

Measuring the Timing Ability and Performance of Bond Mutual Funds

by
Yong Chen
Wayne Ferson
Helen Peters

first submission: April, 2009
this revision: August 24, 2009

* Ferson is the Ivadelle and Theodore Johnson Chair of Banking and Finance and a Research Associate of the National Bureau of Economic Research, Marshall School of Business, University of Southern California, 3670 Trousdale Parkway Suite 308, Los Angeles, CA. 90089-0804, ph. (213) 740-5615, ferson@marshall.usc.edu, www-ref.usc.edu/~ferson/. Peters is Professor of Finance at Boston College, 140 Commonwealth Ave, Chestnut Hill, MA. 02467, (617) 552-3945, helen.peters.1@bc.edu. Chen is Assistant Professor of Finance at Virginia Tech, Blacksburg, VA. 24061, yong.chen@vt.edu. We are grateful to participants in workshops at Baylor University, Brigham Young, Boston College, DePaul, the University of Connecticut, the University of Florida, the University of Georgia, Georgetown, the Hong Kong University of Science and Technology, the University of Iowa, the University of Massachusetts at Amherst, Michigan State University, the University of Michigan, the National University of Singapore, Nanyang Technological Institute, New York University, Singapore Management University, Southern Methodist University, the University of Oklahoma, the University of Kansas, Virginia Tech, York University and the Office of Economic Analysis, Securities and Exchange Commission for feedback, as well as to participants at the following conferences: The 2008 China International Conference in Finance, the 2005 Global Finance Conference at Trinity College, Dublin, the 2006 Journal of Investment Management Conference, the 2008 ISCTE/NOVA Business School conference, the 2005 Southern Finance and the 2005 Western Finance Association Meetings. We also appreciate helpful comments from an anonymous referee and from Vasantha Chigurupati, Walter Dolde, James Hilliard, Scott Linn, Ricardo Rodriguez, Antonios Sangvinatos, Bryan Schmutz, M. Abdullah Sahin, Nadia Vozlyublennaia and Hong Yan.

Measuring the Timing Ability and Performance of Bond Mutual Funds

ABSTRACT

This paper evaluates the ability of bond funds to "market time" nine common factors related to bond markets. Timing ability generates nonlinearity in fund returns as a function of common factors, but there are several non-timing-related sources of nonlinearity. Controlling for the non-timing-related nonlinearity is important. Funds' returns are more concave than benchmark returns, and this would appear as poor timing ability in naive models. With controls, the timing coefficients appear neutral to weakly positive. Adjusting for nonlinearity the performance of many bond funds is significantly negative on an after-cost basis, but significantly positive on a before-cost basis.

JEL classification: C15; C31; G12

Keywords: Mutual funds; Market timing; Bond funds; Investment performance evaluation

1. Introduction

The amount of academic research on bond fund performance is small in comparison to the economic importance of bond funds. Recently the total net assets of U.S. bond funds has been about 1/6 the amount in equity-style mutual funds and similar to the value of hedge funds. Large amounts of additional fixed-income assets are held in professionally managed portfolios outside of mutual funds, for example in pension funds, trusts and insurance company accounts. The turnover of a typical bond mutual fund far exceeds that of a typical equity fund (e.g. Moneta, 2008), suggesting active portfolio management. Thus, it is important to understand the performance of bond fund managers.

Elton, Gruber and Blake (EGB, 1993, 1995) and Ferson, Henry and Kisgen (2006) study US bond mutual fund performance, concentrating on the funds' risk-adjusted returns. They find that the average performance is slightly negative after costs, and largely driven by funds' expenses.

This might suggest that investors would be better off selecting low-cost passive funds, and EGB draw that conclusion. However, conceptually at least, performance may be decomposed into components, such as timing and selectivity. If investors place value on timing ability, for example a fund that can mitigate losses in down markets, they would be willing to pay for this insurance with lower average returns. This is one of the first papers to comprehensively study the ability of US bond funds to time their markets.¹

Timing ability on the part of a fund manager is the ability to use superior information about the future realizations of common factors that affect bond market returns.² Selectivity refers to the use of security-specific information. If common factors explain a significant part of the variance of bond returns, consistent with term structure studies such as Litterman and Sheinkman (1991), then a significant fraction of the potential performance of bond funds might be attributed to timing. However, measuring the timing ability of bond funds is a subtle problem.

Traditional models of market timing ability rely on convexity in the relation between the fund's returns and the common factors.³ This paper looks at timing ability and performance after adjusting for four potential biases. First, there might be a nonlinear relation between economic factors and a fund's benchmark portfolio. Second, interim trading, where the fund rebalances more frequently than the return observation interval, can generate nonlinearity. Third, stale pricing

¹ Brown and Marshall (2001) develop an active style model and an attribution model for fixed income funds, isolating managers' bets on interest rates and spreads. Comer, Boney and Kelly (2009) study timing ability in a sample of 84 high quality corporate bond funds, 1994-2003, using variations on Sharpe's (1992) style model. Aragon (2005) studies the timing ability of balanced funds for bond and stock indexes.

² We do not explicitly study "market timing" in the sense recently taken to mean trading by investors in a fund to exploit stale prices reflected in the fund's net asset values. But we will see that these issues can affect measures of a fund manager's ability.

³ The alternative approach is to directly examine managers' portfolio weights and trading decisions to see if they can predict returns and factors (e.g. Grinblatt and Titman, 1989). Comer (2006), Moneta (2008) and Wang (2007) are early steps in this direction for bond funds. Of course, weight-based approaches cannot capture market timing that occurs between weight reporting intervals, which can be up to six months in length.

that is correlated with economic factors – “systematic” stale pricing – can generate nonlinearity. Finally, funds’ exposures to the factors may vary due to publicly observed conditioning variables. Most of these issues have been treated in studies cited below, but this is the first paper to combine them all and the first to consider systematic stale pricing.

We study monthly returns for more than 1400 bond funds during 1962-2007 and find that controlling for non-timing-related nonlinearity is important. Funds’ returns are typically more concave, in relation to a set of nine bond market factors, than are unmanaged benchmarks. Thus, funds would appear to have poor (negative) market timing ability in naive models. When we introduce the controls for non-timing-related nonlinearities the overall distribution of the timing coefficients appears neutral to weakly positive. After adjusting for the nonlinearity in funds’ returns, the performance of many bond funds is significantly negative on an after-cost basis but significantly positive on a before-cost basis.

The rest of the paper is organized as follows. Section 2 describes the models and methods. Section 3 describes the data. Section 4 presents our empirical results and Section 5 offers some concluding remarks.

2. Models and Methods

A traditional view of fund performance separates timing ability from security selection ability, or selectivity. Timing is closely related to asset allocation, where funds rebalance the portfolio among asset classes and cash. Selectivity means picking good securities within the asset classes. Like equity funds, bond funds engage in activities that may be viewed as selectivity or timing. Bond funds may attempt to predict issue-specific supply and demand or changes in credit risks associated with particular bond issues. Funds can also attempt to exploit liquidity differences across individual bonds. These trading activities can be classified as security selection. In addition, managers may adjust the interest rate sensitivity (e.g., duration) of the portfolio to time changes in interest rates. They may vary the allocation to asset classes differing in credit risk or

liquidity, and tune the portfolio's exposure to other economic factors. Since these activities relate to anticipating market-wide factors, they can be classified as market timing.

Classical models of market-timing use convexity in the relation between the fund's return and the "market" return to identify timing ability. In these models the manager observes a private signal about the future performance of the market and adjusts the market exposure of the portfolio. If the response is assumed to be a linear function of the signal as modelled by Admati, Bhattacharya, Ross and Pfleiderer (1986), the portfolio return is a convex quadratic function of the market return as in the regression model of Treynor and Mazuy (1966). If the manager shifts the portfolio weights discretely, as modelled by Merton and Henriksson (1981), the convexity may be modelled with call options. We modify the classical setup to control for nonlinearities that are unrelated to bond fund managers' timing ability.

2.1 Nonlinearity Unrelated to Timing

There are many reasons apart from timing ability that a fund's return could have a nonlinear relation to a market factor. We group these reasons into four general categories. This section explains the intuition for each effect.

First, the underlying assets held by a fund may have a nonlinear relation to market factors. For example, a fund that holds call options bears a convex relation to the underlying asset (Jagannathan and Korajczyk, 1986). Pseudo-timing caused by this kind of nonlinearity is treated by Glosten and Jagannathan (1994) for equity mutual funds and by Fung and Hsieh (2001) and Bondarenko (2004) for hedge funds. Nonlinearity is very likely in bond funds. Even simple bond returns are nonlinearly related to interest rate changes. Callable and convertible bonds contain explicit option components. As we show in Table 3 below, unmanaged bond benchmark returns are convex functions of common factors. Thus, to measure timing ability it is important to control for this nonlinearity.

A second, related cause of nonlinearity is "interim trading," studied by Goetzmann,

Ingersoll and Ivkovic (2000), Ferson and Khang (2002) and Ferson, Henry and Kisgen (2006).

This refers to a situation where fund managers trade more frequently than the fund's returns are measured. With mutual funds interim trading definitely occurs. This could allow the manager to generate fake timing by increasing exposure after seeing high returns in the first half of the month. This is also related to derivatives holdings, because derivatives may often be replicated by high frequency trading. If we can control for interim trading we also control for these derivatives.⁴

A third potential reason for nonlinearity unrelated to timing ability is stale pricing of a fund's assets. Thin or nonsynchronous trading has long been known to bias downward the estimates of beta for a portfolio (e.g., Fisher (1966), Scholes and Williams, 1977). If the degree of stale pricing varies over time in relation to a market factor, such "systematic" stale pricing can create spurious concavity or convexity in the measured return.

A fourth reason for nonlinearity unrelated to timing ability arises if there is public information about future asset returns. As shown by Ferson and Schadt (1996), even if the conditional relation between the fund and a benchmark return is linear, the conditional portfolio betas may be correlated with market returns because of their common dependence on the public information, and the unconditional relation may therefore be nonlinear. Examples of public information effects on equity fund timing coefficients are provided by Ferson and Warther (1996), Becker, et al (1999), Christopherson, Ferson and Turner (1999) and Ferson and Qian (2004).

In summary, in order to measure the market timing ability of bond funds we need to control for nonlinearity in the benchmark assets and for nonlinearity that may arise from interim trading, public information effects and systematic stale pricing. In the following subsections we modify the classical market timing model to allow for nonlinearity in the benchmark returns and we develop controls for interim trading, public information effects and systematic stale pricing.

⁴ Brown et al. (2004) explore arguments that incentives and behavioral biases can induce managers without superior information to engage in option-like trading within performance measurement periods.

2.2 Classical Market Timing Models

The classical market-timing regression of Treynor and Mazuy (1966) is:

$$r_{pt} = a_p + b_p f_t + \lambda_p f_t^2 + u_t \quad (1)$$

where r_{pt} is the fund's portfolio return, measured in excess of a short-term Treasury bill. With equity market timing as considered by Treynor and Mazuy, f_t is the excess return of the stock market index. Treynor and Mazuy (1966) argue that $\lambda_p > 0$ indicates market-timing ability. The logic is that when the market is up, the successful market-timing fund will be up by a disproportionate amount. When the market is down, the fund will be down by a lesser amount. Therefore, the fund's return bears a convex relation to the market factor.

It seems natural to replace the equity market excess return with changes in the systematic factors in bond markets, like interest rate levels and spreads. However, if a factor is not an excess return, the appropriate sign for the timing coefficient might not be obvious. For example, bond returns move in the opposite direction as interest rates, so a signal that interest rates are about to rise means bond returns are likely to be low. We show in the Appendix that market timing ability still implies a positive coefficient on the squared factor.

Stylized market-timing models confine the fund to a single risky-asset portfolio and cash. This makes sense from the perspective of the Capital Asset Pricing Model (CAPM, Sharpe, 1964). Under that model's assumptions there is two-fund separation and all investors hold the market portfolio and cash. But two-fund separation is generally limited to single-factor term structure models, and there is no central role for a "market portfolio" of bonds in most fixed income models. In practice, however, bond funds often manage to a "benchmark" portfolio that defines the peer group or investment style. We use style-specific benchmarks to replace the market portfolio, and funds are assumed to time the factors by anticipating their impact on the benchmark returns.

