Timing Ability of Government Bond Fund Managers: Evidence from Portfolio Holdings

Jing-Zhi Huang, Ying Wang

MANAGEMENT SCIENCE
Vol. 60, No. 8, August 2014, pp. 2091–2109
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2013.1843

Timing Ability of Government Bond Fund Managers: Evidence from Portfolio Holdings

Jing-Zhi Huang
Smeal College of Business, Pennsylvania State University, University Park, Pennsylvania 16802, jxh56@psu.edu

Ying Wang
School of Business and Center for Institutional Investment Management, University at Albany, State University of New York, Albany, New York 12222, ywang@albany.edu

This study examines the ability of government bond fund managers to time the bond market, based on their monthly or quarterly holdings of Treasury securities during the 1997–2006 period. We find that, on average, government bond funds exhibit significantly positive timing ability at the one-month horizon under an unconditional holdings-based timing measure. However, our results indicate that managers' actions based on public information can explain the documented positive timing ability—namely, the average government bond fund has neutral or even slightly negative conditional market timing ability once public information is controlled for. Nonetheless, we find evidence that fund managers specializing in Treasury securities can better interpret public information than general government bond fund managers do.

Keywords: government bond funds; market timing; bond fund holdings; macroeconomic news announcements
History: Received December 17, 2011; accepted September 12, 2013, by Wei Xiong, finance. Published online in *Articles in Advance* March 10, 2014.

1. Introduction

The market timing ability of equity fund managers has been extensively discussed in the literature.[1] Yet little attention has been paid so far to the same ability of bond fund managers even though bond funds play an increasingly important role in investors' portfolios. For instance, according to Investment Company Institute (2010), the total assets under management by nonmunicipal bond funds are around $1.75 trillion at the end of 2009 and account for 15.7% of total mutual fund assets, as compared with 44.6% for equity funds.

Most bond fund managers are in fact believed to be timers (see, e.g., Elton et al. 2003). Indeed, there has been anecdotal evidence suggesting that bond fund managers time the market or base asset allocation decisions on their views of future market conditions, especially future interest rate changes.[2] To some extent this is not surprising because, as the benchmark for other fixed-income securities, Treasury yields are among the key subjects of forecasts of economists and market strategists. Nonetheless, the empirical evidence on the ability of managers to time bond market factors, in particular, the Treasury market factor, is inconclusive so far. One possible reason is that the evidence documented so far is all based on studies using either a return-based timing measure or a return attribution approach with annual or quarterly sector weights (e.g., Boney et al. 2009, Chen et al. 2010, Comer 2006). An alternative approach is to use so-called holdings-based measures constructed from funds' actual portfolio holdings (which are very different from sector weights provided by a data vendor such as Morningstar). However, this approach has yet been applied to bond funds even though there are reportedly a number of advantages of using such measures to evaluate timing abilities of equity fund managers (e.g., Daniel et al. 1997, Jiang et al. 2007, Elton et al. 2012, Ferson and Mo 2013).[3]

[1] Studies that find no or negative timing ability include Treynor and Mazuy (1966), Chang and Lewellen (1984), Henriksson (1984), Graham and Harvey (1996), Ferson and Schadt (1996), Becker et al. (1999), Jiang (2003), and Elton et al. (2012). However, Bollen and Busse (2001) and Mamaysky et al. (2008) document significant market timing ability using daily and monthly fund returns, respectively, and Jiang et al. (2007) find similar results using holdings-based performance analysis. Busse (1999) and Cao et al. (2013) provide evidence on volatility and liquidity timing abilities, respectively.

[2] See, for example, McDonald (2006), where Bill Gross of PIMCO and George Fischer of Fidelity, two top government bond fund managers, discuss their views about future interest rate changes and the portfolio rebalances they have done in anticipation of the fluctuations.

[3] To the best of our knowledge, the only other analysis of bond funds, since our study, that uses holdings is Cici and Gibson (2012), who examine the performance of corporate bond funds.

In this study, we examine the market timing ability of government bond fund managers based on their monthly or quarterly actual portfolio holdings, using a holdings-based timing measure developed by Jiang et al. (2007). Specifically, we evaluate the ability of fund managers to time the Treasury bond market, based on their holdings of Treasury securities.

As pointed out by Daniel et al. (1997), returns generated from portfolio holdings are more appropriate than actual fund returns in determining whether managers have such abilities. In particular, as shown by Jiang et al. (2007), holdings-based timing measures are free from potential biases associated with the passive timing and interim trading effects illustrated in Goetzmann et al. (2000), Ferson and Khang (2002), and Ferson et al. (2006). In addition, one can use the holdings information to construct portfolio betas as weighted average betas of individual securities held by a fund by utilizing return observations of individual securities at a higher frequency (e.g., daily). As such, holdings-based timing tests have higher statistical power than their return-based counterparts (Ferson and Khang 2002, Jiang et al. 2007). Furthermore, the use of monthly or quarterly holdings data can mitigate the potential issue of window dressing; Elton et al. (2010) find no evidence of semiannual or quarterly window dressing and further point out that funds concerned about having window dressing detected by investors would not report monthly holdings. Last, although the focus of our study is not on alpha per se, it is worth mentioning that Elton et al. (2011) show that alphas computed using holdings are better than those computed from returns-based regressions and that "the more frequently the holdings data are available, the greater the benefit" (p. 341).

We focus on government bond funds and their Treasury security portfolios for several reasons. First, because such funds mainly hold Treasuries and the returns of individual Treasuries are highly correlated, the main mechanism by which government bond fund managers can deliver superior performance is to engage in market timing rather than security selection activities. As such, those managers are the ones most likely capable of timing the Treasury bond market. Additionally, government bond fund managers need to roll over their positions on a regular basis to maintain the target investment horizon or sometimes the target duration. The timing of such activities is very important and can potentially have significant effect on fund performance. In this aspect, bond funds already have the active component embedded in their investment strategy, which is different from equity and other funds where there is no required rollover of positions.

Next, because of their high liquidity and low transaction costs, Treasury securities are more suitable than are less liquid fixed-income securities for timers and data on daily Treasury returns are also more reliable than those for less liquid securities—the latter property is essential to the implementation of a holdings-based timing measure. Additionally, the potential stale pricing problem of bond fund assets (an issue raised by Chen et al. 2010) should not be of a concern for our analysis.

Last, a manager of Treasury security portfolios is primarily concerned with the interest rate risk, unlike a manager of other types of bond funds (e.g., corporate bond funds) whose returns are also subject to other market risks such as default risk. As a result, we can focus on fund managers' ability to time the bond market return factor only and carry out the analysis using a return generating process where the bond market return itself is the explanatory variable. More specifically, we can specify bond fund returns as a linear function of the bond market return factor rather than a nonlinear function of non-return factors (such as the level, slope, and curvature of yield curves). Such a specification naturally addresses another concern raised by Chen et al. (2010) that non-timing-related nonlinearity in fund returns can arise as a result of the nonlinear relationship between bond returns and interest rate changes.

Using a sample of 146 government bond funds constructed from survivor-bias free Morningstar and the Center for Research in Security Prices (CRSP) mutual fund databases over 1997–2006,[4] we find that, on average, government bond fund managers possess significantly positive (unconditional) timing ability at the one-month forecasting horizon under the holdings-based Treynor and Mazuy (1966) timing measure, based on a bootstrap approach to statistical inference. This result is robust to alternative ways of estimating security betas and adjusting portfolio betas, the length of data records, different time periods, and the use of alternative holdings-based timing measures.

Given the evidence on the impact of macroeconomic variables and news on bond prices and returns (e.g., Fama and Bliss 1987, Balduzzi et al. 2001, Ludvigson and Ng 2009, Huang and Shi 2010), we also conduct a conditional analysis (Ferson and Schadt 1996) to understand what specific information may be used to help managers time the bond market. We present evidence that managers do adjust their portfolio betas in response to public information and, furthermore, that the timing ability can be explained

[4] The widely used Thomson database does not include holdings data on bonds. See Elton et al. (2012) on other advantages of Morningstar holdings data over Thomson holdings data.

by historical bond betas or economic news on nonfarm payroll, factory orders, and personal consumption. Namely, the average government bond fund is found to have neutral or even slightly negative conditional market timing ability.

Additionally, we investigate the scope of funds' rollover activities to better understand this unique aspect of government bond fund management. Our results indicate that the majority of funds in our sample do roll over their positions in Treasuries and that less than one third of such activities involve a passive rollover strategy—that may induce negative artificial timing (§4.4.4). Also, funds with fewer rollover activities tend to exhibit more significantly positive timing ability, providing evidence consistent with the presence of the negative artificial timing effect.

We also conduct a subsample analysis to compare Treasury bond funds (specializing in Treasury securities) with general government funds (investing in a combination of Treasury, mortgage-backed, and agency securities) to gain a more complete picture of fund timing ability. We find that the average Treasury bond fund engages in fewer rollover activities and has more significantly positive timing ability than the average general government fund but that the former does not possess timing ability beyond public information. These findings imply that on average Treasury bond fund managers have better abilities in interpreting public information and, perhaps as a result, have less rollover than general government bond funds.

The literature on bond fund performance goes back at least to Cornell and Green (1991), Blake et al. (1993), and Elton et al. (1995). These studies document that, in general, bond funds underperform the relevant benchmarks. Ferson et al. (2006) incorporate dynamic term structure models into the performance analysis of bond funds.

In a comprehensive study that examines whether general bond fund managers can time nine different factors related to various markets, Chen et al. (2010) show that it is important to control for four potential non-timing-related sources of nonlinearity in fund returns: passive timing, interim trading, stale pricing, and public information. They find that without such controls, managers appear to have negative timing ability under the standard return-based timing measures; with these controls, "the timing coefficients appear neutral to weakly positive" (p. 72) (under the assumption that managers can time only one factor each time). An alternative approach is to evaluate the timing ability of bond funds by examining their portfolio sector weights or attribution returns (Brown and Marshall 2001). For example, Comer (2006) finds that general government bond funds may possess the ability to time the government/credit or mortgage sector. Using a variation of Sharpe's (1992) style analysis and attribution returns, Boney et al. (2009) find evidence of negative timing ability of high-quality corporate bond funds that can shift assets between bonds and cash. Moneta (2009) documents that based on sector weights, the average bond fund exhibits neutral timing ability across different sectors and credit quality ratings. And all three aforementioned studies focus on unconditional timing models.

The remainder of this paper is organized as follows. Section 2 describes data used in our empirical analysis. Section 3 discusses the empirical methodology and, in particular, the holdings-based timing measures and the bootstrap approach used to derive statistical inference. Section 4 presents empirical results. Section 5 concludes.

2. Data

In this section, we discuss first how to construct our sample of government bond mutual funds along with data on their holdings of Treasury securities. We then describe the set of economic news releases used in the analysis of conditional timing ability of fund managers.

2.1. Mutual Fund Data

We use two mutual fund databases in this study. The first is the Morningstar Principia Pro Plus for Mutual Funds data set, which contains stock and bond holdings for all mutual funds existing at any time between January 1997 and December 2006. Before 2004, mutual funds were required by law to disclose their holdings twice every calendar year. Effective May 2004, the SEC increased the mandatory disclosure frequency from semiannually to quarterly. Many funds, however, voluntarily disclose their holdings monthly, especially since 2003.[5] As such, some funds may have unequally spaced holding observations if they disclose holdings more frequently in recent years. The database also includes information on prospectus objectives, Morningstar ratings, duration, and the total number of holdings for each fund. It is worth noting that we extract all the data from Morningstar historical monthly CDs, so our sample is free of survivorship bias. The second database we use is the CRSP survivor-bias-free mutual fund database, which contains quarterly or monthly fund holdings data over July 2003–December 2006 and information on fund characteristics such as total net assets (TNAs),

[5] Along with government bond funds, this more frequent disclosure is also observed among traditional (Elton et al. 2010) and alternative equity funds (Huang and Wang 2013). Following Elton et al. (2012), we refer to those funds with reported holdings for at least eight months in a calendar year as funds with monthly holdings.

expense ratios, and turnover rates over the 1962–2006 period.

