UNIVERSITY of GUELPH

DEPARTMENT OF ECONOMICS AND FINANCE

DISCUSSION PAPER 2012-06

Is there an optimal forecast combination? A stochastic dominance approach to the forecasting combination puzzle.

Mehmet Pinar

Thanasis Stengos

M. Ege Yazgan

APRIL, 2012

CME

College of Management and Economics | Guelph Ontario | Canada | N1G 2W1
www.uoguelph.ca/economics

Is there an optimal forecast combination? A stochastic dominance approach to forecast combination puzzle.

Mehmet Pinar [*]
University of Guelph and
Fondazione Eni Enrico Mattei

Thanasis Stengos[†]
University of Guelph

M. Ege Yazgan[‡]
Istanbul Bilgi University

April 16, 2012

Abstract

The forecast combination puzzle refers to the finding that a simple average forecast combination outperforms more sophisticated weighting schemes and/or the best individual model. The paper derives optimal (worst) forecast combinations based on stochastic dominance (SD) analysis with differential forecast weights. For the optimal (worst) forecast combination, this index will minimize (maximize) forecasts errors by combining time-series model based forecasts at a given probability level. By weighting each forecast differently, we find the optimal (worst) forecast combination that does not rely on arbitrary weights. Using two exchange rate series on weekly data for the Japanese Yen/U.S. Dollar and U.S. Dollar/Great Britain Pound for the period from 1975 to 2010 we find that the simple average forecast combination is neither the worst nor the best forecast combination something that provides partial support for the forecast combination puzzle. In that context, the random walk model is the model that consistently contributes with considerably more than an equal weight to the worst forecast combination for all variables being forecasted and for all forecast horizons, whereas a flexible Neural Network autoregressive model and a self-exciting threshold autoregressive model always enter the best forecast combination with much greater than equal weights.

JEL Classifications: C12; C13; C14; C15; G01

Key Words: Nonparametric Stochastic Dominance,
Mixed Integer Programming; Forecast combinations; Forecast combination puzzle

[*]mehmet.pinar@feem.it
[†]tstengos@uoguelph.ca
[‡]eyazgan@bilgi.edu.tr

1 Introduction

Since the seminal work of Bates and Granger (1969), combining forecasts of different models, instead of relying on forecasts of individual models has come to be viewed as an effective way of improving the accuracy of predictions regarding a certain target variable. A significant number of theoretical and empirical studies, e.g. Timmerman (2006) and Stock and Watson(2004) have been able to show the superiority of combined forecasts over single-model based predictions.

The central question on which the literature focused on, naturally, is to determine the optimal weights used in the calculation of combined forecasts. In combined forecasts the weights attributed to each model depends on the model's out of sample performance. As time moves on forecast errors used for the calculation of optimal weights changes, and thus the weights themselves vary over time. However, in empirical applications, numerous papers (Clemen (1989), Stock and Watson (1999, 2001, 2004), Hendry and Clements (2004), Smith and Wallis (2009), Huang and Lee (2010), Aiolfi et al. (2010)) have found that the equal weighted forecast combination often outperforms estimated optimal forecast combinations. This finding is frequently referred as "forecast combination puzzle" by Stock and Watson (2004).[1] Overall, even though different optimal forecast combination weights are derived for static, dynamic, or time-varying situations, most empirical findings suggest that the simple average forecast combination outperforms more sophisticated weighting schemes and/or the best individual model.

In this paper, we will follow an approach for the combination of forecasts based on stochastic dominance (SD hereafter) analysis and we test whether a simple average combination of forecasts would outperform forecast combinations with more elaborate weights or not. In this context, we will examine whether an equally weighted forecast combination is optimal. Instead of assigning arbitrary equal weights to each forecast, we use stochastic dominance efficiency analysis (SDE hereafter) to propose a weighting scheme that dominates the equally weighted forecast combination or is alternatively dominated by it.[2]

[1]Smith and Wallis (2009) found that the finite sample error is the reason behind the forecast combination puzzle. On the other hand, Aiolfi et al. (2010) suggested that parameter or model instability is another reason why simple average forecast combination outperforms the best individual forecast model (see Diebold and Pauly (1987), Clements and Hendry (1998, 1999, 2006), Timmermann (2006) for further discussion of model instability and Elliot and Timmermann (2005) forecast combinations for time varying data.

[2]Mostly, stochastic dominance comparisons are made pairwise in the literature. Barrett and Donald (2003) developed pairwise stochastic dominance comparisons that relied on Kolmogorov-Smirnov type tests developed within a consistent testing environment. This offers a generalization to Anderson (1996), Beach and Davidson (1983), Davidson and Duclos (2000) who have looked at second order stochastic dominance

The main contribution of the paper is the derivation of an optimal (worst) forecast combination based on SD analysis with differential forecast weights. For the optimal (worst) forecast combination, this index will minimize (maximize) the forecasts errors by combining time-series model based forecasts for a given probability level. By weighting each forecast differently, we will find the optimal (worst) forecast combination that do not rely on arbitrary weights.[3]

In our empirical applications using two exchange rate series on weekly data for the Japanese Yen/U.S. Dollar and U.S. Dollar/Great Britain Pound for the period from 1975 to 2010 we find that the simple average forecast combination is neither the worst nor the best forecast combination something that provides partial support for the forecast combination puzzle. [4] In that context, the random walk (RW hereafter) model is the model that consistently contributes with considerably more than an equal weight to the worst forecast combination for all variables being forecasted and for all forecast horizons. For the optimal forecast combination, the best forecasting model (i.e., the model which gets relatively more weight than other forecast models) always includes a flexible Neural Network autoregressive (hereafter NNETTS) model and a self-exciting threshold

using tests that rely on pair-wise comparisons made at a fixed number of arbitrary chosen points. This is not a desirable feature since it introduces the possibility of a test inconsistency. Linton et al. (2005) propose a subsampling method which can deal with both dependent samples and dependent observations within samples. This is appropriate for conducting SD analysis for model selection among many forecasts. In this context, comparisons were available for pairs where one can compare one forecast with respect to another forecast and conclude whether one forecast dominates the other one. In other words, one can find the best individual model by comparing all forecasts. Lately, multi-variate (multidimensional) comparisons, in our case forecast combinations, have become more popular. In an application to optimal portfolio construction in finance, Scaillet and Topaloglou (2010), hereafter ST, use SD efficiency tests that can compare a given portfolio with an optimal diversified portfolio constructed from a set of assets. In a related paper, Pinar, Stengos and Topaloglou (2010) use a similar approach to construct an optimal Human Development Index. The same methodology is applied in Agliardi et al. (2011), where an optimal country risk index is constructed following SD analysis with differential component weights, yielding an optimal hybrid index for economic, political, and financial risk indices that do not rely on arbitrary weights as rating institutions do.

[3]To achieve stochastic dominance we maximize the difference between two cumulative distribution functions. This maximization results in the worst forecast combination constructed from the set of forecast models in the sense that it reaches the maximum value of absolute forecast errors for a given probability level. A minimization of the difference would result in the best case forecast scenario, where the forecast combination now achieves the minimum value of absolute forecast errors. We expand on this point in the next section.

[4]To obtain full support for the puzzle we would expect the equally weighted combination to be the best, on the other hand to refute the puzzle, in our empirical applications, we would expect the simple average combination to be the worst.

autoregressive (hereafter SETAR) model for all series and for all forecast horizons

The remainder of the paper is as follows. In section 2 we define the notion of stochastic dominance and discuss the general hypothesis for stochastic dominance at any order. Section 3 describes the data, time-series forecasting models and forecast methods used in our paper and presents the empirical analysis. We use the ST methodology to find the optimal (worst) forecast combination for macroeconomic variables for different forecast horizons. Finally, section 4 concludes.

2 Hypothesis, Test Statistics and Asymptotic Properties

Let $\hat{y}_{i,t+h}$ be the forecast of the ith forecasting model for y_t generated at time t for the period of $t+h$ $(h \geq 1)$. We have n different time-series forecasting models and we denote the absolute forecast error of the ith forecasting model as $\hat{\varepsilon}_{i,t+h} = |y_{t+h} - \hat{y}_{i,t+h}|$. The equally weighted forecast combination can be obtained as the simple average of individual forecasts derived from n different models, i.e. $\hat{y}^{ew}_{t+h} = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_{i,t+h}$. The absolute forecast error of the equally weighted forecast combination is given by $\hat{\varepsilon}^{ew}_{t+h} = \left|y_{t+h} - \hat{y}^{ew}_{t+h}\right|$. Now consider an alternative weighing scheme as $\hat{y}^{w}_{t+h} = \sum_{i=1}^{n}\lambda_i\hat{y}_{i,t+h}$, where $\sum_{i=1}^{n}\lambda_i = 1$. Therefore $\hat{\varepsilon}^{w}_{t+h} = \left|y_{t+h} - \hat{y}^{w}_{t+h}\right|$. The forecast combination literature asserts that there exists a combination of forecasts that delivers smaller $\hat{\varepsilon}^{w}_{t+h}$ than all of the individual constituent models' $\hat{\varepsilon}_{i,t+h}$'s, hence there exists a solution to the following optimization problem apart from any of λ_i-'s taking on value 1 with remaining λ_i's equal to 0's, i.e. one of the i-s is the best forecaster.

$$\underset{(\lambda_i)}{\text{Arg Min}}\ \hat{\varepsilon}^{w}_{t+h} = \left|y_{t+h} - \sum_{i=1}^{n}\lambda_i\hat{y}_{i,t+h}\right| \qquad s.t. \sum_{i=1}^{n}\lambda_i = 1 \tag{1}$$

The forecast combination puzzle refers to a situation in which $\hat{\varepsilon}^{ew}_{t+h}$ constitutes the minimum (absolute) forecast error.

