ANNALS OF ECONOMICS AND FINANCE **11-1**, 1–33 (2010)

On the Economic Value of Return Predictability *

Yufeng Han †

University of Colorado Denver
E-mail: Yufeng.Han@ucdenver.edu

Recent studies provide strong statistical evidence challenging the existence of out-of-sample return predictability. The economic significance of return predictability is also controversial. In this paper, we find significant economic gains for dynamic trading strategies based on return predictability when appropriate portfolio constraints are imposed. We find that imposing appropriate portfolio constraints is critical for obtaining economic profits, which seems to explain the contradictory findings about economic significance in the literature. We also compare the performance of several predictive models including the VAR, the VAR-GARCH, and the (semi)nonparametric models and find that the simple VAR model performs similarly to other more complex models.

Key Words: Return predictability; Economic value; Model specification; Portfolio constraint.

JEL Classification Number: G11.

1. INTRODUCTION

Are stock returns predictable? This question has been one of the most actively research topics in finance because of its important theoretic and practical implications. Many studies from early studies including Fama and Schwert (1977) and French, Schwert, and Stambaugh (1987) to recent stud-

* The author wishes to thank an anonymous referee, Kerry Back, Michael Brandt, Phil Dybvig, Heber Farnsworth, Wayne Ferson, Robert Hansen, Ron Kaniel, Neal Maroney, Stefan Nagel, Brad Paye, Pedro Santa-Clara, John Scruggs, Gorge Tauchen, Janathan Taylor, Jiang Wang, Guofu Zhou, Heng-fu Zou (the editor), and seminar participants at the 2007 Asset Pricing Mini Conference at Washington University in St. Louis, China International Finance Conference 2007, the Center for Research in Security Prices 2006 Forum, MFA 2005 meeting, University of New Orleans, and Washington University in St. Louis for helpful comments. The author also thanks Cynthia Fransen for many helpful editorial comments.

† **Correspondence:** Yufeng Han, The Business School, University of Colorado Denver, PO Box 173364, Campus Box 165, Denver, CO 80217-3364. Tel: (303) 315-8411.

1529-7373/2010

ies such as Guo (2006) and Boudoukh, Michaely, Richardson, and Roberts (2007), show that return predictability is statistically significant,[1] whereas other recent studies such as Goyal and Welch (2008) and Ang and Bekaert (2007) provide comprehensive statistical evidence that strongly challenges the presence of out-of-sample return predictability. Despite the statistical controversy, however, Kandel and Stambaugh (1996) provide ex ante evidence that economic gains from return predictability can be significant even if the statistical evidence is rather weak.[2]

In this paper, we investigate the *ex post* economic value of predicting market returns. In particular, we employ dynamic trading strategies that are based on continuously updated estimates of the future returns out of sample and compare performance of the portfolios formed from the dynamic strategies to the performance of the benchmark portfolios that assume returns are not predictable. Even though ex ante evidence supports the economic importance of return predictability, ex post evidence is perhaps more relevant especially in light of the strong statistical evidence against return predictability. Unfortunately, the ex post evidence so far in the literature is far from conclusive. For example, Pesaran and Timmermann (1995), Breen, Glosten, and Jagannathan (1989), Marquering and Verbeek (2004), and Giannetti (2007) report significant economic profits from return predictability, whereas Handa and Tiwari (2006), Cooper, Gutierrez, and Marcum (2005), and Cooper and Gulen (2006) find that the economic profits are unstable and questionable.

We find that the economic value of return predictability is significant. In our analysis, the dynamic portfolios outperform the benchmark portfolios based on a number of performance measures including the Sharpe ratio, the Graham and Harvey (1997) risk-adjusted abnormal return, and the certainty equivalent rate of return criterion. We also find that imposing no-short-sale constraint (portfolios weights are restricted between zero and one) is crucial. Unconstrained dynamic portfolios do not outperform the benchmark portfolios, nor do dynamic portfolios that allow limited short selling. We further show that switching portfolios whose weights can only

[1]The list of studies in this area is enormous. Some other studies are Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988a), Campbell and Shiller (1988b), Fama and French (1988), Fama and French (1989), Ferson (1989), Harvey (1989), Schwert (1989), Jegadeesh (1990), Ferson and Harvey (1991), Cochrane (1991), Hodrick (1992), Bekaert and Hodrick (1992), Marathe and Shawky (1994), Lamont (1998), Lewellen (1999), Lettau and Ludvigson (2001), Santos and Veronesi (2006), Cremers (2002), Avramov (2002), and Guo and Savickas (2006). See Campbell (2000) and Goyal and Welch (2008) for more extensive surveys.

[2]Xu (2004) also shows that small levels of predictability (e.g., 2% R^2) can be economically significant. Other studies including Balduzzi and Lynch (1999) and Campbell, Chan, and Viceira (2003) also demonstrate potentially significant ex ante economic gains of predicting future returns.

be either zero or one even outperform the no-short-sale constrained portfolios. These results suggest that imposing appropriate portfolio constraints is critical for obtaining significant economic gains. The results seem to also suggest that predictive variables and predictive models appear to provide useful predictive information about the sign of the market expected excess returns but perhaps not about the magnitude of the market expected excess returns.[3] Further research in this direction may be of interest.

Our findings do not contradict with the findings of Goyal and Welch (2008). Information in the predictive models may be statistically rather weak, but the limited information appears to be enough to generate significant economic gains provided that appropriate restrictions are imposed on the portfolio weights. Similar approaches have been used by Campbell and Thompson (2008) and Wachter and Warusawitharana (2009) who both show that return predictability becomes statistically significant after imposing restrictions on the predictive regression slope coefficients. However, they do not examine the economic significance under their restrictions.

Our findings also provide novel reconciliation of a number of contradictory results reported in previous studies. For example, the reason that Breen, Glosten, and Jagannathan (1989) and Pesaran and Timmermann (1995) find out-of-sample economic significance is because that each uses the switching strategy. Moreover Marquering and Verbeek (2004) find significant economic gains from predicting market returns because they constrain from short-selling the market portfolio, whereas Handa and Tiwari (2006) find no consistently significant superior performance because they allow short-selling.

Another contribution of the paper is that we focus on the model specification of the underlying data-generating process, which has received very little attention in the literature. A number of papers including Cremers (2002), Avramov (2002), Masih, Mansur, Masih, and Mie (2008), Aiolfi and Favero (2003), etc., recognize the uncertainty in choosing the best predictors and take a model averaging approach. The focus of these papers is on the choice of the predictors while assuming a linear regression model as the data-generating process for the market returns. The focus of this paper instead is on the different forms of specifications for the data-generating process, such as linear models vs nonlinear models. Our goal is to see if a more sophisticated model can produce better performance, and we compare the models using the out-of-sample portfolio performance measures. Pesaran and Timmermann (1995) and Granger and Pesaran (2000) argue

[3] A possible explanation, attributable to Merton (1980), may be too much noise in observed returns to accurately estimate expected returns, even if the predictive relation holds. Torous and Valkanov (2000) similarly argue that even if returns are predictable the noise in the predictive regression may overwhelm the signal of the conditional variables.

that economic criteria such as portfolio performance measures used in this paper are more appropriate than the statistical measures of forecast accuracy when comparing different return predictability models. We consider four predictive models. The first is a simple vector autoregressive (VAR) model, a special case of which is the predictive regression model used in most studies. In the second model, we add a GARCH feature to the VAR model to accommodate time-varying volatility, yielding the VAR-GARCH model. The third model is the seminonparametric (SNP) model, proposed by Gallant and Tauchen (1989), which uses Hermite polynomial expansions to approximate the underlying data-generating process and is thus, capable of capturing many features of the data. The SNP model nests the VAR and VAR-GARCH models as special cases. Fourth is a generalized SNP model which allows for more non-linearity and is almost nonparametric (NLNP model). We find that all the predictive models perform similarly. In particular, the more sophisticated predictive models such as VAR-GARCH, SNP and NLNP do not consistently perform better than the VAR model. These results contrast with the findings of Carlson, Chapman, Kaniel, and Yan (2004) who report significant utility cost associated with ignoring volatility dynamics (e.g., GARCH feature). However, their results are based on calibration analysis and simulations, which raises concerns with the real-world relevance of their findings. Our results are also different from those of Marathe and Shawky (1994) who find that a nonlinear model substantially improve the ability of dividend yield to predict market returns both statistically and economically.

Our findings that the more sophisticated predictive models do not yield better portfolio performance may seem puzzling. One explanation for the lack of improvement is that portfolio performance is fairly insensitive to the specification of the underlying data-generating process. This agrees with findings in Pástor and Stambaugh (2000) and Tu and Zhou (2004), both showing that similar portfolio performance can be obtained despite different specifications of the data-generating processes. In particular, Tu and Zhou (2004) show that normality assumption works well in the portfolio choice problem. We extend their results to the case when returns are predictable.

The remainder of this article is organized as follows. Section 2 discusses the group of predictive models that incorporate return predictability. Section 3 describes the predictive variables and discusses the estimation results of the predictive models. Section 4 discusses investors' optimal portfolio choice problems. Section 5 conducts the out-of-sample portfolio analysis to examine the performance of the predictive models. Section 6 concludes.

2. SPECIFICATION OF THE PREDICTIVE MODELS

The first order vector autoregressive (VAR) model has been extensively used in the literature to model return predictability of the market portfolio. It captures the basic notion that the market return is a (linear) function of the predictive variables. However, the choice of the first order is arbitrary and for convenience. In this paper, we will use statistical model selection criteria to choose the best order. The general specification of the VAR model is given as follows:

$$\mathbf{y}_t = \Phi_0 + \sum_{i=1}^{L_\mu} \Phi_i \mathbf{y}_{t-i} + \epsilon_t, \tag{1}$$

where $\mathbf{y}_t$ is the state vector including the excess returns on the market portfolio and predictors at time t, ϵ_t is a vector of normally distributed disturbances with a zero vector of means and variance-covariance matrix Σ, and L_μ denotes the order of autoregression. As pointed out earlier, L_μ is always set to one in the empirical studies. Furthermore, many studies often use a further simplified predictive linear regression model, which only considers the return equation in the VAR model. However, this regression model is subject to estimation bias discussed by Ferson, Sarkissian, and Simin (2003), and Stambaugh (1999).

The VAR model assumes the disturbances ϵ_t are independently identically distributed. Stock returns, however, exhibit prominent conditional heteroscedasticity (see, e.g., Engle, 1982). Therefore, a natural extension of the VAR model to deal with conditional heteroscedasticity is to incorporate GARCH features. The extended VAR-GARCH model captures predictability in both the first and the second moments of stock returns. It should be noted that the predictive variables also display conditional heteroscedasticity. For example, the variance of T-bill yield is known to vary with the level of the yield. The VAR-GARCH model not only captures the conditional heteroscedasticity in the market returns, but also those in the predictive variables. We believe our paper presents a novel application of the VAR-GARCH model.

