Robust Optimization of Forecast Combinations

Thierry Post*, Selçuk Karabatı[†] & Stelios Arvanitis[‡]

June 29, 2018

Abstract

A methodology is developed for constructing robust forecast combinations which improve upon a given benchmark specification for all symmetric and convex loss functions. The optimal forecast combination asymptotically almost surely dominates the benchmark and, in addition, minimizes the expected loss function, under standard regularity conditions. The optimum in a given sample can be found by solving a large Convex Optimization problem. An application to forecasting of changes of the S&P 500 Volatility Index shows that robust optimized combinations improve significantly upon the out-of-sample forecasting accuracy of simple averaging and unrestricted optimization.

Key words: Forecast combinations, Stochastic Dominance, Asymptotic theory, Convex Optimization, Volatility forecasting

JEL code: C44, C54, C61.

*Post is Professor of Finance at the Graduate School of Business of Nazarbayev University; Astana, Kazakhstan; e-mail: thierrypost@hotmail.com.

[†]Karabatı is Professor of Operations Management at Koç University; 34450 Sarıyer/Istanbul, Turkey; e-mail: skarabati@ku.edu.tr.

[‡]Arvanitis is Assistant Professor at the Department of Economics of Athens University of Economics and Business; Athens; Greece; email: stelios@aueb.gr.

1 Introduction

A body of literature starting with [Bates and Granger \(1969\)](#) and [Newbold and Granger \(1974\)](#) aims to improve forecast accuracy by combining multiple individual forecast models; [Clemen \(1989\)](#) provides a review and annotated bibliography of the early work in this area. Forecast combinations with estimated optimal weights however often trail simple averages out of sample, due to estimation error; see [Smith and Wallis \(2009\)](#), [Claeskens, Magnus, Vasnev and Wang \(2016\)](#) and [Chan and Pauwels \(2018\)](#), among others. The present study aims to contribute to this literature by developing a methodology for constructing robust combinations which stochastically dominate a given benchmark specification.

Stochastic Dominance (SD; [Hadar and Russell \(1969\)](#), [Hanoch and Levy \(1969\)](#) and [Rothschild and Stiglitz \(1970\)](#)) was originally developed to compare risky choice alternatives using utility functions. [Jin, Corradi & Swanson \(2017\)](#) extend the use of this stochastic order to forecast comparison based on loss functions. An SD relation arises if a given forecast achieves a lower expected loss than a second forecast for all permissible loss functions, allowing for a robust classification.

Whereas the previous study tests for dominance relations among a set of given forecast models in the spirit of [Schmid and Trede \(1998\)](#), [Anderson \(1996\)](#), [Davidson and Duclos \(2000\)](#), [Barrett and Donald \(2003\)](#) and [Linton, Maasoumi and Whang \(2005\)](#), the present study seeks to combine given models to build a dominance relation in the spirit of [Kuosmanen \(2004\)](#), [Roman, Darby-Dowman and Mitra \(2006\)](#) and [Post, Karabati and Arvanitis \(2017\)](#).

This application of SD seems even more promising than the application to forecast model comparison. Optimization introduces a greater need for robustness and, in addition, optimization enhances the discriminatory power of stochastic orders.

Optimization based on a given loss function fuels over-fitting to the data and generally exhibits limited predictive ability out of sample. The over-fitting problem is reduced by requiring improvements for all permissible loss functions, including ones which are more robust against outliers than the standard Gaussian.

Whereas SD comparisons tends to suffer from low discriminatory power due to incomparability of the evaluated forecast models, the proposed methodology can construct forecast combinations which exhibit the desired SD relation.

The optimal forecast combination weights are found using numerical optimization. A Convex Optimization (CO) problem is obtained by applying a linearization of partial moments in the spirit of [Rockafellar and Uryasev \(2000\)](#). Similar linearizations have been used in optimization with standard second-degree SD (SSD) constraints ([Dentcheva and Ruszczyński \(2003\)](#), [Kuosmanen \(2004\)](#) and [Roman, Darby-Dowman and Mitra \(2006\)](#)).

Relevant large-sample properties of the optimal forecast combination can be obtained using functional central limit theorems for the relevant estimator of the joint distribution function, by analogy to the existing asymptotic theory for SD relations in [Linton, Maasoumi and Whang \(2005\)](#), [Scaillet and Topaloglou \(2010\)](#), [Linton, Post and Whang \(2014\)](#) and [Post, Karabati and Arvanitis \(2017\)](#).

The limit theory generalizes results from [Post, Karabati and Arvanitis \(2017\)](#) by allowing for data-dependent slack variables to use results from the econometric literature about set identification. These generalized results can also be applied in portfolio optimization with standard SSD constraints.

A preliminary analysis reveals that the classes of General Loss (GL) functions and Convex Loss (CL) functions considered in [Jin, Corradi & Swanson \(2017, Definition 2.1\)](#) have limited discriminatory power because they allow for exotic asymmetric loss functions. The focus in the present study is therefore on a class of Symmetric Convex Loss (SCL) functions, which includes the standard Gaussian, Laplacian and Huber loss functions as special cases. The associated stochastic order is tentatively referred to as SCLSD.

An application to forecasting of daily changes of the Chicago Board Options Exchange (CBOE) S&P 500 Volatility Index (VIX) shows that robust optimized combinations improve significantly upon the out-of-sample forecasting accuracy of simple averaging and unrestricted optimization.

2 Theoretical Concepts

2.1 Preliminaries

A random variable X is forecast using $M \geq 2$ distinct forecast models, generating point forecasts $\mathbf{Y} := [Y_1 \cdots Y_M]$. The forecasts could be constructed, for example, using predictive regression, analysts forecasts or market prices of securities. The forecast models are evaluated based on their forecast errors $\mathbf{U} := X\mathbf{1}'_M - \mathbf{Y}$. The joint cumulative distribution function (CDF) of the errors

is denoted by $\mathcal{F} : \mathcal{Z}^M \rightarrow [0, 1]$, where $\mathcal{Z} := [a, b]$, $-\infty < a < 0 < b < +\infty$.

Convex combinations of the forecast models can be created using mixing weights which are represented by $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$, $\boldsymbol{\Lambda} := \{\boldsymbol{\lambda} \in \mathbb{R}^M : \boldsymbol{\lambda}' \mathbf{1}_M = 1; \boldsymbol{\lambda} \geq \mathbf{0}_M\}$. The errors of a given forecast combination can be expressed as $\mathbf{U}\boldsymbol{\lambda} = \mathbf{X} - \mathbf{Y}\boldsymbol{\lambda}$. The marginal CDF is $\mathcal{F}_{\boldsymbol{\lambda}}(u) = \int_{\{\mathbf{U} : \mathbf{U}\boldsymbol{\lambda} = u\}} \mathcal{F}(\mathbf{U}) d\mathbf{U}$.

It is straightforward to generalize the analysis from the unit simplex $\boldsymbol{\Lambda}$ to a general polytope $\mathbf{K}$, for example, to correct for possible biases of the forecast models. Any polytope can be formulated as the convex hull of its vertices. Therefore, the individual forecasts may be replaced with the most extreme feasible combinations. An alternative, equivalent approach is obtained by formulating the polytope as the intersection of halfspaces. This formulation however requires a generalization of the proposed notation based on $\mathbf{X} - \mathbf{Y}\boldsymbol{\lambda} = \mathbf{U}\boldsymbol{\lambda} - \mathbf{X}(\boldsymbol{\lambda}' \mathbf{1}_M - 1)$ instead of $\mathbf{X} - \mathbf{Y}\boldsymbol{\lambda} = \mathbf{U}\boldsymbol{\lambda}$.

2.2 Stochastic order

Let $\mathcal{L}$ be the class of SCL functions $L : \mathcal{Z} \rightarrow \mathbb{R}_+$ which achieve a minimum at $L(0) = 0$, increase as the error moves away from zero and obey symmetry: $L(E) = L(|E|)$. This class includes the standard Gaussian, Laplacian and Huber loss functions as special cases.

Instead of pursuing a complete order for a single ‘optimal’ loss function $L \in \mathcal{L}$, the analysis relies on a partial order which considers the entire class of loss functions:

Definition 1 (SCLSD): *Forecast combination $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$ stochastically dominates forecast combination $\boldsymbol{\tau} \in \boldsymbol{\Lambda}$ for symmetric convex loss functions if*

$$\mathbb{E}_{\mathcal{F}} [L(\mathbf{U}\boldsymbol{\lambda})] \leq \mathbb{E}_{\mathcal{F}} [L(\mathbf{U}\boldsymbol{\tau})] \quad \forall L \in \mathcal{L}. \quad (1)$$

In other words, SCLSD occurs if the first combination achieves a lower expected loss than the second combination for every permissible loss function. If this dominance relation can be established, then the analyst does not have to choose a specific loss function to rank the two forecasts combinations. It seems particularly comforting that the ranking arises also for loss functions which are robust to outliers.

The class $\mathcal{L}$ is closely related to standard Second-degree Stochastic Dominance (SSD). Specifically, $U(x) := -L(-x)$, $x \leq 0$, is an increasing and concave

utility function and the minimization of $\mathbb{E}_{\mathcal{F}} [L(E)]$ is equivalent to the maximization of the expectation of $\mathbb{E}_{\mathcal{F}} [U(-|E|)] = -\mathbb{E}_{\mathcal{F}} [L(|E|)]$. SCLSD in terms of forecast error E thus corresponds to SSD in terms of negative absolute forecast error $(-|E|)$.

