

The Alpha of a Market Timer*

Georges HÜBNER[†]

HEC Management School – University of Liège, Belgium
School of Business and Economics, Maastricht University, the Netherlands
Gambit Financial Solutions Ltd, Belgium
EDHEC, France

December 2010

Abstract

Portfolio managers claim to be able to generate abnormal returns through either superior asset selection or market timing. The Treynor and Mazuy (TM) model is the mostly used return-based approach to isolate market timing skills, but all existing corrections of the regression intercept can be manipulated by a manager who can trade derivatives. We revisit the TM model by applying the original option replication approach proposed by Merton. We exploit both the linear and the quadratic coefficients of the TM regression to assess the replicating cost of the cheapest option portfolio with the same convexity. The application of the new correction on two samples of market timing funds delivers particularly encouraging empirical results. The portfolio replication approach reveals that the performance of market timing funds increases with their convexity level, and the effect is larger and significant for positive market timers. All other classical corrections of the TM model underestimate the necessary adjustment for the fund's convexity, leaving positive timers with negative performance and vice-versa. This bias explains the converging conclusion of most studies based on the TM model that market timers do not outperform the market. Furthermore, inadequate correction methods weaken the link between the magnitude of market timing and the associated performance. Such results suggest that a correction of alpha based on an arbitrage argument clarifies the role of market timing in the generation of performance.

Keywords: Performance measurement, market timing, Treynor and Mazuy, option replication, mutual fund performance.

JEL codes: G10, G12

*I thank Thomas Lejeune and Arnaud Cavé for research assistance and Deloitte (Luxembourg) for financial support. All errors remaining of mine.

[†]Corresponding author. University of Liège, HEC Management School, Department of Finance. Rue Louvrex 14, Bldg N1, B-4000 Liège, Belgium. Phone: +32 42327428. E-mail: G.Hubner@ulg.ac.be

1 Introduction

The assessment of the performance of actively managed funds is an ongoing theoretical and empirical challenge. According to the hypothesis of efficient financial markets, active fund managers should not take advantage of proprietary skills that would generate reproducible positive abnormal returns. For that purpose, traditional performance measures such as the Sharpe (1966) and Treynor (1965) ratios, respectively in the total and systematic risk frameworks, were initially designed to capture performance of portfolios delivering stationary returns in a mean-variance environment. When non-linear portfolio strategies became very popular amongst hedge fund managers, and as the benchmarking of these funds often reveals to be a difficult task, some measures have been developed aiming at capturing these nonlinearities. Some influential ones are the Sortino ratio (Sortino and Price, 1994) and the Omega (Keating and Shadwick, 2002).

In the mutual fund literature, a benchmark can generally be identified. It either leads the specification of a control portfolio, or to regression-based procedures for the estimation of required portfolio returns. Portfolio managers claim to be able to generate returns above their benchmark through superior selectivity of securities or asset classes (asset selection), or by successfully anticipating market movements (market timing). There has been a clear tendency to design tests aiming at detecting the ability to perform superior asset selection. Performance measures derived from Jensen's alpha and its multi-factor extensions have been extensively used in this context, as evidenced in the comprehensive study by Kothari and Warner (2001). To date, only two return-based approaches have been used on a regular basis for the detection of market timing skills: the piecewise linear regression of Henriksson and Merton (1981) (henceforth HM) and the quadratic regression of Treynor and Mazuy (1966) (henceforth TM).¹ Empirical evidence based on these measures have regularly delivered a disappointing picture regarding market timing skills (see a.o. Kryzanowski *et al.*, 1996; Becker *et al.*, 1999; Bollen and Busse, 2004; Comer *et al.*, 2009). However, a more detailed holding-based approach introduced by Jiang *et al.* (2007) concludes that the average performance attributable to timing skills is positive, and could amount up to an annual 0.6%. The extent of this abnormal return is economically relevant but this contrasting results with pervasive evidence is also empirically puzzling.

As emphasized by Ingersoll *et al.* (2007), it is extremely easy to manipulate all performance measures listed above by altering the distribution of returns, or by dynamically trading securities to curb the distribution of returns. Such option-like characteristics displayed by the pattern of portfolio returns has long been explicitly associated with spurious market timing effects (Jagannathan and

¹A noteworthy exception is the Positive Period Weighting Measure proposed by Grinblatt and Titman (1989), but it has lead to few further developments in the literature on fund performance.

Korajczyk, 1986). The mere separation between the regression intercept and the market timing coefficient in the HM and the TM models is not sufficient to discriminate between genuine and false market timing skills. For instance, Lee and Rahman (1990) separate asset selection and market timing by considering that the alpha fully represents the former dimension. To better identify pure market timers, three types of corrections have been proposed: a variance correction approach (Grinblatt and Titman, 1994); an approximation based on the squared benchmark returns (Bollen and Busse, 2004); and a synthetic option pricing approach (Merton, 1981). All three methods, however, remain prone to manipulations because a manager who has access to a complete derivatives market can easily alter the market timing coefficient without affecting the regression intercept (alpha) to a proportional extent. Based on this statement, Ingersoll *et al.* (2007) propose a manipulation-proof performance measure, but its identification is contingent on the characterization of the investor's preferences.

In this paper, we revisit the Treynor and Mazuy model by applying, almost literally, the original option replication approach proposed by Merton in the context of the HM model. Unlike all previous attempts to alter the TM regression alpha on the basis of the quadratic term alone, we exploit information on *both* the linear *and* the quadratic coefficients of the regression to assess the replicating cost of the option for a passively managed portfolio with non-linear benchmark exposures. The intuition goes as follows. If the manager has free access to a complete derivatives market on its benchmark, as suggested by Ingersoll *et al.* (2007), she can indeed curb the payoff of her portfolio in many possible ways. Considering the benchmark case where asset return distributions are stationary and Gaussian, we can always approximate the returns of the option portfolio up to the second (quadratic) order through a Taylor series expansion. Then, only the option delta, gamma and theta matter. Anyone can mix these option "greeks" to get the desired linear and quadratic sensitivities, but the same greeks will also entirely drive the periodic holding cost of this portfolio. Amongst all portfolios involving derivatives that mimic a fund's pattern of returns, the cheapest one provides the minimum performance required from the manager who claims to time the market. The actual performance of the active fund, mixing the impact of asset selection and market timing skills, is obtained by difference. To illustrate our approach, we show that most market timing performance measures would mistakenly assign nonzero performance to a pure passive fund involving options, while our correction properly sets the adjusted performance of an uninformed portfolio to zero.

This approach has the potential to overcome the performance manipulation issue. We derive a proper performance adjustment in a preference-free setup or, more precisely, in an economy where preferences are consistent with mean-variance analysis. The availability of a complete market enables us to carry out an arbitrage argument, unlike in Ingersoll *et al.* (2007) whose correction involves replicating a non-traded option on the quadratic market return, resulting in market incompleteness.

There is a potential lack of precision to our approach however. We create a portfolio whose quadratic sensitivity to the benchmark returns reproduces the pattern of the TM regression. But both equations involve a noise component: the specification error in the TM regression, and the approximation error in the Taylor series expansion of the replicating portfolio. The former error can be reduced by extending the reasoning to a multivariate setup, with the advantage that one does not need to identify interaction terms as in the extension of the TM model proposed by Lehmann and Modest (1987). The latter issue is more tricky, because the sign and magnitude of the error cannot be determined *a priori*. Furthermore, the quest for the cheapest-to-replicate option might lead to situations where the cheapest option displays unacceptable approximation errors. We carry out an experiment whose results suggest that the choice of option maturing in 6 months provide a reasonable trade-off between approximation error and replication cost when returns are measured on a weekly interval.

The application of the new correction delivers particularly encouraging empirical results. From a sample of 1,262 equity mutual funds, we focus on two subsamples of funds with positive and negative convexity of their market sensitivity, respectively. If market timing skills exist, one would expect the performance of convex portfolios to increase with their convexity level. Concave funds that voluntarily adopt a contrarian stance might deliver superior performance to compensate for their gamma risk, but one would expect the emphasis of market timing skills to be much less pronounced because of the spurious effect induced by positive cash inflows when markets are up, which mechanically reduce the fund's beta when markets are up (Warther, 1995; Edelen, 1999). The application of the option portfolio replication approach yields the predicted pattern of results. All other attempts to correct the TM model with existing adjustment methods dramatically fail. They underestimate the necessary correction for the fund's convexity, leaving positive timers with negative adjusted performance and vice-versa. This bias explains the converging conclusion of most studies based on the TM model that market timers do not outperform the market. Besides the issue of their biased level, inadequate corrections appear to weaken the link between the magnitude of market timing and the associated performance. Such results suggest that a correction of alpha based on an arbitrage argument clarifies the role of market timing in the generation of performance. This will undoubtedly improve the diagnostics regarding the market timing skills of portfolio managers.

The paper is organized as follows. The second section introduces the alternative approaches to account for market timing in performance evaluation, including the option replicating portfolio correction proposed in this paper. In section 3, we discuss the adjustments necessary to cope with estimation or approximation issues, and deliver associated potential solutions. Section 4 presents the results of an empirical study on a sample of mutual funds, with which the impacts of alternative adjustment methods are tested. Section 5 concludes.

2 The option-based approach to market timing

In the scope of the one-factor model for asset returns, the literature identifies two major specifications allowing to expose a manager's market timing abilities. The Henriksson and Merton (1981) model considers that the manager switches the portfolio's beta depending on the sign of the market return. A good market timer increases the market exposure when the return is positive, and keeps it lower otherwise. Treynor and Mazuy (1966) propose to add a quadratic term to the one-factor model. The coefficient of this term reflects the convexity achieved by the manager in her exposure to the market portfolio. If it is positive, the manager gradually increases her beta as the market goes up, indicating that she displays good timing abilities. The regression equations of the Henriksson and Merton (1981) and the Treynor and Mazuy (1966) models are represented in equations (1) and (2), respectively:

$$r_t = \alpha_{\text{HM}} + \beta_{\text{HM}} r_{mt} + \gamma_{\text{HM}} (-r_{mt}^+) + \varepsilon_t \quad (1)$$

$$r_t = \alpha_{\text{TM}} + \beta_{\text{TM}} r_{mt} + \gamma_{\text{TM}} r_{mt}^2 + \varepsilon_t \quad (2)$$

where $r_t \equiv R_t - R_f$ is the portfolio excess return over the risk-free rate, r_{mt} is the market portfolio excess return, and $-r_{mt}^+ = \max(-r_{mt}, 0)$.

The HM model translates the behavior of a manager who succeeds in switching her market beta from a high level equal to β_{HM} when the market return exceeds the risk-free rate to a low level of $(\beta_{\text{HM}} - \gamma_{\text{HM}})$ otherwise. Admati *et al.* (1986) show that under the standard assumption of a joint normal distribution of asset returns, the TM model is consistent with a manager whose target beta varies linearly with her forecast for the expected market rate of return. In both models, a negative value of gamma induces negative market timing.

The literature has proposed many extensions of these two models in the context of multiple risk factors or benchmark portfolios², but so far there has been no widely adopted alternative approach to market timing performance evaluation under a single factor specification.

2.1 The classical view of performance as an additive measure

The HM model and the TM models aim to capture timing and selectivity skills using one single equation. Under both specifications, the constant of the regression is supposed to capture the asset selection skills of the manager. The coefficient of the second variable (the truncated positive market return for HM and the squared market return for TM) reveals her ability to time the market, but it does not deliver a measure of excess return due to market timing by itself. So far, the literature

²See Pesaran and Timmermann (1994) for the HM model, and Bollen and Busse (2001, 2004), Comer (2006), Comes *et al.* (2006), Comer *et al.* (2009) for recent extensions of the TM model.

on mutual fund performance has produced three alternative ways to measure the contribution of market timing to active returns: one based on the variance of market returns, another one of the squared market returns, and a final one based on the value of an option.

The work of Admati *et al.* (1986) characterizes the properties of an active market timing portfolio managed optimally if returns are multivariate normal and the representative investor exhibits a constant absolute risk aversion (CARA) utility function. It is assumed that the manager, when receiving information about market returns, linearly adjusts her beta to the timing signal. They show that the optimal portfolio returns are characterized by a quadratic function that can be measured by equation (2). In this particular instance of the TM model, the reward for the manager's market timing ability is adequately represented by the variance of the market portfolio returns (Grinblatt and Titman, 1994), leading to equation (3) for the total return attributable to the manager's performance, denoted π :

$$\pi_{\text{TM},v} = \alpha_{\text{TM}} + \gamma_{\text{TM}}\sigma_m^2 \quad (3)$$

where σ_m^2 is the variance of the market portfolio returns.

Coles *et al.* (2006) and derive the corresponding formulation for the performance return in the HM model:

$$\pi_{\text{HM},v} = \alpha_{\text{HM}} + \gamma_{\text{HM}}(\bar{r}_m^+ - f_{r_m>0}\bar{r}_m) \quad (4)$$

where $\bar{r}_m^+ = \frac{1}{T} \max(r_{mt}, 0)$ and $f_{r_m>0}$ denotes the frequency of positive market returns.³

In the framework of Admati *et al.* (1986), the identification market timing premium is a step towards the identification of the coefficients of the regression, which corresponds to the optimal timing portfolio when the portfolio manager makes linear beta adjustments to timing signals (Grinblatt and Titman, 1994). This framework is thus restrictive. It entails an analytical connection between the regression coefficients α_{TM} and β_{TM} in equation (2).