For the purpose of this study, we restrict our sample to funds that are classified by both the CRSP and Morningstar as a government bond fund and exclude index, index enhanced, inflation-linked, and mortgage-backed bond funds and funds of funds. We also combine multiple share classes of the same fund (to be included in the sample) into a single fund.

We first select from CRSP those bond funds with Standard and Poor's (S&P) investment objectives of government general (GGN), government intermediate-maturity (GIM), and government short-maturity (GSM)—which invest in securities backed by the federal government, its agencies, or instrumentalities, with average maturities of over 10, 5 to 10, and 1 to 5 years, respectively. This results in an initial sample of 337 unique government bond funds from CRSP. We next select from Morningstar those funds with prospectus objectives of general government and Treasury (Morningstar bases a fund's prospectus objective on the wording in the prospectus sent by the mutual fund's distributor or underwriter). According to Morningstar, general government bond funds pursue income by investing in a combination of mortgage-backed securities, Treasuries, and agency securities, whereas Treasury bond funds seek income by generally investing at least 80% of their assets in U.S. Treasury securities. We obtain an initial sample of 313 unique government bond funds from Morningstar.

We then merge the above two initial samples by fund CUSIPs and ticker symbols and generate a matched sample of 311 funds. The 28 funds not in the matched sample, including two in the Morningstar sample and 26 in the CRSP sample, do not have sufficient holdings information available over the sample period and thus are dropped from the sample.[6] Untabulated results indicate that there are no statistically significant differences in general fund characteristics, such as the TNA, duration, time-to-maturity, turnover, expense ratio, and the annual fund return among the two initial and matched samples. As such, the matched sample of 311 funds is a representative sample of government bond funds covered by either CRSP or Morningstar.

Finally, to ensure robust statistical inference, we require a fund to have a minimum of 20 holding observations available to be included in the sample. Applying this filter to the matched sample results in our final sample of 146 unique funds used in the baseline analysis. Recall that mutual funds are not required to report their holdings on a quarterly basis until May 2004 and our sample ends in December 2006, so funds in the matched sample that are filtered out in this last stage tend to have a short history and thus be small. Indeed, untabulated results indicate that the matched and final samples on average have similar fund characteristics except the fund size. However, as mentioned in §4.1, our main findings are robust to the minimum number of holdings observations used for the sample selection. Additionally, we find no evidence that funds with larger size are more likely to successfully time the market (see §4.2).

Next, we obtain each sample fund's holdings of U.S. Treasury bills, notes, and bonds only—because of the purpose of this study—although these funds may hold non-Treasury securities.[7] We do not include funds' holdings of Treasury inflation-protected securities (TIPS) and separate trading of registered interest and principal securities (STRIPS) because data on their daily returns, necessary for estimating their individual betas, are not available. However, the exclusion of these two types of securities should not impose systematic biases on our results because they account for less than 5% of total holdings.

Panel A of Table 1 reports summary statistics of fund characteristics. As indicated in the table, among our sample of 146 government bond funds, by S&P investment objectives, 60 are GGN funds, 44 are GIM funds, and 42 are GSM funds; by Morningstar prospectus objectives, 110 are general government bond funds and 36 are Treasury bond funds. For funds with multiple share classes, we aggregate the TNAs of individual share classes; we compute other fund-level characteristics (such as Morningstar rating, duration, expense ratio, turnover rate, and return) as the TNA-weighted averages across multiple share classes of the same fund. The fund characteristics reported are first averaged over time for each fund and then averaged across funds. On average, a government bond fund has TNAs of \$413 million, a Morningstar rating of 3 (out of 5), duration of four years, an annual turnover rate of 1.64, an annual expense ratio of 0.86%, and an annual return of 5.2%. In addition, consistent with their definitions, GGN funds have the longest durations, followed by

[6] These funds do not have long enough history, especially after May 2004 when mutual funds are required to report their holdings quarterly. For instance, the 2 funds in the Morningstar sample and 6 (12) funds in the CRSP sample have fewer than two (three) years of history, in the early part of our sample period. Also, recall that the CRSP database has no information available on fund holdings prior to July 2003.

[7] Although the CRSP mutual fund database reports detailed holdings information on individual Treasury securities, Morningstar does not have the holdings of Treasury bills available. Morningstar, however, reports quarterly cash holdings (including any fixed-income securities that mature in fewer than 12 months) for each fund. Therefore, we use cash holdings as a proxy for Treasury bill holdings for those funds whose holdings are only available from Morningstar.

Table 1 Summary Statistics of Fund Characteristics

		Fund investment objectives				
		By CRSP			By Morningstar	
	All funds	GGN	GIM	GSM	General	Treasury
Panel A: Government bond funds						
Number of funds	146	60	44	42	110	36
Total net assets ($ million)	412.82	516.69	288.79	394.38	401.23	471.72
Morningstar rating	3.05	2.73	3.14	3.41	3.05	3.07
Duration (years)	3.99	5.56	3.85	1.85	3.84	4.76
Turnover rate (per year)	1.64	1.85	1.63	1.35	1.67	1.47
Expense ratio (%/year)	0.86	0.99	0.84	0.67	0.89	0.70
Annual return (%)	5.22	5.64	5.46	4.36	5.17	5.47
Total number of holdings	176	318	90	63	202	43
Panel B: Treasury security holdings of government bond funds						
Sharpe ratio	0.58	0.56	0.62	0.57	0.58	0.58
Total net assets ($ million)	203.55	243.74	125.85	208.05	168.05	420.34
Duration (years)	5.29	7.81	4.81	2.19	5.37	4.91
Investment in T-bills (%)	8.18	7.85	6.65	10.25	8.35	7.31
Investment in T-notes (%)	24.83	15.95	26.67	35.17	19.68	51.92
Investment in T-bonds (%)	16.30	23.38	10.25	7.34	13.85	29.88
Total investment in Treasury securities (%)	49.31	47.17	43.58	52.75	41.88	89.11
Average number of Treasury securities held	8	9	9	7	8	12
NOBS—holdings	41	40	43	39	42	38
Panel C: TTMs (year) of bonds at the fund and portfolio levels						
Bond funds	6.54	9.61	5.82	2.89	6.46	6.94
Treasury security portfolios	8.04	12.94	6.64	2.52	8.17	7.41
Equally weighted:						
Initiating-buy bond portfolios (mean)	7.80	11.29	7.38	3.26	7.89	7.35
Terminating-sale bond portfolios (mean)	6.55	10.13	5.83	2.20	6.70	5.80
Initiating-buy bond portfolios (median)	7.63	10.20	7.04	2.90	7.84	6.38
Terminating-sale bond portfolios (median)	6.13	9.49	5.48	2.02	6.42	4.20
Value-weighted:						
Initiating-buy bond portfolios (mean)	7.72	11.27	7.27	3.13	7.82	7.24
Terminating-sale bond portfolios (mean)	6.52	10.19	5.74	2.10	6.66	5.84
Initiating-buy bond portfolios (median)	7.26	10.57	6.72	2.85	7.73	6.18
Terminating-sale bond portfolios (median)	5.80	9.25	5.02	1.93	6.41	3.92

Notes. This table reports summary statistics on the characteristics of government bond funds (panel A), their holdings of Treasury securities (panel B), and the average time-to-maturity (TTM) of holdings at both the fund and portfolio levels (panel C), by investment objectives. Funds are divided into three subgroups by CRSP S&P investment objectives, government general (GGN), government intermediate-maturity (GIM), and government short-maturity (GSM), or two subgroups of general government (General) and Treasury bond funds by Morningstar prospectus objectives. The characteristics are first averaged over time for each fund and then averaged across funds. The TTM is obtained similarly, for both equally- and value-weighted cross-sectional averages. "NOBS—holdings" denotes the cross-sectional average number of observations with portfolio holdings information. Initiating-buy and terminating-sale positions refer to new purchases and selling of Treasury securities, respectively. Only funds with a minimum of 20 holding observations during the 1997–2006 period are included in the sample.

GIM and GSM funds. Finally, GGN funds have higher expense ratios than GIM and GSM funds; Treasury bond funds have relatively lower expense ratios compared with general government bond funds.

Panel B of Table 1 reports summary statistics on Treasury security holdings of our sample funds. On average, government bond funds have 41 reported holding observations. In particular, 135 out of 146 funds report portfolio holdings at a quarterly or higher frequency; more than 40% of funds report monthly portfolio holdings from 2003 through 2006. Interestingly, a typical government bond fund in our sample holds 176 securities, of which only 8 are Treasury securities but account for about 49% of the fund's TNAs. Other portfolio compositions include 3% TIPS and STRIPS, 6% agency securities, 38% mortgage-backed securities, and 3% other securities. However, Treasury bond funds, on average, invest 89% of their assets in 12 Treasuries and thus are more concentrated on Treasuries compared with general government bond funds. Similarly, we find that GSM funds are more concentrated than GIM and GGN funds are. Finally, an average Treasury security portfolio has a Sharpe ratio of 0.58, TNAs of $204 million, and duration of 5.3 years.

One potential concern is that holding observations might cluster in a short time period for some funds. To have a better understanding of the data structure, we report the number of calendar years that our sample funds span. We find that only two funds have life spans of 2–3 years. Out of the remaining 144 funds, 13 have life spans of 3–5 years with an average of 30 holding observations, 35 have life spans of 5–8 years with an average of 36 holding observations, and 96 have life spans of 8–10 years with an average of 45 holding observations. Overall, the results indicate that about 90% of funds in our sample have holding observations spread out over five years.

Panel C of Table 1 presents summary statistics on time-to-maturities of bonds at both the fund and Treasury security portfolio levels. We discuss this panel in detail in §4.4.4.

2.2. Data on Individual Treasury Securities

To implement a holdings-based timing measure for our sample funds, we need to estimate betas of individual Treasury securities using their daily returns. To obtain data on these returns from the CRSP, we merge holdings data with the CRSP daily Treasury database in the following way: We first use CUSIPs, which are available for those individual Treasury securities that also appear in the CRSP holdings data. Then for those Treasury securities whose CUSIPs are not available, we match them with the CRSP database by coupon, maturity, and security type and, in fact, obtain a unique match for this sample of Treasury securities. Last, we also check fund holdings with SEC mutual fund fillings from EDGAR and match the Treasuries with incomplete information.

2.3. Data on Macroeconomic News Announcements

In §4.3 we examine whether fund managers change fund betas in response to macroeconomic news announcements. We use survey data on macroeconomic variables from Bloomberg Professional for this analysis. We begin with the list of 26 commonly followed macroeconomic news announcements considered in Balduzzi et al. (2001). For each variable, we need data on both its actual released figures and market expectations (the median of preannouncement estimates by surveyed economists). Table 2 provides the list of 15 macro variables that have sufficient survey data available from Bloomberg. Announcements on all these variables are released monthly except for initial jobless claims, whose figures come out weekly.

Because these economic variables have different units, we follow Balduzzi et al. (2001) and Flannery and Protopapadakis (2002) and use the standardized surprise measure given by

$$S_{i,t_a} = (A_{i,t_a} - F_{i,t_a})/\sigma_i, \quad i = 1, \ldots, 15, \tag{1}$$

where A_{i,t_a} and F_{i,t_a} denote the actual value and median forecast of the ith variable on announcement day t_a, respectively, and σ_i represents the standard deviation of the surprise $(A_{i,t_a} - F_{i,t_a})$.

