Supplemental file for the paper titled "On Nonparametric Statistical Process Control Of Univariate Processes"

To save some space in the printed paper, some numerical results are presented in this supplemental file. First, in Figures 2 and 3 of the printed paper, we have presented results of the eight control charts when $M = 500$, their parameters are chosen to be the optimal ones for detecting the shift of size 0.6, and these parameters are then used in all other cases. In Figures S.1 and S.2 here, we present the corresponding results when $M = 200$ and other parameters are chosen in exactly the same way as that of Figures 2 and 3 of the printed paper.

In Figures 2 and 3 of the printed paper and Figures S.1 and S.2 here, we have presented results of the eight control charts when their parameters are chosen to be the optimal ones for detecting the shift of size 0.6 and these parameters are then used in all other cases. In Figures S.3-S.6 here, we present the corresponding results when the parameters of the eight charts are chosen to minimize the OC ARL values for detecting each individual shift, while their IC ARL values are all fixed at 500. The remaining setup of the example considered here is exactly the same as that of the examples considered in Figures 2 and 3 of the printed paper and Figures S.1 and S.2 here.

In Figure 5 of the printed paper, we investigate the impact of the IC sample size M on the optimal OC ARL values of the P-CUSUM chart. Intuitively, the value of M may also have an impact on the variability of the run length distribution of the P-CUSUM chart. To investigate this issue, we compute the sample standard deviation of the 10,000 run length values (denoted as SDRL) obtained from the 10,000 replications in each case considered in Figure 5 of the printed paper, and the results are presented in Figure S.7 here. From the figure, it can be seen that (i) the SDRL value tends to be smaller when M increases, and (ii) its values are stable when $M \geq 200$ and are almost identical when $M \geq 500$ in most cases considered.

Based on the results presented in Figures 6 and 7 in the printed paper, we made the following three practical guidelines for selecting the value of p in Section 3.3 of the printed paper. (i) p can be chosen smaller when m is larger. (ii) In cases when we do not have any prior information about the process distribution, then we can choose $p = 10$. In such cases, the P-CUSUM chart should perform reasonably well. (iii) If we know that the process distribution is quite symmetric, or that it is skewed but the potential shift is in the direction of the longer tail, then p can be chosen as small as 5. Here, we provide some explanations of the three practical guidelines. First, the above

