Journal Of Investment Management, Vol. 11, No. 2, (2013), pp. 73–91
© JOIM 2013

www.joim.com

MUTUAL FUND'S NET ECONOMIC ALPHA: DEFINITION AND EVIDENCE

Sharon Garyn-Tal[a] *and Beni Lauterbach*[b]

It is sometimes argued that existing methodologies for assessing mutual fund's performance are unfair, as fund's return is taken net of expenses and benchmark return is gross of expenses. Examining over 1000 U.S. non-specialized mutual funds in 2001–2009, we find that the above-mentioned problem is minute—the net economic alpha, an alpha that accounts for the actual costs of investing in benchmarks via exchange-traded funds (ETFs), yields similar fund's ranking and classification as the traditional methods. Also interesting is that the average net economic alpha is only slightly negative, suggesting that the mutual funds industry is not inferior.

1 Introduction

The classical measures of investment performance compare the return of a managed portfolio with the return of a benchmark portfolio. The benchmark portfolio should represent an equivalent and feasible investment alternative to the managed portfolio being evaluated.

Jensen (1968) uses the market portfolio of the CAPM to advocate Jensen's alpha. A positive alpha implies that the manager earns an abnormal return relative to the alternative of holding the benchmark portfolio. Current practice follows Fama and French (1996) and Carhart (1997), who suggest evaluating performance using three- or four-factor models derived from empirically observed patterns in stock portfolio returns. In general, performance evaluations based on these models typically conclude that mutual funds deliver disappointing net returns.

In existing comparisons of a managed portfolio return with that of a factor model or benchmark it is regularly assumed that the benchmark and/or factors have no costs. This generates a bias against the evaluated managed portfolio, as the measured

[a]Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology, Haifa 32000, Israel. E-mail: sharon.tal.garyn@gmail.com
[b]School of Business Administration, Bar-Ilan University, Ramat-Gan 52900, Israel. E-mail: beni.lauterbach@biu.ac.il

Not for Distribution

return of the managed portfolio, mutual fund for example, is evaluated net of the funds' expenses and trading costs. Fama and French (2009) on the bottom of page 6 write that "…if the question is whether mutual funds are better for investors than passive investments, benchmark returns, like fund returns, should be net of costs."

In accordance with Fama and French (2009) recommendation, we compute the net economic alpha by adding to the traditional alpha as extracted from a benchmark model regression the cost of mimicking the fund's systematic risk via exchange-traded funds (ETFs).[1] The motivation for the net economic alpha is that the alternative to buying a mutual fund—buying ETFs on the benchmarks—is not costless (in fact, the costs of mimicking multiple benchmarks or the long-short portfolios of Fama and French are non-negligible). Our suggested net economic alpha should better represent the net economic (i.e., the relative to alternatives) excess return of the fund.

Comparing the standard alphas with their corresponding net economic alphas, we find little differences in fund's classification and ranking. In our data, comprising more than 1,000 mutual funds over 2001–2009, the impact of the net economic alpha is minute. However, we also observe that net economic alphas change a bit the outlook of mutual fund performance. The mean net economic alpha of mutual funds in our sample is typically small, between 0 and -1% per year. Moreover, examining the subperiods of our sample, the mean net economic alphas are becoming more positive with time. In the last subperiod net economic alphas are positive on average. This may imply that the mutual fund industry is only slightly inferior to ETFs and that it (the mutual fund industry) is becoming more and more competitive (relative to ETFs) with time.

2 Net economic alphas' definition

Consider, for simplicity, a mutual fund with a beta relative to the market of 1. Traditional practice regresses its net of fees excess return on the excess return of the market index. The intercept of that regression is commonly termed the alpha, α_P, and serves as an estimate of the excess return on the fund due to its manager's skill. However, α_P underestimates the managerial contribution because the alternative to the fund, investing in an ETF mimicking the market, is also costly as ETFs also have expense ratios. Thus, a more precise and relevant measure of the manager's contribution to the fund's return is α_P + the expense ratio on the "matching" ETF. This measure, termed by us the net economic alpha, measures the net return of the fund relative to the net return of its ETF alternative. It essentially subtracts the net return of the fund from the net return of an ETF (or in more general cases an ETF portfolio) that mimics the fund's systematic risk.

Given the above discussion, the net economic alpha can be computed as:

$$\alpha_{NET,P,N} = \alpha_{P,N} + \sum_{i=1}^{N} \phi_i |\beta_{P,i}|, \qquad (1)$$

where $\alpha_{P,N}$ is the intercept of the traditional regression of the (net of fee) excess return of the fund on the excess returns of the factors in the multifactor model, ϕ_i is the expense ratio for an ETF that follows benchmark i, and $\beta_{P,i}$ is benchmark i's coefficient in the above-mentioned traditional regression. A formal derivation of Equation (1) is offered in the Appendix.

While we have developed the net economic formula based on Fama and French (2009) recommendation, a question may arise whether our netting out procedure is preferable to a grossing up procedure, where fund return is grossed up by its expense ratio before the traditional regression. Cesari and Panetta (2002) and Otten and

Not for Distribution

Bams (2007), among others, have used the grossing up procedure. In our opinion, the grossing up procedure is less suitable for evaluating fund's performance from the investor's viewpoint. This is because the net return approach directly measures and analyzes the net returns to investors, and from these net returns the true net excess return ("alpha") to fund investors should be extracted. This is basically the insight of Fama and French (2009) comment.

3 Data and benchmark models

3.1 Sample and variables

We collect monthly return data on equity funds that are classified by Lipper as "non-specialized open-end" equity funds. The sample period is 2001–2009 (as Lipper classification is available starting in 2001) and return data are from CRSP (The Center for Research in Security Prices). Our choice to focus on the new millennium data is also motivated by the recognition that comparison with multifactor models and ETFs that afford mimicking these factors became feasible only in the 2000s.

We start with all available funds on the CRSP database in 2001–2009, and exclude funds that changed their investment policy (Lipper classification), funds for which Total Net Assets (TNA) is missing, and funds with less than 12 return observations. Our empirical inference is based on funds with full return data within an analyzed subperiod. (We divide the overall sample period into three 3-years subperiods.) However, because survivorship bias is an issue, we also present some results for funds with at least 12 months of return data.

Mutual funds typically have more than one class of shares. We compute fund's return as a value-weighted average of the returns on the minimum and maximum TNA classes of that fund.

