1-2008

Estimating the dynamics of mutual fund alphas and betas

Harry MAMAYSKY

Matthew SPIEGEL

Hong ZHANG
Singapore Management University, hongzhang@smu.edu.sg

Estimating the Dynamics of Mutual Fund Alphas and Betas

Harry Mamaysky
Old Lane LP

Matthew Spiegel
Yale School of Management

Hong Zhang
INSEAD, Singapore

This article develops a Kalman filter model to track dynamic mutual fund factor loadings. It then uses the estimates to analyze whether managers with market-timing ability can be identified *ex ante*. The primary findings are as follows: (i) Ordinary least squares (OLS) timing models produce false positives (nonzero alphas) at too high a rate with either daily or monthly data. In contrast, the Kalman filter model produces them at approximately the correct rate with monthly data; (ii) In monthly data, though the OLS models fail to detect any timing among fund managers, the Kalman filter does; (iii) The alpha and beta forecasts from the Kalman model are more accurate than those from the OLS timing models; (iv) The Kalman filter model tracks most fund alphas and betas better than OLS models that employ macroeconomic variables in addition to fund returns. (*JEL* G12, G14, G23, C01, C12, C13, C52, C53)

A great deal of attention has gone into understanding mutual fund returns and trading strategies.[1] What most studies have in common is the maintained hypothesis that their returns can be represented by a static ordinary least squares (OLS) model.[2] Although this may be a reasonable conjecture for some funds, it seems unlikely to be true for many others.

The authors thank Robert Engle, Wayne Ferson, Will Goetzmann, and Geert Rouwenhorst for their comments. We would like to thank participants at the Rutgers Conference honoring David Whitcomb, the 2004 Meetings of the American Finance Association, and seminar participants at Boston College and the University of Wisconsin-Madison. Hong Zhang thanks the INSEAD Alumni Fund (IAF) for financial support. We also thank two anonymous referees and the editor Cam Harvey for their comments that led to the article's investigation of market-timing measures. Address correspondence to Matthew Spiegel, Yale School of Management, P.O. Box 208200, New Haven, CT 06520, or e-mail: matthew.spiegel@yale.edu.

[1] Examples of the former include Lehmann and Modest (1987), Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), Brown and Goetzmann (1995), Carhart (1997), Daniel et al. (1997), Wermers (2000), Pástor and Stambaugh (2002), and Teo and Woo (2001). Examples of the latter include Ferson and Schadt (1996), Brown and Goetzmann (1997), and Ferson and Khang (2002).

[2] One exception is Grinblatt and Titman (1994). The methodology they use avoids a direct comparison against a specific portfolio, and instead uses an "endogenous" benchmark. However, their technique requires knowledge of the fund's actual composition, which may not always be available. Ferson and Khang (2002) extend the technique to condition the portfolio betas on exogenous variables such as macroeconomic data.

doi:10.1093/rfs/hhm049

Investors presumably employ portfolio managers to move assets into and out of various sectors and securities as part of a dynamic strategy. This is especially true if they hope to employ managers with market-timing skills. If so, detecting managerial timing ability with a static statistical model may lead to false positives at an unexpectedly high rate. The analysis presented here confirms that this is indeed a problem and shows how a dynamic Kalman filter model can be employed to ameliorate it successfully and find mangers who appear to have a genuine ability to time the market.

This article extends the mutual fund performance literature along the lines of Ferson and Schadt (1996) (hereafter FS). The FS technique is designed to estimate the manager's implicit strategy with respect to macro variables and then allow for the resulting correlations when judging performance.[3] However, in contrast to FS, our goal is to allow for portfolio shifts due to factors unobservable by the econometrician. This is accomplished by assuming that assets are reallocated on the basis of some unobserved factor, and then estimating the system of equations via a Kalman filter. Of course, one can also include the macroeconomic factors FS use, thereby allowing for both observable and unobservable factors in the specification.

This article demonstrates the Kalman filter model's ability to handle dynamic factor loadings by estimating it on U.S. mutual fund data. The resulting alpha and beta time series show that many funds do indeed follow highly dynamic strategies. Using the monthly data, a subset of them (perhaps as high as 20%) is also shown to possess market-timing ability. This contrasts with Bollen and Busse (2001) (hereafter BB) who, using the OLS models proposed by Treynor and Mazuy (1966) and Henriksson and Merton (1981), do not find such ability with monthly data. The difference lies in the Kalman filter model's ability to adapt to a fund's current loading on market risk in a way that a rolling OLS model cannot.

The results presented here also question the usefulness of using daily data for analyzing mutual fund performance. As BB note, the daily data do potentially offer researchers additional power.[4] But, this potential also comes with a number of microstructure problems like stale pricing and bid–ask bounce. Tests presented here indicate that such problems may indeed impact market-timing estimates.

On average, as a fund's estimated market-timing skill increases with regard to current period returns, the analogous skill estimates

[3] Several recent articles have adopted this technique for performance evaluation. For example, see Christopherson, Ferson, and Glassman (1998), and Blake, Lehmann, and Timmermann (2002).

[4] BB create passive characteristic matched control portfolios for the managed funds in their dataset. They then find that with daily data the managed funds produce larger market-timing parameters than the control funds. In contrast, using monthly data they fail to detect any statistical difference between the control portfolios and the managed funds. From this the paper concludes that with daily data one can find evidence of market-timing ability among fund managers. The monthly data, in their view, do not generate the same result owing to a lack of power generated by the infrequent sampling interval.

using lagged returns decrease. In the absence of microstructure problems, the lagged returns should produce uncorrelated parameter estimates because lagged returns should be uncorrelated with current returns. However, if microstructure issues have generated a false-positive market-timing ability parameter, the lagged return coefficients then act to offset it. The results presented here are consistent with this hypothesis: a nonzero current-period market-timing parameter is associated with lagged timing parameters of about equal size and opposite sign.

As noted by Jagannathan and Korajczyk (1986) an examination of the Treynor and Mazuy (1966) (hereafter TM) and Henriksson and Merton (1981) (hereafter HM) parameter estimates shows that they tend to trade off errors in the market-timing parameter with errors in the estimated fund alphas.[5] In this case, high market-timing estimates are highly correlated with low estimated alphas. This implies that testing a model's factor forecasts one factor at a time may miss problems in other factors. Conversely, it is also possible for a model's overall forecasting power to be good even if the underlying factor estimates appear questionable. To check for these possibilities this article uses a test suggested by Bollen and Busse (2005) and an omnibus test developed here that controls for all of a model's factors at once.

The omnibus test, rather than conducting the usual sorts, creates fund-by-fund zero-alpha zero-beta portfolios by going long the fund and short the factors based on each model's forecasts. An ideal model should produce out-of-sample factor loadings centered on zero for these portfolios. The TM, HM, and four-factor Kalman models fail this test. One might conjecture that the OLS-based HM and TM models would do well if estimated only on funds with low turnover rates. They do not. In contrast, though, the out-of-sample portfolio statistics produced by the one-factor Kalman model cover zero within the 95% bootstrapped confidence interval for the sample as a whole and each turnover tercile. This implies that the one-factor Kalman filter model does not create spurious parameters but instead provides a useful signal regarding a fund's timing ability and its future factor loadings.

The final test in the article looks at the degree to which using conditioning information, as in FS, adds to the model's ability to fit the data within sample. Overall, the conditioning information does not improve the model's fit (as measured by the R^2 statistic). But this is not true of every fund. The number of funds with significant parameter values somewhat exceeds the number that would be produced by chance. From an economic point of view, these findings indicate that even though some

[5] The HM model adds $\gamma r_{mt} I\{r_{mt} > 0\}$ to the standard factor model as a means of detecting market-timing ability. Here, r_{mt} is the market return net of the risk-free rate, and $I\{r_{mt} > 0\}$is an indicator function that equals 1 when $r_{mt} > 0$ and zero otherwise. The TM model replaces $\gamma r_{mt} I\{r_{mt} > 0\}$ with γr_{mt}^2.

funds condition on the type of macro information tested here, many do not. For those that do not, the Kalman filter picks up the time variation in their betas and alphas via estimates of the unobserved factor's value. The tests in this article suggest that up to 20% of all mutual funds (using a 5% critical value) exhibit investment strategies with some dependence on the lagged treasury-bill rate, and on the market dividend yield. Of course, the other funds may be conditioning on macro information not included in this article's tests, a possibility that offers intriguing avenues for future research.

The remainder of the article proceeds as follows. Section 1 describes the data used to estimate the model. Section 2 derives the Kalman filter model that this article proposes and an alternative empirical specification for tracking dynamic alphas and betas. Section 3 conducts a series of out-of-sample tests to check for the reliability of the estimates generated by the models examined here. Section 4 develops and provides an omnibus test of each model's overall ability to provide accurate forecasts. Section 5 examines the reasonableness of the estimated parameter dynamics produced with the Kalman filter model. Section 6 explores the impact of adding macroeconomic factors like those used in FS to the model. Section 7 concludes.

1. Data Description and Model Estimation

Monthly mutual fund data from 1970 to 2002 come from Center for Research in Security Prices (CRSP). A fund is included only if it has more than 48 months of return data. Daily mutual fund data come from a now-defunct firm called Wall Street Web (WSW).[6] The data begin on July 25, 1962 and end on October 5, 2004. Once multiple-share class funds are consolidated, up to 1998, the overlap between the CRSP and WSW data averages just over 90%. After that the overlap declines, but only because the WSW database ceases to include new funds. Examining the total returns for funds that are in both the CRSP daily database and the WSW database shows that they are identical up to some small rounding errors. Finally, a comparison of the equally weighted monthly fund returns from both files yields a correlation coefficient of 0.9977 before 1999.

Other data include the market, Treasury bill, Fama-French factors, momentum (MOM) factor, and CRSP stock decile returns. The empirical section comparing the Kalman filter to the FS conditional model also uses the lagged dividend yield on the market. The monthly dividend yield equals the CRSP value-weighted index total return with dividends, minus the return without dividends. The lagged dividend yield is then calculated as the average of these monthly values in the previous calendar year.

[6] The database is currently housed at the Yale School of Management's International Center for Finance.

Although many studies such as Grinblatt and Titman (1993), Daniel et al. (1997), and Cohen, Coval, and Pastor (2005) have used the mutual fund holdings data with great success, this database is not used here. The reason for this disparity is that papers based on the holdings data seek to detect, prior to fees and other expenses, whether managers have the ability to select stocks that will produce above-market, risk-adjusted returns. Here, however, the focus is on whether managers have market-timing ability and if this ability can be reliably detected. Unfortunately, the relative infrequency with which fund holdings are sampled makes their use in this article's context difficult. Nevertheless, tests were conducted to see if these data could be used to detect market-timing ability among fund mangers. Because all the results were negative, they are not reported here.

2. Derivation of the Kalman Filter Model

Every statistical model generates false positives. In fact, the statement that a coefficient is "significant at the $X\%$ level" recognizes that $X\%$ of the time the statistical model will yield a false-positive result. However, while one must accept that any model will yield false positives, the model should not do so more frequently than one expects, given a particular critical value. Studies including those by Jagannathan and Korajczyk (1986) and Bollen and Busse (2001) show that with real data the HM and TM models generate false positives on the model's timing coefficient (labeled γ in this study) at a rate that is far too high.[7] Thus, the question is whether it is possible to produce an alternative model capable of detecting market timing that does not suffer from this problem.

A generic problem with the OLS model is that the constant factor-loading assumption is likely to be violated for managed portfolios, especially if managers are attempting to time the market. This means that such models are inherently misspecified and are thus likely to produce unexpected results—a tendency to produce false positives, for example. One potential solution is to derive a dynamic model and use it to test for managerial market-timing ability.

If fund managers are to outperform the market on a risk-adjusted basis, they must receive some sort of private signal that forecasts returns. To accommodate this, one needs to start with a general equilibrium model of asset returns with asymmetric information such as Admati (1985). Extending the basic setting to a multiple-period framework, from a particular fund manager's perspective the return on asset i can be described by a linear factor model with constant factor loadings:

$$r_{it} - r_f = \alpha_{it} + \beta_i'(r_{mt} - r_f) + \varepsilon_{it}. \tag{1}$$

[7] In addition to the standard factors, the HM and TM models produce a market-timing coefficient. Those interested in further details should consult the original articles.

The risk-adjusted abnormal return α_{it} depends upon the current value of the manager's signal. Technically, therefore, it should include a parameter indicating the signal upon which it is based. However, for the sake of notational simplicity it is not displayed here. Under the null hypothesis (which will be formally developed later on) the manager's signal does not forecast stock returns, and the α-terms are zero. Throughout this article it is assumed that the errors are normally distributed and independent over time. Note that although returns change over time, their loadings on the economywide risk-factor returns (here, the r_m's) remain constant.[8] If the r_m's are known, estimates of a security's loadings on the economy's risk factors can be obtained by regressing security returns on factor returns.