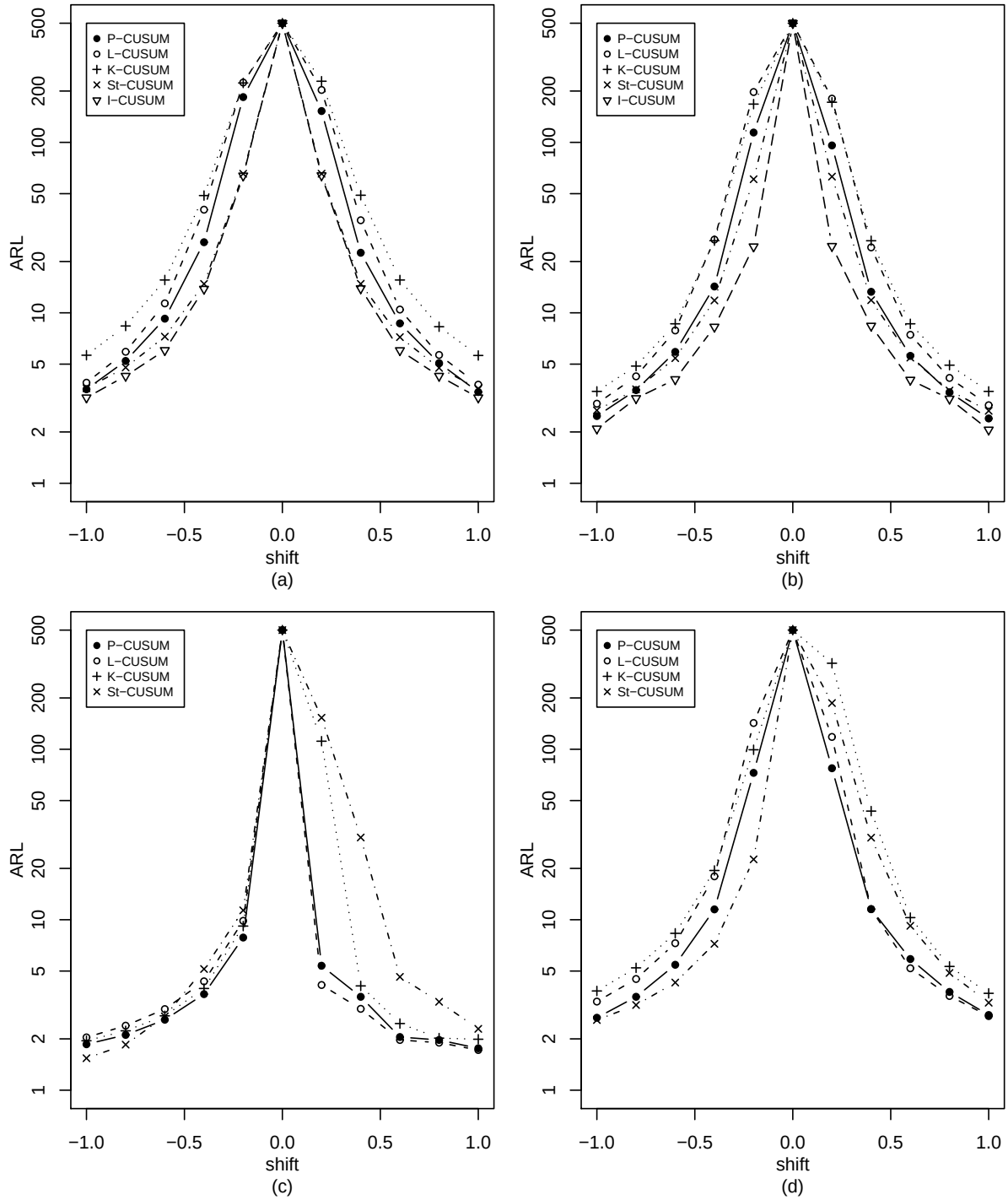

Figure S.1: OC ARL values of five control charts when the IC ARL is 500, $M = 200$, $m = 5$, and the actual IC process distribution is the standardized version of $N(0, 1)$ (plot (a)), $t(4)$ (plot (b)), $\chi^2(1)$ (plot (c)), and $\chi^2(4)$ (plot (d)). Procedure parameters of the control charts are chosen to be the ones that minimize their OC ARL values when detecting the shift of 0.6. Scale on the y-axis is in natural logarithm.

