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

May 2012

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 as well as monthly 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, results are robust to controls for stale pricing and 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, Andrew Karolyi, Gabriel Power, 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 as well as monthly 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, results are robust to controls for stale pricing and 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 5 525 B$ in December 1998 to 11 622 B$ in December 2011[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 forcefuly emphasized by Lehmann and Modest (1987). Furthermore, the many ways to measure the performance 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 squared 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 2012 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 three different classes of informed managers (daily, occasional and monthly 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, the occasional timers receive similar signals twice per month on random days, and the monthly timers receive signals every month. 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 generalizations, Christopherson, Ferson and Glassman (1998) also introduce time-varying alphas, while Ferson and Qian (2004) consider time-varying market timing parameters. 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

[2] Their global performance measure combines asset selectivity and market timing skills.

be problematic as they rely on unconditional measures 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 six models: the unconditional CAPM, the unconditional multi-index or style benchmark of Sharpe (1992), which is popular in practice, the unconditional multi-index timing model of Comer (2006), a variant of Sharpe (1992) which allows a timing coefficient for each style benchmark, the conditional model from Christopherson, Ferson and Glassman (1998), the conditional model from Ferson and Qian (2004) and the BiGARCH model of Beaulieu, Coggins and Gendron (2009).

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 (2002), 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, market timing measures are relatively inefficient in both detecting performance and ranking when estimated with a data sampling frequency different from the active trading frequency. The global measures proposed by Bollen and Busse (2004) generally fare better, a finding amplified when the manager's active trading frequency is much lower than the measurement frequency. For the daily timers, the daily and monthly market timing measures still work relatively well, due to the high returns induced by timers even when they have a low timing ability parameter. On the other hand, for the occasional and monthly timers, the global measure shows more robustness to differences in trading and data sampling frequencies. The timing activities of these managers generate a considerable selectivity component that favors the global measure.

The results are similar across performance models, a conclusion also reached by Farnsworth, Ferson, Jackson and Todd (2002) in a different simulation setup. In particular, conditional measures do not generally improve upon unconditional ones. Furthermore, all models lack power to detect and rank performance significantly at low induced ability levels. We finally find that the results are robust to controls for stale pricing in daily data and similar for performance detection versus ranking.

Overall, these findings expand on the analysis of Goetzmann, Ingersoll and Ivković (2000), which focuses only on the monthly measurement of daily timers. They also revisit the conclusion of Bollen and Busse (2001) on the benefits of using daily instead of monthly data, as we show that the frequency of informed trading could be crucial in evaluating these benefits. They add to the analytical results of Lehmann and Timmermann (2007) on the difficulty of separating market timing ability from total performance, and to the methodological exploration of Chen, Ferson and Peters (2010), who investigate the effects of assets with nonlinearities, interim trading, public information and stale pricing in the context of timing in bond mutual funds. Ultimately, we argue that the frequency of data sampling for performance measurement, and in particular how it matches with the frequency of informed trading of the manager, deserves more attention in the evaluation process.

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 three types of fictitious portfolio managers, denoted daily timers, occasional timers and monthly 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:

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

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 return of asset i at day t_p;
$\overline{R}_i$ = The full-sample geometric mean of the daily returns of asset i;
Φ = An independent $N(0,1)$ random number;
σ_i = The full-sample standard deviation of the daily 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 geometric 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. The ability level γ can thus be thought as representing the proportion of the trading signal that is informative, with the remaining proportion being noise.

Apart from considering timers with varying ability levels, the above equation allows us to create three types of timers to account for three 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 shifts its portfolio weights 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 composed 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). These dates are then used for each ability level γ. 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. The third type, denoted the *monthly timers*, receives their signals at the end of each month on next month’s return. The monthly timers are similar to the occasional timers in that they get relatively infrequent information, but similar to the daily timers in that they have access to it on a regular time interval. For a given ability level, a higher frequency of informed trading should lead to a higher performance, so that the daily timer should earn the highest return and the monthly timer should earn the lowest return.

Equipped with their private 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 these major asset classes 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.

For each ability level γ, for $\gamma = 0$ to $\gamma = 0.49$ by increment of 0.01, we simulate the signals needed to form the daily returns of 1000 daily, occasional and monthly timers. We also composed these daily returns to obtain the monthly returns of the three types of timers. Using both daily and monthly data sampling frequencies, we then assess their performance with the measures described in the next section. 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%. [4]

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 six models: the unconditional CAPM, the unconditional multi-index or style benchmark of Sharpe (1992), the unconditional multi-index timing model of Comer (2006), the conditional model from Christopherson, Ferson and Glassman (1998) with time-varying alphas and betas, the conditional model from Ferson and Qian (2004) with time-varying betas and market timing parameters, 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 (1966) 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 stale pricing, 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

[4] This choice of reference portfolio weights result in returns that match closely the returns of the different timers when their ability level is $\gamma = 0$ (see the descriptive statistics in table 1).

associated with the market premium at days t and t-1. In this context, the diffusion process for the daily returns of a timer's managed portfolio can therefore be written as follows:

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

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 Treynor and Mazuy (1966), the CAPM market timing measure is $\beta_{\gamma 3}$. The CAPM global performance measure of Bollen and Busse (2004) is given by $\alpha_{\gamma} + \beta_{\gamma 3} \cdot \overline{r_{r,t}^{2}}$, 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 variable. Hence, $\beta_{\gamma 2} = 0$ and the CAPM market timing and global performance measures still 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 style analysis, is popular in practice. Rather than using a single reference portfolio as risk factor, we regress the returns of a simulated timer on the returns of the three market indexes while restricting the sum of their coefficients to be equal to 1, obtaining a portfolio that reflects the average style of the managed portfolio. When using daily data, still accounting for stale pricing, the diffusion process becomes:

$$r_{\gamma,t} = \alpha_\gamma + \beta_{\gamma 1} r_{s,t} + \beta_{\gamma 2} r_{b,t} + (1 - \beta_{\gamma 1} - \beta_{\gamma 2}) r_{m,t} + \beta_{\gamma 3} r_{r,t-1} + \beta_{\gamma 4} r_{r,t}^2 + \varepsilon_{\gamma,t}\ , \qquad (3)$$

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 return 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. $\beta_{\gamma 4}$ and $\alpha_\gamma + \beta_{\gamma 4} \cdot \overline{r_{r,t}^2}$, respectively. With monthly data, we eliminate the t-1 variable. The Multi-Index market timing and global performance measures still corresponds to $\beta_{\gamma 4}$ and $\alpha_\gamma + \beta_{\gamma 4} \cdot \overline{r_{r,t}^2}$, with parameters α_γ and β_γ estimated by OLS with Newey and West (1987) standard errors.

2.2.3. UNCONDITIONAL MEASURES FROM COMER (2006)

Based on Lehmann and Modest (1987), Comer (2006) proposes a model where the squared returns of every market index are included, so that the manager can time each index differently. Comer (2006) argues that such specification will better capture the timing ability for each individual asset class. Implementing this suggestion as an extension to the Multi-Index model, the diffusion process for the daily returns becomes:

$$r_{\gamma,t} = \alpha_\gamma + \beta_{\gamma 1} r_{s,t} + \beta_{\gamma 2} r_{b,t} + (1 - \beta_{\gamma 1} - \beta_{\gamma 2}) r_{m,t} + \beta_{\gamma 3} r_{r,t-1} + \beta_{\gamma 4} r_{s,t}^2 + \beta_{\gamma 5} r_{b,t}^2 + \beta_{\gamma 6} r_{m,t}^2 + \varepsilon_{\gamma,t} \quad (4)$$

where:

$r_{s,t}^2$ = The squared excess return of the stock index;

$r_{b,t}^2$ = The squared excess return of the bond index;

$r_{m,t}^2$ = The squared excess return of the money market index.

Hereafter we refer to this model as the Comer model. Even if there are multiple timing measures, for convenience and like Comer (2006), we report only the market timing performance measure $\beta_{\gamma 4}$, i.e. the timing parameter of the stock index, as it's the most important one. In our setup, if a manager cannot time the stock market well, then his performance will be low irrespective of the timing of the bond or money markets. The global performance measure still needs to consider every index, so it becomes $\alpha_\gamma + \beta_{\gamma 4} \cdot \overline{r_{s,t}^2} + \beta_{\gamma 5} \cdot \overline{r_{b,t}^2} + \beta_{\gamma 6} \cdot \overline{r_{m,t}^2}$. With monthly data, we eliminate the lag variable and the Comer market timing and global performance measures remains respectively $\beta_{\gamma 4}$ and $\alpha_\gamma + \beta_{\gamma 4} \cdot \overline{r_{s,t}^2} + \beta_{\gamma 5} \cdot \overline{r_{b,t}^2} + \beta_{\gamma 6} \cdot \overline{r_{m,t}^2}$. All parameters are estimated by OLS with Newey and West (1987) standard errors.

2.2.4. 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:

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

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

[5] The information variables that show the most explanatory power for the daily reference portfolio premium are the variation in the three-month Treasury bill yields between *t-2* and *t-1* $\left[Z_{1,t-1}\right]$ and a measure of liquidity defined as the yield difference between AA commercial papers and three-month Treasury bills (Gatev and Strahan, 2006) $\left[Z_{2,t-1}\right]$.

conditional alpha and beta are provided by $a_{\gamma 0}$ and $b_{\gamma 0}$, respectively. The CFG market timing and global measures correspond respectively to $\beta_{\gamma 3}$ and $a_{\gamma 0} + \beta_{\gamma 3} \cdot \overline{r_{r,t}^2}$, with the t-1 variable excluded when the parameters are estimated with monthly data.

2.2.5. CONDITIONAL MEASURES FROM FERSON AND QIAN (2004)

In the spirit of the CFG model, Ferson and Qian (2004) (FQ, subsequently) propose a model where beta and the timing parameter are linear functions of the information variables [$z_{i,t-1}$]. The diffusion process is:

$$r_{\gamma,t} = \alpha_\gamma + b_{\gamma 0} r_{r,t} + \sum_{i=1}^{N} b_{\gamma i} \cdot (z_{i,t-1} \otimes r_{r,t}) + \beta_{\gamma 2} r_{r,t-1} + c_{\gamma 0} r_{r,t}^2 + \sum_{i=1}^{N} c_{\gamma i} \cdot (z_{i,t-1} \otimes r_{r,t}^2) + u_{\gamma,t} \quad (6)$$

Hereafter we refer to this model as the FQ model, and the parameters are estimated by OLS with Newey and West (1987) standard errors. The $b_{\gamma i}$ $c_{\gamma i}$ are respectively measuring the sensitivity of the conditional beta and the timing coefficient to $z_{i,t-1}$. Like in the CFG model, the average conditional beta and timing coefficient are $b_{\gamma 0}$ and $c_{\gamma 0}$, respectively. The FQ market timing and global measures correspond respectively to $c_{\gamma 0}$ and $\alpha_\gamma + c_{\gamma 0} \cdot \overline{r_{r,t}^2}$, with the t-1 variable excluded from the estimation with monthly data.

