Return Predictability, Economic Profits, and Model Mis-Specification: How Important are the Better Specified Models?*

Yufeng Han†

Tulane University

December 2006

Abstract

This paper addresses the question whether investors can profit from return predictability in the real world while focusing on the impact of the data-generating process (DGP). We estimate an array of predictive models ranging from the simplest VAR to nonparametric ones and evaluate their out-of-sample portfolio performance with various predictive variables. We find that despite the significant statistical improvement the better specified predictive models do not consistently outperform the VAR. Another striking finding is that investors appear to be better off predicting only the sign but not the magnitude of the market expected excess returns.

Keyword: return predictability, economic value, model mis-specification, data-generating process, VAR, GARCH, seminonparametric model, Model selection criteria, recursive estimation, out of sample, portfolio performance, no-short-sale, switching strategy, market timing

*This paper is based on the second essay in my dissertation. I wish to thank Kerry Back, Michael Brandt, Robert Hansen, Pedro Santa-Clara, John Scruggs, Neal Maroney, Janathan Taylor, Guofu Zhou, and seminar participants at the CRSP 2006 Forum, MFA 2005 meeting, University of New Orleans, and Washington University in St. Louis for helpful comments. All remaining errors are solely mine.

†**Correspondence:** Yufeng Han, A. B. Freeman School of Business, Tulane University, 7 McAlister Drive, New Orleans, LA 70118. Tel: (504) 865-5479. Fax: (504) 862-8327. Email: yhan@tulane.edu.

Return Predictability, Economic Profits, and Model Mis-Specification: How Important are the Better Specified Models?

Abstract

This paper addresses the question whether investors can profit from return predictability in the real world while focusing on the impact of the data-generating process (DGP). We estimate an array of predictive models ranging from the simplest VAR to nonparametric ones and evaluate their out-of-sample portfolio performance with various predictive variables. We find that despite the significant statistical improvement the better specified predictive models do not consistently outperform the VAR. Another striking finding is that investors appear to be better off predicting only the sign but not the magnitude of the market expected excess returns.

Introduction

Can an investor profit from predicting the market using the public information in the real world? This interesting question has received much attention recently due to extensively documented evidence that many economic variables such as the dividend yield and the term spread predict future stock returns over time.[1] However the answer to this question is still largely unsettled. For example, Kandel and Stambaugh (1996) and Campbell, Chan, and Viceira (2002) evaluate the economic benefits via *ex ante* calibration and find that the economic gains are significant. In contrast, studies that evaluate the *ex post* performance of return predictability find mixed results. For example, Breen, Glosten, and Jagannathan (1989) and Pesaran and Timmermann (1995) find that return predictability yields significant economic gains out of sample. On the other hand, Cooper, Gutierrez, and Marcum (2001) and Cooper and Gulen (2001) fail to find any economic significance. Similarly, while Jacobsen (1999) and Marquering and Verbeek (2001) find that the economic gains of exploiting return predictability are significant in the mean-variance framework, Handa and Tiwari (2004) find that the economic significance of return predictability is questionable.

We try to address this question from the perspective of data-generating process and model misspecification. We evaluate the importance of the data-generating process in two dimensions. First, does a better specified model improve the portfolio performance of return predictability and thus generate better economic value? In other words, do statistical gains automatically translate to economic gains? Second, how important are the specifications of the data-generating process in determining the economic profits of return predictability relative to other factors such as the choice of predictive variables and portfolio constraints? Ultimately, we are interested in how investors can improve the out-of-sample portfolio performance or economic profits when they try to profit from predicting the market with the predictive variables?

One problem in the study of return predictability is that the predictive model can be misspecified. The literature has been focused on the choice of predictive variables (see, e.g. Pesaran and Timmermann, 1995; Bossaerts and Hillion, 1999; Avramov, 2002; Cremers, 2002; Roskelley, 2004), which is certainly

[1] See, e.g., Fama and Schwert (1977), Keim and Stambaugh (1986),Campbell (1987), French, Schwert, and Stambaugh (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), Harvey (1991), Ferson and Harvey (1991), Cochrane (1991), Hodrick (1992), Bekaert and Hodrick (1992), Lamont (1998), Lewellen (1999), Lettau and Ludvigson (2001), Shanken and Tamayo (2001), Santos and Veronesi (2001), Cremers (2002), Avramov (2002), and Goyal and Welch (2003).

very important, but little attention has been paid to the specification of the underlying data-generating process. In most studies, the predictive model is simply specified as a first-order vector autoregression (VAR) or a further simplified predictive linear regression, which is subject to estimation bias as discussed by Ferson, Sarkissian, and Simin (1999) and Stambaugh (1999). Nevertheless, this simple model is the most likely misspecified. For example, the specification of a constant variance for stock returns is firmly rejected by the extensive empirical evidence that stock volatilities are time-varying. To extend the VAR model to incorporate time-varying volatility, we add a GARCH feature to the model, which we refer to as the VAR-GARCH model. To our best knowledge, our analysis seems to be the first application of the VAR-GARCH model to the area of return predictability although it has been applied to other areas such as dynamics of interest rates and exchange rates where the focus is on the GARCH effect.[2] This extension also incorporates another important source of predictability not present in the VAR model, that is, the predictability of the conditional second moments of stock returns. Therefore, the VAR-GARCH model allows us to consider the combined effect of predicting both expected return and volatility and to obtain more accurate estimation of the conditional distributions of returns, which *a priori* may result in better portfolio decisions.

Nevertheless, both the VAR and VAR-GARCH models are restricted by two assumptions. The first is that both models assume the (conditional) distributions of returns are normally distributed.[3] As a result, only the mean (VAR) and variance (VAR-GARCH) need to be modeled. This assumption is also inconsistent with the data and empirical tests. The second assumption is that stock returns are linear functions of predictive variables. However, it is likely that the true relation between stock returns and predictive variables is nonlinear, and hence the estimated conditional distributions of returns are likely biased. To overcome these limitations, we consider the seminonparametric (SNP) model proposed by Gallant and Tauchen (1989), which relies on Hermite polynomial expansions to approximate the underlying data-generating process and thus is capable of capturing many features of the data. In addition, the SNP model nests the VAR and VAR-GARCH models as special cases, thus facilitating model comparison and selection. As in Bossaerts and Hillion (1999), to further guard again possible model mis-specification, we use statistical model selection criteria to choose the best specification for each type of models.

[2] Recently Tang (2003) uses the VAR-GARCH model as the true data-generating process to evaluate the performance of Bayesian model averaging methodology introduced to the literature by Avramov (2002) and Cremers (2002).

[3] The error terms in GARCH models are often assumed to be normally distributed, even tough GARCH models can incorporate non-normal error terms.

Based on the above proposed predictive models, we conduct portfolio analysis for three different types of risk averse investors, whose preferences are represented by the power utility, who allocate funds to the market portfolio (S&P 500 index) and the 1-month Treasury bill to maximize their expected utilities. The first investor believes that asset returns are predictable by the economic variables and forms her optimal portfolios based on the proposed predictive models. As a benchmark, the second investor believes that asset returns can only be described as random walks and chooses her optimal portfolio weights according to a simple normal I.I.D model. However, because of estimation errors, we do allow the IID benchmark investor to update the estimates of the mean and variance recursively. As another benchmark, the third investor believes that market returns have certain dynamics but cannot be predicted by other economic variables. Therefore, this market-dynamics investor will estimate certain dynamic models using the market returns but will not use any predictive variables. We also consider the passive buy-and-hold strategy as an additional benchmark. The portfolio analysis is conducted both in sample and out of sample with the focus on the latter, and we consider two cases: 1) no restriction on the portfolio weights; 2) no-short-sale constraint which restricts the portfolio weights between 0 and 1. The portfolio performance is measured *ex post* by various performance measures including the Sharpe ratio, a measure proposed by Graham and Harvey (1997), and the certainty equivalent rate of return (CER).

We consider a number of variations of predictive variables, although incorporating model uncertainty in the aspect of choosing predictive variables is beyond the scope of this study. Among the numerous empirically identified predictive variables, the most popular one is the dividend yield; many studies including Aït-Sahalia and Brandt (2001), Barberis (2000), and Balduzzi and Lynch (1999), use the dividend yield as the predictor. Other variables we choose are 3-month T-bill yield, term spread (yield difference between a 10-year Treasury Bond and a 1-year Treasury Bill), and default spread (yield difference between Moody's AAA-rated and BAA-rated corporate bonds).[4]

We find several interesting results. First, the VAR model is clearly misspecified, adding the GARCH feature substantially improves the goodness-of-fit of the model, and the SNP model tends to be the best overall statistical model. For example, with the T-bill yield as the predictor, the Bayesian information criterion (BIC, smaller is better) for the VAR, VAR-GARCH, SNP, and an overfitted SNP that incorporates

[4]The potential drawback is the problem of data snooping as the same data used to identify the predictive variables is used to test the predictive power of the predictive variables. This problem is mitigated to some extent by conducting out-of-sample testing, but would not be completely solved unless a substantially long period of new data become available.

more nonlinearity is 0.92, 0.29, 0.27, and 0.35, respectively. Clearly, the reduction of BIC is the most when the GARCH feature is added. Therefore, conditional heteroscedasticity is an indispensable feature of the data. Modeling nonlinearity or higher moments also helps, but the additional improvement is not as dramatic as modeling the second moment. Despite the significant statistical improvement over the VAR, however, VAR-GARCH and SNP predictive models do not consistently perform better than the VAR measured by various portfolio performance measures, either in sample or out of sample. For example, with the T-bill yield as the predictor, the four predictive models yield in-sample Sharpe ratios of 0.55, 0.56, 0.53, and 0.51 and out-of-sample Sharpe ratios of 0.49, 0.49, 0.48, 0.47, respectively under the no-short-sale constraint. In particular, incorporating GARCH feature, although resulting in much statistical improvement, does not always yield superior portfolio performance. Our results are in sharp contrast to the findings of Carlson, Chapman, Kaniel, and Yan (2004) who examine the utility cost of ignoring volatility dynamics through calibration analysis and find that volatility related specification errors are economically significant. Among others, the biggest difference is that Carlson, Chapman, Kaniel, and Yan (2004) employ *ex ante* simulation studies, and thus the relevance to the real world performance is unclear, whereas we evaluate the *ex post* portfolio performance. On the other hand, the misspecified VAR model often performs well and sometimes the best, whereas the best statistical model many times performs the worst. These results suggest that even though the VAR model is misspecified, it may be the preferred model to use when studying the portfolio performance of return predictability due to its simplicity and good out-of-sample performance. However, it should be noted that the specifications of the best VAR model often differ from the one normally assumed in the literature, i.e., the order of the best VAR model is often higher than the first order.

The result that the better specified predictive models will not necessarily yield better portfolio performance than a simple VAR model is a bit striking. Two possible explanations may account for the lack of performance improvement for the better specified models. First, portfolio performance seems fairly insensitive to the underlying data-generating process. In other words, strategies based on two very different data-generating processes may nevertheless yield very similar performance. For example, Pástor and Stambaugh (2000) and Tu and Zhou (2003) both show that similar portfolio performance may be obtained from different data-generating processes. Among others, this study differs from those two in two important aspects. First, we examine predictive models whereas Pástor and Stambaugh (2000) compare different pricing models and Tu and Zhou (2003) examine the impact of data-generating process uncertainty. Second, we evaluate *ex post* portfolio performance, whereas both studies evaluate *ex ante* performance. The second explanation, as argued by Aït-Sahalia and Brandt (2001), is that different objective functions may place different emphases on the various features of the conditional return distribution. Therefore the ranking by statistical criteria may not be the same as the ranking by financial criteria. This highlights the importance of choosing the right criteria for different problems.

Second, we find strong in-sample evidence that the predictability investor would do much better than the IID investor and the market-dynamics investor. More specifically, portfolios based on the predictive models significantly outperform the benchmark portfolios based on a fixed-weight strategy, a passive buy-and-hold strategy and dynamic strategies based on modeling the market returns alone. This is true for all variations of predictive variables considered except the dividend yield alone and to a less extent default spread alone. However, evidence of superior out-of-sample performance is very limited. Indeed, no portfolio based on any predictive model with any variation of predictive variables outperforms the benchmark portfolios when the portfolio weights are not constrained. Superior performance is possible only when the no-short-sale constraint is imposed. For example, under the no-short-sale constraint, the VAR-GARCH model of the T-bill yield has a Sharpe ratio of 0.56 and a risk-adjusted abnormal return of 2.59% per annum, compared to 0.43 and 0.74% respectively for the passive buy-and-hold strategy, whereas the same model produces a Sharpe ratio of 0.08 and a risk-adjusted abnormal return of -4.67% without the constraint. Even with the no-short-sale constraint, however, not all variations of predictive variables outperform the benchmarks. The T-bill yield alone generates superior portfolio performance, whereas the default spread, term spread, and in particular the dividend yield, do not seem to possess any predictive power alone measured by portfolio performance. On the other hand, the default spread and term spread seem to complement the T-bill yield in that combining each with the T-bill yield produces stronger portfolio performance than the T-bill yield alone, which suggests that these two variables may provide additional useful information beyond what the T-bill yield provides. For example, the VAR-GARCH model of the T-bill yield and default spread combination produces a Sharpe ratio of 0.62 and a risk-adjusted return of 3.63% per annum, both of which are higher than those produced by the T-bill yield alone (see above). It is worth noting that the dividend yield, however, has no predictive power either in sample or out of sample, which suggests that future research should avoid using the dividend yield as the predictive variable.[5] Finally, the out-of-sample results are very robust. We obtain similar results when we vary the number of years between re-estimation in the recursive estimation, the investment horizon, the degree of risk aversion, and the utility function. For example, when we assume instead mean-variance preference, the results are similar.