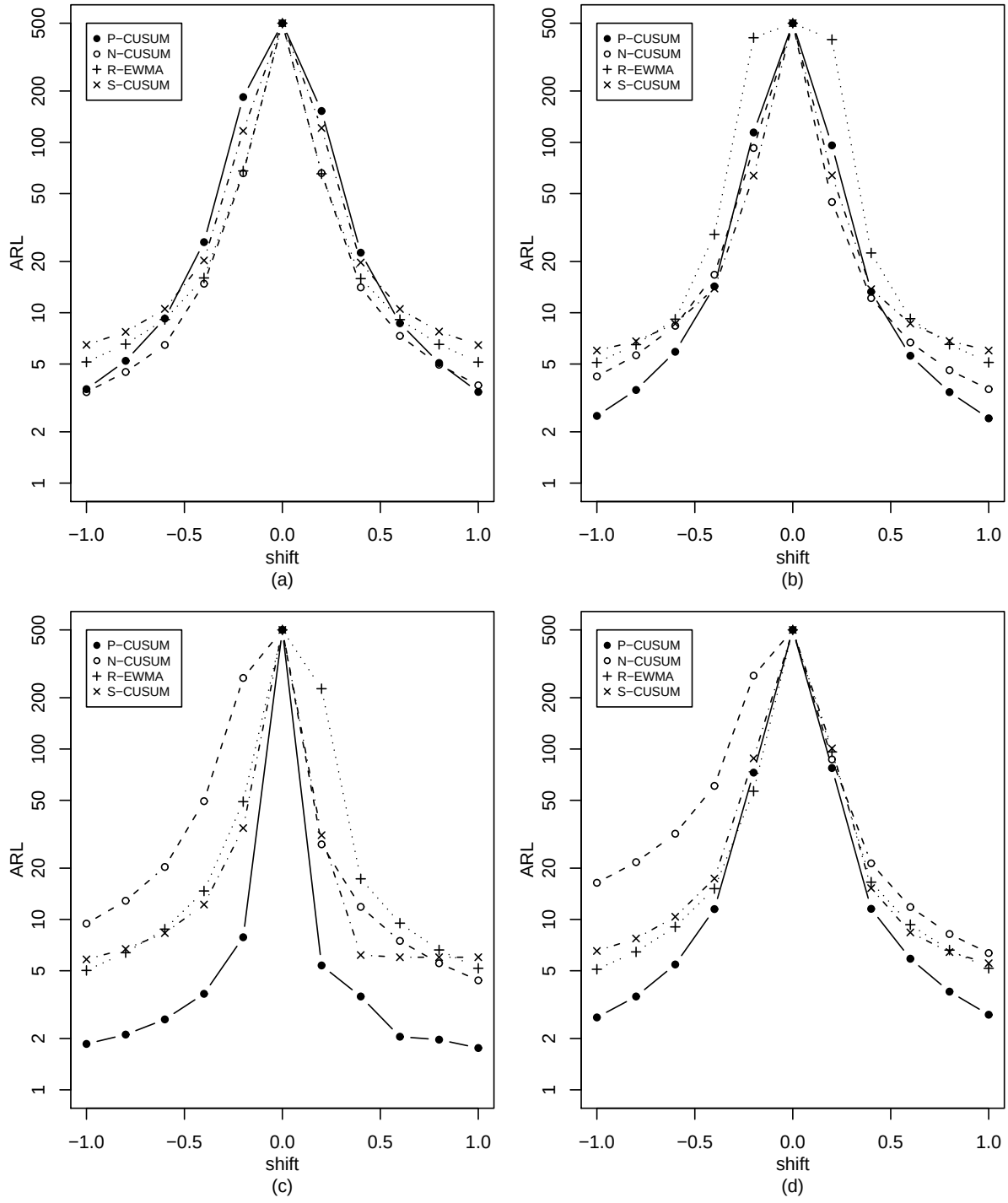

Figure S.2: OC ARL values of four control charts when the IC ARL is 500, $M = 200$, $m = 5$, and the actual IC process distribution is the standardized version of $N(0, 1)$ (plot (a)), $t(4)$ (plot (b)), $\chi^2(1)$ (plot (c)), and $\chi^2(4)$ (plot (d)). Procedure parameters of the control charts are chosen to be the ones that minimize their OC ARL values when detecting the shift of 0.6. Scale on the y-axis is in natural logarithm.

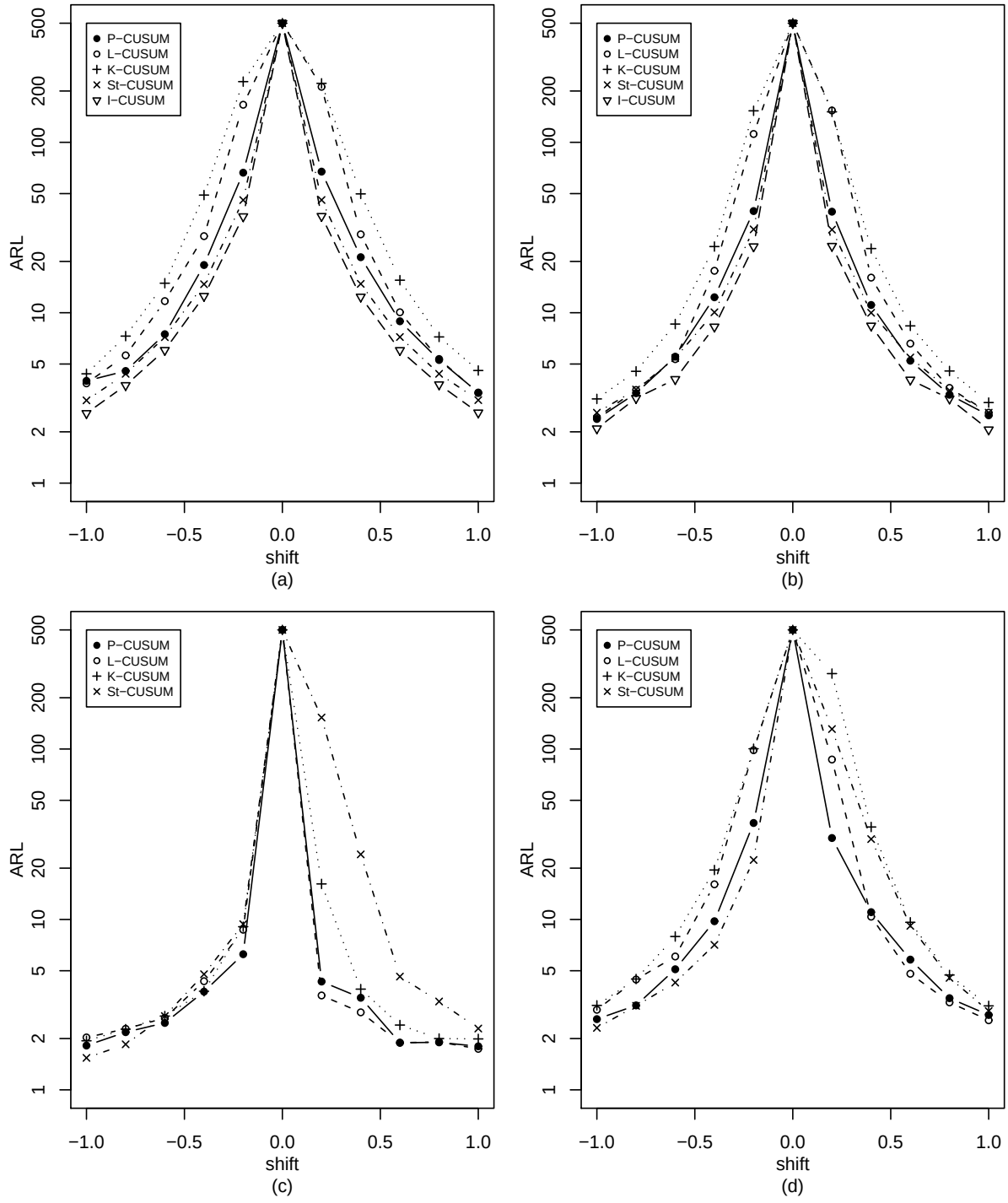

Figure S.3: Optimal OC ARL values of five control charts when the IC ARL is 500, $M = 500$, $m = 5$, and the actual IC process distribution is the standardized version of $N(0, 1)$ (plot (a)), $t(4)$ (plot (b)), $\chi^2(1)$ (plot (c)), and $\chi^2(4)$ (plot (d)). Procedure parameters of the control charts are chosen to be the ones that minimize their OC ARL values when detecting each individual shift. Scale on the y-axis is in natural logarithm.

