A Reference Point Logit Model for Estimating Substitution Probabilities Using Point of Sale Data

Luis E. Castro, University of Miami, Coral Gables, USA
Yuan Ren, Shanghai Dianji University, Shanghai, China
Nazrul I. Shaikh, University of Miami, Coral Gables, USA

ABSTRACT

This article presents a practical approach to estimate the substitution probabilities between products at a retail store by using the store’s point of sale data and prospect theory based structural restrictions on the consumer choice behavior. The prospect theory-based reference dependent preference structure imposed on the consumer choice behavior (a) accounts for how consumers make their original choice as well as how they substitute, (b) eliminates the IIA and IPS assumptions that the standard utility theory based models impose on consumer choice, and (c) alleviates the need for inventory information for estimating the substitution probabilities. Simulations and empirical studies have been used to show that the estimates of the substitution probabilities are efficient and are robust to stock-out rates.

KEYWORDS
Assortment Planning, IIA, IPS, Reference Dependent Preferences, Substitution Probabilities

1. INTRODUCTION

Assortment planning helps retailers decide what products and how many units of each product to stock. It presents an opportunity for retailers to improve profitability through reducing stock-out linked abandonment, decreasing the frequency and depth of markdowns, decreasing the variance in forecasts, and improving the overall efficiency of their supply chain. Several researchers have shown the link between improved assortment plans and improvements in financial metrics. For example, Kök and Fisher (2007) report gains of $155 million in profitability for a grocery chain through improved product availability. Rajaram (2001) reports a 40% increase in profits for a catalog retailer through reduced markdown, and Fisher et al. (2001) report a 100% increase in profits for a large retailer through assortment optimization. Besides these immediate benefits, results also suggest that improvements in assortments lead to an increase in the satisfaction levels amongst buyers which has long-term implications for the retailer (Broniarczyk et al. 1998; Sloot et al. 2006).

Accurate estimates of the demand side parameters, namely the future demand for products and the substitution probabilities among products are two key inputs for generating good assortment plans. However, the task of estimating these two quantities can be difficult due to the limitations in the data collected at most retail stores. Information on non-purchase occasions and whether the product that was sold was what the consumer actually wanted or if it was a substitute is seldom collected by retailers. Therefore, the recorded sales of products that encounter stock-outs are lower than the demand for those products (and the data presents truncated demand information) while the recorded

DOI: 10.4018/IJISSCM.2018100102
sales of products that are used as substitutes are exaggerated. Cooper et al. (2006) show that ignoring
the truncation and/or the exaggeration in sales while using historical sales data for making future
assortment plans could lead to a systematic increase in the gap between what the consumers want
versus what the retailers think they want and this gap impacts the retailer’s revenues. Estimation of
the demand side parameters therefore relies on supplementing the sales data with extra information
such as the inventory data and non-purchase rates to account for the unobserved truncations as well
as substitutions. However, these datasets are oftentimes unavailable or even inaccurate. Our objective
in this paper is to develop a practical approach to estimate the substitution probabilities between
products using store level point of sale (POS) data that is supplemented with structural restrictions on
the consumer’s choice behavior. The structural restrictions account for the impact of the unobserved
truncation and substitutions.

In this paper, we propose a prospect theory-based reference dependent preference structure to
model a consumer’s choice. Prior research has established that (a) prospect theory offers a realistic
framework for modeling a consumer’s choice when there is limited information and/or time pressure
(Thaler 1980), and (b) the inclusion of reference dependency in preferences unifies the concepts of
consumption utility and gain-loss utilities under one framework (Köszegi and Rabin 2006) thereby
making the classic utility a special case of the prospect theory based preference. The reference
dependent preference structure also enables the relaxation of two contentious (in assortment planning
literature) assumptions of independence of irrelevant alternatives (IIA) and invariant proportion of
substitution (IPS) that classic utility theory-based models such as the multinomial logit (MNL) models
impose on a consumer’s choice behavior.

The proposed reference dependent preference structure makes two modifications to the classic
reference point models. First, it makes the reference “point” to be product dependent and yields a
set of reference points, one for each product in the consumers’ choice set. This modification enables
us to explicitly incorporate into the model the assumption that when consumers substitute, they are
likely to give a higher preference to products that are similar to their original choice. These product-
specific reference points also make the substitution probabilities product dependent.

Next, the proposed reference dependent preference structure uses the link between substitutability
and similarity of products (Gaur and Honhon, 2006) to define the reference points in terms of
substitution probabilities. The proposed model can therefore estimate the substitution probabilities
between products without the need for tracking and tagging stock-outs and alleviates the need of
inventory information for estimation.

The paper is organized into five sections. Section 2 provides a brief literature review. Section
3 provides information about the proposed reference point logit (RPL) model and the estimation
procedure. The results are presented in Section 4. We use simulations to establish: (1) that the
parameters of RPL are identified; (2) that RPL leads to efficient as well as robust estimates; (3)
precision of the substitution probabilities estimated using RPL converge asymptotically to the
precision obtained under the hypothetical condition of availability of complete information about the
consumers’ choices process; and (4) RPL performs better than the MNL as well as other models that
rely on exogenously imposed substitution structures. We also exemplify the use of RPL through two
empirical studies that use actual POS data on coffee and potato snack categories. Though researchers
such as Karabati et al. (2009) have addressed the issue of estimating the demand side parameters
using POS data, issues such as identification and efficiency are not addressed for finite size data.
Only claims of “satisfactory” as against efficient estimates have been made. We present the conditions
under which the RPL estimates are efficient under the assumption of finite sample size. Section 5
outlines the conclusions, limitations and topics for future research.
Retail Management and Electronic Retailing
Kijpokin Kasemsap (2018). *Supply Chain Management Strategies and Risk Assessment in Retail Environments* (pp. 139-158).
[www.igi-global.com/chapter/retail-management-and-electronic-retailing/193302?camid=4v1a](www.igi-global.com/chapter/retail-management-and-electronic-retailing/193302?camid=4v1a)

Sustainable Supply Chains
[www.igi-global.com/chapter/sustainable-supply-chains/161153?camid=4v1a](www.igi-global.com/chapter/sustainable-supply-chains/161153?camid=4v1a)