Let us define Mean Absolute Forecast Errors (MAFEs) of different models as $\overline{\hat{\varepsilon}^{ew}_{t+h}} = \frac{1}{n}\sum_{i=1}^{n}|\hat{\varepsilon}_{i,t+h}|$ which is obviously identical to $\hat{\varepsilon}^{ew}_{t+h}$ defined above. Similarly defining Weighted Absolute Forecast Errors (WAFE's) as $\overline{\hat{\varepsilon}^{w}_{t+h}} = \sum_{i=1}^{n}\lambda_i\,|\hat{\varepsilon}_{i,t+h}|$ which is identical to $\hat{\varepsilon}^{w}_{t+h}$ defined above. Hence the optimization problem Equation (1) can also be written as a minimization problem of the sum of WAFE's.

$$\underset{(\lambda_i)}{\text{Arg Min}}\ \overline{\hat{\varepsilon}^{w}_{t+h}} = \sum_{i=1}^{n}\lambda_i\,|y_{t+h} - \hat{y}_{i,t+h}| \qquad s.t. \sum_{i=1}^{n}\lambda_i = 1 \tag{2}$$

The optimization problem above derives the weights that minimize the overall absolute forecast error, placing emphasis only on the first moment. However, SDE analysis allows for all moments as it examines the whole distribution function of absolute forecast errors. In this current paper, we test the efficiency of MAFE's by testing whether the cumulative distribution function of mean absolute forecast errors, $\overline{\hat{\varepsilon}_{t+h}^{ew}}$, is stochastically efficient or not.

We denote by $F(|\hat{\varepsilon}_{t+h}|)$, the continuous cdf of $\boldsymbol{\varepsilon_{t+h}} = (|\hat{\varepsilon}_{1,t+h}|, ..., |\hat{\varepsilon}_{i,t+h}|, ..., |\hat{\varepsilon}_{n,t+h}|)'$. Let us consider a forecast combination $\boldsymbol{\lambda} \in \mathbb{L}$ where $\mathbb{L} := \{\boldsymbol{\lambda} \in \mathbb{R}_+^n : \boldsymbol{e}'\boldsymbol{\lambda} = 1\}$ with $\boldsymbol{e}$ being a vector of ones. This means that all the different absolute forecast errors from different models have positive weight and that the forecast combination weights sum to one. Let us denote by $G(z, \boldsymbol{\lambda}; F)$ the cdf of the forecast combination $\boldsymbol{\lambda}'\boldsymbol{\varepsilon_{t+h}}$ at point z given by $G(z, \boldsymbol{\lambda}; F) := \int_{\mathbb{R}^n} \mathbb{I}\{\boldsymbol{\lambda}'\boldsymbol{\varepsilon_{t+h}} \leq z\} dF(\boldsymbol{\varepsilon_{t+h}})$. Let us denote the equally weighted forecast combination $\boldsymbol{\tau}$ which is a special case of $\boldsymbol{\lambda}$, being vector of $\frac{1}{n}$'s. Therefore $G(z, \boldsymbol{\tau}; F)$ is the cdf of the equally-weighted forecast combination, $\overline{\hat{\varepsilon}_{t+h}^{ew}}$, or $\boldsymbol{\tau}'\boldsymbol{\varepsilon_{t+h}}$.

For any two distributions we say that the hybrid combination $\boldsymbol{\lambda}$ dominates the distribution of some other hybrid combination $\boldsymbol{\tau}$ stochastically at first order (SD1) if, for any point z of the distribution $G(z, \boldsymbol{\tau}; F) \geq G(z, \boldsymbol{\lambda}; F)$. In general, the dominant combination refers to a "best outcome" case as there is more mass to the right of z with $G(z, \boldsymbol{\lambda}; F)$ than with $G(z, \boldsymbol{\tau}; F)$. In the context of the present analysis, since the distribution of outcomes refers to forecast errors, the "best outcome" dominant case corresponds to a forecast combination with the largest forecast errors and as such it would yield the worst possible forecast combination. More precisely, in the context of our analysis, if z denotes an absolute forecast error level, then the inequality in the definition means that the value (mass) of the cdf of forecast errors with $\boldsymbol{\lambda}$ at point z is no larger than the value (mass) of the cdf of absolute forecast errors with $\boldsymbol{\tau}$. In other words, there is at least as high a proportion of absolute error levels in $\boldsymbol{\lambda}$ as in $\boldsymbol{\tau}$. If the forecast combination $\boldsymbol{\lambda}$ dominates the equal weighted forecast combination (i.e. MAFEs) $\boldsymbol{\tau}$ at first order, then there is always less value (mass) of absolute forecast errors in the cdf of τ than that of $\boldsymbol{\lambda}$. In that case, $\boldsymbol{\lambda}$ yields the worst forecast combination. If τ were to refer to the the equally weighted forecast combination, we could establish that this could not have been the worst forecast combination as the alterative given by $\boldsymbol{\lambda}$ would be that.

More precisely, to achieve stochastic dominance we maximize the following objective function: Max $[G(z, \boldsymbol{\tau}; F) - G(z, \boldsymbol{\lambda}; F)]$. This maximization results in the worst forecast combination $\boldsymbol{\lambda}$ constructed from the set of forecast models in the sense that it reaches the maximum value of absolute forecast errors for a given probability level. Following the traditional definition of stochastic dominance we center our discussion around the worst forecast combination, that would correspond to the "best outcome" scenario, consistent with the highest mass of forecast errors. Any dominant combination in that case would imply a worse forecast combination than the one that it dominates. If there is no other alternative weighting scheme that dominates the equally weighted forecast combination, then the latter would constitute the worst forecast combination. Of course, if we were to apply the optimization criterion on the cdf's of the transformed absolute forecast errors, where the latter are obtained by multiplying the original absolute forecast errors by minus one, that would result in minimizing instead of maximizing the difference between the cdf's. In that case, the stochastic dominant forecast combinations would correspond to the optimal best-case scenario forecast combination that would be consistent with minimal forecast errors. In the empirical section we will derive worst and best case forecast combinations and we will assess the SDE of the equally-weighted combination.

The general hypotheses for testing the stochastic dominance efficiency of order j of $\boldsymbol{\tau}$, hereafter SDE_j, can be written compactly as:

$$\begin{aligned} H_0^j &: \mathcal{J}_j(z, \boldsymbol{\tau}; F) \leq \mathcal{J}_j(z, \boldsymbol{\lambda}; F) \quad \text{for all}\, z \in \mathbb{R}\, \text{and for all}\, \boldsymbol{\lambda} \in \mathbb{L}, \\ H_1^j &: \mathcal{J}_j(z, \boldsymbol{\tau}; F) > \mathcal{J}_j(z, \boldsymbol{\lambda}; F) \quad \text{for some}\, z \in \mathbb{R}\, \text{or for some}\, \boldsymbol{\lambda} \in \mathbb{L}. \end{aligned}$$

where

$$\mathcal{J}_j(z, \boldsymbol{\lambda}; F) = \int_{\mathbb{R}^n} \frac{1}{(j-1)!}(z - \boldsymbol{\lambda}'\boldsymbol{\varepsilon_{t+h}})^{j-1}\mathbb{I}\{\boldsymbol{\lambda}'\boldsymbol{\varepsilon_{t+h}} \leq z\}dF(\boldsymbol{\varepsilon_{t+h}}) \tag{3}$$

and $\mathcal{J}_1(z, \boldsymbol{\lambda}; F) := G(z, \boldsymbol{\lambda}; F)$ [5]. Under the null hypothesis H_0^j there is no forecast combination $\boldsymbol{\lambda}$ constructed from the set of absolute forecast errors from the model based forecasts that dominates the equal weighted index $\boldsymbol{\tau}$ (MAFEs) at order j. In this case, the function $\mathcal{J}_j(z, \boldsymbol{\tau}; F)$ is always lower that the function $\mathcal{J}_j(z, \boldsymbol{\lambda}; F)$ for all possible forecast combinations $\boldsymbol{\lambda}$ for any absolute forecast error level z. In this case as discussed earlier, we interpret this result to mean that the equally-weighted forecast combination yields the worst forecast scenario. Under the alternative hypothesis