2.3 Nonlinearity in Benchmark Assets

We model nonlinearity in the relation between the benchmark asset returns and the common factors with a nonlinear regression:

$$r_{Bt} = a_B + b_B(f_t) + u_{Bt} \quad (2)$$

where $b_B(f)$ is a nonlinear function of the factor changes. Appendix A.1 derives the generalization of the market-timing regression that incorporates the nonlinear benchmark:

$$r_{pt} = a_p + b_p [b_B(f_t)] + \lambda_p [b_B(f_t)]^2 + u_{pt} \quad (3)$$

The intuition of Equation (3) is that with no timing, the nonlinearity of the benchmark return determines the nonlinearity of the fund's return. If $\lambda_p=0$ the nonlinearity of the fund's return simply scales that of the benchmark through the second term of the regression. A successful timer's return is convex relative to the benchmark, and thus $\lambda_p > 0$. We combine equations (2) and (3), and estimate the model by the Generalized Method of Moments (Hansen, 1982).

One of the forms for $b_B(f)$ that we consider is a quadratic function, which has an interesting interpretation in terms of systematic coskewness. Asset-pricing models featuring systematic coskewness are developed, for example, by Kraus and Litzenberger (1976). Equation (1) is, in fact, equivalent to the quadratic "characteristic line" used by Kraus and Litzenberger. Under their interpretation the coefficient on the squared factor changes does not measure market timing, but measures the systematic coskewness risk. Thus, a fund's return can bear a convex relation to a factor because it holds assets with coskewness risk. Equations (2) and (3) allow the benchmark to have coskewness risk.

2.4 Interim Trading

Interim trading means that fund managers trade more frequently than the fund's returns are measured. This can lead to incorrect inferences about market timing ability, as discussed by Bhattacharya and Pfleiderer (1983) and shown empirically by Jiang, Yao and Yu (2007) and also to incorrect inferences about total performance, as shown by Ferson and Khang (2002). Ferson, Henry and Kisgen (2006) propose a solution using a continuous-time asset pricing model. The time-aggregated model prices all portfolio strategies that may trade within the period as nonanticipating functions of the state variables in the model. Thus, a manager with no ability will not record abnormal performance. If the manager wastes resources by interim trading that generates trading costs, the portfolio return will be low and this should be detected as negative performance. If the continuous-time model can price derivatives by replication with dynamic strategies, the use of derivatives is also covered by this approach.

Ferson, Henry and Kisgen show that the time-aggregated stochastic discount factor (SDF) from a broad class of popular term structure models is approximately:⁵:

$${}_{t-1}m_t = \exp(a - A_t^r + b'A_t^x + c'[x(t) - x(t-1)]). \quad (4)$$

In Equation (4) $x(t)$ is the vector of state variables in the model at time t and $x(t) - x(t-1) = f_t$ are the factor changes. The terms $A_t^x = \sum_{i=1, \dots, 1/\Delta} x(t-1 + (i-1)\Delta)\Delta$ approximate the integrated levels of the state variables over the period from $t-1$ to t . The monthly measurement period is divided into $(1/\Delta)$ intervals of length $\Delta = \text{one trading day}$. A_t^r is the time-averaged level of the short-term interest rate. The empirical "factors" in the SDF thus include the usual discrete monthly changes in the state variables f_t , and also include their time averaged levels and the time-averaged short term interest rate.

⁵ A stochastic discount factor is a random variable, ${}_{t-1}m_t$, that "prices" assets through the equation $E_{t-1}\{{}_{t-1}m_t(1+R_t)\} = 1$.

With the approximation $e^y \approx 1+y$, which is accurate for numerically small y , the SDF is linear in the expanded set of empirical factors. Since a linear SDF is equivalent to a beta pricing model, this motivates including the time-averaged variables $\{A_t^x, A_t^r\}$ as additional term in the regression to control for interim trading.

2.5 Public Information

Conditional timing models control for public information effects by allowing funds' betas to vary over time with public information. Ferson and Schadt (1996) and Becker, et al. (1999) find that conditional timing models for equity funds are better specified than models that do not control for public information. In particular, Ferson and Schadt (1996) propose a conditional version of the market timing model of Treynor and Mazuy (1966):⁶

$$r_{pt} = a_p + b_p f_t + C_p'(Z_{t-1} f_t) + \Lambda_p f_t^2 + u_t \quad (5)$$

The interaction term $C_p'(Z_{t-1} f_t)$ controls for nonlinearity due to the public information, Z_{t-1} . We include similar interaction terms in this paper to control for public information effects.

2.6 Stale Prices

Thin or nonsynchronous trading in a portfolio biases estimates of the portfolio beta and a similar effect occurs when the measured value of a fund reflects stale prices, possibly due to illiquid assets (e.g. Getmansky, Lo and Makarov, 2004). If the extent of stale pricing is related to a common factor we call it *systematic stale pricing*. To address systematic stale pricing we use a simple model generalizing Getmansky, Lo and Makarov. Let r_{pt} be the true return on a fund's assets. The true

⁶ Ferson and Schadt (1996) also derive a conditional version of the market timing model of Merton and Henriksson (1981), which views successful market timing as analogous to producing cheap call options. This model is considerably more complex than the conditional Treynor-Mazuy model, but they find that it produces similar results.

return would be the observed return if no prices were stale. We assume r_{pt} is independent over time with mean μ . The measured return on the fund, r_{pt}^* , is given by:

$$r_{pt}^* = \theta_t r_{pt-1} + (1-\theta_t) r_{pt} \quad (6)$$

where the coefficient $\theta_t \in [0,1]$ measures the extent of stale pricing at time t . Getmansky, Lo and Makarov allow $K > 1$ lagged returns and K different θ coefficients in Equation (6) to capture multiperiod smoothing in hedge fund returns, but they assume that the smoothing coefficients are constant over time. Our model allows time-varying smoothing coefficients in order to capture systematic stale pricing, but we restrict to a single lag, $K=1$. With K lags, the measured returns should have a K -th order moving average structure. In our bond fund portfolio returns we find significant first order autocorrelations, but the second order autocorrelations are insignificant, suggesting a first-order moving average structure.

Assume that the factor change f_t is normal with mean μ_f and variance σ_f^2 , independent and is identically distributed over time and is measured without error. To model systematic stale pricing consider a regression of θ_t on the absolute factor changes, $g_t = |f_t| - E(|f_t|)$:

$$\theta_t = \delta_0 + \delta_1 g_t + \varepsilon_t \quad (7)$$

where we assume that ε_t is independent of the other variables in the model. A relation between stale pricing and absolute factor changes can be motivated from the results of Karpoff (1987), Fleming and Remolona (1999), Brandt and Kavajecz (2004) and Xing, Zhang and Zhou (2007). We are interested in moments of the true return like $\text{Cov}(r_p, f_t^2)$, which measures the timing ability. Using the facts that $\text{Cov}(r_p, g_t^2) = \text{Cov}(r_p, f_t^2) - 2E(|f|)\text{Cov}(r_p, g_t)$ and Stein's (1973) Lemma, which under normality implies $\text{Cov}(r_p, |f_t| f_t^2) = \text{Cov}(r_p, |f_t|^3) = 3\sigma_f^2 \text{Cov}(r_p, f)$, simple calculations relate the moments of the observed

variables to the moments of the unobserved variables as follows:

$$E(r_p^*) = \mu - \delta_1 \text{Cov}(r_p, g) \quad (8a)$$

$$\text{Cov}(r_{pt}^*, f_t) = \text{Cov}(r_p, f)(1 - \delta_0) - \delta_1 E(r_p g f) \quad (8b)$$

$$\text{Cov}(r_{pt}^*, f_{t-1}) = \text{Cov}(r_p, f) \delta_0 \quad (8c)$$

$$\text{Cov}(r_{pt}^*, f_t^2) = \text{Cov}(r_p, f^2)(1 - \delta_0) + \delta_1 \sigma_f^2 \text{Cov}(r_p, g) - \delta_1 [3\sigma_f^2 \text{Cov}(r_p, f) - E\{f\} \text{Cov}(r_p, f^2)] \quad (8d)$$

$$\text{Cov}(r_{pt}^*, f_{t-1}^2) = \text{Cov}(r_p, f^2) \delta_0 \quad (8e)$$

$$\text{Cov}(r_{pt}^*, g_t) = \text{Cov}(r_p, g)[1 - \delta_0 + 2\delta_1 E(f)] - \delta_1 \text{Cov}(r_p, f^2) \quad (8f)$$

$$\text{Cov}(r_{pt}^*, g_{t-1}) = \text{Cov}(r_p, g) \delta_0 \quad (8g)$$

Equation (8b) captures the bias in the measured covariance with the factor. Equation (8d) shows the measured covariance with the squared factor return, which is a biased estimator of market timing.

Equations (8d) and (8e) reveal how to control for a biased timing coefficient due to systematically stale prices. The sum of the covariances of the measured return with the squared factor changes and the lagged squared factor changes delivers the correct timing coefficient when $\delta_1 = 0$. This is similar to the bias correction for betas in the model of Scholes and Williams (1977). When δ_1 is not zero the sum must be adjusted by subtracting a stale pricing adjustment term. All of the expressions in the system (8) are valid if we replace f_t by $b_B(f_t)$ and modify the definition of g_t accordingly.

2.7 Combining the Effects

The general form of the model is a system including Equations (2), (8) and (9):

$$r_{pt}^* = a + \beta' X_t + \lambda_p \text{Sp} [b_B(f_t)^2 + b_B(f_{t-1})^2] + u_{pt} \quad (9)$$

where the Sp term uses the moments identified in (8) to adjust the timing coefficient for systematic stale pricing:

$$Sp = \text{Cov}(r_p, b_B(f)^2) / \{\text{Cov}(r_p, b_B(f)^2) + \delta_1 \sigma_f^2 \text{Cov}(r_p, g) - \delta_1 [3\sigma_f^2 \text{Cov}(r_p, b_B(f)) - E(|f|) \text{Cov}(r_p, b_B(f)^2)]\} \quad (10)$$

In this model, Λ_p is proportional to $\text{Cov}(r_p, b_B(f)^2)$ and thus measures the timing ability after adjusting for all of the biases. X_t is a vector of control variables observed at t or before. The nonlinear function $b_B(f_t)$ is included in X_t to control for nonlinearity in the underlying benchmark assets and the lagged $b_B(f_{t-1})$ is also included. The time-averaged factor and short-term interest rate, $\{A^f, A^r\}$, are included in X_t to control for interim trading. The products of $b_B(f_t)$ and the lagged state variables (levels of the factor) control for public information effects.

The combined model generalizes the classical market timing regression (1) in three essential respects. The first is the additional control variables X_t . The second is the introduction of the nonlinear functions $b_B(f)$ and the third is the additional moment conditions and lagged terms to control for systematic stale pricing. We evaluate empirically the impact of these generalizations below.⁷

3. The Data

We first describe our sample of bond funds. We then describe the interest rate and other economic data that we use to construct the factors relative to which we study timing ability. Finally, we describe the funds' style-related benchmark returns.

⁷ Adding moments for $E(|f|)$, σ_f^2 , $E(f)$ and $E(r_p f g)$ we have ten moments. The ten parameters are: $\{\mu, \bar{\delta}_0, \bar{\delta}_1, \text{Cov}(r_p, f), \text{Cov}(r_p, f^2), \text{Cov}(r_p, g), E(r_p f g), \sigma_f^2, E(f)$ and $E(|f|)\}$ and the system is exactly identified. Equation (8e) is not used. The regression errors in (2) and (9) are orthogonal to a constant and the right hand side variables in the regressions.

