Forecasting Market Returns: Bagging or Combining?

Steven J. Jordan
Econometric Solutions
econometric.solutions@yahoo.com

Andrew Vivian
Loughborough University
a.j.vivian@lboro.ac.uk

Mark E. Wohar
University of Nebraska-Omaha and Loughborough University
mwohar@mail.unomaha.edu

May 17, 2016

Forecasting Market Returns: Bagging or Combining?

Abstract

This paper provides a rigorous and detailed analysis of the methods of bagging, which addresses both model and parameter uncertainty. We provide a multi-country study of bagging, of which there are very few to date, that examines out-of-sample forecasts for the G7 and a broad set of Asian countries. We find that, when portfolio weight restrictions are applied, bagging generally improves forecast accuracy and generates economic gains relative to the benchmark. Bagging also performs well compared to forecast combinations in this setting. We incorporate data mining critical values for appropriate inference on bagging and combination forecast methods. We provide new evidence that the results for bagging cannot be fully explained by data mining concerns. Finally, forecasting gains are highest for countries with high trade openness and high FDI. The potentially substantial economic gains could well be operational given the existence of index funds for most of these countries.

1. INTRODUCTION

Most *out-of-sample* return forecasting evidence rests on major developed countries and is especially focused on US data. The evidence generally suggests that it is difficult to consistently outperform a simple benchmark. For example, Goyal and Welch (2003) state: "By assuming that the equity premium was `like it always has been,' a trader would have performed at least as well in most of our samples." The majority of the international literature considers several macro variables and fundamental predictor variables based on dividends and earnings;[1] it provides mixed evidence on the extent of out-of-sample (OOS) predictability.[2] However, there has been very little prior international evidence on amalgamating information from these different predictor variables (Jordan and Vivian, 2011, is a notable exception that considers one simple forecast combination technique). Our paper addresses this topic in detail using the bagging method that explicitly accounts for model uncertainty and parameter instability.

In the stock return context, given that predictive ability of individual variables is unstable, it is important to control for model uncertainty and parameter instability (Bossaerts and Hillion, 1999; Paye and Timmermann, 2006; Rapach and Wohar, 2006; amongst many others). According to Inoue and Kilian, 2008 p511: "Bagging involves generating a large number of bootstrap resamples of the original forecasting problem, applying a pretest model selection rule to each of the resamples, and averaging the forecasts from the models selected by the pretest on each bootstrap sample." Bagging explicitly accounts for the model uncertainty by allowing the predictive model to change not only over time but also across bootstrap samples for the same time period, by effectively averaging the parameter estimates across bootstrap samples in each time period bagging addresses the parameter instability issue as well. Bagging has recently been implemented for forecasting economic variables with applications to US inflation

---

[1] Bossaerts and Hillion (1999) use fundamental predictors. Rapach, Wohar, and Rangvid (2005) examine macro predictors. Rapach and Wohar (2009) use the dividend-price ratio. Giot and Petitjean (2011) use fundamental predictors. Jordan and Vivian (2011) use fundamental predictors.

[2] Jordan (2012) examines one particular type of time-variation in aggregate returns, long-term reversals. He finds these reversals can be explained by asset pricing models with time-varying risk factors and time-varying alphas; this suggests that macroeconomic factors are important determinants of time variation in equity returns.

(Inoue and Kilian, 2008) and US employment growth (Rapach and Strauss, 2010) but there is, to our knowledge, very little empirical evidence on the effectiveness of bagging for a dataset of international stock market returns nor in a multi-country setting more broadly.[3]

The main contribution of this paper is to provide a rigorous and detailed analysis of the bagging method (Inoue and Kilian, 2008) for forecasting stock returns in Asia and the G7. The bagging method has also been used for binary prediction problems where Lee and Yang (2006) considered an asymmetric loss function with an application to forecasting the sign of the US stock market return. The branch of bagging examined by Inoue and Kilian (2008) and followed in this paper applies to linear regression models with continuous variables (predictors and predictand) where it is designed to reduce mean squared forecast errors. Results of the bagging method relative to the no-predictability benchmark (a random-walk with drift model) [4] are compared to the performance of forecast combination methods which have received little attention in stock return forecasting applications until recent studies of the US (Rapach, Strauss and Zhou, 2010) and the G7 countries (Jordan and Vivian, 2011).[5] We extend prior literature on bagging in important dimensions. Firstly, we incorporate data mining critical values for appropriate inference on bagging (and combination forecast methods), which may be particularly important for the set of G7 countries. Secondly, we estimate the utility gain that can be derived from bagging continuous variables. Thirdly, we provide comprehensive evidence on bagging from 17 countries compared to prior single country studies; this also enables us to examine cross-country determinants of forecast ability.

First, we examine whether or not data mining can account for the evidence of predictability that we report? Prior literature focuses on the G7 applying bivariate regressions multiple times. Rapach and Wohar (2006) demonstrate that data mining can partly account for evidence on aggregate return

---

[3] Bagging was also implemented to estimate the appropriate restrictions on univariate US stock return predictability regressions by Hillebrand et al., (2013) but this setting is rather different to ours where we are using it in a multivariate predictive setting to tackle model uncertainty in an analogous way to Inoue and Kilian (2008) and Rapach and Strauss (2010).

[4] We utilize an historical average benchmark to be consistent with prior literature. Robustness results using the AR(p) benchmark provide similar results to the historical average.

[5] Forecast combination methods have been well established as effective in improving forecast accuracy in many disparate applications (Clemen, 1989), but perhaps surprisingly received little attention for stock returns until recently.

forecastability in the US. This suggests data mining could be a concern. The data mining issue is more acute for the G7 countries as these countries have been investigated more often than the Asian countries.

Second, we investigate whether or not bagging generates utility gains in an asset allocation exercise. Prior empirical investigations of bagging to US macroeconomic series do not examine the economic value of bagging forecasts, while contemporaneous work does so in single country settings (e.g. Jordan et al., 2016 for Canadian industries in terms of a sector rotation strategy). From a practical perspective, our results suggest implementing equity index trading strategies based upon the new bagging method in medium and smaller markets could help investors to time-vary their portfolio allocations between debt and equity. Hence, we provide new evidence that bagging can generate economic value which is robust to reasonable trading costs and cannot be fully accounted for by data mining concerns.

Third, we examine aggregate stock return forecasting for 11 Asian countries and the G7 (Japan is in both groups) compared to prior international studies of equity market return predictability that generally utilize a similar set of large economies consisting of the major developed countries.[6] Major developed countries stock returns are highly correlated with each other; for example in the G7 monthly correlations are about 0.7 amongst country pairs. This suggests that data from these countries is only partly independent from each other and may reflect one (or a small number of) common effect(s).[7] In contrast, Asian financial markets have much lower correlations with each other and thus broaden evidence compared to a sample exclusively of G7 countries. Moreover, Asian economies are of global interest given that they produce about 30% of global economic output; make up over 40% of the world population, are (generally) growing rapidly, and whose financial markets are emerging as an important

---

[6] Bossaerts and Hillion (1999) use data from 1969-1995 for G7 and 7 other developed European countries. Rapach, Wohar, and Rangvid (2005) use data from mid 1970s to late 1990s for G7 and 5 other developed European countries. Rapach and Wohar (2009) study the G7 countries. Giot and Petitjean (2011) use data starting from early 1950s until 2005 for 10 developed countries. Jordan and Vivian (2011) use data from 1927-2009 for 7 developed countries. McMillan and Wohar (2011) use less sophisticated measures of economic value but include the G-7 countries and 4 Asian. Guidolin, Hyde, McMillan and Ono (2009) study the G-7.

[7] Guidolin et al. (2009) find G7 countries return forecasting performance could be broadly split into two groups those that are Anglo-Saxon and those which are Continental European. Thus, results that hold for G7 markets may not transfer well to other markets and thus examining countries outside of Anglo-Saxon and Continental European warrants investigation.

investment class despite their higher volatility of stock returns and of economic fundamentals.[8] Asian countries also differ in terms of economic, institutional, and cultural characteristics in comparison to the US and other major developed countries.[9] Forecast accuracy for a range of different predictands varies across countries with differences in such characteristics (e.g. advanced versus developing countries, Loungani, 2001; information disclosure and cultural traits, Bilinski et al., 2013; legal setting, Chen et al., 2010), although rarely is the channel or mechanism through which this occurs formally outlined (notable exceptions include Dovern et al., 2015; Loungani et al., 2013 who link forecast accuracy of professional economists to information rigidities and efficient use of information). In terms of a stock return application the characteristics outlined can impact the speed of the market response, e.g. market development, information frictions and equity market liquidity should all impact how quickly news is incorporated into stock prices.[10] The greater the delay in reaction to news the more accurately the next periods return can be forecast. Whilst this area of analysis is at an early stage it still enables us to further our understanding beyond that provided in prior analysis that considered alternative settings.[11] By providing analysis on both Asian and G7 markets, we can provide comprehensive evidence on bagging and whether country characteristics impact aggregate return forecastability across countries.

2. DATA DESCRIPTION

We examine two samples of countries: (1) the G7 countries for comparison to prior work and (2) a subset of Asian countries for out-of-sample testing. Our eleven Asian countries include: China (CH),

---

[8] Recent forecasting studies purely on Asia include Chen et al. (2012), Chen et al., (2016) and Qin et al. (2008) which have a multi-country focus, while single country studies include Chang et al. (2011) for Taiwan, Prskawetz et al. (2007) for India, Lewis-Beck and Tien (2012) for Japan and Wang et al. (2015) for China.

[9] Compared to the G7 countries, our sample of Asian countries consist of firm that have: (1) smaller size, (2) higher B/M, and (3) more negative B/M firms. Culturally, our Asian sample of countries is more accepting of power inequalities and less individualistic than the G7 countries.

[10] Theoretically, imperfect information models (Sims, 2003; Woodford, 2002) indicate there could be sluggish adjustment to indicator variables.

[11] For example Jordan, Vivian and Wohar (2014b) provided some evidence on the cross-country variation in market return forecasts for European countries from bivariate predictive regressions (and a simple average combination). The current paper considers a larger number of countries, a broader set of factors and focuses on bagging but includes more while including a wider range of combination methods.

Hong Kong (HK), India (IN), Indonesia (ID), Japan (JP), Korea (KO), Malaysia (MY), Philippines (PH), Singapore (SG), Thailand (TH) and Taiwan (TW), over a balanced sample period of 1995-2011 employing monthly data. The selection criteria are to include Central, East, South, and Southeast Asian countries for which: (1) there is a Datastream (DS) Total Market Index and (2) there is data available from at least 1995.[12]

The data is primarily from Thomson Datastream. We collect monthly data from January 1995 to June 2011. The start date enables inclusion of a wide number of Asian countries for which there is virtually no prior OOS forecasting evidence, with the exception of Japan.[13] It also enables a reasonably sized OOS test period. We have 96 monthly observations (some combination methods use 36 observations for optimization) before beginning the OOS forecasting period in 2003. G7 countries were subsequently included at the suggestion of an Associate Editor. We obtain data for 10 predictor variables including 8 of the variables used by Goyal and Welch (2008, hereafter GW).

We include the following fundamental variables from GW: i) Dividend–price ratio (log), (DP): Difference between the log of dividends paid on the market index and the log of market index price, where dividends are measured using a one-year moving sum, ii) Dividend yield (log), (DY): Difference between the log of dividends and the log of one month lagged market index price, iii) Earnings–price ratio (log), (EP): Difference between the log of earnings on the market index and the log of stock prices, where earnings are measured using a one-year moving sum and iv) Book-to-market ratio, (BM): Ratio of book value to market value for the market index.

We include the following two macroeconomic variables from GW: i) Risk-free rate, (RF): Interest rate on a low risk short-term security and ii) Inflation, (INFL): Calculated from CPI; since inflation rate data are generally released in the following month, we use one month lagged inflation data.

---

[12] The following countries did not have a DS Total Market Index: Afghanistan, Bhutan, Brunei, Burma, Cambodia, East Timor, Laos, Macau, Maldives, Mongolia, Nepal, Tajikistan, Turkmenistan, and Uzbekistan. For the following countries Market index data (not DS index) started after 1995: Bangladesh, Kazakhstan, and Vietnam. At the suggestion of a referee we excluded two small illiquid markets (Pakistan and Sri Lanka) and included Taiwan using alternative data from Datastream, such as the policy interest rate as a proxy for the risk-free rate.
[13] In particular, there is very little prior OOS forecasting evidence on the economic value of return forecasts (portfolio allocation evidence) in our sample of countries, except for Japan.

We include the following two technical variables from GW: i) Stock variance, (SVAR): Sum of squared weekly returns on the market index and ii) Net equity expansion, (NTIS): Ratio of twelve-month moving sums of net issues by listed stocks to total market capitalization of index. We consider two new variables in this context, which are also of interest to technical traders: i) Price Pressure, (PRES): Calculated as the ratio of the number of rising stocks in the previous month divided by the number of falling stocks and ii) Change in Volume, (CVm): Calculated as the monthly change in the volume of traded stocks (in the index).

Table 1 provides a summary of descriptive statistics for our sample countries. We report the mean and the standard deviation for each independent variable used and for the aggregate market return. There are several interesting comparisons. First, the average nominal returns (RET) vary substantially across countries from -0.0015 (-0.15% per month or -1.8% per year compounded) in Japan, up to 0.0111 (1.11% per month or 14.2% per year compounded) in Indonesia. The standard deviation of returns also varies substantially across countries from 0.0422 for the UK to 0.1048 for China. This means the Sharpe ratio (return per unit of risk) varies dramatically across countries from over 0.09 in the US to a negative -0.029 in the Philippines. The wide variation across countries exists for most of the variables we study.

[INSERT TABLE 1 AROUND HERE]

Table 2 contains the correlation matrix for returns across our sample countries. There is substantial difference in correlation between the different country pairs. Interestingly the cross-correlations in Asian countries' returns are modest, on average about 0.40; the lowest correlation is between China and Indonesia at 0.195. This is interesting since the G7 developed markets typically used in prior literature tend to have correlations above 0.70 on average; this is important since G7 studies may be capturing just a single (or small number) of common effect(s).

