Predicting the U.S. Stock Market Return: Evidence from the Improved Augmented Regression Method

Jurdi, Doureige and Kim, Jae

La Trobe University, Department of Economics and Finance

20 May 2019

Predicting the U.S. Stock Market Return: Evidence from the Improved Augmented Regression Method

Doureige Jurdi*, Jae Kim

Department of Economics and Finance
La Trobe University - Australia

Abstract

We examine whether the stock market return is predictable from a range of financial indicators and macroeconomic variables, using monthly U.S. data from 1926 to 2012. We adopt the improved augmented regression method for parameter estimation, statistical inference, and out-of-sample forecasting. By employing moving sub-sample windows, we evaluate the time-variation of predictability free from data snooping bias and report changes in predictability dynamics over time. Although we may find statistically significant in-sample predictability from time to time, the associated effect size estimates are fairly small in most cases. We also find weak predictability of the stock market return from multi-step ahead (out-of-sample) forecasts. In addition, we find that mean variance investors realize sporadic economic gains in utility based on predictive regression forecasts relative to naive model historic average forecasts.

1. Introduction

Stock return predictability has been at the center of theoretical and empirical research in finance. The empirical literature on stock return predictability is extensive, with many financial indicators proposed as potential predictors. These predictors include valuation ratios such as the dividend-to-price ratio (Rozeff, 1984; Campbell and Shiller, 1988a; Fama and French, 1988; Hodrick, 1992;

*Corresponding author: Doureige Jurdi. Email: d.jurdi@latrobe.edu.au

Lewellen, 2004), earnings-to-price ratio (Campbell and Shiller, 1988a, 1998b; Lewellen, 2004), and book-to-market ratio (Kothari and Shanken, 1997; Pontiff and Schall, 1998; Lewellen, 2004). Other possible predictors include inflation (Nelson, 1976; Fama and Schwert, 1977 and Campbell and Vuolteenaho 2004), short term interest rates (Campbell, 1987; Hodrick, 1992), and technical indicators (Neely et al. 2014).

The accumulated evidence on in-sample and out-of-sample predictability of stock return is mixed and contentious. Bossaerts and Hillion (1999) investigate predictability in international stock markets and conclude that predictability is mainly an in-sample phenomenon that disappears out-of-sample. They argue that in-sample predictability is due to non-stationarity of predictors and raise concerns about the stability of predictive regression. Campbell and Thompson (2008) argue that U.S. stock returns contain a significant unpredictable component leading to imprecise parameter estimates of the predictive regression. By imposing restrictions on the intercept and slope of the predictive regression and truncating the equity return premium to the positive domain, they demonstrate that the predictive regression forecasts perform better than the historical average. They also show that investors derive positive utility gains by using the predictive regression forecasts. Based on a comprehensive study of S&P500 equity premium using a host of predictors, Welch and Goyal (2008) report there is significant evidence of in-sample predictability, but they find weak out-of-sample evidence. They show that the out-of-sample equity premium forecast based on a bivariate predictive regression fails to outperform the simple average historical forecast. A kitchen sink forecast that includes multiple predictors also fails to beat the historical average forecast. Rapach and Wohar (2006) correct for data mining bias by conducting a bootstrap inferential procedure, and report there is less disparity between in-sample and out-of-sample forecasts in two out of nine predictors used in their study.

A test for return predictability is typically conducted using a predictive regression, where the equity premium is expressed as a linear function of lagged predictors. One of the important econometric issues affecting sound inference based on the predictive regression is endogeneity, which occurs when a predictor is strongly correlated contemporaneously with the equity premium. Stambaugh (1999) shows that the coefficient estimator of the predictor in the predictive regression is biased upwards leading to substantial over-statement of return predictability. Amihud and Hurvish (2004) and Amihud et al. (2008, 2010) propose the augmented regression method (ARM) for bias-reduced estimation and statistical inference in a predictive regression with multiple predictors. They show that the conventional least-squares (LS) method seriously overstates predictability, whereas their ARM provides a more accurate estimate and statistical inference in small samples. Kim (2014) proposes an improved version of the ARM (IARM), conducted with bias-correction of higher accuracy and a stationarity-correction. Monte Carlo simulations in Kim (2014) reveal that IARM has: firstly, better size and power properties for statistical inference than the ARM; and secondly, it generates more accurate out-of-sample forecasts.

The purpose of this paper is to evaluate the predictability of return on the U.S. stock market, using the long monthly data set provided by Amit Goyal . Our predictors include valuation ratios: the dividend-yield, price-earnings ratio, and the book-to-market ratio. We also consider macroeconomic variables: inflation rate, Treasury bill rate, term-spread, and default spread, and moving average technical indicators. Neely et al. (2014) assert that technical indicators have significant in-sample and out-of-sample forecasting power. These indicators strongly capture declining equity return premium near troughs of the business cycle using predictive regression models and the diffusion index approach. Neely et al. (2014) contend that macroeconomic variables complement technical variable by picking up rising equity risk premium near the cyclical trough.

