Nonparametric monitoring of equal predictive ability

Vasyl Golosnoy [a],[*], Yarema Okhrin [b]

[a] *Institute of Statistics and Econometrics, University of Kiel, Olshausenstr. 40, 24118 Kiel, Germany*
[b] *Department of Statistics, University of Augsburg, Universitätsstr. 16, 86159 Augsburg, Germany*

1. Introduction

Testing for equal predictive ability (EPA) is of great importance for evaluating the performance of competing forecasting models. The existing tests (cf. Diebold and Mariano, 1995; West, 1996; Giacomini and White, 2006) are statistical tools which provide decisions as to whether the EPA hypothesis holds, based on a fixed sample of historical observations. However, it is also of practical interest to decide about EPA validity on the current edge by using innovations from the considered forecasting models. This paper introduces a novel sequential approach for EPA online monitoring, which relies on two nonparametric control procedures. We suggest statistical tools suitable for early signalling about violations of the EPA hypothesis given a large number of alternative forecasting rules or forecasters. A signal from our monitoring scheme indicates that the initially assumed EPA hypothesis may not hold any more.

In order to formalize the sequential monitoring problem, consider K competing forecasting rules or forecasters initially assumed to have the same predictive ability. Let $\mathbf{d}_t = (d_{1,t}, \ldots, d_{K,t})'$ be the K-dimensional orthogonal vector of losses from these forecasting rules. The losses are assumed to be independent in time and distributed as $\mathbf{d}_t \sim \mathcal{F}_K(\boldsymbol{\mu}_t, \mathbf{I})$, where the K-variate distribution function $\mathcal{F}_K$ has continuous support. The corresponding moments are the expectation $E(\mathbf{d}_t) = \boldsymbol{\mu}_t$ and the identity covariance matrix $Cov(\mathbf{d}_t) = \mathbf{I}$. Moreover, the function $\mathcal{F}(\cdot)$ is assumed to have finite and positive Fisher information and absolutely continuous derivatives (Hajek et al., 1999, Section 6.1.7). These assumptions would guarantee the applicability of the central limit theorem to the rank-based statistics suggested below. Such mild requirements cover a rather broad family of distribution functions. The EPA hypothesis implies the same expected loss $\mu_{k,t} = \mu_t$ for all K models, i.e., $\boldsymbol{\mu}_t = \mu_t \mathbf{1}$. Assuming $\mu_t = 0$ without loss of generality, the distribution of losses under the EPA is

$$\mathbf{d}_t \sim \mathcal{F}_K(\mathbf{0}, \mathbf{I}). \tag{1}$$

---

* Corresponding author.
   E-mail addresses: vgolosnoy@stat-econ.uni-kiel.de (V. Golosnoy), yarema.okhrin@wiwi.uni-augsburg.de (Y. Okhrin).

Then, under the validity of EPA at $t=0$, a sequential monitoring procedure should provide a decision at every $t \geq 1$, $t \in \mathbb{N}$ as to whether the EPA hypothesis $H_{0,t}$ still holds:

$$H_{0,t} : E(\mathbf{d}_t) = \mathbf{0}, \quad H_{a,t} : E(\mathbf{d}_t) \neq \mathbf{0}. \tag{2}$$

The EPA hypothesis remains valid until an unknown change point $\tau \geq 1$, $\tau \in \mathbb{N}$. The no-change case is characterized by $\tau = \infty$. The ultimate goal is to get a signal, which is interpreted as a failure of the EPA hypothesis, as soon as a change occurs. Additionally, it is desired to have no (false) signals for a long time if no change has really happened, i.e., for all $t < \tau$.

Nonparametric sequential procedures for differentiating between the null and alternative hypothesis in (2) are required in situations where the distribution function $\mathcal{F}(\cdot)$ remains unknown. Such methods should be favoured, because they are robust with respect to misspecifications of the distribution functions. Moreover, nonparametric approaches, contrary to parametric ones, benefit from a larger number K of alternative forecasting models or competing forecasters. The stated problem arises in the application, where many models or many experts make point forecasts of some publicly available measure. The possible examples are forecasts of exchange rates, stock market indices, basic macroeconomic indicators etc. Moreover, it is presumed that the forecasted variable is observable. Thus we do not confront measurement error problems.

This paper introduces nonparametric sequential tools suitable for online monitoring of the EPA validity for a large number K of alternative forecasting rules. Since the distribution of the innovations $d_{k,t}$ is the same for all k under the EPA, their realizations can be seen as K realizations of the i.i.d. univariate random variable δ_t. The corresponding EPA distribution of δ_t is $\delta_t \sim \mathcal{F}_1(0,1)$. It is symmetric around zero with the moments $E(\delta_t) = 0$ and $V(\delta_t) = 1$. Relying on these properties of the δ_t distribution, we elaborate two nonparametric sequential procedures (control charts) for monitoring the EPA validity.

First, we suggest a chart using the ordinary sign test statistic for detecting location changes. Such a chart is suitable for detecting any deviations from the EPA hypothesis. However, this chart cannot be the best possibility for detecting some types of shifts because the underlying ordinary sign test exhibits a low statistical power. Shifts affecting the symmetry of the δ_t distribution around zero can also be detected by a procedure relying on the Wilcoxon signed rank test. The Wilcoxon test exhibits a better power compared to the ordinary sign test in such situations. For this reason we exploit the second chart, based on the Wilcoxon test statistic, for detecting shifts violating the EPA by destroying the symmetry of the δ_t distribution. The statistics for both tests are asymptotically normally distributed under the EPA. Note that any other nonparametric test for symmetry or location can also be used in our methodology if its statistic could be sufficiently well approximated by the Gaussian distribution under the EPA.

The exponentially weighted moving average (EWMA) control charts (Montgomery, 2005, p. 405ff) are suitable for the EPA monitoring. Two EWMA schemes, based on the ordinary sign and the Wilcoxon rank test statistics, are started simultaneously. The design of this simultaneous chart is determined in accordance with the approach of Woodall and Ncube (1985). A signal would occur if any of control statistics crossed the corresponding critical limits for the first time. This would imply that the EPA hypothesis might not hold any more. The asymptotic critical limits for $K \to \infty$ are calculated for practically important parameter values of the simultaneous chart. The use of the asymptotic limits simplifies the application of our procedure to empirical problems.

The ability of our sequential procedure to detect deviations from the EPA is evaluated in an extensive Monte Carlo simulation study. We find that changes affecting the symmetry of the δ_t distribution (e.g., one strategy gets better or worse than all other alternatives) can on average be detected much quicker than changes of comparable size which do not affect distribution symmetry (e.g., one forecasting strategy gets better, but another one gets worse to the same extent). Shifts affecting symmetry are primarily detected by the Wilcoxon chart, while other shifts are predominantly signalled by the ordinary sign chart. Practical recommendations for control chart design are given based on a certain cost function for delays in shift detection. Moreover, we investigate the robustness of our procedure against suboptimal choices of the control chart parameters.

The rest of this paper is organized as follows. Section 2 discusses the nonparametric statistical tests required for EPA monitoring. Section 3 introduces the parsimonious sequential control procedure for online checking of EPA validity. Section 4 evaluates the performance of the suggested procedure in a Monte Carlo simulation study. Section 5 provides recommendations about optimal control chart design and investigates robustness issues. The concluding Section 6 presents a summary of the paper.

2. Testing validity of EPA

2.1. Ordinary sign test for location

Consider the distributional properties of the absolute innovations $|d_{k,t}|$ for detecting any change in the expectation vector $E(\mathbf{d}_t) = \mathbf{0}$ which would violate the EPA hypothesis. Since $E(d_{k,t}) = 0$, any violation of the EPA such that $E(d_{k,t}) \neq 0$ for some k would increase the average quantity of $|d_{k,t}|$s which are larger than some location measure. The idea of taking absolute values is crucial for our analysis. It can be illustrated with the example where $E(d_{1,t}) = -E(d_{K,t}) = a > 0$ whereas $E(d_{k,t}) = 0$ holds for $k = 2, \ldots, K-1$. Such a shift can be detected with the ordinary sign test for the absolute values, because realizations of $|d_{1,t}|$ and $|d_{K,t}|$ would tend to be larger than realizations of $|d_{k,t}|$s for all other k s. We exploit the median as a popular robust location measure. The median of absolute innovations would divide the set with K realizations $|d_{k,t}|$s into

two approximately equal parts under the EPA. Evidence that the number of elements in each part appear to be strongly unequal would indicate that the EPA hypothesis could be violated. Next, we formalize this idea by adapting the ordinary sign test for decisions concerning the EPA validity.

