Analyzing ordinal data from a split-plot design in the presence of a random block effect

Heidi Arnouts[1] and Peter Goos[1,2]

*

†

[1]University of Antwerp, Belgium

[2]University of Leuven, Belgium

Abstract

Many industrial experiments involve some factors that are hard to change. In this situation, experimenters often choose to perform an experiment with restricted randomization, such as a split-plot or a strip-plot experiment. In this paper, we discuss the analysis of an experiment concerning the adhesion between steel tire cords and rubber. Besides an ordinal response, the experiment also involves one hard-to-change factor. Therefore, the experimenters performed a split-plot experiment. An additional complication of the experiment is that there is also a blocking factor. A proper analysis of the experiment requires the inclusion of random effects in the model to account for its split-plot nature and its blocked nature. The need for random effects and the ordinal response necessitate the use of a mixed cumulative logit model.

Keywords: split-plot design, random and fixed blocks, cumulative logit regression, generalized linear mixed model

1 Introduction

Many industrial experiments involve some factors that are hard to change. Due to time and/or budget constraints, in this situation, a completely randomized experiment is no longer desired and experimenters resort to experiments with restricted randomization,

*Dr. Arnouts is a guest lecturer at the Faculty of Applied Economics and lecturer at Odisee. Her email address is heidi.arnouts@uantwerpen.be

†Dr. Goos is Full Professor at the Faculty of Bioscience Engineering and the Leuven Statistics Research Centre (LStat) of the University of Leuven and at the Faculty of Applied Economics and StatUa Center for Statistics of the University of Antwerp. He is a Senior Member of the American Society for Quality. His email address is peter.goos@biw.kuleuven.be.

such as split-plot and strip-plot experiments.

In the last decades, a lot of research effort has been spent on the design and analysis of these types of experiments (see, for example, Arnouts, Goos and Jones (2010), Bingham, Schoen and Sitter (2004), Bingham and Sitter (1999), Goos and Gilmour (2012), Goos, Langhans and Vandebroek (2006), Goos and Vandebroek (2001, 2003, 2004), Jones and Goos (2007, 2012), Letsinger, Myers, and Lentner (1996), Mee and Bates (1998), Mylona, Goos and Jones (2014), Panigua-Quiñones and Box (2009), Parker, Kowalski and Vining (2006, 2007a, 2007b), Vivacqua and Bisgaard (2009), and Vining, Kowalski and Montgomery (2005)). Most of these papers assume that the response in the experiment is normally distributed, which implies the use of a linear mixed model to analyse the experimental data. However, in many industrial experiments, the response is categorical, which requires the use of another model. In situations with non-normal response data, it has become standard to use generalized linear mixed models (see, for example, Robinson, Myers and Montgomery (2004), Robinson et al. (2006), and Goos and Gilmour (2012)). In this article, we focus on an experiment involving an ordinal response. This kind of response has received virtually no attention in the split-plot literature. To a large extent, this is due to the fact that the ordinal scale is often ignored and the data are analysed using a linear mixed model, with which practitioners are more familiar. Ignoring the ordinal character of data and treating the data as though it were interval-scale data is a common mistake mentioned in many books on basic statistics when discussing the possible measurement scales of data and their consequences. In the case of ordinal data, it is basically incorrect to use a linear model in the responses. Due to the availability of modern software and powerful PCs, it is also no longer necessary to be restricted to linear models. It is possible now, in various software packages, to use more sophisticated models that adequately deal with the nature of ordinal responses.

In this paper, we discuss an experiment that studies the adhesion between steel tire cords and rubber. In the experiment, different types of steel cords are produced and subsequently vulcanized into a block of rubber. After that, the cords are pulled out of the rubber and the rubber coverage of the steel cords is rated visually on an ordinal, integer-valued, scale from 1 up to 7. Besides the ordinal response, the experiment also involves one hard-to-change factor which made the experimenters use a split-plot design. An additional complication of the experiment is the fact that there is also a blocking factor, i.e. the block of rubber in which the steel cords are vulcanized. A proper analysis of any split-plot experiment with an additional blocking factor requires the inclusion of random effects in the model. The various complications of the experiment suggest the use of a model such as a mixed cumulative logit model from the class of generalized linear mixed models.

The aim of this paper is to provide a tutorial on how to analyse data from this type of experiment. Therefore, the paper is constructed as follows. In the next section, we describe the steel tire cord experiment, the motivating example for this paper, in detail.

In Section 3, we present a diagram, called a Hasse diagram, to identify the structure of the experimental units. The cumulative logit model, which is used in this paper to deal with the ordinal response in the experiment, is discussed in Section 4. Finally, in Section 5, the data of the experiment are analysed.

2 The Steel Tire Cord Experiment

The steel tire cord experiment was conducted at a large company producing steel cords, mainly for the automotive industry. In this specific situation, the steel cords were used in the production of tires. Here, it is crucial that there is a strong adhesion of the steel cord to the rubber. More particularly, the adhesion of the steel cord to the rubber should be greater than the cohesion of the rubber.

A typical steel cord experiment involves the production of a roll of steel tire cord; a construction of several thin wires reinforcing the casing of the tire, while keeping it flexible enough to resist shocks and to improve comfort. The adhesion of steel cord to rubber is possible through a brass coating (typically, 70% copper, 30% zinc, with chemical additives) applied to the wire. The brass coating reacts chemically with the sulfur in the tire's rubber.

