Jensen's alpha and the market-timing puzzle

Sebastian Bunnenberg[1] | Martin Rohleder[2] | Hendrik Scholz[3] | Marco Wilkens[2]

[1]Finance, ESB Business School, Reutlingen University, Reutlingen, Germany

[2]Finance and Banking, University of Augsburg, Augsburg, Germany

[3]Finance and Banking, Friedrich-Alexander-University Erlangen-Nürnberg, Nürnberg, Germany

Correspondence
Sebastian Bunnenberg, Finance, ESB Business School of Reutlingen University, Alteburgstraße 150, 72762 Reutlingen, Germany.
Email: sebastian.bunnenberg@reutlingen-university.de

Abstract

Theory predicts that market-timing activities bias Jensen's alpha (JA). However, empirical studies have failed to find consistent evidence of this bias. We tackle this puzzle in a nested model analysis and show that the bias contains an exogenous market component that is unrelated to market-timing skill. In a comprehensive empirical analysis of US mutual funds, we find that the timing-induced bias in JA is mainly driven by this market component, which is uncorrelated with measured timing activities. Measures of total performance that allow for timing activities are virtually identical to JA, even if timing activities are present in the evaluated fund. Hence, we conclude that JA is a sufficient measure of total performance.

JEL CLASSIFICATION
G11, G23

KEYWORDS
market-timing, mutual fund performance, stock selection, total performance

1 | INTRODUCTION

Jensen's (1968) alpha (JA) and its multifactor variants are still among the most widely used approaches to measure the economic value that portfolio managers add for their clients. However, criticizing his own model, Jensen (1972) shows that JA is biased downward when applied to funds that successfully engage in market-timing activities: A manager who perfectly times the market may have a statistically significant, negative JA. Treynor and Mazuy (1966) (TM) and Henriksson and Merton (1981) (HM) suggest performance models that allow the separation of funds' selection activities from their timing activities. While these models are still standard approaches in the literature,[1] the impact of market-timing on JA is hardly ever discussed. Moreover, the empirical studies of this subject find no evidence of a significant market-timing bias, leaving a disparity between theoretical and empirical research. We address this puzzle and provide consistent evidence why the timing-induced bias in JA is usually irrelevant. Thus, our key contribution is the demonstration that JA is an adequate empirical measure of total performance for mutual funds.

To briefly illustrate how timing activities bias JA, Figure 1 shows scatterplots of portfolio excess returns over market excess returns in a stylized stochastic simulation. Assuming a single-factor model, we draw 36 market excess returns from a normal distribution with a mean of 0.5% and a volatility of 4.536%.[2] The portfolio excess returns follow either a TM market-timing strategy—that is, exposure to market risk depends linearly on the market excess return—or an HM market-timing strategy, which implies that the portfolio manager chooses between two levels of market risk, depending on the sign of the market risk premium. In both plots, a linear regression of portfolio excess returns against market excess returns as depicted has a positive intercept, that is, JA. In our study, we analyze if JA unbiasedly reflects the joint economic value that results from selection activities and timing activities present in portfolios that exhibit such timing qualities.

Figure 1 illustrates that the presence of market-timing activities violates the assumption of a linear relation between portfolio and market excess returns, which is implicit within JA. Grant (1977) relates the impact of this violation to parameters

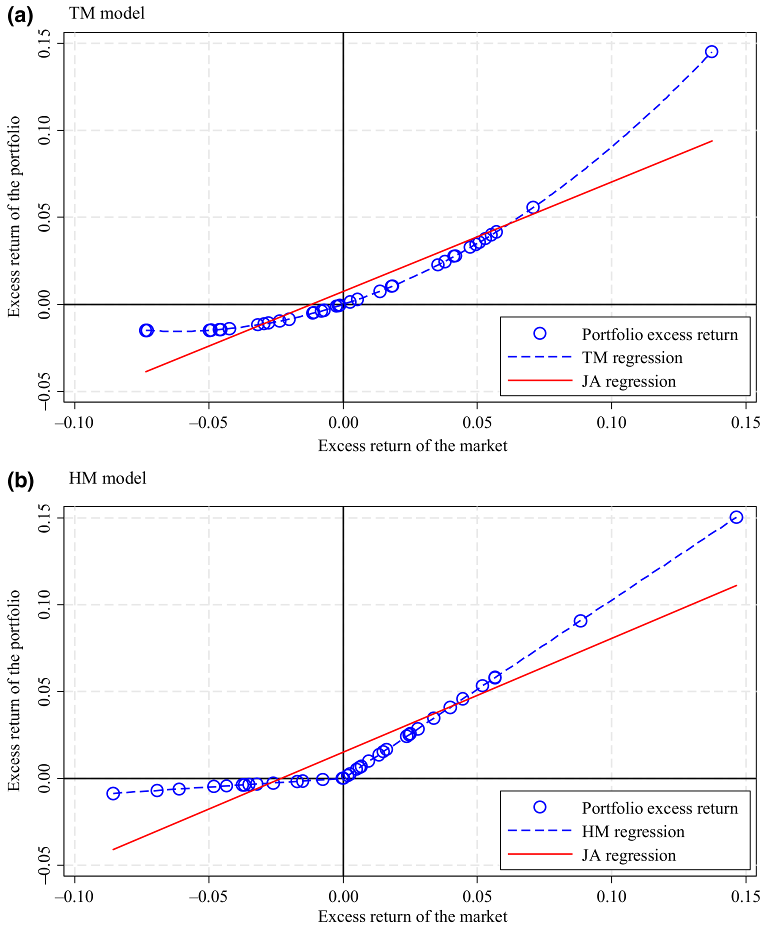

FIGURE 1 Stylized illustrations of the bias in JA as a result of timing activities according to the TM and HM models. For these illustrations, we sample 36 market excess returns from an identical and independent normal distribution with an expected value of 0.5% and a standard deviation of 4.536%. Using these sampled market excess returns, we generate two portfolios which follow market-timing activities as imposed by the TM and HM models. (a) The portfolio excess returns follow a TM market-timing strategy: The exposure to market risk depends linearly on the market excess return, resulting in a quadratic relation between portfolio and market excess returns. (b) The portfolio excess returns follow a HM market-timing strategy: We vary between a low level and a high level of exposure to market risk with the sign of the market excess return. In this case, the relation between portfolio and market excess return is kinked for a market excess return of 0 [Colour figure can be viewed at wileyonlinelibrary.com]

of the distribution of the market excess return. Admati and Ross (1985) and Dybvig and Ross (1985) show that, under conditions of market equilibrium, the sign of JA is unpredictable in the presence of successful market-timing activities. Grinblatt and Titman (1989) attribute the timing-induced bias in JA to a false estimation of a fund's exposure to market risk during the evaluation period, which results from the varying exposure to market risk implied by timing activities. However, none of these studies addresses the problem empirically.

Grinblatt and Titman (1994) provide empirical evidence by comparing the measured performance of JA with that of the TM model for a small sample of 279 mutual funds. They find that the choice of model has little impact on the results. However, they do not incorporate the HM model into their considerations. Coles, Daniel, and Nardari (2006) compare JA with TM and HM total performance using bootstrapped mutual fund returns. They, too, conclude that the choice of model has no significant impact on measured total performance. However, neither study analyzes why the theoretically predicted market-timing bias is not observed in the empirical results, leaving this an unresolved puzzle.

We contribute a solution of the market-timing puzzle in mutual fund performance in three ways. First, using measures of timing performance and total performance based on the TM and HM models, we take a nested model perspective on the issue, similar to Aragon and Ferson (2006) and Chen, Ferson, and Peters (2010). We show that the market-timing bias is determined by two components: the extent of market-timing activity, which is a fund-specific component, and the market conditions experienced by the fund, which are an exogenous component beyond the control of the fund manager.

Furthermore, we show that the relation between both components is multiplicative so that the correlation between these components is highly relevant to the magnitude of the empirically realized bias.

Second, we analyze the timing-induced bias in JA via simulations. In a stochastic simulation with normally distributed market excess returns, the bias in JA is small and mostly insignificant if timing activities according to the TM model are present. If a portfolio exhibits timing activities according to the HM model, JA is even unbiased with respect to the true HM total performance. In a bootstrap simulation, we document how the empirical distribution of the market excess returns affects these results: For TM-timing activities, JA tends to be negatively biased and may largely underestimate the true total performance. For HM-timing activities, the bias in JA is very small and exhibits only little dispersion. In both cases, however, the bias is hardly ever statistically significant.

Third, we perform an empirical analysis of the timing-induced bias in the Carhart (1997) variant of JA using a comprehensive sample of 1,289 actively managed US domestic equity funds with monthly and daily returns from January 1987 through December 2014 obtained from the CRSP and Morningstar. As total performance is factually identical for JA and the TM and HM models in the cross section of our fund sample, we find hardly any empirical evidence of this bias in our sample. Minor differences remain, even at a statistically significant level, but they are too small to be of economic relevance. In a rolling window analysis, we are able to explain this puzzling result: We find nearly no evidence of the timing-induced bias in JA in our empirical study because both components of the bias are small and they are weakly correlated with each other. In cases in which a significant bias exists, this is mostly caused by the exogenous component and not by the timing skill of fund managers. Thus, the timing-induced bias in JA is economically irrelevant and JA and its multifactor variants are a sufficient measure of total performance for mutual funds.

In contrast, measures of selection performance and timing performance differ systematically between the TM and HM models, and by a much higher margin. Hence, the separation of total performance into selection and timing is highly sensitive to the choice of timing model, creating a critical dual hypothesis problem. As the true type of timing activities adopted by the manager is usually unknown for empirical data, the separation by either model leads to potentially false inferences.

The remainder of this paper is organized as follows. In Section 2, we introduce the performance models used in our analysis and measures of timing performance and total performance as abnormal in-sample returns. In Section 3, we derive the components of the market-timing bias via a nested model approach and test their impact via different simulations. In Section 4, we discuss the results of our empirical analysis of the timing-induced bias in JA. Section 5 concludes the paper.

2 | PERFORMANCE MEASUREMENT OF TIMING ACTIVITIES

2.1 | JA and a generalized model of timing activities

For brevity, we assume a single-factor market as in Jensen (1968).[3] In such a market, the population of fund returns in excess of the risk-free rate, er_{it}, is generated by

$$er_{it} = \alpha_i + \beta_i er_{mt} + \varepsilon_{it} \tag{1}$$

with i and t denoting the fund and the period, respectively, er_{mt} being the contemporaneous excess return on the market portfolio, and ε_{it} being an independently and identically distributed (i.i.d) residual term with an expected value of 0. The coefficient β_i is the exposure to market risk of fund i, and α_i is the performance as measured by JA: The average return generated by the management of fund i that is not attributable to its exposure to market risk. Equation (1) imposes a fund-specific but constant exposure to market risk β_i, thereby excluding any timing activity by definition. Under this assumption, a fund's α_i represents its ex post average abnormal return resulting from stock selection, henceforth selection performance.