2.2.6. 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 (7) and (8) describe the diffusion processes for the excess returns of the reference portfolio and the managed portfolio with ability level γ, respectively. Equation (9) 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).

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

$$r_{r,t} = a_r + \sum_{i=1}^{N} a_{ri} z_{i,t-1} + e_{m,t} \ , \tag{7}$$

$$r_{\gamma,t} = a_{\gamma 0} + \sum_{i=1}^{M} a_{\gamma i} z_{i,t-1}^{a} + \frac{h_{\gamma r,t}}{h_{r,t}}(a_r + \sum_{i=1}^{N} a_{ri} z_{i,t-1}) + \beta_{\gamma 2} r_{r,t-1} + \beta_{\gamma 3} r_{r,t}^{2} + e_{\gamma,t} \ , \tag{8}$$

$$H_t = \begin{bmatrix} h_{\gamma,t} & h_{\gamma r,t} \\ h_{\gamma r,t} & h_{r,t} \end{bmatrix} = C^{'}C + A^{'}e_{t-1}e_{t-1}^{'}A + B^{'}H_{t-1}B + G^{'}\eta_{t-1}\eta_{t-1}^{'}G \ , \tag{9}$$

where:

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).

This system of equations is estimated by quasi-maximum likelihood with robust standard errors following Bollerslev and Wooldridge (1992).[8] With daily data, we define the conditional BiGARCH market timing and global measures as respectively $\beta_{\gamma 3}$ and $a_{\gamma 0} + \beta_{\gamma 3} \cdot \overline{r^2_{m,t}}$. We do not consider an estimation with monthly data since the GARCH specification is more appropriate for high-frequency returns and a large number of observations (Nelson, 1990).

2.2.7. ROBUSTNESS CHECK: STALE PRICING TIMING MODELS

As a further control for the stale pricing issues discussed by Scholes and Williams (1977), Chen, Ferson and Peters (2010) suggest augmenting the diffusion process for day *t* with a squared return at day *t*-1. Thus, their market timing measure using daily returns becomes the sum of the coefficients on the squared returns for days *t* and *t*-1. As a robustness check, we implement this so-called 'stale pricing timing' version for four of the six models. (The two exceptions are the Comer and FQ models because of their already large number of timing related parameters.) For example, for the stale pricing timing version of the unconditional CAPM, the diffusion process can be written as follows:

$$r_{\gamma,t} = \alpha_\gamma + \beta_{\gamma 1} r_{r,t} + \beta_{\gamma 2} r_{r,t-1} + \beta_{\gamma 3} r^2_{r,t} + \beta_{\gamma 4} r^2_{r,t-1} + \varepsilon_{\gamma,t} \ , \tag{10}$$

[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.

[8] As for other complex multivariate systems, estimation of the BiGARCH specification does not always reach convergence. To facilitate comparison, we report results across performance models using the subsample of simulated timers for which BiGARCH convergence is reached. In unreported results, we also compile the results for the other models using the full simulated sample and obtain similar findings.

With the control for stale pricing timing, the CAPM market timing measure is now given by $\beta_{\gamma 3} + \beta_{\gamma 4}$, while the CAPM global performance measure becomes $\alpha_\gamma + (\beta_{\gamma 3} + \beta_{\gamma 4})\overline{r_{r,t}^2}$. It is straightforward to obtain the stale pricing timing version of the other models.

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, using the simulated returns of 1000 daily, occasional and monthly timers, we estimate every performance measure with daily and monthly data sampling frequencies. If the measure properly accounts 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 across the evaluations with two statistics: 1- We compute a *t*-statistic on the significance of the mean performance value; 2- We compile the proportion of significant *t*-statistics at the 5% threshold.

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:

$$IC = \frac{\sum_{i=1}^{k} 2\left(\overline{Rank}_i - \frac{k+1}{2}\right)^2}{k(k+1)/12}, \tag{11}$$

where k is the number of ranked timers and $\overline{Rank}_i$ is, for timer i, the average of two ranks: 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-squared distribution with k-1degrees of freedom. For each performance measure and each ability level γ, we summarize the results across the evaluations with two statistics: 1- We provide the mean p-value associated with the *IC* statistics; 2- We report the proportion of significant p-values at the 5% threshold.

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 reference portfolio r_r assumes an allocation of 33.3% in each 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[9] available on Bloomberg. For each index, excess returns over the one-month Treasury bill are computed. 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 descriptive statistics on the variables used in this study.

Panel A reports daily and monthly statistics (in percentage) on the excess returns of the reference portfolio and the stock, bond and money indexes, as well as the returns of the risk-free asset and the values of the lagged information variables. Although the sample contains the recent "subprime" recession, the market data show the expected risk-return trade off. The daily (monthly) mean excess returns are 0.021%

[9] For an example on using this rate to form a money market index, see McCauley (2001). The daily return is computed as $\frac{\left(1+LIBOR_{t-1}/260\right)^{65}}{\left(1+LIBOR_t/260\right)^{64}}-1$, where 65 and 260 are the average number of trading days in three months and a year, respectively.

(0.357%) for the stock index, 0.008% (0.182%) for the bond index and 0.02% (-0.026%) for the money market index. In addition, the daily (monthly) standard deviations of excess returns are 1.366% (4.549%) for the stock index, 0.256% (1.155%) for the bond index and 0.009% (0.116%) for the money market index.

Panel B examines the daily and monthly portfolio excess returns of the daily timers, who trade every day, the occasional timers, who trade twice per month on random days, and the monthly timers, who trade every month. The statistics are the averages and standard deviations (in parenthesis) across all 1000 simulations when ability level γ equals 0.00, 0.05, 0.10, 0.20 and 0.40.[10] First, the results highlight that, even with lower ability levels, the daily timers enjoy a larger number of opportunities to time the indexes, resulting in a higher mean excess return than the occasional or monthly timers. For example, looking at daily or monthly data, an ability level of 0.05 for a daily timer produces similar excess returns (annualized mean and standard deviation of approximately 5.3% and 6.2%, respectively) than an occasional timer with an ability of 0.20 or a monthly timer with an ability of 0.40. Second, as the ability level is increasing, the mean excess return goes up and the standard deviation goes down. When the ability level is high, noise is less important and the number of times that the managers take the best decisions is high, resulting in simulated timers performing well and acting similarly across simulations[11].

4. EMPIRICAL RESULTS

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

4.1. DESCRIPTIVE STATISTICS OF THE PERFORMANCE MEASURES

[10] These ability levels are chosen as an illustration, but our full simulations let the level vary from $\gamma = 0$ to $\gamma = 0.49$ by 0.01.

[11] A notable exception occurs for the monthly standard deviations of the daily timers, which increase with the ability level. At a high ability level, the daily frequency of informed trading generates numerous high daily returns. Their composition to monthly returns produces some large right-tail returns (positive asymmetry), which result in a high standard deviation.

Table 2 gives descriptive statistics on the performance measures of the simulated market timers with daily, occasional or monthly transactions. The market timing (denoted T) and global (denoted G) measures for the six models described in section 2 are estimated with either daily data or monthly data. The table provides the mean and standard deviation (in parenthesis) of the evaluations for five ability levels γ (0, 0.05, 0.10, 0.20 and 0.40) chosen to illustrate the effect of γ on the performance. The market timing measure indicates whether the relation between the timer's returns and the reference portfolio's returns is linear (T = 0), convex (T > 0) or concave (T < 0). A convex relation represents a favourable timing evaluation: the timer's return increases more for a given positive change in the reference portfolio return than it decreases for a negative reference portfolio return change of similar magnitude. The global measure captures both the alpha and timing components of the timer's mean return. It can be interpreted as an excess risk-adjusted return of the timer's portfolio over the benchmark implicit in the performance model. For example, for $\gamma = 0.05$, the CAPM global measure for the daily timer in table 2 indicates a monthly evaluation of 0.252% (an annualized return of around 3%), which represents the risk-adjusted version of the corresponding mean excess return of 0.441% (an annualized value around 5.3%) taken in panel B of table 1.

Table 2 shows the following results. First, similar to the average simulated returns shown in panel B of table 1, the average performance evaluations consistently increase with the ability level γ. Importantly, for all models, the market timing and global measures are reassuringly close to zero when $\gamma = 0$ (i.e. when the trading signal reflects only noise). They increase gradually to reveal highly convex relations with large excess returns when $\gamma = 0.4$. The only exception is the difficult-to-estimate BiGARCH evaluations for the monthly timers, which are counterintuitive.

Second, consistent with the frequency of informative signals (and also reflected in the simulated returns summarized in panel B of table 1), the average evaluations show superior performance for the daily timers than the occasional or monthly timers. For example, for the ability level $\gamma = 0.4$, the CAPM global measures indicate monthly excess risk-adjusted returns of 2.051% for the daily timers, 0.453% for the occasional timers and 0.304% for the monthly timers. Put differently, to obtain a CAPM monthly global performance of 0.25%, the timers need signals with informative proportion of $\gamma \approx 0.05$ (daily timer), $\gamma \approx 0.23$ (occasional timer) and $\gamma \approx 0.33$ (monthly timer).

Third, looking across models for a given data sampling frequency, the average evaluations are relatively similar at equivalent ability levels, except for the lower market timing values of the Comer model (attributed to our reporting of only the stock market portion of the timing). There is nevertheless some evidence of higher market timing evaluations for the FQ and BiGARCH models with daily data. The results for these conditional models are in accordance with a common finding in the literature that conditional performance measures tend to produce higher evaluation than their unconditional counterparts for mutual funds (Ferson and Schadt, 1996; Beaulieu, Coggins and Gendron, 2009). But the results in table 2 also show that such finding cannot be generalized to all conditional models and both data sampling frequencies.