For the empirical analysis we also need data on several other variables. The Fama–French–Carhart four factors returns are taken from French website: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

We plan to define a fund's style by identifying the benchmark that yields the highest correlation with respect to the fund's return. Thus, a list of benchmarks should be suggested. Based on a review of the Internet information sheet of Fidelity and Vanguard non-specialized open-end equity funds, we find that the main benchmarks used for evaluating these funds' performance are the S&P500 or an appropriate Russell index (Russell 1000, Russell 1000 growth, Russell 1000 value, Russell midcap, Russell midcap growth, Russell midcap value, Russell 2000, Russell 2000 growth, Russell 2000 value). The Russell indices data and returns are extracted from the Russell investments website (www.russell.com). S&P500 returns are collected from the CRSP database.

3.2 Data characteristics

The data comprises nine years (2001–2009) of funds' returns and is divided into three subperiods: 2001–2003, 2004–2006, and 2007–2009.

There are 1,037 funds with full return data throughout 2001–2003 and 1,545 funds with at least 12 return observations within this subperiod. Of the 1,037 full-data funds, 188 are classified by Lipper as Large-Cap Core Funds (LCCE), 123 are Large-Cap Growth Funds (LCGE), 66 are Large-Cap Value Funds (LCVE), 52 are Mid-Cap Core Funds (MCCE), 85 are Mid-Cap Growth Funds (MCGE), 34 are Mid-Cap Value Funds (MCVE), 68 are Multi-Cap Core Funds (MLCE), 68 are Multi-Cap Growth Funds (MLGE), 98 are Multi-Cap Value Funds (MLVE), 89 are Small-Cap Core Funds (SCCE), 100 are Small-Cap Growth Funds (SCGE), and 66 are Small-Cap

Not for Distribution

Value Funds (SCVE). The 2001–2003 average Total Net Assets (TNA) of all funds is 667.9 million dollars. The average expense ratio (management fee) charged by fund managers is 1.31% (0.74%) and the weighted (by TNA) average expense ratio (management fee) is 1.24% (0.74% as well).[2] Khorana *et al.* (2009) study fees charged by mutual funds in 2002. They document (in Table 2, page 1288) an average management fee of 0.62% per year and an average expense ratio of 1.11% per year for U.S. equity funds. These fees are slightly lower than the fees we report above.

There are 1,547 funds with full return data in 2004–2006 and 2,065 funds with at least 12 return observations within this subperiod. The number of funds in each Lipper classification is available from the authors. However, there are at least 64 funds in each of the above-specified 12 Lipper groups. The 2004–2006 average TNA of all funds is 971.2 million dollars. The average expense ratio (management fee) charged by fund managers is 1.37% (0.78%).

There are 1,312 funds with full return data throughout 2007–2009 and 1,982 funds with at least 12 return observations. In 2007–2009 we have at least 48 funds in each of the Lipper groups. The 2007–2009 average TNA of all funds is 759.6 million dollars. The average expense ratio (management fee) charged by fund managers is 1.27% (0.73%).

3.3 The benchmark models

We apply five different benchmark models for evaluating a fund's alpha. The models differ in the benchmark-system used in the regression.

(1) M – the CAPM regression is

$$R_{P,t} - R_{f,t} = \alpha_P + \beta_P(R_{M,t} - R_{f,t}) + \epsilon_{P,t}. \quad (2)$$

The excess market return is the value-weighted return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate (from Ibbotson Associates).

(2) S – the fund's style—the style benchmark is determined by checking the correlations between a fund's return and the return on each of the following ten indices—Russell 1000, Russell 1000 growth, Russell 1000 value, Russell midcap, Russell midcap growth, Russell midcap value, Russell 2000, Russell 2000 growth, Russell 2000 value, and S&P500. The index that yields the highest correlation with respect to the fund's return is defined as the fund's style. The regression equation for fund P is

$$R_{P,t} - R_{f,t} = \alpha_P + \beta_P(R_{S,t} - R_{f,t}) + \epsilon_{P,t}. \quad (3)$$

(3) 3F – the Fama and French three-factor model is

$$R_{P,t} - R_{f,t} = \alpha_P + \beta_{P,1}(R_{M,t} - R_{f,t}) + \beta_{P,2}HML_t + \beta_{P,3}SMB_t + \epsilon_{P,t}. \quad (4)$$

The factors are: (1) the excess market returns is estimated as in model 1 above, (2) the performance of value stocks relative to growth stocks (HML, High Minus Low), and (3) the performance of small stocks relative to big stocks (SMB, Small Minus Big).

(4) 4F – Carhart's (1997) four-factor model is

$$R_{P,t} - R_{f,t} = \alpha_P + \beta_{P,1}(R_{M,t} - R_{f,t}) + \beta_{P,2}HML_t + \beta_{P,3}SMB_t + \beta_{P,4}MOM_t + \epsilon_{P,t}. \quad (5)$$

The model is based on the Fama and French three-factor model and the additional momentum factor (MOM) of Carhart (1997), constructed as the

Not for Distribution

average return on the two high prior return portfolios minus the average return on the two low prior return portfolios.

(5) 4F-S – a five-factor model, where the fifth benchmark is the fund's style. Adding fund's style as a fifth factor may appear redundant, as all style returns are highly correlated with market return. For example, in our sample, the correlations of style returns with the market return vary between 0.79 and almost 1. Nevertheless, we believe there is a style-specific behavior. Thus, to eliminate the dependency of the style factor, we employ the Hunter *et al.* (2009) methodology. Specifically, we regress the style factor excess return on the other four factors, and use the intercept plus residual of this regression as an estimate of an orthogonal style factor S*. (Thus, factor S* is uncorrelated with R_M.) Then, we estimate fund P alpha via the regression:

$$\begin{aligned} R_{P,t} - R_{f,t} = \alpha_P &+ \beta_{P,1}(R_{M,t} - R_{f,t}) \\ &+ \beta_{P,2} HML_t + \beta_{P,3} SMB_t \\ &+ \beta_{P,4} MOM_t + \beta_{P,5} R_{S^*,t} \\ &+ \epsilon_{P,t}. \end{aligned} \tag{6}$$

We may be the first to estimate the above five-factor model. Nevertheless, Hunter *et al.* (2009) have previously suggested to add a group factor to the Fama–French–Carhart model. A more popular five-factor model that adds the liquidity factor (see Pastor and Stambaugh, 2003) is not attempted because we could not find ETFs that mimic the liquidity factor. Without such ETFs, computing net economic alphas (our main task) is impossible.

3.4 *Variables constructed for net economic alpha estimation*

Since we are interested in net economic alphas, we need to assess the expense ratios of ETFs replicating fund styles, and expense ratios of ETFs replicating the Fama–French–Carhart four factors.[3]

Assessing the expense ratios of ETFs, we ignore brokerage commissions because all kinds of discounts and online trading are available, and we omit the bid—ask spread because it is negligible for heavily traded ETFs. This is consistent with existing literature (see Dellva, 2001; Kostovetsky, 2003). As for mutual funds expenses, we ignore redemption fees and loads. We hope that these ignored costs of ETFs and mutual funds balance each other off.

