A Comprehensive Look at the Return Predictability of Variance Risk Premia

Suk-Joon Byun*
Bart Frijns†
Tai-Yong Roh‡

Abstract:
The discrepancy between the in-sample and out-of-sample predictability of common return predictors for equity premiums has been widely discussed in the literature. There is also a growing amount of evidence that the variance risk premium predicts the excess returns of various assets in-sample. We examine the out-of-sample predictability of variance risk premiums (VRP) and the economic significance of the gains obtainable from using that predictability in market timing. We find strong evidence that VRP significantly predicts equity premiums out-of-sample and a simple market-timing strategy produces a certainty equivalent return (CER) of 1.89% per year. We also show that the VRP-based predictability model for international equity returns outperform the no-predictability benchmark in economic terms. We extensively examine out-of-sample predictability of VRP for other asset class such as equity portfolios, bonds, currencies and commodity indices. We find strong out-of-sample forecasting ability of VRP and its economic significance for equity portfolios and currency markets, but not for bond and commodity markets.

JEL classification: G12; G14

Keywords: Return predictability; Out-of-sample predictability; Variance risk premium; Economic significance of predictability; Macroeconomic uncertainty; Asset allocation

* Graduate School of Finance, KAIST (Korea Advanced Institute of Science and Technology) Business School, Hoegiro, Dongdaemoon-gu, Seoul, 130-722, Korea. Phone: +82-2-958-3968
E-mail: sjbyun@business.kaist.ac.kr
† Department of Finance, Auckland University of Technology, Private Bag 92006, 1142 Auckland, New Zealand. Phone: +64-9-921-9999 ext. 5706
E-mail: bart.frijns@aut.ac.nz
‡ Department of Finance, Auckland University of Technology, Private Bag 92006, 1142 Auckland, New Zealand. Phone: +64-22-167-9214
E-mail: tai-yong.roh@aut.ac.nz

A Comprehensive Look at the Return Predictability of Variance Risk Premia

Abstract

The discrepancy between the in-sample and out-of-sample predictability of common return predictors for equity premiums has been widely discussed in the literature. There is also a growing amount of evidence that the variance risk premium predicts the excess returns of various assets in-sample. We examine the out-of-sample predictability of variance risk premiums (VRP) and the economic significance of the gains obtainable from using that predictability in market timing. We find strong evidence that VRP significantly predicts equity premiums out-of-sample and a simple market-timing strategy produces a certainty equivalent return (CER) of 1.89% per year. We also show that the VRP-based predictability model for international equity returns outperform the no-predictability benchmark in economic terms. We extensively examine out-of-sample predictability of VRP for other asset class such as equity portfolios, bonds, currencies and commodity indices. We find strong out-of-sample forecasting ability of VRP and its economic significance for equity portfolios and currency markets, but not for bond and commodity markets.

JEL classification: G12; G14

Keywords: Return predictability; Out-of-sample predictability; Variance risk premium; Economic significance of predictability; Macroeconomic uncertainty; Asset allocation

1. Introduction

Starting from the work of Bollerslev et al. (2009), several studies have shown the empirical success of variance risk premiums (VRP)—the difference between model-free implied variances and realized variances—in predicting aggregate U.S. stock market returns at short horizons. (Bollerslev et al., 2009; Drechsler and Yaron, 2011; Bollerslev et al., 2014). This measure seems to be attractive to investors due not only to its good statistical performance in predicting stock market returns at short horizons, but also to the fact that it can avoid the issues of spurious regressions and biased estimates when forecasting excess returns with common predictors, which follow a near unit root process (e.g., Stambaugh, 1999; Ferson et al., 2003). Bollerslev et al. (2014) show that VRP are still viable predictors of returns in the international equity market. A number of papers have extended the scope of the research to find the predictive relation between VRP and the returns of various assets, such as bonds and currency (Mueller et al., 2011; Aloosh, 2012; Londono and Zhou, 2014).

However, most papers focusing on the return predictability of VRP point out the predictive relationship in-sample, not out-of-sample. These studies investigate the robustness of the return predictability of VRP in the aspect of finite sample bias, inclusion of alternative variables or various proxies for VRP, but they do not discuss their economic significance. Out-of-sample performance is very important in that this is one of the big issues widely criticized in the literature focusing on the existence of the return predictability (Goyal and Welch, 2003, 2008; Butler et al., 2005; Campbell and Thomson, 2008; Maio, 2014). A comprehensive analysis conducted by Goyal and Welch (2007) shows that common return predictors work poorly out-of-sample, generating low or negative out-of-sample R^2. As a result, it has become mandatory for researchers who develop new variables for predicting returns to conduct an out-of-sample analysis as a robustness check (Maio, 2014; Cooper and Priestley, 2009; Rangvid, 2006; Møller and Rangvid, 2013).

Therefore, we naturally raise the question as to whether the return predictability of VRP still holds out-of-sample, and gains obtainable from using that predictability are economically significant. Our study mainly examines the out-of-sample evidence for VRP as a robust predictor of equity premium, constructs simple trading strategies based on the out-of-sample forecasting power of VRP for excess equity returns, compares the performance of the trading strategies with those associated with alternative predictors. To the best of our knowledge, this comprehensive study is the first to examine the out-of-sample return predictability of VRP in a formal setting and its economic significance by constructing trading strategies using the conditioning information.

We perform out-of-sample tests used in Goyal and Welch (2008)'s comprehensive study and construct a parametric trading strategy based on one-month ahead out-of-sample predictability. By adapting Campbell and Thompson's (2008) method of exploiting return predictability, our trading strategy changes the portfolio weight of risky assets monthly, based on their predicted excess returns from the regression and return variance. As the risky asset becomes more mean-variance efficient to an investor, the trading strategy is to acquire additional risky assets. We assess the performance of the trading strategy carefully by looking at several measures of portfolio performance used in the literature.

We find strong evidence that VRP predicts excess market returns both in-sample and out-of-sample. The *t*-statistics associated with the slope coefficient on VRP is 5.11 and out-of-sample coefficient of determination (in %) is 5.80. The gains obtainable from using that predictability are economically significant. The stock market timing strategies based on VRP significantly outperform the buy-and-hold strategy as well as stock market timing strategies based on other popular predictors. Specifically, it produces a Sharpe ratio of 0.32 (versus 0.18 for the passive strategy) and a certainty equivalent return (CER) gain of 1.89% per year, which measures the extra utility generated by the market timing strategy if an investor utilizes it instead of simply holding risky assets. Most of the stock market timing strategies based on other predictors do not produce a Sharpe ratio that is significantly larger than a Sharpe ratio of the buy-and-hold strategy, and they produce a negative CER gain. The results are robust against the issues of parameter uncertainty, the sensitivity of forecasting schemes, and the market friction such as transaction costs or borrowing costs.

As a robustness test, we extend the methodology, applied to monthly U.S market data, to international equity market data. Bollerslev et al. (2014) show that a global risk premium, which is a weighted average of the individual country variance risk premia, results in strong in-sample predictability power for other 7 countries. We find strong out-of-sample predictability of the global VRP for most countries analyzed in Bollerslev et al. (2014). We provide new evidence that the out-of-sample forecasting power of VRP for individual countries generates significant economic gains for investors who engage in asset allocation strategies in international equity markets. Specifically, if an investor who seeks a mean-variance efficient portfolio uses the conditional mean estimates derived from the VRP-based predictability model in place of the no-predictability benchmark, the investor can increase the Sharpe ratio of his portfolio by roughly 0.42 and get extra utility gain larger than 1.1% per year.

The return predictability of VRP for the excess returns of equity portfolios and other assets in different financial markets is also examined. Specifically, we analyze with 6 representative equity portfolios, 12 foreign currencies, 6 commodity indices and bonds with various maturities and default risk. The results for equity portfolios show that VRP positively predict the excess returns of the portfolios and the predictive relationship holds out-of-sample. Further, we also apply the asset allocation framework, which is applied to international equity markets, to assess the economic significance. The asset allocation strategies using the return predictability of VRP at equity portfolio level yield higher annualized Sharpe ratio than those associated with the no-predictability benchmark (0.69 versus 0.56).

Among various assets in different markets, the VRP positively predict currency returns and negatively predict excess returns of long-term bonds with low default risk. We show that the predictive relationship between VRP and future currency returns exists for 9 out of the 12 countries and the robustness of the in-sample results is supported by the strong out-of-sample performance. The gains obtainable from using that predictability are also economically significant for most countries, but smaller in magnitude than those associated with equity markets. On the other hand, the relationship between VRP and future excess returns of Treasury bonds and Aaa-rated corporate bonds holds for out-of-sample weakly. However, its economic significance does not exist.

Our extensive work applied to various assets can be linked to the growing body of literature that has found a role for VRP as a fundamental factor driving movements in various financial markets around the world. Based on the theoretical framework developed by Bollerslev et al. (2009), the risk factor embedded in VRP captures general macroeconomic uncertainty and varies independently from the consumption growth risk, which is the main focus of long-run risk models (Bansal and Yaron, 2004). Mueller et al. (2011) find a predictive relation between VRP and excess bond returns. Londono and Zhou (2014) and Aloosh (2012) study the link between VRP and excess foreign exchange returns. Wang et al. (2013) conduct similar studies with data on credit spreads. We comprehensively re-examine the predictive relationship between the VRP and the excess returns of various assets both in-sample and out-of-sample. Furthermore, we study the economic significance of the predictive power of VRP for the excess returns of various assets.

Our analysis about the return predictability of VRP at the equity portfolio level also contributes to the literature on portfolio allocation in equity markets. Fleming et al. (2001) investigate volatility timing in equity markets. Karstanje et al. (2013) evaluate the economic value of liquidity timing in equity markets. Our work is more closely related to the latter, which analyzes the economic significance of return predictability rather than forecasting volatility.

The remainder of this paper proceeds as follows. In Section 2, we briefly explain the theoretical background of the return predictability of VRP and re-examine the in-sample predictability of VRP and other well-known predictors for excess stock returns. Section 3 analyzes out-of-sample performance and the performance of market timing strategies for the stock index. and individual portfolios. In Section 4, we extend our analysis to the equity portfolios and other assets in different financial markets such as currency, commodity, and bond markets. Section 5 sets forth the summary and conclusions.

2. Data Description and Review of In-Sample Evidence for Equity Premium

In this section, we briefly review the return predictability of common predictors for equity premium and revisit the in-sample return predictability for equity premium with recent data, covering from 1990 to 2013. The sample period includes three NBER recession periods. The basic predictive regressions are specified as

$$r^e_{t,t+q} = a_q + b_q x_t + \ u_{t,t+q}$$

where $r^e_{t,t+q}$ is the excess market return over q periods and x_t is the forecasting variable known at time t. We use the monthly excess market return defined as the difference between the return on S&P 500 composite index and the one-month Treasury bill rate.

2.1. Variance Risk Premium

Since the introduction of the measure called VRP by Bollerslev et al. (2009), VRP is regarded as the state variable linked to uncertainty about economic fundamentals.[4]

There are two major empirical findings reported in the literatures. First, VRP predicts future stock returns at short horizons strongly, not at long horizons. Second, there exists a pattern that the degree of predictability is the largest at 3-month or 4-month horizons, as indicated by t-statistics and R^2.

[4] Theoretical channels that justify the short run return predictability of VRP for equity premium have been provided by introducing additional process related to higher moments of economic fundamentals: A simple economy with additional consumption volatility of volatility process (Bollerslev et al., 2009) or Long-Run Risk model with Jump process (Drechsler and Yaron, 2011), combined with the Epstein and Zin (1989) form of representative agent's preferences. See p.4466-69 in Bollerslev et al. (2009) and p. 9-24 in Drechsler and Yaron (2011) for more details.

To implement our main empirical test, we use a proxy for VRP defined as $VRP_t \equiv IV_t - RV_{t-1,t}$, by following Bollerslev et al. (2009). Using this proxy means that $RV_{t-1,t}$ is a proxy for $E_t^P[Var_{t,t+1}]$. For forecasting purposes, this proxy is more appropriate than other proxies used in other studies (e.g., Mueller et al., 2011) since the VRP measure is available at time t (information set), implying that we can avoid uncertainty or errors related to estimation. The data is from Hao Zhou's website[5].

Table 1 tabulates the presents the mean, standard deviation, skewness, and excess kurtosis of returns and predictors. We also report AR (1) coefficients and unit root test statistics (Augmented Dickey-Fuller test) to check the persistency of predictors explicitly. VRP is positively skewed and very leptokurtic compared to other predictors. Specifically, as indicated by figure 1, extremely volatile movement of VRP during the recent financial crisis leads to extremely high kurtosis of VRP. The time-series of VRP is less persistent and non-unit-root process, as indicated by the AR (1) coefficient of 0.26.

Table 2 provides the pattern on the degree of predictability for each predictor. It shows that the predictive slope associated with VRP is significantly positive for 1, 3, 6, and 12-month horizons. The $\bar{R}^2$ (Adjusted R-squared) of the regression equation soars to 11.0% at 3-month horizons and decreases as the forecasting horizons increases.

The overall pattern of the degree of predictability is consistent with the implication from the calibrated theoretical model developed by Bollerslev et al. (2009). However, the results that the forecasting power of VRP is not only significant at the monthly horizon, but also at relatively long horizons such as 12 months are different from the results reported by Bollerslev et al. (2009). Therefore, one-month ahead out-of-sample return predictability of VRP for the stock index should be tested to check the robustness of the in-sample results with 1-month horizon.

In the next sections, we briefly revisit the empirical evidence regarding the in-sample predictability of other common predictors for the aggregate equity premium over the past 20 years.

[5] https://sites.google.com/site/haozhouspersonalhomepage/

2.2. Other Predictors

To highlight the empirically stylized fact that VRP predicts equity premium at short horizons and shows strong out-of-sample forecasting power, we selected the following alternative equity premium predictors for comparison, based on the comprehensive study of Goyal and Welch (2008). Specifically, we use all the predictors used for monthly regressions in their study. The predictors are Dividend Price Ratio (**d/p**), Dividend Yield (**d/y**), Earnings Price Ratio (**e/p**), Dividend Payout Ratio (**d/e**), Stock Variance (**svar**), Book to Market Ratio (**b/m**), Net Equity Expansion (**ntis**), Treasury Bills (**tbl**), Long Term Yield (**lty**), Long Term Rate of Return (**ltr**), Term Spread (**tms**), Default Yield Spread (**dfy**), Default Return Spread (**dfr**), and Inflation (**infl**). We do not include Cross-Sectional Premium (**csp**) since the monthly series of **csp** are only available until 2002. The monthly series of the common predictors in our study are available from Amit Goyal's website[6].

Table 1 reports the persistency of the alternative predictors. Most predictors have first-order autocorrelations above 0.95. We conduct ADF (Augmented Dickey-Fuller) unit-root test to categorize each predictor into unit-root process variables (Near-unit-root predictors) and non-unit-root process variables (Non-unit-root predictors). The last column of Table 1 reports p-value for null hypothesis that the predictors have a unit root. Based on the p-value of the ADF test, near-unit-root predictors are **d/p**, **d/y**, **b/m**, **ntis**, **tbl**, **lty** ,and **tms**. Non-unit-root predictors are **e/p**, **d/e**, **svar**, **ltr**, **dfy**, **dfr**, and **infl**.

In a sharp contrast to in-sample predictability results associated with VRP at 1-month horizon, the slope coefficient estimates associated with alternative predictors are significant at the 5% level only for **svar**, implying that most common predictors has one-month forecasting power on excess market returns. Further, predictors categorized as non-unit-root predictors do not seem to predict to predict the future excess market returns at short horizons.

At long horizons longer than 1-year, most near-unit-root predictors have strong forecasting power on excess market returns. We easily find a well-known pattern reported in the literature that the degree of

[6] http://www.hec.unil.ch/agoyal/

predictability associated with near-unit-root predictors increases as forecasting horizons increase. Some unit-root predictors such as **d/p**, **d/y**, **b/m** even predict future excess market returns significantly at 3-month or 6-month horizons, but the degree of predictability is much less than that associated with VRP. The results suggest that the persistency of predictors has a big impact on the pattern on the degree of predictability.

Overall, we mainly reconfirm that under recent 24 years, in-sample forecasting power and the pattern of the degree of predictability associated with VRP (and other well-known predictors) are consistent with the results reported by the previous studies. VRP predicts short-term future excess stock returns positively. Furthermore, VRP outperforms other common predictors at short-horizons.

In the next section, we investigate whether the return predictability of VRP for the stock index reported from the in-sample analysis still holds in an out-of-sample analysis and whether this measure is economically significant.

3. Market Timing Strategy Based on the Out-of-sample Forecasting Power: Stock Index

3.1. Out-of-sample Regressions

In this section, we conduct statistical tests designed to assess the out-of-sample forecasting power of VRP and alternative predictors.

We verify that the in-sample predictability of VRP for the excess market return is stronger at short horizons than are the other predictors that we analyzed. The analysis in this section deals with common concerns expressed in the literature regarding the economic validity of predictive models. Investors who allocate their wealth using predictive models are concerned that the predictive models work well for the future, not the past. We analyze the statistical measures used by Goyal and Welch (2008) to question the out-of-sample predictive ability of the return forecasting models based on VRP. Those measures diagnose whether the predictive models are stable to use as a basis for the market timing strategy.

