On the Comparison of Run Orders of Unreplicated 2^{k-p}-designs in the presence of a time-trend

Kayode S. Adekeye and **Joachim Kunert**

Fachbereich Statistik

Universität Dortmund

Dortmund, Germany.

Abstract

The response from a factorial experiment carried out in a time sequence may be affected by uncontrollable variables that are highly correlated with the time in which they occur. In such a situation, one possibility is to randomize the run order of the experiment. Another possibility is to use a systematic run order that is robust against time-trends. Since randomized run orders make the time trend part of the error, it can be hoped that systematic run orders will be more effective to identify truly active factors. In this paper, a simulation study is used to compare the performances of the randomized and the systematic run orders. The response from an experiment where we have observed a strong time-trend is used to demonstrate the influence of a realistic time trend on the run orders under consideration. The performance of the run orders is then measured by taking the probabilities of false rejection and the probabilities of detection of active contrasts. Our results show that the randomized run order managed to keep the nominal level, while the systematic did not. Additionally, when there were active factors, then the systematic run orders did not achieve more power than did the randomized run order.

Key words: Active contrasts, Probability of false rejection, Probability of effect detection.

1 Introduction

When engineers or physical scientist perform factorial or fractional factorial experiments, they usually have one machine or one pilot plant and they are therefore compelled to conduct their experimental runs in a sequence over time. However, then the observations may be affected by a trend. For instance, the response from the experiment may be affected by uncontrollable variables that are highly correlated with the time in which they occur (see, e.g. Bailey, Cheng and Kipnis, 1992). In such a situation, run orders are usually randomized before the experiment is performed. However, any particular random run order may or may not be adequate and hence randomization may lead to run orders where the estimates of factor effects of interest are adversely affected by the presence of trend. Therefore, a systematic run order in the presence of time-trend may improve the efficiency with which factor effects are estimated. It is therefore pertinent to consider systematic run orders in which the estimates for factor effects of interest are trend resistant.

On the other hand, there are authors who do not even accept that randomizing the run order of a factorial design is a useful precaution against time trends, see e.g. de León, Grima and Tort-Martorell (2003). It is not clear that the randomization argument really works for saturated fractional factorial designs: each design with n runs has $n-1$ contrasts (i.e. main effects or interactions) that may become influenced by the time trend. Note that there are only $(n-1)n/2$ possible run orders for each column. So there must always be some columns of the design that are heavily influenced by the time trend.

In this paper, three possible run orders of a fractional factorial design are considered. These are the standard, the randomized and a systematic run order for a non-replicated 2^{k-p} experiment, where $k-p=4$ and therefore the number of runs equals 16. This allows to estimate 15 main effects or interactions. The analysis was done with a simple version of the half-normal plot (Daniel, 1959), where we have used 1.5 times the median of the absolute values of the contrast estimates as an estimate of the variance. This estimate was called $\hat{\sigma}_M^2$ by Kunert (1997).

There are two objectives of our study. We firstly want to compare the performance of the run orders when there are no active contrasts. A contrast is said to be active if it has a true effect on the behavior of the response. The second objective is to determine the power with which truly active contrasts are identified. The performance of the three types of run order are measured in a simulation study by identification of **probability of false rejection** of non-active contrasts and **probability of effect detection** of active contrasts. The response that we considered for the study, is based on two data sets produced in a physical experiment that we normally use in our courses on experimental design to demonstrate the problem of time trends (see also Toutenburg, Gössl and Kunert, 1998, p. 99 - 104). The variable of interest is the run time of a ball-bearing in a funnel. This increases over time, even if the experimental conditions are left unchanged. For details about the present experiment see Adekeye (2004).

2 Simulation Study

The purpose of the simulation study is to compare the behavior of a randomized run order with a systematic order (which is a linear trend resistant design) and the standard run order of an unreplicated fractional factorial designs. For each of the run orders under consideration, we based our simulation study on 10,000 simulations of a 16 run experiment. In our study, a design identified some contrasts as active whenever the largest of the absolute values of the test statistics was greater than a given (simulated) critical value at a desired α level of significance. The proportion of simulated designs with false rejection is an estimate of the probability of false rejection (**PFR**) while the proportion of simulated designs with correct detection of an active effect estimates the probability of effect detection (**PED**).