[5] Defining

$$\begin{aligned} \mathcal{J}_2(z, \boldsymbol{\lambda}; F) &:= \int_{-\infty}^{z} G(u, \boldsymbol{\lambda}; F)du = \int_{-\infty}^{z} \mathcal{J}_1(u, \boldsymbol{\lambda}; F)du, \\ \mathcal{J}_3(z, \boldsymbol{\lambda}; F) &:= \int_{-\infty}^{z}\int_{-\infty}^{u} G(v, \boldsymbol{\lambda}; F)dvdu = \int_{-\infty}^{z} \mathcal{J}_2(u, \boldsymbol{\lambda}; F)du, \end{aligned}$$

and so on. Davidson and Duclos (2000) (Equation (2))states that

$$\mathcal{J}_j(z, \boldsymbol{\lambda}; F) = \int_{-\infty}^{z} \frac{1}{(j-1)!}(z-u)^{j-1}dG(u, \boldsymbol{\lambda}, F),$$

which can be rewritten as in the text.

H_1^j, we can construct a forecast combination $\boldsymbol{\lambda}$ that for some absolute forecast error level z, the function $\mathcal{J}_j(z, \boldsymbol{\tau}; F)$ is greater than the function $\mathcal{J}_j(z, \boldsymbol{\lambda}; F)$. Thus, when $j = 1$, the equally-weighted absolute forecast errors (MAFEs) $\boldsymbol{\tau}$ is stochastically inefficient (i.e., it does not yield the worst forecast scenario) at first order if and only if some other forecast combination $\boldsymbol{\lambda}$ dominates it at some absolute forecast error level z. Alternatively, the equally-weighted forecast combination $\boldsymbol{\tau}$ is stochastically efficient (i.e., the worst forecast combination) at first order if and only if there is no other forecast combination $\boldsymbol{\lambda}$ that dominates it at all absolute error levels.

We obtain SD at first and second order when $j = 1$ and $j = 2$, respectively. The hypothesis for testing SD of order j of the distribution of the equally weighted forecast combination $\boldsymbol{\tau}$ over the distribution of an alternative forecast combination $\boldsymbol{\lambda}$ takes analogous forms, but for a given $\boldsymbol{\lambda}$ instead of several of them.

The empirical counterpart of (3) is simply obtained by integrating with respect to the empirical distribution $\hat{F}$ of F, which yields:[6]

$$\mathcal{J}_j(z, \boldsymbol{\lambda}; \hat{F}) = \frac{1}{T-h} \sum_{t=1}^{T-h} \frac{1}{(j-1)!} (z - \boldsymbol{\lambda}' \boldsymbol{\varepsilon}_{t+h})^{j-1} \mathbb{I}\{\boldsymbol{\lambda}' \boldsymbol{\varepsilon}_{t+h} \leq z\} \tag{4}$$

We consider the weighted Kolmogorov-Smirnov type test statistic

$$\hat{S}_j := \sqrt{T-h} \frac{1}{T-h} \sup_{z, \boldsymbol{\lambda}} \left[\mathcal{J}_j(z, \boldsymbol{\tau}; \hat{F}) - \mathcal{J}_j(z, \boldsymbol{\lambda}; \hat{F}) \right], \tag{5}$$

and a test based on the decision rule:

$$\text{“ Reject} \quad H_0^j \quad \text{if} \quad \hat{S}_j > c_j \text{ ”},$$

where c_j is some critical value. (The derivation of the test is given by ST (2010)).

In order to make the result operational, we need to find an appropriate critical value c_j. Since the distribution of the test statistic depends on the underlying distribution, this is not an easy task, and we decide hereafter to rely on a block bootstrap method to simulate p-values (see Appendix).[7]

[6]This can be rewritten more compactly for $j \geq 2$ as:

$$\mathcal{J}_j(z, \boldsymbol{\lambda}; \hat{F}) = \frac{1}{T-h} \sum_{t=1}^{T-h} \frac{1}{(j-1)!} (z - \boldsymbol{\lambda}' \boldsymbol{\varepsilon}_{t+h})_+^{j-1}.$$

[7]The asymptotic distribution of $\hat{F}$ is given by $\sqrt{T-h}(\hat{F} - F)$ which tends weakly to a mean zero Gaussian process $B \circ F$ in the space of continuous functions on R^n (see e.g. the multivariate functional central limit theorem for stationary strongly mixing sequences stated in Rio (2000)). ST (2010) derive the limiting behavior by using the Continuous Mapping Theorem (as in Lemma 1 of Barrett and Donald (2003)), see ST (2010) Lemma 2.1.

The test statistic $\hat{S}_1$ for first order stochastic dominance efficiency is derived using mixed integer programming formulations (see Appendix) [8].

3 Empirical Analysis

3.1 Data, Forecasting Models, and Forecast Methodology

In this section we apply the SDE testing methodology to obtain optimal (worst) forecast combinations on Japanese Yen/U.S. Dollar and U.S. Dollar/Great Britain Pound exchange rate returns data. We use log first differences of exchange rate levels. The exchange rate series data are expressed at a weekly frequency for the period between 1975:1-2010:52.[9] The use of weekly data avoids the so-called weekend effect, as well as other biases associated with nontrading, bid-ask spread, asynchronous rates and so on, which are often present in higher frequency data. To initialize our parameter estimates we use weekly data between 1975:1 - 2006:52. We then generate pseudo out-of-sample forecasts of 2007:1 - 2010:52. Parameter estimates are updated recursively by expanding the estimation window by one observation forward and, thereby, reducing the pseudo out-of-sample test window by one period.

In our out-of-sample forecasting exercise we concentrate only on univariate models and consider three types of linear and four types of nonlinear univariate models. While linear models consist of the random walk (RW), autoregressive (AR) and autoregressive moving average (ARMA) models, nonlinear models comprise the following models: logistic smooth transition autoregressive (LSTAR), self-exciting threshold autoregressive (SETAR), Markov-Switching autoregressive (MS-AR) and autoregressive neural networks (NNETTS).

In the RW model $\hat{y}_{t+h}$ is equal to the the value of y_t at time t. ARMA model is

$$y_t = \alpha + \sum_{i=1}^{p} \phi_{1,i} y_{t-i} + \sum_{i=1}^{q} \phi_{2,i} \varepsilon_{t-i} + \varepsilon_t \tag{6}$$

where p and q are selected to minimize Akaike Information Criterion (AIC) and with a maximum lag of 24. After estimating the parameters of equation (6) one can easily produce h-step ($h \geq 1$) forecasts by the following recursive equation:

[8]In this paper we only test first order SD in the empirical applications below. Since there are alternative weighting schemes that dominates the given one at the first order, we do not move to the second one

[9]The daily noon buying rates in New York City certified by the Federal Reserve Bank of New York for customs and cable transfers purposes are obtained from the FREDÂ® Economic Data system of Federal Reserve Bank of St Louis (http://research.stlouisfed.org). The weekly series is generated by selecting the Wednesday series (if a Wednesday is a holiday then the following Thursday is used).

$$\hat{y}_{t+h|t} = \alpha + \sum_{i=1}^{p} \hat{\phi}_{1,i} \hat{y}_{t+h-i} + \sum_{i=1}^{q} \hat{\phi}_{2,i} \hat{\varepsilon}_{t+h-i} \,. \tag{7}$$

When $h > 1$, to obtain forecasts we iterate on a one-period forecasting model, by feeding the previous period forecasts as regressors into the model. That means when $h > p$ and $h > q$, $y_{t+h-i|t}$ is replaced by $\hat{y}_{t+h-i|t}$ and ε_{t+h-i} by $\hat{\varepsilon}_{t+h-i|t} = 0$. An obvious alternative to iterating forward on a single-period model would be to tailor the forecasting model directly to the forecast horizon, i.e., estimate the following equation by using the data up to t.

$$y_t = \alpha + \sum_{i=0}^{p} \phi_{1,i} y_{t-i-h} + \sum_{i=0}^{q} \phi_{2,i} \varepsilon_{t-i-h} + \varepsilon_t \tag{8}$$

for $h \geq 1$ and use the fitted values of this regression to produce h-step ahead forecast directly [10]

Being a special case of ARMA, estimation and forecasts of AR model can be simply obtained by setting $q = 0$ in (7) and (8).