Table 2 **Macroeconomic News Releases and Their Impact on Portfolio Betas**

(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Mean		Median	
	News announcement	Variable	t	$p(t)$	t	$p(t)$
1	Civilian unemployment (% level)	*UNEM*	0.16	0.36	0.29	0.27
2	Consumer price index (% change)	*CPI*	−0.11	0.60	−0.11	0.59
3	Durable goods orders (% change)	*DGNO*	0.68	0.15	0.76	0.14
4	Index of leading indicators (% change)	*LEI*	1.12	0.00	1.33	0.00
5	Initial jobless claims—weekly (thousands)	*INJC*	−0.04	0.51	−0.11	0.57
6	Merchandise trade balances ($ billions)	*TB*	0.34	0.22	0.51	0.13
7	Nonfarm payrolls (change in thousands)	*NFP*	−0.66	0.94	−0.60	0.90
8	Producer price index (% change)	*PPI*	0.13	0.37	0.01	0.47
9	Capacity utilization (% level)	*CU*	−0.26	0.02	−0.16	0.03
10	Industrial production (% change)	*IP*	0.04	0.48	0.03	0.50
11	Consumer confidence (% level)	*CC*	−0.18	0.65	−0.13	0.60
12	Factory orders (% change)	*FO*	0.83	0.06	0.83	0.07
13	NAPM index (index value)	*NAPM*	−0.44	0.81	−0.36	0.74
14	Personal consumption (% change)	*PCE*	−0.67	0.96	−0.69	0.95
15	Personal income (% change)	*PI*	0.01	0.45	0.04	0.43

Notes. This table shows the list of 15 macroeconomic news releases (column (2)) used in the conditional timing analysis and reports the results on the impact of the releases on Treasury security portfolio betas. Columns (4)–(7) report the cross-sectional mean and median of the Newey–West t-statistics (t) and their bootstrapped p-values ($p(t)$) for the coefficient estimates from 15 univariate regressions (7), each of which is against the one-month lagged standardized surprise of a particular news release as defined by Equation (1).

3. Methodology

In this section, we first discuss how to construct holdings-based Treynor and Mazuy (1966) and Henriksson and Merton (1981) timing measures and then describe the bootstrap approach to be used to derive statistical inference.

3.1. Holdings-Based Timing Measures

Given our focus on a fund's holdings of Treasury securities, the first step is to estimate their individual betas. Let b_{it} denote the beta of individual Treasury security i as of the portfolio holding date t. We estimate b_{it} with daily returns over the past one year using the following one-factor model:

$$r_{i\tau} = a_{it} + b_{it} Term_\tau + e_{i\tau}, \quad \tau = t - k + 1, \ldots, t, \tag{2}$$

where $r_{i\tau}$ is the excess return of Treasury security i on day τ, $Term_\tau$ is the term premium defined as the difference between the long-term government bond return and the risk-free rate on day τ and representing the bond market return factor, and k is the number of daily observations on or before the portfolio holding date t in the estimation. We use the Lehman long-term government bond index return from Datastream as a proxy for the long-term government bond return and one-month T-bill rate from CRSP as a proxy for the risk-free rate.

Note that the specification (2) is a special case of the Fama and French (1993) two-factor model of bond returns and is also used in Elton et al. (2012).[8] Default risk—the other bond market risk factor used in Fama and French—does not enter Equation (2) because we focus on Treasury securities. Moreover, because we use the bond market excess return, *Term*, rather than yield curve factors, as an underlying risk factor in Equation (2), we need not consider nonlinear terms related to yield curve factors as additional independent variables. As a result, we can control for one source of non-timing-related nonlinearity identified by Chen et al. (2010). In addition, we need not use duration, a measure relying on the assumption that the entire bond market can be described by a single unobservable "level" factor.

To ensure robust estimation of b_i, we require a Treasury security to have a minimum of 20 daily return observations; otherwise we set its beta to be the matching portfolio beta. We construct matching portfolios in a similar way to Fama (1984) except that we base on duration (from Morningstar) instead of maturity in constructing these portfolios.[9] For the cash portion in the Morningstar database, we set its beta to be that of the matching portfolio with average duration of no more than one year.

Next, we compute the beta of each fund's Treasury security portfolio at time t (the beginning of period $t+1$), $\hat{\beta}_t$, as follows:

$$\hat{\beta}_t = \sum_{i=1}^{N_t} w_{it} \hat{b}_{it}, \tag{3}$$

where w_{it} is the normalized portfolio weight of Treasury security i at time t such that $\sum_{i=1}^{N_t} w_{it} = 1$, $\hat{b}_{it}$ is the beta estimate of Treasury security i from (2), and N_t is the number of Treasuries held by each fund as of the portfolio holding date t (when the holding info is available). Such constructed portfolio betas are referred to as "bottom–up" betas in Elton et al. (2011, 2012).

We can then construct two holdings-based timing measures in the spirit of Jiang et al. (2007). We estimate the first one, a holdings-based Treynor and Mazuy (1966; hereafter, TM) timing measure, denoted by γ, from the following regression:

$$\hat{\beta}_t = \alpha + \gamma Term_{t+h} + \eta_t, \tag{4}$$

where $\hat{\beta}_t$ is the time-t portfolio beta from (3) and $Term_{t+h}$ is the term premium over the period $[t, t+h]$ with a forecasting horizon of h periods (e.g., one, three, or six months). We obtain the other timing measure from the following specification:

$$\hat{\beta}_t = \alpha + \gamma_{\mathrm{HM}} I_{\{Term_{t+h}>0\}} + \eta_t, \tag{5}$$

where $I_{\{\cdot\}}$ is an indicator function and γ_{HM} denotes the holdings-based Henriksson and Merton (1981) timing measure. If managers possess positive timing ability, we expect the γ and γ_{HM} coefficients estimated, respectively, from (4) and (5) to be significantly positive. In the empirical analysis that follows, we focus on the TM timing measure and consider the Henriksson–Merton timing measure in robustness checks. Also, the holdings-based timing measures estimated from Equations (4) and (5) are unconditional ones. We will consider conditional timing measures by controlling for public information in §4.3.

As mentioned earlier, holdings-based timing measures have several advantages over the traditional returns-based measures (Ferson and Khang 2002, Jiang et al. 2007, Elton et al. 2012). For instance, the former do not suffer from any biases induced by the interim trading or dynamic trading effect because they use only ex ante information on portfolio holdings. They are also robust to the passive timing effect that could be an important concern for government bond funds given the convexity of bond prices to interest rate changes.

For funds with evenly spaced holding observations, we compute the t-statistics for $\hat{\gamma}$ based on Newey and West (1987) heteroskedasticity- and autocorrelation-consistent standard errors with six lags. As noted before, some funds may have unequally spaced holding observations because of a mix of monthly and quarterly holdings data. In such cases, we use a slightly modified Newey–West approach to correct for potential biases associated with unequally spaced time-series data. For example, for the TM timing

[8] Elton et al. (2012) examine the timing ability of equity funds. However, because some of their sample funds can take significant positions in bonds, the authors use a multifactor return generating model that includes the excess return on a bond index (the Lehman U.S. Government/Credit index) as the only bond market factor in their analysis.

[9] Specifically, we sort Treasury securities into seven portfolios based on 12-month duration intervals: from 1–12, 13–24, 25–36, 37–48, 49–60, 61–120, and greater than 120 months from the quote date. The matching portfolio returns are then the equally weighted averages of holding period returns of individual Treasury securities held in the portfolio.

measure, multiplying both sides of (4) by a dummy variable, d_t, we have

$$d_t\hat{\beta}_u = \alpha d_t + \gamma d_t Term_{u+h} + d_t \eta_u, \quad t = 1, 2, \ldots, T;\ u \subset \{1, \ldots, T\}, \quad (6)$$

where d_t is equal to 1 if fund holdings are available at the end of month t and 0 otherwise. We then follow the standard procedure to compute the Newey–West t-statistics for $\hat{\gamma}$.

3.2. Statistical Inference: Bootstrapping

Following Kosowski et al. (2006), we adopt a bootstrap approach to derive statistical inference in our analysis. This approach addresses two issues associated with statistical inference based on the cross-sectional distributions of holdings-based timing measures and their t-statistics: violation of the independent and identically distributed assumption across funds and the finite sample property of the test statistics. These issues are especially more of a concern for government bond funds (than for equity funds). First, Treasury securities are relatively homogeneous and their returns are highly correlated. As a result, the portfolio betas are highly correlated and the timing statistics are not independent and identically distributed across funds. Second, our final sample is relatively small, so the finite sample distributions for the cross-sectional statistics, particularly those extreme percentiles, may differ from their asymptotic counterparts.

In each bootstrap, we generate a random sample of cross-sectional timing measures (gammas) under the null hypothesis that no fund has timing ability. Specifically, we randomly resample $Term_{t+h}$ while fixing each fund's $\hat{\beta}_t$, re-estimate each fund's timing coefficient, and obtain a bootstrapped sample of gammas and their pivotal t-statistics. We repeat this procedure for 2,000 times to obtain the distributions of bootstrapped statistics and then calculate the bootstrapped p-values. For example, the distribution of gammas (or t-statistics) for the top timer is constructed as the empirical distribution of the maximum γ (or t-statistic) across the 2,000 bootstraps. Under the null hypothesis of no timing ability, we expect the distributions of bootstrapped statistics to approximate the empirical distributions of timing statistics. If the estimated timing measures or their t-statistics are consistently higher than their corresponding bootstrapped values are; i.e., if the bootstrapped p-values are close to 0, we conclude that fund managers have significantly positive timing ability not purely due to luck. Similarly, if the estimated timing measures or their t-statistics are consistently lower than their corresponding bootstrapped values are; i.e., if the bootstrapped p-values are close to 1, we argue that fund managers have significantly negative timing ability.

4. Empirical Analysis of Market Timing

In this section, we first conduct the baseline holdings-based (unconditional) TM timing tests. We then examine fund characteristics versus market timers in §4.2. In §4.3 we consider conditional timing measures. Section 4.4 conducts robustness checks. Finally, we provide an estimate of the economic value of market timing in §4.5.

4.1. Holdings-Based Treynor–Mazuy Timing Tests

In the baseline analysis, we apply the holdings-based TM timing measures as described in §3.1 to a sample of government bond funds with a minimum of 20 holding observations during the 1997–2006 period.[10] Statistical inference is based on the bootstrap approach described in §3.2.

Table 3 presents the results of our baseline bootstrap analysis. Specifically, the table reports the cross-sectional statistics of holdings-based TM timing measures and their Newey–West t-statistics, together with the bootstrapped p-values, for the forecasting horizons of one, three, and six months, respectively. In particular, we base statistical inference on the bootstrapped p-values for the pivotal t-statistics since the distributions of bootstrapped t-statistics are likely to have fewer problems associated with high variance or survival bias and thus have better statistical properties than the distributions of bootstrapped timing estimates (Kosowski et al. 2006). Indeed, note from the table that the kurtosis of the pivotal t-statistics is much smaller than that of the non-pivotal timing measures, $\hat{\gamma}$, for all horizons.