Fourth, the standard deviations of the performance evaluations consistently decrease with the ability level γ across all models and both data sampling frequencies, so that the timers show more similar evaluations across simulations when noise is a less important part of the trading signals. Consistent with the frequency of signals, the decrease is larger for the daily timers than the other timers. The standard deviations are also notably higher for the conditional models than the unconditional ones. This finding is consistent with the notion that information variables with low forecasting power can make the estimation more imprecise. In our setup, the manager's trading signal only accounts for them indirectly through their predictive ability for the future returns.

4.2. Formal Comparison of Performance Evaluation Detection

To convey our main findings clearly, we first present a detailed analysis of the results for the CAPM performance measures (shown in table 3). Then, we expand our analysis to the other models (summarized in table 4) for further confirmation.

Table 3 shows *t*-statistics on the mean performance values (panel A) and the proportions of significant *t*-statistics at the 5% threshold (panel B) of the measures (market timing, denoted T, and global, denoted G) evaluated from the CAPM. The three types of market timers with ability levels varying from 0 to 0.49 are considered. Either daily or monthly data are used in the estimation and shaded statistics indicate significance at least at the 5% level in panel A, and proportions greater than 95% in panel B.

The main findings from panels A and B are as follow. First, the market timing measure T estimated with a data sampling frequency similar to the manager's active trading frequency allows a better assessment of the performance. For the daily timers, the timing measure requires a lower ability level γ before capturing significantly the performance when estimated with daily data than with monthly data. In contrast, for the occasional and monthly timers, the monthly timing measure now requires lower ability levels than the daily timing measure for capturing a significant timing performance. Thus, for the managers who trade once or twice per month, the daily timing measures are less able to detect performance than the monthly timing measures.

Second, the global measure G is useful in detecting the performance of our timers as we obtain significant values at lower ability levels γ than with the market timing measure. This improved detection is particularly important for the daily evaluations of the occasional and monthly timers, suggesting than the use of a global (as opposed to market timing) performance measure is needed when the manager's active trading frequency is much lower than the measurement frequency. Simply put, the market timing measures are 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 and detect significant performance at a relatively similar ability level.

Third, while the previous two findings can be equivalently seen in panels A and B, it takes a relatively high ability level before 95% of the estimations detect significant performance (panel B) in contrast to only detecting significant performance on average (panel A). In particular, both the market timing and global measures lack the power to detect significant performance at low ability levels and thus might erroneously conclude that no timing ability exists in such cases. To illustrate this point, we can return to the example of a CAPM monthly risk-adjusted excess return of 0.25% (an annualized value of around 3%) highlighted in table 2, which corresponds to a daily timer with $\gamma \approx 0.05$, an occasional timer with $\gamma \approx 0.23$ and a monthly timer with $\gamma \approx 0.33$. Panel B of table 3 shows that the proportions of the estimations that detect significant performance for this excess return are between 22.7% and 76.0% for the daily timer, 16.6% and 90.3% for the occasional timer, and 7.2% and 93.6% for the monthly timer, depending on the measures (market timing or global) and the data (daily or monthly) considered.

Table 4 summarizes the results for every models of section 2 by showing the ability levels at which the *t*-statistic on the mean performance values first becomes significant at the 5% threshold (panel A) and the ability levels at which the proportion of significant *t*-statistics becomes greater than 95% (panel B). '#N/A' indicates that the ability level is greater than $\gamma = 0.49$, the maximum level used in our simulations.

Table 4 confirms that the three main findings for the CAPM in table 3 holds for the other evaluation models. First, the results for the market timing measures show the importance of matching the trading frequency of the timers with the data sampling frequency for model estimation. Second, the global measures outperforms the market timing measures in detecting significant performance. Third, all models fail to detect significant performance at low ability levels.

Further comparing the models, the three unconditional models (CAPM, Multi-Index and Comer) perform relatively similarly, with comparable ability levels across timers and data frequencies, but the three conditional models (CFG, FQ and BiGARCH) generally require higher ability levels for significant performance detection. This poor performance suggests that considering conditioning variables that controls for public information at the possibly wrong frequency might be more hurtful than helpful, an issue that calls for further investigation.[12] The FQ model requires the highest ability levels in most cases, except in its daily timing measures of the occasional and monthly timers. The BiGARCH model produces the best daily timing measure for the daily timers, but obtains among the highest ability levels in the other cases. Finally, there is no evidence that the random information arrival for the occasional timers is problematic for performance detection compare to the fixed arrival for the monthly timers.

4.3. FORMAL COMPARISON OF OBSERVED VERSUS EXPECTED RANKINGS

[12] 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.

We now turn to the ranking ability of the performance measures. We again first present a detailed analysis of the results for the CAPM performance measures (shown in table 5) and expand our analysis to the other models (summarized in table 6) for further confirmation.

Table 5 shows, for the CAPM performance measures (market timing and global), the average *p*-values of tests on the equality between the observed performance ranking and the performance ranking expected from the pre-selected ability levels (panel A) and the proportions of *p*-values inferior to the 5% threshold (panel B). The tests are based on the index of coincidence or *IC* statistic proposed by Friedman (1920), with the tabulated γ identifying the highest ability level used in the tests. Specifically, in the γ = 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. The results tabulated in the $\gamma = 0.14$ row then represent the average across simulations of the *p*-values of the tests. 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. Either daily or monthly data are used in the estimation and shaded statistics indicate significance at least at the 5% level in panel A, and proportions greater than 95% in panel B.

Table 6 summarizes the ranking results for every models of section 2 by showing the ability levels at which the average *p*-value on the *IC* statistic becomes significant at the 5% threshold (panel A) and the ability levels at which the proportion of significant *p*-values becomes greater than 95% (panel B). '#N/A' indicates that the ability level is greater than $\gamma = 0.49$, the maximum level used in our simulations.

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 measure estimated with daily data requires 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. For the occasional and monthly timers, the monthly timing measures require lower ability levels than the daily timing measures for ranking correctly at the 5% significance level. Furthermore, the ranking *p*-values of the market timing measures become significant at much higher ability levels than the ones of the global measures. Compare to panel A, panel B highlights the relatively high ability levels needed before 95% of the tests conclude that the timers of different ability are ranked correctly. Table 6 reaches these same conclusions across the different models presented in section 2, and show slightly better ranking results for the unconditional models than the conditional ones.

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 with the daily timing measures, the CAPM, Multi-Index, Comer, CGF, FQ and BiGARCH models detect significant performance with ability levels starting at 0.08, 0.08, 0.08, 0.08, 0.10 and 0.07, respectively (see table 4), while they rank the timers correctly with ability levels starting at 0.15, 0.15, 0.15, 0.15, 0.16 and 0.14, respectively (see table 6). 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).

4.4. RESULTS FOR STALE PRICING TIMING MODELS

As a robustness check proposed by Chen, Ferson and Peters (2010) and discussed in section 2.2.7, table 7 presents the results of the 'stale pricing timing' version of the four of our six models, namely the CAPM, Multi-index, CFG and BiGARCH models. Panel A focuses on performance detection while panel B examines performance ranking. Results are reported in the same format than panel A of table 4 for detection, and panel A of table 6 for ranking.

Overall, our main findings are robust to the stale pricing timing version of the models. Generally, the stale pricing timing version slightly improves the market timing measure T, reducing the ability level needed for significant performance detection and ranking. The reductions are no more than 0.01 for the daily timers and are larger for the occasional timers. However, the inclusion of the stale pricing timing controls does not improve (and sometimes worsens) the results for the global measure, which remains the most powerful one.

4.5. DISCUSSION

In summary, 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 lower than the measurement frequency. Finally, all models lack power to detect and rank performance significantly at low ability levels, with conditional measures that account for information at another frequency that the one considered by the market timers appearing the most problematic.

Numerous performance studies focus their attention on the model or benchmark choices and on the use of conditional versus unconditional measures. In our particular setup, the model choice has little impact, a conclusion also reached by Farsworth, Ferson, Jackson and Todd (2002) for stochastic discount factor performance measures. Instead, our results emphasize the importance of a much less investigated issue. They suggest that the frequency of data sampling for performance measurement, and in particular how it matches with the frequency of informed trading of the manager, deserves more attention in the evaluation process. This finding expands the analysis of Goetzmann, Ingersoll and Ivković (2000), which focuses on the monthly measurement of daily timers. It also revisits the conclusion of Bollen and Busse (2001) on the benefits of using daily instead of monthly data in performance measurement, as our analysis shows that the frequency of informed trading could be crucial in evaluating these benefits.

Our results are also favorable to the global performance measure advanced by Bollen and Busse (2004), as it outperforms the more common Treynor-Mazuy market timing measure in all situations. While the global measure is design to captures both asset selectivity and market timing, the latter ability is the exclusive focus of the Treynor-Mazuy measure. When there is an important mismatch between informed trading and measurement frequencies, we find that the timing activities of the managers generate instead a considerable selectivity component that favors the global measure. However, even when there is no such mismatch, we find that the global measure is still better, potentially due to the functional form of the trading signal and its non-linear mapping into portfolio weights. These findings add to the analytical results of Lehmann and Timmermann (2007) on the difficulty of separating market timing ability from total performance, and to the methodological exploration of Chen, Ferson and Peters (2010), who investigate the effects of assets with nonlinearities, interim trading, public information and stale pricing in the context of timing in bond mutual funds.

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 six different models. We study either market timing measures with a squared 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, occasional timers, who trade two times per month on random days, and monthly timers, who trade at the end of each month.

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 work relatively well. However, for the occasional and the monthly timers, the global performance measures perform much better as they show more robustness to differences in trading and data sampling frequencies. Nevertheless, for low pre-selected ability levels, all performance measures lack power. 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, and by controls for stale pricing in the timing measures.

Overall, we conclude that while the performance model is an important choice in the evaluation process, the frequency of data sampling for measurement, and in particular how it matches with the frequency of informed trading of the manager, deserves more attention.

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.

Lehmann, B., & Timmermann, A. (2007). Performance Measurement and Evaluation. In *Handbook of Financial Intermediation and Banking*, eds. Boot, A. & Thakor, A. Elsevier.

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.