The first measure is R_{OS}^2, which measures the proportional reduction in the mean squared error for the OLS model with the predictor relative to the historical mean model. R_{OS}^2 is computed as

$$R_{OS}^2 \;\; = 1 - \frac{MSE_A}{MSE_N}$$

where $MSE_A = \frac{1}{T}\sum_{t=1}^{T} e_{A_t}^2$ denotes the mean squared error for the OLS model with the predictor and $MSE_N = \frac{1}{T}\sum_{t=1}^{T} e_{N_t}^2$ denotes the mean squared error for the historical mean model. T is the number of observations of the out-of-sample regressions.

The second measure is McCracken's (2007) *F*-statistic, which is designed to test statistically whether the OLS model with the predictor can beat a historical mean model in terms of forecasting performance. The null hypothesis of this statistics test is that the unrestricted model, typically the model based on the predictive regression, does not have better predictive power for excess returns than the restricted model (the historical mean model). An alternative view is that this forecasting variable contains additional information and could be used to obtain a better forecast. This measure is calculated as

$$MSE - F = T \times \left(\frac{MSE_N - MSE_A}{MSE_A}\right)$$

We use the critical value derived by McCracken (2007) to obtain statistical inference for the MSE-F statistics that we compute.

The third measure is ENC, which was also designed as a statistical test and proposed by Clark and McCracken (2001):

$$ENC = \left(\frac{\sum_{t=1}^{T}(e_{N_t}^2 - e_{N_t} \cdot \; e_{A_t})}{MSE_A}\right)$$

We also use the critical value derived by Clark and McCracken (2001) to obtain statistical inference for the ENC statistics we compute.

The fourth measure is R_{OS-CT}^2, a measure modified from R_{OS}^2, proposed by Campbell and Thompson (2008). The method is similar to the way we calculate R_{OS}^2, but restricts the sign of the predicted expected excess return estimate. We set the predicted value to zero whenever it is negative and obtain corresponding residuals to calculate the statistics using the formula for R^2_{OS}. In this way, we avoid the situation of a negative equity premium, which is inconsistent with the theory.

We use an initial sample of 120 months (January 1990 to December 1999) to conduct the first predictive regression. The period for the out-of-sample analysis is from January 2000 to December 2013.

Table 3 summarizes the results for the out-of-sample predictive regressions. The results show that VRP has strong out-of-sample predictive power. The value of R^2_{OS} for VRP is positive and the magnitude (5.80) is similar with the evaluated in-sample $\bar{R}^2$ (4.90). The values of the R^2_{OS-CT} statistics (4.00) are also similar to that of the R^2_{OS} statistics, which means that the results are robust against the theoretical restriction. The values of the MSE-F and ENC statistics for VRP indicate that we reject the null hypothesis that the expected squared forecasting error of the historical mean model is lower than that associated with the predictive regression based on VRP (at the 5% level).

On the contrary, no predictors other than VRP significantly outperform the historical mean model based on the four criteria, R^2_{OS}, R^2_{OS-CT}, ENC and MSE-F. Specifically, most common predictors have negative values or positive values close to zero for the R^2_{OS}. The evaluated ENC statistics and MSE-F statistics associated with **d/p**, **d/y** and **b/m** are positive, but there are no cases for the three predictors that both evaluated ENC statistics and MSE-F statistics are statistically significant at the 5% level. VRP is the only variable that has strong out-of-sample performance based on the four criteria.

To analyze the forecasting performance pattern as to whether the model based on the predictive regression outperforms or underperforms the historical mean model over the sample period, we follow Goyal and Welch (2008) to plot the time-series of the difference between the cumulative sum of squared prediction errors of a historical mean model and those of a model with predictive variables. We mainly focus on the predictors showing either positive R^2_{OS}. Figure 2 plots the difference between the cumulative sum of squared prediction errors (SSE) of conditioning models and a historical mean model for VRP, **d/p**, **d/y**, and **b/m**. An increase in the cumulative SSE difference indicates better performance of the model with predictors; a decrease in the cumulative SSE difference indicates better performance of the historical mean model.

The model based on VRP significantly outperforms the historical mean model (unconditional model) from the beginning of the sample (2000), and the magnitude of the outperformance is more amplified since the start of the Global Financial Crisis. We also calculate the average of the difference between squared forecasting errors for conditioning model and those for the historical model across economic states to check the robustness of our results. The average for the NBER expansion is 0.0010 and the average for the NBER recession is 0.0028, indicating that the outperformance is not solely driven by the outperformance in turbulent periods.

However, as indicated by Figure 2, the outperformance associated with other predictors (**d/p**, **d/y**, and **b/m**) are not significant in turbulent periods. As a predictive model for excess market returns, these models are not valid in bad states.

To summarize, VRP appears to be a robust predictor of excess market returns according to our four OOS test statistics. This feature is quite impressive in that other common predictors couldn't survive the OOS tests. Further, the model conditioning on VRP captures equity premium more precisely in bad economic states. In sum, the out-of-sample tests associated with VRP indicate that the model based on the predictive power of VRP can be used for constructing a market timing strategy for the stock index. In the next section, we construct a market timing strategy that exploits the predictive models' out-of-sample forecasting power for the stock index, and we assess the economic significance of the market timing strategy for the stock index.

3.2. Construction of Market Timing Strategies and Performance Measures

In this section, we construct the market timing strategies that exploit the out-of-sample forecasting power of the predictive models and analyze the portfolio performance of those investment strategies.

Given the evidence that the return predictability of VRP for the stock index is robust out-of-sample, we construct a trading strategy based on the one-month ahead out-of-sample predictability (Breen et al., 1989; Goyal and Santa-Clara, 2003; Campbell and Thompson, 2008, among others).

At each time t, we conduct a one-month predictive regression on the excess market return based on conditioning information available up to t,

$$r_s^e = a + bx_{s-1} + u_s, \qquad s = 1, \dots, t$$

where x_s is the value of the predictor at time s. Then we can extract the forecasted excess returns for the next period at each time, $\widehat{r_{t+1}} = \hat{a} + \hat{b}x_t$

The market-timing trading strategy allocates portfolio weights to the stock market index and the risk-free asset based on the procedures used by Campbell and Thompson (2008) and Ferreira and Santa-Clara (2011). These trading guidelines reflect an investor's optimal decision to exploit the predictive relationship more than trading guidelines that simply shift an investor's portfolio entirely to stocks and T-bills.

The portfolio weights are derived from an optimization problem of an investor with a mean-variance objection function specified as

$$U(R_{p,t+1}) = E(R_{p,t+1}) - \frac{\gamma}{2} Var(R_{p,t+1})$$

where γ represents the level of relative risk aversion. The portfolio weight for the stock market index is specified as

$$\omega_t = \frac{\hat{r}_{t+1}}{\gamma \widehat{Var}(R_{t+1})}$$

where $\hat{r}_{t+1}$ is the fitted excess return from the predictive regressions, γ is fixed at three, and $\widehat{Var}(R_{t+1})$ is the variance of the return on the risky asset, computed based on the time-series of recent five-year monthly returns. We also set the parameters for ω_t in order to avoid a situation involving high leverage or a large short sale. We constrain the portfolio weights [-0.5, 1.5].

Using the portfolio weight and return series of a risky asset and a risk-free asset, the time-series of the realized returns of the market-timing trading strategy can be derived from

$$R_{p,t+1} = \omega_t R_{t+1} + (1 - \omega_t) R_{f,t+1}$$

In order to compare the performance of the market timing trading strategy to the passive trading strategy that simply holds the risky asset (buy-and-hold), we compute the average returns, standard deviations, skewness, excess kurtosis, FF alpha (Fama-French 3 factors alpha), FFC alpha (Fama-French-Carhart 4 factors alpha) and Sharpe ratios associated with both the active strategy and the passive strategy. FF alpha and FFC alpha assess if the economic relevance of predictors is linked to existing risk factors. We calculate the p-values associated with the alpha by a bootstrap method[7] used by Anderson et al. (2012).

We also calculate a simple variant of Sharpe ratio corrected by a skewness adjustment factor. Zakamouline and Koekebakker (2009) devised a performance measure, which is a generalized form of Sharpe ratio reflecting the investor's preferences to higher moments of distribution. We do not use the most generalized form of the measure which can be applied under any utility function and any distribution, but we use ASSR (Adjusted for skewness Sharpe ratio) under a CRRA utility function. It is calculated as,

$$ASSR = SR\sqrt{1 + \frac{Skew}{3}SR}$$

where SR is the standard Sharpe ratio. This metric measures attractiveness of the strategy in a mean–variance–skewness framework

We compute the certainty equivalent return (CER), which is a well-known utility-based performance measure, by following Campbell and Thompson (2008) and Ferreira and Santa-Clara (2011). The CER can be derived by taking the difference between the value of utility from the active trading strategy and the value of utility from the passive strategy (buy-and-hold),

$$CE = \mathrm{E}\left(R_{p,t+1}\right) - \mathrm{E}\left(\tilde{R}_{p,t+1}\right) + \frac{\gamma}{2}\left[\mathrm{Var}(\tilde{R}_{p,t+1}) - \mathrm{Var}\left(R_{p,t+1}\right)\right]$$

where $R_{p,t+1}$ represents the returns of the active strategies, $\tilde{R}_{p,t+1}$ represents the returns of the passive strategies, and γ is fixed at three. The CER can be interpreted as the management fee that an investor would be willing to pay to have access to the predictive regression forecasts instead of the

[7] See p.89-90 in Anderson et al. (2012) for more details.

historical average forecasts. This measure is similar to the Sharpe ratio, but we give weights for the average return and the volatility of return with proper levels of risk aversion of a particular investor.

The Omega is a simple generalization of the gain–loss ratio, developed by Keating and Shadwick (2002). It is calculated as the probability weighted ratio of gains versus losses for some threshold return target,

$$\Omega(r) = \frac{\int_r^{\infty}\big(1 - F(x)\big)dx}{\int_r^{\infty} F(x)dx}$$

where F is the cumulative distribution function, r is the threshold and partition defining the gain versus the loss. A larger ratio indicates that the asset provides more gains relative to losses for some threshold r and so would be preferred by an investor. We set $r = 0.002$ by reflecting average value of risk-free rate from 2000 to 2013.

Sortino ratio is simply a reward-to-downside risk ratio. It is calculated as,

$$\mathrm{S}(r) = \frac{R - r}{\sqrt{\int_{-\infty}^{r}(r - x)^2 f(x)dx}}$$

where R is the portfolio average realized return, r is the threshold and partition defining the upside and downside for the investment strategy under consideration. The term in the denominator is the square root of the downside semi-variance. When return distributions are near symmetrical and r is close to the distribution median, Sortino ratio and Sharpe ratio will produce similar results. However, as skewness increases and r vary from the median, results can be expected to show dramatic differences.

Following Thornton and Valente (2012), we also calculate the GISW statistics suggested by Goetzmann et al. (2007) as a performance measure to take into account possible portfolio manipulation issues. Since the Sharpe ratio and CER measures are based only on the mean and variance of the portfolio, it is possible to manipulate such moments to get high values from the performance measures by using high leverage or tilting away from the benchmark. We can interpret the GISW statistics as being similar to the CER. A positive GISW indicates that the active trading strategy outperformed the buy-and-hold strategy. GISW is calculated as

$$GISW = \frac{1}{1-\gamma}\left[ln\left(\frac{1}{T}\sum_{t=0}^{T-1}\left[\frac{R_{p,t+1}}{1+r_{f,t+1}}\right]^{1-\gamma}\right) - ln\left(\frac{1}{T}\sum_{t=0}^{T-1}\left[\frac{\tilde{R}_{p,t+1}}{1+r_{f,t+1}}\right]^{1-\gamma}\right)\right]$$

where T is the number of samples and γ is set at three.

The reason we assess portfolio performance not only with standard measures such as Sharpe ratio and CER, but also with various measures is to carefully look at the payoffs of each strategy with various angles. We take account into preference on higher moments, downside risk aversion, gain-loss preference and portfolio manipulation issue.

3.3. Performance of Active Strategies

Table 4 presents the mean, standard deviation, skewness, excess kurtosis, and evaluated portfolio measures of the monthly returns of the buy-and-hold strategy and the market timing strategy conditioning on the forecasting power of the predictors. The period for the analysis is the same as that of the out-of-sample analysis. With an initial sample of 120 months (January 1990 to December 1999) to conduct the first predictive regression, the market timing strategy starts at January 2000.

The market timing strategy conditioning on VRP generates an average return of 0.61% per month, a standard deviation of 4.92% per month, skewness of 0.2, and excess kurtosis of 5.24. The buy-and-hold strategy generates an average return of 0.40% per month, a standard deviation of 4.52% per month, skewness of -0.54, and excess kurtosis of 3.80. Briefly, it is difficult to conclude which one is better based on the moments of the trading strategies since the strategy conditioning on VRP has a higher average return and is less negatively skewed, but is more volatile and leptokurtic.

The evaluated performance measures strongly indicate that the market timing strategy conditioning on VRP significantly outperforms the buy-and-hold strategy. The market timing strategy conditioning on VRP has a Sharpe ratio of 0.32, whereas the buy-and-hold strategy has a Sharpe ratio of 0.18. The higher Sharpe ratio of the market timing strategy conditioning on VRP is mainly due to its higher mean return than that of the buy-and-hold strategy.

The market timing strategy conditioning on VRP yields a CER of 1.89% per year, meaning that an investor can benefit from the extra utility generated by this market timing strategy if the investor chooses that strategy instead of simply holding the risky asset. Other portfolio measures also support the economic significance of VRP. Omega. Sortino and ASSR measures associated with the market timing strategy conditioning on VRP are also significantly higher than those associated with the passive strategies, indicating the results are robust if we consider the issue of preference on skewness, downside risk aversion, gain-loss preference. The evaluated GISW statistics are also positive and significant (2.05% per year). Therefore, the results are also free from portfolio performance manipulation issues. To summarize, the market timing strategy conditioning on VRP clearly outperforms the passive strategy and is thus economically significant.

On the contrary, most of the market timing strategies conditioning on other predictors are more negatively skewed and leptokurtic than the buy-and-hold strategy, which is less attractive for an investor. Moreover, those strategies produce significantly lower Sharpe ratios than the buy-and-hold strategy and produce negative CERs and negative GISWs, indicating underperformance against the buy-and-hold strategy. One exception is the market timing strategy conditioning on **e/p**. This strategy produces a monthly Sharpe ratio of 0.47, a CER of 4.47%, and a GISW of 4.58%.

To look at the direct linkage between statistical significance of out-of-sample test and economic gains from the predictive relationship, we check Campbell and Thompson (2008)'s prediction that .a mean-variance investor can increase monthly expected portfolio return by a proportional factor of R_{OS}^2/S^2 (S: Unconditional Sharpe ratio of the risky asset) from a conditional model. Therefore, the implied value of ratio between the expected returns of portfolio using the model conditioning on VRP and those associated with the no-benchmark case is 21.48. However, if we calculate with the ratio by using realized returns in our sample, the ratio is 0.53. Even if the calculation of the ratio is based on expected returns, the magnitudes of the difference are very large. One reason for the huge difference between the implied ratio from the prediction by Campbell and Thompson (2008) and the actual ratio might be due to restriction on the weight on risky asset, as pointed by Campbell and Thompson (2008). Further, the difference is also amplified by smaller Sharpe ratio (0.18) of stock index in our sample period (from 2000 to 2013) than long-term unconditional average (0.37 since 1871). A long series of data should be needed to have clear look about Campbell and Thompson (2008)'s prediction, which states the direct linkage between statistical significance of out-of-sample test and economic gains from the predictive relationship.

To get closer look at the outperformance of the conditioning models over the passive strategy, we graph the empirical distribution associated with the passive strategy and two market timing strategies outperforming the passive strategy, the strategy based on VRP and **e/p**. The figure 3 shows that the strategy based on VRP has lighter left tail than the passive strategy and the strategy based on **e/p**. Even though the probability associated with large upside movements is relatively low, the strategy based on VRP has given quite stable payoff during our sample periods.

Specifically, the outperformance of strategy based on VRP over the passive strategy is more pronounced during the NBER recession. The mean return of the strategy based on VRP is 0.08% in the NBER recession period whereas the mean return of the passive strategy is -2.89%. The mean return of the strategy based on **e/p** is also lower than 0 (-1.6%). For investors' perspective, the strategy based on VRP is most attractive one among the strategies in our analysis because the payoffs are not much affected by economic conditions.

Overall, the results of this section show that the VRP is quite useful for constructing a market timing strategy for the stock market and that it significantly outperforms the passive strategy that simply holds stocks. The outperformance is robust against any possible issues related to performance measure manipulation, preference on skewness, downside risk aversion, gain-loss preference. Moreover, the strategy based on the predictive power of VRP outperforms most of the market timing strategies based on the predictive power of alternate predictors.

3.4. Robustness Checks

In this sub-section, we perform several checks to establish the robustness of our main results.

3.4.1. Length of Initial Estimation Sample and Rolling Scheme

We consider robustness checks on two issues arising from the forecasting scheme used to construct the market timing strategies. First, since our main results are based on an expanding window with initial length of 120 months, it might be problematic if there was a structural change or regime shift during the sample period that changed the predictive relationship between excess returns and the forecasting variables. We mitigate this problem by using a rolling scheme that uses only the most recent data.

