Generic Conditions for Forecast Dominance – Online Appendix

Fabian Krüger* Johanna F. Ziegel†

December 13, 2019

This document is structured as follows: Section A contains further analytical examples related to the Gaussian setup in Section 4 of the paper. Section B provides details on the hypothesis tests used in Section 6 of the paper. Section C contains a result on forecast dominance for quantiles (by contrast, the paper treats expectiles, and the mean in particular).

A Further examples for the Gaussian setup

Example A.1. Consider a setup in which the forecast X_{tj} and the realization Y_t form a bivariate time series process that is observed at time $t = 1, \ldots, T$, with the understanding that X_{tj} is the forecast of Y_t given some information set. Suppose that the joint process for X_{tj} and Y_t is described by the following bivariate auto-regression with Gaussian innovations:

$$\begin{pmatrix} X_{tj} \\ Y_t \end{pmatrix} = \begin{pmatrix} 0 & a_j \\ 0 & a_Y \end{pmatrix} \begin{pmatrix} X_{t-1,j} \\ Y_{t-1} \end{pmatrix} + \varepsilon_t, \tag{1}$$

where

$$\varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_j^2 & \tau_{Yj} \\ \tau_{Yj} & \tau_Y^2 \end{pmatrix} \right).$$

The example implies that given Y_{t-1}, both X_{tj} and Y_t are independent of $X_{t-1,j}$. This restriction can be relaxed but is assumed for simplicity. Furthermore, the process in (1) is strictly stationary if $|a_Y| < 1$, which we assume here. The unconditional joint distribution of X_{tj} and Y_t is Gaussian with mean zero and covariance matrix given by

$$\begin{pmatrix} \tau_j^2 + \tau_Y^2 a_j^2/(1-a_Y^2) & \tau_{Yj} + \tau_Y^2 a_j a_Y/(1-a_Y^2) \\ \tau_{Yj} + \tau_Y^2 a_j a_Y/(1-a_Y^2) & \tau_Y^2/(1-a_Y^2) \end{pmatrix}.$$

Hence the present time series example matches the setup of Equation (5) in the paper. Example 2.1 of Ehm and Krüger (2018) is obtained as a special case if $a_A = a_B = a_Y$ and

*Karlsruhe Institute of Technology, Department of Economics, Blücherstraße 17, 76185 Karlsruhe, Germany. Email: fabian.krueger@kit.edu.

†University of Bern, Institute of Mathematical Statistics and Actuarial Science, Alpeneggstrasse 22, 3012 Bern, Switzerland. Email: johanna.ziegel@stat.unibe.ch.

$\tau_{Yj} = \tau_j^2, j \in \{A, B\}$, such that both forecasts are auto-calibrated. In the latter situation, the forecast j for which τ_j is greater dominates its competitor. To obtain a simple example without auto-calibration, let $0 < a_Y < 1$, and assume that forecast A neglects any time series dependence in Y_t, such that $a_A = 0$ and $\tau_{YA} > 0, \tau_A^2 > 0$. By contrast, assume that forecast B sets $X_{tB} = Y_{t-1}$, corresponding to an erroneous random walk assumption, with $a_B = 1, \tau_{YB} = \tau_B^2 = 0$. If it holds that $a_Y \tau_Y^2/(1 - a_Y^2) \leq \tau_{YA} \leq \tau_A^2 \leq \tau_Y^2/(1 - a_Y^2)$, then the conditions of Case 2a are satisfied, and A dominates B.

Example A.2. Consider the following parameter setup: $\beta_A = 1.2, \beta_B = 0.9, \rho_{YA} = \rho_{YB} = 0.5, \sigma_Y = 1, \mu_Y = 0$. In this setup, the necessary condition for dominance is satisfied (for either dominance of A over B, or vice versa), but neither of the four cases applies. Figure 1 below plots the difference in expected scores against the auxiliary parameter θ from Proposition 4.1 in the paper. The figure shows that the difference in expected scores switches signs in this example, i.e., a dominance relation does not exist.