Consider the absolute values of the realizations $|d_{k,t}|$, $k=1,\ldots,K$, which can be seen as K draws of the random variable $|\delta_t|$. The median of the random variable $|\delta_t|$ under the EPA is denoted by $\mathrm{med}(|\delta_t|) > 0$. Recall that for the standard normal distribution $\mathrm{med}(|\delta_t|) = \Phi^{-1}(3/4) \approx 0.6745$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulated distribution function, while, for the standard Cauchy distribution with undefined variance, $\mathrm{med}(|\delta_t|) = 1$. This suggests that the measure $\mathrm{med}(|\delta_t|)$ should refer to the interval $[\frac{2}{3},1]$, which accounts for possible heavy tails of the distribution $\mathcal{F}(\cdot)$. In this study we use $\mathrm{med}(|\delta_t|) = \Phi^{-1}(\frac{3}{4})$, but in practice it should be estimated from the historical sample under the EPA validity.

Next, define the vector $\boldsymbol{\phi}_t = (\phi_{1,t},\ldots,\phi_{K,t})'$ with the elements

$$\phi_{k,t} = |d_{k,t}| - \mathrm{med}(|\delta_t|), \quad k = 1,\ldots,K. \tag{3}$$

If the measure $\mathrm{med}(|\delta_t|)$ is correctly specified, 50% of the observations in $\boldsymbol{\phi}_t$ should on average be smaller than zero and 50% greater than zero under the EPA hypothesis. Note that any EPA violation would increase the expected proportion of positive ϕ_ks, so that a change would only happen in one direction. In practice, however, it is reasonable to consider the problem as two-sided, because the median $\mathrm{med}(|\delta_t|)$ could be misspecified in situations where the true distribution function $\mathcal{F}(\cdot)$ is unknown. Robustness of the median as a location measure makes this approach resistent against situations where the tails of $\mathcal{F}(\cdot)$ are heavier than that of the normal distribution.

The ordinary sign test (Gibbons, 1971, p. 100) tests EPA validity by counting the number of positive elements $N_t \in [0,K]$ of the vector $\boldsymbol{\phi}_t$. Under the EPA the statistic $S_t^* = (2N_t - 1 - K)/\sqrt{K}$ is asymptotically distributed as $S_t^* \xrightarrow{L} \mathcal{N}(0,1)$ for $K \to \infty$. This result is motivated by the normal approximation to the binomial distribution with continuity correction (cf. Gibbons, 1971). Then, relying on the central limit theorem, the distribution of S_t^* converges to the limiting standard normal distribution at the rate $K^{-1/2}$. Further, we exploit the statistic S_t suggested by Borges (1970):

$$S_t = 4^{1/6}\sqrt{K+1/3} \int_{1/2}^{q} [s(1-s)]^{-1/3}\, ds, \quad q = \frac{N_t + 1/6}{K + 1/3}, \quad S_t \xrightarrow{L} \mathcal{N}(0,1), \tag{4}$$

because its distribution converges to normality more quickly at the rate K^{-1}.

2.2. Wilcoxon signed rank test for symmetry

The EPA hypothesis implies that the distribution of δ_t is symmetric around zero. Changes affecting the symmetry of the distribution of δ_t constitute an important class of shifts for timely detection. For example, such a shift occurs when a single forecasting model gets better (or worse) than others. The Wilcoxon signed rank test is an established nonparametric method for testing distribution symmetry (Gibbons, 1971, p. 106ff). It is more powerful than the ordinary sign test for this task.

Let $R_{k,t}$ be the ascending rank of the absolute innovations $|d_{k,t}|$ at time point t. The products $1_{(d_{k,t} > 0)}R_{k,t}$, where $1_{(\cdot)}$ is an indicator function, form the Wilcoxon signed rank statistic T_t^*, which is defined by

$$T_t^* = \sum_{k=1}^{K} 1_{(d_{k,t} > 0)} R_{k,t}. \tag{5}$$

The moments of T_t^* are given under the EPA hypothesis as

$$E(T_t^*) = \frac{K(K+1)}{4}, \quad V(T_t^*) = \frac{K(K+1)(2K+1)}{24}. \tag{6}$$

The standardized statistic $T_t = (T_t^* - E(T_t^*))/\sqrt{V(T_t^*)}$ is distributed as $T_t \xrightarrow{L} \mathcal{N}(0,1)$ for $K \to \infty$. The exact distribution of the Wilcoxon test statistic could be obtained numerically for each given value of K.

3. Sequential procedure

The validity of the EPA hypothesis is checked by sequential monitoring for changes in the expectations of the processes $\{S_t\}$ and $\{T_t\}$. A pair of univariate exponentially weighted moving average (EWMA) control charts is used for this purpose. Their control statistics are given for $t \geq 1$ by

$$Z_{1,t} = (1-\lambda)Z_{1,t-1} + \lambda S_t, \tag{7}$$

$$Z_{2,t} = (1-\lambda)Z_{2,t-1} + \lambda T_t, \tag{8}$$

where the memory parameter $\lambda \in (0,1]$. The starting values are selected to be $Z_{1,0} = Z_{2,0} = 0$. Both charts have the same parameter λ, because S_t and T_t follow the same asymptotic distribution for $K \to \infty$ under the EPA. Since both $Z_{1,t}$ and $Z_{2,t}$ have the same EPA asymptotic distribution, so a unique asymptotic critical limit $c = c_1 = c_2$ can be applied. The simultaneous use of these two charts is appropriate for timely detection of various EPA violations. The Wilcoxon-based

scheme is more suitable for quick detection of symmetry destroying shifts, while the ordinary sign scheme captures symmetry preserving changes as well.

A signal is given at time $t = L \geq 1$ if $|Z_{1,t}| > c$ or $|Z_{2,t}| > c$ occurs for the first time, whereby the time of the signal is further denoted as the run length L. The EPA distribution of the run length for the joint chart $L = \min\{L_1, L_2\}$ is required in order to select the critical limit c. Here L_1 and L_2 denote the run lengths for the schemes based on the ordinary sign and Wilcoxon statistics, respectively.

The average run length (ARL), defined as the expectation $E_\tau(L)$ for a given change point τ, is an important operational characteristic of a control chart. The in-control ARL for $\tau = \infty$, i.e., under the EPA, is further denoted by $E(L)$, while the out-of-control ARL for the given shift at $\tau = 1$ is hereafter denoted as $E_1(L)$. The critical limit c is usually determined by setting the in-control ARL, related to the test size, to a large desired value. A well-performing chart provides small out-of-control ARLs, which are related to the statistical power in the conventional test theory. In general, the out-of-control ARLs should decrease with the increase of the shift size. Woodall and Ncube (1985) suggest to approximate the in-control ARL $E(L)$ of the joint bivariate scheme as a function of the ARL for a single scheme $E(L_1)$ by

$$E(L) \approx \left(1 - \left(\frac{E(L_1) - 1}{E(L_1)} \right)^2 \right)^{-1}. \tag{9}$$

The approximation (9) requires mutual independence of $Z_{1,t}$ and $Z_{2,t}$. The properties $Corr(T_t, S_t^*) = 0$ and $Corr(T_t, S_t) = 0$ hold due to $E(d_{k,t} | d_{k,t}|) = 0$ for all symmetric continuous distributions. Consequently, $Corr(Z_{1,t}, Z_{2,t}) = 0$ under the EPA. Moreover, Woodall and Ncube argue that the approximation in (9) is also suitable in the bivariate case for weakly dependent control statistics.

Solving (9) immediately yields $E(L_1) \approx E(L) + \sqrt{E(L)^2 - E(L)}$. Thus, the critical limit c can be obtained for a given λ value by setting the in-control ARL $E(L) = \xi$, where ξ is the desired average (large) number of periods without false signals, and solving numerically $E(L_1 | c, \lambda) = \xi + \sqrt{\xi^2 - \xi}$ for c with a given λ value. The regula falsi numerical iteration method is used for calculating c. It stops if both conditions $|E(\hat{L}) - \xi| < 0.1$ and $|\Delta c| < 10^{-4}$ are satisfied, where Δc is a change in c at the final iteration. These conditions guarantee that the calculated critical limits provide the desired in-control ARL ξ with a maximal error of 0.5%, which is negligibly small in practical applications. The asymptotic critical limits c with $K \to \infty$ for different values of $E(L) = \xi$ and λ are reported in Table 1.

