

Detecting dominance in stated choice data and accounting for dominance-based scale differences in logit models

Bliemer, Michiel C.J.; Rose, John M.; Chorus, Caspar G.

DOI

[10.1016/j.trb.2017.05.005](https://doi.org/10.1016/j.trb.2017.05.005)

Publication date

2017

Document Version

Accepted author manuscript

Published in

Transportation Research. Part B: Methodological

Citation (APA)

Bliemer, M. C. J., Rose, J. M., & Chorus, C. G. (2017). Detecting dominance in stated choice data and accounting for dominance-based scale differences in logit models. *Transportation Research. Part B: Methodological*, 102, 83-104. <https://doi.org/10.1016/j.trb.2017.05.005>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Detecting dominance in stated choice data and accounting for dominance-based scale differences in logit models

Michiel C.J. Bliemer^{1,*}, John M. Rose², Caspar G. Chorus³

¹ Institute of Transport and Logistics Studies, The University of Sydney

² Institute for Choice, University of South Australia

³ Faculty of Technology, Policy and Management, Delft University of Technology

*corresponding author, T +61.2.91141840, E michiel.bliemer@sydney.edu.au

Abstract – Stated choice surveys have been used for several decades to estimate preferences of agents using choice models, and are widely applied in the transportation domain. Different types of experimental designs that underlie such surveys have been used in practice. In unlabelled experiments, where all alternatives are described by the same generic utility function, such designs may suffer from choice tasks containing a dominant alternative. Also in labelled experiments with alternative specific attributes and constants such dominance may occur, but to a lesser extent. We show that dominant alternatives are problematic because they affect scale and may bias parameter estimates. We propose a new measure based on minimum regret to calculate dominance and automatically detect such choice tasks in an experimental design or existing dataset. This measure is then used to define a new experimental design type that removes dominance and ensures the making of trade-offs between attributes. Finally, we propose a new regret-scaled multinomial logit model that takes the level of dominance within a choice task into account. Results using simulated and empirical data show that the presence of dominant alternatives can bias model estimates, but by making scale a function of a smooth approximation of normalised minimum regret we can properly account for scale differences without the need to remove choice tasks with dominant alternatives from the dataset.

Keywords: stated choice experiments, dominant alternatives, discrete choice, regret, scale

1. Introduction

Discrete choice models based on utility theory are widely used to analyse behaviour and preferences of agents (e.g., travellers) in order to estimate willingness-to-pay measures (e.g., value of travel time savings) and to predict market shares (e.g., mode shares). Stated choice surveys are often used to collect data for estimating the coefficients that describe behaviour. For example, in the transportation field, there are many such surveys for investigating behaviour in mode choice (e.g., Bekhor and Shiftan, 2010) and route choice (e.g., Hensher, 2006), but also parking choice (e.g., Axhausen and Polak, 1991) and vehicle type and fuel choice (e.g., Hess et al., 2012). For an extensive review of stated preference studies in the transportation literature, see Bliemer and Rose (2011). In this paper we look at the negative impact of having dominant alternatives in a stated choice survey and we provide different solutions for dealing with this problem.

We call an alternative *dominant* if it is better than (or equal to) any other alternative in the choice set with respect to all attributes. An alternative is said to be *dominated* if there exists another alternative in the choice set that is dominant. A dominated alternative is expected never to be chosen (zero probability). A pairwise comparison of attribute levels between alternatives can be made in unlabelled experiments, i.e., surveys in which alternatives do not have specific ‘brand’ value (e.g., route A, route B) and in which the coefficients are generic across alternatives and have a known sign. According to the literature review in Bliemer and Rose (2011), most studies conducted in the transportation literature to date have been unlabelled experiments. Nevertheless, there are many exceptions that use labelled experiments in which the alternatives may have alternative specific attributes and coefficients. The labels of the alternatives typically carry ‘brand’ value (e.g., car, train) represented by an alternative specific constant. In this paper we initially focus on unlabelled experiments since the issue of dominance arises mainly with this type of experiments. In a later section we extend our methodology to define dominance in labelled experiments, although dominance here is generally much less of an issue.

Whilst it is well known that the presence of dominant alternatives in the dataset can lead to significantly biased parameter estimates in model estimation (Huber et al., 1982), there is surprisingly little research on the exact nature and magnitude of the impacts and how best to resolve the problem. The most common approach is simply to remove choice tasks with dominant alternatives during the survey design stage. Some deliberately put a dominant alternative in a survey to test attention or understanding of a respondent, and eliminate the choice task later in model estimation (or even remove all observations from the respondent in case he or she fails to choose the dominant alternative). In many cases this is a manual exercise in which the analyst reviews the choice tasks and eliminates the ones with dominant alternatives. Such choice tasks may easily be overlooked at the design stage and end up in the survey used during data collection. Then the analyst has the choice to remove them from the dataset during the data cleaning stage, or keep them in and account for them in estimating choice models.

The contributions of our study are as follows. First, we define a dominance measure based on regret minimization. Secondly, we propose a new design methodology that automatically detects problematic choice tasks by embedding our dominance measure and generates an experimental design without dominance issues. Thirdly, we study the impact of the presence of dominant alternatives in the dataset on parameter estimates in a simple multinomial logit context and show that mainly scale is affected. Finally, we propose a novel discrete choice model that corrects for the presence of dominant alternatives by automatically adjusting the scale for each choice task based on a smooth approximation of our newly proposed dominance measure.

The paper is structured as follows. In Section 2 we provide a brief literature review on dominant alternatives in stated choice studies and show that such alternatives easily occur in experimental designs. In Section 3 we describe a measure expressed in terms of regret that can be used to detect dominant alternatives in an experimental design or dataset. Further, we describe a new efficient experimental design methodology that we term D^* -efficiency, which aims to generate a design with maximum information while avoiding dominant alternatives. Section 4 proposes a novel regret-scaled multinomial logit model aiming to correct for scale differences due to dominant alternatives. Section 5 describes eight experimental designs with varying numbers of problematic choice tasks in a simple route choice case study, and use

these in order to simulate choice observations as well as to collect empirical data in a real-world survey. Section 6 describes the simulation results, while Section 7 discusses outcomes from the empirical dataset. Section 8 extends our regret measure to include utility functions with nonlinear effects and labelled experiments. Finally, Section 9 concludes with a discussion, recommendations, and limitations of this study.

2. Dominancy in stated choice studies

2.1 Literature review

Analysts often include dominant alternatives on purpose in order to determine whether a respondent pays attention to or understands the survey. For example, in the DATIV study in Denmark in 2004 (Burge and Rohr, 2004), nine choice tasks were generated for an unlabelled experiment with two alternatives and two attributes (travel time and travel cost) and choice task six contained on purpose a dominant alternative. If the respondent failed to choose the dominant alternative, all choices from this respondent were removed from the dataset. Also in a value of time study in the Netherlands such a dominant alternative was imposed in one of the choice tasks (Van de Kaa, 2006). Bradley and Daly (1994) collected data using a design in which the first choice task contained a dominant alternative as a lead-in into the survey, which also allows the interviewer to check whether the respondent has understood the choice task. They estimate a discrete choice model with a separate scale parameter for each choice task. The scale parameter of the first choice task with the dominant alternative is much larger compared to the other scale parameters, a result that is in line with behavioural intuition and discrete choice theory (see Section 4). Foster and Mourato (2002) use dominant alternatives to test for consistency of responses. Also Johnson and Mathews (2001) and many others include a dominant alternative in the survey to test for consistency. It could be argued that including a dominant alternative in a survey could be problematic and may actually lead to inconsistency in subsequent choice tasks, since the respondent may no longer take the survey seriously. Therefore, putting a choice task with a dominant alternative at the end of the survey is possibly better than at the beginning of the survey.

Hensher et al. (1988) states that dominant alternatives often occur when generating experimental designs. Experimental designs differ with respect to structure and combinations of attribute level combinations and can also be related to efficiency. The efficiency of a design describes the reliability with which parameters of a given model can be estimated, and several efficiency measures have been proposed in the literature. Walker et al. (2015) generate several experimental designs (e.g., random, orthogonal, D-efficient) and illustrate that it is necessary to check for dominant alternatives and remove such choice tasks in all design types. Hence, existing experimental design strategies are not immune to dominancy. Crabbe and Vandebroek (2012) propose adjusting prior information in order to significantly reduce the likelihood of generating dominant alternatives in Bayesian D-efficient designs (although they cannot be avoided completely). Altering prior information may reduce the occurrence of dominant alternatives, but is not desirable since the analyst is artificially changing assumptions on preferences of respondents. Huber and Zwerina (1996) propose a utility balancing approach that limits (but not necessarily prevents) the number of dominant alternatives. However, one has to be careful since a high level of utility balance may lead to efficiency losses.

As stated in Huber et al. (1982, page 91), “[d]ominance is not easily modeled by most choice models”. In most cases, the analyst will remove choice tasks with dominant alternatives from

the dataset before model estimation, motivated by the idea that no information is obtained from choices of dominant alternatives (Hensher et al., 1988). As discussed above, experimental designs created using common design techniques often include dominant alternatives (by accident or on purpose). Therefore, there is often the need to check for them in the survey design stage as well as the data cleaning stage.

In practice it is often not too difficult to manually detect and remove problematic choice tasks in a survey during the survey design stage. However, as choice tasks become more complex and multiple designs of the survey may exist with different sets choice tasks, it may become more difficult for the analyst to detect them, so experimental designs may need to be computer generated with appropriate dominance constraints on choice tasks in place. In practice, most analysts post-process the experimental designs and remove choice tasks with dominant alternatives. This leads in almost all cases to a loss of orthogonality (if the design was orthogonal in the first place), efficiency, and attribute level balance of the design. It is therefore desirable to make the dominance check an integral part of the experimental design methodology.

Next we illustrate that experimental designs for surveys with unlabelled alternatives are very likely to contain dominant alternatives without proper attention.

2.2 Likelihood of dominant alternatives in unlabelled experiments

Consider a stated choice survey with M unlabelled alternatives described by A attributes. Furthermore, suppose that each attribute has L levels. An L^{MA} factorial experimental design contains all possible choice tasks described by combinations of attribute levels (Louviere et al., 2000), although some of these choice tasks will essentially be the same by simply re-arranging the order of the alternatives in the survey (because they are unlabelled).

Making assumptions on the respondents' preferences, the analyst can determine dominant alternatives. The fewer attributes are present in the survey, the higher the likelihood that dominance will occur. Also, having more alternatives and fewer attribute levels increases the chance of dominant alternatives. Table 1 illustrates the fraction of choice tasks without a dominant alternative in an L^{MA} factorial design. For example, 89.3 per cent of choice tasks in a $3^{3 \times 3}$ full factorial design has a dominant alternative, leaving only 350 unique choice tasks¹ without a dominant alternative². Suppose that we would like to create a design consisting of six choice tasks. In total $\binom{350}{6} = 350! / (6!(350-6)!) \approx 2.445 \cdot 10^{12}$ unique designs without dominant alternatives can be created. While this is a very large number of possible designs to choose from, since the probability of picking a choice task without a dominant alternative is $1 - 89.3 = 10.7$ per cent, this means that the probability of randomly generating a design without any dominant alternatives is negligible (around 0.00015 per cent). For certain design dimensions it is not even possible to find choice tasks that do not contain a dominant alternative (e.g., a $2^{3 \times 2}$ design, $2^{4 \times 2}$ design, $2^{4 \times 3}$ design, and $3^{4 \times 2}$ design). The choice tasks

¹ In total there are 2,100 choice tasks without a dominant alternative, but most of them are permutations of 350 unique choice tasks by re-ordering alternatives. The number of possible permutations for three alternatives is $3! = 6$, such that the number of unique combinations is $2100 / 6 = 350$.

² Note that in order to check dominance for all cases in Table 1, only relative differences between attributes matter. Hence, the *number* of choice tasks without a dominant alternative does not depend on the exact marginal utilities of each attribute, we only require the signs to be known without stating whether they are positive or negative.

without dominant alternatives corresponding to the shaded design dimensions in Table 1 are shown in Table 2, in which we use two or three route alternatives with two (travel time and toll cost) or three attributes (travel time, fuel cost, toll cost), and two or three levels (10, 15, and 20 minutes travel time and \$1, \$2, or \$3 costs, where the middle level is omitted in case of two levels only). Clearly, requiring non-dominance is typically a rather strict constraint on the experimental design.