However, both VAR and VAR-GARCH models assume (conditional) normality for the distributions, an assumption firmly rejected by the data, and a linear relation between returns and predictive variables, an assumption unlikely to be true. To further relax these two restrictions, we consider the seminonparametric (SNP) model proposed by Gallant and Tauchen (1989). The SNP model relies on the Hermite polynomial expansions to approximate the conditional density of the underlying data-generating process. Because of polynomial expansions, the conditional distribution is no longer normal, and the moments are non-linear functions of the predictors. Another relevant advantage of the SNP model is that it nests both the VAR and VAR-GARCH models as degenerated cases, which makes it easy to compare and select different type of models. To facilitate estimation and model comparison, all models including the VAR and VAR-GARCH models are estimated using the procedure proposed by Gallant and Tauchen (1997).

The SNP model is specified as follows. Let $f(\mathbf{y}|\mathbf{x},\theta)$ denote the conditional density of the state vector $\mathbf{y}$ conditioned on the lagged values of $\mathbf{y}$, denoted by $\mathbf{x}$. Then

$$f(\mathbf{y}|\mathbf{x},\theta) \propto [P(\mathbf{z})]^2 \phi(\mathbf{y}|\mu_x, \Sigma_x), \tag{2}$$

where

$$\mathbf{z} = R_x^{-1}(\mathbf{y} - \mu_x), \quad \mu_x = b_0 + Bx, \quad \Sigma_x = R_x R_x',$$
$$Vec(R_x) = \rho_0 + \sum_{i=1}^{L_r} \rho_i |\mathbf{y} - \mu_x| + \sum_{j=1}^{L_g} Diag(G_j) Vec(R_j),$$

and $P(\mathbf{z})$ is the multivariate Hermite polynomials with degree K_z. The GARCH specification used in the SNP model is more akin to the one suggested by Nelson (1991). Note that because of the rich parameterizations in multivariate GARCH, we restrict the GARCH to a diagonal specification. We can easily see that when K_z is zero, the Hermite polynomial degenerates to a constant, and thus, the SNP model degenerates to the VAR-GARCH model; when Σ_x is constant, the SNP model further degenerates to the Gaussian VAR model. For financial data, it may be necessary to consider a more general model where the coefficients of the polynomial $P(\mathbf{z})$ are polynomials of degree K_x in x because of the extraordinary heteroscedasticity. This model is non-linear and nonparametric. Collectively, the parameters L_μ, L_g, L_r, K_z, and K_x uniquely identify the SNP model,[4] and hence, we use "$[L_\mu/L_g/L_r/K_z/K_x]$" to denote the specification of a predictive model. For example, [1/0/0/0/0] denotes VAR(1), while [1/1/1/0/0] denotes VAR(1)-GARCH(1,1).

One advantage of the framework is that we can systematically select the best model specification for each type of the models (VAR, VAR-GARCH, SNP, and the generalized SNP) using statistical criteria. Another advantage is that the models are nested, which allows us to compare and select the best overall model specification across the different types of models.

[4]Two additional parameters, I_z and I_x are used to reduce the cross-interaction terms in the polynomials when y is multivariate. The highest orders for cross-interaction terms are $K_z - I_z$ and $K_x - I_x$, respectively. We include these two parameters in the specification search.

This systematical approach is far superior to the ad hoc assumption that the data follow certain processes such as VAR(1). To this end, we use Schwartz's Bayesian Information Criterion (BIC), defined as

$$BIC = \frac{-2\mathcal{L}_k + k \log n}{n}, \tag{3}$$

where $\mathcal{L}_k$ is the log likelihood function with k parameters, and n is the number of observations. Additional statistical criteria are also considered including Akaike's Information Criterion (AIC), and Hannan and Quinn Criterion (HQ), defined as

$$AIC = \frac{-2\mathcal{L}_k + 2k}{n}, \quad \text{and} \quad HQ = \frac{-2\mathcal{L}_k + k \log \log n}{n}. \tag{4}$$

Because all the model selection criteria are negatively related to the log likelihood functions, smaller numbers indicate better model specifications. However, different model selection criteria balance differently the tradeoff between complexity of the model and overfitting. BIC has the most severe penalty for rich parameterizations, whereas AIC has the least severe penalty, and HQ is in between. It turns out that the generalized SNP model is always rejected by the BIC because of its rich parameterizations, but sometimes is favored by the AIC. In the sequel we denote the best SNP model as OPT, and the best generalized SNP model as NLNP. On this note, both Bossaerts and Hillion (1999) and Pesaran and Timmermann (1995) emphasize using statistic criteria to choose the best predictive models. Among others, the key difference between those two studies and this study is that they are confined to linear regression models, whereas we consider a more broad class of models including both linear and non-linear ones.

3. ESTIMATION OF THE PREDICTIVE MODELS

3.1. Data description

In recent years, empirical literature has identified many economic variables that seem to have predictive power over stock and bond returns. These variables include dividend yield, Treasury-bill yield, term spread, default spread, consumption to wealth ratio (Lettau and Ludvigson, 2001), investment to capital ratio (Cochrane, 1991), dividend to earnings ratio (Lamont, 1998), debt to equity ratio (Schwert, 1989), and lagged returns, just to name a few. Among these predictive variables, the dividend yield is the most popular one, partly because of theoretic support, and the T-bill yield, term spread, and default spread are also widely used. In our empirical analysis, we use these four predictive variables as examples to illustrate our analysis, but the same analysis can be easily carried out with other predictive variables.

We use the S&P 500 composite index as the proxy for the market portfolio. Monthly returns on the S&P 500 index and 30-day Treasury bill are obtained from the CRSP and are converted to continuously compounded (log) returns. Excess returns in percentage are used to fit various predictive models and converted to decimal returns for portfolio optimization. The dividend yield (DVYD) defined as the sum of the dividends paid on the S&P 500 index over the past 12 months divided by the current level of the index, the three-month Treasury-bill yield (TBYD), the term spread (TRSD), defined as the difference in yields between the ten-year Treasury bond and one-year Treasury-bill, and the default spread (DFSD), defined as the difference in yields between Moody's AAA bonds and BAA bonds, are obtained from the DRI (now Global Insight). Monthly observations from January 1947 to December 1998, spanning 624 months, are collected except for the term spread, which is only available from April 1953, a total of 549 observations.

TABLE 1.

Descriptive statistics of data

Panel A: Descriptive Statistics

	Mean	Std. Dev.	Skewness	Kurtosis	Jarque Bera	Autocorrelations ρ_1	ρ_3	ρ_6	ρ_{12}
EXRN	3.654	14.267	−0.640	5.747	238.9	0.022	0.012	−0.059	0.042
RFT1M	4.782	0.869	1.045	4.527	174.2	0.955	0.922	0.877	0.808
DVYD	3.950	1.211	0.460	2.585	26.6	0.989	0.962	0.919	0.835
TBYD	4.954	2.998	0.973	4.208	136.5	0.988	0.952	0.909	0.847
DFSD	0.908	0.426	1.532	5.345	387.4	0.976	0.918	0.850	0.720
TRSD	0.718	0.990	−0.137	3.463	6.6	0.960	0.841	0.710	0.517
EXRN(47:1–78:12)	2.296	13.638	−0.261	3.725	127.7	0.034	0.072	−0.075	0.085
EXRN(79:1–98:12)	5.825	15.228	−1.099	4.828	284.6	0.004	−0.061	−0.041	−0.037

Panel A in Table 1 reports the mean, standard deviation, and other statistics about the market excess returns (EXRN), the returns on the 30-day Treasury bill (RFT1M), and the predictive variables. As expected, the excess returns exhibit negative skewness and excess kurtosis; Jarque-Bera statistics also indicate that the excess returns, riskfree rates, and the predictive variables are far from normally distributed. Also reported are the autocorrelation coefficients up to lag 12. The excess returns have very little autocorrelations, whereas the predictive variables are highly autocorrelated, with the first order autocorrelation coefficients as high as 0.989.

Panel B: Correlations

EXRN	DVYD	TBYD	DFSD	TRSD	EXRN	DVYD	TBYD	DFSD
		1953:4–1998:12					1947:1–1998:12	
1.000	-0.069	-0.153	0.053	0.165	1.000	-0.030	-0.140	0.043
	1.000	0.409	0.497	-0.191		1.000	0.011	0.228
		1.000	0.643	-0.408			1.000	0.664
			1.000	0.105				1.000
				1.000				
		1953:4–1978:12					1947:1–1978:12	
1.000	-0.005	-0.266	0.095	0.280	1.000	0.045	-0.225	0.074
	1.000	-0.025	0.150	0.111		1.000	-0.425	-0.077
		1.000	0.343	-0.508			1.000	0.399
			1.000	0.386				1.000
				1.000				
		1979:1–1998:12						
1.000	-0.115	-0.167	-0.005	0.080				
	1.000	0.807	0.801	-0.328				
		1.000	0.681	-0.703				
			1.000	-0.179				
				1.000				

Panel A of this table shows the descriptive statistics for the continuously compounded excess return (annualized in percentage) on the S&P 500 composite index (EXRN), continuously compounded return on the one-month T-bill (RFT1M), dividend yield (DVYD), three-month Treasury bill yield (TBYD), default spread (DFSD), and term spread (TRSD). For all the variables except the term spread, the data is sampled monthly from January 1947 through December 1998, with a total of 624 observations. For TRSD, the data is only available from April 1953. The whole sample period is divided into two subperiods: the first subperiod is from 1947:1 to 1978:12 (or 1953:4 - 1978:12 for TRSD); the second subperiod is from 1979:1 to 1998:12. Panel B shows the correlations of the excess returns with the predictive variables for the whole sample period and the two subperiods.

Panel A also reports statistics of the market excess returns for two subperiods, 1947:1–1978:12 and 1979:1–1998:12. These two subperiods are quite different; the market excess returns are much higher on average, more volatile, and more skewed in the second subperiod than in the first subperiod. In the out-of-sample analysis, the first subperiod serves as the base period for estimating the predictive models, and the second subperiod serves as the evaluation period.