[INSERT TABLE 2 AROUND HERE]

3. METHODOLOGY

3.A. Assessing the Impact of Individual Variables

Individual predictive regression models are used to estimate the linkage between the dependent, lagged dependent, and a potential predictor variable (including its lags). Define $\Delta RI_t = RI_t - RI_{t-1}$, where RI_t is the log-level of the total stock return index (stock price index that includes reinvested dividends) at month t. [14] In addition, define $y_{t,t+h}^h = (1/h)\sum_{j=1}^h \Delta RI_{t+j}$ so that $y_{t,t+h}^h$ is the (approximate) monthly growth rate of the stock return index from time t to t + h, where h is the forecast horizon. In this section we outline the general models for h step ahead forecasts, however in the empirical analysis we purely focus on 1 step ahead forecasts (h=1). These predictive regression models take the form of (1) below:

$$y_{t,t+h}^h = \alpha + \lambda\ x_{i,t} + \varepsilon_{t+h}^h \tag{1}$$

This model can be employed to estimate h-step ahead forecasts of stock returns using a recursive expanding window. $y_{t,t+h}^h$, is linear in the potential predictor variables (i.e., X_t). The parameter α is a constant. The parameter λ capture the effect of the potential predictor variable. Finally, ε_{t+h}^h is an error term. For each country's stock return, 10 regression models are estimated one for each of the 10 explanatory variables.

The models generated are used to conduct 1 step-ahead out-of-sample forecasts of stock returns. These forecasts are then compared to the respective benchmark model, which takes the form of the historical average (a random walk with drift model). The form of the benchmark model is the same as in (1) where all the λ_j's = 0.

---

[14] This section draws upon Rapach and Strauss (2010).

3.B. Assessing the Impact of Variable Groups

3.B.1. Forecast Combinations

The method of forecast combining is considered to be a useful technique for "...sharing strengths of different forecasting procedures..." (Yang, 2004:205) and is an alternative method of imposing "...structure on high-dimensional forecasting models" (Stock and Watson, 2004:1). Empirical as well as theoretical evidence (see Clemen and Winkler, 1986; Rapach and Strauss, 2010; Stock and Watson, 2003; 2004; and Yang, 2004) indicates that forecast combining generally improves the predictive ability of models because it includes more variables or potential predictors, thus increasing the amount of information used in generating forecasts. Recent work by Rapach, Strauss and Zhou (2010) demonstrates combination forecasts are also useful for US stock index returns. The forecast combination methods used in this paper include mean, median, trimmed mean, Discounted Mean Squared Forecast Errors (MSFE), and Cluster (C), and Principal Components (PC) combinations. According to Rapach and Strauss (2008, 2010), these combination methods can be described below in (2):

$$\hat{y}^h_{CB,t+h|t} = \sum_{i=1}^{n} w_{i,t}\, \hat{y}^h_{i,t+h|t} \tag{2}$$

where $\hat{y}^h_{CB,t+h|t}$ is the combined forecast of the variable of interest from individual regression models (1) and $w_{i,t}$ is the weight of the individual regression forecasts, $\hat{y}^h_{i,t+,h|t}$ is the individual forecast. The weights sum to unity.

It is now well known that combining forecasts generally improves OOS forecast performance in a wide range of applications including aggregate financial variables. Hence, we discuss each method intuitively and refer the reader to Rapach and Strauss (2010) for full details of its estimation. We investigate a total of eight combination methods from four different types. Firstly, three simple combination strategies are examined: Mean, Median, and the Trimmed mean, where the trim excludes the lowest and highest forecasts from the average. Second, we use two discount forecast combination models. By decreasing the discount rate, more recent observations can be weighted more heavily. As in

Stock and Watson (2004), we use use δ = 1.0 (no discounting of past forecasts) and δ = 0.9 (more weight on most recent forecasts) resulting in Stock and Watson's discounted mean-square forecast error (DMSFE) methods DMSFE(1) and DMSFE(0.9). Thirdly, we use cluster combinations to control for forecast persistence (see Aiolfi and Timmermann, 2006). Utilizing a hold-out period, the individual ARDL models are ranked by their mean square forecast error (MSFE) and clusters are created by consecutively adding the next lowest MSFE ARDL model to the cluster. Clusters are formed on the MSFE over a rolling window. Within the cluster with previous best (PB) MSFE an average of ARDL model forecasts are taken to generate the cluster combination forecast. We utilize this method with two clusters [C(2,PB)] and three clusters [C(3,PB)].[15] Finally, we apply a principal component method to the individual ARDL model forecasts. We estimate the number of components with the ICp3 criterion up to a maximum of 4 (Bai and Ng, 2002). See Stock and Watson (2004) for further details. We label this final combination method as PC(C,3B).

3.C. Construction of Bagging Forecasts

Recall that $y_{t,t+h}^{h}$ is the monthly stock return from time t to t + h, where h is the forecast horizon. Let $x_{i,t}$ denote one of n potential predictors of stock returns (so that i = 1, . . ., n). We consider 10 potential predictors of stock returns (n = 10) in our analysis.

We compute bagging forecasts of stock returns at horizon h using the bagging-augmented pretesting procedure (BA) of Inoue and Kilian (2008). The procedure begins with the general model:

$$y_{t,t+h}^{h} = \mu + \sum_{i=1}^{n} \delta_i x_{i,t} + \xi_{t+h}^{h}, \tag{3}$$

where ξ_{t+h}^{h} is an error term characterized by autocorrelation of degree h − 1. Suppose we are interested in forming a forecast of y_{t+h}^{h} at time t . The pretesting procedure involves estimating (3) via ordinary least

---

[15] C(P,PB) uses P clusters. PB denotes that we employ the Previous Best conditional combination strategy

11

squares (OLS) using data from the start of the available sample through time t and computing the t-statistics corresponding to each of the potential predictors.[16] The $x_{i,t}$ variables with t-statistics less than 1.645 in absolute value are dropped from (3), and the model is estimated a second time using only significant predictors.

Bagging can be implemented for the pretesting procedure via a moving-block bootstrap. More specifically, a large number (B) of pseudo samples of size t for the left-hand-side and right-hand-side variables in (5) are generated by randomly drawing blocks of size m (with replacement) from the observations of these variables available from the beginning of the sample through time t. For each pseudo-sample, we estimate (5) using the pseudo-data and OLS, the (pretesting) procedure determines the predictors to include in the forecasting model, the model is re-estimated using the pseudo-data, and a forecast of y^h_{t+h} is formed by plugging the actual included $x_{i,t}$ values into the re-estimated version of the forecasting model (and again setting the error term equal to its expected value of zero). The bagging model forecast corresponds to the average of the B forecasts for the bootstrapped pseudo samples.[17]

Dividing the complete available sample of T observations for Δy_t and $x_{i,t}$ (i = 1, . . . , n) into an in-sample portion comprised of the first R observations and an out-of-sample period comprised of the last P observations, we can form a series of P − (h − 1) recursive simulated out-of-sample forecasts using the bagging procedure.[18] We denote this series by $\left\{ \hat{y}^h_{BA,t+h|t} \right\}^{T-h}_{t=R}$.

3.D. Statistical Tests

Tests for encompassing and equal forecast accuracy

In the case of nested models, Clark and McCracken (2001) and McCracken (2007) develop a set of asymptotics that allow for an out-of-sample test of equal population-level predictive ability between

---

[16] The t -statistics for the OLS estimates of δ_i in (5) are computed using Newey and West (1987) heteroscedasticity and autocorrelation consistent (HAC) standard errors based on a lag truncation of h − 1.

[17] Following Inoue and Kilian (2008), we use *m = h*. We use *B = 1000*.

[18] "Recursive" indicates that the forecasts are generated using an expanding estimation window.

two nested models. They show that, in the context of linear, OLS-estimated models, a number of different statistics can be employed to test for equal forecast accuracy and forecast encompassing, despite the fact that the models are nested. Based on Monte Carlo simulations, Clark and McCracken (2001, 2004) indicate that *ENC–NEW* is the most powerful statistic, followed by their *ENC-t, MSE – F* and the *MSE – T* statistics. These rankings suggest that the forecast encompassing statistics, especially *ENC – NEW*, can have important power advantages over test statistics based on relative MSFE. We report results for the most powerful statistic ENC-NEW, which is an F type test and is related to the Harvey et al. (1998) statistic designed to test for forecast encompassing. It has been shown through extensive Monte Carlo simulations in Clark and McCracken (2001, 2004) that the ENC-NEW statistic has power advantages over the original Diebold and Mariano (1995) statistic as well as the Harvey et al. (1998) ENC-t statistic which applies to nested forecasting models.

Under the null hypothesis, the restricted model forecasts encompass the unrestricted model forecasts, while under the one-sided (upper-tail) alternative hypothesis the restricted model forecasts do not encompass the unrestricted model forecasts. Clark and McCracken (2001) note that the limiting distribution of the ENC-NEW statistic is non-standard and pivotal for one step ahead forecasts (h = 1) considered in this paper. Clark and McCracken (2004) recommend basing inferences the *ENC – NEW* statistics on a bootstrap procedure, given that the statistics are not in general asymptotically pivotal (when h>1). The bootstrap procedure we employ is similar to the one in Clark and McCracken (2004), which is a version of the Kilian (1999) bootstrap procedure, and is discussed in detail in Rapach and Weber (2004) and Rapach and Wohar (2006).

Clark and West (2006, 2007) demonstrate that the ENC-t test can be viewed as an adjusted test for equal MSE. In the Clark and West framework the null hypothesis is a random walk and the alternative hypothesis is of a predictive regression. If the null hypothesis of a random walk is true then it will have a lower mean-squared error relative to the alternative (despite the fact the alternative include an additional variable) due to the fact that there is sampling error associated with estimating the alternative model. The Clark and West t-test adjusts the forecast error of the alternative model to take account of this sampling

13

error. The CW-t statistic proposed by Clark and West (2006, 2007) is equivalent to the Harvey et al. (1998) ENC-t test for forecast encompassing as considered in such studies as Clark and McCracken (2001, 2005).

3.E. Measuring Economic Value

Our final set of empirical tests deal with the economic value of forecasts. We analyze if portfolio allocations could have been improved by following the regression model rather than the historical average benchmark. We consider a mean-variance optimizing investor. We take the return forecast from the historical average benchmark and compare it to an alternative return forecast from i) bagging and ii) combination forecast methods.

Recall that y_{t+1} is the log stock return. Define Y_{t+1} as the stock return ($Y_{t+1} = \exp(y_{t+1})$-1. A mean-variance optimizing investor has objective function:

$$O = E(Y_p) - \frac{\gamma}{2}\sigma_{Y_p}^2 \approx E(y_p) + \frac{1}{2}\sigma_{y_p}^2 - \frac{\gamma}{2}\sigma_{y_p}^2 \tag{4}$$

where O is the objective, Y_p is the portfolio return, y_p is the portfolio log return, and γ is the coefficient of relative risk aversion. Such an investor will choose a portfolio weight, $\omega_{t,b}$ $\left(\omega_{t,z}\right)$ of the risky asset under the prediction from the historical average benchmark (the alternative forecast model [bagging or forecast combination]):[19]

$$\omega_{t,b} = \left(\frac{1}{\gamma}\right)\left(\frac{E(Y_{t+1,b}) - Y_f}{\sigma_t^2}\right) \tag{5}$$

$$\omega_{t,z} = \left(\frac{1}{\gamma}\right)\left(\frac{E(Y_{t+1,z}) - Y_f}{\sigma_t^2}\right) \tag{6}$$

---

[19] Weights are recalculated in every time period and portfolio allocations adjusted accordingly.

Y_f is the risk-free rate. We use 5-year rolling monthly data to estimate volatility $\left(\sigma_t^2\right)$; however, estimating volatility using alternative horizons has very little impact on the utility *gain* since σ_t^2 is the same in the benchmark weight, $\omega_{t,b}$, and the alternative model weight placed, $\omega_{t,z}$ (see the denominator in equations 12 and 13 above).

The utility gain (ΔO) from using the regression model rather than the historical average benchmark is:

$$\Delta O = Y_z - Y_b - \frac{\gamma}{2}\left(\sigma_{Y_z}^2 - \sigma_{Y_b}^2\right) \tag{7}$$

Second, we implement Goetzmann et al.'s (2007) performance measure (referred to as GISW):

$$GISW = \frac{1}{1-\Gamma}\left[\ln\left(\frac{1}{T}\sum_{t=0}^{T-1}\left(\frac{1+Y_{t+1,z}}{1+Y_{t+1,f}}\right)^{1-\Gamma}\right) - \ln\left(\frac{1}{T}\sum_{t=0}^{T-1}\left(\frac{1+Y_{t+1,b}}{1+Y_{t+1,f}}\right)^{1-\Gamma}\right)\right] \tag{8}$$

$$\text{where: } \Gamma = \frac{\ln[E(1+Y_m)]-\ln(1+Y_f)}{\text{Var}[\ln(1+Y_m)]}$$

GISW measures the average performance of a portfolio relative to the risk-free rate; it is a certainty equivalent measure of abnormal performance. The parameter Γ is set to reflect the overall reward (return) to risk (variance) ratio for each country based upon our sample data. This reduces the possibility of manipulation and incorrect inference.

3.F. Data Mining Robust Critical Values

When testing the predictive ability of a large number of financial variables, Lo and MacKinlay (1990) and Inoue and Kilian (2005) note that data mining is a serious concern when one is dealing with stock return predictability regardless of whether the tests are in-sample or out-out-sample. Inoue and Kilian (2005) note that an important way in which data mining can be controlled for is to use appropriate critical values, which explicitly account for the possibility of data mining. Here we employ the data mining bootstrap used in Rapach and Wohar (2006). While Rapach and Wohar (2006) consider individual

predictive regressions, we focus on combinations and bagging in this paper. We calculate the two out-of-sample statistics (MSE-F and CW-t) and the economic value measures for all models (each of the forecast combination techniques and bagging) in turn. For each pseudo sample we store the maximal values for each metric. We repeat this (whole) process 5000 times, to generate an empirical distribution for each of the maximal out-of-sample statistics and for each of the maximal economic value measures. After ordering the empirical distribution for each maximal statistic, the 4,500th, 4,750th, and 4,950th values serve as the 10%, 5%, and 1% critical values for each maximal statistic.