Every permissible loss function $L \in \mathcal{L}$ is a positive mixture of singularity functions $L_z(x) := (z + |x|)_+$, $z \in \mathcal{Z}^-$, where $()_+$ is the positive part and $\mathcal{Z}^- := [\min(a, -b), 0]$. Using this insight, the stochastic order can be reformulated as follows:

Proposition 2.2.1 (SCLSD): *Forecast combination $\lambda \in \Lambda$ stochastically dominates forecast combination $\tau \in \Lambda$ iff*

$$\mathbb{E}_{\mathcal{F}} [L_z(\mathbf{U}\lambda)] \leq \mathbb{E}_{\mathcal{F}} [L_z(\mathbf{U}\tau)] \quad \forall z \in \mathcal{Z}^-. \quad (2)$$

For some fixed and given benchmark forecast combination $\tau \in \Lambda$, the set of dominant combinations is given by

$$\Lambda_{\mathcal{F}}^{\succ} := \{\lambda \in \Lambda : \mathbb{E}_{\mathcal{F}} [L_z(\mathbf{U}\lambda)] \leq \mathbb{E}_{\mathcal{F}} [L_z(\mathbf{U}\tau)] \quad \forall z \in \mathcal{Z}^-\}. \quad (3)$$

This set is non-empty, closed and convex under the maintained assumptions about the joint distribution, loss functions and opportunity set.

2.3 Optimal combinations

An optimization problem is proposed to construct a feasible forecast combination which dominates the benchmark. The goal function is the reduction of expected loss for some given loss function, or $G_{\mathcal{F},\lambda} := \mathbb{E}_{\mathcal{F}} [L(\mathbf{U}\tau)] - \mathbb{E}_{\mathcal{F}} [L(\mathbf{U}\lambda)]$, $L \in \mathcal{L}$. The following optimization problem is proposed:

$$\max_{\lambda \in \Lambda_{\mathcal{F}}^{\succ}} G_{\mathcal{F},\lambda}. \quad (4)$$

If the loss function is strictly convex, the solution will be unique and the optimal combination will be efficient (not dominated by a nonequivalent alternative). In case of local linearity of the loss function, multiple optimal solutions

may exist. In this case, a second optimization problem could be solved to detect secondary improvement possibilities and avoid a solution which is inefficient.

Although optimization problem (4) is convex, the number of constraints is uncountable and, furthermore, the constraints contain the intractable maximum part operator $(\cdot)_+$. Fortunately, an empirical counterpart of the problem can be reformulated as a standard CO problem by linearizing the $(\cdot)_+$ operator using additional model variables; see Section 3.4.

3 Empirical Counterparts

The present section introduces the empirical counterpart of the forecast optimization problem and derives a limit theory for the optimal forecast combinations. The assumption framework is gradually refined to obtain increasingly strong results. The results can be used for constructing statistical inference procedures for the optimal forecasts in the spirit of Diebold and Mariano (1995) and analyzing the effect of using different CDF estimators, notably special cases of the general BEL ICDF.

To derive the limit theory, some additional notation is introduced. In what follows, $\|\cdot\|$ denotes the Euclidean norm, $\ell^\infty(A)$ the space of real-valued bounded functions on a set A equipped with the sup norm, and $\rightsquigarrow$ convergence in distribution. $\bar{B}_{\mathbf{\lambda}}(\eta)$ denotes the closed Euclidean ball in $\mathbb{R}^M$ centered at $\mathbf{\lambda}$ with radius equal to $\eta > 0$.

3.1 Empirical optimization problem

In practice, the CDF $\mathcal{F}$ is latent and the analyst has access to a discrete time series of realizations x_t , $\mathbf{y}_{t-1} := [y_{1,t-1} \cdots y_{M,t-1}]$, $\boldsymbol{\varepsilon}_t := x_t \mathbf{1}'_M - \mathbf{y}_{t-1}$, $t = 1, \dots, T$. Let $\mathbf{E} := [\boldsymbol{\varepsilon}_1 \cdots \boldsymbol{\varepsilon}_T]'$, $\mathcal{Z}_T := \{\boldsymbol{\varepsilon}_t \boldsymbol{\tau}\}_{t=1}^T$ and $\mathcal{Z}_T^- := \{-|\boldsymbol{\varepsilon}_t \boldsymbol{\tau}|\}_{t=1}^T$.

Special attention is given to non-parametric and discrete estimators, which fit well in the distribution-free assumption framework and allow for discrete mathematics and numerical optimization. Using the data set, a discrete estimator for the CDF is constructed:

$$F_T(\mathbf{U}) := \sum_{t=1}^T p_t \mathbb{I}(\boldsymbol{\varepsilon}_t \leq \mathbf{U}). \quad (5)$$

The estimator is essentially a multinomial with atoms at the observations.

The probability weights are collected in a probability vector $\mathbf{p} := [p_1 \cdots p_T]'$, $\mathbf{1}'_T \mathbf{p} = 1$, $\mathbf{p} \geq \mathbf{0}_T$.

An obvious specification is the empirical cumulative distribution function (ECDF), or $\mathbf{p} = T^{-1} \mathbf{1}_T$. The probability weights can alternatively be estimated using the implied CDF (ICDF) estimators of Generalized Methods of Moments (Back and Brown (1993) and Generalized Empirical Likelihood (Qin and Lawless (1994)), which allows for accounting for side information and dynamic patterns which are not captured by the ECDF.

The limit theory in Section 3 will pay special attention to the Blockwise Empirical Likelihood (BEL) methodology (Kitamura (1997)), extending earlier work by Post, Karabati and Arvanitis (2017) in the context of portfolio optimization with standard SSD constraints.

To approximate the dominant set $\Lambda_{\mathcal{F}}^{\succ}$, it is important to account for boundary problems which may arise in case of binding inequalities (see, for example, Andrews and Soares (2010)). These problems are addressed here by introducing random slacks in the right-hand sides of the empirical dominance constraints. Using a stochastic function $c_T : \mathcal{Z}^- \rightarrow \mathbb{R}_+$ to specify the permissible slacks, the following empirical version of $\Lambda_{\mathcal{F}}^{\succ}$ is employed:

$$\Lambda_{\mathcal{F}_T}^{\succ}(c_T) := \{\boldsymbol{\lambda} \in \Lambda : \mathbb{E}_{\mathcal{F}_T} [L_z(\mathbf{U}\boldsymbol{\lambda})] \leq \mathbb{E}_{\mathcal{F}_T} [L_z(\mathbf{U}\boldsymbol{\tau})] + c_T(z) \quad \forall z \in \mathcal{Z}^-\}. \quad (6)$$

The empirical counterpart of the goal function is simply $\mathcal{G}_{\mathcal{F}_T, \boldsymbol{\lambda}} = \mathbb{E}_{\mathcal{F}_T} [L(\mathbf{U}\boldsymbol{\tau})] - \mathbb{E}_{\mathcal{F}_T} [L(\mathbf{U}\boldsymbol{\lambda})]$. Maximizing this goal function subject to the empirical dominance constraints gives the following optimization problem:

$$\max_{\boldsymbol{\lambda} \in \Lambda_{\mathcal{F}_T}^{\succ}(c_T)} \mathbb{E}_{\mathcal{F}_T} [L(\mathbf{U}\boldsymbol{\tau})] - \mathbb{E}_{\mathcal{F}_T} [L(\mathbf{U}\boldsymbol{\lambda})]. \quad (7)$$

3.2 Empirical process

A central assumption for the limit theory is that the SCLSD empirical process $\sqrt{T}(\mathbb{E}_{\mathcal{F}_T} [L_z(\mathbf{U}\boldsymbol{\lambda})] - \mathbb{E}_{\mathcal{F}} [L_z(\mathbf{U}\boldsymbol{\lambda})])$ converges weakly to a well-defined, zero-mean Gaussian process:

Assumption 3.1. As $T \rightarrow \infty$

$$\sqrt{T} (\mathbb{E}_{\mathcal{F}_T} (L_z (X_t - \mathbf{Y}_t \boldsymbol{\lambda})) - \mathbb{E}_{\mathcal{F}} (L_z (X_t - \mathbf{Y}_t \boldsymbol{\lambda}))) \rightsquigarrow \mathcal{G}(\boldsymbol{\lambda}, z), \text{ in } \ell^\infty(\boldsymbol{\Lambda} \times \mathcal{Z}^-), \quad (8)$$

where $\mathcal{G}$ is a zero mean Gaussian process on $\boldsymbol{\Lambda} \times \mathcal{Z}$ with continuous sample paths and covariance kernel $K_{\mathcal{G}} : (\boldsymbol{\Lambda} \times \mathcal{Z}^-)^2 \rightarrow \mathbb{R}$ positive definite.

Lower-level assumptions can be specified to justify this assumption in the context of the ECDF or the more general BEL ICDF. A first assumption is that the bounded joint process $(\text{vec}(X_t, \mathbf{Y}_t))_{t \in \mathbb{Z}}$ is stationary and strongly mixing with mixing coefficients $(\alpha_k)_{k \in \mathbb{N}}$ which satisfy $\alpha_k = O(k^{-r})$ for $r > 1$. Common strictly stationary versions of multivariate ARMA, GARCH and stochastic volatility models obey this assumption.

