The Journal of Choice Modelling 18 (2016) 51–70

Contents lists available at ScienceDirect

The Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

On the robustness of random regret minimization modelling outcomes towards omitted attributes

Sander van Cranenburgh [a,*], Carlo G. Prato [b]

[a] Delft University of Technology, Transport and Logistics Group, Faculty of Technology, Policy & Management, Jaffalaan 5, 2628 BX Delft, The Netherlands

[b] The University of Queensland, School of Civil Engineering, Brisbane, St Lucia, QLD 4072, Australia

ARTICLE INFO

Article history:
Received 17 June 2015
Received in revised form
19 April 2016
Accepted 20 April 2016
Available online 19 May 2016

Keywords:
Random Regret Minimization
Omitted attributes
Choice elasticity
Forecasting
Alternative specific constants

ABSTRACT

As discrete choice models may be misspecified, it is crucial for choice modellers to have knowledge on the robustness of their modelling outcomes towards misspecification. This study investigates the robustness of Random Regret Minimization (RRM) modelling outcomes towards one sort of model misspecification: the omission of relevant attributes. We explore the effect of omitted attributes (orthogonal and correlated) in the context of labelled and unlabeled data. In the context of labelled data, we show that – just as in RUM models – in RRM models Alternative Specific Constants (ASCs) can be used to capture the average effect of omitted attributes. However, in contrast to RUM models, ASCs in RRM models are choice set composition specific. This implies that in order to achieve consistent parameter estimates when the choice set composition varies across choice observations, different sets of ASCs need to be estimated for each unique choice set composition. In the context of unlabeled data, we show – using Monte Carlo simulations – that RRM models are fairly robust towards the presence of an orthogonal omitted attribute, though not as robust as the linear-additive RUM model. Specifically, we find that: (1) Aggregate Demand Elasticities (ADEs) implied by RRM models are less robust towards the presence of an orthogonal omitted attributes than those implied by linear-additive RUM models, and (2) Average Sample Effects (ASEs) implied by RRM models are – in the presence of an omitted orthogonal attribute – more sensitive towards misspecification in terms of the underlying decision rule than those implied by its linear-additive RUM counterpart.

1. Introduction

Recently, Random Regret Minimization (RRM) based discrete choice models (Chorus, 2010) have been proposed as a counterpart of the canonical Random Utility Maximization (RUM) based choice models. Since their introduction, RRM models have increasingly been used to explain and predict a wide variety of choices such as departure time, route, mode-destination, activity, on-line dating, health-related and policy choices (e.g. Chorus et al., 2011; Kaplan and Prato, 2012; Thiene et al., 2012; Boeri et al., 2013; Hess et al., 2014; Mai et al., 2015 and see Chorus et al., 2014 for a recent overview of applications). RRM models postulate that decision makers aim to minimize regret, which is experienced when one or more non-chosen alternatives outperform the chosen one in terms of one or more attributes. Distinguishing properties of RRM

* Corresponding author.
E-mail addresses: S.vanCranenburgh@tudelft.nl (S. van Cranenburgh), c.prato@uq.edu.au (C.G. Prato).

http://dx.doi.org/10.1016/j.jocm.2016.04.004
1755-5345/

models are that they capture context effects and accommodate for semi-compensatory behaviour (Chorus, 2012).

When estimating choice models, it is crucial that all attributes relevant to the choice are included in the model. Omitting a relevant attribute results in inconsistent estimators, implying that the probability that the parameter estimates are getting closer to the true population parameter values does not increase with increasing sample size. In turn, an omitted attribute will result in larger finite sample bias, possibly leading to erroneous modelling outcomes such as poor market share forecasts and inaccurate estimates of demand elasticities (Washington et al., 2003). In practice however, relevant attributes may be omitted from choice models for various reasons. This happens, for instance, because the choice modeller is not aware of all attributes that matter, is not able to measure all attributes, or uses a third-party data set which does not contain all attributes relevant to the choice.