To assess relevant ETFs' expense ratio, we examine first all ETFs that are relevant for our one-factor models (the CAPM and the Style-factor models). These are ETFs that follow the S&P500 and the various Russell indices. When we find several ETFs that follow the same index, we choose the most veteran ETF—the index-linked ETF that started trading first. In our view, the most veteran-specific-index-linked ETF is preferred because it was the first available alternative.[4]

A noteworthy technical point is that if an ETF is leveraged relative to its benchmark, we consider the benchmark mimicking effective expense ratio as the ETF's expense ratio divided by level of leverage. Thus, for example, Direxion Daily Large Cap Bear 3x Shares charges an expense ratio of 0.95%. However, since this ETF is leveraged -X3 relative to its benchmark, we assess the benchmark mimicking effective expense ratio as only 0.95% divided by 3.

Table 1 specifies our chosen ETFs. Based on a review of the Internet information sheet of the chosen ETFs, we assess that a long (short) position in the S&P500—the market factor—costs 0.10% (0.90%). This is because SPDRs S&P500 charges 0.10% for providing a long position in the S&P500 index, and ProShares Short S&P500

Not for Distribution

Table 1 A list of ETFs following various investment styles and asset pricing factors.

Panel A. Simple benchmarks

Benchmark	ETF name	Symbol	Long/short and leverage	Total net assets (TNA)	Inception date	ETFs expense ratio (%)	Benchmark mimicking effective expense ratio (%)
S&P500	SPDRs S&P500	SPY	Long	99.88M	1/22/1993	0.10	0.10
	ProShares Short S&P500	SH	Short	1.84B	6/19/2006	0.90	0.90
Rusell 1000	iShares Russell 1000 Index Fund	IWB	Long	6.84B	5/15/2000	0.15	0.15
	Direxion Daily Large Cap Bear 3x Shares	BGZ	Short (-X3)	163.61M	11/5/2008	0.95	0.32
Russell 1000 Growth	iShares Russell 1000 Growth Index Fund	IWF	Long	16.19B	5/22/2000	0.20	0.20
	ProShares UltraShort Russell 1000 Growth	SFK	Short (-X2)	2.09M	2/20/2007	0.95	0.48
Russell 1000 Value	iShares Russell 1000 Value	IWD	Long	12.28B	5/2/2000	0.20	0.20
	ProShares UltraShort Russell 1000 Value	SJF	Short (-X2)	1.69M	2/20/2007	0.95	0.48
Russell Midcap	iShares Russell Midcap Index Fund	IWR	Long	6.68B	7/17/2001	0.21	0.21
	Direxion Daily Midcap Bear 3x Shares	MWN	Short (-X3)	12.19M	1/8/2009	0.95	0.32
Russell Midcap Growth	iShares Russell Midcap Growth Index Fund	IWP	Long	3.41B	7/17/2001	0.25	0.25
	ProShares UltraShort Russell MidCap Growth	SDK	Short (-X2)	1.96M	2/20/2007	0.95	0.48
Russell Midcap Value	iShares Russell Midcap Value Index Fund	IWS	Long	2.98B	7/17/2001	0.26	0.26
	ProShares UltraShort Russell MidCap Value	SJL	Short (-X2)	1.44M	2/20/2007	0.95	0.48

Not for Distribution

Table 1 (*Continued*)

Benchmark	ETF name	Symbol	Long/short and leverage	Total net assets (TNA)	Inception date	ETFs expense ratio (%)	Benchmark mimicking effective expense ratio (%)
Russell 2000	iShares Russell 2000 Index Fund	IWM	Long	14.82B	5/22/2000	0.26	0.26
	ProShares Short Russell 2000	RWM	Short	351.73M	1/23/2007	0.95	0.95
	ProShares UltraShort Russell 2000	TWM	Short (-X2)	260.37M	1/23/2007	0.95	0.48
Russell 2000 Growth	iShares Russell 2000 Growth Index	IWO	Long	3.85B	7/24/2000	0.25	0.25
	ProShares UltraShort Russell 2000 Growth Index	SKK	Short (-X2)	9.25M	2/20/2007	0.95	0.48
Russell 2000 Value	iShares Russell 2000 Value Index	IWN	Long	3.83B	7/24/2000	0.40	0.40
	ProShares UltraShort Russell 2000 Value	SJH	Short (-X2)	4.51M	2/20/2007	0.95	0.48

Panel B. Factors

Benchmark	ETF name	Symbol	Factor replication	Factor type	Total net assets (TNA)	Inception date	ETFs expense ratio (%)	Benchmark mimicking effective expense ratio (%)
SMB	iShares Russell 2000 Index Fund	IWM	—	—	14.82B	5/22/2000	0.26	
	ProShares UltraShort Russell 2000	TWM	—	—	260.37M	1/23/2007	0.95	
	iShares Russell 1000 Index Fund	IWB	—	—	6.84B	5/15/2000	0.15	
	Direxion Daily Large Cap Bear 3x Shares	BGZ	—	—	163.61M	11/5/2008	0.95	
			IWM+BGZ/3	SMB long				0.58
			IWB+TWM/2	SMB short				0.63

Not for Distribution

Table 1 (*Continued*)

Benchmark	ETF name	Symbol	Factor replication	Factor type	Total net assets (TNA)	Inception date	ETFs expense ratio (%)	Benchmark mimicking effective expense ratio (%)
HML	iShares Russell 3000 Value Index	IWW	—	—	293.4M	7/24/2000	0.25	
	ProShares UltraShort Russell 1000 Value	SJF	—	—	1.69M	2/20/2007	0.95	
	ProShares UltraShort Russell 2000 Value	SJH	—	—	4.51M	2/20/2007	0.95	
	iShares Russell 3000 Growth Index	IWZ	—	—	346.05M	7/24/2000	0.25	
	ProShares UltraShort Russell 1000 Growth	SFK	—	—	2.09M	2/20/2007	0.95	
	ProShares UltraShort Russell 2000 Growth Index	SKK	—	—	9.25M	2/20/2007	0.95	
			IWW+(SFK+SKK)/4	HML long				0.73
			IWZ+(SJF+SJH)/4	HML short				0.73
Momentum	Russell Contrarian ETF	CNTR	—	—	4.61M	5/17/2011	0.37	
	Russell Small Cap Contrarian ETF	SCTR	—	—	6.50M	10/4/2011	0.45	
	Russell 1000 High Momentum ETF	HMTM	—	—	5.27M	5/25/2011	0.20	
	Russell 2000 High Momentum ETF	SHMO	—	—	5.02M	5/25/2011	0.30	
			(HMTM+SHMO)/2 -(CNTR+SCTR)/2	Momentum long				0.66
			(CNTR+SCTR)/2 -(HMTM+SHMO)/2	Momentum short				0.66

Not for Distribution

charges 0.90% for providing a short position in the S&P500. Using the same logic and procedure, a long (short) position in Russell 1000 costs 0.15% (0.32%), a long (short) position in Russell 1000 growth costs 0.2% (0.48%), a long (short) position in Russell 1000 value costs 0.2% (0.48%), a long (short) position in Russell midcap costs 0.21% (0.32%), a long (short) position in Russell midcap growth costs 0.25% (0.48%), a long (short) position in Russell midcap value costs 0.26% (0.48%), a long (short) position in Russell 2000 costs 0.26% (0.48%), a long (short) position in Russell 2000 growth costs 0.25% (0.48%), and a long (short) position in Russell 2000 value costs 0.4% (0.48%).