Further lower-level assumptions include a rate restriction of $o(T^{\frac{1}{2}})$ on the block size, boundedness of vector-valued moment function $\mathbf{g}$, unbiasedness of the moment conditions, that is, $\mathbb{E}_{\mathcal{F}}[\mathbf{g}(\mathbf{x})] = \mathbf{0}$, and limiting behavior for their variance of the form $\text{Var} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{g}(X_t) \right] \rightarrow V_g$ which is positive definite.

A final assumption involves the use of the bracketing number $\mathcal{N}_\epsilon$ of the set $\{L_z(X - \mathbf{Y}\boldsymbol{\lambda}), (\boldsymbol{\lambda}, z) \in (\boldsymbol{\Lambda} \times \mathcal{Z}^-)\}$, or the minimum number of relevant ϵ -small brackets that can cover the set. The limiting behavior of the bracketing number as $\epsilon \rightarrow 0$ can be used to construct measures of set complexity. When the bracketing integral $\int_0^1 \sqrt{\frac{\mathcal{N}_\epsilon}{\epsilon}} d\epsilon$ is finite, then arguments similar to those in the proof of Lemma 3.1.2 of Post, Karabati and Arvanitis (2017) imply the validity of Assumption 3.1 with the following covariance kernel

$$K_{\mathcal{G}}((\boldsymbol{\lambda}_1, z_1), (\boldsymbol{\lambda}_2, z_2)) = \sum_{i=0}^{\infty} \kappa_i \text{Cov}_{\mathcal{F}}(L_{z_1}(X_0 - \mathbf{Y}_0 \boldsymbol{\lambda}_1), L_{z_2}(X_i - \mathbf{Y}_i \boldsymbol{\lambda}_2)) - \sum_{i=0}^{\infty} \kappa_i \mathbb{E}_{\mathcal{F}}[\mathbf{g}'(X_0) L_{z_1}(X_i - \mathbf{Y}_i \boldsymbol{\lambda}_1)] V_g^{-1} \sum_{i=0}^{\infty} \kappa_i \mathbb{E}_{\mathcal{F}}[\mathbf{g}(X_0) L_{z_2}(X_i - \mathbf{Y}_i \boldsymbol{\lambda}_2)] ,$$

$$\text{for } \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \boldsymbol{\Lambda}, z_1, z_2 \in \mathcal{Z}^-, \kappa_i = \begin{cases} 1, & i = 0 \\ 2, & i > 0 \end{cases}.$$

Since every $L \in \mathcal{L}$ can be characterized as $\int_{\mathcal{Z}^-} w(z) L_z dz$ for some positive mixture function $w : \mathcal{Z}^- \rightarrow [0, 1]$, with $\int_{\mathcal{Z}^-} w(z) dz = 1$, the CMT directly implies the following result:

Lemma 3.2. Under Assumption 3.1, as $T \rightarrow \infty$,

$$\sqrt{T} (\mathbb{E}_{\mathcal{F}_T} (L(X_t - \mathbf{Y}_t \boldsymbol{\lambda})) - \mathbb{E}_{\mathcal{F}} (L(X_t - \mathbf{Y}_t \boldsymbol{\lambda}))) \rightsquigarrow \mathcal{G}_L(\boldsymbol{\lambda}), \text{ in } \ell^\infty(\boldsymbol{\Lambda}), \quad (9)$$

where $\mathcal{G}_L(\boldsymbol{\lambda}) := \int_{\mathcal{Z}^-} w(z) \mathcal{G}(\boldsymbol{\lambda}, z) dz$ is a zero-mean Gaussian process with co-

variance kernel $K_{\mathcal{G}_L}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) := \int_{\mathcal{Z}^- \times \mathcal{Z}^-} w(z) w(z^*) K_{\mathcal{G}}((\boldsymbol{\lambda}_1, z), (\boldsymbol{\lambda}_2, z^*)) dz dz^*$.

Due to the bilinearity of the covariance kernel $K_{\mathcal{G}}$, in the case of the BEL-ICDF estimator, the covariance kernel is given by

$$K_{\mathcal{G}_L}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \sum_{i=0}^{\infty} \kappa_i \text{Cov}(L(X_0 - \mathbf{Y}_0 \boldsymbol{\lambda}_1), L(X_i - \mathbf{Y}_i \boldsymbol{\lambda}_2)) \\ - \sum_{i=0}^{\infty} \kappa_i \mathbb{E}_{\mathcal{F}}[g'(X_0) L(X_i - \mathbf{Y}_i \boldsymbol{\lambda}_1)] V_g^{-1} \sum_{i=0}^{\infty} \kappa_i \mathbb{E}_{\mathcal{F}}[g(X_0) L(X_i - \mathbf{Y}_i \boldsymbol{\lambda}_2)] .$$

In the special case of the ECDF estimator, the expression is further reduced to

$$K_{\mathcal{G}_L}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \sum_{i=0}^{\infty} \kappa_i \text{Cov}_{\mathcal{F}}(L(X_0 - \mathbf{Y}_0 \boldsymbol{\lambda}_1), L(X_i - \mathbf{Y}_i \boldsymbol{\lambda}_2)) .$$

3.3 Consistency Properties

The present subsection deals with consistency properties of the empirical dominance classification, the optimal value of the goal function and the optimal values of the portfolio weights. In what follows, $\boldsymbol{\lambda}_T$ denotes an optimal forecast combination, or a solution to the empirical optimization problem (7).

The following proposition establishes that false dominance classifications asymptotically almost never occur:

Proposition 3.3. *If Assumption 3.1.1 holds and $\sup_{z \in \mathcal{Z}^-} c_T(z) \rightarrow 0$, then as $T \rightarrow \infty$, $\boldsymbol{\lambda}_T \in \Lambda_{\mathcal{F}}^{\succ}$ with probability converging to one.*

This result does not suffice to guarantee that $\boldsymbol{\lambda}_T$ converges to an optimal solution of the latent optimization problem (4). To achieve stronger results, several additional assumptions and concepts are introduced.

Some optimal solutions may have contact points, that is, $z \in \mathcal{Z}^-$ for which $\mathbb{E}_{\mathcal{F}}[L_z(\mathbf{U}\boldsymbol{\tau})] = \mathbb{E}_{\mathcal{F}}[L_z(\mathbf{U}\boldsymbol{\lambda})]$. To deal with resulting boundary problems, the assumptions on the slack process are tightened: $c_T(z)$ is assumed to be strictly positive and become asymptotically uniformly negligible at a slower than $\sqrt{T}$ rate.

To establish the asymptotic approximation of $\Lambda_{\mathcal{F}}^{\succ}$ by $\Lambda_{\mathcal{F}_T}^{\succ}(c_T)$ under these assumptions, a stochastic version of Painleve-Kuratowski set convergence (see Molchanov (2006, Appendix B)) is used. A sequence of non-empty closed subsets $\Lambda_T \subseteq \Lambda$ is said to Painleve-Kuratowski converge to a closed subset $\Lambda^* \subseteq \Lambda$, if and only if any $\boldsymbol{\lambda} \in \Lambda^*$ is simultaneously a cluster point of some sequence $\boldsymbol{\lambda}_T$ and a limit point in probability of some sequence $\boldsymbol{\lambda}_T^*$, with $\boldsymbol{\lambda}_T, \boldsymbol{\lambda}_T^* \in \Lambda_T$ for

all T . More generally, if Λ_T and/or Λ^* are random closed subsets of Λ (in the sense of Molchanov (2006, Chapter 1)), then Λ_T Painleve-Kuratowski converges to Λ^* in probability, if and only if the aforementioned limit and cluster point properties hold with probability converging to one.

The solution to (4) may not be not unique. In case of multiple optimal solutions, several subsequences of λ_T might approximate different solutions. In order to ensure a unique solution, strict convexity is assumed for the loss function L in the goal function.

The following theorem establishes various additional consistency properties using the aforementioned assumptions:

Theorem 3.4. *Suppose that $\sup_{z \in \mathcal{Z}} c_T(z) \rightarrow 0$ and $\sqrt{T} \inf_{z \in \mathcal{Z}} c_T(z) \rightarrow \infty$ in probability. Then, under Assumption 3.1, as $T \rightarrow \infty$, a) $\Lambda_{\mathcal{F}_T}^{\succ}(c_T)$ converges in probability to $\Lambda_{\mathcal{F}}^{\succ}$ in the Painleve-Kuratowski sense and, consequently, b) $\mathbb{E}_{\mathcal{F}_T}[L(\mathbf{U}\tau)] - \mathbb{E}_{\mathcal{F}_T}[L(\mathbf{U}\lambda_T)]$ converges in probability to $\max_{\lambda \in \Lambda_{\mathcal{F}}^{\succ}} G_{\mathcal{F},\lambda}$. If moreover L is strictly convex then, c) $\lambda_T \rightsquigarrow \lambda_{\mathcal{F}}^*$ where $\lambda_{\mathcal{F}}^*$ is the unique solution to (4).*

The assumption of strict convexity applies for the piecewise quadratic loss function which underlies the MSFE goal function which is used in the application section. However, it does not hold in the piecewise linear case of the MAFE goal function. In the latter case, multiple optimal forecast combinations may exist and point convergence cannot be established. Nevertheless, it can still be established that the optimal solution features asymptotically vanishing decision errors for the dominance classification and the optimal value of the goal function; Proposition 3.3 and Theorem 3.4b do not require strict convexity.