We employ Kims (2014) IARM as a general predictive model in which the stock market return is a function of a lagged predictor of unknown order. To avoid the data-snooping bias and evaluate time-variation of predictability, a moving-sample window of 10 years is used for model estimation and statistical inference. For the evaluation of out-of-sample forecasts, a range of different window lengths is adopted. We contribute to the extant literature in the following ways. First, the IARM has so far found limited empirical applications, and this is the first study applying the IARM to a long U.S. data set with a range of predictors. Many past studies draw inferences about predictability without bias-correction, which may lead to spurious statistical significance. In addition, they often assume a lag order of 1, which may result in model mis-specification. The IARM provides bias-corrected estimation and inference for a predictive regression of a lag order higher than one. Second, we employ moving sample windows, which effectively capture the dynamics of stock market returns that are regularly affected by a range of shocks including business cycles and macroeconomic events. Previous studies estimate the predictive regression over the whole sample period or by using subsamples, which dampens the return dynamics. Third, we evaluate out-of-sample predictability based on multi-step ahead forecasts, in contrast to most past studies that restrict their attention to out-of-sample predictability of horizon 1. We also pay attention to the effect size of return predictability to assess the magnitude of predictability. We consider both univariate and multivariate predictive models, in view of the empirical evidence that the stock returns are more predictable with multivariate models (see Ang and Bekaert, 2007).

Our analysis reveals that the evidence of predictability is rather weak, for both in-sample and out-of-sample. Concerning all predictors considered here, statistically significant in-sample predictability occurs only temporarily and sporadically. The effect size estimates of in-sample predictability are also fairly small in most cases. Similarly, we find little evidence for out-of-sample predictability and inconsistencies in reported utility gains. The results from multivariate models are found to be consistent with those of univariate models. Finally, we report that the accuracy of out-of-sample forecasts deteriorates as the forecasting window size increases, which further supports our findings of little predictability of stock return.

The paper is organized as follows: Section 2 presents the IARM of Kim (2014), and Section 3 provides the data details. Section 4 presents the empirical results, and Section 5 concludes.

2. Methodology

In this section, we present the IARM developed by Kim (2014) and forecast evaluation methods employed. We also discuss the methods of evaluating the economic gain of out-of-sample forecasts. For brevity, we present the model and discuss its main features, since the full details are available from the references given.

Predictive Regression A simple predictive regression model for asset returns is written as

$$y_t = \alpha + \beta_1 X_{t-1} + u_t \tag{1}$$

$$X_t = \theta + \rho_1 X_{t-1} + v_t \tag{2}$$

where y_t is typically excess stock returns, X_t is a predictor and $(u, v)'$ is a vector of error terms. It is assumed that the error terms are independent and identically distributed as a bivariate normal with covariance matrix $\sum = vech(\sigma_u^2, \sigma_{u,v}, \sigma_v^2)$. Under $H_0 : \beta_1 = 0$, the predictor X has no predictive power of y_t. Stambaugh (1999) shows that the OLS estimator of β_1 is biased in a finite sample. This is because the bias in the estimation of ρ_1 carries over to the estimation of β_1 through the covariation between the two error terms $\sigma_{u,v}$. The bias formula Stambaugh (1999) derived indicates that the bias is larger when the predictor X is more persistent, when the contemporaneous correlation between the two error terms is higher, and when the sample size n is smaller. Stambaughs (1999) bias formula suggests a bias-corrected estimator for β_1 as

$$\hat{\beta}_s = \hat{\beta}_1 + \frac{\sigma_{u,v}}{\sigma_v^2} \frac{(1 + 3\widehat{\rho_1})}{n}. \tag{3}$$

Amihud et al. (2010) propose the augmented regression method (ARM) for bias-corrected estimation and inference, extending the earlier studies by Amihud and Hurvich (2004) and Amihud et al. (2008). They consider the case of AR(p) predictor given as

$$y_t = \alpha + \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p} + u_t \tag{4}$$

$$X_{1,t} = \theta_1 + \rho_{11} X_{1,t-1} + \cdots + \rho_{1,p} X_{1,t-p} + v_{1,t} \tag{5}$$

with an assumption that the two error terms are related as $u_t = \phi v_{1,t} + e_t$ where e_t is independent of u_t and v_t. The equation (4) is then written as

$$y_t = \alpha + \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p} + \phi' v_{1,t} + e_t \tag{6}$$

The ARM is conducted by augmenting the regression (4), as in (6), with the bias-corrected residuals from (5). Amihud et al. (2010) construct a proxy v_t^c for v_t as

$$v_t^c = X_t - \hat{\theta^c} - \hat{\rho_1^c} X_{t-1} - \cdots - \hat{\rho_p^c} X_{t-p}$$

where $\hat{\theta^c}$ and $\hat{\rho_1^c}$ are the bias-corrected estimators of $\hat{\theta}$ and $\hat{\rho}_1$ based on the Stine and Shaman (1989) formulae. The bias-corrected estimator $\hat{\beta}^c \equiv (\hat{\beta}_1^c, \ldots, \hat{\beta}_p^c, \hat{\phi}^c)$ for $\beta \equiv (\beta_1, \ldots, \beta_p, \phi)$ is obtained from OLS regression of (6) using v_t^c in place of v_t. Amihud et al. (2010) also provide a formula for the covariance estimator for β^c, based on which the null hypothesis of no in-sample predictability of y ($H_0 : \beta_1 = \cdots = \beta p = 0$) is tested.

Kims (2014) IARM improves the ARM in three ways. First, an improved bias-corrected estimator $\hat{\rho_i^c}$s for ρ_is is adopted, which is of higher order accuracy than the one used by Amihud et al. (2010). The bias-corrected estimator converges to the true values at the rate of n^{-2} which is faster than the rate of convergence in Amihud et al. (2010) which is of the order n^{-1}. Second, the IARM adopts a matrix formula which makes bias-correction and covariance matrix estimation computationally simpler for a higher order case. Third, the IARM implements Killians (1998) stationarity-correction to ensure that bias-corrected estimators of ρs always satisfy the condition of stationarity. Without this correction, the bias-corrected estimators for $\rho_i s$ may render the predictive model explosive (see Lewellen, 2004). Kim (2014) presents Monte Carlo results showing that the IARM has better size and power properties in small samples than the original version of the ARM.