LSTAR model is

$$y_t = \left(\alpha_1 + \sum_{i=1}^{p} \phi_{1,i} \, y_{t-i} \right) + d_t \left(\alpha_2 + \sum_{i=1}^{q} \phi_{2,i} \, y_{t-i} \right) + \varepsilon_t \tag{9}$$

with $d_t = (1 + \exp\{-\gamma(y_{t-1} - c)\})^{-1}$. ε_t are regarded as normally distributed i.i.d. variables with zero mean and α_1, α_2, $\phi_{1,i}$, $\phi_{2,i}$, γ and c are simultaneously estimated by maximum likelihood.

In LSTAR model, while direct forecast can be obtained as in ARMA case, which is also the case for all the subsequent nonlinear models [11], it is not possible to apply any iterative scheme to obtain multistep ahead forecasts as in the linear models. This impossibility follows from the general fact that conditional expectation of a nonlinear function is not necessarily equal to function of that conditional expectation, and one cannot iteratively derive the forecasts for $h > 1$ by plugging in previous forecasts (see, for example, Kock and Terasvirta 2011) [12]. We, therefore, use Monte Carlo integration scheme suggested by Lin and Granger (1994) to numerically calculate conditional expectations and, then, produce forecasts iteratively.

[10] Which approach is best -the direct or the iterated- is an empirical matter since it involves trading off estimation efficiency against robustness to model misspecification, see Elliott and Timmerman (2008). Marcellino, Stock and Watson (2006) address these points empirically using a data set of 170 US monthly macroeconomic time series. They find that the iterated approach generates the lowest MSE-values, particularly if long lags of the variables are included in the forecasting models and if the forecast horizon is long.

[11] This requires replacing y_t by y_{t+h} on the left hand side in equation (4) and running the regression using data up to time t to fitted values for corresponding forecasts

[12] Indeed, d_t is convex in y_{t-1} whenever $y_{t-1} < c$ and $-d_t$ is convex whenever $y_{t-1} > c$. Therefore, by Jensen's inequality, naive estimation under-estimates d_t if $y_{t-1} < c$ and over-estimates if $y_{t-1} > c$.

When $|\gamma| \to \infty$ LSTAR model approaches two-regime SETAR model, which is also included in our forecasting models. Alike LSTAR and most nonlinear models, in forecasting with SETAR, it is not possible to use simple iterative scheme to generate multi period forecasts. In this case, we employ a version of the Normal Forecasting Error (NFE) method suggested by Al-Qassam and Lane (1989) to generate multistep forecasts[13]. NFE is an explicit form recursive approximation to calculate higher step forecasts under normality assumption of error terms and is shown by De Gooijer and De Bruin (1998) to perform reasonably accurate compared with numerical integration and Monte Carlo method alternatives.

The two-regime MS-AR model that we consider here is

$$y_t = \alpha_s + \sum_{i=1}^{p} \phi_{s,i} y_{t-i} + \varepsilon_t \tag{10}$$

where s_t is a two-state discrete Markov chain with $S = \{1, 2\}$ and $\varepsilon_t \sim$ i.i.d. $N(0, \sigma^2)$. We estimate MS-AR by using the maximum likelihood algorithm expectation-maximization.

Although MS-AR models may encompass complex dynamics, point forecasting is less complicated in comparison to other non-linear models. The h-step forecasts from the MS-AR model is

$$\begin{aligned}\hat{y}_{t+h|t} = {} & P\left(s_{t+h} = 1 \mid y_t, ..., y_0\right) \left(\alpha_{s=1} + \sum_{i=1}^{p} \phi_{s=1,i} \hat{y}_{t+h-i}\right) \\ & + P\left(s_{t+h} = 2 \mid y_t, ..., y_0\right) \left(\alpha_{s=2} + \sum_{i=1}^{p} \phi_{s=2,i} \hat{y}_{t+h-i}\right)\end{aligned} \tag{11}$$

where $P\left(s_{t+h} = i \mid y_t, ..., y_0\right)$ is the ith element of column vector $\mathbf{P}^h \hat{\xi}_{t|t}$. $\hat{\xi}_{t|t}$ represent the filtered probabilities vector and $\mathbf{P}^h$ is the constant transition probabilities matrix (see, for example, Hamilton 1994). Hence, multistep forecasts can be obtained iteratively by plugging in $1, 2, 3, \ldots$ period forecasts similar to the iterative forecasting method of AR processes.

NNETTS the autoregressive single hidden layer feed-forward neural network model[14] suggested in Terasvirta (2006), is defined as

$$y_t = \alpha + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{h} \lambda_j d\left(\sum_{i=1}^{p} \gamma_i y_{t-i} - c\right) + \varepsilon_t \tag{12}$$

where d is the logistic function defined above such that $d = \left(1 + \exp\{-x\}\right)^{-1}$. Estimation of an ARNN model may, in general, be computationally challenging. Here we follow QuickNet method,

[13] A detailed exposition of approaches for forecasting from a SETAR model can be found in van Dijk et al. (2003)

[14] See Franses and Dijk (2000) for a review of feed-forward type neural network models

a kind of "relaxed greedy algorithm", suggested by White (2006). Forecasting procedure with NNETTS, on the other hand, is identical to that of LSTAR.

To obtain pseudo out-of-sample forecasts for a given horizon,h, the models are estimated by running regressions with data up through the date $t_0 < T$, where t_0 refers to the date where the estimation is initialized (2006:52 in both of the exchange rate return applications) and T to the final date in our data. The first h horizon forecast is obtained by using the coefficient estimates from this first regression. Next, the time subscript is advanced, and the procedure is repeated for $t_0 + 1$, $t_0 + 2$,...,$T - h$ to obtain $N_f = T - t_0 - h - 1$ distinct h-step forecasts.[15]

For each of h-step forecasts, we calculate N_f absolute forecast errors for each our models that we use in our applications.

3.2 Results for the efficiency of forecast combinations

This section presents our findings of the tests for SD1 efficiency of the equally-weighted forecast combination. We find that the equally weighted forecast combination constitutes neither the optimal nor the worst forecast combination. We obtain the best and worst forecast combinations of the model based forecasts for Japanese yen/U.S. dollar and U.S. dollar/Great Britain pound exchange rate forecasts by computing the weighting scheme on each forecast model which offers the optimal (worst) forecast combination that minimizes (maximizes) absolute forecast errors.

3.2.1 Japanese yen/U.S. dollar exchange rate application

First, we start our empirical analysis with the weekly Japanese yen/U.S. dollar exchange rate forecasts for different forecast horizons. We proceed with testing whether the equally-weighted forecast combination of the forecasting models for different horizons is the worst forecast combination or there are alternative weights on forecast models that stochastically dominate the equally-weighted forecast combination, $\boldsymbol{\tau}$, in the first order sense (e.g. for which $G(z, \boldsymbol{\tau}; F) > G(z, \boldsymbol{\lambda}; F)$), where absolute forecast errors are maximized. Three panels of Table 1 present the results for the worst case scenario (i.e., forecast combination of models that maximizes the absolute forecast error) for different forecast horizons. In panel A, B and C of Table 1, we present the results for the weekly (i.e. 1 to 24 steps ahead), monthly (i.e. 28 to 52 steps ahead) and quarterly (i.e. 65 to 104 step ahead) forecast horizons respectively. In panel A of Table 1, for one step ahead forecast horizon (i.e. h=1), each time-series model produces 180 one step ahead forecasts with their associated absolute forecast errors. The average of these absolute forecast errors are given by the correspond-

[15] N_f differs between 208 and 105 for different h values between 1 and 104.

ing MAFE's. There are 204 alternative forecast combinations that dominate the equally-weighted forecast combination (i.e. MAFE's) in the first order sense and the first row in panel A of Table 1 summarizes the results. This table presents the average weights of the 204 alternative forecast combinations that dominate the equally-weighted forecast combination. We find that the RW model has the highest contribution with 76.6% weight. On the other hand, SETAR, NNETTS, ARMA, MS, and AR contribute with weights of 10.9%, 9.2%, 2.4%, 0.5% and 0.4% respectively. We replicate the same exercise for weekly (step 1 to 24), monthly (step 28 to 52) and quarterly (step 65 to 104) forecast horizons in panel A, B and C respectively. The models that receive more than the arbitrarily assigned equal weight are highlighted in all the Tables.[16] Overall, we find that the RW model consistently contributes to the worst forecast combination with weight well above 0.143 and also receives the highest weight relative to other models for all of the forecast horizons. Moreover, the NNETTS and SETAR models contribute to the worst forecast combination with weights that are above 0.143 for some forecast horizons and below 0.143 for some other forecast horizons. Finally, the AR, ARMA, LSTAR and MS models always contribute with weights that are less than 0.143 for all forecast horizons.