Second concern is that 10-years for the initial in-sample regression seems to be a very small period, especially in predictive regressions for monthly stock returns that are known to be very noisy. Most paper assumes at least 20 years of data are needed to begin OOS forecasts (Goyal and Welch, 2008; Campbell and Thompson, 2008). Due to the fact that VRP data is available from 1990 to 2013 (24 years of data), we need to achieve an appropriate balance between a reasonable sample to produce the first forecasts and a still long enough period for the OOS test by using 10 year of initial sample. We make a robustness check by analyzing our results with and initial length of 180 months (15 years).

Therefore, we conduct the out-of-sample statistical test and construct market timing portfolios using the forecasting power of VRP for total 4 cases, either rolling or expanding window with initial length of 120 and 180 months

Table 5 reports the out-of-sample statistical test results and portfolio performance associated with each forecasting scheme. This table also includes the results for our basic forecasting scheme (Expanding, initial length of 120 months), reported in table 3 and 4. For all 4 cases, the R_{OS}^2 statistics is much higher than zero and statistically outperform historical mean models, as indicated by evaluated ENC and MSE-F statistics. Second, the market timing strategies conditioning on VRP outperform the buy-hold strategies for all 4 cases. The market timing strategies generate CERs and GISWs larger than 1% per year. We find a pattern that the cases using the scheme with expanding window show stronger out-of-sample forecasting power and produce better portfolio performance than the cases with rolling window.

Overall, the out-of-sample forecasting power of VRP on excess market returns are robust against issues arising from selecting our main forecasting scheme which has relatively short length of initial estimation sample and uses expanding window.

3.4.2. Parameter Uncertainty

We also check the issue about parameter uncertainty. As Connor (1997) noted, a mean-variance optimizer tends to severely overweigh those securities with positive estimation errors in their expected-return forecasts and severely under-weigh those with negative estimation errors. If the time of a positive estimation error and the time of a good market state coincide, we might wrongly conclude the existence of outperformance of the market timing strategy based on the forecasting variable.

To take this issue into account, we adjust the parameter estimates according to the Bayesian setup, as suggested by Connor (1997).[8] The results are similar to those following our main scheme. Table 6 shows that the market timing strategy based on VRP with a parameter uncertainty adjustment has a annualized Sharpe ratio of 0.34, a CER of 2.49%, and a GISW of 2.68.

It is worth noting that the results of the market timing strategies conditioning on **e/p** have negative values for the CER and negative values for the GISW statistics with the parameter uncertainty adjustment, whereas they have significant positive values for the CER and GISW without the parameter uncertainty adjustment. Since the parameter uncertainty adjustment prevents problems of overweighing, it is possible that the outperformance of the market timing strategies conditioning on **e/p** is due to the coincidental timing of positive estimation errors in their expected-return forecasts and a good market state, especially during the period from 2001 to 2007.

3.4.3. Transaction Costs and Borrowing Costs

[8] See p. 3150 in Thornton and Valente (2012) for more details.

We also examine the impact of transaction costs on the economic gains of the market timing strategies conditioning on VRP. Anderson et al. (2012) show that one market timing strategy, called risk parity strategy, generates too high transaction costs relative to its benefits. Since the VRP signal is less persistent than other common predictors, there are more variation of the weights on equity index for the strategies associated with VRP. Figure 4 shows the time-series weights for VRP and **d/y**, which is a typical near-unit-root predictor, for comparison. The weights for VRP shows quite sizable variations whereas the weights for **d/y** are not changed frequently and 1.5 in most times.

By following Anderson et al. (2012), we assume turnover-induced trading costs of 0.1% to estimate total trading costs arising from our market timing strategies. Specifically, let $\widetilde{\omega}_t$ be the weight on risky asset after reflecting the price movement of risky asset and the risk-free rate over a single period at each time t, it is given by

$$\widetilde{\omega}_t = \frac{\omega_{t-1}(1+r_t)}{(1-\omega_{t-1})\left(1+r_{f,t}\right)+\omega_{t-1}(1+r_t)}$$

Therefore, the turnover ,which is required to meet target weight ω_t discussed in section 3.2, is given by

$$x_t = |\widetilde{\omega}_t - \omega_t|$$

Trading cost-adjusted returns are given by

$$r_t' = r_t - x_t c$$

where c denotes the rate of turnover-induced trading costs. We set $c = 0.001$.

We also examine the effect of borrowing cost on the profitability of the market timing strategy conditioning on VRP based on the assumption used by Anderson et al. (2012) since we allow leverage between 100% and 150%. We assume that if the weight on risky asset at time $t-1$ exceeds 100% ($\omega_t > 1$), the implied financing cost can be calculate as,

$$r_{f,t}' = \frac{(\omega_{t-1}-1)}{\omega_{t-1}} r_{f,t}^B + \frac{1}{\omega_{t-1}} r_{f,t}$$

where $r_{f,t}^{B}$ is the borrowing rate. The implied financing cost is the value-weighted average of financing costs for levered position and those for self-financed position. Then, borrowing cost-adjusted excess returns are given by

$$r'_{e,t} = r_t - r'_{f,t}$$

We use the U.S. three-month Eurodollar deposit rate as a proxy for the borrowing rate. The Eurodollar deposit rate data is from the Federal Reserve Economic Data at the Federal Reserve Bank of St. Louis.

The annualized Sharpe ratio of the market timing strategies conditioning on VRP decreases from 0.32 to 0.28 and the CER decreases from 1.89% to 1.24% when we consider the impact of transaction costs. The annualized Sharpe ratio of the market timing strategies conditioning on VRP decreases from 0.32 to 0.28 and the CER decreases from 1.89% to 1.34% when we consider the effect of borrowing cost. Further, the annualized Sharpe ratio of the market timing strategies conditioning on VRP decreases from 0.32 to 0.25 and the CER decreases from 1.89% to 0.69% when we consider the effect of both borrowing cost and transaction cost. (Untabulated)

The evaluated values of CER and Sharpe ratio indicate that the effect of borrowing cost and transaction cost are not influential and the economic significance of the return predictability are still valid.

Overall, the results of this section show that the issues of parameter uncertainty, the sensitivity of forecasting schemes, and the market friction such as transaction costs or borrowing costs do not affect the conclusion that the out-of-sample forecasting power of VRP for stock index excess returns is economically significant.

3.5. International Evidence

3.5.1. Global VRP and the Return Predictability for non-U.S countries

Bollerslev et al. (2014) shows the in-sample predictability of country-specific VRPs exists for a set of seven non-U.S countries, although the magnitude of the predictability and the statistical significance observed for non-U.S countries are albeit weaker than those observed for the United States. They also introduced a global-VRP which shows stronger in-sample predictability than country-specific VRPs in the non-U.S countries. Motivated by the empirical evidence reported by Bollerslev et al. (2014), we examine OOS predictability of VRP in an international context to check whether our results are an outcome of an elaborate data snooping process.

We apply the same methodology, applied to the U.S market, to study OOS predictability of the global VRP (GVRP) for the seven non-U.S countries. The predictor we focus on here is GVRP, not the country-specific VRPs because GVRP has unique feature that predicts market excess returns for each individual country as a global variable and provides more accurate predictions than the country-specific VRPs. The seven non-U.S countries are France (CAC 40), Germany (DAX 30), Japan (Nikkei 225), Switzerland (SMI 20), Netherlands (AEX), Belgium (BEL 20), the United Kingdom (FTSE 100).

Due to lack of availability of intraday data for each market, we use the sum of the daily squared returns over a month to construct end-of-month realized variances $RV_{t-1,t}$ for each of the countries. We obtained the corresponding end-of-month model-free implied volatilities ($IV_{i,t}$) 1/2 for the S&P 500 (VIX) from the CBOE, the CAC (VCAC), the DAX (VDAX) FTSE (VFTSE), SMI (VSMI), AEX (VAEX), and BEL (VBEL) were obtained from Datastream whereas the Japanese volatility index (VXJ) were obtained directly from the Center for the Study of Finance and Insurance at Osaka University. Country-specific VRPs are constructed by the taking difference between model-free implied variance and realized variance. GVRP is constructed by the weighted sum of the country-specific VRPs,

$$GVRP_t \equiv \sum_{i=1}^{8} w_t^i VRP_t^i$$

where $i = 1,2, \dots .,8$ refers to each of the eight countries included in our analysis. The end-of-month market capitalizations data from Thomson Reuters Institutional Brokers¨ Estimate System (IBES) via Datastream is used for the weights. Since, most model-implied variance in international data are available after 2000, the initial estimation period is from January 2000 to December 2009. For out-of-sample test and constructing market time strategies, we use data from 2010 to 2014. We use dollar denominated returns rather than local currency denominated returns to assess economic significance by constructing market timing strategies in international equity market.

Panel A of Table 7 summarizes the results from the one-month ahead predictive regressions of GVRP on excess returns of the equity index for the seven non-U.S countries. The results show that GVRP significantly predicts excess returns of equity index for 5 out of 7 countries. These slope coefficient estimates associated with GVRP are significant for 4 out of 7 countries at the 5% level and significant at the 10% level for Germany. One notable exception is Japan. The slope coefficient estimates for Japan are even negative and insignificant. In sum, these results reconfirm that GVRP is a common factor imbedded in the expected excess return of equity index in international market.

Panel B of Table 7 indicates that these statistically significant in-sample return predictabilities associated with the 5 countries still hold for out-of-sample, as indicated by positive R_{OS}^2 larger than 2.5% and statistically significant values of MSE-F, ENC statistics at the 5% level (One exception: ENC statistics is significant at the 10% level for Germany). Even though out-of-sample forecasting power of GVRP for non-U.S countries is less than that associated with VRP for U.S market, it is still significant at the 5% level.

We also construct the market timing trading strategies conditioning on GVRP for each individual country. Panel C of Table 7 shows that the market timing strategies exploiting predictive power of GVRP outperform corresponding passive strategies for the 5 countries. The difference between annualized Sharpe ratios of the active strategies and the passive strategies for the 5 countries are 0.44 on average. The CERs are 3.51% per year on average. We find the outperformance is especially strong for the U.K, France and Belgium, as indicated by CERs larger than 3.5% per year.

3.5.2. Asset Allocation Framework

To further examine whether return predictability for GVRP has economic significance in a real world setting, we follow Thornton and Valente (2012) to construct market timing strategies using the risk-free asset and the 7 equity index (except Japan) examined above. This analysis is a unified approach to examine the economic significance of GVRP in global equity market. We exclude markets returns for Japan since there is no in-sample predictability of GVRP for Japan[9]. The strategy is based on the asset allocation framework that constructs a mean-variance efficient portfolio.

For each month, an investor derives optimal weights of each asset that minimizes the conditional variance of portfolio return subject to achieving a target conditional mean. Specifically, let r_{t+1} denote the 7×1 vector which is consist of portfolio returns, μ_{t+1} is conditional expectation of r_{t+1} derived from either the model conditioning on VRP or the historical mean model and $\sum_{t+1} = E_t[(r_{t+1} - \mu_{t+1})(r_{t+1} - \mu_{t+1})']$ is the conditional covariance-covariance matrix of r_{t+1}. The conditional covariance-covariance matrix is calculated with recent 10-year (120 months) data. Let w_t be a 7×1 vector of portfolio weights. The asset allocation problem can be expressed as,

$$\min_{w_t} \quad w_t' \ \textstyle\sum_{t+1} \ w_t$$

$$\mathrm{s.t} \qquad \mu_{t+1} = \mu_c$$

The optimal portfolio weights for an investor are represented as,

$$w_t = \frac{\mu_c}{\mu_{t+1}' \sum_{t+1}^{-1} \mu_{t+1}} \textstyle\sum_{t+1}^{-1} \mu_{t+1}$$

To avoid extreme leverage or large short sale, we also set bounds for w_t between -50% and 150%. So returns of trading strategies are calculated by

$$R_{p,t+1} = w_t' r_{t+1} + (1 - w_t' 1) r_{f,t+1}$$

We calculate CER and GISW statistics to evaluate the economic significance of the out-of-sample forecasting power of VRP. In this case, we use the benchmark for calculation of CER, GISW with the returns of the market timing strategy using the historical mean as a conditional mean in the scheme above.

---

[9] As in Barroso and Santa-Clara (2012), only assets which are predictable (in-sample) should be included in the OOS portfolio optimization exercise. We also conduct a test with data including Japan and the results are qualitatively similar.

Table 8 reports the results for both the historical mean model and the model conditioning on GVRP at each target conditional mean level. We change the target conditional mean level from 0.001 to 0.005. The highest of unconditional mean of returns among 7 countries is 0.005 (Belgium) and the lowest is -0.001 (Japan) with first 120 months (10 years) data, meaning that the level of the target conditional mean is reasonable. The results show that returns of the market timing strategies based on the model conditioning on GVRP are more negatively skewed, less volatile, and less leptokurtic than the market timing strategies associated with the historical mean model for most cases of the target conditional mean level. Therefore, the market timing strategy conditioning on GVRP has a more attractive moments profile for an investor. The attractiveness of the strategy is also evidenced by the positive values of CER and GISW larger than 1% per year on average, meaning that GVRP generates larger economic gain than the historical mean model to an investor who faces an asset allocation problem with international equity index. The average value of CER is 1.13% per year and GISW is 1.11% for 5 cases of the target conditional mean level. with first 120 months (10 years) data

Overall, the in-sample predictability of GVRP reported by Bollerslev et al. (2014) still holds for out-of-sample, showing that out-of-sample forecasting power associated with VRP is robust in international data. Further, we can construct profitable market timing strategies exploiting return predictability of GVRP using equity index of various countries.

4. An Extended Analysis: Equity Portfolios

In this section, we comprehensively investigate the predictability of VRP on excess returns of equity portfolios.

4.1. Statistical Analysis

We select Small (the first decile portfolio sorted by size), Big (the tenth decile portfolio sorted by size), Growth (the first decile portfolio sorted by book-to-market), Value (the tenth decile portfolio sorted by book-to-market), Loser (the first decile portfolio sorted by momentum), and Winner (the tenth decile portfolio sorted by momentum) as testing assets. These testing assets are widely discussed in the literature. The portfolio return data are from the web page of Kenneth French.

Panel A of Table 9 summarizes the results from the one-month ahead predictive regressions for excess returns of the six portfolios. The results show that VRP significantly predicts all of the excess returns of the portfolios. The slope coefficient estimates are all positive and statistically significant at the 5% level. These results imply that VRP is a common factor imbedded in the expected excess return of assets in the equity market. One interesting result is that the coefficient estimate of VRP for the Loser portfolio is much larger than that of the Winner portfolio, which might indicate that we can find predictability evidence of VRP in zero-cost momentum-based strategies.

Panel B of Table 9 summarizes the results from the out-of-sample predictive regressions for excess returns of the six equity portfolios. The results show that VRP has strong out-of-sample predictive power for the excess returns of all 6 portfolios. Specifically, R_{OS}^2 and R_{OS-CT}^2 are all positive and the values of MSE-F and ENC statistics are statistically significant at the 5% level for all cases.

We also construct the market timing strategies for each equity portfolio using the same method that we applied to the stock index in order to check whether the statistical significance implies economic significance. Panel C of Table 9 shows that the market timing strategies exploiting predictive power of GVRP outperforms corresponding passive strategies for 4 out of 6 cases. The difference between annualized Sharpe ratios of the active strategies and the passive strategies for the 4 countries are 0.13 on average. The CERs are 2.23% per year on average. It is quite unusual that the market timing strategy for Winner and Small do not outperform the corresponding passive strategies even though VRP strongly predicts out-of-sample.

4.2. Asset Allocation Framework

To further examine the economic significance of the out-of-sample forecasting power of VRP for excess returns of assets in the equity market, we also apply the asset allocation framework used in international stock markets with a risk-free asset and the six equity portfolios examined above.

Table 11 reports the results for both the historical mean model and the model conditioning on VRP at each target conditional mean level. We set the target conditional mean to 0.006, 0.008, and 0.01. The lowest unconditional mean of the returns of the six portfolios is 0.003 (Loser), and the highest is 0.013 (Winner), meaning that the level of the target conditional mean is reasonable. The results show that for all the cases of the target conditional mean, the returns of the market timing strategy based on the predictive model with VRP are less negatively skewed and have a smaller standard deviation and smaller kurtosis than the market timing strategies associated with the historical mean model, which means that the strategy based on the predictive model with VRP has a more attractive moments profile for an investor.

The positive values of the CER and GISW indicate that the predictive model based on VRP generates larger economic gains than the historical mean model. Specifically, the average value of CER is 1.68% per year and GISW is 1.80% for 3 cases of the target conditional mean level. VRP is also useful to investors for constructing profitable trading strategies while facing an asset allocation problem with their equity portfolios.

Overall, the return predictability of VRP for the stock market exists at the aggregate level as well as at the equity portfolio level, indicating that VRP is a factor driving common movements of the equity market. The out-of-sample forecasting power of VRP and its economic significance also holds at the equity portfolio level.