A typical adhesion experiment in this context studies various tire cords, various coating compositions (e.g., inorganic and organic phosphates and sulfur-containing rubber vulcanization accelerating agents) and thicknesses, and various ways to apply the coating. Commonly, various experimental runs are performed for each roll of steel tire cord, each with different coating compositions, thicknesses and/or application techniques. As a result, a typical steel tire cord experiment involves a split-plot design, where factors defining the steel cord are hard-to-change or whole-plot factors and factors defining the brass coating are easy-to-change or subplot factors.

The experiment we discuss in this article involved five experimental factors, which, for reasons of confidentiality, are denoted by A, B, C, D and E. Factor A is related to the steel tire cord production, whereas the factors B, C, D and E are related to the coating.

The experimenters were interested in the main effects and all two-factor interaction effects of the five factors on the adhesion of the steel cord to the rubber. When setting up the experiment, they took into account the fact that the first experimental factor, A, was hard to change and, therefore, they chose to perform a split-plot experiment. All five factors were studied at two levels. The available budget allowed them to perform 32 runs, i.e. to produce 32 different pieces of steel cord, for which the hard-to-change factor A was independently set eight times (in other words, eight rolls of steel tire cord were produced). The remaining four (easy-to-change) experimental factors were reset independently at each run, i.e. 32 times in total. As a result, the design utilized for the first phase of the experiment is a split-plot design involving eight whole plots of four runs. Figure 1 provides a graphical representation of the organization of the experiment. In the figure, the abbreviation WP refers to the whole plots of the experiment.

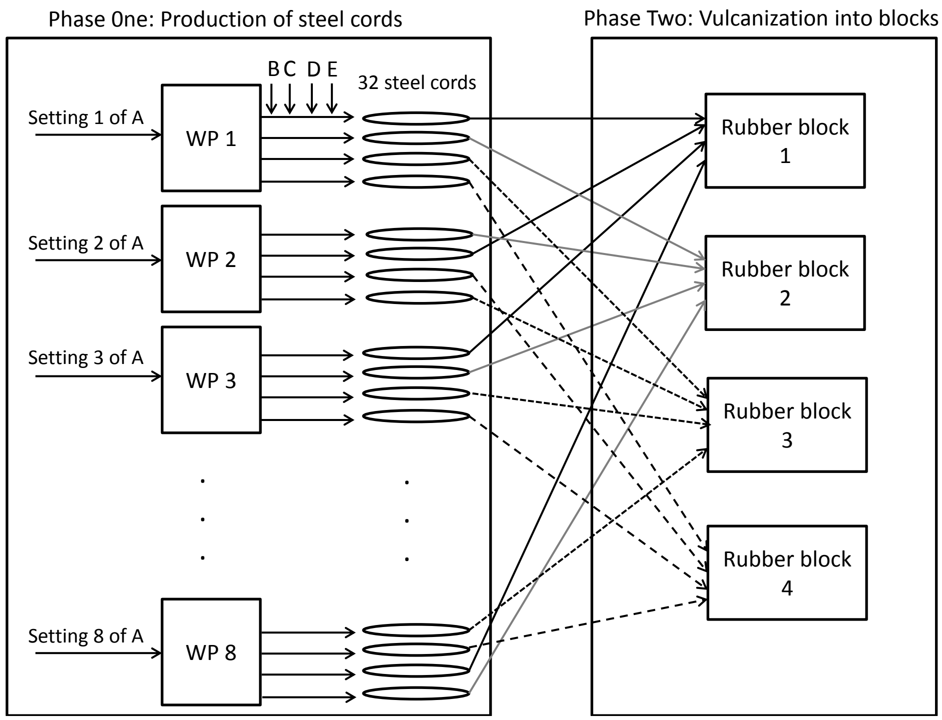

Figure 1: Graphical representation of the steel tire cord experiment.

In a next step of the experiment, the 32 produced steel cords were vulcanized into a block of rubber after which, the cords were pulled out of the blocks. The mold used to produce the rubber blocks is designed in such a way that it produced four test blocks with the appropriate dimensions. Each block could contain up to 15 steel cords. Given that the experiment involved 32 different steel cords, each of the four blocks of rubber eventually contained only eight steel cords. Because of the blocking, the adhesion qualities of the steel tire cords that were vulcanized into the same block were possibly correlated, whereas those of steel tire cords vulcanized into different blocks were not.

Finally, the adhesion quality was judged on the basis of the rubber coverage of the steel cords after they have been pulled out the rubber block. When the adhesion of the steel cord to the rubber was indeed greater than the cohesion of the rubber, the steel cord should still be covered by a large amount of rubber after being pulled out. The amount of rubber coverage was rated visually on an ordinal, integer-valued, scale from 1 to 7, with 7 representing the most desirable outcome. In this article, we show how to develop a model suitable for that ordinal response, taking into account the split-plot nature of the experiment and the additional blocking factor.

