A Variance Component Model for Studying Consumer Behaviour

Adriaan W. Hoogendoorn  
CentER Applied Research  
P.O. Box 90153  
5000 LE Tilburg, the Netherlands  
a.w.hoogendoorn@kub.nl

Dirk Sikkel  
CentERdata  
P.O. Box 90153  
5000 LE Tilburg, the Netherlands  
d.sikkel@kub.nl

1. Introduction

The study of consumer behaviour is a broad field that covers several scientific areas ranging from economics and marketing to psychology and anthropology. Our view in this field is focussed on numbers of product purchases that consumers make. We will try to describe the purchasing process by use of a stochastic model.

Classical theory on consumer behaviour assumes a Poisson process for product purchasing at the individual level. Within the class of Poisson based models the Negative Binomial Distribution (NBD) model, introduced by Ehrenberg in 1959 is one of the most popular models (see Ehrenberg, 1988). The NBD model mixes the Poisson process with a Gamma distribution to allow heterogeneity between individuals. The model can be used as a practical tool to evaluate product sales (see Morrison and Schmidtlein, 1988). However, Poisson based models assume that the time between events (purchases) have an exponential distribution. Since the coefficient of variation for exponential distributed interpurchase times is equal to one, the Poisson assumption implies that the process is assumed to have a certain (high) rate of regularity. The assumed rate of regularity may not be reflected by the empirical data. In its turn this may lead to undesired properties of the parameter estimates (see Hoogendoorn and Sikkel, 1999).

2. A model that models the regularity of the purchasing process

As an alternative we suggest to use a model where the regularity of the process is embedded as a free parameter into the model that we can estimate from the data. Therefore we will assume that the purchasing process is a renewal process where the interpurchase times $X_{i1}, X_{i2}, \ldots$ can be written as

\begin{equation}
\log(X_i) = \mu + U_i + E_i.
\end{equation}

In model (1) the parameter $\mu$ is an overall mean and $U_i$ and $E_i$ are random variables that are stochastically independent and normally distributed with zero mean and standard deviations $\sigma_u$ and $\sigma_e$ respectively. We will refer to model (1) as the Variance Component (VC) model (see Aalen and Husebye (1991)). Thus the VC model splits the total variance into a component between individuals and a component within individuals (in time). The value of $\sigma_u$ corresponds to the amount of
heterogeneity between individuals. The value of \( \sigma_0 \) corresponds to the regularity of the process, being deterministic when it is zero and more and more irregular for higher values of \( \sigma_0 \).

3. Estimating the VC model from count data

One complication in estimating the parameters of model (1) comes from the fact that we have count data. The data consist of numbers of purchases per week reported by the households in a panel survey. We developed a method that generalised moment estimators based on the asymptotic properties known from renewal theory (see Cox, 1962).

The number of parameters in the VC model is three. In order to formulate generalised moment estimators for the model we therefore need three empirical quantities from the data. The question is what three quantities will we use? The first two quantities that come to mind are the (sample) mean and variance of \( M(t) \), the number of purchases in a time period of length \( t \). For the third quantity we examined two alternatives. The first alternative is the penetration. This is the fraction \( p \) of consumers that bought at least one product in a time period of length \( t \). The second alternative is a quantity that we call internal correlation. Therefore we split \( M(t) \) into the variables \( M(0,t_1) \) and \( M(t_1,t) \), the number of events in time interval \([0,t_1]\) and \([t_1,t]\) respectively. When there is no individual heterogeneity the relation between \( M(0,t_1) \) and \( M(t_1,t) \) will be weak, because there is not much dependence between what happens in \([0,t_1]\) and what happens in \([t_1,t]\). When there is individual heterogeneity, this implies a relation between \( M(0,t_1) \) and \( M(t_1,t) \), because at the individual level they share the same latent variable. The internal correlation is the correlation between \( M(0,t_1) \) and \( M(t_1,t) \). We will discuss and compare the behaviour of the alternative methods to the use of Poisson based models using both synthetic and empirical data.


**3. Modèle des composants de la variance pour étudier la conduite des consommateurs.**

Nous étudions la conduite des consommateurs avec une modèle statistique qui tient compte de la regularité du processus des acquérir. La modèle statistique est une modède composant de la variance qui sépare la variance entre des individus (hétérogénité) et à l’intérieur des individus (dans la dimension du temps).