3.4 Rates and Limiting Distributions

The limit distributions of the optimal forecast combinations are derived using the above results and a local quadratic approximation for $\mathbb{E}_{\mathcal{F}_T}[L(\cdot)]$ in a neighborhood of $\lambda_{\mathcal{F}}^*$. To address complications which stem from the possibility that $\lambda_{\mathcal{F}}^*$ lies on the boundary of $\Lambda_{\mathcal{F}}^{\succ}$, the following additional assumption is introduced:

Assumption 3.5. *Suppose that the following hold:*

1. For some $\eta > 0$, and any $\lambda \in \Lambda_{\mathcal{F}}^{\succ} \cap \bar{B}_{\lambda_{\mathcal{F}}^*}(\eta)$,

$$\begin{aligned} & T \left(\mathbb{E}_{\mathcal{F}_T} \left[L \left(\mathbf{U} \left(\lambda_{\mathcal{F}}^* + \frac{1}{\sqrt{T}} \mathbf{u} \right) \right) \right] - \mathbb{E}_{\mathcal{F}_T} [L(\mathbf{U} \lambda_{\mathcal{F}}^*)] \right) \\ &= \sqrt{T} \mathbb{E}_{\mathcal{F}_T} [s_t(\lambda_{\mathcal{F}}^*)]^T \mathbf{u} + \frac{1}{2} \mathbf{u}^T \mathbb{E}_{\mathcal{F}_T} [H_t(\mu_{\mathcal{F}}^*)] \mathbf{u}, \end{aligned} \quad (10)$$

where $s_t(\cdot)$ is a random $M \times 1$ vector function and $H_t(\cdot)$ a random $M \times M$ matrix function defined on $\Lambda_{\mathcal{F}}^{\succ} \cap \bar{B}_{\lambda_{\mathcal{F}}^*}(\eta)$, $\mathbf{u} := \sqrt{T}(\lambda - \lambda_{\mathcal{F}}^*)$ and $\mu_{\mathcal{F}}^*$ is some element of $\Lambda_{\mathcal{F}}^{\succ} \cap \bar{B}_{\lambda_{\mathcal{F}}^*}(\eta)$ lying on the ray that connects λ and $\lambda_{\mathcal{F}}^*$.

2. As $T \rightarrow \infty$,

$$\sqrt{T} \mathbb{E}_{\mathcal{F}_T} [s_t(\lambda_{\mathcal{F}}^*)] \rightsquigarrow s_{\lambda_{\mathcal{F}}^*} \sim N(\mathbf{0}, V_{\lambda_{\mathcal{F}}^*}), \quad (11)$$

with $V_{\lambda_{\mathcal{F}}^*}$ a pd $M \times M$ matrix, and for any $\lambda_T \rightarrow \lambda_{\mathcal{F}}^*$,

$$\mathbb{E}_{\mathcal{F}_T} [H_t(\lambda_T)] \rightsquigarrow H_{\lambda_{\mathcal{F}}^*}, \quad (12)$$

with $H_{\lambda_{\mathcal{F}}^*}$ a pd $M \times M$ matrix.

Assumption 3.5.1 can be derived from continuous second-order (and potentially one-sided) differentiability of L on $[a, b]$ and the relevant first-order Taylor expansion of L with second-order remainders obtained via the Mean Value Theorem, in the sense of Andrews (1999). In this case, $s_t(\lambda) = \frac{dL}{dx}(\mathbf{U}(\lambda)) Y_t$ and $H_t(\lambda) = \frac{d^2 L}{dx^2}(\mathbf{U}(\lambda)) Y_t Y_t^T$, where the derivatives involved may be one-sided and, furthermore, (10) essentially holds for any $\eta > 0$, given the convexity of $\Lambda_{\mathcal{F}}^{\succ}$.

Under the aforementioned continuous differentiability conditions, and if $\mathcal{F}_T$ is the BEL-ICDF estimator, under the stationarity, mixing and boundedness conditions that justify the validity of Assumption 3.1, both (11) and (12) in Assumption 3.5.2 follow with

$$\begin{aligned} V_{\lambda_{\mathcal{F}}^*} &= \sum_{i=0}^{\infty} \kappa_i \text{Cov} \left(\frac{dL(X_0 - \mathbf{Y}_0 \lambda_{\mathcal{F}}^*)}{dx} Y_0, \frac{dL(X_i - \mathbf{Y}_i \lambda_{\mathcal{F}}^*)}{dx} Y_i \right) \\ &- \sum_{i=0}^{\infty} \kappa_i \text{Cov} \left[\mathbf{g}(X_0), \frac{dL(X_i - \mathbf{Y}_i \lambda_{\mathcal{F}}^*)}{dx} Y_i \right] V_g^{-1} \sum_{i=0}^{\infty} \kappa_i \text{Cov} \left[\mathbf{g}(X_0), \frac{dL(X_i - \mathbf{Y}_i \lambda_{\mathcal{F}}^*)}{dx} Y_i \right] \\ H_{\lambda_{\mathcal{F}}^*} &= \mathbb{E} \left[\frac{d^2 L(X_0 - \mathbf{Y}_0 \lambda_{\mathcal{F}}^*)}{dx^2} Y_0 Y_0^T \right], \end{aligned}$$

with $\kappa_i = \begin{cases} 1, & i = 0 \\ 2, & i > 0 \end{cases}$, due to arguments similar to the ones in Kitamura (1997).

Analogously, for the ECDF estimator,

$$V_{\lambda_{\mathcal{F}}^*} = \sum_{i=0}^{\infty} \kappa_i \text{Cov} \left(\frac{dL(X_0 - Y_0 \lambda_{\mathcal{F}}^*)}{dx} Y_0, \frac{dL(X_i - Y_i \lambda_{\mathcal{F}}^*)}{dx} Y_i \right),$$

while $H_{\lambda_{\mathcal{F}}^*}$ is as above. In both cases, given the strict convexity of L , the limiting $H_{\lambda_{\mathcal{F}}^*}$ matrix is positive definite if the elements of Y_0 are linearly independent.

The main result of the limit theory now follows:

Theorem 3.6. *Under Assumptions 3.1 and 3.5, if λ_T solves (7) and $\lambda_T \rightsquigarrow \lambda_{\mathcal{F}}^*$, then, as $T \rightarrow \infty$,*

$$\sqrt{T}(\lambda_T - \lambda_{\mathcal{F}}^*) \rightsquigarrow \arg \min_{\mathbf{u} \in \Lambda_{\infty}^{\succeq}} \frac{1}{2} \left(\mathbf{u} + H_{\lambda_{\mathcal{F}}^*}^{-1} s_{\lambda_{\mathcal{F}}^*} \right)^T H_{\lambda_{\mathcal{F}}^*} \left(\mathbf{u} + H_{\lambda_{\mathcal{F}}^*}^{-1} s_{\lambda_{\mathcal{F}}^*} \right), \quad (13)$$

where $\Lambda_{\infty}^{\succeq}$ is the non-empty closed and convex Painleve-Kuratowski limit of $\sqrt{T}(\Lambda_{\mathcal{F}}^{\succeq} \cap \bar{B}_{\lambda_{\mathcal{F}}^*}(\eta) - \lambda_{\mathcal{F}}^*)$.

The theorem implies standard $\sqrt{T}$ rates and an asymptotic distribution obtained from the linear projection via the norm $\sqrt{(\cdot)^T H_t(\lambda_{\mathcal{F}}^*)(\cdot)}$ of the zero-mean normal random vector $-H_{\lambda_{\mathcal{F}}^*}^{-1} s_{\lambda_{\mathcal{F}}^*}$ to $\Lambda_{\infty}^{\succeq}$. The latter is a non-empty closed and convex subset of $\mathbb{R}^M$ and it necessarily contains zero. When $\lambda_{\mathcal{F}}^*$ is an interior point of $\Lambda_{\mathcal{F}}^{\succeq}$ and therefore $\Lambda_{\infty}^{\succeq} = \mathbb{R}^M$, a more familiar result is obtained:

$$\sqrt{T}(\lambda_T - \lambda_{\mathcal{F}}^*) \rightsquigarrow -H_{\lambda_{\mathcal{F}}^*}^{-1} s_{\lambda_{\mathcal{F}}^*} \sim N \left(0, H_{\lambda_{\mathcal{F}}^*}^{-1} V_{\lambda_{\mathcal{F}}^*} H_{\lambda_{\mathcal{F}}^*}^{-1} \right). \quad (14)$$

The limiting distribution depends on the system of inequalities in (2) and thereby on the benchmark τ , because the asymptotic variance depends on $\lambda_{\mathcal{F}}^*$ and the limiting cone $\Lambda_{\infty}^{\succeq}$ depends on $\lambda_{\mathcal{F}}^*$ and $\Lambda_{\mathcal{F}}^{\succeq}$.

Theorem 3.6 can be used to analyze the relative asymptotic efficiency of alternative CDF estimators (for example, ECDF vs. BEL-ICDF). Furthermore, using (14), one can construct asymptotic inference procedures in the spirit of Paragraph 1.1 of [Diebold and Mariano \(1995\)](#).