The critical limits from Table 1 could be used starting from $K \geq 20$. Since the distributions of the statistics T_t and S_t are only normal for $K \to \infty$, the goodness of the approximated critical limits c should be evaluated depending on the memory parameter λ for finite values of K. The differences between the exact critical limits for finite $K = 10, 20, 50$ and for their asymptotic counterparts are plotted in Fig. 1 for the in-control ARL $\xi = 20$ as a function of $\lambda \in [0.01, 0.9]$.

Fig. 1 shows that the differences between the asymptotic and finite K critical limits are negligible for $\lambda \leq 0.7$ and increase starting from $\lambda \geq 0.8$. The normality approximation fails for large λs ($\lambda \lesssim 1$) where the statistics $Z_{1,t}$ and $Z_{2,t}$ take only a limited number of distinct values due to the construction of the Wilcoxon and ordinary sign tests. We illustrate the impact of discreteness in the worst case situation, namely for the no-memory Shewhart chart with $\lambda = 1$. The out-of-control ARLs $E_1(L_1)$ and $E_1(L_2)$ of both S_t and T_t schemes are plotted as functions of the realizations N_t and T_t^* in Fig. 2 for a symmetry preserving shift with the size a, which is formally defined in Section 4.

Fig. 2 confirms that it is hardly reasonable to use the procedures for λs close to unity. The discreteness has an especially dramatic impact on the performance of the S_t chart, based on the ordinary sign test, where the step function is clearly observed. Consequently, the critical limits cannot be uniquely related to the desired ARLs because of the discrete control statistic for $\lambda = 1$. This phenomenon causes problems by controlling the statistical size of the procedure, which are most

Table 1
Asymptotic critical limits c for $K \to \infty$, various in-control ARL ξ and smoothing parameters λ, calculated with 10^6 replication with a precision $\xi \pm 0.1$.

| λ	$\xi = 20$	$\xi = 30$	$\xi = 40$	$\xi = 50$	$\xi = 60$	$\xi = 100$
0.01	0.04971	0.06091	0.06976	0.07710	0.08342	0.10210
0.03	0.13781	0.16321	0.18274	0.19762	0.20982	0.24336
0.05	0.21505	0.25017	0.27523	0.29438	0.30962	0.35073
0.10	0.37857	0.42781	0.46166	0.48689	0.50687	0.55970
0.15	0.51667	0.57526	0.61468	0.64384	0.66702	0.72833
0.20	0.63978	0.70528	0.74925	0.78190	0.80775	0.87591
0.30	0.85938	0.93644	0.98819	1.02636	1.05641	1.13772
0.40	1.05958	1.14680	1.20490	1.24846	1.28294	1.37508
0.50	1.24977	1.34657	1.41191	1.46030	1.49869	1.60182
0.60	1.43718	1.54334	1.61533	1.66888	1.71150	1.82590
0.70	1.62563	1.74195	1.82058	1.87946	1.92656	2.05303
0.80	1.81888	1.94642	2.03264	2.09763	2.14905	2.28868
0.90	2.02083	2.16075	2.25563	2.32728	2.38400	2.53761
1.00	2.23638	2.39080	2.49548	2.57386	2.63689	2.80603
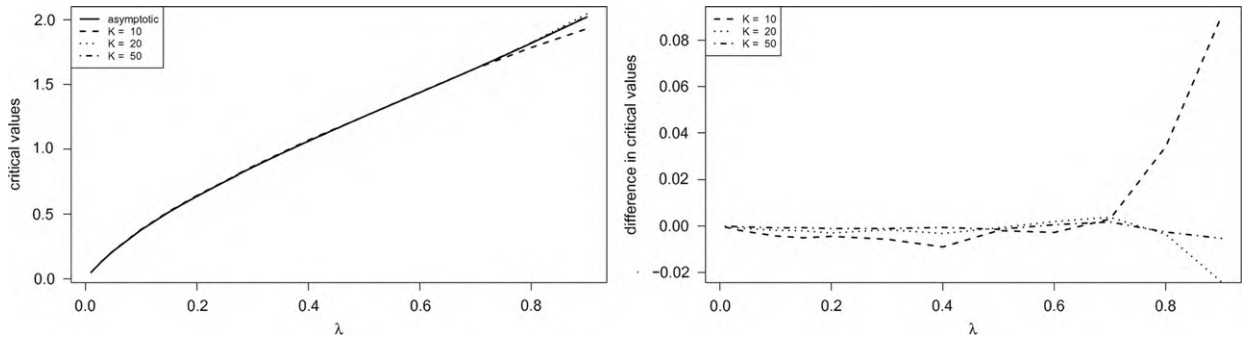
Fig. 1. Asymptotic and finite $K=10$, 20, 50 critical limits c (left) and differences between the asymptotic and finite K critical limits (right) as a function of λ calculated for in-control ARL $\xi = 20$.
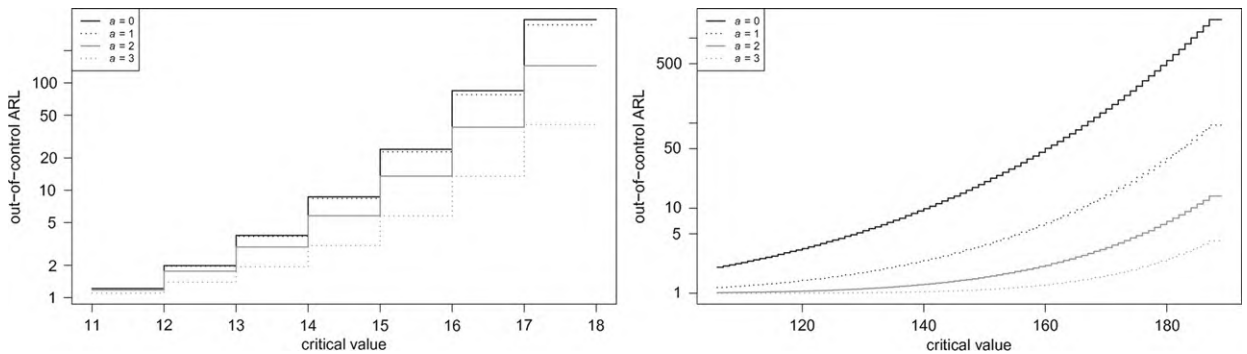

Fig. 2. Out-of-control ARLs for the single charts based on S_t (left) and T_t (right) with $\lambda = 1$, $K = 20$ as a function of N_t and T_t^* for symmetry preserving shift of size a.

pronounced for smaller values of K. The evidence suggests that the common advice to take large smoothing values $\lambda \lesssim 1$ for the quickest detection of large shifts (Montgomery, 2005, p. 412) could not be applied in our case, at least for practically relevant situations with $K \leq 50$.

4. Performance evaluation

The goodness of the monitoring procedure is evaluated by calculating the out-of-control ARLs $E_1(L)$. We model the shifts in the expectation vector $E(\mathbf{d}_t)$ by generating losses from the normal distribution with $d_{k,t} \sim \mathcal{N}(\alpha_k, 1)$ or $\mathbf{d}_t \sim \mathcal{N}(\boldsymbol{\alpha}, \mathbf{I})$ with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)'$.

The different specifications of α_ks reflect various types of shifts violating the EPA hypothesis. Two types of shifts, investigated in this study, are assessed by modeling α_ks with a single shift parameter a. The resulting out-of-control ARLs $E_1(L|a)$ can be interpreted as a detection delay for the given type and size of a shift a at time point $\tau = 1$. A shift of type I, preserving the symmetry of the distribution of δ_t, is modeled by setting the out-of-control expectation for the kth model, $k = 1, \ldots, K$, as

$$\alpha_k = a\left(\frac{k-1}{K-1} - \frac{1}{2}\right), \quad \alpha_k \in [-a/2, a/2]. \tag{10}$$

Such deviations from the EPA imply that if one model improves, another one worsens to the same extent. Consequently, a shift of type I can hardly be detected with the chart using the Wilcoxon statistic, because the distribution of δ_t remains symmetric around zero. The chart based on the ordinary sign statistic is, however, suitable for the detection of such shifts.