The general approach used in the simulation study for comparing the run orders is as follows. (1) When we assume that none of the contrasts is active, the original experimental data is used to estimate the contrast effects of the run order. (2) If a contrast is assumed to be active, it is used to modify the data.

Let y_j, where $j = 1, 2, ...n$ and $n = 2^{k-p}$, denote the response from a 2^{k-p} experiment. Further, let m be a constant. To modify the data, the constant m is added to the experimental response y_j for all runs j where the active factor i is at the high level. More precisely,

$$y_j^{(m)} = \begin{cases} y_j + m, & \text{if factor } i \text{ is at the high level } (+) \text{ in run } j \\ y_j, & \text{if factor } i \text{ is at the low level } (-) \text{ in run } j \end{cases} \quad (1)$$

where $i \in \{1, 2, ..., n-1\}$ represents the factor used to modify the data. If we assume two or more factors to be active, then $y_j^{(m)}$ is derived by adding m for each of the active factors that is at level (+) in the run j. The vector with entries $y_j^{(m)}$, $1 \leq j \leq n$, is called modified observations.

Our analysis is based on an easy formal version of the half normal plot. The test statistic to test whether factor i is active if there are b, say, contrasts in all, is given by

$$t_i = \frac{\hat{\beta}_i}{\hat{\sigma}}, 1 \leq i \leq b. \quad (2)$$

where $\hat{\sigma}$ is the estimate of the common standard deviation σ of the contrasts. In this study, the statistic used to estimate σ is based on the median of absolute contrasts. The estimate is given by

$$\hat{\sigma} = 1.5(median|\hat{\beta}_i|). \quad (3)$$

We say that there is an active factor in the design if the largest of the $|t_i|$ is larger than a critical value that depends on the number b of contrasts considered.

2.1 Probabilities of False Rejection and Effect Detection

The probability of false rejection (PFR) describes the proportion of designs that falsely declare the presence of an active contrast. This is synonymous with the type

1 error i.e. the level of the test. A test is valid, if the true probability of false rejection is not larger than the nominal PFR α.

The probability of effect detection (PED) measures the proportion of designs that rightly declare presence of active contrasts. If there are active contrasts, it is the probability of making a correct decision i.e. the power of the test.

The probabilities of false rejection and effect detection are, therefore, estimated by the proportion of designs in the study with $max(|t_i|) > C(b, \alpha)$, where $C(b, \alpha)$ is an appropriate critical value which depends on the number b of contrasts plotted in the half normal plot and on the desired α level of significance.

Hence, the probabilities of false rejection and effect detection of active contrasts are

$$PFR = pr(max|t_i| > C(b,\ \alpha)),$$

when $m = 0$ is used to modify the data, and

$$PED = pr(max|t_i| > C(b,\ \alpha)),$$

when $m > 0$ is used to modify the data.

We used two approaches to estimate the probabilities of false rejection and effect detection of active contrasts for the three run orders under study. The first approach could only be used for the randomized run order, while the second approach could be used for the standard and systematic run orders as well. In the first approach, the experimental data will be used directly and the test-statistics are calculated from several realizations of the randomized ordering. In the second approach, artificial data are generated from a time series model that is derived from the experimental data.

2.1.1 Probability of False Rejection (PFR)

The two approaches that we used to determine the probability of false rejection are called Approaches **1** and **2**. The first approach (**Approach 1**) will be used for the randomized run order while the second approach (**Approach 2**) is used for the standard and the systematic run order.

Approach 1: PFR for randomized run order

In this approach, we created 10,000 artificial designs by permuting the rows of the model design matrix in standard run order. For each of the permuted designs, we used the same vector of responses. Using this fixed vector, the estimates for the contrasts ($\hat{\beta}_i$) are computed for each randomized run order. The PFR is then estimated by comparing the maximum of the absolute test statistics given in Equation (2) with $C(b, \alpha)$ for each design. The algorithm of the procedure is as follows:

(i) Permute the rows of the model design matrix.

(ii) For the permuted model design matrix, obtain the estimate of the contrasts using the experimental results in the original order.

(iii) Compute the test statistics using Equation (2).

(iv) Determine the maximum of the absolute test statistics ($max|t_i|$) obtained in (iii).

(v) Repeat step (i) to (iv), 10,000 times

(vi) Determine the proportion of designs with the maximum in step (iv) greater than $C(b, \alpha)$.

