of the risk-free rate $er_{i,t}$ against the excess returns of the market index $er_{m,t}$:

$$er_{i,t} = \alpha_i + \beta_i er_{m,t} + \epsilon_{i,t} \tag{1}$$

The coefficient α_i measures the stock selection activities of fund i. The possibilty of timing activities is ruled out by the assumed constant exposure to market risk, expressed by the static beta coefficient β_i.

The model proposed by Treynor and Mazuy (1966) explicitly allows for dynamic changes in the fund's exposure to market risk and models these changes in linear dependence on the market return:

$$\begin{aligned} er_{i,t} &= \alpha_i^{TM} + \left(\beta_i^{TM} + \gamma_i^{TM} er_{m,t}\right) er_{m,t} + \epsilon_{i,t}^{TM} \\ &= \alpha_i^{TM} + \beta_i^{TM} er_{m,t} + \gamma_i^{TM} er_{m,t}^2 + \epsilon_{i,t}^{TM} \end{aligned} \qquad (2)$$

The coefficient α_i^{TM} measures selection performance, the coefficient β_i^{TM} captures the fund's average exposure to market risk and γ_i^{TM} measures its adjustment in dependence on the market excess return. Positive values of γ_i^{TM} indicate succesful timing activities.

Henriksson and Merton (1981) capture the time-varying exposure to market risk by interpreting successful market timing activities as "protective" put options[5] on the market return. They model the dynamics of beta in a binary dependence of the market return:

$$er_{i,t} = \alpha_i^{HM} + \beta_i^{HM} er_{m,t} + \gamma_i^{HM} \max(0, -er_{m,t}) + \epsilon_{i,t}^{HM} \qquad (3)$$

Again, the coefficient α_i^{HM} captures selection activities of the fund manager, the coefficient β_i^{HM} the average exposure to market risk in times of positive market excess returns and the coefficient γ_i^{HM} the timing activities, figuratively spoken the deleveraging by executing the protective puts on the market in times of negative market excess return. Successful timing activities are therefore indicated by positive values of γ_i^{HM}.

---

5 As in Henriksson and Merton (1981) we here consider these options as puts. The model can also be formulated if one interprets timing activies as calls on the market index, as done e.g. by Henriksson (1984). This has only formal consequences for the further results.

While the theoretical understanding of these measures is unambiguous, it is a different question of how the empirically measured fund performance can be interpreted: If market timing activities are present, the economic interpretation of the JA α_i in Equation (1) clearly differs empirically from the interpretation of the coefficients α_i^{TM} and α_i^{HM} in Equation (2) and (3), as the JA reflects selection and timing activities of the fund.[6] We therefore consider the JA as a measure of total performance in terms of the ex-post total average abnormal return. The interpretation of the regression constants of the TM and the HM models as measures of selection activities also only holds theoretically as this requires the respective model to be the return-generating process of the evaluated fund. Since this may not be the case empirically, the constants in the TM and the HM model do not have to exclusively reflect selection performance. In accordance with the literature,[7] we nevertheless regard α_i^{TM} and α_i^{HM} as measures of the ex-post average abnormal return due to selection activities (henceforth “selection performance”).[8]

Besides issues related to the specification of the performance measurement model, other problems concerning the interpretation of the results of these models have been found: Jagannathan and Korajczyk (1986) show that portfolios which include options or other derivatives may show positive timing performance, though their trades are not based on superior timing information. Comer (2006) reports artificial timing for hybrid funds when performance is measured without including bond market factors. Mattallín-Sáez (2003) analyzes theoretically how the omission of benchmark variables can lead to the measurement of artificial timing and finds empirical support for his results. Goetzmann et al. (2007) demonstrate how fund managers can actively manipulate performance measures and create

---

6 See Grant (1977) and Grinblatt and Titman (1989, 1995).

7 See e.g. Grinblatt and Titman (1994), Bollen and Busse (2005) and Goetzmann et al. (2007).

8 The same arguments apply to the coefficients γ_i^{TM} and γ_i^{HM} in Equations (2) and (3) in respect of the detection of timing activities.

artificial results by employing dynamic trading strategies. In respect to these findings we point out that our empirical study does not aim to judge funds in terms of the economic value they add for investors: We specifically focus on the impact of the choice of performance model on the results of the performance analysis.

2.2. Abnormal return due to selection and timing activities

Grinblatt and Titman (1989, 1995) show that the expected excess return of a portfolio can be decomposed into three components:

$$\mathrm{E}(er_{i,t}) \quad = \quad \mathrm{E}(\beta_{i,t})\mathrm{E}(er_{m,t}) + \mathrm{Cov}(\beta_{i,t}, er_{m,t}) + \mathrm{E}(\epsilon_{i,t}) \tag{4}$$

E(·) stands for the expected value, Cov(·) for the covariance, and $\beta_{i,t}$ for a dynamic exposure to market risk[9]. The first term on the right-hand side of Equation (4) is the risk premium paid for the expected exposure to market risk. Accordingly, the remaining terms reflect the abnormal return: The expected value of the fund residuals $\mathrm{E}(\varepsilon_{i,t})$ is the abnormal return due to selection activities (the selection performance). It analytically corresponds to the regression constant α_i^{TM} (α_i^{HM}) in the TM (HM) model. The covariance between the dynamic exposure to market risk and the market excess return $\mathrm{Cov}(\beta_{i,t}, er_{m,t})$ is the abnormal return due to market timing activities (henceforth "timing performance"). Without any assumption concerning the return-generating process of a fund, one can derive these covariances for the TM and the HM model which yields the respective timing performance tim_i in terms of abnormal return:[10]

---

9 It is important to note that $\beta_{i,t}$ fundamentally differs from the coefficient β_i in (1): Since the JA rules out timing activities, β_i is necessarily static, while $\beta_{i,t}$ allows for market timing. In the TM model, the dynamic exposure to market risk follow a process which is assumed to be $\beta_{i,t}^{TM} = \beta_i^{TM} + \gamma_i^{TM}\, er_{m,t}$. In the HM model, the equivalent process is $\beta_{i,t}^{HM} = \beta_i^{HM} + \gamma_i^{HM}\, \mathrm{I}(-er_{m,t} > 0)$ where I(·) denotes an indicator function.

10 See Grinblatt and Titman (1994) for the TM model and Appendix A for the HM model.

$$tim_i^{TM} = \gamma_i^{TM}\mathrm{Var}(er_m) \quad (5)$$
$$tim_i^{HM} = \gamma_i^{HM}\left[\mathrm{P}(er_m < 0)\mathrm{E}(er_m) - \mathrm{E}\left[\min(0, er_m)\right]\right] \quad (6)$$

E(·) and Var(·) denote the expected value and the variance of the indicated variables and $P(er_m < 0)$ the probability of negative market excess returns. Equations (5) and (6) also apply in a multifactor world, e.g., when using multifactor models as suggested by Fama and French (1993) and Carhart (1997).[11]

Since both the selection and the timing performance of a fund are defined as return quantities, we are now able to directly compare the results of the TM and the HM model. Furthermore, we can compare all three models in terms of their ex-post total performance. In case of the JA this is the coefficient α_i in Equation (1). From (4) it follows that the total performance measures of the TM model tot_i^{TM} and of the HM model tot_i^{HM} are:

$$tot_i^{TM} = \alpha_i^{TM} + \gamma_i^{TM}\mathrm{Var}(er_m) \quad (7)$$
$$tot_i^{HM} = \alpha_i^{HM} + \gamma_i^{HM}\left[\mathrm{P}(er_m < 0)\mathrm{E}(er_m) - \mathrm{E}\left[\min(0, er_m)\right]\right] \quad (8)$$

Bollen and Busse (2005) use a similar approach when measuring ex-post total performance performance with the TM model:[12] They substitute the variance of the market excess returns in (7) with the mean squared market excess returns which results in[13] $tot_i^{TM} = \alpha_i^{TM} + \gamma_i^{TM}\,\overline{er}_{m,t}^2$. Since they use daily returns which implies $E(er_m)^2 \approx 0$, their approach is virtually identical to (7) as the computational formula for the variance ($\mathrm{Var}(er_m) = E(er_m^2) - E(er_m)^2$) shows.

---

[11] If one allows for simultaneous timing activities in several market factors, computing timing performance according to (5) and (6) implicitly ignores possible cross terms of timed market factors. This is not uncommon in the literature: Grinblatt and Titman (1994) and Comer (2006) do not consider cross terms to obtain a feasible model. Kryzanowski et al. (1997) assume orthogonal timing signals in their multifactor timing model which rule outs cross terms of the timed market factors.