Table 1 – Dominancy in choice tasks in an L^{MA} factorial design

<i>L</i>	<i>M</i>	<i>A</i>	Dominant (%)	Unique non-dominant tasks
2	2	2	87.5	1
2	2	3	71.9	9
2	2	4	57.0	55
2	3	2	100	0
2	3	3	97.7	2
2	3	4	90.6	64
2	4	2	100.0	0
2	4	3	100.0	0
2	4	4	99.1	25
3	2	2	77.8	9
3	2	3	55.6	162
3	2	4	38.3	2,025
3	3	2	99.2	1
3	3	3	89.3	350
3	3	4	72.9	24,025
3	4	2	100.0	0
3	4	3	98.6	310
3	4	4	91.1	159,300
4	2	2	71.9	36
4	2	3	47.3	1080
4	2	4	30.1	22,896
4	3	2	97.7	16
4	3	3	82.2	7,760
4	3	4	61.8	1,069,056
4	4	2	99.9	1
4	4	3	95.9	28,355
4	4	4	-- *	-- *

* Not calculated since there exist $4^{4 \times 4} \approx 4,295,000,000$ choice tasks

Table 2 – Choice tasks without dominant alternatives in several L^{MA} factorial designs

Task	Route A			Route B			Route C		
	Time	Fuel	Toll	Time	Fuel	Toll	Time	Fuel	Toll
<i>Unique non-dominant choice tasks in a $2^{2 \times 2}$ factorial design</i>									
1	10	–	3	20	–	1	–	–	–
<i>Unique non-dominant choice tasks in a $2^{2 \times 3}$ factorial design</i>									
1	10	1	2	10	3	1	–	–	–
2	10	1	2	20	1	1	–	–	–
3	10	1	2	20	3	1	–	–	–
4	10	3	3	20	1	1	–	–	–
5	10	3	3	20	1	3	–	–	–
6	10	3	3	20	1	1	–	–	–
7	10	3	3	20	1	3	–	–	–
8	10	3	3	20	3	1	–	–	–
9	20	1	3	20	3	1	–	–	–
<i>Unique non-dominant choice tasks in a $2^{3 \times 3}$ factorial design</i>									
1	10	3	3	20	1	3	20	3	1
2	10	1	3	10	3	1	20	1	1
<i>Unique non-dominant choice tasks in a $3^{2 \times 2}$ factorial design</i>									
1	15	–	3	20	–	2	–	–	–
2	10	–	3	20	–	2	–	–	–
3	15	–	3	20	–	1	–	–	–
4	15	–	2	20	–	1	–	–	–
5	10	–	3	20	–	1	–	–	–
6	10	–	2	20	–	1	–	–	–
7	10	–	3	15	–	2	–	–	–
8	10	–	3	15	–	1	–	–	–
9	10	–	2	15	–	1	–	–	–
<i>Unique non-dominant choice tasks in a $3^{3 \times 2}$ factorial design</i>									
1	10	–	3	15	–	2	20	–	1

3. Detecting dominant alternatives in unlabelled experimental designs

In this section a simple measure is proposed that, together with assumptions on preferences, can be used to assess whether a choice task contains a dominant alternative in an unlabelled experiment and to generate experimental designs without dominance.

3.1 Dominancy measure

Consider a choice model with systematic utilities V_{nsj} for each respondent $n \in \{1, \dots, N\}$, alternative $j \in \{1, \dots, J\}$ and each choice task $s \in \{1, \dots, S\}$. Assume that each alternative has

attributes indexed by $k \in \{1, \dots, K\}$. We further assume that the systematic utilities are given by a linear function $V_{nsj}(\mathbf{x}_{nsj} | \boldsymbol{\beta}_n) = \boldsymbol{\beta}'_n \mathbf{x}_{nsj}$ (the prime indicates the transpose operator) in which for each respondent n and each choice task s , alternative j is represented by a set of attribute levels given by a $K \times 1$ vector $\mathbf{x}_{nsj} = [x_{nsjk}]_{k=1, \dots, K}$ called a profile. Preferences of respondent n are given by a $K \times 1$ vector of coefficients, $\boldsymbol{\beta}_n = [\beta_{nk}]_{k=1, \dots, K}$. For each respondent n , choice task s is defined by the $1 \times JK$ vector consisting of profiles, $\mathbf{x}_{ns} = [x'_{ns1}, \dots, x'_{nsJ}]$. As shown in Table 2, an experimental design for respondent n is the collection of choice tasks given by $S \times JK$ matrix $\mathbf{x}_n = [\mathbf{x}_{n1}, \dots, \mathbf{x}_{nS}]'$, where each row represents a choice task. Note that in many cases design \mathbf{x}_n will be the same for all respondents (called a homogeneous design), but in some cases these levels may vary across respondents (heterogeneous design). For example in a pivot design the attribute levels are based on respondent specific reference levels (Rose et al., 2008).

We define dominance of an alternative as follows. An alternative j is said to dominate alternative i for respondent n in choice task s if for each attribute k the utility of alternative j is larger than (or equal to) the utility that would be obtained if the level of that attribute in alternative j would be replaced by its level in alternative i , ceteris paribus (keeping all other attribute levels in alternative j the same). This is a fairly general definition of dominance that can also be applied to nonlinear utility functions as we will show in Section 8.1.

We can formulate this definition in terms of profiles. Define $\Delta_{ns, j \leftarrow i, k}$ as the difference in utility between alternative j and alternative i in choice task s for respondent n by only comparing differences in attribute k . Consider alternatives j and i with profiles \mathbf{x}_{nsj} and \mathbf{x}_{nsi} , respectively. Then $\Delta_{ns, j \leftarrow i, k}$ is defined as

$$\Delta_{ns, j \leftarrow i, k} = \beta_{nk} (x_{nsjk} - x_{nsik}). \quad (1)$$

An alternative j with profile \mathbf{x}_{nsj} is said to dominate an alternative i with profile \mathbf{x}_{nsi} if and only if

$$\Delta_{ns, j \leftarrow i, k} \geq 0, \quad \text{for all } k. \quad (2)$$

Alternative j would strictly dominate alternative i if the inequality sign in (2) would be strict for at least one attribute k . If the profiles of i and j are identical, then $\Delta_{ns, j \leftarrow i, k} = 0$ for all k and the inequalities in (2) hold by definition.

An alternative j is said to be dominant in choice task s for respondent n if and only if

$$\Delta_{ns, j \leftarrow i, k} \geq 0, \quad \text{for all } k, \text{ for all } i \neq j. \quad (3)$$

Alternative j is strictly dominant if the inequality in (2) strictly holds for at least one attribute k and one other alternative i .

In order to determine whether $\Delta_{ns, j \leftarrow i, k} \geq 0$ the analyst needs to look at the difference in the levels of attribute k between alternatives i and j , and needs to know the (expected) sign of

coefficient β_{nk} for each attribute (and for each respondent). The exact value of the coefficient is not relevant. If the analyst expects that an attribute has a negative impact on utility (e.g., travel time, toll cost), then the analyst can simply use $\beta_{nk} = -1$ (or any other negative value), while for attributes with a positive impact on utility (e.g., in-flight entertainment, on-board wifi) one can use $\beta_{nk} = 1$ (or any other positive value). This is useful for removing dominant alternatives at the survey design stage where exact information about coefficient values is typically not available.

In case the (expected) sign of a coefficient is unknown, we can assume $\beta_{nk} = 0$, which means that $\Delta_{ns,j \leftarrow i,k} = 0$ and hence we implicitly assume no trade-offs on this attribute. In order to avoid dominant alternatives this means that we will require trade-offs on other attributes, that is, we need at least one attribute k for which holds that $\Delta_{ns,j \leftarrow i,k} < 0$. Setting multiple coefficients equal to zero will make it more difficult (and perhaps even impossible) to find designs without dominant alternatives. One could argue that if one does not know the sign of coefficients of multiple attributes, then there will be no apparent dominant alternatives and therefore checking for dominancy is not needed.

The conditions in (3) can be combined into the following measure:

$$R_{nsj} = \sum_{i \neq j} \sum_{k=1}^K \max \{0, -\Delta_{ns,j \leftarrow i,k}\} = 0. \quad (4)$$

Value R_{nsj} can be seen as the regret that respondent n attaches to selecting alternative j over all other alternatives in choice task s . More specifically, we use conceptualization of regret as proposed in the context of the random regret minimization (RRM) model (Chorus et al., 2008; Chorus, 2010).³ If this regret is zero, then alternative j is better than (or equally good as) alternative i in a pairwise comparison between all attributes. Therefore, in order for the respondent to make trade-offs, R_{nsj} need to be strictly positive, in which a respondent always feels some regret choosing an alternative. If $R_{nsj} = 0$ for all alternatives j , then all their profiles are identical.

Define also for each respondent n and each choice task s the minimum regret per choice task, R_{ns} , and the minimum regret per design, R_n ,

$$R_{ns} = \min_j \{R_{nsj}\}, \quad (5)$$

$$R_n = \min_s \{R_{ns}\}. \quad (6)$$

If $R_{ns} = 0$, then choice task s contains a strictly dominant alternative or it contains identical alternatives. In both cases no trade-offs between attributes need to be made, hence we

³ This conceptualization of regret differs from the one proposed in classical regret based models such as Regret Theory (Loomes and Sugden, 1982); these conventional theories postulate that regret is a function of the relative *utilities of alternatives* and can only exist in the context of uncertainty. In contrast, RRM postulates that regret is a function of the relative *values of attributes* and that it arises – also in the absence of uncertainty – when the decision maker has to put up with a relatively poor performance on one or more attributes to arrive at a relatively strong performance on other attributes.

typically would like to avoid such choice tasks in the dataset (since they provide no information). Hence, in this section, we make no distinction between the two cases. However, as will be discussed in Section 4, the two cases have a very different impact on scale in estimation. If $R_n = 0$, then the experimental design contains at least one dominant alternative.

While a deterministic model postulates that rational respondents always select a dominant alternative, this may not be the case in a stochastic model with random utilities. In this case, each respondent is assumed to maximise random utility given by $U_{njs} = V_{njs} + \varepsilon_{njs}$, where ε_{njs} is a random unobserved error term following a certain probability distribution. Even in a random utility framework it may be difficult to explain why a respondent would choose a dominated alternative in an unlabelled choice experiment. Assuming that the analyst has a correct understanding of the coefficients of the respondent, it may be that the error is confounded with one of the attributes, i.e., the respondent may relate a small travel time to a trip on a motorway (a characteristic not included as an attribute in the model and therefore assumed to be in the error term), and may not like driving on motorways. Then even if the respondent has a negative marginal utility for travel time, he or she may still choose the alternative with a higher travel time. Hence, the error will never be exactly equal to zero, but will likely be close to zero. Understanding that dominance is related to the error term, which in turn is related to the scale parameter in a logit model, is the starting point for scaling each choice task with respect to regret in Section 4.

3.2 Efficient experimental designs without dominant alternatives

None of the existing experimental design techniques rule out the existence of dominant alternatives in one or more choice tasks. We therefore propose a constrained experimental design method that automatically checks for dominance (i.e., strictly dominant alternatives and identical alternatives that are not strictly dominated) within the design.

Assume that X_n denotes the set of all possible experimental designs for respondent n that satisfy the analysts design dimensions and possibly attribute level balance and orthogonality. A D-optimal design is a matrix with attribute levels $\mathbf{x}_n \in X_n$ that minimises the determinant of the asymptotic variance-covariance (AVC) matrix under the assumption of a vector of prior coefficients $\boldsymbol{\beta}_n$ (e.g., Huber and Zwerina, 1996). Such a design maximises the (Fisher) information obtained from the choice tasks. Since in most cases one cannot guarantee to have found the optimal design (as this would require evaluating all possible designs), these designs are often referred to as D-efficient instead of D-optimal. Besides minimising the determinant, one can also minimise the trace of the AVC matrix (resulting in an A-efficient design), or minimise the maximum sample size required for statistically significant parameter estimates (resulting in an S-efficient design, see Rose and Bliemer, 2013).

We define a D*-optimal (efficient) design as a design that maximises the determinant of the Fisher information matrix under the restrictions that (i) the design contains no dominant alternatives as defined in (3) and (4), and (ii) does not contain choice task replications. Permutations of profiles in a choice task result in an identical choice task. This is not necessarily a problem, and sometimes such choice tasks are included in the design on purpose in order to assess consistent choice behaviour respondents (see Section 1.2). However, most analysts would prefer to include unique attribute level combinations in each choice task and avoid any replications.

To formulate mathematically, let $I(\mathbf{x}_n | \boldsymbol{\beta}_n)$ denote the Fisher information matrix that depends on the experimental design and prior coefficients $\boldsymbol{\beta}_n$. These prior values can be best guesses from the literature, a pilot study, or expert judgement (Bliemer and Collins, in press). Then the D^* -optimal design for respondent n is the matrix $\mathbf{x}_n \in X_n$ that solves the following nonlinear programming problem:

$$\begin{aligned} & \max_{\mathbf{x}_n \in X_n} |I(\mathbf{x}_n | \boldsymbol{\beta}_n)| \\ \text{subject to: } & R_n(\mathbf{x}_n | \boldsymbol{\beta}_n) > 0, \\ & \mathbf{x}_n \text{ does not contain choice task replications,} \end{aligned} \tag{7}$$

where $|\cdot|$ denotes the matrix determinant. Note that such a design cannot be generated if $\boldsymbol{\beta}_n = \mathbf{0}$, i.e., if the analyst has no information regarding the coefficients, not even the sign, since in that case by definition $R_n(\mathbf{x}_n | \mathbf{0}) = 0$. In case the analyst only knows the signs, one can set values close to zero for the priors, i.e. $\beta_{nk} = -0.001$ or $\beta_{nk} = 0.001$. This enables computation of minimum regret and these small deviations from zero will only have little effect on the Fisher information matrix.