A simpler approach is proposed by Bollen and Busse (2004) and Comer *et al.* (2009). They build upon the fact that, in the HM and TM model, a perfect market timer *ex post* would make her beta time-varying according to market conditions by setting, respectively, $\beta_{\text{HM},t} = \beta_{\text{HM}}\mathbf{1}_{\{r_{m,t}>0\}}$ and $\beta_{\text{TM},t} = \beta_{\text{TM}} + \gamma_{\text{TM}}r_{m,t}$. Accordingly, the total performance of a manager combining asset selection and market timing skills results from averaging the periodic market returns in the model with time-varying betas:

$$\pi_{\text{HM},a} = \alpha_{\text{HM}} + \gamma_{\text{HM}}\bar{r}_m^+ \quad (5)$$

$$\pi_{\text{TM},a} = \alpha_{\text{TM}} + \gamma_{\text{TM}}\bar{r}_m^2 \quad (6)$$

³The authors transform the original HM model by estimating the regression $r_t = \alpha_{\text{HM}} + \beta_{\text{HM}}r_{mt} + \gamma_{\text{HM}}r_{mt}^+ + \varepsilon_t$, but the outcome remains the same regarding the determination of the manager's total performance.

Equations (5) and (6) are simplified versions of (4) and (3), respectively. The advantage of this formulation is that it is no longer restricted to correspond to an optimal behavior as in the Admati *et al.* (1986) framework, but the cost is a potentially high upward bias in the performance as measured with these metrics. When returns are computed with a daily frequency, the difference between the corresponding equations does not appear to matter empirically (Krimm *et al.*, 2008). However, as the expected market return becomes significant with longer measurement intervals, the upward bias in equations (5) and (6) may start to matter.

All three formulations for π share a common additive structure. They leave the α unaffected, considering it as the contribution of selectivity to performance. The market timing coefficient multiplied by a premium constitute the second term of the sum. Note that the market beta does not intervene explicitly in the formula. In the Admati *et al.* (1986) approach, this coefficient represents an output of the model. It does not appear in the empirical approach of Bollen and Busse (2004) either.

In the next section, we defend an alternative approach that explicitly accounts for the role of the linear market beta in total performance with market timing. As for the original option-based approach proposed by Merton (1981), we associate the market timing return to an option. Our analysis focuses on the cost, expressed in returns, of replicating the pattern of the HM or TM regression using a mix of options and risk-free instruments only.

2.2 The replicating cost of building market timing portfolios

Both the Admati *et al.* (1986) and the Bollen and Busse (2004) frameworks use an *ex post* approach to the detection of market timing. They aim at directly measuring the contribution of market timing skills in the total portfolio return, irrespective of how the manager tried to achieve this result. When considered *ex ante*, the skill of the manager is related to her capacity to create an option on the market portfolio.

2.2.1 The Henriksson and Merton model reinterpreted

The idea of calling upon option trading strategies to assess the market timing abilities of a portfolio manager is introduced by Merton (1981). In the context of the HM model – developed in his sequel paper coauthored with Henriksson – he shows that the portfolio manager’s global performance, combining her timing and selectivity abilities, expressed in realized returns and denoted π , can be represented as

$$\pi_{\text{HM},o} = \alpha_{\text{HM}} + \gamma_{\text{HM}} e^{R_f \Delta t} P(M, \Delta t, e^{R_f \Delta t}) \quad (7)$$

where $P(M, \Delta t, e^{R_f \Delta t})$ is a put with remaining time to maturity equal to the time interval and strike price of $e^{R_f \Delta t}$ written on the market portfolio M , whose price is normalized to 1

The reasoning underlying Merton's result is that the portfolio return obtained in equation (1) can be achieved by taking at the same time a long position of β_{HM} in the underlying index and of γ_{HM} in a put on the same index that will only pay off if the index return is lower than the riskless rate. The remaining amount $1 - \beta_{\text{HM}} - \gamma_{\text{HM}}$ is invested (if positive) or borrowed (if negative) at the risk-free rate. The cost of adopting such a strategy is the initial put premium, which is a lump sum amount $P(S, \Delta t, e^{R_f \Delta t})$ per unit of gamma, and that should be capitalized to represent the realized opportunity cost of the strategy versus a linear one.

Ingersoll *et al.* (2007) adapt (7) to the TM framework and get

$$\pi_{\text{TM},o} = \alpha_{\text{TM}} + \gamma_{\text{TM}} e^{2R_f \Delta t} \left(e^{\sigma_m^2 \Delta t} - 1 \right) \quad (8)$$

where R_f is the continuous interest rate and σ_m^2 is the volatility of the market portfolio returns. The authors interpret the second term of this expression as the payoff for the fraction γ_{TM} of a derivative security that pays the square of the excess market return. At the same time, Ingersoll *et al.* (2007) emphasize that this definition of total performance is easily prone to manipulation.

There are two major differences between the original Merton (1981) interpretation of the HM model and the transformation proposed by Ingersoll *et al.* (2007). The first one is that, once we do not strictly follow the Admati *et al.* (1986) optimization framework, the value of γ_{TM} in the TM formulation can be fully independent of the level of β_{TM} . A market timer can indeed freely choose her beta and, at the same time, achieve a level of performance that depends on her skill (or luck) in timing the market. In the HM model this is not so: the level of γ_{HM} reflects the difference between the high market beta and the low market beta. Therefore the value of β_{HM} is present, although implicitly, in equation (7). The second reason for the difference is that the Merton interpretation rests on the cost of replicating a protective put on the market. The quadratic option proposed in equation (8) is not tradable, and so the reasoning is held in a context of incomplete markets. Below, we follow the very same reasoning as Merton (1981) under the scope of the TM model.

2.2.2 The Treynor and Mazuy model revisited

Following the idea of Ingersoll *et al.* (2007), we examine the properties of a passive portfolio strategy that would exhibit a pattern of returns similar to equation (2). In the absence of any managerial skill, there is a direct connection between the levels of beta and gamma coefficients on the one side, and the intercept of the regression on the other side. The difference between the resulting "passive" level of alpha and the actual alpha delivers the performance attributable to the manager.

Consider a simple self-financing investment strategy consisting in creating a long position in an index with a positive or a negative convexity in returns and lending or borrowing at the risk-free rate. A position involving a long call option written on this index has a positive delta and a positive gamma. Similarly, a position involving a short put option has a positive delta and a negative gamma. In principle, one can find an option whose time-to-maturity and moneyness match the desired sensitivities to the underlying index.

Formally, rewrite equation (2) of the actively managed portfolio to replicate in terms of total returns:

$$\begin{aligned} R_t - R_f &= \alpha_{\text{TM}} + \beta_{\text{TM}} (R_{mt} - R_f) + \gamma_{\text{TM}} (R_{mt} - R_f)^2 + \varepsilon_t \\ \Rightarrow R_t &\simeq \alpha'_{\text{TM}} + \beta_{\text{TM}} R_{mt} + \gamma_{\text{TM}} R_{mt}^2 + \varepsilon_t \end{aligned} \quad (9)$$

where $\alpha'_{\text{TM}} = \alpha_{\text{TM}} + (1 - \beta_{\text{TM}}) R_f$ and the second line follows from that $R_f^2 \simeq 0$ and $R_{mt} R_f \simeq 0$.

To reproduce the pattern of (9) with a positive convexity, we create a passive portfolio consisting in a long position $w > 0$ in a call option on the index S , the remainder of the portfolio $(1 - w)$ being held in the risk-free asset. Unlike in the HM model, in which the maturity and moneyness of the option are constrained by the return generating function, the option characteristics are left open. As before, we normalize to 1 the price of the market portfolio at time t .

Using the second order Taylor series expansion, the rate of return during the interval Δt of the portfolio can be expressed using the "greeks", i.e. the partial derivatives of the call with respect to selected variables:

$$R_t^{(\tau, \kappa)} = w \left(\Delta_{\tau, \kappa} R_{mt} + \frac{1}{2} \Gamma_{\tau, \kappa} R_{mt}^2 + \Theta_{\tau, \kappa} \right) + (1 - w) R_f + o(\Delta t) \quad (10)$$

where τ and κ are the time-to-maturity and the strike price (expressed as a multiple of the spot price) of the option, respectively, and $\Delta_{\tau, \kappa} \equiv \frac{\partial C(M, \tau, \kappa)}{\partial M}$, $\Gamma_{\tau, \kappa} \equiv \frac{\partial^2 C(M, \tau, \kappa)}{\partial M^2}$ and $\Theta_{\tau, \kappa} \equiv \frac{\partial C(M, \tau, \kappa)}{\partial t}$ are the option delta, gamma and theta. Note that, in the context of constant interest rates and market volatility over the time interval, the remaining greeks (rho and vega) do not show up. The remaining term $o(\Delta t)$ results from the higher orders of the Taylor series expansion and from the potential rho and vega effects.

Besides the residual term, equations (9) and (10) look very similar. The linear and quadratic coefficients are equal if the following conditions set of is satisfied:

$$\beta_{\text{TM}} = w \Delta_{\tau, \kappa} \quad (11)$$

$$\gamma_{\text{TM}} = w \frac{1}{2} \Gamma_{\tau, \kappa} \quad (12)$$

In general, the system of two equations with three unknowns $\Delta_{\tau, \kappa}$, $\Gamma_{\tau, \kappa}$ and w is indefinite, as there exists a continuum of pairs (τ, κ) that satisfy conditions (11) and (12). Nevertheless,

as this portfolio is a passive one, the objective of this replicating portfolio is to minimize the cost of replication, i.e. to maximize the constant rate of return, that we call $\alpha^{(\tau,\kappa)} = w_{\tau,\kappa}\Theta_{\tau,\kappa} + (1 - w_{\tau,\kappa})R_f$. The solution of this problem is given by solving the following program:

$$\alpha^{(\tau^*,\kappa^*)} = \max_{\tau,\kappa} (w_{\tau,\kappa}\Theta_{\tau,\kappa} + (1 - w_{\tau,\kappa})R_f) \quad (13)$$

$$\text{s.t. } \frac{2\Delta_{\tau^*,\kappa^*}}{\Gamma_{\tau^*,\kappa^*}} = \frac{\beta_{\text{TM}}}{\gamma_{\text{TM}}} \quad (14)$$

where $w_{\tau,\kappa} \equiv \frac{\beta_{\text{TM}}}{\Delta_{\tau,\kappa}} > \beta_{\text{TM}}$.

The outperformance of the active portfolio that yields the returns (9) over the replicating portfolio is obtained by subtracting the value of $\alpha^{(\tau^*,\kappa^*)}$:

$$\begin{aligned} \pi_{\text{TM}}^* &= \alpha'_{\text{TM}} - \alpha^{(\tau^*,\kappa^*)} \\ &= \alpha_{\text{TM}} + (w_{\tau^*,\kappa^*} - \beta_{\text{TM}})R_f - w_{\tau^*,\kappa^*}\Theta_{\tau^*,\kappa^*} \end{aligned} \quad (15)$$

$$= \alpha_{\text{TM}} + \beta_{\text{TM}} \left[\left(\frac{1}{\Delta_{\tau^*,\kappa^*}} - 1 \right) R_f - \frac{\Theta_{\tau^*,\kappa^*}}{\Delta_{\tau^*,\kappa^*}} \right] \quad (16)$$

As $w_{\tau^*,\kappa^*} > \beta_{\text{TM}}$ and $\Theta_{\tau^*,\kappa^*} < 0$ (Merton, 1981), both terms between brackets are positive. The first one reflects the incremental risk-free return resigned by the passive portfolio through investing a proportion in options higher than the original exposure to the market index. The second term, as in Merton (1981), stands for the time value of the option lost because of the passage of time. In order to generate a neutral performance, the value of α_{TM} in the regression ought to be negative.

The case of a negative convexity involves $\gamma_{-\text{TM}} < 0$ (where the index -TM means negative market timer) bears a similar analysis. The replicating portfolio involves going short an amount of $-w$ puts on the index and investing $1 + w$ in the riskless asset. The performance of the "contrarian" market timer, denoted $\pi_{-\text{TM}}^*$, is obtained by the following equation:

$$\pi_{-\text{TM}}^* = \alpha_{-\text{TM}} - (w_{\tau^*,\kappa^*} + \beta_{-\text{TM}})R_f + w_{\tau^*,\kappa^*}\Theta_{\tau^*,\kappa^*} \quad (17)$$

where, this time, the second and third term are both negative. In order to post a zero aggregate performance, the manager must deliver a positive alpha from the TM regression.

2.3 Specification issues

The portfolio replicating approach of the TM quadratic return generating process creates a potential bias related to the difference in residual terms between equations (9) and (10). There are two potential issues: the specific risk in the quadratic equation and the approximation risk in the Taylor series expansion.

2.3.1 Specific risk in the quadratic return generating model

The replication process presented in the TM model suffers from the lack of identification of the residual term. An actively managed portfolio producing a quadratic pattern of returns such as in (9) is likely to bear a significant specific risk, represented by ε_t in the equation. The portfolio replication approach entails the presence of a numerical approximation risk, which is not to be confused with specific risk. The time variation of $o(t)$ in (10) is also limited as this term reflects a bias rather than a noise component. Unless the magnitude of specific risk in the TM equation is sufficiently low, which would statistically correspond to a high significance level of the quadratic regression, the replicating portfolio cannot be associated to any close to a no-arbitrage relation. The performance measure π_{TM}^* might not faithfully reflect the true performance of the market timer in this context.

Therefore, the specification issue has a particular importance in this context. One has to ensure that the return generating process employed reaches a very high explanatory power. The proposed multi-factor extensions of the original market model employed by Treynor and Mazuy are of interest.

Bollen and Busse (2004) apply the Fama and French (1993) - Carhart (1997) four-factor model to get a more precise return-generating process, and they add a single term to reflect the ability to anticipated variations of the market returns. Comer (2006) and Comer *et al.* (2009) extend further this specification to a set four bond indexes retrieved from the Blake *et al.* (1993) 6-factor bond model. This overlay enables them to capture the potential risk exposures of hybrid funds, with a reported average R-squared reaching more than 98% for the two samples they study.