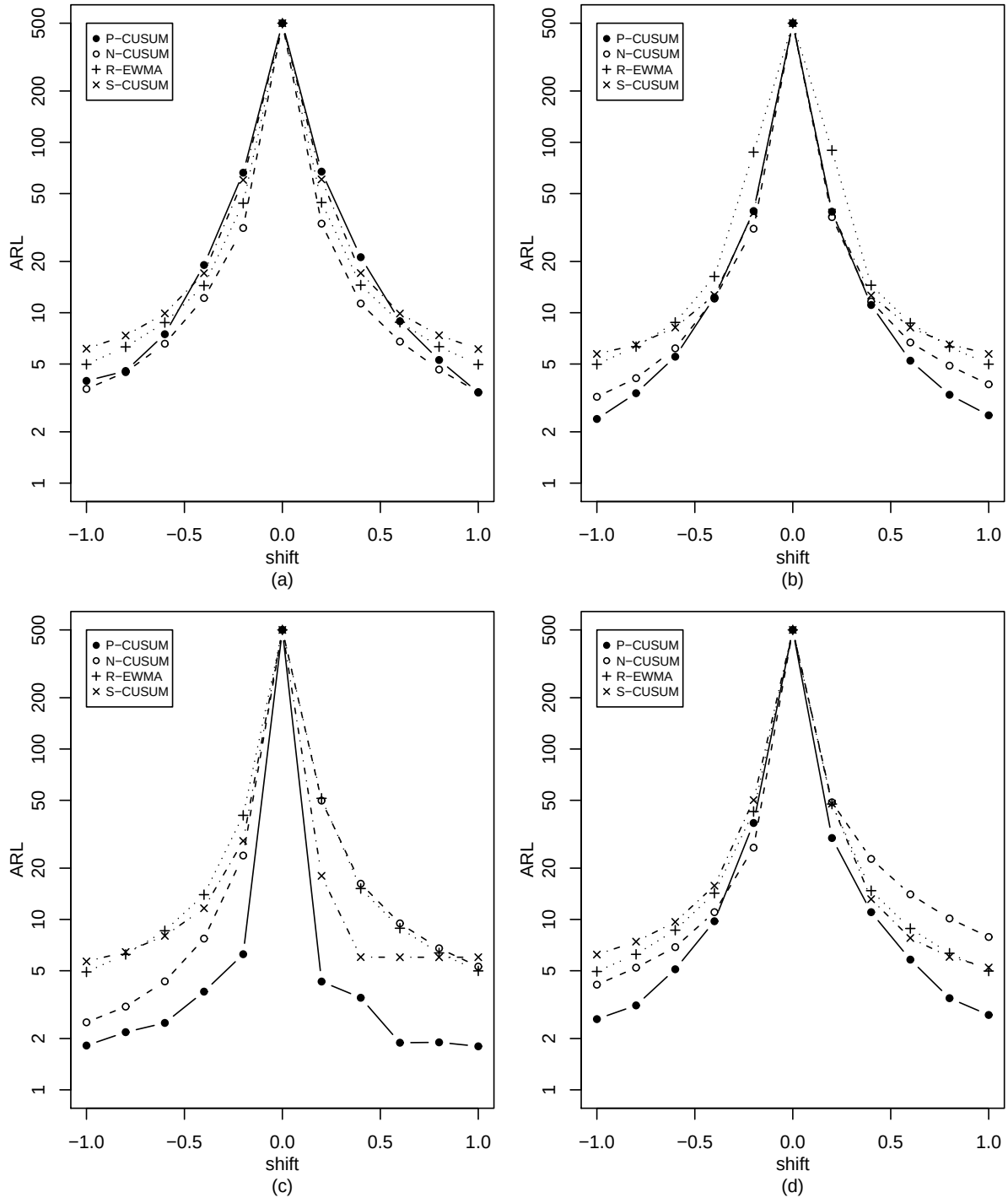

Figure S.4: Optimal OC ARL values of four control charts when the IC ARL is 500, $M = 500$, $m = 5$, and the actual IC process distribution is the standardized version of $N(0, 1)$ (plot (a)), $t(4)$ (plot (b)), $\chi^2(1)$ (plot (c)), and $\chi^2(4)$ (plot (d)). Procedure parameters of the control charts are chosen to be the ones that minimize their OC ARL values when detecting each individual shift. Scale on the y-axis is in natural logarithm.