Forecast Evaluation

The predictive regression model given in (4) and (5) is able to generate out-of-sample dynamic forecasts for future stock market returns. Kim (2014) provides the Monte Carlo evidence that the IARM forecasts are more accurate than those generated from OLS and ARM estimation. In this paper, we evaluate the accuracy of the IARM forecasts for stock return using the Theil-U statistic, which is given by

$$U = \frac{\sqrt{\sum_{t=1}^{h}(y_t - \hat{y}_t)^2}}{\sqrt{\sum_{t=1}^{h} y_t^2} + \sqrt{\sum_{t=1}^{h} \hat{y}_t^2}}$$

where y_t and $\hat{y}_t$ denote the stock market return realization and its IARM forecasts respectively, while h is the out-of-sample forecasting period. The statistic $U = 0$ when the forecast is perfect $y_t = \hat{y}_t$ for all t and U = 1 when $\hat{y}_t = 0$ for all t (crude or naive forecasts). Jordan et al. (2014) note that the Theil's U is closely related to the out-of-sample R^2, which is the measure of forecast accuracy used in previous studies. That is, the out-of-sample $R^2 = 1 - U^2$.

We then examine the economic significance of stock market return forecasts by evaluating utility gains of a mean-variance investor who allocates her portfolio between a risk-free investment in Treasury bills and the stock market on a monthly basis. Recent studies including Campbell and Thompson (2008), Rapach et al. (2010) and Neely et al. (2014) evaluate the economic utility of equity premium forecasts and report significant utility gains for a mean-variance investor. The importance of evaluating economic gains is discussed in Cenesizoglu and Timmermann (2012), who argue that economic value measures provide additional insights about forecast evaluation beyond statistical measures because out-of-sample forecasts (mean squared forecast error measures) of the equity premium are not strongly correlated with investor utility. To evaluate the utility gained from employing the predictive regression model forecasts against a historic average model, the investors utility is given as

$$U(r_p) = E(r_p) - \frac{1}{2}\sigma_p^2$$

where r_p and ${}^2_p$ are the mean and variance of the return of the portfolio. The investor forecasts the stock market return using the predictive regression and decides to allocate ${}_{PR,t}$ to the stock market at the end of period t into period t+1 as

$$\omega_{PR,t} = \left(\frac{1}{\gamma}\right)\left(\frac{\hat{r}_{t+1}}{\hat{\sigma}^2_{t+1}}\right)$$

where $\hat{r}_{t+1}$ is the forecasted stock market return using (4) estimated using Kim (2014) IARM and $\hat{\sigma^2_{t+1}}$ is the forecast of the variance of stock returns calculated using a rolling window of 10 years (120 months). Instead, the investor may allocate $\omega_{HA,t}$ using forecasts from a historical average model as

$$\omega_{HA,t} = \left(\frac{1}{\gamma}\right)\left(\frac{\bar{r}_{t+1}}{\hat{\sigma}^2_{t+1}}\right)$$

where $\bar{r}_{t+1}$ is the historic average of the stock market return. The investors utility gain is the certainly equivalent annualized return given as

$$CER_{p,gain} = CER_{p,PR} - CER_{p,HA} \tag{7}$$

where $CER_p = \hat{u}_p - \frac{1}{2}\gamma\hat{\sigma^2_p}$, u_p and $\hat{\sigma^2_p}$ are the estimates of the sample mean and variance of portfolio returns. $CER_{p,gain}$ in (7) is the difference between the utility gain (Certainty Equivalent Return) ensuing to using the predictive regression forecast relative to the historical average forecast in the asset allocation decision. The value of $CER_{p,gain}$ represents portfolio management fees that the investor is willing to pay in order to obtain access to forecasts generated by the IARM model relative to the information made available by the historical average model.

3. Data and Computational Details

As mentioned earlier, we use the data compiled by Amit Goyal, monthly from January 1926 to December 2012. The stock market return is calculated as the difference between the S&P500 continuously compounded returns (including dividends) and the risk-free rate. The variables used as predictors are:

(1) Dividend yield (DY): difference between the log of the 12 months moving sum of aggregate dividends paid on the S&P500 constituents and the log of the lagged prices of the S&P500.

(2) Price earnings ratio (PE): difference between the log of prices and log of 12 months moving sum of earnings.

(3) Book-to-market ratio (BM): ratio of book value to market value for the Dow Jones Industrial Average (DJIA).

(4) Term spread (TERM): difference between the long-term yield on government bonds and yield on three months Treasury bills.

(5) Default spread (DEF): difference between BBB and AAA-rated corporate bond yield.

(6) Inflation (INF): growth in one month Consumer Price Ratio with one month lag.

(7) Risk-free rate (INT): interest rate on a three-month Treasury bills.

Further details on the data sources and how these predictors are constructed are available in Welch and Goyal (2008).

The financial ratios (DY, BM, PE) are commonly used by investors and financial analysts to make investment recommendations and financial decisions. According to the mispricing view, when stocks are under-priced (over-priced), these ratios (DY, BM) are high (low) because they have the market price in the denominator and they predict high (low) returns. The PE ratio carries the opposite interpretation because the market price is the numerator. Rational asset pricing theory provides a different view. The theory suggests that financial ratios predict returns because they capture information about the risk premium and track stochastic discount rates (for discussion see Lewellen, 2004). Campbell and Shiller (1998, 2001) show that low market dividend yield and high market P/E ratio reliably predict declines in stock prices. Unlike dividends, which tend to be persistent, earnings are subject to sharp surges and declines. Siegel (2002) shows that an increase in the PE ratio due to market optimism- i.e., a sharp increase in prices - delivers lower future real return than a sharp decline in earnings; however, an increase in the PE ratio that is deteriorating earnings are followed by average stock returns. Pontiff and Schall (1998) explain that the BMs ability to predict aggregate stock returns is because the book value proxies for future cash flow. Fama and French (1992) argue that systematic differences in average stock returns are due to risks that are captured by the BM ratio.