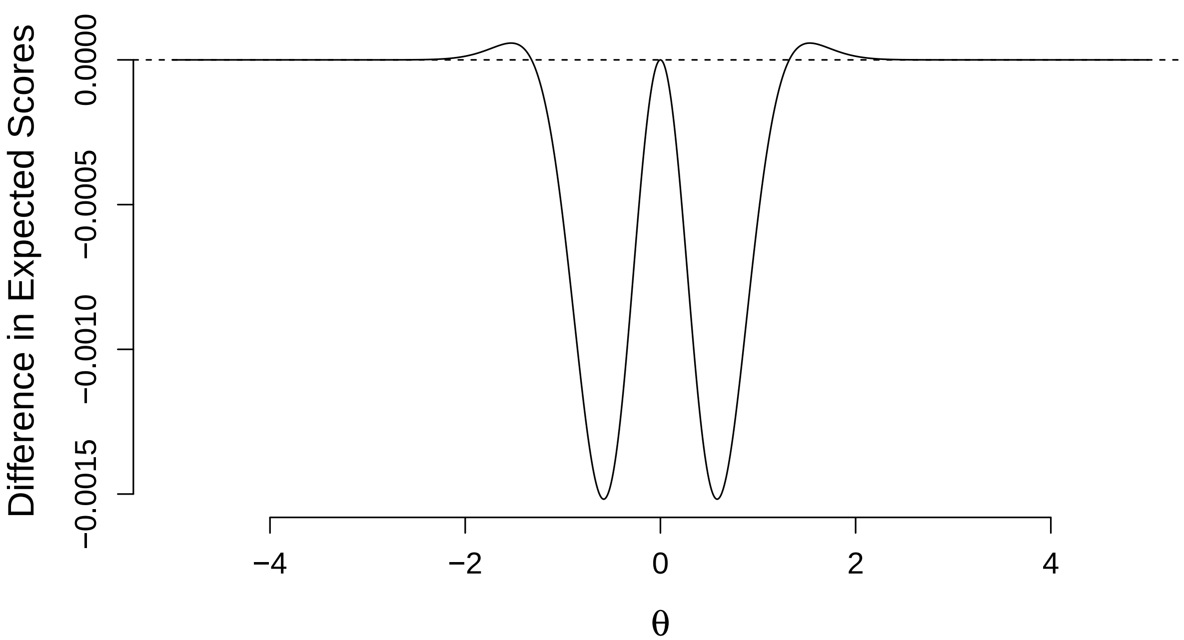

Figure 1: The figure plots the difference in expected scores for methods A and B against the auxiliary parameter θ. The relevant expression for the expected scores is provided in Proposition 4.1 of the paper.

B Details on Hypothesis Testing Methods

B.1 Testing for Second-Order Stochastic Dominance

This section explains the subsampling-based test for second order stochastic dominance (SOSD) by Linton et al. (2005, henceforth LMW). The concept to be tested is defined in Definition 2 on p. 738 of LMW.

Test statistic

We are interested in the integral of the cumulative distribution function (CDF) F up to a point z:

$$D(z) = \int_{-\infty}^{z} F(u)du, \tag{2}$$

c.f. LMW's Equations (3) and (4), where $s = 2$ in our case (the notation we use in the following is generally different from LMW). Replacing F by its empirical estimate, we get

$$\begin{aligned}
\hat{D}(z) &= \int_{-\infty}^{z} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_i \leq u)du, \\
&= \frac{1}{n} \sum_{i=1}^{n} \underbrace{\int_{-\infty}^{z} \mathbf{1}(X_i \leq u)du,}_{\mathbf{1}(X_i \leq z)(z - X_i)} \\
&= \frac{1}{n} \sum_{i: X_i \leq z} (z - X_i).
\end{aligned}$$

While LMW mainly consider symmetric tests for SOSD, we are interested in the asymmetric hypothesis that A is smaller than B according to SOSD, that is,

$$\text{Hypothesis: } D_A(z) \leq D_B(z) \ \forall \ z \in \mathcal{Z}, \text{ with strict inequality for some } z,$$

where $D_j(z)$ is as defined at (2), but referring to distribution $j \in \{A, B\}$, and $\mathcal{Z}$ is the support of interest. As noted by LMW on p. 740 (end of first paragraph, 'It is also sometimes of interest...'), their test can be used for this case as well. Furthermore, we are interested in the case of only two random variables to be compared. Simplifying Equation (5) of LMW, the test quantity of interest is hence

$$\sup_{z \in \mathcal{Z}} [D_B(z) - D_A(z)].$$

Its empirical analogue is given by

$$\max_{z \in G} \ [\hat{D}_B(z) - \hat{D}_A(z)], \tag{3}$$

where G is a set of grid point representing $\mathcal{Z}$ (see last sentence of Section 3 in LMW), and the estimates $\hat{D}_j(z)$ have been discussed above.

Subsampling

LMW propose to use sub-sampling in order to obtain critical values for the test statistic. The idea of subsampling is to compute the test statistic on rolling subsamples of size $b < n$, where n is the sample size. Subsampling is attractive here because there is no need to impose the hypothesis being tested (which is required in the bootstrap scheme, and would be difficult to do in the case of SOSD). Following p. 744 of LMW, let $t_{n,b,i}$ denote the test statistic at (3), computed based on the b observations with indices $i, i+1, \ldots, i+b-1$. These observations are an adjacent block in the original time series of data. The subsampling p-value of the SOSD hypothesis is then given by

$$\frac{1}{n-b+1} \sum_{i=1}^{n-b+1} \mathbf{1}(\sqrt{b}\; t_{n,b,i} > \sqrt{n}\; T_n),$$

where T_n is the test statistic computed from the original sample. Note that T_n is scaled by the full sample size $\sqrt{n}$, whereas the subsample test statistics $t_{n,b,i}$ are scaled by $\sqrt{b}$, the size of each subsample.