Similarly, an A^* -optimal, S^* -optimal or other efficient designs can be defined, where the asterisk indicates that the design is dominance constrained. Also other more advanced designs such as Bayesian D^* -optimal designs can be defined by a direct extension of Bayesian D -optimal designs (Sándor and Wedel, 2001). Such Bayesian efficient designs are more robust against misspecification of prior coefficients. In order to check for dominance, one can use the sign of the mean of the Bayesian prior distribution.

Traditional column based algorithms, i.e., relabelling and swapping techniques described in Huber and Zwerina (1996), modify columns in matrix \mathbf{x}_n and will generally struggle generating designs without dominant alternatives. Since the dominance constraint is on the entire choice task, a row based algorithm that modifies a row in matrix \mathbf{x}_n will therefore be more useful. Federov (1972) proposed a row based algorithm for generating efficient designs, which was modified by Cook and Nachtsheim (1980). This modified Federov algorithm can be used to first construct a candidature set that consist of all (or a select of) choice tasks that do not contain dominant alternatives. For example, when generating a fractional factorial $3^{3 \times 3}$ design, we first determine the 350 unique choice tasks without dominance and without replications (see Table 1). Then we randomly select S choice tasks from this set to form a design, and keep replacing rows in the design with rows in the candidature set until the best design has been found. Note that the number of designs that can be created by selecting S tasks out of 350 is typically very large (see Section 2.2). Therefore the algorithm is usually terminated once the Fisher information no longer improves for a certain number of iterations.

We implemented a column based as well as a row based algorithm in Ngene version 1.1 (ChoiceMetrics, 2012), which that take the constraint $R_n > 0$ into account and avoids replications of the same choice task. We use these algorithms to generate the D^* -efficient designs in this paper. While a row based algorithm can easily avoid dominant alternatives, it is more difficult to generate attribute level balanced designs. If attribute level balance is required, our algorithm selects new choice tasks from the candidature set such that attribute level balance is satisfied or only marginally violated.

4. Regret-scaled multinomial logit model

4.1 Choice task based scaling

Suppose that the analyst decides not to remove choice tasks with dominant alternatives from the dataset and wishes to estimate a simple discrete choice model assuming a decision maker n selecting alternative j that maximise the random utility $U_{nsj} = V_{nsj} + \varepsilon_{nsj}$, where ε_{nsj} is a random, unobserved (by the analyst) component of the utility.

Now, there are two reasons why one would expect the variance of the error term (which is definitionally linked to the inverse of the scale of utility) to decrease strongly when a choice task contains a dominating alternative, leading to more deterministic behaviour and hence choice probabilities close to 1 for the dominant alternative. First, in a conventional logit model with non-random parameters, a substantial portion of the error variance consists of unobserved taste heterogeneity across respondents. Now, if we are willing to assume that this heterogeneity does not involve sign differences across respondents (e.g., we assume that despite differences in cost sensitivity, everyone dislikes higher costs), such taste heterogeneity would not lead to heterogeneity in choice outcomes when a dominant alternative is present. That is, irrespective of differences in weights assigned to attributes by respondents everyone would choose the dominant alternative as it by definition performs best on all attributes (we for the moment ignore other sources of error; see directly below). So, the unobserved taste heterogeneity which in non-dominated choice tasks constitutes a sizeable portion of error variance leading to a distribution of choices across alternatives, is expected to vanish in choice tasks with a dominant alternative, leading to a much larger scale of utility, i.e., a much smaller error term variance, and hence much more pronounced differences in choice probabilities.

Second, it is expected that choice tasks containing a dominant alternative are perceived by individuals as being much easier to respond to than non-dominated choice tasks. The reason is that, in the latter choice tasks, a trade-off needs to be made by the respondent whereas in the former this is not the case. As a consequence, we expect that there will be less behavioural idiosyncracies or mistakes in the dominated choice tasks than in other tasks. Since such idiosyncracies by definition end up in the error term, this is another reason why it is expected that the variance of the error term will be strongly diminished when a choice task contains a dominant alternative.

In order to account for these expected differences in scale across observations which are due to the presence of dominant alternatives in some choice tasks, one can estimate a model with choice task specific scale parameters, such as in Bradley and Daly (1994). Scale parameters for choice tasks with a strictly dominant alternative are expected to be large in contrast to other choice tasks. On the other hand, a choice task with identical alternatives will also not allow any trade-offs, but is expected to have a very small scale parameter, since the choice will be mostly based on the unobserved component (i.e., the user chooses more or less randomly since all alternatives are the same). Indeed, these expectations are in line with the results of Bradley and Daly, who estimated 14 scale parameters on top of four regular coefficients in the utility function. One of the scale parameters (corresponding to a base choice task) needs to be set to one, and all other scales are relative to this base. Clearly, such a choice task specific scaling significantly increases the number of parameters to be estimated. In order to avoid having to estimate a separate parameter per choice task, we adopt

a parametric approach in which we make scale a function of our dominance measure introduced in Section 3.1.

Under the assumption that ε_{nsj} are independently and identically extreme value type I distributed with variance $\frac{1}{6}\pi^2\lambda_{ns}^{-2}$, we obtain an extension of the well-known multinomial logit (MNL) model (McFadden, 1974) in which this variance of the error term in choice task s is inversely related to the scale in the choice task, λ_{ns} . The probability of respondent n selecting alternative j in choice task s is then given by

$$P_{nsj} = \frac{\exp(\lambda_{ns}V_{nsj})}{\sum_{i=1}^J \exp(\lambda_{ns}V_{nsi})}. \quad (8)$$

In a (homoscedastic) MNL logit model, $\lambda_{ns} = 1$ for all s . In case of a strictly dominant alternative in choice task s , the variance of error ε_{nsj} is expected to diminish, which corresponds to an increase in the scale parameter.

There is a formal equivalence between Luce models (Luce, 1959) and standard MNL models. The probabilities in Equation (8) are consistent with a set-dependent Luce model (Marley et al., 2008), where the scale can depend on (all) the options in the choice set, and this class of models could be called set-dependent MNL models. As with the MNL model with a scale that does not depend on the choice set, the Luce model cannot produce a probability of one for a dominant alternative except in the limit as the scale goes to infinity.

In our study, we will relate scale parameter λ_{ns} to minimum regret R_{ns} in order to make it heteroscedastic. However, there are two concerns in using minimum regret R_{ns} as a descriptor for λ_{ns} . First of all, R_{ns} is bounded from below by zero, but the upper bound depends on the attribute level ranges. For interpretability reasons we prefer an upper bound that does not rely on the levels, similar to the entropy upper bound of (independent of attribute level range) used in the model of Swait and Adamowicz (2001) for scaling choice tasks according to complexity. Secondly, R_{ns} as defined in Equation (5) is not ‘smooth’, since it involves minimum and maximum operators. This typically leads to numerical problems in model estimation, and it also does not discriminate between a choice task with a strictly dominant alternative (with a very high scale parameter) and a choice task with identical alternatives (with a very low scale parameter). We address these two concerns in the next subsections.

4.2 Normalised minimum regret

In order to address the first issue, we simply normalise the minimum regret by the average regret in the choice set. Hence, our normalised minimum regret M_{ns} becomes

$$M_{ns} = \frac{R_{ns}}{J^{-1} \sum_{j=1}^J R_{nsj}}. \quad (9)$$

Note that $R_{njs} \geq 0$ for all alternatives j such that $M_{ns} \geq 0$. Suppose that choice task s contains a strictly dominant alternative for respondent n . This means that $R_{ns} = 0$ and there exists a dominated alternative j for which $R_{njs} > 0$. As a result, a choice task with a strictly dominant alternatives yields $M_{ns} = 0$. Now suppose that choice task s does not contain any strictly dominant alternatives for respondent n , such that $R_{njs} > 0$ for all alternatives j . Since R_{ns} is the minimum over these values, R_{ns} can never be greater than $\frac{1}{J} \sum_j R_{njs}$, hence $M_{ns} \leq 1$. The upper bound of $M_{ns} = 1$ is reached when all alternatives have the same positive regret R_{njs} , making each alternative equally attractive. In summary, it holds that $M_{ns} \in [0, 1]$.

In the extreme case where the profiles of all alternatives in a choice task are identical, i.e., $R_{njs} = 0$ for all alternatives j , the normalised minimum regret in (9) is undefined (zero divided by zero). Clearly such choice tasks should be prevented at all times, but as we will show in Section 4.3, the smooth approximation of the normalised minimum regret is properly defined in this extreme case and will not lead to numerical problems.

It is interesting to note how normalised minimum regret M_{ns} relates to entropy E_{ns} , which is defined by Shannon (1948) as (our notation):

$$E_{ns} = -\sum_{j=1}^J P_{nsj} \log P_{nsj}. \quad (10)$$

This entropy value is bounded by $E_{ns} \in [0, -\frac{1}{2} \log(\frac{1}{2})]$. Entropy is used as a proxy for choice task complexity in Swait and Adamowicz (2001). Typically a low (high) normalised minimum regret also means a low (high) entropy, and vice versa. For example, if a choice task has a strictly dominant alternative (i.e., a relatively easy choice) such that one alternative is chosen with a probability equal to 1, then $M_{ns} = 0$ and $E_{ns} = 0$. On the other hand, if all alternatives are different on every attribute but probabilities and regrets are identical (i.e., a relatively difficult choice), then both M_{ns} and E_{ns} are maximised.

An important difference is that entropy depends on choice probabilities, which makes it dependent on the model assumptions. Normalised minimum regret only depends on the utility function and not on a specific type of discrete choice model. This means that the entropy metric is much less sensitive to dominance than the regret metric, especially when attribute differences across alternatives are small. To illustrate this, consider a simple route choice example in which Route A is described by a travel time of 10 minutes and a travel cost of \$1, while Route B has a travel time of 11 minutes and the same travel cost of \$1, making route A a strictly dominant alternative. Assume a linear utility function and negative coefficients for time and cost. If one uses an MNL model, then the probabilities will be almost identical, yielding a high value for E_{ns} . In contrast, M_{ns} will be equal to zero.

4.3 Smooth approximation of minimum regret

In order to resolve the second issue, we replace the maximum operator with the ‘soft maximum’ operator in order to approximate the non-smooth minimum regret R_{ns} function by

a smooth function \tilde{R}_{ns} (we denote all smooth approximations with a tilde). The soft maximum for a series of values a_1, \dots, a_z is defined as follows (see e.g., Cook, 2011):

$$\max_z \{a_z\} \approx \frac{1}{\xi} \log \left(\sum_{z=1}^z \exp(\xi a_z) \right), \quad (11)$$

where $\xi > 0$ defines the ‘hardness’. The approximation becomes exact if $\xi \rightarrow \infty$. In this paper we will use $\xi = 10$ since Figure 1 illustrates that using this value results already in a reasonably good approximation of the maximum operator. Using a larger value would theoretically yield a better approximation but may lead to numerical problems due to the use of the exponential function in Equation (11).