4.3. Zero-cost strategies

To further examine this issue, we also investigate the predictability of VRP on the payoffs of the zero-cost strategies based on size (SMB), book-to-market (HML), and momentum (WML). This analysis might reveal a time-varying source of financial anomalies, as shown by Wu et al. (2010), who report the predictability evidence of VRP on the payoffs of zero-cost accruals-based strategies. SMB denotes the trading strategy that takes a long position for the Small portfolio and a short position for Big. HML takes a long position for the Value portfolio and a short position for Growth, while WML takes a long position for Winner and a short position for Loser. The portfolio return data are from the web page of Kenneth French.

Panel A of Table 8 shows that VRP predicts only zero-cost momentum-based strategies weakly, with a negative slope coefficient (*t*-statistics: -1.65). This negative slope coefficient is consistent with the empirical fact that WML is procyclical whereas VRP is countercyclical.

Panel B of Table 8 summarizes the results from the out-of-sample predictive regressions for SMB, HML, and WML. The results show that VRP has no out-of-sample predictability for SMB, HML, or WML. The in-sample predictability of VRP on WML does not hold out-of-sample. The case for WML implies that an out-of-sample test should be conducted as a robustness check to analyze the time-varying source of financial anomalies.

5. An Extended Analysis: Other Assets in Different Financial Markets

Recent empirical evidence shows that VRP predicts the excess returns of other assets in-sample and suggests that VRP captures aggregate economic uncertainty level. By following the work in Section 3 that focuses on the equity market, we undertake an extensive analysis of in-sample and out-of-sample tests to the bond market, commodity market, currency market, and credit derivative (credit default swap) market in an effort to better understand the role of VRP as a common factor driving risk premiums in various asset markets.

5.1. Bond Market

First, we investigate the information contained in VRP for bond excess returns. Not only we investigate the existence of out-of-sample forecasting power of VRP on bond markets, but also examine the pattern of the degree of predictability, the degree of out-of-sample forecasting power, and its economic significance, based on two dimensions: Bonds' maturity and default risk. To look at the effect of the bonds' maturity, we mainly analyze with short-term default free fixed income securities such as T-bills, short-term treasury bonds with maturity less than 5 years. We will draw full implication about the effect of maturity by combining with the results associated with long-term treasury bonds in the section for long-term bonds (Section 5.1.2). We use long-term bonds to examine the effect of default risk because most aggregate corporate bond indices are constructed by using long-term bonds. We cover from long-term treasury bonds as safest assets to High Yield corporate bond index as most speculative ones.

5.1.1. Short-Term Bonds

In the short-term bond category, we analyzed with the holding period excess returns of two to six-month T-bills and two to five-year Treasury bonds by following Mueller et al. (2012). However, our analysis is different from Mueller et al. (2012) in two ways. First, we analyze with one-month holding period excess returns of Treasury bonds with maturity longer than 1-year whereas Mueller et al. (2012) analyzed with 1-year holding period excess return of Treasury bonds. To analyze with 1-month holding period excess return of short-term default-free bonds, we use monthly series of total returns of U.S BENCHMARK DS GOVT. INDEX for 2, 3 and 5-year, provided by Datastream.

Second, Mueller et al. (2012) use a proxy for $E_t^P[Var_{t,t+1}]$ derived from the HAR-RV model, which is a parametric method proposed by Corsi (2009), whereas we use the one-month lagged values of the realized variance ($RV_{t-1,t}$), as a proxy for $E_t^P[Var_{t,t+1}]$. They show that VRP significantly predicts short-term bond excess returns, with a positive slope coefficient and that the predictive relationship still holds after controlling other factors that predict bond excess returns.

Therefore, our analysis reexamines the information contained in VRP for short-term bond excess returns with a different proxy for VRP, which is more appropriate for forecasting purpose, and studies the information contained in VRP for one-month holding excess return of intermediate-term bonds additionally.

We also select a one-month holding period excess returns of 2, 4 and 6-month Fama-Bliss T-bills in the short-term bond category. The sample is from January 1990 to December 2013. We conduct the out-of-sample test and examine the economic significance based on an expanding window with initial length of 120 months.

Panel A of Table 12 shows that VRP does not seem to predict one-month holding period excess returns of short-term bonds. The slope coefficients are negative, but not statistically significant for 5 out of 6 cases. One notable exception is 2-month T-bill. VRP positively predicts future excess returns of 2-month T-bill, but it is not statistically significant (*t*-statistics: 1.58). Even though the slope coefficients on VRP become more negative as bonds' maturity of short-term bonds becomes longer, it is difficult to regard those patterns associated with bonds' maturity as meaningful patterns since the slope coefficients are not statistically significant in most cases.

It is quite surprising that our results associated with T-bills are quite different from the results reported by Mueller et al. (2012), which show that VRP in their study positively predicts one-month holding period excess returns of T-bills for most cases. The main reason for the discrepancy between our results and the results reported by Mueller et al. (2012). is that we use a proxy for VRP that is different from the one used by Mueller et al. (2012). The results indicate that more studies are needed to ascertain the in-sample predictive relationship between VRP and excess returns of short-term bonds, with various proxies for VRP.

Panel B of table 12 shows that weak in-sample return predictability for 2 month T-bill does not hold for out-of-sample, as indicated by negative values of R^2_{OS} and MSE-F statistics. Panel C of Table 12 indicates that the market timing trading strategies conditioning on VRP for short-term bonds are not profitable. For all cases, either the Sharpe ratio of the active strategy is less than the Sharpe ratio of the passive strategy or the value of CER is less than 0.2% per year.

Overall, VRP does not have forecasting power for short-term bond excess returns based on in-sample analysis. There seems to be weak positive relationship between VRP and the excess returns of fixed income securities with very short maturity, but the predictive relationship does not hold for the out-of-sample analysis.

5.1.2. Long-Term Bonds

In the long-term bond category, we employ U.S Treasury bond and following corporate bond indices: Barclays U.S Treasury Long Index, Barclays U.S Treasury Aggregate Corporate Aaa Long Index, Barclays U.S Treasury Aggregate Corporate Baa Long Index, and Barclays U.S Treasury Corporate High Yield Index. Having returns of these bond indices with wide range of credit ratings is essential for drawing implication about default risk.

In Panel A of table 12, the slope coefficients obtained with excess returns of four long-term bonds show a distinguishing pattern that as default risk of a bond increases, the negative relationship between VRP and future bond excess returns becomes weaker. The slope coefficients are significant at the 5% level for Treasury bonds (*t*-statistics: -2.28) and Aaa-rated bonds (t-statistics: -2.24). On the other hand, there seems to exist a positive relationship between VRP and future excess returns of high yield bonds, but not statistically significant (*t*-statistics: 1.56).

Combined with the results in section 5.1.1, we find more profound negative relationship between default-free bond risk premia and VRP. One possible channel is from Bansal and Shaliastovich (2012)'s finding that bond risk premia rise with uncertainty about expected inflation and fall with uncertainty about expected (consumption) growth. They provide theoretical justification for the empirical results under the long-run risk model. If VRP is more related to uncertainty about real economy rather than inflation, the negative relationship between default-free bond risk premia and VRP will be generated.

Panel B of table 12 indicates that the strong in-sample return predictability associated with low default risk bonds still holds for out-of-sample. The value of R_{OS}^2 is positive for Treasury bonds (0.88) and Aaa-rated bonds (0.85). The values of the MSE-F for Treasury bonds and Aaa-rated bonds are statistically significant at the 5% level and The values of the ENC statistics are statistically significant at the 10% level. For Baa-rated bonds and high yield bonds, there is no evidence for significant out-of-sample forecasting power.

To further analyze the out-of-sample test results, we plot the difference between the cumulative sum of squared prediction errors of a historical mean model and those of the model conditioning on VRP for long-term bonds. The figure 5 indicates that the models conditioning on VRP in long-term bonds underperforms the historical mean model out-of-sample for most sample period of our analysis. For the case of long-term treasury bonds and Aaa-rated corporate bonds, the model conditioning on VRP outperforms the no-predictability benchmark only in the last 26 months of our sample period.

Panel C of Table 12 indicates that the market timing trading strategies using the weak out-of-sample forecasting power of VRP for Treasury bonds outperform the corresponding passive strategy as indicated by a CER of 1.28% per year, but not for Aaa-rated bonds as indicated by a CER of -0.23% per year.

Overall, there is an increasing pattern between the slope coefficients (in-sample) associated with VRP and the default risk of long-term bonds. The slope coefficients are statistically significant for long-term bonds with low default risk. However, those predictive relationships are weak for out-of-sample and the economic significance is also weak.

5.1.3. CDS Indices

In the previous section, we find that credit risk of bonds affects the relationship between VRP and future bond excess returns. The slope coefficients increase with credit risk of bonds. However, our analysis with corporate bonds still gives unclear look at the relationship between credit risk and return predictability since the returns of corporate bonds have non-credit risk components such as tax, liquidity, and interest rate risk.

To clearly look at the effect of credit risk on the return predictability of VRP, we use credit default swap (CDS) indices [10] as an alternative asset, which are standardized vehicles for hedging or speculating against market-wide credit risk in a highly liquid and cost-efficient way. Therefore, the returns of CDS indices mainly contain credit-related components.

To implement the research goal in this section by using data on the CDS indices, we must first overcome the issue that these indices have a relatively short history. Both the Dow Jones High Yield CDX index (CDX.NA.HY) and the Dow Jones Investment Grade CDX index (CDX.NA.IG), published by Markit Group Limited, which markets the CDX indices, were launched in April 2004. Thus, we cannot have any data for the out-of-sample test and for the returns of the market timing strategy even if we conduct the initial regression with first 120 months (10 years) data.

In order to avoid such a small sample issue, we have selected the CDX HY five-year total return index (Bloomberg ID: DBCDXHY5), offered by Deutsche Bank, the data for which is available from January 1997. This index is a total return version of the High Yield CDX index, which is an equal-weighted daily index composed of 100 high-yield entities. The total return version of the CDX index mimics the wealth of an investor who rolls his or her long credit risk position into the relevant on-the-run CDS index contract. Even though most studies use the CDS indices offered by Markit Group Limited, the correlation coefficient between the returns time-series of the High Yield CDX index from Deutsche

[10] Credit default swaps (CDS) are single-name over-the-counter credit derivatives that provide default insurance. The buyer of a CDS makes quarterly payments over the life of the contract in exchange for protection against a default event such as bankruptcy, failure to pay, or a debt-restructuring event for the reference entity.

Whereas single-name CDS is based on a single reference entity, CDS indices, which are synthetically constructed of various single-name CDSs, are widely referenced variables representing the credit market.

Bank and those from Markit Group Limited is 0.97, indicating that it is a minor issue to use data from Deutsche Bank rather than from Markit Group Limited.

The initial estimation period is from January 1997 to December 2006. Therefore, we have 7-year (January 2007 to December 2013) monthly return data for construction of the market timing strategies.

Panel A of Table 13 summarizes the results from the one-month ahead predictive regressions for the CDS returns. The results show that VRP significantly predicts CDS returns at a one-month horizon. For comparison, we also run the predictive regressions with Default Yield Spread (**dfy**) and Default Return Spread (**dfr**), which are main variables capturing overall credit market conditions. Interestingly, **dfy** and **dfr** do not predict CDS returns significantly at a one-month horizon, as indicated by the insignificant slope coefficient estimate (*t*-statistics: 0.19). This might be due to the weak short run forecasting power of the predictors, which have near-unit-root process, on excess returns of risky assets (equity).

Panel B of Table 13 summarizes the results from the out-of-sample predictive regressions. The results show that VRP has positive and strong out-of-sample predictive power for CDS returns. The OLS model with VRP significantly reduced the mean squared error for one-month ahead CDS returns relative to the historical mean model. Specifically, we reject the null hypothesis that the expected squared forecasting error of the historical mean model for CDS returns is lower than that associated with the predictive regression of VRP at the 5% level (MSE-F: 3.33, ENC: 2.14).

The results from the predictive regressions for excess CDS returns indicate that VRP can be used to construct a stock market timing strategy for CDS returns. The results for the CDS strategy are displayed in Panel C of Table 13. The market timing strategy conditioning on VRP produces a significantly higher Sharpe ratio than the passive strategy (0.84 versus 0.61) and a CER gain of 4.67% per year.

Overall, there is a tendency that VRP predict negatively on the excess returns of safe assets such as long-term default-free bonds and VRP predict positively on the excess returns of assets with high credit risk. The strong in-sample predictive relationships for cases of long-term treasury bonds, long-term

Aaa-rated corporate bonds, and CDS indices associated with speculative grades still holds for out-of-sample.

5.2. Currency

Now we move to the currency market. As shown by Aloosh (2012), the global variance risk premium, which is constructed by the end-of-last month market capitalization weighted average of the VRP of individual countries, predicts the excess foreign exchange return both in-sample and out-of-sample. Londono and Zhou (2014) conduct a comprehensive study with 22 countries of foreign exchange data and report that variance risk premium in stock market positively predicts the excess foreign exchange return in-sample. They also provide a theoretical framework with a consumption-based international asset pricing model for explaining their findings.

We reexamine the information contained in VRP for the currency market and its economic significance with longer time-series of currency returns that including the 1990s, which are excluded in the previous studies (Aloosh, 2012; Londono and Zhou, 2014) and larger set of countries than those associated with Aloosh (2012). We set the VRP extracted only from the U.S. market rather than using one constructed from the weighted average of VRPs extracted from markets in various countries. Given that the weighted average version of VRP and the VRP of the U.S. market move very closely[11], the results are qualitatively similar if we analyze with the weighted average version of VRPs in place of the VRP extracted from the U.S. market.[12]

We consider one-month returns of zero-cost investments constructed by taking long one-month forward contracts of foreign currencies from the perspective of a U.S. investor (Barroso and Santa Clara,

---

[11] The correlation coefficient between the weighted average version of VRP and the VRP of the U.S. market is 0.93.

[12] Table 3 of Aloosh (2012) indicates that the results with the weighted average version of VRP and the VRP of the U.S. market are qualitatively similar.

2012; Maio, 2014) as follows:

$$R_{F,t+1} = \frac{S_{t+1}}{F_{t.t+1}} - 1$$

$R_{F,t+1}$ can be regarded as currency returns. $F_{t.t+1}$ is the forward exchange rate agreed upon at time t for a transaction at the next period $t+1$ (price of one foreign currency unit in Dollars), and S_{t+1} is the spot exchange rate at time $t+1$.

We select the countries for our analysis, based on the availability of one-month forward exchange rate and spot exchange rate data from Datastream: Japan (JPY), the Great Britain (GBP), the Euro Area (EUR), Switzerland (CHF), Canada (CAD), Australia (AUD), Hong Kong (HKD), Sweden (SEK), New Zealand (NZD), Singapore (SGD), South Africa (ZAR), Denmark (Denmark). The sample period is from January 1990 to December 2013.

Panel A of table 14 reports that there exists significantly positive relationship between VRP and future currency returns in-sample for 9 out of the 12 countries. The three exceptional cases (Japan, Singapore, Hong Kong) are from Asian countries. The results are quite consistent with the results of Londono and Zhou (2014), which show strong in sample return predictability of VRP in exchange rate returns for the countries we analyzed. One notable exception is the case of Japan. Londono and Zhou (2014) report significant negative relationship between future returns of JPY and VRP whereas there is no significant relationship in our analysis, as indicated by low value of the *t*-statistics of the slope efficient (0.06).

Panel B of table 14 shows that reported strong in-sample return predictability of VRP preserves for out-of-sample for all 9 cases, as indicated by the value of R^2_{OS} above 1% and statistically significant values of the MSE-F statistics at the 5% level and ENC statistics at the 10% level (Statistically significant at the 5% level for 6 out of the 9 cases)

Figure 5 plots the difference between the cumulative sum of squared prediction errors of a historical mean model and those of the model conditioning on VRP for currency markets. The key difference between the results for the stock index and the results for currency markets is that the models conditioning on VRP in currency markets underperforms the historical mean model out-of-sample before the Global Financial Crisis. In the case of the stock index, the models conditioning on VRP start to outperform the historical mean model during the recession period in early 2000s. On the other hand, the statistical outperformance of the models conditioning on VRP over the no-predictability benchmark for currency markets is mainly driven by good performance during the Global Financial Crisis.

Overall, VRP positively predicts the currency returns and the strong in-sample predictive relationship still holds for out-of-sample. The statistical forecasting power of VRP for currency markets mainly comes from good performance during the Global Financial Crisis.

5.3. Commodity Index

Finally, we investigate the information contained in VRP for excess returns of commodity indices. We mainly use the return on the S&P GSCI index as a proxy for commodity returns at the aggregate level (Maio, 2014). The index currently comprises 24 commodities from all commodity sectors: energy products, industrial metals, agricultural products, livestock products and precious metals. We also select commodity indices associated with above six commodity sectors.

Panel A of Table 15 shows the results from in-sample predictive regressions for excess returns of the commodity index. The results show that VRP predicts excess returns of commodity only for the Energy sector (*t*-statistics: 1.99), but there is no return predictability of VRP for other commodity. One interesting result is that the slope coefficient estimate of VRP for or the Precious Metal sector is positive. If we expect Silver and Gold, which consist of Precious Metal, as safe assets and behave like default-free bonds, this positive slope coefficient is quite unusual and there might be specific risk factors affecting the predictive relation between Precious Metal and VRP.

Panel B of Table 15 indicates that the strong in-sample return predictability associated with Energy sector does not hold for out-of-sample, as indicated by negative values of R^2_{OS}, ENC and MSE-F statistics.