The apparently large performance gap between the unconstrained portfolios and no-short-sale constrained portfolios in the out-of-sample analysis may not come as a surprise since the unconstrained portfolios often have very large long or short positions, presumably due to estimation errors. However, when we examine the portfolio performance under other constraints, in particular Regulation T of the Federal Reserve Board, we find that the predictive models still underperform the benchmarks even though the weights are much less variable. A close examination of the weights under the no-short-sale constraint reveals that about 70% times, the weights are either 0 or 1, which leads us to examine the switching strategy that switches between the market portfolio and the 1-month T-bill depending on whether the forecasted expected excess returns are positive or negative. The switching portfolios indeed produce even stronger performance than the no-short-sale constrained portfolios. For example, the switching portfolio based on the VAR model of the T-bill yield and default spread combination produces a risk-adjusted abnormal return of 5.31% per annum, which is 1.24% higher than that produced under the no-short-sale constraint, and a Sharpe ratio of 0.74 higher than that of 0.65 under the constraint. These results suggest that investors are better off predicting the sign not the magnitude of the market expected excess returns. In other words, the predictive variables may provide useful information about the sign of the market expected excess returns but fail to provide any useful information about the magnitude due to large estimation errors. A possible explanation is related to the fact that, as pointed out by Merton (1980), there is too much noise in the observed returns to accurately estimate the expected returns even if the predictive relation holds. Indeed, Torous and Valkanov (2000) argue that even if returns are predictable, the noise in the predictive regression may overwhelm the signal of the conditional variables. Alternatively, it is likely that models considered here are misspecified despite our effort to minimize specification error by using the flexible seminonparametric specification.

Another possible explanation is that the predictive relationship is unstable and changes over time,

[5] Ang and Bekaert (2006) find that the short rate is the only one having the statistically significant coefficient in a predictive regression with dividend yield and earning yield. Both Ang and Bekaert (2006) and Goyal and Welch (2003) find that dividend yield has no predictability in statistical tests.

which causes large estimation errors. We find some evidence supporting this. For example, we find that the correlations between the market excess returns and predictive variables are unstable and the optimal specifications of the predictive models change in different periods. Further evidence comes from the subperiod performance analysis. We find that in the 25 years from 1979 to 2003, the predictive models always outperform the benchmarks measured by accumulative performance, but the level of outperformance decreases gradually. This is due to the poor performance of the predictive models in some periods, for example, the last ten years from 1994 to 2003, and in particular the last recession period around 2001. Additional evidence is provided by previous studies. For example, Pesaran and Timmermann (1995) find better predictive performance in the volatile period of 1970's, and Handa and Tiwari (2004) find poor predictive performance in the most recent period after 1989 whereas favorable results during the periods of 1959–1973 and 1974–1988. Abhyankar and Davies (2002) find that the predictive ability of T-bill yield is low prior to the 1951 Treasury Accord Act period, high during the period of 1950-1975, and has disappeared in the last two decades.

Our findings may explain the mixed results obtained by previous studies. For example, the reason that Breen, Glosten, and Jagannathan (1989), Pesaran and Timmermann (1995), and Abhyankar and Davies (2002) find out-of-sample economic significance is because all the three studies use the switching strategy. Marquering and Verbeek (2001) find significant economic gains from predicting market returns because the short-selling on the market portfolio is prohibited, whereas Handa and Tiwari (2004) find no consistently significant superior performance because (limited) short-selling is allowed in their study. On the other hand, none of the studies examine the portfolio performance of the two cases simultaneously - by comparing the performance of the unconstrained and no-short-sale constrained portfolios side by side, we are able to show that predictive variables such as the T-bill yield can only predict the direction of market movements but not the magnitude, which is a very interesting and new result.

The remainder of this article is organized as follows. Section 1 discusses the group of predictive models that incorporate return predictability. Section 2 describes the predictive variables and discusses the estimation results of the predictive models. Section 3 discusses investor's optimal portfolio choice problem. Section 4 conducts in-sample as well as out-of-sample portfolio analysis to examine the performance of the various predictive models. Section 5 concludes.

1 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 (1999) 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., French, Schwert, and Stambaugh, 1987; 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, even though it has been used in other areas.

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 nonlinear functions of the predictors. Another relevant advantage of the SNP nonlinear model is that it nests both the VAR and VAR-GARCH models as degenerated cases, which makes it easy to compare and select different 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

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

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. It is easy to 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 a polynomial of degree K_x in x because of the extraordinary heteroscedasticity. This model is fully nonlinear and nonparametric. Collectively, the parameters L_μ, L_g, L_r, K_z, and K_x uniquely identify the SNP model,[6] and hence we use "$[L_\mu L_g L_r K_z K_x]$" to denote the specification of a predictive model. For example, [10000] denotes VAR(1), while [11100] denotes VAR(1)-GARCH(1,1).

One advantage of our framework is that we can systematically select the best model specification for

[6]There are two additional parameters, I_z and I_x, used to reduce the cross-interaction terms in the polynomials when y is multivariate. The highest order for cross-interaction terms is $K_z - I_z$ and $K_x - I_x$ respectively.

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. This systematical approach is far superior to the *ad hoc* assumption that the data follows certain process 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. Since the BIC tends to be conservative, additional statistical criteria are also considered including Akaike's Information Criterion (AIC) and Hannan and Quinn Criterion (HQ) defined as

$$\begin{aligned} AIC &= \frac{-2\mathcal{L}_k + 2k}{n}, \\ HQ &= \frac{-2\mathcal{L}_k + k \log\log n}{n}. \end{aligned} \tag{4}$$

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. It is easy to see that BIC has the most severe penalty for rich parameterization, whereas AIC has the least severe penalty, and HQ is in between. It turns out that the fully nonlinear and nonparametric model is always rejected by the BIC due to its rich parameterization but sometimes is favored by the AIC. In the sequel we denote the best SNP model as OPT and the fully nonlinear and nonparametric 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 restrict to linear regression models, whereas we consider a more broad class of models including both linear and nonlinear ones.

2 Estimation of the Predictive Models

2.1 Data Description

In recent years, empirical literature has identified many economic variables that have predictive power over stock and bond returns. These variables include term spread, dividend yield, Treasury-bill yield, 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 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 3-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. Monthly data 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.

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 1st order autocorrelation coefficients as high as 0.989.

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 skewed in the second subperiod than in the first sub-period. 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 testing 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.[7] One interesting result in Panel B is that the correlations are not stable over time. For example, the correlation between the excess return 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. For the default spread and term spread, the correlations are much weaker in the second sub-period. On the other hand, the correlation between the excess return and T-bill yield is relatively stable and remains considerable. These differences in correlations are consistent with our 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 shows 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. It is easy to 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. 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 1960's. As we will see later these observations have direct implications on the portfolio performance.

[7] 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.

2.2 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 predictive variables is selected according to the Bayesian Information Criterion (BIC). However, because the BIC tends to be conservative, we also consider other statistic criteria such as Akaike's Information Criterion (AIC) and Hannan-Quinn Criterion (HQ).

Table 2 reports the best specifications for the whole sample period (Panel A: Full Sample Period) and the first subperiod (Panel B: Estimation Period).[8] Several interesting results emerge. First, the predominantly used VAR model is clearly misspecified. 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 in the whole sample period 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 drastically 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, in the full sample period, 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.

[8]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.

Third, a polynomial of degree 4 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.

Finally, comparing Panel A and B, we find that in some cases the best specifications are different in the two periods. For example, the best GARCH specification for T-bill yield and default spread combination in the whole sample period is GARCH(1,1), whereas in the estimation period (1947:1–1978:12) it is GARCH(2,1). Even when the specifications remain the same, the parameter estimates are often different. This result suggests that the relationships between the market returns and predictive variables may change over time.

3 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 estimate of the conditional distributions of future stock returns.

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 future excess return and portfolio weight on the market portfolio, respectively, and $r_{f,t+1}$ is the return on the riskless asset.

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 Brent method with analytic derivatives.

In the presence of transaction costs, the investor will choose the optimal portfolio weights taking into consideration the costs associated with rebalancing the weights. Assume the proportional transaction cost is τ for the market portfolio, and there is 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,

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

In addition to the one-period problem, we also examine portfolio performance at longer investment horizons, for example, 3 months or 6 months. In general, if the investor has an investment horizon of $\hat{T}$ periods, then the forecast of the one-period-ahead excess return $r_{t+1|t}$ is replaced by the forecast of the cumulative excess return over the $\hat{T}$ periods, $R_{t+\hat{T}|t}$,

$$R_{t+\hat{T}|t} = r_{t+1|t} + r_{t+2|t} + \cdots + r_{t+\hat{T}|t}, \tag{11}$$

and the investor's problem is to solve

$$\max_{\omega_t} \frac{1}{N} \sum_{i=1}^{N} \frac{W_t^{1-\gamma}[\omega_t e^{R^i_{t+\hat{T}|t} + r_{f,\hat{T}}} + (1-\omega_t) e^{r_{f,\hat{T}}}]^{(1-\gamma)}}{1-\gamma}. \tag{12}$$

4 *Ex Post* Portfolio Performance of the Predictive Models

Having determined the best specifications for the predictive models, the predictability investor forecasts and generates sample draws from the one-step-ahead conditional return distributions, conditioning on the previous realized returns and predictors, and then uses the sample draws to find the optimal portfolio weights as described in Section 3. 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 $\mathcal{GH}2$). Note that we compute and compare *ex post* CERs or sample CERs instead of *ex ante* CERs used in studies such as Kandel and Stambaugh (1996), Pástor (2000), and Pástor and Stambaugh (2000). The sample CERs are calculated by taking the average of the realized utilities over the period considered;

$$\mu(W_0(1 + r_{ce})) = \frac{1}{T} \sum_{t=1}^{T} \mu(W_0(1 + r_{p,t})), \tag{13}$$

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 1-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 find the weight ω to solve the following problem:

$$V_m = \mathbf{Var}(\omega r_{pt} + (1-\omega) r_{ft}) = \omega^2 V_p + (1-\omega)^2 V_f + 2\omega(1-\omega)\mathbf{Cov}(r_{pt}, r_{ft}), \tag{14}$$

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

$$\mathcal{GH}2 = \omega \bar{r}_p + (1-\omega)\bar{r}_f - \bar{r}_m, \tag{15}$$

where $\bar{r}_p$, $\bar{r}_f$, and $\bar{r}_m$ are the average returns on the managed portfolio, 1-month T-bill, and the market portfolio respectively. $\mathcal{GH}2$ is related to the Sharpe ratio[9] but unlike the Sharpe ratio, it also quantifies the outperformance. Note that when the average return is lower than the riskfree rate, however, the Sharpe ratio will be negative and can no longer be used to rank performance, and $\mathcal{GH}2$ will overestimate the outperformance.

4.1 In-Sample Portfolio Performance Analysis

The first step in our empirical portfolio analysis is to conduct the in-sample tests. If the predictive models cannot outperform the benchmarks in the in-sample tests, then there is no need to conduct further out-of-sample tests. The in-sample tests are conducted using the whole sample period from 1947:1 (or 1953:4 when the term spread is involved) to 1998:12 as the estimation period as well as the testing period. Table 3 reports the in-sample performance results. Results are reported for both cases where no constraint is imposed on the portfolio weights and where no short sale is allowed. The benchmark strategies are the fixed-weight strategy where the returns are perceived IID and therefore the weight is always rebalanced to keep constant, and the passive buy-and-hold strategy where the weight is determined at the beginning of the period and no rebalance of the portfolio is required thereafter. In addition, we consider two dynamic strategies from the market-dynamics investor, which are based on the VAR and VAR-GARCH of the market returns, respectively.

[9] $\mathcal{GH}2$ 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 there is no direct mathematical relation between $\mathcal{GH}2$ and the Sharpe ratio.

The most prominent result from Table 3 is that the predictive models outperform the benchmarks across the board. T-bill yield, default spread, term spread, and various combinations of the three variables all yield superior performance with every predictive model except the OPT and NLNP models of the default spread where the performance is close to that of the benchmarks. In addition, imposing the no-short-sale constraint often reduces the portfolio performance. For example, the VAR model of the T-bill yield produces a Sharpe ratio of 0.58, a risk-adjusted abnormal return of 4.40% per annum and a CER of 10.29% per annum, in comparison with 0.29, 0.48%, and 6.14%, respectively, of the fixed-weight strategy, or 0.32, 0.79%, and 6.43%, respectively, of the market VAR model. Imposing the no-short-sale constraint reduces the performance to 0.55, 4.07%, and 8.85%, respectively. However, the dividend yield is an exception - it fails to outperform the benchmarks except in its VAR model. This is in sharp contrast to other predictive variables and suggests that dividend yield does not seem to have any predictive power over the market returns. This result is consistent with the weak correlation between the dividend yield and the market returns. Note that the default spread alone also has somewhat weak performance, which is consistent with its low overall correlation with the market returns. On the other hand, T-bill yield seems to yield the best performance and thus to be the most powerful predictor, followed by term spread and then default spread. In addition, term spread and default spread seem to complement the T-bill yield as the combination of each one with the T-bill yield produces even stronger performance.[10] For example, with the VAR model, T-bill yield and default spread combined yield a Sharpe ratio of 0.69, a risk-adjusted abnormal return of 5.98%, and a CER of 13.04% per annum, all of which are higher than the performance numbers of the T-bill yield cited above.