Overall, for all forecast horizons, we found that the equally-weighted forecast combination does not constitute the worst forecast combination. There are many other forecast combinations that dominate the equally-weighted forecast combination for all different forecast horizons that maximize absolute forecast errors. Now using the same weekly Japanese yen/U.S. dollar exchange rate forecasts, we proceed with testing whether the equally-weighted forecast combination of the forecasting models for different horizons is the optimal forecast combination or there exist alternative weights on forecasting models that stochastically dominate the equally-weighted forecast combination, τ, in the first order sense (e.g. for which $G(z, \boldsymbol{\tau}; F) > G(z, \boldsymbol{\lambda}; F)$), where now absolute forecast errors are minimized.[17] Three panels of Table 2 present the results for the best case scenario (i.e. forecast combination of models that minimizes absolute forecast errors) for different forecast horizons. In panel A, B and C of Table 2, we present the results for the weekly (i.e. 1 to 24 steps ahead), monthly (i.e. 28 to 52 steps ahead) and quarterly (i.e. 65 to 104 step ahead) forecast horizons respectively. For example, for the one step ahead forecast, there are 208 given MAFE's and there exist alternative forecast combinations for all case that dominate the equally-weighted forecast combination in the first order sense. Table 2 summarizes the results. This Table presents

[16]In our application, there are 7 time-series forecasting models and in order to obtain the mean absolute forecast errors, each model gets equal weights, a weight of approximately 0.143 (i.e. 1/7).

[17]As proposed earlier in the previous section, we multiply each absolute forecast error with minus one to use SDE analysis to test whether the equally-weighted forecast combination is optimal or not.

the average weights of the alternative forecast combinations that dominate the equally-weighted forecast combination. For the one step ahead forecast, NNETTS is the model with the highest contribution with a weight of 59%. On the other hand, SETAR, MS, ARMA, AR, RW, and LSTAR models take weights of 22.8%, 10.5%, 5.6%, 1%, 0.9% and 0.2% respectively. Overall, NNETTS and SETAR contribute with more than equal weights (i.e. a weight of 0.143) for the construction of the optimal forecast combination at all forecast horizons. Moreover, NNETTS model always contributes the most for all forecast horizons for the optimal forecast combination, whereas ARMA model contributes more than equal weights for some horizons and less than equal weights for some other. However, the AR, LSTAR (except one forecast horizon), MS and RW models make part of the optimal forecast combination with weights that are less than the equal weight of 0.143 for all forecast horizons.

For the application to weekly Japanese yen/U.S. dollar exchange rate forecasts, we find that the equally-weighted forecast combination is neither the optimal nor the worst forecast combination. We find that different models contribute with differential weights for the optimal forecast combination at different forecast horizons. Some of the models have robust contributions to the optimal (worst) forecast combination. The RW model always contributes with more (less) than equal weights to the worst (optimal) forecast combination. Moreover, the NNETTS and SETAR models contribute with more than equal weights to the optimal forecast combination for all forecast horizons. On the other hand, the AR, LSTAR and MS models always contribute less than equal weights both for the optimal and worst forecast combination for all forecast horizons. The upshot of our analysis is that the simple average forecast combination is neither the worst nor the best forecast combination something that provides partial support for the forecast combination puzzle in that there are many other forecast combinations that are less efficient than the equally weighted one.

3.2.2 U.S. dollar/Great Britain pound exchange rate application

In this subsection, we have a similar application as previous subsection, but in this case we obtain the optimal (worst) forecast combination for the foreign exchange rate of U.S. dollar/Great Britain pound forecasts for different time horizons. The three panels of Table 3 and 4 present the results for the worst and optimal forecast combinations for different forecast horizons respectively. In panel A, B and C of Table 3 and 4, we present the results for the weekly (i.e. 1 to 24 steps ahead), monthly (i.e. 28 to 52 steps ahead) and quarterly (i.e. 65 to 104 step ahead) forecast horizons respectively. Tables 3 and 4 present the average weights of the alternative forecast combinations that dominate the equally-weighted forecast combination for the worst case forecast combination (i.e. forecast combination that maximizes absolute forecast errors) and for the best-case combination (i.e. forecast combination that minimizes absolute forecast errors) respectively. For the worst case, for all forecast horizons, the RW model contributes the most with at least a weight of 68.7%. The NNETTS and SETAR models contribute to the worst forecast combination with more than equal weights for some forecast horizons (for only six and two forecast horizons respectively), while the AR, ARMA, LSTAR and MS models always contribute with a weight that is less than 6%. For the optimal forecast combination in Table 4, the NNETTS and SETAR (except one forecast horizon) models always contribute with more than equal weights (i.e. a weight of 0.143) for all horizons. Moreover, ARMA contributes with more than an equal weight to the optimal forecast combination for some forecast horizons. Finally, the AR, LSTAR, MS and RW models contribute less than equal weights to the optimal forecast combination for all forecast horizons.

Overall, for the application to weekly U.S. dollar/Great Britain pound exchange rate forecasts, the findings are very similar to those of Japanese yen/U.S. dollar exchange rate application. We find that the equally-weighted forecast combination is neither optimal nor the worst forecast combination. We find that the NNETTS and SETAR models contribute with more than equal weights to the optimal forecast combination for all forecast horizons and ARMA contributes with more than equal weight to the optimal forecast combination for some forecast horizons. On the other hand, the RW model always contributes with the highest weight to the worst forecast combination. As before, we find that the simple average forecast combination is neither the worst nor the best forecast combination something that provides partial support for the forecast combination puzzle.

4 Conclusion

This paper presents stochastic dominance efficiency tests at any order for time dependent data. We study tests for stochastic dominance efficiency of an equal weighted forecast combinations with respect to all possible forecast combinations constructed from a set of time-series model forecasts. We proceed to test whether stochastic dominance efficiency confirms or contradicts the use of the equally weighted forecast combination. The results from the empirical analysis indicate that the equally weighted forecast combination is neither optimal nor the worst forecast combination for different forecast horizons. We can construct many alternative forecast combinations that dominate the equal weighted forecast combinations which are consistent with the least (most) absolute forecast errors. We construct the optimal (worst) forecast combination for different forecast horizons for weekly Japanese yen/U.S. dollar and U.S. dollar/Great Britain pound foreign exchange rate forecasts. We found that the RW model is the main contributor to the worst forecast combinations for both applications. On the other hand, the NNETTS and SETAR models contribute more than equal weights to the optimal forecast combination for both weekly Japanese yen/U.S. dollar and U.S. dollar/Great Britain pound exchange rate forecasts. Overall, there is also an agreement for both applications that the ARMA model contributes more than equal weight for the optimal forecast combinations for some forecast horizons, whereas the AR, LSTAR and MS models always contribute with a weight that is less than equal weights for both the worst and optimal forecast combinations for all forecast horizons.

In summary, we find that the equally-weighted forecast combination is neither optimal nor the worst forecast combination, something that provides partial support for the forecast combination puzzle in that there are many other forecast combinations that are less efficient than the equally weighted one. However, at the same time some other time-series models receiving more weights can lead to an improvement on forecasting when compared with the equally weighted combination. We should mention that we only applied stochastic dominance efficiency analysis with seven time-series models to weekly Japanese yen/U.S. dollar and U.S. dollar/Great Britain pound exchange rate data. Therefore, one should further test whether the equally-weighted forecast combination is optimal or not by using other variables of interest and also by using different time-series model forecasts. Another possible future application is to test whether the average of survey based forecast combinations are optimal. One may expect that there may be some cases where the equally-weighted forecast combination outperforms all other possible forecast combinations. Therefore, SDE of the optimal forecast combination should be tested periodically.

References

[1] Al-Qassam, M. S. and J. A. Lane (1989) Forecasting exponential autoregressive models of order 1. Journal of Time Series Analysis 91 (2), 95-113.

[2] Abadie, A. (2002) Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models, Journal of the American Statistical Association, 97, 284-292.

[3] Agliardi, E., Agliardi, R., Pinar, M. Stengos, T. and N. Topalaglou (2010) Economic, political and financial risk in emerging markets: a stochastic dominance approach for a country risk index, mimeo

[4] Aiolfi, M., Capistran C., and A. Timmermann (2010) Forecast Combinations, In Forecast Handbook, edited by M. Clements and D. Hendry. Oxford, Oxford University Press.

[5] Anderson, G. (1996) Nonparametric tests of stochastic dominance in income distributions, Econometrica, 64, 1183-1193.

[6] Barrett, G., and S. Donald (2003) Consistent tests for stochastic dominance, Econometrica, 71, 71-104.

[7] Bates, J.M. and C.W.J. Granger (1969) The Combination of Forecasts, Operations Research Quarterly, 20, 451-468.

[8] Beach, C., and R. Davidson (1983) Distribution-free statistical inference with Lorenz curves and income shares, Review of Economic Studies, 50, 723-735.