Overall, there is no significant predictive relationship between VRP and the commodity indices.

6. Conclusion

We mainly examine the out-of-sample forecasting power of VRP for excess equity returns and its economic significance. We find strong evidence that VRP predicts excess returns of U.S stock index out-of-sample and that it is possible to construct a profitable market timing strategy based on the predictive power of VRP for excess equity returns. Our results show that the market timing strategy based on VRP produces a annualized Sharpe ratio of 0.32 (versus 0.18 for the passive strategy) and a certainty equivalent return (CER) gain of 1.89% per year. The market timing strategy based on VRP outperforms the strategies associated with alternate predictors. We verify that the results are not affected by the issues of parameter uncertainty, the sensitivity of forecasting schemes, and market friction such as transaction costs or borrowing costs. The out-of-sample predictability of VRP in an international context is also examined. We provide a profitable global market timing strategy in international equity markets by using country-specific VRPs.

We extensively examine the forecasting power of VRP for other asset class such as equity portfolios, bonds, currencies and commodity indices. We find strong in-sample evidence that VRP positively predicts excess returns of 6 representative equity portfolios and currency returns for 9 out of the 12 countries in our analysis. For all cases showing the in-sample predictive relationships, we find strong out-of-sample forecasting power of VRP and those are economically significant. For bond markets, we find that the negative relationship between VRP and future excess returns of Treasury bonds and Aaa-rated corporate bonds in-sample holds for out-of-sample weakly. However, its economic significance does not exist. There is no significant predictive relationship between VRP and the excess returns of the commodity indices.

References

Aloosh, A. 2012. Global variance risk premium and forex returns., Working Paper, BI Norwegian Business School.

Anderson, R., Bianchi, S., and Goldberg, L.2012. Will my risk parity strategy outperform? *Financial Analysts Journal* 58:75-93.

Bansal, R., and Shaliastovich, I. 2012. A long-run risks explanation of predictability puzzles in bond and currency markets. *Review of Financial Studies* 26:1-33.

Bansal, R., and Yaron, A. 2004. Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance* 59:1481-1509.

Barroso, P. and Santa-Clara, P. 2012. Beyond the carry trade: Optimal currency portfolios, Working Paper, Nova School of Business and Economics.

Bollerslev, T., Marrone, J., Xu, L., Zhou, H., 2014. Stock Return Predictability and Variance Risk Premia: Statistical Inference and International Evidence. *Journal of Financial and Quantitative Analysis*, forthcoming.

Bollerslev, T., Tauchen, G., and Zhou, H. 2009. Expected stock returns and variance risk premia. *Review of Financial Studies* 22:4463 - 4492.

Breen, W., Glosten, L., and Jagannathan, R. 1989. Economic significance of predictable variations in stock index returns. *Journal of Finance* 44:1177-1189.

Bystrom, H. 2008. Credit default swaps and equity prices: The iTraxx CDS index market. N. Wagner (Ed.), Credit Risk — Models, Derivatives, and Management 69–84.

Campbell, J. 1991. A variance decomposition for stock returns. *Economic Journal* 101:157-179.

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

Campbell, J. and Thompson, S. 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21:1509-1531.

Clark, T. and McCracken, M. 2001. Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics* 105:85-110.

Connor, G. 1997. Sensible return forecasting for portfolio management. *Financial Analysts Journal* 53:44-51.

Cooper, I. and R. Priestley. 2009. Time-varying risk premiums and the output gap. *Review of Financial Studies* 22:2801-2833.

Corsi, F. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7:174-196.

Dickey, D.A. and W.A. Fuller. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74:427-431.

Drechsler, I., and Yaron, A. 2011. What's vol got to do with it. *Review of Financial Studies* 24:1-45.

Fama, E.F. 1981. Stock returns, real activity, inflation, and money. *American Economic Review* 71: 545–565.

Fama, E. F., K. R. French. 1988. Dividend yields and expected stock returns. *Journal of Financial Economics* 22:3-25.

Fama, E. F., K. R. French. 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25:23-49.

Fama, E. F., G. W. Schwert. 1977. Asset returns and inflation. *Journal of Financial Economics* 5: 115-146.

Ferreira, M. and Santa-Clara, P. 2011. Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics* 100:514-537.

Giammarino, F., P. Barrieu. 2009. A semiparametric model for the systematic factors of portfolio credit risk premia. *Journal of Empirical Finance* 16:655–670.

Glosten, L., R. Jagannathan, D. Runkle. 1993. On the relation between expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48:1779-1801.

Goetzmann, W., J. Ingersoll, M. Spiegel, and I. Welch. 2007. Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies* 20:1503–46.

Goyal, A. and Santa-Clara, P. 2003. Idiosyncratic risk matters, *Journal of Finance* 58:975-1006.

Goyal, A. and Welch, I. 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21:1455-1508.

Guo, H. 2006. On the out-of-sample predictability of stock returns. *Journal of Business* 79:645-670.

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

Jagannathan, R., Z. Wang. 1996. The conditional CAPM and the cross-section of expected returns. *Journal of Finance* 51:3-54.

Jondeau, E., and M. Rockinger. 2008. The economic value of distributional timing. Mimeo, Swiss Finance Institute.

Kandel, S., R. F. Stambaugh. 1990. Expectations and volatility of consumption and asset returns. *Review of Financial Studies* 3:207–232.

Karstanje, D. Sojli, W. W. Tham., M. van der Wel. 2013. Economic valuation of liquidity timing. *Journal of Banking and Finance Studies* 37:5073–5087.

Keating, C., and W. F. Shadwick. 2002. A universal performance measure. *Journal of Performance Measurement* 6:59–84.

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

Londono, J. M., and H. Zhou. 2014. Variance Risk Premiums and the Forward Premium Puzzle, Working Paper, Federal Reserve Board

McCracken, M. 2007. Asymptotics for out of sample tests of Granger causality. *Journal of Econometrics* 140:719-752.

Maio, P. 2014. "Don't Fight the Fed!". *Review of Finance* 18:623-679.

Maio, P. and Santa-Clara, P. 2012. Multifactor models and their consistency with the ICAPM. *Journal of Financial Economics* 106:586-613.

Merton, R.C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* 29:449-470.

Møller, S. V., and J. Rangvid. 2013. End-of-the-year economic growth and time-varying expected returns, Working Paper.

Mueller, P., A. Vedolin, and H. Zhou. 2011. Short-run bond risk premia. Federal Reserve Board, Working Paper.

Thornton, D. and G.Valente. 2012. Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. *Review of Financial Studies* 25:3141–3168.

Rangvid, J. 2006. Output and expected returns, *Journal of Financial Economics* 81:595-624.

Wang, H., H. Zhou, and Y. Zhou. 2013. Credit default swap spreads and variance risk premia. *Journal of Banking and Finance* 37: 3733-3746.

Wu, J., L. Zhang, and X. Zhang. 2010. The q-Theory approach to understanding the accrual anomaly. *Journal of Accounting Research* 48: 177–223.

Zakamouline, V., and S. Koekebakker.. 2009. Portfolio performance evaluation with generalized Sharpe ratios: Beyond the mean and variance. *Journal of Banking and Finance Studies* 33: 1242–1254.

Zhou, H. 2010. Variance risk premia, asset predictability puzzles, and macroeconomic uncertainty, Working Paper, Federal Reserve Board

Table 1 Summary Statistics for Monthly Stock Excess Returns and Return Predictors

This table reports the mean, standard deviation, skewness, kurtosis, and the first-order autocorrelation of the predictors and the stock index return. We also compute augmented Dickey–Fuller (Dickey and Fuller, 1979) statistics for each predictor to explicitly identify non-unit root process. The predictors are Variance Risk Premium (*VRP*), Dividend Price Ratio (**d/p**), Dividend Yield (**d/y**), Earnings Price Ratio (**e/p**), Dividend Payout Ratio (**d/e**), Stock Variance (**svar**), Book to Market Ratio (**b/m**), Net Equity Expansion (**ntis**), Treasury Bills (**tbl**), Long Term Yield (**lty**), Long Term Rate of Return (**ltr**), Term Spread (**tms**), Default Yield Spread (**dfy**), Default Return Spread (**dfr**), and Inflation (**infl**). The sample period is from January 1990 to December 2013.

	Mean (%)	Std (%)	Skew	Kurt	AR (1)	Unit root test (p-value)
Return	0.49	4.32	-0.79	4.62	0.07	0.00***
	Predictors					
VRP	18.13	20.01	-2.45	38.93	0.26	0.00***
d/p	-3.91	0.30	0.16	2.32	0.98	0.35
d/y	-3.91	0.30	0.12	2.33	0.99	0.44
e/p	-3.12	0.38	-2.03	8.76	0.98	0.00***
d/e	-0.80	0.44	2.54	11.52	0.98	0.00***
svar	0.00	0.00	6.83	65.88	0.71	0.00***
b/m	0.29	0.09	0.19	2.53	0.97	0.20
ntis	0.01	0.02	-0.82	4.09	0.98	0.16
tbl	0.03	0.02	-0.05	1.81	0.99	0.23
lty	0.06	0.02	0.05	2.48	0.98	0.42
ltr	0.01	0.03	-0.01	5.52	0.03	0.00***
tms	0.05	0.01	0.07	2.49	0.97	0.16
dfy	0.01	0.00	3.13	15.89	0.96	0.00***
dfr	0.00	0.02	-0.44	11.31	0.03	0.00***
infl	0.00	0.00	-1.39	15.38	0.45	0.00***

Table 2 In-sample Univariate Predictive Regressions for the Market Excess Returns

This table summarizes the results for multiple-horizon univariate predictive regressions for the market excess returns at horizons of 1, 3, 6, 12, 24, 36, and 48-months ahead. For each regression, β denotes the slope estimates, and we report Newey-West t-statistic (in parentheses). The bold t-statistics figures signify statistical significance at the 5% levels. $\bar{R}^2$ (%) denotes the adjusted coefficient of determination.

Panel A : Short term												
	1 month			3 months			6 months			12 months		
	β	t	$\bar{R}^2$	β	t	$\bar{R}^2$	β	t	$\bar{R}^2$	β	t	$\bar{R}^2$
Non-unit-root												
VRP	4.90	**5.11**	5.95	12.96	**7.09**	11.00	16.29	**4.30**	7.54	14.05	**2.08**	2.64
e/p	0.01	0.64	1.26	0.02	0.48	0.44	0.02	0.43	0.25	0.06	0.82	1.99
d/e	0.00	0.08	0.78	0.01	0.42	0.13	0.03	0.83	0.65	0.05	1.31	1.55
svar	-1.31	**-2.10**	3.05	-2.58	-1.63	2.57	-0.87	-0.40	-0.22	1.20	0.52	0.02
ltr	0.03	0.46	0.83	-0.10	-0.51	-0.01	0.15	0.73	-0.21	0.16	0.57	-0.02
dfy	-0.62	-0.53	1.14	-1.04	-0.33	0.17	0.45	0.10	-0.34	3.32	0.63	0.56
dfr	0.23	0.88	1.56	0.38	0.84	0.49	0.55	0.77	0.24	0.83	1.24	0.53
infl	0.28	0.32	0.83	1.55	0.55	0.31	-2.83	-0.91	0.31	-7.64	**-1.96**	2.18
Near-unit-root												
d/p	0.01	1.52	1.77	0.05	1.87	2.93	0.10	**2.42**	5.87	0.21	**3.68**	14.06
d/y	0.02	1.80	1.98	0.05	**2.07**	3.29	0.10	**2.50**	6.15	0.22	**3.80**	14.86
b/m	0.03	1.19	1.21	0.12	1.74	1.86	0.31	**2.38**	5.24	0.65	**3.14**	11.80
ntis	0.19	0.95	1.64	0.69	1.19	3.20	1.50	1.48	6.74	2.38	1.58	8.23
tbl	-0.02	-0.16	0.79	-0.08	-0.21	-0.09	-0.29	-0.49	-0.05	-0.87	-1.02	1.18
lty	-0.05	-0.31	0.81	-0.12	-0.31	-0.07	-0.30	-0.43	-0.18	-0.14	-0.12	-0.08
tms	-0.01	-0.07	0.78	0.02	0.04	-0.14	0.37	0.38	-0.18	2.22	1.49	2.89

	Panel B : Long term											
	24 months			36 months			48 months			60 months		
	β	t	$\bar{R}^2$	β	t	$\bar{R}^2$	β	t	$\bar{R}^2$	β	t	$\bar{R}^2$
	Non-unit-root											
VRP	17.32	1.63	1.12	10.28	0.83	-0.31	0.54	0.03	-0.70	-7.83	-0.44	-0.85
e/p	0.07	0.60	0.37	0.11	0.81	0.98	0.05	0.33	-0.41	-0.05	-0.31	-0.77
d/e	0.15	**2.72**	6.18	0.21	**3.42**	8.36	0.36	**4.81**	18.79	0.54	**5.29**	38.00
svar	3.03	0.91	-0.34	2.79	0.71	-0.55	2.85	0.53	-0.54	8.10	1.36	0.13
ltr	0.32	0.70	-0.56	0.67	1.01	-0.38	0.29	0.36	-0.65	0.35	0.39	-0.95
dfy	10.06	1.68	1.97	10.27	1.61	1.10	16.94	**2.39**	3.27	26.22	**3.55**	7.58
dfr	0.86	1.00	-0.38	1.12	1.05	-0.43	1.74	1.31	-0.17	1.63	0.95	-0.62
infl	-7.95	-1.61	0.38	-5.27	-0.89	-0.45	-7.44	-1.28	-0.25	-5.64	-0.68	-0.78
	Near-unit-root											
d/p	0.44	**4.71**	26.15	0.63	**6.29**	36.04	0.81	**7.71**	48.02	1.00	**9.97**	65.99
d/y	0.45	**4.74**	26.65	0.63	**6.28**	36.09	0.81	**7.60**	47.69	0.98	**9.68**	64.25
b/m	1.09	**3.23**	13.76	1.37	**3.71**	14.37	1.72	**4.59**	17.85	2.30	**6.30**	27.30
ntis	2.85	1.69	4.48	3.58	**2.14**	4.49	3.91	1.87	4.26	2.91	1.10	1.46
tbl	-3.32	**-2.19**	6.98	-4.49	**-2.61**	7.90	-4.46	**-2.23**	5.32	-2.99	-1.05	1.11
lty	-0.37	-0.20	-0.63	1.95	0.77	0.04	5.55	1.67	3.62	9.45	**2.41**	9.18
tms	8.24	**3.12**	17.19	12.97	**5.96**	29.42	15.15	**6.88**	31.55	14.79	**4.53**	25.43

Table 3 Out-of-sample Evaluation Statistics for the One-month Ahead Predictability Associated with the Market Excess Returns

This table summarizes the performance of the out-of-sample test for the one-month ahead predictability associated with the excess stock market returns. R_{OS}^2 denotes the out-of-sample coefficient of determination (in %). $\mathrm{MSE-F}$ (McCracken's (2007) F-statistic) and ENC statistics developed by Clark and McCracken (2001) test for null hypothesis that using the predictors does not significantly improve on a forecast based solely on the historical average return. The numbers in bold signify that the null hypothesis associated with $\mathrm{MSE-F}$ or ENC is rejected at the 5% levels. R_{OS-CT}^2 (%) represents the out-of-sample coefficient of determination that restricts the non-negative fitted expected excess return, as proposed by Campbell and Thompson (2008). The total sample is from Jan. 1990 to Dec. 2013, and the estimation period for the first regression is from Jan. 1990 to Dec. 1999

	In Sample (1 month)		Out-of-Sample			
	β_1	t_{β_1}	R_{OS}^2	$\mathrm{MSE-F}$	ENC	R_{OS-CT}^2
VRP	4.90	5.11	5.80	**13.3**	**10.18**	4.00
d/p	0.01	1.52	0.73	**1.24**	0.88	0.73
d/y	0.02	1.80	0.98	**1.66**	1.09	0.98
e/p	0.01	0.64	-1.83	-3.02	**2.55**	2.66
d/e	0.00	0.08	-3.50	-5.67	-1.53	-0.87
svar	-1.31	-2.10	-3.60	-5.84	**2.57**	-3.29
b/m	0.03	1.19	0.03	0.06	0.13	0.03
ntis	0.19	0.95	-0.39	-0.65	0.65	-0.50
tbl	-0.02	-0.16	-1.45	-2.40	-0.64	-0.17
lty	-0.05	-0.31	-1.45	-2.40	-0.75	-0.35
ltr	0.03	0.46	-0.96	-1.60	-0.74	-0.87
tms	-0.01	-0.07	-0.97	-1.62	-0.75	-0.94
dfy	-0.62	-0.53	-2.85	-4.66	**2.95**	0.16
dfr	0.23	0.88	-3.77	-6.11	-2.26	-2.75
infl	0.28	0.32	-1.71	-2.82	-0.92	-1.67

Table 4 Evaluated Performance Measures for the Market-Timing Strategies Based on the Out-of-sample Forecasting Power for the Market Excess Returns

This table summarizes the performance measures for the market timing strategies based on the out-of-sample forecasting power for the excess stock market return. “Buy-hold" denotes the passive strategy associated with holding the market portfolio. “Mean" denotes the average return (in %); "Std” is the standard deviation (in %); "Skew” is the skewness; "Kurt” is the kurtosis; FF alpha is the Fama-French 3 factors alpha; FFC alpha is the Fama-French-Carhart 4 factors alpha. We calculate the p-values associated with the alpha by a bootstrap method used by Anderson et al. (2012). “SR" represents the monthly Sharpe ratio; “CER" represents the extra utility generated from the market timing strategy instead of the buy-and-hold strategy; “GISW" is a manipulation-proof measure of performance developed by Goetzmann et al. (2007); “ASSR” is a variant of Sharpe ratio adjusted for skewness under a CRRA utility function; “Sortino” is a reward-to-downside risk ratio; “Omega” is a simple generalization of the gain–loss ratio. The total sample is from Jan. 1990 to Dec. 2013, and the estimation period for the first regression is from Jan. 1990 to Dec. 1999.