As shown by the results for the one-month horizon (in panel A), the bootstrapped p-values for the t-statistics of timing measures, $p(t)$, are below 5% for the mean, median, maximum, the 75th, and 99th percentiles and below 10% for the 95th percentile. In particular, the bootstrapped p-value for the average t-statistic across 146 government bond funds is well below 1%, indicating that, on average, government bond fund managers exhibit significantly positive timing ability at the one-month horizon. Note, however, that $p(t)$ is 0.29 for the 90th percentile fund (despite its positive t-statistic). This indicates that we cannot reject the hypothesis that the positive timing ability of the 90th percentile fund manager (ranked by the t-statistics of timing measures) is

[10] For robustness checks, we also perform holdings-based Henriksson–Merton timing tests in Equation (5), redo the baseline analysis using both a smaller sample of 109 funds with at least 30 holding observations for each fund, and examine a bigger sample of 195 funds with at least 10 observations for each fund. We find that in all these cases the results—untabulated but included in an earlier version of the paper—are qualitatively similar to those from the baseline analysis.

Table 3 **Holdings-Based Treynor–Mazuy Timing Tests**

	Min	1%	5%	10%	25%	Mean	Median	75%	90%	95%	99%	Max	Stdev	Skew	Kurto
						Panel A: One-month horizon									
$\hat{\gamma}$	−27.49	−4.60	−2.20	−1.57	−0.58	0.21	0.19	0.77	2.12	3.07	3.93	33.86	3.89	1.99	55.68
$p(\hat{\gamma})$	(0.35)	(0.07)	(0.02)	(0.05)	(0.03)	(0.13)	(0.01)	(0.47)	(0.11)	(0.15)	(0.85)	(0.12)	(0.65)	(0.19)	(0.68)
t	−3.16	−2.57	−1.67	−1.35	−0.56	0.25	0.22	1.05	1.64	2.35	4.05	5.74	1.28	0.57	2.14
$p(t)$	(0.49)	(0.38)	(0.12)	(0.15)	(0.00)	(0.00)	(0.01)	(0.02)	(0.29)	(0.08)	(0.04)	(0.03)	(0.22)	(0.07)	(0.05)
						Panel B: Three-month horizon									
$\hat{\gamma}$	−23.63	−3.90	−2.53	−1.58	−0.62	−0.22	−0.03	0.48	1.12	1.82	2.73	7.45	2.33	−6.87	71.20
$p(\hat{\gamma})$	(0.02)	(0.24)	(0.77)	(0.75)	(0.47)	(0.02)	(0.53)	(0.88)	(0.45)	(0.11)	(0.62)	(0.39)	(1.00)	(0.02)	(0.99)
t	−2.82	−2.64	−1.80	−1.49	−0.86	0.05	−0.04	0.92	1.68	2.10	3.25	3.62	1.26	0.36	−0.10
$p(t)$	(0.26)	(0.39)	(0.12)	(0.21)	(0.30)	(0.14)	(0.48)	(0.23)	(0.32)	(0.41)	(0.31)	(0.49)	(0.55)	(0.20)	(0.61)
						Panel C: Six-month horizon									
$\hat{\gamma}$	−4.27	−3.76	−2.44	−1.59	−0.50	0.48	0.13	0.74	1.28	1.71	7.86	62.43	5.40	10.66	122.71
$p(\hat{\gamma})$	(0.00)	(0.46)	(0.98)	(1.00)	(0.78)	(0.00)	(0.27)	(0.11)	(0.34)	(0.28)	(0.03)	(0.00)	(0.73)	(0.00)	(0.75)
t	−3.46	−3.17	−2.18	−1.74	−0.96	0.25	0.27	1.22	2.37	2.75	3.77	5.28	1.57	0.17	−0.08
$p(t)$	(0.53)	(0.73)	(0.84)	(0.90)	(0.98)	(0.37)	(0.20)	(0.13)	(0.04)	(0.32)	(0.46)	(0.19)	(0.05)	(0.54)	(0.74)

Notes. This table reports the cross-sectional distributions of the holdings-based Treynor–Mazuy timing measures ($\hat{\gamma}$) and their Newey–West t-statistics (t) for government bond funds with a minimum of 20 holding observations during the 1997–2006 period. Treasury security betas are estimated using the past one-year daily returns. Panels A–C report the results for the forecasting horizons of one, three, and six months, respectively. The bootstrapped p-values for the timing measures and their Newey–West t-statistics, $p(\hat{\gamma})$ and $p(t)$, are shown in the parentheses underneath, respectively. "Stdev," "Skew," and "Kurto" denote the cross-sectional standard deviation, skewness, and excess kurtosis, respectively.

purely due to luck. On the other hand, panels B and C show that managers generally do not exhibit positive timing ability at the three- and six-month horizons. The result suggests that fund managers in our sample have difficulty timing the market over longer horizons.

Note that in panel A of Table 3, the standard p-values and their bootstrapped counterparts can be very different for some funds. The reason is that, as alluded in Kosowski et al. (2006), the empirical cross-sectional distribution of gammas can be nonnormal even if individual fund gammas are normally distributed. To further illustrate this point, Figure 1 plots the kernel density functions of the distributions of bootstrapped t-statistics for funds at various points in the cross section, as well as the actual t-statistics of timing measures. For example, panel A of the figure reports the actual t-statistic versus the distribution of the bootstrapped t-statistics of the top-ranked fund. We can see that the actual top-fund t-statistic of 5.74 (the dashed line in panel A) lies well within the right-tail rejection region of the bootstrapped distribution, indicating that the significantly positive timing ability of the fund (the Vanguard Intermediate Term U.S. Treasury Fund) is not purely due to luck. Similarly, at the mean, median, the 75th, 95th, and 99th percentiles, the bootstrapped distributions reject the null hypothesis of no significantly positive timing ability for managers. In general, the results suggest that government bond fund managers, on average, exhibit significantly positive timing ability at the one-month horizon and that the timing ability of top fund managers cannot simply be attributed to luck.

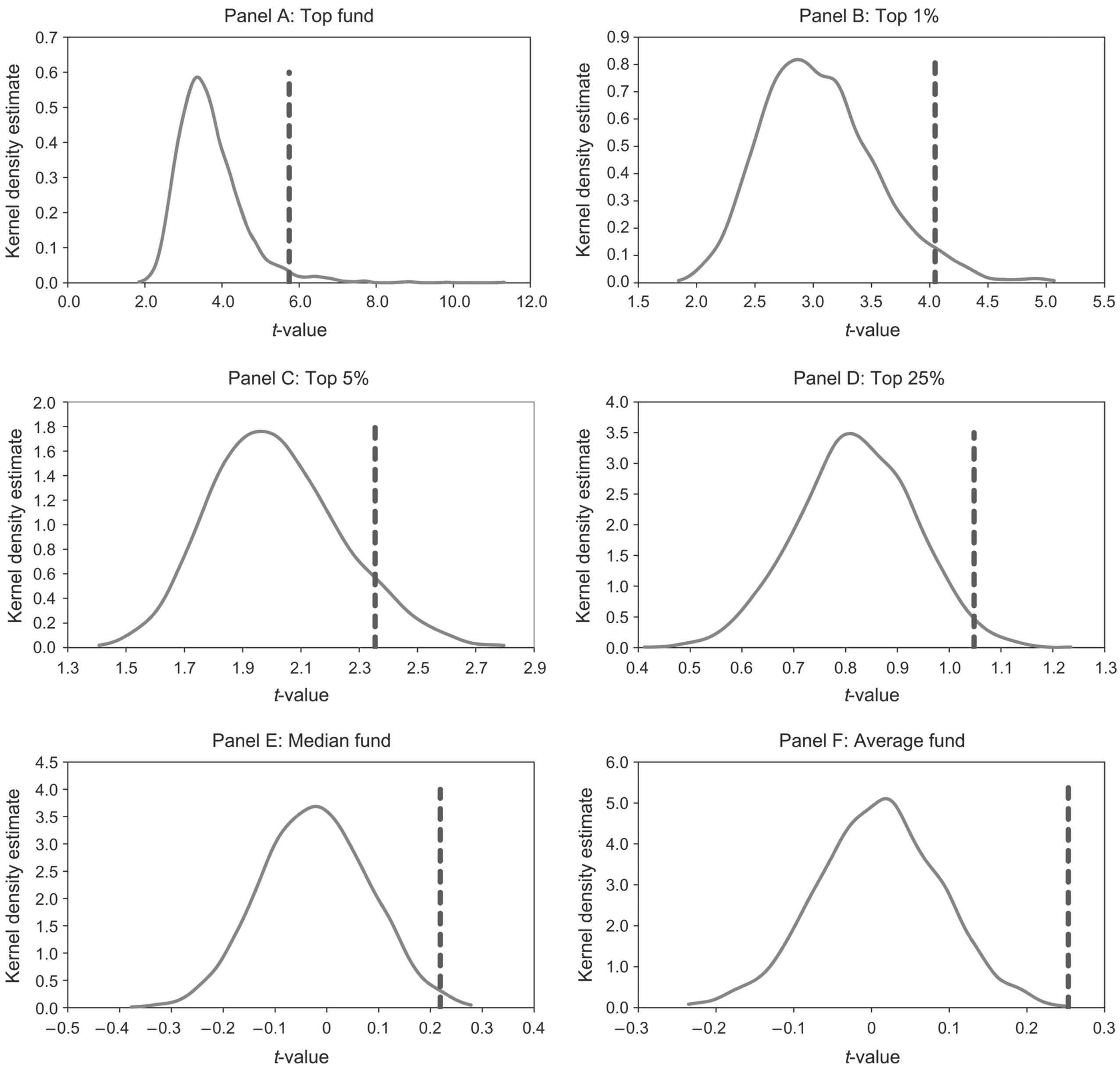
4.2. Fund Characteristics and Market Timing

Although we document evidence of positive (unconditional) timing ability for the average fund, we also observe a wide variation of timing performance across funds. In this subsection we examine potential links between fund characteristics and market timing.

We first perform holdings-based timing tests for fund subgroups with different investment objectives to see if market timing ability is associated with fund objectives. Recall from §2 that we can divide the full sample into either three subgroups (GGN, GIM, and GSM) under the CRSP–S&P classification system or two subsamples (general government and Treasury bond funds) based on Morningstar prospectus objectives. Untabulated results indicate that GSM funds have the most significant market timing ability, followed by GIM and GGN funds, and that Treasury bond fund managers (specializing in Treasury securities) exhibit superior positive timing ability compared with general government bond fund managers.

Next we consider seven other important fund characteristics. The first three are the characteristics of government bond funds per se, i.e., Morningstar rating, expense ratio, and turnover rate. The next four are related to Treasury security portfolios of government bond funds, namely the Sharpe ratio, TNAs, value-weighted duration, and percentage of assets held in Treasuries. We rank 146 government bond funds into five quintile portfolios by their Newey–West t-statistics of holdings-based TM timing measures from Table 3. The fund characteristics are first averaged over time for each fund and then averaged

Figure 1 **Actual vs. Bootstrapped t-Statistics of Holdings-Based Treynor–Mazuy Timing Measures for Individual Funds**



Notes. This figure plots kernel density estimates of the bootstrapped t-statistics of holdings-based Treynor and Mazuy (1966) timing measures (solid lines), as well as the actual t-statistics of timing measures (dashed vertical lines), for individual funds at various percentile points in the cross section. Panels A–F show the results for the marginal fund at the maximum, top 1%, top 5%, top 25%, median, and mean of the distribution, respectively. The x- and y-axis show the t-statistics and the kernel density estimates, respectively. The result obtained is based on a final sample of 146 government bond funds with a minimum of 20 holding observations during the 1997–2006 period. Treasury security betas are estimated using the past one-year daily returns. The forecasting horizon used in the estimation is one month.

across funds within each quintile. Results (untabulated) indicate that funds with the most significantly positive timing ability have higher Morningstar ratings and Sharpe ratios, lower expense ratios, larger TNAs, and higher concentrations of holdings of Treasury securities than do those with the most perverse timing ability.