The work of Lehman and Modest (1987) provides the foundations for extending the TM market timing model to a multi-factor setting. Considering that the manager can anticipate the variations of a number K of indexes, they propose to add the same number of squared returns to the regression, but also all the two-by-two interaction terms. This would leave us with a number $K(K-1)/2$ market timing terms to compute, with presumably a high number of insignificant explanatory variables, inducing issues of overspecification. In his 8-factor model (4 for stocks and 4 for bonds), Comer (2006) defends a more parsimonious solution. He discards all the interaction terms, and considers two market timing factors: one for the stock market and one for the bond market.

Our portfolio replicating approach bears a multi-factor extension. Consider that the linear return generating specification features K risk factors, but only a subset $L \leq K$ are prone to a market timing behavior. For each of these L factors, we can isolate the linear and quadratic sensitivities and apply the same reasoning as in the previous section. For each underlying index, the weight to be invested in options is first determined. Once all the weights are set, they are summed up with the linear betas on the $K-L$ factors with no market timing, and the remainder of the portfolio

value is invested in the risk-free asset. Formally, the model writes:

$$R_t \simeq \alpha + \left(1 - \sum_{i=1}^K \beta_i\right) R_f + \sum_{i=1}^K \beta_i R_{it} + \sum_{i=1}^L \gamma_i R_{it}^2 + \varepsilon_t \quad (18)$$

and the associated market timing adjusted performance is computed as:

$$\pi^* = \alpha + \sum_{i=1}^L (w_{\tau_i^*, \kappa_i^*} - \beta_i) R_f - \sum_{i=1}^L w_{\tau_i^*, \kappa_i^*} \Theta_{\tau_i^*, \kappa_i^*} \quad (19)$$

which is a straight generalization of equation (15) to a multi-index model. Note that we do not require any correlation structure between the risk factors, as each replicating portfolio is considered in isolation.

2.3.2 Approximation risk in the portfolio replication model

Equation (10) holds approximately when portfolio rebalancing occurs at sufficiently high frequency. When returns are measured on a biweekly or a monthly basis, for instance, the risk of the hedged portfolio can become substantially large (Gilster, 1997). But besides the issue of rebalancing frequency, the Taylor series approximation is not likely to mimic portfolio dynamics in a satisfactory manner for any kind of option. Gilster (1990) documents that replicating portfolios for near-to-maturity options, when rebalanced in discrete-time, exhibit significant systematic risk with the underlying index returns. Considered in the TM quadratic regression framework, this would influence the linear beta of the replicating portfolio over the underlying returns. This beta would not simply correspond to the product $w\Delta_{\tau, \kappa}$ as in equation (11), but would be greater, in spite of the correction brought by the quadratic term in the regression. To control for this bias, it is necessary to make sure that the correlation between the approximation error and the index returns is limited.

The option maturity choice must be related to the cost of the replicating portfolio. Even though this is not generally warranted, the pair (τ^*, κ^*) that maximizes expression (13) is likely to combine high moneyness and low maturity in the case of a call option. But whenever the relation between the portfolio replicating cost and the maturity is monotonically positive, the optimal solution is given by setting the option maturity equal to the return measurement frequency, i.e. $\tau^* = \Delta t$. This case collapses to the Henriksson and Merton framework, in which the market timer creates an option that exactly matures at the next point in time. Formally, the Taylor series approximation of the HM model up to the second order is the same as equation (10), but the approximation error is not acceptable at any reasonable tolerance.

The selection of the option moneyness can also lead to unacceptable return patterns. Presumably, the cheapest long-term call option to replicate with a target ratio of delta over gamma, as in equation (14), will also display a low moneyness. If it is too small, the corresponding pattern of returns will

display a large number of very small returns (when the index returns are negative or reasonably positive) and a small number of very large returns (when the index returns are largely positive, leading the option to become in-the-money or close). The best quadratic function matching this pattern could correspond to the TM specification with the desired parameters, but the fit of the curve would be very poor.

Therefore, the strike price and time-to-maturity of the cheapest to replicate option should also respect some tolerance regarding the quality of the quadratic fit regarding the individual coefficients. Combined with the systematic risk of the residual of the approximation, and considering that residuals are i.i.d. and normally distributed, this translates into supplementing the optimization problem (13) with two additional constraints:

$$\left| \beta_{\text{TM}} - \beta^{(\tau^*, \kappa^*)} \right| \leq \text{tol}_\beta \quad (20)$$

$$\left| \gamma_{\text{TM}} - \gamma^{(\tau^*, \kappa^*)} \right| \leq \text{tol}_\gamma \quad (21)$$

where $\beta^{(\tau^*, \kappa^*)}$, $\gamma^{(\tau^*, \kappa^*)}$ and η_t are obtained by the reverse quadratic regression of the replicating portfolio returns on the underlying index:

$$R_t^{(\tau^*, \kappa^*)} = c^{(\tau^*, \kappa^*)} + \beta^{(\tau^*, \kappa^*)} R_{mt} + \gamma^{(\tau^*, \kappa^*)} R_{mt}^2 + \eta_t \quad (22)$$

As they only involve individual parameter values, the constraints are fully compatible with the multi-factor specification (18). The aim of the replicating portfolio is to mimic the linear and quadratic portfolio exposure to a single index through an option written on this very index, the other exposures being accounted for by other options. Hence constraints (20) and (21) can be used in conjunction with (19).

It is also important not to impose restrictions on the significance of the reverse quadratic regression compared to the TM specification. The latter model represents the return generating process of an actively managed portfolio, that potentially mixes asset selection and market timing skills. If the manager puts strong emphasis on selectivity, the residual term ε_t of the quadratic regression will presumably be volatile, and specific risk will be high.⁴ As the reverse quadratic regression reproduces a mere fitting exercise on passive portfolio returns, there is no expected connection with the significance of the original market timing model.

The reverse regression is obviously an important quality check of the option replicating approach. The difference between equations(22) and (10) taken at point (τ^*, κ^*) yields the following inequality:

$$\left| c^{(\tau^*, \kappa^*)} - \alpha^{(\tau^*, \kappa^*)} \right| \leq \text{tol}_\beta |\bar{R}_m| + \text{tol}_\gamma \bar{R}_m^2 = \text{tol}_\alpha \quad (23)$$

⁴Nothing in our setup precludes the use of the appraisal ratio to assess the risk-adjusted performance of the manager. It simply entails replacing the regression alpha by expression (19) at the numerator.

If the linear quadratic coefficients are arbitrarily close to the original TM values, then the constant of the regression $c^{(\tau^*, \kappa^*)}$ will also approach the theoretical value of the constant rate of return $\alpha^{(\tau^*, \kappa^*)}$. The difference between $c^{(\tau^*, \kappa^*)}$ and $\alpha^{(\tau^*, \kappa^*)}$ is an empirical matter. Expression $\alpha^{(\tau^*, \kappa^*)}$ represents an ex-ante cost of replicating the market timer's portfolio with options, and corresponds to the original Merton (1981) interpretation. After the option replicating strategy has been systematically implemented, the outcome of regression (22) depends on the realizations of the index returns. This may induce deviations from the effects foreseen by the Taylor series expansion. For instance, a very large outlier may have incurred during the estimation period. Effects of a higher order than two, captured neither by the delta nor by the gamma, might then become significant in option returns. They impact the estimates of $\beta^{(\tau^*, \kappa^*)}$ and $\gamma^{(\tau^*, \kappa^*)}$ in the quadratic fitting equation. Being the intercept of this regression, $c^{(\tau^*, \kappa^*)}$ represents the ex post return of the option portfolio.

3 The price of the market timer's convexity

Having presented an alternative characterization of the performance of a market timer, we start this section with a comparative analysis of each performance measure proposed in the literature. We focus on the Treynor and Mazuy (1966) model. The purpose of this section is not to study the performance of individual funds. Rather, our goal is to identify whether and to what extent these performance measures would detect performance when it does not exist, i.e. when the portfolio is truly a passive one. After this illustrative but instructive exercise, we analyze the characteristics of the option that best replicates the quadratic pattern of portfolio returns; "best" meaning here the cheapest option that generates returns that are sufficiently close to the ones of the active portfolio under review.

3.1 The performance of a passive portfolio

As in the Ingersoll *et al.* (2007) study, we adopt the point of view of a passive portfolio manager who decides to replicate the behavior of a market timer. For this purpose, she constructs a simple strategy involving only the riskless asset and call and put options. The manager attempts to replicate a quadratic regression similar to (9) with an excellent fit. Since the portfolio involves neither any asset selection nor any market timing skills, it should post zero aggregate performance. A performance metric that reports positive or negative performance in these circumstances could thus be manipulated.

We provide an illustration that replicates actual market conditions in three cases of positive convexity (long a call) in the return generating process. Every week from January 1999 till September

2008⁵, we construct a portfolio of a constant weight w invested in an option on the S&P500 index, and the rest being invested in the 3-months T-Bill. One week later, the option and money market positions are unfolded at market price, and a new analogous portfolio is created with the same weights. We make the simulation as realistic as possible by pricing the options under the Black-Scholes-Merton formula using the GARCH(1,1) forecasted volatility of the market index as input. Besides, we account for the weekly variations in the T-Bill rate to compute the return on the money market position.

For each case, we fix the moneyness, time-to-maturity and portfolio weight of the option as constants. Every week, we recompute the delta, gamma and theta of the newly created option, and get an average over the whole sample period. With the time series of portfolio returns, we build the TM regression with total returns and estimate the α_{TM} , β_{TM} and γ_{TM} coefficients. From these values, we compute the four performance measures designed around the TM model: $\pi_{TM,v}$ (Grinblatt and Titman, 1994); $\pi_{TM,a}$ (Bollen and Busse, 2004); $\pi_{TM,o}$ (Ingersoll *et al.*, 2007) and π_{TM}^* or π_{-TM}^* , as respectively proposed in equations (15) and (17). To improve precision, we adopt the weekly return on the risk-free asset as the risk-free rate. Table 1 reports the results of this illustrative application.

Insert Table 1 here

The options have been chosen so as to display three levels of maturity and moneyness. The weight invested in options is set at $w = 60\%$. The choice of portfolios for the illustration must be such that the ratio $\frac{2\Delta}{\Gamma}$, which is known *ex ante* from the choice of option characteristics, remains very close to the *ex post* ratio $\frac{\beta_{TM}}{\gamma_{TM}}$ retrieved from the quadratic regression of portfolio total returns on the market index. Otherwise, the Taylor series expansion presented in equation (10) behaves too remotely from the regression (9). Under the market conditions that were prevailing during the sample period, it appears that only options with a strong convexity displayed this property⁶. The difference between the two ratios is limited, as it oscillates from 2 to 19% relative to each other.

The linear alpha for each portfolio obtained with the market model is also reported. As expected from passive portfolios with a positive convexity, the alpha is negative and significant in all three cases. The alpha from the quadratic regression is deeply negative, and requires a strong correction in order to avoid the threat of an easy manipulation. The additive corrections brought by the traditional approaches are too limited. The adjustment brought by taking the time value of the

⁵We remove the financial crisis from this analysis as the condition on the good fit (high R^2) of the quadratic regression cannot be satisfied when the extreme returns of October to December 2008 are left in the sample. Keeping these returns makes the link between the quadratic beta and the option convexity too loose for a proper treatment, whatever the option maturity and moneyness considered.

⁶Detailed results are available upon request.

option into account in the expression for π_{TM}^* are of an adequate order of magnitude. It leaves a performance value reasonably close too zero. Although these examples are illustrative and are not meant to provide statistical evidence, they emphasize at least that there are circumstances, when the quadratic mode specification is good, where the association of option strategies to convex portfolio returns *à la* Treynor-Mazuy brings useful insight regarding their true return generating process.

3.2 The choice of the right option

We discuss the theoretical features of the best replicating option, then assess the empirical limitations affecting this choice. The discussions focuses on a portfolio that achieves positive market timing, bearing in mind that a similar reasoning would hold for the mirror case of a negative market timer, which involves the sale of a put.

3.2.1 Option choice without constraints

Implementing a strategy that consists of replicating a portfolio with a long call option involves the choice of this option. As the underlying asset is determined by the selection of the index in the TM model, the choice collapses to setting the moneyness κ and the time-to-maturity τ of the option. The contract must respect a constraint, namely the target level of the ratio of the delta over its gamma in equation (14). Then, amongst all eligible options, the best is the one that minimizes the cost of replication, i.e. that maximizes expression $w_{\tau,\kappa}\Theta_{\tau,\kappa} + (1 - w_{\tau,\kappa})R_f$. Unfortunately, the function to maximize depends on the pair (τ, κ) through the option theta, but also through the weight invested in the option in the passive portfolio $w_{\tau,\kappa} \equiv \frac{\beta_{\text{TM}}}{\Delta_{\tau,\kappa}}$.

The partial derivatives of delta and gamma with respect to time are usually called Charm and Color, respectively (Garman, 1992). In the Black-Scholes-Merton world, they bear an analytical form and their behavior is well-known. Unfortunately, even in such a controlled environment, their sign is erratic. Haug (2003) shows an example where the Charm is negative for ITM and positive for OTM calls, but at the same time the Color is negative for near ATM and positive for far OTM or ITM options. Overall, the evolution of the ratio of delta over alpha over time (and so their derivative with respect to time-to-maturity) is indeterminate.