The second interesting result is that the performance differences among the four predictive models are generally small, and no consistent pattern can be found. This striking result is unexpected because the four models are statistically very different. In particular, even though the VAR model is clearly misspecified, the portfolio performance of the VAR model is on par with other better specified models. On the other hand, the OPT model - the best overall statistical model - does not perform quite as well as others and often the worst. In addition, the VAR-GARCH model often performs behind the VAR model despite the significant statistical improvement of the GARCH feature. Finally, the in-sample performance is very robust to different performance measures.

[10]Unlike the default spread, combining the dividend yield with T-bill yield does not produce stronger performance, and thus the results are not reported.

4.2 Out-of-Sample Portfolio Performance Analysis

Although it is encouraging that the in-sample tests report strong outperformance for the predictive models, these results are subject to look-ahead bias and other estimation problems as the estimation period is also used as the testing period, and thus the results may not be relevant to the real-world performance. To evaluate the real-world performance of the predictive models, we shall assume that the investors only have historical records of the market excess returns and predictive variables.

To perform the out-of-sample tests, we use the last 20 years in the sample period as the testing period, i.e. from 1979:1 to 1998:12, and the period before 1979:1 as the estimation period. The period from 1979 to 1998 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 1990's. We conduct two types of out-of-sample tests. In the first test, the estimation period is fixed and the investor does not update the estimates of the models, whereas in the second test, the estimation is repeated every five years with an expanding window of periods - recursive estimation. For example, the predictability investor initially uses the data from the estimation period (e.g. from 1947:1 to 1978:12) to search for the best specifications, generates sample draws from the forecasted one-period-ahead return distributions, and forms the optimal portfolios. After five years, however, the investor repeats the search for the best specification using data from the estimation period plus the most recent five years (expanding window), and uses the new specification and parameter estimates to choose the optimal portfolio weights. The investor repeats the same procedure every five years. This test is motivated by results in Table 1 and Table 2, which show that the relationships between the market returns and predictive variables are unstable over time, and the best specifications may change over different periods. Accordingly, we replace the fixed-weight strategy with the expanding-window strategy as the benchmark. We also include the dynamic strategies based on modeling the market returns alone.

The results of these two tests are similar, with the recursive estimation results being slightly better in general but worse in a few cases. In the sequel, we focus on the recursive estimation. Table 4 reports the performance results of the recursive estimation. In sharp contrast to the in-sample results, when no portfolio constraint is imposed, no predictive models yield performance superior to the benchmarks regardless of the combinations of predictive variables present. Indeed, every predictive model with each combination of predictive variables significantly underperforms the benchmarks with negative measures of the CER and $\mathcal{GH}2$. For example, the VAR-GARCH model of the default spread yields a Sharpe ratio of 0.23, a $\mathcal{GH}2$ of -2.29% per annum, and a CER of -5.11% per annum versus 0.45, 0.95%, and 9.40%, respectively, of the expanding-window strategy. Oddly enough, the most powerful predictor in the in-sample analysis, the T-bill yield, produces the worst performance - e.g. a Sharpe ratio of 0.08, a $\mathcal{GH}2$ of -4.67%, and a CER of -35.30% in the VAR-GARCH model. A close examination of the table suggests that the main reason for the failure is the tremendously high volatility of the dynamic portfolios, even though the average returns are often higher than those of the benchmarks. For example, the VAR-GARCH model of the T-bill yield mentioned above has an average return of 9.88%, which is comparable to the benchmark portfolios, but the standard deviation is 38.67%, about six times higher than those of the benchmarks.

On the other hand, perhaps the most interesting result in Table 4 is that imposing the no-short-sale constraint substantially improves the portfolio performance of the predictive models. This is also in sharp contrast to the in-sample results where imposing the constraint reduces the performance. For example, the same VAR-GARCH model of the T-bill yield mentioned above produces a Sharpe ratio of 0.56, a risk-adjusted abnormal return ($\mathcal{GH}2$) of 2.59% per annum, and a CER of 9.70% per annum under the no-short-sale constraint, which are higher than those of the expanding-window benchmark (0.45, 0.95%, and 9.40%) and of the passive buy-and-hold strategy (0.43, 0.74%, and 9.32%). However, even with the help of the constraint, only a few combinations of predictive variables manage to outperform the benchmarks, such as TBYD, TBYD and DFSD combination, and TBYD and TRSD combination. Again, the combinations of T-bill yield with other predictive variables produce stronger performance than any of the variables alone. For example, when both the T-bill yield and default spread are present, the VAR-GARCH model yields a Sharpe ratio of 0.62, a risk-adjusted abnormal return of 3.63% per annum, and a CER of 10.72% per annum, higher than those of the T-bill yield alone cited above. It is clear that imposing no-short-sale constraint drastically lowers the volatilities of the dynamic portfolios without severely reducing the average returns, which seems to account for the much improved performance. For example, the VAR-GARCH model of the T-bill yield has a standard deviation of only 5.94%, lower than those of the benchmarks and much lower than what the model has when no constraint is imposed (38.67%).

Focusing on the constrained portfolio performance, there seems to be little consistent differences among the four types of predictive models, a result similar to that in the in-sample analysis. Similarly, the best statistical SNP model (OPT) often performs the worst, whereas the misspecified VAR model performs quite well. Adding the GARCH feature to the VAR model improves performance in some cases but not consistently. 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 is 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 (2003). Both show that different data-generating processes may unnecessarily yield different portfolio performance. Among others, this study differs from those two in two important aspects. First, we examine predictive models whereas Pástor and Stambaugh (2000) compare different pricing models and Tu and Zhou (2003) examine the impact of data-generating process uncertainty. Second, we evaluate *ex post* portfolio performance, whereas both studies evaluate *ex ante* performance.

4.3 Further Investigation of the Out-of-Sample Performance

The large performance difference between the constrained portfolios and unconstrained portfolios in the out-of-sample analysis warrants further investigation. We first examine the robustness of the out-of-sample results in several dimensions, and then we examine the performance of other types of constraints.

For robustness, we first change the number of years between re-estimation in the recursive estimation analysis from five years to two years. Second, we increase the investment horizon from one month to three months or six months. Third, we change the relative risk aversion coefficient from four to ten. In all three cases, the results (not reported) are qualitatively similar. We further consider three levels of transaction costs: 0.25%, 0.50%, and 1.00%, representing low, medium, and high transaction costs. Results reported in Table 5 show that with the low level of transaction cost, there is virtually no impact on the performance. This is not surprising 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. For the high level of transaction cost, however, the negative effect is apparent. For example, the performance of the combination of T-bill yield and default spread becomes worse than the benchmarks when transaction costs are high.

Finally, we change the investor's preference from power utility to mean-variance utility. Table 6 reports the portfolio performance under this new preference. Interestingly, the results are similar to those of the power utility. For example, with the portfolio weights restricted, T-bill yield and the combination of T-bill yield and default spread outperform the benchmarks in every predictive model. However, the performance measures are slightly lower. For example, 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, whereas 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. On the other hand, the performance of the unconstrained portfolios is stronger under the mean-variance preference than under the power utility. For example, the VAR model of the term spread and the combination of T-bill yield and default spread outperforms the benchmarks, but in most other cases, the predictive models still underperform the benchmarks.

A close examination of the weights of the unconstrained portfolios shows that the weights vary widely. For example, the VAR-GARCH model of the TBYD and DFSD combination has a maximal weight of 2.63 and a minimal weight of -7.57, while the VAR model of this combination has a maximal weight of 8.11 and a minimal weight of -9.35. These wide variations are likely due to estimation errors. Therefore it may be of no surprise that imposing no-short-sale constraint improves the performance. In fact, one would assert that any constraint should work as long as it restricts the portfolio weights to a reasonable range. In Table 7 we further compare the portfolio performance under some other 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 ψ% is the 50% margin requirement. We also consider 100% and 200% margins. Another constraint considered here allows borrowing up to 100%, but excludes short selling. As shown in Table 7, imposing various constraints indeed improves the performance over the unconstrained portfolios, but Regulation T based constraints fail to outperform the benchmarks for the most part. On the other hand, allowing limited borrowing but no short selling yields superior performance to the benchmarks. For example, the combination of T-bill yield and default spread 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 4. For example, the combination of the T-bill yield and default spread has 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 ($0 \leq w \leq 2$).

4.4 Market Timing - Switching Strategy

Because the no-short-sale constraint yields the best portfolio performance so far, we further examine the portfolio weights under this constraint. It is of interest to note 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 is achieved if the portfolio weights are restricted to either pure cash position or pure equity position. The evidence suggests that it is necessary to make a finer distinction about the predictive ability of the predictive variables: ability to predict the magnitude of the market expected excess return and the ability to predict the sign of the market expected excess return. No predictive variables seem to have the ability to predict the magnitude of the market expected excess return out of sample due to estimate errors and other problems, but a few variables such as the T-bill yield appear to have the ability to predict the sign or direction of changes. In other words, T-bill yield may be used to predict whether the market will go up or down, but it can not tell investors by how much the market will move up or down. To further support this conjecture, we examine the performance of switching portfolios. By construction, switching portfolios switch from the all-equity position to the pure-cash position or vice versa depending on wether the forecasted expected excess returns are positive or negative. Therefore, switching portfolios only time the direction of the market movement - Henrikssson and Merton's (1981) type of market timing.

Table 8 reports the performance of the switching portfolios. Interestingly, all combinations except dividend yield and default spread generate superior performance to the benchmarks in at least one predictive model, and many outperform the benchmarks in all predictive models. In addition to the benchmarks used in the previous tables, we add a random switching portfolio as another benchmark. The random switching portfolio invests similarly to the switching portfolios considered here, but 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 portfolios underperform the other benchmarks significantly, with an average Sharpe ratio of 0.28 and 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%. Again, the performance of the four predictive models is close, and no model consistently outperforms the others.

Furthermore, the performance of the switching portfolios is stronger than that of the not-short-sale constrained 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-à-vis 0.55 and 2.45% per annum for the no-short-sale constrained portfolio. There is a 1.28% increase in risk-adjusted abnormal return moving from the no-short-sale constraint to the switching strategy, which is 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 portfolios beat the corresponding no-short-sale constrained portfolios. In most cases, this ratio is larger than one, indicating more frequently the switching portfolios have higher returns. The largest improvement is with the term spread, whose portfolio performance changes from underperforming to outperforming. Results in this table indeed suggest that the investor will be better off not predicting the magnitude of the market movement at all but focusing on the direction of the movement.

In Table 9, we compare the market timing performance of the unconstrained, no-short-sale constrained, and switching portfolios under various predictive models and predictive variables. Specifically, we examine the coefficient of the squared market excess returns in the quadratic regression proposed by Treynor and Mazuy (1966),

$$r_t = \alpha + \beta_1 r_{mt} + \beta_2 r_{mt}^2 + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2). \tag{16}$$

A significantly positive estimate of β_2 indicates successful market timing. Furthermore, under the conditions provided by Admati, Bhattacharya, Pfleiderer, and Ross (1986), $\beta_2 var(r_{mt})$ measures the abnormal return of market timing. The three benchmark portfolios, expanding window IID, market AR model, and market GARCH model, all have significant and positive β_2 coefficient, but the abnormal timing performance is small and close to zero. The third column in Table 9 reports the correlations between the market returns and the weights of the measured portfolios, which are essentially zeros for the three benchmark portfolios. However, some interesting observations emerge when T-bill yield is the predictor. On the one hand, the unconstrained portfolios do not possess any market timing ability even though the portfolio weights are positively correlated with the market returns. On the other hand, the no-short-sale constrained portfolios have significantly positive coefficient β_2 and positive abnormal returns. Furthermore, the switching portfolios have even higher abnormal returns of market timing. For example, the switching portfolio of VAR-GARCH model has an abnormal return of 1.39 % versus 1.22 % of the constrained portfolio of the same model. Stronger market-timing results are obtained when both T-bill yield and default spread are present. For example, the coefficient, abnormal return, and correlation are 1.76, 4.01 %, and 0.4, respectively for the constrained portfolio of the VAR model, and 2.08, 4.75 %, and 0.17, respectively for the switching portfolio of the same model. However, other variations of predictive variables do not seem to have strong market timing ability even with the switching portfolios.