	Moments					
	Mean (%)	Std (%)	Skew	Kurt	FF alpha (p-value)	FFC alpha (p-value)
Buy Hold	0.40	4.52	-0.54	3.80		
VRP	0.61	4.92	0.20	5.24	**0.004** (0.01)	**0.004** (0.01)
d/p	0.19	4.56	-1.39	9.85	0.002 (0.16)	0.002 (0.14)
d/y	0.27	4.43	-0.97	7.76	0.003 (0.10)	0.003 (0.09)
e/p	0.68	3.82	-0.06	3.24	**0.007** (0.00)	**0.007** (0.00)
d/e	0.19	4.21	-0.43	4.14	0.002 (0.10)	0.002 (0.11)
svar	0.03	4.99	-1.23	7.22	0.001 (0.24)	0.001 (0.23)
b/m	0.18	4.61	-0.56	4.13	0.002 (0.13)	0.002 (0.14)
ntis	0.04	4.11	-0.50	4.82	0.001 (0.25)	0.001 (0.24)
tbl	-0.11	3.63	-0.34	6.41	-0.001 (0.45)	-0.001 (0.45)
lty	-0.19	3.26	-0.38	7.68	-0.002 (0.60)	-0.002 (0.59)
ltr	-0.07	4.63	-0.53	4.09	0.000 (0.34)	0.000 (0.34)
tms	0.00	4.16	-0.60	4.80	0.000 (0.26)	0.000 (0.26)
dfy	0.15	3.85	-0.35	5.78	0.001 (0.16)	0.001 (0.15)
dfr	0.10	4.76	-0.63	4.31	0.002 (0.18)	0.001 (0.20)
infl	0.03	4.59	-1.38	8.01	0.001 (0.25)	0.001 (0.25)
	Performance Measure					
	SR	CER (%)	GISW (%)	ASSR	Sortino	Omega
Buy Hold	0.18			0.18	0.04	1.05
VRP	0.32	1.89	2.05	0.32	0.09	1.14
d/p	0.02	-2.53	-2.83	0.02	0.00	1.00
d/y	0.08	-1.41	-1.54	0.08	0.01	1.02
e/p	0.47	4.47	4.58	0.47	0.13	1.16
d/e	0.02	-2.04	-1.99	0.02	0.00	1.00

svar	-0.09	-5.22	-5.57	-0.09	-0.03	0.96
b/m	0.02	-2.70	-2.72	0.02	0.00	1.00
ntis	-0.10	-3.61	-3.59	-0.10	-0.03	0.95
tbl	-0.26	-4.80	-4.74	-0.26	-0.08	0.89
lty	-0.38	-5.31	-5.24	-0.38	-0.11	0.85
ltr	-0.17	-5.79	-5.82	-0.17	-0.05	0.94
tms	-0.14	-4.22	-4.21	-0.14	-0.04	0.95
dfy	-0.01	-2.01	-1.94	-0.01	-0.01	0.98
dfr	-0.05	-4.01	-4.07	-0.05	-0.02	0.97
infl	-0.10	-4.56	-4.85	-0.10	-0.03	0.95

Table 5 Out-of-sample Assessment of Stock Return Predictability of VRP and Evaluated Performance Measures for the Market-Timing Strategies Conditioning on VRP: Under Various Forecasting Schemes as a Robustness Check

This table summarizes the out-of-sample performance (Panel A), and the performance measures (Panel B) for the market timing strategies based on the out-of-sample forecasting power of *Variance Risk Premium* for the excess stock market returns. The results are based on both an expanding window with initial length of 120, 180 months, and a rolling window with initial length of 120, 180 months. The numbers in bold signify that the null hypothesis associated with $\mathrm{MSE}-\mathrm{F}$ or ENC is rejected at the 5% levels.

	Expanding		Rolling	
	120 m	180 m	120 m	180 m
Panel A: Out-of-Sample Analysis				
R^2_{OS}	6.13	9.86	5.99	9.87
$\mathrm{MSE}-\mathrm{F}$	**10.97**	**11.81**	**10.7**	**11.83**
ENC	**8.61**	**8.8**	**8.69**	**9.13**
R^2_{OS-CT}	3.96	5.72	2.77	5.78
Panel B: Performance of the Market Timing Strategy				
Mean(%)	0.61	1.14	0.47	0.94
Std(%)	4.92	4.5	4.35	4.28
Skew	0.2	0.24	0.17	0.39
Kurt	5.24	6.77	6.69	7.35
SR	0.32	0.78	0.24	0.66
CER (%)	1.89	5.51	1.10	3.53
GISW (%)	2.05	5.69	1.23	3.73

Table 6 Evaluated Performance Measures for the Market-Timing Strategies Based on the Out-of-sample Forecasting Power for the Market Excess Returns: Constructed under Parameter Uncertainty as a Robustness Check

This table summarizes the performance measures for the market timing strategies conditioning on the out-of-sample forecasting power for the excess stock market returns. The market timing strategies are constructed under *parameter uncertainty* by implementing the procedure developed by Connor (1997). “Mean" denotes the average return (in %); "Std” is the standard deviation (in %); "Skew” is the skewness; "Kurt” is the kurtosis; “SR" represents the monthly Sharpe ratio; “CER" represents the extra utility generated from the market timing strategy instead of the buy-and-hold strategy; “GISW" is a manipulation-proof measure of performance developed by Goetzmann et al. (2007).The predictors are Variance Risk Premium (*VRP*), Dividend Price Ratio (**d/p**), Dividend Yield (**d/y**), Earnings Price Ratio (**e/p**), Dividend Payout Ratio (**d/e**), Stock Variance (**svar**), Book to Market Ratio (**b/m**), Net Equity Expansion (**ntis**), Treasury Bills (**tbl**), Long Term Yield (**lty**), Long Term Rate of Return (**ltr**), Term Spread (**tms**), Default Yield Spread (**dfy**), Default Return Spread (**dfr**), and Inflation (**infl**).The total sample is from Jan. 1990 to Dec. 2013, and the estimation period for the first regression is from Jan. 1990 to Dec. 1999.

	Moments				Performance Measures		
	Mean (%)	Std (%)	Skew	Kurt	SR	CER (%)	GISW (%)
VRP	0.59	4.42	0.47	6.32	0.34	2.49	2.68
d/p	0.63	6.74	-0.58	3.90	0.24	-1.66	-2.04
d/y	0.68	6.65	-0.65	4.02	0.27	-0.93	-1.34
e/p	0.20	6.14	-0.65	4.32	0.02	-5.50	-5.83
d/e	0.22	3.93	-0.41	5.62	0.05	-1.27	-1.21
svar	0.24	4.81	-1.24	7.62	0.06	-2.35	-2.65
b/m	-0.01	2.49	-0.66	18.68	-0.24	-2.32	-2.24
ntis	0.00	3.55	-0.44	5.82	-0.16	-3.34	-3.27
tbl	-0.09	3.47	-0.44	7.57	-0.25	-4.35	-4.29
lty	-0.03	3.11	-0.19	8.76	-0.21	-3.18	-3.09
ltr	0.06	4.35	-0.59	4.53	-0.08	-3.75	-3.76
tms	0.25	4.71	-0.59	4.05	0.06	-2.14	-2.17
dfy	0.41	5.15	-1.11	6.60	0.17	-0.94	-1.28
dfr	0.03	4.51	-0.63	4.17	-0.10	-4.36	-4.39
infl	-0.04	4.89	-1.02	6.19	-0.14	-5.85	-6.08

Table 7 In-sample and Out-of-sample Predictive Regressions of GVRP on the Excess Returns of Equity index for Non-U.S countries

This table shows the in-sample, the out-of-sample performance and performance measures for market-timing strategies based on the out-of-sample forecasting power of *Global Variance Risk Premium* for the excess returns of equity index for 7 non-U.S countries. Panel A summarizes the in-sample performance of one-month ahead predictive regressions of *Global Variance Risk Premium*. Panel B summarizes performance of out-of-sample test for the one-month ahead predictability of *Global Variance Risk Premium*. Panel C summarizes performance measures for market-timing strategies based on the out-of-sample forecasting power for the excess returns of equity index for 7 non-U.S countries. The numbers at bold mean statistical significance at the 5% levels. The sample period is from January 2000 to December 2014 and the estimation period for the first regression is from January 2000 to December 2009.

		UK	*Belgium*	*Japan*	*Netherlands*	*Germany*	*France*	*Swiss*
Panel A : In-Sample Analysis								
β_1		4.04	4.17	-0.34	2.86	3.17	3.46	2.78
t_{β_1}		**3.33**	**2.35**	-0.28	1.58	**1.78**	**1.99**	**2.25**
$\bar{R}^2$ (%)		6.84	8.09	-0.73	2.48	1.13	2.53	6.03
Panel B : Out-of-Sample Analysis								
R^2_{OS} (%)		7.10	5.51	-1.44	3.21	2.66	3.78	4.18
$\mathrm{MSE-F}$		**4.59**	**3.50**	-0.85	**1.99**	**1.64**	**2.36**	**2.62**
ENC		**2.72**	**2.06**	-0.39	**1.10**	0.92	**1.30**	**1.48**
R^2_{OS-CT} (%)		5.27	3.80	-0.15	1.76	1.87	2.05	2.94
Panel C : Performance of Market Timing Strategy								
Active	Mean(%)	0.81	0.69	-0.27	0.22	0.77	0.39	0.97
	Std(%)	3.57	3.24	1.75	1.45	1.91	2.28	3.58
	SR	0.79	0.73	-0.53	0.51	0.47	0.59	0.88
Passive	Mean(%)	0.68	0.65	0.50	0.56	0.77	0.39	0.97
	Std(%)	4.89	5.81	4.00	6.01	6.72	6.66	4.61
	SR	0.48	0.39	0.43	0.32	0.40	0.20	0.72
CER (%)		3.65	4.58	-6.84	2.02	1.36	7.06	0.90

Table 8 Evaluated Performance Measures for Market-timing Strategies Based on the Out-of-sample Forecasting Power of GVRP: International Market Data

This table shows the evaluated performance measures for market-timing strategies based on the out-of-sample forecasting power of GVRP and the *historical mean model* (HM) for the excess returns of equity index for 7 countries (excluding Japan) using *the asset allocation framework*. Panel A summarizes the results for the case that target expected return set to be 0.001. Panel B summarizes the results for the case that target expected return set to be 0.002. Panel C summarizes the results for the case that target expected return set to be 0.003. Panel D summarizes the results for the case that target expected return set to be 0.004. Panel E summarizes the results for the case that target expected return set to be 0.005. "We set the target expected return set based on the evaluated values of average market excess returns for 7 countries. "Mean" denotes the average return (in %); "Std" is the standard deviation (in %); "Skew" is the skewness; "Kurt" is the kurtosis; "SR" represents the monthly Sharpe ratio; "CER" represent extra utility generated from the market timing strategy instead of the strategy based on the historical mean model and "GISW" is a manipulation-proof measure of performance developed by Goetzmann et al. (2007). The sample period is from January 2000 to December 2014 and the estimation period for the first regression is from January 2000 to December 2009.

	Moments				Performance Measure		
	Mean (%)	Std (%)	Skew	Kurt	SR	CER (%)	GISW (%)
			Panel A : μ_c = 0.001				
HM	0.06	0.51	0.33	6.35	0.42		
VRP	0.10	0.38	-0.20	5.81	0.91	0.49	0.49
			Panel B : μ_c = 0.002				
HM	0.13	1.03	0.33	6.45	0.42		
VRP	0.20	0.77	-0.20	5.80	0.91	1.03	1.02
			Panel C : μ_c = 0.003				
HM	0.19	1.54	0.33	6.45	0.42		
VRP	0.31	1.15	-0.20	5.80	0.91	1.60	1.58
			Panel D : μ_c = 0.004				
HM	0.27	2.09	0.21	6.46	0.45		
VRP	0.38	1.59	-0.48	6.56	0.82	1.64	1.61
			Panel E : μ_c = 0.005				
HM	0.40	2.76	0.07	6.63	0.50		
VRP	0.42	2.04	-0.80	7.66	0.71	0.87	0.83

Table 9 In-sample and Out-of-sample Predictive Regressions of VRP on the Excess Returns of Individual Portfolios

This table shows the in-sample performance, the out-of-sample performance, and the performance measures for the market timing strategies based on the out-of-sample forecasting power of *Variance Risk Premium* for the excess returns of the *Small, Big, Growth, Value, Winner, and Loser* portfolios. Panel A summarizes the in-sample performance of the one-month ahead predictive regressions of *Variance Risk Premium*. Panel B summarizes the performance of the out-of-sample test for the one-month ahead predictability of *Variance Risk Premium*. Panel C summarizes the performance measures for the market timing strategies based on the out-of-sample forecasting power for the excess returns of *Small*, *Big*, *Growth*, *Value*, *Winner, and Loser* portfolios. The numbers in bold signify statistical significance at the 5% levels. The sample period is from Jan. 1990 to Dec. 2013, and the estimation period for the first regression is from Jan. 1990 to Dec. 1999.

		Small	*Big*	*Growth*	*Value*	*Loser*	*Winner*
Panel A: In-Sample Analysis							
β_1		5.26	4.98	6.79	5.76	11.52	6.48
t_{β_1}		**2.48**	**5.05**	**7.36**	**2.82**	**3.87**	**3.97**
$\bar{R}^2$ (%)		2.97	6.30	8.40	3.94	5.98	5.33
Panel B: Out-of-Sample Analysis							
R^2_{OS} (%)		1.24	5.93	9.15	2.63	5.95	5.39
$\mathrm{MSE-F}$		**2.11**	**10.59**	**16.92**	**4.54**	**10.64**	**9.58**
ENC		**1.63**	**9.61**	**15.97**	**2.85**	**7.37**	**8.22**
R^2_{OS-CT} (%)		0.63	4.52	6.44	2.24	3.46	3.82
Panel C: Performance of the Market Timing Strategy							
Active	Mean(%)	1.02	0.48	1.36	0.93	0.72	1.22
	Std(%)	6.52	4.99	7.88	6.72	5.03	8.61
	SR	0.45	0.22	0.53	0.39	0.39	0.42
Passive	Mean(%)	1.14	0.30	0.32	0.66	0.31	0.98
	Std(%)	6.81	4.44	10.95	6.47	4.95	7.16
	SR	0.49	0.11	0.05	0.27	0.10	0.39
CER (%)		-0.70	1.20	22.87	2.55	4.88	-1.25

Table 10 In-sample and Out-of-sample Predictive Regressions of VRP on SMB, HML, and WML

This table shows the in-sample performance, the out-of-sample performance, and the performance measures for the market timing strategies based on the out-of-sample forecasting power of *Variance Risk Premium* for the zero-cost strategies based on size (SMB), book-to-market (HML), and momentum (WML). Panel A summarizes the in-sample performance of one-month ahead predictive regressions of *Variance Risk Premium*. Panel B summarizes the performance of the out-of-sample test for the one-month ahead predictability of *Variance Risk Premium*. The numbers in bold signify statistical significance at the 5% levels. The sample period is from Jan. 1990 to Dec. 2013, and the estimation period for the first regression is from Jan. 1990 to Dec. 1999.

	WML	*HML*	*SMB*
Panel A: In-Sample Analysis			
β_1	-5.04	-1.03	0.28
t_{β_1}	-1.65	-0.53	0.13
$\bar{R}^2$ (%)	1.15	-0.21	-0.33
Panel B: Out-of-Sample Analysis			
R^2_{OS} (%)	0.40	-4.92	-8.14
$\mathrm{MSE-F}$	0.67	-7.87	-12.65
ENC	0.64	0.44	-4.25
R^2_{OS-CT} (%)	0.46	-4.71	-6.22

Table 11 Evaluated Performance Measures for Market Timing Strategies Based on the Out-of-sample Forecasting Power of VRP: Individual Portfolio Level

This table shows the performance measures for the market timing strategies based on the out-of-sample forecasting power of *Variance Risk Premium* and the historical mean model (HM) for the excess returns of *Small, Big, Growth, Value, Winner,* and *Loser* portfolios using *the asset allocation framework*. Panel A summarizes the results for the case in which the target expected return was set to be 0.006. Panel B summarizes the results for the case in which target expected return was set to be 0.008. Panel C summarizes the results for the case in which the target expected return was set to be 0.01. “Mean" denotes the average return (in %); "Std” is the standard deviation (in %); "Skew” is the skewness; "Kurt” is the kurtosis; “SR" represents the monthly Sharpe ratio; “CER" represents extra utility generated from the market timing strategy instead of the strategy based on the historical mean model, and “GISW" is a manipulation-proof measure of performance developed by Goetzmann et al. (2007). The sample period is from Jan. 1990 to Dec. 2013, and the estimation period for the first regression is from Jan. 1990 to Dec. 1999.