In addition to these valuation ratios, we use the Treasury bill rate (INT) and inflation (INF) as potential predictors. Fama and Schwert (1977) and Ferson (1989) detect a negative relationship between short-term Treasury bills and stock market returns. Inflation affects stock returns in two ways. Firstly, a rise in inflation increases the discount rate resulting in lower stock returns. Secondly, it increases future dividend leading to higher stock returns. Campbell and Shiller (1988b) argue that the second effect prevails in the long run.[1] Boudoukh and Richardson (1993) find that stock returns (nominal or real) and inflation are negatively related. They report that the relationship holds at the 5-year horizon but not the 1-year horizon.

We also use a moving average technical indicator as a predictor. This measure is constructed as the relative difference between two moving average series of length S and L calculated at time t using the most recent price observation

[1] Several studies empirically find a negative correlation between inflation and stock market return. This evidence is anomalous because it goes against Fishers hypothesis (Fisher, 1930). Modigliani and Cohn (1979) attribute this anomaly to investors money illusion. Fama (1981) explains this anomaly by proposing the proxy hypothesis, which predicts that the decline in real economic activity reflected by the stock market is induced by inflation. Geske and Roll (1983) propose a reverse causality explanation and argue that declines in real economic activity increase fiscal deficit and Fed money supply leading to an increase in inflation. Campbell and Vuolteenaho (2004) use Campbell and Shillers (1988) log-linear dynamic valuation framework to show that high inflation is positively correlated with long-term real dividend growth and mispricing. Others contend that the relationship is due to changes in expected returns and risk aversion (see Blanchard, 1993; Fama and French, 2002, and others).

as

$$R_t = \frac{\delta_{t,S} - \delta_{t,L}}{\delta_{t,L}} \tag{8}$$

where $\delta_{t,S} = \frac{1}{S}\sum_{j=1}^{S} p_{t-S+j}$ and $\delta_{t,L} = \frac{1}{L}\sum_{j=1}^{L} p_{t-L+j}$ are the short-term and long-term moving average measures, respectively; and p_t is the observed index level at time t. $\delta_{t,S}$ and $\delta_{t,L}$ are constructed using daily price observations obtained from the Bloomberg database from 1927 to 2012. R_t is calculated and monthly time series is generated and matched with Goyals dataset. According to moving average trading rules, a rise (decline) in the index price level is predictable when $\delta_{t,S} > \delta_{t,L}(\delta_{t,S} < \delta_{t,L})$.

For in-sample testing, we employ a moving sub-sample window of 10 years (120 observations), which moves every month . That is, we take the sample of the first 10 years from January 1926 to December 1935 and conduct the IARM estimation and test for no predictability. We then move one month forward to conduct the test for the sample from February 1926 to January 1936. This continues to the end of the whole data set. This approachs main advantage is that we can observe time-variation of return predictability. In addition, according to Hsu and Kuan (2005, p. 608), it serves as an effective guard against data snooping bias, and an alternative to the reality check test of White (2000). In each sub-sample window, the lag order p in (4) and (5) is treated as unknown and is selected using Akaikes information criterion with the maximum order set to 6. We also generate 12-step ahead out-of-sample dynamic forecasts from each window and report the U statistics and utility gain forecasts discussed in the previous section. For forecasting, we employ a range of different window lengths, ranging from 6 months to 2 years.

4. Empirical Results

For the simplicity of exposition, we present and discuss the results of the valuation ratios (DY, PE and BM) and the moving average technical indicator since other predictors (INF, TERM, and DEF) exhibit qualitatively similar results. For each predictor considered, we report three time-series plots. The first plot reports the p-value for the F-test of $H_{01} : \beta_1 = \quad = \beta_p = 0$; the second the effect size estimate of $\beta \equiv \beta_1 + ... + \beta_p$, and its 95% confidence interval; and the third the U statistic. Note that the confidence interval for β is determined by using the bias-reduced covariance matrix for $\hat{\beta}^c \equiv (\hat{\beta}_1^c, \ldots, \hat{\beta}_p^c, \varphi^c)$ detailed in Kim (2014). These statistics are plotted from January 1936 (corresponding to the first window of January 1926 to December 1935) to January 2013 (corresponding to the last window of January 2003 to December 2012). In the first graph, the p-values are plotted with three horizontal lines representing the conventional levels of significance of 0.01, 0.05, and 0.10. In the second, the effect size estimates and their 95% confidence intervals are plotted with the horizontal line of 0: if the interval does not cover 0, the null hypothesis that effect size is 0 is rejected at 5% level of significance. In the third, the U statistics are plotted with the horizontal line of 0.7, which approximately corresponds to the out-of-sample R^2 value of 0.5. Figure 1 reports the case of the dividend-yield. The top plot of Figure 1 shows that the null hypothesis of no predictability is often rejected at the 5% significance level through out the sample period. By allowing for a moving sub-sample window, we can identify periods of predictability in the time series. The in-sample evidence shows that the null hypothesis of no predictability is rejected during various phases of the time-series. These phases occur during the mid-2000s, the mid-1990s, the late 1960s and in early 1970s. This indicates that the information content of the dividend yield is not always relevant to predicting aggregate stock market returns. The total effect size of DY is time-varying. It has a range between -3.54% to 18.78% over the estimation period. To assess the out-of-sample predictability of DY, we plot the U statistic in the bottom plot of Figure 1. The plot shows that U values are close to unity during most phases of the time series. There are few exceptions when U is less than 0.7. This indicates that the out-of-sample forecasts based on IARM may not be better than those generated by a naive forecasting model during various phases of the time-series.