To conclude, the design configuration used for the steel tire cord experiment along with the ordinal responses is shown in Table 1. In this design configuration, the higher-order effects ABC, ADE and BCDE are confounded with the rubber blocks and the interaction effects BE and CDE together with the main effect of A are used to define the whole plots. It is possible to construct more a informative design configuration using an optimal experimental design methodology (for more information on optimal experimental design, see Goos and Jones (2011)), but this is not the scope of this paper.

3 Hasse diagram

To properly analyse the data from the experiment, it is crucial to identify the structure in the experimental units in order to determine the random effects necessary in the model. Goos and Gilmour (2012) suggest the use of Hasse diagrams to gain understanding of the structure of experimental units. Since the treatment structure in the steel tire cord experiment is clear, we only visualize the structure in the experimental units. A Hasse diagram is a simple graph in which the nodes represent the grouping factors (such as blocks and whole plots) involved in an experiment and the edges represent the nesting relationships between the grouping factors. The following rules hold when constructing Hasse diagrams:

- the Hasse diagram starts with a node called the *universe*, U, representing the entire experiment,
- if a grouping factor G_1 is nested within a grouping factor G_2, the node for G_1 appears below the node for G_2 with an edge connecting both nodes,
- if grouping factors G_1 and G_2 are crossed, their nodes appear at the same level,
- the Hasse diagram ends with the smallest experimental unit, the individual run.

Next to each node, two numbers are mentioned, i.e. the number of levels of the corresponding grouping factor and, between brackets, the corresponding degrees of freedom. The Hasse diagram for the steel tire cord experiment is shown in Figure 2. In this experiment, the whole plots are crossed with the rubber blocks, and the experimental runs are nested within the whole plots and the rubber blocks. The general recommendation is that each node in the Hasse diagram, except for the universe and the individual runs, implies a set of random effects in the model. This recommendation is generally accepted for models with normal responses. Although subject to debate, Gilmour (2011) provides a randomization-based argument for including a random effect for the individual runs too, in case the response is not normal. Following his recommendation and given the ordinal nature of the response, the model for the steel tire cord experiment should therefore contain three sets of random effects, i.e. one set for the whole plots, one set for the rubber

Table 1: Design configuration of the steel tire cord experiment and response

Run	Whole Plot	Rubber Block	A	B	C	D	E	Response
1	1	1	−1	−1	−1	−1	−1	2
2	1	2	−1	−1	1	1	−1	3
3	1	3	−1	1	1	−1	1	4
4	1	4	−1	1	−1	1	1	5
5	2	2	1	−1	−1	−1	−1	4
6	2	1	1	−1	1	1	−1	3
7	2	4	1	1	1	−1	1	2
8	2	3	1	1	−1	1	1	3
9	3	1	−1	1	1	−1	−1	4
10	3	2	−1	1	−1	1	−1	5
11	3	3	−1	−1	−1	−1	1	4
12	3	4	−1	−1	1	1	1	3
13	4	2	1	1	1	−1	−1	2
14	4	1	1	1	−1	1	−1	3
15	4	4	1	−1	−1	−1	1	1
16	4	3	1	−1	1	1	1	5
17	5	4	−1	1	−1	−1	−1	4
18	5	3	−1	1	1	1	−1	3
19	5	2	−1	−1	1	−1	1	6
20	5	1	−1	−1	−1	1	1	3
21	6	3	1	1	−1	−1	−1	4
22	6	4	1	1	1	1	−1	5
23	6	1	1	−1	1	−1	1	6
24	6	2	1	−1	−1	1	1	3
25	7	4	−1	−1	1	−1	−1	2
26	7	3	−1	−1	−1	1	−1	3
27	7	2	−1	1	−1	−1	1	4
28	7	1	−1	1	1	1	1	5
29	8	3	1	−1	1	−1	−1	5
30	8	4	1	−1	−1	1	−1	7
31	8	1	1	1	−1	−1	1	2
32	8	2	1	1	1	1	1	3

blocks and one set for the individual runs.

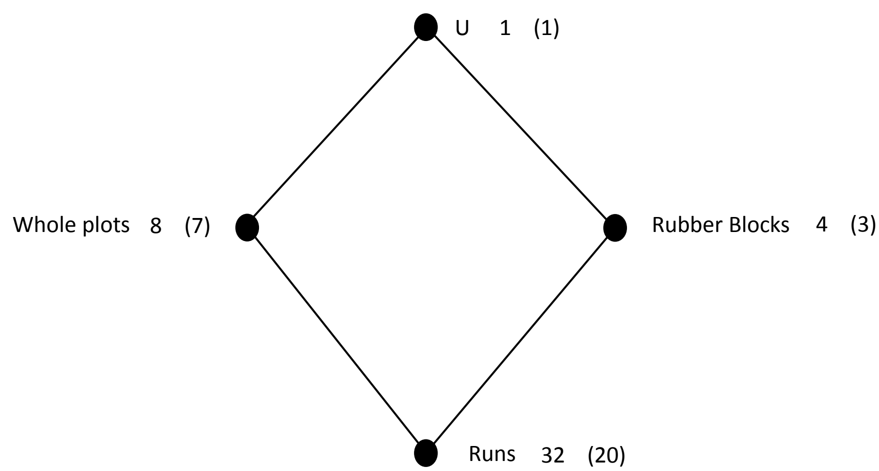

Figure 2: Hasse diagram for the steel tire cord experiment.