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 the Data Series and Simulated Returns

Panel A: Data Series

	Daily Data				Monthly Data			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
A. Indexes:								
Reference Portfolio	0.011	0.446	-3.022	3.846	0.194	1.603	-6.645	3.865
Stock Market	0.021	1.366	-9.000	11.510	0.357	4.549	-18.493	11.060
Bond Market	0.008	0.256	-1.261	1.335	0.182	1.155	-3.431	3.643
Money Market	0.002	0.009	-0.066	0.103	-0.026	0.116	-0.271	0.441
Risk-Free Asset	0.010	0.007	0.000	0.022	0.208	0.146	0.000	0.483
B. Lagged Instruments:								
Variation in 3-Month Rates	-0.001	0.072	-0.810	0.760	-0.014	0.241	-0.890	0.450
Liquidity Measure	0.423	0.529	-0.030	3.730	3.029	1.762	0.220	5.930

Panel B: Simulated Returns

	Daily Timer		Occasional Timer		Monthly Timer	
γ	Daily Data	Monthly Data	Daily Data	Monthly Data	Daily Data	Monthly Data
0.00	0.011	0.218	0.011	0.215	0.010	0.212
	(0.500)	(1.879)	(0.501)	(1.792)	(0.501)	(1.751)
0.05	0.021	0.441	0.013	0.259	0.012	0.241
	(0.500)	(1.783)	(0.496)	(1.786)	(0.494)	(1.734)
0.10	0.033	0.696	0.015	0.311	0.013	0.270
	(0.499)	(1.802)	(0.490)	(1.782)	(0.487)	(1.713)
0.20	0.057	1.222	0.020	0.416	0.016	0.339
	(0.495)	(2.001)	(0.477)	(1.778)	(0.470)	(1.663)
0.40	0.101	2.202	0.031	0.658	0.023	0.492
	(0.488)	(2.461)	(0.451)	(1.745)	(0.440)	(1.591)

NOTE: This table presents descriptive statistics (in percentage) of the daily and monthly series from January 1, 2003 to July 1, 2009. The daily data include 1694 observations while the monthly data result in 79 observations. Panel A displays the mean, standard deviation, minimum and maximum for the indexes (in A) and for the lagged instruments (in B). The indexes refer to the excess returns of the reference portfolio, the stock market index, the bond market index and the money market index, and the return on the risk-free asset (the one-month Treasury bill). The lagged instruments are the two information variables used in the conditional models, namely the variation in the 3-month rates and the liquidity measure computed as the difference between the yield on AA commercial papers and the short-term Treasury bills. Panel B shows the mean and standard deviation (in parenthesis) of the 1000 simulated excess returns of the timers with ability levels γ equal to 0.00, 0.05, 0.10, 0.20 and 0.40. The daily, occasional and monthly timers trade respectively every day, twice a month on random days and at the end of every month. The simulation procedure is described in section 2.

Table 2: Estimates of Performance Evaluation for Simulated Timers

		Daily Timer				Occasional Timer				Monthly Timer			
		Daily Data		Monthly Data		Daily Data		Monthly Data		Daily Data		Monthly Data	
Model	γ	T	G	T	G	T	G	T	G	T	G	T	G
CAPM	0.00	-0.017	0.000	-0.034	0.021	-0.004	0.000	-0.013	0.024	0.015	0.000	-0.055	0.014
		(0.500)	(0.005)	(0.519)	(0.110)	(0.279)	(0.005)	(0.423)	(0.093)	(0.236)	(0.005)	(0.426)	(0.096)
	0.05	0.435	0.010	0.456	0.252	0.038	0.002	0.092	0.072	0.026	0.001	-0.004	0.043
		(0.454)	(0.005)	(0.516)	(0.107)	(0.272)	(0.005)	(0.419)	(0.093)	(0.231)	(0.005)	(0.420)	(0.092)
	0.10	0.908	0.022	1.001	0.521	0.071	0.005	0.190	0.115	0.029	0.003	0.122	0.077
		(0.367)	(0.005)	(0.547)	(0.120)	(0.268)	(0.005)	(0.400)	(0.093)	(0.233)	(0.005)	(0.418)	(0.091)
	0.20	1.649	0.046	1.906	1.057	0.131	0.010	0.392	0.213	0.045	0.006	0.261	0.144
		(0.200)	(0.005)	(0.428)	(0.109)	(0.232)	(0.004)	(0.375)	(0.088)	(0.219)	(0.004)	(0.394)	(0.087)
	0.40	2.157	0.091	2.845	2.051	0.259	0.021	0.749	0.453	0.099	0.014	0.628	0.304
		(0.030)	(0.003)	(0.154)	(0.071)	(0.145)	(0.003)	(0.150)	(0.068)	(0.152)	(0.003)	(0.209)	(0.065)
Multi-Index	0.00	-0.017	0.000	-0.001	0.042	-0.003	0.000	0.019	0.045	0.015	0.000	-0.023	0.035
		(0.501)	(0.005)	(0.512)	(0.108)	(0.280)	(0.005)	(0.423)	(0.091)	(0.237)	(0.005)	(0.420)	(0.097)
	0.05	0.435	0.010	0.488	0.259	0.037	0.002	0.125	0.092	0.026	0.001	0.029	0.063
		(0.454)	(0.005)	(0.509)	(0.106)	(0.273)	(0.005)	(0.419)	(0.091)	(0.231)	(0.005)	(0.415)	(0.092)
	0.10	0.908	0.022	1.030	0.513	0.070	0.005	0.224	0.133	0.028	0.003	0.154	0.098
		(0.368)	(0.005)	(0.544)	(0.116)	(0.268)	(0.005)	(0.400)	(0.091)	(0.233)	(0.004)	(0.413)	(0.092)
	0.20	1.649	0.046	1.935	1.018	0.130	0.009	0.429	0.227	0.043	0.006	0.292	0.163
		(0.200)	(0.005)	(0.429)	(0.105)	(0.232)	(0.004)	(0.376)	(0.085)	(0.219)	(0.004)	(0.389)	(0.088)
	0.40	2.156	0.091	2.880	1.975	0.256	0.021	0.789	0.461	0.094	0.014	0.658	0.321
		(0.030)	(0.003)	(0.157)	(0.070)	(0.145)	(0.003)	(0.151)	(0.067)	(0.151)	(0.003)	(0.210)	(0.064)
Comer	0.00	-0.002	0.000	-0.003	0.043	0.000	0.000	-0.001	0.045	0.003	0.000	-0.005	0.035
		(0.055)	(0.005)	(0.058)	(0.107)	(0.035)	(0.005)	(0.056)	(0.089)	(0.031)	(0.005)	(0.058)	(0.096)
	0.05	0.048	0.010	0.054	0.259	0.004	0.002	0.013	0.092	0.003	0.001	0.004	0.064
		(0.050)	(0.005)	(0.062)	(0.104)	(0.034)	(0.005)	(0.056)	(0.090)	(0.031)	(0.005)	(0.058)	(0.091)
	0.10	0.099	0.022	0.117	0.511	0.007	0.005	0.026	0.132	0.003	0.003	0.019	0.097
		(0.040)	(0.005)	(0.069)	(0.114)	(0.034)	(0.005)	(0.053)	(0.090)	(0.032)	(0.004)	(0.057)	(0.091)
	0.20	0.178	0.046	0.223	1.014	0.013	0.009	0.054	0.226	0.003	0.006	0.040	0.162
		(0.022)	(0.005)	(0.059)	(0.104)	(0.029)	(0.004)	(0.052)	(0.085)	(0.029)	(0.004)	(0.053)	(0.087)
	0.40	0.230	0.091	0.343	1.971	0.026	0.021	0.099	0.459	0.006	0.014	0.088	0.318
		(0.003)	(0.003)	(0.025)	(0.070)	(0.019)	(0.003)	(0.027)	(0.066)	(0.019)	(0.003)	(0.034)	(0.064)

Table 2 (Continued)

		Daily Timer				Occasional Timer				Monthly Timer			
		Daily Data		Monthly Data		Daily Data		Monthly Data		Daily Data		Monthly Data	
Model	γ	T	G	T	G	T	G	T	G	T	G	T	G
CFG	0.00	-0.016	0.000	-0.011	0.022	-0.008	0.000	-0.017	0.024	0.020	0.000	-0.023	0.015
		(0.512)	(0.005)	(0.493)	(0.107)	(0.343)	(0.004)	(0.440)	(0.091)	(0.283)	(0.003)	(0.402)	(0.095)
	0.05	0.433	0.010	0.337	0.251	0.042	0.002	0.074	0.071	0.032	0.001	0.030	0.044
		(0.458)	(0.005)	(0.494)	(0.105)	(0.333)	(0.004)	(0.430)	(0.091)	(0.277)	(0.004)	(0.397)	(0.091)
	0.10	0.894	0.022	0.711	0.519	0.075	0.004	0.139	0.114	0.033	0.002	0.150	0.078
		(0.373)	(0.005)	(0.529)	(0.119)	(0.327)	(0.004)	(0.421)	(0.092)	(0.281)	(0.004)	(0.383)	(0.091)
	0.20	1.630	0.046	1.349	1.055	0.144	0.008	0.284	0.211	0.047	0.005	0.278	0.145
		(0.223)	(0.005)	(0.437)	(0.110)	(0.279)	(0.004)	(0.391)	(0.088)	(0.262)	(0.003)	(0.375)	(0.087)
	0.40	2.099	0.091	2.044	2.061	0.282	0.018	0.535	0.450	0.103	0.011	0.628	0.306
		(0.040)	(0.003)	(0.202)	(0.072)	(0.177)	(0.003)	(0.207)	(0.069)	(0.181)	(0.003)	(0.243)	(0.066)
FQ	0.00	-0.013	0.000	0.008	0.028	-0.016	0.000	-0.037	0.021	0.034	0.000	0.024	0.026
		(0.535)	(0.006)	(1.117)	(0.216)	(0.776)	(0.008)	(0.828)	(0.125)	(0.537)	(0.006)	(0.783)	(0.164)
	0.05	0.518	0.011	0.146	0.200	0.136	0.003	0.043	0.064	0.068	0.001	0.066	0.052
		(0.525)	(0.006)	(1.130)	(0.217)	(0.757)	(0.008)	(0.819)	(0.124)	(0.530)	(0.006)	(0.736)	(0.156)
	0.10	1.030	0.023	0.357	0.417	0.253	0.005	0.112	0.103	0.097	0.003	0.209	0.090
		(0.562)	(0.008)	(1.313)	(0.233)	(0.730)	(0.007)	(0.829)	(0.135)	(0.536)	(0.006)	(0.749)	(0.161)
	0.20	2.071	0.050	0.673	0.848	0.494	0.011	0.248	0.190	0.189	0.006	0.325	0.152
		(0.378)	(0.006)	(1.067)	(0.194)	(0.619)	(0.007)	(0.750)	(0.129)	(0.504)	(0.005)	(0.738)	(0.158)
	0.40	3.132	0.099	0.594	1.642	0.916	0.024	0.489	0.412	0.438	0.014	0.668	0.306
		(0.091)	(0.004)	(0.487)	(0.107)	(0.357)	(0.005)	(0.500)	(0.111)	(0.394)	(0.004)	(0.545)	(0.121)
BiGARCH	0.00	0.066	0.001			0.009	-0.001			-0.033	-0.001		
		(0.431)	(0.007)			(0.201)	(0.004)			(0.242)	(0.004)		
	0.05	0.522	0.011			0.052	0.000			-0.041	-0.001		
		(0.407)	(0.007)			(0.205)	(0.004)			(0.235)	(0.004)		
	0.10	1.020	0.023			0.083	0.001			-0.047	-0.001		
		(0.329)	(0.006)			(0.211)	(0.004)			(0.238)	(0.004)		
	0.20	1.908	0.045			0.173	0.003			-0.077	-0.001		
		(0.250)	(0.006)			(0.217)	(0.004)			(0.228)	(0.004)		
	0.40	2.993	0.087			0.308	0.008			-0.097	-0.001		
		(0.250)	(0.005)			(0.198)	(0.004)			(0.203)	(0.003)		