3.1 Bond Funds

The mutual fund data are from the Center for Research in Security Prices (CRSP) mutual fund data base, and include returns for the period from January of 1962 through March of 2007. We select open-end funds whose stated objectives indicate that they are bond funds.⁸ We exclude money market funds and municipal securities funds. We subject the fund data to a number of screens as described in Appendix A.3. We group the funds into equally-weighted portfolios according to eight mutually exclusive investment styles: Index, Global, Short-term, Government, Mortgage, Corporate, High Yield and Other.⁹

We also form subgroups according to various fund characteristics, measured at the end of the previous year. Since the characteristics are likely to be associated with fund style, we form characteristic groups within each of the style classifications. The characteristics include fund age, total net assets, percentage cash holdings, percentage of holdings in options, reported income yield, turnover, load charges, expense ratios, the average maturity of the funds' holdings, and the lagged return for the previous year. The Appendix A.4 provides the details.

Summary statistics for the broad style-grouped funds' returns are reported in Panel A of

⁸ Prior to 1990 we consider funds whose POLICY code is B&P, Bonds, Flex, GS or I-S or whose OBJ codes are I, I-S, I-G-S, I-S-G, S, S-G-I or S-I. We screen out funds during this period that have holdings in bonds plus cash less than 70% at the end of the previous year. In 1990 and 1991 only the three digit OBJ codes are available. We take funds whose OBJ is CBD, CHY, GOV, MTG or IFL. If the OBJ code is other than GOV, we delete those funds with holdings in bonds plus cash totalling less than 70%. After 1991 we select funds whose OBJ is CBD, CHY, GOV, MTG, or IFL or whose ICDI_OBJ is BQ, BY, GM or GS, or whose SI_OBJ is BGG, BGN, BGS, CGN, CHQ, CHY, CIM, CMQ, CPR, CSI, CSM, GBS, GGN, GIM, GMA, GMB, GSM or IMX. From this group we delete fund years for which the POLICY code is CS or the OBJ code is I-G.

⁹ Global funds are coded SI_OBJ=BGG or BGN. Short-term funds are coded SI_OBJ=CSM, CPR, BGS, GMA, GBS or GSM. Government funds are coded OBJ=GOV, POLICY=GS, ICDI_OBJ=GS, or SI_OBJ=GIM or GGN. Mortgage funds are coded ICDI_OBJ=GM, OBJ=MTG or SI_OBJ=GMB. Corporate funds are coded as OBJ=CBD, ICDI_OBJ=BQ, POLICY=B&P or SI_OBJ=CHQ, CIM, CGN or CMQ. High Yield funds are coded as ICDI_OBJ=BY, SI_OBJ=CHY or OBJ=CHY or OBJ=I-G and Policy=Bonds. Index funds are identified by searching for the word "index" in the fund name. Other funds are defined as funds that we classify as bond funds (see the previous footnote), but which meet none of the above criteria.

Table 1. The mean returns are between 0.37% and 0.74% per month. The standard deviations of return range between 0.46% and 1.85% per month. The first-order autocorrelations range from 14% for Index funds to 30% for Short Term funds. The minimum return across all of the style groups in any month is -7.3%, suffered in October of 1979 by the Corporate bond funds. The maximum return is almost 11%, also earned by the Corporate funds, in November of 1981.

Table 1 also reports the second order autocorrelations of the fund returns. The stylized stale pricing model assumes that all the assets are priced within two months. This implies that the measured returns have an MA(1) time-series structure, and the second order autocorrelations should be zero. The largest second order autocorrelation in the panel is 14.9%, with an approximate standard error of $1/\sqrt{T} = 1/\sqrt{171} = 7.6\%$. For the portfolio of all funds, where the number of observations is the greatest, the second order autocorrelation is -1%, with an approximate standard error of $1/\sqrt{543} = 4.3\%$. Thus, the autocorrelations provide no strong evidence against the single-lag structure used in the model for systematic stale pricing.¹⁰

3.2 Bond Market Factors

We use daily and weekly data to construct monthly empirical factors. Most of the data are from the Federal Reserve (FRED) and the Center for Research in Security Prices (CRSP) databases. The daily interest rates are from the H.15 release. The factors reflect the term structure of interest rates, credit and liquidity spreads, exchange rates, a mortgage spread and two equity market factors. The Appendix A.5 provides the details.

Table 2 presents summary statistics for the monthly series starting in January of 1962 or later, depending on data availability, and ending in December 2007. Missing values are excluded and the units are percent per year (except for the US dollar index and Equity values, represented as

¹⁰ The relation between the autocorrelations and the lag structure of stale pricing is derived assuming that the “true” returns are not serially correlated. This assumption is likely to be false, especially for the short term bond funds where we observe the largest autocorrelation. We examine the sensitivity of our results for the short term bond funds to this issue below.

the price/dividend ratio). Panel A presents the levels of the variables and panel B presents the monthly first differences. The time-averaged values used as controls for interim trading effects look similar to the levels in Panel A and are not shown.

The average term structure slope was positive, at just over 80 basis points during the sample period. The average credit spread was about one percent, but varied between 32 basis points and about 2.8%. The average mortgage spread over Treasuries was just over 2% and the liquidity spread averaged 0.4% for the period starting in 1971.

In their levels the variables shown in Table 2 are highly persistent time series, as indicated by the first order autocorrelation coefficients. Five of the nine autocorrelations exceed 95%. Moving to first differences, the series look more like innovations.¹¹ We use the first differences of these variables to represent the factor changes in our analysis.

3.3 Style Index Returns

We form style-related benchmark returns for the mutual funds using two alternative methods. The first method assigns a benchmark based on a fund's most recently-declared style. This has the advantage that the benchmark is determined *ex ante* and nothing has to be estimated. It has the disadvantage of relying on the fund's self-declared style. If a fund strategically misrepresents its style or is more accurately represented as a hybrid style, then the benchmark will be inaccurate.

We select seven benchmarks based on funds' declared styles. Global funds are paired with the Lehman Global Bond Index. Short-term bond funds are paired with a portfolio of US

¹¹ Given the relatively high persistence and the fact that some of the factors have been studied before exposes us to the risk of spurious regression compounded with data mining, as studied by Ferson, Sarkissian and Simin (2003). However, Ferson, Sarkissian and Simin (2008) find that biases from these effects are largely confined to the coefficients on the persistent regressors, while the coefficients on variables with low persistence are well behaved. Thus, the slopes on our persistent control variables and thus the regression intercepts may be biased, but the market timing coefficients should be well behaved.

Treasury bond returns with less than or equal to 48 months to maturity. This best matches the reported maturities of their holdings. Mortgage funds are paired with the Lehman US Mortgage Backed Securities Index. Corporate bond funds are paired with the Lehman US AAA Credit Index, while High-yield funds are paired with the Merrill Lynch High-Yield US Master Index.¹² For two of the styles we use combinations of Lehman bond indexes, weighted in proportion to their contributions to the Lehman US Aggregate bond index. Because Government style funds hold significant amounts of mortgage-backed securities in the latter part of the sample period, we pair the Government bond funds with a combination of the Lehman mortgage backed index and a long-term Treasury bond index.¹³ Finally, we pair the catch-all Other bond funds and the Index funds with a combination of the Treasury bond index, the mortgage backed index and the corporate bond index.¹⁴

Our second method for forming benchmark portfolios follows Sharpe (1992). Historical returns are used to estimate a tracking portfolio of passive asset class returns for each of the 160 fund groups based on style and characteristics. This has the advantages of not relying solely on the fund's self-declared styles and of allowing hybrid styles. It has the disadvantage that the portfolio weights must be estimated, and the estimates will be imprecise in cases with a limited sample of

¹² We splice the Blume, Keim and Patel (1991) low grade bond index returns prior to 1991, with the Merrill Lynch High Yield US Master Index returns after that date.

¹³ We splice the Ibbotson Associates 20 year government bond return series for 1962-1971, with the CRSP greater than 120 month government bond return after 1971.

¹⁴ The aggregate weights are from Lehman (2006), in percent:

Year	Government	Credit	Mortgage
1976	49	42	5
1986	56	19	23
1996	52	17	30
2001	34	27	35

We use the most recent ex ante weight to form the benchmark in a given year. Prior to 1977 we pair the Government funds with the long-term Treasury return index and we pair the Other and Index funds with the Corporate bond index.

returns. If the portfolio weights for a particular period are estimated using any future returns data, there may be a look-ahead bias in the analysis for the future period. The details of this approach are discussed in the Appendix A.6.

Sharpe's approach to estimating the style portfolio weights differs from the more traditional, regression-based approach to forming mimicking portfolios by imposing that the weights cannot be negative. This is realistic for mutual funds, where we are unlikely to find large short positions. Because the passive asset class returns are correlated, the regression approach delivers both extreme and negative weights. For example, with a regression approach we get weights as large as 130% and as small as -130%, and seventeen out of 63 weights are negative.

Panel B of Table 1 presents the Sharpe style-index weights for the broad style-based portfolios. The weights present sensible patterns, suggesting that both the style classification and Sharpe's procedure are reasonably valid. The Global funds load most heavily on global bonds. Short term funds have most of their weight in bonds with less than 48 months to maturity. Mortgage funds have their greatest weights on mortgage backed securities. Corporate funds have more than 70% of their weight in high or low grade corporate bonds. High yield funds have 77% of their weight in low grade corporate bonds. Government funds do load highly on mortgage-backed securities (21%), consistent with the observations of Comer (2006) and Moneta (2008) that their mortgage-related holdings are substantial.

4. Empirical Results

We first examine the empirical relations between the factors and passive investment strategies proxied by the style benchmarks. We also evaluate the effects of the controls for nonlinearities on broad portfolios of the mutual funds. We then apply the models to individual funds.

4.1 Factor Model Regressions

We begin the empirical analysis with regressions of the style benchmark returns on changes and squared changes in the factors, looking for convexity or concavity. This suggests what would happen given a naive application of the timing regression (1) for funds, if funds simply held the benchmark portfolios. If the benchmark returns are nonlinearly related to the factors, it suggests that controls for nonlinearities could be important.

Table 3 summarizes the t-ratios for the regression coefficients on the squared factors. Only t-ratios that exceed 1.6 in absolute value are shown. Panel A of Table 3 shows that out of 72 cases (8 style benchmarks times 9 factors) there are 25 heteroskedasticity-consistent t-ratios with absolute values larger than 1.6. Nine of the absolute t-ratios are above 2.0. Thus, there is significant evidence of nonlinearity in the benchmark returns. Most of the large coefficients are positive, indicating convexity. Using the Sharpe style benchmarks (not shown in the table) the evidence for convexity is even stronger. This implies that we would measure positive timing ability, based on regression (1), if funds simply held the benchmarks. Thus, controlling for nonlinearity is likely to be important for measuring the timing ability of bond funds.

Comparing the funds with the benchmarks, we see the effects of active management. The coefficients for the mutual funds are summarized in Panel B of Table 3. Here we find 11 absolute t-ratios larger than two and 19 in excess of 1.6. This is more than expected by chance; thus, the mutual fund returns are also significantly nonlinearity related to the factors.¹⁵ The coefficients on the squared factors are negative in about half of the cases. Thus, the fund returns appear more concave in relation to the factors than do the benchmark returns. This might suggest poor market timing ability on the part of the mutual funds, but it could also reflect derivatives, interim trading, public information or stale pricing effects that are not found in the benchmarks. Note that marginally significant t-ratios are found for the index funds on three of the nine factors, as shown

¹⁵ Using a simple binomial model assuming independence, the t-ratio associated with finding 11 "rejections," when the probability of observing a rejection is 5%, is $(11/72 - .05)/(.05(.95)/72)^{0.5} = 4.00$.

in the first row of Panel B. As index funds are unlikely to be actively timing, this suggests that much of the nonlinearity is unrelated to timing.

We also examine regressions like Table 3 where the fund style returns net of the benchmark returns are the dependent variables. The evidence of concavity is much stronger in these regressions, as would be expected. We find 27 or 28 absolute t-ratios larger than 1.6 and all but one or two are negative, depending on the type of style benchmark. Thus, simply measuring fund returns net of a benchmark will not control for the nonlinearities. If we naively ran the regression (1) using fund returns net of benchmark returns on the left hand side, we would find strong evidence of negative timing ability.