[12] Their approach is also used by Comer et al. (2007) and Huui and Derwall (2008).

[13] In contrast to Comer et al. (2007), Huui and Derwall (2008) and the approach used here, Bollen and Busse (2005) interpret α_i^{TM} as the cost of timing activties and denote the right-hand side of the following equation as timing performance.

In contrast to our approach to evaluate timing activities, other studies appraise timing activities from an ex-ante view using option pricing models, as done by Merton (1981) for the HM model and by Goetzmann et al. (2007) for the TM model. We apply an ex-post view because it directly measures the return contribution that timing activities have added to the total performance of funds. Moreover, this allows us to identify the timing return component unambiguously based on ex-post returns. in contrast to the ex-ante view in which the results may be influenced by a potential misspecification of the option pricing model applied.

3. Impact of model choice on measured performance

In this section we first derive analytical expressions for the differences in selection, timing and total performance between the JA, the TM and the HM models. These derivations directly relate the three models with each other and do not rely on any assumptions concerning the investment activities of a fund manager. In order to describe the difference in total performance between the JA and the TM model, we at first explain the additional factor within the TM model $er_{m,t}^2$ with a one-factor regression based on Equation (1):[14]

$$er_{m,t}^2 = \alpha_{er_m^2} + \beta_{er_m^2} er_{m,t} + \epsilon_{er_m^2,t} \tag{9}$$

Substituting the right hand side of Equation (9) into the TM model in Equation (2) shows its relation to the JA:

$$\begin{aligned} er_{i,t} &= \alpha_i^{TM} + \beta_i^{TM} er_{m,t} + \gamma_i^{TM}\left(\alpha_{er_m^2} + \beta_{er_m^2} er_{m,t} + \epsilon_{er_m^2,t}\right) + \epsilon_{i,t}^{TM} \\ &= \underbrace{\alpha_i^{TM} + \gamma_i^{TM}\alpha_{er_m^2}}_{\alpha_i} + \underbrace{\left(\beta_i^{TM} + \gamma_i^{TM}\beta_{er_m^2}\right)}_{\beta_i} er_{m,t} + \underbrace{\epsilon_{i,t}^{TM} + \gamma_i^{TM}\epsilon_{er_m^2,t}}_{\epsilon_i} \end{aligned} \tag{10}$$

[14] For a comparable approach explaining omitted factors in performance measures based on factor models see Pástor and Stambaugh (2002).

The JA of a fund α_i equals the sum of the selection performance within the TM model α_i^{TM} and the timing coefficient γ_i^{TM} times the "Jensen alpha" of the squared market returns $\alpha_{er_m^2}$. Thus the difference between the total performance of the TM model according to Equation (7) and the JA is:

$$\begin{aligned} tot_i^{TM} - \alpha_i^{JA} &= \alpha_i^{TM} + \gamma_i^{TM}\mathrm{Var}(er_m) - \alpha_i^{TM} - \gamma_i^{TM}\alpha_{er_m^2} \\ &= \gamma_i^{TM}\left[\mathrm{Var}(er_m) - \alpha_{er_m^2}\right] \end{aligned} \qquad (11)$$

This difference is the timing coefficient γ_i^{TM} times the difference between the variance of the market excess return and the "Jensen alpha" of the squared market excess return. The difference in total performance can therefore be divided into a fund-specific component γ_i^{TM} and a market component which is the term within the squared brackets in (11). This allows for a separate judgement of these components when analyzing the potential magnitude of this difference.

Analogue we can relate the JA with the total performance based on the HM model according to (8). First we explain the additional factor within the HM model $\max(0, -er_{m,t})$ with the market excess returns according to Equation (1):

$$\max(0, -er_{m,t}) = \alpha_{\max(0,-er_m)} + \beta_{\max(0,-er_m)} er_{m,t} + \epsilon_{\max(0,-er_m),t} \qquad (12)$$

We relate the HM model to the JA by substituting $\max(0, -er_{m,t})$ in (3) with the right-hand side of (12):

$$\begin{aligned} er_{i,t} &= \alpha_i^{HM} + \beta_i^{HM} er_{m,t} + \gamma_i^{HM}\left[\alpha_{\max(0,-er_m)} + \beta_{\max(0,-er_m)} er_{m,t} \right. \\ &\quad \left. + \epsilon_{\max(0,-er_m),t}\right] + \epsilon_{i,t}^{HM} \\ &= \underbrace{\alpha_i^{HM} + \gamma_i^{HM}\alpha_{\max(0,-er_m)}}_{\alpha_i} + \underbrace{\left[\beta_i^{HM} + \gamma_i^{HM}\beta_{\max(0,-er_m)}\right]}_{\beta_i} er_{m,t} \\ &\quad + \underbrace{\epsilon_{i,t}^{HM} + \gamma_i^{HM}\epsilon_{\max(0,-er_m),t}}_{\epsilon_i} \end{aligned} \qquad (13)$$

The JA α_i equals the sum of the selection performance based on the HM model α_i^{HM} and the timing coefficient γ_i^{TM} times the "Jensen alpha" of the absolute negative market excess returns $\alpha_{\max(0,-er_m)}$. Based on this, the difference between the total performance of the HM model as according to (8) and the JA is:

$$\begin{aligned} tot_i^{HM} - \alpha_i &= \alpha_i^{HM} + \gamma_i^{HM}\left[\mathrm{P}(er_m < 0)\mathrm{E}(er_m) - \mathrm{E}[\min(0, er_m)]\right] \\ &\quad -\alpha_i^{HM} - \gamma_i^{HM}\alpha_{\max(0,-er_m)} \\ &= \gamma_i^{HM}\left[\mathrm{P}(er_m < 0)\mathrm{E}(er_m) - \mathrm{E}[\min(0, er_m)] - \alpha_{\max(0,-er_m)}\right] \end{aligned} \tag{14}$$

The difference in (14) is the product of the fund-specific timing coefficient γ_i^{HM} and a market component which is the term in squared brackets in the second line of (14). The difference between the total performance based on the HM model and the JA hence reveals the same structure as the difference between the total performance based on the TM model and the JA. Finally, it is obvious that in the absence of market timing activities, all three measures will result in the same evaluation of total performance.

Subtracting (14) from (11) yields the difference in total performance between the TM and the HM models:

$$\begin{aligned} tot_i^{TM} - tot_i^{HM} &= \gamma_i^{TM}\left[\mathrm{Var}(er_m) - \alpha_{er_m^2}\right] - \gamma_i^{HM}\left[\mathrm{P}(er_m < 0)\mathrm{E}(er_m)\right. \\ &\quad \left. -\mathrm{E}[\min(0, er_m)] - \alpha_{\max(0,-er_m)}\right] \\ &= \underbrace{\gamma_i^{TM}\mathrm{Var}(er_m) - \gamma_i^{HM}\left[\mathrm{P}(er_m < 0)\mathrm{E}(er_m) - \mathrm{E}[\min(0, er_m)]\right]}_{\text{Difference in timing performance}} \\ &\quad \underbrace{+\gamma_i^{HM}\alpha_{\max(0,-er_m)} - \gamma_i^{TM}\alpha_{er_m^2}}_{\text{Difference in selection performance}} \end{aligned} \tag{15}$$

Again the difference in total performance depends on fund-specific components and market components. Additionally, Equation (15) also reveals the differences in selection performance and in timing performance between the TM and the HM models. It is important to stress that these results hold regardless of any assumptions concerning the distribution of the market returns or the actual timing activities of a given fund.

Previous studies show that the bias of the JA due to timing activities depends on the mean and the variance of the market excess return.[15] Assuming identically and independently, normally distributed market excess returns in a one-factor setting, the difference in total performance between the TM model and the JA as stated in Equation (11) respectively the HM model and the JA as stated in Equation (14) can then be further simplified to:[16]

$$tot_i^{TM} - \alpha_i = \gamma_i^{TM} \mathrm{E}(er_m)^2 \quad (16)$$

$$tot_i^{HM} - \alpha_i = 0 \quad (17)$$

Under this distributional assumption, the difference in total performance between the TM model and the JA increases with the expected absolute market excess return. According to (17), the JA exactly coincides with the total performance based on the HM model, irrespective of the degree of timing activities of the evaluated fund. This is an innovative finding which contradicts previous literature showing the JA to be biased in the presence of timing activities. As a consequence from Equation (17), the difference between the total performance based on the TM and the HM models is the same as the one between the TM model and the JA stated in Equation (16).