4 Numerical Optimization

4.1 Optimization problem

The empirical application will focus on the case with constant slacks: $c_T(z) = c_T$ $\forall z \in \mathcal{Z}^-$. In this case, the empirical dominance conditions in (6) can be reduced to the following finite system

$$\mathbb{E}_{\mathcal{F}_T} [L_z(\mathbf{U}\boldsymbol{\lambda})] \leq \mathbb{E}_{\mathcal{F}_T} [L_z(\mathbf{U}\boldsymbol{\tau})] + c_T \quad \forall z \in \mathcal{Z}_T^-. \quad (15)$$

The partial moments are evaluated only at the elements of $\mathcal{Z}_T^-$ rather than the entire support $\mathcal{Z}^-$. This reduction is allowed because $\mathbb{E}_{\mathcal{F}_T} [L_z(\mathbf{U}\boldsymbol{\tau})]$ by construction is a piece-wise linear, convex function of the threshold $z \in \mathcal{Z}^-$ with kinks only at the elements of $\mathcal{Z}_T^-$; consequently, if the constraints are satisfied for all $z \in \mathcal{Z}_T^-$, then they are also satisfied for all $z \in \mathcal{Z}^-$. A similar result arises in [Bawa, Bodurtha, Suri and Rao \(1985, Section IC.1\)](#) for standard SD criteria for pairwise and multiple comparison.

Let $\mathbf{z}$ a $(T \times 1)$ vector with the ranked elements of $\mathcal{Z}_T^-$ and $\boldsymbol{\sigma}$ a $(T \times 1)$ vector of corresponding values of $(\mathbb{E}_{\mathcal{F}_T} [L_z(\mathbf{U}\boldsymbol{\tau})] + c_T)$, $z \in \mathcal{Z}_T^-$.

Using this notation, the empirical dominance conditions (15) can be rewritten as follows:

$$\mathbf{p}'(\mathbf{1}_T \mathbf{z}' + |\mathbf{E}\boldsymbol{\lambda}| \mathbf{1}_T')_+ \leq \boldsymbol{\sigma}'. \quad (16)$$

A convenient linearization is obtained in the spirit of [Rockafellar and Uryasev \(2000\)](#):

$$\begin{aligned} \mathbf{p}'\boldsymbol{\Theta} &\leq \boldsymbol{\sigma}' \\ -\boldsymbol{\Theta} - \mathbf{E}\boldsymbol{\lambda}\mathbf{1}_T' &\leq -\mathbf{1}_T \mathbf{z}' \\ -\boldsymbol{\Theta} + \mathbf{E}\boldsymbol{\lambda}\mathbf{1}_T' &\leq -\mathbf{1}_T \mathbf{z}' \\ \boldsymbol{\Theta} &\geq \mathbf{0}_{T \times T} \end{aligned} \quad (17)$$

Here, $\boldsymbol{\Theta}$ are additional model variables which capture the element-wise positive parts $(\mathbf{1}_T \mathbf{z}' + |\mathbf{E}\boldsymbol{\lambda}| \mathbf{1}_T')_+$ if system (16) is solvable.

The empirical counterpart of the goal function, can be formulated as $\mathcal{G}_{\mathcal{F}_T, \boldsymbol{\lambda}} = \mathbf{p}'\mathbf{L}(\mathbf{E}\boldsymbol{\lambda})$, using the vector-valued function $\mathbf{L}(\boldsymbol{\epsilon}) := [L(\epsilon_1) \cdots L(\epsilon_T)]'$.

The full optimization problem follows:

$$\begin{aligned}
& \max \mathbb{E}_{\mathcal{F}_T} [L_z(\mathbf{U}\boldsymbol{\tau})] - \mathbf{p}'\mathbf{L}(\mathbf{E}\boldsymbol{\lambda}) \\
& \text{s.t. } \mathbf{p}'\boldsymbol{\Theta} \leq \boldsymbol{\sigma}' \\
& -\boldsymbol{\Theta} - \mathbf{E}\boldsymbol{\lambda}\mathbf{1}'_T \leq -\mathbf{1}_T\mathbf{z}' \\
& -\boldsymbol{\Theta} + \mathbf{E}\boldsymbol{\lambda}\mathbf{1}'_T \leq -\mathbf{1}_T\mathbf{z}' \\
& \boldsymbol{\Theta} \geq \mathbf{0}_{T \times T} \\
& \boldsymbol{\lambda} \in \Lambda
\end{aligned} \tag{18}$$

Since $T \gg M$, the number of variables and constraints is $\mathcal{O}(T^2)$ and increases at a quadratic rate with the number of time-series observations. However, the problem is perfectly tractable for samples of hundreds of observations with standard computer hardware and software. For very samples of high-frequency data, high-performance platforms and specialized solver software are recommended.

4.2 Special cases

For important special cases of the goal function, the CO problem reduces to a standard Linear Programming (LP) or Convex Quadratic Programming (CQP) problem.

The goal function based on loss function $L_1(U) = |U|$ is the Mean Absolute Forecast Error (MAFE), $\mathcal{G}_{\mathcal{F},\boldsymbol{\lambda}} = \mathbb{E}_{\mathcal{F}} [L(\mathbf{U}\boldsymbol{\tau})] - \mathbb{E}_{\mathcal{F}} [|\mathbf{U}\boldsymbol{\lambda}|]$. A simple way to include this goal function assumes that the zero vector $\boldsymbol{\varepsilon}_T = \mathbf{0}'_M$ has been included in error data matrix $\mathbf{E}$ (which is inconsequential for system (17)). Using $\mathbf{e}_i$ for the i -th unit vector of proper dimensions, so that $(\boldsymbol{\Theta}\mathbf{e}_T)$ is the last column of $\boldsymbol{\Theta}$, the goal function can be linearized as follows:

$$\mathbb{E}_{\mathcal{F}_T} [|\mathbf{U}\boldsymbol{\lambda}|] = \mathbf{p}'\mathbf{L}_1(\mathbf{E}\boldsymbol{\lambda}) = \mathbf{p}'\boldsymbol{\Theta}\mathbf{e}_T, \tag{19}$$

Similarly, a convex quadratic goal function is obtained when the goal is the Mean Squared Forecast Error (MSFE) $\mathcal{G}_{\mathcal{F},\boldsymbol{\lambda}} = \mathbb{E}_{\mathcal{F}} [(\mathbf{U}\boldsymbol{\lambda})^2]$, which is based on $L_2(U) = U^2$:

$$\mathbb{E}_{\mathcal{F}_T} [(\mathbf{U}\boldsymbol{\lambda})^2] = \mathbf{p}'\mathbf{L}_2(\mathbf{E}\boldsymbol{\lambda}) = \mathbf{p}'\boldsymbol{\Theta}^2\mathbf{e}_T. \tag{20}$$

4.3 Hardware and software

The empirical analysis in Section 5 focuses on the special cases of MAFE (19) and MSFE (20). In these cases, the stochastic optimization problem reduce to a LP or CQP problem with limited computational complexity.

The linear and quadratic optimization problems are modeled and solved with IBM ILOG CPLEX Optimization Studio 12.8.0.0 on a computer with 2 x Intel(R) Xeon(R) CPU E5-2695 2.10GHz processors and 512GB memory. Particularly for the quadratic problems, the CPLEX's 'Numerical Emphasis' option is activated to guarantee convergence to optimal solutions.

5 Application

5.1 VIX forecasting

The proposed methodology is applied to the forecasting of daily log returns to the VIX. The VIX is a leading measure of implied volatility (IV) of short-term S&P 500 stock index option prices. Futures and options contracts written upon this index provide trading instruments related to market volatility.

The literature about IV predicting starts with [Harvey and Whaley \(1992\)](#) who forecast daily changes of IV of S&P100 index options with a view to trading option positions on the basis of the forecasts. This approach should not be confused with using IV estimates for forecasting volatility. A survey article by [Poon and Granger \(2003\)](#) concludes that IV estimates often provide more accurate volatility forecasts than competing econometric model-based forecasts.

A data set from January 02, 1990, through November 16, 2017 (7,024 trading days), is taken from the FRED database of the St. Louis Federal Reserve Bank. The first 750 days are used for creating initial forecasts and the data set used for optimization therefore starts on December 18, 1992. A separate analysis is made for the years 2007 and 2008 to account for the possible effects of the Global Financial Crisis (GFC).

Seven base forecasts are considered based on stylized facts which were documented already by [Harvey and Whaley \(1992\)](#) before the start of the sample period: the mean-reversal for short horizons and the decreasing pattern of market volatility as the number of remaining days in the week falls (higher VIX on

Mondays and lower VIX on Friday).

Six forecasts are based on simple linear univariate regressions based on a single instrument variable and a 500-day moving estimation window. The six instrument variables are (1) one-day lagged return, (2) moving average (MA) of the past two days, (3) MA of the past 5 days, (4) MA of the past 10 days, (5) MA of the past 20 days and (6) MA of the past 60 days. The seventh forecast is the MA in the past 500 days on the same weekday (MAW).

For estimating the optimal weights, the analysis uses a rolling estimation window with a length of $T = 250$ days. In each estimation window, the optimal forecast combination is constructed based on the seven base forecasts.

The objective is to minimize either MAFE or MSFE. The objective function is minimized both with and without the SCLSD constraints (17).

The benchmark for the SCLSD constraints is the Equal Weighted Average (EWA) of the seven forecasts. Optimality conditions for the EWA (under the population distribution) are discussed in [Capistran and Timmermann \(2009\)](#), [Hsiao and Wan \(2014\)](#) and [Chan and Pauwels \(2018\)](#). A possible alternative benchmark is $Y = 0$, which is motivated by the Random Walk Hypothesis. This alternative benchmark is however clearly inferior to the EWA in this application.

5.2 Forecast accuracy measurement

The forecast accuracy of the forecast combination is evaluated out of sample based on the forecast error for the first trading day after the end of the estimation window ('Day $T + 1$ ').