In Panel B we assess the expense ratio of ETFs that follow the common asset pricing factors. The expense ratio of a long position in the SMB factor is composed of the expense ratio charged for investing in a long position in a small stocks portfolio (via iShares Russell 2000 Index Fund that charged an expanse ratio of 0.26%) plus the expense ratio charged for investing in a short position in a big stocks portfolio (via Direxion Daily Large Cap Bear 3x Shares that charges an expense ratio of 0.95%). It is important to note that there are no short-sell restrictions on the simple investor when she buys this Direxion ETF. Hence, the short position is readily available, and its costs are $1/3 * 0.95\%$ (because this ETF is leveraged -X3 relative to its benchmark). Thus, we assess that a long position on the SMB factor costs $0.26\% + 0.95\%/3 = 0.58\%$.[5] Short position in the SMB factor is composed of a long position in the big stocks portfolio (via iShares Russell 1000 Index Fund that charged an expense ratio of 0.15%), and a short position in the small stocks portfolio (via ProShares UltraShort Russell 2000 that charges an expense ratio of 0.95%). (Again, since this ETF is leveraged -X2 relative to its benchmark, we assess the benchmark mimicking effective expense ratio as only 0.95% divided by 2). Thus, we assess that a short position in the SMB factor costs 0.63%.

In the same spirit, we assess that a long (short) position in the HML factor costs 0.73% (0.73%).

To mimic the momentum factor we use a series of four relevant ETFs: Russell 1000 High Momentum ETF, Russell 2000 High Momentum ETF, Russell Contrarian ETF, and Russell Small Cap Contrarian ETF. This is because the composition of the momentum factor a-la French is the average of the returns on two (big and small) high prior return portfolios minus the average of the returns on two low prior return portfolios. Accordingly, a long (short) position in the momentum factor costs 0.66% (0.66%). Note that the momentum factor is the only one in this study that investors cannot readily imitate. Because there are no short ETFs available to mimic the momentum factors requires that the investor actually shorts some ETFs (the ETFs with the minus sign).

4 Empirical results

4.1 Funds' alpha estimates

We fit our benchmark models over three subperiods: 2001–2003, 2004–2006, and 2007–2009. Table 2 reports alpha estimates in 2001–2003, 2004–2006, and 2007–2009, for each of the five benchmark frameworks. Panel A summarizes alpha estimates for all funds that survived at least 12 months within a subperiod, while Panel B focuses on funds that survived throughout the subperiod.

The mean alphas of our benchmark models are mostly negative. Consistent with previous evidence, mutual funds tend to deliver disappointing net returns. The negative excess returns can be explained, in most cases, by the funds' expense ratios of about 1.3%—see Section 3.2, suggesting that typically the before-fees average

Not for Distribution

Table 2 Alphas and net economic alphas generated by five benchmark models of funds' return.

Panel A. Mean alphas of funds that survived at least 12 months within the subperiod

	2001–2003 (1,545 funds)		2004–2006 (2,065 funds)		2007–2009 (1,982 funds)	
Model	Annual $\bar{\alpha}$ (%)	Annual $\bar{\alpha}_{NET}$ (%)	Annual $\bar{\alpha}$ (%)	Annual $\bar{\alpha}_{NET}$ (%)	Annual $\bar{\alpha}$ (%)	Annual $\bar{\alpha}_{NET}$ (%)
M	0.12	0.22*	−2.67**	−2.55**	−0.26*	−0.16
S	−1.90**	−1.70**	−1.26**	−1.06**	−0.62**	−0.42**
3F	−3.92**	−3.31**	−0.89**	−0.40**	−1.21**	−0.71**
4F	−4.05**	−3.56**	−0.50*	0.09	−1.20**	−0.67**
4F-S	−2.61**	−1.80**	−1.04**	−0.30**	−0.80**	−0.11

Panel B. Mean alphas of funds that survived the entire 36 months subperiod

	2001–2003 (1,037 funds)		2004–2006 (1,547 funds)		2007–2009 (1,312 funds)	
Model	Annual $\bar{\alpha}$ (%)	Annual $\bar{\alpha}_{NET}$ (%)	Annual $\bar{\alpha}$ (%)	Annual $\bar{\alpha}_{NET}$ (%)	Annual $\bar{\alpha}$ (%)	Annual $\bar{\alpha}_{NET}$ (%)
M	0.88**	0.98**	−2.50**	−2.38**	0.23*	0.33**
S	−1.36**	−1.16**	−0.98**	−0.77**	−0.13	0.07
3F	−3.79**	−3.30**	−0.70**	−0.21	−0.71**	−0.23
4F	−3.68**	−3.07**	−0.30**	0.27*	−0.71**	−0.20
4F-S	−2.25**	−1.46**	−0.81**	−0.08	−0.36**	0.31**

The benchmark models are: 1. *M*: the CAPM—the market factor, 2. S: the fund's style—the fund's style is determined by identifying the benchmark that generates the highest correlation with respect to the fund's return, 3. 3F: Fama and French three-factor model, 4. 4F: Carhart's four-factor model, and 5. 4F-S: a five-factor model that includes Carhart's four-factor model and the style factor. We denote $\bar{\alpha}$ as the average standard alpha, and $\bar{\alpha}_{NET}$ as the average net economic alpha. $\bar{\alpha}$ is computed as follows: first, for each fund, the intercept from the regression of net of fees fund excess return on the excess return on the appropriate indices is extracted. Second, the fund's annualized alpha is calculated as $(1 + intercept)^{12} - 1$. Then, we take the average of the annualized alpha of all funds and obtain $\bar{\alpha}$. $\bar{\alpha}_{NET}$ is computed as follows: for each fund P, we add to the annualized alpha, α_P, the cost of mimicking the fund's systematic risk via ETFs: $\alpha_{NET,P} = \alpha_P + \sum_{i=1}^{N} \phi_i|\beta_{P,i}|$, where ϕ_i is the expense ratio of ETFs following benchmark i and $\beta_{P,i}$, $i = 1..N$ are extracted from the regression of net of fees fund excess return on the excess return on the appropriate indices. Last, we take the average of the net economic alpha of all funds and obtain $\bar{\alpha}_{NET}$. We also test the null hypothesis H0: $\alpha = \alpha_{NET}$, and find that alpha is different from the net economic alpha at the 1% significance level for all benchmark-models.