In order to test the robustness of Equations (16) and (17) with respect to the distribution of the market excess returns, we use a bootstrap approach to estimate the market components of the difference in total performance in Equations (11) and (14). We randomly draw a sample of 60 monthly excess returns of the value-weighted index of all stocks listed on the NYSE, AMEX

---

[15] See, e.g., Grant (1977) and Grinblatt and Titman (1989).
[16] Appendix B contains a detailed derivation.

or NASDAQ above the one-month Treasury Bill rate from January 1993 to December 2006.[17] Based on this random sample, we calculate the market components of the difference in total performance in Equations (11) and (14). We repeat this process 2,500 times. To better understand the effects of the market climate on the difference in total performance, we plot the simulated market components against the mean, the standard deviation, the skewness and the kurtosis of the bootstrapped market returns.

[Insert Figure 1 about here]

The resulting non-normal distribution of the sampled market excess returns affects the market component of the difference in total performance between the TM model and the JA in Equation (11). Figure 1a reveals that the market component is the smallest for mean market excess returns of zero and that its range grows with the absolute value of the mean market excess returns. It roughly follows the quadratic form indicated by Equation (16), indeed it shows more negative values for positive mean market excess returns. The standard deviation of the bootstrapped market excess returns has no influence on it (Figure 1b), in contrast to the skewness (kurtosis), with which the market component slightly ascends (descends), as Figure 1c (Figure 1d) shows. Irrespective of the market parameters the market component is usually small: 90% of the simulated values are between +2 and -4 bp as indicated by the dashed lines.

[Insert Figure 2 about here]

Figure 2 shows scatter plots of the market component of the difference between the HM total performance and the JA against descriptive statistics of the bootstrapped market excess returns. In Figure 2a, we find that the market component is approximately zero for a mean market excess return of zero. Both are clearly negatively related (Figure 2a) with the range of

[17] We are grateful to Kenneth R. French for providing this data on his webpage.

the market component growing with the absolute value of the mean market excess return. The standard deviation of the market excess return (Figure 2b) shows practically no influence on the market component, while there exists a weakly positive (negative) relation with the skewness (kurtosis) of the bootstrapped market excess returns in Figure 2c (Figure 2d). Within the analysis, 90% of the simulated market components are between ca. +5 and -30 bp.

To sum up, the bootstrap analysis shows that the mean market excess return within the respective evaluation period may have a strong impact one the differences in total performance between the three models. However, both market components tend to disappear when the mean market excess return is close to zero, which will typically be much more likely for longer evaluation periods or when using daily fund returns for performance evaluation. The effects of the higher moments are comparably weak. Furthermore, the market component of the difference in total performance between the TM model and the JA is substantially smaller than the one between the HM model and the JA. Since both market components are scaled by the degree of measured timing activities – the coefficients γ^{TM} and γ^{HM} of the fund – , the question of how large these differences actually are will be answered based on an empirical analysis of funds.

4. Empirical analysis

4.1. Data description

We use monthly fund returns as reported in the CRSP Survivor-Bias-Free US Mutual Fund Database 2006, out of which we select 6,853 funds based on three criteria: First, the fund has to belong to one of the six Standard & Poor's Fund Objective categories that imply it is an equity fund.[18] Since we conduct the empirical analysis for the single fund categories separately, funds that change between these six categories are excluded. The strict usage of the Standard & Poor's Fund Objective categories limits the evaluation period to January 1993 to December 2006. Second, we require at least 36 continuously reported returns for a fund to be included in the sample to yield reliable results when estimating mean, risk and alphas of individual funds. This potentially creates a survivorship bias[19] in our sample which is indeed hardly avoidable when applying regression based measures on single funds. As we do not aim to judge the economic value added by fund managers but to show the relations between the models applied, we find this drawback to be acceptable. Third, the time series of the fund's returns may show no gaps or obviously implausible values. We use the same market index as in the bootstrap analysis in the previous section.

A large number of empirical studies using the TM and the HM approaches to measure fund performance use monthly fund returns.[20] However, Bollen and Busse (2001) find that using daily returns leads to a considerably larger proportion of funds with significant timing performance. Goetzmann et al. (2000) simulate time series of portfolio returns based on different timing intervals and document the relevance of the evaluation period: If the interval of the timing activities of a fund differs from the interval for which portfolio returns are determined, the fund's measured timing performance will be biased downwards and might not be detected. In our empirical analysis, we use monthly returns to show the relations between

---

[18] Following Pástor and Stambaugh (2002) we include the following objective categories by Strategic Insight (formerly Standard & Poor's) in our sample: AGG (Aggressive), GMC (Growth MidCap), GRI (Growth Income), GRO (Growth), ING (Income Growth), and SCG (SmallCap Growth). We do not consider sector funds in order to avoid benchmark specification problems.

[19] Detailed studies of the problem of survivorship bias can be found in Brown and Goetzmann (1995), Elton et al. (1996), and Carhart et al. (2002).

[20] See, e.g., Grinblatt and Titman (1994), Cai et al. (1997), Kryzanowski et al. (1997), Beckers et al. (1999), Comer (2006), Chen and Liang (2007) and Jiang at al. (2007)

the different measures. However, based on our analytical results we can expect empirical findings on the relations between these measures to also hold when daily returns are applied.

4.2. Empirical results

We estimate the regression equations (1), (2) and (3) for each fund[21] and calculate their timing performance according to Equations (5) and (6) and the total performance of each fund according to Equations (7) and (8). Table 1 presents descriptive statistics of the selection, timing and total performance for the total sample as well as for each fund category.

[Insert Table 1 about here]

Irrespective of the performance measure applied, the funds have on average underperformed in terms of total performance. This result accords with the vast majority of empirical studies of fund performance that use returns net of management fees and costs. Therefore, it is surprising that three fund categories show a positive mean[22] total performance which significantly differs from zero at the 1% level.[23] In the case of the categories GMC and SCG, the weak representation of the respective investment universe in the market index used could be an explanation for this result. The mean selection performance within the total sample is weakly negative and also differs significantly across the single categories.[24] Finally, the average fund exhibits little or no timing performance according to the TM and the HM

[21] We have also applied these measures using the four factor model suggested by Carhart (1997). As the results lead to the same conclusions we do not present them here for reasons of brevity.

[22] This applies to the median total performance of these categories as well.

[23] The mean total performance differs significantly for the six fund categories at the 1% level. The null hypothesis (H_0) of identical median total performance over all categories is also rejected at the 1% level. Tests of the mean and the median of each individual categories' total performance against the value of the total sample also reject the H_0 at a 1% level for all categories but GMC, for which the H_0 cannot be rejected in either case.

[24] Again we test the mean selection performance of all six categories simultaneously and of each individual category against the total sample for identity. The null hypothesis is always rejected at the 1% level. This also applies for the median selection performance with the exception of the category GRI, whose median selection performance does not differ significantly from the one of the total sample.

measure. Again we find the differences between the single fund categories to be mostly significant.[25]

Comparing the results between the different measures shows that the descriptive statistics of the total performance are nearly identical. This is an indication that all three measures yield very similar inferences about the total performance of funds. The descriptive statistics of the selection and timing performance obviously differ between the TM and the HM model. The means of the selection performance (timing performance) of the TM model are higher (lower) for all categories, as well as for the total sample, than the ones of the HM model. Moreover, the standard deviations of both performance components are higher for the HM model in all categories.

[Insert Table 2 about here]

The proportions of significant total performance in Panel A of Table 2 reflect the descriptive statistics of total performance: For all categories there are more funds showing a significant positive (negative) total performance, the higher (lower) the mean of total performance within this category is. The similarity of these results for the different models is again striking. The proportions of funds showing a significant alpha coefficient generally reflect those of funds showing a significant total performance but tend to be slightly lower. In the case of the gamma coefficients this difference is much more pronounced with only a small number of funds possessing significant timing performance. Finally, the comparison of the TM and the HM models shows that the latter yields a smaller proportion of significant alpha and gamma

[25] The null hypotheses of an identical mean or median timing performance over all six categories are both rejected at the 1% level. The results of the tests of each category against the total sample are partly mixed but tend to indicate significant differences as well. The details can be supplied by the authors upon request.

coefficients, which stands in contrast to the nearly identical proportions of significant total performance compared to the other measures.

