The Performance of Market Timing Measures in a Simulated Environment☆

Stéphane Chrétien [a,*], Frank Coggins [b], Félix d'Amours [c]

[a] *Investors Group Chair in Financial Planning*
Associate Professor of Finance, Finance, Insurance and Real Estate Department
Faculty of Business Administration, Laval University, CIRPÉE, GReFA, and LABIFUL

[b] *Associate Professor of Finance, Department of Finance*
Faculté d'administration, Université de Sherbrooke, CIRPÉE and GReFA,

[c] *Doctoral student, Finance, Insurance and Real Estate Department*
Faculty of Business Administration, Laval University

January 2012

Preliminary – Please do not quote without the authors' permission

Abstract

Using simulations controlling for the manager's ability to time the equity, bond and money markets, we compare daily and monthly market timing and global performance measures in terms of performance detection and ranking. Our main results highlight the joint importance of the trading frequency of the fictitious timer and the data sampling frequency for model estimation. Specifically, daily timing measures are superior to those estimated monthly for daily timers, but inferior for occasional, twice-per-month timers. Global measures perform better than timing measures as they show more robustness to differences in trading and data sampling frequencies. In this experiment, we also find that conditional measures do not improve upon unconditional ones, and that conclusions are similar for performance detection versus ranking.

JEL Classification: G12, G23

Keywords: Portfolio Performance Measurement; Market Timing; Simulations

---

☆ We would like to thank Marie-Claude Beaulieu, Patrick Savaria and participants at the ASAC 2010 Conference in Regina, SK, the 2011 Mathematical Finance Days in Montreal, QC, the 2011 ACFAS Conference in Sherbrooke, QC, and the *Université de Sherbrooke* for helpful comments. We gratefully acknowledge financial support from the *Institut de Finance Mathématique de Montréal* (Chrétien and Coggins), the Faculty of Business Administration at Laval University (Chrétien and d'Amours), the *Faculté d'administration*, *Université de Sherbrooke* (Coggins) and the Investors Group Chair in Financial Planning (Chrétien). Stéphane Chrétien is also grateful to Kalok Chan (Department Head) and the Department of Finance at the Hong Kong University of Science and Technology, where part of this research was conducted while he was a Visiting Associate Professor of Finance.
[*] Corresponding author. Pavillon Palasis-Prince, 2325, rue de la Terrasse, Quebec City, QC, Canada, G1V 0A6. Voice: 1 418 656-2131, ext. 3380. Fax: 1 418 656 2624. Email: stephane.chretien@fsa.ulaval.ca.

The Performance of Market Timing Measures in a Simulated Environment

Abstract

Using simulations controlling for the manager's ability to time the equity, bond and money markets, we compare daily and monthly market timing and global performance measures in terms of performance detection and ranking. Our main results highlight the joint importance of the trading frequency of the fictitious timer and the data sampling frequency for model estimation. Specifically, daily timing measures are superior to those estimated monthly for daily timers, but inferior for occasional, twice-per-month timers. Global measures perform better than timing measures as they show more robustness to differences in trading and data sampling frequencies. In this experiment, we also find that conditional measures do not improve upon unconditional ones, and that conclusions are similar for performance detection versus ranking.

JEL Classification: G12, G23

Keywords: Portfolio Performance Measurement; Market Timing; Simulations

1. INTRODUCTION

The U.S. mutual fund industry continues to grow significantly, from total net assets of 5 791 B$ in December 1998 to 10 349 B$ in December 2008[1], in spite of the so-called "dotcom" financial crisis of 2000, the "9/11" crisis of 2001 and the "subprime" crisis of 2008. The vast majority of the funds advocate active management strategies to generate added value compared to their benchmark index. The performance evaluation of these funds is one of the most long-standing issues in finance, starting with the classic contributions of Jensen (1968), Sharpe (1966), Treynor (1965) and Treynor and Mazuy (1966).

There are now a large number of ways to measure the performance, with, for example, more than 100 ways compiled by Cogneau and Hübner (2009a, 2009b). Yet the empirical results are difficult to reconcile as the performance evaluation may change significantly across models and other methodological choices, as first emphasized by Lehmann and Modest (1987). Furthermore, the many ways to measure the performane produce results that are inevitably subject to the benchmark choice or "bad model" problems. (See Roll, 1978; Dybvig and Ross, 1985a, 1985b; Green, 1986; Chen and Knez, 1996; Fama, 1998; and Ahn, Cao and Chrétien, 2009.) Ultimately, such issues call for the developpment of strategies to evaluate the performance measures themselves.

This paper uses simulations, controlling explicitly for the manager's ability, to evaluate the performance of performance measures. The main advantage of such an experiment is that it allows us a cleaner comparison of the quality of different performance measures, a difficult task with real mutual funds as their true ability is unknown. Given the large number of existing ways to measure the performance, this paper focuses on measures of the *market timing* activities of portfolio managers that include a square benchmark return term in the spirit of Treynor and Mazuy (1966). We study the measures in terms of their ability to both significantly detect and correctly rank the performance of simulated portfolio managers. Our simulation setup and choice of models further reflect three important considerations.

[1] Statistics from the 2009 Investment Company Fact Book, p. 9.

First, as a mismatch between the frequency of informed trading and the frequency of timing measurement is potentially problematic (Goetzmann, Ingersoll and Ivković, 2000), we consider two different classes of informed managers (daily and occasional timers) as well as daily and monthly timing and global performance measurements. In our setup, the daily timers receive signals every day on future returns while the occasional timers receive similar signals twice per month on random days. The signals can be at worst random to at best perfect, depending on pre-specified managers' ability levels, and motivate the timers to trade. The resulting portfolio returns are then evaluate daily and monthly to identify the ability levels needed for the different measures to detect significantly the performance or rank correctly the timers.

While performance studies using monthly data are widespread, there is relatively little evidence on the impact of using daily data. Bollen and Busse (2001) examine the market timing ability of equity funds and argue that the daily performance measures produce estimates that are more precise than their monthly counterparts, with a greater number of funds with positive evaluation. Bollen and Busse (2004) furthermore show that detecting persistence in the best equity funds is possible when they are evaluated with daily data. They do so by proposing a global performance measure that complements the market timing measure in the Treynor-Mazuy CAPM-based framework.[2] The findings of these studies suggest that measurement frequency and the use of global performance measures are issues that need to be further examined.

Second, as market timing activities induce time-varying risk exposures, we examine conditional versus unconditional measures. Conditional measures have been first proposed by Chen and Knez (1996) and Ferson and Schadt (1996) to account for public information and time-varying betas in performance evaluation. In a generalization, Christopherson, Ferson and Glassman (1998, hereafter CFG) also introduce time-varying alphas. Ferson and Schadt (1996) show that the evaluation can be biased when time-varying betas are assumed fixed as they are in unconditional measures. In this light, the findings of Bollen and Busse (2001, 2004), for example, could be problematic as they rely on unconditional measures

[2] Their global performance measure combines stock picking and market timing skills.

to evaluate market timing strategies that have time-varying betas. Comparing conditional and unconditional measures thus appear important in our context.

One article that considers both conditional measures and daily data is Beaulieu, Coggins and Gendron (2009). They propose measures based on a bivariate GARCH framework that estimates the time-varying betas and volatilities as functions of the public information aggregated in past error terms.[3] Their results show that GARCH-type performance evaluations are usually higher than their competitors and significantly decrease the number of extreme (positive or negative) performance compared to daily unconditional measures. These findings suggest that some results of Bollen and Busse (2001) on the impact of using daily versus monthly data can be attributed to a misspecified periodic risk assessment.

Based on this literature, this paper examines Treynor-Mazuy-type market timing and global performance measures based on four models: the unconditional CAPM, the unconditional multi-index or style benchmark of Sharpe (1992), which is popular in practice, the conditional CFG model from Christopherson, Ferson and Glassman (1998) and the BiGARCH model of Beaulieu, Coggins and Gendron (2009). Ultimately, we feel that the selection of these two unconditional and two conditional models relevant for estimation with daily and monthly data is consistent with our first two considerations.

As our third consideration, in an attempt to generate realistic market timing strategies, we design our portfolio construction to emulate a typical asset allocation choice faced by balanced mutual funds. Specifically, based on their simulated trading signals, our timers allocate their portfolios between three asset classes, namely a stock index, a bond index and a money market index, so that the maximum allocation in a single class is 50%. We choose this strategy in order to get portfolios that more closely resemble real-life portfolios compared to the in-and-out, stock-index-versus-risk-free-asset-only portfolio choices common in the academic literature. We feel that the resulting comparison of performance measures should be more relevant. Studies dealing specifically with balanced funds are relatively rare. Treynor and Mazuy (1966), Becker, Ferson, Myers and Schill (1999), Ferson and Qian (2004), Aragon

[3] See McCurdy and Morgan (1992) for another financial application of this framework. They also provide a BiGARCH software that is used for part of our results.

(2004), Comer (2006) and Comer, Larrymore and Rodriguez (2009) are some examples using real data. Given that the skills involved in balanced funds management is precisely to time the evolution of different markets, we propose a first look at the ability of market timing and global performance measures for such type of funds.

Some authors have studied the performance of performance measures with simulations, but in different contexts than ours. Goetzmann, Ingersoll and Ivković (2000) show that monthly market timing measures applied to fictitious managers changing their risk exposure daily are biased downwards. They however do not consider conditional measures and occasional-type traders, and focus exclusively on the timing of the stock market. Kothari and Warner (2001) and Kosowski, Timmermann, Wermers and White (2006) investigate more specifically equity asset selectivity performance measures. While the first study shows that selectivity measures are severely biased, the second reveals that there are nevertheless skilled managers whose performance cannot be attributed to luck.