This proportion estimates the PFR for the randomized run order.

Permuting the rows of either the standard or the systematic run order will result in a loss of the features of both run orders. Therefore, the above algorithm can only be used for the randomized run ordering. Hence we need an alternative approach to determine their PFR, such that the features of both the standard and systematic run orders will be retained.

The alternative approach, **Approach 2**, uses an artificial set of data generated from a model fitted to the experimental data. Therefore, before **Approach 2** can be employed, we need to first fit an appropriate model to the experimental data set.

Approach 2: PFR for systematic and standard run orders

In this approach, sets of data will be generated from an appropriate model fitted to the original data set. Here, the response is not assumed to be fixed, instead we assume that the model design matrix is fixed. This is to protect the trend resistance property of the run order for the design under study. Using the data generated from the fitted model, we then estimate the effects of the contrasts. The PFR is taken to be the proportion of designs with maximum absolute test statistic that is greater than the simulated critical value for a desired α level.

The following steps give the algorithm for **Approach 2**:

(i) Generate a set of data from the time series model fitted to the experimental results.

(ii) Obtain the estimates of the contrasts from the model design matrix using the generated data in step (i) and the run order under consideration.

(iii) Compute the test statistics using Equation (2).

(iv) Determine the maximum of the absolute test statistics ($max|t_i|$) obtained in (iii).

(v) Repeat step (i) to (iv), 10,000 times.

(vii) Determine the proportion of designs with the maximum in step (iv) greater than $C(b, \alpha)$. This proportion estimates the PFR for the run order under study.

It should be noted that both Approaches **1** and **2** have the same fundamental objective. That is, to estimate the proportion of designs with maximum absolute test statistics greater than the critical value when there is no active contrast. By implication, this is the proportion of designs that falsely declare an active contrast.

2.1.2 Probability of Effect Detection (PED)

To determine the PED for the standard, randomized and systematic run orders, we implored the modified data approach by using the setting of the columns in the model design matrix to get a new data set. In this section, again two approaches were used. We called them approaches **A** and **B**. Approach **A** is for the randomized run ordering only while Approach **B** is for the standard and systematic run orders as well. The procedures for the two approaches are similar to those presented in Section 2.1.1 with the introduction of an additional step after step (i) of the algorithm. The additional step is to modify the data using Equation (1). In Approach **A**, the experimental results were modified by adding some constant m whenever the corresponding column setting in the model design matrix is at its high level (+) and zero when the setting is at its low level (-), where $m \geq 1$. The new data set (that is, $y_j^{(m)}$) along with the model design matrix of the run order under study is then used to get the proportion of designs that rightly declare active contrasts. In Approach **B**, the data generated from the model fitted to the original data are used along with the run orders under consideration to obtain the PED. An algorithmic description of the two approaches in steps is presented as follows:

Approach A: PED for randomized run ordering

(i) Permute the rows of the model design matrix.

(ii) For each factor of the model design matrix assumed to be active, modify the data using Equation (1).

(iii) For the permuted model design matrix, determine the estimates of the contrasts ($\hat{\beta}_i$) using the new data in (ii).

(iv) Compute the half normal plot test statistic using Equation (2).

(v) Determine the maximum of the absolute test statistics ($max|t_i|$) obtained in (iv).

(vi) Repeat step (i) to (v), 10,000 times.

(viii) Determine the proportion of designs with the maximum in step (v) greater than $C(b, \alpha)$. This is taken to be the PED for the randomized run order.

Approach B: PED for standard and systematic run orderings

(i) Generate a set of data from the model fitted to the series of the experimental results.

(ii) Use the active factors to modify the generated data using Equation (1).

(iii) Obtain the estimates of the contrasts ($\hat{\beta}_i$) using the modified generated data in (ii).

(iv) Compute the half normal plot test statistics t_i using Equation (2).

(v) Determine the maximum of the absolute test statistics ($max|t_i|$).

(vi) Repeat step (i) to (v), 10,000 times.

(viii) Determine the proportion of designs with the maximum in step (v) greater than $C(b, \alpha)$. This is taken to be the PED for the run order under study.