[Insert Table 3 about here]

The correlation matrix in Table 3 shows that the measures are closely related at the level of individual funds: The Pearsson as well as the Spearman correlation coefficients indicate a nearly perfect correlation between all three measures of total performance and an only slightly smaller correlation between the measures of selection or timing performance. Within the TM and the HM models, selection and timing performance are negatively correlated and total performance seems to be much more driven by the respective selection performance than timing performance.

[Insert Table 4 about here]

The descriptive statistics of the differences in total performance as shown in Table 4 particularly confirm that the total performance is nearly identical in the three measures applied: The mean and median differences in total performance are factually negligible. Irrespective of the performance measure applied, 90% of the differences in total performance are within an interval from ca. -2.5 to ca. 2.5 basis points (bp) for the total sample which is only a little more than ± one standard deviation of the difference around the mean. The high kurtosis confirms that the differences are concentrated around the mean in all categories. Altogether the choice of model shows only a very limited influence on the measurement of a fund’s total performance in our sample. This statement is further supported by the results of simple linear cross sample regressions of the total performance measures as presented in Table 5.

[Insert Table 5 about here]

The total performance is explained nearly perfectly as indicated by the high adjusted R^2 of more than 99% in all models we estimate. We test whether the regression constants differ significantly from zero and whether the slope coefficients differ significantly from one. Within our sample, the total performance is frequently translated parallel between the measures applied, by indeed not more than 0.6 bp, which we consider economically irrelevant. We also find evidence of a dilation of total performance which again is very weakly pronounced. In all cases, the cross-sectional regressions confirm once more that the differences in total performances are negligible.

[Insert Table 6 about here]

Table 6 shows the descriptive statistics of the difference in selection performance and the difference in timing performance between the TM and the HM models. We find that the average differences of 5 bp scatter strongly around the mean. The statistics of mean tendency are very similar to each other, while their means (medians) show opposite signs. This holds for all fund categories as well as the total sample and shows that while there are no systematic differences concerning the total performance of funds, selection and timing performance differ systematically between the TM and the HM models. We further assess these relations by conducting cross-sectional regressions explaining the selection (timing) performance of the HM model with the selection (timing) performance of the TM model.

[Insert Table 7 about here]

The adjusted R^2 for the regressions of selection and timing performance range from 64% to 95% in the different fund categories and are still at a considerably high 89% for the total sample in the case of selection performance and 86% for the total sample of timing performance. This indicates a distinct linear relationship between the results. In almost all cases, we find the selection performance of the HM model to be significantly shifted downwards by about 3 to 10 bp per month compared to the selection performance of the TM model. Moreover, the timing performance of the HM model is significantly shifted upwards by roughly the same amount. Furthermore, the selection and timing performance measures of the HM model are significantly dilated as compared to the TM model: With one exception all slope coefficients significantly differ from one. In our sample, the selection and timing performance measured with the HM model is systematically larger in absolute terms than the selection and timing performance measured with the TM model. As the adjusted R^2 and the correlation coefficients in Table 3 show, this indeed has a rather small impact on the fund ranking

5. Conclusion

This paper analyzes the ex-post measured performance of funds according to the JA, the TM and the HM approach. Focusing on the respective selection, timing and total performance in terms of return quantities allows us to directly compare these three models. We first analytically derive the relations between these quantities without any assumptions concerning the market factor and its distribution. We find that the differences in selection, timing and total performance depend on a fund-specific and on a market component. Using a bootstrap approach, we show that our theoretical results are approximately confirmed for realistic distributions of market returns. In particular, we show the differences in total performance to increase with the absolute mean and the standard deviation of the market excess return. However, under realistic market conditions, the JA of a fund is very similar to the fund's total performance according to the TM and the HM approaches.

The empirical analysis of a broad sample of US equity mutual funds confirms that the differences between the JA and the total fund performance based on the TM and the HM approach are negligibly small and mostly insignificant in statistical and economic terms.

Thus, practically, the JA can be interpreted as an ex-post measure of total fund performance. Furthermore, we show that the components of selection and timing performance according to the TM and the HM approach clearly differ. However, cross-sectional regressions of selection (and timing) components reveal that the differences between these two approaches are systematic as they show a similar relation. The correlation between the results of the different approaches shows that its impact on fund rankings based on selection (and timing) performance is almost negligible.

Appendix A: Derivation of the contribution of timing activities to the average abnormal return of funds

The timing performance as the contribution of timing activities to the total abnormal return of a fund is the covariance between the timed market factor and the fund's exposure to that factor.[26] Within the TM model, this covariance tim_i^{TM} is:[27]

$$tim_i^{TM} = \gamma_i^{TM} \sigma_{er_m}^2$$

Within the HM model, the dynamic exposure to the timed market factor $\beta_{i,t}^{HM}$ can be stated as[28]

$$\beta_{i,t}^{HM} = \beta_i^{HM} + \gamma_i^{HM} \frac{\max(0, -er_{m,t})}{er_{m,t}}$$

We derive the covariance of $\beta_{i,t}^{HM}$ and the timed market factor using the general definition of the covariance:

$$\begin{aligned}
tim_i^{HM} &= \mathrm{Cov}\left[\beta_{i,t}^{HM}, er_m\right] \\
&= \mathrm{Cov}\left[\beta_i^{HM} + \gamma_i^{HM} \frac{\max(0, -er_m)}{er_m}, er_m\right] \\
&= \gamma_i^{HM} \mathrm{Cov}\left[\frac{\max(0, -er_m)}{er_m}, er_m\right] \\
&= \gamma_i^{HM} \left[\mathrm{E}\left[\max(0, -er_m)\right] - \mathrm{E}\left[\frac{\max(0, -er_m)}{er_m}\right] \mathrm{E}(er_m)\right] \\
&= \gamma_i^{HM} \left[\mathrm{P}(er_m < 0)\mathrm{E}(er_m) - \mathrm{E}\left[\min(0, er_m)\right]\right]
\end{aligned}$$

---

26 See Grinblatt and Titman (1989), Grinblatt and Titman (1995).
27 See Grinblatt and Titman (1994).
28 See Coles et al. (2006) for a derivation of the covariance if one considers the protective options in the market as calls.

Appendix B: Derivation of the differences in total performance

Assuming a single market factor whose returns are normally distributed allows us to express the constants in Equations (9) and (12) in dependence of the distributional parameters of the market factor. We can state the expected values of the constants in the respective OLS estimation as:

$$\begin{aligned} \mathrm{E}(\alpha_{er_m^2}) &= \mathrm{E}(er_m^2) - \frac{\mathrm{Cov}(er_m, er_m^2)}{\mathrm{Var}(er_m)}\mathrm{E}(er_m) \\ \mathrm{E}[\alpha_{\max(0,-er_m)}] &= \mathrm{E}[\max(0,-er_m)] - \frac{\mathrm{Cov}[er_m, \max(0,-er_m)]}{\mathrm{Var}(er_m)}\mathrm{E}(er_m) \end{aligned}$$

In order to further simplify these equations, we need to derive the respective covariances. If we assume the distribution of the market return as

$$er_m \sim N(\mu_{er_m}, \sigma_{er_m}^2)$$

with the accordant density and distribution functions

$$\begin{aligned} f_{er_m}(x) &= \frac{1}{\sqrt{2\pi\sigma_{er_m}^2}} \exp\left[-\frac{(x-\mu_{er_m})^2}{2\sigma_{er_m}^2}\right] \\ F_{er_m}(x) &= \int_{-\infty}^{x} f_{er_m}(x)dx \end{aligned}$$

We can derive both covariances using the general definition of a covariance:

$$\begin{aligned} \mathrm{Cov}(er_m, er_m^2) &= \mathrm{E}(er_m^3) - \mathrm{E}(er_m)\mathrm{E}(er_m^2) \\ &= \int_{-\infty}^{\infty} er_m^3 f_{er_m}(x)dx - \mu_{er_m}(\sigma_{er_m}^2 + \mu_{er_m}^2) \\ &= \mu_{er_m}^3 + 3\mu_{er_m}\sigma_{er_m}^2 - \mu_{er_m}\sigma_{er_m}^2 - \mu_{er_m}^3 \\ &= 2\mu_{er_m}\sigma_{er_m}^2 \end{aligned}$$

$$
\begin{aligned}
\mathrm{Cov}[er_m, \max(0, -er_m)] &= \mathrm{E}(er_m \max(0, -er_m)) - \mathrm{E}(er_m)\mathrm{E}(\max(0, -er_m)) \\
&= \int_{-\infty}^{\infty} er_m \max(0, -er_m) f_{er_m}(x)dx \\
&\quad -\mu_{er_m} \int_{-\infty}^{\infty} \max(0, -er_m) f_{er_m}(x)dx \\
&= -\int_{-\infty}^{0} er_m^2 f_{er_m}(x)dx + \mu_{er_m} \int_{-\infty}^{0} er_m f_{er_m}(x)dx \\
&= -\sigma_{er_m}^2 F_{er_m}(0)
\end{aligned}
$$