Even when Equation (1) describes each individual stock's return accurately, it may not extend to a portfolio of such stocks. Consider a fund that holds securities A and B. At any time t the portfolio's return (r_{Pt}) equals

$$r_{Pt} = w_{A,t-1} r_A + w_{B,t-1} r_B \tag{2}$$

where the w terms equal the fraction of the portfolio invested in each asset. Using this and Equation (1), it is straightforward to see that portfolio returns are also linear in the factor returns r_{mt}'s. However, unless the returns on A and B at time t happen to be the same, the portfolio weights for securities A and B will be different at time $t+1$ than they were at time t. Thus, while time $t+1$ portfolio returns remain linear in the $r_{m,t+1}$'s, the weights attached to each factor's return will have changed from the time t weights. Clearly, even in this simple example without any active trading by the fund manager, security returns and a portfolio's returns may not be described well by the same model, especially a linear factor model with constant coefficients.

Now suppose one wishes to estimate the alphas and betas of the above portfolio, rather than the alphas and betas of its constituent securities. In this case, an OLS estimate of the portfolio's loadings on the r_{mt}'s can produce answers that are quite far from the portfolio's true loadings on the factor returns in question. How far off they will be depends upon the covariance of factor loadings with each other and market returns, as well as the degree to which they vary over time. Expressions for the exact magnitude of the error can be found in both Grinblatt and Titman (1989) or FS.

To address the above problem, a statistical model needs to allow explicitly for variation in the fund's portfolio weights over time. A

[8] Many studies like those of Ferson and Harvey (1991, 1993) and Ferson and Korajczyk (1995) question whether individual security loadings are constant. However, this does not qualitatively alter this article's conclusion that fund loadings change over time. If anything, such intertemporal variation in the underlying securities will only add to the importance of allowing for time variation in the mutual funds themselves.

portfolio's time t return equals the weighted average of the returns from the underlying I assets:

$$\begin{aligned} r_{Pt} - r_{ft} &= w'_{t-1}(\alpha_t + \beta'(r_{mt} - r_{ft}) + \varepsilon_t) - k_t \\ &= \alpha_{Pt} + \beta'_{Pt}(r_{mt} - r_{ft}) + \varepsilon_{Pt}, \end{aligned} \tag{3}$$

where the variables α_{Pt}, β_{Pt}, and ε_{Pt} are defined by

$$\begin{aligned} \alpha_{Pt} &\equiv w'_{t-1}\alpha_t - k_t, \\ \beta_{Pt} &\equiv \beta w_{t-1}, \\ \varepsilon_{Pt} &\equiv w'_{t-1}\varepsilon_t, \end{aligned} \tag{4}$$

with w, α, and ε, the I by 1 vectors containing their corresponding firm-specific elements w_i, α_i, and ε_i. The β-term represents a matrix with I columns containing the vectors β_i. Finally, k equals the transactions costs incurred by the portfolio, which for mathematical tractability are assumed to be proportional to the funds under management. In Equation (3), if the Capital Asset Pricing Model (CAPM) or Arbitrage Pricing Theory (APT) holds period by period, then α_{Pt} equals a vector of zeros for all t and all managers. If a model such as Admati (1985) holds, individual managers may use their information to produce nonzero alphas. Again, one should keep in mind that the α-terms are manager and signal dependent.

Note that the α in Equation (3) can derive from a manager's ability to forecast either cross-sectional stock returns or intertemporally time the market. In the former case the portfolio betas may or may not be time dependent and may or may not be correlated with alpha. In the latter case they must be. As will be seen, the Kalman filter model developed below can handle either case, although the market-timing application is the one pursued here.

Equation (3) is the main focus of the econometric analysis in this article, and so merits some discussion. So far two important assumptions have been employed:

1. The evolution of portfolio wealth must satisfy an intertemporal budget constraint;
2. All stocks have constant betas.

These two assumptions together imply that portfolio returns will satisfy a linear factor model, but with time-varying coefficients, and with a heteroskedastic innovation term. This suggests that linear-factor, constant-coefficient models for portfolio returns, a common paradigm for empirical work in asset pricing, are misspecified.[9]

Without information about a fund's holdings, and the alphas and betas of the underlying assets, the empirical system in Equations (3) and (4) cannot be estimated. However, these problems can be overcome by adding some assumptions. With the proper specification of the dynamics governing a fund's portfolio weights, it is not necessary to know the individual weights, alphas, and betas.

Let F_t represent some signal (normalized to have an unconditional mean of zero) that the fund uses to trade. Once again, for notational simplicity, the subscript identifying the signal's recipient is suppressed. Assume that it follows the AR(1) process through time (although more general specifications are possible):

$$F_t = \upsilon F_{t-1} + \eta_t. \tag{5}$$

The $\upsilon \in [0,1)$ coefficient measures the degree to which the signal's value persists over time, and η_t represents an independently and identically distributed (i.i.d.) innovation.

If the signal F has value, then one expects it to influence both the fund's present holdings and future expected stock returns. Statistically, these dual impacts can be represented by assuming that the portfolio weights follow

$$w_{it} = \overline{w}_i + l_i F_t, \tag{6}$$

and that stock alphas equal

$$\alpha_{it} = \overline{\alpha}_i F_{t-1}. \tag{7}$$

Here $\overline{w}_i$ represents the steady-state fraction of the strategy invested in a given security. Alternatively, $\overline{w}_i$ can depend upon any set of observable variables, in which case it may be time dependent. The variable l_i is stock i's loading on a common unobservable factor F_t which shifts the portfolio weights from their steady-state values. This formulation holds exactly under Admati's (1985) model and is generally consistent with Blake, Lehmann, and Timmermann's (1999) empirical finding of mean reversion in fund weightings across securities among U.K. pension funds. Finally, $\overline{\alpha}_i$ represents the degree to which a stock's expected return is predictable by the signal F. If the signal has no value, then all the $\overline{\alpha}_i$ terms equal zero. Also, the present specification ensures that the steady-state alpha values equal zero.

Now use Equations (4) and (7) in the above formulation. Define $\overline{w}$, l, and $\overline{\alpha}$ as the I by 1 vectors with elements $\overline{w}_i$, l_i, and $\overline{\alpha}_i$ respectively, and

[9] It is possible that the stocks have time-varying betas and that fund managers maintain constant portfolio loadings by trading appropriately. Although this is technically possible, the article will provide empirical evidence that high-turnover funds also have greater intertemporal variation in their betas.

one finds that

$$\begin{aligned}\alpha_{Pt} &= \overline{w}'\overline{\alpha}F_{t-1} + l'\overline{\alpha}F_{t-1}^2 - k_t \\ &= \overline{\alpha}_P F_{t-1} + b_P F_{t-1}^2 - k_t,\end{aligned} \tag{8}$$

for the appropriately defined $\overline{\alpha}_P$ and b_P. Similarly, one has

$$\begin{aligned}\beta_{Pt} &= \beta\overline{w} + \beta l F_{t-1} \\ &= \overline{\beta}_P + c_P F_{t-1},\end{aligned} \tag{9}$$

for the appropriately defined $\overline{\beta}_P$ and c_P.

The model's derivation has assumed that managers vary their portfolio's market sensitivity in response to their signals, while the underlying securities have time-invariant betas. Alternatively, one might think that the opposite case holds and that fund managers react to changing factor loadings in their portfolio by rebalancing toward their preferred risk profile. In this case the c_P in (9) will be equal to zero, and if a manager is talented he should generate positive parameter estimates for $\overline{\alpha}_P$ and b_P. Thus, empirically, the model can accommodate funds that select stocks in an attempt to both generate excess returns and maintain a constant factor loading.

The $\overline{\alpha}_i$, $\overline{\alpha}_P$, and b_P each play a unique economic role in the analysis. In Equation (7), $\overline{\alpha}_i \neq 0$ implies that a given fund's signal has a systematic relationship with the instantaneous excess returns of individual stocks in an economy. Therefore, one can add an indicator variable to the $\overline{\alpha}_i$ that indicates that the coefficient is both stock *and* fund dependent. However, the point of having nonzero $\overline{\alpha}_i$'s is to allow the fund's α_P to depend systematically on the fund's trading strategy F. This dependence comes about through a linear term, $\overline{\alpha}_P$, and a quadratic term, b_P. There is no constant alpha term in α_P because in the long run all alphas are assumed to be zero (their unconditional value). The linear term $\overline{\alpha}_P$ simply measures the degree to which a given fund's strategy is actually related to the instantaneous alphas of individual stocks. Because F can be positive or negative, a nonzero $\overline{\alpha}_p$ does not indicate either under- or overperformance. The quadratic term b_P, in contrast, indicates exactly this—it measures the degree to which a fund is able to go systematically long (short) in response to discovering a set of positive (negative) alpha stocks.[10] Note that this is a sufficient, though not necessary, condition for a given fund to exhibit

[10] Intuitively, b_P can be thought of as the covariance between a fund's security weights (w_t) and the underlying security alphas.

occasional (as opposed to systematic) risk-adjusted outperformance. A weaker and necessary condition is that a fund's α_P is persistent and occasionally positive (which is obtained when $\overline{\alpha}_P \neq 0$ and when $\upsilon > 0$).

The empirical model derived above is very flexible. For example, if one assumes that η_t has a variance of zero, or that υ equals zero, the FS specification can be reproduced. What is important, however, is that the model can still be estimated without these assumptions. Also note that nowhere does the econometrician need data on the actual portfolio weights used to produce the observed returns.[11]

Equations (3), (5), (8), and (9) can be estimated via extended Kalman filtering. To obtain the observation equation, use Equations (8) and (9) in (3) to eliminate α_{Pt} and β_{Pt} and produce:

$$\begin{aligned} r_{Pt} - r_{ft} = {} & b_P F_{t-1}^2 - k_t + \overline{\beta}_P (r_{mt} - r_{ft}) \\ & + \left[\overline{\alpha}_P + c_P (r_{mt} - r_{ft})\right] F_{t-1} + \varepsilon_{Pt} \end{aligned} \tag{10}$$

after some minor algebra. Owing to the F_{t-1}^2 term, standard Kalman filtering techniques will fail, as the conditional variance of $r_{Pt} - r_{ft}$ will no longer be independent of the estimated values of F_{t-1}. The standard solution is to use a first-order Taylor expansion around the conditional expectation of F_{t-1}, or

$$\begin{aligned} F_{t-1}^2 \approx {} & 2\mathbb{E}[F_{t-1} | r_{P,t-1} - r_{f,t-1}, F_{t-2}] F_{t-1} \\ & - \mathbb{E}[F_{t-1} | r_{P,t-1} - r_{f,t-1}, F_{t-2}]^2 \end{aligned} \tag{11}$$

to replace the F_{t-1}^2 term in Equation (10), where $\mathbb{E}$ is the expectations operator in (11).[12] Equation (5) then forms the state equation. Note that the vector c_P has n elements (one for each risk factor) but only $n - 1$ degrees of freedom. Thus, in the scalar case (as in the CAPM) it can be normalized to one when estimating the model. In the case where n is greater than 1, at least one element's value must be fixed or some other normalization must be applied. The other fact needed for estimation is that the variance of ε_{Pt}, conditional on time $t - 1$ information, is given by

$$\mathrm{var}_{t-1}(\varepsilon_{Pt}) = \sum_{i=1}^{I} w_{i,t-1}^2 \mathrm{var}_{t-1}(\varepsilon_{it}). \tag{12}$$

[11] Of course, other modeling choices are possible, and this is an interesting area for future research. For example, some portfolio strategies lead to known security weightings. Examples include the Fama-French-Carhart momentum-, growth-, and size-based portfolios. In such cases the portfolio alpha and beta in Equation (4) may be calculated directly, as long as alphas and betas of individual stocks are known.

[12] For details about extended Kalman filtering, see Harvey (1989).

This follows from the last equation in (4), and from the fact that all ε_{it}'s are independent. Estimation is conducted by maximizing the log likelihood function and so one needs some way to define when the algorithm has or has not converged. This article defines convergence as having occurred if the R^2 measure increases by at least 0.01, and if the parameters $\overline{\alpha}_p$ and b_P do not hit a boundary of 10.

The system specified in Equations (5) and (11) embeds an important timing convention. The alphas and betas that determine time t returns are known at time $t-1$ (assuming that k_t is deterministic). Therefore, any covariance between a portfolio's time t alphas and time t market returns indicates that at time $t-1$ the portfolio manager makes investment decisions that successfully anticipate market returns at time t. The same is true for time t betas and time t market returns. Whether a portfolio manager has this ability or not will affect the interpretation of our results later on.