4. OUT-OF-SAMPLE STOCK RETURN FORECASTS

Could investors' actually utilize fundamental-price based models in order to benefit from more accurate predictions of future stock returns? This issue is of importance to both practitioners and academics alike. Asset managers, economic policymakers, as well as pension providers and contributors all need accurate estimates of future market returns.

In this section, we examine a range of fundamental-price ratios as well as macro and technical variables for a sample of Asian countries. Following Rapach, Strauss, and Zhou (2010) and Stock and Watson (2004) we consider if various combining forecasts or bagging methods can improve forecast accuracy over individual models. The historical average model is used as the benchmark.

4.A   OOS Forecast Accuracy (Individual Regression Forecasts)

Table 3 shows 1-month forecast results for individual predictive regression models.[20] Overall, the individual predictive models have mixed results. In the Asian markets, some predictors provide dismal forecasts, such as stock variance (SVAR), net equity issuance (NTIS), and change in volume at the monthly frequency (CVM). However, the performance of individual fundamental-price ratios, e.g., dividend-price ratio (DP), dividend yield (DY), earnings price (EP), and book-to-market (BM), with the

---

[20] We focus on one-step-ahead forecasts due to space considerations.

16

exception of EP, show evidence of predictability. DP, DY, and BM show predictability above that of the benchmark model in 8 of 11 countries. We implement the Clark and McCracken (2001) ENC-NEW encompassing test and the Clark and West (2007) CW-T test of equal forecast accuracy. Predictability in our Asian sample is robust even if very high hurdles associated with data snooping adjustments are made, however data-snooping significance in fundamentals is found only under the ENC-NEW. Technical indicators, e.g., price pressure as measured by rising stocks against falling stocks (PRES), also demonstrate some predictability. PRES demonstrates predictability in 6 of 11 countries using standard statistical significance tests and in 2 of 11 countries after adjusting for data-snooping bias. The risk-free rate (RF) and inflation (INFL) are the only two predictors that demonstrate predictability under both the CW-T test and data-snooping bias adjustments. Thus, it appears that investors interested in Asian markets should specifically consider these variables in forecasting models.

[INSERT TABLE 3 AROUND HERE]

Predictability in the G7 countries is not found to be robust in our tests. Fundamental ratios do not perform well in the G7 sample. The lack of robust statistical evidence of predictability in the G7 countries contrasts sharply with the robust predictability found in our sample of Asian countries. The evidence for predictability in G7 countries is even more dismal once data mining adjustments are made. When data mining adjustments are used, there is virtually no evidence of predictability in the G7 countries.

There is also considerable variation in predictability across countries. Although there is virtually no predictability in the G7, there is evidence at the 5% significance level for DE (Germany), JP, and the US. In the Asian sample, we find there is no evidence of OOS predictability in China as no variable demonstrates predictability when data-snooping bias is controlled. Predictability is also weak for the other large Asian markets (HK, India, Japan, Korea, and Taiwan) where out of 10 predictor variables only one exhibits some predictability after data-snooping bias adjustments. However, four countries (Indonesia, Malaysia, Singapore, and Thailand) have robust predictability, even after data-snooping bias.

17

4.B   OOS Forecast Accuracy (Bagging and forecast combinations)

This section contains the forecast results from bagging and forecast combinations. Table 4 reports results for the 1-month forecast horizon. We find consistent evidence in our Asian sample that combining forecasts improves forecast accuracy in our sample of Asian countries even after adjustments are made for data-snooping bias. This is true for seven of eight methods examined, the lone exception is the principal component method (PC(C,3B)), and in 10 of 11 countries there are consistent improvements in forecast accuracy, the sole exception is China. The lack of predictability in China may be due to structural reforms in its equity market that substantially increase the proportion of tradable shares after 2005, which is near the start of our forecasting period (see Li et al., 2011 for further information on the impact of this reform). Combination forecasts provide consistently large gains in HK, India, Indonesia, Malaysia, Singapore, and Thailand. The finding that combination forecasts consistently perform well is strengthened by the fact they are subject to greater parameter estimation error than the benchmark. Clark-West (2007) emphasize that parameter estimation error leads to an expectation that the mean-squared prediction error from an alternative model (here the forecast combination) is larger than that from a parsimonious model (here the benchmark).

<p style="text-align:center">[INSERT TABLE 4 AROUND HERE]</p>

Once more, in the G7 set of countries the results for predictability is not promising. There is little to no evidence with predictability only documented for JP and the US for combination methods and only for JP for bagging methods. Again, this is in complete contrast to the results found for Asia.

We note that the various classes of forecast combination methods yield broadly similar results; hence in the subsequent analysis we only report results from one method in each class: Mean, DMSFE(1), C(2,PB) and PC(C,3B); we can confirm results from the other combination techniques in the subsequent tests are qualitatively similar to those of the same combination class that is reported.

In terms of forecast errors, the bagging method provides mixed results for the majority of Asian countries; in general it underperforms the benchmark. For several countries (China, HK, Korea, Singapore and Taiwan) bagging forecast errors are 5% larger than those from the benchmark, a

substantial magnitude. However, bagging does provide substantial improvement for four countries: Indonesia, Japan, Malaysia, and Thailand. These results for bagging are in contrast to prior literature that provides favorable results for the bagging method when applied to macroeconomic forecasts (Inoue and Kilian, 2008; Rapach and Strauss, 2010).

6. ECONOMIC SIGNIFICANCE OF STOCK RETURN FORECASTS

6.A   Economic Gains (Bagging and Combination Forecasts)

Table 5 and Table 6 present the results for the bagging method and various combination methods for economic gains. Tables 5 and 6 provide results for the utility gain method and the manipulation proof method, respectively.

The utility gain method for bagging indicates large outperformance of the benchmark is available in several Asian countries, including HK, ID, KO, SG, TH, and TW. For Singapore, annualized gains in utility of more than 10 percentage points across most models are possible. For other countries gains are more modest, but still economically large, e.g., one can easily gain more than 5 percentage points in HK, ID, and MY. For the average forecast combinations (AV COMB), utility gains are generally about 1 to 10 percentage points for 9 of the 11 countries. In the G7 countries, modest utility gains are generally available. However, more than 5% can be realized in DE, FR, and the UK. The evidence suggests that bagging performance is robust to data mining adjustments.

Economic gains with combination forecasts exist for China. Overall, combination methods consistently provide economic value across countries and these results are robust to data snooping bias in the Asian sample of countries. Before data bias adjustments, almost every combination model performs well across our sample of countries. After data snooping bias adjustments evidence of predictability weakens substantially; the most consistently performing method is the cluster approach PC(3,PB) with economic gains in only 4 of 11 markets. So far, the results are broadly consistent with the statistical test results in Table 4. The major difference between Table 4 and Table 5 is that bagging demonstrates large and significant economic gains in 9 of 11 countries. Some of the gains are very large, e.g., 18.73% in SG

and 15.63% in TW. We report gains when relative risk aversion is 3, i.e., P(UG), and when relative risk aversion is 5 , i.e., P(UG5). The results are consistent for both values.

The bagging method can generate value for investors in G7 markets and this is statistically significant in most markets except the US and Canada. The statistical evidence for bagging is most consistent for Germany and France although Japan and UK also have some support that gains are significant using the approach we implement to address data mining concerns. In terms of the combination methods, the results are generally, quite similar to those for bagging. Positive gains can be earned in most countries; however, these are typically smaller than those from the bagging method and statistical evidence of outperformance largely evaporates when measures to adjust for data mining are implemented.

[INSERT Table 5 AROUND HERE]

Table 6 provides results for a manipulation proof measure of economic gain. The GISW certainty equivalent gains confirm the conclusions drawn from Table 5. For the G7, the GISW results are much larger than under the utility gain measure. Overall, GISW gains can be large both in Asian countries (e.g., 14.62 for the mean combination in SG) and exist across almost all countries.[21] Again, bagging is a reliable model for attaining economic gains. After the manipulation proof adjustments, bagging outperforms the benchmark in every country and does so at a significant level in 6 of the 11 Asian countries and in 4 of the 7 G7 countries. Unlike the combination models, gains are available by using the bagging method even after controlling for data snooping bias. As in Table 5, we report gains when relative risk aversion is 3, i.e., GISW, and when relative risk aversion is 5, i.e., GISW5. The results are again broadly consistent for both values of risk aversion.

[INSERT Table 6 and 7 AROUND HERE]

---

[21] In some countries, e.g., TH, the risk-return tradeoff is negative, meaning that the risk-free rate is higher than the average stock return. This will affect the manipulation-proof measure since it looks at the return on the model relative to the risk-free rate. Then it takes this to the power of 1-risk-return tradeoff. But in the case of Thailand this is negative, so this inverts the relationship, which will drive the undesirable negative result. So a caveat to using the manipulation proof method is that one must be careful to ensure there is a positive equity premium; if not, then one should interpret the method's results cautiously.

In Table 7 we examine the impact of accounting for trading costs on the utility gain measure. We report results for trading costs of 10 basis points (on the portion of the portfolio adjusted), although results are qualitatively similar if 30 basis points are used or if the GISW measure is implemented. The results in Table 7 are very similar to Table 5. In fact, generally the net impact on economic value is less than 0.2% p.a. It should be noted that re-balancing costs are incurred by the historical average allocations as well, but even if these are assumed to be zero, the impact of transaction costs is not large because the average change in portfolio allocation per period is small.

The most striking part of the economic value results (in Tables 5-7) is that for bagging these are in clear contrast to those under the forecast accuracy approach (see Table 4). In contrast to its poor forecast accuracy performance, bagging performs well in economic tests across countries and usually is a top or comparable performer when compared to the combination predictor results. We shall explore reasons for this difference in Tables 8. Overall, the economic value results are consistent for both the utility gain and the manipulation proof measure methods, with the prior results slightly stronger. For the G7 especially, both bagging and forecast combination methods provide more consistent evidence of economic gains than of forecast accuracy.

6.B   Discussion and Further Exploration

Section 6.A demonstrates that both bagging and combination forecasts would enable an investor to tilt their portfolio to increase utility and to make certainty equivalent gains. Simply, when the regression model predicts high (low) returns, the portfolio is tilted towards equities (T-bills). Our empirical results suggest there is some economic value from forecasting returns with macro and technical variables that could potentially be exploited by practitioners. Our results also indicate that noise may be an issue and that consistent results are best obtained with a combination of many forecasts or by simulating the relationship utilizing in-sample data.

The economic value results differ from forecast accuracy results. A possible reason for this is that they weight extreme predictions differently. Forecast accuracy penalizes these large errors heavily under a

21

squared error rule; however, in the economic value exercise a restriction on portfolio weight ameliorates the impact of large forecast errors. The results for bagging were different when we conducted economic value tests (performed well) compared to forecast accuracy measures (performed poorly). The economic value tests employ a restriction that the portfolio weight is between 0 and 1.5. The forecast accuracy tests do not have such a portfolio weight restriction. In Table 8, we explore whether the bagging method performed poorly in the forecast accuracy tests due to lack of restrictions. The restricted model (here winsorization of extreme forecasts) will reduce the variance of the forecast compared to the unrestricted model and if it moves the unrestricted model 'closer' to the true model will generate gains in forecasting by reducing forecast error. Thus, we apply equivalent restrictions on the forecast accuracy tests as we did in the economic value tests. We require that the forecast return must be between the risk-free rate (Y_f) and 1.5 times the risk aversion coefficient (γ) times the variance of the return (σ_t^2) plus the risk-free rate (Y_f):

$$Y_f \leq E(Y_{t+1,b}) \leq Y_f + 1.5(\gamma)\sigma_t^2 \ .$$

We rerun the forecast accuracy tests of Table 4 with these restrictions imposed, which are reported in Table 8. For the Asian subset of countries the results with restrictions applied are given in Panel A of Table 8. Almost every single model outperforms the benchmark for more than one-half of the countries, with many methods doing considerably better when restrictions apply. Bagging now performs strongly and the performance is very robust to data mining bias. The average combination forecasts generally perform better and are more robust to data mining bias. Our results of this table indicate that applying restrictions on the forecast return may be an important improvement to forecast performance tests in Asian markets. This finding is consistent with the restricted model being 'closer' to the true model than the unrestricted model since in this case gains in forecasting occur due to reduced estimation error of the forecast. The lack of restrictions on Bagging in our initial forecast accuracy tests (Table 4) are thus an important factor for the differing results between the accuracy tests and the economic gain tests (Tables 5-7).

Panel B of Table 8 also provides some evidence that the bagging procedure can beat the benchmark in 5 (4) of the G7 countries using the RMSFE (RMAFE) measure. There is also consistent statistical evidence in favor of bagging in 5 of 7 countries using conventional critical values but using data mining critical values this is confined to the CW-t test for Germany, UK and US. There is also evidence that combinations are generally able to beat the benchmark in terms of point estimates. However, the statistical evidence is weak. Only for the US is consistent evidence found that the benchmark is beaten by a statistically significant amount (corroborating Rapach et al., 2010). The US results are robust to the data-mining procedure we implement using the CW-t test. The evidence of statistical significance from combinations is not widespread. Apart from the US there is some sporadic evidence for Japan in favor of combinations, but there is relatively little statistically significant support for combinations from other G7 countries.

[INSERT Table 8 AROUND HERE]

7. FURTHER ANALYSIS

7.A   Encompassing tests and the optimal weights

Table 9 provides results from the encompassing test of whether the combination forecast is encompassed by the bagging forecast. It reports the optimal weight on a combination forecast when being used with bagging to obtain a composite return forecast. The statistical significance of the HLN (1997) test is denoted by stars.

