www.joim.com

Journal Of Investment Management, Vol. 17, No. 3, (2019), pp. 47–64

RETURN PREDICTABILITY AND MARKET-TIMING: A ONE-MONTH MODEL

Blair Hull[a], *Xiao Qiao*[b] *and Petra Bakosova*[c]*

We use weighted least squares to combine 15 diverse variables to build a predictive model for the one-month-ahead market excess returns. We transform our forecasts into investable positions to form a market-timing strategy. From 2003 to 2017, our strategy had 16.6% annual returns with 0.92 Sharpe ratio and 20.3% maximum drawdown. In comparison, the S&P 500 had annual returns of 10%, 0.46 Sharpe ratio, and maximum drawdown of 55.2%. We also combine our one-month model with the six-month model of Hull and Qiao (2017). The combined model had 15% annual returns, Sharpe ratio of 1.12, and maximum drawdown of 14%.

1 Introduction

Return predictability is a central issue in financial economics. Evidence of stock return predictability is supported by numerous academic research papers. Early work includes Campbell and Shiller (1988a, 1988b), and Fama and French (1989) demonstrate that the price–dividend ratio and bond spreads can predict business-cycle frequency return variation. More recently, several papers have presented convincing arguments in favor of predictability (Cochrane, 2005; Rapach *et al.*, 2010; Bollerslev *et al.*, 2009; Jiang *et al.*, 2015). Compared to the large body of work on statistical evidence of return predictability,

*The views expressed are those of the individual authors and do not necessarily reflect official positions of Paraconic Technologies US Inc. or Hull Tactical. All errors are our own.

[a]Founder and Managing Partner, Ketchum Trading, LLC, Founder and Chairman, Hull Investments, LLC. Phone: (312)356-4444. Mailing address: 141 W. Jackson Blvd., Suite 1650, Chicago, IL, 60604, USA. E-mail: blairh@hullinvestmentsllc.com.

[b]Co-Head of Research, Paraconic Technologies US Inc. Phone: (267)320-1601. Mailing address: 200 E 39th St Apt 4E, New York, NY 10016, USA. Email: xqiao@uchicago.edu.

[c]Chief Operating Officer, Hull Tactical Asset Allocation, LLC. Phone: (312)356-3150. Mailing address: 141 W. Jackson Blvd., Suite 1650, Chicago, IL, 60604, USA. E-mail: pbakosova@ketchumtrading.com.

there is relatively little work on the portfolio construction of market-timing strategies based on the proposed return predictors, or the economic importance for an investor who has the resources to engage in a market-timing strategy. Understanding the practical economic significance of market-timing allows us to view predictability from the investor's perspective rather than that of the econometrician.

Hull and Qiao (2017) illustrate the economic importance of return predictability through a market-timing strategy. They select from a set of 20 well-known return predictors, and combine these variables using correlation screening, a variable selection method. They focus on forecasting six-month market excess returns. In the market-timing strategy proposed by Hull and Qiao (2017), the investor can achieve a Sharpe ratio of 0.85 from 2001 through 2015 by taking tactical positions in the S&P 500, compared to a buy-and-hold Sharpe ratio of 0.21 over the same period.

Following the work of Hull and Qiao (2017), we present another novel market-timing strategy. Similar to Hull and Qiao (2017), we combine a large set of forecasting variables to allow for a diverse information set, and construct a reliable market-timing model using these variables. Unlike Hull and Qiao (2017), our forecasting horizon is one month rather than six months. Furthermore, we consider several variables not used in Hull and Qiao (2017) to show that additional variables proposed in the literature not included in Hull and Qiao (2017) could also be used to build an ensemble forecasting model. The new variables include change in inflation, commodity price, housing starts, exchange rates, FRB Loan Officer Survey, delinquencies, and the change in unemployment rate. We discuss the variables in more detail in the data section.

Our goal in this paper is to construct a market-timing strategy based on a one-month forecasting model. Our focus is centered on the performance improvement that our market-timing model can provide over a buy-and-hold strategy of the S&P 500, rather than demonstrating the success of a statistical model. All of the variables we consider have been proposed in the return predictability literature, so the in-sample performance of a forecasting model combining these predictors will necessarily be strong. Instead, we focus on out-of-sample economic importance of return predictability by focusing on investment strategies derived from forecasting models.

We consider 15 return predictors that have been proposed in the return predictability literature. Using weighted least squares (WLS) with stepwise variable selection, we build a forecasting model for next month's market excess returns. We estimate our forecasting model at the end of every month using an expanding window. Within each month, we hold the model parameters constant and form forecasts using updated predictor values each day.

Each return predictor has clear economic intuition. Scaled price ratios capture information about future returns contained in prices (Campbell and Shiller, 1988a). Macroeconomic variables such as change in inflation, industrial production, housing starts, new orders and new shipments, Baltic Dry Index, and change in unemployment rate reflect prevailing macroeconomic states (Chen *et al.*, 1986; Jones and Tuzel, 2012; Bakshi *et al.*, 2011; Flannery and Protopapadakis, 2002). Credit risk premium and slope of the interest rate term structure also reveal macroeconomic conditions and are associated with time-varying returns (Fama and French, 1989). Time-series momentum appears to be associated with investor under-reaction in the short term (Moskowitz *et al.*, 2012).

Combining distinct return predictors leads to improved forecasting power at the expense of economic interpretation. Each time we estimate the model, weights allocated to predictors may change as the relative importance of each variable waxes and wanes through the business cycle. The exact nature of these weights may be difficult to pin down. The merit of our forecasting model comes from combining different predictors. Variable combination allows us to capture a diverse information set and form superior forecast compared to using these predictors in isolation (Rapach *et al.*, 2010).

Based on our statistical model, we construct a monthly market-timing strategy by transforming forecast values into investable positions. We place the following restrictions on our transformation: (1) Avoid market exposure in disadvantageous forecasts. We want to be 0% invested in the S&P 500 and 100% invested in T-bills when the equity premium forecast is zero or negative. (2) Attain full market exposure, or 100% invested in the S&P 500, when the forecast of the equity premium is equal to its historical average. (3) Greater than 100% exposure to the S&P 500 when the equity premium forecast is above its historical average to take advantage of potentially higher returns. We limit the maximum exposure to be 150% long to avoid overleveraging.

We transform our equity premium forecast into investable positions while satisfying the above conditions, and scale the positions to be inversely proportional to the root mean square error (RMSE) of our statistical model. The inverse RMSE scaling can be interpreted as our confidence in our forecasts: If the RMSE of our fitted model is large, we are less confident and want to decrease our position, whereas if the RMSE is small, we are more confident in our model and increase our position.