2.1 False rejections and the Kalman filter model

Unlike an OLS model, the Kalman filter model generates a dynamic estimate of a fund's beta. By correlating the fund's time-varying market beta with the market return, one has a natural measure of a fund's market-timing ability. Using this measure, Table 1 presents the results from estimating a one- and four-factor Kalman filter model on the daily CRSP size deciles. Table 2 repeats the exercise with monthly data.

The CRSP size deciles clearly have no market-timing ability, and thus a properly specified model should yield false positives at the corresponding critical rate. Using daily data and a 5% critical value the Kalman filter model yields false positives at a rate of 58% for the one-factor model and 30% for the four-factor model. These numbers, while bad, are roughly comparable to the rates produced by the TM and HM models (tables available from the authors). However, with monthly data the Kalman filter's performance improves dramatically. In Table 2 the one-factor model produces false positives at a rate of 3%, and the four-factor model at a rate of 6%. Both of these figures are in line with the rate one expects from a properly specified statistical model. In comparison, the rates for the TM and HM models are 17% and 14%, respectively.

The difference in the rate of false positives generated by the Kalman filter model in the monthly and daily data may come from microstructure issues. When using daily data, the Kalman filter may be detecting pseudo market timing generated by things such as bid–ask bounce, stale pricing, and other factors known to generate serial correlation in the data.[13] In contrast, these

[13] For example, suppose a price drop in a stock reduces its subsequent trading volume and an increase raises its subsequent trading volume. In this case, stale pricing will be a bigger problem after a price drop and a smaller problem after an increase. As a result, the stock will (on average) appear to be less correlated with the overall market when overall returns are down and more correlated when returns are up. The HM, TM, and Kalman models will then interpret this pattern as evidence of market-timing ability.

Table 1
Daily Market-Timing Ability of CRSP Size Deciles: Kalman Filter Estimates

Periods	1-Dec	2-Dec	3-Dec	4-Dec	5-Dec	6-Dec	7-Dec	8-Dec	9-Dec	10-Dec
A1. Kalman 1-factor market-timing parameters of the 10 VW size portfolios.										
1970	0.0387	0.0464	0.0415	0.0288	0.0267	−0.0054	−0.0153	0.0136	−0.0440	0.0609
1975	−0.1588	−0.1668	−0.2056	−0.2226	−0.2477	−0.2586	−0.2632	−0.1987	−0.1317	0.2628
1980	−0.1377	−0.1717	−0.1568	−0.1645	−0.1676	−0.1593	−0.1613	−0.1225	−0.0637	0.1919
1985	−0.0840	−0.0555	−0.0369	−0.0540	−0.0542	−0.0384	−0.0557	−0.0645	−0.0419	0.0716
1990	0.0383	−0.1501	−0.1335	−0.1559	−0.1032	−0.1143	−0.1398	−0.0136	−0.0337	0.0417
1995	−0.1468	−0.1753	−0.1362	−0.1515	−0.1405	−0.1053	−0.0801	−0.0619	−0.0347	0.0492
2000	−0.0712	−0.1296	−0.1274	0.0253	0.0309	0.0324	0.0352	0.0363	0.0328	0.0182
1970	(1.36)	(1.64)	(1.47)	(1.02)	(0.95)	(0.19)	(0.53)	(0.48)	(1.54)	(2.11)
1975	(3.17)	(2.93)	(2.82)	(2.85)	(2.82)	(2.98)	(2.93)	(3.09)	(2.89)	(2.91)
1980	(2.87)	(2.96)	(3.21)	(3.30)	(2.98)	(2.90)	(3.05)	(3.18)	(2.29)	(3.09)
1985	(2.99)	(1.93)	(1.31)	(1.87)	(1.94)	(1.36)	(1.94)	(2.25)	(1.47)	(2.47)
1990	(1.37)	(2.83)	(2.85)	(2.94)	(3.02)	(2.95)	(2.89)	(0.48)	(1.20)	(1.47)
1995	(2.92)	(3.02)	(2.91)	(3.20)	(3.15)	(3.29)	(3.63)	(2.12)	(1.22)	(1.74)
2000	(2.49)	(2.85)	(3.41)	(0.89)	(1.10)	(1.15)	(1.25)	(1.27)	(1.15)	(0.64)
B1. Kalman 4-factor timing parameters of the 10 VW size portfolios.										
1970	0.0168	0.0625	0.0544	0.0093	−0.0491	0.0543	−0.1155	−0.0057	−0.0943	0.0778
1975	0.0364	0.0093	−0.0091	0.0236	−0.0546	0.0949	−0.0123	0.0352	0.0664	0.0833
1980	0.0172	−0.0275	−0.0563	−0.0564	−0.0482	0.0504	−0.0042	0.1063	0.0861	0.0088
1985	−0.0634	−0.0859	−0.0414	−0.0215	−0.0095	0.0624	−0.0116	0.0073	0.0327	−0.0318
1990	−0.0436	0.0259	−0.1374	−0.0433	−0.0248	−0.0239	0.0107	0.0627	−0.0070	−0.0119
1995	−0.0648	−0.0943	−0.0811	−0.0341	0.0103	−0.0009	−0.0117	−0.0023	−0.0061	−0.0057
2000	−0.0545	−0.1574	−0.1329	−0.0849	−0.0168	−0.0118	−0.0328	−0.0490	−0.0207	−0.0292
1970	(0.59)	(2.14)	(1.95)	(0.33)	(1.75)	(1.95)	(2.89)	(0.20)	(2.82)	(2.44)
1975	(1.28)	(0.33)	(0.31)	(0.83)	(1.92)	(2.98)	(0.44)	(1.25)	(2.26)	(2.91)
1980	(0.61)	(0.97)	(1.95)	(2.05)	(1.68)	(1.79)	(0.15)	(3.18)	(3.07)	(0.30)
1985	(2.28)	(2.89)	(1.47)	(0.77)	(0.33)	(2.14)	(0.41)	(0.25)	(1.17)	(1.12)
1990	(1.55)	(0.91)	(2.85)	(1.54)	(0.87)	(0.84)	(0.38)	(2.17)	(0.25)	(0.42)
1995	(2.27)	(3.02)	(2.91)	(1.20)	(0.36)	(0.04)	(0.41)	(0.07)	(0.21)	(0.20)
2000	(1.94)	(2.85)	(3.41)	(3.05)	(0.59)	(0.41)	(1.14)	(1.72)	(0.74)	(1.03)

Within each 5-year window, Panel A applies the one-factor and four-factor Kalman models (the four-factors are MKT, SMB, HML, and MOM) to estimate a beta time series based on the excess daily returns from the 10 CRSP size-sorted deciles using NYSE-AMEX-NASDAQ stocks. A market-timing parameter is estimated as the correlation between the estimated beta time series and the market return. T-Statistics are in parentheses. About 58% of all portfolios under the one-factor Kalman model and 30% under the four-factor Kalman model have significant timing coefficients at the 5% level.

data problems are likely to have little impact on the monthly data. This in turn allows the Kalman filter model to avoid detecting market-timing ability when none obviously exists. To examine this possibility, regressions of the form

$$r_{it} - r_{ft} = \alpha_0 + \alpha_1 SMB_t + \alpha_2 HML_t + \alpha_3 MOM_t + \sum_{j=0}^{19} (\beta up_j MKT_{t-j} DUP_{t-j} + \beta dn_j MKT_{t-j} DDN_{t-j}) \quad (13)$$

were run. In this equation r_{it} equals the return on asset i in period t, and r_{ft} is the corresponding risk-free rate. The α-terms are estimated parameters for the Fama-French-Carhart factors small-minus-big (SMB), high-minus-low (HML), and momentum (MOM). The βup and βdn terms

Table 2
Monthly Market-timing Ability of the CRSP Size Deciles: Kalman Filter Estimates

	Decile									
Periods	1	2	3	4	5	6	7	8	9	10
A1. Kalman 1-factor timing parameters of the 10 VW size portfolios (from CRSP).										
1970	0.1827	0.1466	0.1640	0.1488	0.1969	0.1980	0.0860	0.2082	0.1747	−0.2013
1975	0.0638	0.0296	−0.0293	−0.0179	−0.0918	−0.1854	−0.1460	−0.0008	0.1382	0.2151
1980	−0.1707	−0.1715	−0.2215	−0.0513	−0.2204	−0.1349	−0.0202	0.0384	0.0413	0.0009
1985	0.0351	−0.1730	−0.1365	−0.0866	−0.3424	0.0288	−0.0922	0.1304	0.0726	0.1751
1990	−0.0195	−0.2426	−0.0441	−0.1475	−0.2981	−0.1583	−0.1080	−0.0367	0.0340	0.0255
1995	0.0326	−0.1200	−0.2312	−0.1579	−0.1677	0.0211	−0.1384	−0.2444	−0.1892	−0.0838
2000	−0.0103	−0.0862	−0.0606	0.0126	−0.0804	0.0777	0.0313	0.0535	0.0926	0.1701
1970	(1.42)	(1.14)	(1.27)	(1.14)	(1.54)	(1.54)	(0.66)	(1.62)	(1.36)	(1.58)
1975	(0.48)	(0.23)	(0.23)	(0.14)	(0.69)	(1.42)	(1.11)	(0.00)	(1.07)	(1.67)
1980	(1.33)	(1.33)	(1.72)	(0.39)	(1.72)	(1.05)	(0.15)	(0.29)	(0.32)	(0.01)
1985	(0.27)	(1.33)	(1.05)	(0.66)	(2.66)	(0.22)	(0.71)	(1.00)	(0.56)	(1.36)
1990	(0.15)	(1.92)	(0.33)	(1.14)	(2.39)	(1.21)	(0.83)	(0.28)	(0.25)	(0.19)
1995	(0.25)	(0.92)	(1.78)	(1.21)	(1.30)	(0.16)	(1.07)	(1.92)	(1.46)	(0.65)
2000	(0.08)	(0.66)	(0.46)	(0.10)	(0.62)	(0.59)	(0.24)	(0.41)	(0.71)	(1.33)
B1. Kalman 4-factor timing parameters of the 10 VW size portfolios (from CRSP).										
1970	0.1177	0.2409	0.0617	−0.0994	0.1085	0.1838	−0.0554	0.0900	0.0882	−0.1171
1975	0.1900	0.0615	0.0741	0.1161	0.1047	0.1689	0.1659	0.1130	−0.0759	0.0439
1980	0.1185	−0.1256	0.1729	0.0406	0.3125	0.1488	−0.0231	0.0652	−0.1788	0.0957
1985	−0.0214	0.1116	−0.0796	−0.0910	−0.1342	0.0789	0.0766	−0.1399	0.1204	−0.2970
1990	0.2092	−0.0358	0.0766	0.1049	0.1076	0.0941	0.0422	−0.0510	0.2160	−0.1427
1995	−0.2253	−0.1917	−0.1361	0.0214	−0.2345	0.0026	−0.1406	0.0147	−0.0016	−0.0232
2000	−0.0822	−0.2651	−0.1996	−0.2069	−0.2618	0.1519	−0.0319	−0.1547	−0.0657	0.2350
1970	(0.90)	(1.92)	(0.47)	(0.76)	(0.83)	(1.42)	(0.43)	(0.69)	(0.68)	(0.90)
1975	(1.46)	(0.47)	(0.57)	(0.88)	(0.79)	(1.30)	(1.27)	(0.87)	(0.59)	(0.33)
1980	(0.90)	(0.96)	(1.33)	(0.31)	(2.39)	(1.14)	(0.18)	(0.50)	(1.39)	(0.73)
1985	(0.16)	(0.85)	(0.60)	(0.69)	(1.02)	(0.60)	(0.59)	(1.07)	(0.92)	(2.39)
1990	(1.62)	(0.27)	(0.59)	(0.81)	(0.83)	(0.73)	(0.32)	(0.39)	(1.67)	(1.09)
1995	(1.78)	(1.50)	(1.05)	(0.16)	(1.85)	(0.03)	(1.09)	(0.11)	(0.01)	(0.18)
2000	(0.63)	(2.10)	(1.54)	(1.62)	(2.10)	(1.16)	(0.24)	(1.19)	(0.50)	(1.85)

Within each 5-year window, Panel A applies the one-factor and four-factor Kalman models (the four-factors are MKT, SMB, HML, and MOM) to estimate a beta time series based on the excess monthly returns from the 10 CRSP size-sorted deciles using NYSE-AMEX-NASDAQ stocks. The market-timing parameter is estimated as the correlation between estimated beta time series and the market return. The t -statistics are in parentheses. About 3% of all portfolios under the one-factor Kalman model and 6% under the four-factor Kalman model have significant timing coefficients at the 5% level.

represent estimated up- and down-market betas. They multiply the market excess return (MKT) and dummies. The *DUP* dummy equals 1 if the MKT is positive and zero otherwise, and the *DDN* dummy equals 1 if the MKT is negative and zero otherwise. If betas are really the same in up and down markets, then tests of the hypothesis that $\sum_{i=0}^{k} \beta up_i = \sum_{i=0}^{k} \beta dn_i$ should produce insignificant results. Note that the HM market-timing parameter can be recovered from Equation (13) by subtracting βdn_0 from βup_0.