The first out-of-sample forecasts are made for December 15, 1993, based on an estimation window which ranges from December 18, 1992, through December 14, 1993. The next forecast is for December 16, 1994, and shifts the estimation window by one day, starting on December 21, 1992, and ending on December 15, 1993; and so forth.

To measure the out-of-sample goodness, the analysis reports the intercept (c_0), slope (c_1) and coefficient of determination (R^2) of the Mincer-Zarnowitz (MZ) regression of actual on forecasted values, together with the [Diebold and Mariano \(1995\)](#) statistics for the MAFE or MSFE of the optimized combination being equal to that of the EWA benchmark. Given that we use one-day ahead forecasts of returns and a long time series and the VIX is relatively close to a random walk process, corrections for serial correlation and small-sample bias

(Harvey, Leybourne and Newbold (1997)) have a minimal effect on the DM statistics. Accounting for the nested nature of the models (Clark and West (2007)) also does not change the conclusions.

Special attention is given to the predicted sign of the VIX change. The analysis reports the fraction of correct predictions for the sign of the VIX change ('Sign'). The analysis also studies the return-reward trade-off of a simple conditional volatility trading strategy which takes a long position in the volatility when an increase is predicted and a short position when a decrease is predicted. Since 2004, volatility trading can be implemented using VIX futures contracts. Prior to 2004, delta-neutral straddles on the underlying S&P 500 index would be the tradable instrument.

The strategy is very profitable on paper, with an annualized gross Sharpe ratio of around 2 (using the EWA forecast combination). However, the returns are computed here without accounting for financing costs, transactions costs, basis risk (for futures contracts), gamma risk (for straddles) and slippage due to delayed implementation of the daily strategy and hence materially overstates the net Sharpe ratio which is feasible in practice.

Since the net returns of the strategy are difficult to estimate without considering the available trading facilities, the focus here is on the incremental effects of using optimized forecast combinations, using the (annualized) Information Ratio (IR) relative to the trading returns based on the EWA forecast.

5.3 Results

Table I summarizes the out-of-sample forecast accuracy of the five competing forecast combinations: EWA, MAFE minimization, MAFE minimization with SCLSD constraints, MSFE minimization and MSFE minimization with SCLSD constraints.

EWA yields an MZ R^2 of about 2.7%. The limited forecast success is not surprising given the noise from unpredictable daily news flows and changes in market conditions at a daily horizon. The R^2 is however materially higher than the values shown in the analysis of S&P100 implied volatility for 1985-1989 in Harvey and Whaley (1992, Table 4). Furthermore, the predictive ability is potentially economically relevant, witness the aforementioned Sharpe Ratio.

The slope of the MZ regression of $c_1 = 2.143$ suggests that the EWA is biased out of sample. A closer inspection reveals that the bias stems from a relatively

high sensitivity to extreme observations of some of the univariate regression forecasts (namely those based on short-term reversal), which makes the EWA (which is less sensitive to outliers) too conservative.

The minimization of MAFE and MSFE improves significantly upon the statistical goodness of EWA, witness the DM statistics for MAFE and MSFE. The MZ regressions reveal that these improvements stem mostly from reducing the bias of the EWA and the MZ R^2 does not improve materially.

The minimization of MAFE leads to higher accuracy than the minimization of MSFE. The optimal combination weights reveal that the higher accuracy is achieved by placing more emphasis on day-of-the-week effects (Y_7 is the univariate regression forecasts based on the MAW variable).

In terms of economic significance, the minimization of MAFE and MSFE leads to negative IRs which amounts to deterioration of the Sharpe ratio relative to the EWA. The IR of -0.142 for MAFE minimization is a striking contrast to the statistical goodness measures. A closer inspection reveals that the optimized models are less successful at forecasting large VIX increases than the EWA and thus provide an inferior basis for conditional volatility trading.

The forecasting accuracy of the optimized models deteriorates sharply during the GFC subsample. In this subsample, all accuracy measures point at underperformance relative to EWA: The MZ R^2 drops below 2 percent, the DM statistics turn negative and the IRs show large negative values.

The optimized combinations strongly benefit from imposing the SCLSD constraints by every criterion. The MZ R^2 improves materially, reflecting improvements upon EWA beyond bias correction. The DM statistics further increase and indicate highly significant reductions in MAFE and MSFE. The IRs turn from negative values to positive values. The optimal combination weights show the moderating effect of the SCLSD constraints. The results are more robust during the GFC period; in this period, the optimal combination weights resemble those of EWA.

Whereas unconstrained optimization of forecast combinations struggles to improve upon simple averaging, robust optimization based on SCLSD constraints yields material improvements in statistical and economic goodness in this application.

[Insert Table 1 about here.]

References

- [1] Anderson, G., 1996, Nonparametric tests of stochastic dominance in income distributions, *Econometrica*, 64, 1183-1193.
- [2] Andrews, D. W. , 1999, Estimation When a Parameter is on a Boundary, *Econometrica* 67, 6, 1341-1383.
- [3] Andrews, D. W., and G. Soares, 2010, Inference for parameters defined by moment inequalities using generalized moment selection, *Econometrica*, 78(1), 119-157.
- [4] Back. K. and D.P. Brown, 1993, Implied Probabilities in GMM Estimators, *Econometrica* 61, 971-975.
- [5] Barrett, G. and S. Donald, 2003, Consistent tests for stochastic dominance, *Econometrica* 71, 71-104.
- [6] Bates, JM and CWJ Granger, 1969, The Combination of Forecasts, *Operations Research Quarterly* 20, 451-468.
- [7] Beer, Gerald, 1993, *Topologies on closed and closed convex sets*, Springer Science & Business Media, vol. 268.
- [8] Beer, G., Rockafellar, R. T., & Wets, R. J. B., 1992, A characterization of epi-convergence in terms of convergence of level sets. *Proceedings of the American Mathematical Society*, 116(3), 753-761
- [9] Capistran, C., and A. Timmermann, 2009, Forecast combination with entry and exit of experts, *Journal of Business & Economic Statistics* 27, 428-440.
- [10] Chan, F. and L.L. Pauwels, 2018, Some theoretical results on forecast combinations, *International Journal of Forecasting* 34, 64-74.
- [11] Chernozhukov, V., Hong, H., and E. Tamer, 2007, Estimation and confidence regions for parameter sets in econometric models, *Econometrica*, 75(5), 1243-1284.
- [12] Claeskens, G., J.R. Magnus, A.L. Vasnev and W. Wang, 2016, A simple theoretical explanation of the forecast combination puzzle, *International Journal of Forecasting* 32, 754-762.
- [13] Clark, T. E. and K.D. West, 2007, Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138, 291-311.
- [14] Clemen, R.T., 1989, Combining forecasts: A review and annotated bibliography, *International Journal of Forecasting* 5, 559-583.

- [15] Cortisoz, J., 2007, On the Skorokhod representation theorem. *Proceedings of the American Mathematical Society*, 135(12), 3995-4007.
- [16] Davidson, R. and J.-Y. Duclos, 2000, Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality, *Econometrica* 68, 1435-1464.
- [17] Diebold, F. and R. Mariano, 1995, Comparing predictive accuracy, *Journal of Business & Economic Statistics* 13, 253-265.
- [18] Harvey, C.R. and R.E. Whaley, 1992, Market volatility prediction and the efficiency of the S&P 100 index option market, *Journal of Financial Economics* 64, 43-73.
- [19] Harvey, D., S. Leybourne and P. Newbold, 1997, Testing the equality of prediction mean square errors, *International Journal of Forecasting* 13, 281-291.
- [20] Hsiao, C. and S. K. Wan, 2014, Is there an optimal forecast combination?, *Journal of Econometrics* 178, 294-309.
- [21] Jin, S., V. Corradi and N.R. Swanson, 2017, Robust Forecast Comparison, *Econometric Theory* 33, 1306-1351.
- [22] Kitamura, Yuichi, 1997, Empirical likelihood methods with weakly dependent processes, *Annals of Statistics* 25, 2084-2102.
- [23] Knight, K., 1999, Epi-convergence in distribution and stochastic equi-semicontinuity. Unpublished manuscript, 37, 28-72.
- [24] Lawrence, M. J., R. H. Edmundson and M.J. O'Connor (1986), "The accuracy of combining judgmental and statistical forecasts," *Management Science* 32, 1521-1532.
- [25] Liang, H., G. Zou, A. T. K. Wan, and X. Zhang, 2011, Optimal weight choice for frequentist model average estimators, *Journal of the American Statistical Association* 106, 1053-1066.
- [26] Linton, O.B., E. Maasoumi and Y.-J. Whang. 2005, Consistent Testing for Stochastic Dominance under General Sampling Schemes, *Review of Economic Studies* 72, 735-765.
- [27] Linton, O.B., Th. Post and Y.-J. Whang, 2014, Testing for the Stochastic Dominance Efficiency of a Given Portfolio, *Econometrics Journal* 17, 59-74.
- [28] Makridakis, S. and R.L. Winkler, 1983, "Averages of Forecasts: Some Empirical Results", *Management Science* 29, 987-996.
- [29] Molchanov, I., 2006, *Theory of random sets*. Springer Science & Business Media.