Panel B in Table 1 reports the correlation matrices of the market excess returns and the predictive variables in the whole sample period and the two subperiods.[5] On the one hand, most of the correlations are not stable over time. For example, the correlation between the excess return

[5] To include the term spread, the whole sample period starts from April, 1953. With other predictive variables, the whole sample period starts from January, 1947.

and dividend yield is about -0.03 for the whole period, but is positive (0.045) in the first subperiod and negative (-0.115) in the second subperiod. The correlation of default spread with the excess return changes from 0.074 in the first subperiod to -0.005 in the second subperiod. Finally, the correlation between the excess return and term spread is strong in the first subperiod but becomes much weaker in the second subperiod. On the other hand, the correlation of T-bill yield with the excess return is relatively stable and remains considerable. These differences in correlations are consistent with our subsequent portfolio performance results that T-bill yield is the strongest predictor, followed by term spread and default spread, and dividend yield does not seem to have any predictive power at all. Interestingly, the correlations of the dividend yield with T-bill yield and default spread increase from negative in the first subperiod to positive in the second subperiod, whereas the correlation between default spread and term spread decreases to negative in the second subperiod. Other correlations also change considerably over the two subperiods.

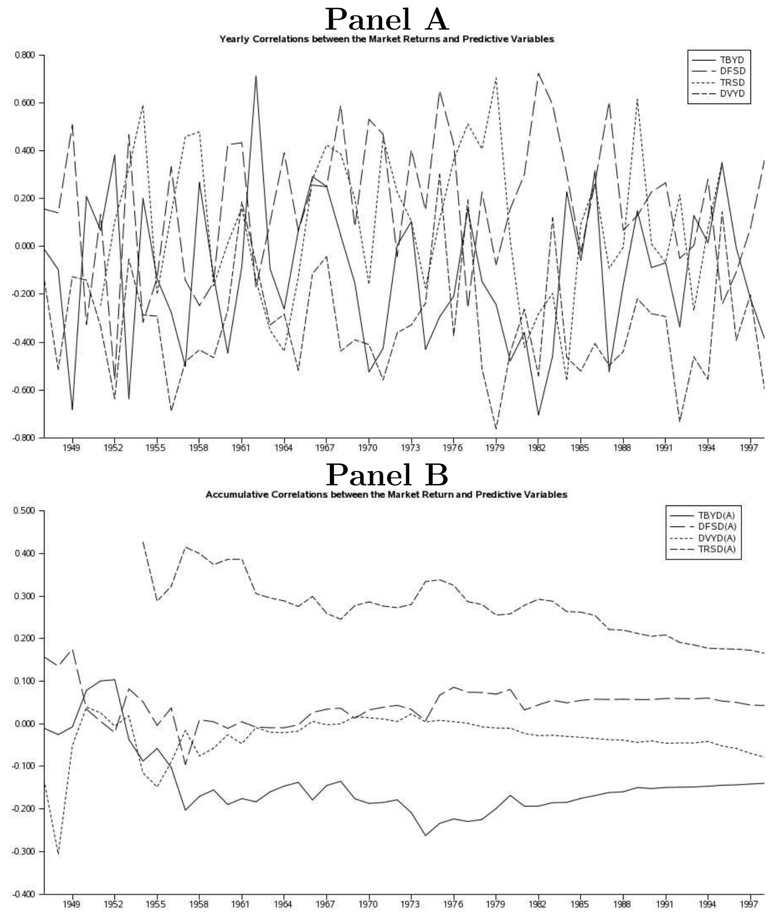
Figure 1 plots the correlations between the market returns and the four predictive variables. Panel A plots the yearly correlations from 1947 (or 1953 for the term spread) to 1998. We can easily see that the correlations vary widely from year to year. Panel B plots the accumulative correlations, which are much smoother. Three observations can be made from the accumulative correlations. First, all the correlations are trending lower. Second, the correlations are more volatile before year 1980 than after 1980. Finally, the dividend yield and default spread have almost zero correlations after year 1960. As we will see later, these observations have direct implications on the portfolio performance.

3.2. Specification search and model estimation

The empirical literature on return predictability has been using the VAR(1) model or a further simplified predictive regression model as the data-generating process. However, little attention has been paid to investigate whether the assumed model is appropriate. In this subsection, we examine various types of predictive models and try to identify the best model according to an array of statistical criteria.

We use the monthly time series of excess returns and predictive variables to search for the best specification for each of the following models: VAR, VAR-GARCH, SNP, and generalized SNP. We also consider various combinations of the four predictive variables. The best specification for each predictive model and each combination of the predictive variables is selected according to the Bayesian Information Criterion (BIC). However, because the BIC criterion tends to reject models with rich parameterizations, we also use other statistic criteria such as Akaike's Information Criterion (AIC), and Hannan-Quinn Criterion (HQ).

FIG. 1. Correlations between the market return and the predictive variables.



Panel A: Yearly correlations. Correlations are calculated each year using the monthly returns within the year. Panel B: Accumulative correlations. Correlations are calculated each year using monthly returns from the beginning of the sample period up to the current year.

TABLE 2.

Optimal predictive models

Optimal Predictive Models and Model Selection Criteria										
	Model	Obj	BIC	AIC	HQ	Model	Obj	BIC	AIC	HQ
	DVYD					TBYD				
VAR	[1/0/0/0/0]	0.40	0.45	0.41	0.43	[2/0/0/0/0]	0.85	0.92	0.87	0.89
GARCH	[2/2/1/0/0]	0.19	0.28	0.22	0.24	[2/2/1/0/0]	0.20	0.29	0.23	0.25
OPT	[2/2/1/4/0]	0.06	0.20	0.10	0.14	[2/2/1/4/0]	0.13	0.27	0.17	0.21
NLNP	[2/2/1/4/1]	-0.02	0.25	0.06	0.14	[2/2/1/4/1]	0.11	0.35	0.18	0.25
	DFSD					TRSD				
VAR	[2/0/0/0/0]	1.29	1.36	1.31	1.33	[2/0/0/0/0]	1.50	1.58	1.53	1.55
GARCH	[2/1/1/0/0]	0.85	0.93	0.88	0.90	[2/1/1/0/0]	1.21	1.30	1.24	1.26
OPT	[2/1/1/4/0]	0.76	0.89	0.80	0.83	[2/1/1/4/0]	1.15	1.29	1.19	1.23
NLNP	[2/1/1/4/1]	0.73	0.95	0.80	0.86	[2/1/1/4/1]	1.09	1.33	1.17	1.23
	TBYD & DFSD					TBYD & TRSD				
VAR	[2/0/0/0/0]	0.61	0.75	0.65	0.69	[2/0/0/0/0]	0.42	0.57	0.47	0.51
GARCH	[2/1/1/0/0]	-0.30	-0.14	-0.25	-0.21	[2/1/1/0/0]	0.03	0.21	0.09	0.14
OPT	[2/1/1/4/0]	-0.42	-0.20	-0.35	-0.29	[2/1/1/4/0]	-0.11	0.16	-0.02	0.05
NLNP	[2/1/1/4/1]	-0.50	-0.07	-0.36	-0.25	[2/1/1/4/1]	-0.23	0.32	-0.05	0.09
	TRSD & DFSD					TBYD & TRSD & DFSD				
VAR	[2/0/0/0/0]	1.34	1.50	1.39	1.43	[2/0/0/0/0]	0.18	0.45	0.27	0.34
GARCH	[2/1/1/0/0]	0.68	0.87	0.74	0.79	[2/1/1/0/0]	-0.50	-0.21	-0.41	-0.33
OPT	[2/1/1/4/0]	0.57	0.82	0.65	0.72	[2/1/1/4/0]	-0.69	-0.26	-0.55	-0.44
NLNP	[2/1/1/4/1]	0.47	0.95	0.62	0.75	[2/1/1/4/1]	-0.85	0.12	-0.54	-0.28

This table reports for each combination of predictive variables the best specifications for the four predictive models and the various model selection criteria used. The estimation is done with the full sample period. Obj is the value of the objective function and is defined as Obj $= -(\frac{1}{n}\sum_{t=1}^{n}\log[f(y_t|x_{t-1},\theta)]$, and BIC, AIC, and HQ are the Schwartz Bayesian Information Criterion, the Akaike Information Criterion, and the Hannan-Quinn Criterion, respectively. We use the following notation to denote the various models, "$[L_\mu/L_g/L_r/K_z/K_x]$", where L_μ denotes the order of the vector autoregression [VAR(L_μ)], L_g and L_r denote the order of the GARCH model [GARCH(L_g, L_r)], K_z denotes the order of the Hermite polynomial expansion, and K_x is the order of polynomial in x for the coefficients in the Hermite polynomial.

Table 2 reports the best specifications in each of the four models for a number of combination of predictive variables.[6] Several interesting results emerge. First, the predominantly used VAR model is clearly mis-specified. Adding the GARCH specification substantially improves the goodness-of-fit over the VAR model, as demonstrated by the much smaller values for all the criteria. For example, incorporating conditional heteroscedasticity reduces the BIC from 0.92 to 0.29 for the T-bill yield, from 1.36 to 0.93 for the default spread, and from 0.75 to -0.14 for the T-bill yield and default spread combination. Adding the SNP specification also improves the fit, but the improvement is not nearly as drastic as adding the GARCH feature. For example, the BIC is reduced from 0.29 to 0.27 for the T-bill yield, from 0.93 to 0.89 for the default spread, and from -0.14 to -0.20 for the combination of the T-bill yield and default spread.

Second, the first order VAR model is not even the best VAR specification for most combinations of predictive variables, whereas the second order VAR model often is. For example, all combinations of predictive variables but the dividend yield have the VAR(2) as the best VAR specification. This observation suggests that it seems inadequate to use the VAR(1) model as the data-generating process for the excess returns and predictive variables.

Third, a polynomial of degree four is often the best choice for the SNP model, which is also the best specification overall because it yields the smallest BIC, and often the smallest HQ, as well. As expected, the overfitted SNP model (NLNP), often has higher BIC value than the OPT model, but the smallest AIC value.

Fourth, two or more predictors provide better fit than any single one of them does. While the improvement is small for T-bill yield, it is considerable for the term spread and default spread. For example, T-bill yield and term spread combined yield a BIC value of 0.16 for the OPT model, whereas T-bill alone yields 0.27, and term spread alone 1.29, respectively.

4. PORTFOLIO CHOICE UNDER THE PREDICTIVE MODELS

Assume a risk-averse investor has a preference over wealth represented by a utility function $u(W)$, where W is her wealth. The investor chooses her asset allocation policy between a risky asset (the market portfolio), and a riskless asset (30-day Treasury Bill), to maximize her expected utility given her estimates of the conditional distributions of future stock returns.

[6] We do not show the combination of dividend yield with any other predictive variables because dividend yield, as we will show later, does not have any predictive power.

Specifically, the investor solves the following one-period optimization problem at time t:

$$\max_{\omega_t} \mathbf{E}[u(W_{t+1})|\mathcal{F}_t] = \max_{\omega_t} \int u(W_{t+1})\pi(r_{t+1}|\mathcal{F}_t)\mathrm{d}r_{t+1}, \tag{5}$$

s.t.

$$W_{t+1} = W_t[\omega_t e^{r_{t+1}+r_{f,t+1}} + (1-\omega_t)e^{r_{f,t+1}}],$$

where r_{t+1} and ω_t are the excess return at $t+1$ and portfolio weight on the market portfolio at time t, respectively, and $r_{f,t+1}$ is the return on the riskless asset at time $t+1$.