One can use Equation (13) to test for microstructure issues in the data by looking at the sign βup_0 minus βdn_0 ($\Delta\beta_0$) and then comparing this value to the sum of the βup_j minus βdn_j for the subsequent j's ($\Delta\beta_{1-19} = \sum_{j=1}^{19} \beta up_j - \beta dn_j$). If there are no microstructure problems, then the $\Delta\beta_{1-19}$ should be independent of $\Delta\beta_0$. Conversely, if there are

Table 3
Impact of Microstructure Issues on Mutual Fund Daily Timing

	One-factor model				Four-factor model			
A. The relationship between current and lagged beta differences (5-year non-overlapping windows)								
Decile	$\Delta\beta_0$	$\Delta\beta_{1-19}$	T_{NW}	T_{OLS}	$\Delta\beta_0$	$\Delta\beta_{1-19}$	T_{NW}	T_{OLS}
1	−0.31	0.22	(9.34)	(9.95)	−0.16	0.10	(4.37)	(4.67)
2	−0.16	0.15	(9.95)	(10.86)	−0.05	0.05	(5.89)	(6.19)
3	−0.11	0.06	(6.56)	(6.59)	−0.02	0.02	(4.62)	(3.45)
4	−0.06	0.06	(7.50)	(7.13)	0.00	0.01	(0.87)	(0.84)
5	−0.03	0.02	(2.36)	(2.80)	0.01	0.00	(0.55)	(0.50)
6	0.00	−0.02	(2.41)	(2.44)	0.03	−0.01	(1.93)	(2.05)
7	0.02	−0.03	(6.44)	(6.08)	0.04	−0.03	(6.40)	(4.87)
8	0.04	−0.05	(8.03)	(7.81)	0.06	−0.05	(6.47)	(6.45)
9	0.06	−0.05	(7.89)	(8.67)	0.09	−0.05	(5.44)	(5.97)
10	0.11	−0.11	(8.58)	(10.97)	0.14	−0.12	(9.63)	(11.91)
Corr		−0.99	[0.00]			−0.98	[0.00]	
B. The relationship between current and lagged beta differences (1-year non-overlapping windows)								
Decile	$\Delta\beta_0$	$\Delta\beta_{1-19}$	T_{NW}	T_{OLS}	$\Delta\beta_0$	$\Delta\beta_{1-19}$	T_{NW}	T_{OLS}
1	−0.42	0.18	(8.10)	(8.29)	−0.26	0.15	(9.06)	(9.00)
2	−0.20	0.04	(3.59)	(3.61)	−0.10	0.06	(8.07)	(8.29)
3	−0.12	0.03	(3.41)	(3.34)	−0.05	0.03	(3.89)	(3.75)
4	−0.07	0.02	(2.56)	(2.69)	−0.02	0.00	(0.36)	(0.37)
5	−0.03	0.01	(1.61)	(1.58)	0.00	0.02	(2.46)	(2.54)
6	0.00	0.00	(0.59)	(0.57)	0.03	0.00	(0.17)	(0.16)
7	0.03	−0.02	(2.78)	(2.67)	0.05	−0.02	(2.49)	(2.49)
8	0.07	−0.02	(2.43)	(2.43)	0.08	−0.03	(3.40)	(3.60)
9	0.12	−0.02	(2.23)	(2.36)	0.12	−0.04	(4.82)	(4.64)
10	0.28	−0.13	(6.80)	(9.62)	0.27	−0.16	(8.72)	(10.83)
Corr		−0.97	[0.00]			−0.99	[0.00]	

Within each 5-year nonoverlapping window the following one- and four-factor regressions are applied to the daily data: $r_{it} - r_{ft} = \alpha_0 + \alpha_1 SMB_t + \alpha_2 HML_t + \alpha_3 MOM_t + \sum_{j=0}^{19} (\beta up_j MKT_{t-j} DUP_{t-j} + \beta dn_j MKT_{t-j} DDN_{t-j})$. For each fund the summation of lagged beta difference $\Delta\beta_{1-19}$ is defined as $\sum_{j=1}^{19} \beta up_j - \beta dn_j$. Similarly, $\Delta\beta_0$ is defined as $\beta up_0 - \beta dn_0$ when regressions without lagged market returns are applied to each fund in the 5-year in-sample period (the results are the same if $\beta up_0 - \beta dn_0$ is calculated from the regression with all 19 lags). The table sorts funds into deciles by $\Delta\beta_0$, and reports the average value of $\Delta\beta_{1-19}$ in Panel A. The T_{NW} are the Newey-West t-statistics for each $\Delta\beta_{1-19}$ with 11 lags. The T_{OLS} column reports t-statistics as the OLS t-statistic. The Corr row reports the Spearman rank correlation coefficient across deciles. The p-value for the rank correlation is in square brackets. Panel B repeats the same process with sample periods of one year. All domestic equity mutual funds (ICDI-OBJ: AG, BL, GI, IN, LG, PM, SF, and UT) are included in the dataset.

microstructure problems, then $\Delta\beta_{1-19}$ should substantially offset $\Delta\beta_0$ as the problems work their way through the data.

Table 3 estimates Equation (13) using daily mutual fund returns. As one can see, the estimated $\Delta\beta_0$ values are strongly negatively correlated with the estimated $\Delta\beta_{1-19}$ values whether one uses a one- or five-year rolling window or a one- or four-factor model to create the timing estimates.[14] It is particularly worth noting how close the magnitudes are for the estimated $\Delta\beta_0$ and $\Delta\beta_{1-19}$ parameters for the top decile HM gamma funds. In Panel

[14] Shorter windows were also used and produced similar results. The data using windows of one year are reported here to avoid raising questions about the reliability of the results, given the large number of parameters being estimated.

A, using five-year windows and the one-factor model, other than the sign the figures are identical (0.11). Using the four-factor model they are close: 0.14 for $\Delta\beta_0$ and -0.12 for $\Delta\beta_{1-19}$. Using one-year windows the problem is somewhat ameliorated but not entirely eliminated. For the highest decile groups $\Delta\beta_{1-19}$ offsets about half of $\Delta\beta_0$'s value. Thus, the data seem to indicate that in the absence of microstructure issues one would need to believe that a fund's current market-timing ability is generally offset largely by its past systematic mistiming.[15] Because this seems unlikely, to avoid microstructure issues the remainder of the article works primarily with the monthly data.

3. Out-of-Sample Tests

Although both Jagannathan and Korajczyk (1986) and Bollen and Busse (2001) show that neither the TM nor the HM models can apparently be used to determine reliably if funds exhibit in-sample timing ability, the same may not be true out-of-sample.[16] Similarly, though the Kalman model does not yield in-sample false positives at an unexpectedly high rate (with monthly data), it may still yield unreliable out-of-sample forecasts. Essentially the argument is that if a model produces false rejections at too high a rate, its parameter estimates may still contain some informational value. If so, this should show up in out-of-sample tests. Managers with larger parameter values should be more likely to exhibit out-of-sample timing ability if, in fact, the parameter values themselves contain some information.

Table 4 examines the degree to which the HM model's in-sample timing parameter corresponds to a set of out-of-sample statistics. Panel A ranks funds by their HM in-sample market-timing parameter using a one-factor model, and Panel B does the same using a four-factor model. On the basis of these ranks, decile portfolios are constructed and the four-factor HM and TM models, along with the lagged factors (as in (13)), are then run on the out-of-sample data. The results indicate that there is essentially no out-of-sample predictability. Consider the most encouraging statistic in the table, the Spearman rank correlation coefficient across deciles for the one-factor model with the out-of-sample HM timing parameter. It

[15] Ang, Chen, and Xing (2002) have shown that many individual stocks have different up- and down-market betas. However, if funds are generating significant "market-timing" parameters with these stocks, then the model is still producing false positives. The γ-parameter does not come from the manager's ability to rebalance the portfolio in anticipation of the market's next move but rather from the stock's inherent risk characteristics. More important, though, for this article, if the Ang et al. result is primarily responsible for the HM and TM false positives, then the lagged returns in Equation (13) should not yield significant market-timing parameters, the sum of which is of opposite sign and magnitude near that of the current period's parameter.

[16] Again, with the caveat, Bollen and Busse (2001) find that some managers produce higher timing estimates than one might expect relative to a bootstrapped control portfolio.

Table 4
Four-Factor Adjusted Return and Market-timing Coefficients for Mutual Fund Deciles Across Estimated HM Market-Timing Skill (1970–2002) with Monthly Data

γ sorted deciles	α (bp)	β	γ_{HM}	γ_{TM}	T_α	T_β	$T_{\gamma HM}$	$T_{\gamma TM}$
			A. Deciles sorted by 1-factor HM market-timing ability.					
Decile 1	−12.37	0.97	−0.0370	−0.0628	(1.34)	(48.01)	(0.64)	(0.31)
Decile 2	−2.10	0.94	0.0205	0.1014	(0.30)	(60.11)	(0.46)	(0.64)
Decile 3	−0.16	0.93	−0.0212	−0.1005	(0.03)	(75.28)	(0.60)	(0.80)
Decile 4	−3.35	0.92	0.0058	0.0111	(0.67)	(84.40)	(0.19)	(0.10)
Decile 5	−2.41	0.92	0.0271	0.0733	(0.58)	(100.90)	(1.04)	(0.80)
Decile 6	−1.74	0.90	0.0196	0.0289	(0.44)	(103.71)	(0.80)	(0.33)
Decile 7	−7.25	0.90	0.0209	−0.0979	(1.93)	(109.06)	(0.89)	(1.17)
Decile 8	−4.92	0.89	0.0268	0.0138	(1.17)	(96.51)	(1.02)	(0.15)
Decile 9	−2.05	0.88	0.0822	0.3018	(0.44)	(86.92)	(2.86)	(2.96)
Decile 10	11.24	0.74	−0.0367	−0.4710	(1.20)	(35.16)	(0.60)	(2.11)
D10−D1	17.36	−0.10	0.0110	−0.1603	(1.29)	(3.47)	(0.13)	(0.54)
Rank correlation	0.55	−0.82	0.37	−0.19	[0.10]	[0.00]	[0.30]	[0.59]
			B. Deciles sorted by 4-factor HM market-timing ability.					
Decile 1	−9.44	0.92	0.0040	0.0386	(1.49)	(66.23)	(0.10)	(0.27)
Decile 2	−7.63	0.93	−0.0192	−0.0530	(1.63)	(90.42)	(0.66)	(0.51)
Decile 3	−5.46	0.91	−0.0120	−0.0939	(1.34)	(101.99)	(0.47)	(1.04)
Decile 4	−4.39	0.92	0.0545	0.1495	(1.29)	(122.81)	(2.57)	(1.99)
Decile 5	1.81	0.90	0.0110	−0.0477	(0.49)	(111.60)	(0.48)	(0.59)
Decile 6	−4.02	0.91	0.0279	0.0839	(0.95)	(97.61)	(1.05)	(0.89)
Decile 7	−3.71	0.91	0.0434	0.1108	(1.00)	(111.62)	(1.88)	(1.35)
Decile 8	−1.39	0.90	0.0497	0.2127	(0.32)	(93.41)	(1.81)	(2.20)
Decile 9	−2.24	0.91	0.0063	−0.1144	(0.45)	(83.32)	(0.20)	(1.03)
Decile 10	8.66	0.78	−0.0524	−0.5142	(0.90)	(36.01)	(0.84)	(2.25)
D10−D1	10.60	0.00	−0.0450	−0.2820	(1.28)	(0.27)	(0.87)	(1.54)
Rank correlation	0.78	−0.63	−0.02	−0.35	[0.01]	[0.05]	[0.96]	[0.33]

Each January a market-timing ability measure (γ) is estimated using the previous 60 months of return data. Gammas are estimated with the HM model using the market return and its lagged value or the four Carhart (1997) factors and their lagged values plus the appropriate timing variable. On the basis of these gammas, funds are sorted into decile portfolios. The out-of-sample returns are then regressed on the four-factors and their lagged values using the HM or TM model to generate out-of-sample gamma values. These are reported in the γ_{HM} and γ_{TM} columns, respectively. Next the same out-of-sample return series is regressed on the four Carhart factors and their lagged values, but not the market-timing variable, to generate the out-of-sample α and market β values reported in their respective columns. Corresponding t -statistics are reported in the last 4 columns. Line "D10–D1" reports the same parameters for the difference between decile 10 and decile 1 funds. The last line reports the Spearman rank correlation for parameters across the 10 deciles, as well as the corresponding p -values. The data uses all domestic equity mutual funds (ICDI-OBJ: AG, BL, GI, IN, LG, PM, SF, and UT).

equals 0.37 but is not significant at the 10% level. The other market-timing rank correlation statistics are not even that strong. In fact, the other three are actually negative. This implies, first of all, that the in-sample HM timing statistic is negatively correlated with a fund's out-of-sample TM timing statistic. Second, it also implies that the in-sample, four-factor HM timing statistic is negatively correlated with *its own* out-of-sample counterpart.[17]

[17] Similar results were found when portfolios were created with the TM market-timing parameter and are not reported here for the sake of brevity.