3 Data and Designs used

We now use the algorithms presented in Section 2 to determine the performance of the three run orders under study. The original data are the run times from two repetitions of 16 runs from a funnel experiment. The funnel experiment was introduced by Gunther (1993) to serve as a tool for teaching experimental design. A ball bearing is made to rotate in a funnel. The time that it runs before it falls out of the bottom of the funnel is measured. It was observed by Toutenburg, Gössl and Kunert (1998) that there is a time trend in this experiment. Even with a constant setting of the factors influencing the speed of the ball bearing, the run times increase over time. For more details, see Adekeye (2004).

As model data to check the performance of the different run orders, we used the run times of 16 consecutive experiments, all with the same setting. This was repeated at another occasion, with another 16 consecutive runs with a slightly varied setting of the factors. The data sets are presented in Table 1 and called example 1 and example 2. They are observations of a true time trend. In both experiments the time trend is clearly visible: The run times get considerably larger when the ball bearing has run several times.

The model matrix for the standard ordering and the systematic run order (which is a linear time trend resistant design) are presented in Tables 2 and 3, respectively.

Table 1: Run times from two repetitions of the funnel experiment with 16 consecutive runs each

Example 1	22.13	23.49	23.32	24.26	23.70	23.92	24.07	24.09
	25.06	25.36	24.32	24.97	25.03	26.09	25.40	26.02
Example 2	21.35	21.36	22.31	21.98	23.07	23.29	22.89	23.71
	23.18	23.73	24.30	23.30	23.68	23.49	23.51	24.19

Note that the trend resistant design does not allow estimation of some two-factor interactions, the corresponding columns are confounded with the time trend. These two factor interactions are therefore not estimable.

Table 2: Standard run orders for 2^{5-1} fractional factorial design

Run No.	Contrast														
	Main factor					Two factor interactions									
	A	B	C	D	E	AB	AC	BC	DE	AD	BD	CE	CD	BE	AE
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	+	+	+	-	-	+	+	+	+	-	-	-	-	-	-
3	+	+	-	+	-	+	-	-	-	+	+	+	-	-	-
4	+	+	-	-	+	+	-	-	-	-	-	-	+	+	+
5	+	-	+	+	-	-	+	-	-	+	-	-	+	+	-
6	+	-	+	-	+	-	+	-	-	-	+	+	-	-	+
7	+	-	-	+	+	-	-	+	+	+	-	-	-	-	+
8	+	-	-	-	-	-	-	+	+	-	+	+	+	+	-
9	-	+	+	+	-	-	-	+	-	-	+	-	+	-	+
10	-	+	+	-	+	-	-	+	-	+	-	+	-	+	-
11	-	+	-	+	+	-	+	-	+	-	+	-	-	+	-
12	-	+	-	-	-	-	+	-	+	+	-	+	+	-	+
13	-	-	+	+	+	+	-	-	+	-	-	+	+	-	-
14	-	-	+	-	-	+	-	-	+	+	+	-	-	+	+
15	-	-	-	+	-	+	+	+	-	-	-	+	-	+	+
16	-	-	-	-	+	+	+	+	-	+	+	-	+	-	-

Table 3: Linear time-trend resistant 2^{5-1} fractional factorial design

Run No.	Contrast										
	Main factor					Two-factor interactions					
	A	B	C	D	E	BC	BD	BE	CD	CE	DE
1	+	+	+	+	+	+	+	+	+	+	+
2	-	+	-	-	-	-	-	-	+	+	+
3	-	-	+	-	-	-	+	+	-	-	+
4	+	-	-	+	+	+	-	-	-	-	+
5	-	-	-	+	-	+	-	+	-	+	-
6	+	-	+	-	+	-	+	-	-	+	-
7	+	+	-	-	+	-	-	+	+	-	-
8	-	+	+	+	-	+	+	-	+	-	-
9	-	-	-	-	+	+	+	-	+	-	-
10	+	-	+	+	-	-	-	+	+	-	-
11	+	+	-	+	-	-	+	-	-	+	-
12	-	+	+	-	+	+	-	+	-	+	-
13	+	+	+	-	-	+	-	-	-	-	+
14	-	+	-	+	+	-	+	+	-	-	+
15	-	-	+	+	+	-	-	-	+	+	+
16	+	-	-	-	-	+	+	+	+	+	+

The design with the systematic run order in Table 3 is linear trend free. For the construction, we used methods put forward by Cheng and Jacroux (1988). Their Lemma 3.1 shows that at most 11 columns of a 16-run factorial design can be linear trend free. Their Lemma 3.2 shows how we can find these 11 columns. Using Cheng and Jacroux (1988) algorithm 3.3, we could place 4 factors in such a way that their 4 main effects and all 6 two-factor interactions between them are linear trend free. This leaves one other trend free column. If we identify a fifth factor with the 11th trend free column, namely the four-factor interaction of the other four factors, then the main effect of this fifth factor (factor A in Table 3) is also linear trend free. All two factor interactions of this factor, however, are confounded with the linear trend. We therefore have only 11 contrasts that can be used for the half normal plot.