4.5 Subperiod Performance of the Switching Portfolios

Since Table 1 and Table 2 suggest that the relationships between the market returns and predictive variables are unstable over time, one needs to be careful about interpreting the results. For example, the results may be specific to the testing 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 switching portfolios. Table 10 reports the results of the switching portfolios for three variations of predictive variables that have shown to outperform the benchmarks: T-bill yield and the combination of T-bill yield with default spread and term spread, respectively. We also extend the sampling period from 1998 to 2003 to include the latest recession period. This 5-year period is truly out of sample as most studies focus on periods before 2000. Two different sets of subperiod results are shown. Panel A shows the period-by-period performance, while Panel B shows the accumulative performance from January 1979 (the start of the testing period). In Panel A, the top rows report the performance of the expanding-window benchmark portfolio. 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 in the second and third 5-year periods, reaches the highest in the fourth 5-year period (1994–1998) with a Sharpe ratio of 1.10, and then sharply drops to -0.27 in the last five years (1999–2003). On the other hand, the performance of the switching portfolios of the predictive variables is very strong in the first 5-year period and remains strong in the rest periods until the last 5-year period when the performance deteriorates considerably. It is especially intriguing that the performance of the switching portfolios of the TBYD and DFSD combination remains rather stable over the first four 5-year periods, which is remarkable considering how low the Sharpe ratio of the benchmark is in the first 5-year period. For example, with the VAR model the Sharpe ratios are 0.86, 0.74, 0.72 and 0.74 for the first four periods. As a result, the switching portfolios outperform the benchmark portfolio in the first three periods but underperform the benchmark in the fourth and fifth periods, as shown by the last column that reports the difference in $\mathcal{GH}2$ between the switching portfolios and the benchmark portfolio. The difference is the highest in the first period, decreases over time, and becomes negative in the last two 5-year periods. Similarly, for the other two variations of predictive variables, there is at least one period when the switching portfolios underperform the benchmarks. It should be pointed out that in the last recession period, the Sharpe ratios are negative and thus the performance measures can no longer be used to compare the performance, otherwise, we would make incorrect inference. But simply comparing the mean and standard deviation shows that the switching portfolios underperform the benchmark.

On the other hand, the accumulative performance of the switching portfolios is superior to that of the benchmark portfolio in every period including the booming period of late 90s and the most recent recession, which demonstrates the robustness of the outperformance to different time periods. However, the outperformance is not stable but decreases over time. For example, during the first 5-year period from 1979 to 1983, the NLNP model of the TBYD and DFSD combination has a risk-adjusted abnormal return of 14.73% over the benchmark, but it only yields 3.01% risk-adjusted abnormal return during the 25-year period from 1979 to 2003.

As a robustness check, we also repeat the analysis with an extended period. Specifically, we use the first fourteen years from 1947 to 1958 as the initial estimation period, and then we recursively estimate all the predictive models every five years in the next 45 years (1959–2003) and form optimal portfolios accordingly. Table 11 reports the accumulative performance of the switching strategies of TBYD and the combination of TBYD and DFSD. If an investor starts to invest from 1959 and follows the switching strategy, she may underperform the expanding-window benchmark for the first five and ten years, but if she holds the switching portfolios longer, she will outperform the benchmark significantly. For example, if she estimates a NLNP model of TBYD and DFSD, follows the switching strategy, and holds the switching portfolio for 45 years, she will have an average return of 9.05% per annum, a Sharpe ratio of 0.39, and a risk-adjusted abnormal return of 4.62% per annum, whereas the IID benchmark investor will have an average return of 5.81%, a Sharpe ratio of 0.03, and a risk-adjusted abnormal return of -0.53%. The difference in the risk-adjusted abnormal return is 5.40% per annum. Put it differently, if the initial investment is $10,000, then an IID benchmark investor will have accumulated $136,340 at the end of year 2003, whereas the investor who follows the switching strategy will have accumulated $266,110 at the end of year 2003.

5 Conclusion

Can an investor profit from predicting the market using the public information in real world? This interesting question has attracted much attention recently. However, the answer to this question is still largely unsettled. One potential problem is that the predictive model, which is specified as a first-order vector autoregression (VAR) or a further simplified predictive linear regression, can be misspecified. In this paper, we address the following two questions. First, does a better specified model improve the out-of-sample portfolio performance of return predictability? Second, how important are the specifications of the data-generating process in determining the economic profits of return predictability relative to other factors such as the choice of predictive variables?

We propose using the VAR-GARCH model to incorporate conditional heteroscedasticity and predictability of the second conditional moment, and SNP model to allow non-normally distributed shocks and nonlinearity in the relationship between the market returns and predictive variables. We use statistical model selection criteria to choose the best specification for each of the four types of predictive models - VAR, VAR-GARCH, SNP and a generalized SNP model and compare their goodness of fit. We then conduct extensive in-sample and out-of-sample analysis to evaluate the portfolio performance of the predictive models, using the widely documented predictive variables such as the dividend yield, T-bill yield, term spread, and default spread.

We find first that the VAR model is clearly misspecified and allowing conditional heteroscedasticity substantially improves the goodness of fit. Allowing non-normality and nonlinearity further improves the goodness of fit, but not nearly as drastic as adding the GARCH feature. However, the portfolio analysis reveals that there are no clear advantages for the better specified predictive models - indeed not a single model consistently outperforms the others. On the other hand, the misspecified VAR model often performs on par with the others, whereas the best overall statistical model (OPT) many times performs the worst. These results suggest that it is important to choose the right model selection criterion; for the purpose of asset allocation, one should use an appropriate financial criterion. Our results also suggest that even though the VAR model is misspecified, it may be the preferred model to use when studying the portfolio performance of return predictability due to its simplicity and comparable out-of-sample performance. However, it should be noted that the specifications of the best VAR model often differ from the one normally assumed in the literature, i.e., the order of the best VAR model is often higher than the first order.

Second, while we find strong in-sample performance supporting predictability, the evidence of out-of-sample predictability is fairly weak. In particular, we only find evidence of superior portfolio performance when no short-selling is imposed. In addition, only certain predictive variables such as the T-bill yield is capable of producing superior performance. A close examination of the portfolio weights under the no-short-sale constraint leads us to consider Henriksssson and Merton's (1981) type of market timing where a switching strategy is used. The switching portfolios produce even stronger performance than the no-short-sale constrained portfolios, which suggests that investors are better off predicting the sign not the magnitude of the market expected excess returns. In other words, the predictive variables may provide useful information to signal either a buy or sell but nothing more. Possible explanations are large estimation errors, model mis-specification, and time-varying predictive relationships.

Our analysis can be extended in a number of interesting directions. First, we do not consider dynamic hedging demands. A dynamic strategy that hedges future changes in the investment opportunity set may perform quite differently. 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 see that if the outperformance is robust to this uncertainty. Third, we examine four economic variables that are believed to be powerful predictors, but there are other potentially powerful predictive variables that may be worth investigating. Fourth, 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. Fifth, since different predictive models perform differently in different periods, it may be beneficial for investors to use a model averaging approach. Unlike Avramov (2002) and Cremers (2002) who average over a set of predictors, a set of data-generating processes such as the ones considered here would be averaged. Finally, our results suggest that the predictive relations seem unstable, and we employ a strategy that re-estimates the model periodically. But an alternative strategy is to explicitly model time-varying parameters or structural breaks. It is well known that regime shift happens 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 and three of eight in CRSP equal-weighted returns.

Pesaran and Timmermann (2002) find that a linear predictive model that incorporates structural breaks has improved out-of-sample statistical forecasting power.

References

Abhyankar, A., and P. R. Davies, 2002, "Return Predictability, Market Timing and Volatility: Evidence from the Short Rate Revisited," working paper, University of Warwick.

Admati, A. R., S. Bhattacharya, P. Pfleiderer, and S. A. Ross, 1986, "On Timing and Selectivity," *Journal of Finance*, 41(3), 715–730.

Aït-Sahalia, Y., and M. W. Brandt, 2001, "Variable Selection for Portfolio Choice," *Journal of Finance*, 56, 1297–1351.

Ang, A., and G. Bekaert, 2006, "Stock Return Predictability: Is it There?," *The Review of Financial Studies*.

Avramov, D., 2002, "Stock Return Predictability and Model Uncertainty," *Journal of Financial Economics*, 64(3), 431–467.

Balduzzi, P., and A. W. Lynch, 1999, "Transaction Costs and Predictability: Some Utility Cost Calculations," *Journal of Financial Economics*, 52, 47–78.

Barberis, N., 2000, "Investing for the Long Run when Returns are Predictable," *Journal of Finance*, 55, 225–264.

Bekaert, G., and R. J. Hodrick, 1992, "Characterizing Predictable Components in Excess Returns on Equity and Foreign Exchange Markets," *Journal of Finance*, 47(2), 467–509.

Bossaerts, P., and P. Hillion, 1999, "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?," *The Review of Financial Studies*, 12(2), 405–428.

Breen, W., L. R. Glosten, and R. Jagannathan, 1989, "Economic Significance of Predictable Variations in Stock Index Returns," *Journal of Finance*, 44(5), 1177–1189.

Campbell, J. Y., 1987, "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18, 373–399.

Campbell, J. Y., Y. L. Chan, and L. M. Viceira, 2002, "A Multivariate Model of Strategic Asset Allocation," *Journal of Financial Economics*, forthcoming.

Campbell, J. Y., and R. J. Shiller, 1988a, "The Dividend Price Ratio and Expectations of Future Dividends and Discount Factors," *The Review of Financial Studies*, 1, 195–228.

——— , 1988b, "Stock Prices, Earning, and Expected Dividends," *Journal of Finance*, 43, 661–676.

Carlson, M., D. A. Chapman, R. Kaniel, and H. Yan, 2004, "Specification Error, Estimation Risk, and Conditional Portfolio Rules," working paper, University Texas at Austin.

Cochrane, J. H., 1991, "Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations," *Journal of Finance*, 46, 209–235.

Cooper, M., and H. Gulen, 2001, "Is Time-series Based Predictability Evident in Real-time?," working paper.

Cooper, M., R. C. Gutierrez, Jr., and W. Marcum, 2001, "On the Predictability of Stock Returns in Real Time," *Journal of Business*, Forthcoming.

Cremers, M., 2002, "Stock Return Predictability: a Bayesian Model Selection Perspective," *The Review of Financial Studies*, 15(4), 1223–1250.

Engle, R. F., 1982, "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of the United Kingdom Inflation," *Econometrica*, 22, 427–432.

Fama, E., and W. Schwert, 1977, "Asset Returns and Inflation," *Journal of Financial Economics*, 5, 115–146.

Fama, E. F., and K. R. French, 1988, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–25.

——— , 1989, "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics*, 25, 23–49.

Ferson, W. E., 1989, "Changes in Expected Security Returns, Risk, and the Level of Interest Rates," *Journal of Finance*, 44, 1191–1217.

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

Ferson, W. E., S. Sarkissian, and T. Simin, 1999, "Spurious Regressions in Financial Economics," working paper, Boston College, McGill University, and University of Washington.

French, K., W. Schwert, and R. Stambaugh, 1987, "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 19, 3–19.

Gallant, R., and G. Tauchen, 1989, "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Process: Asset Pricing Applications," *Econometrica*, 57, 1091–1120.

———, 1997, *SNP: A Program for Nonparametric Time Series Analysis.* Version 8.7.

Goyal, A., and I. Welch, 2003, "Predicting the Equity Premium With Dividend Ratios," *Management Science*, 49(5), 639–654.

Graham, J. R., and C. R. Harvey, 1997, "Grading the Performance of Market-Timing Newsletters," *Financial Analysts Journal*, 53(6), 54 – 68.

Handa, P., and A. Tiwari, 2004, "Does Stock Return Predictability Imply Improved Asset Allocation and Performance?," working paper, University of Iowa.

Harvey, C., 1989, "Time-Varying Conditional Covariances in Tests of Asset Pricing Models," *Journal of Financial Economics*, 24, 289–317.

Harvey, C. R., 1991, "The Specification of Conditional Expectations," working paper, Duke University.

Henrikssson, R. D., and R. C. Merton, 1981, "On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills," *Journal of Business*, 54(4), 513–533.

Hodrick, R. J., 1992, "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *The Review of Financial Studies*, 5, 357–386.

Jacobsen, B., 1999, "The Economic Significance of Simple Time Series Models of Stock Return Predictability," working paper, University of Amsterdam.

Jegadeesh, N., 1990, "Evidence of Predictable Behavior of Security Returns," *Journal of Finance*, 45(3), 881–898.

Kandel, S., and R. F. Stambaugh, 1996, "On the Predictability of Stock Returns: An Asset-Allocation Perspective," *Journal of Finance*, 51(2), 385–424.

Keim, D. B., and R. F. Stambaugh, 1986, "Predicting Returns in the Stock and Bond Markets," *Journal of Financial Economics*, 17, 357–390.

Lamont, O., 1998, "Earnings and Expected Returns," *Journal of Finance*, 53, 1563–1588.

Lettau, M., and S. Ludvigson, 2001, "Consumption, Aggregate Wealth and Expected Stock Returns," *Journal of Finance*, 56(3), 815–849.

Lewellen, J., 1999, "The Time-Series Relations Among Expected Return, Risk, and Book-to-Market," *Journal of Financial Economics*, 54, 5–43.

Marquering, W., and M. Verbeek, 2001, "The Economic Value of Predicting Stock Index Returns and Volatility," working paper, Tilburg University.

Merton, R. C., 1980, "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.

Nelson, D., 1991, "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59(2), 347–370.

Pástor, Ľ., 2000, "Portfolio Selection and Asset Pricing Models," *Journal of Finance*, 55, 179–223.

Pástor, Ľ., and R. Stambaugh, 2000, "Comparing Asset Pricing Models: An Investment Perspective," *Journal of Financial Economics*, 56, 335–381.

Pesaran, M. H., and A. Timmermann, 1995, "Predictability of Stock Returns: Robustness and Economic Significance," *Journal of Finance*, 50(4), 1201–1228.

Pesaran, M. H., and A. Timmermann, 2002, "Market Timing and Return Predictability under Model Instability," *Journal of Empirical Finance*, 8, 495–510.