The smooth approximation for the regret of alternative j , as defined in Equation (4), is given by

$$\tilde{R}_{nsj} = \frac{1}{\xi} \sum_{i \neq j} \sum_{k=1}^K \log \left(1 + \exp(-\xi \Delta_{ns, j \leftarrow i, k}) \right). \quad (12)$$

The smooth approximation for R_{ns} in Equations (5) and (6) can be calculated in the same way by taking the ‘soft minimum’. Since $\min_z \{a_z\} = -\max\{-a_z\}$, we can use Equation (12) again to calculate the following smooth approximations:

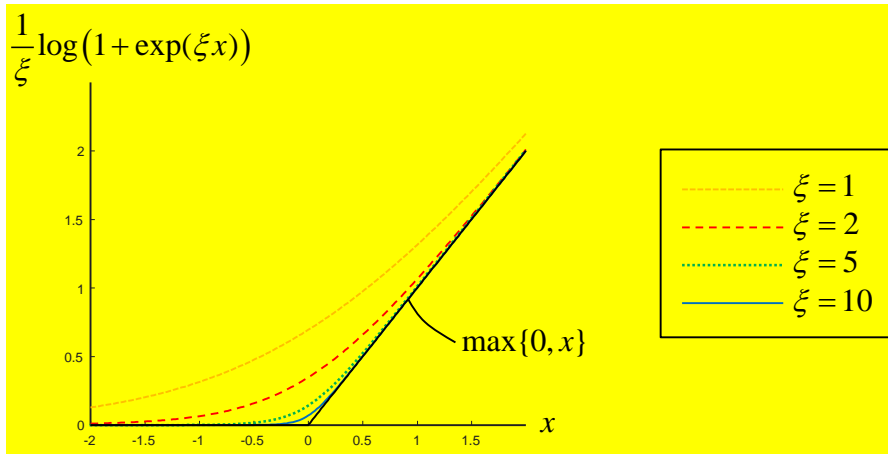


Figure 1 – Smooth approximation for different levels of ‘hardness’

$$\tilde{R}_{ns} = -\frac{1}{\xi} \log \sum_{j=1}^J \exp(-\xi \tilde{R}_{nsj}), \quad (13)$$

$$\tilde{R}_n = -\frac{1}{\xi} \log \sum_{s=1}^S \exp(-\xi \tilde{R}_{ns}). \quad (14)$$

It is interesting to note that Equation (12) is identical to the formulation of regret for an alternative as formulated in Chorus (2010) and using a moderate hardness of the soft maximum of $\xi = 1$, resulting in

$$\tilde{R}_{nsj} = \sum_{i \neq j} \sum_{k=1}^K \log\left(1 + \exp\left(\beta_{nk}(x_{nsik} - x_{nsjk})\right)\right). \quad (15)$$

Furthermore, regret for a choice task as stated in Equation (13) is identical to the random regret logsum derived by Chorus (2012) in the case of $\xi = 1$. Our generalisation with respect to hardness ξ and nonlinear utility functions (including interactions between attributes, see Section 8.1) can also be applied in a random regret choice modelling context. Van Cranenburgh et al. (2015) provide an alternative derivation and interpretation of hardness ξ in the regret formulation.

The smooth approximation of normalised minimum regret M_{ns} , denoted by \tilde{M}_{ns} , can be calculated for each respondent n and for each choice task s using Equation (9) by replacing R_{ns} with \tilde{R}_{ns} and R_{nsj} with \tilde{R}_{nsj} . If choice task s for respondent n has a strictly dominant alternative j , then $\tilde{M}_{ns} \rightarrow 0$ approaches zero for a sufficiently large ξ . In case all alternatives have an identical positive regret R' , then $\tilde{M}_{ns} = 1 - \log(J)(\xi R')^{-1}$, which approaches one for sufficiently large values of ξ . Hence, for finite ξ it holds that $\tilde{M}_{ns} \in (0, 1)$. Finally, consider the case in which all alternatives are represented by identical profiles, i.e., $R_{nsj} = 0$ for all alternatives j . While M_{ns} in Equation (10) is undefined in this case, it can be shown that in case of identical profiles $\tilde{M}_{ns} = 1 - (K(J-1)\log(2))^{-1} \log(J)$, which equals 0.5 when $J = K = 2$.

4.4 Scaling using smooth approximations of normalised minimum regret

Now that we have normalised minimum regret and also derived a smooth approximation, we can relate scale λ_{ns} to \tilde{M}_{ns} in such a way that scale decreases with increasing normalised minimum regret. Two obvious choices would be an exponential or a power function. We propose the following power function:

$$\lambda_{ns} = \tilde{M}_{ns}^{-\gamma}, \quad (16)$$

where γ is a coefficient that needs to be estimated. If $\gamma = 0$, then the probabilities in Equation (8) are consistent with the homoscedastic MNL model. Given how λ_{ns} and \tilde{M}_{ns} are related, it is expected that $\gamma \geq 0$. We also tested other functional forms, such as $\lambda_{ns} = \exp(-\gamma \tilde{M}_{ns})$, but a power function seems to work best, especially since $\lambda_{ns} \rightarrow \infty$ if $\tilde{M}_{ns} \rightarrow 0$. We call our choice model in Equation (8) with scale determined as in Equation (16) a regret-scaled multinomial logit (RS-MNL) model.

5. Simulated and empirical datasets

In order to demonstrate how dominance can be excluded from surveys, how it can be taken into account in estimation, and how it affects results when not taken into account appropriately, we created four experimental designs for a simple route choice study. Then we used these designs to simulate choices and also to create an online survey to collect actual choice data from respondents.

5.1 Simple route choice case study

In order to demonstrate the impact of dominance, we consider a simple route choice case study in which there are two unlabelled alternatives (Routes 1 and 2) with a generic linear utility function considering two attributes, namely travel time and travel cost:

$$V_{njs} = \beta_T T_{njs} + \beta_C C_{njs}, \quad (17)$$

where β_T and β_C are the coefficients for time and cost, respectively, such that the value of travel time savings (VTTS) is given by β_T / β_C . We assume a homogeneous population such that these coefficients are the same for all respondents, and four different levels for each attribute, namely $T_{njs} \in \{10, 15, 20, 25\}$ (minutes) and $C_{njs} \in \{1, 2, 3, 4\}$ (Australian dollars). We generate homogeneous and heterogeneous designs in which each respondent faces eight choice tasks.

In order to assess dominance, we need to know the signs of the coefficients. We assume that coefficients β_T and β_C are both negative. Further, in order to generate efficient experimental designs, we assume the following prior values (best guesses) for these coefficients: $\beta_T = -0.2$ and $\beta_C = -1.2$, such that the VTTS is \$10 per hour.

5.2 Experimental designs

We generate eight experimental designs as listed in Table 3, namely three heterogeneous designs (denotes R1, R2, and R3) in which each respondent faces a different set of randomly generated choice tasks (we create in total 2500 different sets), and five homogeneous designs (denoted O1, O2, E1, E2, and E3) constructed using orthogonality and/or efficiency criteria in which each respondent faces the same choice tasks. For more information on generating experimental designs for stated choice studies we refer to Huber and Zwerina (1996) and Rose and Bliemer (2009).

According to Table 1, there are 36 unique choice tasks without dominant alternatives, such that there exist $\binom{36}{8} = 30,260,340$ unique designs consisting of eight choice tasks without a dominant alternative. As shown in Table 3, existing methods for generating experimental designs, including random designs (R1, R2), (near-)orthogonal designs (O1, O2), and D-efficient designs (E1), are not able to rule out choice tasks with dominant alternatives (each respondent faces at least two such choice tasks in each of the designs). Designs R3, E2, and E3 have been created by rejecting designs that include choice tasks with a dominant alternative (detected using our regret measure). The D-error reported in Table 3 is a measure

for the efficiency of a design for a single respondent⁴ and is computed as $|I(\mathbf{x}_n | \boldsymbol{\beta}_n)|^{-1/K}$ assuming priors $\boldsymbol{\beta}_n = (-0.2, -1.2)'$ for all respondents n . The lower the D-error, the more (Fisher) information is captured per choice task and the smaller standard errors in estimation will be. Some of the designs are attribute level balanced, while others are not. Attribute level balance ensures that the respondent sees all attribute levels an equal number of times throughout the survey and that the data covers the range of levels for each attribute equally, which is often seen as a desirable property.

Table 3 – Generated experimental designs

	Experimental design							
	R1	R2	R3	O1	O2	E1	E2	E3
D-error	0.121	0.127	0.089	0.304	0.076	0.057	0.064	0.053
Dominant choice tasks	68%	72%	0%	100%	63%	25%	0%	0%
<i>Design property</i>								
Random choice tasks	■	■	■					
(Near) orthogonal				■	■			
Attribute level balanced	■			■	■	■	■	
D-efficient					■	■	■	■
No strict dominance			■				■	■

Design R1 is generated using a column based algorithm in which we create choice tasks for each respondent by taking random permutations of levels within each column of an initial design matrix that has attribute level balanced columns $(10,10,15,15,20,20,25,25)'$ and $(1,1,2,2,3,3,4,4)'$ for the time and cost attributes, respectively. On average 68 per cent of the choice tasks contain a dominant alternative. Design R2 is generated using a row based algorithm in which we randomly select attribute levels for each attribute and each alternative. This design does not satisfy attribute level balanced and contains on average 72 per cent choice tasks with a dominant alternative (which is consistent with the 71.9 per cent listed in Table 1). Design R3 randomly selects choice tasks from a candidate set with 36 unique choice tasks without any dominance. This design also does not satisfy attribute level balance.

Designs O1 and O2 are both near-orthogonal (i.e., the design is attribute level balanced and attribute levels are uncorrelated, but in order to limit the size of the design not all pairwise attribute level combinations are present). Design O1 has a low D-efficiency (high D-error) while design O2 has a high D-efficiency (low D-error). All eight choice tasks in design O1 have a dominant alternative, while this is the case in five out of eight choice tasks in O2.

Designs E1 and E2 both maximise D-efficiency under the constraint of attribute level balance. Design E1 contains two choice tasks with a dominant alternative. Further, note that choice tasks 2 and 3 are essentially the same, as well as 1 and 6, 4 and 5, and 7 and 8, therefore this design contains only four unique combinations of attribute levels and four replications. Design E2 is a D^* -efficient design (i.e., without dominance or choice task

⁴ Since the random designs are heterogeneous, the D-error per respondent is calculated as the average D-error obtained for all 2500 sets combined, multiplied by 2500.

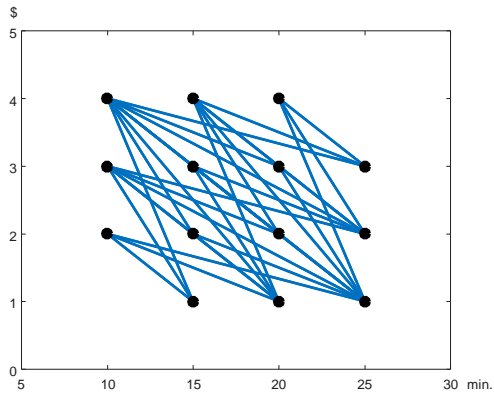
replications). We also generate design E3 which is a D^* -efficient design without requiring attribute level balance (and hence the D-error goes further down).

In Figure 2 we have visually represented the choice tasks in the homogeneous experimental designs, with travel time and cost on the horizontal and vertical axis, respectively. Each profile in the design is represented with a black dot and each choice task is represented by a line between two dots. All possible choice tasks without a dominant alternative are shown in Figure 2(a), i.e., all lines need to have a negative slope (running from north-west to south-east or vice versa). Dashed (red) lines indicate a choice task with a dominant alternative, while solid (blue) lines indicate a choice task without a dominant alternative.

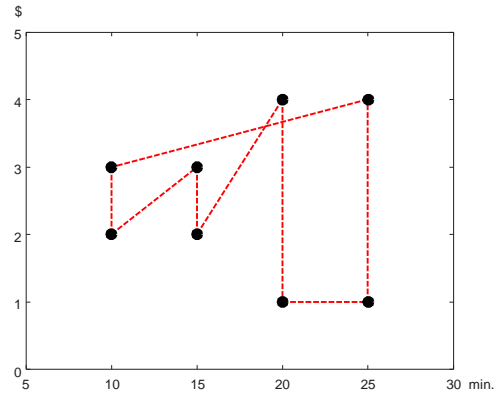
Designs O1, O2, E1, and E2 are clearly attribute level balanced, since each attribute level appears exactly twice. All choice tasks in the design O1 in Figure 2(b) contain a strictly dominant alternative. Design O2 in Figure 2(c) contains three solid (blue) line segments and five dashed (red) lines. Design E1 in Figure 2(d) shows only four lines, since each choice task is replicated twice. Design E2 in Figure 2(e) shows eight solid (blue) lines, such that there are no replications nor dominant alternatives. Design E3 as visualised in Figure 2(f) shows that without the requirement of attribute level balance, profiles are pushed towards the edges since this increases trade-offs and thereby efficiency.

All eight designs will be used to simulate choices in order to create datasets and compare model estimates. For our empirical analysis we concentrate on the four attribute level balanced homogeneous designs (O1, O2, E1, and E2) that differ in dominance levels (namely 100, 63, 25, and 0 per cent of choice tasks with dominant alternatives). These four designs are listed in Table 4 including the associated MNL probabilities consistent with $\beta_T = -0.2$ and $\beta_C = -1.2$. The shading in the table indicates choice tasks with a dominant alternative.