In order to choose b in practice, LMW advice to plot the subsampling p-value against b, and choose b from a segment over which p is stable (see last paragraph of Section 5.2). We follow this advice in our Section 6.1 (see especially our Figure 1).

Implementation in R

The R function `lmw_plot` in our replication package `gcfdtools` implements the test described above and creates a plot of the test's p-value against the subsample size. Further information is available via the function documentation (`?lmw_plot`).

B.2 Testing Forecast Dominance

We adopt the test by Ziegel et al. (2018) in order to test for forecast dominance. While Ziegel et al. consider a family of scoring functions that is relevant in the context of forecasting Expected Shortfall (see their Proposition 2.1), we apply their test to the class of Bregman scoring functions defined in our Equation (1), with associated elementary scores defined in our Equation (3). We refer to Ziegel et al. (2018, Section 3.2) for a description of the test's implementation, noting that their symbol η (threshold parameter of the elementary score) corresponds to θ in our notation. Following their implementation, we use a geometric distribution with parameter $1.36\; n^{-1/3}$ in the stationary bootstrap scheme (Politis and Romano, 1994) on which the test is based. Furthermore, we use an equally spaced grid of 100 points for θ, where the lower and upper end of the grid are determined by the empirical minimum and maximum of the set $\{x_{tA}, x_{tB}\}_{t=1}^{n}$ consisting of all empirical forecasts. This choice is computationally simple and performs well in the simulation study by Ziegel et al. (2018). We use $10,000$ bootstrap replications.

R Implementation

The R function `dom_test` in our replication package `gcfdtools` implements the test. Please see the documentation at `?dom_test` for details.

C Forecast Dominance for Quantiles

In this section we present a result akin to Theorem 3.1 in the paper, but referring to quantile forecasts as opposed to mean forecasts. Quantile forecasts play a major role in applications, notably in financial risk management where quantiles are referred to as 'Value-at-Risk'. Let $\tau \in (0, 1)$ denote the level of the quantile. To simplify the presentation, we assume that the τ-quantile of Y and the conditional τ-quantiles of Y we consider are all unique and that the distribution and the conditional distributions of Y are continuous at their τ-quantiles. Consistent scoring functions for quantiles at level τ are of the form

$$S(x, y) = \{\mathbf{1}_{(y<x)} - \tau\}\{g(x) - g(y)\}, \tag{4}$$

where $\mathbf{1}_{(E)}$ is the indicator function of the event E, and g is an increasing function. Analogous to the case of the mean (Definition 2.1 in the paper), one forecast dominates another if it attains a lower expected score for all functions covered by Equation (4).

Definition C.1. Auto-calibrated quantile forecast Forecast X is auto-calibrated for Y if $\mathbb{E}\left(\mathbf{1}_{(Y<X)}\middle|X\right) = \mathbb{P}(Y < X|X) = \tau$ almost surely.

This notion of auto-calibration is analogous to the one for mean forecasts (Definition 3.1 in the paper), and corresponds to the null hypothesis of a Mincer-Zarnowitz type regression for quantile forecasts as proposed by Guler et al. (2017); see also Nolde and Ziegel (2017, Definition 3). Note that auto-calibration of a quantile forecast is equivalent to independence of the random variables X and $Z = \mathbf{1}_{(X>Y)}$, where $\mathbb{P}(Z = 1) = \tau$.

Theorem C.1. *Suppose X_A and X_B are both auto-calibrated quantile forecasts. Then A dominates B if and only if one of the following equivalent conditions hold*

1. $$\mathbb{P}(X_A > Y > \theta) \geq \mathbb{P}(X_B > Y > \theta) \tag{5}$$

   for every $\theta \in \mathbb{R}$.

2. *The distribution $\mathcal{L}(Y|X_A > Y)$ stochastically dominates $\mathcal{L}(Y|X_B > Y)$.*