In contrast, the TM and HM models allow for time-varying exposure to market risk. For both models, a generalized data-generating process in the population is represented by

$$er_{it} = \delta_i + \zeta_i er_{mt} + \gamma_i f(er_{mt}) er_{mt} + \eta_{it} \tag{2}$$

The residual η_{it} is again assumed to be i.i.d. with an expected value of 0. The exposure to market risk of fund i is $\zeta_i + \gamma_i f(er_{mt})$ and varies with er_{mt}, thereby explicitly allowing for market-timing activities. The constant δ_i is the average return of fund i in excess of the premium for the dynamic exposure to market risk[4] and is often interpreted as a measure of selection activities.[5]

The TM and HM models vary in the specification of $f(er_{mt})$: The TM model assumes that fund managers linearly increase their exposure to market risk with the market risk premium. Hence, it is based on $f(er_{mt}) = er_{mt}$ and the resulting data-generation process is quadratic in er_{mt}:

$$er_{it} = \delta_i^{TM} + \zeta_i^{TM} er_{mt} + \gamma_i^{TM} er_{mt}^2 + \eta_{it}^{TM} \tag{3}$$

The sensitivity ζ_i^{TM} is the exposure to market risk, given $er_{mt} = 0$. The coefficient γ_i^{TM} reflects the adjustment of the fund's market risk exposure when $er_{mt} \neq 0$. Positive values of γ_i^{TM} indicate successful market timing.

The HM model assumes a binary dynamic of market risk exposure: Depending on the realization of the market risk premium with respect to a given reference point, the fund management varies between a low and a high level of market risk. Commonly, this reference point is $er_{mt} = 0$, which results in the binary variable $f(er_{mt}) = \frac{\max(0, er_{mt})}{er_{mt}}$, and the data-generating process is

$$er_{it} = \delta_i^{HM} + \zeta_i^{HM} er_{mt} + \gamma_i^{HM} \max(0, er_{mt}) + \eta_{it}^{HM} \tag{4}$$

The coefficient ζ_i^{HM} indicates exposure to market risk in times of negative market returns. The coefficient γ_i^{HM} measures timing activity as the change in exposure to market risk in times of positive market returns. Positive values of γ_i^{HM} indicate successful market timing.

2.2 | Measuring timing performance and total performance

It is important to note that the coefficients δ_i and γ_i in Equation (2) differ not only in terms of their economic interpretation but also in terms of units: While δ_i is a return, γ_i is not. To determine the contribution of timing activities as measured abnormal return in-sample, the timing performance tim_i, we refer to Grinblatt and Titman (1989). Taking the expectation of Equation (2), they show that timing performance is the covariance Cov(·) between the market excess return, er_{mt}, and fund i's dynamic exposure to market risk, $\zeta_i + \gamma_i \cdot f(er_{mt})$[6]:

$$\begin{aligned} tim_i &= \mathrm{Cov}[er_{mt}, \zeta_i + \gamma_i f(er_{mt})] \\ &= \gamma_i \cdot \mathrm{Cov}[er_{mt}, f(er_{mt})] \end{aligned}$$

Given tim_i, we can identify the total performance tot_i as the economic value added by the fund management through their selection and timing activities.[7]

$$tot_i = \delta_i + \gamma_i \mathrm{Cov}[er_{mt}, f(er_{mt})]$$

As selection and timing performance are each expressed as abnormal return in-sample, this must also apply to tot_i. We implement these measures for the TM model and the HM model as follows[8] :

$$tim_i^{TM} = \gamma_i^{TM} \mathrm{Var}(er_{mt}) \tag{5}$$

$$tim_i^{HM} = \gamma_i^{HM} [\mathrm{E}(\max(0, er_{mt})) - \mathrm{E}(er_{mt}) \Pr(er_{mt} > 0)] \tag{6}$$

$$tot_i^{TM} = \delta_i^{TM} + \gamma_i^{TM} \mathrm{Var}(er_{mt}) \tag{7}$$

$$tot_i^{HM} = \delta_i^{HM} + \gamma_i^{HM} [\mathrm{E}(\max(0, er_{mt})) - \mathrm{E}(er_{mt}) \Pr(er_{mt} > 0)] \tag{8}$$

where E(·), Var(·), and Pr(·) denote expectations, variances, and probabilities, respectively. An online appendix shows derivations of timing and total performance for conditional models, in which selection and timing activity vary with the economic cycle as in Kacperczyk, van Nieuwerburgh, and Veldkamp (2014).

Finally, to test tim_i and tot_i for statistical significance, we calculate their standard errors as follows

$$\mathrm{SE}(tim_i) = \mathrm{SE}(\gamma_i) \cdot \mathrm{Cov}[er_{mt}, f(er_{mt})] \tag{9}$$

$$\mathrm{SE}(tot_i) = \sqrt{\mathrm{SE}(\delta_i)^2 + \mathrm{SE}(tim_i)^2 + 2\,\mathrm{Cov}[er_{mt}, f(er_{mt})]\,\mathrm{Cov}(\delta_i, \gamma_i)} \tag{10}$$

where SE(·) denotes the standard error of the given coefficient.[9] The covariance Cov(δ_i, γ_i) is the estimated covariance between the corresponding coefficients in Equation (2). With these standard errors, we test timing and total performance for statistical significance in one-sample t tests.

3 | THE TIMING-INDUCED BIAS IN JA

3.1 | Nested model analysis of JA

We first analyze how timing activities affect JA by considering the fact that JA is nested within Equation (2): Under the restriction of no timing activities, that is, $\gamma_i = 0$, Equation (2) is identical to JA in Equation (1). Furthermore, timing activities constitute a nonlinear relationship between er_{mt} and er_{it}, so JA imposes an incorrect functional form on fund excess returns if timing activities are present. However, we are able to derive the population JA of a fund given its return-generating process follows Equation (2) with $\gamma_i \neq 0$.[10]

For this purpose, we apply a "maximum squared correlation portfolio" approach similar to Aragon and Ferson (2006), section 2.2.1.2, and Chen et al. (2010), section 4.4. To replicate timing activities, we identify the passive investment in the market factor that exhibits the highest squared correlation with $y_t = f(er_{mt}) \cdot er_{mt}$. For this purpose, we estimate the following regression via ordinary least squares (OLS):

$$y_t = a_h + b_h er_{mt} + \Delta_{ht} \tag{11}$$

The coefficient b_h in Equation (11) is the holding in the market portfolio that best approximates the nonlinear payout from market-timing activities, $f(er_{mt}) \cdot er_{mt}$. The intercept a_h is the mean approximation error, and the residual Δ_{ht} is the demeaned approximation error in period t. Economically, a_h can be interpreted as the component in the expected value of y_t that cannot be replicated by a passive investment in the market. While conditional mean independence does not hold for Equation (11), that is, $\mathrm{E}(\Delta_{ht}|er_{mt}) \neq 0$, estimating Equation (11) via OLS correctly identifies the passive investment that exhibits the highest squared correlation with $y_t = f(er_{mt}) \cdot er_{mt}$ in the sample.[11]

Next, we substitute $f(er_{mt}) \cdot er_{mt}$ in Equation (2) with Equation (11) and rearrange the result to depict its relationship with JA in Equation (1).

$$er_{it} = \underbrace{(\delta_i + \gamma_i a_h)}_{\alpha_i} + \underbrace{(\zeta_i + \gamma_i b_h)}_{\beta_i} er_{mt} + \underbrace{(\eta_{it} + \gamma_i \Delta_{ht})}_{\varepsilon_{it}} \tag{12}$$

Imposing the restriction $\gamma_i = 0$ in the performance measurement on a portfolio that exhibits timing activities induces a bias in all coefficients of the restricted model. The restricted model is JA, so the measured JA, α_i, is identical to the intercept of the return-generating process, δ_i, plus the average approximation error from Equation (11), a_h, scaled by the extent of timing activities, γ_i.

$$\alpha_i = \delta_i + \gamma_i a_h \tag{13}$$

Thus, JA necessarily reflects selection and timing activities, if a fund exhibits the latter: The term $\gamma_i \cdot a_h$ reflects the component in the expected portfolio return that cannot be achieved by a passive investment in the market.[12] Accordingly, we conclude that JA is a measure of total performance in empirical applications, even if its results are biased.[13] We decompose this timing-induced bias into two components by relating JA to the true total performance given market-timing activities, tot_i :

$$\alpha_i - tot_i = \gamma_i [a_h - \mathrm{Cov}[er_{mt}, f(er_{mt})]] \tag{14}$$

The difference between JA and the true total performance tot_i stems from different valuations of the measured timing activities γ_i: Within JA, timing activities are evaluated by the intercept a_h. Within a performance model that allows for timing activities, their value is determined according to the covariance between the market excess return and its transformation that defines the function form of timing activities. Thus, this difference can be decomposed into two components: The first component is the fund-specific extent of timing activities, γ_i, and endogenous. The second component is represented by the term in square brackets and depends on the assumed characteristics of timing activities in the model used—that is, the specification of $f(er_{mt})$—and the realizations of er_{mt}. This component is exogenous and identical for all funds in a given period and a given $f(er_{mt})$. We define this difference to be the timing-induced bias in JA.

It is important to note that Equation (14) holds without error in the population and within any sample. The timing-induced bias in JA grows linearly with the extent of measured market-timing activities, γ_i. Furthermore, if no timing activities are present, JA is identical with the total performance tot_i. Finally, because the bias is the product of these two components, its magnitude is strongly determined by their dependence.[14]

3.2 | Stochastic simulation

We depict the distributional properties of the timing-induced bias in JA using stochastic and bootstrap simulations. In the stochastic simulation, we initially draw 1,000 realizations of the excess returns of the market, er_{mt}, from an independent and identical normal distribution (i.i.n.d.) with an expected value of 0.025% and a standard deviation 1%.[15] With these simulated data, we generate a portfolio $er_{pt}^{TM}(er_{pt}^{HM})$ that replicates timing activities according to the TM (HM) model as follows:

$$er_{pt}^{TM} = \delta_p^{TM} + \zeta_p^{TM} er_{mt} + \gamma_p^{TM} er_{mt}^2 + \eta_{pt}^{TM} \tag{15}$$

$$er_{pt}^{HM} = \delta_p^{HM} + \zeta_p^{HM} er_{mt} + \gamma_p^{HM} \max(0, er_{mt}) + \eta_{pt}^{HM} \tag{16}$$

The simulated residuals η_{pt}^{TM} and η_{pt}^{HM} are i.i.n.d. with an expected value of 0% and a standard deviation $\sigma_{\eta_p}^{TM} = \sigma_{\eta_p}^{HM} = 0.41\%$.[16] The selection performance is set to $\delta_p^{TM} = \delta_p^{HM} = 0$, and the market risk exposure is set to $\zeta_p^{TM} = \zeta_p^{HM} = 1$.[17] Parameters γ_p^{TM} and γ_p^{HM} are set to achieve an expected timing performance of 2.5% per annum (p.a.). This level corresponds to the median estimated timing performance of all funds with statistically significant positive market-timing skill in our empirical analysis.[18] All parameters of the simulation are subject to sampling error, with the only exception being the in-sample-mean of the simulated residuals: To exclude the impact of luck as defined by Fama and French (2010), we restrict the mean residuals to be 0.

We refer to the portfolio in Equation (15) as the "TM portfolio" and to the portfolio in Equation (16) as the "HM portfolio." We estimate JA and the TM (HM) model for the TM (HM) portfolio to measure JA, α_p, and the TM (HM) total performance, tot_p^{TM} (tot_p^{HM}). Finally, we calculate the timing-induced bias in JA as their difference, that is, $\alpha_i - tot_p^{TM}$ ($\alpha_i - tot_p^{HM}$). We replicate this process 1,000 times.

We report summary statistics over all replications in Table 1a, the results of one-sample (α_p and tot_p) and paired two-sample ($a_p - tot_p$) t tests in Table 1b, rank correlations in Table 1c, and p-values of hypothesis tests in Table 1d. For the paired t tests, we calculate standard errors of the differences in total performance as

$$\mathrm{SE}(\alpha_p - tot_p) = \sqrt{\mathrm{SE}(\alpha_p)^2 + \mathrm{SE}(tot_p)^2 - 2\mathrm{SE}(\alpha_p)\mathrm{SE}(tot_p)\mathrm{Corr}(\varepsilon_{pt}, \eta_{pt})} \tag{17}$$

We include the correlation between JA residuals and TM/HM residuals, $\mathrm{Corr}(\varepsilon_{pt}, \eta_{pt})$, to account for the fact that the estimation error may be strongly correlated across the performance models. As this correlation is in fact highly positive in our simulation, standard errors calculated as above tend to be lower and therefore increase the criticality of the t test.