This allows the regression constants to be stated as follows:

$$
\begin{aligned}
\mathrm{E}\left(\alpha_{er_m^2}\right) &= \mathrm{E}\left(er_m^2\right) - \frac{2\mu_{er_m}\sigma_{er_m}^2}{\sigma_{er_m}^2}\mu_{er_m} \\
&= \mu_{er_m}^2 + \sigma_{er_m}^2 - 2\mu_{er_m}^2 \\
&= \sigma_{er_m}^2 - \mu_{er_m}^2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}\left[\alpha_{\max(0,-er_m)}\right] &= \mathrm{E}\left[\max(0, -er_m)\right] + \frac{\sigma_{er_m}^2 F_{er_m}(0)}{\sigma_{er_m}^2}\mu_{er_m} \\
&= -\mathrm{E}\left[\min(0, er_m)\right] + F_{er_m}(0)\mu_{er_m} \\
&= -\sigma_{er_m}^2 f_{er_m}(0)
\end{aligned}
$$

Substituting the expected values into (11) and (14) yields the differences between the total performance based on the TM (HM) approach and the JA under the assumption of a normally distributed market factor:

$$
\begin{aligned}
tot_i^{TM} - \alpha_i^{JA} &= \gamma_i^{TM}\left(\sigma_{er_m}^2 - \alpha_{er_m^2}^{JA}\right) \\
&= \gamma_i^{TM}\mu_{er_m}^2
\end{aligned}
$$

$$
\begin{aligned}
tot_i^{HM} - \alpha_i &= \gamma_i^{HM}\left[\mathrm{P}(er_m < 0)\mathrm{E}(er_m) - \mathrm{E}[\min(0, er_m)] + \sigma_{er_m}^2 f_{er_m}(0)\right] \\
&= \gamma_i^{HM}\left[F_{er_m}(0)\mu_{er_m} - \sigma_{er_m}^2 f_{er_m}(0)\right. \\
&\quad \left. -\mu_{er_m} F_{er_m}(0) + \sigma_{er_m}^2 f_{er_m}(0)\right] \\
&= 0
\end{aligned}
$$

References

Becker, Connie, Ferson, Wayne E., Myers, David H., and Schill, Michael J. (1999). Conditional market timing with benchmark investors. *Journal of Financial Economics,* 52 (1), 119-148.

Bollen, Nicolas P. B. and Busse, Jeffrey A. (2001). On the timing ability of mutual fund managers. *Journal of Finance,* 56 (3), 1075-1094.

Bollen, Nicolas P. B. and Busse, Jeffrey A. (2005). Short-Term persistence in mutual fund performance. *Review of Financial Studies,* 18 (2), 569-597.

Brown, Stephen J. and Goetzmann, William N. (1995). Performance persistence. *Journal of Finance,* 50 (2), 679-698.

Cai, Jun, Chan, K. C., and Yamada, Takeshi (1997). The performance of Japanese mutual funds. *Review of Financial Studies,* 10 (2), 237-274.

Carhart, Mark M. (1997). On persistence in mutual fund performance. *Journal of Finance,* 52 (1), 57-82.

Carhart, Mark M., Carpener, Jennifer N., Lynch, Anthony W., and Musto, David K. (2002). Mutual fund survivorship. *Review of Financial Studies,* 15 (5), 1439-1143.

Chen, Yong and Liang, Bing (2007). Do market timing hedge funds time the market? *Journal of Financial and Quantitative Analysis,* 42 (4), 827-856.

Coles, Jeffrey L., Daniel, Naveen D., and Nardari, Federico (2006). Does the choice of model or benchmark affect inference in measuring mutual fund performance? Working paper, *W. P. Carey School of Business, Arizona State University*.

Comer, George (2006). Hybrid mutual funds and market timing performance. *Journal of Business,* 79 (2), 771-797.

Comer, George, Larrymore, Norris, and Rodriguez, Javier (2007). Controlling for fixed-income exposure in portfolio evaluation: Evidence from hybrid mutual funds. *Review of Financial Studies,* forthcoming.

Elton, Edwin J., Gruber, Martin J., and Blake, Christopher R. (1996). Survivorship bias and mutual fund performance. *Review of Financial Studies,* 9 (4), 1097-1120.

Fama, Eugene F. and French, Kenneth R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics,* 33 (1), 3-56.

Fletcher, Jonathan and Forbes, David (2002). UK unit trust performance: Does it matter which benchmark or measure is used? *Journal of Financial Services Research,* 21 (3), 195-218.

Glosten, L. R. and Jagannathan, Ravi (1994). A contingent claim approach to performance evaluation. *Journal of Empirical Finance,* 1 (2), 133-160.

Goetzmann, William N., Ingersoll Jr., Jonathan, and Ivkovic, Zoran (2000). Monthly measurement of daily timers. *Journal of Financial and Quantitative Analysis,* 35 (3), 257-290.

Goetzmann, William, Ingersoll, Jonathan, Spiegel, Matthew, and Welch, Ivo (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies,* 20 (5), 1503-1546.

Grant, Dwight (1977). Portfolio performance and the cost of timing decisions. *Journal of Finance,* 32 (2), 837-846.

Grinblatt, Mark and Titman, Sheridan (1989). Portfolio performance evaluation: Old issues and new insights. *Review of Financial Studies,* 2 (3), 393-421.

Grinblatt, Mark and Titman, Sheridan (1994). A study of monthly mutual fund returns and performance evaluation techniques. *Journal of Financial and Quantitative Analysis,* 29 (3), 419-444.

Grinblatt, Mark and Titman, Sheridan (1995). Performance Evaluation. In: Jarrow, Robert A., Maksimovic, Vojislav and Ziemba, William T. (Eds.), *Handbooks in Operations Research and Management Science Vol. 9, Finance*, 581-608. Amsterdam, New York: Elsevier.

Henriksson, Roy D. and Merton, Robert C. (1981). On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business,* 54 (4), 513.

Henriksson, Roy D. (1984). Market timing and mutual fund performance: An empirical investigation. Journal of Business, 57 (1), 73-96.

Huij, Joop and Derwall, Jeroen (2008). "Hot hands" in bond funds. *Journal of Banking and Finance,* 32 (4), 559-572.

Jagannathan, Ravi and Korajczyk, Robert A. (1986). Assessing the market timing performance of managed portfolios. *Journal of Business,* 59 (2), 217.

Jensen, Michael C. (1968). The performance of mutual funds in the period 1945-1964. *Journal of Finance,* 23 (2), 389-416.

Jensen, Michael C. (1972). Optimal Utilization of Market Forecasts and the Evaluation of Investment Performance. In: Szego, G. P. and Shell, K. (Eds.), *Mathematical Methods of Investment and Finance*, 310-335. Amsterdam: Elsevier.

Jiang, George J., Yao, Tong, and Yu, Tong (2007). Do mutual funds time the market? Evidence from portfolio holdings. *Journal of Financial Economics,* 86 (3), 724-758.

Kryzanowski, Lawrence, Lalancette, Simon, and To, Minh Chau (1997). Performance attribution using an APT with prespecified macrofactors and time-varying risk premia and betas. *Journal of Financial and Quantitative Analysis,* 32 (2), 205-224.

Lehmann, Bruce N. and Modest, David M. (1987). Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons. *Journal of Finance,* 42 (2), 233-265.

Matallin-Saez, J. C. (2003). Asymmetric relation in omitted benchmarks and market timing in mutual funds. *Applied Economics Letters,* 10 (12), 775-778.

Merton, Robert C. (1981). On market timing and investment performance. I. An equilibrium theory of value for market forecasts. *Journal of Business,* 54 (3), 363-406.

Newey, Whitney K. and West, Kenneth D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica,* 55 (3), 703-708.