Finally, on a related point, we investigate whether these seven fund characteristics contain any information about market timing ability. Specifically, based on time-series averages of fund characteristics, we rank funds into five quintiles by each characteristic and then examine the timing ability of managers within each quintile. Untabulated results indicate that funds with higher Morningstar ratings, lower expense ratios, higher Sharpe ratios, or higher concentrations of holdings of Treasury securities are more likely to successfully time the market. Given that the Morningstar rating and Sharpe ratio are measures of a fund's risk-adjusted returns, the results provide supporting evidence that government bond fund managers enhance performance by timing the market.

4.3. Conditional Market Timing Ability

The evidence so far on the timing ability of government bond fund managers is based on unconditional timing measures. In this subsection, we follow Ferson and Schadt (1996) and investigate the conditional timing ability of funds in our sample.

4.3.1. Impact of Public Information. We first examine whether funds in our sample change beta in reaction to public information. Empirical evidence has documented that government bond returns are predictable (e.g., Keim and Stambaugh 1986, Fama and Bliss 1987, Ilmanen 1995, Cochrane and Piazzesi 2005, Duffee 2011). Furthermore, several recent studies show that the price movements in the Treasury market are mainly driven by macroeconomic news (e.g., Fleming and Remolona 1999, Balduzzi et al. 2001, Andersen et al. 2007). As such, it is natural to ask whether government bond fund managers rely on macroeconomic predictors in making asset allocation decisions.

To address this question, we consider the following time-series regression for each fund i:

$$\hat{\beta}_{it} = a_i^m + b_i^m M_{t-1} + e_{it}, \tag{7}$$

where $\hat{\beta}_{it}$ is the Treasury security portfolio's beta as of month t and M_{t-1} is the vector of one-month lagged predictors to be used. Once M is specified, we can conduct statistical inference using the parametric bootstrap approach described earlier.

We consider two sets of predictors in the analysis that follows. The first one includes four economic variables used in Ilmanen (1995): the term spread (*TERMSP*), the difference between 10-year and three-month Treasury yields; the real bond yield (*REALYLD*), the difference between 10-year bond yield and the one-month lagged year-on-year inflation rate; inverse relative wealth (*INVRELW*), a ratio of past to current real wealth;[11] and bond beta (*BETA*), the regression coefficient of excess long-term government bond returns against excess stock market returns. The first two variables proxy for the overall expected bond risk premium. The last two represent the time-varying risk aversion and time-varying risk, respectively.

Panel A of Table 4 reports the cross-sectional distributions of the Newey–West t-statistics and their bootstrapped p-values for the estimated coefficients on the above four economic variables. The results show that only the mean and median coefficient estimates on the historical bond beta are significantly positive, indicating that managers are most concerned about the time-varying risk of the bond market when making asset allocation decisions. One possible explanation is that the historical bond beta in some way reflects the impact of macroeconomic conditions on the bond market, given that macroeconomic variables usually affect both stock and bond markets (Fama and French 1993).

The first set of predictors is used to mainly capture the information about the bond market contained in the past bond yields. However, the empirical evidence that (Treasury) bond prices are largely driven by macroeconomic news implies that the public information set in bond markets goes beyond what is contained in historical bond yields. As such, we next examine whether fund managers use information contained in macroeconomic news announcements and, if they do, what news items are deemed as important by the managers.[12]

Namely, we use the set of 15 macroeconomic news variables and their surprises described in §2.3 as the second set of beta predictors. To proceed, let M_{t-1} be the surprise of a given macroeconomic news variable at $t-1$. We then estimate Equation (7) for each fund and obtain the cross-sectional distributions of Newey–West t-statistics and their bootstrapped p-values for estimated coefficients on the surprise of this particular news variable. And we repeat this exercise for the remaining 14 news variables.

Columns (4)–(7) of Table 2 report the mean and median t-statistics and their associated bootstrapped p-values for estimated coefficients on all these 15 macroeconomic news variables. As indicated in the table, the average coefficients on the index of leading indicators (LEI) and factory orders (FO) are significantly positive, whereas the average coefficients on nonfarm payrolls (NFP) and personal consumption (PCE) have negative t-values with bootstrapped p-values close to one and thus are significantly negative, consistent with our intuition.[13] These results suggest that government bond fund managers make their asset allocation decisions in response to important news announcements such as LEI, FO, NFP, and PCE.

[11] $INVRELW_t = \sum_{k=1}^{36} 0.1 * 0.9^{k-1} * W_{t-k}/W_t$, where W_t is the real level of stock market (the value-weighted CRSP market index deflated by the consumer price index) at time t, and the numerator is the exponentially weighted average of real stock market levels up to $t-1$.

[12] We thank an anonymous referee for this insight and for suggesting this exercise.

[13] See §3.2 for the interpretation of the significance in case of a bootstrapped p-value close to one. McQueen and Roley (1993) and Balduzzi et al. (2001) document that surprises in nonfarm payrolls have a negative impact on Treasury bond returns.

Table 4 Market Timing and Public Information

	Min	1%	5%	10%	25%	Mean	Median	75%	90%	95%	99%	Max	Stdev	Skew	Kurto
Panel A: Bottom-up betas and macroeconomic predictors															
b_{TERMSP}															
t	−4.10	−3.92	−2.11	−1.42	−0.36	1.22	1.07	2.31	4.80	5.54	8.73	8.92	2.38	0.62	0.69
$p(t)$	(0.80)	(0.95)	(0.82)	(0.77)	(0.67)	(0.52)	(0.37)	(0.61)	(0.11)	(0.63)	(0.01)	(0.10)	(0.03)	(0.34)	(0.15)
$b_{REALYLD}$															
t	−9.55	−9.00	−6.43	−5.10	−3.15	−2.18	−2.01	−0.76	0.17	1.13	2.71	2.95	2.20	−0.61	1.05
$p(t)$	(0.75)	(0.96)	(0.99)	(0.92)	(0.69)	(0.78)	(0.77)	(0.59)	(0.64)	(0.17)	(0.14)	(0.42)	(0.02)	(0.53)	(0.38)
$b_{INVRELW}$															
t	−6.43	−4.99	−3.78	−2.67	−1.09	0.08	0.15	1.40	2.66	3.58	5.13	5.14	2.09	−0.13	0.43
$p(t)$	(0.83)	(0.66)	(0.88)	(0.57)	(0.26)	(0.27)	(0.30)	(0.13)	(0.04)	(0.15)	(0.58)	(0.91)	(0.10)	(0.90)	(0.88)
b_{BETA}															
t	−2.23	−1.69	−1.29	−0.99	−0.32	0.51	0.53	1.30	1.85	2.17	3.17	3.50	1.10	0.07	−0.23
$p(t)$	(0.07)	(0.02)	(0.08)	(0.11)	(0.10)	(0.04)	(0.05)	(0.03)	(0.04)	(0.07)	(0.14)	(0.27)	(0.45)	(0.60)	(0.89)
Panel B: Market timing controlling for macroeconomic predictors															
b_{TERMSP}															
t	−4.10	−3.75	−2.12	−1.35	−0.22	1.27	1.12	2.32	4.66	5.48	8.57	9.08	2.31	0.63	0.80
$p(t)$	(0.77)	(0.89)	(0.81)	(0.60)	(0.41)	(0.38)	(0.29)	(0.56)	(0.20)	(0.62)	(0.01)	(0.09)	(0.14)	(0.31)	(0.12)
$b_{REALYLD}$															
t	−9.40	−8.77	−6.31	−5.02	−3.13	−2.12	−1.93	−0.70	0.23	1.19	2.77	2.97	2.21	−0.60	0.91
$p(t)$	(0.72)	(0.93)	(0.97)	(0.88)	(0.67)	(0.71)	(0.70)	(0.52)	(0.57)	(0.14)	(0.14)	(0.42)	(0.02)	(0.49)	(0.49)
$b_{INVRELW}$															
t	−6.36	−4.91	−3.55	−2.70	−1.08	0.11	0.18	1.40	2.64	3.64	5.21	5.35	2.10	−0.07	0.37
$p(t)$	(0.82)	(0.63)	(0.75)	(0.63)	(0.26)	(0.25)	(0.28)	(0.12)	(0.06)	(0.11)	(0.53)	(0.85)	(0.09)	(0.85)	(0.89)
b_{BETA}															
t	−2.26	−1.71	−1.33	−1.04	−0.33	0.52	0.52	1.30	1.87	2.18	3.17	3.44	1.11	0.05	−0.25
$p(t)$	(0.07)	(0.02)	(0.09)	(0.14)	(0.10)	(0.04)	(0.05)	(0.03)	(0.04)	(0.07)	(0.14)	(0.30)	(0.45)	(0.62)	(0.91)
γ															
t	−3.38	−2.12	−1.74	−1.54	−0.89	−0.24	−0.32	0.45	1.07	1.39	2.19	3.18	1.00	0.27	0.51
$p(t)$	(0.72)	(0.23)	(0.56)	(0.73)	(0.70)	(0.71)	(0.73)	(0.68)	(0.72)	(0.78)	(0.77)	(0.42)	(0.72)	(0.26)	(0.43)
Panel C: Market timing controlling for macroeconomic news releases															
b_{LEI}															
t	−3.96	−2.84	−1.37	−0.86	0.23	1.16	1.13	2.23	3.09	3.65	4.59	4.67	1.55	−0.31	0.25
$p(t)$	(0.77)	(0.50)	(0.08)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.05)	(0.94)	(0.41)
b_{NFP}															
t	−3.41	−3.27	−2.62	−2.26	−1.82	−0.89	−0.83	−0.17	0.59	1.18	2.08	2.13	1.14	0.28	−0.23
$p(t)$	(0.81)	(0.93)	(0.97)	(0.98)	(0.99)	(0.98)	(0.97)	(0.97)	(0.95)	(0.82)	(0.65)	(0.82)	(0.15)	(0.15)	(0.81)
b_{FO}															
t	−2.64	−1.91	−1.38	−0.98	−0.40	0.35	0.29	1.00	1.64	2.05	3.82	3.94	1.10	0.33	0.68
$p(t)$	(0.48)	(0.21)	(0.36)	(0.30)	(0.41)	(0.39)	(0.49)	(0.45)	(0.39)	(0.36)	(0.05)	(0.14)	(0.39)	(0.08)	(0.20)
b_{PCE}															
t	−5.48	−3.06	−1.80	−1.59	−0.91	−0.48	−0.44	0.10	0.49	0.76	1.34	1.44	0.91	−1.38	5.94
$p(t)$	(1.00)	(0.90)	(0.74)	(0.85)	(0.81)	(0.93)	(0.90)	(0.95)	(0.99)	(0.99)	(0.99)	(0.99)	(0.82)	(1.00)	(0.00)
γ															
t	−2.63	−2.47	−2.19	−1.91	−1.26	−0.60	−0.64	0.20	0.75	1.02	1.41	1.86	0.98	0.18	−0.59
$p(t)$	(0.47)	(0.63)	(0.91)	(0.93)	(0.92)	(0.93)	(0.93)	(0.87)	(0.93)	(0.95)	(0.99)	(0.95)	(0.58)	(0.32)	(0.97)

Notes. Panel A reports the cross-sectional distributions of the Newey–West t-statistics (t) and their bootstrapped p-values ($p(t)$) for the coefficient estimates from regression (7) against four one-month lagged macroeconomic variables jointly: term spread (*TERMSP*), real bond yield (*REALYLD*), inverse relative wealth (*INVRELW*), and historical bond beta (*BETA*). Panel B shows results from (8) that augments (7) with the month-$(t+1)$ term premium ($Term_{t+1}$) whose coefficient γ is the holdings-based timing measure. Panel C presents results from (8) against $Term_{t+1}$ and one-month lagged surprises in the index of leading indicators (LEI), nonfarm payrolls (NFP), factory orders (FO), and personal consumption (PCE). In both (7) and (8) the dependent variable $\hat{\beta}_{it}$ is the month-t Treasury security portfolio beta of fund i. "Stdev," "Skew," and "Kurto" denote the cross-sectional standard deviation, skewness, and excess kurtosis, respectively.