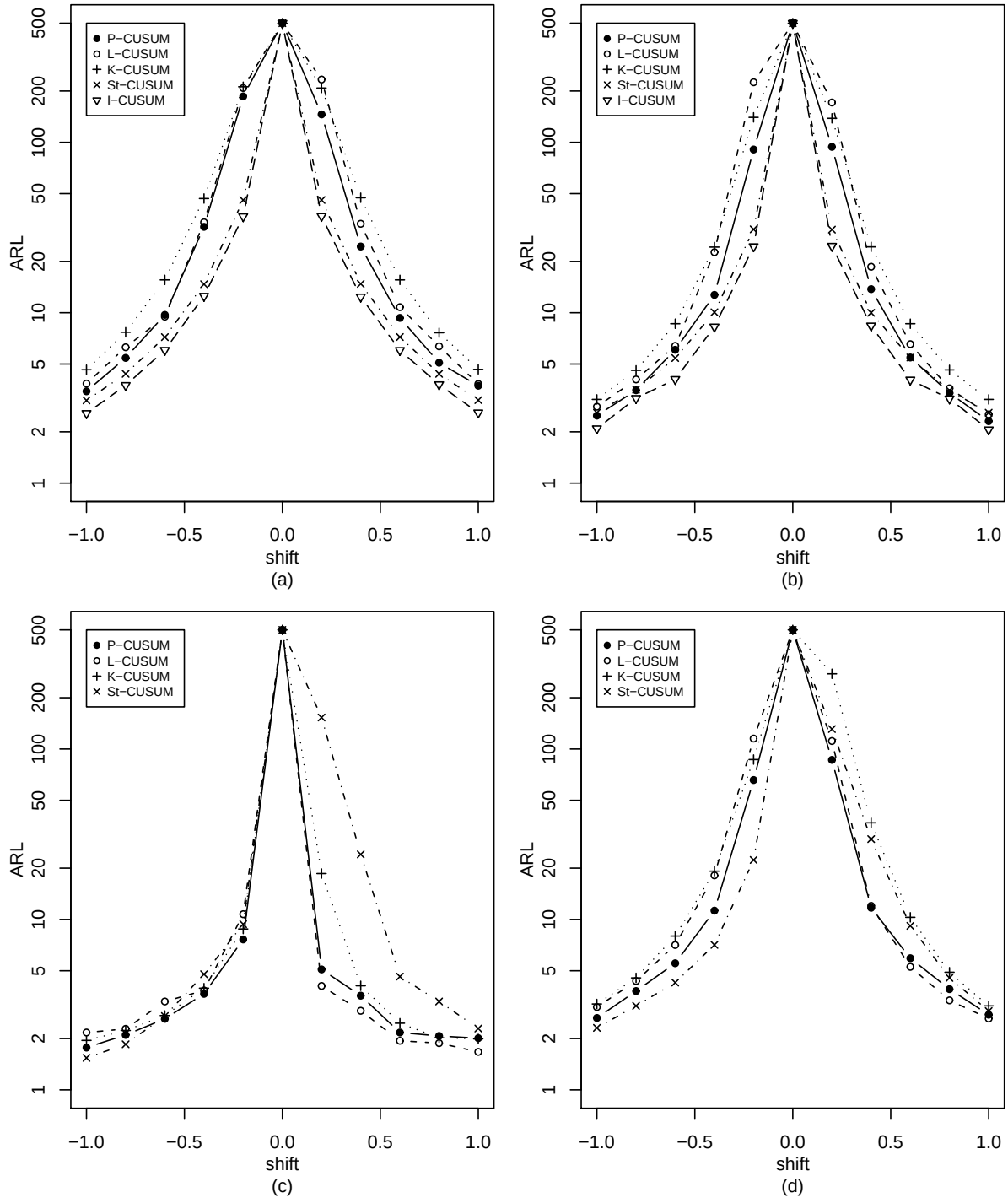

Figure S.5: Optimal OC ARL values of five control charts when the IC ARL is 500, $M = 200$, $m = 5$, and the actual IC process distribution is the standardized version of $N(0, 1)$ (plot (a)), $t(4)$ (plot (b)), $\chi^2(1)$ (plot (c)), and $\chi^2(4)$ (plot (d)). Procedure parameters of the control charts are chosen to be the ones that minimize their OC ARL values when detecting each individual shift. Scale on the y-axis is in natural logarithm.