We illustrate in Figure 1 the sets of parameters that reach different target values of the ratio $\frac{\beta_{\text{TM}}}{\gamma_{\text{TM}}} = \frac{2\Delta_{\tau,\kappa}}{\Gamma_{\tau,\kappa}}$ for options that are priced under the Black-Scholes-Merton model. We take as inputs the average 3-month T-Bill rate and conditional volatility of the S&P500 weekly returns over the Jan. 1999 – Sept. 2008 period, namely 0.066% and 2.289%, respectively. We plot the contour lines corresponding to target values of this ratio taking multiples of 1/3, ranging from 0 to 3. The ratio

represents the fraction of beta over gamma in the TM model; as the average market beta is equal to one, we span in principle values of gamma starting from 1/3 on. We set the range of maturities to 1 to 52 weeks, and the range of moneyness ratios from 0.80 to 1.00.

Insert Figure 1 here

Interpreting the ratio $\frac{2\Delta_{\tau,\kappa}}{\Gamma_{\tau,\kappa}}$ as a reverse indicator of curvature, figure 1 shows that only ITM options (i.e. whose strike $\kappa < 1$) provide potentially meaningful convexity. For long maturity options, the progression of the ratio remains gradual. With a one-year maturity option, it takes a moneyness of ca. 90% to get a beta equal to half the option gamma or, by identity, a TM beta equal to its gamma. On the other hand, the ratio evolves very quickly with short maturity options. When the maturity approaches one week, i.e. the lowest maturity that exceeds the frequency of returns estimation, the value of the ratio literally explodes when the moneyness gets lower than 93%.

The next step of the analysis is to find out the cheapest-to-replicate option among the ones that respect a target ratio value. Following the same example as above, we set $\beta_{\text{TM}} = 1$ and let the value of $\frac{\beta_{\text{TM}}}{\gamma_{\text{TM}}}$ equal to 2 (blue line), 1 (red line) and 0.67 (green line). They correspond, respectively, to the contour lines between the dark and light green regions (ratio = 2), sky and light blue regions (ratio = 1) and light and dark blue regions (ratio = 0.67) in figure 1. The γ_{TM} equals 0.5, 1.0 and 1.5, which are reasonable values for a market timer as shown in our empirical study. For each feasible pair (τ, κ) , we compute the cost of the option replicating portfolio $-\alpha^{(\tau,\kappa)} = -w_{\tau,\kappa}\Theta_{\tau,\kappa} - (1 - w_{\tau,\kappa})R_f$. The lower this cost, the cheaper it is to replicate the option.

Insert Figure 2 here

The replication cost increases with the level of γ_{TM} . This is the natural consequence of increase the convexity of the portfolio payoff, which is done at the expense of the option theta. The comparison of the three lines shows quite small differences between the patterns of the cost function. For $\gamma_{\text{TM}} = 0.5$ (blue line), the cost increases from 1.420% to 1.613%. The start and end points are 2.853% and 3.319% for $\gamma_{\text{TM}} = 1$, and 4.293% and 5.105% for $\gamma_{\text{TM}} = 1.5$, respectively. Thus, the cost increases slightly less than proportionally with the value of gamma and with maturity.⁷

In all illustrated cases, the option replication cost is an increasing function of option maturity. Even though this result cannot be generalized (because option Charm and Color have undefined signs), our realistic example shows that it can happen. It means that, in absence of any constraint on specific or approximation risk control, the cheapest-to-replicate option might have a maturity of one period. As it matches the frequency of returns computation, the option-based portfolio produces the

⁷For a given value of gamma, the cost rises proportionally to the level of beta regardless of the maturity.

returns of the HM model, and is unlikely of being adequately estimated using the TM specification in the reverse regression. Hence, our case illustrates the need to assign constraints (20) and (??) for the reverse regression of option returns.

3.2.2 Option choice with approximation constraints

As indicated above, the option choice without constraints might realistically indicate that the shorter to maturity, the cheaper the option replication. The reverse quadratic regression (22) restricts the feasibility of maturity reduction because the quality of the fit naturally deteriorates as the option maturity decreases. The goal of this subsection is to detect, within the same setup as in the former example, the range of option maturities for which the approximation error induced by the Taylor series expansion is "acceptable", i.e. falls within the tolerance bounds for the alpha, beta and gamma retrieved from the quadratic regression.

We set portfolio beta equal to one and adopt the same set of TM gammas as before, namely 0.5, 1.0 and 1.5. The sample period is Jan. 1999 - Sept. 2008 and we create portfolios with a quadratic exposure on the S&P500 index. In order to ensure the correspondence between the Black-Scholes option prices and the behavior of the time series of index returns, we posit a flat weekly volatility of 2.375%. By using the sample standard deviation of returns in option prices, we avoid introducing a pricing bias in the estimation of regression (22). For each feasible pair (τ, κ) , a portfolio is constituted every week by investing a weight $w_{\tau, \kappa}$ in the option at a price $C(M, \tau, \kappa)$ and $(1 - w_{\tau, \kappa})$ in the risk-free asset. The following week, the option is sold at a price $C(M, \tau - 1, \frac{\kappa}{(1+R_m)})$, the risk-free return is booked on the remaining part, and the portfolio is rebalanced. We estimate the reverse quadratic regression by applying the TM specification to the returns of this portfolio.

We report in Figure 3, the evolution of the difference between the intercept of the reverse regression and the analytical portfolio alpha, i.e. $c^{(\tau, \kappa)} - \alpha^{(\tau, \kappa)}$ (Figure 3a), the reverse regression beta (Figure 3b), and the reverse regression gamma (Figure 3c), as a function of the option time-to-maturity from 1 to 30 weeks. The values for $\gamma_{TM} = 0.5, 1.0$ and 1.5 are printed in green, red and blue, respectively.

Insert Figures 3a, 3b and 3c here

While the analytical replicating cost $-\alpha^{(\tau, \kappa)}$ increases with the option maturity, Figure 3a shows that the regression intercept $c^{(\tau, \kappa)}$ becomes closer to zero as time-to-expiration rises. Because of the poor regression fit for near-maturity options, $c^{(\tau, \kappa)}$ starts at a very negative level (from -6.24% to -8.58%). As the regression significance level increases with maturity, the intercept gradually approaches zero. In the cases illustrated by Figure 3, the two functions intersect at maturities equal to 23 weeks ($\gamma_{TM} = 0.5$), 14.5 weeks ($\gamma_{TM} = 1.0$) and 10.5 weeks ($\gamma_{TM} = 1.5$). Figures 3b and 3c

indicate that the convergence of the coefficients for the linear and quadratic term asymptotically converge to their theoretical values. The speed of convergence typically decreases after a 5-week maturity. The linear coefficient ($\beta^{(\tau,\kappa)}$) remains very close to its target value of 1, with a distance becoming lower than 0.01 when the maturity exceeds 20 weeks. As expected from the imperfect fit of the second order approximation of the option return, the value of $\gamma^{(\tau,\kappa)}$ remains more remote. The coefficient estimate remains systematically upward biased with respect to its target value. The explanation has to be found in the small variability of the observed squared market returns ($\sigma(R_m^2) = 0.90\%$ on a yearly basis) as compared to the returns themselves ($\sigma(R_m) = 17.13\%$). In this OLS setup, the smaller variation in the independent variable translates into a larger standard deviation in the estimated coefficient.

In order to get a more rigorous analysis of the set of option maturities that support reasonable coefficient values for the reverse quadratic regression, we apply equations (20) and (21) to our data set. To reflect the volatility levels of the independent variables, the tolerance levels are adjusted by setting $tol_\beta = \theta/\hat{t}_{R_m}$ and $tol_\gamma = \theta/\hat{t}_{R_m^2}$ for different values of a constant θ . This yields naturally $tol_\alpha = \theta \left(\frac{|\bar{R}_m|}{\hat{t}_{R_m}} + \frac{\bar{R}_m^2}{\hat{t}_{R_m^2}} \right)$ by applying equation (23). We set θ to 0.006, 0.004 and 0.002. These values are chosen so as to produce usable maturity intervals and to analyze how they shrink as the tolerance level decreases. The results are displayed in Table 2.

Insert Table 2 here

For each value of $\Gamma_{\tau,\kappa}$ and θ , the table reports the maturity intervals that respect the tolerance level for each parameter. The last row displays the intersection between these intervals. Interestingly, the intervals become thinner as the convexity of returns diminishes. The regression intercept yields the most severe constraint on the upper bound of the interval because, as shown in Figure 3a, the reverse regression intercept tends to become too large for longer maturities, while its theoretical value is supposed to decrease. For $\theta = 0.006$, a large set of option maturities are acceptable, while the interval becomes an empty set for too low convexity and tolerance levels ($\Gamma_{\tau,\kappa} = 0.5$ and $\theta = 0.002$). Overall, maturities between 18.9 and 26.8 weeks comply with most intervals: any maturity fits for $\Gamma_{\tau,\kappa} = 1.5$, all from 24.8 to 26.8 weeks for $\Gamma_{\tau,\kappa} = 1.0$, and from 18.9 to 23.8 weeks for $\Gamma_{\tau,\kappa} = 0.5$.

The goal of this subsection is to identify the most satisfactory trade-off between the regression fit and the accuracy of the intercept. The simulations indicate that, for reasonable values of the option convexity gamma of the replicating option, maturities around six months induce the best match between the regression results and the Taylor series expansion results of the option replicating strategy.

4 Empirical evidence on market timing revisited

We carry out an empirical analysis with a focus on funds that seem to exhibit a market timing behavior. Our major objective is to determine whether the performance attributable to these funds would be substantially altered by the correction for the option replication cost. To achieve this goal, we perform two analyses. We first apply the Treynor and Mazuy model on a set of market timing funds and examine how various ways to correct the alpha for the timing ability of the manager affect their performance assessment. In a second stage, we study the determinants of the adjusted performance, and attempt to detect whether there exists a genuine "market timing skill" that can be pointed out for different types of funds.

The analysis is performed on a sample of 2,521 U.S. equity mutual funds denominated in U.S. dollars with weekly returns spanning the period January 1999 - September 2008 (508 observations). The choice of weekly data represents a compromise between the superior ability to detect market timing effects with higher frequency data (Bollen and Busse, 2001) and the recent evidence of higher potential bias due to benchmark misspecification with the use of daily fund returns (Coles *et al.*, 2006). The comparability of the parameter estimates is warranted by retaining the 1,242 funds with a full return history over the period.⁸ The fund characteristics and returns are retrieved from *Bloomberg*. To ensure consistency across the various model specifications, all index returns as well as the risk-free rate are extracted from Kenneth French's online data library.

4.1 Reestimating the Treynor-Mazuy performance

The Treynor and Mazuy specification in excess returns, corresponding to equation (??), is estimated on each individual fund during the period using the White heteroskedasticity-consistent estimation method. We further select the "best" market timers by identifying the 100 funds displaying the highest t-stat for the estimated quadratic coefficient (γ_{TM}).⁹ We call it the "High" subsample. Likewise, the 100 "worst" market timers are those displaying the 100 most negative values of the student statistic for the same coefficient γ_{TM} . We call this subsample the "Low" one. The descriptive statistics for the global sample and the three subsamples are reported in Table 3.

Insert Table 3 here

The adoption of the Treynor and Mazuy specification does not bring a sensitive improvement over the market model (MM) overall. Only when the market timing is pronounced, the adjusted

⁸As the objective is not to perform a comprehensive assessment of mutual fund performance, the resulting survivorship bias does not affect the output of the study.

⁹We rejected two funds for which the beta was insignificant or significantly negative.

R-squared increases by 0.5% on average, while the estimated alpha of market timers naturally decreases in absolute value by accounting for the funds' convexity.

The TM statistics for the individual sub-samples High and Low substantially differ from the overall sample on all criteria except for the market beta, which is remarkably similar across samples. Positive market timers record slightly negative excess returns, negative alphas and strong positive gammas. The exact opposite holds for the Low set. Overall, the estimates are more favorable for the Low subsample regarding both excess and abnormal returns, but the negative convexity of the TM regression is more pronounced as well. The average excess return, alpha and beta from the whole sample are close to the ones of the combined High+Low sub-samples featuring the 100 best and 100 worst market timers. However, because the absolute values of alpha and gamma of market timing funds are often greater and more significant than for the rest of the sample, the High+Low sub-sample exhibits highly dispersed values for these parameters.

Henceforth, we focus on the High, Low, and High+Low subsamples. For each fund, the replicating option is priced under the Black-Scholes-Merton formula. As we deal with retrospective performance evaluation, we use the in-sample stock market volatility and average interest rate. They are set to a yearly 17.16% and 3.495%, respectively. Consistently with our discussion in the previous section, a constant option maturity of 0.5 year (26 weeks) is taken for all options. The option moneyness κ_i for each fund i then equates $\frac{2\Delta_{0.5,\kappa_i}}{\Gamma_{0.5,\kappa_i}} = \frac{\beta_{TM,i}}{\gamma_{TM,i}}$. In case the estimated coefficient $\gamma_{TM,i}$ is positive, the alpha is adjusted through the cost of replicating a call option as in equation (15). If the gamma is negative, the replicating portfolio involves a put option, and the alpha is corrected according to equation (17).