From 2003 to 2017, the one-month market-timing model achieves annualized returns of 16.6% with a Sharpe ratio of 0.92. In the same period, buy-and-hold S&P 500 has 10% annualized returns and a 0.46 Sharpe ratio. The maximum drawdown of the market-timing strategy is 20% compared to 55% for buy-and-hold. Furthermore, the one-month strategy works even better when combined with Hull and Qiao's (2017) six-month model. A portfolio of equal investments in Hull and Qiao's (2017) six-month model and our one-month model results in 15% annualized returns with a Sharpe ratio of 0.92, and a maximum drawdown of only 14%.[1] The improvement in risk-adjusted return for the combined model comes from diversification across models, as the one-month and six-month models complement each other. Drawdowns for each strategy tend to occur at different times. Combining the two market-timing strategies forms a more stable and robust strategy.

Our paper demonstrates a practical application of return predictability. Many academic studies of return predictability focus on longer horizons of one or more years (Cochrane, 2008; Welch and Goyal, 2008). A smaller number of articles examine forecasting horizons shorter than one year. Bollerslev *et al.* (2009) investigate forecasting using the variance risk premium (VRP) at different horizons, and document a 1.07% adjusted R^2 at the monthly frequency. Moskowitz *et al.* (2012) document that past 12-month market returns are a positive indicator of next month market return. All of these studies focus on the statistical model rather than the investment implications of return predictability. We emphasize the investor's perspective, which is shared in Hull and Qiao (2017).

Our paper is similar to Breen *et al.*'s (1989) who also try to illustrate the economic importance of return predictability from the investor's perspective. Breen *et al.* (1989) calculate the value of market-timing to be worth 2% of the total assets under management. The authors construct their market-timing strategy using the negative relationship between stock returns and the nominal interest rate. We also put forward a market-timing model, but we consider a significantly larger set of predictors. We not only focus on the economic significance of market-timing, but also provide in-depth discussions of the statistical model and the implementation of an investment strategy.

This paper is organized as follows. We list our data sources and describe our statistical model in Section 2. In Section 3, we explain how to translate statistical results in Section 2 into a market-timing strategy. We proceed with an analysis of model combination in Section 4. Section 5 concludes and suggests potential future research.

2 Forecasting the one-month equity risk premium

2.1 Data and variables

The primary sources we use to obtain our data are Bloomberg, the Federal Reserve Board, and the Archival Economic Data from the Federal Reserve Bank of St. Louis (ALFRED).

Motivated by Chen *et al.* (1986), we include macroeconomic variables which may proxy for future investment opportunities as forecasting variables: change in inflation, industrial production, credit risk premium, and the slope of the interest rate term structure. Whereas Chen *et al.* (1986) focused on the cross-section of stocks returns, our focus is on forecasting the aggregate stock market. We add changes in commodity prices, the dollar exchange rate, as well as other variables that may reveal prevailing business conditions. The full set of forecasting variables we consider is as follows:

1. Change in Inflation (UI): Monthly change in the inflation rate net of the change in the risk-free rate. Inflation is calculated as the percentage change in the Consumer Price Index from ALFRED and the risk-free rate is the three-month Treasury bill rate from Bloomberg. Positive change in inflation is associated with higher future market excess returns. Our variable is similar to "unexpected inflation" in Chen *et al.* (1986), calculated as the difference between inflation and its expectation using the Fama and Gibbons (1984) method.
2. Industrial Production (IP): Monthly change in the industrial production index published by ALFRED. IP serves as a leading indicator for the real economy, and is positively associated with future market excess returns (Chen *et al.*, 1986).
3. Credit Risk Premium (CRP): The difference between the BAA and AAA corporate bond yields, also known as the default spread, obtained from Bloomberg. Fama and French (1989) find that a higher CRP is associated with higher future market excess returns.
4. Slope of the Interest Rate Term Structure (STS): The difference between the 10-year Treasury note and the three-month Treasury bill yields from Bloomberg. This quantity is sometimes called the term spread. STS is positively associated with future market excess returns (Fama and French, 1989).
5. Commodity Price (CP): We use the monthly change in the S&P GSCI Index to track the movements in oil price, since the fluctuations in the GSCI are predominantly driven by oil price changes. Historically commodity prices and stock prices have been negatively correlated. We expect an increase in CP to be associated with lower future market excess returns (Black *et al.*, 2014; Casassus and Higuera, 2011).

6. Housing Starts (HS): Monthly difference in the housing starts index from ALFRED. HS is another leading indicator for market cycles, and should be positively correlated with future market excess returns (Flannery and Protopapadakis, 2002).
7. Exchange Rate (EVUSD): Monthly change in the US Dollar Index (DXY) from Bloomberg. Higher dollar exchange predicts lower future market excess returns (Bahmani-Oskooee and Sohrabian, 1992).
8. FRB Loan Officer Survey (LOAN): We compute what fraction of banks have been tightening lending conditions compared to the previous quarter as a proxy for the change in lending conditions among US banks. This variable is from the Federal Reserve Board (Lown *et al.*, 2000). An increase in LOAN indicates credit tightening, which we expect to be associated with lower future economic activity and lower future excess returns.
9. Delinquencies (DL): Annual change in delinquencies from the Federal Reserve Board. Similar to LOAN, this is another variable that we use to capture the macroeconomic conditions of banks, and complements the information contained in LOAN. Higher DL forecasts lower future excess returns (Lown *et al.*, 2000).
10. New Orders New Shipments (NONS): The value of new orders and new shipments excluding defense and aircraft. We exclude defense and aircraft spending to focus on the core capital goods which reveal the underlying trend in real economic activity. Higher NONS is associated with lower future market excess returns (Jones and Tuzel, 2012).
11. Baltic Dry Index (BD): Monthly percent change in the Baltic Dry Index from Bloomberg. We use BD to track global shipping as one measurement of macroeconomic conditions. BD is a leading indicator of economic cycles, and is positively correlated with future excess returns. (Bakshi *et al.*, 2011).
12. National Association of Purchasing Managers (NAPM) Survey: The difference between the manufacturing survey new orders and the prices paid indexes from ALFRED. An increase in NAPM signifies an expansion which is a leading positive indicator for stocks.
13. Change in Unemployment Rate (UR): Monthly change in the unemployment rate from ALFRED. Increases in the unemployment rate indicate deteriorating macroeconomic conditions in the future and are associated with low future market excess returns (Flannery and Protopapadakis, 2002).
14. Momentum (MOM): Monthly percent change in the price return of the S&P 500. Moskowitz *et al.* (2012) suggest that past 12-month returns are a positive signal for next month's returns. We use past one-month return to capture shorter trend in the data. MOM positively predicts future excess returns.
15. PCA of Price Ratios (PRC): We include the first principal component of the cyclically adjusted price-to-earnings ratio (CAPE), cyclically adjusted price to total yield (dividends plus buybacks), and the price-to-book ratio. This composite variable is negatively correlated with future market excess returns. Shiller (2000) proposed the first use of CAPE, and Hull and Qiao (2017) reduce the dimension of price-based variables through PCA.