Figure 2 reports the case of the PE ratio. The first plot of Figure 2 shows that the null hypothesis of no predictability is rejected at different periods of the sample. For example, at the 5% significance level, the null hypothesis is rejected prior to 1940, and between 2008 and 2010. In the first period, the effect size ranges between -15% and -5% whereas the second period the effect size is close to zero. The plot of the U statistic rarely goes below 0.7 during phases of the time series.

Figure 3 reports the case of the BM ratio. In this case, in-sample predictability persists from the 1930s till the late 1940s with an effect size ranging from 1% to 5%. After 2008, the effect size is between 1% and 10%. We also identify other episodes of in-sample predictability occur sporadically during the sample period. Like other financial predictors already discussed, out-of-sample forecasts based on the BM ratio are sporadic and weak.

Figure 4 reports the case of the relative moving average method in (8) using S=1 and L=50. The p-values of the no predictability hypothesis show irregular yet extended periods of in-sample predictability with a relatively larger effect size than any other valuation ratio considered in this study. The range of the effect size in plot 2 is from-366% to 249%. However, the out-of-sample predictability evidence is weak as shown by the U statistic plot. This outcome contrasts with findings of Neely et al. (2014) who report strong evidence of the out-of-sample predictive ability of technical indicators. [2]

Overall, we find evidence of time-varying predictability which is stronger in-sample that out-of-sample. The results, however, agree with other findings in the literature which question the out-of-sample predictability based on the estimation of predictive models using the full sample period or on dividing the sample period into sub-periods (Welch and Goyal, 2007).

We then compute the utility gain (Certainty Equivalent Return) that a

[2]We note that the analysis using technical variables in our study is limited compared to that of Neely et al. (2014). This is because the IARM cannot be conducted with a dummy variable (for buy and sell signal) as a predictor. Other continuous technical indicators may be used with the IARM, but we are limited by data availability (e.g., volume). We, therefore, leave this potential line of investigation to future research.

mean-variance investor with a risk aversion coefficient of 3 would obtain by relying on excess return forecasts generated by (4) relative to the historical average model. Monthly utility gains are computed using moving sub-sample window of 10 years (120 observations), which moves every month. Similar to Campbell and Thompson (2008), the model restricts the investor from shorting stocks or taking more than 50% leverage. Figure 5 plots the monthly utility gains (CER) obtained by (7). In all three individual cases that use valuation ratios to predict the stock market return, the financial predictors fail to consistently deliver positive utility gains throughout the sample period. Nonetheless, there are occasional realizations of positive utility gains of up to 0.4%, indicating that the investor would be willing to pay up to 40 basis points to have access to the forecasts generated by the IARM relative to the historical average forecast. The utility gain results obtained by the moving average predictor are qualitatively similar to other predictors despite the larger magnitude of spikes in utility gains across the sample period.

To examine whether the use of multivariate models can improve predictability, we consider the model of the form

$$\begin{aligned} y_t &= \alpha + \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p} + \upsilon_t \\ X_{1,t} &= \theta_1 + \rho_{11} X_{1,t-1} + \cdots + \rho_{1,p} X_{1,t-p} + \nu_{1,t} \\ X_{2,t} &= \theta_2 + \rho_{21} X_{2,t-1} + \cdots + \rho_{2,p} X_{2,t-p} + \nu_{2,t} \\ X_{3,t} &= \theta_3 + \rho_{31} X_{3,t-1} + \cdots + \rho_{3,p} X_{3,t-p} + \nu_{3,t} \end{aligned} \tag{9}$$

where X_1 denotes a fundamental (one of DY, PE, BM), and X_2 and X_3 denote the interest rate and inflation rate, respectively. The IARM for the multivariate model of the above form can be conducted as an extension to the univariate case detailed in Section 2 (for details, see Kim, 2014). The results are presented in Figure 6. In comparison with those reported in Figures 1-3, the multivariate models show similar results, where time variation of the return predictability of DY, PE, and BM are similar to those observed in the univariate case. We note that similar results are obtained for the effect size, forecast accuracy, and economic gains.

Finally, as a further check on the predictability of the stock market return, we examine how forecast accuracy changes under different window lengths. In Figure 7, we plot the mean of Theil's U values using window lengths ranging from 6 months to 20 years. Predictive regression models are fitted, and forecasts are generated from the estimated predictive model with h = 12 for all cases, except when the window length is 6 months where the mean return of the past 6 months are used as forecasts. There is a strong tendency that the forecast accuracy improves as the length of the rolling window decreases. For all cases, except for PE, the smallest mean U values are achieved when the window length is 6 months with the mean return as the forecasts. This means that the forecasts from the predictive model are found to be inferior to the naive forecasts, which further supports the evidence of weak predictability of U.S. stock return from a range of predictors.