Alternatively, a shift of type II, which destroys the symmetry of the distribution of δ_t, is detected by

$$\alpha_k = \frac{a}{2}\frac{k-1}{K-1}, \quad \alpha_k \in [0, a/2]. \tag{11}$$

A shift of type II implies that all models, except the first one, improve (or worsen) but to different extents. This type of shift could be detected by both the Wilcoxon and the sign control charts, whereas the former is preferable due to its higher power. Fig. 3 visualizes the impact of change $\boldsymbol{\alpha}(a)$ in the expectation vector $E(\mathbf{d}_t)$ as a function of $k = 1, \ldots, K$ for both symmetry preserving type I and symmetry destroying type II shifts. Note that the change profiles in Fig. 3 could be easily generalized for non-linear alterations in $E(\mathbf{d}_t)$.
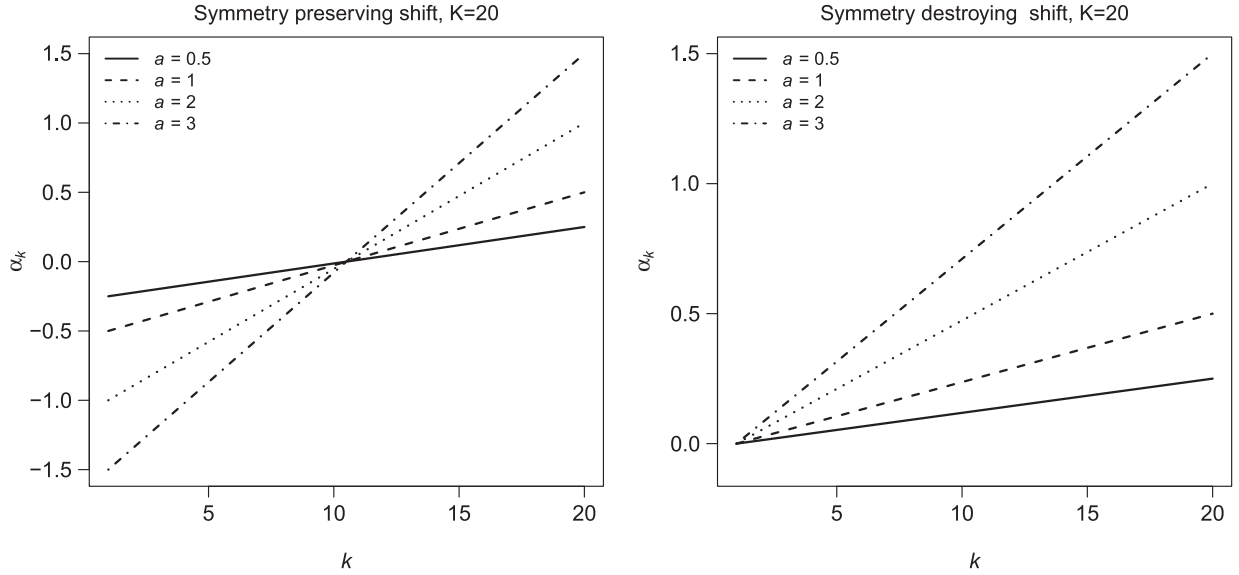
Fig. 3. Change α_k in the expectation $E(d_{k,t})$ as a function of $k=1,\ldots,K$ for symmetry preserving shift I (left) and symmetry destroying shift II (right) with $K=20$, $a=0$, 1, 2, 3.

Delays in shift detection, measured by the out-of-control ARL $E_1(L|a)$, are calculated for the shift a with $a \in \mathcal{A} = \{0.5,1,1.5,2,3\}$. Note that the resulting mean shifts of type I are larger than the mean shifts of type II. The smoothing parameter λ is chosen from the set $\mathcal{L}$ such that $\lambda \in \mathcal{L} = \{0.01,0.03,0.05,0.1,0.15,0.2,0.3,\ldots,1.0\}$. We consider the in-control ARLs $\xi = 20,50$ and the number of competing models $K=10$, 20, 50. The asymptotic critical limits c are taken from Table 1. The out-of-control ARLs $E_1(L)$ are reported in Table 2 for shifts of type I and in Table 3 for shifts of type II for the given (K,ξ,λ) parameters. The aforementioned results were obtained by 10^6 replications in a Monte-Carlo simulation study.

Table 2 reports the out-of-control ARLs for type I shifts, where the δ_t distribution remains symmetric around zero. The shifts are detected more quickly for larger K values. The optimal choice of λ depends on the shift size a, confirming the conventional result that the optimal smoothing parameter should increase with increasing shift magnitude. We found no great influence of the preselected in-control ARLs on the optimal λ values. These results are in general supported by the evidence from Table 3, where shifts of type II destroy the symmetry around zero required by the EPA. Note that shifts of type II are detected on average much more quickly than comparable type I shifts.

The found evidence speaks against using charts with λs close to unity. Fig. 1 reports that the asymptotic critical limits function properly only for $\lambda \le 0.8$. As shown in Tables 2 and 3 for a zero shift of size $a=0$, the actual in-control ARLs almost coincide with the required in-control ARLs only for $\lambda \le 0.8$. Using asymptotic critical limits for $\lambda > 0.8$, however, leads to very strong misspecifications of the in-control ARLs. This evidence confirms that the asymptotic critical limits are not appropriate for λs close to unity. Calculating the exact critical limits for each given K is rather computationally demanding. Balakrishnan et al. (2009) provide the exact critical limits for a special case $\lambda = 1$, however, they cover only a limited set of in- and out-of-control ARLs which could be achieved due to the discreteness of the control statistics. Since it is of practical interest to have critical limit which does not depend on the number of strategies K, the asymptotic critical limits for $K \to \infty$ are advantageous in applications. For this reason we consider further only memory parameters satisfying $\lambda \le 0.8$.

The benefits from using both charts simultaneously arise from the quick detection of different types of EPA violations. The chart based on the Wilcoxon statistic should be preferred for detecting shifts destroying symmetry, while the chart based on the ordinary sign test is more suitable for shifts where symmetry remains preserved. Signal proportions from the Wilcoxon T_t-chart are provided in Table 4, confirming our conjecture for the considered shift types. In particular, the majority of symmetry destroying shifts is signaled by the Wilcoxon chart, while symmetry preserving shift are primarily detected by the ordinary sign chart.

5. Selecting memory parameter λ

The size and type of a shift to occur is usually unknown. To make a recommendation concerning the choice of the smoothing parameter λ, we suggest to minimize the aggregated costs arising due to delays in the shift detection. Let the type of a coming shift be known. A unit of time delay in the shift detection causes costs assessed by the function $f(a)$ for each shift size a, which determines the shift vector $\boldsymbol{\alpha}$. The number of forecasting models K, the in-control ARL ξ and the smoothing parameter λ are given in advance. Then the aggregated cost function $\mathcal{Q}(K,\xi,\lambda)$ is defined as a definite integral over plausible shift sizes a. The integration goes over the cost function $f(a)$ for a shift size $a \in \mathcal{A}$, weighted by the

Table 2

Out-of-control ARL: symmetry preserving change at $\tau = 1 : \alpha_k = a((k-1)/(K-1) - \frac{1}{2}) \in [-a/2, a/2]$.