The results in Panel A of Table 9 are for Asian countries. These indicate that the weight on the combination forecast is typically less than 0.5 (this is true for 30 of the 44 observations) suggesting it is optimal to put greater weight on the bagging forecast. The HLN test is generally insignificant except for India and Taiwan. Thus, in 9 of 11 Asian countries the encompassing tests indicate that the combination forecast does not add new information beyond that contained in the bagging forecast. Panel B of Table 9 gives results for the G7 countries. Of these, 15 of the 28 optimal weights are below 0.5 suggesting that in the majority of cases more weight is placed on the bagging forecast. Only in Canada are weights typically

23

above 1 observed for the combination forecasts. Medium weights France, Italy, Japan and US are typically less than 0.5. In Germany and the UK the weights are typically low or even negative. Canada is the only country where the null hypothesis of the HLN test can be rejected. That is, only in Canada do the combination forecasts consistently add information beyond that contained in the bagging forecasts. This indicates that for the G7 countries except Canada, the combination forecasts do not contain incremental information beyond the bagging forecast. In both sets of countries there is little statistically significant evidence that combination forecasts add new information to that contained in the bagging forecast. Overall, this implies the bagging forecast should be used by itself (or as the senior partner in a composite forecast) in practical applications of stock return forecasting.

[INSERT Table 9 AROUND HERE]

7.B   Cross-country evidence

Table 10 examines cross-country evidence on forecasting performance. A wide range of country-level variables are employed including economic development, trade, financial development, information availability, institutional setting. Most of these characteristics outlined can impact the speed of the market response, e.g. market development, information frictions and equity market liquidity should all impact how quickly news is incorporated into stock prices.[22] The greater the delay in reaction to news the more accurately the next periods return can be forecast.

Panel A reports results from bivariate cross-sectional regressions for RMSFE by these variables across our sample of 17 countries. Many variables linked to legal institutional setting (anti-self dealing, property rights, common law), market development (GDP, FDI, market development) and investor sophistication (internet usage, mobile phone usage) are not significantly linked to RMSFE. Nonetheless there are several variables that are significantly related to RMSFE. Results suggest that trade openness (Trade / GDP) is an important determinant of RMSFE. The higher the trade openness is the lower the

---

[22] Theoretically, imperfect information models (Sims, 2003; Woodford, 2002) indicate there could be sluggish adjustment to indicator variables.

24

RMSFE and hence the better the forecast accuracy. Perhaps as outlined in Jordan et al. (2014a) there is information flow between trading partners (especially importers) from other countries but the market fails to full incorporate this information immediately. Finally there is a positive relationship between stock turnover (trading volume / equity capitalization) and RMSFE for most combinations. This indicates that the higher the turnover and the more liquid the equity market the higher the RMSFE, which indicates reduced precision of forecasts since there is less ability to forecast based on stale data.

Table 10 Panel B examines the relationship between utility gains and these country-level variables across our sample of 17 countries. Empirical results indicate that Trade openness and FDI are pervasively related to utility gain, as well as forecast accuracy. Panel C of Table 11 examines the relationship between GISW manipulation proof certainty equivalent gains and the country-level variables. At the 5% significance level, there continues to be evidence of predictability via trade-to-GDP (TRADE) and foreign direct investments (FDI). However, there seems to be no consistently strong relationship between several country-level groups of variables and forecast performance. There is little linkage between investor sophistication and forecast performance established and sporadic significance for proxies linked to legal structure or equity market development. This suggests that good forecast performance can be achieved regardless of a country's ranking by these characteristics.

[INSERT Table 10 AROUND HERE]

Trade openness is the only pervasive factor that predicts across all the measures of forecast accuracy / economic value we consider at the 5% significance level. Trade openness is statistically significant at the 5% level by all forecast metrics for the bagging method and by all metrics for three of the four combination methods. We can conclude that there is strong support that trade openness is related to predictability. It is plausible that for countries with greater trade links there is a larger information set of potentially relevant information, which takes more time to be fully absorbed into market data by investors with limited processing power. This is consistent with the view that information relevant to equity markets travels via trading links (Jordan et al., 2014a).

25

7.C   Summary of other Tests

We also conducted other tests in the process of producing this paper. The main ones revolved around i) whether bagging did well on alternative measures of forecast performance (e.g. sign prediction), ii) which measure of forecast performance was most closely related to economic value and iii) whether the empirical results for forecast accuracy were driven by prediction of volatility rather than prediction of mean. These results suggested that: i) bagging's performance for alternative measures of forecast performance was mixed, ii) of the forecast performance measures RMSFE and RMAFE are generally the most closely related to economic value and iii) results for prediction of return volatility are not as favorable as for the mean. These results are available upon request.

8. CONCLUSION

This study focuses on the bootstrap aggregating method (bagging), which explicitly addresses model and parameter instability as well as providing inference that explicitly addresses data mining concerns. The bagging method has been recently applied to forecasting macroeconomic series using linear regressions with continuous variables. In this study we apply the bagging method to forecast market stock returns out-of-sample (OOS) in a multi-country setting. We study the G7 which is commonly studied in past return forecasting literature and a sample of 11 Asian countries that covers countries with differing characteristics and for which there is relatively little prior literature on return forecasting. [23] In our tests, we consider three types of predictors (fundamental ratios, macro variables, and technical variables); we examine if bagging and combining methods, proposed to overcome poor return predictability, can improve forecast accuracy or increase economic value. We apply bagging methods to

---

[23] Balsara, Chen, Zheng (2007) demonstrate that an ARIMA forecasting model outperforms forecasts based on a random walk model and Goh, Jiang, Tu, and Wang (2013) find Chinese stock returns can be forecast by both Chinese and US economic variables since 2001. However, neither of these papers examines bagging nor forecast combinations.

international stock return forecasts, which provides new and broader insights on whether or not bagging is effective internationally.[24] Our findings are as follows.

Firstly, we find for stock return forecasts in both G7 and in Asian markets that bagging is of economic importance and can lead to sizable gains from an asset allocation strategy. The empirical results indicate that economic gains from bagging are generally statistically different from those from the benchmark and are robust to reasonable trading costs. Further the economic gains from bagging tend to be greater than those from forecast combinations.[25] Thus, bagging methods may prove useful to practitioners.

We examine forecast accuracy metrics when restrictions (i.e., winsorization of extreme forecast values) on forecasts implied from the economic value exercises are applied.[26] When these restrictions are applied, bagging generally performs well for standard forecast accuracy metrics (RMSFE, RMAFE), whilst the unrestricted results are weaker.[27] Consequently, although bagging is a technique to reduce forecast variance, we find it is necessary to winsorize extreme forecasts in order for it to perform well in standard forecast accuracy tests. Thus these restrictions appear to be an important factor that can help reconcile the initially differing results between statistical and economic gain tests. Overall, we provide evidence that bagging methods are effective in a range of small, medium, and large-sized Asian markets with different characteristics to the US and other G7 countries. Thus, bagging could generate gains universally.

Secondly, we provide statistical evidence that explicitly accounts for data mining issues. We adapt the approach of Rapach and Wohar (2006) which is based on the pioneering work of White (2000) and Inoue and Kilian (2005). To our knowledge this process for generating data mining robust critical

---

[24] To our knowledge there is no prior multi-country study of bagging applied to international (non-US) stock returns.

[25] Forecast combination techniques have been widely applied in many different disciplines but only recently applied to US stock return forecasts (Rapach, Strauss and Zhou, 2010).

[26] We restrict the stock return forecast to be consistent with portfolio weights between 0 and 1.5. This winsorises the stock return forecast to be in the range: $Y_f \leq E(Y_{t+1,b}) \leq Y_f + 1.5(\gamma)\sigma_t^2$, where Y_f is the risk-free rate, γ is the coefficient of risk aversion and σ_t^2 is equity index return volatility.

[27] Della Corte, Sarno and Valente (2010) and Cenesizoglu and Timmermann (2012) find forecasts may have economic value even though statistical tests provide no evidence of superior performance.

27

values has not previously been applied to forecast combinations or to bagging. These data mining robust critical values place a very high hurdle on empirical tests. Interestingly we find that even after accounting for data mining there is some evidence that bagging beats the historical average in both forecast accuracy tests and economic value tests.

Finally, we explore if predictability is related to specific country characteristics. Somewhat surprisingly, legal origin, economic or financial development has no significant relation to predictability, which contrasts with prior studies that suggest such variables may impact the effective operation of financial markets in various settings. However, we find strong support that predictability is related to trade-to-GDP and to an extent to foreign direct investment. These results provide further evidence that trading links can impact forecast performance (Jordan et al., 2014a).

BIBLIOGRAPHY

Aiolfi, M. and Timmermann, A. (2006), "Persistence in forecasting performance and conditional combination strategies." *Journal of Econometrics*, Vol. 135, No. 1-2, pp. 31-53.

Bai, J., and Ng, S. (2002), "Determining the number of factors in approximate factor models." *Econometrica*, Vol. 70, No. 1, pp. 191–221.

Balsara, N. J., Chen, G. and Zheng, L. (2007), "The Chinese stock market: an examination of the random walk model and technical trading rules." *Quarterly Journal of Business and Economics*, Vol. 46, No. 2, pp. 43-63.

Bilinski, P., Lyssimachou, D., & Walker, M. (2012). Target price accuracy: International evidence. *The Accounting Review*, Vol. 88, No. 3, pp. 825-851.

Bossaerts, P. and Hillion, P. (1999), "Implementing statistical criteria to select return forecasting models: what do we learn?" *Review of Financial Studies*, Vol. 12, No. 2, pp. 405-428.

Campbell, J. and Thompson, S. (2008), "Predicting excess stock returns out of sample: can anything beat the historical average?" *Review of Financial Studies*, Vol. 21, No.4, pp.1509-1531.

Cenesizoglu, T. and Timmermann, A. (2012), "Do return prediction models add economic value?" *Journal of Banking and Finance*, Vol. 36, No. 11, pp. 2974–2987.

Chang, C. L., Franses, P. H., and McAleer, M. (2011). "How accurate are government forecasts of economic fundamentals? The case of Taiwan".*International Journal of Forecasting*, Vol. 27, No. 4, pp. 1066-1075.

Chen, C. J., Ding, Y., & Kim, C. F. (2010). "High-level politically connected firms, corruption, and analyst forecast accuracy around the world." *Journal of International Business Studies*, Vol. 41 No. 9, pp. 1505-1524.

Chen, C.W.S., Gerlach, R., Hwang, B.B.K., and McAleer, M., (2012), Forecasting Value-at-Risk using nonlinear regression quantiles and the intra-day range. *International Journal of Forecasting*, Vol. 28, No. 3 (July – September), pp. 557-574.

Chen, Q., Costantini, M., & Deschamps, B. (2016). "How accurate are professional forecasts in Asia? Evidence from ten countries." *International Journal of Forecasting*, Vol. *32, No.* (1), pp. 154-167.

Clark T. and McCracken, M. W. (2001), "Tests of equal forecast accuracy and encompassing for nested models," *Journal of Econometrics*, Vol. 105, No. 1, pp. 85-110.

Clark T. and McCracken, M. W. (2004). Evaluating long-horizon forecasts. Manuscript, University of Missouri – Columbia.

Clark T. and McCracken, M. W. (2005), "Evaluating direct multistep forecasts," *Econometric*

Reviews 24, 369-404.

Clark, T. E. and K. D. West (2006), "Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis," *Journal of Econometrics* 135, 155-186.

Clark, T. E. and K. D. West (2007), "Approximately normal tests for equal predictive accuracy in nested models," *Journal of Econometrics* 138, 291-311.

Clemen, R.T. and Winkler, R.L. (1986), "Combining economic forecasts." *Journal of Business and Economic Statistics*, Vol. 4, No. 1, pp. 39-46.

Clemen, R. T. (1989), "Combining forecasts: a review and annotated bibliography", *International Journal of Forecasting,* Vol. 5, No. 4, pp. 559-583.

Della Corte P., Sarno, L. and Valente, G. (2010), "A century of equity premium predictability and the consumption–wealth ratio: An international perspective", *Journal of Empirical Finance*, Vol. 17, No. 3, pp. 313-331.

Diebold, F. X., Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, Vol 13, pp. 253–263.

Dovern, J., Fritsche, U., Loungani, P., and Tamirisa, N. (2015). "Information rigidities: Comparing average and individual forecasts for a large international panel". *International Journal of Forecasting*, Vol. 31, No. 1, pp. 144–154.

Giot, P. and Petitjean, M. (2011), "On the statistical and economic performance of stock return predictive regression models: an international perspective." *Quantitative Finance*, Vol. 11, No. 2, pp. 175-193.

Goetzmann, W., Ingersoll, J., Speigel, M., and Welch, I. (2007), "Portfolio Performance Manipulation and Manipulation-proof Performance Measures." *Review of Financial Studies*, Vol. 20, No. 5, pp. 1503-1546.

Goh, J. C., Jiang, F., Tu, J. and Wang, Y., (2013), "Can US economic variables predict the Chinese stock market?" Pacific-Basin Finance Journal, 22: 69-87.

Goyal A and Welch I (2003), "Predicting the equity premium with dividend ratios", *Management Science*, Vol. 49, No. 5, pp. 639-654.

Goyal A and Welch I (2008), "A comprehensive look at the empirical performance of the equity premium prediction." *Review of Financial Studies*, Vol. 21, No.4 pp.1455-1508.

Guidolin, M., Hyde, S., McMillan, D. G., and Ono, S., (2009), "Non-linear predictability in stock and bond returns: when and where is it exploitable?" *International Journal of Forecasting*, Vol. 25, No. 2, pp. 373-399.

Harvey, D. I., Leybourne, S. J., and Newbold, P. (1998), "Tests for forecast encompassing." *Journal of Business and Economic Statistics*, Vol. 16, No. 2, pp. 254–259.

Hillebrand, E., Lee, T. H., and Medeiros, M. C. (2013), "Bagging constrained equity premium predictors." *Essays in Nonlinear Time Series Econometrics*, (Festschrift for Timo Teräsvirta).