5. Conclusion

This paper examines whether the U.S. stock market return (equity premium) is predictable from a range of financial and economic predictors using monthly data from 1926. We employ the improved augmented regression method of Kim (2014), which is an improved version of the method proposed earlier by Amihud et al. (2008, 2010). The method corrects for the small sample bias (the Stambaugh bias) in the estimation of predictive coefficients, providing improved statistical inference and out-of-sample forecasting in small samples. The method is applied to a general predictive model where the unknown lag order is allowed to take a value higher order than 1 and estimated using an information criterion. We test the in-sample predictive ability of stock return using moving sub-sample windows of 10 years and evaluate time variation of predictability free from possible data-snooping bias. We also assess out-of-sample forecast accuracy of a range of financial, economic and technical indicator variables for the U.S. stock returns, using sub-sample windows of different lengths.

Our results show time-varying evidence of predictability both in-sample and out-of-sample. The total size effect varies significantly over the time-series. The out-of-sample forecast accuracy is found to be low for all predictors considered. Hence, forecasts based on the IARM are unable to beat forecasts derived from a naive method during various phases of the time-series. Similar findings are also observed for multivariate predictive regressions. Mean-variance investors however realize economic gains in utility using the IARM forecasts relative to naive model historic average forecasts during various phases of the times-series. Overall, our results are relatively consistent with those of Welch and Goyal (2008), who find weak evidence of in-sample and out-of-sample predictability for the U.S. stock market. We note that our results of little predictive ability may be attributable to the correction of the endogeneity bias in a predictive model of a general order, which has not been extensively considered in previous studies.

References

Amihud, Y., & Hurvich, C. M. (2004). Predictive regression: a reduced-bias estimation method. Journal of Financial and Quantitative Analysis, 39(4), 813-841.

Amihud, Y., Hurvich, C. M., & Wang, Y. (2008). Multiple-predictor regressions: hypothesis testing. Review of Financial Studies, 22, 413-434.

Amihud, Y., Hurvich, C. M., & Wang, Y. (2010). Predictive regression with order-p autoregressive predictors. Journal of Empirical Finance, 17, 513-525.

Ang, A., & Bekaert, G. (2007). Stock return predictability: is it there? Review of Financial Studies, 20, 651707.

Blanchard, O. J. (1993). Movements in the equity premium. Brookings Papers on Economic Activity, 2, 75118.

Bossaerts, P., & Hillion, P. (1999). Implementing statistical criteria to select return forecasting models: What do we learn? Review of Financial Studies, 12, 405-428.

Campbell, J. Y. (1987). Stock returns and the term structure. Journal of Financial Economics, 18, 373-399.

Campbell, J., & Shiller, R. J. (1998). Valuation ratios and the long-run stock market outlook. The Journal of Portfolio Management, 24, 11-26.

Campbell, J. Y., & Shiller, R. J. (1988a). The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies, 1, 195-228.

Campbell, J. Y., & Shiller, R. J. (1988b). Stock prices, earnings, and expected dividends. Journal of Finance, 43, 661676.

Campbell, J., & Shiller, R. J. (2001). Valuation ratios and the long-run stock market outlook: An Update. Cowles Foundation Discussion Paper No. 1295. Yale University.

Campbell, J., & Thompson, S. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? The Review of Financial Studies, 21(4), 1509-1531.

Campbell, J., & Vuolteenaho, T. (2004). Inflation illusion and stock prices. American Economic Review, Papers and Proceedings, 94, 19-23.

Cenesizoglu, T., & Timmermann, A. (2012). Do return prediction models add economic value? Journal of Banking & Finance, 36(11), 2974-2987.

Fama, E. (1981). Stock returns, real activity, inflation, and money. American Economic Review, 71, 545-565.

Fama, E., & French, K. (2002). The equity premium. Journal of Finance, 57, 637-659. Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. Journal of Financial Economics, 22, 3-25.

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

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. Journal of Finance, 47, 427-465.

Fama, E., & Schwert, G. (1977). Asset returns and inflation. Journal of Financial Economics, 5, 115-146.

Fisher I. (1930). The Theory of Interest, Macmillan, New York.

Geske, R., & Roll, R. (1983). The fiscal and monetary linkage between stock returns and inflation. Journal of Finance, 38, 1-34.

Goyal, A., & Welch, I. (2003). Predicting the equity premium with dividend ratios. Management Science, 46, 639-645.

Hodrick, R. J. (1992). Dividend yields and expected stock returns: alternative procedures for inference and measurement. Review of Financial Studies, 5(3), 357-386.

Hsu, P. O., & Kuan, K. C. (2005). Re-examining profitability of technical analysis with data snooping checks. Journal of Financial Econometrics, 3, 606-628.

Jordan, S.J., Vivian, A.J., & Wohar, M.E. (2014). Forecasting returns: New European evidence. Journal of Empirical Finance, 26, 76-95.

Kilian, L. (1998). Small sample confidence intervals for impulse response functions. Review of Economics and Statistics, 80, 218230.

Kim, J. H. (2014). Predictive regression: An improved augmented regression method. Journal of Empirical Finance, 26, 13-25.

Kim, J. H. (2014). VAR.etp: VAR modelling: estimation, testing, and prediction. R package version 0.7. URL http://CRAN.R-project.org/package=VAR.etp

Kothari, S. P., & Shanken, J. (1997). Book-to-market, dividend yield, and expected market returns: a time series analysis. Journal of Financial Economics, 44, 169-203.

Lewellen, J. (2004). Predicting returns with financial ratios. Journal of Financial Economics, 74, 209-235.

Modigliani, F., & Cohn, R. (1979). Inflation, rational valuation and the market. Financial Analysts Journal, 35, 24-44.

Neely, C., Rapach, D., Tu, J., & Guofu, Z. (2014). Forecasting the equity risk premium: The role of technical indicators. Management Science, 60, 1772-1791.

Nelson, C. (1976). Inflation and rates of return on common stocks. Journal of Finance, 31, 471-483.