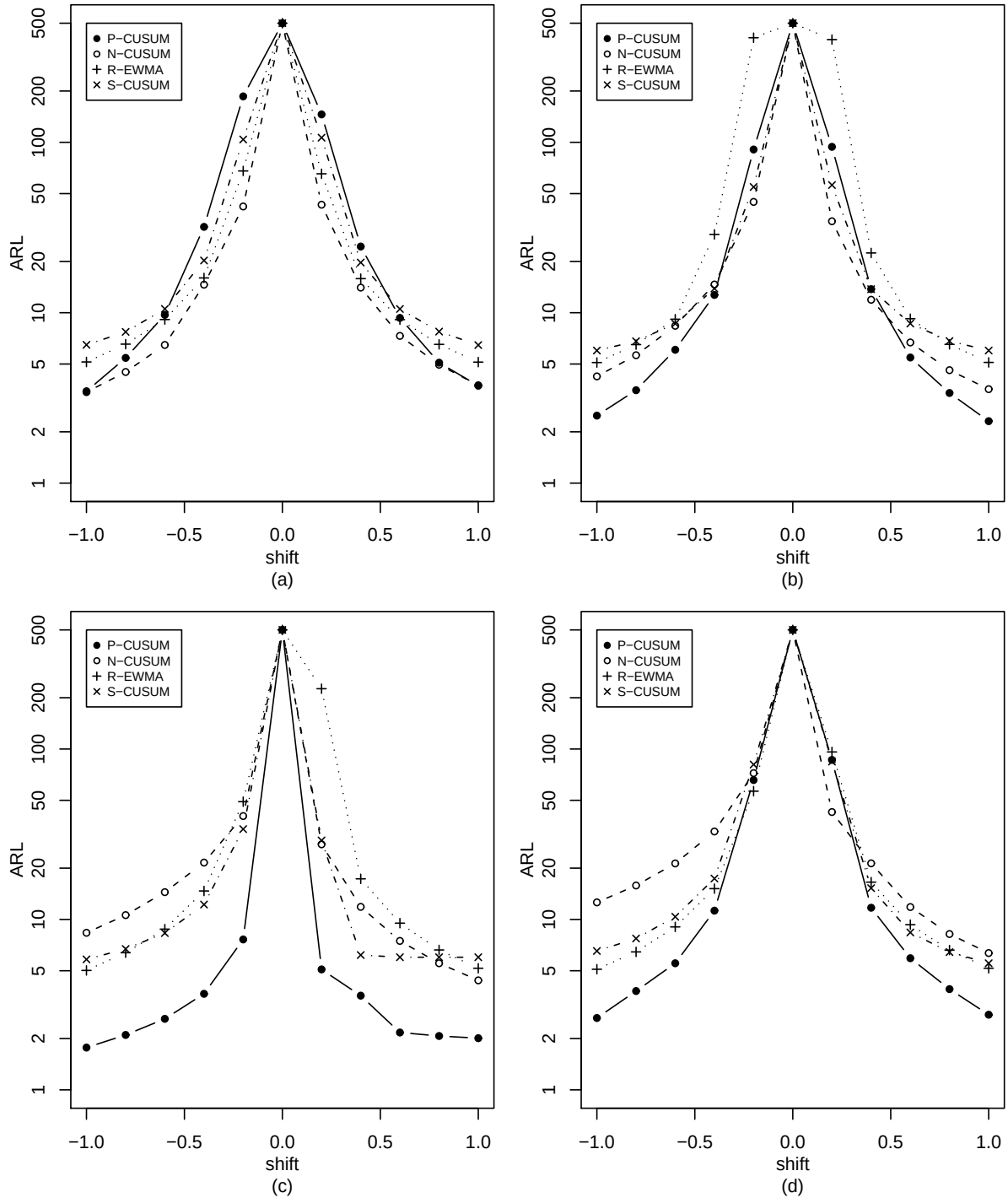

Figure S.6: Optimal OC ARL values of four control charts when the IC ARL is 500, $M = 200$, $m = 5$, and the actual IC process distribution is the standardized version of $N(0, 1)$ (plot (a)), $t(4)$ (plot (b)), $\chi^2(1)$ (plot (c)), and $\chi^2(4)$ (plot (d)). Procedure parameters of the control charts are chosen to be the ones that minimize their OC ARL values when detecting each individual shift. Scale on the y-axis is in natural logarithm.