	Moments				Performance Measure		
	Mean (%)	Std (%)	Skew	Kurt	SR	CER (%)	GISW (%)
Panel A: μ_c = 0.006							
HM	0.75	3.79	-0.30	4.10	0.60		
VRP	0.82	2.77	-0.22	5.60	0.73	2.04	2.14
Panel B: μ_c = 0.008							
HM	1.07	5.64	-0.31	3.87	0.56		
VRP	0.97	4.14	-0.12	5.36	0.68	1.48	1.57
Panel C: μ_c = 0.010							
HM	1.25	7.12	-0.35	3.75	0.53		
VRP	1.21	5.49	-0.09	5.11	0.66	3.25	3.50

Table 12 Statistical and Economical Significance of the Forecasting Power of VRP for the Bonds

This table shows the in-sample performance, the out-of-sample performance, and the performance measures for the market timing strategies based on the out-of-sample forecasting power of *Variance Risk Premium* for the excess returns of *bonds*. "T-bill 2" represents the returns of zero coupon bonds with a maturity of two months; "T-bill 4" represents the returns of zero coupon bonds with a maturity of four months; "T-bill 6" represents the returns of zero coupon bonds with a maturity of six months; "2y T-bond" represents returns of U.S BENCHMARK DS GOVT. INDEX for 2-year, provided by Datastream; "3y T-bond" represents returns of U.S BENCHMARK DS GOVT. INDEX for 3-year, provided by Datastream; "5y T-bond" represents returns of U.S BENCHMARK DS GOVT. INDEX for 5-year, provided by Datastream; "Treasury" represents the returns of Barclays U.S Treasury Long Index; "Aaa" represents the returns of Barclays U.S Treasury Aggregate Corporate Aaa Long Index; "Baa" represents the returns of Barclays U.S Treasury Aggregate Corporate Baa Long Index; "HY" represents the returns of Barclays U.S Treasury Corporate High Yield Index. Panel A summarizes the in-sample performance of the one-month ahead predictive regressions of *Variance Risk Premium*. Panel B summarizes the performance of the out-of-sample test for the one-month ahead predictability of *Variance Risk Premium*. Panel C summarizes the performance measures for the market timing strategies based on the out-of-sample forecasting power for the excess returns of *other assets*. The sample period for the bond returns is from Jan. 1990 to Dec. 2013, estimation period for the first regression is from Jan. 1990 to Dec. 1999. The numbers in bold signify statistical significance at the 5% levels.

	Long-term Bonds (Sorted by Default Risk)				Short-term Bonds (Sorted by Maturity)					
	Treasury	Aaa	Baa	HY	T-bill 2	T-bill 4	T-bill 6	2y T-bond	3y T-bond	5y T-bond
Panel A: In-Sample Analysis										
β_1	-2.76	-2.85	-0.85	2.23	0.02	-0.01	-0.02	-0.04	-0.07	-0.52
t_{β_1}	**-2.28**	**-2.24**	-1.17	1.56	1.58	-0.25	-0.39	-0.15	-0.23	-0.79
$\bar{R}^2$ (%)	3.36	3.44	-0.06	2.99	10.13	0.54	-0.2	0.43	0.6	0.88
Panel B: Out-of-Sample Analysis										
R^2_{OS} (%)	0.88	0.85	-0.84	-3.49	-1.81	-1.83	-3.15	-3.83	-2.69	-3.66
$\mathrm{MSE-F}$	**1.5**	**1.43**	-1.4	-5.67	-2.98	-3.02	-5.13	-6.19	-4.39	-5.93
ENC	1.59	1.52	-0.35	-1.31	**4.77**	-1.1	-1.91	-2.28	-1.85	-2.46
R^2_{OS-CT} (%)	1.17	0.55	-0.83	-3.08	2.15	-0.83	-1.31	-0.78	-1.37	-1.62

Panel C: Performance of the Market Timing Strategy											
Active	Mean(%)	0.88	0.66	0.74	0.75	0.17	0.19	0.22	0.35	0.45	0.49
	Std(%)	4.32	4.75	4.11	4.18	0.18	0.2	0.24	0.74	1.12	1.9
	SR	0.57	0.36	0.49	0.49	0.14	0.46	0.83	0.86	0.9	0.59
Passive	Mean(%)	0.63	0.52	0.7	0.68	0.17	0.18	0.2	0.3	0.37	0.45
	Std(%)	3.1	3.39	2.88	2.97	0.17	0.18	0.21	0.51	0.76	1.32
	SR	0.53	0.36	0.64	0.6	0.1	0.34	0.69	0.89	0.95	0.75
CER (%)		1.28	-0.23	-1.01	-0.64	0.03	0.09	0.2	0.55	0.87	0.11

Table 13 Statistical and Economical Significance of the Forecasting Power of VRP for Credit Markets

This table shows the in-sample performance (Panel A), the out-of-sample performance (Panel B), and the performance measures (Panel C) for the market timing strategies based on the out-of-sample forecasting power of key variables in credit markets for the total returns of CDX NA High Yield Index. The key variables in credit markets are Variance Risk Premium (*VRP*), Default Yield Spread (**dfy**), and Default Return Spread (**dfr**). The sample period is from Jan. 1997 to Dec. 2013, estimation period for the first regression is from Jan. 1997 to Dec. 2006. The numbers in bold signify statistical significance at the 5% levels..

		VRP	**dfy**	**dfr**
Panel A: In-Sample Analysis				
β_1		2.51	0.15	0.03
t_{β_1}		2.56	0.19	0.19
$\bar{R}^2$ (%)		3.97	-0.72	-0.76
Panel B: Out-of-Sample Analysis				
R^2_{OS} (%)		3.81	-9.23	-10.88
$\mathrm{MSE-F}$		**3.33**	-7.09	-8.24
ENC		**2.14**	-2.72	-3.52
R^2_{OS-CT} (%)		3.01	-3.27	-10.22
Panel C: Performance of the Market Timing Strategy				
Active	Mean(%)	1.16	0.56	0.66
	Std(%)	4.76	4.75	5.05
	SR	0.84	0.41	0.45
Passive	Mean(%)	0.61	0.61	0.61
	Std(%)	3.46	3.46	3.46
	SR	0.61	0.61	0.61
CER (%)		4.67	-2.47	-1.77

Table 14 Statistical and Economical Significance of the Forecasting Power of VRP for the Currency Markets

This table shows the in-sample performance (Panel A), the out-of-sample performance (Panel B), and the performance measures (Panel C) for the market timing strategies based on the out-of-sample forecasting power of *Variance Risk Premium* for one-month returns of zero-cost investments constructed by taking long one-month forward contracts of foreign *currencies* from the perspective of a U.S. investor. We select the following countries: Japan (JPY), the Great Britain (GBP), the Euro Area (EUR), Switzerland (CHF), Canada (CAD), Australia (AUD), Hong Kong (HKD), Sweden (SEK), New Zealand (NZD), Singapore (SGD), South Africa (ZAR), Denmark (Denmark). The sample period is from Jan. 1990 to Dec. 2013, estimation period for the first regression is from Jan. 1990 to Dec. 1999. The numbers in bold signify statistical significance at the 5% levels.

		JPY	GBP	EUR	CHF	CAD	AUD	HKD	SEK	NZD	SGD	ZAR	DKK
Panel A: In-Sample Analysis													
β_1		0.10	2.11	1.62	1.91	1.80	3.12	0.15	2.80	3.25	1.35	4.87	1.74
t_{β_1}		0.06	**3.29**	**2.04**	**2.35**	**2.98**	**2.64**	1.53	**3.41**	**2.69**	1.94	**2.93**	**1.96**
$\bar{R}^2$ (%)		-0.64	3.05	0.81	1.13	1.33	3.28	2.97	2.36	3.41	2.49	2.89	0.87
Panel B: Out-of-Sample Analysis													
R^2_{OS} (%)		-4.22	3.23	1.28	1.30	2.28	3.85	-7.65	3.46	2.75	2.70	2.03	1.56
$\mathrm{MSE-F}$		-6.80	**5.61**	**2.18**	**2.21**	**3.91**	**6.72**	-11.93	**6.02**	**4.76**	**4.66**	**3.48**	**2.66**
ENC		-2.42	**3.95**	1.23	1.35	**2.28**	**3.77**	**6.33**	**3.58**	**2.84**	**2.84**	**3.80**	1.62
R^2_{OS-CT} (%)		-1.96	1.23	1.02	0.86	1.64	3.18	0.45	2.34	2.28	1.62	2.74	1.37
Panel C: Performance of the Market Timing Strategy													
Active	Mean(%)	-0.20	0.17	0.17	1.34	2.23	7.44	-0.05	3.89	8.84	1.18	17.32	2.96
	Std(%)	1.36	2.57	2.08	5.48	10.14	13.11	0.55	8.56	16.16	4.83	22.27	8.56
	SR	-0.51	0.23	0.28	0.24	0.22	0.57	-0.10	0.45	0.55	0.24	0.78	0.35
Passive	Mean(%)	-0.17	0.12	0.22	3.45	2.96	5.92	-0.40	2.87	7.33	1.21	16.37	2.87
	Std(%)	2.81	2.56	3.12	11.17	9.03	13.24	0.49	11.87	13.93	5.69	21.53	10.50
	SR	-0.21	0.17	0.24	0.31	0.33	0.45	-0.82	0.24	0.53	0.21	0.76	0.27
CER (%)		0.74	0.58	0.34	-0.69	-1.05	1.57	0.35	2.04	0.50	0.10	0.46	0.65

Table 15 Statistical and Economical Significance of the Forecasting Power of VRP for the Commodity Markets

This table shows the in-sample performance (Panel A), the out-of-sample performance (Panel B), and the performance measures (Panel C) for the market timing strategies based on the out-of-sample forecasting power of *Variance Risk Premium* for the excess returns of *commodity* indices. “GSCI” represents the returns on S&P GSCI index (aggregate level). The following sub-sector commodity indices are also included for our analysis: “Energy”, “Industrial Metals”, “Agriculture”, “Livestock”, “Precious Metal”. The sample period is from Jan. 1990 to Dec. 2013, estimation period for the first regression is from Jan. 1990 to Dec. 1999. The numbers in bold signify statistical significance at the 5% levels.

		Aggregate	Component				
		GSCI	Energy	Industrial Metals	Agriculture	Livestock	Precious Metal
Panel A: In-Sample Analysis							
β_1		0.00	0.00	0.00	0.00	0.00	0.00
t_{β_1}		1.25	**1.99**	1.53	0.67	-0.71	1.39
$\bar{R}^2$ (%)		0.46	2.64	0.65	-0.62	0.03	0.88
Panel B: Out-of-Sample Analysis							
R^2_{OS} (%)		-1.40	-1.50	0.37	-0.03	-0.51	-0.56
$\mathrm{MSE-F}$		-2.31	-2.48	0.63	-0.05	-0.84	-0.93
ENC		-0.49	-0.84	1.09	0.54	-0.01	1.06
R^2_{OS-CT} (%)		-1.44	-1.42	0.10	-0.84	-0.25	0.86
Panel C: Performance of the Market Timing Strategy							
Active	Mean(%)	0.49	3.35	11.38	3.73	0.87	1.05
	Std(%)	5.01	14.29	19.72	9.85	5.44	17.92
	SR	0.22	0.10	0.48	0.18	-0.20	-0.05
Passive	Mean(%)	0.76	12.55	10.54	2.95	0.42	13.92
	Std(%)	6.79	31.01	22.64	21.85	13.84	18.96
	SR	0.31	0.34	0.38	0.05	-0.11	0.63
CER (%)		0.46	2.16	2.70	6.49	2.88	-12.29

Figure 1 The Monthly Time-Series for the Variance Risk Premium

This figure plots the monthly time-series for the Variance Risk Premium. The sample is from Jan. 1990 to Dec. 2013. The shaded areas indicate NBER recession periods.

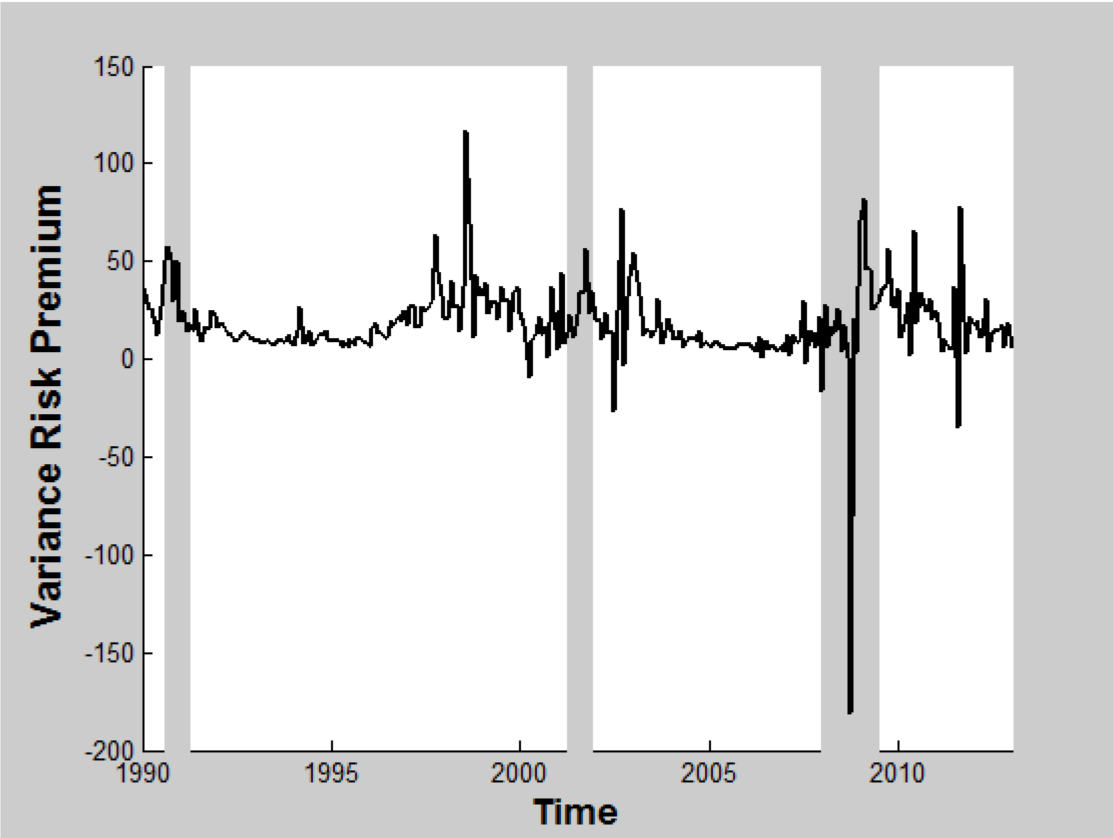


Figure 2 Out-of-sample Performance of the Monthly Predictive Regression for Stock Index: Difference in the Cumulative Sum of Squared Errors (SSE) between the Active and the Passive Strategy

We plot the difference in the cumulative sum of squared errors (SSE) for long-term bond returns. The difference in the cumulative sum of squared errors (SSE) is defined by the difference between the cumulative squared prediction errors of a historical mean model and those of a model with predictive variables. An increase in the cumulative SSE difference indicates better performance of the model with predictors; a decrease in the cumulative SSE difference indicates better performance of the historical mean model. The sample is from Jan. 1990 to Dec. 2013, and the estimation period for the first regression is from Jan. 1990 to Dec. 1999. The shaded areas indicate NBER recession periods.

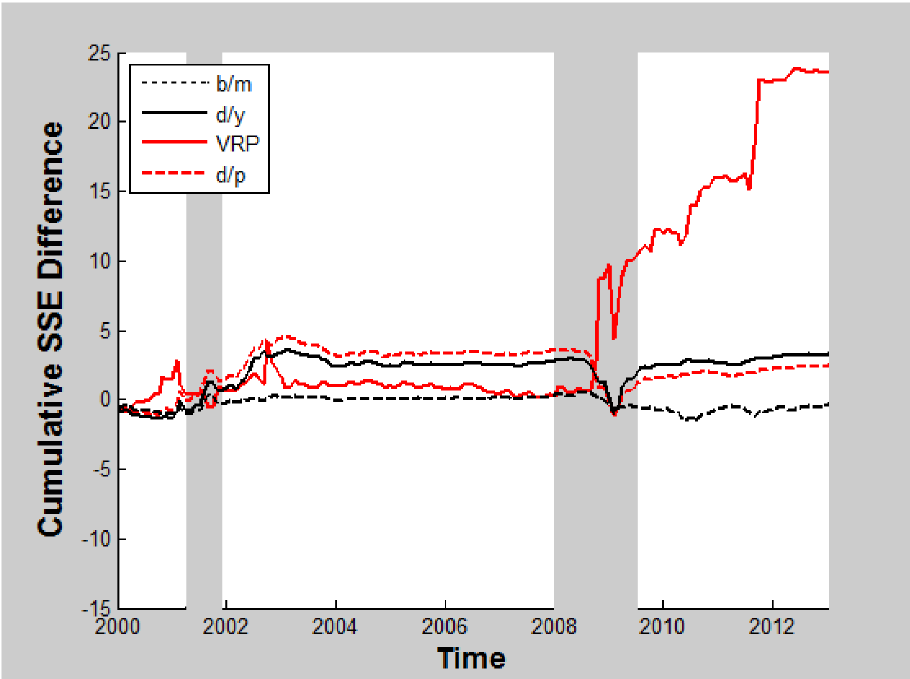


Figure 3 Empirical Distribution for Portfolio Returns Premium: Stock Index

We graph the empirical distribution associated with monthly returns of the passive strategy and two market timing strategies outperforming the passive strategy, the strategy based on VRP and **e/p**.