Inoue, A., and Kilian, L. (2005), "In-Sample or out-of-sample tests of predictability: which one should we use?" Econometric Reviews, Vol. 23, No. 4, pp. 371–402.

Inoue, A. and Kilian, L. (2008), "How useful is bagging in forecasting economic time series? A case study of U.S. CPI inflation." *Journal of the American Statistical Association*, Vol. 103, No. 482, pp. 511–522.

Jordan, S. J. (2012), "Time-varying Risk and Long-term Reversals." *Journal of International Business Studies*, Vol. 45, No. 2, pp. 123-142.

Jordan, S. J. and Vivian, A. J. (2011), "Forecasting stock returns internationally: can fundamental-price models beat the historical average?" *IFABS 2011 Conference (Rome) Paper*

Jordan, S. J., Vivian, A. J. and Wohar, M. E. (2014a), "Sticky prices or economically-linked economies: The case of forecasting the Chinese stock market." *Journal of International Money and Finance,* Vol 41, pp. 95-109.

Jordan, S. J., Vivian, A. J., & Wohar, M. E. (2014b). "Forecasting returns: new European evidence". *Journal of Empirical Finance*, Vol. 26, pp. 76-95.

Jordan, S. J., Vivian, A. J. and Wohar, M. E. (2016), "Can commodity returns forecast Canadian sector stock returns?" *International Review of Economics & Finance* Vol. 41, pp. 172-188.

Kilian, L. 1999. Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-horizon Regressions? *Journal of Applied Econometrics*, Vol. 14, No. 5, pp. 491-510.

La Porta, R., Lopez-de-Silanes F. and Shleifer, A. (2008) The Economic Consequences of Legal Origins. *Journal of Economic Literature*, Vol. 46, No. 2, pp. 285-332.

Lee, T. H., and Yang, Y., (2006). "Bagging binary and quantile predictors for time series." *Journal of Econometrics*, Vol. 135 No. 1-2, pp. 465–497.

Levich, R. M. (2001). *International Financial Markets*. 2nd Edition. Mc-Graw Hill. New York.

Lewis-Beck, M. S., and Tien, C. (2012). Japanese election forecasting: Classic tests of a hard case. *International Journal of Forecasting*, Vol. 28, No. 4, pp.797-803.

Li, K., Wang, T., Cheung, Y. L., and Jiang, P. (2011). Privatization and Risk Sharing: Evidence from the Split Share Structure Reform in China. *Review of Financial Studies*, Vol. 24, No. 7, pp. 2499-2525.

Lo, A. W., MacKinlay, A. C. (1990). Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies,* Vol. 3, pp. 431-467.

Loungani, P. (2001). "How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth". *International Journal of Forecasting*, Vol. 17, pp. 419–432.

Loungani, P., Stekler, H., & Tamirisa, N. (2013). "Information rigidity in growth forecasts: some cross-country evidence". *International Journal of Forecasting*, Vol. 29, pp. 605–621.

McCracken, M. W. (2007), "Asymptoptics for Out-of-Sample Tests of Granger Causality," *Journal of Econometrics*, Vol. 140, No.2, pp. 719-752.

McMillan, D.G. and Wohar, M. E. (2011), "Sum of the parts stock return forecasting: international evidence.", *Applied Financial Economics*, Vol. 21, No. 12, pp. 837-845.

Newey, W. K., and West, K. D. (1987), "A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix." *Econometrica*, Vol. 55, No. 3, pp. 703–708.

Paye, B. S., and Timmermann, A. (2006). "Instability of return prediction models." *Journal of Empirical Finance*, Vol 13, No. 3, pp. 274-315.

Prskawetz, A., Kögel, T., Sanderson, W. C., and Scherbov, S. (2007). The effects of age structure on economic growth: An application of probabilistic forecasting to India. *International Journal of Forecasting*, Vol. 23, No. 4, pp. 587-602.

Qin, D., Cagas, M. A., Ducanes, G., Magtibay-Ramos, N., and Quising, P. (2008). "Automatic leading indicators versus macroeconometric structural models: A comparison of inflation and GDP growth forecasting". *International Journal of Forecasting*, Vol. *24,* No. 3, pp. 399-413.

Rapach, D. E., Strauss, J. K. (2008), "Forecasting U.S. employment growth using forecast combining methods. " *Journal of Forecasting*, Vol. 27, No. 1, pp. 75-93.

Rapach, D. E., Strauss, J. K. (2010), "Bagging or combining (or both)? An analysis based on forecasting U.S. employment growth." *Econometric Reviews*, Vol. 29, No. 5-6, pp. 511-533.

Rapach, D. E., Strauss, J. K., and Zhou, G. F. (2010), "Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy." *Review of Financial Studies*, Vol. 23, No. 2. pp. 821-862.

Rapach, D. E. and Weber, C. E. (2004). Financial Variables and the Simulated Out-of-sample Forecastability of U.S. Output Growth Since 1985: An Encompassing Approach. *Economic Inquiry*, 42, 717-738.

Rapach, D.E. and M.E. Wohar (2006), "In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining" *Journal of Empirical Finance* 13 (March): 231-247.

Rapach, D. E. and Wohar, M. E. (2009), "Multi-period portfolio choice and the intertemporal hedging demands for stocks and bonds: International evidence." *Journal of International Money and Finance,* Vol. 28, No. 3, pp. 427-453.

Rapach, D. E., Wohar, M. E. and Rangvid J. (2005) "Macro Variables and International Stock Return Predictability." *International Journal of Forecasting*, Vol. 21, No. 1, pp. 137-166.

Sims, C. A. (2003). "Implications of rational inattention". *Journal of Monetary Economics*, Vol. 50, No. 3, pp. 665-690.

Stock, J. H., and Watson, M. W. (2003), "Forecasting output growth and inflation: The role of asset prices." *Journal of Economic Literature* , Vol. 41, No. 3, pp. 788–829.

Stock, J. and Watson, M. (2004), "Combination forecasts of output growth in a seven-country data set", *Journal of Forecasting*, Vol. 23, No. 6, pp. 405-430.

Wang, X., Wu, C., and Xu, W. (2015). "Volatility forecasting: The role of lunch-break returns, overnight returns, trading volume and leverage effects". *International Journal of Forecasting*, Vol. 31 No. 3, pp.609-619.

White, H. (2000), "A reality check for data-snooping", *Econometrica*, Vol. 68, No. 5, pp. 1097-1126.

Woodford M (2002), 'Imperfect Common Knowledge and the Effects of Monetary Policy', in E Phelps, A Alchian and C Holt (eds), Knowledge, Information, and Expectations in Modern Macroeconomics: In Honour of Edmund S. Phelps, Princeton University Press, Princeton, pp 25–58.

Yang, Y. (2004), "Combining forecasting procedures: some theoretical results." *Econometric Theory*, Vol. 20, No.1,pp. 176-222.

Table 1: Descriptive Statistics (1995:1-2011:6)

		CH	HK	IN	ID	KO	MY	PH	SG	TH	TW	CN	DE	FR	IT	JP	UK	US
RET	**Mean**	0.0099	0.0076	0.0099	0.0111	0.0073	0.0052	0.0049	0.0047	0.0022	0.0033	0.0090	0.0061	0.0075	0.0049	-0.0015	0.0066	0.0073
	SD	0.1048	0.0754	0.0847	0.0937	0.0926	0.0726	0.0723	0.0643	0.1002	0.0775	0.0450	0.0568	0.0539	0.0614	0.0519	0.0422	0.0471
DP	**Mean**	-3.556	-3.530	-4.215	-3.831	-4.124	-3.664	-4.164	-3.755	-3.681	-3.858	-3.865	-3.873	-3.519	-3.548	-4.548	-3.427	-4.087
	SD	0.477	0.220	0.300	0.402	0.327	0.347	0.500	0.277	0.578	0.587	0.237	0.305	0.247	0.433	0.391	0.180	0.234
DY		-3.549	-3.525	-4.207	-3.822	-4.119	-3.661	-4.161	-3.752	-3.681	-3.856	-3.858	-3.869	-3.514	-3.546	-4.550	-3.423	-4.082
		0.469	0.218	0.305	0.411	0.313	0.346	0.505	0.277	0.579	0.587	0.233	0.299	0.241	0.422	0.387	0.178	0.237
EP		-2.374	-2.690	-2.780	-2.696	-2.572	-2.761	-2.777	-2.751	-2.542	-2.919	-2.884	-2.753	-2.731	-2.822	-3.604	-2.745	-3.018
		0.430	0.244	0.267	0.330	0.331	0.263	0.257	0.325	0.417	0.345	0.175	0.241	0.227	0.308	0.439	0.257	0.218
BM			0.617	0.473	0.349	0.902	0.560	0.639	0.676	0.508	0.471	0.498	0.524	0.551	0.640	0.646	0.451	0.353
			0.118	0.133	0.109	0.217	0.122	0.211	0.120	0.158	0.120	0.078	0.163	0.151	0.226	0.159	0.125	0.098
RF		0.003	0.003	0.007	0.012	0.005	0.003	0.007	0.002	0.004	0.003	0.003	0.002	0.003	0.003	0.000	0.004	0.003
		0.002	0.002	0.006	0.013	0.004	0.002	0.003	0.001	0.004	0.002	0.001	0.001	0.001	0.002	0.000	0.002	0.002
INFL		0.031	0.089	0.005	0.010	0.003	0.002	0.005	0.001	0.003	0.001	0.079	0.081	0.080	0.079	0.000	0.082	0.025
		0.047	0.005	0.007	0.012	0.001	0.001	0.003	0.002	0.002	0.001	0.008	0.006	0.006	0.008	0.001	0.007	0.012
SVAR		0.045	0.033	0.038	0.039	0.040	0.028	0.031	0.026	0.042	0.035	0.021	0.026	0.026	0.028	0.027	0.022	0.023
		0.017	0.012	0.012	0.015	0.016	0.016	0.009	0.010	0.015	0.011	0.008	0.010	0.010	0.009	0.008	0.009	0.009
NTIS		0.023	0.005	0.007	0.010	0.007	0.004	0.005	0.004	0.007	0.006	0.002	0.002	0.003	0.005	0.002	0.002	0.002
		0.081	0.012	0.028	0.034	0.016	0.011	0.020	0.019	0.017	0.017	0.012	0.016	0.010	0.019	0.006	0.009	0.007
PRES		1.012	0.977	0.988	0.982	0.974	0.998	0.954	0.972	0.959	0.987	1.013	0.995	0.984	0.970	0.953	1.036	1.030
		0.276	0.231	0.231	0.253	0.229	0.272	0.262	0.238	0.257	0.259	0.154	0.169	0.173	0.210	0.176	0.238	0.176
CVm		0.125	0.059	0.067	0.138	0.080	0.080	0.093	0.064	0.159	0.067	0.027	0.054	0.043	0.059	0.027	0.029	0.026
		0.608	0.331	0.354	0.574	0.391	0.427	0.463	0.331	0.884	0.403	0.175	0.363	0.254	0.346	0.192	0.218	0.160

Notes: This table shows the sample mean and standard deviations. The top row gives the country code. The first column gives the code for each variable. Mean is the sample average, SD is the standard deviation, Results are reported for the following countries: CH is China, HK is Hong Kong, IN is India, ID is Indonesia, JP is Japan, KO is Korea, MY is Malaysia, PH is Philippines, SG is Singapore, TH is Thailand, TW is Taiwan, CN is Canada, DE is Germany, FR is France, IT is Italy, JP is Japan, UK is United Kingdom and US is United States. RET is the stock return, DP is log dividend-price ratio, DY is log dividend-yield, EP is log earnings-price ratio, BM is book-to-market ratio, RF is the risk-free rate, INFL is inflation, SVAR is stock variance, NTIS is net equity issuance, PRES is price pressure measured by numbers of rising stocks divided by number of falling stocks, and CVm is change in volume at monthly frequency.

Table 2: Return Correlation Matrix

	CH	HK	IN	ID	KO	MY	PH	SG	TH	TW	CN	FR	BD	IT	JP	UK	US
CH	1.000																
HK	0.620	1.000															
IN	0.343	0.469	1.000														
ID	0.195	0.404	0.405	1.000													
KO	0.307	0.513	0.384	0.462	1.000												
MY	0.424	0.589	0.351	0.523	0.385	1.000											
PH	0.351	0.553	0.360	0.594	0.460	0.584	1.000										
SG	0.474	0.782	0.490	0.537	0.523	0.594	0.671	1.000									
TH	0.380	0.517	0.350	0.563	0.608	0.559	0.624	0.615	1.000								
TW	0.472	0.565	0.451	0.331	0.525	0.473	0.414	0.559	0.486	1.000							
CN	0.413	0.673	0.465	0.534	0.459	0.485	0.506	0.626	0.493	0.511	1.000						
FR	0.234	0.551	0.428	0.461	0.462	0.364	0.395	0.572	0.366	0.521	0.713	1.000					
BD	0.251	0.562	0.413	0.462	0.438	0.420	0.406	0.585	0.410	0.538	0.687	0.904	1.000				
IT	0.125	0.417	0.440	0.465	0.409	0.286	0.322	0.476	0.329	0.451	0.613	0.852	0.765	1.000			
JP	0.273	0.433	0.435	0.449	0.487	0.248	0.334	0.497	0.394	0.437	0.540	0.556	0.521	0.491	1.000		
UK	0.328	0.620	0.427	0.497	0.536	0.378	0.445	0.625	0.475	0.497	0.751	0.845	0.799	0.748	0.554	1.000	
US	0.352	0.650	0.417	0.490	0.512	0.424	0.477	0.646	0.479	0.524	0.821	0.787	0.791	0.664	0.539	0.827	1.000
AV. ALL	0.346	0.558	0.414	0.461	0.467	0.443	0.468	0.580	0.478	0.485	0.581	0.563	0.559	0.491	0.449	0.585	0.588
AV. ASIA	0.396	0.557	0.400	0.446	0.463	0.498	0.512	0.583	0.523	0.475							
AV. G7											0.688	0.776	0.744	0.689	0.533	0.754	0.738

Notes: This table shows the return correlations (Pearson) between different country pairs. Results are reported for the following countries: CH is China, HK is Hong Kong, IN is India, ID is Indonesia, KO is Korea, MY is Malaysia, PH is Philippines, SG is Singapore, TH is Thailand, TW is Taiwan, CN is Canada, DE is Germany, FR is France, IT is Italy, JP is Japan, UK is United Kingdom and US is United States. AV. denotes the average correlation between other countries and the country given in the column header.