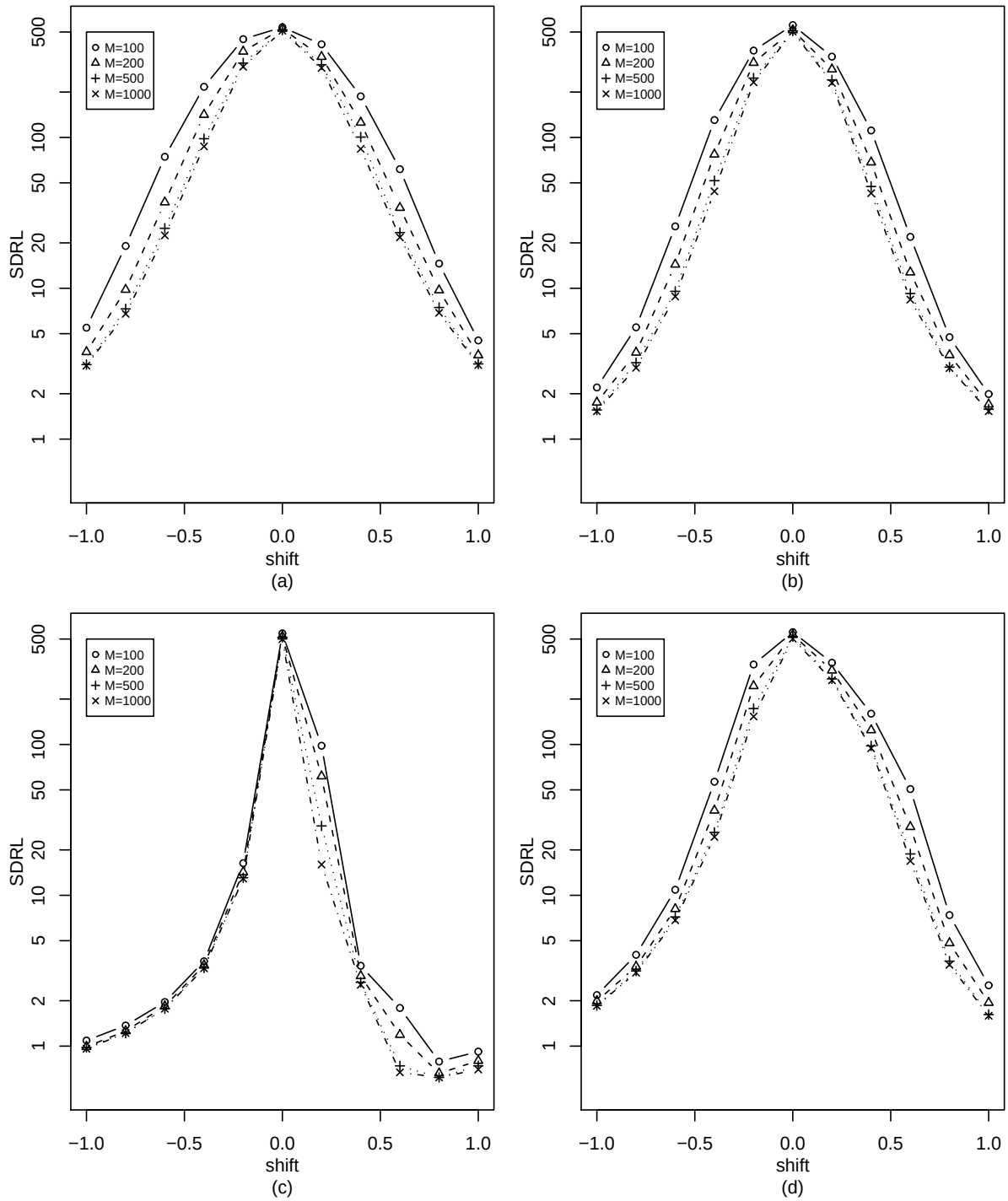

Figure S.7: Sample standard deviation values of the run length distribution of the P-CUSUM chart, computed from 10,000 replicated simulations, in cases considered in Figure 5 of the printed paper. Scale on the y-axis is in natural logarithm.

Table S.1: Values of $E_2(\widetilde{X}^2(n))$, $E_{10}(\widetilde{X}^2(n))$, and $RD = (E_{10}(\widetilde{X}^2(n)) - E_2(\widetilde{X}^2(n)))/E_2(\widetilde{X}^2(n)))$ in cases when the IC distribution is $N(0,1)$, the mean shift is 0.4, $p = 2$ or 10, and m changes among 1, 5, 10, and 20.

m	$E_2(\widetilde{X}^2(n))$	$E_{10}(\widetilde{X}^2(n))$	RD
1	1.000	9.000	8.000
5	1.386	9.638	5.954
10	1.870	10.434	4.580
20	2.836	12.028	3.241

guideline (i) can be explained intuitively as follows. Assume that the process is OC at the n-th time point. Then, it can be checked that

$$E_p\left(\widetilde{X}^2(n)\right) = \sum_{l=1}^{p} \frac{f_l^{(1)}(1-f_l^{(1)}) + m(f_l^{(1)} - f_l^{(0)})^2}{f_l^{(0)}},$$

where E_p denotes the expectation when p categories are used in the P-CUSUM chart, $\widetilde{X}^2(n)$ is the Pearson's chi-square statistic on which the P-CUSUM chart is based (cf., its definition in Section 2 of the paper), and $\mathbf{f}^{(0)} = (f_1^{(0)}, f_2^{(0)}, \ldots, f_p^{(0)})'$ and $\mathbf{f}^{(1)} = (f_1^{(1)}, f_2^{(1)}, \ldots, f_p^{(1)})'$ are the IC and OC distributions of $\mathbf{Y}_j(n)$. From the above expression, when m is larger, there is indeed more information in $\widetilde{X}^2(n)$ about the mean shift. In the case when the IC distribution is $N(0,1)$ and the mean shift is 0.4, $\mathbf{f}^{(0)} = (0.5, 0.5)'$ and $\mathbf{f}^{(1)} = (0.345, 0.655)'$ when $p = 2$, and