In this approach, we use random instead of fixed effects for the rubber blocks. This is the approach we prefer since the rubber blocks used in the experiment represent a sample from the entire population of all possible blocks and we wish to generalize the results to the entire population.

For some experimental situations it is not always clear what the corresponding Hasse diagram should be. Therefore, Grossmann (2014) has developed an algorithm that automates the construction of Hasse diagrams.

4 Cumulative Logit Model

In the steel cord experiment, the rubber coverage response (measuring the quality of the adhesion) is measured on an ordinal, integer-valued, scale from 1 to 7. Clearly, when the response is categorical, the assumption of normality, crucial for a linear regression analysis, is violated. The generalized linear model (GLM) framework, which allows the use of any distribution from the exponential family, offers an alternative in this scenario. The exponential family includes the normal, binomial, multinomial, Poisson, geometric, negative binomial, exponential, gamma, and inverse normal distributions.

A generalized linear model has three components: a response variable distribution, a linear predictor that involves the effects of the experimental factors, and a link function that connects the linear predictor to the natural mean of the response variable (Myers, Montgomery and Vining (2002)). One of the assumptions of a GLM is that the observations are independent. Clearly, this assumption does not hold for the steel tire cord experiment where the responses are correlated due to their grouping in whole plots and rubber blocks. To capture that correlation, the linear predictor needs to contain random effects for the whole plots and the blocks, on top of the fixed effects of the experimental factors. This implies the use of a generalized linear mixed model (GLMM), which is a statistical model that extends the class of generalized linear models by incorporating normally distributed random effects (Robinson, Myers and Montgomery (2004); Robinson et al. (2006); and Goos and Gilmour (2012)). Instead of specifying a distribution for the response, as in the case of a GLM, we now have to specify a distribution for the response conditional on the random effects.

For the data analysis of the steel tire cord experiment, we apply the cumulative logit model, as proposed in Goos and Gilmour (2012). The ordinally amount of rubber coverage Y_{ij} of the observation in whole plot i ($i = 1, \ldots, 8$) and rubber block j ($j = 1, \ldots, 4$), conditional on the random effects δ_i, γ_j and ε_{ij} (for the concision of the paper we have omitted this from the equation), is assumed to follow a multinomial distribution with parameters 1 and $\boldsymbol{\pi}_{ij}$, where $\boldsymbol{\pi}_{ij} = [\pi_{1ij}, \ldots, \pi_{7ij}]'$ with $\pi_{kij} = P(Y_{ij} = k)$, i.e. the probability that the observation in whole plot i and rubber block j has a response equal to k. The link function used is the cumulative logit link function.

This leads to the cumulative logit model in which the cumulative logits are defined as

$$\begin{aligned}
\text{logit}_{1i} &= \ln\left(\frac{P(Y_{ij} > 1)}{P(Y_{ij} \leq 1)}\right) = \ln\left(\frac{\pi_{2ij} + \ldots + \pi_{7ij}}{\pi_{1ij}}\right) = \beta_{10} + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \delta_i + \gamma_j + \varepsilon_{ij}, \\
\text{logit}_{2i} &= \ln\left(\frac{P(Y_{ij} > 2)}{P(Y_{ij} \leq 2)}\right) = \ln\left(\frac{\pi_{3ij} + \ldots + \pi_{7ij}}{\pi_{1ij} + \pi_{2ij}}\right) = \beta_{20} + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \delta_i + \gamma_j + \varepsilon_{ij}, \\
&\vdots \\
\text{logit}_{6i} &= \ln\left(\frac{P(Y_{ij} > 6)}{P(Y_{ij} \leq 6)}\right) = \ln\left(\frac{\pi_{7ij}}{\pi_{1ij} + \ldots + \pi_{6ij}}\right) = \beta_{60} + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \delta_i + \gamma_j + \varepsilon_{ij},
\end{aligned} \tag{1}$$

where the parameters β_{k0} ($k = 1, \ldots, 6$) are the intercept parameters representing the overall levels falling into each response category, $\boldsymbol{x}_{ij}$ represents the model expansion of the levels of the treatment factors for the ith whole plot and the jth rubber block (as an example for the main-effects-and-two-factor-interaction-effects model for the steel tire cord experiment $\boldsymbol{x}_{ij} = \begin{bmatrix} A_{ij} & B_{ij} & C_{ij} & D_{ij} & E_{ij} & (A*B)_{ij} & \ldots & (D*E)_{ij} \end{bmatrix}$), $\boldsymbol{\beta}$ is a vector of fixed factor effects, δ_i is the random effect of the ith whole plot, γ_j is the random effect of the jth rubber block, and ε_{ij} is the random error of the individual run in block j and whole plot i, as suggested by Gilmour (2011). Furthermore, we assume that δ_i follows a normal distribution with mean zero and variance σ^2_δ, γ_j is normally distributed with a mean of zero and a variance of σ_γ^2, ε_{ij} is normally distributed with a mean of zero and a variance of σ_ε^2, and all random variables δ_i, γ_j and ε_{ij} are pairwise independent. The quantity $\boldsymbol{x}'_{ij}\boldsymbol{\beta}$ can be viewed as a continuous latent variable quantifying the quality of the adhesion between the steel cord and the rubber.