NOTE: This table shows descriptive statistics on the estimates of the market timing (T) and global (G) performance evaluations for the daily, occasional and monthly simulated timers. The daily data include 1694 observations while the monthly data result in 79 observations. The performance models are the unconditional CAPM, the unconditional Multi-Index model, the unconditional multi-index timing model of Comer (2006), the conditional model of Christophersen, Ferson and Glassman (1998) (CFG), the conditional model of Ferson and Qian (2004) (FQ) and the conditional BiGARCH model. In each case, the mean and standard deviation (in parenthesis) of the estimates for ability levels γ equal to 0.00, 0.05, 0.10, 0.20 and 0.40 are reported. The simulation procedure and performance evaluation measures are described in section 2. The estimation data are presented in table 1.

Table 3: Detailed Evaluation Ability of the CAPM Performance Measures

Panel A: *t*-statistics on the mean performance values

	CAPM											
	Daily Timer				Occasional Timer				Monthly Timer			
	Daily Data		Monthly Data		Daily Data		Monthly Data		Daily Data		Monthly Data	
γ	T	G	T	G	T	G	T	G	T	G	T	G
0	-0.035	-0.028	-0.065	0.194	-0.013	0.011	-0.030	0.262	0.065	-0.082	-0.130	0.146
0.01	0.235	0.456	0.209	0.654	0.055	0.127	0.016	0.367	0.033	-0.012	-0.082	0.237
0.02	0.353	0.796	0.335	0.980	-0.006	0.181	0.062	0.413	0.160	0.081	-0.143	0.345
0.03	0.590	1.323	0.595	1.490	0.063	0.342	0.143	0.573	0.093	0.169	-0.005	0.413
0.04	0.826	1.817	0.744	1.982	0.078	0.414	0.212	0.655	0.084	0.198	0.061	0.444
0.05	0.959	2.219	0.883	2.354	0.139	0.520	0.221	0.773	0.115	0.217	-0.009	0.468
0.06	1.261	2.758	1.138	2.925	0.117	0.594	0.269	0.838	0.078	0.317	0.093	0.581
0.07	1.585	2.954	1.351	3.025	0.200	0.693	0.277	0.934	0.124	0.333	0.113	0.579
0.08	1.850	3.220	1.554	3.307	0.206	0.796	0.360	1.030	0.130	0.410	0.176	0.649
0.09	2.183	3.702	1.693	3.797	0.266	0.865	0.413	1.089	0.304	0.598	0.109	0.892
0.1	2.472	4.260	1.830	4.335	0.264	0.989	0.475	1.234	0.126	0.602	0.292	0.851
0.11	3.028	4.681	2.310	4.699	0.291	1.099	0.515	1.329	0.004	0.545	0.317	0.791
0.12	3.387	5.056	2.386	5.103	0.267	1.111	0.538	1.349	0.147	0.705	0.319	0.976
0.13	3.713	5.694	2.545	5.730	0.360	1.314	0.689	1.538	0.144	0.793	0.402	1.056
0.14	4.265	6.185	2.931	6.218	0.377	1.505	0.761	1.718	0.145	0.867	0.407	1.137
0.15	4.684	6.864	3.170	6.699	0.415	1.618	0.766	1.858	0.094	1.005	0.481	1.284
0.16	5.173	7.165	3.336	7.154	0.502	1.650	0.849	1.870	0.174	1.049	0.589	1.314
0.17	5.938	7.925	3.617	7.920	0.511	1.874	0.881	2.095	0.137	0.986	0.598	1.243
0.18	6.432	8.347	3.903	8.379	0.539	1.913	0.973	2.127	0.198	1.173	0.563	1.453
0.19	7.514	9.817	4.219	9.843	0.575	2.168	1.014	2.412	0.196	1.247	0.615	1.539
0.2	8.262	9.645	4.454	9.661	0.567	2.200	1.044	2.412	0.205	1.392	0.662	1.661
0.21	10.001	10.619	5.451	10.676	0.651	2.435	1.144	2.674	0.174	1.413	0.717	1.696
0.22	10.541	11.365	5.455	11.284	0.693	2.500	1.331	2.726	0.140	1.497	0.828	1.766
0.23	11.392	11.886	5.886	11.818	0.741	2.700	1.326	2.922	0.233	1.756	1.019	2.029
0.24	12.899	12.370	6.225	12.334	0.736	2.823	1.421	3.009	0.210	1.820	1.080	2.127
0.25	15.079	13.486	7.315	13.949	0.853	3.003	1.609	3.188	0.222	1.895	1.148	2.188
0.26	17.850	15.173	8.286	15.533	0.865	3.179	1.768	3.425	0.259	2.108	1.219	2.397
0.27	19.411	15.044	8.989	15.533	0.906	3.311	1.770	3.519	0.261	2.201	1.272	2.497
0.28	22.997	15.794	8.850	16.357	0.940	3.454	1.953	3.629	0.357	2.424	1.122	2.791
0.29	25.348	17.361	9.664	17.723	0.983	3.907	2.007	4.138	0.319	2.450	1.452	2.748
0.3	27.538	19.201	10.251	19.640	1.127	4.039	2.298	4.188	0.279	2.497	1.422	2.764
0.31	30.469	17.909	11.156	18.343	1.135	4.054	2.464	4.226	0.316	2.735	1.681	3.033
0.32	34.126	19.988	11.350	20.380	1.234	4.371	2.554	4.534	0.328	3.123	1.996	3.419
0.33	37.630	21.280	12.706	21.640	1.309	4.463	2.896	4.576	0.422	2.918	1.784	3.209
0.34	39.468	21.478	12.615	22.044	1.478	4.999	2.913	5.126	0.449	3.115	2.021	3.396
0.35	46.884	21.695	13.401	22.336	1.442	5.037	3.543	5.159	0.407	3.461	2.327	3.762
0.36	49.443	24.201	15.551	25.016	1.566	5.207	3.201	5.361	0.492	3.771	2.437	4.059
0.37	50.316	23.949	14.513	24.423	1.848	5.620	3.760	5.812	0.541	3.781	2.579	4.100
0.38	58.876	24.732	16.315	24.975	1.689	5.549	3.947	5.697	0.550	4.037	2.736	4.257
0.39	68.250	25.871	17.371	26.849	1.715	6.334	4.290	6.464	0.595	4.422	2.843	4.781
0.4	71.110	27.881	18.422	29.015	1.781	6.463	5.010	6.656	0.652	4.416	3.003	4.693
0.41	83.681	28.278	19.782	28.660	1.901	6.936	5.039	7.016	0.631	4.906	3.309	5.173
0.42	88.296	28.507	20.137	29.237	1.870	7.112	5.430	7.142	0.673	5.046	3.760	5.288
0.43	98.962	30.131	21.515	30.948	2.157	7.289	5.382	7.504	0.760	5.513	4.232	5.779
0.44	113.278	32.910	22.744	33.767	2.201	7.589	5.712	7.716	0.761	5.843	4.368	6.162
0.45	113.225	33.006	23.073	34.015	2.163	7.986	6.793	8.029	0.908	5.819	4.573	6.135
0.46	130.402	36.836	26.809	37.792	2.345	8.299	6.642	8.475	1.058	6.470	5.172	6.693
0.47	140.549	37.334	25.645	38.351	2.385	8.546	8.308	8.618	1.055	6.839	5.505	7.148
0.48	152.452	35.956	26.758	36.773	2.489	8.805	8.086	8.818	1.142	6.845	6.186	7.118
0.49	170.969	39.373	28.203	40.686	2.543	9.826	8.493	9.929	1.232	7.542	6.608	7.662

Table 3 (Continued)