- [30] Newbold, P. and C.W.J. Granger, 1974, Experience with Forecasting Univariate Time Series and the Combination of Forecasts, *Journal of the Royal Statistical Society* A137, 131-165.
- [31] Poon, S-H. & C.W.J. Granger, 2003, Forecasting volatility in financial markets: a review, *Journal of Economic Literature* 41, 478-539.
- [32] Post, T., S. Karabati and S. Arvanitis, 2017, Portfolio Construction Based on Stochastic Dominance and Empirical Likelihood, forthcoming in *Journal of Econometrics*.
- [33] Rockafellar, R. T., & Wets, R. J. B., 2009, *Variational analysis* (Vol. 317). Springer Science & Business Media.
- [34] Qin, J. and J. Lawless, 1994, Empirical likelihood and general estimating equations, *Annals of Statistics* 22, 300-325.
- [35] Scaillet, O., and N. Topaloglou, 2010, Testing for stochastic dominance efficiency, *Journal of Business & Economic Statistics* 28, 169-180.
- [36] Schmid, F. and Trede, M., 1998, A Kolmogorov-type test for second order stochastic dominance, *Statist. Probab. Lett.* 37, 183-193.
- [37] Smith, J. and K. Wallis, 2009, A Simple Explanation of the Forecast Combination Puzzle, *Oxford Bulletin of Economics and Statistics* 71, 331-355.
- [38] Winkler, R. L., 1981, Combining probability distributions from dependent information sources, *Management Science* 27, 479-488.
- [39] van der Vaart, A.W., 2000, *Asymptotic Statistics*, Cambridge University Press.

Appendix

Proof of Proposition 3.3. Consider the event $\mathcal{A}_T := \left\{ \boldsymbol{\lambda} \notin \Lambda_{\mathcal{F}_T}^{\succ} (c_T), \forall \boldsymbol{\lambda} \notin \Lambda_{\mathcal{F}}^{\succ} \right\}$. Due to Lemma 3.2 and $\sup_{z \in \mathcal{Z}^-} c_T(z) \rightarrow 0$, it is found that $\mathbb{P}(\mathcal{A}_T) \rightarrow 1$, as $T \rightarrow \infty$. By construction $\mathcal{A}_T \subseteq \left\{ \boldsymbol{\lambda}_T \in \Lambda_{\mathcal{F}}^{\succ} \right\}$ and the result follows since by construction $\boldsymbol{\lambda}_T \in \Lambda_{\mathcal{F}_T}^{\succ} (c_T)$. ■

Proof of Theorem 3.4. In what follows $\mathcal{D}_{\boldsymbol{\lambda}}(\mathcal{F}, z) := \mathbb{E}_{\mathcal{F}}[L_z(\mathbf{U}\boldsymbol{\tau})] - \mathbb{E}_{\mathcal{F}}[L_z(\mathbf{U}\boldsymbol{\lambda})]$ is the achieved reduction of expected loss and $\mathcal{Z}_{\boldsymbol{\lambda}}^{\leftarrow} := \{z \in \mathcal{Z}^- : \mathcal{D}_{\boldsymbol{\lambda}}(z) = 0\}$ the ‘contact set’ for $\boldsymbol{\lambda}$. Assumption 3.1 and the CMT imply that $\mathcal{D}_{\boldsymbol{\lambda}}\left(\sqrt{T}(\mathcal{F}_T - \mathcal{F}), z\right)$ weakly converges to $\mathcal{G}(\boldsymbol{\tau}, z) - \mathcal{G}(\boldsymbol{\lambda}, z)$ w.r.t. to the product topology of continuous (w.r.t. $\boldsymbol{\lambda}$) epi-convergence (w.r.t. z) on the product of the relevant spaces of usc real valued functions (see, e.g., Knight (1999)). This product space is metrizable as complete and separable (see again Knight (1999)). Hence, via Skorokhod representations which are applicable due to Theorem 1 in Cortissoz (2007) and by using the same notation, for any $\boldsymbol{\lambda}$ and any sequence $\boldsymbol{\lambda}_T \rightarrow \boldsymbol{\lambda}$, we

have that $\mathcal{D}_{\lambda_T} \left(\sqrt{T} (\mathcal{F}_T - \mathcal{F}), z \right)$ converges to $\mathcal{G}(\tau, z) - \mathcal{G}(\lambda, z)$ almost surely, w.r.t. to the topology of epi-convergence. Notice that,

$$\mathcal{D}_{\lambda} \left(\sqrt{T} \mathcal{F}_T, z \right) = \mathcal{D}_{\lambda} \left(\sqrt{T} (\mathcal{F}_T - \mathcal{F}), z \right) + \sqrt{T} \mathcal{D}_{\lambda} (\mathcal{F}, z),$$

and since $\mathcal{D}_{\lambda} (\mathcal{F}, \cdot)$ is locally Lipschitz locally uniformly w.r.t. λ (more) imply that almost surely, $\mathcal{D}_{\lambda} \left(\sqrt{T} \mathcal{F}_T, z \right)$ converges w.r.t. to the product topology of continuous (w.r.t. λ) hypo-convergence (w.r.t. z) to

$$\Gamma(\lambda, z) := \begin{cases} \mathcal{G}(\tau, z) - \mathcal{G}(\lambda, z), & \lambda \in \Lambda_{\mathcal{F}}^{\succ}, z \in \mathcal{Z}_{\lambda}^{\leftarrow} \\ +\infty, & \lambda \in \Lambda_{\mathcal{F}}^{\succ}, z \notin \mathcal{Z}_{\lambda}^{\leftarrow} \\ -\infty, & \lambda \notin \Lambda_{\mathcal{F}}^{\succ} \end{cases}.$$

Then, Proposition 3.2 of Molchanov (2006, Chapter 5, p. 337) and the relationship between epi-convergence and continuous convergence (see, e.g., Rockafellar and Wets (2009)) imply that $\inf_{z \in \mathcal{Z}^-} \mathcal{D}_{\lambda} \left(\sqrt{T} \mathcal{F}_T, z \right)$ hypoconverges almost surely to $\inf_{z \in \mathcal{Z}^-} \Gamma(\lambda, z)$. Notice that

$$\Lambda_{\mathcal{F}_T}^{\succ}(c_T) \subseteq \left\{ \lambda \in \Lambda : \inf_{z \in \mathcal{Z}^-} \mathcal{D}_{\lambda} \left(\sqrt{T} \mathcal{F}_T, z \right) \geq -\sqrt{T} \sup_{z \in \mathcal{Z}^-} c_T \right\}$$

and analogously

$$\Lambda_{\mathcal{F}_T}^{\succ} = \left\{ \lambda \in \Lambda : \inf_{z \in \mathcal{Z}^-} \Gamma(\lambda, z) > -\infty \right\}, \text{ almost surely.}$$

Using this and (the dual part of) Theorem 3.1 of Beer, Rockafellar and Wets (1992), the result in part (a) follows, because, for any $\lambda \in \Lambda_{\mathcal{F}}^{\succ}$ for which $\mathcal{Z}_{\lambda}^{\leftarrow} \neq \emptyset$, due to the Portmanteau Lemma and the properties of c_T ,

$$\mathbb{P} \left(\inf_{z \in \mathcal{Z}^-} \mathcal{D}_{\lambda} \left(\sqrt{T} \mathcal{F}_T, z \right) \geq -\sqrt{T} \sup_{z \in \mathcal{Z}^-} c_T(z) \right) \rightarrow \mathbb{P} \left(\inf_{z \in \mathcal{Z}_{\lambda}^{\leftarrow}} (\mathcal{G}(\tau, z) - \mathcal{G}(\lambda, z)) \geq -\infty \right)$$

which equals to one, while if $\lambda_T \rightsquigarrow \lambda \notin \Lambda_{\mathcal{F}}^{\succ}$ (or the latter is a cluster point), since $\Lambda_{\mathcal{F}}^{\succ}$ is open and by construction $\inf_{z \in \mathcal{Z}^-} \mathcal{D}_{\lambda_T} \left(\sqrt{T} \mathcal{F}, z \right)$ almost surely diverges to $-\infty$ faster than $-\sqrt{T} \sup_{z \in \mathcal{Z}^-} c_T(z)$, we have that any such λ cannot lie inside the limiting set. For part (b), notice that, since Λ is compact and separable, the Painleve-Kuratowski convergence on the hyperspace of its closed non-empty subsets is metrizable (see Molchanov (2006, Appendix B)) and the relevant metric space is complete and separable. Lemma 3.2, implies as previously the epi-convergence in probability of $\sum_{t=1}^T p_t L(X_t - Y_t \lambda)$ to $\mathbb{E}(L(X_t - Y_t \lambda))$ as functions defined on Λ . Then again Theorem 1 in Cortisoz (2007) implies joint Skorokhod representations for both the sets involved in the convergence in part (a) and the functions involved in the previous one. This and the compactness of

Λ enables again the use of Proposition 3.2 of [Molchanov \(2006, Ch. 5, p. 337\)](#) and then part (b) follows from Lemma 3.2, and part (c) follows from the fact that $\mathbb{E}(L(X_t - Y_t \lambda))$ has a unique minimizer and Theorem 7.31 of [Rockafellar and Wets \(2009\)](#). ■