Table 3: Individual Predictive Models Forecast Performance – 1 Month Horizon
Panel A: Asian Countries

		CH	HK	IN	ID	JP	KO	MY	PH	SG	TH	TW
HA (MSFE)		0.008	0.004	0.008	0.006	0.003	0.004	0.002	0.003	0.003	0.006	0.004
DP	**RMSFE|ENC-NEW**	1.038	0.998**	0.973**	0.995** $$	1.046	0.994	0.984** $$	0.963**	0.995*	0.975**	1.027**
	RMAFE|CW-T	1.012	1.009	1.006**	0.957**	1.016	0.991	0.959*	0.943**	0.947	0.963*	1.001
DY	**RMSFE|ENC-NEW**	1.052	0.966**	0.960** $$	0.981** $$	1.051	0.990**	0.987** $$	0.951** $$	0.970**	0.972**	1.030**
	RMAFE|CW-T	1.021	1.008*	1.000**	0.948**	1.011	0.973	0.962**	0.929**	0.936*	0.966**	1.001
EP	**RMSFE|ENC-NEW**	1.079	1.020	0.974**	1.000	1.029	0.985*	1.017	0.991*	1.026	0.989**	1.062
	RMAFE|CW-T	1.037	1.001	0.993**	1.005	1.014	0.989*	1.007	0.980	1.027	0.980*	1.012
SVAR	**RMSFE|ENC-NEW**	1.003**	1.007	1.006*	1.010	1.015	1.064	1.044	1.018	1.023	0.999	1.028
	RMAFE|CW-T	1.021	0.988	0.996	1.002	1.006	1.034	1.024	1.010	1.035	0.981	1.002
NTIS	**RMSFE|ENC-NEW**	1.010*	0.994	0.998	0.995	1.001	1.009	1.005	1.012	1.102	1.003	1.002
	RMAFE|CW-T	1.013	1.005	0.998	0.991	1.000	1.003	1.002	1.007	1.035	1.003	1.000
BM	**RMSFE|ENC-NEW**		0.983** $$	1.000	0.972** $$	0.999**	1.052**	0.960** $$	0.986*	0.944** $$	0.954** $$	0.954**
	RMAFE|CW-T		1.052**	1.020	1.013** $$	0.971	1.050	0.990**	1.002	0.985**	0.966**	0.973**
RF	**RMSFE|ENC-NEW**	1.050	1.044	1.016	1.003	0.970**	0.976**	0.943** $$	1.019	0.992*	0.945** $$	1.001
	RMAFE|CW-T	1.030	1.040	1.008	0.998	0.983** $$	0.978**	0.943** $$	1.009	0.990*	0.944** $$	0.993
INFL	**RMSFE|ENC-NEW**	1.000	0.977**	1.001	1.010	0.982**	0.960**	0.927** $$	1.047	0.916** $$	0.904** $$	0.928** $$
	RMAFE|CW-T	0.999	0.979**	0.995	1.008	0.993	0.979** $$	0.979**	1.018	1.057** $$	0.951** $$	0.979** $$
PRES	**RMSFE|ENC-NEW**	1.003*	0.988	1.002	0.956**	0.977**	1.003	0.917** $$	1.016**	0.933** $$	1.050	0.999
	RMAFE|CW-T	1.020	1.003	1.007	0.991**	0.978*	0.979	0.968** $$	1.006*	0.982**	1.017	1.001
CVM	**RMSFE|ENC-NEW**	1.004	0.997**	1.004	1.009	1.014	1.008	1.040	1.026	1.034	1.003	1.004
	RMAFE|CW-T	1.000	1.009	0.998	0.996	1.013	1.001	1.017	1.012	1.020	0.997	1.003

Panel B: G7 Countries

		CN	DE	FR	IT	JP	UK	US
HA (MSFE)		0.002	0.002	0.002	0.003	0.003	0.002	0.002
DP	**RMSFE | ENC-NEW**	1.048	1.037	1.022	1.033	1.046	1.039	1.073
	RMAFE | CW-T	**0.988**	1.009	**0.999**	1.007	1.016	**0.997**	1.005
DY	**RMSFE | ENC-NEW**	1.052	1.036	1.023	1.045	1.051	1.026	1.045
	RMAFE | CW-T	**0.986**	1.003	**0.993**	1.015	1.011	**0.995**	**0.990**
EP	**RMSFE | ENC-NEW**	1.027	1.023	1.005	1.038	1.029	1.059	1.108
	RMAFE | CW-T	**0.997**	**0.993**	1.002	**0.997**	1.014	**0.988**	1.039
SVAR	**RMSFE | ENC-NEW**	1.029	1.055	1.034	1.022	1.015	1.035	1.066
	RMAFE | CW-T	**0.995**	1.000	1.005	**0.989**	1.006	1.014	1.040
NTIS	**RMSFE | ENC-NEW**	1.015	1.005	1.009	1.008	1.001	**0.995**	1.015
	RMAFE | CW-T	1.013	1.005	1.002	1.006	1.000	1.005	1.003
BM	**RMSFE | ENC-NEW**		1.055	1.039	1.058	**0.999** **	1.081	1.043
	RMAFE | CW-T		**0.996**	1.003	1.022	**0.971**	1.005	**0.998**
RF	**RMSFE | ENC-NEW**	1.016	**0.980** ** $$	**0.988** **	1.006	**0.970** **	1.073	1.049
	RMAFE | CW-T	1.007	**0.979** ** $$	**0.974**	1.001	**0.983** ** $$	1.053	1.045
INFL	**RMSFE | ENC-NEW**	1.074	1.155	1.106	1.029	**0.982** **	1.108	**0.926** ** $$
	RMAFE | CW-T	1.079	1.091	1.057	1.016	**0.993**	1.084	**0.946** **
PRES	**RMSFE | ENC-NEW**	**0.990** *	**0.977** **	**0.987** *	1.005	**0.977** **	1.001	1.002
	RMAFE | CW-T	**0.997**	**0.983** *	**0.977**	1.004	**0.978** *	1.002	**0.995**
CVM	**RMSFE | ENC-NEW**	1.016	1.025	1.011 **	1.027	1.014	1.008 **	1.026
	RMAFE | CW-T	1.004	1.023	1.011	1.012	1.013	1.001	1.013

Notes: This table shows the forecast results for individual predictive regression models at the one-month forecast horizon. The top row gives the country code. The second row [HA(MSFE)] reports the mean-squared forecast error for the historical average benchmark model. The other rows report the ratios of the mean-squared forecast errors for each predictive regression model to the mean-squared forecast error of the Historical Average model. Bold typeface denotes if the mean-squared forecast error from the individual model is smaller than the historical average model. RMSFE is the relative mean-squared forecast error. RMAFE is relative mean absolute forecast error. ** and * shows significance at the 5% and 10% level respectively using conventional critical values. $$ and $ shows significance at the 5% and 10% level respectively using data-mining robust critical values. We implement the Clark-McCracken (2001) ENC-NEW encompassing test and the Clark-West (2007) CW-T test of equal forecast accuracy. Results are reported for the following countries: CH is China, HK is Hong Kong, IN is India, ID is Indonesia, JP is Japan, KO is Korea, MY is Malaysia, PH is Philippines, SG is Singapore, TH is Thailand, TW is Taiwan, CN is Canada, DE is Germany, FR is France, IT is Italy, JP is Japan, UK is United Kingdom and US is United States. Individual predictors are denoted as follows: DP is log dividend-price ratio, DY is log dividend-yield, EP is log earnings-price ratio, SVAR is stock variance, NTIS is net equity issuance, BM is book-to-market ratio, RF is the risk-free rate, INFL is inflation, PRES is price pressure measured by numbers of rising stocks divided by number of falling stocks, and CVM is change in volume at monthly frequency.

Table 4: Bagging and Combination Forecast Performance – 1 Month Horizon
Panel A: Asian Countries

		CH	HK	IN	ID	JP	KO	MY	PH	SG	TH	TW
BA	**RMSFE | ENC-NEW**	1.127	1.002 ** $$	1.001	0.944 ** $$	0.988 ** $	1.062 ** $$	0.888 ** $$	1.021 **	1.070 ** $$	0.939 ** $$	1.070 ** $$
	RMAFE | CW-T	1.078	1.083 *	1.001	0.985 ** $$	0.980 **	1.031 *	0.956 ** $$	1.000	1.077 ** $	0.951 ** $$	1.032 ** $
Mean	**RMSFE | ENC-NEW**	1.010	0.974 **	0.979 *	0.958 **	0.993	0.986	0.948 **	0.984 *	0.955 **	0.967 **	0.986 *
	RMAFE | CW-T	1.005	0.993 **	0.987 ** $	0.968 ** $$	0.988	0.990 **	0.963 ** $$	0.981 **	0.978 ** $$	0.969 ** $$	0.984 *
Median	**RMSFE | ENC-NEW**	1.011	0.969 **	0.990	0.980 *	0.990	0.987	0.961 **	0.990	0.969 **	0.971 **	0.979 *
	RMAFE | CW-T	1.003	0.989 ** $$	0.995 **	0.984 ** $$	0.992 ** $	0.989 *	0.973 ** $$	0.992 ** $	0.981 ** $$	0.973 ** $$	0.985 ** $
Trimmed	**RMSFE | ENC-NEW**	1.006	0.971 **	0.976 *	0.962 **	0.990	0.987	0.954 **	0.988	0.961 **	0.972 **	0.985 *
	RMAFE | CW-T	1.003	0.990 ** $$	0.986 ** $$	0.972 ** $$	0.989 *	0.989 ** $	0.966 ** $$	0.986 *	0.974 ** $$	0.971 ** $$	0.983 *
DMSFE(1)	**RMSFE | ENC-NEW**	1.009	0.974 **	0.978 *	0.958 **	0.992	0.986	0.948 **	0.984 *	0.955 **	0.967 **	0.986 *
	RMAFE | CW-T	1.005	0.994 **	0.987 ** $	0.966 ** $$	0.987	0.990 **	0.963 ** $$	0.981 *	0.978 ** $$	0.969 ** $$	0.984 *
DMSFE(0.9)	**RMSFE | ENC-NEW**	1.006	0.972 **	0.978 *	0.959 **	0.988	0.985	0.944 **	0.982 *	0.952 **	0.965 **	0.985 *
	RMAFE | CW-T	1.003	0.990 ** $	0.987 ** $	0.968 ** $$	0.986 *	0.990 ** $	0.962 ** $$	0.979 ** $	0.976 ** $$	0.968 ** $$	0.985 *
C(2,PB)	**RMSFE | ENC-NEW**	0.989	0.988	0.994 *	1.001	0.970 **	0.987	0.936 ** $	1.001	0.956 **	0.953 **	0.988 *
	RMAFE | CW-T	0.994	1.000 **	0.995	0.985	0.975 ** $$	0.990	0.956 ** $$	0.985	0.976 ** $$	0.952 ** $$	0.997 *
C(3,PB)	**RMSFE | ENC-NEW**	0.986 *	0.989 *	0.990 **	1.004 *	0.965 **	0.992	0.929 ** $$	1.002	0.937 ** $	0.945 **	0.987 **
	RMAFE | CW-T	0.995 *	1.000	0.999 *	0.990	0.975 ** $$	0.998	0.954 ** $$	0.983	0.963 ** $$	0.946 ** $$	0.998 *
PC(C,3B)	**RMSFE | ENC-NEW**	1.011 *	1.011 **	1.071 **	0.991 **	1.069	1.056	1.281	1.032	1.072 **	1.125 ** $$	1.131
	RMAFE | CW-T	1.028	1.012	1.077	0.993 *	1.027	1.042	1.125	0.983	1.081	1.022 *	1.044
AV COMB	**RMSFE | ENC-NEW**	0.998	0.971 **	0.985 *	0.970 **	0.990	0.988	0.958 **	0.991	0.957 **	0.956 **	0.994
	RMAFE | CW-T	1.001	0.989 ** $	0.989 *	0.970 ** $	0.985 *	0.993	0.969 ** $$	0.981	0.984 ** $$	0.960 ** $$	0.992

Panel B: G7 Countries

		CN	DE	FR	IT	JP	UK	US
BA	RMSFE| ENC-NEW	1.128	1.088**	1.068**	1.113	**0.988** ** $	1.051*	1.089
	RMAFE| CW-T	1.088	1.117	1.047	1.039	**0.980** **	1.012	1.040
Mean	RMSFE| ENC-NEW	1.009	1.002	**0.998**	1.014	**0.993**	1.015	1.013
	RMAFE| CW-T	1.000	**0.994**	**0.998**	1.002	**0.988**	1.003	**0.993**
Median	RMSFE| ENC-NEW	1.010	1.001	1.001	1.021	**0.990**	1.001	1.001
	RMAFE| CW-T	1.000	**0.994**	**0.996**	1.009	**0.992** ** $	**0.999**	**0.988**
Trimmed	RMSFE| ENC-NEW	1.002	1.003	**0.994**	1.011	**0.990**	1.014	1.015
	RMAFE| CW-T	**0.995**	**0.995**	**0.991**	1.002	**0.989** *	1.001	**0.992**
DMSFE(1)	RMSFE| ENC-NEW	1.009	1.002	**0.999**	1.014	**0.992**	1.015	1.014
	RMAFE| CW-T	1.000	**0.994**	**0.998**	1.002	**0.987**	1.003	**0.993**
DMSFE(0.9)	RMSFE| ENC-NEW	1.008	1.002	**0.999**	1.013	**0.988**	1.013	1.009
	RMAFE| CW-T	**0.999**	**0.993**	**0.997**	1.001	**0.986** *	1.001	**0.993**
C(2,PB)	RMSFE| ENC-NEW	1.001	**0.995**	1.004	1.003	**0.970** **	1.019	**0.983** *
	RMAFE| CW-T	**0.992**	**0.991**	**0.997**	**0.992**	**0.975** ** $$	**0.999**	**0.988** ** $
C(3,PB)	RMSFE| ENC-NEW	1.011	1.031	1.001	1.012	**0.965** **	1.054	**0.961** **
	RMAFE| CW-T	**0.991**	1.003	**0.996**	**0.987**	**0.975** ** $$	1.010	**0.979** ** $$
PC(C,3B)	RMSFE| ENC-NEW	1.147	1.057**	1.215*	1.080	1.069	1.201	1.037
	RMAFE| CW-T	1.069	1.080	1.156	1.052	1.027	1.088	1.017
AV COMB	RMSFE| ENC-NEW	1.015	**1.000**	**1.000**	1.015	**0.990**	1.021	1.001
	RMAFE| CW-T	1.002	**0.998**	1.008	1.002	**0.985** *	1.000	**0.991**

Notes: This table shows the forecast results for bagging and forecast combination methods at the one-month forecast horizon. The top row gives the country code. The other rows report the ratios of the mean-squared forecast errors for each ARDL models to the mean-squared forecast error of the historical average benchmark. Bold typeface denotes if the mean-squared forecast error from the individual model is smaller than the historical average benchmark. RMSFE is the relative mean-squared forecast error. RMAFE is relative mean absolute forecast error. ** and * shows significance at the 5% and 10% level respectively using conventional critical values. $$ and $ shows significance at the 5% and 10% level respectively using data-mining robust critical values. We implement the Clark-McCracken (2001) ENC-NEW encompassing test and the Clark-West (2007) CW-T test of equal forecast accuracy.