The integration in eq. (5) can be evaluated numerically via Monte Carlo simulation. Thus, the optimization problem can be written as

$$\max_{\omega_t} \frac{1}{N}\sum_{i=1}^{N} u(W_t[\omega_t e^{r^{(i)}_{t+1|t}+r_{f,t+1}} + (1-\omega_t)e^{r_{f,t+1}}]), \tag{6}$$

where $r^{(i)}_{t+1|t}$ are the sample draws from the forecasted one-step-ahead future conditional distribution of stock returns, generated from the underlying predictive models, and N is the number of simulations. If we assume that the investor's preference over wealth is determined by the constant relative risk averse power utility, then the optimization problem is

$$\max_{\omega_t} \frac{1}{N}\sum_{i=1}^{N} \frac{W_t^{1-\gamma}[\omega_t e^{r^{(i)}_{t+1|t}+r_{f,t+1}} + (1-\omega_t)e^{r_{f,t+1}}]^{(1-\gamma)}}{1-\gamma}, \tag{7}$$

where γ is the investor's relative risk aversion coefficient. The optimization is solved numerically by the Brent method with analytic derivatives. In the similar spirit of Campbell and Thompson (2008), we restrict the weights being between zero and one, which means the investor is prohibited from short selling the market portfolio or buying the market portfolio on margin.

In the presence of transaction costs, the investor will choose the optimal portfolio weights, taking into consideration the costs associated with rebalancing the weights. We assume the proportional transaction cost is τ for the market portfolio, and assume no transaction cost in trading the riskless asset. The investor's wealth is given by

$$W_{t+1} = W_t(1-f_t)[\omega_t e^{r_{t+1}+r_{f,t+1}} + (1-\omega_t)e^{r_{f,t+1}}], \tag{8}$$

where the transaction cost at time t, f_t, is given by

$$f_t = \tau|\omega_t - \hat{\omega}_t|, \tag{9}$$

and $\hat{\omega}_t$ is the inherited portfolio weight from the previous period,

$$\hat{\omega}_t = \frac{\omega_{t-1} e^{r_t + r_{f,t}}}{\omega_{t-1} e^{r_t + r_{f,t}} + (1 - \omega_{t-1}) e^{r_{f,t}}}. \tag{10}$$

5. OUT-OF-SAMPLE PORTFOLIO PERFORMANCE OF THE PREDICTIVE MODELS

5.1. Portfolio performance measures

Having determined the best specifications for the predictive models, the predictability investor forecasts the one-step-ahead conditional return distributions, conditioning on the previous realized returns and predictors, and then find the optimal portfolio weights as described in Section 4. To measure the performance of the portfolios formed in this manner, we use several performance measures including the Sharpe ratio, certainty equivalent rate of return (CER), and a measure proposed by Graham and Harvey (1997) (henceforth $\mathcal{GH}2$). The sample CERs are calculated by taking the average of the realized utilities over the period considered;

$$\mu(W_0(1 + r_{ce})) = \frac{1}{T} \sum_{t=1}^{T} \mu(W_0(1 + r_{p,t})), \tag{11}$$

where r_{ce} is the sample CER, $r_{p,t}$ is the realized portfolio return at time t, and $\mu(\cdot)$ is the utility function. $\mathcal{GH}2$ is a measure of risk-adjusted abnormal returns, which is suitable for diversified portfolios only. In a nutshell, $\mathcal{GH}2$ is the abnormal return that the measured portfolio would have earned if it had the same risk (volatility) as the market portfolio. More specifically, we first lever up or down the measured portfolio with the one-month T-bill (riskfree asset) so that the levered portfolio has the same risk (volatility) as the market portfolio. We then compare the average return of the levered portfolio with that of the market portfolio. It amounts to finding the weight ω to solve the following problem:

$$V_m = \mathbf{Var}(\omega r_{pt} + (1-\omega) r_{ft}) = \omega^2 V_p + (1-\omega)^2 V_f + 2\omega(1-\omega)\mathbf{Cov}(r_{pt}, r_{ft}), \tag{12}$$

where V_m, V_p, and V_f are the variances of the market returns, managed portfolio returns, and riskfree rates, respectively. $\mathcal{GH}2$ is then given as

$$\mathcal{GH}2 = \omega \bar{r}_p + (1 - \omega)\bar{r}_f - \bar{r}_m, \tag{13}$$

where $\bar{r}_p$, $\bar{r}_f$, and $\bar{r}_m$ are the average returns on the managed portfolio, one-month T-bill, and the market portfolio, respectively. $\mathcal{GH}2$ is related to the Sharpe ratio[7] but unlike the Sharpe ratio, it also quantifies the outperformance. Note that when the average return is lower than the riskfree rate, the Sharpe ratio will be negative and can no longer be used to rank performance, and $\mathcal{GH}2$ will overestimate the outperformance.

5.2. Out-of-sample portfolio performance analysis

We use the last 20 years in the sample period, i.e., from 1979:1 to 1998:12, as the evaluation period for the out-of-sample tests. This period is an interesting period as it contains some recession periods (1980:1–1980:7, 1981:7–1982:11, and 1990:7–1991:3), and the longest boom period in the 1990s. To strike a balance between capturing changes in parameters and computational burden, we estimate the predictive models recursively with a five-year moving window - the estimation is repeated every five years with an expanding window of periods. For example, the predictability investor first estimates the predictive models using the data from the initial estimation period (from 1947:1 to 1978:12), and then over the next five years, she forms the optimal portfolios based on the forecasted one-period-ahead return distributions from the predictive models. After five years she repeats the estimation of the predictive models using data from the initial estimation period plus the past five years (expanding window). The estimation is repeated every five years. This procedure is motivated by results in Table 1, which show that the relationships between the market returns and predictive variables are unstable over time, and thus the best specifications and the parameter estimates may change over time.

Table 3 reports the performance results of the recursive estimation. The first benchmark strategy is the passive buy-and-hold strategy (labeled "Passive") where the weight is determined at the beginning of the evaluation period using the historical distribution of returns, and no rebalance of the portfolio is required thereafter. The other benchmark strategies require portfolio rebalance. For example, the unconditionally optimal strategy (labeled "Unconditional"), which is based on the assumption that the returns are I.I.D. normally distributed, requires monthly rebalance to keep the weight from drifting away from the optimal weight. Rebalance is also necessary because the I.I.D investor recursively updates the mean and variance of market returns. In addition, we include two other benchmark strategies that only model the dynamics of the market returns, the autoregressive model (labeled "AR") and AR-GARCH model (labeled "GARCH"), respectively. It is interesting to note that the performance of the four bench-

[7] $\mathcal{GH}2$ is similar to the well-known $\mathcal{M}2$ measure except that $\mathcal{GH}2$ does not assume that the riskfree rate is constant over time. $\mathcal{M}2$ is directly related to the Sharpe ratio as $\mathcal{M}2 = \sigma_m(\mathcal{SR}_p - \mathcal{SR}_m)$, whereas no direct mathematical relation exists between $\mathcal{GH}2$ and the Sharpe ratio.

TABLE 3.

Out-of-sample performance of the predictive models

	Out-Of-Sample Portfolio Performance Tests									
	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$	r_{ce}	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$	r_{ce}
	Benchmarks									
Passive	10.18	7.47	0.43	0.74	9.32					
Unconditional	9.96	6.06	0.45	0.95	9.40					
AR	9.63	6.27	0.43	0.68	9.04					
GARCH	9.90	9.33	0.32	-1.01	8.58					
	DVYD					TBYD				
VAR	7.78	7.44	0.11	-4.09	6.95	10.38	6.31	0.55	2.45	9.78
GARCH	8.69	8.95	0.20	-2.84	7.50	10.23	5.94	0.56	2.59	9.70
OPT	7.64	8.05	0.09	-4.48	6.68	10.21	6.24	0.53	2.16	9.62
NLNP	8.09	6.39	0.18	-3.06	7.49	9.80	5.65	0.51	1.89	9.32
	DFSD					TRSD				
VAR	7.85	10.31	0.09	-4.47	6.21	11.46	11.49	0.39	0.14	9.36
GARCH	11.00	11.05	0.37	-0.26	9.15	11.73	11.67	0.41	0.40	9.56
OPT	9.54	11.39	0.23	-2.36	7.46	10.89	10.89	0.36	-0.32	8.97
NLNP	9.91	8.71	0.34	-0.65	8.77	11.52	11.60	0.40	0.17	9.37
	TBYD & DFSD					TBYD & TRSD				
VAR	11.89	7.60	0.65	4.07	11.05	12.98	9.85	0.61	3.46	11.51
GARCH	11.52	7.35	0.62	3.63	10.72	12.40	9.84	0.55	2.58	10.92
OPT	10.16	6.35	0.51	1.89	9.58	11.07	10.90	0.38	-0.08	9.15
NLNP	11.84	6.73	0.73	5.21	11.17	13.58	8.75	0.76	5.66	12.44
	TRSD & DFSD					TBYD & TRSD & DFSD				
VAR	10.73	11.02	0.35	-0.60	8.78	10.89	6.90	0.57	2.87	10.19
GARCH	10.84	10.93	0.36	-0.40	8.92	10.38	8.08	0.43	0.64	9.39
OPT	10.27	10.98	0.30	-1.22	8.33	10.24	9.53	0.35	-0.55	8.75
NLNP	11.37	11.89	0.37	-0.18	9.13	9.02	8.49	0.25	-2.07	7.80

Predictive sample draws of the excess returns $\tilde{\mathbf{r}}_{t+1|t}$ are generated at each month t from the one-step-ahead conditional distributions of the predictive models, conditioned on the observed out-of-sample data $\mathbf{y}_t$. The predictive models are estimated recursively every five years. The realized portfolio returns r_{pt} are calculated from the observed excess returns r_t and the riskfree rates. The average returns (r_p), standard deviations (σ_p), and three performance measures, Sharpe ratio ($\mathcal{SR}$), Graham-Harvey measure ($\mathcal{GH}2$), and CER (r_{ce}), are reported for each combination of the predictive variables and each predictive model.

mark strategies except "GARCH" is quite similar, whereas "GARCH" strategy performs slightly worse than the others.