$$\mathbf{f}^{(0)} = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)'$$

$$\mathbf{f}^{(1)} = (0.046, 0.061, 0.071, 0.079, 0.088, 0.097, 0.108, 0.121, 0.140, 0.189)'$$

when $p = 10$. In such cases, when m changes among 1, 5, 10, and 20, $E_2(\widetilde{X}^2(n))$, $E_{10}(\widetilde{X}^2(n))$, and their relative difference $RD = (E_{10}(\widetilde{X}^2(n)) - E_2(\widetilde{X}^2(n)))/E_2(\widetilde{X}^2(n)))$ are listed in Table S.1. From the table, it can be seen that, when m increases, the relative difference between $E_2(\widetilde{X}^2(n))$ and $E_{10}(\widetilde{X}^2(n))$ gets smaller. This explains why the OC ARL values of the P-CUSUM chart when $p = 2$ get closer to those when $p = 10$ as m increases, as seen in Figures 6 and 7 of the printed paper. The guidelines (ii) and (iii) imply that selection of p also depends on the shape of the process distribution, which can be demonstrated by the calculation shown in Table S.2. In the table, we consider two IC distributions: $N(0,1)$ and the standardized version of $\chi^2(1)$, as in Figures 6 and 7 of the printed paper. We also assume that the mean shift in the original process observations $\mathbf{X}(n)$ is 0.2, and p is 2 or 5. The table presents the IC distribution $\mathbf{f}^{(0)}$, the OC distribution $\mathbf{f}^{(1)}$, and the relative distance between $\mathbf{f}^{(0)}$ and $\mathbf{f}^{(1)}$, defined by $Q = \sum_{l=1}^{p}(f_l^{(1)} - f_l^{(0)})^2/f_l^{(0)}$, of the

Table S.2: The IC distribution $\mathbf{f}^{(0)}$, the OC distribution $\mathbf{f}^{(1)}$, and the relative distance between them, defined by $Q = \sum_{l=1}^{p}(f_l^{(1)} - f_l^{(0)})^2/f_l^{(0)}$, of the categorized data $\mathbf{Y}_j(n)$ when the IC distribution is $N(0,1)$ or the standardized version of $\chi^2(1)$, and p is 2 or 5.

IC Distribution	p	$\mathbf{f}^{(0)}$	$\mathbf{f}^{(1)}$	Q
$N(0,1)$	2	(0.5,0.5)'	(0.421,0.579)'	0.025
	5	(0.2,0.2,0.2,0.2,0.2)'	(0.149,0.716,0.196,0.218,0.261)'	0.036
$\chi^2(1)$	2	(0.5,0.5)'	(0.322,0.678)'	0.127
	5	(0.2,0.2,0.2,0.2,0.2)'	(0,0,0.486,0.271,0.243)'	0.843

categorized data $\mathbf{Y}_j(n)$. It can be seen from the table that, when the IC distribution is $N(0,1)$ which is symmetric, Q changes its value from 0.025 to 0.036 when p increases from 2 to 5. However, when the IC distribution is the standardized version of $\chi^2(1)$ which is skewed, Q changes its value from 0.127 to 0.843. The increase in the Q value is about 7 times in the latter case. Therefore, the effect of p on the P-CUSUM chart does depend on the shape of the IC process distribution. By similar arguments, we can also explain why the P-CUSUM chart with $p = 2$ performs better than the chart with $p = 3$ in some cases shown in Figure 7 of the printed paper. For instance, in the case when the IC process distribution is the standardized version of $\chi^2(1)$ (cf., Figure 7(c)), $p = 2$, and the mean shift in the original process observations $\mathbf{X}(n)$ is 0.4, the IC distribution $\mathbf{f}^{(0)}$ and the OC distribution $\mathbf{f}^{(1)}$ of the categorized data $\mathbf{Y}_j(n)$ are $(0.5, 0.5)'$ and $(0, 1)'$, respectively, and the resulting Q value is 1.0. After p increases from 2 to 3 and other parameters remain unchanged, $\mathbf{f}^{(0)}$ and $\mathbf{f}^{(1)}$ become $(1/3, 1/3, 1/3)'$ and $(0, 0.457, 0.543)'$. The resulting Q value is 0.511 in this case, which is much smaller than the Q value when $p = 2$. Therefore, the shift would be relatively easier to detect by the P-CUSUM chart with $p = 2$. By the way, the above justifications with Tables S.1 and S.2 are all based on theoretical results. In applications, observed values of $\widetilde{X}^2(n)$ would be different from its mean values presented in Table S.1. Similarly, although $\mathbf{f}^{(0)}$ presented in Table S.2 have equal probabilities for each group, its estimator $\widehat{\mathbf{f}}^{(0)}$ obtained from an IC data would not have this property due to estimation error.