Our data set is from 06/01/1990 to 02/28/2017. To avoid overfitting, we do not optimize over variable transformations. We normalize all our predictors by dividing by their standard deviations using a backward-looking window of 500 days. This way, the coefficients on the forecasting model can be interpreted as the impact on future returns from a one-standard deviation change in the predictor. Our results are robust to changing the backward-looking window length between 250 and 1,000 days. For macroeconomic variables with revisions, we use the real-time revision history from ALFRED: for each model refit, we only take the latest unrevised observation available to avoid look-ahead bias.

2.2 *Statistical model*

Consider a forecasting regression of target Y_{t+1} using a set of predictors $Z_{j,t}$:

$$Y_{t+1} = \alpha + \sum_{j=1}^{m} \beta_j Z_{j,t} + \varepsilon_{t+1}, \tag{1}$$

where α is the regression intercept, β_j are the forecasting coefficients, and ε_{t+1} is the forecast error. In our application, Y_{t+1} are the one-month-ahead equity risk premia and $Z_{j,t}$ are return predictors which are known this month. The most commonly used and perhaps the best understood method to estimate the model above is using Ordinary Least Squares (OLS). OLS has the following objective function:

$$\min_{\beta_j} \sum_{t=1}^{T} \left(Y_{t+1} - \alpha - \sum_{j=1}^{m} \beta_j Z_{j,t} \right)^2. \tag{2}$$

We search over possible values for the forecasting coefficients and intercept, α and β_j, to minimize the sum of squared residuals.

By construction, OLS puts equal weight on all data points. Standing at time t, Y_{t-1} receives the same weight as Y_{t-12}, which means that the observation one year ago is just as important in estimating model coefficients as the observation last month. While weighting observations equally is simple and easy to understand, as the economy evolves, more recent data may be more relevant compared to older data. To allow for recent data to have a larger impact on our model, we consider weighted least square (WLS) with the following objective function:

$$\min_{\beta_j} \sum_{t=1}^{T} \rho^{T-t} \left(Y_{t+1} - \alpha - \sum_{j=1}^{m} \beta_j Z_{j,t} \right)^2, \tag{3}$$

where ρ is the decay factor which we set to 0.99 at the monthly frequency. This value implies a half-life of approximately 60 months. Varying the value of ρ such that the half-life of the weights ranges between 30 and 120 months does not materially affect our results. OLS is a special case of WLS if we set ρ to be one.

We consider 15 forecasting variables, which means each time we fit the model we need to estimate 15 forecasting coefficients plus the intercept term for a total of 16 parameters. We limit the number of estimated parameters through variable selection, which leads to more parsimonious models and generally results in better out-of-sample forecasting properties. Parsimonious models are also easier to interpret and attribute performance. It is easier to understand which variables contribute to forecasting results when there are fewer variables to consider.

We estimate the WLS specification which incorporates a bidirectional stepwise procedure. Variables are chosen based on the Akaike Information Criterion (AIC). The bidirectional stepwise selection combines forward selection, which starts with no variables in the model and adds variables that capture the largest improvement, and backward elimination, which starts with all candidate variables and removing the least significant variables. One feature of the stepwise WLS estimation is that the number of selected variables will change, as predictors come in and out of the selected set. In comparison, in WLS without variable selection, all of the variables always get non-zero weights, even if they only contribute marginally in a given sample. Hull and

Qiao (2017) use correlation screening as their variable selection technique because using overlapping six-month market returns leads to inflated t-statistics and results in misleading traditional likelihood function calculations and AIC values.

We estimate the stepwise WLS at the end of each month. Starting on 03/31/2003, we use 06/01/1990 to 03/31/2003 to estimate our model. We obtain the model parameters on 03/31/2003, which we hold constant from 04/01/2003 to 04/30/2003. For every day in this month, we use the updated return predictors, along with the fixed model parameters, to produce one-month equity risk premium forecasts on a daily frequency. Specifically, we use parameter values from 03/31/2003 and predictor values from 04/01/2003 to form our position for 04/02/2003. On 04/30/2003, we re-estimate our model using an expanding window, from 06/01/1990 to 04/30/2003, to obtain new parameter values which we use for next month's equity premium forecasts. We continue to re-estimate our model monthly and make one-month equity premium forecasts every day until the end of the sample.

2.3 *Variable selection*

In Figure 1 we look at the identity of the selected variables. On the vertical axis is the contribution of each explanatory variable towards the total explained variance (Lindeman *et al.*, 1980; Chevan and Sutherland, 1991). The stepwise WLS puts zero weight on marginal variables that do not add substantially to the model. Of the 15 variables we consider, typically only about five to seven are selected at any given time. There are significant changes in the number and identify of the selected variables. In contrast, without variable selection, WLS reduces the weight put on marginal variables that do not contribute to the explanatory power of the model, but does not remove them from the model.

Consider variable X which does not add any forecasting power to the model. If we use WLS with variable selection, the marginal contribution of X

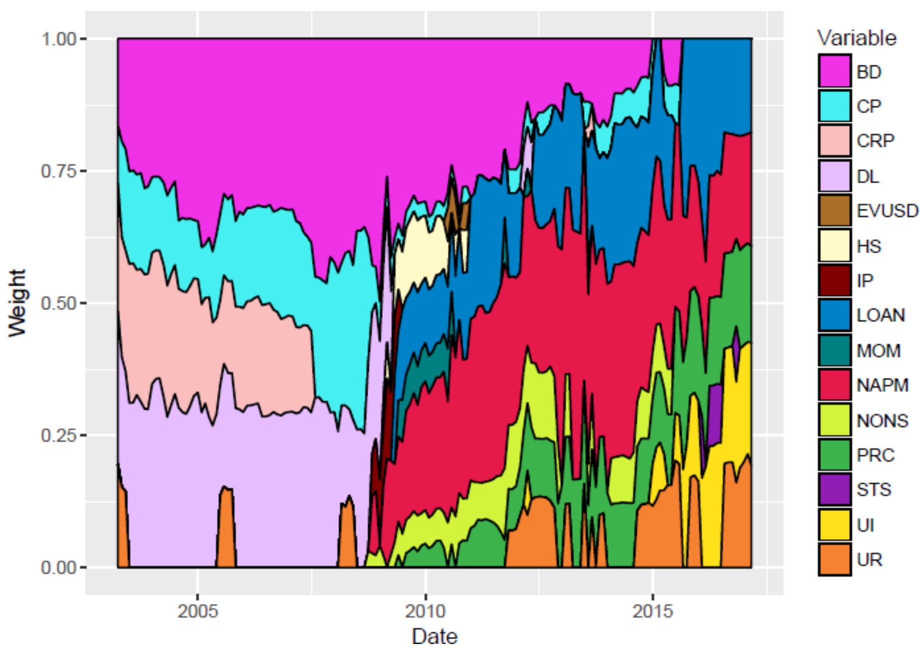

Figure 1 Forecast Contribution of Selected Variables, 03/31/2003–02/28/2017.