*The coefficient is different from zero at the 5% significance level.

**The coefficient is different from zero at the 1% significance level.

performance of mutual funds might be fair and adequate.

Next, we take into consideration the costs of mimicking the funds' risk via ETFs and construct the net economic alpha. As expected, the mean net economic alphas are higher than the traditional alphas and in general closer to zero, portraying mutual funds in a more positive light.

The differences between the alpha and net economic alpha columns vary with the complexity of the benchmark model. The three- to five-factor benchmark models naturally entail higher mimicking costs than the one-factor portfolio benchmark models. For example, in Table 2, the mean mimicking costs of one-factor model is 0.1–0.2% per year. In contrast, for the four (five) factor model the mean mimicking costs increase

Not for Distribution

to about 0.5–0.6% (0.7–0.8%, respectively) per year, which may be nonnegligible. The more complicated the model gets, the higher are the alternative mimicking costs (the costs of mimicking fund's risk via ETFs) and the wider are the discrepancies between raw alphas and net economic alphas.

Using a *t*-test of differences, we find that the mean alpha is statistically significantly different from the mean net economic alpha at the 1% significance level for all five benchmark models and throughout the three subperiods: 2001–2003, 2004–2006, and 2007–2009. Yet, the magnitude of the mean difference between alphas and net economic alphas is quite modest—less than 0.8% per year. Thus, the mean economic impact of transfer to net economic alpha calculations appears quite limited.

Table 2 affords a few additional interesting observations. First, note that the CAPM yields the highest alpha and net economic alpha in 2001–2003 and 2007–2009, and the lowest alpha and net economic alpha in 2004–2006. In contrast, all our other models yield, in all three subperiods, negative alphas. It appears that the CAPM alphas vary most (and perhaps too much) from period to period.

Second, the effect of survivorship bias can be observed by comparing Panels A and B. The exclusion of nonsurviving funds increases surviving funds mean alpha by 0.43%, 0.21%, and 0.47% per year in 2001–2003, 2004–2006, and 2007–2009 subperiods, respectively. Grinblatt and Titman (1989) find that the mean survivorship bias is in the order of 0.5% per year or less.

Last and most important, in Panel B of Table 2, the average (across models) of the net economic alphas is −1.60% in 2001–2003, −0.63% in 2004–2006, and 0.06% in 2007–2009 (the averages in Panel A are slightly lower). These findings suggest that the mutual fund industry is becoming more and more competitive (relative to ETFs) with time. In fact, in the last subperiod, on average, there did not appear any noticeable performance difference between our sample mutual funds and their comparable-risk ETFs. The nice performance scores of mutual funds in the last subperiod may also emanate from the previously documented abilities of mutual funds managers to excel during financial crises. For example, Kosowski (2011) finds that mutual funds risk-adjusted performance is positive during recession periods. However, in our sample, the improvement in the mean mutual funds performance has already started in the expansionary 2004–2006 subperiod. Thus, consistent with economic theory, the growing competition from ETFs apparently induced a prudent reorganization and improvement process in the mutual fund industry.

4.2 *Differences in funds classification and ranking due to net economic alpha calculation*

In each subperiod, 2001–2003, 2004–2006, and 2007–2009, and for each of our five benchmark models, we compare the net economic alphas of our sample funds with the corresponding traditional alphas. Table 3 reports the comparison results. Panel A focuses on fund's classification differences (a sign difference between alpha and net economic alpha), Panel B reports fund's relative ranking differences of more than 10%, and Panel C (D) compares the 20 best (worst) performing funds.

Panel A involves the following calculation procedure. First, each fund is classified according to each benchmark model's alpha as either a good fund (positive alpha) or a poor fund (negative alpha). Then, for each alpha, we calculate the corresponding net economic alpha. Based on

Not for Distribution

the net economic alpha, the fund is again classified into a good fund (positive net economic alpha) or a bad fund (negative net economic alpha). Next, for each benchmark model, we compare funds classifications based on alpha with corresponding classifications based on net economic alpha. If benchmark model i's alpha and benchmark model i's net economic alpha agree

Table 3 The impact of net economic alphas on classifications and rankings.

Panel A. Classification differences—alpha versus net economic alpha

Benchmark model	Classification difference frequency: α versus α_{NET} (%)
M	0.8
S	3.1
3F	6.0
4F	6.2
4F-S	9.2
Average	5.1

Based on each alpha, a fund is classified as either a good fund (alpha is greater than zero) or a bad fund (alpha is lower than zero). Then, for each alpha, we calculate the corresponding net economic alpha. Based on the net economic alpha, the fund is again classified as a good fund (positive net economic alpha) or a bad fund (negative net economic alpha). If a fund is classified as a good (bad) fund according to both benchmark model i's alpha and benchmark model i's net economic alpha, then there is no classification difference between models i's alpha and net economic alpha. On the other hand, if a fund is classified as a good (bad) fund according to benchmark model i's alpha and is classified as a bad (good) fund according to benchmark model i's net economic alpha, then there is a classification difference between models i's alpha and net economic alpha. Panel A reports the 2001–2003, 2004–2006, and 2007–2009 average frequency of classification differences between alphas and net economic alphas. The benchmark models are: 1. M: the CAPM—the market factor, 2. S: the fund's style—the fund's style is determined by identifying the benchmark that generates the highest correlation with respect to the fund's return, 3. 3F: Fama and French three-factor model, 4. 4F: Carhart's four-factor model, and 5. 4F-S: a five-factor model that includes Carhart's four-factor model and the style factor. We use all funds with full return data throughout all 36 months of a subperiod.

Panel B. Ranks differences (of more than a decile)—alpha versus net economic alpha

Benchmark model	Rank difference frequency: α versus α_{NET} (%)
M	0.0
S	0.0
3F	0.2
4F	0.3
4F-S	0.5
Average	0.2

For each benchmark model, the best performing fund with the highest alpha is ranked at the first place, the second best performing fund with the second highest alpha is ranked at the second place, and so on. Then, based on the corresponding net economic alpha, the funds are ranked again. If fund P is ranked in the k-th place according to benchmark model i's alpha and is ranked in the l-th place according to benchmark model i's net economic alpha, then if $|k - l| < 0.1N$ then there is no ranking difference between model i's alpha and net economic alpha. On the other hand, If $|k - l| \geq 0.1N$, then there is a ranking difference between model i's alpha and net economic alpha. We compute ranking differences in each subperiod (2001–2003, 2004–2006, and 2007–2009) and report averages across the three subperiods.