Each cumulative logit has its own intercept which increases with k, the cumulative logit model however assumes that the influence of each experimental factor is the same for each logit. This assumption is called the proportional odds asssumption. From the model in Equation (1), we can derive the following probabilities for each of the seven outcomes:

$$\begin{aligned}
\pi_1 &= \frac{e^{\text{logit}_{1ij}}}{1+e^{\text{logit}_{1ij}}}, \\
\pi_1 + \pi_2 &= \frac{e^{\text{logit}_{2ij}}}{1+e^{\text{logit}_{2ij}}}, \\
&\vdots \\
\pi_7 &= 1 - (\pi_1 + \pi_2 + \ldots + \pi_6).
\end{aligned} \tag{2}$$

To estimate the mixed cumulative logit model in Equation (1), we use the SAS procedure GLIMMIX with residual or restricted maximum likelihood (REML) to estimate the variance components. For the significance tests, we use the Kenward-Roger degrees of freedom option. More details about REML and Kenward-Roger degrees of freedom can be found in Letsinger et al. (1996) and Goos et al. (2006).

5 Data Analysis

In this section, we first analyze the original data involving the 7-level ordinal response. Next, we present an alternative, simplified analysis using just four outcomes categories for the ordinal response.

5.1 Original Data Analysis

In the first step of the data analysis, the model contained all the main effects of the five experimental factors, the two-factor interaction effects, a random effect associated with each of the whole plots, and a random effect associated with each of the rubber blocks.

At first, we looked at the estimation of the three different variance components in the model, i.e. the whole-plot variance σ_δ^2, the block variance σ_γ^2, and the error variance σ_ε^2. A first striking result from Table 2 is the fact that the block variance is estimated to be zero, suggesting that the responses of observations coming from the same rubber block are not more alike than observations coming from different rubber blocks. This could mean that there is little variation between the different rubber blocks, which could be viewed as

Table 2: Point estimates and standard errors of the whole-plot variance, the block variance and the error variance.

Variance	Estimate	Stand. Err.
WP	4.5353	9.7073
Block	0.0000	-
Error	27.9532	17.8780

a desirable result for further experimentation. However, another possible explanation is mentioned by Gilmour and Goos (2009) and Goos and Gilmour (2012). A block variance estimated to be zero can also be due to the small number of blocks in the experiment and thus the small number of degrees of freedom to estimate this variance component. A possible solution could be to perform a Bayesian analysis which takes into account the fact that the experiment contains little information concerning the block variance and uses a priori information concerning that variance. Such a Bayesian analysis is, however, beyond the scope of this paper.

Next, the variance ratio $\sigma_\delta^2/\sigma_\varepsilon^2$ is estimated to be only $4.5353/27.9532 = 0.1622$, suggesting a weak correlation between observations conducted in the same whole plot. Based on the results in Table 2, we decided to leave the random block effect out of the model and to redo the analysis. This led to the estimated parameters and significances in Table 3. To identify significant effects, we use a significance level equal to 0.10. Our significance level is more liberal than usual due to the small size of the experiment and the ordinal nature of the response.

The table clearly shows that the intercepts of the cumulative logit model indeed increase with k, i.e. the outcome category. In the initial model, there are two significant effects involving the hard-to-change factor A, namely the interaction effect between the hard-to-change factor A and the easy-to-change factors B and E. The interaction effect between the easy-to-change factors C and E is also significant. To further fine tune this model, we apply a backward elimination procedure and eliminate the non-significant effects, starting with the one that has the highest p-value. In Table 3, the main effect of E turns out to be least significant with a p-value of about 0.95, but, since its interaction effect with A is significant, we leave the main effect in the model to preserve the marginality of the model. The first effect that is eliminated from the model is therefore the interaction effect between D and E.

Further fine tuning eventually leads to the estimated model in Table 4 and the estimated variance components in Table 5. It turns out that there are only three significant effects in

Table 3: Parameter estimates, standard errors and significance test results for a cumulative logit model containing all main effects and two-factor interaction effects.

Effect	Estimate	Stand. Err.	DF	Sign.
Intercept 1	-12.7848	3.4416	11	0.0034
Intercept 2	-8.9368	2.4330	11.000	0.0037
Intercept 3	-4.1352	1.8393	11.000	0.0460
Intercept 4	0.1885	1.6451	11.000	0.9108
Intercept 5	6.2548	2.2281	11.000	0.0171
Intercept 6	12.0943	3.5424	11.000	0.0058
A	-0.4369	1.3119	4.026	0.7557
B	-0.2293	1.0701	8.323	0.8355
C	0.6331	1.0672	8.194	0.5690
D	1.0648	1.0889	8.581	0.3549
E	-0.0591	1.0642	8.119	0.9571
A*B	-2.844	1.2233	11.000	0.0402
A*C	0.6713	1.073	8.359	0.5482
A*D	1.0401	1.0907	8.759	0.3659
A*E	-2.6388	1.2058	11.000	0.0511
B*C	-1.2766	1.0796	8.430	0.2693
B*D	0.8724	1.0785	8.497	0.4406
B*E	-0.7289	1.3247	4.181	0.6102
C*D	-1.1667	1.0941	8.841	0.3145
C*E	2.2027	1.1377	9.849	0.0821
D*E	-0.4912	1.0791	8.458	0.6604

Table 4: Parameter estimates, standard errors and significance test results for the final cumulative logit model.