Proof of Theorem 3.6. Since $\lambda_T \rightsquigarrow \lambda_{\mathcal{F}}^*$ then $\lambda_T \in \bar{B}_{\lambda_{\mathcal{F}}^*}$ with probability tending to one. Due to Lemma 3.2, the CMT, the strict convexity of L , and the Portmanteau Theorem we have that for any $\varepsilon > 0$,

$$\lim_{T \rightarrow \infty} \mathbb{P} \left(\min_{\lambda \in \bar{B}_{\lambda_{\mathcal{F}}^*}(\eta) \cap \Lambda} \mathbb{E}_{\mathcal{F}_T} [L(U\lambda)] < \min_{\lambda \in \Lambda_{\lambda_{\mathcal{F}}^*}^{\neq}(\varepsilon)} \mathbb{E}_{\mathcal{F}_T} [L(U\lambda)] \right) = 1,$$

where $\Lambda_{\lambda_{\mathcal{F}}^*}^{\neq}(\varepsilon) = \left\{ \lambda \in \bar{B}_{\lambda_{\mathcal{F}}^*}(\eta) \cap \Lambda_{\lambda_{\mathcal{F}}^*}^{\neq} : |\lambda - \lambda_{\mathcal{F}}^*| > \varepsilon \right\}$ and $\Lambda_{\lambda_{\mathcal{F}}^*}^{\neq}$ is the complement of $\Lambda_{\lambda_{\mathcal{F}}^*}^{\leq}$ in Λ . Hence, λ_T belongs to $\Lambda_{\lambda_{\mathcal{F}}^*}^{\leq}$ with probability converging to one, and thereby with the same probability it asymptotically solves

$$\begin{aligned} & \max \mathbb{E}_{\mathcal{F}_T} [L(U\tau)] - \mathbb{E}_{\mathcal{F}_T} [L(U\lambda)] \\ & \text{s.t. } \mathbb{E}_{\mathcal{F}} [L_z(U\lambda)] \leq \mathbb{E}_{\mathcal{F}} [L_z(U\tau)] \quad \forall z \in \mathcal{Z}^-. \\ & \lambda \in \bar{B}_{\lambda_{\mathcal{F}}^*}(\eta) \end{aligned} \tag{21}$$

This and Assumption 3.5 imply that $\mathbf{u}_T := \sqrt{T}(\lambda_T - \lambda_{\mathcal{F}}^*)$ is the minimizer of

$$T \left(\mathbb{E}_{\mathcal{F}_T} \left[L \left(U \left(\lambda_{\mathcal{F}}^* + \frac{1}{\sqrt{T}} \mathbf{u} \right) \right) \right] - \mathbb{E}_{\mathcal{F}_T} [L(U\tau)] - \mathbb{E}_{\mathcal{F}_T} [L(U\lambda_{\mathcal{F}}^*)] + \mathbb{E}_{\mathcal{F}_T} [L(U\tau)] \right)$$

over $\sqrt{T}(\Lambda_{\lambda_{\mathcal{F}}^*}^{\leq} \cap \bar{B}_{\lambda_{\mathcal{F}}^*}(\eta) - \lambda_{\mathcal{F}}^*)$. Using this and Assumptions 3.5.1-2 we have that for $\mu_{\mathcal{F}}^*$ necessarily converging to $\lambda_{\mathcal{F}}^*$ with probability one,

$$\sqrt{T} \mathbb{E}_{\mathcal{F}_T} [s_t(\lambda_{\mathcal{F}}^*)]^T \mathbf{u}_T + \frac{1}{2} \mathbf{u}_T^T \mathbb{E}_{\mathcal{F}_T} [H_t(\mu_{\mathcal{F}}^*)] \mathbf{u}_T \leq O_p(1),$$

hence

$$\sqrt{T} \mathbb{E}_{\mathcal{F}_T} [s_t(\lambda_{\mathcal{F}}^*)]^T \mathbf{u}_T + \frac{1}{2} \mathbf{u}_T^T (H_{\lambda_{\mathcal{F}}^*} + o_p(1)) \mathbf{u}_T \leq O_p(1),$$

which then due to Assumption 3.5.2 implies that

$$(\|\mathbf{u}_T\|^2 + 2\|\mathbf{u}_T\|) O_p(1) (1 + o_p(1)) + O_p(1) \leq O_p(1),$$

and thereby $\|\mathbf{u}_T\| (1 + o_p(1)) \leq O_p(1)$, which establishes uniform tightness for $\mathbf{u}_T$. Notice then that due to [Beer \(1993, Exercise 10 in Exercise Set 5.2\)](#), $\sqrt{T}(\Lambda_{\lambda_{\mathcal{F}}^*}^{\leq} \cap \bar{B}_{\lambda_{\mathcal{F}}^*}(\eta) - \lambda_{\mathcal{F}}^*)$ converges in the Painlevé-Kuratowski sense to the

non-empty, closed and convex $\Lambda_{\infty}^{\succeq}$, while due to Assumption 3.5.2, the sup distance between $\sqrt{T}\mathbb{E}_{\mathcal{F}_T} [s_t(\boldsymbol{\lambda}_{\mathcal{F}}^*)]^T \mathbf{u} + \frac{1}{2}\mathbf{u}^T \mathbb{E}_{\mathcal{F}_T} [H_t(\boldsymbol{\lambda}_{\mathcal{F}}^*)] \mathbf{u}$ and $\sqrt{T}\mathbb{E}_{\mathcal{F}_T} [s_t(\boldsymbol{\lambda}_{\mathcal{F}}^*)]^T \mathbf{u} + \frac{1}{2}\mathbf{u}^T \mathbb{E}_{\mathcal{F}_T} [H_t(\mu_{\mathcal{F}}^*)] \mathbf{u}$ is $o_p(1)$, locally uniformly on $\sqrt{T}(\Lambda_{\mathcal{F}}^{\succeq} \cap \bar{B}_{\boldsymbol{\lambda}_{\mathcal{F}}^*}(\eta) - \boldsymbol{\lambda}_{\mathcal{F}}^*)$. Furthermore, as noted above, if we let generally $\mathbf{u} \in \mathbb{R}^M$, the function $\sqrt{T}\mathbb{E}_{\mathcal{F}_T} [s_t(\boldsymbol{\lambda}_{\mathcal{F}}^*)]^T \mathbf{u} + \frac{1}{2}\mathbf{u}^T \mathbb{E}_{\mathcal{F}_T} [H_t(\boldsymbol{\lambda}_{\mathcal{F}}^*)] \mathbf{u}$ weakly converges to the quadratic function $\frac{1}{2}(\mathbf{u} + H_{\boldsymbol{\lambda}_{\mathcal{F}}^*}^{-1} s_{\boldsymbol{\lambda}_{\mathcal{F}}^*})^T H_{\boldsymbol{\lambda}_{\mathcal{F}}^*} (\mathbf{u} + H_{\boldsymbol{\lambda}_{\mathcal{F}}^*}^{-1} s_{\boldsymbol{\lambda}_{\mathcal{F}}^*})$. Hence Corollary 5.58 of [van der Vaart \(2000\)](#) is applicable, which establishes the sought result. ■

Table 1: VIX forecasting: Shown are the intercept, slope and R-squared of the Mincer-Zarnowitz regression of actual on forecasted values; the Diebold-Mariano statistic for MAFE of the optimized combination being equal to that of the Equal-Weighted Average (DM1); the Diebold-Mariano statistic for MSFE (DM2); the fraction of correct sign predictions; the (annualized) Information Ratio (IR) for a simple strategy of long (short) volatility when an increase (decrease) is predicted. Separate results are shown for the full sample (1994-2017), the Global Financial Crisis period (2007-2008) and the sample excluding the crisis period.

Period		Accuracy Measures					Average weights								
		c_0	c_1	R^2	DM1	DM2	Sign	IR	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7
94-17	EWA	0.000	2.163	2.704			0.554		0.143	0.143	0.143	0.143	0.143	0.143	0.143
	MAFE	0.000	1.250	2.747	3.554	2.665	0.558	-0.142	0.024	0.040	0.122	0.073	0.064	0.104	0.573
	MAFE SD	0.000	1.520	3.001	4.920	3.528	0.562	0.027	0.045	0.075	0.177	0.095	0.084	0.126	0.399
	MSFE	0.000	1.158	2.615	2.864	2.357	0.557	-0.009	0.036	0.083	0.165	0.121	0.026	0.082	0.487
	MSFE SD	0.000	1.353	2.742	3.700	2.698	0.560	0.093	0.056	0.091	0.186	0.114	0.061	0.110	0.381
GFC	EWA	0.003	1.961	3.079			0.554		0.143	0.143	0.143	0.143	0.143	0.143	0.143
	MAFE	0.002	1.191	1.996	-0.709	-0.483	0.536	-0.613	0.116	0.214	0.160	0.009	0.050	0.013	0.438
	MAFE SD	0.003	1.511	2.846	-0.254	0.386	0.550	-0.382	0.111	0.266	0.168	0.052	0.093	0.119	0.190
	MSFE	0.002	0.935	1.967	-1.536	-0.388	0.552	-0.792	0.175	0.501	0.163	0.033	0.037	0.001	0.090
	MSFE SD	0.003	1.118	2.213	-1.169	-0.176	0.546	-0.282	0.176	0.344	0.114	0.058	0.095	0.108	0.105
94-17 ex GFC	EWA	0.000	2.209	2.657			0.554		0.143	0.143	0.143	0.143	0.143	0.143	0.143
	MAFE	0.000	1.255	2.853	3.936	3.003	0.560	-0.086	0.016	0.024	0.118	0.079	0.065	0.112	0.586
	MAFE SD	0.000	1.521	3.023	5.158	3.593	0.563	0.072	0.038	0.057	0.178	0.099	0.083	0.127	0.418
	MSFE	0.000	1.197	2.724	3.577	2.744	0.557	0.016	0.023	0.044	0.165	0.129	0.025	0.090	0.524
	MSFE SD	0.000	1.395	2.838	4.348	3.087	0.561	0.185	0.046	0.068	0.193	0.119	0.058	0.111	0.406