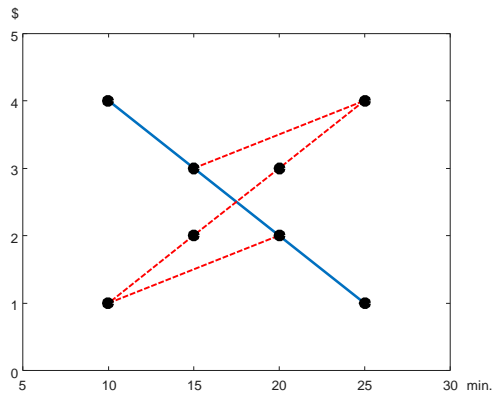
Even though the probabilities in the MNL model would suggest that the probability of choosing Route 2 is 0.77 for the first choice task in design O1, a decision maker would under these assumptions be expected to always choose Route 2. Hence we would expect that the observed probabilities will be (close to) 0.00 and 1.00 for routes 1 and 2, respectively. This discrepancy is due to the difference between the assumptions in the (homoscedastic) MNL model and the actual (heteroscedastic) behaviour. Such a discrepancy between the modelled and actual choice probabilities could be diminished by increasing scale λ_1 in our RS-MNL model.



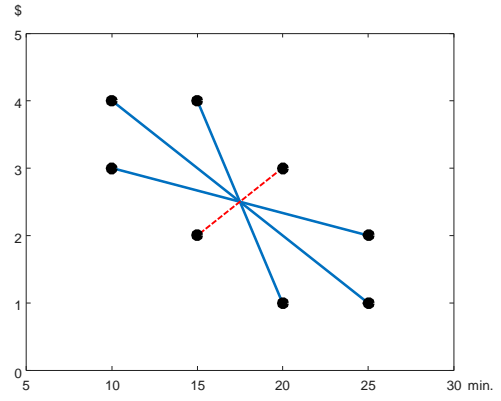
(a) Choice tasks without dominant alternatives



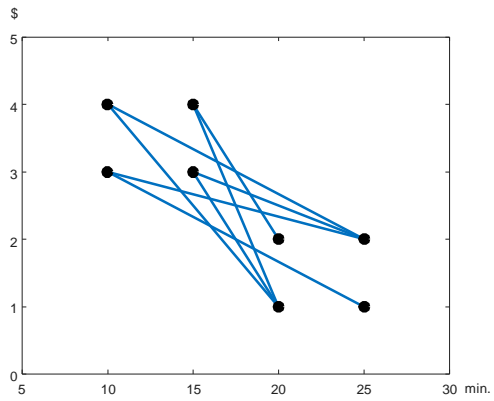
(b) Design O1



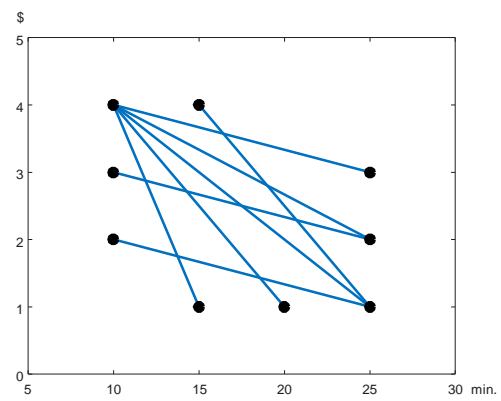
(c) Design O2



(d) Design E1



(e) Design E2



(f) Design E3

Figure 2 – Visualisation of choice tasks

Table 4 – Balanced homogeneous experimental designs (grey shading indicates a choice task containing a dominant alternative)

s	j	Design O1			Design O2			Design E1			Design E2		
		T_{sj}	C_{sj}	P_{sj}	T_{sj}	C_{sj}	P_{sj}	T_{sj}	C_{sj}	P_{sj}	T_{sj}	C_{sj}	P_{sj}
1	1	15	3	0.23	20	2	0.60	10	4	0.35	10	4	0.17
1	2	15	2	0.77	10	4	0.40	25	1	0.65	20	1	0.83
2	1	25	1	0.97	10	4	0.60	25	1	0.65	25	2	0.35
2	2	25	4	0.03	25	1	0.40	10	4	0.35	10	4	0.65
3	1	20	4	0.03	25	1	0.35	10	3	0.86	15	4	0.20
3	2	20	1	0.97	15	3	0.65	25	2	0.14	20	2	0.80
4	1	10	2	0.90	10	1	0.96	25	2	0.14	20	1	0.80
4	2	15	3	0.10	20	2	0.04	10	3	0.86	15	3	0.20
5	1	10	3	0.23	20	3	0.10	20	1	0.93	15	3	0.69
5	2	10	2	0.77	15	2	0.90	15	4	0.07	25	2	0.31
6	1	20	1	0.73	15	2	0.10	15	4	0.07	25	2	0.14
6	2	25	1	0.27	10	1	0.90	20	1	0.93	10	3	0.86
7	1	15	2	0.97	25	4	0.10	20	3	0.10	10	3	0.65
7	2	20	4	0.03	20	3	0.90	15	2	0.90	25	1	0.35
8	1	25	4	0.01	15	3	0.96	15	2	0.90	20	1	0.93
8	2	10	3	0.99	25	4	0.04	20	3	0.10	15	4	0.07

5.3 Simulated choices

In this section we generate datasets by simulating choices consistent with an MNL model, except when there is a dominant alternative. In such a choice task, there are no trade-offs to be made and therefore, for reasons explained in Section 4.1, we assume that the actual behaviour will be that all respondents choose the dominant alternative. This simulation setup is therefore similar to Rose et al. (2013), who simulate datasets to determine the impacts of wrong model assumptions (although they did not look at the case of dominant alternatives).

Let y_{nsj} denote a choice indicator that equals one if respondent n chooses alternative j in choice task s , and zero otherwise. Assuming an MNL model and that the true coefficients are $\beta_T = -0.2$ and $\beta_C = -1.2$, we simulate these observations by randomly drawing ε_{nsj} from an extreme value type I distribution with variance $\frac{1}{6}\pi^2$ independently for each alternative, choice task, and respondent. In case there is no dominant alternative in choice task s (i.e., $R_{ns} > 0$), then $y_{nsj} = 1$ if $V_{nsj} + \varepsilon_{nsj} \geq V_{nsi} + \varepsilon_{nsi}$ for all i , and zero otherwise. In case the choice task does contain a dominant alternative (i.e., $R_{ns} = 0$), then $y_{nsj} = 1$ for alternative j that has minimum regret $R_{nsj} = 0$, and zero otherwise. Note that none of the homogeneous experimental designs in Table 4 have identical alternatives in a single choice task, so there will be only one such dominant alternative. In contrast, the heterogeneous random experimental designs contain some choice tasks with identical alternatives. In that case, we randomly select an alternative.

We simulate choices for 2,500 respondents for each of the eight designs, such that in total there are $2500 \times 8 = 20,000$ choice observations in each dataset. Further, using a jack-knifing technique we create five subsamples consisting of 500 respondents each to confirm results on smaller samples.

5.4 Empirical choices

We used the four balanced and homogeneous experimental designs (O1, O2, E1, and E2) to create an internet survey. In total 360 respondents were asked to participate in the survey, in which each respondent faced 16 choice tasks originating from two of the four designs, thereby obtaining in total 5,760 choice observations (1,440 per experimental design).

In total six different combinations of experimental designs can be made (O1-O2, O1-E1, O1-E2, O2-E1, O2-E2, E1-E2), and the order can be reversed, such that each respondent saw one of twelve different versions of the survey. We emphasized in the survey that the choice tasks were computer generated in order to prepare the respondent for possible ‘silly’ choice tasks because of dominant alternatives.

The observed choice probabilities are listed in Table 5 in which the shading again indicates a choice task with a dominant alternative. It is interesting to see that the choice probabilities only reach 1.000/0.000 in one case (namely the second choice task in the design O1 in which both routes have the same travel time, but one route has a cost of \$1 while the second route has a cost of \$4). In all other cases, at least one respondent did not choose the dominant alternative (i.e., they chose a dominated alternative). Taking a closer look at the data, there are 40 respondents that chose one dominated alternative, five respondents that chose two dominated alternatives, two respondents that chose three dominated alternatives, three respondents that chose four dominated alternatives, and one respondent that chose seven dominated alternatives (out of 16). We will refer to these choice observations as spurious choices. Hence, out of 5,760 choice observations there are 75 spurious choices (1.3 per cent). There were no respondents that consistently chose routes with longer travel times and higher costs, so we can conclude that all respondents perceive time and cost as a disutility in general. The 40 respondents may have made a mistake due to fatigue, especially since the ‘mistake’ occurred mostly near the end of the survey; or they may have been annoyed by the seemingly unreasonable choice task leading to a ‘protest’ response. The 11 respondents that did not choose the strictly dominant alternative multiple times may not have taken the survey seriously and may have selected their preferred option in a somewhat random fashion.

6. Results from simulated dataset

6.1 Estimates for the multinomial logit model

Using the data simulated in Section 5.3, we first estimate coefficients in an MNL model for each of the eight experimental designs. We use BIOGEME (Bierlaire, 2003) for all model estimations in this paper. The estimation results on the dataset of 2,500 simulated respondents are summarised in Table 6.

First, we consider the estimates based on the simulated choices for random designs R1 and R2 that contain a large number of dominant alternatives. The estimates for the time and cost coefficients are significantly inflated (due to scale) compared to the ‘true’ values -0.2 and -1.2. While the VTTS values seem reasonable, they are actually statistically different from \$10/hr (at the 95 per cent significance level). Hence, not only is the scale different due to dominant alternatives, also the ratios of coefficients are affected. In contrast, the coefficients for random design R3, which does not contain any dominant alternatives, are fairly close to the ‘true’ values. Further, its corresponding VTTS is not statistically different from \$10/hr. The model fit measures (log-likelihood values and adjusted ρ^2) for designs R1 and R2 are

much better than for design R3. It is important to note that this does not mean that one should use design R1 or R2 instead of design R3. Model fits can only be compared within datasets and not across datasets. The model fit merely indicates how well the model can correctly predict choice observations. In datasets with many dominant alternatives, it is quite easy to predict the choices, and hence the model fit is high. One should clearly prefer using design R3 despite its seemingly low model fit, since more information on trade-offs is captured and model estimates are not biased.

Table 5 – Observed choice probabilities in empirical dataset (grey shading indicates a choice task containing a dominant alternative)

<i>s</i>	<i>j</i>	Design O1	Design O2	Design E1	Design E2
1	1	0.011	0.689	0.356	0.206
1	2	0.989	0.311	0.644	0.794
2	1	1.000	0.583	0.756	0.517
2	2	0.000	0.417	0.244	0.483
3	1	0.017	0.356	0.233	0.061
3	2	0.983	0.644	0.767	0.939
4	1	0.967	0.961	0.939	0.889
4	2	0.033	0.039	0.061	0.111
5	1	0.011	0.050	0.056	0.578
5	2	0.989	0.950	0.944	0.422
6	1	0.983	0.050	0.633	0.367
6	2	0.017	0.950	0.367	0.633
7	1	0.994	0.033	0.078	0.500
7	2	0.006	0.967	0.922	0.500
8	1	0.028	0.978	0.972	0.956
8	2	0.972	0.022	0.028	0.044

Table 6 – MNL estimates on simulated datasets (2,500 respondents, 20,000 observations per design)

	Design R1		Design R2		Design R3		Design O1	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Time (β_T)	-0.374	0.006	-0.396	0.007	-0.196	0.004	--*	--*
Cost (β_C)	-2.140	0.034	-2.240	0.036	-1.187	0.020	--*	--*
VTTS	\$10.49/hr		\$10.60/hr		\$9.90/hr		--*	
LL	-4988.8		-4901.5		-11367.4		--*	
Adj. ρ^2	0.640		0.646		0.180		--*	

	Design O2		Design E1		Design E2		Design E3	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Time (β_T)	--*	--*	-0.247	0.003	-0.197	0.003	-0.192	0.003
Cost (β_C)	--*	--*	-1.461	0.018	-1.185	0.017	-1.171	0.015
VTTS	--*		\$10.15/hr		\$9.98/hr		\$9.84/hr	
LL	--*		-6808.1		-10135.9		-9393.1	
Adj. ρ^2	--*		0.509		0.269		0.322	

* We could not obtain estimates for this dataset (unidentifiable model according to BIOGEME)

Next we estimate the coefficients using the data from near-orthogonal designs O1, and O2. As expected, we were not able to estimate the model for design O1 since no information on trade-offs is collected from choice tasks with dominant alternatives. Maybe more surprising is the fact that we could also not estimate the model using design O2, despite containing three choice tasks without a dominant alternative. Although the percentage of choice tasks with dominant alternatives is smaller in this dataset than the datasets related to designs R1 and R2, the heterogeneous random designs contain more variation in the data and hence sufficient information is captured in order to estimate the coefficients.