Effect	Estimate	Stand. Err.	DF	Sign.
Intercept 1	-8.6064	2.2443	19.000	0.0011
Intercept 2	-5.9626	1.4886	19.000	0.0008
Intercept 3	-2.8234	1.1369	19.000	0.0225
Intercept 4	0.0888	1.0479	19.000	0.9334
Intercept 5	4.2527	1.3086	19.000	0.0042
Intercept 6	8.1204	2.0209	19.000	0.0007
A	-0.3020	0.8476	4.598	0.7373
B	-0.1510	0.7263	10.760	0.8392
C	0.4259	0.7240	10.650	0.5686
E	-0.0346	0.7210	10.540	0.9626
A*B	-1.9220	0.8008	14.140	0.0307
A*E	-1.7760	0.7855	13.340	0.0411
C*E	1.4792	0.7574	11.960	0.0746

Table 5: Point estimates and standard errors of the whole-plot variance σ_δ^2 and the error variance σ_ε^2 in the final cumulative logit model.

Variance	Estimate	Stand. Err.
WP	1.5270	3.6518
Error	11.9819	7.2300

the final model, i.e. the two-factor interaction effects between A and B, A and E, and C and E. Besides that, the final model also contains the main effects of these experimental factors. As shown in Table 5, the different variance components are estimated slightly differently in the reduced model than in the original model. However, the variance ratio $\sigma_\delta^2/\sigma_\varepsilon^2$ remains small with an estimated value of only $1.5270/11.9819 = 0.1274$. In a last step of the model selection procedure, we re-introduced a random effect for the rubber block. This is because it may happen that, in the reduced model, the smaller number of fixed effects frees up degrees of freedom for estimating certain variance components, resulting in nonzero estimates. However, the reduced model with random block effect again led to an estimated block variance equal to zero.

As a final step in the data analysis, we use the model estimates of Table 4 to create a prediction formula for the probability that an observation results in a response level of at least five, which is desirable. The prediction formula was implemented in the Prediction

Profiler in the JMP software. This gave us some more insight in the level settings of the five experimental factors that maximize this probability. From Figure 3, it is clear that a maximum probability of 94.77% is obtained when the easy-to-change factors B, C and E are set at their high level while the hard-to-change factor A is set at its low level.

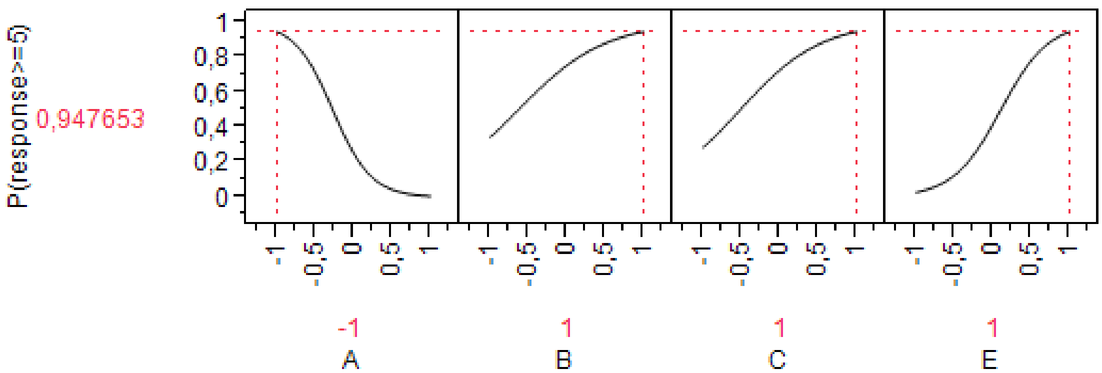

Figure 3: Maximum probability that the response level is at least five

5.2 Alternative Data Analysis

When there is a highly uneven distribution of the responses over the different categories, this may lead to problems with fitting the mixed cumulative logit model. Therefore, it is sometimes interesting to group some response outcomes to get a more even distribution. Another possible beneficial consequence of this grouping is that fewer intercepts need to be estimated (remember that each cumulative logit has its own intercept) and that more degrees of freedom become available for estimating the remaining fixed effects and variance components.

When we take a look at the frequencies of the seven outcomes in the steel tire cord experiment in Table 6, there is indeed a quite uneven distribution of the responses over the various outcomes. Although this did not lead to fitting problems in the previous section, we grouped some of the response outcomes in order to have a larger number of degrees of freedom available to estimate the fixed effects and the variance components. The steel tire cord experiment involves only 32 observations, i.e. 31 degrees of freedom. In the analysis of Section 5.1., six degrees of freedom were necessary to estimate the different intercepts, leaving at most 25 degrees of freedom to estimate the remaining 15 factor effects and 3 variance components.

Table 6: Frequency table of the response outcome in the steel tire cord experiment

Response	Total Freq.
1	1
2	5
3	10
4	7
5	6
6	2
7	1

Table 7: Point estimates and standard errors of the whole-plot variance, the block variance and the error variance when response outcomes are grouped.