For both portfolios, the chosen parameters add economic value of 2.5% p.a. for investors on average, as indicated by the mean total performance tot_p^{TM} and tot_p^{HM} and the results of the t test in the first line of Table 1d. The TM total performance and the HM total performance each scatter with a standard deviation of 16 basis points (bp) p.a. For both portfolios, the total performance is always statistically insignificant at the 5% level, as Table 1b shows. The same applies to JA.

In comparison to the TM (HM) total performance, JA of the TM (HM) portfolio exhibits a similar distribution in terms of location and spread: The mean JA do not differ significantly from 2.5% p.a., as Table 1d shows, and the higher moments are fairly alike. On average, JA is marginally higher than the TM (HM) total performance, but the mean bias $\alpha_p - tot_p^{TM}$ ($\alpha_p - tot_p^{HM}$) of 0.23 bp p.a. (0.59 bp p.a.) is insignificant.

However, the timing-induced bias scatters slightly less than the total performance does. For the TM portfolio, 90% of the bias in JA falls between ± 24 bp p.a. For the HM portfolio, this interval is smaller from −18 bp p.a. to +20 bp p.a. Relative to the simulated total performance of 2.5% p.a., these intervals indicate a bias in JA of less than 10% of the expected total performance. Furthermore, the timing-induced bias in JA is hardly ever statistically significant at the 5% level according to paired two-sample t tests as reported in Table 1b. Finally, the high correlation in Table 1c indicates that rankings of the simulated portfolios according to JA and according to their true total performance are similar.

Our stochastic simulation shows that JA is unbiased on average if a portfolio exhibits timing activities as imposed by the TM or HM model. With a simulated timing performance of 2.5% p.a., this even applies for an economically high level of timing skill. Any existing bias tends to be small—respectively small enough to be economically irrelevant in the most cases—and statistically insignificant. Thus, JA can be expected to correctly capture total performance in our stochastic simulation.

TABLE 1 Descriptive statistics of the timing-induced bias in JA in stochastic simulations

	TM portfolio			**HM portfolio**		
	α_p	tot_p^{TM}	$\alpha_p - tot_p^{TM}$	α_p	tot_p^{HM}	$\alpha_p - tot_p^{HM}$
(a) Descriptive statistics						
Mean	2.4946	2.4923	0.0023	2.4984	2.4924	0.0059
Minimum	1.8476	1.8110	−0.8688	1.9336	1.9142	−0.5866
5th percentile	2.2185	2.2203	−0.2377	2.2748	2.2298	−0.1761
10th percentile	2.2887	2.3076	−0.1616	2.3471	2.3049	−0.1114
25th percentile	2.3988	2.4113	−0.0596	2.4273	2.4113	−0.0346
Median	2.4913	2.4943	0.0010	2.4952	2.4954	0.0008
75th percentile	2.6054	2.5738	0.0647	2.5727	2.5706	0.0524
90th percentile	2.7049	2.6887	0.1642	2.6552	2.6909	0.1348
95th percentile	2.7861	2.7502	0.2423	2.7317	2.7641	0.2033
Maximum	3.1315	3.3287	0.8882	3.2073	3.1672	0.6398
Standard deviation	0.1722	0.1635	0.1504	0.1420	0.1615	0.1138
Skewness	−0.0865	0.0382	−0.0952	−0.0315	0.0146	−0.1535
Kurtosis	3.9788	5.1588	7.6668	5.5603	4.8155	7.0371
(b) Proportions of statistically significant outcomes at the 5% level						
Proportion	0.00	0.00	4.80	0.00	0.00	3.40
(c) Rank correlations						
tot_p	52.07***	100		61.68***	100	
$a_p - tot_p$	54.65***	−30.57*	100	23.62***	−50.69***	100
(d) p-values of hypothesis tests						
H_0: mean 2.5% p.a.	0.0032	0.0014		0.0071	0.0014	
H_0: mean 0			0.0063			0.0010
H_0: identical variance	0.1040			0.0000		

Notes: Table 1 presents descriptive statistics (a), proportions of significant results (b), rank correlations (c), and p-values of hypothesis tests (d) of JA, the true total performance, and the timing-induced bias in JA for two portfolios that replicate timing activities in a stochastic simulation. For the simulation, we draw 1,000 realizations of the market excess return from an independent and identical normal distribution. With these simulated returns, we generate a portfolio that replicates timing activities according to the TM (HM) model, the TM (HM) portfolio. We estimate Equations (1) and (3) (Equation 4) for the TM (HM) portfolio and measure JA, α_p, the TM (HM) total performance, tot_p^{TM} (tot_i^{HM}), and the timing-induced bias in JA as their difference, $\alpha_p - tot_p^{TM}$ ($\alpha_p - tot_p^{HM}$). In (a) we report descriptive statistics for these measures over 1,000 iterations of the simulation. With the exception of skewness and kurtosis, all values in (a) are denoted in percent per annum. In (b), we report the proportion of statistically significant outcomes at the 5%-level using heteroskedasticity and autocorrelation robust standard errors of Newey and West (1987). In (c), we report rank correlations between JA, the TM (HM) total performance, and their difference. Superscripts ***, **, and * indicate statistical significance at the 1%, the 5%, and the 10% level. All values in (b) and (c) are reported in percent.

3.3 | Bootstrap simulation

Next, we use a bootstrap simulation procedure to analyze how the empirical distribution of market excess returns affects the timing-induced bias in JA. The underlying empirical data for our bootstrap simulation are 23,889 daily observations of the CRSP value-weighted index of all US stocks from July 1926 to December 2016. From these data, we bootstrap a sample of 1,000 realizations, which we use to generate the TM (HM) portfolio with the same data-generating process and parameters as for the stochastic simulation.[19]

The further procedure also follows the stochastic simulation: We estimate JA and the TM (HM) model for the TM portfolio (HM portfolio) to measure JA, α_i, and the TM (HM) total performance, tot_i^{TM} (tot_i^{HM}). We then calculate the timing-induced bias in JA as their difference.

We report summary statistics of 1,000 replications of this process in Table 2a, the results of one-sample (α_p and tot_p) and paired two-sample t tests in Table 2b, rank correlations in Table 2c, and p-values of hypothesis tests in Table 2d.

TABLE 2 Descriptive statistics of the timing-induced bias in JA in bootstrap simulations

	TM portfolio			**HM portfolio**		
	α_p	tot_p^{TM}	$\alpha_p - tot_p^{TM}$	α_p	tot_p^{HM}	$\alpha_p - tot_p^{HM}$
(a) Descriptive statistics						
Mean	2.4667	2.5027	−0.0360	2.5109	2.5070	0.0038
Minimum	1.3181	1.4484	−0.9467	1.5902	1.5840	−0.1676
5th percentile	1.9142	1.9664	−0.2713	2.2468	2.2380	−0.0321
10th percentile	2.0480	2.0803	−0.1460	2.3118	2.3004	−0.0176
25th percentile	2.2252	2.2409	−0.0573	2.4165	2.4136	−0.0043
Median	2.4240	2.4572	−0.0077	2.5122	2.5084	0.0015
75th percentile	2.6777	2.7410	0.0116	2.6096	2.6059	0.0134
90th percentile	2.9417	2.9919	0.0537	2.6981	2.6955	0.0318
95th percentile	3.1690	3.2056	0.1005	2.7728	2.7778	0.0443
Maximum	3.8733	4.0214	0.6451	3.0525	3.0613	0.1749
Standard deviation	0.3678	0.3704	0.1316	0.1619	0.1650	0.0266
Skewness	0.5866	0.5395	−2.1230	−0.2400	−0.1823	−0.3217
Kurtosis	3.6238	3.3900	16.0276	4.5930	4.5015	12.2784
(b) Proportions of statistically significant outcomes at the 5% level						
Proportion	0.00	0.00	0.00	0.00	0.00	0.00
(c) Rank correlations						
tot_p	94.60***	100		98.60***	100	
$a_p - tot_p$	10.71***	−13.45***	100	−3.13	−16.01***	100
(d) p-values of hypothesis tests						
H_0: mean 2.5% p.a.	0.0000	0.8170		0.0338	0.1777	
H_0: mean 0			0.0000			0.0000
H_0: identical variance	0.8238			0.5566		

Notes: Table 2 presents descriptive statistics (a), proportions of significant results (b), rank correlations (c), and p-values of hypothesis tests (d) of JA, the total performance, and the timing-induced bias in JA for two portfolios that replicate timing activities in a bootstrap simulation. For the bootstrap simulation, we draw 1,000 realizations from the daily excess returns of the CRSP value-weighted index of all US stocks from July 1926 to December 2016. With these simulated returns, we generate a portfolio that replicates timing activities according to the TM (HM) model, the TM (HM) portfolio. We estimate Equations (1) and (3) (Equation 4) for the TM (HM) portfolio and measure JA, α_p, the TM (HM) total performance, tot_p^{TM} (tot_i^{HM}), and the timing-induced bias in JA as their difference, $\alpha_p - tot_p^{TM}$ ($\alpha_p - tot_p^{HM}$). In (a), we report descriptive statistics for these measures over 1,000 iterations of the simulation. With the exception of skewness and kurtosis, all values in (a) are denoted in percent per annum. In (b), we report the proportion of statistically significant outcomes at the 5% level using heteroskedasticity and autocorrelation robust standard errors of Newey and West (1987). In (c), we report rank correlations between JA, the TM (HM) total performance, and their difference. Superscripts ***, **, and * indicate statistical significance at the 1%, the 5%, and the 10% level. All values in (b) and (c) are reported in percent.

For both portfolios, the mean true total performance differs insignificantly from 2.5% p.a. The TM total performance scatters between 1.45% p.a. and 4.02% p.a., with a standard deviation of 0.37% p.a. The HM total performance is closer distributed, with a range between 1.58% p.a. and 3.06% p.a. and a standard deviation of 0.17% p.a. In comparison to the stochastic simulation, the distribution of the TM total performance spreads more, is more right-skewed, and exhibits a lower skewness. The HM total performance differs mostly in skewness across both simulations, while the other moments appear fairly similar. The differences in the cross-sectional distributions of the true total performance arise as a result of the non-normal distribution of the bootstrapped market excess returns.

For the TM portfolio, JA is slightly lower than 2.5% p.a. on average. Its standard deviation does not differ significantly from that of the TM total performance. Furthermore, both cross-sectional distributions are alike, as indicated by the higher moments. Table 2b shows that JA and TM total performance are insignificant in all iterations. The mean timing-induced bias in JA is about −4 bp p.a. and significant. Although its kurtosis of 16 suggests distinctive fat tails, the bias in JA remains between −27 bp p.a. and +10 bp p.a. with 90% probability and is always insignificant. Rank correlations imply a factual identity between JA and TM total performance and a negative relation between TM total performance and the timing-induced bias. These findings are consistent with the fact that TM total performance is strongly affected by outliers in the realized market excess returns: Extreme realizations of er_{mt} lead to extreme market exposures due to the quadratic relation between portfolio and market excess returns. They also strongly bias JA, as this is a linear model.