Rapach, D. E., and M. E. Wohar, 2004, "Structural Change and the Predictability of Stock Returns," working paper, St. Louis University and University of Nebraska at Omaha.

Roskelley, K., 2004, "Asset Return Predictability: Economic Inference When the Investor Engages in a Specification Search," working paper, Louisiana Tech University.

Santos, T., and P. Veronesi, 2001, "Labor Income and Predictable Stock Returns," working paper, University of Chicago.

Schwert, G. W., 1989, "Why Does Stock Market Volatility Change Over Time?," *Journal of Finance*, 44(6), 1115–1153.

Shanken, J. A., and A. Tamayo, 2001, "Risk, Mispricing, and Asset Allocation: Conditioning on Dividend Yield," working paper, University of Rochester.

Stambaugh, R. F., 1999, "Predictive Regressions," *Journal of Financial Economics*, 54, 375–421.

Tang, D. Y., 2003, "Asset Return Predictability and Bayesian Model Averaging," working paper, University of Texas, Austin.

Torous, W., and R. Valkanov, 2000, "Boundaries of Predictability: Noisy Predictive Regressions," working paper, UCLA.

Treynor, J. L., and K. Mazuy, 1966, "Can Mutual Funds Outguess the Market?," *Harvard Business Review*, 44, 131–136.

Tu, J., and G. Zhou, 2003, "Data-Generating Process Uncertainty: What Difference does it Make in Portfolio Decisions?," *Journal of Financial Economics*, forthcoming.

Table 1: Descriptive Statistics of Data

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 1-month T-bill (riskfree return), dividend yield (DVYD), and 3-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 predictive variables and excess return for the whole sample period and the two subperiods.

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 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				

Table 2: Optimal Predictive Models

This table reports for each combination of predictive variables the best predictive model and the various model selection criteria used in each category. Panel A reports the in-sample predictive models, while Panel B reports the out-of-sample predictive models. 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 Shwarz Bayesian Information Criterion, the Akaike Information criterion, and the Hannan-Quinn criterion, respectively. We use the following notation to denote the various models, "$[m_1 m_2 m_3 m_4 m_5]$", where m_1 denotes the order of vector autoregression, m_2 and m_3 denote the order of GARCH (GARCH(m_2, m_3)), m_4 denotes the order of the Hermite polynomial expansion, and m_5 is the order of polynomial in x for the coefficients in the Hermite polynomial.

Optimal Predictive Models and Model Selection Criteria																				
	Model	Obj	BIC	AIC	HQ	Model	Obj	BIC	AIC	HQ	Model	Obj	BIC	AIC	HQ	Model	Obj	BIC	AIC	HQ
Panel A: Full Sample Period																				
	DVYD					TBYD					DFSD					TBYD & DFSD				
VAR	[10000]	0.40	0.45	0.41	0.43	[20000]	0.85	0.92	0.87	0.89	[20000]	1.29	1.36	1.31	1.33	[20000]	0.61	0.75	0.65	0.69
GARCH	[22100]	0.19	0.28	0.22	0.24	[22100]	0.20	0.29	0.23	0.25	[21100]	0.85	0.93	0.88	0.90	[21100]	-0.30	-0.14	-0.25	-0.21
OPT	[22140]	0.06	**0.20**	0.10	**0.14**	[22140]	0.13	**0.27**	**0.17**	**0.21**	[21140]	0.76	**0.89**	**0.80**	**0.83**	[21140]	-0.42	**-0.20**	-0.35	**-0.29**
NLNP	[22141]	-0.02	0.25	**0.06**	0.14	[22141]	0.11	0.35	0.18	0.25	[21141]	0.73	0.95	0.80	0.86	[21141]	-0.50	-0.07	**-0.36**	-0.25
	TRSD					TBYD & TRSD					TRSD & DFSD					TBYD & TRSD & DFSD				
VAR	[20000]	1.50	1.58	1.53	1.55	[20000]	0.42	0.57	0.47	0.51	[20000]	1.34	1.50	1.39	1.43	[20000]	0.18	0.45	0.27	0.34
GARCH	[21100]	1.21	1.30	1.24	1.26	[21100]	0.03	0.21	0.09	0.14	[21100]	0.68	0.87	0.74	0.79	[21100]	-0.50	-0.21	-0.41	-0.33
OPT	[21140]	1.15	**1.29**	1.19	**1.23**	[21140]	-0.11	**0.16**	-0.02	**0.05**	[21140]	0.57	**0.82**	0.65	**0.72**	[21140]	-0.69	**-0.26**	**-0.55**	**-0.44**
NLNP	[21141]	1.09	1.33	**1.17**	1.23	[21141]	-0.23	0.32	**-0.05**	0.09	[21141]	0.47	0.95	**0.62**	0.75	[21141]	-0.85	0.12	-0.54	-0.28
Panel B: Estimation Period																				
	DVYD					TBYD					DFSD					TBYD & DFSD				
VAR	[10000]	0.47	0.54	0.49	0.51	[20000]	0.76	0.86	0.79	0.82	[20000]	1.25	1.35	1.28	1.31	[20000]	0.53	0.75	0.61	0.66
GARCH	[11100]	0.27	0.36	0.30	0.32	[22100]	0.31	0.45	0.36	0.39	[21100]	0.96	1.09	1.01	1.04	[22100]	-0.02	0.24	0.07	0.13
OPT	[11140]	0.16	**0.32**	0.22	**0.26**	[22140]	0.24	**0.45**	**0.31**	**0.37**	[21140]	0.90	**1.09**	**0.96**	**1.01**	[22140]	-0.22	**0.15**	-0.10	**0.00**
NLNP	[11141]	0.09	0.39	**0.20**	0.27	[22141]	0.19	0.56	0.32	0.41	[21141]	0.89	1.10	0.96	1.02	[22141]	-0.37	0.38	**-0.12**	0.08
	TRSD					TBYD & TRSD					TRSD & DFSD					TBYD & TRSD & DFSD				
VAR	[20000]	1.54	1.66	1.58	1.61	[20000]	0.52	0.78	0.61	0.68	[10000]	1.40	1.57	1.46	1.50	[10000]	0.39	0.68	0.49	0.57
GARCH	[21100]	1.39	1.54	1.44	1.48	[21100]	0.36	0.66	0.47	0.54	[11100]	0.96	1.17	1.03	1.09	[11100]	-0.04	0.29	0.08	0.16
OPT	[21140]	1.32	**1.55**	**1.40**	**1.46**	[21140]	0.24	**0.66**	0.39	**0.49**	[11140]	0.84	**1.17**	0.96	**1.04**	[11140]	-0.22	**0.27**	-0.05	**0.08**
NLNP	[21141]	1.27	1.67	1.41	1.52	[21141]	0.10	0.88	**0.37**	0.58	[11141]	0.71	1.40	**0.95**	1.13	[11141]	-0.46	0.67	**-0.07**	0.23

Table 3: In-Sample Portfolio Performance of the Predictive Models

Predictive sample draws $\tilde{\mathbf{y}}_{t+1|t}$ are generated at each month t from the one-step-ahead conditional distributions of the predictive models, conditioned on the observed data $\mathbf{y}_t$. The predictive sample draws of the excess returns are then used to solve the portfolio optimization problem at each month t. The realized portfolio returns r_{pt} are calculated from the observed excess returns r_t and the riskfree rates. The average returns, standard deviations, 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. The number of sample draws at each period t is 50000.

	In-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}	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$	r_{ce}	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$	r_{ce}
	Benchmarks																			
Market	8.44	14.17	0.26		5.35															
Fixed	7.03	7.69	0.29	0.48	6.14															
Passive	7.50	10.59	0.25	-0.03	5.78															
AR	7.38	7.92	0.32	0.79	6.43															
GARCH	7.36	9.75	0.26	-0.08	5.91															
	DVYD					TBYD					DFSD					TBYD & DFSD				
	No Constraint					No Constraint					No Constraint					No Constraint				
VAR	7.78	8.46	0.34	1.18	6.71	15.54	18.65	0.58	4.40	10.29	8.73	11.05	0.35	1.26	6.92	20.67	23.09	0.69	5.98	13.04
GARCH	8.33	15.52	0.22	-0.55	4.72	14.86	17.71	0.57	4.29	10.10	8.71	12.95	0.30	0.50	6.18	19.11	20.09	0.71	6.34	13.28
OPT	7.42	15.04	0.17	-1.30	4.00	13.90	16.98	0.53	3.83	9.47	8.28	14.22	0.24	-0.30	5.16	18.39	19.37	0.70	6.18	13.11
NLNP	7.52	10.61	0.25	-0.16	5.86	13.27	15.50	0.54	3.99	9.60	7.88	11.11	0.27	0.15	6.04	19.28	20.26	0.71	6.38	13.34
	No-Short-Sale					No-Short-Sale					No-Short-Sale					No-Short-Sale				
VAR	7.65	8.21	0.34	1.10	6.64	10.36	10.03	0.55	4.07	8.85	8.09	9.03	0.36	1.37	6.88	11.26	9.82	0.65	5.55	9.83
GARCH	7.52	10.11	0.26	0.02	5.99	9.91	9.42	0.53	3.90	8.57	8.33	10.92	0.32	0.79	6.52	11.31	9.95	0.65	5.49	9.83
OPT	7.42	10.21	0.25	-0.17	5.85	9.92	9.95	0.51	3.51	8.42	7.53	10.45	0.26	-0.09	5.84	11.00	10.02	0.61	4.99	9.50
NLNP	7.13	8.39	0.27	0.12	6.08	9.77	9.75	0.50	3.43	8.33	7.56	9.72	0.28	0.22	6.12	10.73	9.46	0.62	5.09	9.39
	TRSD					TBYD & TRSD					TRSD & DFSD					TBYD & TRSD & DFSD				
	No Constraint					No Constraint					No Constraint					No Constraint				
VAR	13.13	14.13	0.55	4.62	10.07	19.49	21.24	0.67	6.26	13.06	14.86	17.33	0.55	4.58	10.25	23.67	25.35	0.73	7.07	14.82
GARCH	11.24	11.52	0.51	4.06	9.21	17.98	20.36	0.62	5.62	11.90	13.26	15.49	0.51	4.04	9.56	21.32	22.57	0.71	6.86	14.15
OPT	10.03	11.19	0.42	2.72	8.05	16.30	18.49	0.59	5.21	11.18	10.83	15.36	0.36	1.83	7.05	19.55	19.99	0.71	6.90	13.98
NLNP	13.19	15.05	0.52	4.20	9.41	19.71	18.36	0.78	7.96	14.84	16.56	18.31	0.61	5.51	11.04	22.57	20.95	0.82	8.52	16.26
	No-Short-Sale					No-Short-Sale					No-Short-Sale					No-Short-Sale				
VAR	9.89	9.11	0.50	3.88	8.59	12.39	9.78	0.72	7.07	10.98	9.76	9.60	0.46	3.31	8.33	12.05	9.56	0.70	6.79	10.70
GARCH	9.70	9.01	0.48	3.66	8.43	11.18	10.10	0.58	5.01	9.66	10.16	10.27	0.47	3.44	8.53	11.56	9.88	0.63	5.75	10.12
OPT	8.76	8.96	0.38	2.17	7.49	10.51	10.41	0.49	3.83	8.88	8.72	9.55	0.35	1.78	7.30	11.39	9.56	0.63	5.80	10.04
NLNP	10.61	9.96	0.53	4.30	9.07	11.60	9.74	0.64	5.93	10.18	11.26	10.37	0.57	4.90	9.60	11.89	9.60	0.68	6.50	10.51

Table 4: Out-Of-Sample Expanding-Window Performance of the Predictive Models

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 re-estimated every 5 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 predictive variable and model combination. Bold face indicates measures higher than those of the benchmarks. The number of sample draws at each period t is 50000.