3.1 Results for Example 1

3.1.1 Estimation of Probability of False Rejection (PFR)

To obtain the proportion of false rejections for the randomized run order, we used the design in Table 2 with $b = 15$ estimable contrasts as a starting point of the randomization. For $b = 15$ and $\alpha = 0.05$, we used a simulated critical value ($C(15, 0.05)$) of 3.70. This was derived from 10,000 simulations with 16 i.i.d. normally distributed observations each (Adekeye, 2004). Therefore, the proportion of simulated designs with $max|t_i| > 3.70$ will estimate the PFR. Permuting the rows of the model design matrix in Table 2, and using the experimental result presented in Table 1 as Example 1, we estimate the contrast effects and follow the algorithm as stated in **Approach 1**. With 10,000 repetitions, the observed proportion of designs where $max|t_i| > 3.70$ was 4.7%. We therefore estimate that indeed approximately 5 % of the randomized run orders will falsely give an active contrast. Note that this is already an important result. It supports the view that randomized orderings will keep the nominal level α.

To determine the PFR and PED for the standard and the systematic run order, we need to generate a large number of data sets from a fitted model. For this data set an $ARIMA(0, 1, 1)$ was fitted to the experimental data. For details about ARIMA modelling, see e.g. Box, Jenkins, and Reinsel (1994) or Pankratz (1983).

In order to obtain the proportion of false rejections for the standard run order, we used the design in Table 2. For this design, $b = 15$ and for $\alpha = 0.05$ we have the same critical values as before. Then 10,000 artificial data sets were generated from the fitted $ARIMA(0, 1, 1)$ model. For details of the fitted model, see Adekeye (2004). Following the algorithm of Approach 2, the observed proportion of simulated designs with $max|t_i| > 3.70$ equals 32.8%. Hence, almost one third of the simulated observations, the standard run order falsely identified an active contrast. This is much too much and the standard run order is clearly not usable in the presence of this time trend.

To obtain the proportion of false rejections for the systematic run order, we used the design in Table 3. For this design the number of contrasts is $b = 11$. Therefore, we had to use another critical value. The simulated critical value ($C(11, 0.05)$) is 3.72 (Adekeye, 2004). Hence, the proportion of simulated designs with $max|t_i| > 3.72$ will estimate the PFR for the systematic run order. Following the algorithm of **Approach 2**, the proportion of simulated designs with $max|t_i| > 3.72$ is 10.8 %. This means that in approximately 10% of the simulated data-sets, the systematic run order falsely identified an active contrast. Hence, the systematic run order performed better than the standard ordering. However, it seems that the systematic run order does not suffice to provide sufficient protection against the realistic trend considered here, which is not just linear.

A plot that displays the empirical distribution function of $max|t_i|$ for the systematic, standard, and randomized run order is presented in Figure 1.