Coles, Daniel and Nardari (2006) propose the study perhaps the closest to ours. Calibrating their simulations on real equity mutual funds returns, they analyze the effectiveness of market timing measures when the unconditional models or reference portfolios are misspecified. Their results show that such misspecification leads to important biases in market timing measures, especially when they are estimated with daily data. In this paper, we look at these issues for unconditional and conditional measures using purely fictitious managers. Our simulation setup is more in line with the one proposed by Farnsworth, Ferson, Jackson and Todd (2000), who study monthly performance measures for equity funds using stochastic discount factors.

Our main empirical results highlight the joint importance of the trading frequency of the fictitious timer and the data sampling frequency for model estimation. In particular, timing measures are relatively inefficient in both detecting performance and ranking when estimated with a data sampling frequency different from the active trading frequency. Global measures generally fare better, a superiority that is amplified when the manager's active trading frequency is much higher than the measurement frequency. Specifically, for the daily timers, the daily market timing measures still work relatively well, with the conditional BiGARCH model being the best. However, for the occasional timers, the global performance measures perform much better as they show more robustness to differences in trading and data sampling frequencies. We finally find that conditional measures do not generally improve upon unconditional ones.

The rest of the paper is divided as follow. The next section provides the theoretical context, including the setup for generating the simulated timers with varying ability and details on the performance measures under investigation. Section 3 describes the methodology for examining if the measures can detect significantly and correctly rank the performance of the simulated portfolios, as well as the data for portfolio construction. Section 4 presents and interprets the empirical results. Section 5 concludes.

2. Theoretical Context

This section first presents how we generate portfolio returns from private signals designed to capture the managers' ability. We then discuss the conditional or unconditional, daily or monthly measures considered for performance evaluation and ranking.

2.1. Returns of Simulated Timers with Varying Ability Levels

The timing experiment investigated in this paper is based on signals that allow two types of fictitious portfolio managers, denoted daily timers and occasional timers, to rank the assets under consideration in terms of their future returns compared to their average returns. This subsection details our setup and highlights the portfolio choices made in an attempt to produce a market timing strategy relevant for balanced mutual funds.

On each trading day, we consider timers who receive an investment signal for each asset with an accuracy that depends on their pre-specified ability to forecast the asset return until the next trading day. Specifically, inspired by Farnsworth, Ferson, Jackson and Todd (2002), we establish the signal as follows:

$$Signal_{i,t-1} = \gamma\left(\sum_{t_p=1}^{T_p} R_{i,t_p} - T_p \cdot \bar{R}_i\right) + (1-\gamma)\Phi \cdot \sqrt{T_p} \cdot \sigma_i \ , \qquad (1)$$

where:

$Signal_{i,t\text{-}1}$ = The signal at day *t*-1 on the return of asset *i* over the next T_p days;
γ = The ability level of the timer, which can vary between 0 and 1;
R_{i,t_p} = The log return of asset *i* at day t_p;
$\overline{R_i}$ = The full-sample mean of the daily log returns of asset *i*;
Φ = An independent *N*(0,1) random number;
σ_i = The full-sample standard deviation of the daily log returns of asset *i*.

According to this equation, a timer with perfect skills (γ = 1) receive a signal for each asset that corresponds precisely to the asset's future return in terms of deviation from its mean. Oppositely, a timer without any skill (γ = 0) receive a completely random signal with a volatility increasing in the asset's standard deviation. A timer with ability level between 0 and 1 thus receives a "mixed" signal that becomes better as γ increases.

Apart from considering timers with varying ability levels, the above equation allows us to create two types of timers to account for two frequencies of informed trading. The first type, denoted the *daily timers*, receives daily signals on the next day's return (so that T_p = 1) and trades every day. The daily timers are thus high-frequency traders. The second type, denoted the *occasional timers*, receives two signals per month on random days and trades only on those two days. Specifically, we draw randomly two days of trading in each month of our sample. The signal for each asset is then based on the cumulative return from one transaction date to the next (so that the value of T_p changes randomly twice per month according to the number of days between two consecutive transaction dates). The occasional timers not only trade more infrequently than the daily timers, but they also receive their information randomly. Hence, they do not have a clear pattern of informed trading, a further difficulty in measuring their ability.

Equipped with their signalling information, the timers then rank the assets according to their signals to form their investment portfolios. To reproduce the investment opportunities faced by balanced mutual funds, we assume that the timers receive signals on three assets, namely a stock index, a bond index and a money market index. In effect, the signals thus help the managers time the evolution of three major asset classes (stock, bond and money markets) and classified them from the most advantageous to the least advantageous. At each transaction date, the timers invest 50% in the first index (or highest signal), 33% in the second index and 17% in the third index (or lowest signal) to form their actively managed portfolios. The return of this portfolio in excess of the risk-free rate is denoted by $r_{\gamma,t}$ to explicitly account for the ability levels γ of the managers under consideration. The reference portfolio excess return used in the performance measures, denoted $r_{r,t}$ hereafter, assumes an allocation of 33.3% in each index, which can be interpreted as the strategic allocation target of the timers, whose active portfolio weights range from 17% to 50%.

2.2. PERFORMANCE MEASURES

To evaluate the performance of the simulated timers, we use Treynor-Mazuy-type market timing and global performance measures based on four models: the unconditional CAPM, the unconditional multi-index or style benchmark of Sharpe (1992), the conditional CFG model from Christopherson, Ferson and Glassman (1998) and the BiGARCH model of Beaulieu, Coggins and Gendron (2009). Each measure is presented in details below.

2.2.1. UNCONDITIONAL MEASURES WITH THE CAPM

Treynor and Mazuy (1996) first propose to measure the market timing ability of managers by adding a quadratic term to the CAPM of Sharpe (1964), Lintner (1965) and Mossin (1966). This idea is the basis of our first unconditional measure. However, with daily returns, Scholes and Williams (1977) indicate that the gradual incorporation of information in prices through non-synchronous trading and other microstructure effects implies that betas at day *t* are better estimated by the sum of the coefficients associated with the market premium at days *t* and *t-1*. In this context, Chen, Ferson and Peters (2010) suggest a daily measure of market timing that also corresponds to the sum of the market timing coefficients for days *t* and *t-1*. The diffusion process for the daily returns of a timer's managed portfolio can therefore be written as follows:

$$r_{\gamma,t} = \alpha_\gamma + \beta_{\gamma 1} r_{r,t} + \beta_{\gamma 2} r_{r,t-1} + \beta_{\gamma 3} r_{r,t}^2 + \beta_{\gamma 4} r_{r,t-1}^2 + \varepsilon_{\gamma,t} \ , \tag{2}$$

where:

$r_{y,t}$ = The excess return of the timer's portfolio with ability level γ at day t;

$r_{r,t}$ = The excess return of the reference portfolio r at day t;

$r_{r,t}^2$ = The squared excess return of the reference portfolio r at day t;

$\varepsilon_{y,t}$ = The error term of the timer's portfolio with ability γ at day t.

Hereafter, for simplicity, we refer to this form as the CAPM model. The parameters α_γ and β_γ are estimated by OLS with Newey and West (1987) standard errors to correct for autocorrelation and heteroskedasticity in error terms. Following Chen, Ferson and Peters (2010), the daily CAPM market timing measure is given by $\beta_{\gamma 3} + \beta_{\gamma 4}$. The CAPM global performance of Bollen and Busse (2004) is given by $\alpha_\gamma + (\beta_{\gamma 3} + \beta_{\gamma 4})\overline{r_{r,t}^2}$, where $\overline{r_{r,t}^2}$ is the average squared excess return of the reference portfolio over the sample. When evaluated with monthly data, the model does not include the t-1 variables. Hence, $\beta_{\gamma 2} = \beta_{\gamma 4} = 0$ and the CAPM market timing and global performance measures correspond respectively to $\beta_{\gamma 3}$ and $\alpha_\gamma + \beta_{\gamma 3} \cdot \overline{r_{r,t}^2}$.

2.2.2. UNCONDITIONAL MEASURES WITH A MULTI-INDEX MODEL

We also analyze the unconditional performance measures with a multi-index model as proposed by Sharpe (1992). This technique, also known as a style analysis, is popular in practice. Rather than use a single reference portfolio as risk factor, we regress the returns of the simulated timers on the returns of three market indexes while restricting the sum of the coefficients to be equal to 1, which reflects the average styles of the managed portfolios. When using daily data, still accounting for the gradual incorporation of information, the diffusion process becomes:

$$
\begin{aligned}
r_{\gamma,t} &= \alpha_\gamma + \beta_{\gamma 1} r_{s,t} + \beta_{\gamma 2} r_{s,t-1} + \beta_{\gamma 3} r_{b,t} + \beta_{\gamma 4} r_{b,t-1} + \beta_{\gamma 5} r_{m,t} \\
&\quad + (1 - \beta_{\gamma 1} - \beta_{\gamma 2} - \beta_{\gamma 3} - \beta_{\gamma 4} - \beta_{\gamma 5}) r_{m,t-1} + \beta_{\gamma 7} r_{r,t}^2 \\
&\quad + \beta_{\gamma 8} r_{r,t-1}^2 + \varepsilon_{\gamma,t}
\end{aligned}, \qquad (3)
$$

where:

$r_{s,t}$ = The excess return of the stock index;

$r_{b,t}$ = The excess return of the bond index;

$r_{m,t}$ = The excess returns of the money market index.