Proof. From Ehm et al. (2016, Corollary 1a), A dominates B if and only if

$$\mathbb{E}\left(S_\theta(X_B, Y)\right) \geq \mathbb{E}\left(S_\theta(X_A, Y)\right), \quad \text{for all } \theta \in \mathbb{R},$$

where

$$S_\theta(x, y) = \left(\mathbf{1}_{(y<x)} - \tau\right)\left(\mathbf{1}_{(\theta<x)} - \mathbf{1}_{(\theta<y)}\right)$$

is the elementary scoring function for quantiles. Using the auto-calibration assumption and the law of iterated expectations yields that

$$\mathbb{E}\left(\left(\mathbf{1}_{(Y<X_j)} - \tau\right)\mathbf{1}_{(\theta<X_j)}\right) = 0$$

for $j \in \{A, B\}$. Hence

$$\begin{aligned}\mathbb{E}\left(S_\theta(X_B, Y)\right) - \mathbb{E}\left(S_\theta(X_A, Y)\right) &= \mathbb{E}\left(\mathbf{1}_{(Y<X_A)}\ \mathbf{1}_{(\theta<Y)}\right) - \mathbb{E}\left(\mathbf{1}_{(Y<X_B)}\ \mathbf{1}_{(\theta<Y)}\right) \\ &= \mathbb{P}(X_A > Y > \theta) - \mathbb{P}(X_B > Y > \theta),\end{aligned}$$

leading to the first statement. The equivalence of the second statement follows because auto-calibration of the forecasts implies that $\mathbb{P}(X_A > Y) = \mathbb{P}(X_B > Y) = \tau$. □

Theorem C.1 characterizes dominance relations among auto-calibrated quantile forecasts. Interestingly, the condition in Equation (5) does not depend on the quantile level τ of interest. Furthermore, the conditions involves the joint distribution of the forecast X_j and the predictand Y. This situation is different from the corresponding condition for mean forecasts in Theorem 3.1, which involves the convex ordering of the forecast distributions but does not depend on the predictand. To understand the interpretation of Equation (5), it is useful to consider the following simple example.

Example C.1. Let $Y = Z + \varepsilon$, where Z, ε are independent and standard normal. The two forecasts are given by $X_A = Z + \Phi^{-1}(\tau)$ and $X_B = \sqrt{2}\ \Phi^{-1}(\tau)$, where Φ is the CDF of the standard normal distribution. Both forecasts are correctly specified given their information sets, whereby the latter is empty in case of X_B. Hence Holzmann and Eulert (2014, Corollary 2) implies that X_A must dominate X_B, which means that condition (5) must be satisfied. Note that

$$\mathbb{P}(X_B > Y > \theta) = \mathbb{P}\left(\sqrt{2}\ \Phi^{-1}(\tau) > Y > \theta\right) = \begin{cases}\tau - \Phi(\frac{\theta}{\sqrt{2}}) & \theta < \sqrt{2}\ \Phi^{-1}(\tau), \\ 0 & \text{else.}\end{cases}$$

Hence for dominance to hold, $\mathbb{P}(X_A > Y > \theta) > 0$ must hold even for some 'large' values $\theta > \sqrt{2}\ \Phi^{-1}(\tau)$. (While the generic condition in Equation (5) states a weak inequality, the probability will be strictly positive for some θ since both X_A and Y are continuously distributed.) Heuristically, due to the variability of X_A there is a nonzero chance that X_A exceeds Y even in cases where Y is large. This situation is in contrast to X_B which does not vary at all.

As illustrated in the example, the condition in Equation (5) requires X_A to be more variable than X_B in a certain sense. Similar to the case of the mean, the auto-calibration assumption rules out artificial (uninformative) variation in the forecasts X_j. Furthermore, note that Equation (5) is satisfied with equality for $\theta \to -\infty$, in which case $\mathbb{P}(X_j > Y > \theta) \to \mathbb{P}(X_j > Y) = \tau$, where the last equality follows from auto-calibration.

Equation (5) is useful in that it yields a better understanding of the conditions under which dominance occurs. Furthermore, the condition can easily be checked for empirical data. However, in contrast to our results for the mean in the paper, the conditions for quantiles are not easily verified in analytical examples where conditional probabilities under inequality constraints are rarely available in closed form.

References

Ehm, W., Gneiting, T., Jordan, A., and Krüger, F. (2016). Of quantiles and expectiles: Consistent scoring functions, Choquet representations and forecast rankings (with discussion and rejoinder). *Journal of the Royal Statistical Society: Series B*, 78:505–562.

Ehm, W. and Krüger, F. (2018). Forecast dominance testing via sign randomization. *Electronic Journal of Statistics*, 12:3758–3793.

Guler, K., Ng, P. T., and Xiao, Z. (2017). Mincer-Zarnowitz quantile and expectile regressions for forecast evaluations under aysmmetric loss functions. *Journal of Forecasting*, 36:651–679.

Holzmann, H. and Eulert, M. (2014). The role of the information set for forecasting–with applications to risk management. *Annals of Applied Statistics*, 8:595–621.

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.

Nolde, N. and Ziegel, J. F. (2017). Elicitability and backtesting: Perspectives for banking regulation (with discussion and rejoinder). *Annals of Applied Statistics*, 11:1833–1874.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89:1303–1313.

Ziegel, J. F., Krüger, F., Jordan, A., and Fasciati, F. (2018). Robust forecast evaluation of Expected Shortfall. *Journal of Financial Econometrics*. Forthcoming.