ARL=20							ARL=50					
λ	$K=10, a$						λ	$K=10, a$				
	0.00000	0.50000	1.00000	1.50000	2.00000	3.00000		0.00000	0.50000	1.00000	1.50000	2.00000	3.00000
0.01	19.60028	19.69670	18.73974	15.27849	11.13029	6.25072	0.01	48.76920	48.59378	41.76606	27.93143	18.00498	9.51465
0.03	19.48447	19.60250	18.67686	15.13175	10.95528	6.10737	0.03	48.45451	48.42780	41.77810	27.45127	17.17714	8.69565
0.05	19.44199	19.60988	18.71932	15.15331	10.89329	5.98944	0.05	48.25434	48.37406	42.07361	27.60007	16.88374	8.24784
0.10	19.38588	19.58132	18.85213	15.32663	10.89157	5.75893	0.10	47.93638	48.34596	43.26938	28.81158	17.06599	7.73066
0.15	19.31884	19.56379	19.02226	15.66777	11.06371	5.63595	0.15	48.21563	48.80227	44.82921	30.73193	17.95080	7.61221
0.20	19.46553	19.74742	19.34340	16.13896	11.42266	5.66675	0.20	48.06597	48.81555	45.69408	32.27298	18.88713	7.66200
0.30	19.43612	19.78508	19.68180	16.84115	12.11053	5.85156	0.30	48.42481	49.42287	47.46127	35.56324	21.30451	8.09213
0.40	19.15104	19.55927	19.65308	17.26319	12.59259	5.85716	0.40	50.25732	51.41965	50.50180	39.89270	24.87445	9.17932
0.50	19.83767	20.20352	20.58492	18.60738	14.05137	6.53431	0.50	50.73270	51.87562	51.34810	41.77188	27.03228	9.96511
0.60	19.68274	20.06781	20.56015	18.88154	14.53373	6.81980	0.60	52.19950	53.38461	53.29022	44.91702	30.35266	11.55486
0.70	20.20169	20.64659	21.25459	19.89705	15.73462	7.47826	0.70	51.99348	53.02937	53.40918	46.66598	33.27552	13.32201
0.80	22.03824	22.57759	23.61001	23.02845	19.23974	9.56083	0.80	48.14458	48.91078	49.11318	43.69659	32.42697	14.13761
0.90	24.22580	24.86369	26.39473	26.80113	23.80666	13.15424	0.90	39.60910	40.06495	40.22762	36.49299	28.42783	13.54859
1.00	24.63640	25.35809	26.82314	27.07672	23.98243	13.15259	1.00	36.78118	37.26984	37.40555	34.42117	27.42484	13.45182
λ	$K=20, a$						λ	$K=20, a$				
	0.00000	0.50000	1.00000	1.50000	2.00000	3.00000		0.00000	0.50000	1.00000	1.50000	2.00000	3.00000
0.01	19.86589	19.87003	18.02964	13.33669	9.11319	4.99974	0.01	49.44389	48.79967	38.11019	22.86020	14.11445	7.32581
0.03	19.75696	19.79917	17.95895	13.14008	8.83811	4.73681	0.03	49.18394	48.79627	38.10761	22.17039	13.25763	6.63920
0.05	19.71942	19.80832	18.01483	13.13145	8.72063	4.60055	0.05	49.21149	48.81802	38.42993	22.07706	12.84287	6.22170
0.10	19.71825	19.85837	18.17834	13.18658	8.59106	4.32325	0.10	49.23137	49.25695	40.05675	22.88216	12.62620	5.67609
0.15	19.74156	19.90862	18.41202	13.47934	8.65399	4.22837	0.15	48.95840	49.36025	41.42315	24.06246	12.86480	5.42151
0.20	19.65251	19.84866	18.55765	13.74099	8.73951	4.11368	0.20	49.23015	49.71936	42.92845	25.64172	13.42666	5.32992
0.30	19.84304	20.11747	19.17873	14.59862	9.30364	4.12181	0.30	49.55791	50.30905	45.39452	28.75324	14.99644	5.42133
0.40	19.73427	20.05567	19.44040	15.27459	9.80510	4.13431	0.40	50.10639	51.14803	47.67042	32.01827	16.97363	5.66620
0.50	19.86695	20.26019	20.00117	16.22846	10.66562	4.39949	0.50	50.90596	51.78168	49.42660	35.05941	19.09688	6.06087
0.60	20.13699	20.57112	20.46927	17.18533	11.59305	4.70273	0.60	50.88171	52.15320	50.68593	37.91130	21.53843	6.60345
0.70	20.25601	20.69634	20.88507	18.01701	12.50629	5.02277	0.70	51.55943	52.85434	52.31339	40.98129	24.46498	7.49700
0.80	19.72516	20.24336	20.64255	18.34281	13.25877	5.33445	0.80	52.39329	53.71415	53.91444	44.48781	27.92917	8.67960
0.90	18.75409	19.17535	19.48190	17.41252	12.72841	5.25436	0.90	51.41379	52.78084	54.06158	47.73196	33.17276	11.68044
1.00	15.54323	15.82759	16.05248	14.77865	11.40693	5.12249	1.00	49.85812	51.39350	52.64140	46.74290	32.92739	11.72995
λ	$K=50, a$						λ	$K=50, a$				
	0.00000	0.50000	1.00000	1.50000	2.00000	3.00000		0.00000	0.50000	1.00000	1.50000	2.00000	3.00000
0.01	19.89476	19.67237	15.83699	9.90225	6.27115	3.38081	0.01	49.74548	47.69886	30.30582	15.83080	9.48796	4.95398
0.03	19.88901	19.69759	15.79007	9.70781	6.08268	3.23253	0.03	49.71965	47.81579	29.92536	14.97641	8.70922	4.42106
0.05	19.92314	19.71931	15.78194	9.60143	5.92691	3.10558	0.05	49.74746	48.07976	30.09454	14.58493	8.25442	4.08829
0.10	19.90656	19.78243	15.89579	9.49526	5.68553	2.87978	0.10	49.58469	48.48320	31.32656	14.47540	7.70191	3.62905
0.15	19.89467	19.83994	16.13000	9.54430	5.57249	2.72681	0.15	49.70227	48.85244	32.84608	14.89394	7.54442	3.37378
0.20	19.88522	19.92806	16.40574	9.71117	5.55243	2.65383	0.20	49.86828	49.26425	34.54924	15.60298	7.57889	3.21880
0.30	19.88479	19.96154	16.94204	10.17160	5.59692	2.44671	0.30	49.82497	49.87007	37.44485	17.42635	7.95397	3.03560
0.40	19.93878	20.11583	17.58823	10.85179	5.87699	2.45063	0.40	49.93258	50.34894	40.06287	19.60381	8.63677	2.96332
0.50	19.90873	20.10312	18.03229	11.50491	6.15550	2.36629	0.50	50.43357	51.01658	42.56432	22.21697	9.72014	3.06935
0.60	20.09332	20.38608	18.70434	12.45404	6.71739	2.45562	0.60	50.56973	51.37337	44.58020	24.78845	10.86653	3.08748
0.70	20.17565	20.51142	19.20984	13.31081	7.28713	2.54636	0.70	50.73887	51.75811	46.44530	27.52804	12.41186	3.31549
0.80	19.82919	20.20361	19.24596	13.82937	7.63564	2.44549	0.80	51.02152	52.18384	48.11942	30.56010	14.30156	3.61187
0.90	19.68829	20.08432	19.46914	14.77178	8.56031	2.69316	0.90	50.89111	52.14974	49.28327	33.27292	16.46080	4.16153
1.00	17.76410	18.09398	17.61118	13.70807	8.25134	2.68951	1.00	41.16804	41.86201	39.68100	27.85028	14.66149	3.78943

corresponding out-of-control ARLs, which represent detection delays. This aggregated cost function is

$$\mathcal{Q}(K,\xi,\lambda) = \int_{a\in\mathcal{A}} f(a) E_1(L|K,\xi,\lambda,a)\, da. \tag{12}$$

The out-of-control ARL $E_1(L|K,\xi,\lambda,a)$ is calculated for the given shift size a. Then the optimal smoothing λ is chosen by minimizing the aggregate cost function $\mathcal{Q}(\cdot)$:

$$\lambda^*(\xi,K) = \underset{\lambda}{\arg\min}\, \mathcal{Q}(K,\xi,\lambda). \tag{13}$$

The weighting function $f(a)$ corresponds to the penalty for a time unit of detection delay given the shift size a. Thus the function $f(a)$ can be interpreted as the utility loss caused by a detection delay. The weighting function is either $f(a)=1$ with an equal penalty for all shift sizes, or $f(a)=a$ with a penalty proportional to the shift size a. The function $f(a)=1$ overweights smaller shifts, while the function $f(a)=a$ underweights smaller shifts in favour of larger ones.

Table 3
Out-of-control ARL: symmetry destroying change at $\tau = 1 : \alpha_k = (a/2)(k-1)/(K-1) \in [0,a/2]$.