Hereafter, we refer to this form as the Multi-Index model. The Multi-Index market timing and global performance measures are then defined in the same way as those of the CAPM model, i.e. $\beta_{\gamma 7} + \beta_{\gamma 8}$ and $\alpha_\gamma + (\beta_{\gamma 7} + \beta_{\gamma 8})\overline{r_{r,t}^2}$, respectively. With monthly data, we eliminate the *t*-1 variables and restrict to one the sum of the coefficients of the three index returns at period *t*. The Multi-Index market timing and global performance measures become $\beta_{\gamma 7}$ and $\alpha_\gamma + \beta_{\gamma 7} \cdot \overline{r_{r,t}^2}$, with parameters α_γ and β_γ still estimated by OLS with Newey and West (1987) standard errors.[4]

2.2.3. CONDITIONAL MEASURES FROM CHRISTOPHERSON, FERSON AND GLASSMAN (1998)

As in Christopherson, Ferson and Glassman (1998) (CFG, subsequently), we define the conditional alpha and beta as linear functions of predetermined financial information variables [$z_{i,t-1}$].[5] These variables are defined as deviations from their sample average, [$z_{i,t-1} = Z_{i,t-1} - E(Z)$]. The diffusion process for daily returns then becomes:

[4] We also investigate the unconditional non-traded factors model examined by Farsnworth, Ferson, Jackson and Todd (2002), a model that included bond factors (Tbill yield, Default, Slope, Convexity) and equity factors (SMB, HML). We do not report its results as the specification appears problematic in our sample and the parameters are poorly estimated and unreliable.

[5] The information variables that show the most explanatory power for the market premium in our sample are the variation in the short-term interest between *t-2* and *t-1* $\left[z_{1,t-1}\right]$ and a measure of liquidity defined as the difference between the yield on AA commercial papers and the short-term interest rate (Gatev and Strahan, 2006) $\left[z_{2,t-1}\right]$.

$$r_{\gamma,t} = a_{\gamma 0} + \sum_{i=1}^{N} a_{\gamma i} \cdot z_{i,t-1} + b_{\gamma 0} r_{r,t} + \sum_{i=1}^{N} b_{\gamma i} \cdot (z_{i,t-1} \otimes r_{r,t}) + \beta_{\gamma 2} r_{r,t-1} + \beta_{\gamma 3} r_{r,t}^2 + \beta_{\gamma 4} r_{r,t-1}^2 + u_{\gamma,t} \quad (5)$$

Parameters are estimated by OLS with the Newey and West (1987) correction. The $a_{\gamma i}$ and $b_{\gamma i}$ are respectively measuring the sensitivity of the conditional alpha and beta to the different $z_{i,t-1}$. The average conditional alpha and beta are provided by $a_{\gamma 0}$ and $b_{\gamma 0}$, respectively. With daily data, the CFG market timing and global measures correspond respectively to $\beta_{\gamma 3} + \beta_{\gamma 4}$ and $a_{\gamma 0} + (\beta_{\gamma 3} + \beta_{\gamma 4})\overline{r_{r,t}^2}$. With monthly data, these CFG conditional measures become $\beta_{\gamma 3}$ and $a_{\gamma 0} + \beta_{\gamma 3} \cdot \overline{r_{r,t}^2}$ since $\beta_{\gamma 2} = \beta_{\gamma 4} = 0$ in equation (5).

2.2.4. Conditional Measures with a BiGARCH Specification

In the CAPM context with a bivariate GARCH conditional specification of the risk measures, we can obtain the performance evaluation through the joint estimation of a system of equations. Equations (6) and (7) describe the diffusion processes for the excess returns of the reference portfolio and the managed portfolio with ability level *γ*, respectively. Equation (8) shows the bivariate GARCH specification of the seconds moments proposed by Engle and Kroner (1995) and Kroner and Ng (1998)[6], and applied to performance measurement by Beaulieu, Coggins and Gendron (2009).

$$r_{r,t} = a_r + \sum_{i=1}^{N} a_{ri} z_{i,t-1} + e_{m,t} \ , \quad (6)$$

$$r_{\gamma,t} = a_{\gamma 0} + \sum_{i=1}^{M} a_{\gamma i} z_{i,t-1}^{a} + \frac{h_{\gamma r,t}}{h_{r,t}} (a_r + \sum_{i=1}^{N} a_{ri} z_{i,t-1}) + \beta_{\gamma 2} r_{r,t-1} + \beta_{\gamma 3} r_{r,t}^2 + \beta_{\gamma 4} r_{r,t-1}^2 + e_{\gamma,t} \ , \quad (7)$$

$$H_t = \begin{bmatrix} h_{\gamma,t} & h_{\gamma r,t} \\ h_{\gamma r,t} & h_{r,t} \end{bmatrix} = C'C + A'e_{t-1}e_{t-1}'A + B'H_{t-1}B + G'\eta_{t-1}\eta_{t-1}'G \ , \quad (8)$$

where:

[6] This model was first introduced by Baba, Engle, Kraft and Kroner (1990) and is known as the BEKK model.

a_r = The constant for the reference portfolio *r*;

a_{ki} = The parameters of sensitivity to the information variables $z_{i,t-1}$ for portfolios $k = \gamma$ or *r*;

$a_{\gamma 0}$ = The mean conditional alpha with GARCH second moments specification;

H_t = The matrix of conditional second moments for the error terms of portfolios *γ* and *r* at *t*;

$h_{\gamma r,t}$ = The conditional covariance between the error terms of portfolios *γ* and *r*;

$h_{r,t}$ = The conditional variance of the error terms of portfolio *r*;

$h_{\gamma,t}$ = The conditional variance of the error terms of portfolio *γ*;

C = The 2×2 triangular matrix with parameters capturing the constant GARCH effect;

A = The 2×2 symmetric matrix with parameters capturing the ARCH effect;

B = The 2×2 symmetric matrix with parameters capturing the GARCH effect;

G = The 2×2 symmetric matrix with parameters capturing the asymmetric ARCH effect;

e_t = The vector of stacked error terms $(e_{r,t}, e_{\gamma,t})'$;

η_t = The vector $(\eta_{r,t}, \eta_{\gamma,t})'$ where $\eta_{r,t}$ = max [$0, -e_{r,t}$] and $\eta_{\gamma,t}$ = max [$0, -e_{\gamma,t}$].

In this form, referred to the BiGARCH model hereafter, the error terms e_t follow a bivariate normal distribution $N(0, H_t)$. A GARCH specification for the second moments is widely used in financial literature and it represents a relevant choice in the conditional approach proposed by Beaulieu, Coggins and Gendron (2009). This system of equations allows to condition on public information the expected reference portfolio premium, as well as the specific risk [$h_{\gamma,t}$] and beta of portfolio *γ*. The beta is represented by the ratio of the conditional covariance between the returns on portfolios *γ* and *r* [$h_{\gamma r,t}$] to the conditional variance of the returns of portfolio *r* [$h_{r,t}$]. The risk measures are implicit functions of all public information aggregated in past error terms[7] and the expected reference portfolio premium depends on the same pre-determined information variables $z_{i,t-1}$ used in the CFG model of equation (5). The conditional alpha is a function of different information variables $z^a_{i,t-1}$, namely the error term $e_{\gamma,t-1}$ as well as dummies for the January and week-end effects (French, 1980).

[7] Since GARCH (1,1) models condition the second moments on the error term and second moment of the previous period, they can be seen as ARCH(∞) models. Accordingly, the risk measures are not only functions of the error term of the previous period, but, recursively, they also become functions of all past error terms.

This system of equations is estimated by quasi-maximum likelihood with robust standard errors following Bollerslev and Wooldridge (1992). With daily data, we define the conditional BiGARCH market timing and global measures as respectively $\beta_{\gamma 3} + \beta_{\gamma 4}$ and $a_{\gamma 0} + (\beta_{\gamma 3} + \beta_{\gamma 4})\overline{r_{m,t}^2}$. We do not consider an estimation with monthly data since the GARCH specification is more appropriate for high-frequency returns and large number of observations (Nelson, 1990, 1992).

3. Comparative Methodology and Data

This section presents the methodology for comparing the performance of the different market timing and global measures as well as the data used for the empirical results.

3.1. Methodology to Compare the Performance of the Market Timing Measures

We examine the performance of the performance measures in two ways. First, we study their ability to detect significant performance. For each ability level γ, we simulate the signals needed to form the returns of 180 daily timers and 180 occasional timers. We then assess their performance with every measure. If the measure properly account for the market timing activities, it should be able to detect significant performance at a low ability level γ. A less effective performance measure should detect a significant performance only at a higher ability level γ. For each performance measure and each ability level γ, we summarize the results by computing a *t*-statistic on the significance of the mean performance value across the 180 evaluations.

Second, we verify if the ranking of the timers according to each performance measure corresponds to the expected classification based on the ability level γ. A good performance measure should rank the timers according to their pre-specified ability, while a bad measure should instead classify them randomly. We validate the ability of the performance measures to correctly rank the timers by using the index of coincidence [*IC*] of Friedman (1920). This test allows to explicitly check whether the ranking based on a performance measure and the one based on the true ability level γ are dissimilar (the null hypothesis), or if they are sufficiently comparable to reject the null. The test is calculated as follows:

$$IC = \frac{\sum_{i=1}^{k} 2\left(\overline{Rank}_i - \frac{k+1}{2}\right)^2}{k(k+1)/12}, \quad (9)$$

where k is the number of ranked timers and $\overline{Rank}_i$ is the average, for timer i, of his rank based on a performance measure and his rank based on the true ability level. If the two rankings are opposite, the average of the ranks for each timer will tend to be equal to the same value, approximately $(k+1)/2$. The *IC* statistic follows a Chi-square distribution with k-1degrees of freedom. For each performance measure and each ability level γ, we summarize the results by providing the mean *p*-value associated with the *IC* statistics across the 180 evaluations.