We plot the contributions of each of the predictors towards the forecast. The total contribution sums up to 100%.

would be too small and it will be eliminated from our model. If we use WLS, we always put a positive weight on X. For out-of-sample prediction, including X only adds noise to our forecast. The forecasts coming from WLS with variable selection will likely be more stable compared to WLS without variable selection.

Some variables were selected to be in the model throughout the sample, whereas others only contributed to the explanatory power of the model in a fraction of the sample. CP was selected in the earlier part of the sample until 2009, and then it was eliminated from the model. BD was important until 2013, and then it was driven out by other variables. On the one hand, CRP and DL were useful in the first half of the sample but not in the second half. On the other hand, NAPM and LOAN were only selected in the second half but not in the first half. In addition, some variables such as EVUSD were almost never selected, but we did not remove them from the pool of candidate variables.

Variables entering and exiting our model may be due to variables containing overlapping information. When one variable is dropped from the model, another variable (or several variables) that may share part of the same information set could come into the model. For example, in 2004 Baltic Dry Index (BD) was temporarily removed from the model and UR (Change in Unemployment Rate) was added. It is likely that both variables contained useful information about the macroeconomic environment, so when BD was dropped from the model, another variable that contained similar information was added.

3 A one-month market-timing strategy

3.1 Trading strategy construction

Now that we have established good statistical forecasts of next month excess market returns, we can translate these statistical results into a trading strategy. We determine the amount invested in the market by transforming the raw equity premium forecasts. When the equity risk premium forecast for the next month is zero, we want to be 0% invested in the market and 100% in T-bills. We also want to be 100% invested when the equity premium forecast is at its historical average. We want to transform our raw equity premium forecasts into how much to invest in the market while satisfying these conditions.

There are different ways we can make such a transformation. We use a simple method that takes into account our confidence in the model. Each month when we estimate the model, we get a value for the root mean square error (RMSE) for the fit. A low RMSE indicates that the model fits well, and a high RMSE indicates that a model which leaves much variation in the data unexplained. Intuitively, when the RMSE is high and the model fits poorly, we are less confident about our forecasts and would like to scale back our position. When the RMSE is low and the model fits the data well, our forecasts are likely to be more precise with lower forecast errors. We want to scale up the resulting position.

In line with the arguments above, we scale our equity premium forecasts by the inverse of the RMSE and multiply it by 5 to satisfy the above conditions. In particular, when the equity premium forecast is at its historical mean, scaling by the RMSE and multiplying by 5 results in a position close to 100%. When the equity premium forecast is zero, we are 0% invested. We restrict our position to be between 0% and 150% invested in the S&P 500. Even if our equity premium forecast were negative, we set our position to be 0% rather than take a short position.

To understand the impact of this transformation on our results, we consider two alternative transformations. First, we simply multiply our forecasts by 100. Second, we divide our forecasts by a four-year moving average of the predicted returns. Both transformations lead to similar results to those of the RMSE transformation.

We include transactions costs and borrowing costs in our analysis. We assume a 0.6 cent per-share fee for trading the SPDR S&P 500 ETF, SPY.[2] We also assume that we can borrow at the 13-week Treasury rate plus 25 basis points. These costs may be somewhat optimistic for retail investors, but they are feasible for a typical professional money manager. Our results are based on trading at the closing price at 4:00 p.m. EST, and assume that we can execute our trades on market close. In 2014, the SPX futures traded $145 billion and SPY traded $21 billion per day, and the closing auction at the NYSE Arca alone averaged more than $422 million per day. At such depth, it is unlikely that slippage will affect the returns from our strategy.

Trading daily may be inefficient because day-to-day adjustment to the position may be small. As an alternative, we consider trading only once a month at the end of the month. Perhaps not surprisingly, this strategy with monthly rebalancing does not perform as well as the strategy that rebalances daily. We have included variables at different frequencies which may also be released at different times of the month. By restricting position adjustment to be only at the end of the month, we do not capture the intra-month changes in some predictors which may provide valuable information. A strategy that considers the tradeoff between transactions costs and signal decay may rebalance more frequently than once a month but less frequently than once a day.

3.2 Strategy performance

Figure 2 presents the cumulative wealth of $1 invested in the market-timing strategy versus invested in the S&P 500. In our sample from 2003 to February of 2017, the market grew from $1 to just over $3, whereas the market-timing strategy would have grown $1 to over $8 in the same period. An important difference between the two cumulative wealth paths is that during the Global Financial Crisis (GFC), the market-timing strategy did not experience a large drawdown, whereas the S&P 500 had a

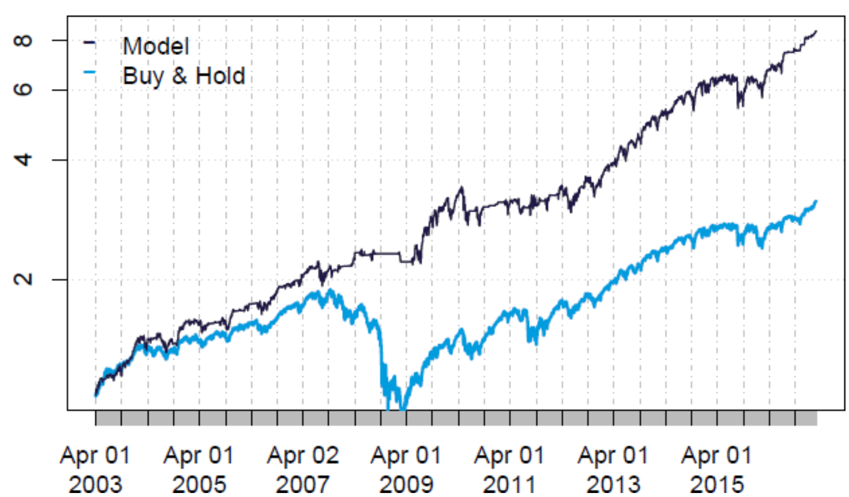

Figure 2 Wealth Accumulation of the One-Month Market-Timing Model and Buy-and-Hold.

We plot the cumulative returns of $1 compounded of our one-month market-timing strategy in black (Model), and cumulative returns of 100% invested in the S & P 500 (Buy & Hold).