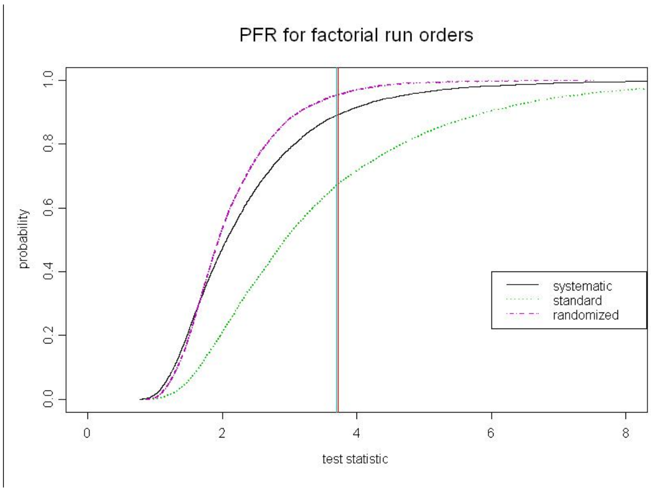


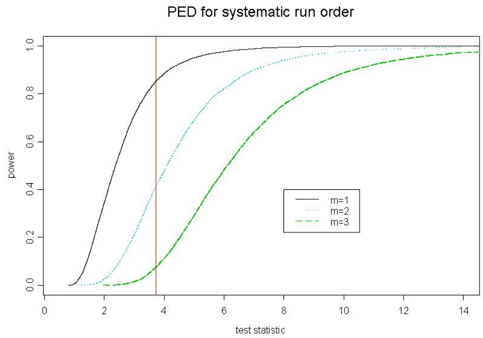
Figure 1: *Empirical distribution function for systematic, standard, and randomized run orders without active contrasts. The (two) vertical lines in the figure represent the critical values. Note that they are very near to each other.*

3.1.2 Estimation of Probability of Effect Detection

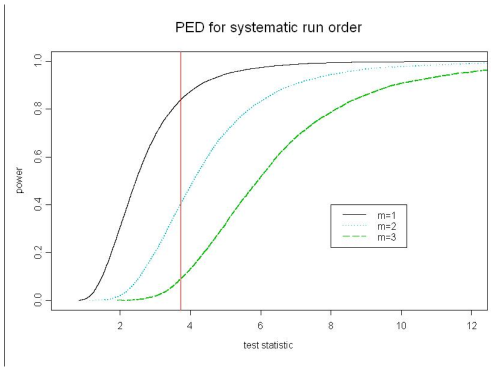
The procedure of **Approach A** of Section 2.1.2 was used to compute the PED for the randomized run order. Again, the rows of the design in Table 2 were permuted 10,000 times. This implies 10,000 randomized run orderings. For each time that the rows of the design were permuted, we used the setting of the first column of the run order to modify the data. This simulated an effect of factor 1. Then the modified data set was used to obtain estimates for the contrasts of all factors and two factor interactions. When adding $m = 1$ to the data, we observed a PED of 7.3%. When adding $m = 2$, we obtained 36%. Similarly by adding $m = 3$, we obtained 86.2 %. We then continued by assuming two or three active contrasts. For simplicity, we assumed that the active contrasts all were of the same size. The results can be seen in the RO column of Example 1 in Table 4. It turned out that the results for two and three active contrasts did not differ much from the results for one active contrast.

The probability of effect detection (PED) for the standard run order and the systematic run order was estimated using the designs in Tables 2 and 3, respectively. Following the algorithm of **Approach B**, we generated data sets using an $ARIMA(0, 1, 1)$ model. The generated data sets were then modified and the modified data were used to determine the proportion of designs with one, two, and three active contrasts that correctly identified at least one contrast as active. For one active contrast, we first add $m = 1$ to the generated data when the first column of the model design matrix under study is at + level and zero otherwise. The new data set is then used to get the maximum of the absolute half normal test statistics which is compared with the simulated critical value 3.70 for the standard run order and 3.72 for the systematic run order. The steps are repeated for $m = 2$ and $m = 3$. For two active contrasts, we add m to the generated data for each + in the first column and for each + in the second column of the model design matrix under study. We proceed similarly for three active contrasts. The empirical PED for one, two, and three active contrasts for both the standard and the systematic run orders are presented in Table 4, in the columns marked SO and Sys. In Figure 2 we also display the empirical distribution function for the systematic run order.

One active contrast



Two active contrasts



Three active contrasts

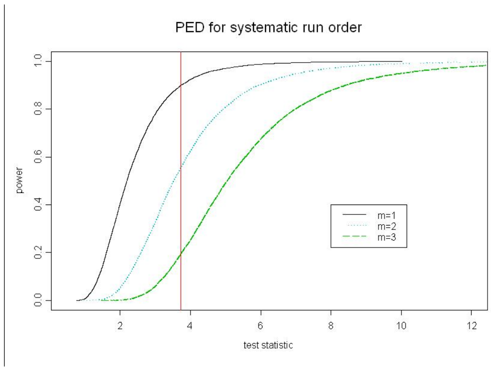


Figure 2: Empirical distribution function of $max|t_i|$ for the systematic run order with different sizes (m) of the active contrasts. The vertical line in each graph represents the critical value

In order to have a fair comparison of the randomized run order with the systematic run order, we did two more series of experiments. First of all, we also used the artificial data to estimate the PFR and PED for the randomized run order. That is, we used 10,000 data sets derived from the ARIMA(0,1,1) as in approach B, but did the modification and the analysis of each with a different random ordering of the design in Table 2 as in approach A. Since this is a combination of the two approaches, we call this the harmonized approach. Here, the results were very similar to what we had for the experimental data.