Finally, looking at model estimates using data from the efficient designs, we can again see that the estimates for the time and cost coefficients for design E1 containing dominant alternatives are somewhat inflated, while they are close to the ‘true’ values for designs E2 and E3 that do not contain dominant alternatives. The VTTS resulting from design E1 is statistically different from \$10/hr at the 95 per cent significance level. The analyst would be better off removing the choice tasks with dominant alternatives from the dataset. On the reduced dataset the coefficients are estimated to be $\beta_T = -0.196$ and $\beta_C = -1.176$ (VTTS is \$10.01/hr. In other words, failure to remove choice tasks with dominant alternatives from the dataset may lead to biased estimates in the MNL model. This suggests that the analyst is better off using design E2 (or E3) instead of design E1, even though design E1 seemingly has a higher efficiency (lower D-error).

The simulation results reported thus far are based on 2,500 simulated respondents, and 20,000 choice observations. Such a sample size is generally larger than those typically used in practice. In order to test the performance of the results on sample sizes more in tune with those used in practice, additional simulation runs assuming 500 respondents were also conducted. The results of these simulations are given in Table A1 in Appendix A. The results of these simulations demonstrate a high level of consistency, suggesting that our results hold for sample sizes typical of most empirical studies.

6.2 Estimates for the regret-scaled multinomial logit model

Using the same simulated dataset, we now estimate our newly proposed RS-MNL model, see Table 7. Again, the coefficients could not be estimated using designs O1 and O2 due to the lack of information on trade-offs in the dataset.

The first thing we observe when comparing the RS-MNL model estimates with the MNL model estimates (Table 6) is that the model fits in the designs without dominant alternatives (i.e., R3, E2, and E3) are the same. In these cases the model parameter γ is not statistically significant and hence the RS-MNL model collapses to the MNL model. This is encouraging and means that the RS-MNL model is able to replicate the ‘true’ values, and can ‘safely’ be used on datasets with and without dominant alternatives. In all other cases, the model fit of the RS-MNL model is superior over the MNL model. This result is of course strongly related to the fact that by design, our choice simulation process mimics the behaviour assumed by the RS-MNL. We present results on subsets of 500 respondents in Table A2 in Appendix A, which confirm our findings on the dataset with 2,500 respondents.

To illustrate the RS-MNL model results we look at design E1. Table 8 shows the scale parameters in the RS-MNL model for each choice task (see Equation (16)) which are between 1.013 and 1.121 for choice tasks without a dominant alternative and equals 4.944 for choice tasks with a dominant alternative. Through this scaling, the RS-MNL model is able to move

the choice probabilities to 0 and 1 for choice tasks with a dominant alternative, while keeping the choice probabilities of the other choice tasks mostly unaffected. We can see that the RS-MNL model is much better able to replicate the simulated choices than the MNL (which has a fixed scale parameter equal to 1). The scaled coefficient for time for choice tasks without dominant alternatives is between $-0.189 \times 1.013 = -0.191$ and $-0.189 \times 1.121 = -0.211$, and for cost between $-1.140 \times 1.013 = -1.155$ and $-1.140 \times 1.121 = -1.278$. These values are much closer to the ‘true’ values than the coefficients for the MNL model. Further, the VTTS of \$9.94/hr is not statistically different from \$10/h. Hence, the RS-MNL model seems to reduce bias in model estimates compared to the MNL model. However, the model estimates for designs R1 and R2 (with a high percentage of choice tasks containing a dominant alternative) still produce values for VTTS that are statistically different from \$10/hr, hence the RS-MNL model is not capable of entirely removing the bias.

Table 7 – RS-MNL estimates on simulated datasets (2,500 respondents, 20,000 observations per design)

	Design R1		Design R2		Design R3		Design O1	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Time (β_T)	-0.123	0.007	-0.120	0.0069	-0.203	0.009	--*	--*
Cost (β_C)	-0.709	0.041	-0.675	0.0397	-1.23	0.053	--*	--*
Exp (γ)	1.223	0.077	1.270	0.0783	-0.070	0.077	--*	--*
VTTS	\$10.43/hr		\$10.63/hr		\$9.89/hr		--*	
LL	-4285.5		-4143.3		-11367.0		--*	
Adj. ρ^2	0.691		0.701		0.180		--*	

	Design O2		Design E1		Design E2		Design E3	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Time (β_T)	--*	--*	-0.189	0.004	-0.191	0.008	-0.200	0.008
Cost (β_C)	--*	--*	-1.140	0.023	-1.147	0.050	-1.214	0.044
Exp (γ)	--*	--*	0.130	0.017	0.060	0.077	-0.072	0.068
VTTS	--*		\$9.94/hr		\$10.01/hr		\$9.86/hr	
LL	--*		-6413.7		-10135.6		-9392.5	
Adj. ρ^2	--*		0.537		0.269		0.322	

*We could not obtain estimates for this dataset (unidentifiable model according to BIOGEME)

Table 8 – Estimated scale parameters and choice probabilities for design E1 (grey shading indicates a choice task containing a dominant alternative)

s	Simulation	MNL model		RS-MNL model	
		λ_s	P_{sj}	λ_s	P_{sj}
1 & 2	0.354	1.000	0.339 (-0.016)	1.013	0.368 (+0.014)
	0.646	1.000	0.661 (+0.016)	1.013	0.632 (-0.014)
3 & 4	0.142	1.000	0.096 (-0.046)	1.081	0.137 (-0.005)
	0.858	1.000	0.904 (+0.046)	1.081	0.863 (+0.005)
5 & 6	0.069	1.000	0.042 (-0.028)	1.121	0.063 (-0.007)
	0.931	1.000	0.958 (+0.028)	1.121	0.937 (+0.007)
7 & 8	0.000	1.000	0.063 (+0.063)	4.944	0.000 (+0.000)
	1.000	1.000	0.937 (-0.063)	4.944	1.000 (-0.000)

7. Results from empirical dataset

Using the choice observations from our online survey as described in Section 5.4, we estimate coefficients in an MNL model as well as our RS-MNL model based on (i) a pooled dataset, (ii) a reduced pooled data set in which observations from respondents that selected one or more dominated alternatives were removed, and (iii) separate datasets for each of the four experimental designs.

7.1 Estimates on pooled dataset

Table 9 summarises the estimates for both models on a pooled dataset of all four experimental designs. The RS-MNL model has a better model fit than the MNL model measured by the log-likelihood value and the adjusted ρ^2 . The exponent γ is positive and significant, which means that scale is not constant over all choice tasks but needs to be adjusted for choice tasks containing dominant alternatives. With $\gamma = 0.055$, the scale parameter λ_s is between 1.055 and 1.078 for all choice tasks that do not contain a dominant alternative, while the scale is between 1.369 and 1.496 for choice tasks that include a dominant alternative. This means that for choice tasks without a dominant alternative, $\lambda_s \beta_T$ is between 0.190 and 0.194, while $\lambda_s \beta_C$ is between 1.132 and 1.348. For choice tasks with a dominant alternative, $\lambda_s \beta_T$ is between 0.246 and 0.269, while $\lambda_s \beta_C$ is between 1.711 and 1.870. The MNL estimates, which assume $\lambda_s = 1$, fall as expected between these values. Clearly, including dominant alternatives in the dataset impacts upon scale and inflates the MNL coefficients. The VTTSs are \$8.83 and \$8.67 per hour respectively for the MNL and RS-MNL model, which are not statistically different.

Table 9 – Estimates on pooled empirical dataset (360 respondents, 5,760 observations)

	MNL		RS-MNL	
	<i>coeff.</i>	<i>s.e.</i>	<i>coeff.</i>	<i>s.e.</i>
<i>Time (β_T)</i>	-0.206	0.0059	-0.159	0.0103
<i>Cost (β_C)</i>	-1.400	0.0335	-1.100	0.0633
<i>Exponent (γ)</i>	--	--	0.055	0.0130
<i>VTTS</i>	\$8.83/hr		\$8.67/hr	
<i>Loglikelihood</i>	-2115.6		-2099.4	
<i>Adjusted ρ^2</i>	0.470		0.473	

7.2 Estimates on reduced pooled dataset

As described in Section 5.4, there were 51 (out of 360) respondents with one or more spurious choices. We cleaned the dataset by removing all choice observations from these 51 respondents (so not only the 1.3 per cent spurious observations), which amounts to a removal of 14.2 per cent of all observations. Table 10 presents the estimates for the MNL and RS-MNL model. The RS-MNL model has a much better model fit than the MNL model. The difference in model fit is much larger than in Table 9, which – in line with intuition – suggests that the presence of the spurious choices actually diminishes the problem of scale inflation.

The γ parameter in the RS-MNL model is significantly higher than the value in Table 9 since much larger scale differences are estimated over different choice tasks. For choice tasks

without a dominant alternative it holds that $1.369 \leq \lambda_s \leq 1.496$, while $11.355 \leq \lambda_s \leq 26.758$ for choice tasks that contain a dominant alternative. Clearly, without the spurious choice observations, the RS-MNL model is much better able to distinguish between choice tasks with and without dominant alternatives, and the VTTS values grow further apart (and become statistically different).

Table 10 – Estimates on reduced pooled empirical dataset (309 respondents, 4,944 observations)

	MNL		RS-MNL	
	<i>coeff.</i>	<i>s.e.</i>	<i>coeff.</i>	<i>s.e.</i>
<i>Time</i> (β_T)	-0.236	0.0072	-0.061	0.0089
<i>Cost</i> (β_C)	-1.610	0.0416	-0.431	0.0634
<i>Exponent</i> (γ)	--	--	0.540	0.0734
<i>VTTS</i>	\$8.80/hr		\$8.45/hr	
<i>Loglikelihood</i>	-1609.4		-1501.0	
<i>Adjusted ρ^2</i>	0.531		0.563	

7.3 Estimates on separate datasets for each experimental design

Next, we estimated models using data from each design separately (without removing spurious choices), see Table 11. All estimates are statistically significant. Looking at the MNL model, the first thing we notice is that the VTTS values are statistically different when using data from different designs. Looking at the coefficients, scale has a clear influence on the estimates, where designs with more dominant alternatives means higher scale and therefore a more deterministic choice. Using these coefficients in prediction will lead to quite different choice probabilities. The RS-MNL model has a better model fit in all four designs, especially in the datasets resulting from designs O1 and E1.

While in our simulated dataset no estimates could be obtained using design O1 (since all choice tasks have a dominant alternative which was by design of the simulated data generating process never chosen), there is no problem estimating the coefficients using the empirical dataset based on the same design. This is due to the fact that the dominant alternative was not always chosen by the respondents; the erroneous choices help identify the model. This indicates that choice tasks with dominant alternatives can actually contain information in empirical applications. Simply removing them would therefore lead to loss of information.

Looking at the estimates for design O2, γ is negative in the RS-MNL model. This means that scale is small when dominancy in a choice task is large. We attribute this counterintuitive result to the existence of spurious choices. However, γ is small such that scale differences are very small.

The estimates obtained from design E2 are quite similar across the MNL and the RS-MNL model. Since this dataset does not contain any dominant alternatives, there do not seem to be scaling issues in the MNL model, such that the models perform similarly and the VTTS values are identical. An exponent of $\gamma = 0.672$ means that there are some scale differences across choice tasks, where the scale parameter takes the values $3.625 \leq \lambda_s \leq 5.297$. In other

words, the relative difference between the highest and the lowest scale is with 46 per cent quite small, such that the results are not so different from the MNL results.

Table 11 – Estimates on empirical datasets (180 respondents, 1,440 observations per design)

	MNL							
	Design O1		Design O2		Design E1		Design E2	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Time (β_T)	-0.385	0.066	-0.256	0.164	-0.183	0.010	-0.143	0.010
Cost (β_C)	-2.660	0.260	-1.510	0.083	-1.180	0.051	-1.170	0.066
VTTs	\$8.68/hr		\$10.17/hr		\$9.31/hr		\$7.33/hr	
LL	-168.3		-506.3		-602.4		-741.5	
Adj. ρ^2	0.831		0.491		0.396		0.256	

	RS-MNL							
	Design O1		Design O2		Design E1		Design E2	
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.
Time (β_T)	-0.748	0.077	-0.610	0.262	-0.019	0.016	-0.032	0.017
Cost (β_C)	-4.680	0.399	-3.290	1.310	-0.128	0.093	-0.258	0.136
Exp (γ)	-0.046	0.003	-0.022	0.003	0.719	0.020	0.672	0.234
VTTs	\$9.59/hr		\$11.12/hr		\$8.88/hr		\$7.33/hr	
LL	-130.7		-505.4		-588.2		-736.4	
Adj. ρ^2	0.866		0.491		0.404		0.259	

8. Extensions

So far we have focussed primarily on dominant alternatives in case of unlabelled experiments and linear utility functions. In this section we extend our regret measure to include nonlinear effects and labelled experiments. We show that dominance checks for these cases often require more information than just the signs of the coefficients. Further, we argue that dominance is much less of an issue in labelled experiments.