From the original estimate of $\alpha_{TM,i}$, two types of adjustments are tested: the variance correction $\gamma_{TM,i}\sigma_m^2$ (Grinblatt and Titman, 1994), and the cost of the replicating option equal to $(w_{0.5,\kappa_i} - \beta_{TM,i})R_f - w_{0.5,\kappa_i}\Theta_{0.5,\kappa_i}$ for a call ($\gamma_{TM,i} > 0$) or $-(w_{0.5,\kappa_i} + \beta_{-TM,i})R_f + w_{0.5,\kappa_i}\Theta_{0.5,\kappa_i}$ for a put ($\gamma_{TM,i} < 0$).¹⁰ The statistics of the corrections and of the resulting performance metrics $\pi_{TM,v}$ and π_{TM}^* are reported in Table 4. For comparison purposes, we also reproduce the statistics for the alphas generated by applying the market model (α_{MM}) as well as those obtained by applying the Fama and French (1993) - Carhart (1997) four-factor model with the size, value and momentum portfolios (α_{4F}).¹¹ The table also provides information on the frequency of significantly positive or negative estimates of performance. For the adjusted values $\pi_{TM,v}$ and π_{TM}^* , we use bootstrapped standard errors.

¹⁰The corrections proposed by Bollen and Busse (2004) and Ingersoll et al. (2007) produce very similar results to the variance correction approach, and are not reported to save space. Detailed results are available upon request.

¹¹We do not present results with one or several market timing coefficients to the four-factor model as this analysis would exceed the scope of this study. Detailed results are available upon request.

Insert Table 4 here

The variance correction approach performs only a limited average adjustment over the initial TM alpha. As a result, the adjusted alpha of market timers tends to reflect the opposite sign of their gamma: a low performance for positive market timers, and a high one for negative market timers. The application of this technique would lead to the conclusion that a large number (almost one third) of the fund managers belonging to the High sample underperform the market. The diagnosis would be more extreme for negative market timers: out of the 100 managers, almost one half (45%) would be supposed to beat the market. The corrected alphas are very close to the initial alphas measured with the market model.

By contrast, the application of the portfolio replication adjustment leads to a large correction, which even results in switching the sign of alpha for the Low subsample. On average, the application of this technique delivers a negative performance for both the positive and negative market performers, even though the average performance is very close to zero in the High subsample. 8.5% the market timers would deliver significant positive alphas, and only 5.5% would display underperformance. These figures are closest to the ones obtained with the four-factor (4F) model, even though the portfolio replication approach would be more generous in the High sample (17 outperformers against only 3 for the 4F model). The four-factor model concludes more often to abnormal (positive or negative) performance with the negative market timers. By contrast, none of them would be supposed to beat the market when the π_{TM}^* measure is used.

The explanation of such a large observed difference between the correction techniques probably comes from the inability of the variance of market returns to reflect the cost (or benefit) of replicating a deep OTM option. By applying a linear penalty or reward for the manager's convexity parameter γ_{TM} , the traditional variance correction technique underestimates the correction to be applied. The theta of the replicating option more properly accounts for the price of convexity. The phenomenon is particularly noticeable for the Low sample, characterized by very negative values of the quadratic term in the TM regression. The variance correction approach yields an average abnormal performance of 5.22% per annum, which is highly unlikely to truly reflect the actual skills of this group of managers. The picture supplied by the option portfolio replication method is much more consistent with the evidence of spurious market timing emphasized by Warther (1995) and Edelen (1999), according to whom negative market timing coefficient mostly result from a fund flows explanation.

Besides its magnitude, the variability of the correction induced by the portfolio replication approach exceeds the one of the variance correction as well. The reason underlying this higher volatility has to be found in the explicit account for the fund's beta (and therefore its cross-sectional

dispersion in each sample) in the portfolio replication technique.

Table 4 also suggests that alphas corrected with the portfolio replication approach are typically slightly negative, even more than the four-factor alphas. Figure 4 provides a synthetic representation of the distribution of performance measure estimates for the High and Low subsamples through a box-and-whisker graph.

Insert Figure 4 here

In both graphs, the distributions of $\pi_{TM,v}$ and α_{MM} appear identical, while the main characteristics of the empirical distributions of α_{4F} and π_{TM}^* are quite similar. We focus on this second finding. Figure 4a (High subsample) shows a similar median, but greater dispersion and, in particular, right skewness of the distribution of π_{TM}^* . In Figure 4b (Low subsample), the distribution of π_{TM}^* is shifted down with respect to the one of α_{4F} . Together with the diverging evidence on significance levels reported in Table 4, these findings suggest that the four-factor model and the Treynor-Mazuy model with the portfolio replication adjustment deliver relatively close ranges of alphas, but they are distributed quite differently across funds. This calls for further analysis of the comovements between each fund’s various performance metrics, and possibly the identification of common patterns that would explain the fund managers’ performance levels. This is done in the next subsection.

4.2 Revisiting the market timing skills

After having justified the relevance of the portfolio replication approach to adjust the performance of a market timer, it is relevant to consider the drivers of this performance. We first examine whether a clear pattern of performance generation can be extracted from the data. This is done through a correlation analysis, as shown in Table 5.

Insert Table 5 here

In both samples, the correlation between $\pi_{TM,v}$ and π_{TM}^* is very high, as expected. However their relationship with the original TM performance metrics α_{TM} and γ_{TM} is very different.

When the variance correction is applied, the convexity effect induced by γ_{TM} either vanishes (High subsample) or negatively correlates with abnormal performance (Low subsample). Thus, for positive market timers, the value of $\pi_{TM,v}$ is almost entirely driven by the TM alpha. Because in the second case γ_{TM} is always negative, the negative correlation implies that a higher market timing effect goes along with higher performance. This can be interpreted as a natural consequence of the inability of the variance adjustment to translate very high values of gamma into an adequate penalty.

The correlation structure of performance measures in the High and Low samples vary to a large extent. The properties of π_{TM}^* for negative market timers are similar, although slightly less pronounced, to the ones of $\pi_{\text{TM},v}$: very high influence of α_{TM} , moderate but significant impact of γ_{TM} . On the other hand, the correlation between the variance-corrected performance $\pi_{\text{TM},v}$ and the initial alpha estimated with the market model α_{MM} is almost perfect. This finding suggests that negative market timers do not "time" the market at all; they merely get exposure to the market portfolio and behave as if they sell call options to enhance their returns through the time value of the option premium. What evidence of Table 4 shows is that the traditional ways to adjust the TM alpha when the gamma is negative are powerless: they simply lead back to the original linear alpha, which is largely positive. They provide a false impression of abnormal performance. In this context, the main value-added of the portfolio replication approach lies in the magnitude of the correction, not in its discrimination between market timing vs. asset selection skills. On average, performance vanishes when a proper correction is applied, as indicated by the Low sample mean of -1.83% shown in Table 4.

In the High subsample however, the covariation between π_{TM}^* and α_{TM} becomes moderate, while the effect of γ_{TM} on π_{TM}^* is stronger, at least in values (the rank correlations are almost identical). The performance metrics obtained with linear asset pricing models, α_{MM} and α_{4F} , hardly covary with the portfolio replication-adjusted performance π_{TM}^* . This finding can be related to the graphical evidence of Figure 4: even though the means of α_{4F} and π_{TM}^* are close, their marginal distributions are remote from each other.

We specifically examine the relation between each fund's risk- and market timing-adjusted performance and its indicator of market timing. The rationale for this investigation is a straightforward consequence of the adjustment in the market timing context. Indeed, if fund managers attempt to time the market, one can expect that the ones who record a significant abnormal performance *indeed* successfully timed the market. In other terms, the market timing skill hypothesis suggests that the adjusted alpha is an increasing function of the absolute value of gamma. Oppositely, when gamma is close to zero, the performance of fund managers should not show any significant relationship with the fund's convexity. Because funds that display no market timing behavior should not be rewarded for their pretended market timing skills, any evidence of a relationship between gamma and a performance measure would indicate an imperfect adjustment for market timing, resulting in a biased estimate of performance.

The jointly test these hypotheses, we carry out a piecewise-linear regression, whereby we regress each fund i 's estimate of performance on its value of $\gamma_{\text{TM},i}$. The candidate performance measures are the adjusted values π_{TM}^* , $\pi_{\text{TM},v}$ and, in order to have a linear benchmark case for validation,

the four-factor performance α_{4F} . For each performance measure, the regression to test is:

$$\text{perf}_i = b_0 + b_1\gamma_{\text{TM},i} + b_2 \max(\bar{g} - \gamma_{\text{TM},i}, 0) + b_3 \max(\underline{g} - \gamma_{\text{TM},i}, 0) + v_i \quad (24)$$

where $\text{perf}_i = \pi_{\text{TM},i}^*$, $\pi_{\text{TM},v,i}$ and $\alpha_{4F,i}$, and $\underline{g} < 0 < \bar{g}$ are the negative and positive angular points, respectively. The hypotheses to test are:

$$H_{01} : b_1 = 0 \text{ and } b_1 - b_2 - b_3 = 0$$

$$H_{02} : b_1 - b_2 \neq 0$$

$$H_A : (b_1 > 0 \text{ or } b_1 - b_2 - b_3 < 0) \text{ and } b_1 - b_2 = 0$$

To properly reflect market timing skills (H_A), the measure should (i) display a significant slope coefficient when market timing is pronounced, either through positive or negative convexity, or both; and (ii) not display a significant slope coefficient when timing is absent, as this would indicate biasedness in the adjustment to the original Treynor-Mazuy alpha. If the market timing skills hypothesis is not accepted, it is either because the measure could be correct but does not reflect any particular skill in market timing (H_{01}), or because the measure could reflect some skill but does not correctly reflect performance (H_{02}), or both.

We choose to split the global High+Low subsample in three parts of equal size. Therefore, \underline{g} and \bar{g} are the 67th observation of the Low subsample and the 33rd observation of the High sample, respectively. Their values are $\underline{g} = -1.577$ and $\bar{g} = 0.650$.¹² The regression results are presented in Table 6. The results with a single switching regime rule, either at \underline{g} or at \bar{g} , are also reported for the sake of completeness. The table directly reports the total slope coefficient for each zone, and significance levels are computed using the Wald test.

Insert Table 6 here

The linear four-factor model achieves very low significance levels. Such an outcome confirms that this kind of model is not adapted to capture market timing effects. Nevertheless, in the global High+Low subsample, the alpha of negative market timing funds increases as the market timing effect is more pronounced, i.e. $b_1 - b_2 - b_3 < 0$. As the slope coefficient is not significantly different from zero for the region where $\underline{g} < \gamma_{\text{TM}} \leq \bar{g}$, the alternative hypothesis H_A appears to be supported. However, we view that as very weak evidence since the significance level achieved by the regression is a very low 2%. We can simply infer that the Fama-French-Carhart alpha tends to somehow reflect the performance of negative market timing funds. But this result holds much stronger for the

¹²Other switching rules, such as maximizing the information criterion of the regression or splitting the sample at the lowest significant value of gamma, produce very similar results.

other two performance measures. Furthermore, this result is not robust to a change in specification: with the Low subsample only, the slope coefficient when $\gamma_{\text{TM}} \leq \underline{g}$ becomes insignificant. The only reasonable conclusion is that one cannot reject hypothesis H_{01} .

The performance measure $\pi_{\text{TM},v}$ obtained from the variance correction method consistently predicts that managers who time the market deliver better performance. Still taking the High+Low sample as benchmark, the slope $b_1 - b_2 - b_3$ for negative gammas is negative and weakly significant, while the slope is close to +1, significant, for positive market timers. Yet, the strongest coefficient is to be found for values of γ_{TM} close to 0. According to the interpretation of this coefficient, we have a confirmation that the variance correction method only partially adjusts the beta, and that the bias is stronger in the sample of negative market timers, whose average gamma is stronger. Hence, the evidence presented in Table 6 provides support for H_{02} for the performance measure $\pi_{\text{TM},v}$.

Evidence regarding the alternative correction method based on option portfolio replication is much more convincing. Both slopes are highly significant for the positive and negative gamma regions. These results are robust through the specifications. In particular, the value of coefficient b_1 in equation (24) is high and stable. In the median zone, the coefficient is close to 0 and insignificant, at least in the global High+Low and in the High subsample. As the regression's intercept is -2% , the measure π_{TM}^* regularly assigns a negative performance with no or little market timing. The results for the Low subsample are much less convincing, but with low significance level. Given that the data for the Low sample is presumably contaminated by the spurious effect of fund flows emphasized by Warther (1995), one would indeed expect the link between negative timing and performance to be much weaker than for positive market timers, who are not exposed to the same bias.

Through Figure 5, we provide a graphical representation to foster the interpretation of the table's results. We plot the observations of each performance measure against the fund's gamma. The fitted lines corresponding to Table 6 are superposed on each graph. We also identify the observations that correspond to significant values of gamma. They are marked as a '+', in red (as opposed to observations corresponding to insignificant gamma values represented as a small 'x', in blue) and those that correspond to significant values of the performance measure itself (circled, in green).¹³

Insert Figure 5 here

In figure 5a, we confirm that the four-factor alpha hardly reflects any market timing skill. The significant values of α_{4F} are scattered throughout the vertical axis, but usually for low values of gamma. As for the other measures, there appears to be a dependence for negative gammas, but obviously not for positive values of the market timing coefficient.

¹³The vertical axis is truncated to $+12\%$ and -12% so as to ease the visual interpretation, even though some values exceed these bounds, but this does not lead to any change in the discussion.

Figure 5b brings some insightful information about $\pi_{\text{TM},v}$. It shows the clustering of significant performance estimates for low but positive values of gamma. The variance correction assigns almost always a significant underperformance to managers who exhibit positive, but moderate market timing. This likely explains the very negative slope observed in the region $\underline{g} < \gamma_{\text{TM}} \leq \bar{g}$. This strong slope, which has no clear economic meaning, also explains the high adjusted R-squared (56.9%) achieved by the piecewise regression in Table 6. The graph shows a large asymmetry between significant performance estimates when the gamma is negative, as the variance correction approach very generously attributes positive performance to fund managers who choose to adopt a concave payoff function. When γ_{TM} is positive, the variance correction provides a weak reward to the fund's convexity. Only one manager, with a gamma smaller than one, achieves positive abnormal returns according to the measure $\pi_{\text{TM},v}$.