3.2. DATA

This study examines the performance of simulated timers who allocate their assets between a stock index, a bond index and a money market index. The stock index is the CRSP value-weighted index of U.S. stocks from the web site of Kenneth R. French. The bond index is the Aggregate U.S. Bond Index from Barclay's Capital. The money market index is derived from the 3-month U.S LIBOR rate[8] available on Bloomberg. The data cover the period beginning on January 2, 2003, and ending on July 31, 2009, for a total of 1,694 daily observations. Table 1 presents some daily descriptive statistics on the variables used in this study.

Panel A examines the portfolio returns of the daily timers, who trade every day, or the occasional timers, who trade twice per month on random days. The statistics are global averages across all 180 simulations for each ability level and all ability levels considered. The daily timers are assumed to have ability levels varying from $\gamma = 0$ to $\gamma = 0.15$ by 0.01 and the occasional timers are assumed to have ability levels varying from $\gamma = 0$ to $\gamma = 0.35$ by 0.01.[9] The results highlight that, even with much lower ability levels considered, the daily timers enjoy a much larger number of opportunities to time the indexes, resulting in a higher mean return than the occasional timers.

[8] For an example on using this rate to form a money market index, see McCauley (2001).

[9] A gamma of 0.35 for our best occasional timers corresponds to a skill level slightly superior to the one for the best funds evaluated by Farnsworth, Ferson, Jackson and Todd (2002). Similarly, a gamma of 0.15 for our best daily timers also generates performances slightly better than those in their article.

Descriptive statistics on the returns of the reference portfolio, the stock index, the bond index, the money market index and the risk-free asset, and on the lagged information variables are shown in Panel B. Although the sample contains the recent "subprime" recession, the data produce the expected risk-return tradeoff between the markets. The mean stock index return (0.0307%) is higher than the mean bond index return (0.0178%), which is in turn higher than the mean money market return (0.0121%). In addition, the standard deviation of returns is greater for stocks (0.0136%) compared to bonds (0.0026%) and money market (0.0016%).

For all variables, the Jarque-Bera tests reject normality at the 1% significance level, which implies asymmetric and/or fat-tails distributions. Also, the Ljung-Box Q tests and Q^2 tests show evidence of autocorrelations in the values of the variables (except for the bond index returns) and their square values at the 1% significance level. Models that take into account the autocorrelation and heteroskedasticity in error terms, like the BiGARCH model, should be relevant.

4. Empirical Results

This section presents the empirical results for the unconditional and conditional monthly performance measures estimated with daily or monthly data.

4.1. Descriptive Statistics of the Performance Measures

Table 2 gives descriptive statistics on the performance measures (market timing and global) of the simulated market timers with daily transactions (panel A) or occasional transactions (panel B). The measures are estimated with either daily data or monthly data and consider timing ability levels between $\gamma = 0$ and $\gamma = 0.15$ for daily timers and between $\gamma = 0$ and $\gamma = 0.35$ for occasional timers. The table provides overall statistics (mean, standard deviation, maximum and minimum) as well as the mean and standard deviation conditional on pre-specified ranges of timing ability.

Panel A shows that the average market timing measures are generally similar across models, with values around 0.75 with daily data and 0.30 with monthly data. The only notable exception is the BiGARCH model which obtains average timing measures that are approximately 0.1 greater than the other models. The result for this conditional model is in accordance with a common finding in the literature that conditional performance measures produce higher evaluation than their unconditional counterparts for mutual funds (Ferson and Schadt, 1996; Beaulieu, Coggins and Gendron, 2009). However, we do not find such a result for the other conditional measures we consider, namely the BiGARCH global measures and the CFG timing and global measures, or when we examine the results for the occasional market timers. Throughout our results, the BiGARCH timing measure is better at evaluating the ability of daily market timers than the other timing measures. Focusing on the global measures, all measures are similar as well, producing average values around 0.017 with daily data and between 0.32 and 0.48 with monthly data. While the daily global performance measures are very close across models, they vary more when estimated with monthly data. Given the samples of 1694 daily observations versus 79 monthly observations for the estimation, the daily performance measures are estimated more precisely than the monthly ones.

For the occasional timers examined in panel B, the results show that the market timing measures estimated with monthly data better capture the managers' timing ability than the ones estimated with daily data. The global measures are similar across models.

4.2. Formal Comparison of Performance Evaluation Detection

Table 3 shows the *t*-statistics of the performance measures (market timing and global) for the *daily* market timers with varying ability levels. Shaded statistics indicate significance at the 5% level.

First, the timing measures estimated with daily data require lower ability levels before capturing significantly the ability of daily market timers than their equivalent estimated with monthly data. This finding suggests that the timing measures estimated with a data sampling frequency the same as the manager's active trading frequency allow a better assessment of the performance.

Second, the global performance measures are useful in detecting the timing ability as we obtain significant values at lower ability levels than with the market timing measures. This more powerful detection is particularly important for the monthly measures, suggesting than the use of global (as opposed to market timing) performance measures is needed when the manager's active trading frequency is much higher than the measurement frequency.

Third, comparing the models, the timing measure of the BiGARCH model requires the smallest ability level, specifically γ = 0.05, for a significant market timing performance. The BiGARCH timing measure thus appears the best specified in our setup. The global measures of all the models perform relatively similarly, with significant detection at γ = 0.05. The BiGARCH model is the only model under comparison for which its global measure does not improve upon its timing measure in detecting an abnormal performance.

Using the format of table 3, table 4 provides the *t*-statistics of the performance measures in the case of the *occasional* market timers, with significant statistics at the 5% level still shaded.

First, in contrast to the results in table 3, the monthly timing measures now require lower ability levels than the daily timing measures for capturing a significant timing performance. Thus, for the occasional managers, who trade twice per month on random days, the daily timing measures are less powerful than the monthly timing measures.

Second, similar to the results for the daily timers, the market timing measures for the occasional timers become significant at higher ability levels than the global measures. However, looking at magnitudes, the global measures now appear much more powerful than the timing measures, with significant detections starting at γ = 0.16 for the best global measures compare to γ = 0.25 for the best timing measures. The timing measures are thus relatively inefficient when estimated with a data sampling frequency different from the active trading frequency. The global measures estimated at any frequency appear able to compensate for this inefficiency.

Third, comparing the models, the most powerful models in properly detecting significant timing ability for the occasional timers are the two unconditional models (CAPM and Multi-Index), which produce similar results. The conditional models (CFG and BiGARCH) perform relatively poorly compare to the unconditional models, a difference that was not observed for the daily timers. This poor performance is particularly strong for the measures estimated with daily data, suggesting that considering conditioning variables that controls for the public information at the wrong frequency might be more hurtful than helpful.[10] This issue calls for further investigation.

4.3. Formal Comparison of Observed Versus Expected Rankings

Tables 5 and 6 show the average *p*-values of tests on the equality between the observed performance ranking from the performance measures (market timing and global) and the expected performance ranking from the ability levels. The tests are based on the index of coincidence statistics *IC* proposed by Friedman (1920). In the results tabulated in the $\gamma = 0.04$ row, the tests look at the performance ranking of the five timers generated by varying the ability levels from $\gamma = 0$ to $\gamma = 0.04$ by 0.01. In each subsequent row, the number of ranked timers is increased by one following the ability level. Thus, for example, in the results tabulated in the $\gamma = 0.14$ row, the tests look at the ranking of the 15 timers generated by varying the ability levels from $\gamma = 0$ to $\gamma = 0.14$ by 0.01. A low *p*-value indicates that a model has a high ability to rank the managers correctly. It is more likely when a large cross-section of ability levels is considered. Tables 5 and 6 provide the average results for the 180 daily and occasional market timers, respectively. Shaded *p*-values indicate significance at the 5% level.

Tables 5 and 6 provide findings on performance ranking similar to the ones reached from tables 3 and 4 on performance detection. In table 5, the timing measures estimated with daily data require lower ability levels before ranking correctly, at the 5% significance level, the ability of daily market timers than their equivalent estimated with monthly data. Also, the global performance measures are useful in ranking correctly the timing ability as they produce significant values at lower ability levels than the market timing measures. Finally, the global measures perform relatively similarly across models, while the timing measure of the BiGARCH model needs the smallest ability level of all the timing measures to rank correctly the daily timers.

In table 6, the monthly timing measures require lower ability levels than the daily timing measures for ranking correctly the occasional timers at the 5% significance level. Furthermore, the ranking *p*-values

[10] As pointed out by Farnsworth, Ferson, Jackson and Todd (2002), it is possible that low correlations between the variables taken into account in the models and the simulated portfolio returns generate greater variability in the performance measures, leading to insignificantly different from zero values.

of the market timing measures become significant at much higher ability levels than the ones of the global measures. Finally, comparing the models, the most powerful models in properly ranking the occasional timers are the two unconditional models (CAPM and Multi-Index), with once again the two conditional models (CFG and BiGARCH) performing relatively poorly.

The most important difference in results between this section and the previous one is that the measures require higher ability levels to rank correctly the timers than to detect significantly their performance. For example, for the daily timers and the daily timing measures, the CAPM, Multi-Index CGF and BiGARCH models detect significant performance with ability levels starting at 0.07, 0.07, 0.08 and 0.05, respectively, while they rank the timers correctly with ability levels starting at 0.14, 0.15, 0.15 and 0.13, respectively. For the occasional timers and the monthly global measures, the CAPM, Multi-Index and CGF models detect significant performance with ability levels starting at 0.17, 0.16 and 0.21, respectively, while they rank the timers correctly with ability levels starting at 0.23, 0.23, and 0.27, respectively. This finding is consistent with the literature on the difficulty of precisely ranking mutual funds (See Roll, 1978; Dybvig and Ross, 1985; Green, 1986; Lehmann and Modest, 1987; Chen and Knez, 1996; and Ahn, Cao and Chrétien, 2009).