Panel B: Proportions of significant *t*-statistics at 5% threshold

	CAPM											
	Daily Timer				Occasional Timer				Monthly Timer			
	Daily Data		Monthly Data		Daily Data		Monthly Data		Daily Data		Monthly Data	
γ	T	G	T	G	T	G	T	G	T	G	T	G
0	4.0%	5.0%	5.8%	7.4%	4.3%	4.6%	4.0%	8.5%	5.6%	4.6%	3.7%	7.6%
0.01	8.5%	9.9%	9.3%	14.3%	4.0%	6.5%	6.0%	10.3%	4.0%	3.6%	3.0%	7.6%
0.02	10.9%	19.6%	11.3%	24.5%	3.5%	6.5%	7.0%	11.5%	7.6%	7.0%	2.0%	10.4%
0.03	16.1%	35.8%	15.3%	45.1%	3.5%	10.3%	7.4%	15.3%	5.9%	7.7%	3.2%	13.2%
0.04	24.0%	54.9%	18.5%	61.5%	2.9%	11.2%	8.5%	18.0%	6.1%	7.6%	3.1%	12.1%
0.05	30.0%	70.9%	22.7%	76.0%	5.2%	13.8%	8.9%	19.0%	6.5%	8.3%	3.2%	14.4%
0.06	39.0%	86.6%	30.4%	88.8%	4.2%	13.2%	9.7%	19.2%	5.8%	10.0%	3.8%	16.8%
0.07	50.9%	90.0%	37.0%	92.5%	5.9%	19.5%	8.4%	25.3%	7.0%	10.5%	3.7%	16.7%
0.08	63.6%	94.9%	46.4%	96.2%	5.8%	20.4%	11.5%	27.3%	6.1%	11.3%	4.6%	16.4%
0.09	73.6%	98.3%	52.0%	98.8%	6.9%	22.2%	10.5%	28.8%	11.2%	15.2%	3.8%	23.8%
0.1	80.7%	99.6%	58.5%	99.9%	7.2%	26.4%	12.8%	34.9%	7.3%	17.6%	7.1%	25.7%
0.11	89.1%	99.7%	74.6%	99.9%	8.0%	29.0%	12.8%	37.3%	5.5%	15.0%	7.5%	20.9%
0.12	94.1%	100.0%	77.9%	100.0%	8.0%	28.7%	14.0%	38.9%	7.0%	20.9%	7.0%	26.2%
0.13	95.6%	100.0%	81.6%	100.0%	7.2%	35.3%	19.7%	46.9%	8.4%	20.6%	8.6%	30.1%
0.14	98.6%	100.0%	89.9%	100.0%	7.9%	44.6%	21.4%	53.1%	6.9%	23.5%	7.9%	32.4%
0.15	98.8%	100.0%	94.5%	100.0%	9.4%	48.1%	21.1%	57.9%	6.6%	25.3%	9.4%	35.3%
0.16	99.2%	100.0%	96.0%	100.0%	10.7%	51.4%	25.0%	61.7%	7.5%	31.5%	12.5%	38.9%
0.17	99.9%	100.0%	97.8%	100.0%	10.8%	59.6%	25.0%	67.6%	7.7%	27.7%	13.4%	37.7%
0.18	100.0%	100.0%	98.7%	100.0%	13.2%	62.1%	31.3%	70.1%	7.8%	35.9%	12.6%	45.6%
0.19	100.0%	100.0%	99.9%	100.0%	13.8%	70.2%	31.2%	76.4%	8.1%	35.4%	16.7%	46.9%
0.2	100.0%	100.0%	99.6%	100.0%	11.8%	71.6%	31.6%	79.1%	9.3%	40.1%	16.4%	49.8%
0.21	100.0%	100.0%	100.0%	100.0%	13.6%	78.5%	38.5%	83.8%	6.6%	42.8%	18.6%	51.0%
0.22	100.0%	100.0%	100.0%	100.0%	15.4%	80.8%	46.2%	85.3%	6.1%	46.1%	22.0%	54.1%
0.23	100.0%	100.0%	100.0%	100.0%	16.6%	87.0%	47.3%	90.3%	7.6%	56.7%	31.6%	67.3%
0.24	100.0%	100.0%	100.0%	100.0%	16.7%	88.4%	52.0%	91.1%	6.4%	59.5%	35.1%	68.6%
0.25	100.0%	100.0%	100.0%	100.0%	22.2%	90.5%	60.1%	93.2%	5.6%	61.1%	37.5%	68.1%
0.26	100.0%	100.0%	100.0%	100.0%	24.6%	93.0%	64.2%	96.0%	7.8%	69.3%	42.2%	77.1%
0.27	100.0%	100.0%	100.0%	100.0%	24.4%	94.8%	66.1%	96.4%	6.4%	69.7%	44.3%	79.2%
0.28	100.0%	100.0%	100.0%	100.0%	25.4%	95.5%	71.6%	96.7%	8.8%	79.3%	35.1%	85.6%
0.29	100.0%	100.0%	100.0%	100.0%	28.9%	97.8%	73.9%	98.9%	7.0%	76.2%	53.8%	84.4%
0.3	100.0%	100.0%	100.0%	100.0%	31.8%	98.7%	79.8%	99.0%	6.6%	80.7%	49.5%	86.1%
0.31	100.0%	100.0%	100.0%	100.0%	32.3%	99.1%	83.7%	99.3%	6.1%	85.1%	62.0%	91.6%
0.32	100.0%	100.0%	100.0%	100.0%	37.3%	98.6%	84.0%	98.8%	6.1%	90.8%	72.7%	95.2%
0.33	100.0%	100.0%	100.0%	100.0%	40.3%	99.1%	89.2%	99.4%	7.2%	90.0%	65.0%	93.6%
0.34	100.0%	100.0%	100.0%	100.0%	44.1%	99.7%	88.9%	99.7%	12.6%	92.9%	77.2%	95.0%
0.35	100.0%	100.0%	100.0%	100.0%	43.7%	99.7%	95.4%	99.7%	7.4%	94.6%	82.7%	97.4%
0.36	100.0%	100.0%	100.0%	100.0%	49.1%	100.0%	92.1%	100.0%	11.8%	97.1%	83.9%	98.4%
0.37	100.0%	100.0%	100.0%	100.0%	56.2%	100.0%	95.0%	100.0%	14.2%	97.3%	87.7%	98.8%
0.38	100.0%	100.0%	100.0%	100.0%	52.9%	99.8%	95.2%	100.0%	13.5%	99.3%	90.5%	99.3%
0.39	100.0%	100.0%	100.0%	100.0%	54.1%	100.0%	97.0%	100.0%	16.7%	99.6%	90.5%	100.0%
0.4	100.0%	100.0%	100.0%	100.0%	53.4%	100.0%	98.4%	100.0%	18.7%	99.3%	91.8%	99.8%
0.41	100.0%	100.0%	100.0%	100.0%	58.2%	100.0%	99.1%	100.0%	18.2%	100.0%	94.6%	100.0%
0.42	100.0%	100.0%	100.0%	100.0%	55.2%	100.0%	98.9%	100.0%	20.3%	100.0%	97.3%	100.0%
0.43	100.0%	100.0%	100.0%	100.0%	65.4%	100.0%	98.4%	100.0%	23.3%	100.0%	98.1%	100.0%
0.44	100.0%	100.0%	100.0%	100.0%	67.3%	100.0%	99.3%	100.0%	21.9%	100.0%	98.1%	100.0%
0.45	100.0%	100.0%	100.0%	100.0%	66.5%	100.0%	100.0%	100.0%	28.4%	100.0%	98.7%	100.0%
0.46	100.0%	100.0%	100.0%	100.0%	75.1%	100.0%	99.4%	100.0%	32.3%	100.0%	99.2%	100.0%
0.47	100.0%	100.0%	100.0%	100.0%	75.7%	100.0%	100.0%	100.0%	32.8%	100.0%	99.8%	100.0%
0.48	100.0%	100.0%	100.0%	100.0%	78.4%	100.0%	99.8%	100.0%	35.1%	100.0%	99.8%	100.0%
0.49	100.0%	100.0%	100.0%	100.0%	79.5%	100.0%	100.0%	100.0%	37.1%	100.0%	100.0%	100.0%

NOTES: This table reports statistics on the significance of the performance evaluations from the market timing (T) and global (G) measures obtained from the unconditional CAPM. Daily, occasional and monthly timers with ability levels varying from $\gamma = 0$ to $\gamma = 0.49$ are evaluated using daily and monthly data. Panel A shows *t*-statistics on the mean performance values across simulations for every ability level γ. Shaded statistics indicate significance at the 5% level (lightest shade), 2.5% level (middle shade) and 1% level (darkest shade). Panel B reports the proportions of significant *t*-statistics at the 5% threshold across simulations. Shaded statistics indicate proportions greater than 95%. The simulation procedure and performance evaluation measures are described in section 2. The statistics are described in section 3.1. The estimation data are presented in table 1.

Table 4: Evaluation Ability of the Performance Measures

Panel A: Ability levels at which the *t*-statistic on the mean performance first becomes significant at the 5% threshold

	Daily Timer				Occasional Timer				Monthly Timer			
	Daily Data		Monthly Data		Daily Data		Monthly Data		Daily Data		Monthly Data	
	T	G	T	G	T	G	T	G	T	G	T	G
CAPM	0.08	0.04	0.09	0.04	0.37	0.16	0.26	0.14	#N/A	0.23	0.31	0.20
Multi-Index	0.08	0.04	0.09	0.03	0.37	0.15	0.25	0.13	#N/A	0.23	0.31	0.18
Comer	0.08	0.04	0.10	0.03	0.43	0.15	0.26	0.13	#N/A	0.23	0.31	0.18
CFG	0.08	0.04	0.11	0.04	0.41	0.18	0.33	0.14	#N/A	0.23	0.32	0.20
FQ	0.10	0.05	#N/A	0.10	0.32	0.20	#N/A	0.22	0.48	0.27	0.46	0.31
BiGARCH	0.07	0.06	---	---	0.41	0.36	---	---	#N/A	#N/A	---	---

Panel B: Ability levels at which the proportion of significant *t*-statistics becomes greater than 95%

	Daily Timer				Occasional Timer				Monthly Timer			
	Daily Data		Monthly Data		Daily Data		Monthly Data		Daily Data		Monthly Data	
	T	G	T	G	T	G	T	G	T	G	T	G
CAPM	0.13	0.09	0.16	0.08	#N/A	0.28	0.37	0.26	#N/A	0.35	0.42	0.34
Multi-Index	0.13	0.08	0.15	0.08	#N/A	0.27	0.37	0.25	#N/A	0.35	0.41	0.32
Comer	0.13	0.08	0.18	0.08	#N/A	0.27	0.38	0.25	#N/A	0.35	0.44	0.32
CFG	0.13	0.09	0.21	0.08	#N/A	0.30	0.49	0.26	#N/A	0.39	0.46	0.35
FQ	0.15	0.11	#N/A	0.17	0.46	0.32	#N/A	0.40	#N/A	0.43	#N/A	0.46
BiGARCH	0.11	0.10	---	---	#N/A	#N/A	---	---	#N/A	#N/A	---	---

NOTE: This table reports the ability levels γ associated with significant performance evaluations from the market timing (T) and global (G) measures for the daily, occasional and monthly timers, using daily and monthly data. Panel A gives the ability levels at which the *t*-statistics on the mean performance values across simulations first become significant at the 5% threshold. Panel B gives the ability levels at which the proportions across simulations of significant *t*-statistics at the 5% threshold first become greater than 95%. '#N/A' indicates that the ability level is greater than $\gamma = 0.49$, the maximum level of the simulations. The performance models are the unconditional CAPM, the unconditional Multi-Index model, the unconditional multi-index timing model of Comer (2006), the conditional model of Christophersen, Ferson and Glassman (1998) (CFG), the conditional model of Ferson and Qian (2004) (FQ) and the conditional BiGARCH model. The simulation procedure and performance evaluation measures are described in section 2. The statistics used to identify the tabulated ability levels are described in section 3.1. The estimation data are presented in table 1.