4.2 Evaluating the Controls

To control for benchmark nonlinearity the function $b_B(f_t)$ is used in Equations (2), (8) and (9). We consider three specifications for $b_B(f)$: Quadratic, exponential and piecewise linear. As described earlier, a quadratic function can be motivated by coskewness. An exponential function can be motivated by a continuous-time model. However, since $e^f \approx 1+f$ when f is numerically small, the linear function well approximates the exponential for small factor changes, and we find no empirically measurable impact of an exponential function, compared with a linear function.

The piecewise linear specification is $b_B(f) = bf + c f^* I(f>0)$, where $I(f>0)$ is an indicator function for a positive change in the factor. A piecewise linear function can be motivated as approximating an option payoff. We use zero as the breakpoint in the piecewise linear function for simplicity and to avoid estimating a breakpoint parameter. This will be useful when we apply the model at the level of the individual funds, where short time series often limit the degrees of freedom.

Our comparisons face a number of challenges related to degrees of freedom. One has to do with multiple factors. In principle funds may try to time multiple factors. A market-timing model with multiple factors would include the cross products as well. In our case with nine factors we

would have 36 cross products plus nine squared factors, and thus timing coefficients on 45 variables. The controls for non-timing-related nonlinearities would proliferate as well, leading to serious problems with overfitting. We would only be able to study a small subset of funds with the longest return histories, leading to a survivor selection bias. We therefore limit the analysis throughout to models that consider one factor at a time. We recognize that, because the variables in our models may be correlated with “missing” factors, it may affect our results, and we caution the reader to interpret our findings accordingly.

We revisit the factor model regressions of Table 3 using the piecewise linear function in place of the quadratic function to capture nonlinearity. The results for the benchmark returns are similar. The coefficient c is often positive, indicating convexity in the benchmarks, and the t -ratio is large in significant fractions of the cases. The coefficients for the fund portfolios are again a mix of positives and negatives, with large t -ratios in a significant fraction of the cases. The funds are typically more concave in the factor changes than are the benchmarks.

Table 4 summarizes the impact of the controls for non-timing-related nonlinearity on the funds grouped into the eight broad style categories, a level of aggregation where we do not expect to find significant timing ability in a well-specified model. The models are estimated one factor at a time. The table summarizes the averages of the timing coefficients¹⁶ taken across the nine market factors and how many of the nine t -ratios are above $+2.0$ or below -2.0 .

In Panel A the benchmark is modelled as a piecewise linear function of the factor changes, while in Panel B the linear function is used. The first column of figures shows the results with no other controls. With controls for the nonlinear benchmark only we find that 25 of the 72 t -ratios are less than -2.0 and ten are greater than 2.0 . The index funds have five of nine t -ratios below -2.0 .

¹⁶ We present the averages across the factors for parsimony. Even with the eight broad fund styles and one factor at a time, there are at least 360 comparisons (8 styles x 9 factors x 5 controls) to models with no controls. We have also examined the effects of the controls on the 160 style and characteristics based fund portfolios and we have cut the data in various ways. These averages present a reasonable summary of the overall impressions from the more detailed analysis, some of which is described below.

When we use a quadratic $b_B(f)$ function (not shown in the table) the results are similar. In Panel B of Table 4 we repeat the analysis without controlling for nonlinear benchmarks. With no controls the results are even more perverse. 34 of the t-ratios for the timing coefficients are below -2.0, and the index funds produce seven of nine t-ratios below -2.0. Thus, controlling for nonlinearity in the benchmarks is important, but not adequate to control for non-timing related nonlinearity in funds' returns.

The columns labeled Public Information, Interim Trading and Stale Pricing introduce these controls alone, while the All Controls column uses all of the controls simultaneously. Under the public information controls the results are similar to the model with the nonlinear benchmark only. The interim trading controls have a larger separate impact on the timing coefficients, and introducing these controls reduces the frequency of large negative t-ratios by about 2/3. The funds' returns appear less concave with the interim trading controls. This suggests that on average funds' interim trading or derivatives positions have the effect of "selling convexity," or inducing concavity in the funds' returns. Purchasing callable bonds, for example, induces concavity because the fund is effectively short the call option. When we control for interim trading effects we remove this concavity from the measure of market timing.

The controls for stale pricing also have the effect of reducing the number of large t-ratios on the timing coefficients; in fact, only four of the 72 are larger than +2.0 and two are smaller than -2.0. However, a comparison of the average coefficient values suggests that this is partly a result of the larger standard errors in the systematic stale pricing model.

In the right hand column of Panel A all of the controls are used together. We find one t-ratio below -2.0 and seven above +2.0. Given that 72 t-ratios are estimated these frequencies are consistent with a well-specified model. The average coefficient values are not extreme, compared with the model with stale pricing only. Overall, the model with all of the controls appears to be a reasonable specification.

The stale pricing model makes the simplifying assumption that the true returns are serially

independent, while the measured returns have serial dependence. The true returns of Short-term bond funds in particular are likely to be serially dependent, and we could mistakenly attribute stale pricing to short-term bond funds. As a check we run the analysis in Table 4 making the extreme assumption that the Short-term bond funds have no stale pricing. With all the controls except stale pricing, the average coefficient for the Short-term bond funds changes from -7.38 to -6.33, and we find no t-ratios for the Short-term bonds above 2.0 and two below -2.0. For our analysis of individual funds below, where Short term bond funds represent about 31% of the cross section, the impact of this misspecification should be smaller.

We repeat the analysis of Table 4, Panel A using a quadratic function to model benchmark nonlinearity in place of the piecewise linear function. All of the impressions are very similar to those in Panel A, so we do not tabulate the results. Table 4 uses the Sharpe style benchmarks. We check the results using the self-declared benchmarks. Assuming the benchmarks are linear in the factors the results are again perverse, suggesting negative timing as in Panel B of Table 4. The results for Panel A are also similar. In an earlier version of the paper the stale pricing model assumed that staleness was related to the value of the factor changes instead of the absolute changes. That model provided less evidence of systematic stale pricing, and the absolute model used here appears to be the better specification.

Table 4 focuses on the effects of the controls at the style group level, but we conduct further experiments that focus on the effects of the controls in the tails of the cross-sectional distributions of the style-and-characteristics-based fund portfolios. We examine the 180 possible cases for each style (9 factors times 20 characteristics-sorted portfolios) and 160 possible cases for each factor. When there are no controls for nonlinearity we find many negative timing coefficients. Introducing the controls for benchmark nonlinearity typically has a substantial impact on the extreme values of the timing coefficients and reduces the incidence of large t-ratios. The controls for interim trading and public information seem to have a smaller impact on the tails.

Note that the controls for stale pricing raise a potential concern. Suppose that the sum

$[b_B(f_t)^2 + b_B(f_{t-1})^2]$ is used as the regressor to identify the timing coefficient while the correct specification uses only $b_B(f_t)^2$. The lagged term appears as a "measurement error." Measurement error is expected to bias the regression coefficient toward zero, which we could misinterpret as an appropriate control. We re-run the analyses of Table 4 replacing the summed term with two separate $b_B(f)^2$ terms, each with its own coefficient, and we estimate the sum of the two coefficients. The results are similar to those recorded in Table 4.

We draw several conclusions from this analysis. First, it is important to control for nonlinearity in funds' benchmarks with respect to the factors, but the other controls are also needed. Second, on average it appears that funds' interim trading or derivatives activities have the effect of "selling convexity," as would be the case when buying callable bonds. Finally, a model that combines the various controls with nonlinear benchmarks produces an overall distribution of the timing coefficients that appears neutral to perhaps slightly positive at the level of fund portfolios.

4.3 Fund level Analysis

We expect to find little timing when funds are grouped into large portfolios, but there could be individual funds with significant timing ability. We estimate the market timing coefficients for each fund with at least 36 monthly returns and study the cross-sectional distributions of the individual timing coefficients. There are at least 1204 eligible funds for each factor. The results using the piecewise linear $b_B(f)$ function, the Sharpe style benchmarks and all of the controls are summarized in Table 5.

The first column of figures in Table 5 lists selected fractiles for the null distributions of estimated timing coefficients. We estimate the null distributions by setting the timing coefficients equal to zero, and simulating funds' returns as the fitted values of regression (9) with this restriction plus the randomly resampled regression residuals. The number of simulated fund returns is equal to the number of actual funds and each simulated fund has the same number of

returns as an actual fund in the sample. Estimating the model on a factor-by-factor basis using the simulated data, we sort the timing coefficients of the funds and determine the critical values at selected fractiles of the null distributions. The fractions in the remaining columns of Table 5 are the fractions of the estimated timing coefficients in the original, unrestricted fund data that exceed the critical values for a given factor. We evaluate the distributions informally.¹⁷

The mutual funds' timing coefficients are mildly skewed toward positive values in the central regions of the distributions. For example, more than 50% of the funds' coefficients exceed the median of the null distribution for six of the nine factors. More than 90% of the funds' coefficients exceed the 10% left-tail critical values for eight of the nine factors. None of these differences is large, however, and the tails conform relatively closely to the null distributions.¹⁸

We repeat the analysis of Table 5 substituting the self-declared style benchmarks and/or the quadratic $b_B(f)$ function. The overall impressions about the cross-sectional distributions of the timing coefficients are similar to Table 5 when the controls for interim trading, public information and systematic stale pricing are in place. The timing coefficients of the individual funds appear slightly more positive than the distribution under the null hypothesis, mainly in the central

¹⁷ To evaluate statistical precision consider a binomial random variable that equals 1.0 with probability p , when a timing coefficient is larger than the critical value for the fractile p . If the correlation of the trials is ρ the variance of the fraction of funds above the critical value is $[p(1-p)\{1/n - (1-1/n)\rho\}]$. The correlation ρ depends on the correlation of the funds' returns. We approximate ρ by estimating the pairwise correlations of all the funds' returns using all pairs with at least 36 months in common. The summary statistics in Table 1 suggest that if two funds' return series are separated by more than a month in calendar time, their average correlation is statistically zero. We therefore scale each contemporaneous correlation by the fraction of the sample where the two fund return series overlap. Here we make the conservative assumption that the correlations at one lag are equal to the contemporaneous correlations, and we set the correlations beyond lag one to zero. The resulting estimate of ρ is 0.083. This implies that the standard deviations of the fractions in Table 5 are approximately 0.14, 0.085, 0.06 and 0.045 respectively, for the 0.50, 0.10, 0.05 and 0.025 fractiles. Of course, the multiple comparisons across factors would further complicate formal statistical inference.

¹⁸ We examine the correlations of the timing coefficients with the ten fund characteristics described earlier. With about 1200 observations, the standard error of these correlations is about 0.03. There are only 22 out of 171 correlations larger than 0.06, which is not statistically significant. The correlations display no obvious patterns across the fund styles or characteristics.

regions of the distributions.

We draw several conclusions about fund-level timing ability. First, when the controls for non-timing-related nonlinearity are in place there is little evidence of significant "negative timing" among bond funds. For some factors more than half of the funds have timing coefficients larger than the median under the null hypothesis of no timing ability, but there are not an unusual number of funds with coefficients in the extreme tails. Overall, the timing ability appears neutral to weakly positive.

4.4 Nonlinearity-adjusted Performance

Most measures of investment performance compare the average return of a managed portfolio over some evaluation period to the average return of a benchmark portfolio that is a feasible unmanaged alternative. To evaluate whether the managed portfolio is attractive relative to the benchmark, the benchmark should be equivalent to the managed portfolio in all expected-return-relevant respects, except that it should not reflect the investment ability of the firm or manager. Aragon and Ferson (2007) call such a portfolio an "*Otherwise Equivalent*" (OE) portfolio. If a fund outperforms its OE portfolio on an after cost basis it may be attractive to investors who would otherwise hold the benchmark.