Otten, Roger and Bams, Dennis (2004). How to measure mutual fund performance: Economic versus statistical relevance. *Accounting and Finance,* 44 (2), 203-222.

Pástor, Lubos and Stambaugh, Robert F. (2002). Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics,* 63 (3), 315-349.

Pesaran, M. Hashem and Timmermann, Allan G. (1994). A generalization of the non-parametric Henriksson-Merton test of market timing. *Economics Letters,* 44 (1), 1-7.

Roll, Richard (1978). Ambiguity when performance is measured by the securities market line. *Journal of Finance,* 33 (4), 1051-1069.

Treynor, Jack L. and Mazuy, Kay K. (1966). Can mutual funds outguess the market. *Harvard Business Review,* 44 (4), 131-136.

Figure 1: Market component of the difference in total performance between the Treynor-Mazuy model and the Jensen alpha

Figure 1a: Scatter plot against mean of bootstrapped market excess returns

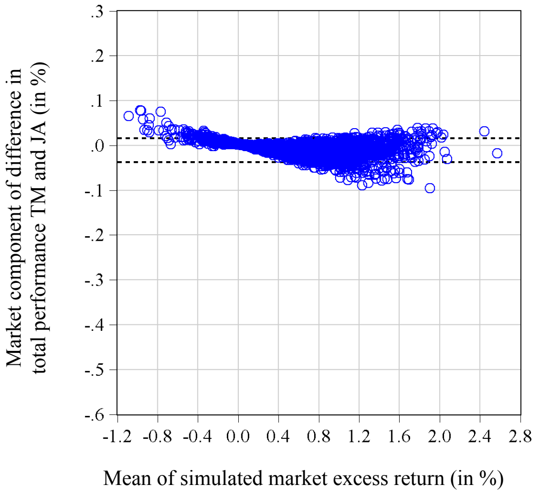

Figure 1b: Scatter plot against standard deviation of bootstrapped market excess returns

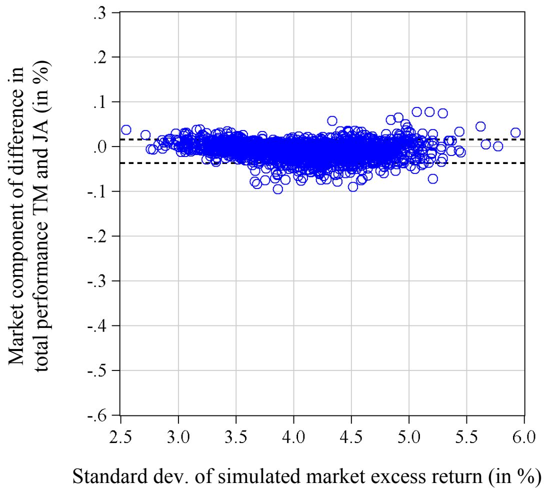

Figure 1c: Scatter plot against skewness of bootstrapped market excess returns

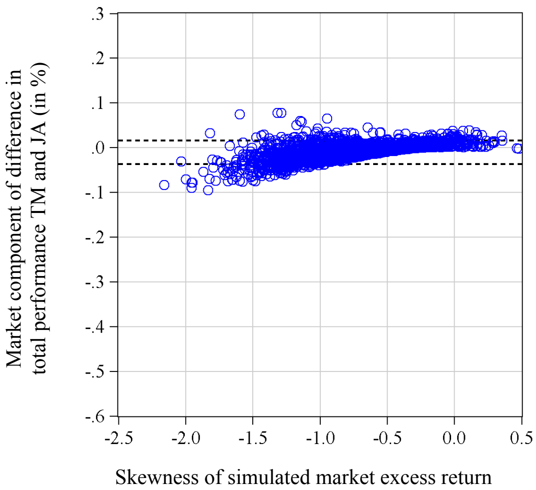

Figure 1d: Scatter plot against kurtosis of bootstrapped market excess returns

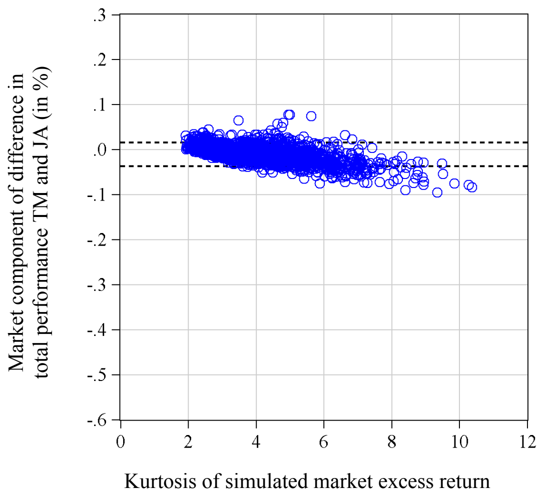

Figure 1 shows scatter plots of the market component of the difference in total performance between the TM model and the JA against the mean (Figure 1a), the volatility (Figure 1b), the skewness (Figure 1c) and the kurtosis (Figure 1d) of the bootstrapped market excess return. The values are based on a bootstrap analysis with 2,500 iterations. For each iteration, a random sample of 60 observations is drawn with replacement from the value-weighted monthly returns of all NYSE, AMEX and NASDAQ stocks between January 1993 and December 2006. Using the random sample, we calculate the market component of the difference between the TM total performance and the JA according to Equation (11). The dashed horizontal lines indicate the 5%- and the 95%-quantile of the difference in total performance.

Figure 2: Market component of the difference in total performance between the Henriksson-Merton model and the Jensen alpha

Figure 2a: Scatter plot against mean of bootstrapeed market excess returns

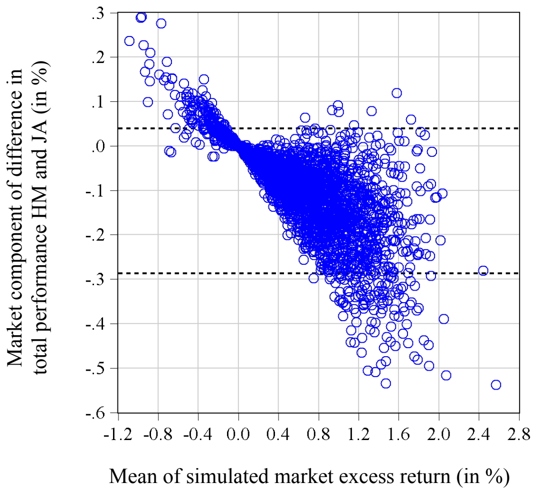

Figure 2b: Scatter plot against standard deviation of bootstrapped market excess returns

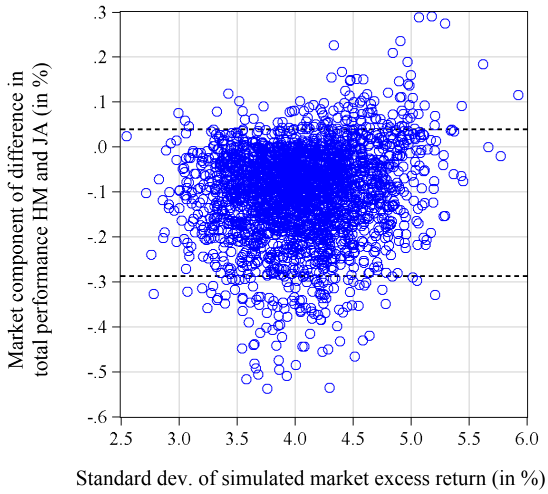

Figure 2c: Scatter plot against skewness of bootstrapped market excess returns

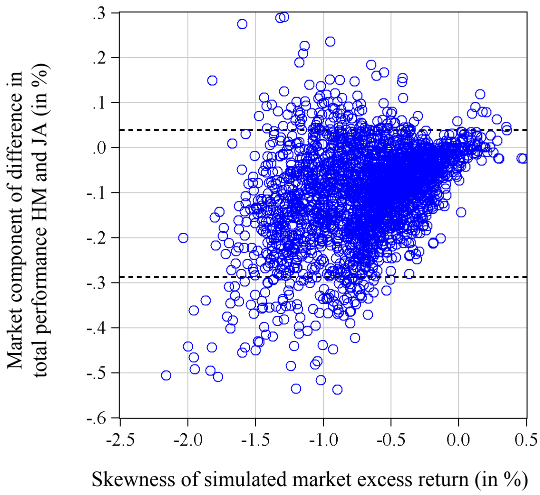

Figure 2d: Scatter plot against kurtosis of bootstrapped market excess returns.