The portfolio replication approach to performance adjustment visually delivers what could be expected from such a measure. Fund managers who choose to adopt a strong nonlinear stance get a reward on average. Those who remain with a convexity parameter close to zero have a large dispersion of abnormal returns, but they are negative on average. Noteworthy is the fact that the majority of significantly negative values of π_{TM}^* also correspond to significant values of γ_{TM} . Thus, the adjustment through portfolio replication often diagnoses that managers who try to time the market but fail to do so record a negative abnormal returns. But the upper right envelope of the observations features significant values of abnormal performance too, indicated that good market timers are rewarded as well. The upward trend for positive gammas is pronounced and the linear fit appears adequate. The visual inspection of figure 5c clearly confirms that the option portfolio replication approach that yields performance measure π_{TM}^* indeed captures, but also reveals the skills of fund managers who actively time the market. Evidence shows that the higher the positive convexity they create, the larger their tendency to replicate on option at a lower cost than the one implied by market prices. Such evidence marks a step towards the reconciliation between regular evidence of low or negative performance of market timers derived by papers using methods parented to the TM model (return-based evidence) and the more positive picture emphasized by Jiang *et al.* (2007) with a more detailed holding-based analysis. These authors attribute the difference to a greater power of their analysis to detect performance. Even though their justification might be right, we provide an alternative, much simpler explanation: perhaps the widely accepted adjustments of alpha for market timing are simply underestimated, and the fund manager's resulting alpha is just systematically too low.

5 Conclusion

The explicit recognition of the isomorphism between the returns of a market timing fund and the pattern of a protective put or a covered call has been done for long. Nevertheless, when performance measurement is involved, it has only been rigorously translated in the Henriksson and Merton (1981) model in a companion paper by Merton (1981). Even though the Treynor and Mazuy (1966) approach remains nowadays the other major – and perhaps dominant – approach for the measurement of a market timer’s performance, the literature has come up with approximate performance corrections. Because none of the solutions proposed so far accounts for the fund’s linear sensitivity (beta) with respect to the benchmark portfolio, the adjustment can only be partial. Our paper has gone back to the basics. We have started from the quadratic return specification underlying the TM model, and identified a passive replicating portfolio with cash and options that delivers the same pattern. The intuition is straightforward: in a mean-variance world where options are priced using the Black-Scholes-Merton formula, any passive portfolio manager could achieve this pattern. A manager’s performance is thus the difference between the regression’s alpha and the replicating cost of his portfolio.

From this rather simple idea, we have had to overcome two issues. They relate to the way to go back and forth from the fund’s returns to the returns of the replicating portfolio. The passage from the TM model to the option replication framework involves some specification error, while the other way around induces a numerical approximation error. We provide workable solutions for both issues. *In fine*, we find no serious hindrance that would endanger the practicability of the replication approach.

The empirical investigation of the properties of this new performance adjustment appears particularly instructive. By focusing on a sample of potential market timing funds, we directly address the fundamental question arising from our theoretical discussion: can we differentiate between active market timing skills and passive convexity exposures obtained with nonlinear trading strategies? Our comparison with other performance measures applied to the same funds emphasize major weaknesses with state-of-the-art metrics. Namely, the linear four-factor alpha does not adequately reflect market timing, while the traditional corrections of the Treynor and Mazuy alpha leave out a biased estimation of the manager’s performance. With a more rigorous treatment of the fund’s returns convexity, not only does the correction appear to have a reasonable magnitude, but we obtain preliminary but insightful evidence of a positive link between the corrected alpha and the magnitude of the Treynor-Mazuy convexity parameter.

We stress that this evidence is preliminary. It is limited with timing and cross-sectional constraints. It emphasizes in-sample performance of several funds, but does not address persistence. We

have also restricted ourselves to a single factor model for the correction, even though the method proposed in this paper is perfectly applicable in a multi-factor asset pricing specification. The parameterization of the replicating option can also probably be discussed. We are aware of these, and possibly many other, limitations. The scope of our empirical analysis remains at the level of the comparison of the properties of a performance measure relative to its peers. In such a context, it would be inadequate to draw inference about the quality of market timing fund managers in a performance assessment context. Rather, the emphasis of the appropriate and proportionate character of a rigorous adjustment to portfolio performance paves the way for the use of such a measure in subsequent analyses of full cross sections active market timing funds in a dynamic and international environment.

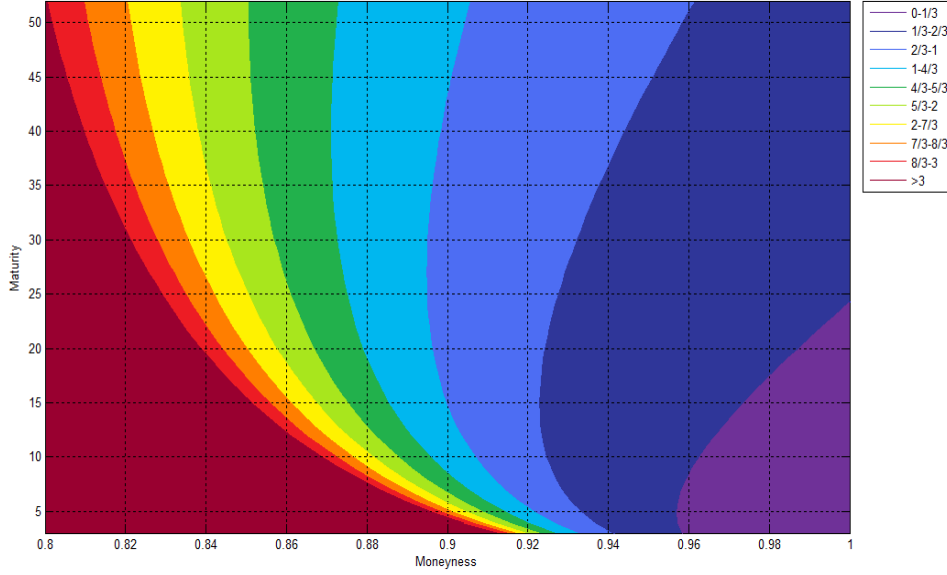
References

- Admati, A., S. Bhattacharya, P. Pfleiderer, and S. Ross. 1986. On Timing and Selectivity. *Journal of Finance* 41: 715–32.
- Becker, C., Ferson, W. E., Myers, D. H., and Schill, M. J., 1999. Conditional market timing with benchmark investors. *Journal of Financial Economics* 52: 119–148.
- Blake, C., E. Elton, and M. Gruber. 1993. The Performance of Bond Mutual Funds. *Journal of Business* 66: 71–403.
- Bollen, N. P. B., and Busse, J. A. 2001. On the Timing Ability of Mutual Fund Managers. *Journal of Finance* 56: 1075–1094.
- Bollen, N. P. B., and Busse, J. A. 2004. Short-Term Persistence in Mutual Fund Performance. *Review of Financial Studies* 18: 569–597.
- Carhart, M., 1997. On Persistence in Mutual Fund Performance. *Journal of Finance* 52: 57–82.
- Coles, J. L., Daniel, N. D., and Nardari, F. 2006. Does the choice of model or benchmark affect inference in measuring mutual fund performance? *Working paper, W.P. Carey School of Business, Arizona State University.*
- Comer, G. 2006. Hybrid Mutual Funds and Market Timing Performance. *Journal of Business* 79:771–97.
- Comer, G., Larrymore, N., and Rodriguez, J. 2009. Controlling for fixed income exposure in portfolio evaluation: Evidence from hybrid mutual funds. *Review of Financial Studies* 22: 481–507.
- Edelen, R. M., 1999. Investor flows and the assessed performance of open-end mutual funds. *Journal of Financial Economics* 53: 439–466.
- Fama, E., and K. French, 1993. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics* 33: 3–56.
- Garman, M. 1992. Charm School. *Risk Magazine* 5 (7): 53–56.
- Gilster, J. E., Jr. 1990. The Systematic Risk of Discretely Rebalanced Option Hedges. *Journal of Financial and Quantitative Analysis* 25: 507–516.
- Gilster, J. E., Jr. 1997. Option Pricing Theory: Is 'Risk-Free' Hedging Feasible? *Financial Management* 26: 91–105.
- Glosten, L. R. and Jagannathan, R. 1994. A contingent claim approach to performance evaluation. *Journal of Empirical Finance* 1: 133–160.
- Grinblatt, M., and Titman, S. 1989. Portfolio performance evaluation: Old issues and new insights. *Review of Financial Studies* 2: 393–421.
- Grinblatt, M., and Titman, S. 1994. A study of monthly mutual fund returns and performance evaluation techniques. *Journal of Financial and Quantitative Analysis* 29: 419–444.
- Haug, E.G., 2003. Know Your Weapon, Part 1, *Wilmott Magazine*, May: 49–57.

- Henriksson, R. D., and R. C. Merton, 1981. On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills. *Journal of Business* 54: 513–33.
- Ingersoll, J., Spiegel, M., Goetzmann, W., and Welch, I., 2007. Portfolio Performance Manipulation and Manipulation-proof Performance Measures. *Review of Financial Studies* 20: 1503–1546.
- Jagannathan, R., and R. Korajczyk, 1986. Assessing the Market Timing Performance of Managed Portfolios. *Journal of Business* 59: 217–35.
- Jiang, G. J., Yao, T., and Yu, T., 2007. Do mutual funds time the market? Evidence from portfolio holdings. *Journal of Financial Economics* 86: 724–758.
- Keating, C. and Shadwick, W. F., 2002. Omega: A Universal Performance Measure. *Journal of Performance Measurement* 6 (Spring): 59–84.
- Kothari, S.P. and Jerold B. Warner. 2001. Evaluating mutual fund performance. *Journal of Finance* 56: 1985–2010.
- Krimm, S, Scholz, H., and Wilkens, M. 2008. Selection, timing and total performance of equity mutual funds: On the relevance of model specification. *Working paper, Catholic University of Eichstätt-Ingolstadt*.
- Kryzanowski, L., Lalancette, S., and 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: 205–224.
- Lehmann, B, and Modest, D. 1987. Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons. *Journal of Finance* 42:233–65.
- Merton, R. C. 1981. On Market Timing and Investment Performance. I. An Equilibrium Theory of Value for Market Forecasts. *Journal of Business* 54: 363–406.
- Pesaran, M. H. and Timmermann, A. G., 1994. A generalization of the nonparametric Henriksson-Merton test of market timing. *Economics Letters* 44: 1–7.
- Sharpe, W. F., 1966. Mutual fund performance. *Journal of Business* 39: 119–138.
- Sortino, F. A., and Price, L. N., 1994. Performance measurement in a downside risk framework. *Journal of Investing* 3 (Fall): 59–65.
- Treynor, J. L., 1965. How to rate management investment funds. *Harvard Business Review* 43 (Jan/Feb): 63–75.
- Treynor, J., and Mazuy, K., 1966. Can Mutual Funds Outguess the Market? *Harvard Business Review* 44 (July–August):131–136.
- Warther, V. A., 1995. Aggregate mutual fund flows and security returns, *Journal of Financial Economics* 39: 209–235.

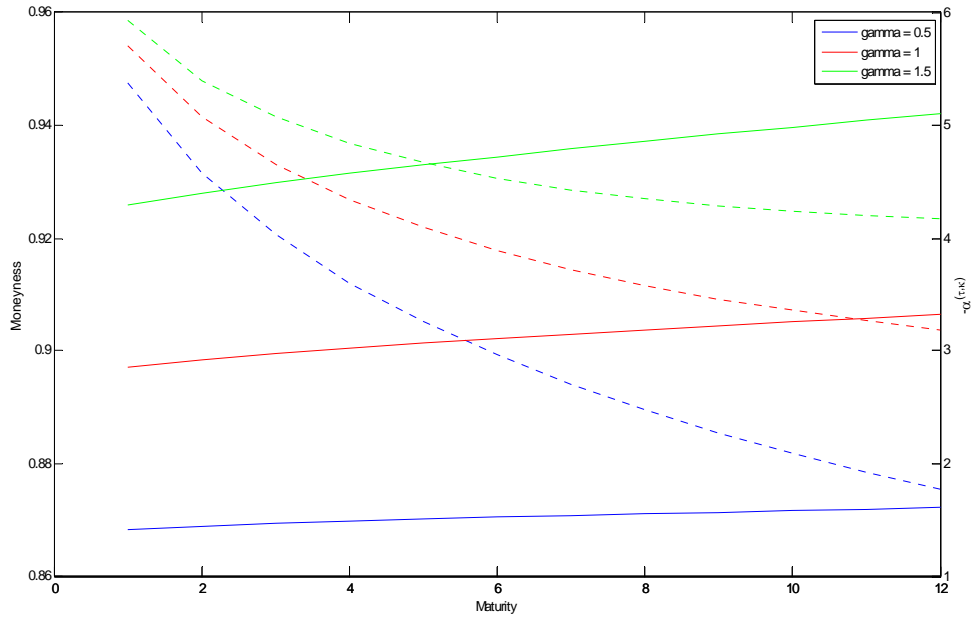
Figures

Figure 1: Target ratio $\frac{\beta_{TM}}{\gamma_{TM}} = \frac{2\Delta_{\tau,\kappa}}{\Gamma_{\tau,\kappa}}$ as a function of option moneyness and maturity



This figure plots the contour lines of the target ratio $\frac{\beta_{TM}}{\gamma_{TM}} = \frac{2\Delta_{\tau,\kappa}}{\Gamma_{\tau,\kappa}}$ as a function of the option moneyness (horizontal axis, expressed as a fraction of the spot underlying price), and maturity (vertical axis, in weeks). We use the Black-Scholes-Merton option pricing formula with as inputs the average 3-month US Treasury Bill rate and the average fitted GARCH(1,1) volatility of the market index between January 1999 to September 2008, which are respectively set to a weekly 0.066% and 2.289%, respectively. The bottom left area corresponds to values of the ratio above 3.