The performance of the predictive variables varies. Dividend yield (DVYD) consistently underperforms the benchmark strategies. Similarly, the default spread (DFSD) and term spread (TRSD) both underperforms the benchmarks as well. However, T-bill yield (TBYD) consistently outperforms the benchmarks. For example, the dynamic strategy based on the VAR-GARCH model of the T-bill yield generates a Sharpe ratio of 0.56, a risk-adjusted abnormal return of 2.59% per annum, and a CER of 9.70% per annum, all of which are higher than those of the benchmark strategies (e.g., 0.45, 0.95%, and 9.40%, respectively, for the unconditionally optimal benchmark strategy). Even though DFSD and TRSD alone and the combination of the two do not generate superior performance, combining each one with TBYD generates performance superior to that of the TBYD alone, suggesting that both DFSD and TRSD provide additional useful information beyond the information contained in TBYD.[8] Furthermore, the combination of TBYD and TRSD generates the strongest performance, suggesting that the term structure of interest rate has a profound impact on the expected stock returns. Finally, the relatively weaker performance of the triple combination of TBYD, TRSD, and DFSD may be due to the difficulty in estimating a high-dimensional complex model.

Unlike the statistical analysis in Table 2, we find no consistent differences in portfolio performance among the four types of predictive models and the performance differences are generally small. In particular, even though the VAR model is clearly mis-specified, the portfolio performance of the VAR model is on par with other better specified models. By contrast, the OPT model - the best overall statistical model - does not perform quite as well as others and is often the worst. Furthermore, adding the GARCH feature to the VAR model improves the performance in some cases (e.g., DFSD), but lowers the performance in other cases. This result is very different from the finding of Carlson, Chapman, Kaniel, and Yan (2004) who show that the utility loss of ignoring volatility dynamics may be economically significant. However, their finding is based on simulation study and the relevance to the real world performance is unclear. The lack of performance difference among different predictive models is similar to the findings of Pástor and Stambaugh (2000) and Tu and Zhou (2004). Both show that different data-generating processes may unnecessarily yield different portfolio performance. Finally, the nonlinear model (NLNP) seems to be able to generate the highest performance in many cases. For example, with

[8]Again, we do not report results of any combination of dividend yield with other predictive variables because adding dividend yield does not produce stronger results.

the combination of TBYD and DFSD and the combination of TBYD and TRSD, NLNP model performs the best.

5.3. Further investigation of the out-of-sample performance

The lack of performance difference among the four statistically very different predictive models is certainly puzzling. In this subsection, we examine the robustness of the out-of-sample results in several dimensions.

First, we examine the robustness to the estimation. In previous analysis, we use recursive estimations where the estimation window has a fixed starting period and expands five years each time. As an alternative, we use rolling estimations where the estimation window rolls over to a new starting period five years later each time and the length of the estimation window is fixed. We obtain very similar results, and thus do not report the results in the paper. As we mentioned earlier, the choice of five-year moving windows is to capture the possible changes in parameters and specifications and at the same time to reduce the computational burden. We would like to re-estimate the models every month, but it presents a daunting computational burden because each time we search for the best specifications, a process cannot be automated. We, however, conduct the analysis using a two-year moving windows and also obtain similar results (not reported).

Second, we examine the robustness to the utility function. We first change the relative risk aversion coefficient from four to ten. we then change the investor's preference from power utility to mean-variance utility. Our results (not reported) seem robust to the changes in the investor's utility function. For example, the performance measures with the mean-variance preference is comparable to those with the power utility. Under the mean-variance preference, T-bill yield generates Sharpe ratios around 0.49, and $\mathcal{GH}2$ ranging from 1.22 to 1.67% per annum, while under the power utility, T-bill yield generates Sharpe ratios from 0.51 to 0.56, and $\mathcal{GH}2$ from 1.89 to 2.59% per annum. Again, no consistent rankings in performance exist among the four predictive models.

Third, we examine the robustness to transaction costs. We consider three levels of transaction costs: 0.25%, 0.50%, and 1.00%, representing low, medium, and high transaction costs, even though we believe that trading the market portfolio (e.g., S&P 500 ETF) probably incurs transaction costs lower than the low transaction cost specified here. Results in Table 4 show that the low level of transaction cost has virtually no impact on the performance. This finding is expected given that the investor incorporates transaction costs into her objective function and optimally chooses the portfolio weights. Similar results are obtained for the medium level of transaction cost, although it starts to show the negative impact. The high level of transaction cost, however, has apparent negative impact on the performance of the combination of T-bill yield and default spread, but has

TABLE 4.

Out-of-sample performance with transaction costs

Out-Of-Sample Portfolio Performance Tests with Transaction Costs										
	r_p(%)	σ_p(%)	$\mathcal{SR}$	$\mathcal{GH}2$	r_{ce}(%)	r_p(%)	σ_p(%)	$\mathcal{SR}$	$\mathcal{GH}2$	r_{ce}(%)
	Benchmarks									
Passive	10.18	7.47	0.43	0.74	9.32					
	TBYD & DFSD					TBYD & TRSD				
	$\tau = 0.25\%$					$\tau = 0.25\%$				
VAR	11.85	7.36	0.67	4.31	11.06	12.31	10.71	0.50	1.78	10.52
GARCH	11.32	7.17	0.61	3.46	10.56	11.91	11.05	0.45	0.99	9.94
OPT	9.88	5.98	0.49	1.66	9.35	11.73	10.97	0.44	0.79	9.78
NLNP	11.47	6.26	0.72	5.16	10.89	13.68	9.03	0.74	5.47	12.47
	$\tau = 0.50\%$					$\tau = 0.50\%$				
VAR	11.21	6.62	0.65	3.99	10.57	12.39	10.88	0.50	1.77	10.56
GARCH	10.03	6.99	0.44	0.90	9.29	12.18	10.76	0.49	1.56	10.36
OPT	9.29	5.80	0.41	0.34	8.79	11.20	11.10	0.38	0.00	9.21
NLNP	10.52	5.87	0.61	3.47	10.02	13.52	9.16	0.72	5.05	12.27
	$\tau = 1.00\%$					$\tau = 1.00\%$				
VAR	9.01	7.04	0.30	-1.31	8.25	13.14	11.49	0.54	2.35	11.01
GARCH	7.69	7.15	0.11	-4.19	6.84	12.91	11.07	0.54	2.34	10.91
OPT	7.56	6.21	0.10	-4.27	6.93	12.59	11.76	0.48	1.46	10.39
NLNP	9.18	5.26	0.43	0.70	8.78	12.74	10.00	0.58	2.96	11.20

The predictive sample draws are generated from the five-year recursive estimation. The optimal portfolio weights are calculated from maximizing the expected power utility in the presence of transaction costs. The transaction costs are 25bps, 50bps, and 100bps. The realized portfolio returns r_{pt} are calculated from the observed excess returns r_t and the riskfree rates. The average returns and standard deviations of the realized portfolio returns, and three performance measures, Sharpe ratio ($\mathcal{SR}$), Graham-Harvey measure ($\mathcal{GH}2$), and CER (r_{ce}), are reported. For brevity, results for only two combinations - T-bill yield and default spread, and T-bill yield and term spread - are reported.

apparently no detectable impact on the performance of the combination of T-bill yield and term spread.

TABLE 5.

Out-of-sample portfolio performance under different constraints

	Out-of-Sample Portfolio Performance under Different Constraints																			
	$r_p(\%)$	σ_p	$\mathcal{SR}$	$\mathcal{GH}2$	r_{ce}	r_p	σ_p	$\mathcal{SR}$	$\mathcal{GH}2$	r_{ce}	r_p	σ_p	$\mathcal{SR}$	$\mathcal{GH}2$	r_{ce}	r_p	σ_p	$\mathcal{SR}$	$\mathcal{GH}2$	r_{ce}
	Benchmarks																			
Market	12.74	15.10	0.38		9.18															
Passive	10.18	7.47	0.43	0.74	9.32															
Expand	9.96	6.06	0.45	0.95	9.40															
AR	9.63	6.27	0.43	0.68	9.04															
GARCH	9.90	9.33	0.32	-1.01	8.58															
	TBYD																			
	No Constraint					$\|w\| \leq 2$					$\|w\| \leq 1$					$0 \leq w \leq 2$				
VAR	13.78	46.63	0.15	-3.62	-75.30	11.19	17.24	0.25	-2.08	6.73	10.33	11.19	0.31	-1.21	8.47	10.91	8.46	0.47	1.29	9.85
GARCH	9.88	38.67	0.08	-4.67	-35.30	10.25	17.88	0.19	-3.01	5.24	10.61	11.03	0.34	-0.76	8.79	10.38	6.68	0.52	2.00	9.71
OPT	-21.45	133.52	-0.21	-9.04	-3804.10	9.62	18.28	0.15	-3.59	4.39	10.18	11.22	0.29	-1.42	8.30	10.71	6.93	0.55	2.45	9.99
NLNP	10.60	23.69	0.16	-3.48	1.41	9.03	17.38	0.12	-3.99	4.35	9.42	11.09	0.23	-2.41	7.58	9.64	6.28	0.43	0.72	9.04
	TBYD & DFSD																			
	No Constraint					$\|w\| \leq 2$					$\|w\| \leq 1$					$0 \leq w \leq 2$				
VAR	18.15	41.86	0.27	-1.80	-15.53	13.96	20.45	0.35	-0.63	7.72	11.57	12.66	0.37	-0.27	9.20	13.93	12.56	0.56	2.61	11.61
GARCH	14.33	29.76	0.25	-2.08	-0.58	12.75	18.91	0.31	-1.17	7.31	11.32	11.72	0.38	-0.15	9.26	12.38	10.72	0.51	1.87	10.65
OPT	9.76	36.62	0.08	-4.66	-13.61	10.01	19.87	0.16	-3.48	4.13	9.12	12.14	0.18	-3.08	6.95	10.97	9.48	0.43	0.62	9.64
NLNP	9.05	39.59	0.05	-5.02	-38.06	11.96	22.59	0.22	-2.47	4.37	11.12	13.50	0.31	-1.12	8.45	14.25	10.72	0.68	4.49	12.57
	TBYD & TRSD																			
	No Constraint					$\|w\| \leq 2$					$\|w\| \leq 1$					$0 \leq w \leq 2$				
VAR	18.26	58.81	0.19	-2.93	-187.73	16.47	20.12	0.48	1.34	10.41	13.34	12.04	0.53	2.24	11.15	15.54	15.53	0.55	2.56	11.90
GARCH	17.93	37.15	0.30	-1.37	-7.07	15.07	19.69	0.41	0.42	9.17	12.49	12.31	0.45	1.02	10.19	14.69	14.13	0.55	2.48	11.69
OPT	11.65	55.24	0.09	-4.54	-153.47	11.94	21.71	0.23	-2.34	4.40	10.95	13.34	0.30	-1.26	8.13	11.90	16.86	0.29	-1.36	7.11
NLNP	20.51	32.84	0.42	0.40	3.18	16.56	21.13	0.46	1.05	9.73	13.63	12.07	0.56	2.59	11.44	15.86	15.49	0.58	2.89	12.23

The predictive sample draws are generated from the recursive estimation with five-year expanding windows. The portfolios are constructed under the corresponding constraint. The *ex post* portfolio returns r_{pt} are calculated from the observed excess returns r_t and the riskfree rates. The average returns (r_p), standard deviations (σ_p), and three performance measures, Sharpe ratio ($\mathcal{SR}$), Graham-Harvey measure ($\mathcal{GH}2$), and CER (r_{ce}), are reported for each predictive variable and model combination.