	ARL=20							ARL=50					
λ	$K=10$, a						λ	$K=10$, a					
	0.00000	0.50000	1.00000	1.50000	2.00000	3.00000		0.00000	0.50000	1.00000	1.50000	2.00000	3.00000
0.01	19.60028	12.64752	7.63773	5.46913	4.37208	3.35591	0.01	48.76920	22.14103	11.95758	8.30353	6.55287	4.95005
0.03	19.50394	12.44870	7.40493	5.26680	4.18553	3.18066	0.03	48.42725	21.31457	11.06435	7.54259	5.89332	4.41727
0.05	19.44930	12.33677	7.24813	5.10314	4.03192	3.04575	0.05	48.21853	21.05438	10.56822	7.09060	5.48762	4.06928
0.10	19.37847	12.25445	7.00802	4.82526	3.76028	2.78760	0.10	47.84236	21.42108	10.09031	6.50503	4.93058	3.59320
0.15	19.32661	12.34330	6.92318	4.67034	3.58676	2.62320	0.15	48.22995	22.36943	10.10059	6.27281	4.65088	3.31160
0.20	19.44528	12.53728	6.94297	4.60200	3.49412	2.53674	0.20	48.05984	23.57464	10.37386	6.23021	4.51484	3.13970
0.30	19.42306	13.11655	7.20929	4.64224	3.44994	2.46446	0.30	48.47600	26.33266	11.37581	6.42159	4.45341	2.95259
0.40	19.17107	13.45056	7.44051	4.66660	3.36097	2.28428	0.40	50.39597	29.78925	13.05560	7.02507	4.66078	2.93849
0.50	19.80030	14.38888	8.09498	4.96925	3.49198	2.28635	0.50	50.69887	32.89277	15.15178	7.92531	5.02916	2.99213
0.60	19.69624	14.94556	8.68479	5.27064	3.60014	2.22755	0.60	52.26348	36.64557	18.04344	9.35186	5.70241	3.14064
0.70	20.20494	15.91038	9.67483	5.89426	3.94716	2.34349	0.70	51.91704	39.23960	21.03479	11.10328	6.57723	3.31605
0.80	22.02916	17.81253	11.11141	6.74439	4.41988	2.45444	0.80	48.13347	39.47621	24.19518	13.62870	8.13344	3.89182
0.90	24.22228	20.05384	12.93563	7.99425	5.20461	2.75284	0.90	39.63152	35.13027	25.02757	15.84429	9.91040	4.62818
1.00	24.65084	21.19967	14.79226	9.59209	6.33229	3.23891	1.00	36.75233	33.81236	26.54786	18.61948	12.49611	6.00820
λ	$K=20$, a						λ	$K=20$, a					
	0.00000	0.50000	1.00000	1.50000	2.00000	3.00000		0.00000	0.50000	1.00000	1.50000	2.00000	3.00000
0.01	19.86589	9.94343	5.58301	3.96975	3.18786	2.43125	0.01	49.44389	16.13112	8.41908	5.87341	4.66915	3.56229
0.03	19.75422	9.68760	5.35587	3.78083	3.02117	2.30156	0.03	49.24400	15.20011	7.66118	5.26555	4.15730	3.19083
0.05	19.76241	9.55594	5.19639	3.63213	2.88959	2.21568	0.05	49.12230	14.76853	7.21119	4.89396	3.83998	2.94355
0.10	19.73294	9.37744	4.91750	3.37976	2.66934	2.09760	0.10	49.16393	14.60205	6.62555	4.36363	3.37667	2.50741
0.15	19.74551	9.39141	4.77761	3.22840	2.53694	1.99894	0.15	49.00175	14.98654	6.39956	4.08573	3.11340	2.29881
0.20	19.66345	9.48501	4.69480	3.11596	2.42786	1.86863	0.20	49.17192	15.62415	6.33688	3.93449	2.95663	2.19926
0.30	19.79758	9.89203	4.68718	3.00073	2.28141	1.67380	0.30	49.59219	17.44140	6.52994	3.81544	2.78865	2.04253
0.40	19.74978	10.36555	4.78227	2.93984	2.16668	1.51550	0.40	50.13081	19.62236	7.00501	3.84710	2.70982	1.88449
0.50	19.89572	10.99932	5.00775	2.95609	2.11409	1.43359	0.50	50.65170	22.10425	7.75047	4.00621	2.69746	1.78266
0.60	20.11189	11.79146	5.36115	3.05715	2.12243	1.39773	0.60	50.94679	24.69163	8.72885	4.26929	2.72928	1.67894
0.70	20.21547	12.55742	5.80744	3.20651	2.15129	1.36731	0.70	51.57735	27.65289	10.13519	4.72342	2.85493	1.63440
0.80	19.74706	12.98276	6.18498	3.33320	2.15077	1.31616	0.80	52.31019	30.80337	12.00106	5.46757	3.12755	1.66016
0.90	18.75939	13.30687	6.79554	3.65734	2.29061	1.33418	0.90	51.45476	33.38447	14.27889	6.50525	3.55500	1.71408
1.00	15.52319	12.12214	6.96564	3.89233	2.40579	1.33702	1.00	49.97204	35.03354	16.57566	7.74305	4.11940	1.79746
λ	$K=50$, a						λ	$K=50$, a					
	0.00000	0.50000	1.00000	1.50000	2.00000	3.00000		0.00000	0.50000	1.00000	1.50000	2.00000	3.00000
0.01	19.89476	6.72243	3.65504	2.62914	2.15896	1.92815	0.01	49.74548	10.30455	5.36844	3.80499	3.09187	2.32106
0.03	19.90820	6.49522	3.47884	2.49554	2.08398	1.78373	0.03	49.72437	9.47694	4.80644	3.38677	2.73009	2.06063
0.05	19.88931	6.33391	3.34115	2.39608	2.01738	1.61976	0.05	49.71101	9.00170	4.45841	3.11790	2.47854	2.01469
0.10	19.87615	6.06601	3.10367	2.21709	1.84619	1.30838	0.10	49.66221	8.43870	3.95608	2.71595	2.18737	1.89756
0.15	19.89123	5.95129	2.95060	2.08442	1.68908	1.16080	0.15	49.59049	8.29734	3.68928	2.50490	2.04681	1.61405
0.20	19.88775	5.92140	2.84613	1.97685	1.56311	1.09089	0.20	49.75942	8.36393	3.52958	2.36838	1.91921	1.37823
0.30	19.88437	5.97732	2.70490	1.81312	1.39192	1.03654	0.30	49.78038	8.84628	3.37484	2.17844	1.69282	1.15107
0.40	19.95708	6.22347	2.64144	1.70619	1.29712	1.01893	0.40	49.87475	9.66628	3.34298	2.04987	1.53754	1.07353
0.50	19.87692	6.53293	2.61710	1.63357	1.23792	1.01154	0.50	50.46588	10.79914	3.41248	1.97653	1.44267	1.04442
0.60	20.10326	6.98913	2.65527	1.59453	1.20291	1.00825	0.60	50.63746	12.15048	3.55175	1.93271	1.37444	1.02881
0.70	20.15095	7.51898	2.74158	1.58329	1.18450	1.00639	0.70	50.61098	13.78556	3.80303	1.93469	1.34029	1.02155
0.80	19.84190	8.02550	2.83788	1.57459	1.16756	1.00511	0.80	51.01141	15.78248	4.20386	1.97940	1.32617	1.01817
0.90	19.70039	8.70812	3.03869	1.59985	1.16549	1.00465	0.90	51.01053	18.05070	4.77285	2.07203	1.32739	1.01609
1.00	17.75387	9.09074	3.30061	1.64974	1.16943	1.00446	1.00	41.09257	18.86873	5.45429	2.22079	1.34496	1.01498

The practically relevant shift magnitudes (e.g., in engineering applications) can be characterized by the interval $a \in [0.04,4]$ (cf. Montgomery, 2005, p. 217ff). Since the out-of-control ARLs $E_1(L|K,\xi,\lambda,a)$ cannot be calculated explicitly, we replace the integral $\mathcal{Q}(\cdot)$ by the sum $\mathcal{Q}^*(\cdot)$ over the set of shifts $\mathcal{A} = \{a_1 = 0.04, a_2 = 0.08, \ldots, a_{n_a} = 4.00\}$, which consists of $n_a = |\mathcal{A}| = 100$ elements. Then the integral in (12) is represented by the sum of the corresponding rectangles:

$$\mathcal{Q}(K,\xi,\lambda) \approx \mathcal{Q}^*(K,\xi,\lambda,n_a) = \sum_{a_i} f(a_i) E_1(L|K,\xi,\lambda,a_i)(a_i^* - a_{i-1}^*), \quad i = 1, \ldots, n_a,$$

where $a_i^* = (a_{i+1} + a_i)/2$, $a_0 = 0$ and $a_{n_a+1} = 4.04$. Consequently,

$$\mathcal{Q}^*(K,\xi,\lambda,n_a) = n_a^{-1}(a_{n_a} - a_0)\sum_{a_i} f(a_i) E_1(L|K,\xi,\lambda,a_i).$$

Table 4
Signal proportions given by T-chart for symmetry preserving $\alpha_k = a((k-1)/(K-1) - \frac{1}{2})$ (left) and symmetry destroying $\alpha_k = (a/2)(k-1)/(K-1)$ (right) shifts at $\tau = 1$.