In order to operationalize the OE portfolio it is necessary to have a model to determine what characteristics of a portfolio should lead to higher or lower expected returns. For example in the CAPM the relevant characteristic is the market beta and the OE portfolio is a combination of the market portfolio and the risk-free asset. The average return in excess of the OE portfolio is Jensen's (1968) alpha.

The task in the present setting is to consider nonlinearity when evaluating performance. The intercept in the Treynor-Mazuy regression (1) has been naively interpreted in some studies as a "timing adjusted" alpha, measuring selectivity. This is only correct in unlikely special cases of stylized "perfect" market timing ability, as discussed by Aragon and Ferson (2007). In general it is

not possible, outside of highly stylized models, to isolate timing from selectivity ability. We use stylized models to estimate timing ability and we estimate total performance – that may include timing and selectivity ability – relative to an OE portfolio that features nonlinearity.

The intercept in Equation (1) does not measure performance relative to an OE portfolio (even if f_t is an excess return) because f_t^2 is not a portfolio's excess return. However, the model can be modified to capture the difference between the return of the fund and that of an OE portfolio, provided that the nonlinearity of the fund can be replicated by trading in market assets (e.g. Glosten and Jagannathan, 1994).

Let r_{h2} be the excess return of the maximum-squared-correlation portfolio for the squared factor changes, f_t^2 . This portfolio is estimated by the regression:

$$f_t^2 = a + \mathbf{H}'\mathbf{r}_t + \mathbf{u}_t \tag{11}$$

where the weights that define the mimicking portfolio r_{h2} are proportional to the regression coefficients, \mathbf{H} . The base assets in \mathbf{r}_t are the seven assets we use to form the Sharpe style benchmarks, excluding the short term Treasury rate, and the returns are in excess of the Treasury rate. The expected value of the excess return, $E(r_{h2})$ is the risk premium associated with the squared factor.¹⁹ Our goal is to form an OE portfolio for each fund that has the same loadings on its style benchmark and r_{h2} as does the fund. This makes the simplifying assumption that style benchmark exposure and convexity are the return-relevant characteristics. The OE portfolio is formed using the following regression:

¹⁹ An alternative approach is to use cross-sectional regressions of returns on betas to estimate mimicking portfolios. While we have a large cross section of mutual funds, using the funds with this approach would contaminate the risk premium estimates with the abnormal returns due to manager ability, if any. We have a small cross section of passive benchmark assets. We therefore use time-series as opposed to cross-sectional regressions to estimate the mimicking portfolios. (See Balduzzi and Robotti, 2008 for comparisons of the two methods.)

$$\mathbf{r}_{pt} = \mathbf{a}_p + \mathbf{b}_p \mathbf{r}_{Bt} + \mathbf{c}_p \mathbf{r}_{Bt-1} + \mathbf{d}_p \mathbf{r}_{h2t} + \mathbf{e}_p \mathbf{r}_{h2t-1} + \mathbf{v}_t \quad (12)$$

The loading on the benchmark, \mathbf{r}_B , is estimated following Scholes and Williams (1977) as $\beta_p = \mathbf{b}_p + \mathbf{c}_p$, and the loading on the hedge portfolio is $\Omega_p = \mathbf{d}_p + \mathbf{e}_p$. The OE portfolio return for fund p is $\beta_p \mathbf{r}_B + \Omega_p \mathbf{r}_{h2}$ and the alpha, measuring the nonlinearity-adjusted performance, is $E(\mathbf{r}_{pt}) - \beta_p E(\mathbf{r}_B) - \Omega_p E(\mathbf{r}_{h2})$.

Table 6 presents the analysis of nonlinearity-adjusted performance. The second and third columns of Panel A summarize the mean excess returns over a short term Treasury and the mean returns over the style benchmarks at selected fractiles of their distributions across the funds. The average excess returns are stated in percent per month. The median fund excess return is 0.46% per month and the median return net of benchmark is -0.07%, consistent with previous studies that find bond funds return less than their benchmarks on average. The distribution is skewed to the left, with 5% of the funds below -0.28% and 5% above 0.10%.

Panel A of Table 6 also reports the fractions of funds with alphas larger than the critical values at the indicated fractiles from the null distribution in which the true alphas are zero. These critical values are similar in construction to the critical values in Table 5. We subtract the fitted alphas from each fund's fitted return and resample the residuals in the simulation. The third column shows the values of the estimated alphas that define the various fractiles of the distribution across funds under the null hypothesis that the true alphas are zero (critical alphas). The critical alphas are very similar across the factors. For simplicity we report here the averaged critical values taken across the nine factors, although the analysis in the remaining columns is based on the factor-specific critical values.

Many funds have negative alphas. Depending on the factor, 78-86% of the funds have alphas below the median value of zero under the null. Between 10-18% of the funds have alphas below -0.156% per month, which is at the left 2.5% tail of the null distribution. The right tails conform

closely to the null distributions. Thus, the table provides evidence of negative performance on an after-cost basis and little evidence that any funds have significant positive performance.

Panel B of Table 6 digs more deeply into the total performance. We first summarize the funds' loadings on the style benchmarks. Funds' loadings vary widely in the cross section, with the 10% tails spanning values between 0.44 and 1.38. This shows that the returns net of benchmark are crude performance measures. They assume that all of the loadings equal 1.0.

Below the benchmark loadings we report the average excess returns of the r_{h2} portfolios, denoted as "hedge premiums." The average hedge premiums vary from -2.93% per month for equity volatility to 3.6% for equity values, but funds' loadings on the two equity-related r_{h2} 's are small and the range is narrow, with 80% of the funds' loadings between -0.12 and +0.09. The rest of the hedge premiums are an order of magnitude smaller than the equity-related premiums, but the range of funds' loadings on the portfolios is typically larger.

If market timing or convexity is valuable, we expect negative return premiums for portfolios that are positively correlated with squared factor changes. Maximum R-squared portfolios may have negative or positive correlation, so the average hedge premiums multiplied by the sign of the correlation of r_{h2} with the squared factor changes should be negative. The correlation is shown in the last line of Table 6. The product is negative for each of the nine factors, excepting the interest rate curvature factor.

An example of the effect of the nonlinearity adjustments on funds' performance is the short rate. A positive loading on the r_{h2} corresponds to concavity with respect to interest rate changes, for which characteristic a higher return is required. This contributes $(0.415)(0.134) = 0.06\%$ per month to the required return of a fund in the upper ten percentile of r_{h2} loadings. Thus, the alpha of such a fund is reduced by 0.06% per month, adjusting for nonlinearity. For a fund at the lower ten percentile of loadings we add $(0.415)(-0.224) = -0.09\%$ per month to the required return, increasing the alpha by 0.09% per month to reflect the value of convexity. These figures are of similar in magnitudes to the -0.067% per month by which the median fund underperforms its benchmark after

costs, but less than half of the 0.2% underperformance at the ten percent left tail of fund returns net of benchmark.

4.5 Before-Cost Performance

Since the OE portfolio pays no trading costs while the bond funds do, the alphas reflect a mixed message. This is consistent with the approach in much of the performance measurement literature. If investors could replicate the OE portfolios at negligible cost, then negative alphas funds are not attractive for investors who would otherwise hold the benchmark. But the costs to replicate the benchmarks may not be negligible. In this section we repeat the analysis with the funds' returns measured on a before-cost basis. If a fund returns more than the OE portfolio on a before cost basis we refer to this, following Aragon and Ferson (2007), as "investment ability." The idea is that investment ability on a before-cost basis may be absorbed by fees and other costs. To obtain the before-cost returns we add back the average expense ratio plus a measure of trading costs. The trading costs for each fund are a round-trip trading cost estimate based on the fund's style multiplied by the average reported turnover of the fund.²⁰

Table 7 presents the results. The median fund return is about 0.11% higher than in Table 6, reflecting a total cost of about 1.3% per year. The median return net of the style benchmark is 0.03% per month, which suggests some investment ability for the median fund. The left-tailed skewness

²⁰ The round trip transaction cost figures are as follows. For Global bond funds we use 31 basis points, based on figures in Biais and Declerk (2006). This is an average of twice the half-spreads from their Table 5 plus the information content from their Table 10, weighted in proportion to the numbers of Eurobonds and Sterling bonds in their sample. For corporate bonds we use 48 basis points and for high yield bonds we use 75 basis points. These figures are averages from Edwards, Harris and Piwowar (2007), Bessembinder, Maxwell and Venkataraman (2006) and Hotchkiss et al. (2009) for intermediate trade sizes. For Government funds we use 12.5 basis points, following Ferson, Henry and Kisgen (2006). For Mortgage funds and Short-term funds we use 20 basis points. For Index funds and Other bond funds, we use the average of these figures, or 34 basis points. We checked these figures with a trader at Smith-Breedon, John Sprow, who suggested the figures for Mortgage and Short-term bond funds and confirmed that the other numbers seemed reasonable for trade sizes typical of mutual funds under average market conditions.

that we saw in Table 6 is no longer observed, suggesting that was driven by skewness in the costs.

The distributions of the before-cost alphas are markedly different from the alphas in Table 6. About 75% of the funds generate before-cost alphas above the median of the distribution generated under the null hypothesis that the alphas are all zero. Between 10-15% of the funds have alphas above 0.149% per month, which is at the upper 2.5% tail of the null distribution. The alphas in the left tails conform more closely to the null distributions. This makes sense in that a fund with significant "negative ability" presents a potentially valuable signal to a competitor who can take the opposite position, while a fund with significantly higher costs presents no such signal.

Comparing tables 6 and 7, the story is similar to what the literature finds for equity funds. After costs a significant number of funds have poor performance, and there is no significant positive performance. Before costs, just the opposite is true.²¹ A significant number of funds have positive investment ability, and the performance is consistent with the null of no ability in the left tails. Overall, the evidence is consistent with the view that bond fund managers have investment ability but investment fees and trading costs absorb the superior returns (and then some), leaving no abnormal positive performance for fund investors.

5. Concluding Remarks

Models of market timing using returns measure convexity in the relation between the fund's return and the common factors. However, convexity or concavity is likely to arise for reasons unrelated to timing ability. We adapt classical market timing models to bond funds by controlling for other sources of nonlinearity, such as the use of dynamic trading strategies or derivatives, portfolio strategies that respond to publicly available information, nonlinearity in the benchmark assets and systematically stale prices. These issues should also arise in equity funds and in hedge funds, so future research should examine similar controls in the context of other types of funds.

²¹ Gutierrez, Maxwell and Xu (2008) recently confirm negative after-cost and positive before-cost performance in a sample of Corporate bond funds using different methods.

We find that controlling for non-timing-related nonlinearity matters, and naive applications of market timing models without these controls would be misleading. Simple returns net of style benchmarks are not likely to be reliable performance measures because funds' loadings on the benchmarks differ substantially from 1.0. Bond funds' returns are typically more concave, in relation to a broad set of bond market factors, than are unmanaged benchmarks. Thus, without controls for non-timing-related nonlinearity, funds would appear to have poor (i.e., negative) market timing ability. When we introduce the controls the distribution of the timing coefficients appears neutral at the fund style-group portfolio level and neutral to slightly positive in the cross-sectional distribution of individual mutual funds.

The impact of nonlinearities on bond funds' total performance is typically small. For example, the value of short-rate convexity could account for less than half of the underperformance at the 10% tail of fund performance. Significant fractions of funds' nonlinearity-adjusted alphas are negative. There is no evidence of positive performance after costs. When we add expense ratios and estimates of trading costs based on fund turnover back to the funds' returns, the opposite is found. About 75% of the funds earn positive before-cost alphas, and the distribution of fund performance is significantly better than would be expected under the null hypothesis that all the alphas are zero. There is no evidence of significant negative performance before costs, suggesting that the left tail skewness in the after cost measures reflects skewness in the costs.