In this context, it is vital for choice modeller to have knowledge on the robustness of their modelling outcomes in order to be able to adequately interpret the outcomes of their models. The robustness of RUM-based choice modelling outcomes towards the omission of a relevant attribute has frequently been studied (Lee, 1982; Yatchew and Griliches, 1985; Bhat and Guo, 2004; Guevara and Ben-Akiva, 2006; Cramer, 2007; Daly, 2008). Contrary to RUM-based choice modelling outcomes however, the robustness of RRM-based choice modelling outcomes towards the omission of a relevant attribute has – to the authors' knowledge – not been studied. Questions such as "are RRM modelling outcomes relatively less robust towards omitted attributes as compared to RUM modelling outcomes, e.g. due to the fact that RRM models account for context effects?", or "are RRM relatively robust towards one sort of omitted attribute, but little robust towards another?" are yet unanswered. This lack of understanding currently hampers adequate interpretation of RRM modelling outcomes.

The objective of this study is to investigate the robustness of RRM modelling outcomes[1] towards the omission of a relevant attribute. Firstly, we show how to account for the effect of omitted attributes in RRM models in the context of labelled data. Specifically, we show that – in contrast to RUM models – Alternative Specific Constants (ASCs) are choice set composition specific in RRM models. This implies that when the choice set composition varies across choice observations different sets of ASCs need to be estimated for each unique choice set composition to achieve consistent parameter estimates in RRM models. Then, we perform several analyses to develop insights on the robustness of RRM modelling outcomes towards omitted attributes in the context of unlabeled data. We investigate the effect of an omitted attribute on implied elasticities using a series of simulation experiments, as well as by conducting hold-out sample analyses on Revealed Preference (RP) data. Synthetic data sets in our simulation experiments were created using various sorts of omitted attributes, as in practice the omitted attribute may be correlated or uncorrelated with the model's observed attributes. Moreover, acknowledging that in real life the true underlying decision rule is inherently unknown, choices in the synthetic data sets were generated using both RRM and RUM Data Generating Processes (DGPs).

The remainder of this paper is structured as follows. Section 2 provides context with a brief overview of RRM models. Section 3 investigates the effect of omitted attributes in the context of labelled data. Section 4 investigates the effect of omitted attributes in the context of unlabeled data. Finally, Section 5 provides a discussion and addresses further research directions.

2. Overview of RRM models

RRM models are based on the premises that, when choosing, the decision maker n minimizes regret. Regret is experienced when a competitor alternative j outperforms the considered alternative i with regard to attribute m. The overall regret of an alternative is conceived to be the sum of all the pairwise regrets that are associated with bilaterally comparing the considered alternative with the other alternatives in the choice set.

The general form of RRM models is given in Eq. (1), where RR_{in} denotes the random regret experienced by decision maker n considering alternative i, R_{in} denotes the observed part of regret, and ε_{in} denotes the unobserved part of regret (one exception to this form is the RRM specification proposed in Chorus et al., 2008). In the core of RRM models is the so-called attribute level regret function $r_{ijmn} = f\left(\beta_m, x_{jmn} - x_{imn}\right)$. This function maps the difference between the levels of attributes m of the competitor alternatives j and the considered alternative i onto regret.

$$RR_{in} = R_{in} + \epsilon_{in}, \quad \text{where } R_{in} = \sum_{j \neq i} \sum_{m} r_{ijmn} \tag{1}$$

Since the first RRM model (Chorus et al., 2008), a number of different types of RRM models have been proposed in the literature. These models differ from one another in terms of their functional form of the attribute level regret function and/or their associated formula for the choice probabilities. Regardless of the exact type of RRM model however, the convex shapes of the attribute level regret functions are such that losses loom larger than equivalently sized gains. As a result of that, RRM models predict that it is relatively ineffective (in terms of minimizing regret) to improve the performance of an alternative in terms of an attribute on which it already performs relatively well. Therefore, having a strong performance on one attribute does not necessarily compensate for having a poor performance on another attribute (which causes much regret).

Table 2-1 shows the attribute level regret functions of different types of RRM models that have recently been proposed in the

[1] In this paper we are dealing with Multinomial Logit (MNL) specifications as the vast majority of RRM (and RUM) models in the literature are estimated in this form.