Out-Of-Sample 5-year Expanding-Window Portfolio Performance Tests																				
	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															
	DVYD					TBYD					DFSD					TBYD & DFSD				
	No Constraint					No Constraint					No Constraint					No Constraint				
VAR	5.25	12.33	-0.14	-7.87	3.04	13.78	46.63	0.15	-3.62	-75.30	7.65	15.40	0.05	-5.11	4.02	18.15	41.86	0.27	-1.80	-15.53
GARCH	7.54	12.66	0.05	-5.09	5.14	9.88	38.67	0.08	-4.67	-35.30	10.29	14.41	0.23	-2.29	7.10	14.33	29.76	0.25	-2.08	-0.58
OPT	4.54	16.22	-0.15	-8.04	0.63	-21.45	133.5	-0.21	-9.04	-3804	10.27	15.57	0.21	-2.57	6.39	9.76	36.62	0.08	-4.66	-13.61
NLNP	0.62	16.57	-0.38	-11.56	-3.45	10.60	23.69	0.16	-3.48	1.41	10.78	11.44	0.34	-0.74	8.86	9.05	39.59	0.05	-5.02	-38.06
	No-Short-Sale					No-Short-Sale					No-Short-Sale					No-Short-Sale				
VAR	7.78	7.44	0.11	-4.09	6.95	10.38	6.31	**0.55**	**2.45**	**9.78**	7.85	10.31	0.09	-4.47	6.21	11.89	7.60	**0.65**	**4.07**	**11.05**
GARCH	8.69	8.95	0.20	-2.84	7.50	10.23	5.94	**0.56**	**2.59**	**9.70**	11.00	11.05	0.37	-0.26	9.15	11.52	7.35	**0.62**	**3.63**	**10.72**
OPT	7.64	8.05	0.09	-4.48	6.68	10.21	6.24	**0.53**	**2.16**	**9.62**	9.54	11.39	0.23	-2.36	7.46	10.16	6.35	**0.51**	**1.89**	**9.58**
NLNP	8.09	6.39	0.18	-3.06	7.49	9.80	5.65	**0.51**	**1.89**	9.32	9.91	8.71	0.34	-0.65	8.77	11.84	6.73	**0.73**	**5.21**	**11.17**
	TRSD					TBYD & TRSD					TRSD & DFSD					TBYD & TRSD & DFSD				
	No Constraint					No Constraint					No Constraint					No Constraint				
VAR	16.90	37.62	0.27	-1.84	-9.57	18.26	58.81	0.19	-2.93	-187.73	13.66	34.33	0.20	-2.87	-11.63	15.35	35.95	0.24	-2.31	-3.34
GARCH	15.33	33.69	0.25	-2.07	-10.37	17.93	37.15	0.30	-1.37	-7.07	13.39	37.87	0.17	-3.25	-22.99	11.93	35.19	0.14	-3.69	-9.57
OPT	11.39	40.83	0.11	-4.18	-39.89	11.65	55.24	0.09	-4.54	-153.5	10.75	36.00	0.11	-4.22	-34.87	10.75	40.86	0.09	-4.42	-18.48
NLNP	12.39	45.71	0.12	-4.03	-82.39	20.51	32.84	0.42	0.40	3.18	8.59	35.60	0.05	-5.12	-39.26	9.12	36.03	0.06	-4.91	-11.65
	No-Short-Sale					No-Short-Sale					No-Short-Sale					No-Short-Sale				
VAR	11.46	11.49	0.39	0.14	9.36	12.98	9.85	**0.61**	**3.46**	**11.51**	10.73	11.02	0.35	-0.60	8.78	10.89	6.90	**0.57**	**2.87**	**10.19**
GARCH	11.73	11.67	0.41	0.40	**9.56**	12.40	9.84	**0.55**	**2.58**	**10.92**	10.84	10.93	0.36	-0.40	8.92	10.38	8.08	0.43	0.64	9.39
OPT	10.89	10.89	0.36	-0.32	8.97	11.07	10.90	0.38	-0.08	9.15	10.27	10.98	0.30	-1.22	8.33	10.24	9.53	0.35	-0.55	8.75
NLNP	11.52	11.60	0.40	0.17	9.37	13.58	8.75	**0.76**	**5.66**	**12.44**	11.37	11.89	0.37	-0.18	9.13	9.02	8.49	0.25	-2.07	7.80

Table 5: Out-Of-Sample Expanding-Window Performance with Transaction Costs

The predictive sample draws are generated from the 5-year expanding-window 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.

Out-Of-Sample 5-year Expanding-Window 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}(%)	r_p(%)	σ_p(%)	$\mathcal{SR}$	$\mathcal{GH}2$	r_{ce}(%)	r_p(%)	σ_p(%)	$\mathcal{SR}$	$\mathcal{GH}2$	r_{ce}(%)
	Benchmarks																			
Market	5.82	15.23	0.38		9.18															
Passive	10.18	7.47	0.43	0.74	9.32															
	TBYD & DFSD					$\tau = 0.25\%$					$\tau = 0.50\%$					$\tau = 1.00\%$				
	No Constraint					No Constraint					No Constraint					No Constraint				
VAR	18.15	41.86	0.27	-1.80	-15.53	17.47	40.35	0.26	-1.90	-12.37	17.27	38.86	0.27	-1.83	-8.95	17.19	37.33	0.28	-1.69	-4.38
GARCH	14.33	29.76	0.25	-2.08	-0.58	13.15	28.57	0.22	-2.55	-0.26	11.64	27.26	0.17	-3.22	-0.22	11.72	23.98	0.20	-2.81	2.99
OPT	9.76	36.62	0.08	-4.66	-13.61	9.40	34.35	0.07	-4.74	-10.60	8.84	32.90	0.06	-4.95	-8.81	5.11	31.58	-0.06	-6.68	-10.10
NLNP	9.05	39.59	0.05	-5.02	-38.06	8.67	35.56	0.05	-5.09	-20.81	7.69	32.00	0.02	-5.46	-12.55	8.29	25.76	0.05	-5.03	-2.20
	No-Short-Sale					No-Short-Sale					No-Short-Sale					No-Short-Sale				
VAR	11.89	7.60	0.65	4.07	11.05	11.85	7.36	0.67	4.31	11.06	11.21	6.62	0.65	3.99	10.57	9.01	7.04	0.30	-1.31	8.25
GARCH	11.52	7.35	0.62	3.63	10.72	11.32	7.17	0.61	3.46	10.56	10.03	6.99	0.44	0.90	9.29	7.69	7.15	0.11	-4.19	6.84
OPT	10.16	6.35	0.51	1.89	9.58	9.88	5.98	0.49	1.66	9.35	9.29	5.80	0.41	0.34	8.79	7.56	6.21	0.10	-4.27	6.93
NLNP	11.84	6.73	0.73	5.21	11.17	11.47	6.26	0.72	5.16	10.89	10.52	5.87	0.61	3.47	10.02	9.18	5.26	0.43	0.70	8.78
	TBYD & TRSD					$\tau = 0.25\%$					$\tau = 0.50\%$					$\tau = 1.00\%$				
	No Constraint					No Constraint					No Constraint					No Constraint				
VAR	18.26	58.81	0.19	-2.93	-187.7	19.63	57.86	0.22	-2.52	-156.9	19.37	55.58	0.22	-2.46	-128.9	18.62	52.02	0.23	-2.44	-80.34
GARCH	17.93	37.15	0.30	-1.37	-7.07	17.63	36.60	0.29	-1.42	-6.20	17.36	35.80	0.29	-1.44	-5.08	18.45	35.62	0.32	-0.96	-3.43
OPT	11.65	55.24	0.09	-4.54	-153.5	4.98	81.43	-0.02	-6.18	-1501	2.61	84.21	-0.05	-6.60	-1766	14.88	51.52	0.15	-3.50	-80.12
NLNP	20.51	32.84	0.42	0.40	3.18	19.79	31.67	0.41	0.29	3.24	20.66	31.95	0.43	0.64	3.75	17.75	30.64	0.35	-0.50	2.03
	No-Short-Sale					No-Short-Sale					No-Short-Sale					No-Short-Sale				
VAR	12.98	9.85	0.61	3.46	11.51	12.31	10.71	0.50	1.78	10.52	12.39	10.88	0.50	1.77	10.56	13.14	11.49	0.54	2.35	11.01
GARCH	12.40	9.84	0.55	2.58	10.92	11.91	11.05	0.45	0.99	9.94	12.18	10.76	0.49	1.56	10.36	12.91	11.07	0.54	2.34	10.91
OPT	11.07	10.90	0.38	-0.08	9.15	11.73	10.97	0.44	0.79	9.78	11.20	11.10	0.38	0.00	9.21	12.59	11.76	0.48	1.46	10.39
NLNP	13.58	8.75	0.76	5.66	12.44	13.68	9.03	0.74	5.47	12.47	13.52	9.16	0.72	5.05	12.27	12.74	10.00	0.58	2.96	11.20

Table 6: Out-Of-Sample Mean-Variance Performance

The predictive sample draws are generated from the 5-year expanding estimation. The optimal portfolio weights are calculated from maximizing the expected quadratic utility with a risk aversion coefficient of 4. 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 predictive variable and model combination. The measures are in bold face when they are higher than those of the benchmarks. The number of sample draws at each period t is 50000.

	Out-Of-Sample Mean-Variance Portfolio Performance Tests															
	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$
	Benchmarks															
Market	12.74	15.10	0.38													
Passive	8.93	5.27	0.38	-0.08												
Expand	8.60	4.36	0.38	-0.02												
AR	7.10	5.58	0.16	-1.04												
GARCH	7.37	7.60	0.15	-1.18												
	DVYD				TBYD				DFSD				TBYD & DFSD			
	No Constraint				No Constraint				No Constraint				No Constraint			
VAR	5.29	12.90	-0.13	-7.73	27.10	52.52	0.39	-0.09					28.91	49.22	**0.45**	**0.85**
GARCH	7.10	11.82	0.02	-5.59	18.50	39.08	0.30	-1.39	9.67	12.87	0.21	-2.61	19.31	32.60	0.38	-0.13
OPT	5.16	17.30	-0.10	-7.36	21.17	43.08	0.33	-0.88	9.92	14.49	0.21	-2.70	18.03	38.40	0.29	-1.49
NLNP	1.59	17.98	-0.30	-10.30	13.06	24.04	0.26	-1.99	10.18	10.56	0.31	-1.15	18.57	38.42	0.30	-1.28
	No-Short-Sale				No-Short-Sale				No-Short-Sale				No-Short-Sale			
VAR	7.48	7.07	0.08	-4.63	9.77	5.78	**0.49**	**1.61**					11.55	7.45	**0.62**	**3.57**
GARCH	8.35	8.52	0.17	-3.30	9.50	5.23	**0.49**	**1.67**	10.26	9.98	0.33	-0.78	11.02	7.02	**0.58**	**3.00**
OPT	7.49	7.82	0.07	-4.72	9.62	5.59	**0.48**	**1.50**	9.09	10.76	0.20	-2.77	9.99	6.04	**0.51**	**1.88**
NLNP	7.96	5.81	0.18	-3.11	9.18	4.88	**0.47**	**1.22**	9.36	7.68	0.31	-1.04	11.48	6.49	**0.70**	**4.78**
	TRSD				TBYD & TRSD				TRSD & DFSD				TBYD & TRSD & DFSD			
	No Constraint				No Constraint				No Constraint				No Constraint			
VAR	21.98	37.79	**0.40**	**0.16**	36.10	63.48	**0.46**	**1.04**	18.13	33.92	0.33	-0.86	22.57	42.84	0.37	-0.36
GARCH	19.36	32.89	0.38	-0.14	25.90	42.20	**0.45**	**0.91**	17.51	35.27	0.30	-1.32	19.58	39.55	0.32	-1.03
OPT	17.72	39.02	0.28	-1.67	28.99	59.96	0.37	-0.33	15.25	34.34	0.24	-2.18	21.71	46.42	0.32	-1.06
NLNP	17.91	45.73	0.24	-2.22	22.59	30.06	**0.52**	**2.01**	10.19	32.18	0.10	-4.29	16.32	37.62	0.25	-2.09
	No-Short-Sale				No-Short-Sale				No-Short-Sale				No-Short-Sale			
VAR	11.12	11.29	0.37	-0.20	12.75	9.37	**0.62**	**3.56**	10.45	10.90	0.32	-0.93	10.61	6.54	**0.56**	**2.72**
GARCH	11.24	11.47	0.38	-0.13	11.91	9.22	**0.54**	**2.35**	10.16	10.74	0.30	-1.27	10.02	7.48	**0.41**	**0.44**
OPT	10.21	10.55	0.31	-1.10	10.46	10.69	0.33	-0.83	9.82	10.67	0.27	-1.71	9.79	9.38	0.31	-1.19
NLNP	11.17	11.40	0.37	-0.19	13.11	8.54	**0.72**	**5.12**	11.00	11.71	0.35	-0.56	8.42	8.08	0.19	-3.01

Table 7: Out-of-Sample Portfolio Performance under Different Constraints

The predictive sample draws are generated from the recursive estimation with 5-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. Bold face indicates measures higher than those of the benchmarks.