Table 5: Detailed Ranking Ability of the CAPM Performance Measures

Panel A: Average p -value of IC tests

	CAPM											
	Daily Timer				Occasional Timer				Monthly Timer			
	Daily Data		Monthly Data		Daily Data		Monthly Data		Daily Data		Monthly Data	
γ	T	G	T	G	T	G	T	G	T	G	T	G
0.01	0.505	0.445	0.493	0.456	0.560	0.543	0.549	0.541	0.596	0.563	0.550	0.561
0.02	0.398	0.325	0.406	0.333	0.452	0.432	0.441	0.429	0.441	0.435	0.467	0.433
0.03	0.362	0.260	0.356	0.270	0.439	0.398	0.430	0.394	0.444	0.426	0.444	0.419
0.04	0.330	0.212	0.330	0.218	0.440	0.381	0.418	0.378	0.449	0.415	0.426	0.411
0.05	0.304	0.164	0.301	0.172	0.436	0.367	0.417	0.363	0.450	0.413	0.432	0.403
0.06	0.271	0.130	0.267	0.135	0.440	0.346	0.411	0.343	0.460	0.393	0.420	0.387
0.07	0.236	0.101	0.231	0.107	0.424	0.335	0.401	0.330	0.467	0.384	0.413	0.380
0.08	0.198	0.081	0.197	0.085	0.421	0.312	0.392	0.308	0.468	0.371	0.400	0.367
0.09	0.167	0.064	0.171	0.067	0.412	0.289	0.374	0.289	0.444	0.344	0.395	0.342
0.1	0.141	0.051	0.148	0.053	0.404	0.269	0.362	0.268	0.444	0.326	0.382	0.324
0.11	0.113	0.041	0.123	0.042	0.398	0.247	0.343	0.247	0.462	0.319	0.365	0.320
0.12	0.091	0.033	0.102	0.034	0.397	0.228	0.335	0.229	0.466	0.304	0.349	0.302
0.13	0.073	0.027	0.084	0.028	0.388	0.209	0.316	0.210	0.469	0.286	0.332	0.284
0.14	0.058	0.022	0.069	0.022	0.382	0.182	0.301	0.181	0.469	0.261	0.316	0.258
0.15	0.046	0.018	0.056	0.018	0.373	0.159	0.287	0.159	0.473	0.241	0.306	0.237
0.16	0.036	0.015	0.046	0.015	0.360	0.141	0.272	0.140	0.471	0.220	0.284	0.217
0.17	0.028	0.012	0.037	0.012	0.351	0.120	0.258	0.120	0.472	0.208	0.271	0.205
0.18	0.022	0.010	0.030	0.010	0.342	0.103	0.243	0.104	0.465	0.193	0.266	0.189
0.19	0.018	0.008	0.024	0.008	0.331	0.087	0.226	0.088	0.461	0.178	0.257	0.173
0.2	0.014	0.007	0.020	0.007	0.323	0.076	0.212	0.077	0.458	0.159	0.244	0.154
0.21	0.011	0.006	0.016	0.006	0.312	0.063	0.199	0.064	0.459	0.142	0.229	0.138
0.22	0.009	0.005	0.013	0.005	0.304	0.052	0.182	0.053	0.464	0.128	0.213	0.124
0.23	0.007	0.004	0.010	0.004	0.292	0.044	0.168	0.044	0.459	0.109	0.198	0.106
0.24	0.006	0.003	0.009	0.003	0.282	0.036	0.151	0.037	0.456	0.095	0.179	0.092
0.25	0.005	0.003	0.007	0.003	0.269	0.029	0.135	0.030	0.453	0.083	0.161	0.079
0.26	0.004	0.002	0.006	0.002	0.259	0.024	0.121	0.025	0.449	0.070	0.148	0.067
0.27	0.003	0.002	0.005	0.002	0.250	0.020	0.108	0.020	0.445	0.059	0.133	0.056
0.28	0.003	0.002	0.004	0.002	0.239	0.016	0.097	0.016	0.434	0.049	0.123	0.047
0.29	0.002	0.001	0.003	0.001	0.231	0.013	0.087	0.013	0.426	0.041	0.112	0.039
0.3	0.002	0.001	0.003	0.001	0.220	0.010	0.077	0.010	0.423	0.035	0.102	0.033
0.31	0.001	0.001	0.002	0.001	0.209	0.008	0.068	0.008	0.418	0.029	0.092	0.027
0.32	0.001	0.001	0.002	0.001	0.200	0.007	0.060	0.007	0.414	0.023	0.080	0.022
0.33	0.001	0.001	0.002	0.001	0.189	0.005	0.051	0.005	0.405	0.019	0.072	0.018
0.34	0.001	0.001	0.001	0.001	0.178	0.004	0.045	0.004	0.397	0.015	0.064	0.014
0.35	0.001	0.000	0.001	0.000	0.167	0.003	0.038	0.003	0.392	0.012	0.055	0.011
0.36	0.001	0.000	0.001	0.000	0.156	0.003	0.033	0.003	0.386	0.010	0.048	0.009
0.37	0.000	0.000	0.001	0.000	0.146	0.002	0.028	0.002	0.375	0.008	0.042	0.007
0.38	0.000	0.000	0.001	0.000	0.137	0.002	0.024	0.002	0.366	0.006	0.035	0.006
0.39	0.000	0.000	0.001	0.000	0.130	0.001	0.021	0.001	0.356	0.005	0.030	0.005
0.4	0.000	0.000	0.000	0.000	0.123	0.001	0.018	0.001	0.346	0.004	0.027	0.004
0.41	0.000	0.000	0.000	0.000	0.117	0.001	0.015	0.001	0.337	0.003	0.023	0.003
0.42	0.000	0.000	0.000	0.000	0.111	0.001	0.013	0.001	0.327	0.002	0.020	0.002
0.43	0.000	0.000	0.000	0.000	0.103	0.001	0.011	0.001	0.316	0.002	0.017	0.002
0.44	0.000	0.000	0.000	0.000	0.097	0.000	0.010	0.000	0.307	0.002	0.014	0.001
0.45	0.000	0.000	0.000	0.000	0.090	0.000	0.008	0.000	0.293	0.001	0.012	0.001
0.46	0.000	0.000	0.000	0.000	0.084	0.000	0.007	0.000	0.277	0.001	0.010	0.001
0.47	0.000	0.000	0.000	0.000	0.078	0.000	0.006	0.000	0.263	0.001	0.008	0.001
0.48	0.000	0.000	0.000	0.000	0.073	0.000	0.005	0.000	0.249	0.001	0.007	0.001
0.49	0.000	0.000	0.000	0.000	0.068	0.000	0.004	0.000	0.235	0.000	0.005	0.000

Table 5 (Continued)