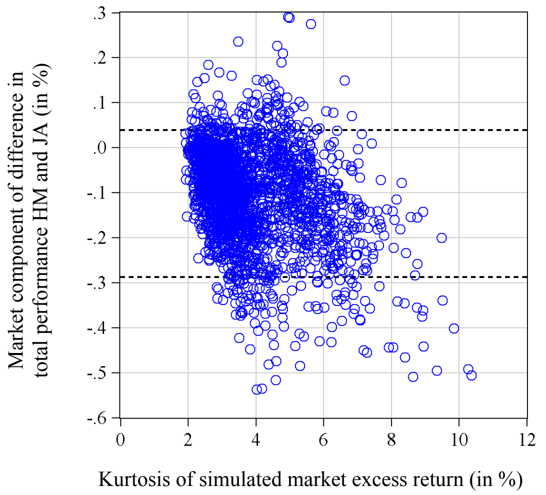

Figure 2 shows scatter plots of the market component of the difference in total performance between the HM model and the JA against the mean (Figure 2a), the standard deviation (Figure 2b), the skewness (Figure 2c) and the kurtosis (Figure 2d) of the bootstrapped market excess return. The values are based on a bootstrap analysis with 2,500 iterations. For each iteration, a random sample of 60 observations is drawn with replacement from the value-weighted monthly returns of all NYSE, AMEX and NASDAQ stocks between January 1993 and December 2006. Using the random sample, we calculate the market component of the difference between the HM total performance and the JA according to Equation (14). The dashed horizontal lines indicate the 5%- and the 95%-quantile of the difference in total performance.

Table 1: Descriptive statistics of the measured performance of the funds

Fund category	Number of funds	JA		TM		HM	
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Panel A: Total performance							
AGG	395	-0.26	0.61	-0.27	0.60	-0.26	0.60
GMC	725	0.07	0.43	0.07	0.43	0.07	0.43
GRI	1,452	-0.07	0.24	-0.07	0.24	-0.08	0.24
GRO	2,514	-0.21	0.34	-0.21	0.34	-0.21	0.33
ING	279	0.08	0.27	0.09	0.27	0.08	0.28
SCG	1,488	0.11	0.59	0.11	0.59	0.12	0.58
Total sample	6,853	-0.07	0.44	-0.07	0.44	-0.07	0.44
Panel B: Selection performance							
AGG	395			-0.31	0.79	-0.38	1.06
GMC	725			0.12	0.42	0.07	0.46
GRI	1,452			-0.10	0.27	-0.14	0.32
GRO	2,514			-0.24	0.36	-0.28	0.42
ING	279			0.09	0.30	0.06	0.37
SCG	1,488			0.19	0.67	0.13	0.91
Total sample	6,853			-0.07	0.50	-0.12	0.62
Panel C: Timing performance							
AGG	395			0.04	0.44	0.12	0.75
GMC	725			-0.05	0.22	0.00	0.35
GRI	1,452			0.02	0.13	0.06	0.21
GRO	2,514			0.02	0.19	0.07	0.30
ING	279			0.00	0.12	0.03	0.21
SCG	1,488			-0.07	0.26	-0.02	0.58
Total sample	6,853			0.00	0.22	0.04	0.41

Table 1 presents the mean and the standard deviation of selection, timing and total performance as measured by the JA, the TM and the HM models. We apply the measures by estimating Equations (1) to (3) for each fund individually, using the value-weighted index of all NYSE, AMEX, and NASDAQ stock returns above the one-month Treasury Bill rate as market index. We calculate the timing performance using Equations (5) and (6) and the total performance as the sum of selection performance and timing performance for both models. The means and standard deviation are calculated over the individual funds in each fund category and for the total sample. All performance measures are stated in percent.

Table 2: Proportions of positive- and negative-measured performance of funds

Fund category	JA				TM				HM			
	+	–	+ +	– –	+	–	+ +	– –	+	–	+ +	– –
Panel A: Total performance												
AGG	31.4	68.6	4.8	22.5	31.6	68.4	4.8	22.8	31.6	68.4	4.8	22.8
GMC	58.8	41.2	16.3	4.1	57.7	42.3	17.1	4.7	57.8	42.2	17.2	3.7
GRI	34.2	65.8	2.6	19.9	34.0	66.0	2.8	19.4	33.7	66.3	2.6	20.3
GRO	24.3	75.7	1.3	28.6	24.1	75.9	1.3	29.1	24.1	75.9	1.5	29.0
ING	62.7	37.3	13.6	1.1	63.1	36.9	12.5	1.4	62.4	37.6	13.3	1.1
SCG	60.4	39.6	9.9	5.1	60.4	39.6	10.4	5.2	60.1	39.9	10.4	5.0
Total sample	39.9	60.1	5.7	17.6	39.7	60.3	5.9	17.8	39.5	60.5	6.0	17.8
Panel B: Alpha coefficient												
AGG					28.4	71.6	4.6	16.7	25.6	74.4	1.8	11.1
GMC					63.0	37.0	11.9	0.6	60.6	39.4	5.1	0.8
GRI					32.1	67.9	2.6	22.9	31.3	68.7	2.4	19.3
GRO					23.7	76.3	2.1	28.6	24.5	75.5	1.0	23.2
ING					65.6	34.4	9.3	2.2	61.6	38.4	5.0	2.2
SCG					65.7	34.3	12.1	2.2	63.0	37.0	3.9	1.5
Total sample					40.7	59.3	5.8	16.9	39.7	60.3	2.6	13.7
Panel C: Gamma coefficient												
AGG					54.4	45.6	4.3	5.1	59.0	41.0	1.8	1.8
GMC					42.5	57.5	0.6	4.1	56.8	43.2	1.2	1.9
GRI					62.3	37.7	4.4	1.7	65.4	34.6	2.1	1.4
GRO					58.6	41.4	4.6	2.9	61.9	38.1	4.4	1.2
ING					49.5	50.5	4.7	1.1	54.1	45.9	2.5	2.2
SCG					35.5	64.5	0.9	5.4	46.6	53.4	0.9	1.2
Total sample					52.1	47.9	3.3	3.4	58.3	41.7	2.6	1.4

Table 2 presents the proportions of positive (+) and negative (–) as well as significantly positive (+ +) and significantly negative (– –) performance measures according to a *t*-test for a significance level of 5%. We calculate the standard deviation of the total performance of the TM and the HM models σ_i^{tot} as follows:

$$\sigma_i^{tot} = \sqrt{\sigma_i^{\alpha\alpha} + \xi^2\sigma_i^{\gamma\gamma} + 2\xi\sigma_i^{\alpha\gamma}}$$

where σ^{jk} stands for the (co-)variance of the indicated regression coefficients and ξ stands for the variance of excess return of the market in the case of the TM model and for the term in squared brackets on the right side of Equation (6) in the case of the HM model. We estimate heteroscedasticity and autocorrelation consistent standard errors for the regression coefficients according to Newey and West (1987). All numbers are in percent.

Table 3: Correlation between the performance measures

	Total performance			Selection performance		Timing performance	
	JA	TM	HM	TM	HM	TM	HM
Total JA	1	0.9986	0.9987	0.9003	0.7540	0.0181	0.0922
Total TM	0.9987	1	0.9982	0.8995	0.7531	0.0194	0.0920
Total HM	0.9988	0.9985	1	0.9125	0.7744	-0.0077	0.0636
Selection TM	0.8975	0.8968	0.9083	1	0.9406	-0.3456	-0.2533
Selection HM	0.7445	0.7439	0.7629	0.9416	1	-0.5227	-0.5013
Timing TM	-0.0511	-0.0471	-0.0759	-0.4842	-0.6547	1	0.9174
Timing HM	-0.0703	-0.0697	-0.0973	-0.4712	-0.7177	0.9260	1

Table 3 shows correlation coefficients for the performance measures of the total sample of 6,853 funds. The lower triangular matrix shows ordinary Pearson correlation coefficients, the upper triangular matrix Spearman rank correlation coefficients between the stated performance measures. All values are in absolute numbers.