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															
VAR	9.63	6.27	0.43	0.68	9.04															
GARCH	9.90	9.33	0.32	-1.01	8.58															
TBYD																				
	$\lvert w\rvert \le 2$					$\lvert w\rvert \le 1$					$\lvert w\rvert \le 0.5$					$0 \le w \le 2$				
VAR	11.19	17.24	0.25	-2.08	6.73	10.33	11.19	0.31	-1.21	8.47	9.37	6.54	0.38	-0.11	8.74	10.91	8.46	**0.47**	**1.29**	**9.85**
GARCH	10.25	17.88	0.19	-3.01	5.24	10.61	11.03	0.34	-0.76	8.79	9.48	6.72	0.39	-0.01	8.81	10.38	6.68	**0.52**	**2.00**	**9.71**
OPT	9.62	18.28	0.15	-3.59	4.39	10.18	11.22	0.29	-1.42	8.30	9.06	6.84	0.32	-1.05	8.36	10.71	6.93	**0.55**	**2.45**	**9.99**
NLNP	9.03	17.38	0.12	-3.99	4.35	9.42	11.09	0.23	-2.41	7.58	8.64	6.81	0.26	-1.97	7.95	9.64	6.28	**0.43**	**0.72**	**9.04**
TBYD & DFSD																				
	$\lvert w\rvert \le 2$					$\lvert w\rvert \le 1$					$\lvert w\rvert \le 0.5$					$0 \le w \le 2$				
VAR	13.96	20.45	0.35	-0.63	7.72	11.57	12.66	0.37	-0.27	9.20	9.59	6.83	0.40	0.16	8.90	13.93	12.56	**0.56**	**2.61**	**11.61**
GARCH	12.75	18.91	0.31	-1.17	7.31	11.32	11.72	0.38	-0.15	9.26	9.82	6.66	0.44	0.83	9.16	12.38	10.72	**0.51**	**1.87**	**10.65**
OPT	10.01	19.87	0.16	-3.48	4.13	9.12	12.14	0.18	-3.08	6.95	7.91	6.63	0.15	-3.53	7.26	10.97	9.48	**0.43**	**0.62**	**9.64**
NLNP	11.96	22.59	0.22	-2.47	4.37	11.12	13.50	0.31	-1.12	8.45	9.79	7.36	0.40	0.12	8.99	14.25	10.72	**0.68**	**4.49**	**12.57**
TBYD & TRSD																				
	$\lvert w\rvert \le 2$					$\lvert w\rvert \le 1$					$\lvert w\rvert \le 0.5$					$0 \le w \le 2$				
VAR	16.47	20.12	0.48	**1.34**	**10.41**	13.34	12.04	0.53	**2.24**	**11.15**	10.43	6.88	0.51	**1.93**	**9.71**	15.54	15.53	0.55	**2.56**	**11.90**
GARCH	15.07	19.69	0.41	0.42	9.17	12.49	12.31	0.45	**1.02**	**10.19**	10.11	7.13	0.45	0.98	9.34	14.69	14.13	0.55	**2.48**	**11.69**
OPT	11.94	21.71	0.23	-2.34	4.40	10.95	13.34	0.30	-1.26	8.13	9.80	6.94	0.42	0.49	9.07	11.90	16.86	0.29	-1.36	7.11
NLNP	16.56	21.13	0.46	**1.05**	**9.73**	13.63	12.07	0.56	**2.59**	**11.44**	11.11	6.60	0.64	**3.85**	**10.46**	15.86	15.49	0.58	**2.89**	**12.23**

Table 8: Out-Of-Sample Switching Portfolio Performance

The predictive sample draws are generated from the 5-year rolling estimation. At each month, the portfolio either invests in the market portfolio or 30-day T-bill depending on wether 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 predictive variable and model combination. $\mathcal{SR}$ and $\mathcal{GH}2$ are in bold face when they are higher than those of the benchmarks. 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 in the recursive estimation (Table 4), and $\mathcal{Z}$, which measures the ratio between the frequency of the switching portfolios beating the corresponding constrained portfolios and that of the constrained portfolio beating the switching portfolios. Bold face indicates measures higher than those of the benchmarks. The number of sample draws at each period t is 50000.

Out-Of-Sample Switching Portfolio Performance Tests												
	r_p(%)	σ_p(%)	$\mathcal{SR}$	$\mathcal{GH}2$	$\Delta_{\mathcal{GH}2}$	$\mathcal{Z}$	r_p(%)	σ_p(%)	$\mathcal{SR}$	$\mathcal{GH}2$	$\Delta_{\mathcal{GH}2}$	$\mathcal{Z}$
Benchmarks												
Market	12.74	15.10	0.38									
Expand	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						TRSD					
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
	TBYD						TBYD & 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
	DFSD						TRSD & DFSD					
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
	TBYD & 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

Table 9: Market Timing Performance

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 $Cov(w, r_{mt})$, measures the correlation between the portfolio weights and the market excess returns. $\star$ denotes positive significance, whereas $*$ denotes negative significance. Three, two, and one star denote 1%, 5%, and 10% significance, respectively.

	β_2	$\mathcal{TM}$	Corr	β_2	$\mathcal{TM}$	Corr	β_2	$\mathcal{TM}$	Corr
Expand	0.13⋆⋆⋆	0.30	0.07						
AR	0.39⋆⋆⋆	0.90	−0.00						
GARCH	0.43⋆⋆⋆	0.98	−0.08						
	TBYD			DFSD			TRSD		
No Constraint:									
VAR	−1.65	−3.77	0.18	1.55⋆⋆⋆	3.55	−0.04	−3.96∗∗	−9.03	0.09
GARCH	−3.30∗∗	−7.54	0.12	9.88	22.55	0.03	−4.89∗∗∗	−11.16	0.07
OPT	−17.57∗∗∗	−40.07	0.12	6.81	15.53	0.05	−9.37∗∗∗	−21.38	0.08
NLNP	−1.50∗	−3.42	0.14	−16.89∗∗	−38.54	0.04	−11.23∗∗∗	−25.62	0.05
No-Short-Sale:									
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		
No Constraint:									
VAR	1.98	4.51	0.16	−1.75	−3.99	0.16	−4.92∗∗∗	−11.22	0.07
GARCH	1.40	3.19	0.16	−1.43	−3.26	0.14	−7.54∗∗∗	−17.2	0.07
OPT	0.41	0.93	0.14	−4.83∗∗	−11.02	0.12	−8.65∗∗∗	−19.73	0.05
NLNP	1.39	3.18	0.12	1.04	2.38	0.13	−7.01∗∗∗	−15.98	−0.02
No-Short-Sale:									
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

Table 10: Out-Of-Sample 5-year Subperiod Switching Portfolio Performance

The predictive sample draws are generated from the 5-year expanding estimation and switching portfolios are formed as described in the text. The testing period is divided into four 5-year periods coincident with the re-estimation. Panel A reports the period-by-period performance for each period, whereas Panel B reports the accumulative performance for each period. 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 ($\Delta_{\mathcal{GH}2}$), and bold face indicates positive numbers.

Out-Of-Sample 5-Year Subperiod Switching Portfolio Performance											
	Panel A: Period-by-Period						**Panel B: Accumulative**				
	r_p(%)	σ_p(%)	$\mathcal{SR}$	$\mathcal{GH}2$	$\Delta_{\mathcal{GH}2}$		r_p(%)	σ_p(%)	$\mathcal{SR}$	$\mathcal{GH}2$	$\Delta_{\mathcal{GH}2}$
Expanding Window Benchmark											
1979 - 1983	9.74	6.92	0.05	0.61		1979 - 1983	9.74	6.92	0.05	0.61	
1984 - 1988	11.45	6.98	0.26	1.18		1979 - 1988	10.59	6.93	0.15	0.72	
1989 - 1993	8.96	4.77	0.43	0.53		1979 - 1993	10.05	6.28	0.22	0.55	
1994 - 1998	10.13	6.52	1.09	0.80		1979 - 1998	10.07	6.33	0.44	0.94	
1999 - 2003	1.32	10.45	-0.27	0.55		1979 - 2003	8.32	7.39	0.23	-0.02	
TBYD & DFSD											
	VAR						VAR				
1979 - 1983	16.88	7.01	0.86	12.77	**12.14**	1979 - 1983	16.88	7.01	0.86	12.77	**12.14**
1984 - 1988	16.10	12.45	0.74	10.17	**9.01**	1979 - 1988	16.49	10.06	0.76	10.92	**10.14**
1989 - 1993	11.13	8.00	0.72	4.19	**3.68**	1979 - 1993	14.70	9.43	0.75	8.67	**8.05**
1994 - 1998	9.31	6.11	0.74	-4.22	-5.06	1979 - 1998	13.35	8.73	0.74	5.31	**4.55**
1999 - 2003	-2.17	15.63	-0.35	-0.71	-1.26	1979 - 2003	10.25	10.60	0.38	2.39	**2.51**
	GARCH						GARCH				
1979 - 1983	13.65	3.03	0.92	13.76	**13.14**	1979 - 1983	13.65	3.03	0.92	13.76	**13.14**
1984 - 1988	17.07	13.30	0.77	10.62	**9.46**	1979 - 1988	15.36	9.62	0.68	9.54	**8.77**
1989 - 1993	9.40	8.59	0.47	0.98	**0.47**	1979 - 1993	13.37	9.30	0.61	6.62	**6.00**
1994 - 1998	9.60	6.13	0.79	-3.58	-4.42	1979 - 1998	12.43	8.62	0.64	3.83	**3.07**
1999 - 2003	0.57	16.36	-0.17	0.49	-0.06	1979 - 2003	10.06	10.68	0.36	2.06	**2.18**
	OPT						OPT				
1979 - 1983	13.08	2.83	0.80	11.93	**11.30**	1979 - 1983	13.08	2.83	0.80	11.93	**11.30**
1984 - 1988	12.12	11.29	0.47	5.07	**3.91**	1979 - 1988	12.60	8.20	0.47	5.94	**5.17**
1989 - 1993	9.62	6.62	0.64	3.17	**2.66**	1979 - 1993	11.61	7.70	0.51	5.06	**4.45**
1994 - 1998	9.36	6.14	0.75	-4.15	-5.00	1979 - 1998	11.05	7.34	0.56	2.67	**1.91**
1999 - 2003	-0.48	16.37	-0.23	0.40	-0.15	1979 - 2003	8.74	9.88	0.26	0.44	**0.55**
	NLNP						NLNP				
1979 - 1983	17.80	6.68	1.03	15.36	**14.73**	1979 - 1983	17.80	6.68	1.03	15.36	**14.73**
1984 - 1988	13.62	10.31	0.65	8.54	**7.38**	1979 - 1988	15.71	8.67	0.79	11.43	**10.66**
1989 - 1993	10.98	7.49	0.75	4.56	**4.05**	1979 - 1993	14.13	8.30	0.78	9.17	**8.55**
1994 - 1998	9.05	6.45	0.66	-5.31	-6.15	1979 - 1998	12.86	7.89	0.75	5.55	**4.79**
1999 - 2003	-0.04	15.10	-0.22	0.44	-0.11	1979 - 2003	10.28	9.85	0.42	2.90	**3.01**

TBYD

	VAR						VAR				
1979 - 1983	11.53	3.20	0.25	3.67	**3.07**	1979 - 1983	11.53	3.20	0.25	3.67	**3.07**
1984 - 1988	11.96	8.95	0.57	6.99	**5.81**	1979 - 1988	11.74	6.69	0.44	5.55	**4.82**
1989 - 1993	8.36	8.54	0.35	-0.55	-1.07	1979 - 1993	10.61	7.35	0.40	3.36	**2.81**
1994 - 1998	18.78	12.83	1.09	0.69	-0.11	1979 - 1998	12.66	9.06	0.63	3.72	**2.79**
1999 - 2003	-1.09	16.43	-0.27	0.71	0.16	1979 - 2003	9.91	11.02	0.34	1.67	**1.69**
	GARCH						GARCH				
1979 - 1983	11.11	1.11	0.45	6.56	**5.95**	1979 - 1983	11.11	1.11	0.45	6.56	**5.95**
1984 - 1988	12.34	9.44	0.58	7.19	**6.01**	1979 - 1988	11.73	6.69	0.44	5.51	**4.79**
1989 - 1993	9.00	8.07	0.45	0.74	**0.21**	1979 - 1993	10.82	7.17	0.44	3.96	**3.41**
1994 - 1998	17.85	12.99	1.01	-0.50	-1.30	1979 - 1998	12.58	9.00	0.63	3.66	**2.73**
1999 - 2003	-0.55	16.40	-0.23	1.27	0.72	1979 - 2003	9.95	10.96	0.34	1.76	**1.78**
	OPT						OPT				
1979 - 1983	11.11	1.11	0.45	6.56	**5.95**	1979 - 1983	11.11	1.11	0.45	6.56	**5.95**
1984 - 1988	11.58	8.96	0.53	6.23	**5.05**	1979 - 1988	11.35	6.35	0.41	4.94	**4.21**
1989 - 1993	8.18	8.54	0.33	-0.80	-1.33	1979 - 1993	10.29	7.15	0.37	2.86	**2.31**
1994 - 1998	18.52	12.86	1.07	0.36	-0.44	1979 - 1998	12.35	8.95	0.60	3.33	**2.39**
1999 - 2003	-0.31	16.62	-0.22	1.57	1.02	1979 - 2003	9.82	10.99	0.33	1.56	**1.58**
	NLNP						NLNP				
1979 - 1983	11.11	1.11	0.45	6.56	**5.95**	1979 - 1983	11.11	1.11	0.45	6.56	**5.95**
1984 - 1988	10.75	8.80	0.45	4.66	**3.48**	1979 - 1988	10.93	6.24	0.35	3.95	**3.23**
1989 - 1993	7.83	8.51	0.29	-1.32	-1.84	1979 - 1993	9.90	7.07	0.32	2.06	**1.51**
1994 - 1998	18.15	12.86	1.04	-0.03	-0.83	1979 - 1998	11.96	8.90	0.56	2.72	**1.79**
1999 - 2003	-0.67	16.34	-0.24	1.13	0.58	1979 - 2003	9.43	10.87	0.30	1.06	**1.08**