Table 5
Four-Factor Adjusted Return and Market-timing Coefficients for Mutual Fund Deciles Across Estimated Kalman Market-timing Skill (1970–2002) with Monthly Data

Mkt.-timing sorted deciles	α (bp)	β	γ_{HM}	γ_{TM}	T_α	T_β	$T_{\gamma HM}$	$T_{\gamma TM}$
A. Deciles sorted by 1-factor Kalman market-timing ability.								
Decile 1	−8.13	0.93	−0.0543	−0.3153	(1.19)	(60.20)	(1.23)	(2.01)
Decile 2	−3.57	0.94	−0.0070	−0.0495	(0.73)	(84.28)	(0.22)	(0.44)
Decile 3	−5.22	0.96	−0.0239	−0.1190	(1.12)	(91.28)	(0.80)	(1.11)
Decile 4	−4.26	0.93	0.0354	0.0167	(1.05)	(102.41)	(1.36)	(0.18)
Decile 5	−3.18	0.90	0.0739	0.2409	(0.80)	(100.39)	(2.92)	(2.66)
Decile 6	1.09	0.91	0.0570	0.1516	(0.30)	(110.39)	(2.44)	(1.81)
Decile 7	−7.68	0.92	0.0373	0.0505	(1.85)	(97.88)	(1.39)	(0.53)
Decile 8	−0.56	0.88	0.0586	0.2146	(0.11)	(77.82)	(1.81)	(1.86)
Decile 9	−0.36	0.89	0.0892	0.2682	(0.08)	(82.23)	(2.91)	(2.44)
Decile 10	−3.44	0.93	0.1019	0.3302	(0.50)	(59.85)	(2.30)	(2.09)
D10–D1	4.69	0.01	0.1561	0.6455	(0.56)	(0.26)	(2.91)	(3.39)
Rank correlation	0.50	−0.53	0.90	0.89	[0.14]	[0.11]	[0.00]	[0.00]
B. Deciles sorted by 4-factor Kalman market-timing ability.								
Decile 1	−5.13	0.94	−0.0424	−0.2106	(0.75)	(60.53)	(0.96)	(1.33)
Decile 2	−3.24	0.90	0.0198	−0.0340	(0.63)	(78.26)	(0.60)	(0.29)
Decile 3	−2.50	0.91	0.0949	0.3171	(0.58)	(92.41)	(3.45)	(3.22)
Decile 4	−6.35	0.93	0.0186	−0.0515	(1.51)	(97.34)	(0.69)	(0.53)
Decile 5	−3.06	0.91	0.0312	0.0780	(0.76)	(100.06)	(1.20)	(0.84)
Decile 6	−2.39	0.94	0.0507	0.0581	(0.57)	(99.47)	(1.89)	(0.61)
Decile 7	−5.96	0.92	0.0538	0.2356	(1.40)	(95.48)	(1.96)	(2.42)
Decile 8	−0.13	0.94	0.0086	−0.0993	(0.03)	(89.74)	(0.29)	(0.93)
Decile 9	−5.48	0.93	0.1159	0.4311	(1.24)	(93.17)	(4.15)	(4.34)
Decile 10	−3.06	0.90	−0.0217	−0.0662	(0.53)	(69.29)	(0.59)	(0.50)
D10–D1	2.07	−0.04	0.0207	0.1443	(0.26)	(2.40)	(0.40)	(0.78)
Rank correlation	0.16	−0.02	0.21	0.30	[0.67]	[0.95]	[0.56]	[0.40]

Each January the one- or four-factor Kalman filter model is estimated using the previous 60 months of return data. On the basis of the estimated parameter values the correlation between the fund's beta time series and the market return is calculated. Call this correlation the fund's forecasted market-timing ability. Funds are then sorted into decile portfolios based on their forecasted market-timing ability. The out-of-sample returns are then regressed on the four-factors and their lagged values in either the HM or TM model to generate out-of-sample gamma values. These are reported in the γ_{HM} and γ_{TM} columns, respectively. Next, the same out-of-sample return series is regressed on the four Carhart factors and their lagged values, but not the market-timing variable, to generate the out-of-sample α and market β values reported in their respective columns. Corresponding t-statistics are reported in the last 4 columns. Line "D10–D1" reports the same parameters for the difference between decile 10 and decile 1 funds. The last line reports the Spearman rank correlation for parameters across the 10 deciles, as well as the corresponding p -values. The data uses all domestic equity mutual funds (ICDI-OBJ: AG, BL, GI, IN, LG, PM, SF, and UT).

Table 5 repeats the exercise in Table 4 but this time sorts portfolios on the basis of their Kalman filter market-timing estimates. The out-of-sample returns from these portfolios are then regressed on the TM and HM models. Now contrast the results in the two tables. Table 5 shows that overall both the one- and four-factor Kalman filter models do a superior job of forecasting the TM and HM timing parameters than the TM and HM models do themselves. Furthermore, the one-factor Kalman filter model is so accurate that it produces both high minus low gamma values and Spearman rank correlations that are also significant at any reasonable level.

Another point of comparison between Tables 4 and 5 is the rank correlation of the alpha columns. From Table 4 it appears that the in-sample market-timing estimates from the HM model predict, if anything, a fund's out-of-sample alpha. Sorting by the in-sample, four-factor HM model's gamma coefficient produces a set of estimated alphas that have, collectively, an out-of-sample rank correlation of 0.78 (p-value of just over 1%). Nevertheless, none of the individual portfolios produces an alpha that is statistically different from zero at the standard critical values. This indicates that funds with positive alphas may be misclassified by the HM model as having market-timing ability when in fact they have (if anything) selection ability. Now consider the results in Table 5. Sorting by the market-timing estimate produced with either Kalman filter model leaves the HM model's estimated alphas unsorted. In this case the rank correlations are 0.50 and 0.16, which are not significant at any reasonable level. This provides evidence that unlike the HM model, the Kalman filter models do not similarly misclassify selection ability as timing ability.

4. Out-of-Sample Multiple Parameter Tests

4.1 Bollen and Busse (2005)

It is possible that the poor out-of-sample performance of the OLS parameter estimates arises here because the tests are using the wrong forecast from the model. An alternative is the measure used in Bollen and Busse (2005),

$$r_{p,\gamma} = \frac{1}{N} \sum_{t=1}^{N} [\alpha_P + \gamma_P f(r_{m,t})] \tag{14}$$

where $f(r_{m,t})$ is the HM or TM market-timing component as defined in footnote 5. This measure has the advantage of using the model's overall return forecast as the benchmark rather than the accuracy of the individual parameters. Thus, if the model's estimation errors tend to cancel out, Equation (14) may reveal an as yet undetected forecasting power. Table 6 examines this possibility using daily data because one might hope that the HM or TM models will produce their best estimates in this case.[18]

To create Table 6, parameters are estimated quarterly and the portfolios rebalanced accordingly. The data is then sorted on the basis of the in-sample estimate of Equation (14). Panel A provides results from the one-factor model. As the α and γ_{TM} columns show, sorting on (14) in-sample then leads to out-of-sample sorts on these two parameters that are in inverse order. Though the rank correlation coefficient on the HM timing measure has the right sign, it is not significant. Also, none of the deciles

[18] The same tests reported in Table 6 were also conducted with monthly data. Because the conclusions to be drawn from the monthly data are similar to those from the daily data, they are not reported here. Interested readers can obtain a copy of the table from the authors.

Table 6
Four-Factor Adjusted Return and Market-timing Coefficients for Mutual Fund Deciles Cross Estimated BB 2004 Measure with Daily Data (1970–2002)

γ sorted deciles	α (bp)	β	γ_{HM}	γ_{TM}	T_α	T_β	$T_{\gamma HM}$	$T_{\gamma TM}$
A. Deciles sorted by 1-factor BB measure for HM market-timing.								
Decile 1	0.63	0.94	0.0085	0.2893	(1.36)	(58.67)	(0.44)	(2.14)
Decile 2	0.48	0.90	0.0064	0.3561	(1.33)	(59.05)	(0.38)	(3.58)
Decile 3	0.17	0.90	0.0100	0.2004	(0.65)	(102.21)	(1.02)	(3.16)
Decile 4	−0.02	0.90	0.0128	0.2853	(0.07)	(93.29)	(1.13)	(3.67)
Decile 5	0.07	0.89	0.0144	0.2024	(0.35)	(110.64)	(1.73)	(2.76)
Decile 6	0.06	0.89	0.0022	−0.0117	(0.31)	(129.84)	(0.24)	(0.18)
Decile 7	−0.11	0.91	0.0076	0.0995	(0.5)	(144.64)	(0.73)	(1.36)
Decile 8	−0.14	0.94	0.0064	0.1329	(0.57)	(109.79)	(0.54)	(1.55)
Decile 9	0.02	0.98	0.0010	−0.2421	(0.07)	(85.04)	(0.06)	(2.64)
Decile 10	0.25	0.99	0.0371	0.7709	(0.5)	(35.09)	(1.06)	(2.8)
D10–D1	−0.38	0.05	0.0286	0.4816	(0.49)	(1.62)	(0.69)	(1.37)
Rank correlation	−0.63	0.62	0.31	−0.07	[0.05]	[0.06]	[0.38]	[0.85]
B. Deciles sorted by 4-factor BB measure for HM market-timing.								
Decile 1	−0.17	0.93	−0.0006	0.1475	(0.45)	(65.75)	(0.04)	(1.35)
Decile 2	0.38	0.89	0.0268	0.4943	(1.32)	(60.37)	(1.68)	(5.11)
Decile 3	0.04	0.91	0.0139	0.3314	(0.16)	(78.23)	(1.03)	(4.2)
Decile 4	0.00	0.90	0.0042	0.0733	(0)	(132.7)	(0.54)	(1.11)
Decile 5	0.24	0.91	0.0031	0.1094	(1.17)	(104.05)	(0.35)	(1.52)
Decile 6	−0.06	0.92	0.0091	0.2343	(0.27)	(126.15)	(0.78)	(3.03)
Decile 7	0.11	0.91	0.0081	0.2002	(0.54)	(125.08)	(0.8)	(3.16)
Decile 8	0.03	0.94	−0.0061	−0.1572	(0.12)	(159.54)	(0.54)	(2.03)
Decile 9	0.08	0.97	−0.0002	−0.0127	(0.32)	(124.02)	(0.02)	(0.16)
Decile 10	0.84	0.95	0.0495	0.6363	(1.92)	(45.66)	(1.87)	(3.43)
D10–D1	1.00	0.03	0.0501	0.4888	(1.73)	(1.27)	(1.64)	(2.37)
Rank correlation	0.45	0.75	0.20	−0.09	[0.19]	[0.01]	[0.58]	[0.82]

At the beginning of each quarter, a market-timing performance measure (BB) is estimated using the previous quarter of return data. First, gammas are estimated with the HM model using the market return and its lagged value or the four Carhart (1997) factors and their lagged values plus the appropriate timing variable. Second, the BB measure is defined as $r_{p,\gamma} = \frac{1}{N}\sum_{t=1}^{N}[\alpha_P + \gamma_P f(r_{m,t})]$, where $f(r_{m,t})$ is the HM or TM market-timing component. Based on these measures funds are sorted into decile portfolios (decile 10 with the highest BB measure), which are rebalanced quarterly. The out-of-sample returns are then regressed on the four-factors and their lagged values using the HM or TM model to generate out-of-sample gamma values. These are reported in the γ_{HM} and γ_{TM} columns, respectively. Next the same out-of-sample return series is regressed on the four Carhart factors and their lagged values, but not the market-timing variable, to generate the out-of-sample α and market β values reported in their respective columns. Corresponding t -statistics are reported in the last 4 columns. Line "D10–D1" reports the same parameters for the difference between decile 10 and decile 1 funds. The last line reports the Spearman rank correlation for parameters across the 10 deciles, as well as the corresponding p -values. The data uses all domestic equity mutual funds (ICDI-OBJ: AG, BL, GI, IN, LG, PM, SF, and UT). Finally, no timing ability can be detected if $\gamma_P f(r_{m,t})$ is added to alpha only when γ_P is significant at the 5% level.

produce HM timing measures that are statistically different from zero at the 5% level. Sorts based on the four-factor model produce results that are little better. Although the alphas and HM timing parameters generate rank correlations of the right sign, again they are statistically insignificant. Furthermore, the TM timing parameter now has the wrong sign on its rank correlation. Overall, there appears to be little evidence that sorts based on (14) provide any better out-of-sample predictability than those based directly on the underlying parameter estimates.