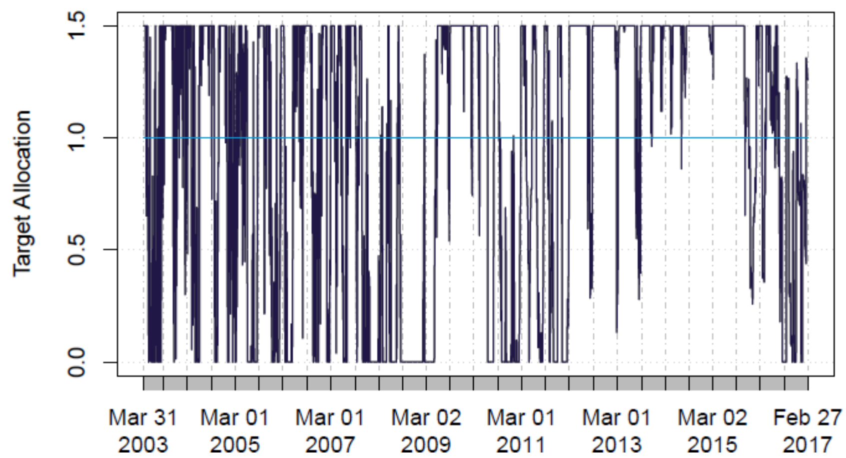

Figure 3 Market-Timing Model Positions.

This figure contains the fraction of assets invested in the S&P 500 for our market-timing model. The strategy is capped between 0% and 150% invested in the market.

drawdown of over 50%. The largest drawdowns of the market-timing strategy occurred in 2010 and 2015.

Figure 3 shows the positions taken by our market-timing model. The market exposure is constrained to be between 0% and 150% (1.5 times levered). There is large variation in the day-to-day positions. In the first four years the position changes were particularly large compared to those in later years. From 2012 to 2015, there is an extended period in which the model called for positive positions of 50% or greater. This aggressive positioning coincides with a period in which market excess returns were large and positive. From Figure 2, we can see the cumulative wealth of the market-timing strategy almost doubled in this period.

Some market-timing models, including the one put forward by Hull and Qiao (2017), tend to be able to avoid large market-wide drawdown and perform especially well when the aggregate market returns are poor. However, in market booms, Hull and Qiao's (2017) model is not fully exposed to the upside gains. In comparison, our one-month market-timing model is able to outperform buy-and-hold returns both in bear and bull markets. In Figure 4, we see that the one-year rolling relative performance to the market is the largest during the Global Financial Crisis, but the relative performance is still positive during the stock market recovery after the GFC. Significant upside capture is one desirable feature of our model that is absent in Hull and Qiao's (2017).

One important measure of risk, especially for practitioners, is the maximum drawdown experienced by a strategy. A drawdown is calculated as the percent change from a strategy's peak to its next closest trough; maximum drawdown is the largest drawdown in the lifetime of the strategy. Maximum drawdown complements volatility to give investors a more complete picture of the stability of the strategy. Figure 5 compares the drawdowns of our market-timing strategy with those of the S&P 500. There are several differences between the two series.

First, the magnitudes of the maximum drawdowns are vastly different between buy-and-hold and our market-timing model. For the S&P 500 buy-and-hold, the maximum drawdown exceeds 50% during the Global Financial Crisis. In comparison, the largest drawdown experienced by the market-timing model in this sample is about

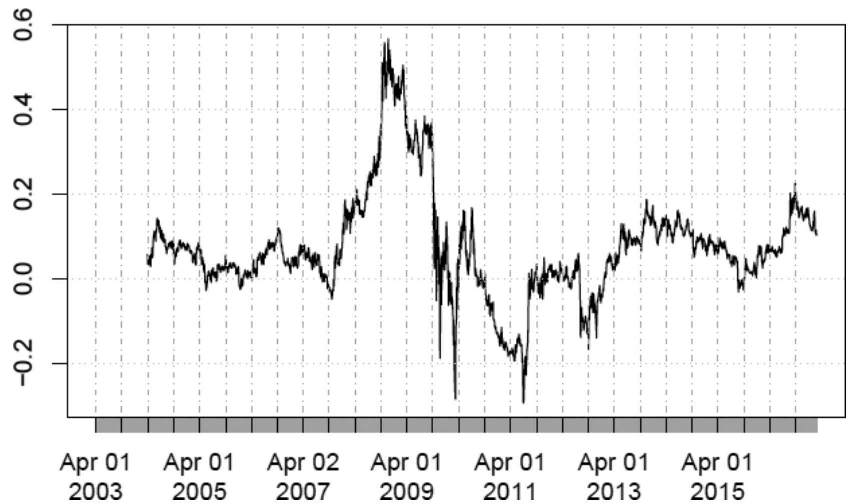

Figure 4 Rolling One-Year Performance Relative to Buy & Hold.

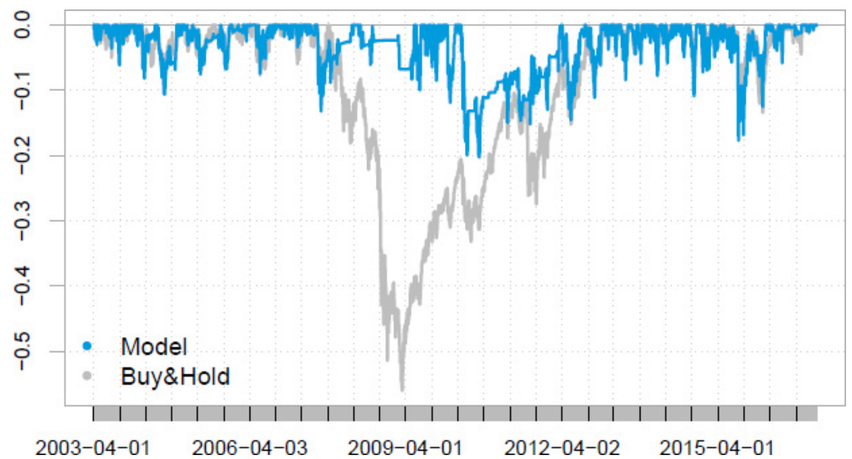

Figure 5 Drawdowns of the One-Month Market-Timing Model and Buy-and-Hold.

We compare the drawdowns of the market-timing model (Model) and 100% invested in the S&P 500 (Buy&Hold).

20%, less than half of the drawdown for the S&P 500.

Second, drawdowns of the market-timing model and those for buy-and-hold do not appear to be strongly correlated. When the S&P 500 had a large drawdown in 2008 and 2009, the market-timing model did not have significant drawdowns in that same period. In fact, during the largest S&P drawdown, the market-timing model had modest drawdowns of less than 10% which only lasted a few months. When the market-timing model had its largest drawdowns of 20% in 2010 and 2015, the S&P 500 did also experience some drawdowns of similar magnitudes, but these are mild episodes compared to the maximum drawdown during the Global Financial Crisis.