|	ARL=50, K=20, α-preserving						ARL=50 K=20, α-destroying				
$\lambda \backslash a$	0.50000	1.00000	1.50000	2.00000	3.00000	$\lambda \backslash a$	0.50000	1.00000	1.50000	2.00000	3.00000
0.01	0.47357	0.30548	0.10671	0.02491	0.00041	0.01	0.92821	0.98768	0.99581	0.99765	0.99751
0.03	0.47201	0.30649	0.11235	0.02857	0.00068	0.03	0.91664	0.98342	0.99366	0.99595	0.99628
0.05	0.46925	0.31033	0.11637	0.03165	0.00098	0.05	0.90817	0.97991	0.99168	0.99465	0.99474
0.10	0.46790	0.31815	0.12467	0.03560	0.00156	0.10	0.89101	0.97196	0.98747	0.99175	0.99363
0.15	0.46127	0.32285	0.13047	0.03802	0.00196	0.15	0.87505	0.96527	0.98334	0.98863	0.99129
0.20	0.45830	0.32707	0.13652	0.03998	0.00218	0.20	0.86073	0.95972	0.98035	0.98648	0.99125
0.30	0.44862	0.33022	0.14553	0.04351	0.00261	0.30	0.83232	0.95009	0.97527	0.98378	0.98880
0.40	0.43903	0.33056	0.15334	0.04666	0.00273	0.40	0.80167	0.93723	0.96766	0.97660	0.97728
0.50	0.42730	0.32718	0.15887	0.04999	0.00273	0.50	0.76824	0.92378	0.96064	0.97215	0.97872
0.60	0.41403	0.32133	0.16363	0.05403	0.00320	0.60	0.72963	0.90139	0.94550	0.95667	0.95761
0.70	0.40591	0.31997	0.17142	0.05899	0.00361	0.70	0.69490	0.88457	0.93836	0.95444	0.96117
0.80	0.39968	0.31783	0.17832	0.06447	0.00399	0.80	0.66090	0.86399	0.92943	0.94974	0.96072
0.90	0.38555	0.31196	0.18676	0.07577	0.00556	0.90	0.61645	0.83195	0.91463	0.94326	0.96065
1.00	0.39501	0.32187	0.19429	0.07987	0.00604	1.00	0.59156	0.80118	0.89328	0.92899	0.95473
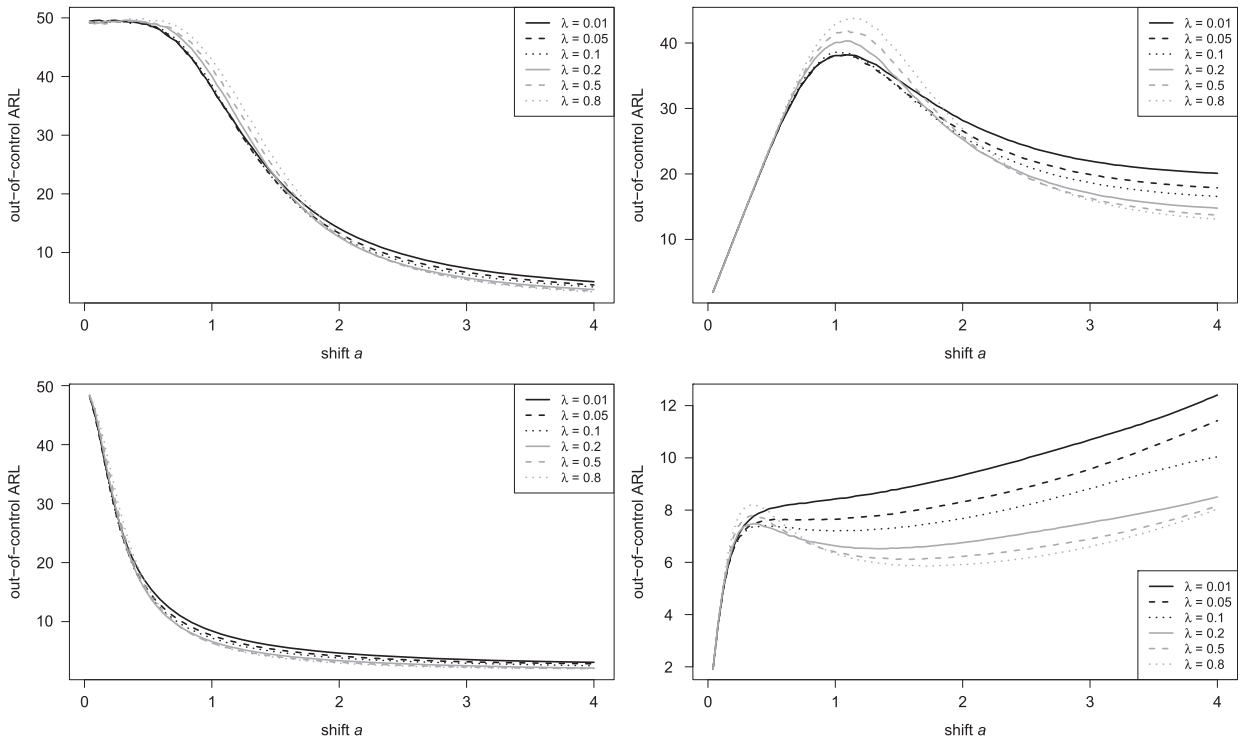
Fig. 4. Weighted costs of delay in detection $E_1(L|K = 20, \xi = 50, \lambda, a)$ (left) and $aE_1(L|K = 20, \xi = 50, \lambda, a)$ (right) as functions of shift a for different values of λ for shift I (above) and shift II (below) with $\lambda \le 0.8$.

Fig. 4 plots the weighted costs of delay in the detection $E_1(L|K = 20, \xi = 50, \lambda, a)$ and $a \cdot E_1(L|K = 20, \xi = 50, \lambda, a)$ in dependence on the shift size a for different values of the smoothing parameter λ. Fig. 4, left, shows the importance of small shift detection for the equal penalty function, whereas shifts of size about $a \approx 1$, see Fig. 4 right, are more costly for the proportional penalties.

The suggested criterion for the choice of λ from the set $\{0.01, \ldots, 0.8\}$ should minimize the aggregate cost function $\mathcal{Q}(\cdot)$ in (12), which corresponds to the area under the curves.

Using the approximation $\mathcal{Q}^*(\cdot)$, we provide evidence about the optimal λs given a cost function $f(a)$. The optimal values of λ for shifts of type I and II are presented in Table 5 for $f(a) = 1$ and in Table 6 for $f(a) = a$. The upper parts of the tables show the results for all shifts $\mathcal{A} = \{0.04, \ldots, 4\}$, while the middle and lower parts report the values for the small shifts $\mathcal{A}_{small} = \{0.04, \ldots, 1\}$ and large shifts $\mathcal{A}_{large} = \{1, \ldots, 4\}$, respectively.

Table 5
Optimal λs for all (upper panel), small (middle panel) and large (lower panel) shifts, $f(a)=1$.

ARL	All shifts I, $\mathcal{A}=\{0.04,\dots,4\}$						All shifts II, $\mathcal{A}=\{0.04,\dots,4\}$					
	20	30	40	50	60	100	20	30	40	50	60	100
K												
10	0.1	0.1	0.1	0.05	0.05	0.05	0.2	0.2	0.15	0.15	0.1	0.1
20	0.1	0.1	0.05	0.05	0.05	0.05	0.4	0.2	0.15	0.15	0.15	0.1
50	0.15	0.1	0.1	0.1	0.05	0.05	0.3	0.3	0.2	0.2	0.2	0.1
	Small shifts I, $\mathcal{A}_{\text{small}}=\{0.04,\dots,1\}$						Small shifts II, $\mathcal{A}_{\text{small}}=\{0.04,\dots,1\}$					
10	0.1	0.1	0.05	0.05	0.03	0.03	0.15	0.1	0.1	0.05	0.05	0.05
20	0.03	0.03	0.03	0.03	0.03	0.01	0.15	0.1	0.1	0.1	0.05	0.05
50	0.01	0.03	0.01	0.03	0.03	0.01	0.2	0.15	0.1	0.1	0.1	0.05
	Large shifts I, $\mathcal{A}_{\text{large}}=\{1,\dots,4\}$						Large shifts II, $\mathcal{A}_{\text{large}}=\{1,\dots,4\}$					
10	0.1	0.1	0.1	0.05	0.05	0.05	0.4	0.3	0.3	0.3	0.2	0.2
20	0.1	0.1	0.1	0.1	0.05	0.05	0.5	0.5	0.5	0.4	0.4	0.3
50	0.15	0.15	0.15	0.1	0.1	0.1	0.7	0.6	0.6	0.6	0.6	0.6

Table 6
Optimal λs for all (upper panel), small (middle panel) and large (lower panel) shifts, $f(a)=a$.