Pontiff, J., & Schall, L. (1998). Book-to-market ratios as predictors of market returns. Journal of Financial Economics, 49, 141160.

R Core Team (2014). R: A language and environment for statistical computing.

R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rapach, D. E., & Wohar, M. E. (2006). In-sample vs. out-of-sample tests of stock return predictability in the context of data mining. Journal of Empirical Finance, 13, 231-247.

Rapach D.E., Strauss J., & Zhou G. (2010). Out-of-sample equity premium prediction: combination forecasts and links to the real economy. Review of Financial Studies, 23, 821862.

Rozeff, M. (1984). Dividend yields are equity risk premiums. Journal of Portfolio Management, 11, 68-75.

Stambaugh, R. F. (1999). Predictive regressions. Journal of Financial Economics, 54, 375-421.

Siegel, J. (2002). Stocks for the Long Run (3rd ed.). New York: McGraw-Hill.

Stine, R. A., & Shaman, P. (1989). A fixed-point characterization for bias of autoregressive estimators. The Annals of Statistics, 17(3), 12751284.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies, 21(4), 1455-1508.

White, H. (2000). A reality check for data snooping. Econometrica, 68, 1097-1126.

Figures:

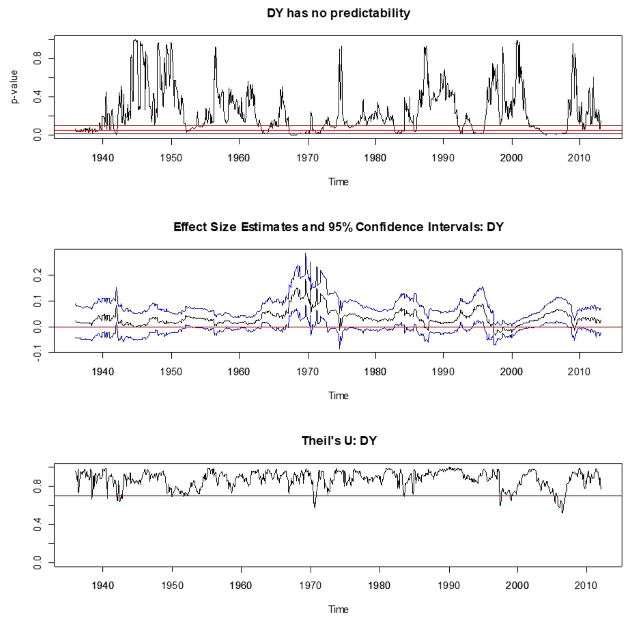