8.1 Nonlinear effects

In Equation (1) we defined $\Delta_{ns,j \leftarrow i,k} = \beta_{nk} (x_{nsjk} - x_{nsik})$, which is applicable to utility functions that are linear in the parameters and linear in the attributes. In this section we discuss extensions to include nonlinear effects and interaction effects for continuous attributes, and nonlinear effects for discrete attributes.

A more general definition for $\Delta_{ns,j \leftarrow i,k}$ is given as follows:

$$\Delta_{ns,j \leftarrow i,k} = V_{nsj}(\mathbf{x}_{nsj} | \boldsymbol{\beta}_n) - V_{nsj}(\mathbf{x}_{nsj} + (x_{nsik} - x_{nsjk})\mathbf{1}_k | \boldsymbol{\beta}_n), \quad (18)$$

where $\mathbf{1}_k$ is a $K \times 1$ vector with zeros, except in row k where it takes on value 1. Profile $\mathbf{x}_{nsj} + (x_{nsik} - x_{nsjk})\mathbf{1}_k$ is identical to profile \mathbf{x}_{nsj} except that the level for attribute k is replaced by the level of that attribute in profile \mathbf{x}_{nsi} . This equation allows for more general nonlinear

utility functions. For example, consider $V_{nsj} = \beta_{n1}x_{nsj1} + \beta_{n2}\log(x_{nsj2}) + \beta_{n3}x_{nsj1}x_{nsj2}$, which includes an interaction effect. Then we can calculate $\Delta_{ns,j\leftarrow i,1} = (\beta_{n1} + \beta_{n3}x_{nsj2})(x_{nsj1} - x_{nsi1})$ and $\Delta_{ns,j\leftarrow i,2} = \beta_{n2}\log(x_{nsj2}/x_{nsi2}) + \beta_{n3}x_{nsj1}(x_{nsj2} - x_{nsi2})$. If the coefficients are all positive (without knowing the exact value), then $\Delta_{ns,j\leftarrow i,k} \geq 0$ if $x_{nsjk} \geq x_{nsik}$ for both attributes. Similarly, if all coefficients are negative, then $\Delta_{ns,j\leftarrow i,k} \geq 0$ if $x_{nsjk} \leq x_{nsik}$. If β_{n1} and β_{n2} are both positive (negative) and β_{n3} is negative (positive), then determining the sign of $\Delta_{ns,j\leftarrow i,k}$ relies on more exact knowledge regarding the values of these coefficients. Hence, in case of nonlinear utility functions, checking for dominance may rely on more knowledge than only knowing the signs of the coefficients.

Often analysts use dummy or effects coding to model nonlinear effects or to handle qualitative attributes. Consider attribute k with L_k levels. Then these coding schemes replace a single coefficient β_{nk} with a series of coefficients $[\beta_{nk}^{(1)}, \dots, \beta_{nk}^{(L_k-1)}]'$, where we assume that the last level L_k is the reference level. This means that if x_{nsjk} is equal to level $l < L_k$ then this contributes $\beta_{nk}^{(l)}$ to utility, and if $l = L_k$ then it contributes 0 (in case of dummy coding) or $-\sum_{l=1}^{L_k-1} \beta_{nk}^{(l)}$ (in case of effects coding) to utility. For example, suppose that x_{nsjk} equals level $l_1 < L_k$ and that x_{nsik} equals $l_2 < L_k$. Then it holds that $\Delta_{ns,j\leftarrow i,k} = \beta_{nk}^{(l_1)} - \beta_{nk}^{(l_2)}$. If $l_2 = L_k$, then $\Delta_{ns,j\leftarrow i,k} = \beta_{nk}^{(l_1)}$ (in case of dummy coding) or $\Delta_{ns,j\leftarrow i,k} = \beta_{nk}^{(l_1)} - \sum_{l=1}^{L_k-1} \beta_{nk}^{(l)}$ (in case of effects coding). In all cases, we can check for dominance by determining the sign of $\Delta_{ns,j\leftarrow i,k}$, which is possible if we have knowledge of the ordering of $\beta_{nk}^{(l)}, l \in \{1, \dots, L_k - 1\}$. Note that requires more information than just the sign of each $\beta_{nk}^{(l)}$. In many cases such information can be easily obtained (e.g., for an attribute ‘comfort’ with levels ‘low’, ‘medium’, and ‘high’) while in other cases such information may not be readily available (e.g., for an attribute ‘colour’ with levels ‘blue’, ‘red’, ‘yellow’).

8.2 Labelled alternatives

Dominance checks can quite easily be applied to labelled experiments as well. Labelled experiments often include alternative specific constants, some attributes may not appear in all alternatives, and coefficients for the same attribute may be different across alternatives.

Let β_{nj0} denote the alternative specific constant for respondent n in alternative j , where at least one of these constants (for a chosen reference alternative) is normalised to zero. Then we can calculate $\Delta_{ns,j\leftarrow i,k} = \beta_{nj0} - \beta_{ni0}$. Similar as in the case of dummy or effects coding, for dominance checks we can determine the sign of $\Delta_{ns,j\leftarrow i,k}$ if we know the ordering of $\beta_{nj0}, j \in \{1, \dots, J\}$. It is important to note that if a dummy coded attribute is included in the utility function of alternative j , the alternative specific constant β_{nj0} is confounded with the reference level of the dummy coded attribute. This means that the constant does not only include the ‘brand’ value of the label, but also includes utility associated with the reference level of the dummy coded attribute. This does not allow proper comparison of alternative specific constants in order to assess dominance. To overcome this issue, any dummy coding should first be converted into effects coding (which can always be easily be done). Effects

coding ensures that the constant is no longer confounded with β_{nj0} and allows proper comparison of the constants.

In case a certain attribute k does not exist in alternative i , we can simply set $x_{nsjk} = 0$, such that $\Delta_{ns,j \leftarrow i,k} = \beta_{nk} x_{nsjk}$. Then we only need to know the sign of β_{nk} in order to check for dominance. If the attribute exists in both alternatives i and j but they have different coefficients β_{njik} and β_{nik} , then $\Delta_{ns,j \leftarrow i,k} = \beta_{njik} x_{nsjk} - \beta_{nik} x_{nsik}$. The sign of $\Delta_{ns,j \leftarrow i,k}$ can be determined if we know the signs of β_{njik} and β_{nik} , as well as the ratio $\beta_{njik} / \beta_{nik}$.

Alternative j cannot be dominant if there exists a $k \in \{0, \dots, K\}$ and an $i \neq j$ for which holds that $\Delta_{ns,j \leftarrow i,k} < 0$. Therefore, we can immediately rule out all alternatives j where $\beta_{nj0} < \max_i \{\beta_{ni0}\}$. In other words, the only possible candidates for dominant alternatives are those with the largest alternative specific constant. For example, consider the four alternatives ‘car route 1’, ‘car route 2’, ‘train’, and ‘stay home’, and assume the alternative specific constants are respectively 2, 2, 1, and 0, respectively. This means that in general ‘car’ is preferred over ‘train’, and that both modes are preferred over (reference alternative) ‘stay home’. Then ‘train’ and ‘stay home’ cannot be dominant, and the only alternatives to check for dominance are ‘car route 1’ and ‘car route 2’. Further, suppose that ‘car route 1’ and ‘car route 2’ include a price attribute (e.g., fuel cost, toll cost), which does not occur in ‘stay home’. Assuming that the coefficient for price is negative, the ‘car’ alternatives cannot be dominant since there is regret on the price attribute compared to ‘stay home’. Hence, in this example there do not exist any dominant alternatives no matter what levels the attributes have. If the ‘stay home’ alternative does not exist and both the ‘car’ and ‘train’ alternatives have time and cost attributes with generic coefficients, then the ‘car’ alternatives could be dominant depending on the attribute levels and checks are needed.

9. Discussion, recommendations, and limitations

9.1 Summary and discussion

In this paper we have discussed the impacts of the existence of choice tasks with dominant alternatives stated choice studies, in particular in unlabelled experiments where such issues can easily arise. In a simple case study with simulated choices we showed that dominant alternatives could lead to biased model estimates due to the discrepancy between actual behaviour (which is heteroscedastic, having a small error term variance in the case of dominant alternatives) and assumed homoscedasticity in the model.

We discussed three ways of dealing with dominant alternatives. First, the analyst could simply make sure that dominant alternatives do not exist in the stated choice data. To this end, we proposed a new D^* -efficient design method, in which we use minimum regret as a measure to detect and eliminate choice tasks that contain a dominant alternative. This regret measure can be used for both unlabelled and labelled experiments with linear and nonlinear utility functions.

Secondly, the analyst can simply clean the data such that (i) choice tasks with a dominant alternative are removed, or (ii) all choice tasks of respondents that fail to choose a dominant alternative are removed. In our empirical analysis we show that a choice task with a dominant alternative may actually contain information, in contrast to common belief. A requirement

seems that the dominant alternative is not chosen by all respondents in the dataset. Removing all choice tasks with dominant alternatives may therefore result in information loss. If we would have used this strategy to clean the dataset of our near-orthogonal design, we would not have been able to estimate the coefficients. If the analyst removes only data from certain respondents, dominant alternatives may still exist in the dataset. Such dominant alternatives typically decreases error variance and as such increases scale, leading to biased estimates in the MNL model.

Thirdly, the analyst can compensate for scale differences in the model. We proposed a regret-scaled (RS-) MNL model, in which scale increases with a decrease in normalised minimum regret. We further proposed a smooth approximation of this normalised minimum regret in order to avoid numerical problems in model estimation. Our simulation and empirical results show that our RS-MNL model improves model fit and seem to appropriately account for scale differences.

9.2 Recommendations

Based on these findings, we would strongly recommend using a (Bayesian) D^* -efficient design instead of a random, (near-)orthogonal or D -efficient design in stated choice studies, since this avoids dominant alternatives particularly in case of unlabelled experiments. Further, when a dataset includes dominant alternatives, we suggest not removing these choice tasks (since they contain some information), but rather adopting our RS-MNL model that automatically accounts for scale differences.

9.3 Limitations

In our study we have focussed on limitations of the MNL model. Clearly, more advanced discrete choice models exist. Therefore, we only demonstrated the impacts on the MNL model assuming homogeneous preferences. However, we argue that dominance has an impact on all homoscedastic models. The theory and methods in our paper can be applied to each individual respondent (indicated by subindex n), and as such can be applied to for example latent class models with discrete groups of heterogeneous users or to mixed logit models with continuous preference heterogeneity. Therefore, results in this paper are expected to translate to more advanced models.

Furthermore, we have only focussed on dominance in isolation. In our empirical dataset, many other behavioural processes may have led to the actual choices, including non-trading, lexicographic, or inconsistent behaviour. We can therefore not guarantee that our observed scale differences are purely the result of the presence of dominant alternatives, but may also be the result of learning, fatigue, and other effects.

Acknowledgments

We would like to thank Tony Marley for his useful comments to earlier drafts of this paper.

References

Axhausen, K.W. and J.W. Polak (1991) Choice of parking: stated preference approach. *Transportation*, Vol. 18, pp. 59-81.

- Bekhor, S., and Y. Shiftan (2010) Specification and estimation of mode choice model capturing similarity between mixed auto and transit alternatives. *Journal of Choice Modelling*, Vol. 3(2), pp. 29-49.
- Bierlaire, M. (2003) BIOGEME: A free package for the estimation of discrete choice models, *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.
- Bliemer, M.C.J., and J.M. Rose (in press) On determining priors for the generation of efficient stated choice experimental designs. *Journal of Choice Modelling*.
- Bliemer, M.C.J., and Rose, J.M. (2011) Experimental design influences on stated choice outputs: an empirical study in air travel choice. *Transportation Research Part A*, 45, pp. 63-79.
- Bradley, M., and A. Daly (1994) Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. *Transportation*, Vol. 21, pp. 167-184.
- Burge, P., and C. Rohr (2004) DATIV: SP design: Proposed approach for pilot study. Tetra-plan in cooperation with RAND Europe and Gallup A/S.
- ChoiceMetrics (2012) *Ngene 1.1.1 User Manual & Reference Guide*, Australia.
- Chorus, C.G. (2010) A new model of random regret minimization. *European Journal of Transport and Infrastructure Research*, 10(2), pp. 181-196.
- Chorus, C.G., T.A. Arentze, and H.J. Timmermans (2008) A random regret-minimization model of travel choice. *Transportation Research Part B*, Vol. 42(1), pp. 1-18.
- Chorus, C.G. (2012) Logsums for utility-maximizers and regret-minimizers, and their relation with desirability and satisfaction. *Transportation Research Part A: Policy and Practice*, Vol. 46(7), 1003-1012.
- Cook, J. (2011) *Basic properties of the soft maximum*. Working document, Department of Biostatistics, University of Texas, Houston, USA.
- Cook, R.D., and C.J. Nachtsheim (1980) A comparison of algorithms for constructing exact D-optimal designs. *Technometrics*, Vol. 22, pp. 315-324.
- Crabbe, M., and M. Vandebroek (2012) Using appropriate prior information to eliminate choice sets with a dominant alternative from D-efficient designs. *Journal of Choice Modelling*, Vol. 5(1), pp. 22-45.
- Federov, V.V. (1972) *Theory of optimal experiments*. Academic Press, New York.
- Foster, V., and S. Mourato (2002) Testing for consistency in contingent ranking experiments. *Journal of Environmental Economics and Management*, Vol. 44, pp. 309-328.
- Hensher, D.A., P.O. Barnard, and T.P. Truong (1988) The role of stated preference methods in studies of travel choice. *Journal of Transport Economics and Policy*, Vol. 22, pp. 45-57.