Even if tests on significant performance appear more powerful than tests on correct ranking, it is interesting to note that the conclusions are similar for the two methodologies. In particular, timing measures are relatively inefficient in both detecting performance and ranking when estimated with a data sampling frequency different from the active trading frequency. Global measures generally fare better, a superiority that is amplified when the manager's active trading frequency is much higher than the measurement frequency. Finally, conditional measures that account for information at another frequency that the one considered by the market timers might be problematic.

5. CONCLUSION

In the literature, several models have been proposed to evaluate the performance of portfolio managers. The objective of our study is to evaluate the performance of those performance models. We focus on selected measures of market timing ability in an environment where the ability to time the stock, bond and money markets is controlled through simulations in a setup inspired by Farnsworth, Ferson, Jackson and Todd (2002). We are interested in the conditional or unconditional performance measures evaluated with daily or monthly data from four different models. We study either market timing measures with a square reference portfolio term, in the spirit of Treynor and Mazuy (1966), or global performance measures following Bollen and Busse (2004). We analyze the ability of the different performance measures to detect significant performance and to rank performance correctly. We consider daily timers, who receive a trading signal every day, and occasional timers, who trade two times per month on random days.

Our results show that the more comprehensive global measures perform better than the more standard timing measures. This finding is particularly true when there is a mismatch between the trading frequency of the simulated timers and the estimation frequency of the performance measures. For the daily timers, the daily market timing measures still work relatively well, with the conditional BiGARCH model being the best. However, for the occasional timers, the global performance measures perform much better as they show more robustness to differences in trading and data sampling frequencies. We finally find that conditional measures do not generally improve upon unconditional ones, and that our conclusions are unaffected by whether we examine performance detection or ranking.

REFERENCES

Ahn, D.-H., Cao, H.H., & Chrétien, S. (2009). Portfolio Performance Measurement: A No Arbitrage Bounds Approach. *European Financial Management*, 15(2), 298-339.

Aragon, G.O. (2005). Timing Multiple Markets: Theory and Evidence from Balanced Mutual Funds, Working Paper, Boston College.

Baba, Y., Engle, R.F., Kraft, D.F., & Kroner, K.F. (1990). Multivariate Simultaneous Generalized ARCH. Working Paper, University of California.

Beaulieu, M.-C., Coggins, F., & Gendron, M. (2009). Mutual Fund Daily Conditional Performance. *Journal of Financial Research*, 32(2), 95-122.

Becker, C., Ferson, W., Myers, D.H., & Schill, M.J. (1999). Conditional Market Timing with Benchmark Investors. *Journal of Financial Economics*, 52, 119-148.

Bollen, N.P.B., & Busse, J.A. (2004). Short-Term Persistence in Mutual Fund Performance. *Review of Financial Studies*, 18(2), 570-597.

Bollen, N.P.B., & Busse, J.A. (2001). On the Timing Ability of Mutual Fund Managers. *Journal of Finance*, 56(3), 1075-1094.

Bollerslev, T., & Woolridge, J.M. (1992). Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances. *Econometric Review*, 11(2), 143-172.

Chen, Z., & Knez, P.J. (1996). Portfolio Performance Measurement: Theory and Applications. *Review of Financial Studies,* 9(2), 511-556.

Chen, Y., Ferson, W., & Peters, H. (2010). Measuring the Timing Ability and Performance of Bond Mutual Funds. *Journal of Financial Economics*, 98(1), 72-89.

Christopherson, J.A., Ferson, W.E., & Glassman, D.A. (1998). Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance. *Review of Financial Studies*, 11(1), 111-142.

Cogneau, P., & Hübner, G. (2009a). The (more than) 100 Ways to Measure Portfolio Performance Part 1: Standardized Risk-Adjusted Measures. *Journal of Performance Measurement*, 13, 56-71.

Cogneau, P., & Hübner, G. (2009b). The (more than) 100 Ways to Measure Portfolio Performance Part 2: Special Measures and Comparison. *Journal of Performance Measurement*, 14, 56-69.

Coles, J., Daniel, N. & Nardari, F. (2006). Does the Choice of Timing Strategy or Timing Index Affect Inference in Measuring Mutual Fund Performance. Working Paper, Arizona State University.

Comer, G. (2006). Hybrid Mutual Funds and Market Timing Performance. *Journal of Business*, 79(2), 771-797.

Comer, G., Larrymore, N., & Rodriguez, J. (2009). Controlling for Fixed-Income Exposure in Portfolio Evaluation: Evidence from Hybrid Mutual Funds. *Review of Financial Studies*, 22(2), 481-507.

Dybvig, P.H., & Ross, S. A. (1985a). The Analytics of Performance Measurement Using a Security Market Line. *Journal of Finance*, 40, 401-416.

Dybvig, P. H. and Ross, S. A. (1985b). Differential Information and Performance Measurement Using a Security Market Line. *Journal of Finance*, 40, 383-399.

Engle, R.F., & Kroner, K.F. (1995). Multivariate Simultaneous Generalized ARCH. *Econometric Theory*, 11(1), 122-150.

Farnsworth, H., Ferson, W.E., Jackson, D., & Todd, S. (2002). Performance Evaluation with Stochastic Discount Factors. *Journal of Business*, 75(3), 473-503.

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Stock Returns, *Journal of Finance*, 47(2), 427-465.

Fama, E. F., & French, K. R. (1993).) Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics*, 33(1), 3-56.

Ferson, W.E., Kisgen, D., & Henry, T. (2006). Evaluating Government Bond Fund Performance with Stochastic Discount Factors. *Review of Financial Studies*, 19(2), 423-456.

Ferson, W.E., & Qian, M. (2004). *Conditional Performance Evaluation, Revisited.* Research Foundation of the Association for Investment Management and Research (AIMR).

Ferson, W.E., & Schadt, R.W. (1996). Measuring Fund Strategy and Performance in Changing Economic Conditions. *Journal of Finance*, 51(2), 425-461.

French, K.R. (1980). Stock Returns and the Weekend Effect. *Journal of Financial Economics.* 8(1), 55-69.

French, K.R. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/. [Access on October 11, 2009].

Friedman, W.F. (1920). *The Riverbank Publications: The Index of Coincidence and its Application in Cryptography*. Laguna Hills, CA: Aegean Park Press.

Gatev, E. & Strahan, P. (2006). Banks' Advantage in Hedging Liquidity Risk: Theory and Evidence from the Commercial Paper Market. *Journal of Finance*, 56(2), 867-892.

Goetzmann, W.N., Ingersoll Jr., J., & Ivković, Z. (2000). Monthly Measurement of Daily Timers. *The Journal of Financial and Quantitative Analysis*, 35(3), 257-290.

Green, R.C. (1986). Benchmark Portfolio Inefficiency and Deviations from the Security Market Line. *Journal of Finance*, 41, 295-312.

Grinblatt, M., & Titman, S. (1994). A Study of Monthly Mutual Fund Returns and Performance Evaluation Techniques. *Journal of Financial and Quantitative Analysis*, 29(3), 419- 444.

Jensen, M.C. (1968). The Performance of Mutual Funds in the Period 1945-1964. *Journal of Finance*, 23(2), 389-415.

Henriksson, R.D. (1984). Market Timing and Mutual Fund Performance: An Empirical Investigation. *The Journal of Business*, 57(1), 73-96.

Investment Company Institute. (2009). *2009 Investment Company Fact Book.* Washington, DC.

Kosowski, R., Timmermann, A. Wermers, R. & White, H. (2006). Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis. *Journal of Finance*, 61(6), 2551-2595.

Kothari, S.P. & Warner, J. B.(2001) , Evaluating Mutual Fund Performance. *Journal of Finance*, 56(5), 1985-2010.

Kroner, K.F., & Ng, V.K. (1998). Modelling Asymetrics Comovements of Assets Returns. *Review of Financial Studies*, 11(4), 817-844.

Lehmann, B., & Modest, D. (1987) Mutual Fund Performance Evaluation: A Comparison of Benchmarks and Benchmark Comparisons. *Journal of Finance*, 42(2), 233-265.

Lintner, J. (1965). The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, 47 (1), 13-37.

McCauley, R.N. (2001). March 2010. *BIS*, *Quarterly Review*, 39-45.

McCurdy, T., & Morgan, I.G. (1992). Evidence of Risk Premiums in Foreign Currency Futures Markets. *Review of Financial Studies*, 5(1), 65-83.

Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica*, 34(4), 768-783.

Nelson, D.B. (1990). ARCH Models as Diffusion Approximations. *Journal of Econometrics*, 45, 7-39.

Nelson, D.B. (1992). Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model. *Journal of Econometrics*, 52, 61-90.

Newey, W.K., & West, K.D. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3), 703-808.

Roll, R.W. (1978). Ambiguity when Performance is Measured by the Securities Market Line. *Journal of Finance*, 33, 1051-1069.

Scholes, M., & Williams, J. (1977). Estimating Betas from Nonsynchronous Data. *Journal of Financial Economics*, 5(3), 309-327.

Sharpe, W.F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *Journal of Finance*, 19(3), 425-442.

Sharpe, W.F. (1966). Mutual Fund Performance. *Journal of Business*, 39, 119-138.