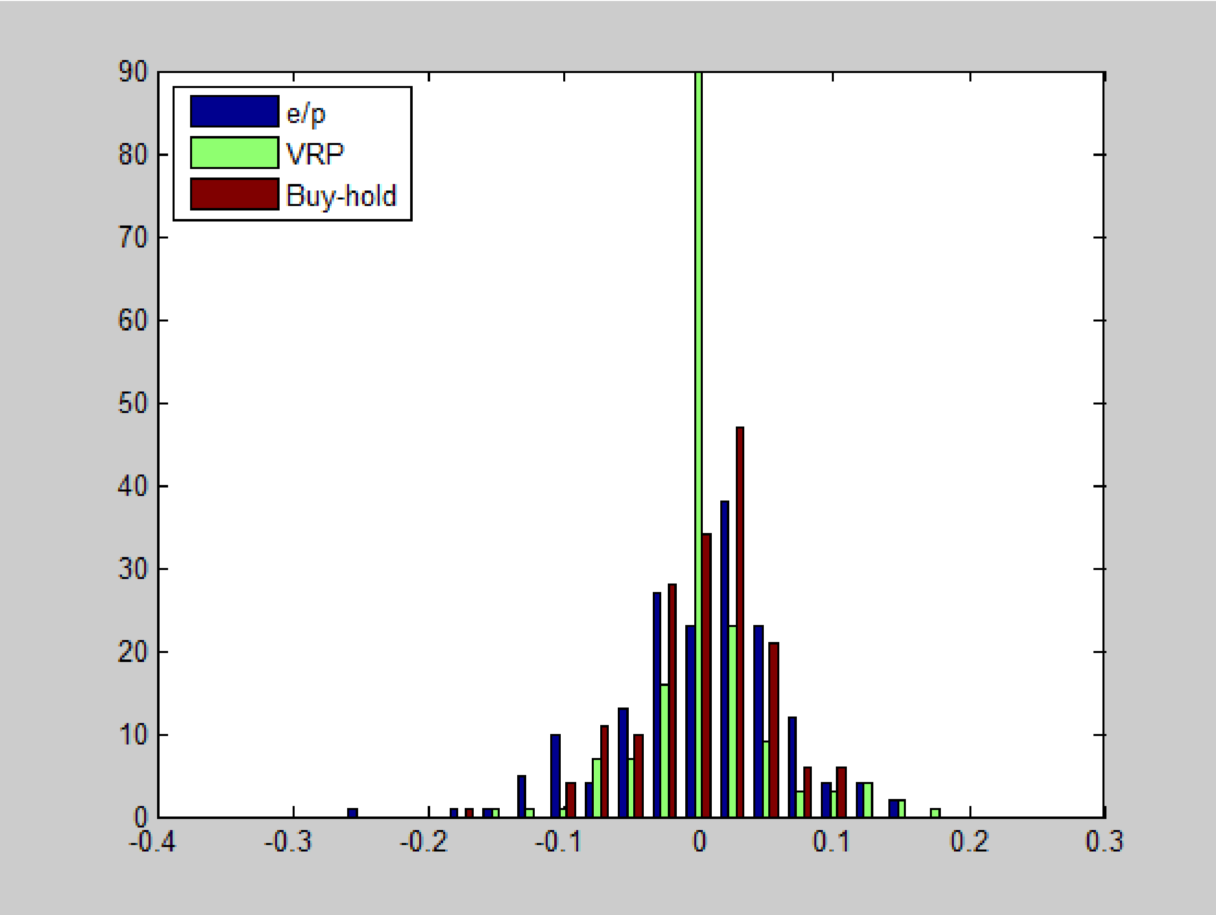
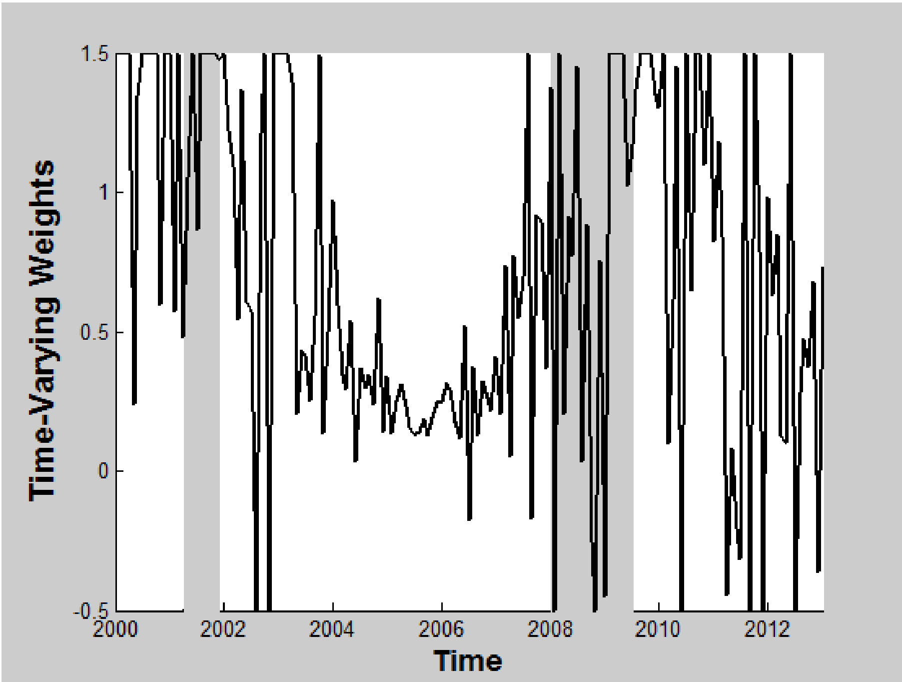


Figure 4 The Portfolio Weights for Market-Timing Strategy: Stock Index

This figure plots his figure plots the portfolio weights associated with the stock index in market-timing strategy based on the forecasting power of VRP (Panel A) and Dividend Yield(Panel B). The sample is from Jan. 2000 to Dec. 2013. The shaded areas indicate NBER recession periods.

Panel A: VRP



Panel B: Dividend Yield (**d/y**)

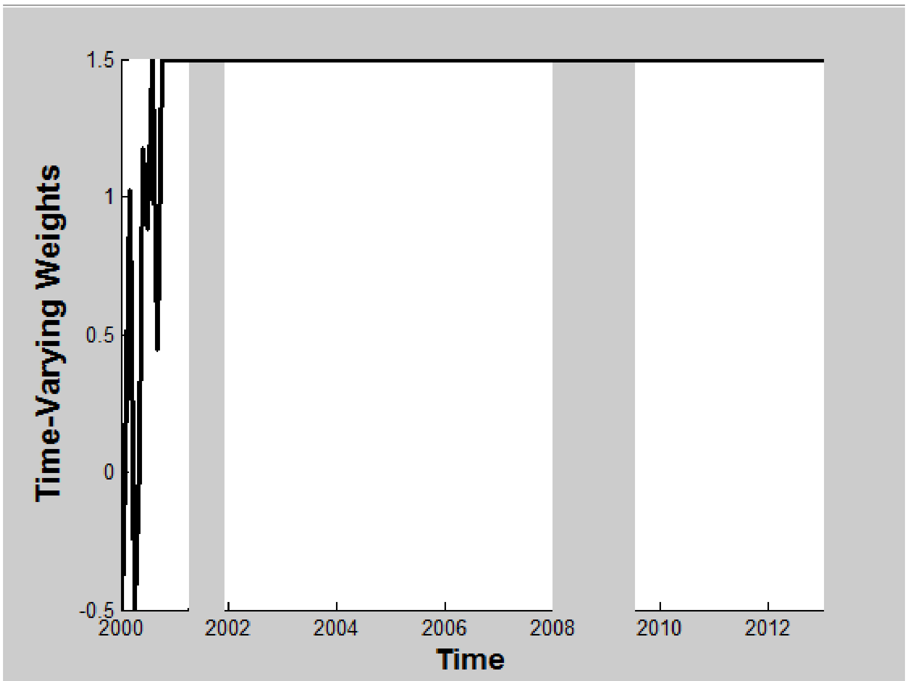


Figure 5 Out-of-sample Performance of the Monthly Predictive Regression for Long-Term Bond Returns: The Difference in the Cumulative Sum of Squared Errors (SSE) between the Active and the Passive Strategy

We plot the difference in the cumulative sum of squared errors (SSE) for long-term bond returns. The difference in the cumulative sum of squared errors (SSE) is defined by the difference between the cumulative squared prediction errors of a historical mean model and those of a model with predictive variables. An increase in the cumulative SSE difference indicates better performance of the model with predictors; a decrease in the cumulative SSE difference indicates better performance of the historical mean model. "HY" denotes Barclays U.S Treasury Corporate High Yield Index, "TR" denotes Barclays U.S Treasury Long Index, "BAA" denotes U.S Treasury Aggregate Corporate Baa Long Index, and "AAA" denotes Barclays U.S Treasury Aggregate Corporate Aaa Long Index. The sample is from Jan. 1990 to Dec. 2013, and the estimation period for the first regression is from Jan. 1990 to Dec. 1999. The shaded areas indicate NBER recession periods.

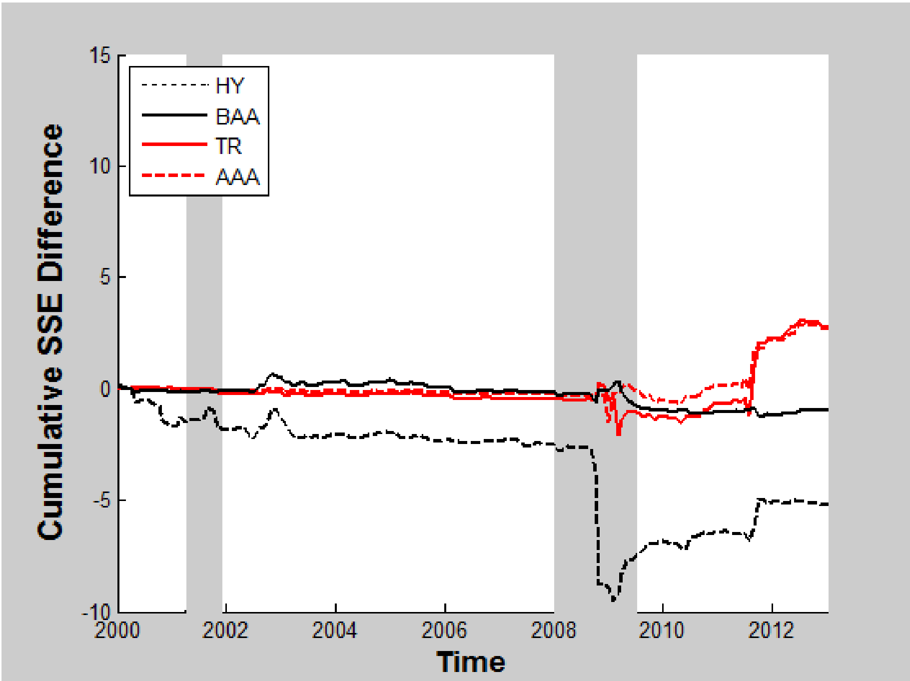


Figure 6 Out-of-sample Performance of the Monthly Predictive Regression for Currency Returns: The Difference in the Cumulative Sum of Squared Errors (SSE) between the Active and the Passive Strategy

We plot the difference in the cumulative sum of squared errors (SSE) for currency returns. The difference in the cumulative sum of squared errors (SSE) is defined by the difference between the cumulative squared prediction errors of a historical mean model and those of a model with predictive variables. An increase in the cumulative SSE difference indicates better performance of the model with predictors; a decrease in the cumulative SSE difference indicates better performance of the historical mean model. Panel A plots the out-of-sample performance of the monthly predictive regressions for currency returns associated with Japan (JPY), the Great Britain (GBP), the Euro Area (EUR) and Switzerland (CHF). Panel B plots the out-of-sample performance of the monthly predictive regressions for currency returns associated with Canada (CAD), Australia (AUD), Hong Kong (HKD) and Sweden (SEK). Panel C plots the out-of-sample performance of the monthly predictive regressions for currency returns associated with New Zealand (NZD), Singapore (SGD), South Africa (ZAR) and Denmark (Denmark). The sample is from Jan. 1990 to Dec. 2013, and the estimation period for the first regression is from Jan. 1990 to Dec. 1999. The shaded areas indicate NBER recession periods.

Panel A: CHF, EUR, JPY, GBP

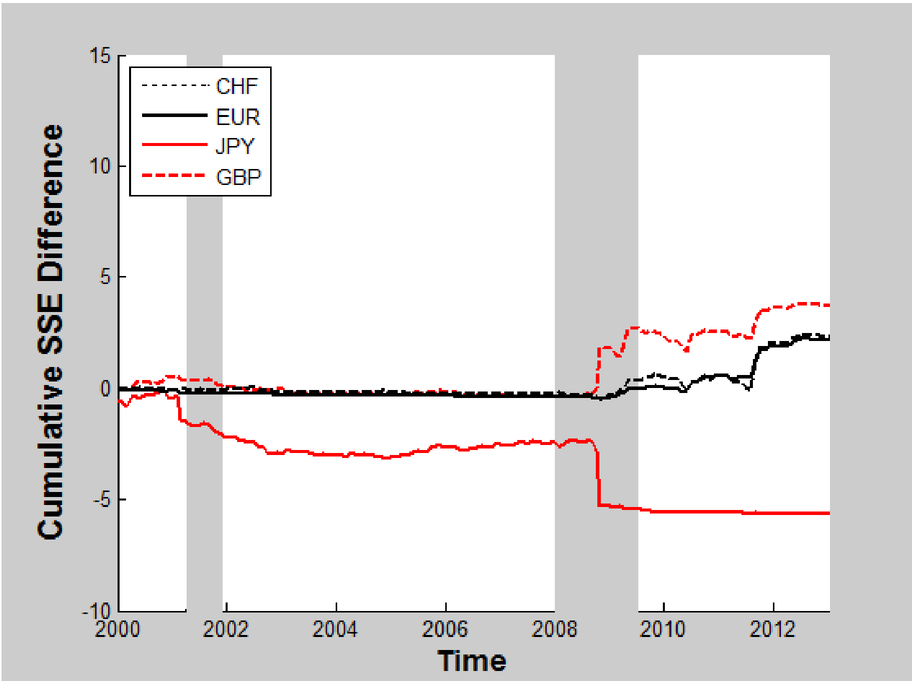
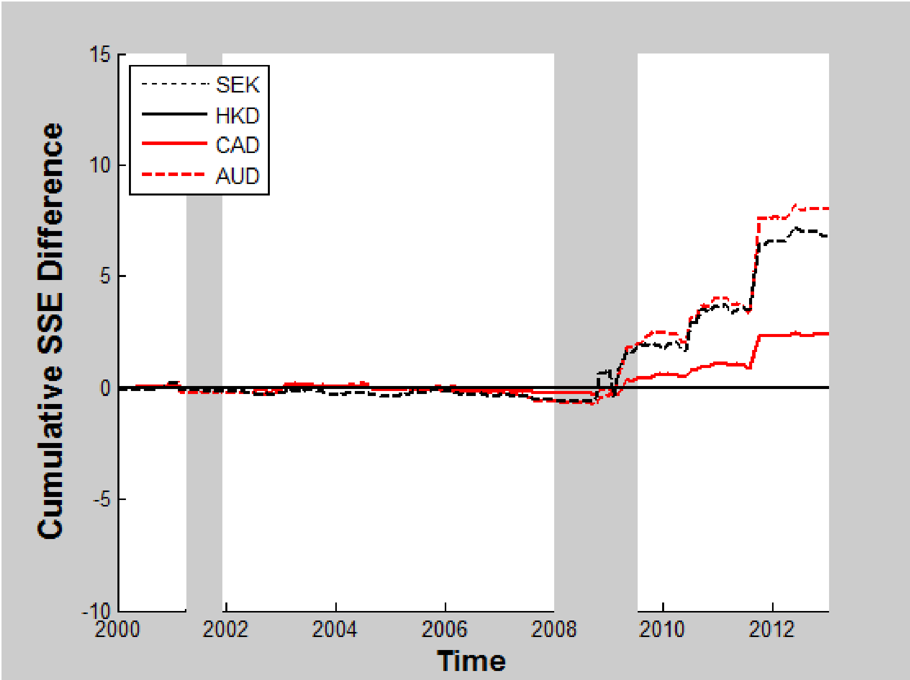


Figure 6 (Continued)

Panel B: SEK, HKD, CAD, AUD



Panel C: DKK, ZAR, NGD, SGD