Results are reported for the following countries: CH is China, HK is Hong Kong, IN is India, ID is Indonesia, JP is Japan, KO is Korea, MY is Malaysia, PH is Philippines, SG is Singapore, TH is Thailand, TW is Taiwan, CN is Canada, DE is Germany, FR is France, IT is Italy, JP is Japan, UK is United Kingdom and US is United States. BA represents the bagging method. Combinations based on measures of central tendency are the Mean, Median and Trimmed (Mean). Combinations based on weighted past forecast errors are DMSFE(1) and DMSFE(0.9). Combinations based on cluster methods are C(2,PB) and C(3,PB). Combination based on principal components are PC(C,3B). AV COMB is the mean of all 8 combination methods.

Table 5: Economic Value of One Month Return Forecasts using Bagging or Forecast Combinations - Utility Gain Method.

Panel A: Asian Countries

			CH	HK	IN	ID	JP	KO	MY	PH	SG	TH	TW
BA	UG	| p(UG)	-0.07	6.29 ** $	-0.68	11.87 ** $$	2.95 **	5.04 ** $	9.93 ** $$	3.84 **	18.73 ** $$	12.11 ** $$	15.63 ** $$
	UG5	| p(UG5)	0.39	4.69 ** $$	-0.41	7.15 ** $$	1.08	2.50 ** $	8.48 ** $$	3.33 ** $$	11.95 ** $$	6.68 ** $$	10.42 ** $$
Mean	UG	| p(UG)	-1.99	4.70 **	2.58 **	5.68 ** $	0.04	2.21 **	7.34 ** $$	0.60	11.82 ** $$	3.49 **	1.76 **
	UG5	| p(UG5)	-1.20	3.43 ** $	1.55 **	3.42 ** $	0.03	1.32 **	4.66 ** $$	0.36	7.10 ** $$	2.10 **	1.06 **
DMSFE(1)	UG	| p(UG)	-1.76	4.45 **	2.74 **	5.61 ** $	0.10	2.19 **	7.34 ** $$	0.52	11.91 ** $$	3.51 **	1.82 **
	UG5	| p(UG5)	-1.06	3.28 **	1.65 **	3.38 ** $	0.06	1.31 **	4.66 ** $$	0.31	7.15 ** $$	2.11 **	1.10 **
C(2,PB)	UG	| p(UG)	3.69 **	1.97	2.87 **	0.18	4.10 **	1.61 **	8.06 ** $$	-2.82	9.82 ** $$	6.28 ** $$	3.89 **
	UG5	| p(UG5)	2.34 **	1.79 **	1.73 **	0.11	2.24 **	0.96 **	5.74 ** $$	-1.70	5.31 ** $$	3.77 ** $$	2.34 **
PC(C,3B)	UG	| p(UG)	4.59 **	2.60	7.20 ** $$	6.01 ** $$	-8.49	-0.72	6.80 ** $	-2.32	10.71 ** $$	1.93	-4.30
	UG5	| p(UG5)	3.03 ** $	-1.82	5.22 ** $$	4.17 ** $$	-6.13	-0.43	5.59 ** $$	-2.93	6.20 ** $$	-4.38	-2.58

Panel B: G7 Countries

			CN	DE	FR	IT	JP	UK	US
BA	UG	| p(UG)	-4.48	8.44 ** $$	5.58 ** $	3.23 **	2.95 **	6.38 ** $$	3.25
	UG5	| p(UG5)	-1.26	5.57 ** $$	2.07 **	-0.89	1.08	2.61 ** $	4.83 ** $$
Mean	UG	| p(UG)	-0.67	2.68 **	0.76	0.90	0.04	-1.74	1.69 **
	UG5	| p(UG5)	-0.85	1.61 **	1.30 **	-0.03	0.03	-4.03	0.44
DMSFE(1)	UG	| p(UG)	-0.69	2.88 **	0.78	0.87	0.10	-1.71	1.76 **
	UG5	| p(UG5)	-0.86	1.72 **	1.24 **	-0.06	0.06	-4.18	0.53
C(2,PB)	UG	| p(UG)	-0.06	3.38 **	2.06 **	3.88 **	4.10 **	0.63	4.20 **
	UG5	| p(UG5)	1.37 **	1.53 **	1.23 **	1.17 **	2.24 **	-1.36	6.27 ** $$
PC(C,3B)	UG	| p(UG)	-8.50	5.57 **	3.97 **	0.10	-8.49	0.70	-1.68
	UG5	| p(UG5)	-6.85	3.21 **	2.57 **	-1.15	-6.13	-0.89	-0.73

Notes: Table 5 reports the economic significance of regression forecasts using the utility gain measure. Results are reported for the coefficient of relative risk aversion of 3 (UG) and of 5 (UG5). Utility gains are in percentage points, which are annualized by multiplying monthly values by 12. Utility gains are calculated for an investor with mean-variance and limits the weight in the risky asset to be no less than 0 and no more than 1.5 (as in Campbell and Thompson, 2008). We forecast the return and use the current market risk-free rate to calculate the excess return. ** and * shows significance at the 5% and 10% level respectively using bootstrapped critical values. $$ and $ shows significance at the 5% and 10% level respectively using data-mining robust bootstrap critical values. CH is China, HK is Hong Kong, IN is India, ID is Indonesia, JP is Japan, KO is Korea, MY is Malaysia, PH is Philippines, SG is Singapore, TH is Thailand, TW is Taiwan, CN is Canada, DE is Germany, FR is France, IT is Italy, JP is Japan, UK is United Kingdom and US is United States. BA is the bagging method. The combination reported based on central tendency is the Mean, the combinations based on weighted past forecast errors is DMSFE(1). The combinations based on cluster methods is C(2,PB) and the combination based on principal components is PC(C,3B).

Table 6: Economic Value of One Month Return Forecasts using Bagging or Forecast Combinations - GISW method.
Panel A: Asian Countries

	CH	HK	IN	ID	JP	KO	MY	PH	SG	TH	TW
BA GISW |p(GISW)	1.88	11.60 ** $$	-0.63	1.84	6.37 ** $	3.67	6.44 **	-1.43	20.48 ** $$	-19.48	15.64 ** $$
GISW5 |p(GISW5)	0.75	5.76 ** $	-0.39	3.95 ** $	6.05 ** $$	2.20	6.70 ** $$	1.33	13.17 ** $$	0.69	12.65 ** $$
Mean GISW |p(GISW)	-0.83	6.85 ** $	3.17 **	-4.47	1.11	2.24 **	3.22 **	-0.36	14.62 ** $$	3.22 **	1.55 **
GISW5 |p(GISW5)	-0.97	3.90 **	1.59 **	0.89 **	0.72	1.33 **	3.39 **	0.09	7.64 ** $$	2.07 **	1.54 **
DMSFE(1) GISW |p(GISW)	-0.64	6.39 **	3.32 **	-6.36	1.27 **	2.23 **	3.31 **	-0.52	14.69 ** $$	3.23 **	1.62 **
GISW5 |p(GISW5)	-0.84	3.73 **	1.69 **	0.42	0.83 **	1.32 **	3.42 **	0.02	7.70 ** $$	2.08 **	1.58 **
C(2,PB) GISW |p(GISW)	4.17 **	3.04 **	2.24 **	-22.03	5.64 **	0.72	1.61	-7.08	10.03 ** $$	5.59 **	3.85 **
GISW5 |p(GISW5)	2.49 **	2.17 **	1.64 **	-4.72	3.65 **	0.84	3.38 **	-2.76	5.14 ** $	3.72 **	2.62 **
PC(C,3B) GISW |p(GISW)	0.78	-4.34	3.62 **	-21.53	-4.11	-1.84	5.35 **	-22.06	12.84 ** $$	-27.95	-4.50
GISW5 |p(GISW5)	2.64	-7.35	4.91 ** $	-4.02	-1.59	-0.60	4.75 **	-12.84	7.76 ** $$	-28.19	-2.48

Panel B: G7 Countries

	CN	DE	FR	IT	JP	UK	US
BA GISW |p(GISW)	3.13	9.15 ** $$	6.15 ** $	4.18 **	6.37 ** $	6.28 **	4.19
GISW5 |p(GISW5)	1.58	5.80 ** $	3.48	1.40	6.05 ** $$	2.31	5.90 ** $
Mean GISW |p(GISW)	0.20	3.00 **	0.71	0.31	1.11	-2.84	1.28 **
GISW5 |p(GISW5)	-1.76	1.64 **	1.25 **	-0.23	0.72	-4.76	0.73 **
DMSFE(1) GISW |p(GISW)	0.20	3.21 **	0.73	0.29	1.27 **	-2.81	1.34 **
GISW5 |p(GISW5)	-1.74	1.76 **	1.20 **	-0.25	0.83 **	-4.97	0.82 **
C(2,PB) GISW |p(GISW)	1.11	2.48 **	2.21 **	3.22 **	5.64 **	1.45 **	4.34 **
GISW5 |p(GISW5)	2.29 **	1.39 **	1.16 **	1.08 **	3.65 **	-1.51	6.81 ** $$
PC(C,3B) GISW |p(GISW)	-4.77	5.25 **	4.74 **	-0.55	-4.11	1.07	-1.58
GISW5 |p(GISW5)	-8.34	3.45 **	3.17 **	-1.16	-1.59	-1.30	-0.63

Notes: This table reports the economic significance of regression forecasts using the GISW certainty equivalent measure of abnormal performance (see Goetzmann, Ingersoll, Spiegel and Welch, 2007), which is robust to manipulation motives of agents. Gains are in percentage points, which are annualized by multiplying monthly values by 12. Gains are reported for portfolios where the weights are based on mean-variance preferences and a relative risk aversion coefficient of 3 (GISW) and 5 (GISW5). The weight in the risky asset is limited to be no less than 0 and no more than 1.5 (as in Campbell and Thompson, 2008). We forecast the return and use the current market risk-free rate to calculate the excess return.
CH is China, HK is Hong Kong, IN is India, ID is Indonesia, JP is Japan, KO is Korea, MY is Malaysia, PH is Philippines, SG is Singapore, TH is Thailand, TW is Taiwan, CN is Canada, DE is Germany, FR is France, IT is Italy, JP is Japan, UK is United Kingdom and US is United States. BA is the bagging method. The combination reported based on central tendency is the Mean, the combinations based on weighted past forecast errors is DMSFE(1). The combinations based on cluster methods is C(2,PB) and the combination based on principal components is PC(C,3B).

Table 7: Economic Value of One Month Return Forecasts using Bagging or Forecast Combinations - Utility Gain Method – Transaction Cost adjusted.