Treynor, J. (1965). How to Rate Management of Investment Funds. *Harvard Business Review*, 43, 63-75.

Treynor, J.L., & Mazuy, K.K. (1966), Can Mutual Funds Outguess the Markets? *Harvard Business Review*, 44(4), 131-136.

Table 1: Descriptive Statistics of Daily Data

	Mean (x100)	S.D. (x100)	Max (x100)	Min (x100)	JB Test		Qtest (K=10)		Q^2 test (K=10)	
Panel A: Descriptive statistics the daily and occasional timers										
Daily timers	0.0369	0.4991	5.7748	-4.5262	31720.90	***	42.34	***	583.35	***
Occasional timers	0.0234	0.4927	5.7748	-4.5262	24814.09	***	54.86	***	1203.86	***
Panel B: Factor and index descriptive statistics										
A. Index:										
Rr	0.0201	0.4494	3.8560	-3.0212	8752.91	***	45.19	***	1758.50	***
Stocks Market	0.0307	1.3663	11.5130	-8.9970	8754.11	***	44.90	***	1654.10	***
Bonds Market	0.0178	0.2536	1.3260	-1.2619	303.82	***	12.93		193.01	***
Money Market	0.0121	0.0135	0.1209	-0.0611	6128.98	***	1986.70	***	415.03	***
Rf	0.0030	0.0015	0.0051	0.0000	100.31	***	16437.53	***	15338.42	***
B. Lagged Instruments:										
Delta on 3M Rates	-0.0006	0.0071	0.7600	-0.8100	93133.69	***	125.77	***	1084.07	***
Liquidity Prime	0.4287	0.5411	3.7300	-0.0300	4916.45	***	13811.87	***	9915.97	***

NOTES: This table presents descriptive statistics of the daily data series. These series include 1694 observations over a period of six and a half years, from January 1, 2003, to July 31, 2009. Panel A shows average statistics for the daily portfolio returns, generated with varying ability levels γ, of the daily timers, who trade every day, or the occasional timers, who trade twice per month on random days. Panel B presents the statistics for the daily index returns (in A) and lagged information variables (in B). The statistics are the mean, standard deviation, minimum, maximum, as well as the values of the Jarque-Bera normality test (JB-Test) and the Ljung-Box tests on the autocorrelation of the variables (Qtest) or the squared variables (Q^2 test) at K=10 lags. The Jarque-Bera and the Ljung-Box statistics follow Chi-square distributions with respectively 2 and 10 degrees of freedoms. *** denotes significance at the 1% level.

Table 2: Performance Evaluation of Simulated Market Timing Portfolios

Panel A: A Descriptive statistics by group of daily timers

		CAPM				Multi-Index				CFG				BiGARCH	
	γ	T_d^{CAPM}	T_m^{CAPM}	G_d^{CAPM}	G_m^{CAPM}	T_d^{MI}	T_m^{MI}	G_d^{MI}	G_m^{MI}	T_d^{CFG}	T_m^{CFG}	G_d^{CFG}	G_m^{CFG}	$T_d^{BiGARCH}$	$G_d^{BiGARCH}$
Mean		0.744	0.304	0.017	0.483	0.709	0.295	0.017	0.438	0.754	0.255	0.017	0.320	0.840	0.018
S.D.		0.388	0.188	0.005	0.171	0.388	0.184	0.007	0.148	0.395	0.204	0.005	0.160	0.348	0.006
Max		2.139	1.014	0.049	1.386	2.101	0.985	0.055	1.264	2.228	0.943	0.048	1.064	2.670	0.050
Min		-1.088	-0.471	-0.013	-0.532	-1.130	-0.453	-0.016	-0.454	-1.292	-0.522	-0.012	-0.474	-0.850	-0.015
G1	$0 \le \gamma \le 0.05$	0.289	0.096	0.005	0.153	0.254	0.093	0.006	0.138	0.309	0.078	0.006	0.095	0.367	0.007
		(0.437)	(0.201)	(0.006)	(0.183)	(0.436)	(0.195)	(0.007)	(0.157)	(0.443)	(0.217)	(0.006)	(0.161)	(0.363)	(0.007)
G2	$0.06 \le \gamma \le 0.1$	0.788	0.322	0.018	0.510	0.754	0.313	0.018	0.463	0.802	0.271	0.018	0.339	0.871	0.019
		(0.393)	(0.189)	(0.005)	(0.175)	(0.392)	(0.184)	(0.007)	(0.151)	(0.395)	(0.204)	(0.005)	(0.165)	(0.351)	(0.006)
G3	$0.11 \le \gamma \le 0.15$	1.245	0.535	0.030	0.852	1.211	0.519	0.030	0.773	1.239	0.452	0.030	0.573	1.377	0.030
		(0.325)	(0.173)	(0.005)	(0.154)	(0.325)	(0.169)	(0.006)	(0.134)	(0.336)	(0.189)	(0.005)	(0.155)	(0.328)	(0.006)

Panel B: Descriptive statistics by group of occasional timers

		CAPM				Multi-Index				CFG				BiGARCH	
	γ	T_d^{CAPM}	T_m^{CAPM}	G_d^{CAPM}	G_m^{CAPM}	T_d^{MI}	T_m^{MI}	G_d^{MI}	G_m^{MI}	T_d^{CFG}	T_m^{CFG}	G_d^{CFG}	G_m^{CFG}	$T_d^{BiGARCH}$	$G_d^{BiGARCH}$
Mean		0.122	0.141	0.009	0.209	0.091	0.140	0.009	0.203	0.080	0.072	0.005	0.168	0.076	0.002
S.D.		0.133	0.151	0.005	0.123	0.133	0.149	0.006	0.113	0.319	0.182	0.010	0.130	0.252	0.004
Max		0.458	0.645	0.028	0.690	0.423	0.640	0.032	0.683	0.998	0.745	0.037	0.696	1.185	0.020
Min		-0.329	-0.508	-0.013	-0.334	-0.364	-0.507	-0.018	-0.307	-0.946	-0.581	-0.026	-0.364	-0.955	-0.016
G1	$0 \le \gamma \le 0.05$	0.048	0.003	0.001	0.015	0.012	0.003	0.001	0.016	0.019	-0.006	0.000	0.011	0.009	-0.001
		(0.136)	(0.145)	(0.005)	(0.121)	(0.137)	(0.143)	(0.006)	(0.112)	(0.303)	(0.178)	(0.009)	(0.130)	(0.254)	(0.004)
G2	$0.06 \le \gamma \le 0.11$	0.083	0.062	0.004	0.096	0.050	0.061	0.004	0.092	0.052	0.031	0.002	0.077	0.032	0.000
		(0.132)	(0.154)	(0.005)	(0.125)	(0.133)	(0.152)	(0.006)	(0.114)	(0.332)	(0.186)	(0.010)	(0.131)	(0.245)	(0.004)
G3	$0.12 \le \gamma \le 0.17$	0.103	0.108	0.007	0.164	0.071	0.107	0.007	0.159	0.053	0.042	0.004	0.124	0.056	0.001
		(0.127)	(0.153)	(0.005)	(0.119)	(0.128)	(0.151)	(0.006)	(0.110)	(0.321)	(0.182)	(0.010)	(0.130)	(0.253)	(0.005)
G4	$0.18 \le \gamma \le 0.23$	0.138	0.164	0.010	0.240	0.108	0.163	0.011	0.234	0.086	0.086	0.006	0.195	0.104	0.003
		(0.120)	(0.144)	(0.005)	(0.113)	(0.120)	(0.143)	(0.005)	(0.104)	(0.302)	(0.175)	(0.009)	(0.122)	(0.241)	(0.004)
G5	$0.24 \le \gamma \le 0.29$	0.165	0.220	0.014	0.322	0.136	0.218	0.014	0.313	0.113	0.111	0.008	0.257	0.119	0.003
		(0.104)	(0.133)	(0.004)	(0.105)	(0.104)	(0.131)	(0.005)	(0.098)	(0.290)	(0.160)	(0.009)	(0.118)	(0.229)	(0.004)
G6	$0.30 \le \gamma \le 0.35$	0.197	0.288	0.018	0.417	0.171	0.286	0.019	0.406	0.159	0.166	0.012	0.342	0.136	0.005
		(0.092)	(0.130)	(0.004)	(0.100)	(0.092)	(0.129)	(0.005)	(0.095)	(0.272)	(0.163)	(0.008)	(0.121)	(0.215)	(0.004)

NOTES: This table shows descriptive statistics on the values of the market timing (T_i) and global (G_i) performance measures estimated with daily data ($i = d$) or monthly data ($i = m$) for different models. The models are the unconditional CAPM (*CAPM*), the unconditional multi-index model (*MI*), the conditional model of Christophersen, Ferson and Glassman (1998) (*CFG*) and the conditional BiGARCH model (*BiGARCH*). Panel A gives the results for all the daily timers with ability levels varying from $\gamma = 0$ to $\gamma = 0.15$ and for three groups (G1, G2, G3) based on increasing levels of ability. Panel B gives the results for all the occasional timers with ability levels varying from $\gamma = 0$ to $\gamma = 0.35$ and for six groups (G1, G2, G3, G4, G5, G6) based on increasing levels of ability. The statistics are the mean, standard deviation, maximum, minimum, as well as the mean and standard deviation conditional on the ability groups. The simulation procedure and performance evaluation measures are described in section 2. The data are presented in table 1.