Not for Distribution

Table 3 (*Continued*)

Panel C. Comparing the 20 best performing funds—alpha versus net economic alpha

Benchmark model	Consent frequency: α versus α_{NET}
M	20.0
S	20.0
3F	18.7
4F	19.0
4F-S	19.0
Average	19.3

We narrow the data and keep only funds ranked as the 20 best performing funds. Then, in each subperiod (2001–2003, 2004–2006, and 2007–2009), we count how many funds appear on both benchmark model i's alpha and net economic alpha lists of best performing funds. Averages of the three subperiods are reported in Panel C.

Panel D. Comparing the 20 worst performing funds—alpha versus net economic alpha

Benchmark model	Consent frequency: α versus α_{NET}
M	20.0
S	20.0
3F	19.3
4F	19.7
4F-S	18.7
Average	19.5

We keep only funds ranked as the 20 worst performing funds. Then, in each subperiod (2001–2003, 2004–2006, and 2007–2009), we count how many funds appear on both benchmark model i's alpha and net economic alpha lists of worst performing funds. Averages of the three subperiods are reported in Panel D.

on fund classification, then there is no classification difference between model i's alpha and model i's net economic alpha. On the other hand, if a fund is classified as a good (poor) fund according to benchmark model i's alpha and is classified as a poor (good) fund according to benchmark model i's net economic alpha, then there is a classification difference. Panel A reports, for each benchmark method, the 2001–2003, 2004–2006, and 2007–2009 average frequency of classification differences between alpha and net economic alpha.

Classification differences are rare for the CAPM (0.8%) and rather infrequent for the four-factor plus the style-factor model (9.2%). In the three-, four-, and five-factor models, the alternative of mimicking a fund's risk with ETFs is more costly than in one-factor models. Hence, naturally, fund classification differences increase with benchmark model complexity. Generally, the emerging differences in funds' classification in Panel A appear small. Only seldom does the net economic alpha methodology change the fund classification suggested by raw alphas.

Next, in Panel B, we focus on ranking differences. Assume N funds are available. For each benchmark model, the best performing fund with the highest alpha is ranked at the first place, the second best performing fund with the second highest alpha is ranked at the second place, and so on.

Not for Distribution

Then, based on the corresponding net economic alpha, the funds are ranked again. Since each period a fund has five different alphas and five different net economic alphas, all funds are ranked ten times. Then, per each benchmark model i, we compare the fund's rank based on alpha with the fund's rank based on its corresponding net economic alpha. Assume fund P is ranked in the k-th place according to benchmark model i's alpha (i.e., according to $\alpha_{(i,P)}$), and assume that the fund is ranked in the l-th place according to benchmark model i's net economic alpha (i.e., according to $\alpha_{(NET,i,P)}$). If $|k - l| < 0.1N$, i.e., the ranking difference is less than 10% of the N existing funds, then we, somewhat arbitrarily, denote that there is a trivial ranking difference between benchmark models i's alpha and benchmark model i's net economic alpha. On the other hand, If $|k - l| \geq 0.1N$, then there is a nontrivial ranking difference between benchmark model i's alpha and benchmark model i's net economic alpha.

Panel B reports the 2001–2003, 2004–2006, and 2007–2009 average frequency of ranking differences between the alphas and net economic alphas. No ranking differences are found for the one-factor models, and minor ranking differences exist for the three-, four-, and five-factor models (0.2–0.5%).

For Panels C and D analysis, we narrow the data sample and keep only funds ranked as 20 best performing funds or as 20 worst performing funds. Each subperiod, each benchmark model yields two lists of funds included in the 20 best performing funds: one list based on benchmark model i's alpha and the other based on benchmark model i's net economic alpha. Similary, two lists of worst performing funds are generated. Then, we count how many funds appear on both benchmark method i's alpha and benchmark method i's net economic alpha lists of best (worst) performing funds, and report the numbers in Panel C (D).

In Panel C, the lists of 20 best performing funds mostly overlap. In fact, for one-factor models, the lists of 20 best funds according to alpha and net economic alpha fully overlap, i.e., the exact same 20 funds appear on both lists. For the three- to five-factor models overlap is also overwhelming, with an average overlap (across the three subperiods) of about 19 (out of 20) funds. In Panel D, very similar results are recorded regarding the list of 20 worst performing funds. Again, the impression is that net economic alpha's impact is minute.

In summary, the evidence in this section suggests that, in our data, alphas and net economic alphas yield almost identical inference, even when multifactor models and models that include pricey factors such as the long–short portfolios are employed. This can also be interpreted as encouraging because it suggests that previous research, based solely on raw alphas, is probably correct, i.e., similar results would be obtained with net economic alphas. Nevertheless, it is arguable that since net economic alphas are the more "decent" and "fair" measures to judge fund's performance, future studies should examine them as well. One should also recall that in Table 2, the net economic alpha analysis suggests that the mean underperformance of mutual funds (relative to ETFs) is slight and that the mutual fund industry relative performance improved over time. Thus, the net economic alphas, computed for the first time in our paper, facilitate some important economic inference.

4.3 Some perspective: differences due to benchmark model choice

To put the net economic alpha findings (Tables 2 and 3 results) into perspective, we examine in this

Not for Distribution

subsection only traditional alphas and record differences in fund's classification (as good or poor) and in fund's relative ranking that are due to the benchmark model choice. Are the differences in inference about fund's performance caused by the benchmark model choice, larger or smaller than the differences in inference due to the net economic alpha refinement?

Panel A of Table 4 reports the 2001–2003, 2004–2006, and 2007–2009 average frequency of funds classification differences between the benchmark models.

The frequency of fund classification differences varies between 6% and 30.8%. Minor classification disagreements exist between the three-factor and the four-factor models (6%). Substantial classification differences emerge when comparing the CAPM and the three-, four-, and five-factor models (28.1–30.8%). The style-factor model generates also substantial (18.9–25.0%) classification differences relative to the three- to five-factor models.

Next, we test the null hypothesis that there is no relation between the funds' classification according to the different benchmark models. If classification is independent across models, all average frequencies of classification differences, reported in Panel A, should equal 50%. We find that in Panel A all classification difference frequencies are lower than 0.5 at the 1% significance level. Thus, there is a positive dependence in funds' classification across our five benchmark models—most of the time our benchmark models tend to agree on a fund's classification.