Last, we examine the predictive power of news announcements over horizons beyond one month. The results (not tabulated) indicate that some news announcements such as NFP show predictive power at the three- and six-month horizons. As such, the combination of this finding and the baseline results seems to suggest that fund managers fail to capture the predictive power beyond the one-month horizon.

4.3.2. Role of Private Information. In addition to public information, private information (e.g., the price discovery through order flow imbalances) might also play a role in the Treasury market (Brandt and

Kavajecz 2005). Below we examine whether fund managers use any private information in their market timing decisions.

To answer this question, we consider the following specification that augments regression (7) with the term premium, $Term_{t+1}$:

$$\hat{\beta}_{it} = a_i^m + b_i^m M_{t-1} + \gamma_i Term_{t+1} + e_{it}. \tag{8}$$

We test for significance of $\hat{\gamma}$ (the holdings-based timing measure) using the parametric bootstrap approach as before. If $\hat{\gamma}$ remains significantly positive after controlling for M_{t-1} (the public information available at $t-1$), then fund managers might have information about long-term government bond returns beyond what is contained in the specification of M_{t-1}; otherwise, there is no evidence indicating that fund managers rely on private information in market timing.

Panel B of Table 4 reports the estimation results, controlling for the four macro predictors used earlier in §4.3.1—namely, with $M = (TERMSP, REALYLD, INVRELW, BETA)$. Note that the average timing measure becomes insignificant, conditional on these four macroeconomic predictors. We obtain similar results when we redo the analysis using $M =$ (LEI, FO, NFP, PCE), the set of the four news variables whose surprises are found earlier to contain information about bottom-up betas. Indeed, as indicated in panel C of the table, the timing measure $\hat{\gamma}$ is actually marginally negative for both the median and average funds. Overall, these results suggest that private information plays little role in government bond fund managers' asset allocation decisions. Although this finding is different from what is documented in equity mutual funds,[14] it is not surprising; as mentioned earlier, it is known that public information plays a major role in the Treasury market.

Note that these findings are different from the pattern for equity funds documented in Ferson and Schadt (1996), who find negative timing unconditionally and neutral conditionally. They attribute the negative timing partially to the patterns of fund flows, which are explored in Ferson and Warther (1996) and confirmed by Ferson and Qian (2004). In our case, fund flows are controlled for because our results are based on holdings. As such, we provide perhaps the first example where fund managers indeed look "worse" under a conditional measure and should not be given credits for using public information to time the market.

The conditional performance analysis done so far is based on the full sample. We next repeat the above analysis using the subsample of Treasury bond funds (funds specializing in Treasury securities) to understand the source of their better unconditional timing ability than general government bond funds' (as documented earlier in §4.2). Untabulated results indicate that $\hat{\gamma}$ is insignificant for the average Treasury bond fund either. This implies that Treasury bond fund managers do not possess timing ability beyond public information but they do have better ability of interpreting public information than do general government bond fund managers. These results also provide more direct evidence supporting the finding of Green (2004) that Treasury market participants have different ability to process the public information release.

To summarize, we find evidence that government bond fund managers adjust their portfolio betas in response to public information (such as the time-varying risk of bond market and macroeconomic news announcements including nonfarm payrolls, factory orders, and personal consumption) and, importantly, that these specific public information variables can explain the average manager's timing ability. Then a natural question is why investors do not replicate this timing strategy on their own. One possible explanation is that certain fund managers have better ability of interpreting public information than investors, as alluded above. Another one is that transaction costs involved in building and managing a fixed-income portfolio are too high for typical non-institutional investors.

4.4. Robustness Checks

This subsection conducts a number of robustness checks on the main finding from the unconditional tests done in §4.1. We first perform return-based timing tests for comparison. We then consider the estimation of individual Treasury security betas in §4.4.2 and control for passive changes in fund betas in §4.4.3. We next examine rollover strategies and their potential impact on timing in §4.4.4. Last, we address the issue of window dressing.

4.4.1. Return-Based Timing Tests. As documented in Chen et al. (2010), taxable bond funds appear to have negative market timing ability under the standard return-based timing measures (without controlling for non-timing-related sources of nonlinearity). Although it is not feasible to directly compare our results to theirs because of different samples, sample periods, and models, we perform standard return-based TM timing tests for our sample funds in this subsection. Namely, we consider the following (unconditional) model:

$$r_{it} = a_i^r + b_i^r \, Term_t + \gamma_i \, Term_t^2 + e_{it}, \tag{9}$$

where r_{it} is the month-t excess return of fund i, $Term_t$ is the term premium for month t as defined in

[14] Empirical evidence suggests that equity fund managers use information about stock market returns not fully captured by those predetermined macroeconomic variables. See Jiang et al. (2007), Mamaysky et al. (2008), Elton et al. (2012), among others.

Table 5 **Return-Based Treynor–Mazuy Timing Tests**

	Min	1%	5%	10%	25%	Mean	Median	75%	90%	95%	99%	Max	Stdev	Skew	Kurto
							Panel A: Based on actual fund returns								
t	−3.70	−2.98	−1.59	−1.24	−0.36	0.37	0.37	1.32	1.87	2.22	3.02	3.28	1.25	−0.39	0.33
$p(t)$	(0.94)	(0.95)	(0.88)	(0.97)	(0.96)	(0.99)	(0.98)	(0.79)	(0.98)	(0.99)	(1.00)	(1.00)	(0.93)	(1.00)	(0.34)
							Panel B: Based on hypothetical portfolio returns								
t	−2.53	−1.84	−1.17	−0.59	0.52	1.50	1.56	2.49	3.41	3.84	4.47	6.21	1.51	−0.16	0.09
$p(t)$	(0.15)	(0.04)	(0.08)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.49)	(0.17)	(0.27)	(0.75)	(0.84)

Notes. This table reports the cross-sectional distributions of the Newey–West t-statistics (t) and their bootstrapped p-values ($p(t)$) for the return-based Treynor–Mazuy timing measures for a sample of government bond funds during the 1997–2006 period. Panels A and B show the results based on actual fund returns and hypothetical portfolio returns constructed from Treasury security holdings of government bond funds, respectively. "Stdev," "Skew," and "Kurto" denote the cross-sectional standard deviation, skewness, and excess kurtosis, respectively.

Equation (2), and γ_i is the return-based timing measure of fund i.[15]

We base the statistical inference of return-based timing tests on a parametric bootstrap approach as well. To proceed, we calculate the cross-sectional statistics of the bootstrapped return-based timing measures and their Newey–West t-statistics. We repeat this procedure 2,000 times to obtain the distributions of bootstrapped statistics and then calculate the bootstrapped p-values.

Panel A of Table 5 reports the cross-sectional distributions of the Newey–West t-statistics (t) and their bootstrapped p-values ($p(t)$) for the return-based TM timing measure. The results show that under the standard return-based timing measure, government bond funds in our sample on average do not exhibit significantly positive timing ability. For example, the mean (median) of t-statistics is 0.37 (0.37) with a bootstrapped p-value of 0.99 (0.98), where the t-value is positive but insignificant.

We then repeat tests in Equation (9) using hypothetical portfolio returns constructed from Treasury security holdings of government bond funds. We assume that funds hold the same Treasury securities until the next portfolio report date. For those securities with the maturity date between the two portfolio report dates, we assume that the amount received on the maturity date will be reinvested at the market risk-free rate until the next portfolio report date. Test results, reported in panel B of Table 5, show that funds exhibit significantly positive timing ability in the Treasury sector. For instance, the t-statistics are significantly positive at the mean, median, the 75th, 90th, and 95th percentiles. This suggests that despite the limited frequency, portfolio holdings provide valuable information to test for the timing ability of funds.

4.4.2. Estimation of Individual Security Betas. Consider first individual Treasury security betas, which are estimated using past one-year daily returns in the baseline analysis. We rerun holdings-based TM timing tests using two alternative estimates of these betas—namely, one based on daily returns over the past three months and the other on monthly returns over the past five years (if an individual Treasury security has fewer than 12 monthly return observations, we set its beta to be the matching portfolio beta as defined before). We find that the evidence of positive timing is robust to the beta estimates used.

Next, we consider benchmark-adjusted individual Treasury security betas to minimize the potential passive timing effects and reduce cross-sectional heteroskedasiticity. Specifically, given security i's beta estimate, $\hat{b}_i$, and its duration-matched benchmark portfolio beta, $\hat{b}_i^b$, the security's adjusted beta is given by $\hat{b}_i^a = \hat{b}_i - \hat{b}_i^b$. We then construct portfolio betas with these adjusted individual betas using (3) and redo the unconditional analysis. Untabulated results indicate that the main conclusion from the baseline analysis still holds after controlling for duration-based matching portfolio betas. For example, the mean (median) of t-statistics is 0.47 (0.51) with a bootstrapped p-value of 0.00 (0.00) for the one-month forecasting horizon.

4.4.3. Controlling for Passive Changes in Fund Betas. In the baseline analysis, we construct holdings-based timing measures as the covariance between time-varying fund betas at the beginning of a holding period and the holding period term premium. However, the time variation in fund beta levels might be solely driven by nonproportional changes in security prices rather than by active trading activities of fund managers. Even though this is not as much of a concern in the Treasury market as in the stock market (given that Treasury security prices are much less volatile than stock prices), we separate active trading

[15] We note that the term premium might not fully explain the returns of government bond funds because these funds on average hold around 49% of their assets in Treasury securities. However, the average adjusted R^2 of this one-factor model is 76% at the fund level, indicating that the bond market return is the most important risk factor. In addition, adding more factors tends to make it more difficult to find evidence of positive timing ability empirically (see, e.g., Elton et al. 2012).

Table 6 Holdings-Based Treynor–Mazuy Timing Tests: Active Changes of Portfolio Betas

	Min	1%	5%	10%	25%	Mean	Median	75%	90%	95%	99%	Max	Stdev	Skew	Kurto
							Panel A: One-month horizon								
t	−3.45	−2.79	−2.19	−1.80	−1.01	0.07	0.17	0.97	1.67	2.41	3.73	4.47	1.43	0.15	0.05
$p(t)$	(0.06)	(0.13)	(0.63)	(0.88)	(0.83)	(0.08)	(0.01)	(0.08)	(0.21)	(0.04)	(0.15)	(0.64)	(0.23)	(0.41)	(0.96)
							Panel B: Three-month horizon								
t	−3.10	−3.02	−2.29	−1.84	−1.09	−0.16	−0.08	0.64	1.28	1.78	3.27	3.93	1.23	0.08	0.27
$p(t)$	(0.13)	(0.34)	(0.66)	(0.82)	(0.88)	(0.83)	(0.47)	(0.94)	(0.96)	(0.88)	(0.45)	(0.51)	(0.89)	(0.45)	(0.67)
							Panel C: Six-month horizon								
t	−4.00	−3.33	−2.12	−1.64	−1.06	−0.10	−0.16	0.83	1.43	1.91	2.65	7.74	1.40	0.94	5.54
$p(t)$	(0.45)	(0.49)	(0.27)	(0.27)	(0.68)	(0.48)	(0.46)	(0.38)	(0.85)	(0.85)	(0.92)	(0.08)	(0.47)	(0.08)	(0.10)

Notes. This table reports the cross-sectional distributions of the Newey–West t-statistics (t) and their bootstrapped p-values ($p(t)$) for the holdings-based Treynor–Mazuy timing measures based on the active changes of portfolio betas over the past three months. We include government bond funds with a minimum of 10 observations of active portfolio beta changes during the 1997–2006 period in the sample. Treasury security betas are estimated using the past one-year daily returns. Panels A–C report the results for the forecasting horizons of one, three, and six months, respectively. "Stdev," "Skew," and "Kurto" denote the cross-sectional standard deviation, skewness, and excess kurtosis, respectively.

from passive portfolio weight changes and rerun the unconditional analysis as a robustness check.