Comparing Table 5 to Table 6, one can see the difference in the out-of-sample performance of the Kalman filter HM models. Unlike the HM model, in-sample sorts based on the Kalman model's market-timing measure sort the funds out-of-sample by their HM and TM timing parameters as well. From a practical standpoint, the Kalman filter results are also easier to utilize. Investors wishing to use the results in Table 6 need to rebalance their portfolio every quarter, which can yield large tax bills and other transactions costs. In contrast, the numbers in Table 5 use monthly data with portfolios that are rebalanced annually. This makes it much easier to exploit any of the Kalman model's results that an investor may wish to take advantage of.

4.2 A general omnibus parameter test

Some of the results presented in Section 3 indicate that the statistical models may balance errors in one variable against those in another. In the case of the TM and HM models this may account for the negative correlation between the alpha and gamma estimates. Thus, the only way to judge a model's overall fit is to control for all of its estimated parameters simultaneously and then examine the resulting portfolio's sample statistics.

To create omnibus out-of-sample tests, this article proposes a three-step process. First, each model is estimated fund by fund. Second, using the estimated parameters a portfolio with a forecasted zero alpha, zero beta, and (where appropriate) zero gamma is created. This is done by going long the fund, taking countervailing positions in the underlying factors, and then subtracting the predicted alpha value. By repeating the above procedure, a time series of returns is produced (1970–2002 for this article's data set). Third, the resulting return sequence is then regressed against the appropriate factor model. A model without any forecasting error should produce portfolios that yield excess returns (alphas) and factor loadings (betas and gammas) of exactly zero. Positive regression parameters indicate that a model has *underestimated* a value, while negative regression parameters imply the opposite.

For the omnibus test described, eight models are tested: the OLS, Kalman, TM, and HM models in both their one- and four-factor forms. In each test the same model is used for all three steps. Thus, if the one-factor OLS model is used to create the portfolio (steps one and two), then the one-factor OLS model is used to determine the out-of-sample distribution of the portfolio's loadings (step three). The only exceptions are the Kalman filter models. The one-factor Kalman filter model is tested out of sample with the one-factor OLS model. Similarly, the four-factor Kalman filter model is paired with the four-factor OLS model. This asymmetric treatment of the Kalman filter models derives from the theoretical properties that an out-of-sample portfolio should possess if the model that created it is properly specified. With a properly specified model the first two steps should generate portfolio returns that have time-invariant parameters equal to zero. A static OLS model should thus be the ideal instrument with which to capture or reject this hypothesis.

Why not use the TM and HM models for the out-of-sample Kalman tests? If the TM and HM models properly describe the data-generation process, then there is no harm in using them. In such cases the out-of-sample portfolio returns they produce should have constant zero-valued parameters, and the step-three model estimates should reflect this. Thus, the use of the TM and HM models for their own step-three testing is both logically consistent and has the advantage of producing a bootstrapped distribution for the out-of-sample market-timing parameter (γ). As discussed earlier, however, these two models tend to produce nonzero gammas at too high a rate even when there is every reason to believe that the portfolios in question have zero gammas. Once the TM and HM models no longer form the null hypothesis, there is no reason to believe that their out-of-sample step three estimates will be unbiased. Thus, using them in step three to test whether the OLS or Kalman models have correctly hedged out each fund's market-timing ability is problematic.[19]

Table 7 reports the distributions from the three-step omnibus tests proposed above. Portfolio returns are bootstrapped with replacement 1000 times. The forecast errors are also broken down by fund turnover. As noted earlier, it seems intuitive that the OLS models should do better when fund turnover is low, and the Kalman filter model when it is high. Unlike the previous set of tables, Table 7 also includes OLS models without the gamma timing parameter. This was done to see if the timing parameter actually helps or hinders the estimation of the other factors.

Each panel in Table 7 includes three sets of parameter statistics. The first is the alpha error, the second the "return-weighted beta error," and the third the gamma error for the HM and TM models. The alpha and gamma errors are simply the estimated alphas and gammas of the supposedly zero-alpha, zero-beta, and zero-gamma portfolios. The return-weighted beta is a variable designed to capture the overall misestimate of the factor loadings within a single statistic. It is constructed by multiplying the factor loadings estimated on the out-of-sample predicted zero-alpha and zero-beta (and zero gamma for the HM and TM models) returns by each factor's average value over the sample period and adding the products together:

$$\text{return weighted beta error} \equiv \sum_i \hat{\beta}_i \overline{r}_i. \tag{15}$$

[19] For the Kalman filter model it is also somewhat unnecessary. The Kalman filter model attempts to forecast the time-varying factor loadings. Assuming it is successful then, by construction, the step-two fund returns should lack any market-timing ability.

Table 7
Out-of-Sample Returns for Zero-Alpha and Zero-Beta Portfolios

	mean	std	5%	10%	50%	90%	95%
Panel A: Results for the 1/3 of all funds with the lowest turnover ratio.							
Alpha error							
OLS 1F	3.35	1.68	0.68	1.20	3.30	5.47	6.19
OLS 4F	−2.70	1.17	−4.62	−4.27	−2.71	−1.20	−0.86
HM 1F	5.16	2.69	0.78	1.68	5.06	8.60	9.49
HM 4F	−8.99	1.89	−12.24	−11.35	−9.03	−6.64	−5.90
TM 1F	9.29	2.57	5.09	6.08	9.28	12.75	13.54
TM 4F	−3.82	1.46	−6.15	−5.62	−3.86	−1.95	−1.45
KAL 1F	1.20	1.88	−2.01	−1.33	1.26	3.55	4.15
KAL 4F	−6.02	2.04	−9.50	−8.49	−6.09	−3.40	−2.61
Return weighted beta error (bp)							
OLS 1F	0.02	0.21	−0.33	−0.24	0.02	0.28	0.36
OLS 4F	2.70	0.55	1.81	2.00	2.73	3.39	3.57
HM 1F	−1.12	0.35	−1.70	−1.57	−1.11	−0.68	−0.57
HM 4F	1.24	0.65	0.17	0.46	1.24	2.08	2.29
TM 1F	1.07	0.44	0.39	0.50	1.07	1.64	1.80
TM 4F	2.88	0.57	1.93	2.16	2.88	3.57	3.84
KAL 1F	−0.11	0.25	−0.54	−0.43	−0.11	0.22	0.30
KAL 4F	1.91	0.83	0.55	0.87	1.91	2.96	3.19
Gamma error							
HM 1F	−0.01	0.01	−0.02	−0.02	−0.01	0.00	0.01
HM 4F	0.04	0.01	0.02	0.03	0.04	0.05	0.05
TM 1F	−0.10	0.04	−0.16	−0.14	−0.10	−0.05	−0.04
TM 4F	0.04	0.03	−0.01	0.00	0.04	0.08	0.09
Panel B: Results for the 1/3 of all funds with middle turnover ratio.							
Alpha error							
OLS 1F	1.37	1.49	−1.14	−0.57	1.34	3.29	3.80
OLS 4F	−4.42	1.25	−6.46	−6.03	−4.46	−2.80	−2.27
HM 1F	9.28	2.89	4.70	5.76	9.14	12.94	14.13
HM 4F	−5.38	2.26	−9.02	−8.19	−5.37	−2.56	−1.54
TM 1F	11.11	2.53	6.95	7.83	11.11	14.36	15.23
TM 4F	−1.64	1.50	−4.28	−3.61	−1.61	0.21	0.78
KAL 1F	−2.64	1.77	−5.60	−4.98	−2.57	−0.42	0.26
KAL 4F	−4.74	1.72	−7.49	−6.89	−4.78	−2.57	−1.96
Return weighted beta error (bp)							
OLS 1F	0.05	0.25	−0.34	−0.25	0.04	0.37	0.48
OLS 4F	−0.59	0.69	−1.74	−1.45	−0.59	0.34	0.56
HM 1F	−1.33	0.44	−2.08	−1.90	−1.33	−0.76	−0.61
HM 4F	−0.87	0.82	−2.27	−1.95	−0.83	0.17	0.44
TM 1F	0.27	0.41	−0.43	−0.27	0.27	0.78	0.91
TM 4F	−0.54	0.65	−1.60	−1.39	−0.54	0.24	0.53
KAL 1F	0.09	0.28	−0.38	−0.27	0.09	0.44	0.55
KAL 4F	−2.19	0.81	−3.48	−3.24	−2.19	−1.15	−0.88

Table 7
(Continued)

	mean	std	5%	10%	50%	90%	95%
Gamma error							
HM 1F	−0.02	0.01	−0.03	−0.03	−0.02	0.00	0.00
HM 4F	0.03	0.01	0.02	0.02	0.03	0.05	0.05
TM 1F	−0.18	0.05	−0.25	−0.23	−0.18	−0.12	−0.10
TM 4F	−0.01	0.04	−0.08	−0.07	−0.02	0.04	0.05
Panel C: Results for the 1/3 funds of all funds with the highest turnover ratio.							
Alpha error							
OLS 1F	11.78	2.28	8.17	8.92	11.75	14.74	15.43
OLS 4F	−4.22	1.45	−6.65	−6.08	−4.26	−2.39	−1.82
HM 1F	15.87	3.00	11.06	12.05	15.72	19.72	20.78
HM 4F	−2.27	2.47	−6.18	−5.34	−2.34	0.84	1.71
TM 1F	14.67	2.64	10.53	11.22	14.47	18.23	19.16
TM 4F	−2.16	1.63	−4.81	−4.29	−2.12	−0.06	0.45
KAL 1F	−2.68	3.05	−7.94	−6.63	−2.58	1.18	2.18
KAL 4F	1.64	2.34	−2.11	−1.44	1.72	4.59	5.58
Return weighted beta error (bp)							
OLS 1F	−0.99	0.38	−1.63	−1.48	−0.98	−0.50	−0.40
OLS 4F	−1.13	0.79	−2.45	−2.19	−1.13	−0.09	0.19
HM 1F	−0.46	0.57	−1.38	−1.18	−0.48	0.23	0.52
HM 4F	−2.03	0.99	−3.65	−3.31	−2.05	−0.78	−0.36
TM 1F	1.33	0.49	0.52	0.73	1.30	1.96	2.21
TM 4F	−1.39	0.82	−2.81	−2.45	−1.38	−0.34	−0.05
KAL 1F	−0.55	0.41	−1.23	−1.05	−0.54	−0.03	0.13
KAL 4F	−3.46	0.97	−5.06	−4.73	−3.47	−2.22	−1.86
Gamma error							
HM 1F	−0.02	0.01	−0.04	−0.03	−0.02	−0.01	0.00
HM 4F	0.06	0.01	0.04	0.05	0.07	0.08	0.08
TM 1F	−0.20	0.06	−0.29	−0.27	−0.19	−0.12	−0.10
TM 4F	0.13	0.06	0.01	0.05	0.13	0.21	0.22

For all domestic equity mutual funds (ICDI-OBJ: AG, BL, GI, IN, LG, PM, SF, and UT) that had at least five years of monthly return data, both 1-factor and 4-factor OLS, TM, HM, and Kalman models are used to forecast a fund's alpha and beta (and for the HM and TM models gamma) at the beginning of each year from 1970 to 2002. These forecasts are then used to construct fund-by-fund zero-alpha and zero-beta (and if appropriate zero gamma) portfolios. Next, the resulting monthly time series for the zero-alpha and zero-beta portfolio is regressed against the market factor (for one-factor models), the four factors (for four-factor models), or the corresponding factors plus market-timing component (for HM and TM models). This process results in risk-adjusted returns (the alpha error) and factor loadings for each zero-alpha zero-beta portfolio. A gamma error is also produced when zero-alpha, zero-beta, and zero-gamma portfolios are produced. The parameter distributions are then calculated by bootstrapping with replacement the above procedure 1,000 times. The return-weighted beta error is defined as: *return-weighted beta error* $\equiv \sum_i \hat{\beta}_i \overline{r}_i$, where $\hat{\beta}_i$ is the estimated value of factor $\boldsymbol{i}$, and $\overline{r}_i$ the factor's average return during the sample period. The Kolmogorov–Smirnov test rejects the hypothesis that the same probability distribution produced any pair of distributions generated by the different models (all p -values virtually zero).

In this equation $\hat{\beta}_i$ is the estimated loading of factor i from the regression, and $\overline{r}_i$ the factor's average return during the sample period. This metric is designed to give greater weight to those factors, which, if misestimated, will yield the largest systematic errors regarding a fund's predicted performance. The closer a model comes to producing return-weighted beta errors of zero, the better it is at predicting a fund's overall future factor risks and returns.