Variance	Estimate	Stand. Err.
WP	0.000	-
Block	0.0000	-
Error	20.8973	12.2536

Taking the frequencies of Table 6 into account, it might be interesting to group response categories 1 and 2, and response categories 5, 6 and 7. This would lead to a more even distribution of the response and result in a cumulative logit model with only three different intercepts to estimate. A drawback of grouping some response outcomes is, of course, the fact that part of the information on the response is lost. For the steel tire cord experiment, this alternative analysis with fewer outcome categories led to the variance component estimates in Table 7 and the fixed effect estimates in Table 8.

From Table 7, we can see that the estimate of the error variance is of the same order of magnitude as that in the original analysis in Table 2. Also, the block variance is estimated to be zero, as was the case in the original analysis. The difference, however, is that the whole-plot variance is now estimated to be zero as well, whereas, in the original analysis, it had a positive estimate. Comparing the results of Table 8 with the ones of Table 3, we notice, for example, that, although there are fewer intercepts to estimate and the degrees of freedom available for testing the factor effects increase, we obtain a smaller number of significant effects. We still find the interaction effects between A and B, and, between A and E, to be significant at the 10% level, but the interaction effect between C and E is no longer significant after merging the outcome categories. This shows that merging outcome categories may make it more difficult to detect significant effects.

Table 8: Parameter estimates, standard errors and significance test results for a cumulative logit model containing all main effects and two-factor interaction effects when response outcomes are grouped.

Effect	Estimate	Stand. Err.	DF	Sign.
Intercept 1	-4.1828	1.5529	14.000	0.0175
Intercept 2	-0.0394	1.3218	14.000	0.9766
Intercept 3	5.5928	1.8288	14.000	0.0085
A	-0.7238	1.0397	13.370	0.4983
B	0.1027	1.0190	12.340	0.9213
C	0.5246	1.0172	12.090	0.6153
D	1.0982	1.0411	13.350	0.3102
E	0.0267	1.0138	11.850	0.9794
A*B	-2.9284	1.1789	14.000	0.0263
A*C	0.9317	1.0328	12.730	0.3837
A*D	0.5251	1.0354	13.230	0.6204
A*E	-2.2391	1.1423	14.000	0.0702
B*C	-1.0680	1.0200	12.140	0.3155
B*D	1.3040	1.0501	13.570	0.2354
B*E	-1.0101	1.0548	13.550	0.3550
C*D	-0.6131	1.0231	12.390	0.5598
C*E	1.5620	1.0580	13.520	0.1628
D*E	0.0109	1.0170	12.380	0.9916

6 Conclusion

The steel tire cord experiment is a real-life example of a split-plot experiment with an additional blocking factor. This requires the introduction of an extra random effect in the original split-plot model, which makes the model for the data analysis similar to that for a strip-plot experiment (See Arnouts, Goos and Jones (2010)). Moreover, there is the additional complication that the response is ordinal and, therefore, we suggest to use a mixed cumulative logit model, which is an example of a generalized linear mixed model.

Analysis of the data suggests that there is no effect of the rubber block used. In addition, there is only a small correlation between the steel tire cords produced from the same whole plot (steel cord roll) and only three fixed effects turn out to have a significant effect on the pull-out force. Finally, the best results regarding the quality of the adhesion are reached when the hard-to-change factor A is set at its low level and the easy-to-change factors B, C and E are set at their high level.

We also looked at two alternative methods to analyse the data of the steel tire cord experiment. However, we do not wish to explicitly report the results of these alternative analyses since we do not recommend these to be used. First, we ignored the recommendations of Gilmour (2011) and used the cumulative logit model without a random effect, ε_{ij}, for the individual runs. For this dataset, the final model of this analysis is the same as the final model of the original analysis in Section 5.1.

Second, in practice, the ordinal nature of the response in the steel tire cord experiment is often ignored, which means that the data are analysed as if the response is continuous, using a linear mixed model. Again, for this dataset, the final model of the linear analysis is the same as the final model in the original analysis. The main difference is that the estimate of the variance ratio $\sigma_\delta^2/\sigma_\varepsilon^2$ is only 0.0754 in comparison to 0.1274 in the original analysis.

We do not recommend ignoring the ordinal nature of the response and using a linear mixed model. On page 5 of his Analysis of Ordinal Categorical Data book, Agresti (2010) lists five major reasons why using a conventional linear model is not a good idea:

"*First, there is usually not a clear-cut choice for the scores. Second, a particular response outcome is likely to be consistent with a range of values for some underlying latent variable, and an ordinary regression analysis does not allow for the measurement error that results from replacing such a range by a single numerical value. Third, unlike the methods presented in this book, that approach does not yield estimated probabilities for the response categories at fixed settings of the explanatory variables. Fourth, that approach can yield predicted values above the highest category score or below the lowest. Fifth, that approach ignores the fact that the variability of the responses is naturally nonconstant for categorical data: for an ordinal response variable, there is little variability at predictor values for which observations fall mainly in the highest category (or mainly in the lowest category), but there is considerable variability at predictor values for which observations tend to be spread among the categories.*"

On top of these five reasons, Agresti (2010) also demonstrates the floor effect which results in spurious interaction effects when treating ordinal outcomes as continuous. This strongly backs up our preference to avoid using the conventional linear model.