Since the systematic run order did not keep the nominal level $\alpha = 5\%$, it is easy for the systematic run order to get a high estimated power. To allow for a fair comparison between the randomized run order and the systematic run order, we did experiments with the systematic run order where a pseudo critical value was used. The pseudo critical value is taken to be the value of the $max|t_i|$ that gives the same proportion of false rejections as was derived by the randomized design. Thus, the pseudo critical value is data driven. Unfortunately, it cannot be determined in practice. To derive the same PFR of 0.0509 obtained with the harmonized approach for the randomized run order, we need a pseudo critical value of 4.60. This pseudo critical value is then used to compute the PED for the systematic run order. The corresponding proportions could be seen in Figure 2, if we moved the vertical line from 3.72 to 4.60.

The PFR and PED for the randomized run order using the harmonized approach and the PFR and PED for the systematic run order using the pseudo critical value are documented in Table 4 in the columns marked with an asterisk. They show that the systematic order does not achieve a higher power than the random ordering, once we corrected the critical value to keep the nominal level.

3.2 Example 2

3.2.1 Estimation of Probability of False Rejection

We now repeated all the analyses of Example 1 with another data set. The experimental data are the ones noted as Example 2 in Table 1.

The design in Table 2 was used to obtain the proportion of false rejection for the randomized run order. Permuting the rows of the model design matrix in Table 2 along with the experimental results of Example 2, we estimate the contrasts and follow the algorithm as stated in **Approach 1**. The obtained PFR value is 4.9%. This verifies once more that approximately 5% of the randomized run orders will falsely give an active contrast.

To evaluate the PFR and PED for the standard and systematic run orders, now data were generated from a $ARIMA(0, 1, 2)$ model that appeared to fit to the original data series. Following the algorithm of **Approach 2**, the observed proportion of simulated data sets with $max|t_i| > 3.70$ for the standard order is 55.5%. This indicates that with data of this kind the standard run order will falsely identify an active contrast more than 50 % of all cases. Similarly for the systematic run order, we used the design in Table 3. Following the algorithm of **Approach 2**, the observed proportion of simulated data sets with $max|t_i| > 3.72$ is 21.3%. This implies that the systematic run order will falsely give an active contrast in about 20% of all cases of data of this kind. Though this is smaller than for the standard order, this is much too large. The systematic run order does not provide sufficient protection against the kind of trend modelled in Example 2.

3.2.2 Estimation of Probability of Effect Detection

The procedure of **Approach A** of Section 2.1.2 was used to compute the PED for the randomized run order. Using a 0.05 level of significance with $b = 15$, adding $m = 1$ to the data whenever the setting of the first column of the permuted design is +, gives a PED of 9.8%, adding $m = 2$ gives 54%, and adding $m = 3$ gives 98.2%. Table 4 presents the obtained empirical PED for the randomized run order for one, two and three active contrasts.

The probability of effect detection (PED) for the standard and systematic run orders was estimated using the data generated from the $ARIMA(0, 1, 2)$ model. Following the algorithm of **Approach B**, we determine the PED for the run orders for one, two, and three active contrasts. The empirical PED is presented in Table 4.

The harmonized approach again yields a PFR of 5%, as it should. However, the PEDs in the column RO* are less than those obtained with Approach A. The pseudo critical value that we should use for the systematic order equals 5.86. This would give exactly the same estimated PFR as obtained with the harmonized approach for the randomized run order. Using this pseudo critical value, we compute the PED with different sizes of m for one, two and three active contrasts for the systematic run order. The results are presented in Table 4. They are not better than those for the harmonized approach.