Panel B: Proportions of p-values from IC tests inferior to the 5% threshold

	CAPM											
	Daily Timer				Occasional Timer				Monthly Timer			
	Daily Data		Monthly Data		Daily Data		Monthly Data		Daily Data		Monthly Data	
γ	T	G	T	G	T	G	T	G	T	G	T	G
0.01	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.02	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.03	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.04	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.05	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.06	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.07	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
0.08	0.3%	6.8%	0.8%	5.7%	0.0%	0.3%	0.0%	0.5%	0.0%	0.7%	0.0%	0.7%
0.09	2.2%	28.7%	1.4%	26.7%	0.0%	1.0%	0.0%	0.8%	0.0%	0.2%	0.0%	0.4%
0.1	7.0%	59.5%	5.4%	54.8%	0.0%	1.0%	0.3%	1.3%	0.0%	0.4%	0.7%	0.7%
0.11	14.8%	84.4%	13.3%	81.9%	0.0%	2.0%	0.5%	2.5%	0.0%	1.1%	0.7%	1.8%
0.12	27.0%	96.3%	23.5%	94.6%	0.2%	4.2%	1.2%	4.2%	0.0%	1.8%	0.9%	1.1%
0.13	41.3%	99.0%	35.1%	98.9%	0.3%	5.5%	1.0%	5.3%	0.0%	2.2%	0.9%	1.8%
0.14	57.9%	99.8%	45.6%	99.8%	0.7%	7.3%	2.2%	6.7%	0.0%	2.9%	1.3%	2.4%
0.15	72.9%	100.0%	58.4%	100.0%	0.5%	9.5%	3.3%	8.8%	0.2%	4.0%	2.2%	4.0%
0.16	82.5%	100.0%	69.4%	100.0%	1.0%	16.6%	3.7%	15.0%	0.0%	5.5%	2.6%	5.7%
0.17	91.7%	100.0%	81.6%	100.0%	1.3%	21.8%	4.7%	20.1%	0.0%	7.5%	2.0%	9.1%
0.18	97.6%	100.0%	88.6%	100.0%	1.2%	28.8%	6.0%	26.0%	0.0%	7.9%	3.5%	8.6%
0.19	99.4%	100.0%	95.7%	100.0%	2.0%	37.4%	7.2%	35.3%	0.0%	11.9%	4.2%	13.7%
0.2	99.8%	100.0%	97.3%	100.0%	2.0%	43.6%	8.2%	42.8%	0.2%	15.2%	6.0%	17.2%
0.21	100.0%	100.0%	98.6%	100.0%	3.2%	51.7%	10.8%	51.4%	0.0%	20.1%	5.7%	19.4%
0.22	100.0%	100.0%	99.7%	100.0%	4.0%	62.2%	14.3%	61.1%	0.0%	24.1%	7.1%	24.9%
0.23	100.0%	100.0%	100.0%	100.0%	4.3%	72.4%	15.5%	72.5%	0.4%	31.1%	8.8%	31.3%
0.24	100.0%	100.0%	100.0%	100.0%	4.3%	79.2%	17.3%	78.7%	0.4%	36.4%	13.2%	36.9%
0.25	100.0%	100.0%	100.0%	100.0%	7.0%	86.0%	22.8%	85.2%	0.4%	41.5%	14.8%	44.2%
0.26	100.0%	100.0%	100.0%	100.0%	6.0%	91.2%	27.6%	91.0%	0.2%	48.8%	17.4%	51.2%
0.27	100.0%	100.0%	100.0%	100.0%	7.0%	95.7%	32.3%	94.2%	0.0%	59.2%	20.8%	62.3%
0.28	100.0%	100.0%	100.0%	100.0%	7.0%	97.5%	36.3%	96.8%	0.2%	68.2%	25.6%	70.2%
0.29	100.0%	100.0%	100.0%	100.0%	7.7%	98.5%	43.4%	98.5%	0.2%	73.7%	26.9%	75.5%
0.3	100.0%	100.0%	100.0%	100.0%	8.0%	99.2%	48.9%	99.2%	0.2%	81.9%	29.8%	83.4%
0.31	100.0%	100.0%	100.0%	100.0%	10.5%	100.0%	54.9%	99.8%	0.4%	86.8%	37.3%	88.1%
0.32	100.0%	100.0%	100.0%	100.0%	11.0%	100.0%	58.9%	100.0%	0.9%	90.9%	44.6%	92.3%
0.33	100.0%	100.0%	100.0%	100.0%	13.1%	100.0%	64.9%	100.0%	0.4%	95.6%	49.0%	96.5%
0.34	100.0%	100.0%	100.0%	100.0%	13.6%	100.0%	71.7%	100.0%	0.7%	97.4%	56.7%	98.0%
0.35	100.0%	100.0%	100.0%	100.0%	14.8%	100.0%	76.7%	100.0%	1.3%	98.2%	61.1%	98.7%
0.36	100.0%	100.0%	100.0%	100.0%	17.0%	100.0%	80.9%	100.0%	1.1%	99.3%	67.3%	99.3%
0.37	100.0%	100.0%	100.0%	100.0%	20.0%	100.0%	85.5%	100.0%	1.8%	99.8%	74.0%	99.8%
0.38	100.0%	100.0%	100.0%	100.0%	23.3%	100.0%	89.0%	100.0%	1.3%	99.8%	78.4%	99.8%
0.39	100.0%	100.0%	100.0%	100.0%	25.6%	100.0%	91.7%	100.0%	2.2%	100.0%	83.2%	100.0%
0.4	100.0%	100.0%	100.0%	100.0%	27.8%	100.0%	93.8%	100.0%	2.4%	100.0%	85.9%	100.0%
0.41	100.0%	100.0%	100.0%	100.0%	30.3%	100.0%	94.7%	100.0%	3.8%	100.0%	90.7%	100.0%
0.42	100.0%	100.0%	100.0%	100.0%	33.6%	100.0%	97.0%	100.0%	5.1%	100.0%	93.6%	100.0%
0.43	100.0%	100.0%	100.0%	100.0%	36.4%	100.0%	97.8%	100.0%	4.9%	100.0%	94.7%	100.0%
0.44	100.0%	100.0%	100.0%	100.0%	38.8%	100.0%	98.8%	100.0%	5.7%	100.0%	96.7%	100.0%
0.45	100.0%	100.0%	100.0%	100.0%	42.3%	100.0%	99.5%	100.0%	6.6%	100.0%	97.6%	100.0%
0.46	100.0%	100.0%	100.0%	100.0%	45.1%	100.0%	99.7%	100.0%	7.5%	100.0%	98.2%	100.0%
0.47	100.0%	100.0%	100.0%	100.0%	46.8%	100.0%	99.8%	100.0%	8.4%	100.0%	99.1%	100.0%
0.48	100.0%	100.0%	100.0%	100.0%	50.4%	100.0%	99.8%	100.0%	8.2%	100.0%	99.6%	100.0%
0.49	100.0%	100.0%	100.0%	100.0%	52.7%	100.0%	99.8%	100.0%	11.0%	100.0%	99.8%	100.0%

NOTE: This table reports statistics on the significance of the performance rankings from the market timing (T) and global (G) measures obtained from the unconditional CAPM. Daily, occasional and monthly timers with ability levels varying from $\gamma = 0$ to $\gamma = 0.49$ are ranked using daily and monthly data. Panel A shows the average p-values across simulations of tests on the equality between the observed ranking and the ranking expected from the pre-selected ability levels. Shaded statistics indicate significance at the 5% level (lightest shade), 2.5% level (middle shade) and 1% level (darkest shade). Panel B shows the proportions of p-values inferior to the 5% threshold across simulations. Shaded statistics indicate proportions greater than 95%. The ranking tests are based on the index of coincidence or IC statistic proposed by Friedman (1920), with the tabulated γ identifying the highest ability level used in the tests. The simulation procedure and performance evaluation measures are described in section 2. The statistics are described in section 3.1. The estimation data are presented in table 1.

Table 6: Ranking Ability of the Performance Measures

Panel A: Ability levels at which the average *p* -value on the IC statistic becomes significant at the 5% threshold

	Daily Timer				Occasional Timer				Monthly Timer			
	Daily Data		Monthly Data		Daily Data		Monthly Data		Daily Data		Monthly Data	
	T	G	T	G	T	G	T	G	T	G	T	G
CAPM	0.15	0.11	0.16	0.11	#N/A	0.23	0.34	0.23	#N/A	0.28	0.36	0.28
Multi-Index	0.15	0.11	0.16	0.11	#N/A	0.23	0.33	0.23	#N/A	0.28	0.36	0.28
Comer	0.15	0.11	0.16	0.11	#N/A	0.23	0.34	0.23	#N/A	0.28	0.39	0.28
CFG	0.15	0.11	0.19	0.11	#N/A	0.24	0.43	0.23	#N/A	0.30	0.38	0.28
FQ	0.16	0.12	#N/A	0.16	0.46	0.27	#N/A	0.29	#N/A	0.33	#N/A	0.38
BiGARCH	0.14	0.13	---	---	0.44	0.35	---	---	#N/A	#N/A	---	---

Panel B: Ability levels at which the proportion of significant *p* -values becomes greater than 95%

	Daily Timer				Occasional Timer				Monthly Timer			
	Daily Data		Monthly Data		Daily Data		Monthly Data		Daily Data		Monthly Data	
	T	G	T	G	T	G	T	G	T	G	T	G
CAPM	0.18	0.12	0.19	0.13	#N/A	0.27	0.42	0.28	#N/A	0.33	0.44	0.33
Multi-Index	0.18	0.12	0.19	0.13	#N/A	0.27	0.41	0.28	#N/A	0.33	0.44	0.34
Comer	0.18	0.12	0.21	0.13	#N/A	0.27	0.45	0.27	#N/A	0.33	0.47	0.33
CFG	0.18	0.13	0.23	0.12	#N/A	0.29	#N/A	0.28	#N/A	0.34	0.44	0.33
FQ	0.19	0.14	#N/A	0.20	#N/A	0.33	#N/A	0.35	#N/A	0.39	#N/A	0.45
BiGARCH	0.17	0.15	---	---	#N/A	0.42	---	---	#N/A	#N/A	---	---

NOTES: This table reports the ability levels γ associated with significant performance rankings from the market timing (T) and global (G) measures for the daily, occasional and monthly timers, using daily and monthly data. Panel A gives the ability levels at which the average *p*-values across simulations of tests on the equality between the observed ranking and the ranking expected from the pre-selected ability levels first become less than the 5% threshold. Panel B gives the ability levels at which the proportions of *p*-values inferior to the 5% threshold across simulations become greater than 95%. '#N/A' indicates that the ability level is greater than $\gamma = 0.49$, the maximum level of the simulations. The ranking tests are based on the index of coincidence or *IC* statistic proposed by Friedman (1920). The performance models are the unconditional CAPM, the unconditional Multi-Index model, the unconditional multi-index timing model of Comer (2006), the conditional model of Christophersen, Ferson and Glassman (1998) (CFG), the conditional model of Ferson and Qian (2004) (FQ) and the conditional BiGARCH model. The simulation procedure and performance evaluation measures are described in section 2. The statistics used to identify the tabulated ability levels are described in section 3.1. The estimation data are presented in table 1.

Table 7: Evaluation and Ranking Abilities of the Stale Pricing Timing Performance Measures

Panel A: Ability levels at which the *t* -statistic on the mean performance first becomes significant

	Daily Timer		Occasional Timer		Monthly Timer	
	T	G	T	G	T	G
CAPM_2	0.07	0.04	0.33	0.17	#N/A	0.23
Multi-Index_2	0.07	0.04	0.33	0.16	#N/A	0.23
CFG_2	0.08	0.04	0.34	0.19	#N/A	0.24
BiGARCH_2	0.06	0.05	#N/A	0.44	#N/A	#N/A

Panel B: Ability levels at which the average *p* -value on the IC statistic becomes significant

	Daily Timer		Occasional Timer		Monthly Timer	
	T	G	T	G	T	G
CAPM_2	0.14	0.11	0.44	0.23	#N/A	0.28
Multi-Index_2	0.14	0.11	0.45	0.23	#N/A	0.28
CFG_2	0.17	0.13	0.49	0.25	#N/A	0.29
BiGARCH_2	0.15	0.11	#N/A	0.39	#N/A	#N/A

NOTES: This table examines the robustness of the results to the stale pricing timing control proposed by Chen, Ferson and Peters (2010). It reports the ability levels γ associated with significant performance evaluations (panel A) and rankings (panel B) from the market timing (T) and global (G) measures for the daily, occasional and monthly timers, using daily data. Panel A gives the ability levels at which the *t*-statistics on the mean performance values across simulations first become significant at the 5% threshold. Panel B gives the ability levels at which the average *p*-values across simulations of tests on the equality between the observed ranking and the ranking expected from the pre-selected ability levels first become less than the 5% threshold. '#N/A' indicates that the ability level is greater than $\gamma = 0.49$, the maximum level of the simulations. The ranking tests are based on the index of coincidence or *IC* statistic proposed by Friedman (1920). The performance models are the stale pricing timing versions of the unconditional CAPM, the unconditional Multi-Index model, the conditional model of Christophersen, Ferson and Glassman (1998) (CFG) and the conditional BiGARCH model. In these versions, a lag term is included in the market timing specification. The simulation procedure and performance evaluation measures are described in section 2. The statistics used to identify the tabulated ability levels are described in section 3.1. The estimation data are presented in table 1.