Appendix

A.1 Market Timing Models

Assume that the fund manager combines a benchmark portfolio with return R_B and a short-term Treasury security or "cash" with known return R_F . The portfolio weight on R_B is $x(s)$, where s is the private timing signal. The managed portfolio return is $R_p = x(s)R_B + [1-x(s)]R_F$. The signal is observed and the weight is set at time $t-1$, the returns are realized at time t , and we suppress the time subscripts when not needed for clarity. In the simplest example the factor changes and the

benchmark's excess returns are related by a linear regression:

$$\mathbf{r}_{Bt} = \mu_B + \mathbf{b}_B \mathbf{f}_t + \mathbf{u}_{Bt} \quad (\text{A.1})$$

where $\mathbf{r}_B = \mathbf{R}_B - \mathbf{R}_F$ is the excess return, $\mu_B = \mathbf{E}(\mathbf{r}_B)$, the factor changes are normalized to have mean zero and \mathbf{u}_{Bt} is independent of \mathbf{f}_t . Assume that the signal $\mathbf{s} = \mathbf{f} + \mathbf{v}$, where \mathbf{v} is an independent, mean zero noise term with variance, σ_v^2 . Assume that the random variables $(\mathbf{r}, \mathbf{f}, \mathbf{s})$ are jointly normal and let $\sigma_f^2 = \text{Var}(\mathbf{f})$. The manager is assumed to maximize the expected value of an increasing, concave expected utility function, $\mathbf{E}\{U(\mathbf{r}_p)|\mathbf{s}\}$. The optimal portfolio weight of the market timer is:²²

$$\mathbf{x}(\mathbf{s}) = \lambda \mathbf{E}(\mathbf{r}_B|\mathbf{s})/\sigma_B^2, \quad (\text{A.2})$$

where $\lambda > 0$ is the Rubinstein (1973) measure of risk tolerance, which is assumed to be a fixed parameter, and $\sigma_B^2 = \text{Var}(\mathbf{r}_B|\mathbf{s})$, which is a fixed parameter under normality.

The implied regression for the managed portfolio's excess return follows from the optimal timing weight $\mathbf{x}(\mathbf{s})$ and the regression (A.1). We have $\mathbf{r}_p = \mathbf{x}(\mathbf{s}) \mathbf{r}_B$, then substituting from (A.1) and (A.2) and using $\mathbf{E}(\mathbf{r}_B|\mathbf{s}) = \mu_B + \mathbf{b}_B [\sigma_f^2/(\sigma_f^2 + \sigma_v^2)](\mathbf{f} + \mathbf{v})$, we obtain equation (1), where $\mathbf{a}_p = \lambda \mu_B^2/\sigma_B^2$, $\mathbf{b}_p = (\lambda \mu_B \mathbf{b}_B/\sigma_B^2) [1 + \sigma_f^2/(\sigma_f^2 + \sigma_v^2)]$ and $\Lambda_p = (\lambda/\sigma_B^2) \mathbf{b}_B^2 [\sigma_f^2/(\sigma_f^2 + \sigma_v^2)]$. The error term \mathbf{u}_{pt} in the regression is a linear function of \mathbf{u}_B , \mathbf{v} , $\mathbf{v}\mathbf{u}_B$, $\mathbf{f}\mathbf{u}_B$ and $\mathbf{v}\mathbf{f}$. The assumptions of the model imply that the regression error is well specified, with $\mathbf{E}(\mathbf{u}_p \mathbf{f}) = \mathbf{0} = \mathbf{E}(\mathbf{u}_p) = \mathbf{E}(\mathbf{u}_p \mathbf{f}^2)$.

The model shows that timing ability implies convexity between the fund's return and the systematic factor changes, independent of the direction of the relation between the factor changes

²² The first order condition for the maximization implies: $\mathbf{E}\{U'(\cdot)|\mathbf{s}\} \mathbf{E}\{\mathbf{r}_B|\mathbf{s}\} + \text{Cov}(U'(\cdot), \mathbf{r}_B|\mathbf{s}) = \mathbf{0}$, where $U'(\cdot)$ is the derivative of the utility function. Using Stein's (1973) lemma, write the conditional covariance as: $\text{Cov}(U'(\cdot), \mathbf{r}_B|\mathbf{s}) = \mathbf{E}\{U''(\cdot)|\mathbf{s}\} \mathbf{x}(\mathbf{s}) \text{Var}(\mathbf{r}_B|\mathbf{s})$. Solving for $\mathbf{x}(\mathbf{s})$ gives the result, with $\lambda = -\mathbf{E}\{U'(\cdot)|\mathbf{s}\}/\mathbf{E}\{U''(\cdot)|\mathbf{s}\} > 0$.

and the benchmark return. That is, since $\lambda > 0$ the coefficient $\Lambda_p \geq 0$, independent of the sign of b_B . If the manager does not receive an informative signal then $\Lambda_p = 0$ because $E(r_B|s)$ and $x(s)$ are constants.

A.2 Nonlinearity

The manager's market-timing signal is now assumed to be $s = b_B(f) + v$, where v is normal independent noise with variance, σ_v^2 . This captures the idea that the manager focuses on the benchmark return implications of information about the factor changes.

The optimal weight function in (A.2) obtains with:

$$E(r_B|s) = \mu_B [\sigma_v^2 / (\sigma_f^{*2} + \sigma_v^2)] + [\sigma_f^{*2} / (\sigma_f^{*2} + \sigma_v^2)] [a_B + b_B(f) + v], \text{ where } \sigma_f^{*2} = \text{Var}(b_B(f)).$$

Substituting as before we derive the nonlinear regression for the portfolio return:

$$r_{pt} = a_p + b_p [b_B(f_t)] + \Lambda_p [b_B(f_t)^2] + u_t, \quad (\text{A.3})$$

with:

$$a_p = \lambda a_B (\mu_B \sigma_v^2 + a_B \sigma_f^{*2}) / \sigma_B^2 (\sigma_f^{*2} + \sigma_v^2),$$

$$b_p = \lambda (\mu_B \sigma_v^2 + 2a_B \sigma_f^{*2}) / \sigma_B^2 (\sigma_f^{*2} + \sigma_v^2),$$

and

$$\Lambda_p = (\lambda / \sigma_B^2) [\sigma_f^{*2} / (\sigma_f^{*2} + \sigma_v^2)].$$

A.3 Screening the Fund Sample

There are a total of 40,390 fund-year records in our initial sample. In order to address back-fill bias we remove the first year of returns for new funds, and any returns prior to the year of fund organization, a total of 2,625 records. Data may be reported prior to the year of fund organization, for example, if a fund is incubated before it is made publicly available (see Elton, Gruber and Blake (2001) and Evans, 2006). We delete 91 fund years for which the POLICY code is CS or the OBJ code is I-G. Extremely small funds are more likely to be subject to back-fill bias. We delete cases where the reported total net assets of the fund is less than \$5 million. This removes 5,698 records. We delete all cases where the reported equity holdings at the end of the previous year

exceeds 10%. This removes 1,017 records. We identify cases where funds report multiple share classes. Multiple classes are identified when two or more ICDI codes for the same year have a common fund name and a different share class code. We retain the share class with the largest Total Net Assets and delete the other share classes. This removes another 10,723 records. After these screens we are left with 20,236 fund-years. The number of funds with some monthly return data in a given year is four at the beginning of 1962, rises to 14 at the beginning of 1973, to 564 by 1993 and is 1,054 at the beginning of 2007.

A.4 Funds Grouped by Characteristics

The fund characteristics include age, total net assets, percentage cash holdings, percentage of holdings in options, reported income yield, turnover, load charges, expense ratios, the average maturity of the funds' holdings, and the lagged return for the previous year. Each year we sort the funds of a given style with nonmissing characteristic data from high to low on the basis of the previous year's value of a characteristic and break them into thirds. We form equally weighted portfolio returns from the funds in the high group and the low group for each month of the next year.

We examine various summary statistics of the fund characteristics grouped by style. The reported turnover is used when we compute our proxy for before-cost returns, and is interesting as an indicator of active management. Not surprisingly the low-turnover Index funds have the lowest turnover, but this increases from 8 to 59% per year over the 1992 to 2006 period. The high-turnover Corporate funds have the highest turnover (increasing from 117 to 202% between 1992 and 2006). Turnover increases over the sample period for all styles except the High-yield funds. If turnover is largely driven by the need to replace maturing bonds one might expect to find that the Short term funds have high turnover. However, the highest-turnover third of each fund style reports more turnover than the lowest third of the Short-term funds, and there is more cross-sectional variation of turnover within a style group than across the fund styles. Turnover does not seem to be

primarily driven by fund style. Thus, the turnover figures suggest active management of the bond funds.

A.5 Bond Market Factor Data

Three factors represent the term structure of Treasury yields: A short-term interest rate, a measure of the term slope and a measure of the curvature of the yield curve. The short-term interest rate is the three-month Treasury rate. The slope of the term structure is the ten-year yield less the one-year yield. The curvature measure is: $y_3 - (y_7 + 2y_1)/3$, where y_j is the j -year fixed-maturity yield.

Since our funds hold corporate bonds subject to default risk and mortgage backed securities subject to prepayment risks, we construct associated factors. Our credit spread series is the yield of Baa corporate bonds minus Aaa bonds, from the FRED. These series are measured as the weekly averages of daily yields. We use the averages of the weeks in the month for our time-averaged version of the spread. For the discrete changes in the spread we use the first differences of the last weekly values for the adjacent months. The first difference series may not be as clean as with daily data, but we are limited by the data available to us. Our mortgage spread is the difference between the average contract rate on new conventional mortgages, also available weekly from the FRED, and the yield on a three-year, fixed maturity Treasury bond. Here we use daily data on the Treasury bond and weekly data on the mortgage yield to construct the time averages and discrete changes.

For market timing we are interested in market-wide fluctuations in liquidity. Our measure follows Gatev and Strahan (2006), who advocate a spread of commercial paper over Treasury yields as a measure of short term liquidity in the corporate credit markets. (See also, Bernanke (1983) who interprets the spread as a monetary policy factor.) We use the yield difference between three-month nonfinancial corporate commercial paper rates and the three month Treasury yield. The commercial paper rates are measured weekly, as the averages over business days.

Some of the funds in our sample are global bond funds, so we include a factor for currency

risks. Our measure is the value of the US dollar, relative to a trade-weighted average of major trading partners, from the FRED. This index is measured weekly, as the averages of daily figures, and we treat it the same way we treat the other weekly data.

Corporate bond funds, and high-yield funds in particular, may be exposed to equity-related factors. We therefore include two equity market factors in our analysis. We measure equity volatility with the VIX-OEX index implied volatility. This series is available starting in January of 1986. We also include an equity market valuation factor, measured as the price/dividend ratio for the CRSP value-weighted index. The dividends are the sum of the dividends over the past twelve months, and the value is the cum-dividend value of the index. The level of this ratio is a state variable for valuation levels in the equity market, and its monthly first difference is used as a factor.

A.6 Sharpe Style Indexes

Following Sharpe (1992) we combine the asset class returns, R_i , using a set of portfolio weights, $\{w_i\}$, to minimize the "tracking error" between the return of the fund group, R_p , and the portfolio, $\sum_i w_i R_i$. The portfolio weights are required to sum to 1.0 and must be non-negative, which rules out short positions:

$$\begin{aligned} & \text{Min}_{\{w_i\}} \text{Var}[R_p - \sum_i w_i R_i], & (\text{A.4}) \\ & \text{subject to: } \quad \sum_i w_i = 1, \quad w_i \geq 0 \text{ for all } i, \end{aligned}$$

where $\text{Var}[\cdot]$ denotes the variance. We solve the problem numerically. The asset class returns include US Treasury bonds of three maturity ranges from CRSP (less than 12 months, less than 48 months and greater than 120 months), the Lehman Global bond index, the Lehman US Mortgage Backed Securities index, the Merrill Lynch High Yield US Master index and the Lehman US Credit Aaa bond index.