Panel A: Asian Countries

			CH	HK	IN	ID	JP	KO	MY	PH	SG	TH	TW
BA UG	| p(UG)		-0.26	5.95 **	-0.72	11.74 ** $$	2.73 **	4.89 ** $	9.76 ** $$	3.09 **	18.47 ** $$	11.72 ** $$	15.43 ** $$
UG5	| p(UG5)		0.26	4.43 ** $	-0.44	7.07 ** $$	0.85	2.41 ** $	8.30 ** $$	2.80 ** $	11.70 ** $$	6.39 ** $$	10.20 ** $$
Mean UG	| p(UG)		-2.08	4.52 **	2.61 **	5.60 ** $	-0.11	2.16 **	7.14 ** $$	0.44	11.68 ** $$	3.44 **	1.70 **
UG5	| p(UG5)		-1.25	3.33 **	1.57 **	3.37 ** $	-0.06	1.30 **	4.52 ** $$	0.27	7.01 ** $$	2.07 **	1.02 **
DMSFE(1) UG	| p(UG)		-1.86	4.27 **	2.77 **	5.52 ** $	-0.06	2.14 **	7.13 ** $$	0.36	11.77 ** $$	3.46 **	1.77 **
UG5	| p(UG5)		-1.12	3.18 **	1.66 **	3.33 ** $	-0.03	1.29 **	4.52 ** $$	0.22	7.07 ** $$	2.08 **	1.06 **
C(2,PB) UG	| p(UG)		3.52 **	1.72	2.86 **	-0.01	3.92 **	1.52 **	7.84 ** $$	-3.04	9.58 ** $$	6.21 ** $$	3.75 **
UG5	| p(UG5)		2.24 **	1.64 **	1.73 **	0.00	2.10 **	0.91 **	5.51 ** $$	-1.83	5.14 ** $$	3.73 ** $$	2.25 **
PC(C,3B) UG	| p(UG)		4.23 **	1.98	7.11 ** $$	5.47 ** $	-8.80	-0.84	6.32 ** $	-2.67	10.58 ** $$	1.61	-4.35
UG5	| p(UG5)		2.80 **	-2.42	5.12 ** $$	3.78 ** $$	-6.37	-0.51	5.16 ** $$	-3.24	6.12 ** $$	-4.65	-2.61

Panel B: G7 Countries

			CN	DE	FR	IT	JP	UK	US
BA UG	| p(UG)		-4.85	8.40 ** $$	5.09 ** $	2.89 **	2.73 **	5.59 ** $$	3.02
UG5	| p(UG5)		-1.57	5.51 ** $$	1.64	-1.21	0.85	1.79 **	4.65 ** $$
Mean UG	| p(UG)		-0.75	2.58 **	0.60	0.84	-0.11	-1.94	1.66 **
UG5	| p(UG5)		-0.92	1.55 **	1.20 **	-0.08	-0.06	-4.17	0.39
DMSFE(1) UG	| p(UG)		-0.76	2.77 **	0.62	0.80	-0.06	-1.91	1.73 **
UG5	| p(UG5)		-0.92	1.66 **	1.14 **	-0.11	-0.03	-4.34	0.48
C(2,PB) UG	| p(UG)		-0.16	3.17 **	1.84 **	3.78 **	3.92 **	0.40	4.05 **
UG5	| p(UG5)		1.28 **	1.39 **	1.08 **	1.10 **	2.10 **	-1.54	6.14 ** $$
PC(C,3B) UG	| p(UG)		-8.72	5.55 **	3.55 **	-0.29	-8.80	0.36	-1.89
UG5	| p(UG5)		-7.07	3.17 **	2.22 **	-1.43	-6.37	-1.23	-0.90

Notes: This table reports the economic significance of regression forecasts using the utility gain measure with transaction costs of 10 basis points incurred on portfolio weight adjustmens. Results are reported for the coefficient of relative risk aversion of 3 (UG) and of 5 (UG5). Utility gains are in percentage points, which are annualized by multiplying monthly values by 12. Utility gains are calculated for an investor with mean-variance and limits the weight in the risky asset to be no less than 0 and no more than 1.5 (as in Campbell and Thompson, 2008). We forecast the return and use the current market risk-free rate to calculate the excess return. ** and * shows significance at the 5% and 10% level respectively using bootstrapped critical values. $$ and $ shows significance at the 5% and 10% level respectively using data-mining robust bootstrap critical values. CH is China, HK is Hong Kong, IN is India, ID is Indonesia, JP is Japan, KO is Korea, MY is Malaysia, PH is Philippines, SG is Singapore, TH is Thailand, TW is Taiwan, CN is Canada, DE is Germany, FR is France, IT is Italy, JP is Japan, UK is United Kingdom and US is United States. BA is the bagging method. The combination reported based on central tendency is the Mean, the combinations based on weighted past forecast errors is DMSFE(1). The combinations based on cluster methods is C(2,PB) and the combination based on principal components is PC(C,3B).

Table 8: Bagging and Combination Forecast Performance – 1 Month Horizon – WITH PORTFOLIO WEIGHT RESTRICTIONS APPLIED

Panel A: Asian Countries

		CH	HK	IN	ID	JP	KO	MY	PH	SG	TH	TW
BA	RMSFE | ENC-NEW	1.000	0.985 **	1.013	0.938 ** $$	0.983 **	0.972 **	0.947 ** $	0.948 ** $	0.920 ** $$	0.934 ** $$	0.993 ** $$
	RMAFE | CW-T	1.002	1.026 *	1.003	0.968 ** $$	0.969 *	0.967 *	0.959 ** $$	0.953 ** $$	0.957 ** $$	0.949 ** $$	0.992 ** $$
Mean	RMSFE | ENC-NEW	1.008	0.975 *	0.998	0.958 **	0.991	0.985	0.950 **	0.955 **	0.953 **	0.963 **	0.986 *
	RMAFE | CW-T	1.004	0.994 **	0.997	0.968 ** $$	0.986	0.990 ** $	0.964 ** $$	0.963 ** $$	0.975 ** $$	0.966 ** $$	0.984 *
DMSFE(1)	RMSFE | ENC-NEW	1.007	0.975 *	0.997	0.958 **	0.991	0.985	0.951 **	0.955 **	0.952 **	0.963 **	0.986 *
	RMAFE | CW-T	1.004	0.994 **	0.996	0.966 ** $$	0.986	0.990 **	0.965 ** $$	0.963 ** $$	0.974 ** $$	0.966 ** $$	0.984 *
C(2,PB)	RMSFE | ENC-NEW	0.985	0.988	0.997	0.991 *	0.979 *	0.986	0.952 **	0.967 **	0.954 **	0.952 **	0.987 *
	RMAFE | CW-T	0.993 *	0.999 **	0.992	0.978	0.977 ** $$	0.990 *	0.958 ** $$	0.965 ** $$	0.966 ** $$	0.952 ** $$	0.996 *
PC(C,3B)	RMSFE | ENC-NEW	0.990 **	0.979 **	0.983 **	0.978 **	1.028	1.012	0.977 **	0.961 **	0.951 **	0.968 ** $$	1.015
	RMAFE | CW-T	1.010	0.984 *	0.987 *	0.966 **	0.997	1.007	0.987 ** $$	0.948 ** $	0.991 ** $$	0.971 ** $$	1.005

Panel B: G7 Countries

		CN	DE	FR	IT	JP	UK	US
BA	RMSFE | ENC-NEW	1.008	0.979 **	0.984 **	1.000	0.983 **	0.982 *	0.974 **
	RMAFE | CW-T	1.026	1.004 ** $	0.995 *	0.994	0.969 *	0.979 ** $	1.001 ** $$
Mean	RMSFE | ENC-NEW	0.988	0.992	0.993	1.002	0.991	1.004	0.976 *
	RMAFE | CW-T	1.003	0.996 *	0.995	1.001	0.986	0.998	0.987 ** $$
DMSFE(1)	RMSFE | ENC-NEW	0.988	0.991	0.993	1.002	0.991	1.004	0.976 *
	RMAFE | CW-T	1.003	0.996 *	0.995	1.001	0.984	0.998	0.986 ** $$
C(2,PB)	RMSFE | ENC-NEW	0.985 *	0.988	0.990	0.994	0.978 **	1.005	0.971 **
	RMAFE | CW-T	0.999	0.991 ** $	0.993	0.993	0.975 ** $$	0.993	0.987 ** $$
PC(C,3B)	RMSFE | ENC-NEW	1.025	0.992 *	0.988 **	1.011	0.988 *	1.003	0.987 *
	RMAFE | CW-T	1.040	1.010	1.016	1.015	0.983 *	0.987	1.007 *

Notes: This table shows forecast results for bagging and forecast combination methods at the one-month forecast horizon when the portfolio weight restrictions are applied. The top row gives the country code. The second row [AR(MSFE)] reports the mean-squared forecast error for the historical average benchmark model. The other rows report the ratios of the mean-squared forecast errors for each predictive regression model to the mean-squared forecast error of the historical average benchmark. Bold typeface denotes if the mean-squared forecast error from the individual model is smaller than the historical average benchmark. RMSFE is the relative mean-squared forecast error. RMAFE is relative mean absolute forecast error. ** and * shows significance at the 5% and 10% level respectively using conventional critical values. $$ and $ shows significance at the 5% and 10% level respectively using data-mining robust critical values. We implement the Clark-McCracken (2001) ENC-NEW encompassing test and the Clark-West (2007) CW-T test of equal forecast accuracy. Results are reported for the following countries: CH is China, HK is Hong Kong, IN is India, ID is Indonesia, JP is Japan, KO is Korea, MY is Malaysia, PH is Philippines, SG is Singapore, TH is Thailand, TW is Taiwan, CN is Canada, DE is Germany, FR is France, IT is Italy, JP is Japan, UK is United Kingdom and US is United States. BA represents the bagging method. The combination reported based on central tendency is the Mean. The combinations based on weighted past forecast errors is DMSFE(1). The combinations based on cluster methods is C(2,PB). The combination based on principal components is PC(C,3B).

Table 9: Encompassing Tests
Panel A: Encompassing of Combination Forecasts – Asia

	CH	HK	IN	ID	JP	KO	MY	PH	SG	TH	TW
Mean	-0.018	0.777	6.112 **	0.003	0.239	0.292	0.415	0.309	-0.376	0.306	0.548 **
DMSFE(1)	0.027	0.774	6.221 **	0.011	0.249	0.289	0.401	0.301	-0.365	0.306	0.549 **
C(2,PB)	1.357	0.408	4.217 *	-0.261	0.614	0.269	0.381	-0.077	-0.494	0.356	0.538 **
PC(C,3B)	0.726	0.580	1.243 **	-0.012	-0.937	-0.038	-0.005	0.030	-0.290	-0.299	0.408 *

Panel B: Encompassing of Combination Forecasts – G7

	CN	DE	FR	IT	JP	UK	US
Mean	1.291 **	-0.034	0.227	0.458	0.239	-1.988	0.439
DMSFE(1)	1.294 **	-0.011	0.233	0.457	0.249	-2.021	0.451
C(2,PB)	1.359 **	0.200	0.354	0.652	0.614	-1.378	0.607
PC(C,3B)	-0.311	-1.770	0.340	0.317	-0.937	-0.626	-1.228

Notes: This table reports results of forecast encompassing tests on pairs of bagging and one of the combination methods. The optimal weight on the forecast combination is reported when generating a forecast comprising the forecast combination method and bagging. For example for CH-DMSFE(1) this indicates a weight of 0.027 on the mean and 0.973 on bagging. Results are reported for the HLN forecast encompassing test (Harvey, Leybourne and Newbold, 1998). The null hypothesis is that the combination return forecast has a weight of zero when combined with the bagging forecast. ** and * shows significance at the 5% and 10% level respectively.

For example for Panel A TW-Mean, ** indicates that we can reject the null hypothesis that the forecast based on the Mean combination has a weight of zero when included with the bagging forecast. Thus, the Mean forecast is not encompassed by the bagging forecast, i.e., the Mean forecast adds new information over and above that contained in the bagging forecast.

Table 10: Cross-sectional Regressions across 17 countries of Forecast measures on Country-level variables

Panel A: RMSFE

	GDP	TRADE	FDI	MKT DEV	TURN	INT USE	MOB PH	ANTI SD	PROP RT	CMN LAW
BA	0.226	-0.486 **	-0.205	-0.053	0.441 *	0.156	-0.035	-0.323	0.208	-0.133
Mean	0.238	-0.548 **	-0.272	-0.230	0.584 **	0.178	0.049	-0.297	0.221	-0.268
DMSFE(1)	0.235	-0.552 **	-0.275	-0.228	0.581 **	0.180	0.053	-0.299	0.221	-0.269
C(2,PB)	0.047	-0.469 *	-0.175	-0.115	0.362	0.065	0.066	-0.318	0.155	-0.316
PC(C,3B)	0.259	-0.556 **	-0.414 *	-0.201	0.433 *	0.475 *	0.108	-0.315	0.318	-0.306

Panel B: UG

	GDP	TRADE	FDI	MKT DEV	TURN	INT USE	MOB PH	ANTI SD	PROP RT	CMN LAW
BA	-0.299	0.764 **	0.513 **	0.346	-0.358	-0.023	0.221	0.445 *	0.064	0.335
Mean	-0.046	0.753 **	0.653 **	0.682 **	-0.260	0.164	0.372	0.397	0.252	0.362
DMSFE(1)	-0.311	0.790 **	0.546 **	0.393	-0.387	-0.018	0.245	0.469 *	0.050	0.324
C(2,PB)	0.305	0.356	0.202	0.121	0.062	-0.020	0.004	0.406	0.023	0.327
PC(C,3B)	-0.214	0.719 **	0.515 **	0.358	-0.259	-0.017	0.253	0.522 **	0.079	0.342

Panel C: GISW

	GDP	TRADE	FDI	MKT DEV	TURN	INT USE	MOB PH	ANTI SD	PROP RT	CMN LAW
BA	-0.166	0.823 **	0.698 **	0.505 **	-0.083	0.266	0.429 *	0.368	0.409	0.415 *
Mean	0.030	0.637 **	0.749 **	0.771 **	-0.066	0.183	0.414 *	0.235	0.336	0.244
DMSFE(1)	-0.210	0.865 **	0.732 **	0.528 **	-0.165	0.211	0.416 *	0.414 *	0.345	0.385
C(2,PB)	0.422 *	0.167	0.205	0.228	0.418 *	0.453 *	0.267	0.227	0.513 **	0.362
PC(C,3B)	-0.010	0.641 **	0.542 **	0.369	0.060	0.360	0.391	0.420 *	0.478 *	0.387

Notes: This table reports the correlations between measures of forecast performance and country characteristics across all 17 countries we examined (11 Asian countries and the 6 non-Asian members of the G7). Panel A reports results for the correlation between country characteristics and relative mean-squared forecast error. Panel B (C) reports results for the correlation between country characteristics and utility gain (GISW) with a coefficient of relative risk aversion equal to 3. The country characteristics are GDP per capita (GDP), TRADE (Total trade/Total GDP), FDI (Foreign Direct Investment/Total GDP), MKT DEV is stock market development (Market Capitalization/GDP), TURN (turnover calculated as Trading Volume/MV), INTUSE (Proportion of Internet Users), MOBPH (Proportion of Mobile Phone Users), ANTISD (Anti Self-Dealing), PROP RT (Index of Property Rights) and CMN LAW (Common law legal tradition dummy). Country characteristics are from the World Development Indicators, with the exception of ANTISD, PROP RT and CMNLAW which are from La Porta et al. (2008). **, * indicate statistical significance at the 5% and 10% level respectively.