Panel B of Table 4 reports the 2001–2003, 2004–2006, and 2007–2009 average frequency of ranking differences between the benchmark models. The methodology is analogous to that employed in Panel B of Table 3. The ranking-difference frequency varies between 12.9% and 62.9%. Small ranking differences emerge when comparing the three- and four-factor models (12.9%). Substantial ranking differences exist when comparing all other models, especially when comparing the CAPM with all other models (56.5–62.9%).

We also generate the lists of 20 "best" funds and 20 "worst" funds according to the five benchmark models using a methodology analogous to that in Panels C and D of Table 3.

In Panel C of Table 4, the average overlap of the 20 best performing funds varies between 5.7 and 16 funds. (Note that the average is computed across our three subperiods.) The highest overlap is obtained when comparing the 20 best funds lists of the three- and four-factor models. The lowest overlap of 5.7 out of 20 best performing funds is obtained when comparing the CAPM with the five-factor model.

In Panel D of Table 4, the average ranking overlap of the 20 worst performing funds vary between 9 and 17.3. The highest overlap (17.3) emerges when comparing the 20 worst funds lists of the three- and four-factor models. The lowest overlap (9) occurs when comparing the style-factor model and the four-factor model.

In sum, the evidence in this subsection shows that fund's classification and ranking depend significantly on the benchmark-system used for their evaluation. Differences in fund's classification and ranking are substantial even for somewhat similar specifications of the benchmark models. This evidence helps interpret Table 3 results. The differences in Table 4, due to benchmark model selection, are much sharper and an order of magnitude higher than the differences in Table 3, due to the net economic alpha introduction. The conclusion that using alphas instead of net economic alphas only slightly distorts the results is reinforced.

Not for Distribution

Table 4 Differences in classification and ranking across five possible benchmark models.

Panel A. Classification differences among the benchmark models

	M (%)	S (%)	3F (%)	4F (%)	4F-S (%)
M	0.0	24.9	28.8	30.8	28.1
S	24.9	0.0	21.9	25.0	18.9
3F	28.8	21.9	0.0	6.0	13.0
4F	30.8	25.0	6.0	0.0	12.5
4F-S	28.1	18.9	13.0	12.5	0.0
Average*	28.1	22.7	17.4	18.6	18.1

*Average does not include signs differences between model i and itself.

The five benchmark models (specified in Table 3) are fitted in three subperiods: 2001–2003, 2004–2006, and 2007–2009 using the samples of funds with full-return data. In this table we abstract from net economic alphas, i.e., only traditional alphas are computed and compared. Panel A reports the frequency of differences in funds' classification between the five benchmark models, i.e., the frequency of cases where benchmark model i's alpha is negative while benchmark model j's alpha is positive.

Panel B. Rank differences (of more than a decile) among the benchmark models

	M (%)	S (%)	3F (%)	4F (%)	4F-S (%)
M	0.0	62.9	56.5	59.8	59.5
S	62.9	0.0	51.0	55.1	44.4
3F	56.5	51.0	0.0	12.9	36.4
4F	59.8	55.1	12.9	0.0	31.8
4F-S	59.5	44.4	36.4	31.8	0.0
Average*	59.7	53.4	39.2	39.9	43.0

*Average does not include differences between model i and itself.

Panel B reports the frequency of 10% differences in funds' rankings between each pair of the five different benchmark models.

Panel C. Comparing the 20 best performing funds

	M	S	3F	4F	4F-S
M	20.0	7.7	9.3	7.7	5.7
S	7.7	20.0	9.7	8.3	8.0
3F	9.3	9.7	20.0	16.0	11.7
4F	7.7	8.3	16.0	20.0	12.7
4F-S	5.7	8.0	11.7	12.7	20.0
Average*	7.6	8.4	11.7	11.2	9.5

*Average does not include numbers on the diagonal (all of which are 20).

Panel C reports averages, across our three subperiods, of the number of funds that are in the list of 20 best performing funds both in the column and in the row models (= common best funds).

Not for Distribution

Table 4 (*Continued*)

Panel D. Comparing the 20 worst performing funds

	M	S	3F	4F	4F-S
M	20.0	13.0	11.3	11.7	11.3
S	13.0	20.0	10.3	9.0	11.0
3F	11.3	10.3	20.0	17.3	13.7
4F	11.7	9.0	17.3	20.0	15.0
4F-S	11.3	11.0	13.7	15.0	20.0
Average*	11.8	10.8	13.2	13.3	12.8

*Average does not include numbers on the diagonal (all of which are 20).

Panel D reports averages, across our three subperiods, of the number of funds that are in the list of 20 worst performing funds both in the column and in the row models (= common worst funds).

5 Summary and conclusion

Using a sample of over 1000 non-specialized open-end equity funds' during 2001–2009, we estimate mutual funds' alphas via five possible benchmark models.

Our main goal is to correct for a bias in traditional alpha estimation—mutual fund return is taken net of expenses while benchmark return is gross of expenses. This traditional practice leads to underestimation of the true alpha of a mutual fund. Following a comment on the issue by Fama and French (2009), we examine a more precise and just alpha measure—the net economic alpha. The net economic alpha is computed by adding to the traditional alpha, the cost of mimicking the fund's systematic risk via ETFs. Theoretically, the net economic alpha should be used for evaluating the real contribution of a fund manager to her investors, since it takes into account the true alternative cost of replicating funds' risk (see Fama and French, 2009, on the bottom of page 6).

This study is the first to compute net economic alphas. Comparing the standard and the net economic alphas, little differences in funds classification and ranking emerge, even for the multifactor models and for models that include pricey factors such as the long–short portfolios. Thus, in our data, the impact of the net economic alpha is minute.

An interesting observation of the study is that the (across models) mean net economic alpha becomes less negative over time; and in the last subperiod the mean net economic alpha is even slightly positive. This may imply that the mutual fund industry is only slightly inferior to ETFs and that the mutual fund industry is becoming more and more competitive with time relative to ETFs. This appears the central "macro" implication of our study, and it should be verified in future studies.

Acknowledgments

We thank Simon Benninga, Doron Kliger, Roger Otten, Haim Reisman, Avi Wohl, Journal of Investment Management reviewer, and participants of the EFMA Symposium on Asset Management (Hamburg, 2012) for their constructive comments. All remaining errors are our own.