5.4. Role of portfolio constraints

As an additional robustness test, we investigate the role of portfolio constrains. In Table 5 we compare the portfolio performance under no constraint and several different constraints. In particular, we impose constraints that are based on Regulation T, which requires 50% margin for purchasing and 150% for short selling. Assuming the interest rates for borrowing and lending are the same, then Regulation T imposes the following restriction, $|w| < 100/\psi$, where $\psi\%$ is the 50% margin requirement. We also consider 100% margins. The last constraint considered here allows borrowing up to 100%, but excludes short selling. As shown in Table 5, without any portfolio constraint, the dynamic portfolios significantly underperform the benchmark portfolios. In particular, the risk-adjusted return $\mathcal{GH}2$ and the certainty equivalent return r_{ce} are mostly negative. Imposing Regulation T based constraints significantly improve the performance over the unconstrained case, especially by drastically reducing the volatilities of the unconstrained portfolios without significantly reducing the average returns. However, Regulation T based constraints fail to outperform the benchmarks for the most part. Disallowing short selling but allowing some borrowing ($0 \leq w \leq 2$) yields superior performance to the benchmarks. For example, the TBYD and DFSD combination outperforms the benchmarks in every predictive model. Nevertheless, the performance of this constraint is still not as good as that of the no-short-sale constraint in Table 3. For example, the TBYD and DFSD combination has respective Sharpe ratios of 0.65, 0.62, 0.51, and 0.73 under the no-short-sale constraint, vis-à-vis 0.56, 0.51, 0.43, and 0.68 under the limited borrowing constraint ($0 \leq w \leq 2$).

5.5. Market timing strategy

Because the no-short-sale constraint yields the best portfolio performance so far, we further examine the portfolio weights under this constraint. We find that at least 70% of the weights are either 0 or 1, which means that more often than not, the optimal weights obtained based on the predictive models may not be correct, and better performance can be achieved if the portfolio weights are restricted to either pure cash position or pure equity position.

To further support this conjecture, we examine the performance of switching portfolios. By construction, switching strategy switches from the all-equity position to the pure-cash position or vice versa, depending on whether the forecasted expected excess returns are positive or negative. Table 6 reports the performance of the switching strategies. To be compatible, the benchmarks also use the switching strategies instead of the optimal strategies. In particular, the benchmark labeled "Unconditional" is the switching strategy that switches between all-equity and pure-cash positions depending on the estimated mean of the excess returns. We also add a random

TABLE 6.

Out-of-sample performance of switching strategies

Out-Of-Sample Portfolio Performance Tests												
	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$	$\Delta_{\mathcal{GH}2}$	$\mathcal{Z}$	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$	$\Delta_{\mathcal{GH}2}$	$\mathcal{Z}$
	Benchmarks											
Unconditional	9.96	6.06	0.45	0.95								
AR	11.58	14.36	0.32	-0.93								
GARCH	12.77	14.77	0.39	0.15								
Random	9.83	10.66	0.28	-1.65								
	DVYD						TBYD					
VAR	8.91	8.89	0.22	-2.45	1.64	1.03	12.92	12.28	0.49	1.56	1.42	1.06
GARCH	10.39	11.87	0.29	-1.42	1.42	1.44	13.71	12.85	0.53	2.16	1.76	1.18
OPT	7.82	9.31	0.10	-4.37	0.11	0.62	12.78	12.73	0.46	1.13	1.45	1.24
NLNP	8.48	9.56	0.16	-3.36	-0.30	0.93	13.28	12.50	0.51	1.85	1.68	0.83
	DFSD						TRSD					
VAR	12.66	9.06	0.63	3.72	1.28	1.25	13.26	11.88	0.53	2.23	-1.22	1.02
GARCH	12.58	9.00	0.63	3.66	1.07	1.38	13.53	12.51	0.53	2.16	-0.42	1.26
OPT	12.35	8.95	0.60	3.33	1.17	1.02	12.58	11.95	0.47	1.32	1.40	1.11
NLNP	11.96	8.90	0.56	2.72	0.83	1.16	14.39	9.52	0.78	6.01	0.35	0.93
	TBYD & DFSD						TBYD & TRSD					
VAR	11.22	14.23	0.30	-1.27	3.20	1.08	11.83	12.26	0.40	0.23	0.83	1.07
GARCH	13.04	14.79	0.41	0.42	0.68	1.39	13.54	12.41	0.53	2.23	2.63	2.13
OPT	10.59	14.16	0.26	-1.91	0.45	1.06	11.70	12.28	0.39	0.05	1.27	1.34
NLNP	12.12	14.15	0.36	-0.28	0.38	1.11	12.90	13.18	0.45	1.02	1.20	1.19
	TRSD & DFSD						TBYD & TRSD & DFSD					
VAR	13.35	8.73	0.74	5.31	1.25	0.80	10.93	9.88	0.41	0.31	-2.55	1.03
GARCH	12.43	8.62	0.64	3.83	0.20	1.18	10.95	10.13	0.40	0.19	-0.46	1.26
OPT	11.05	7.34	0.56	2.67	0.78	0.60	11.87	10.33	0.48	1.42	1.97	0.97
NLNP	12.86	7.89	0.75	5.55	0.34	0.61	11.20	10.30	0.41	0.45	2.52	0.98

The predictive sample draws are generated from the five-year recursive estimation. At each month, the portfolio either invests in the market portfolio or one-month T-bill depending on whether the forecasted expected returns are higher or lower than the riskfree rates. The realized portfolio returns r_{pt} are calculated from the observed excess returns r_t and the riskfree rates. The average returns (r_p), standard deviations (σ_p), and two performance measures, Sharpe ratio ($\mathcal{SR}$), and Graham-Harvey measure ($\mathcal{GH}2$), are reported for each combination of the predictive variables and each predictive model. Also reported are $\Delta_{\mathcal{GH}2}$, which measures the difference in Graham-Harvey measure between switching portfolios and the corresponding no-short-sale constrained portfolios, and $\mathcal{Z}$, which measures the ratio of the frequency of the switching portfolios beating the corresponding constrained portfolios, and that of the constrained portfolios beating the switching portfolios.

switching strategy as another benchmark. For the random switching strategy, the weights are determined by the toss of a coin. We repeat the experiment 5000 times, and the means of the Sharpe ratios and $\mathcal{GH}2$ are reported. On average, the random switching strategy significantly underperforms the other benchmark strategies, with an average Sharpe ratio of 0.28 and a negative risk-adjusted abnormal return (-1.65%). The 90th percentile of the Sharpe ratio is 0.51, and of the $\mathcal{GH}2$ is 1.96%.

All combinations, except dividend yield alone and, to a less extent, default spread alone, are able to generate superior performance to the benchmarks in at least one predictive model, and many outperform the benchmarks in all four models.

Furthermore, the performance of the switching portfolios is stronger than that of the no-short-sale constrained optimal portfolios. For example, the VAR model of the T-bill yield has a Sharpe ratio of 0.63 and a risk-adjusted abnormal return of 3.72% per annum for the switching portfolio, vis-$\grave{a}$-vis 0.55 and 2.45% per annum for the no-short-sale constrained optimal portfolio. Thus moving from the no-short-sale constrained, utility-maximization based strategy to the switching strategy results in a 1.28% increase in the risk-adjusted abnormal return as reported in the second to the last column (labeled $\Delta_{\mathcal{GH}2}$). Most models show a positive improvement in the risk-adjusted abnormal return. The last column (labeled $\mathcal{Z}$) measures how frequently the switching strategies beat the corresponding no-short-sale constrained optimal strategies. In most cases, this ratio is larger than one, indicating that the switching strategies more frequently have higher returns. The largest improvement is with the term spread, whose portfolio performance changes from underperforming to outperforming.

This evidence suggests the necessity of making a finer distinction of the predictive ability of the predictive variables: the ability to predict the magnitude of the market expected excess return and the ability to predict just the sign. Predictive variables do not seem to have the ability to predict the magnitude of the market expected excess return out of sample because of estimation errors and other problems, but appear to have the ability to predict the sign or direction of changes of the market expected excess return.

As for the performance differences of four predictive models, the results are similar - the performance is close among the predictive models for the most part, and not a single model consistently outperforms the others. However, the nonlinear model (NLNP) performs considerably better than the other models with the combination of T-bill yield and term spread. For example, NLNP yields a Sharpe ratio of 0.78 vs. 0.53 of the next best model, and risk-adjusted abnormal return of 6.01% vs. 2.23% of the next best model. Again, the VAR model performs on par with other models, better in some cases, and worse in other cases.

TABLE 7.

Market timing performance

	β_2	$\mathcal{TM}$	Corr	β_2	$\mathcal{TM}$	Corr	β_2	$\mathcal{TM}$	Corr
Unconditional	0.13^{***}	0.30	0.07						
AR	0.39^{***}	0.90	-0.00						
GARCH	0.43^{***}	0.98	-0.08						
	TBYD			DFSD			TRSD		
Optimal:									
VAR	0.66^{***}	1.50	0.10	0.80^{***}	1.83	-0.04	-0.31	-0.70	0.04
GARCH	0.53^{***}	1.22	0.09	0.94^{***}	2.15	0.07	-0.25	-0.58	0.04
OPT	0.38^{*}	0.87	0.09	-0.57	-1.31	0.01	-0.77	-1.75	0.02
NLNP	0.37^{*}	0.85	0.08	0.04	0.09	0.05	-0.47	-1.07	0.03
Switching:									
VAR	0.61^{**}	1.38	0.13	-0.71	-1.61	-0.04	-0.15	-0.33	0.08
GARCH	0.61^{**}	1.39	0.12	-0.01	-0.03	0.09	0.12	0.27	0.11
OPT	0.57^{*}	1.30	0.12	-0.30	-0.68	0.00	-0.09	-0.22	0.07
NLNP	0.56^{*}	1.28	0.11	-0.97	-2.22	0.02	-0.14	-0.33	0.09
	TBYD & DFSD			TBYD & TRSD			TRSD & DFSD		
Optimal:									
VAR	1.76^{***}	4.01	0.14	0.39	0.89	0.12	-0.55	-1.26	0.02
GARCH	1.73^{***}	3.94	0.12	0.04	0.10	0.10	-0.71	-1.62	0.02
OPT	1.45^{***}	3.31	0.10	-0.70	-1.60	0.04	-0.59	-1.35	0.00
NLNP	1.74^{***}	3.97	0.12	1.62^{***}	3.69	0.14	-0.12	-0.28	0.02
Switching:									
VAR	2.08^{***}	4.75	0.17	-0.22	-0.50	0.11	-0.22	-0.50	0.04
GARCH	1.97^{***}	4.51	0.16	-0.71	-1.62	0.12	-0.25	-0.57	0.10
OPT	1.82^{***}	4.15	0.11	-0.54	-1.23	0.08	-0.25	-0.57	0.03
NLNP	1.92^{***}	4.39	0.12	1.84^{***}	4.20	0.16	0.17	0.38	0.06

β_2 is the coefficient of the squared market excess returns in the following regression

$$r_t = \alpha + \beta_1 r_{mt} + \beta_2 r_{mt}^2 + \epsilon_t,$$

where r_t is the excess return on the measured portfolio and r_{mt} is the market excess return. The timing performance measure $\mathcal{TM}$ is defined as $\mathcal{TM} = \beta_2 var(r_{mt})$. Corr., defined as $Cov(w_t, r_{mt})$, measures the correlation between the portfolio weights and the market excess returns. Positive significance at 1%, 5%, and 10% levels is denoted by ***, **, *, respectively.