[9] Clemen, R.T. (1989) Combining Forecasts: A Review and Annotated Bibliography. International Journal of Forecasting, 5, 559-581.

[10] Clements, M.P. and D.F. Hendry (1998) Forecasting Economic Time Series, Cambridge University Press.

[11] Clements, M.P. and D.F. Hendry, 1999, Forecasting Non-stationary Economic Time Series, The MIT Press.

[12] Clements, M.P. and D.F. Hendry (2006) Forecasting with Breaks, In Elliott, G., C.W.J. Granger and A. Timmermann (eds.), Handbook of Economic Forecasting, 1, 605-658. North-Holland.

[13] Davidson, R., and J.-Y. Duclos (2000) Statistical inference for stochastic dominance and for the measurement of poverty and inequality, Econometrica, 68, 1435-1464.

[14] Diebold, F.X., and P. Pauly (1987) Structural Change and the Combination of Forecasts, Journal of Forecasting ,6, 21-40.

[15] De Gooijer, J. G. and P. T. De Bruin (1998) On forecasting SETAR processes, Statistics and Probability Letters 37 (1), 7-14.

[16] Elliott, G., and A. Timmermann (2005) Optimal Forecast Combination under Regime Switching, International Economic Review, 46(4), 1081-1102.

[17] Elliott, G., and A. Timmermann (2008) Economic forecasting. Journal of Economic Literature 46 (1), 3-56.

[18] Franses, P. and D. Dijk (2000) Nonlinear time series models in empirical finance. Cambridge University Press.

[19] Hamilton, J. (1994) Time series analysis. Princeton University Press.

[20] Hendry, D.F. and M.P. Clements (2004) Pooling of Forecasts, Econometrics Journal, 7, 1-31.

[21] Huang, H. and T.-H. Lee (2010) To combine forecasts or to combine information?, Econometric Reviews, 29(5), 534-570.

[22] Kock, A. B. and T. Terasvirta (2011) Forecasting with nonlinear time series models, pp. 61-87. Oxford University Press.

[23] Lin, J. L. and C. W. J. Granger (1994) Forecasting from non-linear models in practice. Journal of Forecasting 13, 1-9.

[24] Linton, O., Maasoumi, E. and Whang, Y.-J. (2005) Consistent testing for stochastic dominance under general sampling schemes, Review of Economic Studies 72, 735-765

[25] Marcellino, M., J. H. Stock, and M. W. Watson (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. Journal of Econometrics 135 (1-2), 499-526.

[26] Pinar, M. , Stengos, T. and N. Topaloglou (2010) On the Construction of the Human Development Index: A Stochastic Efficiency Approach, University of Guelph, manuscript.

[27] Rio, E. (2000) Theorie Asymptotique des Processus Aleatoires Faiblement Dependants, Mathematiques et Applications, 31, Springer-Verlag, Berlin

[28] Scaillet, O., and N. Topaloglou (2010) Testing for stochastic dominance efficiency, Journal of Business and Economic Statistics, 28 (1), 169-180

[29] Smith, J., and K.F. Wallis (2009) A simple explanation of the forecast combination puzzle, Oxford Bulletin of Economics and Statistics, 71, 331-355.

[30] Stock, J. H., M.W. Watson (1999) Forecasting inflation, Journal of Monetary Economics, 44, 293-335.

[31] Stock, J.H. and M.W. Watson (2001) A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series, in R.F. Engle and H. White, eds., Festschrift in Honour of Clive Granger. Cambridge University Press, 1-44.

[32] Stock, J.H. and M.W. Watson (2004), Combination Forecasts of Output Growth in a Seven Country Data Set, Journal of Forecasting, 23, 405-430.

[33] Terasvirta, T. (2006), Forecasting economic variables with nonlinear models, Volume 1 of Handbook of Economic Forecasting, Chapter 8, pp. 413-457. Elsevier.

[34] Timmermann A. (2006), Forecast combinations. In Handbook of Economic Forecasting, Elliott G, Granger CWJ, Timmermann A (eds). North-Holland: Amsterdam; 135-196.

[35] van Dijk, D., P. H. Franses, M. P. Clements, and J. Smith (2003). On SETAR non-linearity and forecasting. Journal of Forecasting 22 (5), 359-375.

[36] White, H. (2006). Approximate nonlinear forecasting methods. In G. Elliott, C. Granger, and A. Timmermann (Eds.), Handbook of Economic Forecasting, Volume 1, Chapter 9, pp. 459-512. Elsevier.

Tables

Table 1: The optimal forecast combination for Japanese Yen/U.S. Dollar exchange rate forecasts

(a) Worst-case: Weekly

Forecast steps ahead	Number of observations	Number of dominating weighting schemes	Average of dominating weighting schemes						
			AR	ARMA	LSTAR	MS	NNETTS	RW	SETAR
1	208	204	0.004	0.024	0.000	0.005	0.092	**0.766**	0.109
2	207	204	0.000	0.069	0.049	0.000	0.113	**0.748**	0.021
3	206	201	0.003	0.025	0.000	0.003	0.076	**0.756**	0.137
4	205	202	0.004	0.017	0.000	0.005	0.065	**0.780**	0.129
5	204	198	0.010	0.016	0.000	0.004	0.077	**0.787**	0.106
6	203	203	0.002	0.097	0.034	0.000	0.115	**0.729**	0.023
7	202	197	0.004	0.012	0.000	0.013	0.074	**0.757**	0.140
8	201	198	0.005	0.023	0.000	0.010	0.105	**0.724**	0.133
9	200	199	0.008	0.029	0.047	0.000	0.114	**0.786**	0.016
10	199	199	0.003	0.032	0.076	0.001	0.093	**0.751**	0.044
11	198	191	0.001	0.029	0.046	0.000	0.097	**0.763**	0.064
12	197	192	0.007	0.026	0.000	0.008	0.071	**0.748**	0.140
13	196	191	0.006	0.035	0.029	0.000	0.119	**0.759**	0.052
14	195	191	0.001	0.018	0.001	0.000	0.091	**0.744**	**0.145**
15	194	194	0.002	0.029	0.000	0.004	0.081	**0.745**	0.139
16	193	192	0.000	0.077	0.026	0.000	0.093	**0.787**	0.017
17	192	187	0.006	0.029	0.000	0.014	0.133	**0.672**	**0.146**
18	191	187	0.004	0.024	0.000	0.008	0.089	**0.794**	0.081
19	190	185	0.003	0.016	0.000	0.008	0.090	**0.723**	**0.160**
20	189	186	0.000	0.032	0.000	0.006	0.113	**0.702**	**0.147**
21	188	182	0.003	0.015	0.003	0.000	0.077	**0.745**	**0.157**
22	187	187	0.007	0.032	0.037	0.002	0.133	**0.778**	0.011
23	186	182	0.009	0.044	0.068	0.000	0.120	**0.718**	0.041
24	185	180	0.006	0.019	0.000	0.003	0.071	**0.762**	0.139

(b) Worst-case: Monthly (4 weeks)

Forecast steps ahead	Number of observations	Number of dominating weighting schemes	Average of dominating weighting schemes						
			AR	ARMA	LSTAR	MS	NNETTS	RW	SETAR
28	181	180	0.004	0.031	0.046	0.000	0.128	**0.774**	0.017
32	177	172	0.014	0.026	0.000	0.005	0.133	**0.675**	**0.147**
36	173	171	0.003	0.046	0.031	0.000	**0.144**	**0.727**	0.049
40	169	168	0.004	0.004	0.000	0.010	0.102	**0.807**	0.073
44	165	161	0.003	0.016	0.000	0.002	0.115	**0.766**	0.098
48	161	160	0.014	0.007	0.001	0.000	0.130	**0.635**	**0.213**
52	157	153	0.000	0.003	0.000	0.011	0.102	**0.780**	0.104

(c) Worst-case: Quarterly (13 weeks)

Forecast steps ahead	Number of observations	Number of dominating weighting schemes	Average of dominating weighting schemes						
			AR	ARMA	LSTAR	MS	NNETTS	RW	SETAR
65	144	144	0.004	0.025	0.039	0.000	**0.180**	**0.694**	0.058
78	131	128	0.006	0.000	0.000	0.001	0.118	**0.724**	**0.151**
91	118	118	0.001	0.002	0.000	0.000	0.104	**0.740**	**0.153**
104	105	104	0.005	0.003	0.000	0.000	**0.166**	**0.762**	0.064

Table 2: The optimal forecast combination for Japanese Yen/U.S. Dollar exchange rate forecasts