Acknowledgment

The first author is a postdoctoral research fellow of the Research Fund Flanders (FWO).

Appendix: SAS code

```
proc glimmix;
class wp block;
model response(descending)=
A
B
C
D
E
A*B
A*C
A*D
A*E
B*C
B*D
B*E
C*D
C*E
D*E
/solution ddfm=kr dist=multinomial link=cumlogit;
random wp block wp*block;
run;
```

References

Agresti, A. (2010). *Analysis of Ordinal Categorical Data*, New York: Wiley.

Arnouts, H., Goos, P. & Jones, B. (2010). Design and analysis of industrial strip-plot experiments, *Quality and Reliability Engineering International* **26**: 127–136.

Bingham, D. R., Schoen, E. D. & Sitter, R. R. (2004). Designing fractional factorial split-plot experiments with few whole-plot factors, *Journal of the Royal Statistical Society, Series C* **53**: 325–339. Corrigendum, **54**, 955–958.

Bingham, D. & Sitter, R. R. (1999). Minimum-abberation two-level fractional factorial split-plot designs, *Technometrics* **41**: 62–70.

Gilmour, S. G. (2011). GLMs and GLMMs in the analysis of randomized experiments, *Presented at the "Optimum Design for Mixed Effects Non-linear and Generalised Linear Models" workshop, Isaac Newton Institute, Cambridge, UK.*, url: http://www.newton.ac.uk/seminar/20110809140014451.

Gilmour, S. & Goos, P. (2009). Analysis of data from non-orthogonal multistratum designs in industrial experiments, *Journal of the Royal Statistical Society: Series C* **58**: 467–484.

Goos, P. & Gilmour, S. (2012). A general strategy for analyzing data from split-plot and multi-stratum experimental designs, *Technometrics* **54**: 340–354.

Goos, P. & Jones, B. (2011). *Optimal Design of Experiments: A Case Study Approach*, New York:Wiley.

Goos, P., Langhans, I. & Vandebroek, M. (2006). Practical inference from industrial split-plot designs, *Journal of Quality Technology* **38**: 162–179.

Goos, P. & Vandebroek, M. (2001). Optimal split-plot designs, *Journal of Quality Technology* **33**: 436–450.

Goos, P. & Vandebroek, M. (2003). D-optimal split-plot designs with given numbers and sizes of whole plots, *Technometrics* **45**: 235–245.

Goos, P. & Vandebroek, M. (2004). Outperforming completely randomized designs, *Journal of Quality Technology* **36**: 12–26.

Grossmann, H. (2014). Automating the analysis of variance of orthogonal designs, *Computational Statistics and Data Analysis* **70**: 1–18.

Jones, B. & Goos, P. (2007). A candidate-set-free algorithm for generating D-optimal split-plot designs, *Applied Statistics* **56**: 347–364.

Jones, B. & Goos, P. (2012). An algorithm for finding D-efficient equivalent-estimation second-order split-plot designs, *Journal of Quality Technology* **44**: 363–374.

Letsinger, J. D., Myers, R. H. & Lentner, M. (1996). Response surface methods for bi-randomization structures, *Journal of Quality Technology* **28**: 381–397.

Mee, R. & Bates, R. (1998). Split-lot designs: Experiments for multistage batch processes, *Technometrics* **40**: 127–140.

Myers, R. H., Montgomery, D. C. & G., V. G. (2002). *Generalized Linear Models: With Applications in Engineering and the Sciences*, New York: Wiley.

Mylona, K., Goos, P. & Jones, B. (2014). Optimal design of blocked and split-plot experiments for fixed effects and variance component estimation, *Technometrics* **56**: 132–144.

Paniagua-Quiñones, C. & Box, G. (2009). A post-fractionated strip-strip-block design for multi-stage processes, *Quality Engineering* **21**: 156–167.

Parker, P. A., Kowalski, S. M. & Vining, G. G. (2006). Classes of split-plot response surface designs for equivalent estimation, *Quality and Reliability Engineering International* **22**: 291–305.

Parker, P. A., Kowalski, S. M. & Vining, G. G. (2007a). Construction of balanced equivalent estimation second-order split-plot designs, *Technometrics* **49**: 56–65.

Parker, P. A., Kowalski, S. M. & Vining, G. G. (2007b). Unbalanced and minimal point equivalent estimation second-order split-plot designs, *Journal of Quality Technology* **39**: 376–388.

Robinson, T. J., Myers, R. H. & Montgomery, D. C. (2004). Analysis considerations in industrial split-plot experiments when the responses are non-normal, *Journal of Quality Technology* **36**: 180–192.

Robinson, T. J., Wulff, S., Khuri, A. I. & Montgomery, D. C. (2006). Robust parameter design using generalized linear mixed models, *Journal of Quality Technology* **38**: 65–75.

Vining, G. G., Kowalski, S. M. & Montgomery, D. C. (2005). Response surface designs within a split-plot structure, *Journal of Quality Technology* **37**: 115–129.

Vivacqua, C. & Bisgaard, S. (2009). Post-fractionated strip-block designs, *Technometrics* **51**: 47–55.