5.6. Market timing performance test

In Table 7, we compare the market timing performance of the no-short-sale constrained optimal strategies and switching strategies under various predictive models and combinations of predictive variables. Specifically, we examine the coefficient of the squared market excess returns in the quadratic regression proposed by Treynor and Mazuy (1966),

$$r_t = \alpha + \beta_1 r_{mt} + \beta_2 r_{mt}^2 + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2). \tag{14}$$

A significantly positive estimate of β_2 indicates successful market timing. Furthermore, under the conditions provided by Admati, Bhattacharya, Pfleiderer, anf Ross (1986), $\beta_2 var(r_{mt})$ measures the abnormal return of market timing. The three benchmark strategies, Unconditionally optimal strategy based on I.I.D. model, market AR model, and market AR-GARCH model, all have a significant and positive β_2 coefficient, but the abnormal timing performance is rather small. The third column in Table 7 reports the correlations between the market returns and the weights of the measured portfolios, which are essentially zeros for all the three benchmarks. However, for T-bill yield (TBYD), both the optimal strategy and switching strategy generate significantly positive coefficient β_2, positive abnormal returns of market timing, and positive correlations. Stronger market-timing performance is obtained when both T-bill yield and default spread (TBYD & DFSD) are present, consistent with the better performance of TBYD & DFSD combination reported in Table 3 and 6. Also of these two cases the switching strategy generates higher market timing performance than does the no-short-sale constrained optimal strategy, which is consistent with the results in Table 6. Other combinations of predictive variables do not seem to have any market timing ability.

Among the four predictive models, the market timing performance is again close and no consistent rankings can be observed, although VAR model seems to perform the best in the cases of TBYD alone and the combination of TBYD and DFSD. For example, In the presence of TBYD and DFSD, the timing coefficient, the abnormal return of market timing, and the correlation are 2.08, 4.75%, and 0.17, respectively, for the switching strategy based on the VAR model, and 1.97, 4.51%, 0.16, respectively, for the same strategy based on the VAR-GARCH model, the second best model in this case. However, NLNP model is the only model that has significant market timing ability for the combination of T-bill yield and term spread.

5.7. Subperiod performance analysis

Since Table 1 suggests that the relationships between the market returns and predictive variables are unstable over time, one must be careful about interpreting the results. For example, the results may be specific to the evaluation period we consider. Therefore, we conduct subperiod performance analysis to examine 1) if the results are specific to the period we choose, and 2) if there are any interesting dynamics of the dynamic strategies.

Table 8 reports the subperiod performance of the switching strategies. For brevity, we only report the results for the combination of T-bill yield and default spread (Panel A) and of T-bill yield and term spread (Panel B). We also extend the sampling period from 1998 to 2003 to include the latest recession period. The top rows report the performance of the I.I.D. benchmark model. In the first five years (1979–1983), the performance of the benchmark is very poor with a Sharpe ratio of 0.05 because of recessions in this period (1980 and 1982). The performance improves subsequently and reaches the highest after 1998 with a Sharpe ratio of 0.44 because of the boom market in 1990s. The Sharpe ratio then drops to 0.23 after 2003 because of the latest recession in 2001. In contrast, the performance of the predictive models in the presence of T-bill yield and default spread is very strong in the first five-year period and remains rather strong in the subsequent periods until after 1998 when the performance deteriorates considerably because of the latest recession. The last column reports the difference in $\mathcal{GH}2$ between the predictive models and the benchmark ($\Delta_{\mathcal{GH}2}$). The difference is always positive, suggesting that the performance of the predictive models is superior to that of the benchmark in every period, including the boom period of late 1990s and the most recent recession period. These results demonstrate the robustness of the outperformance of the predictive models to different time periods. Similar results are obtained for the combination of T-bill yield and term spread.

As a robustness check, we also repeat the analysis with a different starting period. Specifically, we use the first 14 years from 1947 to 1958 as the initial estimation period, and then we recursively estimate the predictive models every five years for the next 45 years (1959–2003). Results not reported show that after 45 years the difference in the risk-adjusted abnormal return is 5.40% per annum between the NLNP model of the combination of T-bill yield and default spread and the I.I.D. benchmark. Put it differently, if the initial investment is \$10,000, then an I.I.D. benchmark investor would have accumulated \$136,340 at the end of year 2003, whereas the predictability investor would have accumulated \$266,110 at the end of year 2003 using the NLNP model.

6. CONCLUSION

While recent studies such as Goyal and Welch (2008) cast strong doubt on the out-of-sample predictability of the market returns, we echo Kandel and Stambaugh (1996)'s argument that the economic value of return pre-

TABLE 8.

Out-of-sample five-year subperiod performance

Out-Of-Sample Five-Year Subperiod Performance Tests											
	r_p(%)	σ_p(%)	$\mathcal{SR}$	$\mathcal{GH}2$	$\Delta_{\mathcal{GH}2}$		r_p(%)	σ_p(%)	$\mathcal{SR}$	$\mathcal{GH}2$	$\Delta_{\mathcal{GH}2}$
Unconditional											
1979 - 1983	9.74	6.92	0.05	0.61							
1979 - 1988	10.59	6.93	0.15	0.72							
1979 - 1993	10.05	6.28	0.22	0.55							
1979 - 1998	10.07	6.33	0.44	0.94							
1979 - 2003	8.32	7.39	0.23	-0.02							
Panel A: TBYD & DFSD						**Panel B: TBYD & TRSD**					
VAR						VAR					
1979 - 1983	16.88	7.01	0.86	12.77	12.16	1979 - 1983	13.72	5.32	0.55	8.12	7.52
1979 - 1988	16.49	10.06	0.76	10.92	10.19	1979 - 1988	11.87	13.54	0.23	1.98	1.26
1979 - 1993	14.70	9.43	0.75	8.67	8.12	1979 - 1993	11.62	12.05	0.33	2.21	1.66
1979 - 1998	13.35	8.73	0.74	5.31	4.38	1979 - 1998	13.26	11.88	0.53	2.23	1.30
1979 - 2003	10.25	10.60	0.38	2.39	2.41	1979 - 2003	10.79	12.72	0.36	2.04	2.06
GARCH						GARCH					
1979 - 1983	13.65	3.03	0.92	13.76	13.16	1979 - 1983	14.30	4.97	0.70	10.38	9.77
1979 - 1988	15.36	9.62	0.68	9.54	8.82	1979 - 1988	12.49	13.48	0.28	2.76	2.04
1979 - 1993	13.37	9.30	0.61	6.62	6.07	1979 - 1993	12.59	12.15	0.41	3.39	2.84
1979 - 1998	12.43	8.62	0.64	3.83	2.90	1979 - 1998	13.53	12.51	0.53	2.16	1.23
1979 - 2003	10.06	10.68	0.36	2.06	2.08	1979 - 2003	10.87	13.17	0.36	1.95	1.97
OPT						OPT					
1979 - 1983	13.08	2.83	0.80	11.93	11.32	1979 - 1983	12.32	2.63	0.59	8.76	8.15
1979 - 1988	12.60	8.20	0.47	5.94	5.22	1979 - 1988	11.17	13.13	0.18	1.21	0.48
1979 - 1993	11.61	7.70	0.51	5.06	4.52	1979 - 1993	10.99	12.18	0.27	1.36	0.81
1979 - 1998	11.05	7.34	0.56	2.67	1.74	1979 - 1998	12.58	11.95	0.47	1.32	0.39
1979 - 2003	8.74	9.88	0.26	0.44	0.46	1979 - 2003	10.03	12.77	0.30	1.10	1.11
NLNP						NLNP					
1979 - 1983	17.80	6.68	1.03	15.36	14.75	1979 - 1983	12.27	2.64	0.57	8.49	7.88
1979 - 1988	15.71	8.67	0.79	11.43	10.71	1979 - 1988	13.34	9.47	0.48	6.18	5.46
1979 - 1993	14.13	8.30	0.78	9.17	8.62	1979 - 1993	12.43	9.06	0.53	5.26	4.71
1979 - 1998	12.86	7.89	0.75	5.55	4.62	1979 - 1998	14.39	9.52	0.78	6.01	5.07
1979 - 2003	10.28	9.85	0.42	2.90	2.92	1979 - 2003	11.43	11.18	0.47	3.72	3.74

The predictive sample draws are generated from the five-year recursive estimation and switching strategies are formed as described in the text. The testing periods are coincident with the re-estimation periods. Panel A reports the results for the combination of T-bill yield and default spread, and Panel B reports the results for the combination of T-bill yield and term spread. The average returns (r_p), standard deviations (σ_p), and three performance measures, Sharpe ratio ($\mathcal{SR}$), Graham-Harvey measure ($\mathcal{GH}2$), and CER (r_{ce}), are reported. The last column reports the performance difference between the predictive models and the benchmark ($\Delta_{\mathcal{GH}2}$).

dictability can be significant even though the statistical evidence may be weak. We show that significant economic gains can be obtained from predicting market returns out of sample using the commonly used predictive variables such as T-bill yield. However, we do find an important caveat; significant economic gains can only be obtained after imposing the appropriate restrictions on the portfolio weights, for example, restricting the weights between zero and one. Unconstrained or inappropriately constrained dynamic trading strategies do not outperform the benchmark strategies based on the assumption that the market returns are IID.

We also examine the impact of model specification, which has received little attention in the literature. We compare the portfolio performance of several predictive models, including the VAR, VAR-GARCH, SNP and a generalized SNP model. We find that the more complex models do not consistently outperform the simple VAR model out of sample, which suggests that a VAR model is adequate for predicting the market returns out of sample.