Table 3: Evaluation Ability of the Performance Measures for the Daily Timers

	CAPM				Multi-Index				CFG				BiGARCH	
γ	T_d^{CAPM}	T_m^{CAPM}	G_d^{CAPM}	G_m^{CAPM}	T_d^{MI}	T_m^{MI}	G_d^{MI}	G_m^{MI}	T_d^{CFG}	T_m^{CFG}	G_d^{CFG}	G_m^{CFG}	$T_d^{BiGARCH}$	$G_d^{BiGARCH}$
0	0.189	0.028	0.037	0.031	0.113	0.024	0.083	0.022	0.214	0.026	0.095	-0.096	0.461	0.371
0.01	0.344	0.183	0.363	0.301	0.268	0.180	0.319	0.307	0.379	0.127	0.420	0.106	0.670	0.643
0.02	0.471	0.435	0.755	0.695	0.386	0.439	0.614	0.730	0.518	0.342	0.805	0.662	0.766	0.890
0.03	0.817	0.528	1.222	0.985	0.741	0.523	0.993	1.032	0.885	0.378	1.262	0.616	1.212	1.285
0.04	0.944	0.857	1.411	1.336	0.856	0.860	1.127	1.389	0.946	0.689	1.447	1.122	1.275	1.499
0.05	1.283	0.887	2.133	1.749	1.204	0.881	1.660	1.850	1.342	0.632	2.169	1.177	1.713	1.923
0.06	1.368	1.400	2.237	2.117	1.287	1.398	1.818	2.209	1.377	1.140	2.283	1.638	1.778	2.162
0.07	1.782	1.301	3.070	2.518	1.698	1.295	2.330	2.644	1.863	0.945	3.067	1.648	2.278	2.550
0.08	2.025	1.795	2.997	2.822	1.934	1.799	2.472	2.936	2.053	1.437	3.079	2.035	2.404	2.822
0.09	2.425	1.881	4.276	3.537	2.342	1.874	3.283	3.743	2.485	1.419	4.315	2.280	3.109	3.601
0.1	2.530	2.189	3.948	3.685	2.440	2.185	3.081	3.827	2.475	1.745	4.045	2.631	2.886	3.379
0.11	3.088	2.398	5.236	4.647	2.998	2.382	3.994	4.842	3.011	1.815	5.176	3.102	3.590	4.169
0.12	3.254	2.775	4.926	4.648	3.151	2.764	3.800	4.820	3.124	2.293	5.023	3.467	3.896	4.464
0.13	3.838	2.993	6.234	5.619	3.738	2.974	5.027	5.858	3.773	2.289	6.080	3.624	4.229	5.196
0.14	4.310	3.650	6.181	6.111	4.205	3.640	4.921	6.340	4.057	2.890	6.189	4.144	4.299	4.902
0.15	4.966	3.769	7.362	6.928	4.850	3.734	6.122	7.108	4.727	2.796	7.185	4.101	5.063	6.160

NOTES: This table presents the *t*-statistics on the significance of the evaluation of the market timing (T_i) and global (G_i) performance measures estimated with daily data ($i = d$) or monthly data ($i = m$) for the *daily* timers with ability levels γ varying from γ = 0 to γ = 0.15 and different models. The models are the unconditional CAPM (*CAPM*), the unconditional multi-index model (*MI*), the conditional model of Christophersen, Ferson and Glassman (1998) (*CFG*) and the conditional BiGARCH model (*BiGARCH*). The simulation procedure and performance evaluation measures are described in section 2. The *t*-statistics are defined in section 3.1. The data are presented in table 1. Shaded statistics indicate significance at the 5% level (lightest shade), the 2.5% level (middle shade) or the 1% level (darkest shade).

Table 4: Evaluation Ability of the Performance Measures for the Occasional Timers

	CAPM				Multi-Index				CFG				BiGARCH	
γ	T_d^{CAPM}	T_m^{CAPM}	G_d^{CAPM}	G_m^{CAPM}	T_d^{MI}	T_m^{MI}	G_d^{MI}	G_m^{MI}	T_d^{CFG}	T_m^{CFG}	G_d^{CFG}	G_m^{CFG}	$T_d^{BiGARCH}$	$G_d^{BiGARCH}$
0	0.317	-0.226	-0.062	-0.143	0.059	-0.226	-0.050	-0.136	0.279	-0.098	0.103	-0.025	0.093	-0.223
0.01	0.281	0.026	0.050	-0.012	-0.003	0.034	0.033	0.026	0.003	0.069	-0.009	0.018	0.117	-0.134
0.02	0.211	0.057	0.151	0.122	-0.053	0.057	0.111	0.123	-0.067	-0.051	-0.068	0.050	-0.085	-0.331
0.03	0.330	-0.007	0.213	0.165	0.087	-0.013	0.161	0.150	0.193	-0.105	0.138	0.090	0.012	-0.235
0.04	0.425	0.021	0.318	0.235	0.157	0.020	0.294	0.250	-0.016	-0.107	-0.047	0.095	-0.008	-0.252
0.05	0.536	0.232	0.400	0.347	0.286	0.233	0.373	0.372	0.063	0.096	0.023	0.245	0.088	-0.064
0.06	0.584	0.246	0.498	0.459	0.303	0.245	0.472	0.477	0.145	0.101	0.142	0.305	0.066	-0.023
0.07	0.568	0.261	0.674	0.598	0.312	0.260	0.543	0.618	0.157	0.095	0.197	0.457	0.305	0.229
0.08	0.662	0.443	0.756	0.733	0.397	0.442	0.714	0.760	0.129	0.179	0.184	0.568	0.110	-0.008
0.09	0.674	0.289	0.865	0.705	0.424	0.295	0.854	0.784	-0.005	0.103	0.118	0.618	0.148	0.046
0.1	0.655	0.530	0.976	0.930	0.417	0.531	0.915	1.001	0.107	0.282	0.212	0.789	0.088	0.065
0.11	0.619	0.685	1.138	1.168	0.388	0.682	1.015	1.201	0.427	0.250	0.546	0.824	0.065	0.107
0.12	0.716	0.668	1.315	1.256	0.467	0.668	1.200	1.297	0.219	0.195	0.376	0.812	0.095	0.096
0.13	0.749	0.450	1.227	1.077	0.501	0.452	1.036	1.155	0.107	0.139	0.292	0.820	0.178	0.270
0.14	0.770	0.521	1.357	1.232	0.509	0.524	1.226	1.275	0.122	0.116	0.395	0.870	0.216	0.247
0.15	0.926	0.775	1.481	1.427	0.662	0.778	1.399	1.516	0.185	0.297	0.334	1.005	0.297	0.274
0.16	0.862	0.880	1.730	1.640	0.608	0.882	1.528	1.698	0.124	0.302	0.382	1.080	0.327	0.341
0.17	0.867	0.983	1.845	1.701	0.608	0.992	1.660	1.811	0.229	0.366	0.545	1.176	0.230	0.343
0.18	0.993	0.959	1.940	1.809	0.743	0.966	1.740	1.915	0.196	0.355	0.508	1.339	0.372	0.505
0.19	0.987	1.061	2.115	1.939	0.745	1.067	1.823	2.037	0.311	0.574	0.700	1.502	0.323	0.536
0.2	1.100	1.114	2.244	2.154	0.851	1.116	2.072	2.272	0.412	0.478	0.731	1.516	0.392	0.569
0.21	1.316	1.014	2.080	1.975	1.029	1.023	1.863	2.094	0.132	0.493	0.579	1.648	0.499	0.679
0.22	1.277	1.198	2.452	2.303	1.014	1.201	2.149	2.421	0.278	0.409	0.695	1.595	0.519	0.612
0.23	1.314	1.514	2.724	2.651	1.051	1.519	2.532	2.786	0.382	0.622	0.873	2.017	0.472	0.611
0.24	1.379	1.378	2.787	2.629	1.103	1.377	2.460	2.722	0.329	0.509	0.845	1.839	0.499	0.623
0.25	1.519	1.692	3.264	3.065	1.254	1.700	3.062	3.214	0.381	0.649	0.953	2.108	0.431	0.597
0.26	1.518	1.484	3.213	2.866	1.221	1.485	3.005	3.001	0.334	0.658	0.904	2.117	0.517	0.756
0.27	1.547	1.786	3.414	3.302	1.296	1.776	3.136	3.459	0.534	0.768	1.037	2.346	0.545	0.733
0.28	1.817	1.734	3.574	3.387	1.531	1.735	3.162	3.485	0.406	0.663	1.044	2.243	0.602	0.808
0.29	1.729	1.891	3.322	3.143	1.455	1.901	3.188	3.305	0.359	0.965	1.076	2.407	0.539	0.945
0.3	2.043	1.832	3.517	3.356	1.760	1.830	3.301	3.419	0.518	0.852	1.232	2.216	0.536	0.898
0.31	2.008	1.825	3.882	3.516	1.710	1.821	3.506	3.649	0.494	0.867	1.306	2.601	0.620	1.099
0.32	2.078	2.301	4.530	4.381	1.781	2.281	4.118	4.489	0.575	0.986	1.448	3.002	0.669	1.041
0.33	2.068	2.189	4.320	3.978	1.773	2.193	4.166	4.160	0.464	1.080	1.353	2.925	0.716	1.287
0.34	2.403	2.508	5.053	4.851	2.068	2.495	4.593	4.929	0.691	1.074	1.709	2.948	0.542	1.180
0.35	2.347	2.761	5.309	5.242	2.033	2.759	4.727	5.385	0.799	1.274	1.937	3.412	0.720	1.313

NOTES: This table presents the *t*-statistics on the significance of the evaluation of the market timing (T_i) and global (G_i) performance measures estimated with daily data ($i = d$) or monthly data ($i = m$) for the *occasional* timers with ability levels γ varying from γ = 0 to γ = 0.35 and different models. The models are the unconditional CAPM (*CAPM*), the unconditional multi-index model (*MI*), the conditional model of Christophersen, Ferson and Glassman (1998) (*CFG*) and the conditional BiGARCH model (*BiGARCH*). The simulation procedure and performance evaluation measures are described in section 2. The *t*-statistics are defined in section 3.1. The data are presented in table 1. Shaded statistics indicate significance at the 5% level (lightest shade), the 2.5% level (middle shade) or the 1% level (darkest shade).