Table 4: Descriptive statistics of the differences in total performance

Fund category	Mean	Median	95% quantile	5% quantile	Standard deviation	Skewness	Kurtosis
Panel A: Difference between TM total performance and JA							
AGG	-0.0023	-0.0008	0.0266	-0.0428	0.0263	-0.3646	21.07
GMC	-0.0052	-0.0002	0.0125	-0.0459	0.0284	-3.9253	30.65
GRI	0.0015	-0.0001	0.0169	-0.0094	0.0139	4.7180	52.96
GRO	-0.0016	-0.0005	0.0167	-0.0209	0.0177	-1.5065	27.87
ING	0.0028	0.0012	0.0266	-0.0111	0.0147	-0.4215	16.64
SCG	0.0010	0.0008	0.0264	-0.0342	0.0296	-0.8531	29.01
Total sample	-0.0006	-0.0001	0.0209	-0.0227	0.0220	-1.4510	39.05
Panel B: Difference between HM total performance and JA							
AGG	-0.0002	-0.0007	0.0398	-0.0321	0.0300	3.7275	33.14
GMC	-0.0013	-0.0014	0.0176	-0.0233	0.0157	2.7718	34.98
GRI	-0.0016	-0.0012	0.0155	-0.0167	0.0127	-1.5527	23.65
GRO	-0.0020	-0.0011	0.0190	-0.0252	0.0186	2.4083	44.37
ING	0.0005	-0.0002	0.0255	-0.0173	0.0149	-0.5801	11.09
SCG	0.0023	0.0002	0.0338	-0.0326	0.0308	1.9974	36.09
Total sample	-0.0007	-0.0009	0.0230	-0.0251	0.0214	2.6517	53.13
Panel C: Difference between TM total performance and HM total performance							
AGG	-0.0021	0.0017	0.0257	-0.0579	0.0322	-3.4641	28.55
GMC	-0.0039	0.0021	0.0216	-0.0546	0.0314	-4.3135	30.79
GRI	0.0031	0.0015	0.0235	-0.0123	0.0161	4.5495	49.85
GRO	0.0004	0.0015	0.0239	-0.0284	0.0209	-2.4166	33.29
ING	0.0023	0.0017	0.0326	-0.0223	0.0149	1.0619	7.96
SCG	-0.0013	0.0021	0.0268	-0.0523	0.0285	-2.8683	30.42
Total sample	0.0001	0.0017	0.0249	-0.0302	0.0239	-3.0125	39.46

Table 4 presents the mean, the median, the 5% and the 95%-quantile, the standard deviation, the (centered) skewness and the (centered) kurtosis of the difference in total performance. We estimate the models in Equations (1) to (3), using the value-weighted index of all NYSE, AMEX, and NASDAQ stock returns above the one-month Treasury Bill rate as market index. We calculate the difference in total performance according to Equations (11), (14) and (15). All values are in percent.

Table 5: Cross-sectional regression of total performance

Fund category	Constant	Slope	Adjusted R^2	Number of funds
Panel A: JA explains TM total performance				
AGG	-3.93E-05 ***	0.994 *	0.998	395
GMC	-6.14E-05 ***	1.012 **	0.996	725
GRI	1.53E-05 ***	1.000	0.997	1,452
GRO	-1.53E-05 ***	1.001	0.997	2,514
ING	2.87E-05 ***	0.999	0.997	279
SCG	1.13E-05	0.999	0.997	1,488
Total sample	-5.85E-06 *	1.001	0.997	6,853
Panel B: JA explains HM total performance				
AGG	-2.63E-05 *	0.991 **	0.998	395
GMC	-9.89E-06	0.995	0.999	725
GRI	-1.62E-05 ***	0.999	0.997	1,452
GRO	-3.15E-05 ***	0.995 ***	0.997	2,514
ING	1.30E-06	1.004	0.997	279
SCG	2.96E-05 **	0.994	0.997	1,488
Total sample	-9.91E-06 ***	0.996 **	0.998	6,853
Panel C: TM total performance explains HM total performance				
AGG	-1.74E-05	0.996	0.997	395
GMC	-5.05E-05 ***	0.979 ***	0.995	725
GRI	3.01E-05 ***	0.997	0.996	1,452
GRO	1.19E-05 *	0.992 ***	0.996	2,514
ING	2.86E-05 ***	1.004	0.997	279
SCG	-1.70E-05	0.994	0.998	1,488
Total sample	3.07E-06	0.994 ***	0.997	6,853

Table 5 presents the results of cross-sectional regressions of total performance. We use a simple linear regression to quantify the relations between the different performance measures as indicated by the panels. Standard errors are heteroscedasticity and autocorrelation consistent according to Newey and West (1987). The columns "Constant" and "Slope" show the respective regression coefficients. The asterisks indicate significance levels of 1% (***), 5% (**), and 10% (*) for significance tests with the H_0 of $x = 0$ in the case of the regression constants and H_0 of $x = 1$ in the case of the slope coefficients.

Table 6: Descriptive statistics of the differences in selection and timing performance

Fund category	Mean	Median	95% quantile	5% quantile	Standard deviation	Skewness	Kurtosis
Panel A: Difference in selection performance between TM and HM models							
AGG	0.0754	0.0517	0.4380	-0.3148	0.3548	7.7425	116.48
GMC	0.0473	0.0717	0.2819	-0.2791	0.1725	-1.3177	6.82
GRI	0.0446	0.0428	0.2122	-0.1340	0.1186	1.2290	13.18
GRO	0.0454	0.0403	0.2883	-0.1866	0.1551	0.5261	8.15
ING	0.0301	0.0225	0.2613	-0.1855	0.1449	1.0031	5.96
SCG	0.0566	0.0563	0.3611	-0.3283	0.3632	12.266	244.11
Total sample	0.0490	0.0461	0.3062	-0.2075	0.2274	12.658	387.18
Panel B: Difference in timing performance between TM and HM models							
AGG	-0.0775	-0.0516	0.2885	-0.4151	0.3408	-8.3036	127.17
GMC	-0.0512	-0.0701	0.2246	-0.2678	0.1568	1.3228	7.50
GRI	-0.0415	-0.0388	0.1171	-0.2088	0.1104	-1.2863	14.24
GRO	-0.0451	-0.0384	0.1621	-0.2748	0.1438	-0.7188	8.95
ING	-0.0278	-0.0191	0.1759	-0.2406	0.1367	-1.1651	6.53
SCG	-0.0579	-0.0558	0.2896	-0.3368	0.3510	-12.8858	262.20
Total sample	-0.0489	-0.0442	0.1879	-0.2925	0.2169	-13.8806	438.58

Table 6 presents the mean, the median, the 5% and the 95% quantile, the standard deviation, the (centered) skewness and the (centered) kurtosis of the differences in selection and timing performance. We estimate the models in Equations (2) and (3), using the value-weighted index of all NYSE, AMEX, and NASDAQ stock returns above the one-month Treasury Bill rate as market index. We calculate the difference in selection and timing performance according to Equation (15). All values are in percent.

Table 7: Cross-sectional regression of selection and timing performance

Fund category	Constant	Slope	Adjusted R^2	Number of funds
Panel A: TM selection performance explains HM selection performance				
AGG	1.51E-04	1.30 ***	0.937	395
GMC	-4.93E-04 ***	1.02	0.859	725
GRI	-3.54E-04 ***	1.10 ***	0.865	1,452
GRO	-2.65E-04 ***	1.08 ***	0.866	2,514
ING	-4.41E-04 ***	1.16 ***	0.862	279
SCG	-1.08E-03 ***	1.27 ***	0.882	1,488
Total sample	-3.71E-04 ***	1.17 ***	0.881	6,853
Panel B: TM timing performance explains HM timing performance				
AGG	5.03E-04 ***	1.68 ***	0.951	395
GMC	7.73E-04 ***	1.54 ***	0.913	725
GRI	2.91E-04 ***	1.55 ***	0.835	1,452
GRO	3.20E-04 ***	1.53 ***	0.882	2,514
ING	2.76E-04 ***	1.40 ***	0.641	279
SCG	1.34E-03 ***	2.04 ***	0.847	1,488
Total sample	5.23E-04 ***	1.69 ***	0.858	6,853

Table 7 presents the results of cross-sectional regressions of selection and timing performance. We use a simple linear regression to quantify the relations between the different performance measures as indicated by the panels. Standard errors are heteroscedasticity and autocorrelation consistent according to Newey and West (1987). The columns "Constant" and "Slope" show the respective regression coefficients. The asterisks indicate significance levels of 1% (***), 5% (**), and 10% (*) for significance tests with the H_0 of $x = 0$ in the case of the regression constants and H_0 of $x = 1$ in the case of the slope coefficients.