- Hensher, D.A. (2006) Towards a practical method to establish comparable values of values of travel times savings from stated choice experiments with differing design dimensions. *Transportation Research Part A*, Vol. 40, pp. 829-840.
- Hess, S., M. Fowler, T. Adler, and A. Bahreinian (2012) A joint model for vehicle type and fuel type choice: evidence from a cross-nested logit study. *Transportation*, Vol. 39, pp. 593-625.
- Huber, J., J.W. Payne, and C. Puto (1982) Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, Vol. 9(1), pp. 90-98.
- Huber, J., and Zwerina, K. (1996) The importance of utility balance and efficient choice designs. *Journal of Marketing Research*, Vol. 33, pp. 307–317.
- Johnson, F.R., and K.E. Mathews (2001) Sources and effects of utility-theoretic inconsistency in stated-preference surveys. *American Journal of Agricultural Economics*, Vol. 83, No. 5, pp. 1328-1333.
- Louviere, J.J., Hensher, D.A., and Swait, J.D. (2000) *Stated Choice Methods: Analysis and Application*, Cambridge University Press, Cambridge.
- Loomes, G., and R. Sugden (1982) Regret Theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 92, 805-824.
- Luce, R.D. (1959) *Individual choice behaviour: A theoretical analysis*. John Wiley and Sons, New York, USA.
- Marley, A.A.J., T.N. Flynn, and J.J. Louviere (2008) Probabilistic models of set-dependent and attribute-level best-worst choice. *Journal of Mathematical Psychology*, Vol. 52, pp. 281-296.
- McFadden, D. (1974) Conditional logit analysis of qualitative choice behaviour. In: Zarembka, P. (ed.). *Frontiers in Econometrics*, Academic Press, New York, 105-142.
- Rose, J.M., and M.C.J. Bliemer (2009) Constructing efficient stated choice designs. *Transport Reviews*, 29(5), pp. 587–617.
- Rose, J.M. and M.C.J. Bliemer (2013) Sample size requirements for stated choice experiments. *Transportation*, Vol. 40(5), pp. 1021-1041.
- Rose, J.M., M.C.J. Bliemer, D.A. Hensher, and A. Collins (2008) Designing Efficient Stated Choice Experiments in the Presence of Reference Alternatives. *Transportation Research Part B*, Vol. 42, pp. 395-406.
- Rose, J.M., S. Hess, and A.T. Collins (2013) What if my model assumptions are wrong? The impact of non-standard behaviour on choice model estimation. *Journal of Transport Economics and Policy*, Vol. 47(2), pp. 245-263.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, Vol. 27(3), pp. 379-423.

Sándor, Z., and M. Wedel (2001) Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research*, Vol. 38, pp. 430-444.

Swait, J., and Adamowicz, W. (2001) Choice environment, market complexity and consumer behaviour: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organization Behavior and Human Decision Processes*, Vol. 86, pp. 141-167.

Van Cranenburgh, S., C.A. Guevara, and C.G. Chorus (2015) New insights on random regret minimization models. *Transportation Research Part A*, Vol. 74, pp. 91-109.

Van de Kaa, E.J. (2006) Assessment of the value of travel time from stated choice surveys: the impact of lexicographic answering. *Proceedings of the European Transport Conference*, Strasbourg, France.

Walker, J.L., Y. Wang, M. Thorhauge, and M. Ben-Akiva (2015) D-efficient or deficient? A robustness analysis of SP experimental design in a VOT estimation context. Presented at the *Annual Meeting of the Transportation Research Board*, Washington DC, USA.

Appendix A

Table A1 – MNL estimates on simulated datasets (500 respondents, 4,000 observations per sample per design)

Sample	Time (β_T)		Cost (β_C)		VTTS	LL	Adj. ρ^2
	coeff.	s.e.	coeff.	s.e.			
<i>Design R1</i>							
1	-0.3506	0.0128	-2.0473	0.0714	10.27	-1060.1	0.6177
2	-0.4039	0.0151	-2.2839	0.0824	10.61	-962.5	0.6528
3	-0.3690	0.0136	-2.1303	0.0760	10.39	-980.3	0.6464
4	-0.3842	0.0144	-2.1990	0.0782	10.48	-972.0	0.6494
5	-0.3671	0.0135	-2.0792	0.0735	10.59	-1009.4	0.6359
Pooled	-0.3740	0.0062	-2.1400	0.0340	10.49	-4988.8	0.6401
<i>Design R2</i>							
1	-0.4028	0.0147	-2.2188	0.0792	10.89	-994.2	0.6414
2	-0.4015	0.0149	-2.1847	0.0783	11.03	-977.9	0.6473
3	0.3698	0.0137	-2.1715	0.0774	10.22	-993.6	0.6416
4	-0.4018	0.0150	-2.3183	0.0830	10.40	-968.1	0.6508
5	-0.4069	0.0151	-2.3240	0.0842	10.50	-962.0	0.6530
Pooled	-0.3957	0.0066	-2.2396	0.0359	10.60	-4901.5	0.6464
<i>Design R3</i>							
1	-0.2109	0.0086	-1.2499	0.0463	10.12	-2240.1	0.1921
2	-0.1890	0.0081	-1.1775	0.0440	9.63	-2266.7	0.1825
3	-0.2058	0.0085	-1.2305	0.0460	10.04	-2256.9	0.1860
4	-0.1868	0.0082	-1.1525	0.0443	9.73	-2302.8	0.1694
5	-0.1880	0.0081	-1.1300	0.0432	9.98	-2296.2	0.1718
Pooled	-0.1957	0.0037	-1.1868	0.0200	9.89	-11367.4	0.1800
<i>Design E1</i>							
1	-0.2483	0.0075	-1.4771	0.0401	10.09	-1348.3	0.5137
2	-0.2516	0.0075	-1.4506	0.0398	10.41	-1366.0	0.5073
3	-0.2449	0.0074	-1.4643	0.0397	10.03	-1361.8	0.5088
4	-0.2493	0.0075	-1.4961	0.0406	10.00	-1331.7	0.5197
5	-0.2420	0.0072	-1.4224	0.0387	10.21	-1397.1	0.4961
Pooled	-0.2471	0.0033	-1.4614	0.0178	10.15	-6808.1	0.5089
<i>Design E2</i>							
1	-0.1990	0.0065	-1.1970	0.0376	9.97	-2016.9	0.2726
2	-0.2002	0.0066	-1.2312	0.0385	9.76	-1995.1	0.2804
3	-0.2155	0.0067	-1.2473	0.0383	10.37	-1955.3	0.2948
4	-0.1801	0.0063	-1.0824	0.0355	9.98	-2117.8	0.2362
5	-0.1926	0.0065	-1.1771	0.0374	9.82	-2039.8	0.2643
Pooled	-0.1971	0.0029	-1.1849	0.0167	9.98	-10135.9	0.2689
<i>Design E3</i>							
1	-0.2096	0.0062	-1.2238	0.0351	10.28	-1803.2	0.3496
2	-0.1836	0.0057	-1.1429	0.0336	9.64	-1915.8	0.3090
3	-0.1895	0.0058	-1.1604	0.0339	9.80	-1891.0	0.3180
4	-0.1910	0.0058	-1.1622	0.0339	9.86	-1886.4	0.3196
5	-0.1880	0.0058	-1.1700	0.0341	9.63	-1888.1	0.3190
Pooled	-0.1920	0.0026	-1.1706	0.0152	9.84	-9393.1	0.3224

Table A2 – RS-MNL estimates on simulated datasets (500 respondents, 4,000 observations per sample per design)

Sample	Time (β_T)		Cost (β_C)		Exp (γ)		VTTS	LL	Adj. ρ^2
	coeff.	s.e.	coeff.	s.e.	coeff.	s.e.			
Design R1									
1	-0.1003	0.0126	-0.5810	0.0753	1.4675	0.1930	10.36	-889.3	0.6793
2	-0.1438	0.0177	-0.8188	0.1010	1.0799	0.1550	10.54	-836.7	0.6982
3	-0.1270	0.0150	-0.7404	0.0876	1.1487	0.1530	10.29	-843.3	0.6959
4	-0.1401	0.0172	-0.8120	0.0990	1.0949	0.1600	10.35	-849.5	0.6936
5	-0.1148	0.0153	-0.6421	0.0881	1.3312	0.1940	10.72	-862.5	0.6889
Pooled	-0.1233	0.0070	-0.7093	0.0405	1.2229	0.0769	10.43	-4285.5	0.6909
Design R2									
1	-0.1132	0.0153	-0.6353	0.0857	1.3598	0.1900	10.69	-833.4	0.6994
2	-0.1159	0.0153	-0.6360	0.0846	1.3202	0.1830	10.93	-822.6	0.7033
3	-0.1130	0.0137	-0.6501	0.0822	1.2932	0.1690	10.43	-838.2	0.6977
4	-0.1307	0.0162	-0.7477	0.0953	1.1617	0.1600	10.48	-826.1	0.7020
5	-0.1307	0.0168	-0.7390	0.0969	1.1816	0.1670	10.61	-821.0	0.7039
Pooled	-0.1196	0.0069	-0.6751	0.0397	1.2697	0.0783	10.63	-4143.3	0.7011
Design R3									
1	-0.2309	0.0224	-1.3684	0.1300	-0.1777	0.1750	10.12	-2239.5	0.1923
2	-0.1766	0.0184	-1.0997	0.1130	0.1287	0.1780	9.63	-2266.4	0.1826
3	-0.2285	0.0208	-1.3690	0.1230	-0.2114	0.1660	10.02	-2256.1	0.1863
4	-0.1812	0.0184	-1.1174	0.1120	0.0593	0.1770	9.73	-2302.7	0.1695
5	-0.1990	0.0196	-1.2018	0.1170	-0.1129	0.1710	9.93	-2296.0	0.1719
Pooled	-0.2029	0.0089	-1.2305	0.0531	-0.0698	0.0771	9.89	-11367.0	0.1800
Design E1									
1	-0.1913	0.0089	-1.1610	0.0495	0.1224	0.0322	9.88	-1271.6	0.5414
2	-0.1910	0.0107	-1.1225	0.0591	0.1378	0.0561	10.21	-1287.3	0.5357
3	-0.1850	0.0105	-1.1327	0.0598	0.1426	0.0581	9.80	-1281.5	0.5378
4	-0.1930	0.0089	-1.1821	0.0497	0.1197	0.0313	9.79	-1256.9	0.5467
5	-0.1832	0.0089	-1.0984	0.0493	0.1339	0.0365	10.01	-1313.3	0.5263
Pooled	-0.1890	0.0041	-1.1400	0.0230	0.1295	0.0172	9.94	-6413.7	0.5373
Design E2									
1	-0.1982	0.0179	-1.1915	0.1200	0.0084	0.1770	9.98	-2016.9	0.2726
2	-0.2091	0.0185	-1.2924	0.1260	-0.0880	0.1670	9.71	-1995.0	0.2805
3	-0.2069	0.0174	-1.1917	0.1100	0.0885	0.1690	10.42	-1955.2	0.2948
4	-0.1606	0.0154	-0.9535	0.0984	0.2343	0.1800	10.11	-2116.9	0.2365
5	-0.1808	0.0169	-1.1006	0.1080	0.1247	0.1710	9.86	-2039.5	0.2644
Pooled	-0.1915	0.0077	-1.1475	0.0500	0.0598	0.0768	10.01	-10135.6	0.2689
Design E3									
1	-0.2322	0.0216	-1.3436	0.1140	-0.1877	0.1630	10.37	-1802.5	0.3499
2	-0.2015	0.0165	-1.2503	0.0983	-0.1760	0.1450	9.67	-1915.1	0.3093
3	-0.1944	0.0163	-1.1892	0.0952	-0.0483	0.1470	9.81	-1891.0	0.3180
4	-0.1944	0.0171	-1.1817	0.0986	-0.0328	0.1550	9.87	-1886.4	0.3196
5	-0.1900	0.0158	-1.1800	0.0946	-0.0208	0.1470	9.66	-1888.1	0.3190
Pooled	-0.1995	0.0077	-1.2142	0.0442	-0.0724	0.0676	9.86	-9392.5	0.3225