(a) Best-case: Weekly

Forecast steps ahead	Number of observations	Number of dominating weighting schemes	Average of dominating weighting schemes						
			AR	ARMA	LSTAR	MS	NNETTS	RW	SETAR
1	208	208	0.010	0.056	0.002	0.105	**0.590**	0.009	**0.228**
2	207	164	0.019	0.109	0.070	0.014	**0.460**	0.022	**0.306**
3	206	195	0.017	0.094	0.002	0.093	**0.551**	0.008	**0.235**
4	205	205	0.014	0.079	0.000	0.107	**0.574**	0.011	**0.215**
5	204	200	0.008	0.100	0.007	0.096	**0.582**	0.011	**0.196**
6	203	186	0.013	0.094	0.119	0.007	**0.415**	0.007	**0.335**
7	202	195	0.014	0.120	0.000	0.115	**0.532**	0.010	**0.209**
8	201	191	0.006	0.119	0.001	0.086	**0.547**	0.016	**0.225**
9	200	181	0.005	**0.166**	0.134	0.001	**0.399**	0.019	**0.276**
10	199	187	0.030	**0.180**	0.079	0.005	**0.435**	0.027	**0.244**
11	198	193	0.018	**0.159**	0.098	0.002	**0.399**	0.011	**0.313**
12	197	186	0.012	**0.156**	0.008	0.067	**0.503**	0.023	**0.232**
13	196	195	0.016	**0.174**	0.113	0.009	**0.430**	0.018	**0.240**
14	195	190	0.007	**0.151**	0.000	0.073	**0.527**	0.019	**0.223**
15	194	190	0.013	**0.143**	0.000	0.061	**0.550**	0.020	**0.213**
16	193	189	0.043	**0.177**	0.112	0.002	**0.368**	0.019	**0.279**
17	192	166	0.015	0.137	0.000	0.096	**0.491**	0.022	**0.239**
18	191	187	0.012	**0.147**	0.000	0.072	**0.581**	0.011	**0.177**
19	190	172	0.022	**0.170**	0.000	0.046	**0.532**	0.020	**0.210**
20	189	179	0.011	0.130	0.000	0.103	**0.506**	0.019	**0.231**
21	188	183	0.017	**0.143**	0.000	0.093	**0.516**	0.018	**0.213**
22	187	179	0.034	**0.168**	0.090	0.002	**0.366**	0.012	**0.328**
23	186	157	0.005	**0.231**	0.111	0.001	**0.406**	0.020	**0.226**
24	185	177	0.039	0.131	0.000	0.114	**0.505**	0.020	**0.191**

(b) Best-case: Monthly (4 weeks)

Forecast steps ahead	Number of observations	Number of dominating weighting schemes	Average of dominating weighting schemes						
			AR	ARMA	LSTAR	MS	NNETTS	RW	SETAR
28	181	172	0.024	0.128	0.071	0.005	**0.427**	0.015	**0.330**
32	177	175	0.013	0.090	0.000	0.069	**0.562**	0.030	**0.236**
36	173	162	0.028	0.114	**0.175**	0.002	**0.405**	0.031	**0.245**
40	169	160	0.060	0.064	0.002	0.099	**0.488**	0.010	**0.277**
44	165	154	0.027	0.101	0.002	0.061	**0.547**	0.017	**0.245**
48	161	152	0.019	0.111	0.006	0.069	**0.553**	0.017	**0.225**
52	157	149	0.009	0.072	0.000	0.082	**0.643**	0.021	**0.173**

(c) Best-case: Quarterly (13 weeks)

Forecast steps ahead	Number of observations	Number of dominating weighting schemes	Average of dominating weighting schemes						
			AR	ARMA	LSTAR	MS	NNETTS	RW	SETAR
65	144	129	0.052	**0.168**	0.074	0.001	**0.366**	0.025	**0.314**
78	131	106	0.044	0.064	0.000	0.067	**0.574**	0.020	**0.231**
91	118	106	0.032	0.048	0.000	0.042	**0.633**	0.014	**0.231**
104	105	101	0.053	0.040	0.000	0.067	**0.597**	0.012	**0.231**

Table 3: The worst forecast combination for U.S. Dollar/Great Britain Pound exchange rate forecasts

(a) Worst-case: Weekly

Forecast steps ahead	Number of observations	Number of dominating weighting schemes	Average of dominating weighting schemes						
			AR	ARMA	LSTAR	MS	NNETTS	RW	SETAR
1	208	208	0.004	0.024	0.000	0.000	0.102	**0.805**	0.065
2	207	207	0.001	0.045	0.000	0.001	0.116	**0.720**	0.117
3	206	203	0.018	0.039	0.004	0.000	0.113	**0.714**	0.112
4	205	203	0.004	0.015	0.000	0.001	0.104	**0.721**	0.155
5	204	204	0.003	0.049	0.000	0.001	0.082	**0.797**	0.068
6	203	195	0.006	0.047	0.000	0.000	0.111	**0.733**	0.103
7	202	199	0.006	0.050	0.000	0.000	0.094	**0.724**	0.126
8	201	201	0.013	0.008	0.000	0.000	0.107	**0.764**	0.108
9	200	200	0.005	0.007	0.000	0.000	0.098	**0.769**	0.121
10	199	199	0.012	0.003	0.000	0.001	0.094	**0.796**	0.094
11	198	198	0.002	0.031	0.000	0.001	0.125	**0.722**	0.119
12	197	197	0.007	0.030	0.000	0.002	0.084	**0.769**	0.108
13	196	194	0.019	0.051	0.000	0.001	0.095	**0.727**	0.107
14	195	195	0.012	0.042	0.000	0.001	0.106	**0.722**	0.117
15	194	194	0.007	0.059	0.000	0.003	0.110	**0.715**	0.106
16	193	192	0.001	0.000	0.009	0.004	0.097	**0.786**	0.103
17	192	190	0.003	0.007	0.000	0.000	0.138	**0.688**	**0.164**
18	191	191	0.002	0.035	0.003	0.000	**0.144**	**0.740**	0.076
19	190	188	0.004	0.025	0.000	0.000	0.102	**0.788**	0.081
20	189	187	0.007	0.013	0.000	0.003	0.087	**0.800**	0.090
21	188	188	0.003	0.048	0.000	0.000	0.114	**0.714**	0.121
22	187	186	0.018	0.038	0.000	0.003	0.100	**0.727**	0.114
23	186	182	0.002	0.033	0.000	0.000	0.095	**0.767**	0.103
24	185	185	0.004	0.027	0.000	0.000	0.119	**0.792**	0.058

(b) Worst-case: Monthly(4 weeks)

Forecast steps ahead	Number of observations	Number of dominating weighting schemes	Average of dominating weighting schemes						
			AR	ARMA	LSTAR	MS	NNETTS	RW	SETAR
28	181	181	0.001	0.011	0.000	0.002	0.102	**0.782**	0.102
32	177	172	0.001	0.054	0.000	0.003	0.092	**0.786**	0.064
36	173	168	0.002	0.032	0.000	0.001	0.093	**0.756**	0.116
40	169	169	0.006	0.006	0.000	0.001	0.118	**0.786**	0.083
44	165	163	0.005	0.042	0.000	0.000	0.112	**0.734**	0.107
48	161	161	0.011	0.037	0.000	0.000	**0.152**	**0.752**	0.048
52	157	157	0.004	0.011	0.000	0.003	**0.158**	**0.728**	0.096

(c) Worst-case: Monthly(4 weeks)

Forecast steps ahead	Number of observations	Number of dominating weighting schemes	Average of dominating weighting schemes						
			AR	ARMA	LSTAR	MS	NNETTS	RW	SETAR
65	144	141	0.005	0.005	0.031	0.005	**0.190**	**0.730**	0.034
78	131	129	0.013	0.018	0.004	0.000	**0.156**	**0.773**	0.036
91	118	117	0.038	0.050	0.012	0.000	**0.192**	**0.687**	0.021
104	105	103	0.040	0.020	0.004	0.003	0.136	**0.747**	0.050

Table 4: The worst forecast combination for U.S. Dollar/Great Britain Pound exchange rate forecasts