Specifically, we calculate the fund beta changes induced by active trading of fund managers from month $t-h$ to month t, $\Delta\hat{\beta}_t$, as follows:

$$\Delta\hat{\beta}_t = \hat{\beta}_t - \hat{\beta}_t^{t-h} = \sum_{i=1}^{N_t} w_{it}\hat{b}_{it} - \sum_{i=1}^{N_{t-h}} w_{it}^{t-h}\hat{b}_{it}, \qquad (10)$$

where $\hat{\beta}_t$ is the fund portfolio beta at the end of month t; $\hat{\beta}_t^{t-h}$ is the hypothetical fund portfolio beta at the end of month t if a fund passively holds all the (Treasury) securities in its portfolio from h months ago; N_t and N_{t-h} are the number of securities held by the fund at the end of month t and $t-h$, respectively; $\hat{b}_{it}$ is the beta estimate of security i at the end of month t; w_{it} is the portfolio weight of security i at the end of month t; and w_{it}^{t-h} is the hypothetical portfolio weight of security i at the end of month t assuming that a fund passively holds its Treasury portfolio from month $t-h$. In particular, w_{it}^{t-h} is defined as follows:

$$w_{it}^{t-h} = \frac{F_{it-h} * P_{it}}{\sum_{i=1}^{N_{t-h}} F_{it-h} * P_{it}}, \qquad (11)$$

where F_{it-h} is the face value of Treasury security i at the end of month $t-h$, and P_{it} is the price of Treasury security i at the end of month t. For a Treasury security with maturity date between months $t-h$ and t, we set its price to be one and its beta to be that of the matching portfolio with average duration of no more than one year as of month t. We measure active trading of fund managers over the past three months ($h = 3$).[16]

We can see that the results based on the active changes of portfolio betas, reported in Table 6, are in general consistent with those based on the level of portfolio betas, despite that the power of the tests based on $\Delta\hat{\beta}_t$ is limited by the number of available observations of portfolio betas' active changes. For example, at the one-month forecasting horizon, the mean (median) of t-statistics is significantly positive with a bootstrapped p-value of 0.08 (0.01).

4.4.4. Rollover Strategies. As mentioned earlier, unlike equity funds or some other bond funds, government bond funds face required rollover of positions in Treasuries. These funds may employ a passive strategy where managers simply roll over their positions around the auction time of new bonds; that is, managers sell their positions in an existing bond right before it is switched from on-the-run to off-the-run and then buy the new on-the-run bond of similar maturity. Because excessive selling before (buying after) auction may depress (inflate) the bond price, such a passive strategy may induce a pattern as "sell low, buy high" or a negative artificial timing.[17] In this subsection, we first try to understand to what extent funds roll over their positions in Treasury securities by utilizing data on fund holdings. We then examine the potential impact of rollover on fund timing ability.

To proceed, for each fund, we obtain first the information on both its buying of Treasury bonds

[16] We keep only those observations where the changes of portfolio holdings over the past three-month period are available. Because of the uneven nature of holding frequency, the average number of holding change observations is smaller. Therefore, we require a fund to have at least 10 observations of three-month portfolio changes to be included in the analysis. The final sample includes 154 funds with an average of 26 three-month portfolio changes.

[17] We thank an anonymous referee for pointing out this passive strategy as an important aspect of government bond fund management and the implication for potential negative artificial timing.

(“initiating-buy”) and its selling of existing bonds (“terminating-sale”) on each date when fund holdings are reported. Next, we calculate the average time-to-maturity (TTM) separately for both initiating buys and terminating sales, on each reporting date, for the given fund. We consider both equally weighted and value-weighted averages in the analysis.

Panel C of Table 1 reports summary statistics on the average and median TTMs, by fund objectives, at both the fund and Treasury portfolio levels, and for both initiating-buys and terminating-sales, under either equally or value-weighted average scheme. We first note that as expected, the average portfolio TTM is longer than that at the fund level except for GSM. Also note that the difference between the equally weighted average TTM and the value-weighted one is rather small for either initiating buys or terminating sales; although the difference between the equally weighted median TTM and the value-weighted one is noticeable, the pattern across fund objectives is similar regardless of the average scheme used. As a result, in the balance of this subsection we focus our discussion on the results for the equally weighted average TTMs and refer to them simply as the average TTM, unless otherwise noted.

We observe from the table that the average TTM of initiating-buys is at least one year longer than that of terminating-sales, regardless of fund objectives. This difference may be related to active trading, duration matching, or passive rollover. A smaller difference should indicate a more likely use of rollover. Similarly, the higher the ratio of the average TTM of initiating-buys to that of terminating-sales, the less likely the use of rollover (both active and passive) and thus the impact of artificial negative timing. This implies that this TTM ratio and the timing measure are more likely positively correlated. Results reported in panel C support this intuition. To see this, we calculate the TTM ratio first, which equals 1.11 (= 11.29/10.13), 1.26, and 1.48 for GGN, GIM, and GSM funds, respectively; and 1.18 and 1.27 for general government bond (GN) and Treasury bond (TB) funds, respectively. Note that this ranking of fund categories by the TTM ratio is consistent with the evidence reported earlier on the association between fund timing measures and fund objectives in §4.2. To explore directly the potential link between the TTM ratio and timing ability, we rank all 146 sample funds into five quintiles by the absolute difference between the TTM of initiating-buys and that of terminating-sales and then examine the timing ability within each quintile. Results (not tabulated) are largely consistent with those based on the fund category-average TTM—namely, funds with larger differences in TTM are in general more likely to better time the bond market return than are funds with lower differences in TTM.

Next, we compare the average TTM of initiating-buys and that of terminating-sales at the individual fund level. Untabulated results from t-tests indicate that out of 146 sample funds, 50 funds (30 GN and 20 TB funds) show statistically significant differences in the average TTM at the 10% level. As such, that more than half of TB funds have such significant differences partially explains why they have a higher TTM ratio (1.27) than do GN funds (1.18). Out of the remaining 96 funds that exhibit no statistical differences in the average TTM, 28.2% of initiating-buy bond positions are within three months of issue dates of purchased bonds and only 2.8% of terminating-sale positions are within three months of maturity dates of bonds sold. These results seem to suggest that the majority of government bond fund managers indeed roll over their positions in Treasuries but not necessarily in a passive manner with the negative artificial timing effect. One possible reason is that liquidity concern may not be an issue for many government bond funds, at least when they roll over their positions in Treasuries.

Last, as another, albeit indirect, way to address the issue of negative artificial timing, we consider government bond index funds—that likely roll over their positions passively and that presumably do not engage in market timing. Specifically, we examine the timing ability of a sample of 12 government bond index funds with a minimum of 20 portfolio holdings over the 1997–2006 period. Results (not tabulated) indicate that index funds have neutral market timing ability. This suggests that the impact of negative artificial timing is not significant here unless it is neutralized by some other passive strategy used by index funds.

Overall, we find evidence suggesting that the majority of sample funds roll over their positions in Treasuries and do so passively in fewer than one third of such activities. We also find that funds doing more rollover tend to show lower timing ability, not inconsistent with the notion of negative artificial timing. Treasury bond funds roll over less often than general bond funds perhaps because the former have better ability of interpreting public information than the latter.

4.4.5. Window Dressing. A number of recent studies have examined the issue of window dressing by institutional investors. For instance, those that focus on mutual funds include Chevalier and Ellison (1997), Musto (2002), Morey and O’Neal (2006), Elton et al. (2010), Agarwal et al. (2013), among others. Several studies examine other types of institutional investors, such as Lakonishok et al. (1991), Sias and Starks (1997), and He et al. (2004). In this subsection, we examine if the unconditional timing ability of government bond funds documented earlier is related to window dressing.

As pointed out and shown in Elton et al. (2010), it is less likely for funds to window dress if they disclose their holdings more frequently. As such, we divide our sample funds into two subsamples based on the average holding reporting frequency over the full sample period: one with average frequency of quarterly or below (the "low-frequency" group) and the other with an average of higher than quarterly frequency (the "high-frequency" group); we then examine whether these two groups of funds have significantly different timing ability.[18] We rerun holdings-based unconditional TM tests for a one-month horizon using each of the two subsamples. If the low reporting-frequency group is found to have a significantly positive timing ability and the high group is not, then the evidence is consistent with window dressing.

As before, we obtain cross-sectional distributions of Newey–West t-statistics and their bootstrapped p-values for the timing measure for a given sample. Untabulated results indicate that both the average low and high reporting-frequency funds show significant timing ability, that the median high reporting-frequency fund also shows significant timing ability, and that the median low reporting-frequency fund has insignificant timing ability. As such, the evidence of the unconditional timing ability documented in the baseline case is unlikely due to window dressing. One reason for this finding—aside from the aforementioned one as provided by Elton et al. (2010)—is that we focus here on managers' ability to time the bond market, and as noted by Agarwal et al. (2013) and others, window dressing represents a false security selection ability.

4.5. Economic Value of Market Timing

Last, we provide an estimate for the economic value of market timing by following Henriksson and Merton (1981), Merton (1981), and Glosten and Jagannathan (1994). Under the specification of the return generating process given in (9), we can estimate the value of market timing as follows:

$$V_{\mathrm{MT}} = \frac{1}{1+r_f} E^{Q}[\gamma Term_t^2] = \gamma(1+r_f)(e^{\sigma_{bm}^2} - 1), \qquad (12)$$

where bond market returns are assumed to be log-normally distributed with the mean of r_f and standard deviation of σ_{bm} under the risk neutral measure Q.

To implement (12), we set γ to 0.21 (0.19), the mean (median) TM timing measure for the one-month forecasting horizon (Table 3); we use $r_f = 3.5\%$ (annual) and $\sigma_{bm}^2 = 0.06\%$ (monthly), which equal, respectively, the average one-month T-bill rate and the variance of monthly returns of the Lehman long-term government bond index over the 1997–2006 period. It follows that V_{MT} equals 0.16% (0.14%) per annum for the average (median) fund. These estimates are higher than typical bid-ask spreads of Treasuries (about 1–3 bps by Fleming 2003); they are also economically meaningful, accounting for 9.4% of the average excess return of our sample funds (1.7% per year, equal to the average annual return of 5.2% from Table 1 minus 3.5%).[19] On the other hand, the above estimates of the economic value are substantially lower than the sample average expense ratio of 86 bps (Table 1); that is, the value generated from market timing alone may not be high enough to cover fees paid by investors, consistent with Chen et al. (2010) who examine a very different sample of funds. However, keep in mind that these estimates are obtained from the model given in (12) that assumes continuous trading and ignores negative artificial timing, among other things (see Aragon and Ferson 2006, Ferson 2009 for more discussion on estimating the value of timing).

5. Conclusion

Because of their importance and high liquidity, Treasury securities are ideal candidates for market timers. Government bond fund managers presumably have the expertise to forecast future bond market conditions. In this article, we examine the ability of a sample of such managers to time the bond market return factor, using holdings-based timing measures and data on funds' monthly or quarterly holdings of Treasury securities.