Third, the two strategies in Figure 5 appear to experience drawdowns at different frequencies. For the S&P 500, there were several large drawdowns on the order of 25–55% which lasted several months or longer each time. In comparison, our market-timing model experienced more frequent drawdowns of about 5–10%, but these rarely lasted for more than a few months. Most of these relatively small drawdowns reversed after just a few weeks.

The two biggest drawdowns for our market-timing strategy occurred in 2010 and 2015. In the summer of 2010 and 2011, our model prescribed 150% long position when the S&P 500 recorded negative returns, and the strategy was close to 0% invested during the market rebounds. The Baltic Dry Index and NAPM were the main drivers for this drawdown. In August 2015, the strategy was levered and was over 100% invested when the market performed poorly. LOAN and NAPM were the main contributing variables for the model signal that month.

While the one-month market-timing model we have put forward is interesting on its own, it becomes more powerful when combined with other models. Investors dislike high volatility and large drawdowns, and value the ability to improve along those dimensions. In the next section, we explore combining the one-month model with another market-timing model.

4 Combining market-timing strategies

Model combination is beneficial both from the perspective of predictive modeling and the portfolio construction. Statistically, combining models that are independently able to forecast returns leads to diversification across models and potentially results in a superior combined forecast (Timmermann, 2006). From the perspective of modern portfolio theory, combining portfolios that are not perfectly correlated leads to new portfolios with better mean–variance properties (Markowitz, 1952).

We explore the benefits of model combination by combining our one-month model with the six-month model laid out in Hull and Qiao (2017). As the focus of Hull and Qiao (2017) is on a six-month strategy, the investment horizon they consider is different from ours. Combining the one-month model with the six-month model has the added benefit of diversifying across different investment horizons.

To combine the two models, we simply invest equal dollar amounts in each. The resulting portfolio is a 50–50 combination of the two strategies. Combining models using equal weights has the advantage that we do not need to estimate portfolio weights on each constituent. With just two time series and a limited history, combining these two strategies using other methods (such as mean–variance optimization, which requires estimating the mean and covariance matrix) may result in limited gains.

Annualized return is the average annual gross return of the strategy over the period from 03/31/2003 to 02/28/2017. Volatility and Sharpe ratio are also annualized. We test for equal Sharpe ratios of a strategy and buy-and-hold using the Ledoit and Wolf (2008) methodology. The *p*-values are shown as LW *P*-Value. CAPM Alpha and Beta are from time-series regressions of each set of strategy excess returns on S&P 500 excess returns. Maximum drawdown is calculated from peak to trough of the cumulative return series.

We compare the performance of different strategies in Table 1. There are some notable differences among the stand-alone strategies before any combination. The six-month model shows annual returns that are about 250 basis points higher compared to buy-and-hold in this period, and the one-month model has higher returns compared to both the six-month model and buy-and-hold returns. Annual market volatility from 2003 to 2017 was 18.4%, whereas the one-month model has a reduced volatility of 16.6% and the six-month volatility is the lowest among the stand-alone strategies at 11.7%. Given the higher annual returns and lower volatility of the market-timing models compared to buy-and-hold, it is

Table 1 Strategy summary statistics.

	Six-month model	One-month model	Combined model	Buy & Hold
Annual Return	12.4%	16.6%	14.9%	9.8%
Volatility	11.7%	16.6%	12.2%	18.4%
Sharpe	0.95	0.92	1.12	0.46
LW P-Value	0.178	0.035	0.008	N/A
CAPM Alpha	9.5%	9.7%	9.7%	0.0%
CAPM Beta	0.18	0.58	0.38	1.00
Max Drawdown	17.4%	20.3%	14.1%	55.2%

not surprising that the Sharpe ratios achieved by the market-timing strategies are markedly higher compared to the Sharpe ratio of buy-and-hold. The one-month model and the six-month model Sharpe ratios are twice as high as that of buy-and-hold.

We evaluate the statistical significance of our market-timing strategies. In our sample period, the VIX ranged from 9% to 90%, so any statistical test must take heteroscedasticity into consideration. Ledoit and Wolf's (2008) test of equal Sharpe ratios accounts for both heteroscedasticity and heavy tails. We can reject the null hypothesis of equal Sharpe ratios for the one-month model versus buy-and-hold; the p-value is 0.035. However, we cannot reject equal Sharpe ratios for the six-month model of Hull and Qiao (2017) compared to buy-and-hold. Hull and Qiao's (2017) model has a much lower volatility compared to buy-and-hold, such that the difference between the two series is still volatile. The inability to reject the null may be due to this volatility. The combined model clearly has a higher Sharpe ratio compared to buy-and-hold, rejecting the null at the 1% level.

Another way to compare the market-timing strategies with buy-and-hold is to calculate the marginal benefit of investing in market-timing strategies relative to investing in buy-and-hold. We run time-series regressions of the six-month and one-month strategies returns on buy-and-hold returns. The regression coefficient is the market beta from the Capital Asset Pricing Model (CAPM), and the intercept is the additional benefit of market-timing relative to holding the market, or alpha. Table 1 shows economically large alphas of 9–10% for the market-timing strategies. The CAPM betas for the market-timing strategies are modest for both the one-month and the six-month models. On average, neither strategy is fully exposed to market risk.

For buy-and-hold, the maximum drawdown was 55% during the Global Financial Crisis in late 2008. In comparison, the market-timing models offer greatly reduced drawdowns: The one-month model has a maximum drawdown of 20% and the six-month model has a maximum drawdown under 18%. Both market-timing models are able to avoid the large market drawdown in the last quarter of 2008.

Now that we have an understanding of how each strategy did relative to one another, we turn our attention to the combined strategy of 50% invested in the six-month model and 50% invested in the one-month model. The combined model has annual returns, CAPM alpha, and CAPM beta which are between the two stand-alone market-timing models. These values should not be surprising as the combined returns must fall between the returns of the two separate strategies. The combined model shows the gains from diversification in its volatility, Sharpe ratio, and maximum drawdown. Strategy volatility and maximum drawdown for the combined model are lower than those for both the one-month and the six-month models. A maximum drawdown of only 14% is less than one-third of the value for buy-and-hold. Although the combined strategy does not have the highest raw returns, it achieves the highest Sharpe ratio of all the strategies. Its Sharpe ratio of 1.12 is 12% higher than the Sharpe ratio of the one-month model and 18% higher compared to that of the six-month model.

Table 2 presents the year-by-year annual returns of the four strategies. The first observation is that the aggregate stock market has done extraordinarily well in this period. Of the 15 years considered from 2003 to 2017, the buy-and-hold strategy only had one negative year—in 2008. Without 2008, the average annual return of the market in this period would be 14%! In comparison, neither the six-month model nor the one-month model has a single negative year.