Figure 2: Evolution of $\alpha^{(\tau,\kappa)}$ as a function of option maturity



This figure reports $-\alpha^{(\tau,\kappa)} = -w_{\tau,\kappa}\Theta_{\tau,\kappa} - (1 - w_{\tau,\kappa})R_f$ in percent, annualized, as a function of option maturity for $\beta_{\text{TM}} = 1$ and a target value of $\frac{\beta_{\text{TM}}}{\gamma_{\text{TM}}} = \frac{2\Delta_{\tau,\kappa}}{\Gamma_{\tau,\kappa}}$ equal to 2 (blue line), 1 (red line) and 0.67 (green line). The dotted lines represent the corresponding option moneyness (expressed as a fraction of the spot underlying price).

Figure 3: Evolution of $c^{(\tau,\kappa)}$, $\beta^{(\tau,\kappa)}$ and $\gamma^{(\tau,\kappa)}$ as a function of option maturity

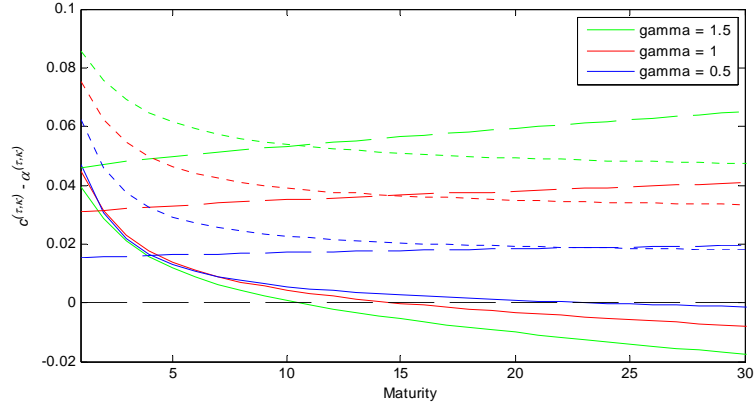


Figure 3a: evolution of $c^{(\tau,\kappa)} - \alpha^{(\tau,\kappa)}$

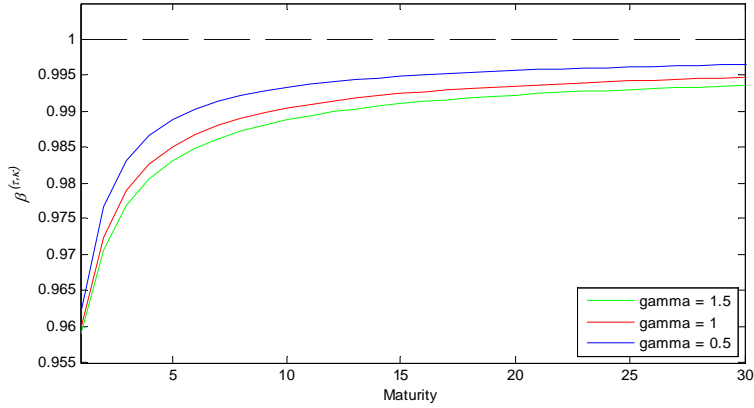


Figure 3b: evolution of $\beta^{(\tau,\kappa)}$

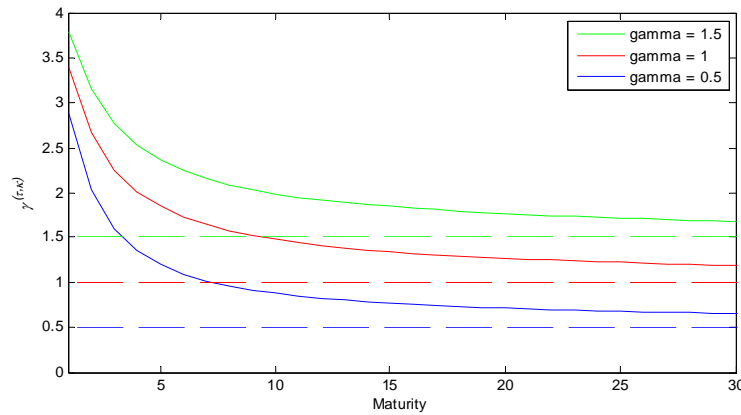


Figure 3c: evolution of $\gamma^{(\tau,\kappa)}$

This figure reports the values of the regression parameters $R_t^{(\tau,\kappa)} = c^{(\tau,\kappa)} + \beta^{(\tau,\kappa)} R_{mt} + \gamma^{(\tau,\kappa)} R_{mt}^2 + \eta_t$ as a function of option maturity and corresponding moneyness for $\beta_{TM} = 1$ and a target value of

$\frac{\beta_{\text{TM}}}{\gamma_{\text{TM}}} = \frac{2\Delta_{\tau,\kappa}}{\Gamma_{\tau,\kappa}}$ equal to 2 (blue line), 1 (red line) and 0.67 (green line). In Figure 3a, we also report $-\alpha^{(\tau,\kappa)} = -w_{\tau,\kappa}\Theta_{\tau,\kappa} - (1 - w_{\tau,\kappa})R_f$ (thin-dotted line) and the difference $c^{(\tau,\kappa)} - \alpha^{(\tau,\kappa)}$ (solid line). In this figure, the values are in percent. In Figures 3b and 3c, the thick-dotted lines represent the corresponding target values.

Figure 4: Box and whisker plot of performance estimates

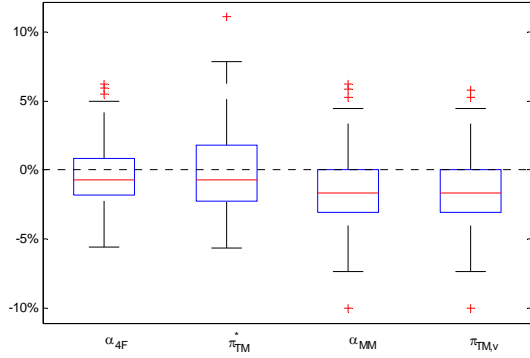


Figure 4a: High subsample

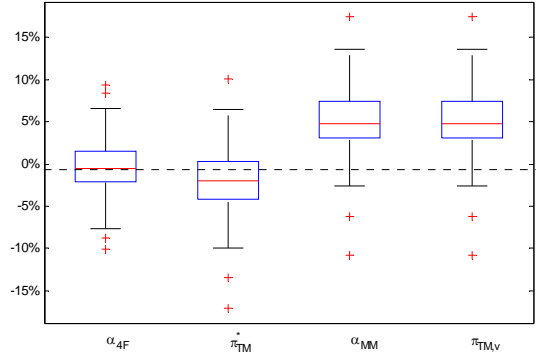


Figure 4b: Low subsample

This figure represents the distribution of the four-factor model alpha (α_{4F}), the portfolio replication-adjusted performance (π_{TM}^*), the market model alpha (α_{MM}) and the variance correction-adjusted performance ($\pi_{TM,v}$) through box and whisker plots. The High and Low subsamples feature the 100 funds with the highest and lowest White heteroskedasticity-consistent estimates of gamma, respectively, extracted from the original set of 1,242 mutual funds with undiscontinued series of weekly returns from January 1999 to September 2008. On each box, the central mark (red line) is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered as outliers, i.e. within 1.5 times the interquartile range, and outliers are plotted individually.

Figure 5: Observed and fitted values of π_{TM}^*

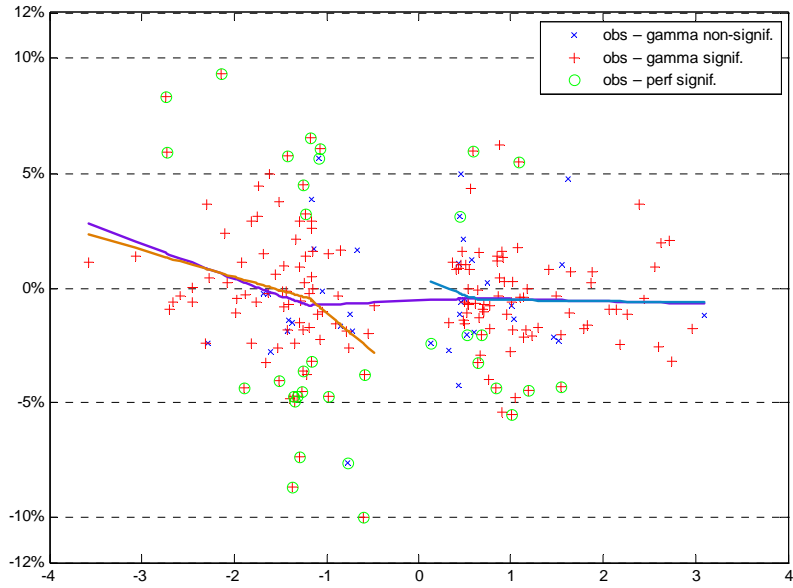


Figure 5a: piecewise estimation of α_{4F} as a function of γ_{TM}

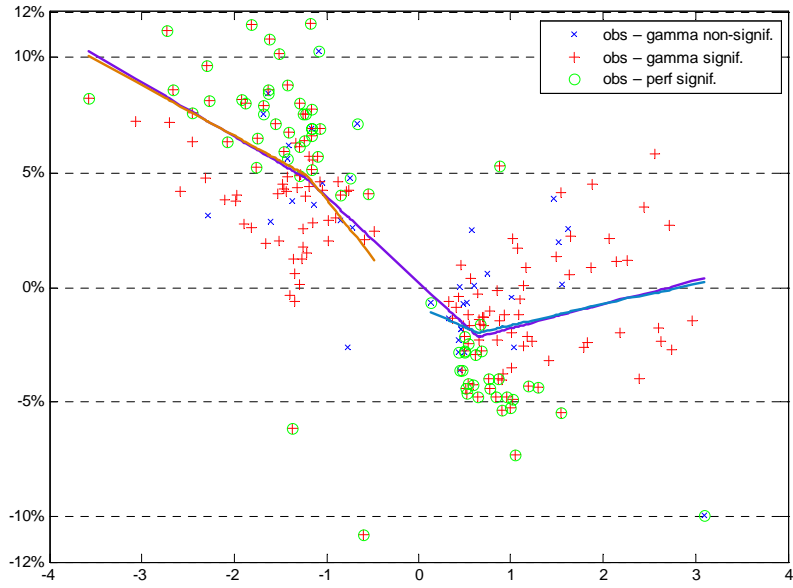


Figure 5b: piecewise estimation of $\pi_{TM,v}$ as a function of γ_{TM}

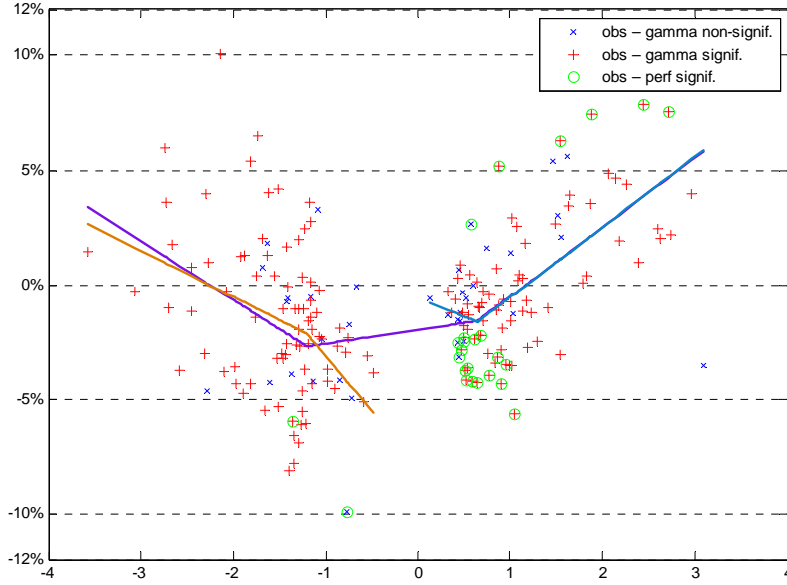


Figure 5c: piecewise estimation of π_{TM}^* as a function of γ_{TM}

This figure reports the observed and fitted values of $\text{perf}_i = b_0 + b_1\gamma_{TM,i} + b_2 \max(\bar{g} - \gamma_{TM,i}, 0) + b_3 \max(\underline{g} - \gamma_{TM,i}, 0) + v_i$ where $\text{perf}_i = \alpha_{4F,i}$ (Figure 5a), $\pi_{TM,v,i}$ (Figure 5b) and $\pi_{TM,i}^*$ (Figure 5c). The break points are $\underline{g} = -1.577$ and $\bar{g} = 0.650$. The combined High+Low subsample features the 100 funds with the highest and lowest White heteroskedasticity-consistent estimates of gamma extracted from the original set of 1,242 mutual funds with undiscontinued series of weekly returns from January 1999 to September 2008. Each observation is reported as an 'x' (in blue) or as a '+' (in red) when the corresponding value of γ_{TM} is insignificant or significant at the 10% confidence level, respectively. It is circled (in green) when the observation is significant at the 10% confidence level. The slope coefficients are equal to b_1 , $b_1 - b_2$ and $b_1 - b_2 - b_3$ when $\gamma_{TM} > \bar{g}$, $\underline{g} < \gamma_{TM} \leq \bar{g}$ and $\gamma_{TM} \leq \underline{g}$, respectively. For the High sample, we remove the third independent variable. For the Low sample, we remove the second independent variable. The coefficients are given in Table 6.