For the HM portfolio, mean JA is slightly, but significantly, above 2.5% p.a. Again, the cross-sectional distributions of JA and HM total performance are very similar, indicating a small timing-induced bias. With 0.4 bp p.a., the mean bias in JA is small, yet significant at the 5% level. It deviates only by about 3 bp p.a., and in 90% of the simulations, the bias is between −3 bp p.a. and +4 bp p.a. Rankings according to JA and HM total performance are also factually identical, as shown by the rank correlation. The timing-induced bias in JA is always insignificant.

Our simulations show that the nature of timing activities and the distribution of the market excess returns affect the timing-induced bias in JA: If portfolio managers follow the stylized timing activities of the TM model, JA is likely to be slightly negatively biased. For empirical distributions with fatter tails, these timing activities may cause an even larger negative bias in JA. If portfolio managers implement timing activities as imposed by the HM model, the timing-induced bias in JA is very small and nonexistent on average. In both cases, a statistically significant bias in JA is unlikely to be found.

4 | EMPIRICAL ANALYSIS

4.1 | Performance measurement with multifactor models

In our empirical analysis, the basic performance model is Carhart's (1997) four-factor model, because it is the most common representation of JA. By including the size factor smb_t and the value factor hml_t, we account for commonly known risk factors in addition to systematic market risk (Fama & French, 1993). Moreover, we incorporate the momentum factor mom_t, thereby accounting for the basic investment strategy of buying past winners while selling past losers (see e.g., Jegadeesh & Titman, 1993). Thus, Carhart's alpha, as shown below, is a more accurate measure of the economic value provided by active investment management:

$$er_{it} = \alpha_i + \beta_{1i} er_{mt} + \beta_{2i} smb_t + \beta_{3i} hml_t + \beta_{4i} mom_t + \varepsilon_{it} \tag{18}$$

The corresponding variants of the TM and the HM models are:

$$er_{it} = \delta_i^{TM} + \zeta_{1i}^{TM} er_{mt} + \zeta_{2i}^{TM} smb_t + \zeta_{3i}^{TM} hml_t + \zeta_{4i}^{TM} mom_t + \gamma_i^{TM} er_{mt}^2 + \eta_{it}^{TM} \tag{19}$$

$$er_{it} = \delta_i^{HM} + \zeta_{1i}^{HM} er_{mt} + \zeta_{2i}^{HM} smb_t + \zeta_{3i}^{HM} hml_t + \zeta_{4i}^{HM} mom_t + \gamma_i^{HM} \max(0, er_{mt}) + \eta_{it}^{HM} \tag{20}$$

The definitions of timing performance in Equations (5) and (6) and total performance in Equations (7) and (8) are not affected by this change and are identical for the performance models above. This also applies to the results of our nested model analysis in Section 3.1, as Appendix C demonstrates.

4.2 | Data

Empirical analyses of timing activity using the TM and HM models are commonly based on monthly returns, see Grinblatt and Titman (1994), Cai, Chan, and Yamada (1997), Kryzanowski et al. (1997), Becker, Ferson, Myers, and Schill (1999), Chen and Liang (2007), and Jiang, Yao, and Yu (2007). However, Bollen and Busse (2001) show that more precise inferences on timing activity can be achieved using daily returns. Goetzmann, Ingersoll, and Ivkovic (2000) find that measures of timing activity are negatively biased if the timing interval differs from the return interval used to measure timing activity. Therefore, we use both monthly and daily fund returns in our empirical study.

We use mutual fund data from the CRSP Survivor-Bias-Free US Mutual Fund Database, matching share classes at the portfolio level using CRSP portfolio numbers and fund names.[20] We identify active US domestic equity mutual funds as all funds flagged as nonindex funds that have invested at least 90% of their TNA in common or preferred stocks and which do not have an international equity CRSP objective code. We match the CRSP data with data from Morningstar Direct using ticker symbols and Committee on Uniform Security Identification Procedures (CUSIP) number, following Berk and van Binsbergen (2015) and Pástor, Stambaugh, and Taylor (2015), to obtain daily fund returns before the availability of CRSP data. We require all funds in our analysis to have a complete and reliable daily return history. Thus, we exclude all funds with erroneous or fragmentary returns.[21] Since we estimate performance at the level of individual funds, we also exclude all funds with fewer than 36 monthly returns available. This approach results in a final sample of 1,289 funds, for which we observe 181,994 monthly and 3,806,242 daily returns from January 2, 1987, to December 31, 2014. For each fund, the observed daily and monthly returns cover the exact same period and differ only with respect to the return frequency. We obtain monthly and daily Fama and French (1993) risk factors and Carhart's (1997) momentum factor from Kenneth R. French's data library.[22]

4.3 | Selection and timing performance in the cross section

We first analyze selection and timing performance by estimating Equations (19) and (20) for each fund. We calculate timing performance according to Equations (5) and (6). Our first analysis, which is reported in Table 3, compares TM and HM selection and timing performance for monthly and daily returns: Table 3a reports descriptive statistics, Table 3b the proportions of significant outcomes, Table 3c rank correlations in the cross section, and Table 3d p-values of various hypothesis tests in the cross section.

The mean selection and timing performance are both mostly negative and statistically significant, which is consistent with the majority of the empirical literature on mutual fund performance: On average, fund managers exhibit neither selection nor

TABLE 3 Selection and timing performance

	Monthly returns				Daily returns			
	Selection		Timing		Selection		Timing	
	TM	HM	TM	HM	TM	HM	TM	HM
(a) Descriptive statistics								
Mean	−0.7142	−0.9023	−0.2048	−0.0259	−0.2626	0.0027	−0.5572	−0.7762
Minimum	−13.2254	−17.7686	−9.9392	−21.5521	−20.7082	−25.5892	−9.8458	−21.5330
5th percentile	−4.8409	−6.2869	−2.4403	−4.1408	−4.9555	−6.2311	−3.6303	−6.9673
10th percentile	−3.8104	−4.9320	−1.7700	−2.8901	−3.5067	−4.7042	−2.7882	−4.9410
25th percentile	−2.0461	−2.5852	−0.9315	−1.3544	−1.8219	−2.3642	−1.4585	−2.4987
Median	−0.6319	−0.7173	−0.1539	−0.0019	−0.3228	−0.1520	−0.3836	−0.4771
75th percentile	0.7876	0.8783	0.5619	1.4473	1.2945	2.1989	0.5033	1.1622
90th percentile	2.2436	2.8043	1.3328	2.9080	3.1573	5.1742	1.4655	3.0019
95th percentile	3.3197	4.0688	1.9750	3.9574	4.4318	6.7138	2.0910	4.2479
Maximum	9.6936	18.3489	5.6016	10.5765	9.8944	16.9643	8.5323	13.6786
Standard deviation	2.6381	3.3407	1.3987	2.5829	2.8327	4.1220	1.8553	3.4836
Skewness	−0.5578	−0.2097	−0.5476	−0.6719	−0.2778	0.0345	−0.4546	−0.6229
Kurtosis	5.4038	5.9513	6.9374	8.4350	5.9865	5.2008	5.6396	6.3444
(b) Proportions of statistically significant outcomes at the 5% level								
Proportion	12.7	10.1	12.5	7.0	18.1	21.0	16.4	18.8
(c) Rank correlations								
Selection HM	88.32***	100			89.73***	100		
Timing TM	−48.93***		100		−58.06***		100	
Timing HM		−70.60***	88.93***	100		−81.12***	94.81***	100
(d) p-values of hypothesis tests								
H_0: mean 0	0.0000	0.0000	0.0000	0.7194	0.0009	0.9812	0.0000	0.0000
H_0: identical mean	0.0000		0.0000		0.0000		0.0000	
H_0: identical variance	0.0000		0.0000		0.0000		0.0000	

Notes: Table 3 presents descriptive statistics (a), proportions of significant results (b), rank correlations (c), and p-values of hypothesis tests (d) of the selection and timing performance measured by the TM and HM models for our full sample of 1,289 funds between January 1987 and December 2014. We estimate Equations (19) and (20) for each fund individually. We calculate timing performance according to Equations (5) and (6). With the exceptions of skewness and kurtosis, all values in (a) are denoted in percent per annum. In (b), we report the proportion of statistically significant selection and timing performance at the 5% level using heteroskedasticity and autocorrelation robust standard errors of Newey and West (1987). In (c), we report rank correlations between measured selection and timing performance. Superscripts ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. All values in (b) and (c) are reported in percent.

timing skill. The only exceptions are the mean HM-timing performance using monthly returns and the HM selection performance using daily returns, which are both insignificant. For monthly returns, roughly 11% of the funds yield statistically significant selection or timing performance using either model. For daily returns, this proportion is slightly higher, as the greater number of daily observation results reduces estimation error. Again, these findings are consistent with the existing literature.

We identify three characteristics in measured selection and timing performance that are relevant for our study: First, the mean selection performance and timing performance differ between the TM and HM models by about 18 bp p.a. for monthly returns and by about 25 bp p.a. for daily returns. In two-sample t tests, these differences are statistically significant with p-values near 0, as shown in Table 3d. The results of the TM model also differ from those of the HM model with respect to dispersion over the cross section: HM selection and timing performance always have a higher standard deviation than TM selection and timing performance. Two-sample F tests as reported in Table 3d show that these differences in standard deviation are statistically significant, with p-values very close to 0. Therefore, the HM model discloses more extreme and more widespread selection and timing performance than the TM model.

Second, the return frequency is a major issue in measured selection and timing performance. For monthly returns, the mean and median selection (timing) performances are much lower (higher) than for daily returns, irrespective of the model. This underlines the importance of the return frequency when measuring timing activities.

Third, Spearman correlations in Table 3c indicate that both models rank funds similarly in terms of selection and timing performance: Correlation coefficients between the TM model and the HM model are around 90%, irrespective of whether we consider selection performance or timing performance. This finding applies to monthly and daily returns. The significant negative rank correlations between selection and timing performance within each of the models allow for two interpretations: Either funds engage in artificial timing activities, in the sense of Jagannathan and Korajczyk (1986), or fund managers vary in their skill, as for Kacperczyk et al. (2014). In both cases, the unconditional models used here are unlikely to produce unbiased measures of selection and timing performance. In total, our results indicate that inferences on selection and timing skills are unlikely to be correct if the nature of the timing activities to be evaluated is unknown: The choice between the TM model and the HM model strongly affects the cross-sectional distribution of measured skill in our sample.

4.4 | Total performance in the cross section

Table 4 compares total performance in the cross section of our mutual funds sample. For each fund, we measure JA as the intercept α_i of Equation (18). For the TM (HM) model, we estimate Equation (19) (Equation 20) and then calculate TM (HM) total performance as in Equation (7) (Equation 8). Table 4a reports descriptive statistics, Table 4b the proportions of significant outcomes, Table 4c rank correlations in the cross section, and Table 4d p-values of various hypothesis tests in the cross section.

The mean total performance is negative, irrespective of the model or return frequency. All means are statistically significant with a very high confidence. This finding is consistent with the general perception in the empirical literature that active management does not add value in excess of a passive benchmark return. For monthly returns, Table 4b shows a significant total performance for about 15% of the funds in our sample; for daily returns, this proportion is marginally higher.

The cross-sectional locations of total performance are virtually identical for all three models: The means do not differ by more than 5 bp, and all differences between the means are insignificant, as the p-values in Table 4d show. The same applies to the standard deviation, which does not differ significantly between the three models. Finally, the percentiles and higher moments of the cross-sectional distributions are very similar for all three models. These finding hold for monthly as well as for daily returns.

Table 4c indicates that the similarity of the cross-sectional distributions arises from very uniform measurements of total performance for each single fund: The relative evaluation of the funds' total performance in our sample is factually identical for all three models, as the high rank correlations show. To illustrate this finding visually, Figure 2 shows pairwise scatterplots of total performance with fitted lines for both monthly and daily return frequencies.