ARL	All shifts I, $\mathcal{A}=\{0.04,\dots,4\}$						All shifts II, $\mathcal{A}=\{0.04,\dots,4\}$					
	20	30	40	50	60	100	20	30	40	50	60	100
K												
10	0.15	0.1	0.1	0.1	0.1	0.05	0.4	0.3	0.3	0.2	0.2	0.15
20	0.2	0.15	0.15	0.1	0.1	0.05	0.5	0.5	0.4	0.4	0.4	0.2
50	0.3	0.2	0.15	0.15	0.1	0.1	0.5	0.5	0.5	0.4	0.4	0.4
	Small shifts I, $\mathcal{A}_{\text{small}}=\{0.04,\dots,1\}$						Small shifts II, $\mathcal{A}_{\text{small}}=\{0.04,\dots,1\}$					
10	0.03	0.05	0.03	0.03	0.03	0.01	0.15	0.1	0.1	0.1	0.05	0.05
20	0.03	0.03	0.03	0.03	0.01	0.01	0.15	0.15	0.1	0.1	0.1	0.1
50	0.01	0.03	0.03	0.03	0.03	0.01	0.3	0.2	0.2	0.15	0.15	0.1
	Large shifts I, $\mathcal{A}_{\text{large}}=\{1,\dots,4\}$						Large shifts II, $\mathcal{A}_{\text{large}}=\{1,\dots,4\}$					
10	0.15	0.1	0.1	0.1	0.1	0.05	0.4	0.3	0.3	0.3	0.3	0.3
20	0.2	0.15	0.15	0.1	0.1	0.1	0.5	0.6	0.6	0.6	0.5	0.5
50	0.3	0.2	0.2	0.15	0.15	0.1	0.8	0.7	0.7	0.7	0.7	0.6

The results from Tables 4 and 5 suggest that the optimal λ should be smaller for the cost function $f(a)=1$ than for the function $f(a)=a$. The optimal λs are larger for situations with symmetry preserving shifts than for symmetry destroying shifts. The latter finding corresponds to the evidence that the out-of-control ARLs are larger for type I shifts than for type II shifts. If there is no prior information about the size and type of the shift, we recommend a smoothing parameter between 0.05 and 0.3 relying on the function $f(a)=1$, and between 0.1 and 0.5 relying on the function $f(a)=a$. For small shifts the corresponding optimal λs belong to the intervals [0.03,0.15] and [0.03,0.2], respectively. For large shifts the optimal λs refers to the intervals [0.1,0.6] and [0.1,0.7] for $f(a)=1$ and $f(a)=a$ functions.

The visual presentation of the aggregate cost functions $\mathcal{Q}^*(K=20,\xi,\lambda)$ provides the information concerning their sensitivity with respect to the choice of λ. Fig. 5 shows the aggregated costs $\mathcal{Q}^*(K=20,\xi,\lambda)$, calculated as $n_a^{-1}(a_{n_a}-a_0)\sum_{a_i}E_1(L|K=20,\xi,\lambda,a_i)$ and $n_a^{-1}(a_{n_a}-a_0)\sum_{a_i}a_iE_1(L|K=20,\xi,\lambda,a_i)$, for different in-control ARLs ξ in dependence on λ. Evidence from Fig. 5 suggests that the aggregate cost criterion is quite insensitive to λs from the recommended intervals, namely [0.05;0.3] for $f(a)=1$ and [0.1;0.5] for $f(a)=a$ functions. Thus our procedure shows robustness with respect to the suboptimal choices of λ for the considered cost functions $f(a)$.

6. Summary

A sequential monitoring approach is developed for online decisions about the validity of the equal predictive ability (EPA) hypothesis for a large number of competing forecasting models or forecasters. It combines two nonparametric EWMA control charts based on the Wilcoxon signed rank test and the ordinary sign test statistics. A signal from the introduced simultaneous control chart would indicate that the EPA validity is statistically rejected starting from this point in time. Our monitoring scheme shows sound detecting properties for important types of shifts violating the EPA.
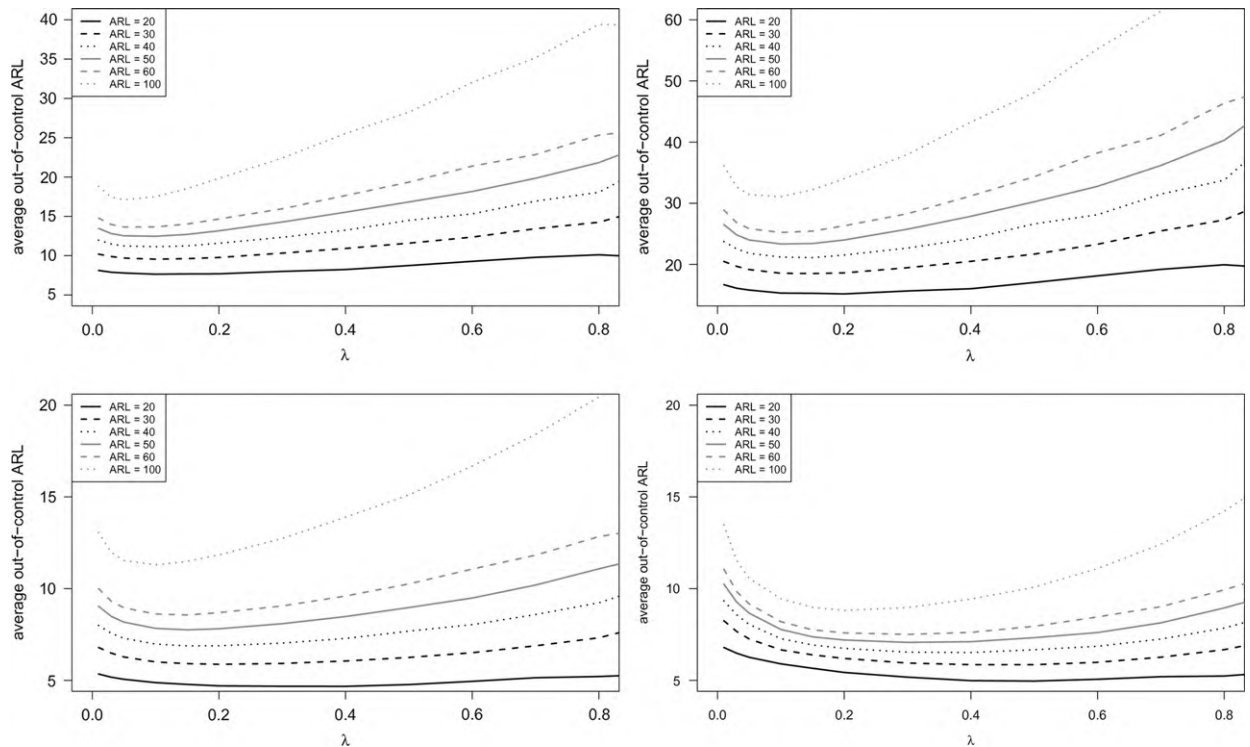
Fig. 5. Aggregated costs of detection delay $n_a^{-1}\sum_{a_i} E_1(L|K=20,\xi,\lambda,a_i)$ (left) and $n_a^{-1}\sum_{a_i} a_i E_1(L|K=20,\xi,\lambda,a_i)$ (right) as functions of λ for different values of ξ for shift I (above) and shift II (below).

Moreover, it is parsimonious and robust against suboptimal choices of the control chart parameters. Recommendations concerning the optimal design of the procedure are provided based on aggregated costs due to delays in shift detections.

References

Balakrishnan, N., Triantafyllou, I.S., Koutras, M.V., 2009. Nonparametric control charts based on runs and Wilcoxon-type rank-sum statistics. Journal of Statistical Planning and Inference 139, 3177–3192.

Borges, R., 1970. Eine Approximation der Binomialverteilung durch die Normalverteilung der Ordnung 1/n. Zeitschift für Wahrscheinlichkeitstheorie und verwandte Gebiete 14, 189–199.

Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 134–144.

Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74, 1545–1578.

Gibbons, J.D., 1971. Nonparametric Statistical Inference. McGraw-Hill, New York.

Hajek, J., Sidak, Z., Sen, P.K., 1999. Theory of Rank Tests, second ed. Academic Press, San Diego, CA, USA.

Montgomery, D.C., 2005. Introduction to Statistical Quality Control, fifth ed. Wiley, New Jersey.

West, K., 1996. Asymptotic inferences about predictive ability. Econometrica 64, 1067–1084.

Woodall, W.H., Ncube, M.M., 1985. Multivariate CUSUM quality control procedures. Technometrics 27, 285–292.