Table 7 Panel A displays the results for low-turnover funds. Assuming that most stock betas remain fairly constant over time, these funds would seem to be ideally suited for OLS models as they are the most likely to have either static or slowly changing factor loadings. For the out-of-sample alpha errors only the one-factor Kalman model encompasses zero within the 90% confidence interval. By contrast, the four-factor OLS model produces portfolio returns with a 90% confidence of −4.62 bps (basis points) to −0.86 bps per month, indicating that it overpredicts fund alphas. The four-factor HM and TM models also produce portfolio returns that are strictly negative (implying they too overpredict the alphas) throughout the 90% confidence interval. Thus, if the goal is to produce an unbiased forecast of a fund's alpha, the one-factor Kalman model appears to provide the best performance.

The middle part of Table 7 Panel A reports the bootstrapped distribution of the return-weighted beta error. Once again only the one-factor Kalman models include zero within the 90% confidence interval.

The final section of the panel displays the results for the gamma errors. Here at least the one-factor HM and four-factor TM models encompass zero within the 90% confidence interval. However, it is worth noting that of all the funds these are the most likely to display the least market-timing ability, and two of the four models still fail to produce forecast errors around zero a reasonable fraction of the time.

Table 7 Panel B examines funds with turnover ratios in the middle third of the data and Panel C the funds with the highest turnover ratios. These are the groups that are more likely to create problems for the OLS models, and the ones that the TM and HM models attempt to address. However, the results do not provide much support for the latter. Of the models tested, only the one-factor Kalman model encompasses zero within the 90% confidence interval for both the out-of-sample alpha and return-weighted beta. The one-factor OLS model manages to do the same, but only for the middle-turnover group. The four-factor TM model comes close, failing to encompass zero only for the high-turnover group's return-weighted beta. However, the gamma error for this group does not cover zero, which is troubling, because that is the parameter on which most applications of this model concentrate.

Overall, three panels of Table 7 indicate that the one-factor Kalman model produces in-sample parameter estimates that successfully forecast a fund's future factor loadings. The other models are far less accurate on this front. Oddly, the OLS models appear to do particularly poorly with low-turnover ratio funds: the one group the model should be best adapted to. A natural conclusion is that even low-turnover funds have sufficiently dynamic factor loadings that the OLS models end up producing biased forecasts.

5. Time Variation in Mutual Fund Betas

Although the Kalman filter's parameter estimates have a number of desirable properties, are they economically reasonable? Mutual funds cannot change their portfolio holdings very quickly, so their estimated factor loadings should not vary too much or too fast. Figure 1 provides some intuition regarding the actual time variation in fund factor loadings and the Kalman filter model's ability to track the observed changes. This figure has three panels, each of which presents a set of comparative graphs for a representative fund by turnover group.

The first fund in Figure 1 is ICDI 660. It has a low annual turnover of 0.38. Not too surprisingly, it also has fairly time-invariant factor loadings. The green lines represent the rolling OLS estimates and the blue lines the Kalman filter estimates. First, note that while a $\sigma_{\beta,CAPM}$ of 0.12 may appear to be high, it actually produces a relatively stable time series. The other thing to note is that the rolling OLS and Kalman filter estimates track each other fairly well, although not perfectly. In general, the Kalman filter estimates appear smoother and less prone to sudden shifts. Given the out-of-sample evidence presented earlier, it seems likely that the sharp fluctuations in the OLS parameter estimates are due to estimation errors of some sort and not real changes in the fund's holdings.

The next set of graphs examine fund ICDI 14 670. It has an average turnover level of 0.69. Here, $\sigma_{\beta,CAPM}$ equals 0.16, but nevertheless the model produces a factor-loading time series that slowly increases throughout the 1990s. Again the model has several interacting variables, and simply looking at the implied variance of the market factor fails to tell the whole story. Also note that here, too, the Kalman filter estimates are at least as smooth as the rolling OLS estimates. For example, the Kalman filter model finds that the fund's estimated alpha is near zero throughout most years except for a few short negative periods. In contrast, the OLS model claims that the fund's alpha has varied dramatically from year to year, a feat that seems highly unlikely given what we know about the forecasting power typically associated with such models.

The third panel in Figure 1 looks at fund (ICDI 8100), which has a high turnover ratio of 1.22. The comparative results between the OLS and Kalman models are qualitatively similar to those in the other two panels. Here, though, both models indicate that the fund's factor loadings have

Fund ICDI	Turnover	$\sigma_{\beta,CAPM}$	$\sigma_{\beta,FF}$
660	0.38	0.12	0.05
14670	0.69	0.16	0.12
8100	1.22	0.23	0.21

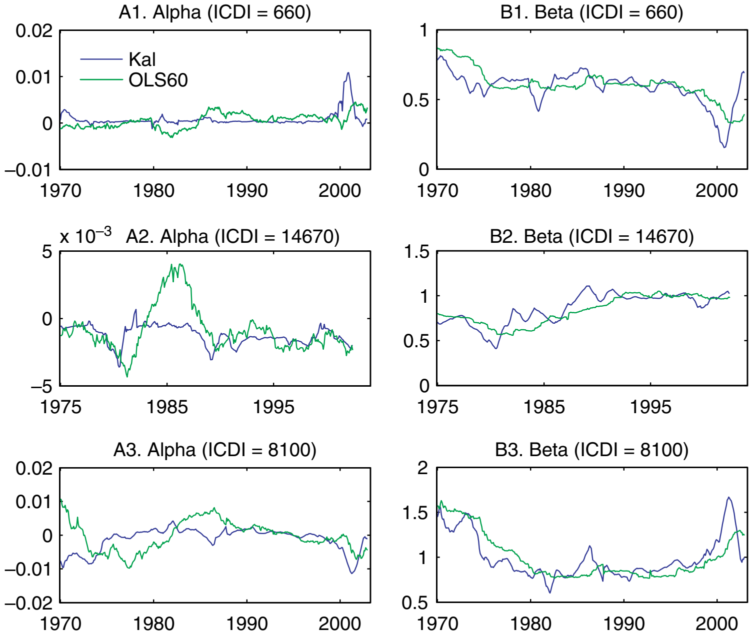

Figure 1
Time series of Kalman model estimated mutual fund alphas and betas.
First, Kalman one- and four-factor models are used to estimate the time series of mutual fund alphas and betas based on available monthly return data in the period from 1970 to 2002. The blue lines plot the estimated time series for one-factor alpha and beta (for the one-factor Kalman model in Panels A and B) for three mutual funds with low, medium, and high turnover ratios. To save space, four-factor exposures and risk-adjusted returns are not plotted. Green lines plot the rolling OLS values of the corresponding parameter. The OLS alpha and beta values in month t are estimated from month $t-60$ to month $t-1$. Fund information is summarized in the table.

varied quite a bit over the years. However, once again the OLS values often tend to fluctuate more than the Kalman filter estimates, especially in regard to the nonmarket factors SMB, HML, and MOM.

Although Figure 1 provides some insight regarding the volatility of the Kalman filter estimates, Table 8 examines the issue more broadly. In this table the Kalman model's parameter estimates are broken down by fund turnover and the factor model used. The penultimate column in each panel displays the average estimated time variation in the fund betas for each turnover group (σ_β), and the last column provides the variance of the residual from fitting an AR(1) to each fund's beta over time. The former measure yields the total variation,

Table 8
Average Cross-Sectional Kalman Parameter Statistics

A. Kalman parameters for 1-factor dynamic model

Turnover	−k	$\overline{\beta}_p$	υ	σ_ε	σ_η^2	$\overline{\alpha}_p$	b_P	σ_β	$\sigma_{\beta,AR(1)}$
Low	0.0007	0.7825	0.5747	0.0224	0.0491	−0.0004	−0.2686	0.1375	0.1284
Media	−0.0004	0.8825	0.5815	0.0224	0.0627	0.0006	−0.1272	0.1529	0.1407
High	−0.001	1.0393	0.543	0.0265	0.1029	0.0226	0.1709	0.1787	0.1604
T-stat									
Low	(4.87)	(94.23)	(72.48)	(20.53)	(20.15)	(0.08)	(7.26)	(61.46)	(32.86)
Media	(1.21)	(98.04)	(75.50)	(27.65)	(18.54)	(0.13)	(3.06)	(64.59)	(31.76)
High	(4.57)	(92.17)	(70.45)	(30.92)	(19.03)	(6.21)	(3.90)	(64.64)	(41.59)

B. Kalman parameters for 4-factor dynamic model

Turnover	$-k$	$\overline{\beta}_{MKT}$	$\overline{\beta}_{SMB}$	$\overline{\beta}_{HML}$	$\overline{\beta}_{MOM}$	υ	σ_ε	σ_η^2	c_{SMB}	c_{HML}	c_{MOM}	$\overline{\alpha}_p$	b_P	σ_β	$\sigma_{\beta,AR(1)}$
Low	−0.0006	0.8667	0.096	0.146	−0.0253	0.5513	0.0141	0.0501	0.1153	−0.2588	−0.0516	−0.0085	0.0857	0.1285	0.1038
Media	−0.0014	0.9257	0.1824	0.0804	0.0094	0.5925	0.0141	0.0545	−0.0823	−0.1336	0.129	0.0032	0.0899	0.1407	0.1101
High	−0.0019	0.9723	0.2832	−0.0828	0.0622	0.5907	0.0173	0.0515	−0.1454	−0.8213	0.2165	0.0137	0.1016	0.1602	0.127
T-stat															
Low	(4.19)	(142.93)	(10.29)	(9.87)	(7.94)	(58.90)	(11.91)	(10.78)	(0.97)	(1.59)	(0.56)	(1.24)	(1.73)	(32.88)	(23.97)
Media	(6.74)	(148.89)	(15.85)	(7.68)	(2.10)	(64.60)	(17.69)	(9.48)	(0.67)	(0.83)	(1.01)	(0.42)	(1.94)	(31.76)	(23.03)
High	(5.04)	(122.13)	(25.94)	(6.44)	(9.49)	(68.15)	(25.60)	(15.44)	(1.24)	(5.43)	(1.76)	(1.79)	(2.03)	(41.59)	(28.51)

This table reports the estimated Kalman parameters for a one- and four-factor model. The Kalman parameters are estimated based on monthly return data from 1970 to 2002. Mutual funds are sorted into three groups according to the average turnover ratio during the same period (the break points occur at annual turnover rates of 0.5596 and 1.0107). Each entry then reports statistics regarding the parameter values within the same turnover group. The last two columns contain the standard deviation of the fund's market exposure (σ_β) and that of the residual from an AR(1) fit of the fund's market exposure ($\sigma_{\beta,AR(1)}$). The cross-section t-statistics are reported in parentheses.

and the latter may be interpreted as the surprise in beta's change from period to period. As both columns indicate, an increase in fund turnover increases the estimated volatility of the fund's market beta. This result also indicates that the conjecture given in footnote 9, that stock betas change over time and that fund managers trade to maintain constant portfolio loadings, is not supported by the data.

Beyond the beta volatility figures, the other parameter estimates in Table 8 also seem to conform to expectations. For the middle-turnover groups the persistence parameter υ is about 0.58 and the standard deviation of the signal's innovation (σ_η^2) equals 0.0627. The combination of the relatively rapid revision of the fund's beta to its long-run mean at a rate of approximately 45% per month and the signal's relatively modest variance implies that though betas can drift somewhat from their long-run means, they do not do so for very long. This in turn leads to the relatively modest values observed in the σ_β column.

6. Comparison with the FS Model

So far all tests have been conducted under the restriction that the fund betas depend only upon some unobservable factor. This section examines the impact on the estimated model when observable conditioning information is added. The tests conducted here use the lagged Treasury-bill rate and the dividend yield on the CRSP value-weighted index. Thus, the equation for β_{Pt} becomes

$$\beta_{Pt} = \beta + F_{t-1} + k_1 z_{1,t-1} + k_2 z_{2,t-1}. \tag{16}$$

If the observable information improves the model's predictive ability, then k_1 and k_2 should differ from zero. This is tested by running the model with and without the conditioning variables on monthly returns during the period 1970–2002. Asymptotically, the likelihood ratio under the null should follow a chi-square distribution with two degrees of freedom. In Figure 2, the bars represent the cross-sectional distribution of the likelihood ratio, and the dashed line traces out a chi-square distribution with two degrees of freedom. Overall, the null hypothesis that k_1 and k_2 are zero cannot be rejected at the traditional 1, 5, or 10% levels. However, for individual funds, the fraction that rejects the null hypothesis at the 1, 5, or 10% level is 10, 20, and 27%, respectively. These numbers are somewhat higher than might be expected by chance, which implies that for some funds the conditioning information appears to improve the model's fit.