The plots confirm a very strong, linear relation between JA and TM as well as HM total performance, irrespective of the return interval. Only a few funds visibly deviate from the plotted fitted lines. Moreover, the fitted lines suggest that measured performance is factually identical at the level of individual funds. Table 5 strengthens this particular finding by testing the pairwise relations of total performance in cross-sectional regressions: In Table 5a, we estimate the cross-sectional regressions using the results of all funds in our sample. In Table 5b, we restrict the sample to those results that exhibit significant timing activities.[23]

The results clearly document that, for all three models, measured total performance is identical at the fund level: The R^2 is practically 100% in all cross-sectional regressions in Table 5a.[24] For the total sample in Table 5a, the joint

TABLE 4 Total performance

	Monthly returns			Daily returns		
	JA	**TM**	**HM**	**JA**	**TM**	**HM**
(a) Descriptive statistics						
Mean	−0.9315	−0.9190	−0.9282	−0.8227	−0.8198	−0.7735
Minimum	−12.8077	−12.7877	−12.742	−15.8031	−15.8615	−16.5651
5th percentile	−4.6718	−4.6903	−4.7047	−4.3941	−4.3702	−4.3898
10th percentile	−3.4821	−3.4308	−3.4219	−3.3100	−3.3082	−3.3165
25th percentile	−1.9842	−1.9871	−2.0101	−1.9450	−1.9560	−1.9272
Median	−0.8040	−0.7952	−0.8133	−0.7449	−0.7361	−0.6958
75th percentile	0.3601	0.3645	0.3511	0.4489	0.4461	0.4810
90th percentile	1.6626	1.7040	1.6603	1.8173	1.8130	1.9109
95th percentile	2.4164	2.4147	2.3734	2.6959	2.7105	2.7940
Maximum	8.9458	8.9130	8.7583	8.3304	8.2944	8.2914
Standard deviation	2.2858	2.2901	2.2767	2.2747	2.2746	2.3060
Skewness	−0.7998	−0.7719	−0.7914	−0.6876	−0.6905	−0.7281
Kurtosis	6.2356	6.0960	6.1914	6.8124	6.8343	7.2324
(b) Proportions of statistically significant outcomes at the 5% level						
Proportion	14.8	15.5	14.9	15.9	16.1	17.1
(c) Rank correlations						
TM	99.74***	100		99.99***	100	
HM	99.65***	99.74***	100	98.59***	98.67***	100
(d) *p*-values of hypothesis tests						
H_0: mean 0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
H_0: identical mean						
With TM	0.8894	1		0.9834	1	
With HM	0.9704	0.9185	1	0.5856	0.6073	1
H_0: identical variance						
With TM	0.9459	1		0.9986	1	
With HM	0.8867	0.8334	1	0.6245	0.6233	1

Notes: Table 4 presents descriptive statistics (a), proportions of significant results (b), rank correlations (c), and *p*-values of hypothesis tests (d) of the total performance measured by JA, the TM model, and the HM model for our full sample of 1,289 funds between January 1987 and December 2014. We estimate Equations (18) to (20) for each fund individually. For the TM and HM models, we calculate total performance according to Equations (7) and (8). With the exceptions of skewness and kurtosis, all values in (a) are denoted in percent per annum. In (b), we report the proportion of statistically significant total performance at the 5% level using heteroskedasticity and autocorrelation robust standard errors of Newey and West (1987). In (c), we report cross-sectional rank correlations between measured total performance. Superscripts ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. All values in (b) and (c) are reported in percent.

hypothesis of pairwise identical total performance is statistically rejected at the 5% level in most of the cases, which is not surprising, given the 1,289 funds in the sample. However, the estimated coefficients indicate that any difference in measured total performance is negligibly small: For monthly returns, shifts in total performance are less than 2 bp p.a. in absolute value, while dilation is at most 1 bp p.a. per percent p.a. of total performance. For daily returns, the constants indicate a shift of HM total performance by about 5 bp p.a. against JA and TM total performance, while dilation is always insignificant.

In Table 5b, we restrict the sample to funds that exhibit significant timing activities. For monthly returns, the joint hypothesis of identical total performance cannot be rejected at the 5% level. For daily returns, JA and TM total performance are factually identical except for a shift of 1 bp p.a., while HM total performance is on average 16 bp p.a. higher than JA and TM total performance after consideration of a light dilation. In total, these results confirm that JA, the TM model, and the HM model disclose factually identical total performance, even at the fund level. They also confirm that there exists no economically relevant bias in JA resulting from timing activities in our sample on average.[25]

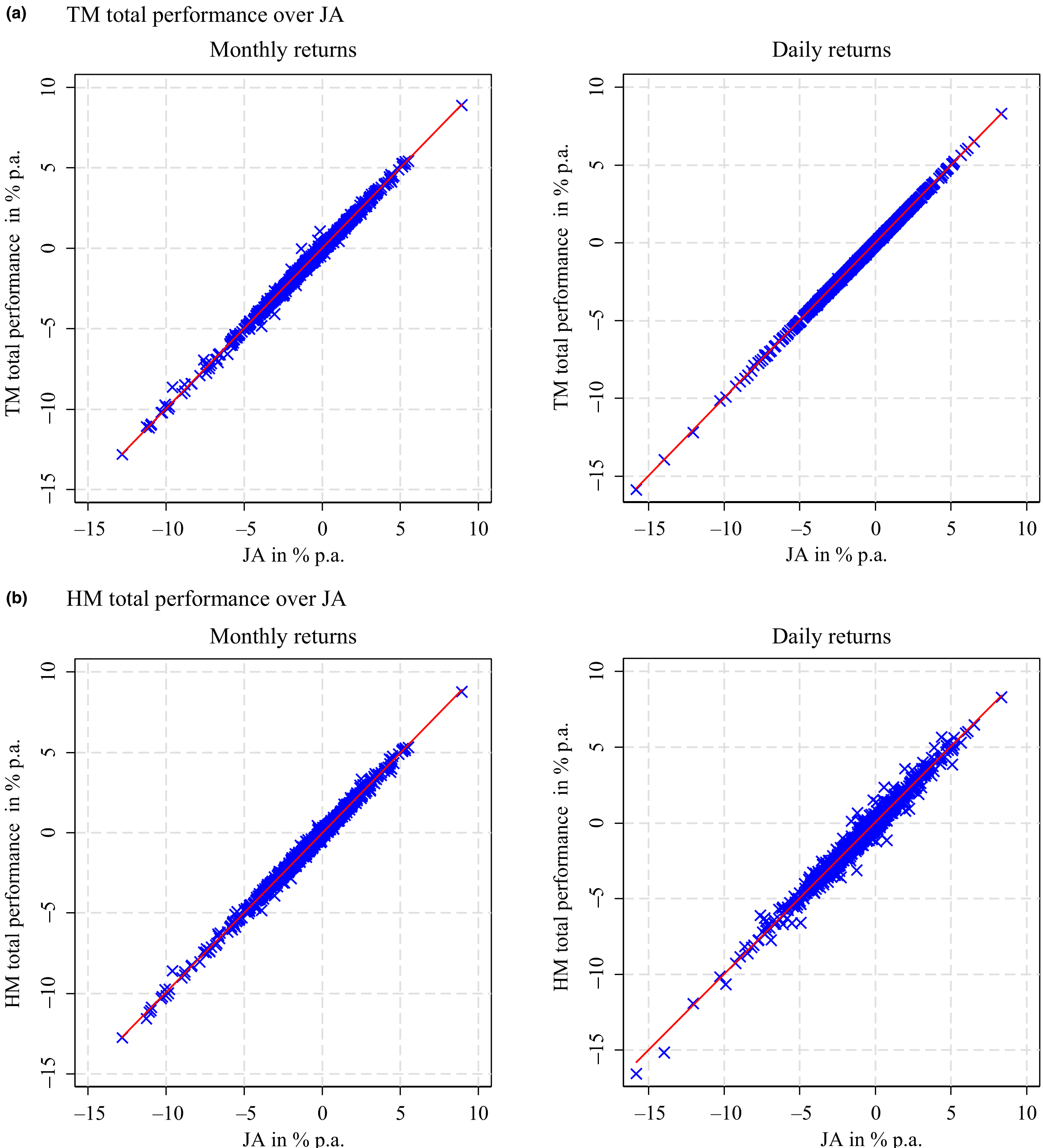

FIGURE 2 Scatter plots of measured total performance. The scatterplots show the measured total performance and the respective fitted lines for our sample of 1,289 funds. (a) The scatterplots show TM total performance over JA. (b) The scatterplots show HM total performance over JA. For the results of the cross-sectional regressions, see Table 5. All axis labels are in percent per annum [Colour figure can be viewed at wileyonlinelibrary.com]

4.5 | The empirical bias in JA

To further analyze why there exists no economically relevant bias in JA in our empirical analysis, we apply a rolling window regression analysis: Consistent with our previous analysis, we match rolling windows for monthly fund returns to those of daily fund returns and estimate all models for individual funds using rolling windows of 36 months (756 days) and a step size of 12 months (252 days) for monthly (daily) fund returns. This procedure results in a total number of

TABLE 5 Cross-sectional regressions of total performance

	Monthly returns					Daily returns				
	N	Const. b_0	Slope b_1	*F*-test	R^2	*N*	Const. b_0	Slope b_1	*F*-test	R^2
(a) All funds										
JA explains TM	1,289	0.0128***	1.0003	0.14	99.67	1,289	0.0027***	0.9999	0.03	99.99
JA explains HM	1,289	−0.0025	0.9937***	2.65	99.54	1,289	0.0519***	1.0033	0.00	97.95
TM explains HM	1,289	−0.0161***	0.9925***	0.00	99.67	1,289	0.0498***	1.0041	0.00	98.10
(b) Funds with significant timing activities only										
JA explains TM	160	0.0530**	1.0088	5.28	98.79	211	0.0098***	0.9988	0.06	99.97
JA explains HM	90	0.0597*	1.0106	25.25	98.56	242	0.1564***	1.0228	0.03	94.76
TM explains HM	185	0.0109	0.9898	7.07	99.11	277	0.1555***	1.0284*	0.00	95.43

Notes: Table 5 presents the results of cross-sectional regressions of total performance for the sample of 1,289 funds. We estimate linear regressions as in $Y_i = b_0 + b_1 X_i$ to quantify the relations between total performance measures as indicated in the first column at the level of individual funds. Standard errors are heteroskedasticity consistent, as per White (1980). Superscripts ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. In the case of the slope coefficient, the null hypothesis is $b_1 = 1$. The column *F*-test shows the *p*-value of a test of the joint null hypothesis $b_0 = 0$ and $b_1 = 1$. All constants are denoted in percent per annum, and columns *F*-test and *R*2 are denoted in percent.

11,547 estimations for each return frequeny. In each window, we measure JA as the intercept α_i of Equation (18). For the TM (HM) model, we estimate Equation (19) (Equation 20) and then calculate TM (HM) total performance as in Equation (7) (Equation 8).

In Table 6, we present descriptive statistics (Table 6a), proportions of significant outcomes (Table 6b), rank correlations (Table 6c), and *p*-values of hypothesis tests (Table 6d) of the empirical differences between JA and TM as well as HM total performance.

The mean differences between JA and the other two measures of total performance are small, positive, and statistically significant at the 5% level, as the *p*-values in the first line of Table 6d show. This indicates that JA slightly overstates the total performance by 1–7 bp in comparison with the TM and HM models. For monthly returns, the cross-sectional distribution of both differences is similar, as indicated by the percentiles, the standard deviation and the higher moments, although the differences in means and standard deviations are significant. For daily returns, the difference between JA and HM total performance scatters much stronger than the difference between JA and TM total performance, leading to a much wider distribution with high extreme values. For both return intervals, rank correlations are around 50% and significant, implying that JA tends to over- or understate TM and HM total performance simultaneously.