Table 5: Ranking Ability of the Performance Measures for the Daily Timers

	CAPM				Multi-Index				CFG				BiGARCH	
γ	T_d^{CAPM}	T_m^{CAPM}	G_d^{CAPM}	G_m^{CAPM}	T_d^{MI}	T_m^{MI}	G_d^{MI}	G_m^{MI}	T_d^{CFG}	T_m^{CFG}	G_d^{CFG}	G_m^{CFG}	$T_d^{BiGARCH}$	$G_d^{BiGARCH}$
0.04	0.324	0.321	0.232	0.246	0.322	0.323	0.272	0.242	0.331	0.350	0.230	0.263	0.315	0.272
0.05	0.283	0.292	0.186	0.208	0.281	0.295	0.225	0.201	0.289	0.324	0.182	0.239	0.262	0.231
0.06	0.256	0.243	0.154	0.169	0.255	0.246	0.185	0.163	0.261	0.282	0.151	0.202	0.226	0.186
0.07	0.215	0.217	0.118	0.135	0.215	0.217	0.149	0.128	0.219	0.262	0.117	0.175	0.182	0.152
0.08	0.185	0.184	0.093	0.108	0.185	0.185	0.119	0.102	0.190	0.227	0.092	0.143	0.153	0.120
0.09	0.149	0.157	0.072	0.083	0.149	0.159	0.090	0.078	0.153	0.201	0.071	0.115	0.119	0.092
0.1	0.126	0.129	0.056	0.065	0.126	0.131	0.073	0.060	0.132	0.171	0.056	0.091	0.101	0.074
0.11	0.099	0.109	0.044	0.050	0.099	0.111	0.057	0.047	0.107	0.151	0.044	0.070	0.078	0.059
0.12	0.081	0.087	0.035	0.040	0.081	0.088	0.046	0.038	0.087	0.122	0.035	0.055	0.062	0.047
0.13	0.064	0.071	0.028	0.032	0.064	0.071	0.036	0.030	0.068	0.103	0.028	0.044	0.048	0.037
0.14	0.050	0.056	0.023	0.025	0.050	0.056	0.029	0.024	0.055	0.083	0.023	0.035	0.039	0.029
0.15	0.039	0.044	0.019	0.021	0.040	0.045	0.023	0.020	0.044	0.068	0.019	0.028	0.031	0.024

NOTES: This table presents the average *p*-values from the *IC* statistics of Friedman (1920) on the significance of the ranking of the market timing (T_i) and global (G_i) performance measures estimated with daily data ($i = d$) or monthly data ($i = m$) for the *daily* timers with ability levels γ varying from γ = 0 to γ = 0.15 and different models. The models are the unconditional CAPM (*CAPM*), the unconditional multi-index model (*MI*), the conditional model of Christophersen, Ferson and Glassman (1998) (*CFG*) and the conditional BiGARCH model (*BiGARCH*). The simulation procedure and performance evaluation measures are described in section 2. The *p*-values and *IC* statistics are defined in section 3.1. The data are presented in table 1. Shaded statistics indicate significance at the 5% level (lightest shade), the 2.5% level (middle shade) or the 1% level (darkest shade).

Table 6: Ranking Ability of the Performance Measures for the Occasional Timers

	CAPM				Multi-Index				CFG				BiGARCH	
γ	T_d^{CAPM}	T_m^{CAPM}	G_d^{CAPM}	G_m^{CAPM}	T_d^{MI}	T_m^{MI}	G_d^{MI}	G_m^{MI}	T_d^{CFG}	T_m^{CFG}	G_d^{CFG}	G_m^{CFG}	$T_d^{BiGARCH}$	$G_d^{BiGARCH}$
0.04	0.446	0.420	0.393	0.381	0.447	0.419	0.413	0.385	0.462	0.475	0.459	0.438	0.492	0.484
0.05	0.415	0.399	0.371	0.356	0.411	0.401	0.381	0.358	0.462	0.445	0.445	0.422	0.479	0.447
0.06	0.412	0.402	0.358	0.351	0.408	0.402	0.357	0.350	0.460	0.435	0.433	0.404	0.468	0.422
0.07	0.403	0.387	0.332	0.327	0.395	0.388	0.344	0.323	0.447	0.431	0.412	0.379	0.439	0.386
0.08	0.394	0.360	0.302	0.300	0.387	0.361	0.314	0.292	0.448	0.427	0.400	0.342	0.448	0.385
0.09	0.381	0.360	0.275	0.285	0.372	0.361	0.285	0.275	0.459	0.426	0.412	0.323	0.446	0.390
0.1	0.372	0.337	0.251	0.256	0.365	0.339	0.265	0.245	0.455	0.407	0.404	0.296	0.450	0.385
0.11	0.362	0.310	0.226	0.223	0.355	0.308	0.234	0.211	0.425	0.398	0.367	0.273	0.456	0.383
0.12	0.356	0.288	0.198	0.197	0.347	0.286	0.208	0.186	0.420	0.394	0.356	0.255	0.459	0.380
0.13	0.347	0.286	0.181	0.184	0.338	0.286	0.198	0.172	0.426	0.403	0.357	0.245	0.456	0.358
0.14	0.337	0.283	0.160	0.167	0.330	0.282	0.179	0.155	0.428	0.403	0.343	0.228	0.449	0.345
0.15	0.323	0.270	0.147	0.154	0.316	0.269	0.164	0.140	0.426	0.391	0.343	0.218	0.429	0.332
0.16	0.317	0.245	0.126	0.131	0.309	0.244	0.140	0.120	0.429	0.384	0.340	0.201	0.416	0.318
0.17	0.309	0.222	0.113	0.119	0.300	0.220	0.127	0.108	0.423	0.365	0.320	0.186	0.414	0.309
0.18	0.295	0.206	0.097	0.105	0.286	0.205	0.108	0.093	0.422	0.354	0.310	0.165	0.400	0.289
0.19	0.283	0.190	0.086	0.093	0.273	0.189	0.097	0.082	0.416	0.331	0.291	0.143	0.391	0.271
0.2	0.271	0.171	0.072	0.077	0.261	0.170	0.084	0.068	0.403	0.318	0.278	0.125	0.378	0.252
0.21	0.251	0.165	0.063	0.068	0.241	0.164	0.075	0.059	0.409	0.307	0.274	0.111	0.361	0.235
0.22	0.239	0.155	0.055	0.060	0.229	0.154	0.066	0.052	0.406	0.308	0.263	0.101	0.345	0.223
0.23	0.223	0.133	0.045	0.048	0.213	0.131	0.054	0.042	0.392	0.287	0.244	0.085	0.334	0.209
0.24	0.217	0.124	0.037	0.042	0.207	0.123	0.046	0.036	0.388	0.285	0.232	0.076	0.326	0.199
0.25	0.204	0.108	0.029	0.032	0.193	0.107	0.037	0.028	0.381	0.270	0.216	0.063	0.322	0.191
0.26	0.193	0.093	0.023	0.026	0.182	0.092	0.030	0.022	0.377	0.259	0.205	0.055	0.313	0.180
0.27	0.182	0.083	0.019	0.022	0.170	0.082	0.025	0.019	0.357	0.244	0.185	0.045	0.302	0.169
0.28	0.168	0.073	0.016	0.017	0.156	0.072	0.021	0.015	0.348	0.236	0.171	0.039	0.292	0.160
0.29	0.156	0.064	0.013	0.014	0.144	0.063	0.017	0.012	0.348	0.217	0.163	0.032	0.283	0.147
0.3	0.140	0.055	0.010	0.012	0.129	0.055	0.014	0.010	0.336	0.201	0.150	0.028	0.280	0.134
0.31	0.126	0.049	0.008	0.010	0.115	0.049	0.011	0.008	0.328	0.192	0.136	0.023	0.267	0.120
0.32	0.116	0.041	0.007	0.008	0.105	0.041	0.009	0.006	0.315	0.177	0.121	0.018	0.256	0.109
0.33	0.106	0.036	0.005	0.006	0.095	0.035	0.007	0.005	0.313	0.164	0.113	0.015	0.247	0.097
0.34	0.098	0.030	0.004	0.005	0.087	0.029	0.006	0.004	0.298	0.153	0.098	0.012	0.246	0.091
0.35	0.088	0.025	0.003	0.004	0.078	0.024	0.004	0.003	0.283	0.139	0.085	0.010	0.235	0.079

NOTES: This table presents the average *p*-values from the *IC* statistics of Friedman (1920) on the significance of the ranking of the market timing (T_i) and global (G_i) performance measures estimated with daily data ($i = d$) or monthly data ($i = m$) for the *occasional* timers with ability levels γ varying from γ = 0 to γ = 0.35 and different models. The models are the unconditional CAPM (*CAPM*), the unconditional multi-index model (*MI*), the conditional model of Christophersen, Ferson and Glassman (1998) (*CFG*) and the conditional BiGARCH model (*BiGARCH*). The simulation procedure and performance evaluation measures are described in section 2. The *p*-values and *IC* statistics are defined in section 3.1. The data are presented in table 1. Shaded statistics indicate significance at the 5% level (lightest shade), the 2.5% level (middle shade) or the 1% level (darkest shade).