References

- Admati, Anat, Sudipto Bhattacharya, Paul Pfleiderer and Stephen A. Ross, 1986, On timing and selectivity, *Journal of Finance* 61, 715-732.
- Aragon, George, 2005, Timing multiple markets: Theory and evidence from balanced mutual funds, working paper, Arizona State.
- Aragon, George, and Wayne Ferson, 2007, Portfolio Performance Evaluation, *Foundations and Trends in Finance*, Now Publishers vol. 2, No. 2, 83-190.
- Balduzzi, Pierluigi, and Cesare Robotti, 2008, Mimicking portfolios, economic risk premia and tests of multi-beta models, *Journal of Business and Economic Statistics* 26, 354-368.
- Becker, C., W. Ferson, D. Myers and M. Schill, 1999, Conditional Market timing with Benchmark investors, *Journal of Financial Economics* 52, 119-148.
- Bernanke, Benjamin, 1983, Nonmonetary effects of financial crisis in the propagation of the Great Depression, *American Economic Review* 73, 257-276.
- Bessembinder, Hendrick, W. Maxwell and K. Venkataraman, 2006, Market transparency, liquidity externalities and institutional trading costs in corporate bonds, *Journal of Financial Economics* 82 (2), 251-288.
- Bhattacharya, S., and P. Pfleiderer, 1983, A note on performance evaluation, Technical Report #714, Stanford University Graduate School of Business.
- Biais, Bruno and Fany Decker, 2006, Liquidity and price discovery in the European corporate bond market, working paper, Toulouse University.
- Blake, Christopher R., Edwin J. Elton and Martin J. Gruber, 1993, The performance of bond mutual funds, *Journal of Business* 66, 371-403.
- Blume, Marshall, D. Keim and J. Patel, 1991, Returns and volatility of low-grade bonds: 1977:1989, *Journal of Finance* 46, 49-74.
- Bollen, N., and J. Busse, 2001, On the timing ability of mutual fund managers, *Journal of Finance* 56, 1175-1094.
- Bondarenko, O., 2004, Market price of variance risk and performance of hedge funds, working paper, University of Illinois at Chicago.
- Brandt, Michael W., and Kenneth A. Kavajecz, 2004, Price Discovery in the U.S. Treasury Market: The Impact of Orderflow and Liquidity on the Yield Curve, *Journal of Finance* 59, 2623-2654.
- Brown, Stephen, David R. Gallagher, Ono W. Steenbeck and Peter L. Swan, 2004, Double or nothing: Patterns of equity fund holdings and transactions, working paper, New York University.

- Brown, David T. and William J. Marshall, 2001, Assessing fixed income manager style and performance from historical returns, *Journal of Fixed Income* 10 (March).
- Christopherson, Jon A, W. Ferson and Andrew L. Turner, 1999, Performance evaluation using Conditional alphas and betas, *Journal of Portfolio Management* 26, 59-72.
- Comer, George, 2006, Evaluating Bond Fund Sector Timing Skill, working paper, Georgetown University.
- Comer, George, Vaneesha Boney and Lynne Kelley, 2009, Timing the investment grade securities market: Evidence from high quality bond funds, *Journal of Empirical Finance* 16, 55-69.
- Edwards, A., L. Harris and M. Piwowar, 2007, Corporate bond market transparency and transactions costs, *Journal of Finance* 62 (3), 1421-1451.
- Elton, Edwin, Martin J. Gruber and Christopher R. Blake, 1995, Fundamental Economic Variables, Expected returns and bond fund performance, *Journal of Finance* 50, 1229-1256.
- Elton, Edwin, Martin J. Gruber and Christopher R. Blake, 2001, A first look at the Accuracy of the CRSP and Morningstar Mutual fund databases, *Journal of Finance* 56, 2415-2430.
- Evans, Richard, 2007, Incubation bias, working paper, University of Virginia.
- Ferson, W., 2008, The problem of alpha and performance measurement, working paper, University of Southern California.
- Ferson, W., Tyler Henry and Darren Kisgen, 2006, Evaluating Government Bond Funds using Stochastic Discount Factors, *Review of Financial Studies* 19, 423-455.
- Ferson, W. and Kenneth Khang, 2002, Conditional performance measurement using portfolio weights: Evidence for pension funds, *Journal of Financial Economics* 65, 249-282.
- Ferson, Wayne E. and Meijun Qian, 2004, Conditional Performance Evaluation Revisited, Research Foundation Monograph of the CFA Institute ISBN 0-943205-69-7, 84 pages.
- Ferson, W. and Rudi Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *Journal of Finance* 51, 425-462.
- Ferson, W., S. Sarkissian and T. Simin, 2003, Spurious regressions in Financial Economics? *Journal of Finance* 58, 1393-1414.
- Ferson, W., S. Sarkissian and T. Simin, 2008, Asset Pricing Models with Conditional Alphas and Betas: The Effects of Data Snooping and Spurious Regression, *Journal of Financial and Quantitative Analysis* 43, 331-354.
- Ferson, W. and Vincent A. Warther, 1996, Evaluating Fund Performance in a Dynamic Market," *Financial Analysts Journal* 52, no. 6, pp.20-28.

- Fisher, Lawrence, 1966, An algorithm for finding exact rates of return, *Journal of Business* 39, 111-118.
- Fleming, M., and E. Remolona, 1999, Price Formation and Liquidity in the U.S. Treasury Market: The Response to Public Information, *Journal of Finance* 54, 1901-1915.
- Fung, W., and D. Hsieh, 2001, The risk in hedge fund strategies: Theory and evidence from trend followers, *Review of Financial Studies* 14, 313-341.
- Gatev, Evan and Phillip Strahan, 2006, Banks' advantage in supplying liquidity: Theory and evidence from the commercial paper market, *Journal of Finance* 61, 867-892.
- Getmansky, Mila, Andrew W. Lo and Igor Makarov, 2004, An econometric model of serial correlation and illiquidity in hedge fund returns, *Journal of Financial Economics* 74, 529-610.
- Glosten, L., and Ravi Jagannathan, 1994, A contingent claims approach to performance evaluation, *Journal of Empirical Finance* 1, 133-166.
- Goetzmann, W., Ingersoll J., Ivkovic Z., 2000, Monthly measurement of daily timers, *Journal of Financial and Quantitative Analysis* 35, 257-290.
- Grinblatt, M., Titman, S., 1989. Mutual fund performance: an analysis of quarterly portfolio holdings. *Journal of Business* 62, 393-416.
- Gutierrez, R., W. Maxwell and D. Xu, 2008, Persistent performance in corporate bond mutual funds, working paper, University of Oregon.
- Hansen, Lars P., 1982, Large sample properties of the generalized method of moments estimators, *Econometrica* 50, 1029-1054.
- Hotchkiss, E., M. Goldstein and E. Sirri, 2009, Transparency and liquidity: A controlled experiment on corporate bonds, *Review of Financial Studies* (forthcoming).
- Jagannathan, R., Korajczyk, R., 1986. Assessing the market timing performance of managed portfolios. *Journal of Business* 59, 217-236.
- Jensen, M., 1968, The Performance of Mutual Funds in the Period 1945–1964, *Journal of Finance* 23, 389–346.
- Jiang, George T., Tong Yao and Tong Yu, 2007, Do mutual funds time the market? Evidence from portfolio holdings, *Journal of Financial Economics* 86(3), 724-758.
- Karpoff, J., 1987, The Relation Between Price Changes and Trading Volume: A Survey, *Journal of Financial and Quantitative Analysis* 22, 109-126.
- Kraus, Alan and Robert Litzenberger, 1976, Skewness preference and the valuation of risky assets, *Journal of Finance* 31, 1085-1100.

Lehman Brothers, 2006, A Guide to the Lehman Brothers Global Family of Indices, Lehman Brothers Fixed Income Research.

Litterman, R. and J. Sheinkman, 1991, Common factors affecting bond returns, *Journal of Fixed Income* 1.

Merton, Robert C. and Roy D. Henriksson. 1981. "On market timing and investment performance II: Statistical procedures for evaluating forecasting skills." *Journal of Business*, 54: 513-534.

Moneta, Fabio, 2008, Measuring bond mutual fund performance with portfolio characteristics, working paper, Boston College.

Morey, R. M. and E.S. O'Neal, 2005, Window dressing in bond mutual funds, *Journal of Financial Research* (forthcoming).

Rubinstein, Mark, 1973, A comparative static analysis of risk premium, *Journal of Business* 19, 425-442.

Scholes, Myron and Joseph Williams, 1977, Estimating betas from nonsynchronous data, *Journal of Financial Economics* 5, 309-327.

Sharpe, William. F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425-42.

Sharpe, William. F., 1992, Asset Allocation: Management style and performance measurement, *Journal of Portfolio Management* 18, 7-19.

Stein, Cotlar, 1973, Estimating the mean of a multivariate normal distribution, *Proceedings of the Prague Symposium on Asymptotic Statistics*, Prague.

Treynor, Jack, and Kay Mazuy, 1966, "Can Mutual funds Outguess the Market?" *Harvard Business Review*, 44, 131-136.

Wang, Ying, 2007, Timing ability of government bond fund managers: Evidence from portfolio holdings, working paper, Penn State University.

Xing, Yuhang, Xiaoyan Zhang, and Xing Zhou, 2007, Systematic liquidity in corporate Bonds, working paper, Rice University, Cornell University and Rutgers University.

Table 1

Mutual Fund Monthly Returns: Summary Statistics. The sample periods for the fund returns are January of the year indicated under Begin through March of 2007. Nobs is the number of nonmissing time series observations, Begno is the number of funds at the start of the sample period and Endno is the number in March of 2007. The returns are percent per month. Mean is the sample mean, std is the sample standard deviation, ρ_1 is the first order sample autocorrelation and ρ_2 is the second order autocorrelation. Panel B presents the portfolio weights of Sharpe style benchmarks associated with each group of funds. The benchmarks are formed from the returns to US Treasury bonds with less than or equal to 12 (le12) 48 (le48) months to maturity, greater than 120 months to maturity (gt120), a high-grade corporate bond index (cb), a low-grade corporate bond index (junk), a global bond index (global) and a mortgage-backed securities index (mort). The weights are estimated using all available months from 1976 through 2007.

Panel A: Equally-weighted Portfolios of Mutual Funds

Style	Begin	nobs	Begno	Endno	Mean	Min	Max	Std	ρ_1	ρ_2
All	1962	543	4	1054	0.617	-5.397	9.770	1.511	0.25	-0.01
Index	1991	195	1	18	0.546	-3.609	3.907	1.122	0.14	-0.07
Global	1994	159	47	54	0.436	-3.483	4.557	1.455	0.16	-0.08
Short Term	1993	171	126	217	0.369	-0.856	2.267	0.455	0.30	0.15
Government	1986	255	1	127	0.547	-3.817	4.318	1.248	0.14	-0.12
Mortgage	1991	195	10	61	0.491	-2.411	3.188	0.854	0.20	0.01
Corporate	1962	543	4	308	0.627	-7.339	10.97	1.691	0.18	-0.03
High Yield	1991	195	34	135	0.737	-6.900	7.535	1.835	0.25	0.08
Other	1964	543	1	134	0.661	-5.103	9.609	1.640	0.22	0.001

Panel B: Sharpe Style Benchmarks

Funds	Weights assigned to asset classes:							Autocorrelations:	
	le12	le48	gt120	cb	junk	global	mort	ρ_1	ρ_2
All	0.00	0.12	0.03	0.35	0.25	0.05	0.19	0.19	-0.03
Index	0.07	0.20	0.21	0.26	0.03	0.02	0.20	0.14	-0.04
Global	0.00	0.00	0.00	0.22	0.17	0.61	0.00	0.20	-0.02
Short Term	0.54	0.19	0.00	0.14	0.02	0.03	0.08	0.21	-0.01
Government	0.06	0.13	0.21	0.36	0.00	0.03	0.21	0.13	-0.03
Mortgage	0.09	0.00	0.02	0.21	0.00	0.00	0.67	0.13	-0.05
Corporate	0.00	0.15	0.07	0.55	0.17	0.02	0.04	0.17	-0.03
High Yield	0.00	0.10	0.00	0.00	0.77	0.00	0.13	0.21	-0.003
Other	0.00	0.00	0.00	0.29	0.39	0.06	0.27	0.20	-0.02

Table 2

Summary Statistics for the bond market factor data. The sample periods begin as indicated under Starts (yyyymm), and all series end in December of 2007. Nobs is the number of time series observations, excluding missing values. The units are percent per year, except the US dollar (an index number) and Equity values (a price to dividend ratio). Mean is the sample mean, std is the sample standard deviation and ρ_1 is the first order sample autocorrelation of the series.