Our analysis can be extended in a number of interesting directions. First, we observe that switching strategies perform better than the no-short-sale constrained strategies. This interesting observation seems to suggest that predicting the sign of the market expected excess return is more profitable than predicting the magnitude. However, further analysis is necessary to support this conjecture, for example, comparing the performance of a model that only predicts the sign to a model that predict both such as ours. Second, we assume perfect foresight and ignore estimation risk. We suspect that taking into account parameter uncertainty may reduce the performance, and, therefore, it remains to be seen that if the outperformance is robust to this uncertainty. Third, we consider predictability at monthly level; it may be of interest considering predictability at quarterly and even annual levels. For example, Lettau and Ludvigson (2001) find that the consumption-wealth ratio is a very powerful predictive variable, which is only available at quarterly and annual level. Finally, our results suggest that the predictive relation seems unstable. An alternative approach is to explicitly model time-varying parameters or structural breaks. It is well known that regime shifts happen with T-bill yield. Rapach and Wohar (2004) find significant evidence of structural breaks in seven of eight predictive regressions of S&P 500 returns. Pesaran and Timmermann (2002) find that a linear predictive model that incorporates structural breaks has improved out-of-sample statistical forecasting power. However, Lettau and Nieuwerburgh (2008) show that shifts in regimes make it hard to exploit return predictability out of sample. Therefore, it remains to see if incorporating regime switching can improve the economic gains.

REFERENCES

Admati, Anat R., Sudipto Bhattacharya, Paul Pfleiderer, and Stephen A. Ross, 1986. On timing and selectivity. *Journal of Finance* **41**, 715 – 730.

Aiolfi, Amrco, and Carlo A. Favero, 2003. Model uncertainty, thick modelling and the predictability of stock returns. *Journal of Forecasting* **24**, 233 – 254.

Ang, Andrew, and Geert Bekaert, 2007. Stock return predictability: Is it there? *The Review of Financial Studies* **20**, 651 – 707.

Avramov, Doron, 2002. Stock return predictability and model uncertainty. *Journal of Financial Economics* **64**, 431 – 467.

Balduzzi, Pierluigi, and Anthony W. Lynch, 1999. Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics* **52**, 47 – 78.

Bekaert, Geert, and Robert J. Hodrick, 1992. Characterizing predictable components in excess returns on equity and foreign exchange markets. *Journal of Finance* **47**, 467 – 509.

Bossaerts, Peter, and Pierre Hillion, 1999. Implementing statistical criteria to select return forecasting models: What do we learn? *The Review of Financial Studies* **12**, 405 – 428.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael Roberts, 2007. On the importance of measuring payout yield: Implications for empirical asset pricing. *Journal of Finance* **62**, 877 – 915.

Breen, William, Lawrence R. Glosten, and Ravi Jagannathan, 1989. Economic significance of predictable variations in stock index returns. *Journal of Finance* **44**, 1177 – 1189.

Campbell, John Y., 1987. Stock returns and the term structure. *Journal of Financial Economics* **18**, 373 – 399.

Campbell, John Y., 2000. Asset pricing at the millennium. *Journal of Finance* **55**, 1515 – 1567.

Campbell, John Y., Yeung Lewis Chan, and Luis M. Viceira, 2003. A multivariate model of strategic asset allocation. *Journal of Financial Economics* **67**, 3 – 40.

Campbell, John Y., and Robert J. Shiller, 1988. The dividend price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies* **1**, 195 – 228.

Campbell, John Y., and Robert J. Shiller, 1988. Stock prices, earning, and expected dividends. *Journal of Finance* **43**, 661 – 676.

Campbell, John Y., and Samuel B. Thompson, 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies* **21**, 1509 – 1532.

Carlson, Murray, David A. Chapman, Ron Kaniel, and Hong Yan, 2004. Specification error, estimation risk, and conditional portfolio rules. Unpublished working paper, University Texas at Austin.

Cochrane, John H., 1991. Production-based asset pricing and the link between stock returns and economic fluctuations. *Journal of Finance* **46**, 209 – 235.

Cooper, Michael, and Huseyin Gulen, 2006. Is time-series based predictability evident in real-time? *Journal of Business* **79**, 1263 – 1292.

Cooper, Michael, Roberto C. Gutierrez, Jr., and William Marcum, 2005. On the predictability of stock returns in real time. *Journal of Business* **78**, 469 – 500.

Cremers, Martijn, 2002. Stock return predictability: a Bayesian model selection perspective. *The Review of Financial Studies* **15**, 1223 – 1250.

Engle, Robert F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of the united kingdom inflation. *Econometrica* **22**, 427 – 432.

Fama, Eugene, and William Schwert, 1977. Asset returns and inflation. *Journal of Financial Economics* **5**, 115 – 146.

Fama, Eugene F., and Kenneth R. French, 1988. Dividend yields and expected stock returns. *Journal of Financial Economics* **22**, 3 – 25.

Fama, Eugene F., and Kenneth R. French, 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* **25**, 23 – 49.

Ferson, Wayne E., 1989. Changes in expected security returns, risk, and the level of interest rates. *Journal of Finance* **44**, 1191 – 1217.

Ferson, Wayne E., and Campbell R. Harvey, 1991. The variation of economic risk premiums. *Journal of Political Economy* **99**, 385 – 415.

Ferson, Wayne E., Sergei Sarkissian, and Timothy Simin, 2003. Spurious regressions in financial economics. *Journal of Investment Management* **1**, 1 – 10.

French, Kenneth, William Schwert, and Robert Stambaugh, 1987. Expected stock returns and volatility. *Journal of Financial Economics* **19**, 3 – 19.

Gallant, Ronald, and George Tauchen, 1989. Seminonparametric estimation of conditionally constrained heterogeneous process: Asset pricing applications. *Econometrica* **57**, 1091 – 1120.

Gallant, Ronald, and George Tauchen, 1997. SNP: A program for nonparametric time series analysis. Version 8.7.

Giannetti, A., 2007. The short term predictive ability of earnings-price ratios: The recent evidence (1994-2003). *The Quarterly Review of Economics and Finance* **47**, 26 – 39.

Goyal, Amit, and Ivo Welch, 2008. A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies* **21**, 1455 – 1508.

Graham, John R., and Campbell R. Harvey, 1997. Grading the performance of market-timing newsletters.*Financial Analysts Journal* **53**, 54 – 68.

Granger, Clive W. J., and M. Hashem Pesaran, 2000. Economic and statistical measures of forecast accuracy. *Journal of Forecasting* **19**, 537 – 560.

Guo, Hui, 2006. On the Out-of-Sample Predictability of Stock Market Returns. *Journal of Business* **79**, 645 – 670.

Guo, Hui, and Robert Savickas, 2006. Understanding stock return predictability. working paper. Unpublished working paper, University of Cincinnati.

Handa, Puneet, and Ashish Tiwari, 2006. Does stock return predictability imply improved asset allocation and performance? *Journal of Business* **79**, 2423 – 2468.

Harvey, Campbell, 1989. Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics* **24**, 289 – 317.

Hodrick, Robert J., 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *The Review of Financial Studies* **5**, 357 – 386.

Jegadeesh, Narasimhan, 1990. Evidence of predictable behavior of security returns. *Journal of Finance* **45**, 881 – 898.

Kandel, Shmuel, and Rober F. Stambaugh, 1996. On the predictability of stock returns: An asset-allocation perspective. *Journal of Finance* **51**, 385 – 424.

Keim, Donald B., and Robert F. Stambaugh, 1986. Predicting returns in the stock and bond markets. *Journal of Financial Economics* **17**, 357 – 390.

Lamont, Owen, 1998. Earnings and expected returns. *Journal of Finance* **53**, 1563 – 1588.

Lettau, Martin, and Sydney Ludvigson, 2001. Consumption, aggregate wealth and expected stock returns. *Journal of Finance* **56**, 815 – 849.

Lettau, Martin, and Stijn Van Nieuwerburgh, 2008. Reconciling the return predictability evidence. *Review of Financial Studies* **21**, 1607 – 1652.

Lewellen, Jonathan, 1999. The time-series relations among expected return, risk, and book-to-market. *Journal of Financial Economics* **54**, 5 – 43.

Marathe, Achla, and Hany A. Shawky, 1994. Predictability of stock returns and real output. *The Quarterly Review of Economics and Finance* **34**, 317 – 331.

Marquering, Wessel, and Marno Verbeek, 2004. The economic value of predicting stock index returns and volatility. *Journal of Financial and Quantitative Analysis* **39**, 407 – 429.

Masih, Rumi, A. Mansur, M. Masih, and Kilian Mie, 2008. Model uncertainty and asset return predictability: an application of bayesian model averaging. *Applied Economics* **1**, 1 – 10.

Merton, Robert C., 1980. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* **8**, 323 – 361.

Nelson, Daniel, 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59**, 347 – 370.

Pástor, Ľuboš, and Robert Stambaugh, 2000. Comparing asset pricing models: An investment perspective. *Journal of Financial Economics* **56**, 335 – 381.

Pesaran, M. Hashem, and Allan Timmermann, 1995. Predictability of stock returns: Robustness and economic significance. *Journal of Finance* **50**, 1201 – 1228.

Pesaran, M Hashem, and Allan Timmermann, 2002. Market timing and return predictability under model instability. *Journal of Empirical Finance* **8**, 495 – 510.

Rapach, David E., and Mark E. Wohar, 2004. Structural breaks and predictive regression models of aggregate u.s. stock returns. *Journal of Financial Econometrics* **4**, 238 – 274.

Santos, Tano, and Pietro Veronesi, 2006. Labor income and predictable stock returns. *Review of Financial Studies* **19**, 1 – 44.

Schwert, G. William, 1989. Why does stock market volatility change over time? *Journal of Finance* **44**, 1115 – 1153.

Stambaugh, Robert F., 1999. Predictive regressions. *Journal of Financial Economics* **54**, 375 – 421.

Torous, Walter, and Rossen Valkanov, 2000. Boundaries of predictability: Noisy predictive regressions. Unpublished working paper, UCLA.

Treynor, Jack L., and Kay Mazuy, 1966. Can mutual funds outguess the market? *Harvard Business Review* **44**, 131 – 136.

Tu, Jun, and Guofu Zhou, 2004. Data-generating process uncertainty: What difference does it make in portfolio decisions? *Journal of Financial Economics* **72**, 385 – 421.

Wachter, J. A., and M. Warusawitharana, 2009. Predictable returns and asset allocation: Should a skeptical investor time the market? *Journal of Econometrics* **148**, 162 – 178.

Xu, Yexiao, 2004. Small levels of predictability and large economic gains. *Journal of Empirical Finance* **11**, 247 – 275.