Tables

Table 1: Comparison of performance measures for three passive portfolios

	Parameters		Option Greeks				Regression outputs					Performance measures				
	κ	τ	Δ	Γ	Θ	$\frac{2\Delta}{\Gamma}$	R_{adj}^2	α_{TM}	β_{TM}	γ_{TM}	$\frac{\beta_{TM}}{\gamma_{TM}}$	α	$\pi_{TM,v}$	$\pi_{TM,a}$	$\pi_{TM,o}$	π_{TM}^*
Call ₁	0.98	2	0.77	9.7	-0.0027	0.159	97.8	-9.88 (-27.2)	0.43 (151)	2.66 (47.2)	0.163	-2.89 (-3.71)	-1.29	-1.12	-1.41	-0.23
Call ₂	1.01	5	0.45	8.5	-0.0024	0.107	96.8	-8.98 (-44.7)	0.27 (121)	2.09 (46.7)	0.132	-3.50 (-5.73)	-1.41	-1.28	-1.51	0.40
Call ₃	1.04	8	0.30	5.6	-0.0019	0.106	93.6	-8.03 (-26.1)	0.20 (84.3)	1.76 (37.9)	0.113	-3.38 (-6.25)	-1.05	-0.94	-1.13	0.15

This table reports the option parameters and sensitivities, the coefficients of the Treynor and Mazuy regression, and the associated performance measures for three passive portfolios written on the S&P500 index with a positive convexity. The returns are computed weekly from January 1999 to September 2008. Each portfolio i has a weight of $w = 60\%$ invested in Call _{i} and 40% invested in the 3-month US Treasury Bill. Every week, the option is priced using the Black-Scholes-Merton pricing formula with the fitted GARCH(1,1) volatility of the market index. κ refers to the initial moneyness of the option as a multiple of the price of the index. τ is the number of weeks until option maturity. The reported option delta, gamma and theta are the average over the sample period. The α_{TM} is obtained directly from the TM model with excess returns, while β_{TM} and γ_{TM} are computed using quadratic regression with total returns. α refers to the intercept of the market model in excess returns. The adjusted R-squared, alpha and pi measures are computed in percentage. All performance measures are annualized. The student t-statistic, when relevant, is displayed in parenthesis.

Table 2: Maturity ranges with tolerance levels

	$\Gamma_{\tau,\kappa} = 0.5$			$\Gamma_{\tau,\kappa} = 1.0$			$\Gamma_{\tau,\kappa} = 1.5$		
	$\theta = 0.6\%$	$\theta = 0.4\%$	$\theta = 0.2\%$	$\theta = 0.6\%$	$\theta = 0.4\%$	$\theta = 0.2\%$	$\theta = 0.6\%$	$\theta = 0.4\%$	$\theta = 0.2\%$
$c^{(\tau,\kappa)} - \alpha^{(\tau,\kappa)}$	[3.26, 33.1]	[4.66, 23.8]	[6.84, 16.1]	≥ 3.60	[5.22, 44.7]	[8.31, 26.8]	≥ 3.37	≥ 4.95	≥ 8.87
$\beta^{(\tau,\kappa)}$	≥ 1.51	≥ 2.96	≥ 9.24	≥ 1.41	≥ 2.65	≥ 7.22	≥ 1.20	≥ 2.00	≥ 4.75
$\gamma^{(\tau,\kappa)}$	≥ 6.98	≥ 11.2	≥ 23.8	≥ 6.71	≥ 10.9	≥ 24.8	≥ 5.30	≥ 8.33	≥ 18.9
$\alpha \cap \beta \cap \gamma$	[6.98, 33.1]	[11.2, 23.8]	\emptyset	≥ 6.71	[10.9, 44.7]	[24.8, 26.8]	≥ 5.30	≥ 8.33	≥ 18.9

This table reports the maturity intervals for which the difference between the coefficients of the reverse quadratic regression $R_t^{(\tau^*, \kappa^*)} = c^{(\tau^*, \kappa^*)} + \beta^{(\tau^*, \kappa^*)} R_{mt} + \gamma^{(\tau^*, \kappa^*)} R_{mt}^2 + \eta_t$ and their respective target values falls below a specific tolerance level. We set $tol_\beta = \theta / \hat{t}_{R_m}$, $tol_\gamma = \theta / \hat{t}_{R_m}^2$ and $tol_\alpha = \theta \left(\frac{|\bar{R}_m|}{\hat{t}_{R_m}} + \frac{\bar{R}_m^2}{\hat{t}_{R_m}^2} \right)$ with three possible values of θ . Tested option maturities range between 1 and 52 weeks. The last row represents the intersection between the intervals set for each coefficient.

Table 3: Descriptive statistics of the samples

	N	Excess returns			R_{adj}^2	alpha			beta			gamma		
		mean	s.d.			mean	s.d.	% signif.	mean	s.d.	% signif.	mean	s.d.	% signif.
All	1242	2.78	19.05	TM	72.0	1.99	0.82	35.1%	0.925	0.273	99.9%	-0.023	0.849	13.3%
				MM	71.9	1.92	0.61	30.1%	0.925	0.275	100%			
High	100	-0.52	18.91	TM	83.3	-4.69	0.43	66.0%	1.005	0.159	100%	1.092	0.686	79.0%
				MM	82.5	-1.41	0.39	30.0%	0.968	0.268	100%			
Low	100	6.15	20.82	TM	71.6	9.59	0.66	88.0%	1.004	0.173	100%	-1.484	0.578	85.0%
				MM	71.2	5.21	0.52	46.0%	1.023	0.175	100%			
High+Low	200	2.81	19.86	TM	77.4	2.45	1.14	77.0%	1.004	0.166	100%	-0.196	1.438	81.5%
				MM	76.8	1.90	0.65	38.0%	0.996	0.227	100%			

This table reports the descriptive statistics of the samples constructed from the original set of 1,242 mutual funds with undiscontinued series of weekly returns from January 1999 to September 2008. Market returns and the risk-free interest rate are obtained from Kenneth French’s website. The alpha, beta and gamma estimates are obtained from the Market Model (MM) and Treynor and Mazuy (TM) regressions with excess returns. The High and Low subsamples feature the 100 funds with the highest and lowest White heteroskedasticity-consistent estimates of gamma, respectively. They are combined in the High+Low subsample. Total returns and alphas are in yearly percentage terms. The R_{adj}^2 column reports the average regression significance across each sample. The ‘% signif.’ column reports the proportion of estimates that are significant at the 10% confidence level.

Table 4: Corrections for market timing

	Linear Models						Treyner-Mazuy Model										
	Market model			Four-factor			Uncorrected alpha			Variance correction				Portfolio replication			
	α_{MM}	<0	>0	α_{4F}	<0	>0	α_{TM}	<0	>0	adjust.	$\pi_{TM,v}$	<0	>0	adjust.	π_{TM}^*	<0	>0
High	-1.41	29	1	-0.48	8	3	-4.69	66	0	+3.21	-1.47	29	1	+4.65	-0.04	7	17
	(0.39)			(0.32)			(0.43)			(0.28)	(0.38)			(0.47)	(0.43)		
Low	5.21	2	44	-0.30	14	9	9.59	1	87	-4.37	5.22	2	45	-11.43	-1.83	4	0
	(0.52)			(0.47)			(0.26)			(0.24)	(0.52)			(0.26)	(0.54)		
H+L	1.90	15.5	22.5	-0.39	11	6	2.45	33.5	43.5	-0.58	1.87	15.5	23	-3.39	-0.93	5.5	8.5
	(0.65)			(0.40)			(1.14)			(0.59)	(0.65)			(1.18)	(0.51)		

This table reports the average values (in yearly percentage terms) of fund performance using the Market Model (MM), the Fama-French-Carhart 4 Factor model (4F), uncorrected and corrected Treynor and Mazuy (TM) specifications. The subsamples are constructed from the original set of 1,242 mutual funds with undiscontinued series of weekly returns from January 1999 to September 2008. The High and Low subsamples feature the 100 funds with the highest and lowest White heteroskedasticity-consistent estimates of gamma, respectively. They are combined in the High+Low (H+L) subsample. Each fund i 's adjustment is obtained by expression $\gamma_{TM,i}\sigma_m^2$ under the variance correction approach, and by expression $(w_{0.5,\kappa_i} - \beta_{TM,i})R_f - w_{0.5,\kappa_i}\Theta_{0.5,\kappa_i}$ for a call ($\gamma_{TM,i} > 0$) or $-(w_{0.5,\kappa_i} + \beta_{-TM,i})R_f + w_{0.5,\kappa_i}\Theta_{0.5,\kappa_i}$ for a put ($\gamma_{TM,i} < 0$) under the portfolio replication approach. A fund's performance $\pi_{TM,v}$ and π_{TM}^* results from adding the corresponding adjustment to the uncorrected alpha α_{TM} . The '>0' and '<0' columns report the percentage of observations that are significantly positive and negative at the 10% confidence level, respectively. Standard deviations (in yearly percentage terms) are reported under parentheses.

Table 5: Correlation analysis

Pearson	High					Low					High+Low				
	α_{MM}	α_{4F}	α_{TM}	$\pi_{TM,v}$	π_{TM}^*	α_{MM}	α_{4F}	α_{TM}	$\pi_{TM,v}$	π_{TM}^*	α_{MM}	α_{4F}	α_{TM}	$\pi_{TM,v}$	π_{TM}^*
α_{4F}	0.672					0.854					0.582				
α_{TM}	0.214	0.221				0.945	0.775				0.860	0.322			
$\pi_{TM,v}$	0.218	0.212	0.763			1.000	0.854	0.945			0.865	0.471	0.926		
π_{TM}^*	0.154	0.162	0.395	0.889		0.988	0.845	0.922	0.988		0.288	0.574	0.132	0.467	
γ_{TM}	-0.032	-0.052	-0.495	0.184	0.595	-0.416	-0.266	-0.690	-0.417	-0.376	-0.706	-0.101	-0.908	-0.684	0.263
Spearman															
α_{4F}	0.633					0.772					0.500				
α_{TM}	0.319	0.265				0.929	0.712				0.852	0.302			
$\pi_{TM,v}$	0.332	0.251	0.788			1.000	0.772	0.930			0.851	0.376	0.936		
π_{TM}^*	0.235	0.192	0.487	0.893		0.983	0.781	0.892	0.983		0.205	0.504	0.150	0.382	
γ_{TM}	-0.132	-0.116	-0.406	0.162	0.488	-0.387	-0.189	-0.664	-0.388	-0.327	-0.727	-0.108	-0.874	-0.684	0.229

This table reports the linear (Pearson) and rank (Spearman) correlation coefficients between performance measures and the quadratic coefficient in the Treynor and Mazuy (TM) regression (γ_{TM}). The subsamples are constructed from the original set of 1,242 mutual funds with undiscontinued series of weekly returns from January 1999 to September 2008. The High and Low subsamples feature the 100 funds with the highest and lowest White heteroskedasticity-consistent estimates of gamma, respectively. They are combined in the High+Low subsample. A fund's performance $\pi_{TM,v}$ and π_{TM}^* results from adding the corresponding adjustment to the uncorrected alpha α_{TM} .

Table 6: Piecewise linear regression analysis

	Four-factor model			Variance correction			Portfolio replication		
	High	Low	H+L	High	Low	H+L	High	Low	H+L
intercept	0.51	-4.52**	-0.60	-0.89	-1.25	-0.20	-0.58	-7.97***	-2.00***
$\gamma_{TM} \leq \underline{g}$		-3.39	-1.52**		-7.91**	-2.36*		-7.56**	-2.58**
$\underline{g} < \gamma_{TM} \leq \bar{g}$	-1.61	-5.60*	0.17	-1.75	-5.06**	-3.72***	-1.62	-4.80**	0.60
$\gamma_{TM} > \bar{g}$	-0.06		-0.09	0.93*		1.07**	3.07***		3.05***
R_{adj}^2	<0	5.98%	2.03%	2.0%	16.8%	56.9%	35.6%	13.3%	24.1%
Hypothesis		H_{01}			H_{02}			H_A	

This table reports the results of the piecewise linear regression model $\text{perf}_i = b_0 + b_1 \gamma_{TM,i} + b_2 \max(\bar{g} - \gamma_{TM,i}, 0) + b_3 \max(\underline{g} - \gamma_{TM,i}, 0) + v_i$ where $\text{perf}_i = \alpha_{4F,i}$ (four-factor model), $\pi_{TM,v,i}$ (variance correction model) and $\pi_{TM,i}^*$ (portfolio replication model). The break points are $\underline{g} = -1.577$ and $\bar{g} = 0.650$. The subsamples are constructed from the original set of 1,242 mutual funds with undiscontinued series of weekly returns from January 1999 to September 2008. The High and Low subsamples feature the 100 funds with the highest and lowest White heteroskedasticity-consistent estimates of gamma, respectively. They are combined in the High+Low subsample. A fund's performance $\pi_{TM,v}$ and π_{TM}^* results from adding the corresponding adjustment to the uncorrected alpha α_{TM} . The slope coefficients are equal to b_1 , $b_1 - b_2$ and $b_1 - b_2 - b_3$ when $\gamma_{TM} > \bar{g}$, $\underline{g} < \gamma_{TM} \leq \bar{g}$ and $\gamma_{TM} \leq \underline{g}$, respectively. For the High sample, we remove the third independent variable. For the Low sample, we remove the second independent variable. Significance levels are estimated using the Wald test. Exponents *, ** and *** denote significance at the 10%, 5% and 1% confidence levels, respectively. The last row reports the hypothesis that is supported by the results.