Figure 1: Predictability of Excess Stock Market Return from the DY

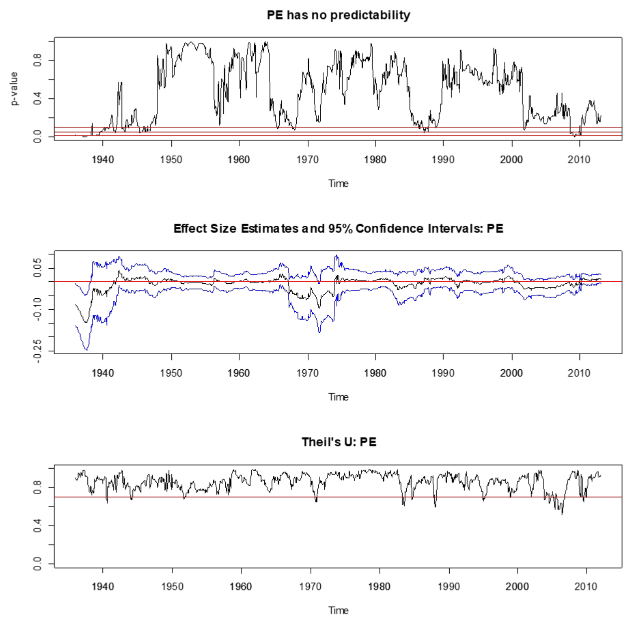

Figure 2: Predictability of Excess Stock Market Return from the PE ratio

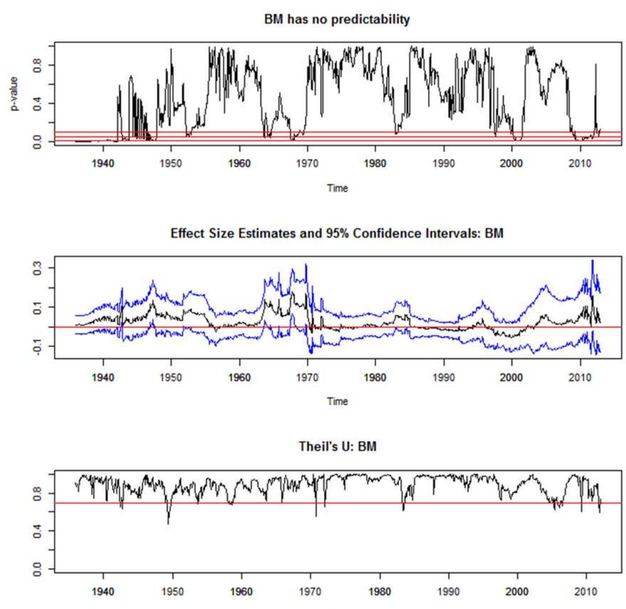

Figure 3: Predictability of Excess Stock Market Return from the BM ratio

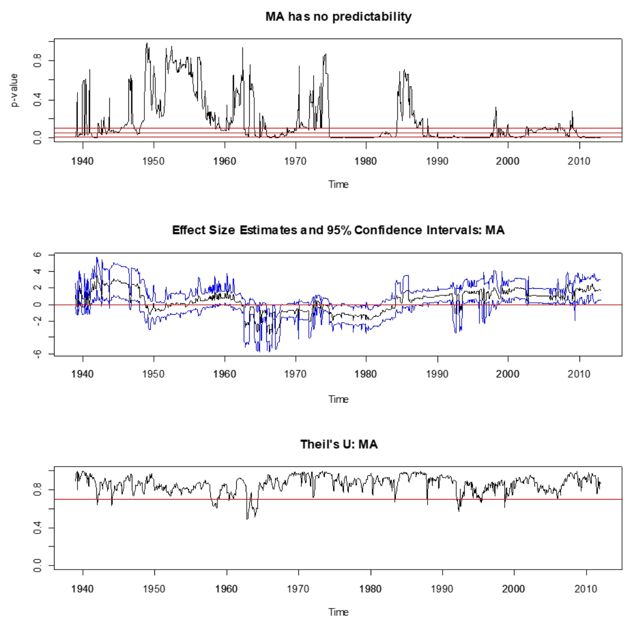

Figure 4: Predictability of Excess Stock Market Return from MA Technical Trading Rule

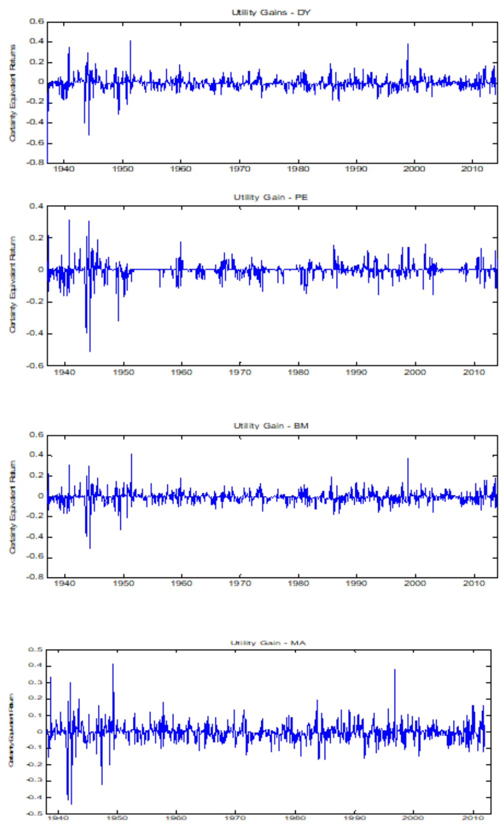

Figure 5: Economic Significance of Excess Stock Market Return Forecast

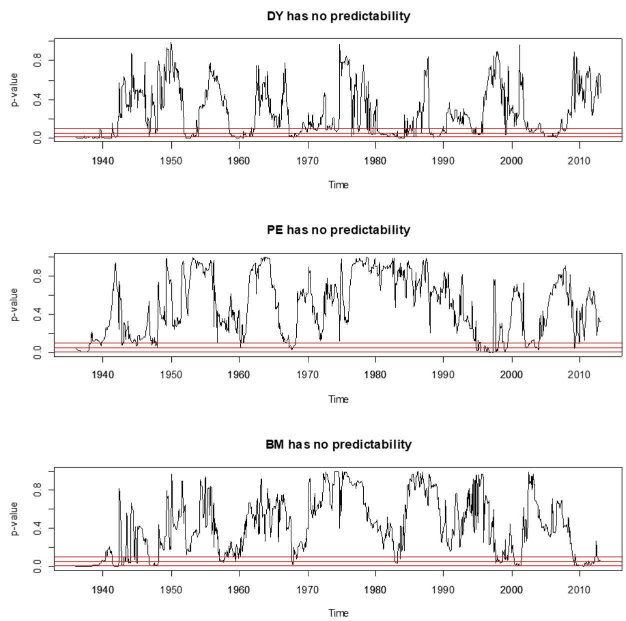

Figure 6: Predictability of Excess Stock Market Return from Multivariate Predictive Model

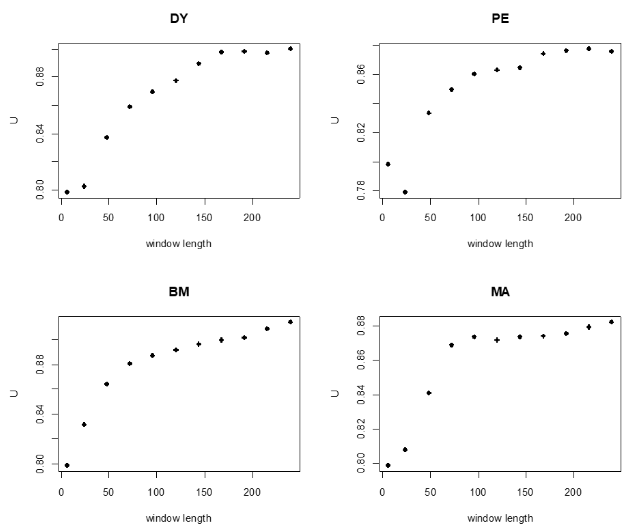

Figure 7: Forecast Accuracy under different rolling window lengths Note: Each graph plots the mean of Theil's U values over the entire sample, under the estimation window lengths (6, 24, 48, 72, 96, 120, 144, 168, 192, 216, 240) in months. When the window length is 6, the predictive regression is not estimated, but forecasts are generated as the mean return over the past six months. For all other cases, the predictive model is fitted in the same way as before. The forecast period h is set at 12 for all cases.