Table 4: Summary of Empirical PFR and PED

Active		Example 1					Example 2				
Contrasts	m	SO	RO	RO*	Sys	Sys*	SO	RO	RO*	Sys	Sys*
	0	0.3283	0.0473	0.0509	0.1076	0.0509	0.5553	0.0494	0.0543	0.2130	0.0543
One	1	0.4168	0.0726	0.0944	0.1469	0.0692	0.5850	0.0979	0.0759	0.1919	0.0392
	2	0.6010	0.3587	0.3749	0.5864	0.3858	0.6519	0.5398	0.2293	0.5536	0.1699
	3	0.7829	0.8620	0.7056	0.9254	0.7841	0.7420	0.9824	0.4709	0.8764	0.4837
	10	1.00	1.00	0.9995	1.00	1.00	0.9970	1.00	0.9887	1.00	1.00
Two	1	0.4338	0.0829	0.1078	0.1601	0.0749	0.5988	0.1150	0.0798	0.2042	0.0355
	2	0.6834	0.4144	0.4159	0.5939	0.3721	0.7006	0.6069	0.2551	0.5356	0.1679
	3	0.8947	0.9013	0.7386	0.9096	0.7587	0.8264	0.9889	0.5059	0.8490	0.4421
	10	1.00	1.00	0.9998	1.00	1.00	0.9999	1.00	0.9905	1.00	0.9995
Three	1	0.3970	0.0778	0.102	0.1004	0.041	0.5486	0.1128	0.0724	0.1409	0.0625
	2	0.6878	0.3703	0.3828	0.4453	0.249	0.6750	0.5585	0.2368	0.4001	0.2313
	3	0.9268	0.8582	0.7051	0.8054	0.5998	0.8641	0.9753	0.4712	0.7230	0.5264
	10	1.00	1.00	0.9995	1.00	1.00	1.00	1.00	0.9880	1.00	1.00

SO ⇒ Standard run order, RO ⇒ Randomized run order, and Sys ⇒ Systematic run order. RO* represents the results obtained from the simulated data for the randomized run order (harmonized approach) and Sys* represents the results obtained by using the pseudo critical value for the systematic run order.

4 Conclusion

The results of the simulation study can be summarized as follows.

- When there are no active contrasts, the randomized run order managed to keep the nominal level. The systematic run order was nearer the nominal level than the standard run order, but both did not manage to keep the nominal level.

- When there are active contrasts, the power of the randomized run order was very similar wether we used the original experimental data or the simulated data from the fitted time series model. This lets us hope that the simulated data mimic the time trend reasonably well.

- When we adapted the critical value such that the systematic order kept the nominal level, then the power of the systematic order decreased considerably. In that case the power was no longer higher than the power derived from the randomized order.

In all, there was no advantage of the systematic run order visible in our study.

Acknowledgement

The research by K. Adekeye was sponsored by a scholarship of the NRW Graduate School of Production Engineering and Logistics at Universität Dortmund. Also, financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

[1] Adekeye, K. S. (2004). Experimental Design For Quality Improvement in The Presence of Time-Trends. Ph.D Thesis. Universität Dortmund. http://eldorado.uni-dortmund.de:8080/FB5/ls2/forschung/2004/Adekeye

[2] Bailey, R. A., Cheng, C. S. and Kipnis, P. (1992). Construction of trend - resistant factorial designs. Statistica Sinica, **2**, 393 – 411.

[3] Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). Time series analysis: Forecasting and control. Prentice Hall: New Jersey.

[4] Cheng, C. S. and Jacroux, M. (1988). The Construction of Trend-Free Run Orders of Two-Level Factorial Designs. Journal of the American Statistical Association, **83**, 1152 – 1158.

[5] Daniel, C. (1959). Use of half normal plots in interpreting factorial two level experiments. Technometrics, **1**, 311 – 341.

[6] de León, G., Grima, P. and Tort-Martorell, X. (2003). Experimental Order in Factorial Design. Paper presented at the 3rd ENBIS-Conference in Barcelona, 21 - 22 August, 2003. Abstract at:
http://www.enbis.org/barcelonaconference/abstracts.html♯114.

[7] Kunert, J. (1997). On the use of the factor-sparsity assumption to get an estimate of the variance in saturated designs. Technometrics, **39**, 81 – 90.

[8] Pankratz, A. (1983). Forecasting with univariate Box-Jenkins models: concepts and cases. John Wiley. New York.

[9] Toutenburg, H., Gössl, R. and Kunert, J. (1998). Quality Engineering. Eine Einführung in Taguchi-Methoden. Prentice Hall, München.