Panel A: Levels of the Factors

Factor	Starts	Nobs	mean	min	max	std	ρ_1
short rate	196201	540	5.847	0.904	16.38	2.842	0.981
term slope	196201	539	0.836	-3.160	3.310	1.115	0.955
curvature	196907	450	0.168	-1.097	0.773	0.286	0.834
credit spread	196201	551	0.989	0.320	2.820	0.418	0.961
mortgage spread	197104	393	2.184	-0.410	5.580	0.821	0.839
liquidity spread	198211	290	0.402	-0.151	2.179	0.345	0.658
US dollar	197101	443	110.3	84.48	167.7	14.77	0.986
Equity Values	196201	540	36.54	16.27	71.18	13.24	0.992
Equity Volatility	198601	252	19.77	10.63	61.41	6.940	0.798

Panel B: First differences of the Factors

short rate	196202	539	0.0041	-4.158	2.611	0.539	0.132
term slope	196202	537	-0.0016	-1.460	2.800	0.334	0.130
curvature	196908	449	-0.0004	-0.807	0.770	0.164	-0.292
credit spread	196202	550	0.0007	-0.550	0.680	0.116	0.063
mortgage spread	197105	392	0.0049	-2.110	3.240	0.461	-0.097
liquidity spread	198212	289	0.0004	-1.845	1.718	0.286	-0.399
US dollar	197102	442	-0.0652	-8.132	8.694	2.318	0.159
Equity Values	196202	539	0.0233	-6.630	6.508	1.615	-0.035
Equity Volatility	198602	251	-0.0319	-15.28	39.03	4.388	-0.193

Table 3

Regressions of bond funds and benchmark index returns on changes in factors and their squares. The sample starts in February of 1962 or later, depending on the factor and fund style, and ends in March of 2007. The t-ratios for the regression coefficients on the squared factor changes are shown when they exceed 1.6 in absolute value; otherwise left blank. These are based on regressions with a single factor and its square.

Style group	short	slope	curve	credit	mort.	liquid	US dollar	equity values	equity volatility
Panel A: Style Benchmarks									
Index		1.93		1.88					2.20
Global		1.64				1.99			3.29
Short Term	1.69	2.69		2.51		1.66			2.94
Government		1.96		1.88					2.22
Mortgage		2.08		1.88		1.74			2.63
Corporate		1.78		1.86		1.83			
High Yield								-1.92	
Other		1.68		1.71		2.51			
Panel B: Funds by Style									
Index	-1.81			-1.90			2.40		
Global	-2.08					-1.75	2.48		
Short Term				-1.97					
Government				-2.23			2.45		2.38
Mortgage				-2.03			1.63		
Corporate						1.85			
High Yield				3.31			2.44	-2.78	-3.58
Other				1.62		2.49			

Table 4

The effects of controls for non-timing-related nonlinearity on funds groups' timing coefficient are summarized. For each fund style group the averages of the coefficients (divided by 100) across nine factors are shown on the first line, the number of t-ratios larger than 2.0 are on the second line and the number of t-ratios less than -2.0, out of the nine ratios, are shown on the third line. The models are estimated one factor at a time on monthly data for 1962-2007, with 543 or fewer observations depending on the fund group and factor combination. The Sharpe Style benchmarks are used. "Only benchmark" is the baseline model with a possibly nonlinear benchmark but no other controls. Public Information introduces only the controls for public information effects, and similarly for the columns labelled Interim Trading and Stale Pricing. The column labeled All Controls uses all of the controls simultaneously for the given factor.

Fund Style	Only Benchmark	Public Information	Interim Trading	Stale Pricing	All Controls
Panel A: Piecewise Linear Benchmarks					
Index	4.91	-3.44	-2.80	-4.0e+4	-6.48
	1	1	1	0	1
	5	5	1	0	0
Global	0.94	5.10	0.75	36.94	-13.45
	0	1	0	1	1
	4	4	1	0	1
Short Term	49.4	-11.4	7.57	-2.88	-7.38
	1	2	0	0	1
	6	6	2	2	0
Government	3.53	0.67	0.45	64.85	-0.85
	2	1	2	0	1
	3	2	2	0	0
Mortgage	2.20	-1.42	1.08	-4.78	1.81
	2	2	2	0	0
	4	3	1	0	0
Corporate	0.79	-15.8	6.85	223	-3.70
	1	1	2	2	1
	0	0	0	0	0
High Yield	0.44	1.29	0.48	1.7e+5	-37.06
	1	2	1	0	1
	3	2	0	0	0
Other	-0.25	0.56	-0.56	1.1e+03	-0.86
	1	1	2	1	1
	0	0	0	0	0

Table 4, continued.

Fund Style	Only Benchmark	Public Information	Interim Trading	Stale Pricing	All Controls
Panel B: Linear Benchmarks					
Index	-8.52 0 7	-48.9 0 6	-38.1 0 1	-2.48 0 0	-7.7e+5 0 1
Global	2.23 0 6	-8.67 0 5	-6.32 0 0	-2.1e+4 1 1	-7.5e+3 1 1
Short Term	8.6e+4 1 7	-3.8e+4 0 6	-1.36 0 2	-2.40 0 1	3.17 2 0
Government	26.0 1 3	-4.6 0 4	3.17 1 2	61.05 3 0	-50.39 0 0
Mortgage	266 0 7	-313 0 7	-159 0 1	1.4e+3 2 0	-9.96 0 0
Corporate	-15.3 0 0	-30.8 0 0	-10.2 1 0	-44.29 1 0	-1.59 0 0
High Yield	-44.6 0 4	-92.0 0 4	-73.3 0 1	2.2e+3 1 0	-7.5e+4 0 0
Other	-1.7e+3 0 0	-2.6e+3 1 0	-1.9e+3 2 0	-1.4e+2 2 0	1.4e+3 1 1

Table 5

Timing coefficients for individual bond funds. The table summarizes the fractions of funds with timing coefficients larger than the critical values for the indicated fractiles from the null distribution in which the true timing coefficients for the indicate factor are zero. The monthly samples start in February of 1962 or later, depending on the factor, and end in March of 2007. The Sharpe Style benchmarks are used and the timing coefficients are adjusted for non-timing-related nonlinearities.

null fractile	short	slope	curve	credit	mortgage	liquidity	exchange	eq. value	eq. vol.
------------------	-------	-------	-------	--------	----------	-----------	----------	-----------	----------

The distributions of the timing coefficients

0.975	0.023	0.027	0.011	0.021	0.023	0.024	0.017	0.015	0.024
0.950	0.032	0.057	0.025	0.033	0.043	0.047	0.035	0.032	0.039
0.900	0.058	0.109	0.073	0.055	0.074	0.107	0.057	0.071	0.068
0.500	0.536	0.564	0.391	0.515	0.491	0.510	0.528	0.566	0.485
0.100	0.923	0.938	0.891	0.933	0.940	0.964	0.919	0.929	0.925
0.050	0.945	0.971	0.953	0.969	0.967	0.981	0.960	0.960	0.961
0.025	0.969	0.986	0.987	0.991	0.985	0.991	0.978	0.974	0.984
# cases	1335	1392	1410	1266	1204	1393	1391	1415	1402

Table 6

Nonlinearity-adjusted performance of individual bond funds. The first three columns of Panel A summarize the mean excess return over a short term treasury, the mean return over a style benchmark, and the values of the estimated alphas under the null hypothesis that the true alphas are zero, taken at various fractiles of the null distribution across funds (critical alphas). These are the averages across factors. The remaining columns report the fractions of funds with alphas larger than the factor-specific critical values. The average excess returns are percent per month. Panel B presents summary statistics for funds' loadings on the style benchmarks, the average risk premiums (hedge premiums) and the distributions of funds' loadings on the hedge portfolios for the squared factors. The monthly samples start in February of 1962 or later, depending on the factor and fund, and end by March of 2007.

fractile	mean return	mean net of bench	critical alphas	short	slope	curve	credit	mort.	liq.	dollar	equity value	equity vol.
Panel A: The distribution of Nonlinearity-Adjusted Alphas												
0.975	0.818	0.171	0.150	0.023	0.031	0.024	0.018	0.020	0.010	0.021	0.023	0.017
0.950	0.715	0.104	0.099	0.053	0.051	0.044	0.029	0.036	0.014	0.048	0.046	0.046
0.900	0.632	0.036	0.061	0.079	0.071	0.074	0.051	0.066	0.030	0.069	0.071	0.068
0.500	0.463	-0.067	-0.000	0.217	0.216	0.205	0.192	0.198	0.143	0.214	0.217	0.220
0.100	0.285	-0.199	-0.064	0.581	0.572	0.616	0.552	0.553	0.468	0.604	0.592	0.597
0.050	0.193	-0.282	-0.108	0.732	0.758	0.795	0.752	0.740	0.659	0.768	0.814	0.778
0.025	0.100	-0.386	-0.156	0.862	0.870	0.892	0.819	0.847	0.812	0.861	0.891	0.866
# cases				1329	1368	1375	1251	1179	1375	1357	1376	1376
Panel B: Premiums and loadings												
Fund Loadings on Style Benchmarks:												
Upper 10%												1.38
Median												0.91
Lower 10%												0.44
Average Hedge premiums				0.415	0.231	-0.332	0.107	0.144	-0.029	-0.244	3.59	-2.93
Fund Hedge Portfolio Loadings:												
Upper 10%				0.134	0.145	0.054	0.130	0.226	0.103	0.064	0.091	0.093
Median				0.005	0.012	0.001	0.003	0.006	-0.001	-0.000	-0.006	0.000
Lower 10%				-0.224	-0.128	-0.065	-0.163	-0.209	-0.129	-0.078	-0.118	0.072
Correlations of Hedge Portfolios: with Squared factor changes				-0.278	-0.383	-0.308	-0.215	-0.456	0.160	0.510	-0.273	0.532

Table 7

Nonlinearity-adjusted performance of individual bond funds gross of transactions costs. Transactions costs are estimated as the average expense ratio of each fund plus an assumed round trip trading cost associated with the fund style, multiplied by the average reported turnover. These costs are added back to the fund return. The first three columns of Panel A summarize the mean excess return over a short term Treasury, the mean return over a style benchmark, and the values of the estimated alphas under the null hypothesis that the true alphas are zero, taken at various fractiles of the null distribution across funds (critical alphas). These are the averages across factors. The remaining columns report the fractions of funds with alphas larger than the factor-specific critical values. The average excess returns are percent per month. The monthly samples start in February of 1962 or later, depending on the factor and fund, and end by March of 2007.

fractile	mean return	mean net of bench	critical alphas	short	slope	curve	credit	mort.	liq.	dollar	equity value	equity vol.
The distribution of before cost Nonlinearity- Adjusted Alphas												
0.975	0.952	0.318	0.149	0.158	0.113	0.144	0.101	0.110	0.141	0.161	0.132	0.155
0.950	0.855	0.239	0.104	0.257	0.198	0.239	0.184	0.197	0.228	0.258	0.251	0.256
0.900	0.768	0.176	0.063	0.430	0.390	0.377	0.360	0.335	0.412	0.409	0.407	0.425
0.500	0.572	0.032	-0.000	0.767	0.754	0.763	0.746	0.730	0.761	0.742	0.753	0.768
0.100	0.389	-0.081	-0.067	0.909	0.909	0.912	0.907	0.907	0.910	0.900	0.906	0.907
0.050	0.303	-0.145	-0.114	0.954	0.941	0.944	0.934	0.940	0.948	0.939	0.945	0.942
0.025	0.241	-0.225	-0.169	0.969	0.958	0.972	0.957	0.956	0.967	0.960	0.966	0.963
# cases				1329	1367	1374	1251	1178	1375	1357	1375	1375