On the level of rolling windows and funds, differences between JA and the TM and HM total performance are mostly insignificant. This especially applies to monthly returns, for which the window size of 36 months may affect statistical significance.[26] For daily returns, however, we also find no statistically significant differences between JA and TM total performance, although each window includes 756 observations. In contrast, 57% of our fund-window estimations exhibit significant differences between JA and HM total performance.

In our next analyses, we demonstrate that this effect is driven exogenously, namely through the market component of the differences. For this purpose, we analyze the components of the bias, that is, the extent of market-timing activities and the exogenous market component. The distribution of these components has a major impact on the bias in JA. However, as both components linearly scale the bias, their correlation is just as important. In Table 7, we analyze the measured timing activities and the exogenous market climate of the bias according to Equation (14).

For monthly returns, the mean timing skills according to both the TM and the HM model, γ_i^{TM} and γ_i^{HM}, are weakly negative and significant. The mean market component with respect to the TM (HM) model is negative (positive) and significant. The distributions of both components are located around 0, which partly explains why we find little average differences between JA and TM as well as HM total performance for monthly returns.

For daily returns, the mean TM- and HM-timing activities are slightly positive and significant. The mean TM market component is also small, while the mean HM market component is much higher, about 7.7% p.a. The TM market component is closely distributed around 0. In contrast, the distribution of mc^{HM} is widely spread and not centered around 0, which explains why the differences between JA and HM total performance for single fund-windows are often significant at the 5% level for daily returns. We stress that these significant differences are driven by the exogenous market component, and not by extensive market-timing skill of fund managers.

TABLE 6 Differences in total performance

	Monthly returns		Daily returns	
	$\alpha_i - tot_i^{TM}$	$\alpha_i - tot_i^{HM}$	$\alpha_i - tot_i^{TM}$	$\alpha_i - tot_i^{HM}$
(a) Descriptive statistics				
Mean	0.0521	0.0096	0.0105	0.0699
Minimum	−3.6222	−3.7327	−0.4895	−3.8109
5th percentile	−0.5372	−0.6231	−0.0774	−0.8418
10th percentile	−0.2929	−0.3826	−0.0413	−0.5339
25th percentile	−0.0826	−0.1429	−0.0096	−0.1699
Median	0.0118	−0.0044	0.0031	0.0315
75th percentile	0.1627	0.1277	0.0255	0.3038
90th percentile	0.4820	0.4120	0.0707	0.7116
95th percentile	0.7740	0.7055	0.1235	1.0557
Maximum	3.2713	3.8134	0.4895	3.9073
Standard deviation	0.4360	0.4725	0.0691	0.6090
Skewness	0.2944	0.5909	0.7850	0.2780
Kurtosis	12.3590	14.9772	13.0725	8.7111
(b) Proportions of statistically significant outcomes at the 5% level				
Proportion	1.97	2.83	0.00	56.98
(c) Rank correlations				
$\alpha_i - tot_i^{HM}$	51.24***	100	44.73***	100
(d) p-values of hypothesis tests				
H_0: mean 0	0.0000	0.0286	0.0000	0.0000
H_0: identical mean	0.0000		0.0000	
H_0: identical variance	0.0000		0.0000	

Notes: Table 6 presents descriptive statistics (a), proportions of significant results (b), rank correlations (c), and p-values of hypothesis tests (d) of the difference between JA and the TM total performance and the difference between JA and the HM total performance. We estimate Equations (18)–(20) for each fund individually using rolling windows of 36 months (756 days) and a step size of 12 months (252 days) for monthly returns (daily returns), resulting in 11,547 fund-window estimations for both return frequencies. For the TM and HM models, we calculate total performance according to Equations (7) and (8), respectively. With the exceptions of skewness and kurtosis, all values in (a) are denoted in percent per annum. In (b), we report the proportion of statistically significant differences in total performance at the 5% level using heteroskedasticity and autocorrelation robust standard errors of Newey and West (1987). In (c), we report cross-sectional rank correlations between differences in total performance. Superscripts ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. All values in (b) and (c) are reported in percent.

Table 7b shows another important reason why we generally find little significant timing-induced bias in JA: The correlation between timing activities γ and the corresponding market climate mc is generally small. Both components, however, multiply with each other in the timing-induced bias of JA. As they are mostly uncorrelated and their distributions are centered on 0, both components tend to cancel each other out and the total bias also becomes small. Economically speaking, fund managers do not vary the extent of their timing activities along with the market climate in which these activities might positively affect measured JA.

5 | CONCLUSION

Does market timing matter in evaluating mutual fund performance? This question is at the center of a long-time debate in the theoretical and empirical literature on investment performance measurement. From a theoretical point of view, market timing is highly relevant: It constitutes a nonlinear relationship between fund returns and market excess returns, which creates a bias in JA. Empirical studies, however, have failed to find evidence that such a bias actually exists.

TABLE 7 Components of the differences in total performance

	Monthly returns				Daily returns			
	$\alpha_i - tot_i^{TM}$		$\alpha_i - tot_i^{HM}$		$\alpha_i - tot_i^{TM}$		$\alpha_i - tot_i^{HM}$	
	γ_i^{TM}	$\boldsymbol{mc^{TM}}$	γ_i^{HM}	$\boldsymbol{mc^{HM}}$	γ_i^{TM}	$\boldsymbol{mc^{TM}}$	γ_i^{HM}	$\boldsymbol{mc^{HM}}$
(a) Descriptive statistics								
Mean	−0.0574	−0.1646	−0.0090	0.5585	0.0515	0.0384	0.0010	7.7497
Minimum	−25.4827	−1.5509	−2.2467	−5.8621	−8.8490	−0.1382	−0.3152	−18.0912
5th percentile	−2.3013	−0.9875	−0.3464	−3.3605	−1.3145	−0.0682	−0.0738	−11.7065
10th percentile	−1.5215	−0.7804	−0.2606	−2.1287	−0.9140	−0.0539	−0.0544	−9.1537
25th percentile	−0.7202	−0.3088	−0.1328	−0.5387	−0.4047	0.0050	−0.0250	−0.4940
Median	−0.1223	−0.1048	−0.0198	0.9422	−0.0002	0.0269	0.0009	8.5857
75th percentile	0.4963	0.0576	0.1038	1.5801	0.4681	0.0548	0.0271	17.8674
90th percentile	1.4225	0.1930	0.2526	2.8324	1.0923	0.1500	0.0564	21.7000
95th percentile	2.4361	0.3989	0.3680	3.6485	1.5944	0.2144	0.0764	23.8039
Maximum	15.8559	1.1648	1.4933	6.1321	10.4676	0.2553	0.2430	28.7215
Standard deviation	1.6371	0.4210	0.2228	2.0301	0.9653	0.0759	0.0476	11.3998
Skewness	0.5810	−0.1999	0.3707	−0.2262	0.3785	0.9443	−0.0187	−0.2844
Kurtosis	16.3781	4.4445	6.3421	3.7147	11.0603	3.8211	5.1595	1.9869
(b) Rank correlations								
mc^{TM}	8.85***	100			18.12	100		
γ_i^{HM}	90.22***		100		93.17***		100	
mc^{HM}		82.76***	1.44	100		−0.79	11.07***	100
(c) p-values of hypothesis tests								
H_0: mean 0	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0211	0.0000

Notes: Table 7 presents descriptive statistics (a), rank correlations (b), and p-values of hypothesis tests (c) of the components of the timing-induced bias in JA according to Equation (14). We estimate Equations (19) and (20) for each fund individually using a rolling window of 36 months (756 days) and a step size of 12 months (252 days) for monthly returns (daily returns), resulting in 11,747 fund-window estimations for both return frequencies. We calculate market components mc^{TM} and mc^{HM} as

$$mc^{TM} = \alpha_h^{TM} - \text{Var}(er_{mt})$$
$$mc^{HM} = \alpha_h^{HM} - [\text{E}(\max(0, er_{mt})) - \text{E}(er_{mt})\,\text{Pr}(er_{mt}{>}0)]$$

where α_h^{TM} (α_h^{HM}) is estimated using the four-factor variant of Equation (9) to explain er_{mt}^2 ($\max(0, er_{mt})$, see Appendix C. With the exceptions of skewness and kurtosis, all values in columns mc^{TM} and mc^{HM} in (a) are denoted in percent per annum. In (b), we report cross-sectional rank correlations between these components. Superscripts ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels. All values in (b) are reported in percent.

We are the first to systematically solve this puzzle by decomposing the timing-induced bias in JA and its multifactor variants into the endogenous extent of timing activities and an exogenous measure of market climate. We find that the magnitude of the two components determines the magnitude of the bias just as much as their correlation. Furthermore, we analyze portfolios that systematically exploit timing information in stochastic and bootstrap simulations. We find that for timing activities as imposed by the TM (HM) model, the timing-induced bias in JA is weakly negative (nonexistent) and hardly ever statistically significant. Finally, we show empirically that JA and measures of total performance according to the TM and HM models are virtually identical for a large sample of US mutual funds. The reason is that empirically, the components of the timing-induced bias in JA are distributed around 0 and nearly uncorrelated. We thus explain comprehensively why empirical studies hardly ever find evidence of this bias, despite its clear theoretical existence. Furthermore, our analysis shows that the separation of total performance into selection and timing performance is sensitive to the assumed model of market timing activities. This causes a critical dual hypothesis problem in empirical research, in which the type of timing activity is usually unknown. Thus, separating total performance into selection and timing performance using popular timing models could lead to false inferences. Overall, we conclude that, for all relevant purposes, JA and its multifactor variants are sufficient measures of mutual fund total performance, even in the presence of timing activity.

ACKNOWLEDGMENTS

The authors thank the anonymous referee for valuable comments, which significantly contributed to improving the quality of the manuscript. We are grateful to Sven Bornemann, Sungju Chun, David Cicero, Philip L. Fazio, Fernando Muñoz Sánchez, Marco Navone, Melissa Porras Prado, and the participants of the 2009 European Finance Association Annual Meeting in Washington D.C., USA; the 2009 Financial Management Association European Conference in Turin, Italy; the 2009 Financial Management Association Annual Meeting in Reno, NV, USA; the Doctoral Consortium of the 2011 European Retail Investment Research Conference in Stuttgart, Germany; and the 2011 Symposium on Finance, Banking, and Insurance in Karlsruhe, Germany, for helpful comments and suggestions on earlier drafts of this manuscript previously titled "Selection, timing and total performance of mutual funds: Wasting time measuring timing". We are responsible for all remaining errors. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ENDNOTES

[1] For a detailed literature overview of recent research using the TM and HM models, see Appendix A.

[2] These values were chosen to resemble the according estimates from the Center for Research in Security Prices (CRSP) value-weighted index of all US stocks from July 1966 to June 2016.

[3] Our results hold for multifactor models if timing activities refer to only one market factor, which is typically assumed, see Appendix B. For studies of simultaneous timing activities with respect to several risk factors see, for example, Grinblatt and Titman (1994), Kryzanowski, Lalancette, and To (1997), Wermers (2000), and Comer (2006).

[4] We interpret δ_i as the average vertical distance between a managed portfolio and a portfolio with the same market-timing activities and assume that any rational investor aims to maximize this distance.