Returns for the six-month model, one-month model, the combined model, and the S&P 500 (Buy & Hold) every year from 2003 to February 2017.

The persistence of strategy performance relative to the buy-and-hold strategy differs between the six-month and the one-month models. The six-month model has long periods of persistent outperformance or underperformance relative to the buy-and-hold benchmark. From 2003 to 2007, the six-month model underperformance buy-and-hold for five consecutive years. On the other hand, from 2011 to 2015, the six-month model outperformed buy-and-hold for five years in a row.

We can readily see the diversification benefits of combining the six-month model and the one-month model. Of the five consecutive underperforming years for the six-month model, the one-month model outperformed buy-and-hold in

Table 2 Year-by-year returns.

	Six-month model	One-month model	Combined model	Buy & Hold
2003	6.42%	35.29%	20.86%	32.93%
2004	5.57%	15.95%	10.76%	10.69%
2005	2.75%	6.41%	4.58%	4.83%
2006	7.50%	15.84%	11.67%	15.85%
2007	7.77%	11.60%	9.69%	5.14%
2008	19.98%	7.76%	13.87%	−36.81%
2009	47.52%	31.85%	39.69%	26.36%
2010	3.49%	0.87%	2.18%	15.06%
2011	10.57%	2.93%	6.75%	1.89%
2012	16.32%	9.93%	13.13%	15.99%
2013	34.77%	47.27%	41.02%	32.31%
2014	14.55%	20.11%	17.33%	13.46%
2015	2.02%	2.69%	2.36%	1.25%
2016	3.72%	26.80%	15.26%	12.00%
2017	2.66%	4.71%	3.69%	9.31%

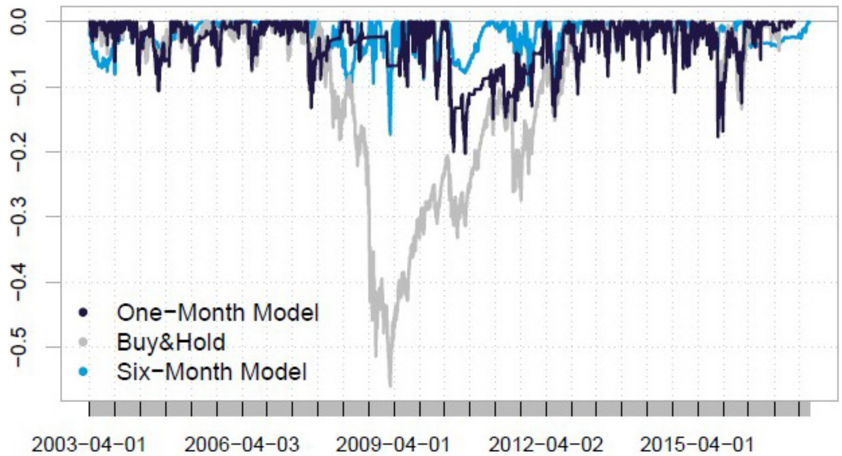

Figure 6 Drawdowns of the Market-Timing Models and Buy-and-Hold.

We overlay the drawdowns of the one-month model, the six-month model, and 100% invested in the S&P 500 (Buy & Hold).

four out of the five years. When we combine the two market-timing models, the persistence in underperformance or outperformance relative to buy-and-hold disappears. The combined model does not show prolonged periods of underperformance or outperformance like those of the six-month model.

Combining the one-month model and the six-month model also provides tail risk diversification. Figure 6 plots the drawdowns of the one-month model, the six-month model, along with those of the S&P 500. It is evident that the two market-timing models suffer drawdowns at different times. In April of 2009, the six-month model had a brief 18% drawdown, whereas the one-month model has a smaller than 5% drawdown at the same time. When the one-month model had a relatively large drawdown of 20% in late 2010, the six-month model had a smaller drawdown of less than 10% which only lasted half as long. In 2015 when the one-month model had a near 20% drawdown, the six-month model only had a minor drawdown and quickly recovered. Figure 6 shows each strategy has drawdowns that occur at different times, suggesting that combining the two market-timing models will bring strategy drawdowns to more moderate levels compared to each individual strategy.

We present the drawdowns of the combined strategy in Figure 7. The combined strategy is less susceptible to drawdowns than both standalone strategies. Although we see many small drawdowns, they are rarely greater than a few percentage points, and the strategy rebounds to a new high more quickly than both the one-month and the six-month models. The largest drawdown of 14% for the combined strategy is smaller than the largest drawdowns for the one-month or the six-month models. In fact, over the 15-year period in this sample, the combined strategy only realizes a 10% or greater drawdown four times.

The volatility of the combined model is 12%, whereas the market volatility over the same period is 18.4%. If the investor were to replace some buy-and-hold allocation in his portfolio with the combined market-timing model, he may want the replacement to have similar levels of volatility. We consider volatility targeting for the combined portfolio by scaling the portfolio positions such that the ex-ante portfolio volatility equals that of the market. The volatility-targeting model

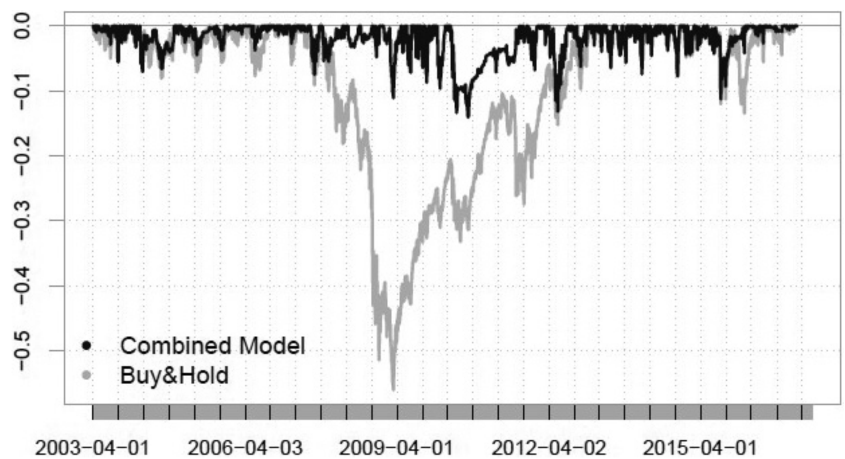

Figure 7 Drawdowns of the Combined Market-Timing Model and Buy-and-Hold.

We compare the drawdowns of the combined market-timing model and 100% invested in the S&P 500 (Buy & Hold).

has a realized volatility of 17% with a compound return of 19%, and maximum drawdown of 22%. Volatility targeting maintains the risk–return properties of the combined model, but offers an investment with similar levels of risk as the S&P 500.