Not for Distribution

Appendix: Net economic alphas in APT frameworks—A formal derivation

Assume, for simplicity, an APT model with one factor—the market (factor M). According to the APT model

$$E(R_P) - R_f = \beta_{P,M}(E(R_M) - R_f). \quad (7)$$

Let the realized excess return of portfolio P be

$$\tilde{R}_P - \tilde{R}_f = \beta_{P,M}(\tilde{R}_M - \tilde{R}_f) + \alpha_{P,M} + \tilde{\epsilon}_{P,M}. \quad (8)$$

We define the last term in the regression ($\alpha_{P,M} + \tilde{\epsilon}_{P,M}$) as the return on the portable alpha portfolio, portfolio X.[6] Portfolio X is a zero investment portfolio that has zero covariance with the benchmark. In equilibrium, $\alpha_{P,M}$ – the expected return of X – should be zero. However, in practice, $\alpha_{P,M}$ is a result of fund manager's activity and talent—a measure of performance.

Suppose that the expense ratio for the benchmark M, defined as the cost of investing in an ETF mimicking the market index M, is a percentage ϕ_M of assets, and that the expense ratio for the fund P is a percentage ϕ_P of assets. Then, the cost of mimicking the benchmarked part of the managed portfolio P appears to be $\phi_M \beta_{P,M}$. However, note that $\phi_M \beta_{P,M}$ implies that when beta is negative the fee for the benchmarked part of the managed portfolio P is negative. This contradicts the reality that a short position, taken by the fund manager relative to factor M (a negative $\beta_{P,M}$), costs at least as much as a long position in M. Thus, if beta with respect to the market, factor M, is negative, the cost of replicating this systematic (benchmarked) part of the portfolio is positive, and should be estimated as $|\beta_{P,M}|\phi_M$. Given that the systematic position of the fund can be replicated at a cost of $\phi_M|\beta_{P,M}|$, the fee charged by the fund on the active (nonsystematic) part of the portfolio equals $\phi_P - \phi_M|\beta_{P,M}|$.[7]

Recall that in the traditional regressions for estimating alpha we use the net of fee excess returns (based on the net of fee prices of the mutual fund) as the dependent variable. This implies that the standard methodology of regressing $\tilde{R}_P - \tilde{R}_f$ on $\tilde{R}_M - \tilde{R}_f$ estimates the net of fee alpha ($\alpha_{P,M}$). Hence, the gross (before fees) excess return of the fund (raw alpha) is $\alpha_{P,M} + \phi_P$.

If we deduct from the raw alpha the fee charged on the active part, we obtain $\alpha_{NET,P,M}$, the net economic alpha:

$$\alpha_{NET,P,M} = \alpha_{P,M} + \phi_P - (\phi_P - \phi_M|\beta_{P,M}|). \quad (9)$$

Thus,

$$\alpha_{NET,P,M} = \alpha_{P,M} + \phi_M|\beta_{P,M}|. \quad (10)$$

Since $\alpha_{NET,P,M}$ takes into account the cost of mimicking the fund's risk, it better represents the fund manager's contribution to her investors.

Similarly, when we consider an APT model with N factors, then

$$\tilde{R}_P = \beta_{P,0} + \sum_{i=1}^{N} \beta_{P,i}\tilde{R}_i + \tilde{\epsilon}_P, \quad (11)$$

and an analogous equation to (10) can be derived:

$$\alpha_{NET,P,N} = \alpha_{P,N} + \sum_{i=1}^{N} \phi_i|\beta_{P,i}|. \quad (12)$$

Another interesting and intuitively appealing point is that $\alpha_{NET,P}$, the net economic alpha, can be approximated by the intercept in a regression of net of fees mutual fund excess return on net of fees factors excess returns. A net of fee factor excess return is the excess return on an ETF that mimics the factor minus the fees of such an ETF. Proofs are available from the authors.

Notes

1 We can show that this procedure is practically equivalent to regressing net of fees mutual fund returns on the net

Not for Distribution

of fees returns on ETFs that mimic the benchmark model factors.

2 Mutual funds' expense ratio includes management fee and 12b1 fee and might also include other expenses. It does not include loads or transaction fees.

3 Five main components make up the ETF expense ratio: index licensing fee, trustee fee, board of directors, custodian, and accountants's fees.

4 Due to the competitiveness of the ETF market, expense ratios are almost identical across similar ETFs. Hence, our specific ETF choice does not affect our main results and conclusions.

5 It may be possible to save on some of these ETF expenses by rebalancing the investor portfolio. However, we prefer to be conservative and examine the most direct and expensive way of mimicking factor returns.

6 Portable alpha strategies have been discussed by Kung and Pohlman (2005), Ezrati (2006), and others.

7 We assume negligible costs for the money market component of the portfolio.

References

Carhart, M. M. (1997). "On Persistence in Mutual Funds Performance," *Journal of Finance* **52**, 57–82.

Cesari, R. and Panetta, F. (2002). "The Performance of Italian Equity Funds," *Journal of Banking and Finance* **26**, 99–126.

Dellva, W. L. (2001). "Exchange-Traded Funds Not for Everyone," *Journal of Financial Planning* **14**, 110–124.

Ezrati, M. (2006). "Portable Alpha: A Glimpse at the Future," *Journal of Investing* **15**, 119–123.

Fama, E. F. and French, K. R. (1996). "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance* **51**, 55–84.

Fama, E. F. and French, K. R. (2009). "Luck versus Skill in the Cross-Section of Mutual Fund Alpha Estimates," Working Paper, University of Chicago.

Grinblatt, M. and Titman, S. (1989). "Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings," *Journal of Business* **62**, 393–416.

Hunter, D., Kandel, E., Kandel, S., and Wermers, R. (2009). "Endogenous Benchmarks," Working Paper, University of Maryland.

Jensen, M. C. (1968). "The Performance of Mutual Funds in the Period 1945–1964," *Journal of Finance* **23**, 389–416.

Khorana, A., Servaes, H., and Tufano, P. (2009). "Mutual Fund Fees Around the World," *Review of Financial Studies* **22**, 1279–1310.

Kosowski, R. (2011). "Do Mutual Funds Perform when It Matters Most to Investors? US Mutual Fund Performance and Risk in Recessions and Expansions," *Quarterly Journal of Finance* **1**, 607–664.

Kostovetsky, L. (2003). "Index Mutual Funds and Exchange-Traded Funds," *Journal of Portfolio Management* **29**, 80–92.

Kung, E. and Pohlman, L. (2005). "An Alpha + Beta Framework," *Journal of Investing* **14**, 9–16.

Otten, R. and Bams, D. (2007). "The Performance of Local versus Foreign Mutual Fund Managers," *European Financial Management* **13**, 702–720.

Pastor, L. and Stambaugh, R. F. (2003). "Liquidity Risk and Expected Stock Returns," *Journal of Political Economy* **111**, 642–685.

Keywords: Mutual funds; performance evaluation

Not for Distribution