[5] Strictly speaking, this interpretation requires perfect market timing in accordance with the respective model and frictionless access to capital markets, as criticized in section 2.2.1 of Aragon and Ferson (2006). Thus, it implicitly rules out dynamics in market risk exposure that are not related to timing information, such as artificial timing as discussed by Jagannathan and Korajczyk (1986).

[6] Grinblatt and Titman (1989) assume a single market factor. Their approach is also applicable to multifactor models by Fama and French (1993) and Carhart (1997), as shown by Lo (2008). Others evaluate timing activity ex ante using option pricing models, such as Merton (1981) for the HM model or Goetzmann, Ingersoll, Spiegel, and Welch (2007) for the TM model. However, these approaches are sensitive to misspecifications of the option pricing model.

[7] More precisely, tot_i is the total contribution of the fund management's investment activities over a passive strategy with the same mean exposure to market risk. These activities need not to be motivated exclusively by selection or timing skills, but may be of any other nature, such as artificial timing.

[8] For the derivation of tim_i^{TM}, see Grinblatt and Titman (1989). For the derivation of tim_i^{HM}, see Coles et al. (2006) and Appendix B. If timing activities take place in several market factors simultaneously, Equations (4) and (5) ignore their cross terms. This approach follows that of Grinblatt and Titman (1994), Comer (2006), and Chen et al. (2010), who do not consider the cross terms of timed market factors to obtain a feasible model.

[9] For JA, the TM, and the HM models, we always estimate heteroskedasticity and autocorrelation robust standard errors as per Newey and West (1987). For brevity, we ignore any correlation between the sampling error in γ_i and the sampling error in $\mathrm{Cov}[er_{mt}, f(er_{mt})]$.

[10] See Appendix C for details on the derivation in a multifactor model. As our results hold for the population, they also hold within any sample of the population.

[11] This interpretation also applies within the population. For the TM model, we estimate Equation (8) with $y_t = er_{mt}^2$, and for the HM model, we estimate Equation (8) with $y_t = \max(0, er_{mt})$.

[12] We suggest an alternative interpretation of the term $\gamma_i a_h$: For the TM as well as for the HM model, it always holds that $f(er_{mt}) \cdot er_{mt} \geq 0$. The coefficient a_h reflects this nonnegativity. The product $\gamma_i a_h$ accounts for the difference induced by a_h in the intercept δ_i, scaled by the level of nonlinearity, γ_i.

[13] Empirically, researchers are rarely able to rule out ex ante the existence of timing activities in observed fund returns. Accordingly, any empirically measured JA is potentially affected by timing activities.

[14] In the online Appendix, we also apply this approach to conditional timing models. The results are qualitatively similar to those of the unconditional models.

[15] These values closely reflect the corresponding empirical parameters of the daily CRSP value-weighted index of all US stocks from July 1926 to December 2016. We allow for estimation error in the parameters of er_{mt}.

[16] This corresponds to the median residual standard deviation of all funds with significant positive market-timing activities in our empirical analysis in Section 4.

[17] We note that JA model fully captures changes in the parameters δ_p^{TM}, δ_p^{HM}, ζ_p^{TM}, and ζ_p^{HM}. Thus, changes in their values do not qualitatively alter our results.

[18] See Section 4.1 for further information on the dataset. Given the i.i.n.d single-factor market, we use Equation (5) to identify the value of γ_p^{TM} that achieves an expected timing performance of 2.5% p.a. Based on Equation (6), we derive the relation between tim_p^{HM} and the parameters of er_m to identify the corresponding value of γ_p^{HM}.

[19] This includes restricting the in-sample mean residual to be zero. As stated earlier, this is the only restriction applied in the simulation.

[20] Monthly returns are weighted using share class total net assets (TNA). Daily share class returns are equally weighted due to the lack of daily TNA.

[21] In the CRSP database, daily returns are available since 1998. For earlier periods, we use data from the Morningstar database. If a fund's daily returns are available in both databases, we use the CRSP data. We assume the data for funds that report only positive returns, returns above 100%, or returns below −100% to be erroneous. We define as fragmentary funds for which returns are obviously missing or for which more than 5% of returns are reported as 0.

[22] See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[23] We test for significance using two-sided *t* test with a significance level of 95% and standard errors according to Newey and West (1987). In the first (second) line of Table 5b, we consider funds that exhibit significant timing activities according to the TM (HM) model. In the third line, we consider funds that exhibit significant timing activities according to either the TM model, the HM model, or both.

[24] Since virtually all the variability in total performance is explained, the estimation error is extremely small. This explains why hypothesis tests of the constant and the slope coefficient reject the null hypothesis, even if there is no economically relevant difference.

[25] The online Appendix exhibits a similar empirical analysis using conditional timing models. The results are qualitatively the same as those in our main analysis using unconditional models.

[26] Repeating this analysis in a "non-rolling" procedure—that is, using the complete monthly time series for each fund—we find that 0.15% (0.62%) of the differences between JA and TM (HM) total performance are significant at the 5% level. This implies that the corresponding results in our rolling window analysis for monthly returns are not purely driven by the window size of 36 months.

REFERENCES

Admati, A. R., & Ross, S. A. (1985). Measuring investment performance in a rational expectations equilibrium model. *Journal of Business*, *58*(1), 1–26.

Amihud, Y., & Goyenko, R. (2013). Mutual fund's R² as predictor of performance. *Review of Financial Studies*, *26*(3), 667–694.

Aragon, G. O., & Ferson, W. E. (2006). Portfolio performance evaluation. *Foundations and Trends in Finance*, *2*(2), 83–190.

Bali, T. G., Brown, S. J., & Caglayan, M. O. (2014). Macroeconomic risk and hedge fund returns. *Journal of Financial Economics*, *114*(1), 1–19.

Becker, C., Ferson, W. E., Myers, D. H., & Schill, M. J. (1999). Conditional market timing with benchmark investors. *Journal of Financial Economics*, *52*(1), 119–148.

Ben-Rephael, A., Kandel, S., & Wohl, A. (2012). Measuring investor sentiment with mutual fund flows. *Journal of Financial Economics*, *104*(2), 363–382.

Berk, J. B., & van Binsbergen, J. H. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, *118*(1), 1–20.

Blake, D., Rossi, A. G., Timmermann, A., Tonks, I., & Wermers, R. (2013). Decentralized investment management: Evidence from the pension fund industry. *Journal of Finance*, *68*(3), 1133–1178.

Bollen, N. P. B., & Busse, J. A. (2001). On the timing ability of mutual fund managers. *Journal of Finance*, *56*(3), 1075–1094.

Bollen, N. P. B., & Busse, J. A. (2005). Short-term persistence in mutual fund performance. *Review of Financial Studies*, *18*(2), 569–597.

Bollen, N. P. B., & Whaley, R. E. (2009). Hedge fund risk dynamics: Implications for performance appraisal. *Journal of Finance*, *64*(2), 985–1035.

Cai, J., Chan, K. C., & Yamada, T. (1997). The performance of Japanese mutual funds. *Review of Financial Studies*, *10*(2), 237–274.

Cao, C., Chen, Y., Liang, B., & Lo, A. W. (2013). Can hedge funds time market liquidity? *Journal of Financial Economics*, *109*(2), 493–516.

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

Chen, Y., Ferson, W. E., & Peters, H. (2010). Measuring the timing ability and performance of bond mutual funds. *Journal of Financial Economics*, *98*(1), 72–89.

Chen, Y., & Liang, B. (2007). Do market timing hedge funds time the market? *Journal of Financial and Quantitative Analysis*, *42*(4), 827–856.

Christoffersen, S. E. K., & Sarkissian, S. (2009). City size and fund performance. *Journal of Financial Economics*, *92*(2), 252–275.

Coles, J. L., Daniel, N., & Nardari, F. (2006). Do model and benchmark specification error affect inference in measuring mutual fund performance? Working paper. Retrieved from http://papers.ssrn.com/abstract_id=891224

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

Comer, G., Larrymore, N., & Rodriguez, J. (2009). Controlling for fixed-income exposure in portfolio evaluation: Evidence from hybrid mutual funds. *Review of Financial Studies*, *22*(2), 481–507.

Dass, N., Nanda, V., & Wang, Q. (2013). Allocation of decision rights and the investment strategy of mutual funds. *Journal of Financial Economics*, *110*(1), 254–277.

Dybvig, P. H., & Ross, S. A. (1985). Differential information and performance measurement using a security market line. *Journal of Finance*, *40*(2), 383–399.

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

Fama, E. F., & French, K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance*, *65*(5), 1915–1947.

Goetzmann, W. N., Ingersoll, J., & Ivkovic, Z. (2000). Monthly measurement of daily timers. *Journal of Financial and Quantitative Analysis*, *35*(3), 247–290.

Goetzmann, W. N., Ingersoll, J., Spiegel, M., & Welch, I. (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies*, *20*(5), 1503–1546.

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

Grinblatt, M., & Titman, S. (1989). Portfolio performance evaluation: Old issues and new insights. *Review of Financial Studies*, *2*(3), 393–421.

Grinblatt, M., & Titman, S. (1994). A study of monthly mutual fund returns and performance evaluation techniques. *Journal of Financial and Quantitative Analysis*, *29*(3), 419–444.

Han, Y. (2006). Asset allocation with a high dimensional latent factor stochastic volatility model. *Review of Financial Studies*, *19*(1), 237–271.

Henriksson, R. D., & Merton, R. C. (1981). On market timing and investment performance: II. Statistical procedures for evaluating forecasting skills. *Journal of Business*, *54*(4), 513–533.

Jagannathan, R., & Korajczyk, R. A. (1986). Assessing the market timing performance of managed portfolios. *Journal of Business*, *59*(2), 217–235.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, *48*(1), 65–91.

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

Jensen, M. C. (1972). Optimal utilization of market forecasts and the evaluation of investment performance. In G. P. Szego & K. Shell (Eds.), *Mathematical methods of investment and finance* (pp. 310–335). Amsterdam: Elsevier.

Jiang, G. J., Yao, T., & Yu, T. (2007). Do mutual funds time the market? Evidence from portfolio holdings. *Journal of Financial Economics*, *86*(3), 724–758.

Kacperczyk, M., van Nieuwerburgh, S., & Veldkamp, L. (2014). Time-varying fund manager skill. *Journal of Finance*, *69*(4), 1455–1484.

Kryzanowski, L., Lalancette, S., & To, M. C. (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.

Lo, A. W. (2008). Where do alphas come from? A new measure of the value of active investment management. *Journal of Investment Management*, *6*(2), 1–29.

Mamaysky, H., Spiegel, M., & Zhang, H. (2008). Estimating the dynamics of mutual fund alphas and betas. *Review of Financial Studies*, *21*(1), 233–264.

Merton, R. 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, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Pástor, L., Stambaugh, R. F., & Taylor, L. A. (2015). Scale and skill in active management. *Journal of Financial Economics*, *116*(1), 23–45.

Treynor, J. L., & Mazuy, K. K. (1966). Can mutual funds outguess the market? *Harvard Business Review*, *44*(4), 131–136.

Wermers, R. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses. *Journal of Finance*, *55*(4), 1655–1703.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 818–838.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Bunnenberg S, Rohleder M, Scholz H, Wilkens M. Jensen's alpha and the market-timing puzzle. *Rev Financ Econ*. 2019;37:234–255. https://doi.org/10.1002/rfe.1033

APPENDIX A

LITERATURE REVIEW

We have reviewed all issues of the *Journal of Finance* (*JF*), the *Journal of Financial Economics* (*JFE*), and the *Review of Financial Studies* (*RFS*) since 2005. As we aim to identify all empirical studies that apply at least one of the TM and HM models, we consider only articles that cite Treynor and Mazuy (1966), Henriksson and Merton (1981), or both. Furthermore, we record if and how these studies report estimations of timing performance—that is, the average abnormal return resulting from market-timing activities— or total performance. The following table summarizes our results.