5 Conclusion

There is a preponderance of positive evidence for return predictability. This paper asserts that we can use our knowledge of return predictability to consistently produce returns which exceed those of buy-and-hold. Our work complements Hull and Qiao (2017) in describing the construction of a market-timing strategy. We build a predictive model for one-month market excess returns, including variables that were proposed in the predictability literature and some novel variables. Weighted least squares regression combines these variables into a forecasting model, allowing for more recent data to have greater influence on the outcome.

The one-month market-timing model doubles the Sharpe ratio of the buy-and-hold strategy (0.92 versus 0.46) from 2003 to 2017 and greatly reduces the maximum drawdown from 55% to 20%. A combination of the one-month model and Hull and Qiao's (2017) six-month model is more efficient than both models separately. A 50–50 portfolio investing equal dollar amounts in the two market-timing strategies results in a higher Sharpe ratio and smaller drawdowns than both the one-month and the six-month strategies.

Throughout this paper, we focused on forecasting future one-month excess market returns. We could try to form forecasts at even shorter horizons, on the order of days or weeks. Candidates from the traditional predictability literature including macroeconomic and fundamental variables may not be well-suited for capturing such short-term market movements. However, expanding the information set to include additional price-based variables or technical indicators may prove fruitful.

We illustrate the power of model combination through combining two market-timing models which operate at different frequencies. The increased Sharpe ratio and decreased drawdowns provide real economic benefits to the investor. Additional models which are not perfectly correlated with these two would likely further improve performance statistics. How to optimally combine models to achieve the highest possible risk-adjusted return may be an interesting research direction.

Another potential research direction is to consider other countries. Although international data has generally shorter history and some variables may be less reliable, these markets would provide out-of-sample tests for our approach. Most of our predictors are specific to the US. It would be interesting to see which predictors work well across countries, and how beneficial variable combination methods are for forecasting returns.

Return predictability has become well-accepted in the academic literature and has shown up in top business school finance courses. The continued stigma associated with market-timing primarily comes from the debate on whether predictability is sufficiently strong for a trading strategy to be built. Such a trading strategy would need to create enough economic value to offset its costs and produce superior returns compared to buy-and-hold. As more strategies such as the one we present here develop, the attitude towards market-timing may change. Just as it was considered irresponsible to participate in market-timing in the last 30 years, it may be considered irresponsible not to participate in market-timing in the next 30 years.

Notes

[1] This combined model has lower volatility compared to the S&P 500. If we matched the S&P volatility, we would obtain higher returns and larger drawdowns.

[2] We could also trade S&P 500 futures instead of SPY. We assume $1.18 per share and a half tick of slippage when trading futures, and find similar results.

Acknowledgment

We would like to thank Alexios Galanos for contributing to the original research on this model, and Alexander Kment for data analysis. We also thank Rick Anderson, Alexander Kment, Jerome Pansera, Jim Lodas, John Rizner, Anlong Li, and Dirk Eddelbuettel for their helpful comments. The views expressed are those of the individual authors and do not necessarily reflect official positions of Paraconic Technologies or Hull Tactical. All errors are our own. This paper is not an offer to sell, or a solicitation of an offer to buy an investment product or services offered by Paraconic Technologies or Hull Tactical. Paraconic Technologies or Hull Tactical does not guarantee the accuracy or completeness of the information contained herein and any information provided by third parties has not been independently verified by Paraconic Technologies or Hull Tactical.

References

Bahmani-Oskooee, M. and Sohrabian, A. (1992). "Stock Prices and the Effective Exchange Rate of the Dollar," *Applied Economics* **24**, 459–464.

Bakshi, G., Panayotov, G., and Skoulakis, G. (2011). "The Baltic Dry Index as a Predictor of Global Stock Returns, Commodity Returns, and Global Economic Activity," AFA 2012 Chicago Meeting Paper.

Black, A. J., Klinkowska, O., McMillan, D., and McMillan, F. J. (2014). "Forecasting Stock Returns: Do Commodity Prices Help?," *Journal of Forecasting* **33**, 627–639.

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

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

Campbell, J. Y. and Shiller, R. J. (1988b). "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance* **43**, 661–676.

Casassus, J. and Higuera, F. (2011). *Stock Return Predictability and Oil Prices*. 406, Instituto de Economia, Pontificia Universidad Catolica de Chile.

Chen, N., Roll, R., and Ross, S. A. (1986). "Economic Forces and the Stock Market," *Journal of Business*, 383–403.

Chevan, A. and Sutherland, M. (1991). "Hierarchical Partitioning," *The American Statistician* **45**, 90–96.

Cochrane, J. H. (2005). *Asset Pricing*. Princeton, NJ: Princeton University Press

Cochrane, J. H. (2008). "The Dog That Did Not Bark: A Defense of Return Predictability," *Review of Financial Studies* **21**, 1533–1575.

Fama, E. G. and Gibbons, M. R. (1984). "A Comparison of Inflation Forecasts," *Journal of Monetary Economics* 13, 327–348.

Fama, E. F. and French, K. R. (1989). "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics* **25**, 23–49.

Flannery, M. J. and Protopapadakis, A. A. (2002). "Macroeconomic Factors Do Influence Aggregate Stock Returns," *Review of Financial Studies* **15**, 751–782.

Hull, B. and Qiao, X. (2017). "A Practitioner's Defense of Return Predictability," *Journal of Portfolio Management* **43**, 60–76.

Jiang, F., Lee, J. A., Martin, X., and Zhou, G. (2015). "Manager Sentiment and Stock Returns."

Jones, C. S., and Tuzel, S. (2012). "New Orders and Asset Prices." *Review of Financial Studies* **26**, 115–157.

Ledoit, O. and Wolf, W. (2008). "Robust Performance Hypothesis Testing with the Sharpe Ratio," *Journal of Empirical Finance* **15**, 850–859.

Lindeman, R. H., Merenda, P. F., and Gold, R. Z. (1980). *Introduction to Bivariate and Multivariate Analysis*. Glenview IL: Scott, Foresman.

Lown, C. S., Morgan, D. P., and Rohatgi, S. (2000). "Listening to Loan Officers: The Impact of Commercial Credit Standards on Lending and Output," *Federal Reserve Bank of New York Economic Policy Review* **6**, 1.

Markowitz, H. (1952). "Portfolio Selection," *Journal of Finance* **7**, 77–91.

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). "Time Series Momentum," *Journal of Financial Economics* **104**, 228–250.

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy," *Review of Financial Studies* **23**, 821–862.

Shiller, R. J. (2000). *Irrational Exuberance*, Princeton University Press.

Timmermann, A. (2006). "Forecast Combinations," *Handbook of Economic Forecasting* **1**, 135–196.

Welch, I. and Goyal, A. (2008). "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies* **21**, 1455–1508.

Keywords: Equity premium; forecasting; predictability; market timing; asset returns; tactical asset allocation