We provide robust evidence that, on average, government bond fund managers possess significantly positive unconditional market timing ability at the one-month horizon. We also present evidence that government bond fund managers react to macroeconomic variables, including both information exacted from bond yields and macroeconomic news announcements. Specifically, our results suggest that when adjusting their portfolio betas, managers rely solely on public information, such as historical bond betas, nonfarm payrolls, the index of leading indicators, factory orders, and personal consumption. On the other hand, average funds have neutral or even slightly negative conditional market timing ability for the level of the bond market return.

[18] We thank an anonymous referee for suggesting this exercise.

[19] The ratio of 9.4% is comparable to what is documented in Jiang et al. (2007) and Elton et al. (2012) on equity fund performance.

We also find evidence that the majority of government bond funds roll over their positions in Treasury securities and that sometimes it is implemented passively. Furthermore, Treasury bond funds (specializing in Treasury securities) are found to less likely roll over their positions than are government general bond funds and exhibit more significantly positive (unconditional) timing ability compared with these funds. These results are consistent with the presence of the negative artificial timing effect induced by a passive rollover strategy and provide evidence that different styles of government bond funds have different abilities to process the public information release.

In conclusion, this study sheds light on the market timing ability of government bond funds and, in particular, on how such funds time the bond market and what specific information is used to time the market. Our findings have implications for performance evaluation of active bond fund managers as well as the investment decision making of investors.

Acknowledgments

For valuable comments and suggestions, the authors are grateful to Wei Xiong (department editor), an associate editor, two anonymous referees, Yakov Amihud, Heber Farnsworth, Wayne Ferson, Jean Helwege, Harrison Hong, Kyu Ho Kang, James Miles, Joris Pinkse, Laura Starks, Quang Vuong, Danielle Xu, Tong Yao, and Hong Zhang; seminar participants at California State University–Fullerton, City University of Hong Kong, Louisiana Tech, Peking University, Pennsylvania State University, Saint Louis University, State Street Global Advisors, University at Albany, University of Hong Kong, and University of Melbourne; and participants at the 2008 Financial Management Association meetings (Grapevine, Texas), 2008 Shanghai Winter Finance Conference, 2009 China International Conference in Finance (Guangzhou, China), and 2009 Northern Finance Association meetings (Niagara, Canada). Jing-Zhi Huang acknowledges a Smeal College Small Research Grant. Ying Wang acknowledges financial support from the Center for Institutional Investment Management at the University at Albany and a Smeal Dissertation Research Grant at Pennsylvania State University.

References

Agarwal V, Gay G, Ling L (2013) Window dressing in mutual funds. Working paper, Georgia State University, Atlanta.

Andersen TG, Bollerslev T, Diebold FX, Vega C (2007) Real-time price discovery in global stock, bond and foreign exchange markets. *J. International Econom.* 73:251–277.

Aragon GO, Ferson WE (2006) Portfolio performance evaluation. *Foundations Trends Finance* 2:83–190.

Balduzzi P, Elton EJ, Green CT (2001) Economic news and bond prices: Evidence from the U.S. Treasury market. *J. Financial Quant. Anal.* 36:523–543.

Becker CL, Ferson WE, Myers DH, Schill MJ (1999) Conditional market timing with benchmark investors. *J. Financial Econom.* 52:119–148.

Blake CR, Elton EJ, Gruber MJ (1993) The performance of bond mutual funds. *J. Bus.* 66:371–403.

Bollen NPB, Busse JA (2001) On the timing ability of mutual fund managers. *J. Finance* 56:1075–1094.

Boney V, Comer G, Kelly L (2009) Timing the investment grade securities market: Evidence from high quality bond funds. *J. Empirical Finance* 16:55–69.

Brandt MW, Kavajecz KA (2005) Price discovery in the U.S. Treasury market: The impact of order flow and liquidity on the yield curve. *J. Finance* 59:2623–2654.

Brown DT, Marshall WJ (2001) Assessing fixed-income fund manager style and performance from historical returns. *J. Fixed Income* 10(4):15–25.

Busse JA (1999) Volatility timing in mutual funds: Evidence from daily returns. *Rev. Financial Stud.* 12:1009–1041.

Cao C, Simin TT, Wang Y (2013) Do mutual fund managers time market liquidity? *J. Financial Markets* 16:279–307.

Chang EC, Lewellen WG (1984) Market timing and mutual fund performance. *J. Bus.* 57:57–72.

Chen Y, Ferson WE, Peters H (2010) Measuring the timing ability and performance of bond mutual funds. *J. Financial Econom.* 98:72–89.

Chevalier JA, Ellison GD (1997) Risk taking by mutual funds as a response to incentives. *J. Political Econom.* 105:1167–1206.

Cici G, Gibson S (2012) The performance of corporate bond mutual funds: Evidence based on security-level holdings. *J. Financial Quant. Anal.* 47:159–178.

Cochrane JH, Piazzesi M (2005) Bond risk premia. *Amer. Econom. Rev.* 95(1):138–160.

Comer G (2006) Evaluating bond fund sector timing skill. Working paper, Georgetown University, Washington, DC.

Cornell B, Green K (1991) Investment performance of low-grade bond funds. *J. Finance* 46:29–48.

Daniel K, Grinblatt M, Titman S, Wermers R (1997) Measuring mutual fund performance with characteristic-based benchmarks. *J. Finance* 52:1035–1058.

Duffee GR (2011) Information in (and not in) the term structure. *Rev. Financial Stud.* 24:2895–2934.

Elton EJ, Gruber MJ, Blake CR (1995) Fundamental economic variables, expected returns, and bond fund performance. *J. Finance* 50:1229–1256.

Elton EJ, Gruber MJ, Blake CR (2011) Holdings data, security returns, and the selection of superior mutual funds. *J. Financial Quant. Anal.* 46:341–367.

Elton EJ, Gruber MJ, Blake CR (2012) An examination of mutual fund timing using monthly holdings data. *Rev. Finance* 16: 619–645.

Elton EJ, Gruber MJ, Brown SJ, Goetzmann WN (2003) *Modern Portfolio Theory and Investment Analysis*, 6th ed. (John Wiley & Sons, Hoboken, NJ).

Elton EJ, Gruber MJ, Blake CR, Krasny Y, Ozelge SO (2010) The effect of holdings data frequency on conclusions about mutual fund behavior. *J. Banking Finance* 34:912–922.

Fama EF (1984) Term premiums in bond returns. *J. Financial Econom.* 13:529–546.

Fama EF, Bliss RR (1987) The information in long-maturity forward rates. *Amer. Econom. Rev.* 77:680–692.

Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. *J. Financial Econom.* 33:3–56.

Ferson WE (2009) The problem of alpha and performance measurement. Working paper, University of Southern California, Los Angeles.

Ferson WE, Khang K (2002) Conditional performance measurement using portfolio weights: Evidence for pension funds. *J. Financial Econom.* 65(2):249–282.

Ferson WE, Mo H (2013) Performance measurement with market and volatility timing and selectivity. Working paper, University of Southern California, Los Angeles.

Ferson WE, Qian M (2004) Conditional evaluation performance: Revisited. *Res. Foundation of the CFA Institute* 2004:1–85.

Ferson WE, Schadt RW (1996) Measuring fund strategy and performance in changing economic conditions. *J. Finance* 51:425–461.

Ferson WE, Warther VA (1996) Evaluating fund performance in a dynamic market. *Financial Analysts J.* 52:20–28.

Ferson WE, Henry TR, Kisgen DJ (2006) Evaluating government bond fund performance with stochastic discount factors. *Rev. Financial Stud.* 19:423–456.

Flannery MJ, Protopapadakis AA (2002) Macroeconomic factors do influence aggregate stock returns. *Rev. Financial Stud.* 15: 751–782.

Fleming MJ (2003) Measuring Treasury market liquidity. *Econom. Policy Rev.* 9:83–108.

Fleming MJ, Remolona EM (1999) Price formation and liquidity in the U.S. Treasury market: The response to public information. *J. Finance* 54:1901–1915.

Glosten LR, Jagannathan R (1994) A contingent claims approach to performance evaluation. *J. Empirical Finance* 1:133–166.

Goetzmann WN, Ingersoll J, Ivković Z (2000) Monthly measurement of daily timers. *J. Financial Quant. Anal.* 35:257–290.

Graham JR, Harvey CR (1996) Market timing ability and volatility implied in investment newsletters' asset allocation recommendations. *J. Financial Econom.* 42:397–421.

Green TC (2004) Economic news and the impact of trading on bond prices. *J. Finance* 59:1201–1234.

He J, Ng L, Wang Q (2004) Quarterly trading patterns of financial institutions. *J. Bus.* 77:493–510.

Henriksson RD (1984) Market timing and mutual fund performance: An empirical investigation. *J. Bus.* 57:73–96.

Henriksson RD, Merton RC (1981) On market timing and investment performance II: Statistical procedures for evaluating forecasting skills. *J. Bus.* 54:513–534.

Huang J-Z, Shi Z (2010) Determinants of bond risk premia. American Finance Association 2011 Denver Meetings Paper, http://dx.doi.org/10.2139/ssrn.1573186.

Huang J-Z, Wang Y (2013) Should investors invest in hedge fund-like mutual funds? Evidence from the 2007 financial crisis. *J. Financial Intermediation* 22:482–512.

Ilmanen A (1995) Time-varying expected returns in international bond markets. *J. Finance* 50:481–506.

Investment Company Institute (2010) *Investment Company Fact Book*, 50th ed. (Investment Company Institute, Washington, DC). http://www.ici.org/pdf/2010_factbook.pdf.

Jiang GJ, Yao T, Yu T (2007) Do mutual funds time the market? Evidence from portfolio holdings. *J. Financial Econom.* 86: 724–758.

Jiang W (2003) A nonparametric test of market timing. *J. Empirical Finance* 10:399–425.

Keim DB, Stambaugh RF (1986) Predicting returns in stock and bond markets. *J. Financial Econom.* 17:357–390.

Kosowski R, Timmermann A, Wermers R, White H (2006) Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. *J. Finance* 61:2551–2595.

Lakonishok J, Shleifer A, Thaler R, Vishny RW (1991) Window dressing by pension fund managers. *Amer. Econom. Rev.* 81: 227–231.

Ludvigson SC, Ng S (2009) Macro factors in bond risk premia. *Rev. Financial Stud.* 22:5027–5067.

Mamaysky H, Spiegel M, Zhang H (2008) Estimating the dynamics of mutual fund alphas and betas. *Rev. Financial Stud.* 21: 233–264.

McDonald M (2006) What Fidelity knows about Bernanke that Gross doesn't. *Bloomberg* (October 23), http://www.bloomberg.com/apps/news?pid=newsarchive&sid=a2SutvNk.ESg.

McQueen G, Roley V (1993) Stock prices, news, and business conditions. *Rev. Financial Stud.* 6:683–707.

Merton RC (1981) On market timing and investment performance part I: An equilibrium theory of value for market forecasts. *J. Bus.* 54:363–406.

Moneta F (2009) Measuring bond mutual fund performance with portfolio characteristics. EFA 2008 Athens Meetings Paper. http://ssrn.com/paper=1100766.

Morey MR, O'Neal ES (2006) Window dressing in bond mutual funds. *J. Financial Res.* 29:325–347.

Musto DK (2002) Investment decisions depend on portfolio disclosures. *J. Finance* 54:935–952.

Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–708.

Sharpe WF (1992) Asset allocation: Management style and performance measurement. *J. Portfolio Management* 18:7–19.

Sias RW, Starks LT (1997) Institutions and individuals at the turn-of-the-year. *J. Finance* 52:1543–1562.

Treynor JL, Mazuy K (1966) Can mutual funds outguess the market? *Harvard Bus. Rev.* 44:131–136.