TBYD & TRSD

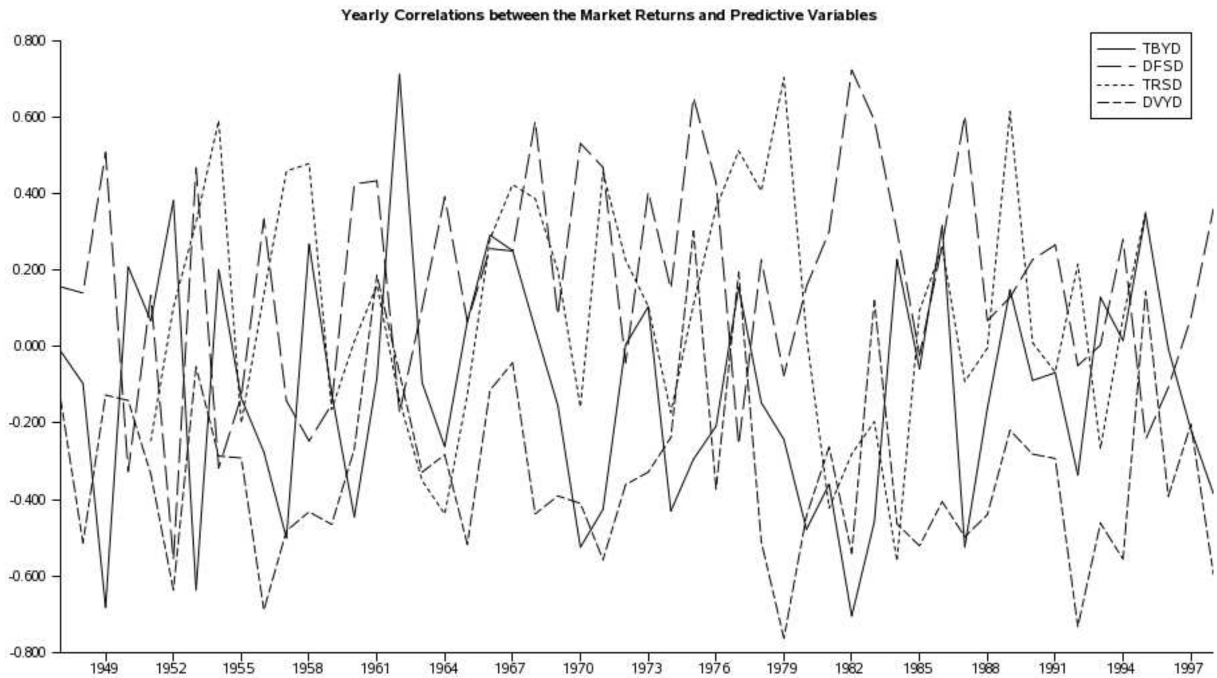
	VAR						VAR				
1979 - 1983	13.72	5.32	0.55	8.12	**7.52**	1979 - 1983	13.72	5.32	0.55	8.12	**7.52**
1984 - 1988	10.03	18.46	0.17	-0.39	-1.57	1979 - 1988	11.87	13.54	0.23	1.98	**1.26**
1989 - 1993	11.13	8.41	0.68	3.73	**3.20**	1979 - 1993	11.62	12.05	0.33	2.21	**1.66**
1994 - 1998	18.17	11.33	1.18	1.97	**1.17**	1979 - 1998	13.26	11.88	0.53	2.23	**1.30**
1999 - 2003	0.89	15.42	-0.16	2.62	2.07	1979 - 2003	10.79	12.72	0.36	2.04	**2.06**
	GARCH						GARCH				
1979 - 1983	14.30	4.97	0.70	10.38	**9.77**	1979 - 1983	14.30	4.97	0.70	10.38	**9.77**
1984 - 1988	10.69	18.47	0.21	0.27	-0.91	1979 - 1988	12.49	13.48	0.28	2.76	**2.04**
1989 - 1993	12.78	9.05	0.82	5.46	**4.93**	1979 - 1993	12.59	12.15	0.41	3.39	**2.84**
1994 - 1998	16.36	13.58	0.86	-2.65	-3.45	1979 - 1998	13.53	12.51	0.53	2.16	**1.23**
1999 - 2003	0.23	15.27	-0.20	1.86	1.31	1979 - 2003	10.87	13.17	0.36	1.95	**1.97**
	OPT						OPT				
1979 - 1983	12.32	2.63	0.59	8.76	**8.15**	1979 - 1983	12.32	2.63	0.59	8.76	**8.15**
1984 - 1988	10.03	18.46	0.17	-0.39	-1.57	1979 - 1988	11.17	13.13	0.18	1.21	**0.48**
1989 - 1993	10.63	10.11	0.52	1.64	**1.11**	1979 - 1993	10.99	12.18	0.27	1.36	**0.81**
1994 - 1998	17.32	11.20	1.12	1.10	**0.30**	1979 - 1998	12.58	11.95	0.47	1.32	**0.39**
1999 - 2003	-0.14	15.43	-0.22	1.48	0.93	1979 - 2003	10.03	12.77	0.30	1.10	**1.11**
	NLNP						NLNP				
1979 - 1983	12.27	2.64	0.57	8.49	**7.88**	1979 - 1983	12.27	2.64	0.57	8.49	**7.88**
1984 - 1988	14.42	13.18	0.57	7.03	**5.85**	1979 - 1988	13.34	9.47	0.48	6.18	**5.46**
1989 - 1993	10.61	8.24	0.64	3.12	**2.59**	1979 - 1993	12.43	9.06	0.53	5.26	**4.71**
1994 - 1998	20.28	10.69	1.46	5.74	**4.94**	1979 - 1998	14.39	9.52	0.78	6.01	**5.07**
1999 - 2003	-0.40	15.86	-0.23	1.29	0.74	1979 - 2003	11.43	11.18	0.47	3.72	**3.74**

Table 11: Accumulative Switching Portfolio Performance over Extended Periods

The predictive sample draws are generated from the 5-year expanding estimation and switching portfolios are formed as described in the text. The testing period is extended to 1959 and divided into nine 5-year periods coincident with the re-estimation. The average returns (r_p), standard deviations (σ_p), Sharpe ratio ($\mathcal{SR}$), Graham-Harvey measure ($\mathcal{GH}2$), CER (r_{ce}), and $\Delta_{\mathcal{GH}2}$ are reported.

Out-Of-Sample 5-Year Subperiod Switching Portfolio Accumulative Performance Over 1959 - 2003																				
	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$	$\Delta_{\mathcal{GH}2}$	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$	$\Delta_{\mathcal{GH}2}$	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$	$\Delta_{\mathcal{GH}2}$	$r_p(\%)$	$\sigma_p(\%)$	$\mathcal{SR}$	$\mathcal{GH}2$	$\Delta_{\mathcal{GH}2}$
	Expanding																			
1959 - 1963	7.28	19.31	0.24	-0.46																
1959 - 1968	7.12	16.63	0.22	-0.31																
1959 - 1973	3.82	16.79	-0.02	0.10																
1959 - 1978	3.35	15.75	-0.08	0.72																
1959 - 1983	4.88	14.40	-0.07	0.52																
1959 - 1988	5.51	13.44	-0.04	0.07																
1959 - 1993	5.78	12.58	-0.01	-0.34																
1959 - 1998	6.45	11.95	0.06	-1.16																
1959 - 2003	5.81	11.71	0.03	-0.78																
TBYD																				
	VAR					GARCH					OPT					NLNP				
1959 - 1963	4.38	8.91	0.19	-1.01	-0.55	1.89	7.53	-0.10	-4.77	-4.31	3.17	6.94	0.07	-2.53	-2.07	4.49	9.75	0.19	-1.08	-0.62
1959 - 1968	4.31	6.28	0.14	-1.28	-0.96	3.06	5.31	-0.07	-3.71	-3.40	3.70	4.89	0.05	-2.26	-1.95	4.36	6.86	0.13	-1.32	-1.00
1959 - 1973	5.86	5.77	0.30	4.13	4.03	5.11	5.10	0.19	2.76	2.66	5.20	4.69	0.23	3.23	3.12	5.63	6.11	0.25	3.43	3.33
1959 - 1978	5.56	5.17	0.18	4.38	3.66	5.36	4.42	0.17	4.18	3.46	5.34	4.09	0.18	4.30	3.58	5.75	5.29	0.21	4.82	4.10
1959 - 1983	6.75	4.89	0.19	4.13	3.60	6.51	4.04	0.17	3.85	3.33	6.49	3.75	0.18	3.98	3.46	6.82	4.80	0.21	4.38	3.86
1959 - 1988	7.62	5.78	0.28	4.82	4.75	7.48	5.35	0.28	4.78	4.71	7.34	5.02	0.27	4.65	4.58	7.48	5.66	0.26	4.53	4.46
1959 - 1993	7.73	6.23	0.29	4.08	4.42	7.70	5.81	0.31	4.34	4.68	7.46	5.64	0.28	3.86	4.20	7.53	6.14	0.26	3.68	4.02
1959 - 1998	9.11	7.44	0.45	4.56	5.73	8.97	7.16	0.45	4.55	5.71	8.84	7.03	0.44	4.41	5.57	8.29	6.37	0.40	3.79	4.96
1959 - 2003	7.97	8.92	0.28	2.97	3.76	7.91	8.70	0.28	2.97	3.75	7.83	8.65	0.27	2.85	3.63	7.29	8.12	0.22	2.13	2.91
TBYD & DFSD																				
	VAR					GARCH					OPT					NLNP				
1959 - 1963	0.80	7.74	-0.24	-6.53	-6.07	7.34	9.98	0.47	2.51	2.97	0.07	7.82	-0.33	-7.69	-7.22	5.57	10.59	0.27	0.03	0.49
1959 - 1968	2.52	5.48	-0.17	-4.85	-4.54	5.79	7.04	0.33	1.00	1.31	2.72	5.71	-0.13	-4.36	-4.05	4.90	7.46	0.19	-0.60	-0.29
1959 - 1973	3.52	7.48	-0.08	-0.68	-0.78	5.26	8.10	0.14	2.08	1.98	3.41	7.27	-0.10	-0.90	-1.00	5.38	6.70	0.19	2.69	2.59
1959 - 1978	3.35	7.58	-0.17	-0.49	-1.21	6.79	8.71	0.25	5.31	4.59	4.61	7.14	0.00	1.84	1.12	6.80	7.67	0.28	5.82	5.10
1959 - 1983	6.05	7.62	0.03	1.87	1.34	8.16	7.94	0.29	5.60	5.07	6.31	6.58	0.07	2.48	1.95	9.00	7.58	0.42	7.38	6.86
1959 - 1988	7.73	8.66	0.20	3.60	3.54	9.64	9.08	0.40	6.61	6.54	7.90	7.81	0.24	4.26	4.20	10.08	8.51	0.48	7.79	7.72
1959 - 1993	8.22	8.57	0.27	3.76	4.10	9.61	9.01	0.41	5.82	6.16	8.31	7.84	0.31	4.30	4.64	10.32	8.37	0.53	7.53	7.87
1959 - 1998	8.35	8.29	0.31	2.57	3.73	9.61	8.69	0.44	4.46	5.62	8.47	7.64	0.36	3.19	4.35	10.19	8.12	0.55	5.98	7.14
1959 - 2003	7.18	9.42	0.18	1.50	2.28	8.60	9.86	0.32	3.53	4.31	7.48	9.05	0.22	2.10	2.88	9.05	9.19	0.39	4.62	5.40

Panel A



Panel B

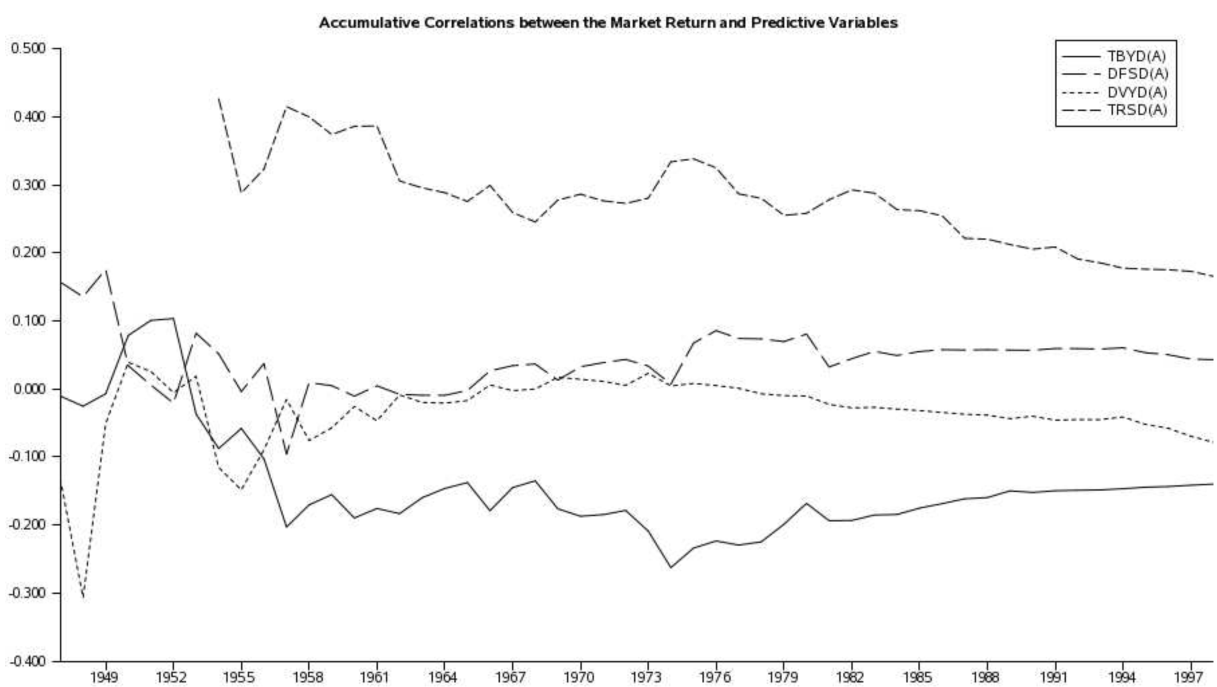


Figure 1: Correlations between the Market Return and Predictive Variables. Panel A: yearly correlations; Panel B: accumulative correlations.