Study	TM	HM	Dataset	Timing/total performance	Remarks
Amihud & Goyenko (2013) (*RFS*)	X	X	Monthly returns of 2,460 US mutual funds, 1988–2010	No	Timing activities are not causal for mutual funds R^2
Bali, Brown, & Caglayan (2014) (*JFE*)	–	X	Monthly returns of 10,305 US hedge funds, 1994–2012	No	Hedge funds successfully time macroeconomic uncertainty
Ben-Rephael, Kandel, & Wohl (2012) (*JFE*)	X	X	Aggregated mutual funds flows from 1984 to 2008	No	An investment strategy based on lagged net flows exhibits significant timing
Blake, Rossi, Timmermann, Tonks, & Wermers (2013) (*JF*)	X	–	Quarterly returns of 2,385 UK pension funds, 1984–2004	Yes	TM total performance is calculated as in Equation (5)
Bollen & Busse (2005) (*JF*)	X	X	Daily returns of 230 US mutual funds, 1985–1995	Yes	TM and HM total performance is not calculated as in Equations (7) and (8)
Bollen & Whaley (2009) (*JF*)	X	–	Monthly returns of 6,158 US hedge funds, 1994–2005	No	Hedge funds significantly change risk loadings
Cao, Chen, Liang, & Lo (2013) (*JFE*)	X	–	Monthly returns of 5,298 US hedge funds, 1994–2009	No	Hedge fund managers time market liquidity
Chen et al. (2010) (*JFE*)	X	–	Monthly returns of 1,054 US bond funds, 1962–2007	Yes	Total performance is derived according to Aragon and Ferson (2006)
Christoffersen & Sarkissian (2009) (*JFE*)	X	–	Monthly returns of 1,917 US mutual funds, 1992–2002	No	Timing abilities of managers vary little with their location
Comer, Larrymore, & Rodriguez (2009) (*RFS*)	X	–	Daily returns of 395 US hybrid funds, 1994–2005	Yes	TM total performance is not calculated as in Equation (7)
Dass, Nanda, & Wang (2013) (*JFE*)	X	X	Monthly returns of 367 US balanced funds, 1992–2009	Yes	Timing performance is based on portfolio weights
Goetzmann et al. (2007) (*RFS*)	X	X	–	Yes	Total performance is based on option pricing theory
Han (2006) (*RFS*)	X	–	Daily returns of 36 US stocks, 1990–2000	Yes	TM-timing performance is calculated as in Equation (5)
Jiang et al. (2007) (*JFE*)	X	X	Monthly returns of 2,294 US mutual funds, 1980–2002	Yes	TM-timing performance is based on option pricing theory
Mamaysky, Spiegel, & Zhang (2008) (*RFS*)	X	X	Monthly and daily returns of US funds, 1970–2002	Yes	TM and HM total performance is not calculated as in Equations (6) and (7)

APPENDIX B

CALCULATING TIMING PERFORMANCE FROM TIMING ACTIVITY

Timing performance is the contribution of timing activity to a fund's total abnormal return. It can be determined analytically by setting the dynamic exposure to market risk to be $\tilde{\zeta}_{it} = \zeta_i + \gamma_i \cdot f(er_{mt})$ in Equation (2) and taking the expectation of

Equation (2) (Grinblatt & Titman, 1989):

$$\begin{aligned}
\mathrm{E}(er_{it}) &= \mathrm{E}(\delta_i + \tilde{\zeta}_{it} er_{mt} + \eta_{it}) \\
&= \delta_i + \mathrm{E}(\tilde{\zeta}_{it} er_{mt}) \\
&= \delta_i + \mathrm{E}(\tilde{\zeta}_{it})\mathrm{E}(er_{mt}) + \mathrm{Cov}(\tilde{\zeta}_{it}, er_{mt})
\end{aligned}$$

The first summand in the last line above is the selection performance, the second summand is the expected premium for the expected exposure to market risk. Thus, the third summand is the contribution of timing activities to the expected return, and timing performance is measured by the covariance between the timed market factor and the fund's exposure to this factor.

For the TM model, this covariance tim_i^{TM} is (Grinblatt & Titman, 1994)

$$tim_i^{TM} = \gamma_i^{TM}\mathrm{Var}(er_{mt})$$

For the HM model, the dynamic exposure to the timed market factor $\tilde{\zeta}_{it}^{HM}$ can be stated as

$$\tilde{\zeta}_{it}^{HM} = \zeta_i^{HM} + \gamma_i^{HM}\frac{\max(0, er_{mt})}{er_{mt}}$$

We derive the covariance of $\mathrm{Cov}(\tilde{\zeta}_{it}^{HM}, er_{mt})$ as follows:

$$\begin{aligned}
tim_i^{HM} &= \mathrm{Cov}\left[\tilde{\zeta}_{it}^{HM}, er_{mt}\right] \\
&= \mathrm{Cov}\left[\zeta_i^{HM} + \gamma_i^{HM}\frac{\max(0, er_{mt})}{er_{mt}}, er_{mt}\right] \\
&= \gamma_i^{HM}\mathrm{Cov}\left[\frac{\max(0, er_{mt})}{er_{mt}}, er_{mt}\right] \\
&= \gamma_i^{HM}\left[\mathrm{E}[\max(0, er_{mt})] - \mathrm{E}\left[\frac{\max(0, er_{mt})}{er_{mt}}\right]\mathrm{E}(er_{mt})\right] \\
&= \gamma_i^{HM}[\mathrm{E}[\max(0, er_{mt})] - \Pr(er_{mt}>0)\mathrm{E}(er_{mt})]
\end{aligned}$$

APPENDIX C

THE TIMING-INDUCED BIAS IN JA IN A MULTIFACTOR MODEL

We apply our nested model analysis in Section 3.1 to the four-factor model of Fama and French (1993) and Carhart (1997). To derive the relation between JA and TM total performance, we first identify the maximum correlation portfolio including the factors smb_t, hml_t, and mom_t:

$$er_{mt}^2 = \alpha_h^{TM} + \beta_{1h}^{TM} er_{mt} + \beta_{2h}^{TM} smb_t + \beta_{3h}^{TM} hml_t + \beta_{4h}^{TM} mom_t + \psi_{ht}^{TM}$$

Then, we substitute the right-hand side of the equation above into the TM model in Equation (19):

$$\begin{aligned}
er_{it} &= \delta_{it}^{TM} + \zeta_{1i}^{TM} er_{mt} + \zeta_{2i}^{TM} smb_t + \zeta_{3i}^{TM} hml_t + \zeta_{4i}^{TM} mom_t \\
&\quad + \gamma_i^{TM}\left(\alpha_h^{TM} + \beta_{1h}^{TM} er_{mt} + \beta_{2h}^{TM} smb_t + \beta_{3h}^{TM} hml_t + \beta_{4h}^{TM} mom_t + \psi_{ht}^{TM}\right) + \eta_{it}^{TM} \\
&= \underbrace{(\delta_{it}^{TM} + \gamma_i^{TM}\alpha_h^{TM})}_{\alpha_i} + \underbrace{(\zeta_{1i}^{TM} + \gamma_i^{TM}\beta_{1h}^{TM})}_{\beta_{1i}} er_{mt} + \underbrace{(\zeta_{2i}^{TM} + \gamma_i^{TM}\beta_{2h}^{TM})}_{\beta_{2i}} smb_t \\
&\quad + \underbrace{(\zeta_{3i}^{TM} + \gamma_i^{TM}\beta_{3h}^{TM})}_{\beta_{3i}} hml_t + \underbrace{(\zeta_{4i}^{TM} + \gamma_i^{TM}\beta_{4h}^{TM})}_{\beta_{4i}} mom_t + \underbrace{(\eta_{it}^{TM} + \gamma_{it}^{TM}\psi_{ht}^{TM})}_{\varepsilon_{it}}
\end{aligned}$$

This equation is identical to JA as in Equation (18). Consequently, JA is the sum of a fund's TM selection performance δ_i^{TM} and its TM-timing activity γ_i^{TM} multiplied by the alpha of the squared market return with respect to the maximum squared correlation portfolio in the market factors α_h^{TM}. Using the TM total performance in Equation (7), we can determine the bias in JA to be:

$$\begin{aligned}\alpha_i - tot_i^{TM} &= \delta_i^{TM} + \gamma_i^{TM}\alpha_h^{TM} - \left[\delta_i^{TM} + \gamma_i^{TM}\mathrm{Var}(er_{mt})\right] \\ &= \gamma_i^{TM}\left[\alpha_h^{TM} - \mathrm{Var}(er_{mt})\right]\end{aligned}$$

Similarly, we relate JA to HM total performance by first estimating:

$$\max(0, er_{mt}) = \alpha_h^{HM} + \beta_{1h}^{HM} er_{mt} + \beta_{2h}^{HM} smb_t + \beta_{3h}^{HM} hml_t + \beta_{4h}^{HM} mom_t + \psi_{ht}^{HM}$$

Then, we substitute the right-hand side of this equation into the HM model in Equation (20):

$$\begin{aligned}er_{it} &= \delta_i^{HM} + \zeta_{1i}^{HM} er_{mt} + \zeta_{2i}^{HM} smb_t + \zeta_{3i}^{HM} hml_t + \zeta_{4i}^{HM} mom_t \\ &\quad + \gamma_i^{HM}\left(\alpha_h^{HM} + \beta_{1h}^{HM} er_{mt} + \beta_{2h}^{HM} smb_t + \beta_{3h}^{HM} hml_t + \beta_{4h}^{HM} mom_t + \psi_{ht}^{HM}\right) + \eta_{it}^{HM} \\ &= \underbrace{\left(\delta_{it}^{HM} + \gamma_i^{HM}\alpha_h^{HM}\right)}_{\alpha_i} + \underbrace{\left(\zeta_{1i}^{HM} + \gamma_i^{HM}\beta_{1h}^{HM}\right)}_{\beta_{1i}} er_{mt} + \underbrace{\left(\zeta_{2i}^{HM} + \gamma_i^{HM}\beta_{2h}^{HM}\right)}_{\beta_{2i}} smb_t \\ &\quad + \underbrace{\left(\zeta_{3i}^{HM} + \gamma_i^{HM}\beta_{3h}^{HM}\right)}_{\beta_{3i}} hml_t + \underbrace{\left(\zeta_{4i}^{HM} + \gamma_i^{HM}\beta_{4h}^{HM}\right)}_{\beta_{4i}} mom_t + \underbrace{\left(\eta_{it}^{HM} + \gamma_{it}^{HM}\psi_{ht}^{HM}\right)}_{\varepsilon_{it}}\end{aligned}$$

This is again identical to JA. Consequently, JA is the sum of a fund's HM selection performance δ_i^{HM} and its HM-timing activity γ_i^{HM} multiplied by the alpha of the protective put α_h^{HM}. Using the definition of the HM total performance in Equation (8), we can determine the bias in JA to be:

$$\begin{aligned}\alpha_i - tot_i^{HM} &= \delta_i^{HM} + \gamma_i^{HM}\alpha_h^{HM} - \left[\delta_i^{HM} + \gamma_i^{HM}\left[\mathrm{E}[\max(0, er_{mt})] - \mathrm{E}(er_{mt})\Pr(er_{mt}>0)\right]\right] \\ &= \gamma_i^{HM}\left[\alpha_h^{HM} - [\mathrm{E}[\max(0, er_{mt})] - \mathrm{E}(er_{mt})]\Pr(er_{mt}>0)\right]\end{aligned}$$