(a) Best-case: Weekly

Forecast steps ahead	Number of observations	Number of dominating weighting schemes	Average of dominating weighting schemes						
			AR	ARMA	LSTAR	MS	NNETTS	RW	SETAR
1	208	198	0.016	0.135	0.000	0.053	**0.470**	0.013	**0.313**
2	207	192	0.013	**0.148**	0.000	0.024	**0.517**	0.014	**0.284**
3	206	176	0.001	**0.174**	0.000	0.108	**0.428**	0.018	**0.271**
4	205	184	0.016	**0.183**	0.000	0.014	**0.501**	0.018	**0.268**
5	204	198	0.003	**0.168**	0.000	0.032	**0.516**	0.012	**0.269**
6	203	191	0.011	**0.206**	0.000	0.044	**0.528**	0.017	**0.194**
7	202	184	0.029	**0.264**	0.000	0.056	**0.389**	0.015	**0.247**
8	201	183	0.050	**0.158**	0.000	0.027	**0.466**	0.014	**0.285**
9	200	183	0.005	**0.232**	0.000	0.049	**0.469**	0.021	**0.224**
10	199	185	0.007	**0.144**	0.000	0.039	**0.537**	0.019	**0.254**
11	198	179	0.009	**0.220**	0.000	0.075	**0.415**	0.015	**0.266**
12	197	192	0.009	**0.171**	0.000	0.053	**0.481**	0.009	**0.277**
13	196	154	0.008	**0.216**	0.000	0.049	**0.480**	0.019	**0.228**
14	195	186	0.000	**0.159**	0.000	0.055	**0.494**	0.014	**0.278**
15	194	186	0.007	**0.159**	0.000	0.044	**0.516**	0.007	**0.267**
16	193	178	0.011	0.142	0.000	0.048	**0.478**	0.014	**0.307**
17	192	178	0.023	0.074	0.004	0.079	**0.495**	0.023	**0.302**
18	191	178	0.005	0.136	0.000	0.048	**0.502**	0.017	**0.292**
19	190	180	0.012	0.083	0.002	0.067	**0.481**	0.006	**0.349**
20	189	177	0.020	**0.148**	0.002	0.033	**0.478**	0.016	**0.303**
21	188	181	0.024	0.108	0.000	0.066	**0.446**	0.014	**0.342**
22	187	172	0.016	**0.188**	0.001	0.029	**0.513**	0.014	**0.239**
23	186	171	0.016	0.119	0.003	0.054	**0.490**	0.010	**0.308**
24	185	177	0.015	0.065	0.000	0.053	**0.556**	0.017	**0.294**

(b) Best-case: Monthly(4 weeks)

Forecast steps ahead	Number of observations	Number of dominating weighting schemes	Average of dominating weighting schemes						
			AR	ARMA	LSTAR	MS	NNETTS	RW	SETAR
28	181	172	0.012	0.072	0.000	0.049	**0.550**	0.015	**0.302**
32	177	175	0.018	0.121	0.000	0.044	**0.495**	0.013	**0.309**
36	173	162	0.011	**0.164**	0.000	0.025	**0.486**	0.017	**0.297**
40	169	160	0.016	0.134	0.000	0.006	**0.465**	0.012	**0.367**
44	165	154	0.028	**0.174**	0.000	0.021	**0.460**	0.019	**0.298**
48	161	152	0.059	0.094	0.000	0.027	**0.446**	0.014	**0.360**
52	157	149	0.046	0.119	0.000	0.034	**0.376**	0.014	**0.411**

(c) Best-case: Quarterly(13 weeks)

Forecast steps ahead	Number of observations	Number of dominating weighting schemes	Average of dominating weighting schemes						
			AR	ARMA	LSTAR	MS	NNETTS	RW	SETAR
65	144	129	0.019	0.057	0.084	0.041	**0.519**	0.029	**0.251**
78	131	106	0.002	**0.204**	0.022	0.001	**0.526**	0.022	**0.223**
91	118	106	0.000	**0.369**	0.073	0.019	**0.451**	0.028	0.060
104	105	101	0.024	0.020	0.002	0.025	**0.693**	0.016	**0.220**

Appendix : Simulating p-values

Block Bootstrap Methods

In this appendix we describe practical ways to compute p-values for testing stochastic dominance efficiency at any order by looking at block bootstrap methods and discuss the theoretical justification for these methods. Block bootstrap methods extend the nonparametric i.i.d. bootstrap to a time series context (see Barrett and Donald (2003) and Abadie (2002) for use of the nonparametric i.i.d. bootstrap in stochastic dominance tests). They are based on "blocking" arguments, in which data are divided into blocks and those, rather than individual data, are resampled in order to mimic the time dependent structure of the original data. We focus on block bootstrap since we face moderate sample sizes in the empirical applications, and wish to exploit the full sample information.

Let b, l denote integers such that $T - h = bl$. The non-overlapping rule (Carlstein (1986)) just asks the data to be divided into b disjoint blocks, the k-th being $\boldsymbol{B}_k = (\boldsymbol{\varepsilon}'_{(k-1)l+1}, ..., \varepsilon'_{kl})'$ with $k \in \{1, ..., b\}$. The block bootstrap method requires that we choose blocks $\boldsymbol{B}_1^*, ..., \boldsymbol{B}_b^*$ by resampling randomly, with replacement, from the set of non-overlapping blocks. If $\boldsymbol{B}_i^* = (\boldsymbol{Y}_{i1}^{*\prime}, ..., \boldsymbol{Y}_{il}^{*\prime})'$, a block bootstrap sample $\{\varepsilon_{t+h}^*;\ t = 1, ..., T\}$ is made of $\{\varepsilon_{11}^*, ..., \varepsilon_{1l}^*, \varepsilon_{21}^*, ..., \varepsilon_{2l}^*, ..., \varepsilon_{b1}^*, ..., \varepsilon_{bl}^*\}$ and we let $\hat{F}^*$ denote its empirical distribution.

Let us define $p_j^* := P[S_j^* > \hat{S}_j]$, where S_j^* is the test statistic corresponding to each bootstrap sample. Then the block bootstrap method is justified by the next statement (the proof is given by ST (2010)).

Proposition 1 *Assuming that $\alpha < 1/2$, a test for SDE_j based on the rule:*

$$\text{“}\mathit{reject}\, H_0^j \quad \mathit{if} \quad p_j^* < \alpha\text{”},$$

satisfies the following

$$\begin{aligned} \lim P[\mathit{reject}\, H_0^j] &\leq \alpha \qquad \mathit{if} H_0^j \text{ is true}, \\ \lim P[\mathit{reject}\, H_0^j] &= 1 \qquad \mathit{if} H_0^j \text{ is false}. \end{aligned}$$

In practice we need to use Monte Carlo methods to approximate the probability. The p-value is simply approximated by $\tilde{p}_j = \frac{1}{R}$, where the averaging is made on R replications. The replication number can be chosen to make the approximations as accurate as we desire given time and computer constraints.

Mathematical formulation of the test statistics

The test statistic $\hat{S}_1$ for first order stochastic dominance efficiency is derived using mixed integer programming formulations. The following is the full formulation of the model:

$$\max_{\mathbf{z},\boldsymbol{\lambda}} \hat{S}_1 = \sqrt{T-h}\frac{1}{T-h}\sum_{t=1}^{T-h}(L_{t+h} - W_{t+h}) \tag{13}$$

$$\text{s.t.} M(L_{t+h} - 1) \leq z - \boldsymbol{\tau}'\boldsymbol{\varepsilon}_{t+h} \leq ML_{t+h}, \qquad \forall t \tag{14}$$

$$M(W_{t+h} - 1) \leq z - \boldsymbol{\lambda}'\boldsymbol{\varepsilon}_{t+h} \leq MW_{t+h}, \qquad \forall t \tag{15}$$

$$\boldsymbol{e}'\boldsymbol{\lambda} = 1, \tag{16}$$

$$\boldsymbol{\lambda} \geq 0, \tag{17}$$

$$W_{t+h} \in \{0,1\},\ L_{t+h} \in \{0,1\}, \qquad \forall t \tag{18}$$

with M being a large constant.

The model is a mixed integer program maximizing the distance between the sum over all forecast combination scenarios of two binary variables, $\frac{1}{T-h}\sum_{t=1}^{T-h} L_{t+h}$ and $\frac{1}{T-h}\sum_{t=1}^{T-h} W_{t+h}$ which represent $G(z,\boldsymbol{\tau};\hat{F})$ and $G(z,\boldsymbol{\lambda};\hat{F})$, respectively (the empirical cdf of $\boldsymbol{\tau}$ and $\boldsymbol{\lambda}$ at absolute forecast error level z). According to inequalities (3b), L_{t+h} equals 1 for each scenario $t+h \in T$ for which $z \geq \boldsymbol{\tau}'\boldsymbol{\varepsilon}_{t+h}$, and 0 otherwise. Analogously, inequalities (3c) ensure that W_{t+h} equals 1 for each scenario for which $z \geq \boldsymbol{\lambda}'\boldsymbol{\varepsilon}_{t+h}$. Equation (3d) defines the sum of all forecast combination weights to be unity, while inequality (3e) disallows for negative weights.

This formulation allows us to test the dominance of the equal weighted forecast combination ($\boldsymbol{\tau}$) over any potential linear forecast combination $\boldsymbol{\lambda}$ of the forecasts based on time series models.

When some of the variables are binary, corresponding to mixed integer programming, the problem becomes NP-complete (non-polynomial, i.e., formally intractable). The problem can be reformulated in order to reduce the solving time and to obtain a tractable formulation (see above and ST (2010), section 4.1 for the derivation of this formulation and details on practical implementation).