Table 9 provides further evidence about the richness of the dynamic coefficient model used in this article. This table shows the R^2's for the fund return regressions using the OLS, the FS two-factor conditional beta

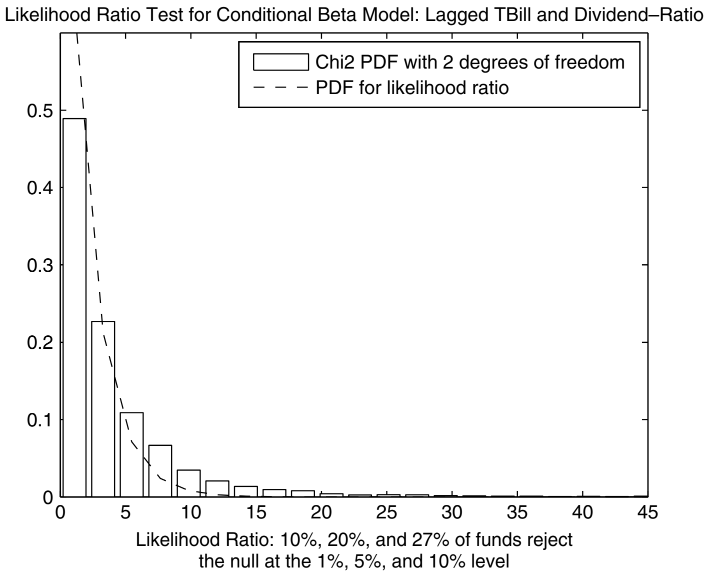

Figure 2
Likelihood ratio tests.
This figure shows the results from a likelihood ratio test in which the Kalman filter model includes conditional information similar to that of Ferson and Schadt (1996). As in their application, lagged macroeconomic information drives the portfolio weights. However, here portfolio weights are also assumed to vary from some unobserved factor following an AR(1) process. As a first-order approximation the portfolio weights are set to $w_{it} = \overline{w}_i + l_i F_t + D_i Z_{t-1}$, where Z_{t-1} is the lagged information. This model is estimated via an extended Kalman filter. For two macro instruments, the estimated system of equations is given by $r_{Pt} - r_{ft} = \alpha_{Pt} + \beta_{Pt}(r_{mt} - r_{ft}) + \beta_{X,t} X_t + \varepsilon_{Pt}$, where $\beta_{Pt} = \overline{\beta}_P + F_{t-1} + k_1 z_{1,t-1} + k_2 z_{2,t-1}$ and $\alpha_{Pt} = -k_t + \overline{\alpha}_P F_{t-1} + b_P F_{t-1}^2$, $F_t = \nu F_{t-1} + \eta_t$, and X_t are other factors (SMB, HML, and MOM). Here z_1 and z_2 are instruments for the lagged treasury-bill rate and the CRSP value-weighted index's dividend yield. The null hypothesis is that $k_1 = k_2 = 0$. The constrained and unconstrained models are estimated using available monthly returns of CRSP equity mutual funds during the period of 1970 to 2002 (a minimum of 60 monthly returns are required). Asymptotically, the likelihood ratio test under the null should follow a chi-square distribution with two degrees of freedom. The bars represent the cross-sectional distribution of the likelihood ratio and the dashed line displays the mathematical values for a chi-square distribution with two degrees of freedom. The fraction of funds that reject the null hypothesis at the 1, 5, or 10% levels are 10, 20, and 27%, respectively.

model, and this article's Kalman filter model across turnover terciles.[20] As can be seen, for each tercile and for the entire sample the FS model provides an improved fit relative to the OLS model (the differences are all statistically significant). Consider, however, how the R^2 statistic changes as one moves across models. The increase when one goes from the OLS to FS model is less than 10% of the increase obtained when moving from the FS to the Kalman model. On the basis of this, it appears that the Kalman model can account for a considerably larger portion of fund return fluctuations than either

[20] The Kalman model used in these tests does not use the FS conditioning variables and looks only at the return series of funds and the market index.

Table 9
A Comparison with Conditional Model

Turnover	R^2_{CAPM}	R^2_{Cond}	R^2_{Kal}	ΔR^2_1	ΔR^2_2	$T_{\Delta R^2_1}$	$T_{\Delta R^2_2}$
Panel A: 1-factor model							
Low	0.7120	0.7326	0.8713	0.0206	0.1387	(31.28)	(40.86)
Media	0.7007	0.7211	0.8683	0.0204	0.1472	(30.85)	(42.01)
High	0.6954	0.7165	0.8769	0.0211	0.1604	(25.90)	(44.73)
All	0.7027	0.7234	0.8722	0.0207	0.1488	(50.18)	(73.59)
Panel B: 4-factor model							
Low	0.8317	0.8426	0.9478	0.0109	0.1053	(30.40)	(33.84)
Media	0.8322	0.8431	0.9498	0.0109	0.1067	(28.49)	(33.18)
High	0.8271	0.8395	0.9517	0.0124	0.1122	(22.34)	(37.39)
All	0.8303	0.8417	0.9498	0.0114	0.1081	(44.76)	(60.19)

The sample used for this table includes all equity mutual funds in the CRSP database that contain at least 60 months of monthly return data from 1970 to 2002. Model parameters are estimated for the unconditional CAPM model, the conditional beta model, and the Kalman filter model. The variables R^2_{CAPM}, R^2_{Cond}, and R^2_{Kal} represent R^2 statistics for each of the models, respectively. Consistent with the OLS model, R^2_{Kal} is defined as $1 - \mathbb{E}(\varepsilon_P(t)^2)/\mathbb{E}((y_t - \overline{y})^2)$, where y_t is the excess portfolio return and $\varepsilon_P(t)$is the residual from the Kalman filter model. The variable $\overline{y}$ equals the mean value of the y_t. Also reported are the cross-sectional means and t -ratios for the improvements of the R^2 statistic: $\Delta R^2_1 = R^2_{Cond} - R^2_{CAPM}$, and $\Delta R^2_2 = R^2_{Kal} - R^2_{cond}$.

the OLS or the FS models. Also, as expected, the greatest improvements relative to the OLS model are for those funds that have the highest turnover rates, and are thus likely to have the most dynamic factor loadings.

Between this article, FS, and Grinblatt and Titman (1989) there is now considerable evidence that mutual fund managers produce portfolios with time-varying betas, and possibly alphas, too. Thus, it is clear that portfolio managers are altering their portfolios in response to some set of economic variables. Why then are k_1 and k_2 statistically indistinguishable from zero for most funds? The model has two ways of fitting a fund's alphas and betas. One way is to use the observable conditioning variables in some manner. Another is to use the estimated lagged values of alpha and beta, and let them change according to an estimated relationship with an unobserved factor following an AR(1) process. Figure 2 indicates that the latter prediction method often dominates, at least when using the lagged Treasury-bill rate and dividend yield on the CRSP value-weighted index. One conclusion may be that a few funds use Treasury-bill rates and the market dividend yield to help manage their assets, though most do not. More practically, one can use the model to identify both those funds with a more macro-based approach to asset allocation and the variables on which they concentrate.

7. Conclusion

Mutual fund managers often trade in the hope of generating superior returns. This trading naturally generates time-varying factor loadings.

However, the standard multifactor OLS is not designed to handle such time variation or to detect whether fund managers have the ability to time the market by appropriately varying their fund's beta. Potential solutions to this problem are the TM and HM models, which include a market-timing parameter in addition to the standard factors. Using these models, BB find that while there is no evidence that fund managers possess market-timing ability in the monthly data, they are able to detect it in the daily data.

In line with BB this article finds that the TM and HM models produce false-positive parameters with daily data at an extraordinarily high rate. However, the analysis here also indicates that the daily parameter estimates are very unstable, with in-sample sorts leading to little out-of-sample predictive power. To address the issue of time variation in mutual fund factor loadings, this article develops a Kalman filter model. This model generates dynamic factor loadings that can allow one to test for a fund's market-timing ability. Although the Kalman filter model does not appear to reliably find market-timing ability in the daily data, it does find such ability in the monthly data. Furthermore, the Kalman filter model's timing ability estimates in the monthly data exhibit good predictive power.

There are, of course, other dynamic models that have been proposed. The conditional model of FS is one example. Their conditional model uses macroeconomic factors to help forecast a fund's factor loadings. The Kalman filter model can accommodate the FS factors. However, tests indicate that for most funds the macroeconomic factors add little in the way of explanatory power.

References

Admati, A. 1985. A Noisy Rational Expectations Equilibrium for Multi-asset Securities Markets. *Econometrica* 53(3):629–57.

Ang, A., J. Chen, and Y. Xing. 2002. Downside Correlation and Expected Stock Returns, http://ssrn.com/abstract=282986, Accessed July 27, 2007.

Blake, D., B. Lehmann, and A. Timmermann. 1999. Asset Allocation Dynamics and Pension Fund Performance. *Journal of Business* 72(4):429–61.

Blake, D., B. Lehmann, and A. Timmermann. 2002. Performance Clustering and Incentives in the UK Pension Fund Industry. *Journal of Asset Management* 3(2):173–94.

Bollen, N., and J. Busse. 2001. On the Timing Ability of Mutual Fund Managers. *Journal of Finance* 56(3):1075–94.

Bollen, N., and J. Busse. 2005. Short-Term Persistence in Mutual Fund Performance. *Review of Financial Studies* 18(2):569–97.

Brown, S., and W. Goetzmann. 1995. Performance Persistence. *Journal of Finance* 50(2):679–98.

Brown, S., and W. Goetzmann. 1997. Mutual Fund Styles. *Journal of Financial Economics* 43(3):373–99.

Carhart, M. 1997. On Persistence in Mutual Fund Performance. *Journal of Finance* 52(1):57–82.

Christopherson, J., W. Ferson, and D. Glassman. 1998. Conditioning Manager Alphas on Economic Information: Another Look at the Persistence of Performance. *Review of Financial Studies* 11:111–42.

Cohen, R., J. Coval, and L. Pastor. 2005. Judging Fund Managers by the Company They Keep. *Journal of Finance* 60(3):1057–96.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring Mutual Fund Performance with Characteristic-Based Benchmarks. *Journal of Finance* 52(3):1035–58.

Ferson, W., and C. Harvey. 1991. The Variation of Economic Risk Premiums. *Journal of Political Economy* 99:385–415.

Ferson, W., and C. Harvey. 1993. The Risk and Predictability of International Equity Returns. *Review of Financial Studies* 6:527–66.

Ferson, W., and R. Korajczyk. 1995. Do Arbitrage Pricing Models Explain the Predictability of Stock Returns? *Journal of Business* 68(3):309–49.

Ferson, W., and R. Schadt. 1996. Measuring Fund Strategy and Performance in Changing Economic Conditions. *Journal of Finance* 51(2):425–62.

Ferson, W., and K. Khang. 2002. Conditional Performance Measurement Using Portfolio Weights: Evidence for Pension Funds. *Journal of Financial Economics* 65:249–82.

Grinblatt, M., and S. Titman. 1989. Portfolio Performance Evaluation: Old Issues and New Insights. *Review of Financial Studies* 2(3):396–422.

Grinblatt, M., and S. Titman. 1992. The Persistence of Mutual Fund Performance. *Journal of Finance* 47(5):1977–84.

Grinblatt, M., and S. Titman. 1993. Performance Measurement Without Benchmarks: An Examination of Mutual Fund Returns. *Journal of Business* 66(1):47–68.

Grinblatt, M., and S. Titman. 1994. A Study of Monthly Mutual Fund Returns and Performance Evaluation Techniques. *Journal of Financial and Quantitative Analysis* 29(3):419–44.

Harvey, A. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Hendricks, D., J. Patel, and R. Zeckhauser. 1993. Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974–1988. *Journal of Finance* 48(1):93–130.

Henriksson, R. D., and R. C. Merton. 1981. On Market-timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills. *Journal of Business* 54(4):513–33.

Jagannathan, R., and R. Korajczyk. 1986. Assessing the Market-timing Performance of Managed Portfolios. *Journal of Business* 59(2):217–35.

Lehmann, B., and D. Modest. 1987. Mutual Fund Performance Evaluation: A Comparison of Benchmarks and Benchmark Comparisons. *Journal of Finance* 42:233–65.

Pástor, L., and R. Stambaugh. 2002. Mutual Fund Performance and Seemingly Unrelated Assets. *Journal of Financial Economics* 63(3):315–50.

Teo, M., and S. Woo. 2001. Persistence in Style-Adjusted Mutual Fund Returns. Working paper, Harvard University, Department of Economics.

Treynor J. L., and K. Mazuy. 1966. Can Mutual Funds Outguess the Market? *Harvard Business Review* 44:131–36.

Wermers, R. 2000. Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses. *Journal of Finance* 55(4):1655–95.