Multivariate Behavioral Research

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hmbr20

Exploratory Tobit Factor Analysis for Multivariate Censored Data

Wagner A. Kamakura & Michel Wedel

Version of record first published: 10 Jun 2010

To cite this article: Wagner A. Kamakura & Michel Wedel (2001): Exploratory Tobit Factor Analysis for Multivariate Censored Data, Multivariate Behavioral Research, 36:1, 53-82

To link to this article: http://dx.doi.org/10.1207/S15327906MBR3601_03

Multivariate Behavioral Research, 36 (1), 53-82

Exploratory Tobit Factor Analysis for Multivariate Censored Data

Wagner A. Kamakura
University of Iowa

Michel Wedel
University of Groningen and University of Michigan

We propose Multivariate Tobit models with a factor structure on the covariance matrix. Such models are particularly useful in the exploratory analysis of multivariate censored data and the identification of latent variables from behavioral data. The factor structure provides a parsimonious representation of the censored data and reduces the dimensionality of the integration required in evaluating the likelihood. In addition, the factor model parameters lend themselves to substantive interpretation and graphical display. The models are estimated with simulated maximum likelihood. Applications to the prescription of pharmaceutical products and the analysis of multi-category buying behavior are provided.

Introduction

Factor analysis has been one of the favored methods for data analysis among behavioral researchers, and is a fundamental tool for psychometricians (c.f., Mulaik, 1972). Factor analysis models were originally developed for continuous (normally distributed) observed variables, but were later extended to binary outcome variables (Bartholomew, 1980), thus allowing behavioral researchers to factor analyze data with continuous or discrete variables. However, some situations require the analysis of data in which each observed variable is a discrete-continuous mixture. This usually happens when analyzing behavioral data. For example, one of the data sets used in the present paper comes from a study pertaining to the prescription of drugs by physicians and contains information on whether or not a physician prescribed a drug and, if so, the volume prescribed. A second example that we provide comes from marketing, where purchase volumes of products in multiple categories are analyzed.

For correspondence regarding this article, the first author may be reached at 108 Pappajohn Business Building, University of Iowa, Iowa City, 52242-1000. The authors are grateful to the editor and two anonymous reviewers for helpful suggestions and comments.

The data contain information on whether a product was purchased, and if so in what volume. Behavioral data, commonly collected and analyzed in the human and social sciences for example in epidemiology (Donovan, 1993), genetics (Waller & Muthén, 1992), health (Mingshan, 1999), psychology (Sher & Wood, 1996), economics (Burket, 1998) and management (Van den Berg & Richardson, 1999), are inherently non-negative and sometimes have a large proportion of zeros. Treating theses zeros as missing data in a factor analysis will obviously produce a biased representation of the data, because the censoring mechanism producing the zeros contains information about the factor pattern. Moreover, analyzing these data with a standard factor model produces biased estimates, given the severe non-normality that results, so that the standard assumptions on the residuals do not hold (Muthén, 1989).

The Tobit Model

This type of data with zero observations is sometimes referred to as mixed type data, having recently attracted much attention in the statistical literature. Examples are the work of Sammel, Ryan and Legler (1997), Sammel and Ryan (1996), Fitzmaurice and Laird (1995), Cox and Wermuth (1992), Arminger and Küsters (1988), and Lance, Cornwell and Mulaik (1988). There, mixed type data are modeled through (bivariate) binomial and normal distributions for the zero and non-zero data values. However, problems in the analysis of mixed type data were already recognized by Tobin (1958). The Tobit model provides the advantage of providing an explicit link between the data-generating mechanism of the zero and non-zero data by offering a variety of specifications of latent variables and censoring mechanisms and restricting the distribution of the non-censored data to have positive support.

Amemiya (1985, Ch. 10) provides a classification of Tobit models that is based on the form of the likelihood function, the number of variables and whether or not each of them drives censoring. The two most common classes of Tobit models are the ones classified by Amemiya as Type-1 and Type-2 models. The Type 1 Tobit is: $y^* = x'\beta + \varepsilon$, with $y = y^*$ if $y^* > 0$ and $y = 0$ otherwise, where y^* is observed only if it is larger than zero and $\varepsilon \sim N(0,\sigma^2)$. The Type 2 Tobit is defined as: $y_1^* = x_1'\beta_1 + \varepsilon_1$, $y_2^* = x_2'\beta_2 + \varepsilon_2$, where only the sign of y_1^* is observed. Further, $y_2 = y_2^*$ is observed if $y_1^* > 0$ and $y_2 = 0$ otherwise. Both models thus explicitly link the distributions of the zero and nonzero data through the censoring mechanism, with the Type-2 model specifying a different latent variable for the censoring mechanism. The type-2 model is thus the more general of the two and is applicable in particular in cases where a different data-generating

mechanism is thought to underlie the zero and nonzero data parts. The increasing availability of micro-level behavioral data has greatly stimulated the interest in these Tobit models in the behavioral sciences (e.g., Amemiya, 1985, Chapter 10; Jones & Possnett, 1991; DeSarbo & Choi, 1999).

The estimation schemes that have been proposed for Tobit models involve two-stage least-squares (Heckman, 1976), nonlinear least-squares (Wales & Woodland, 1980), Markov Chain Monte Carlo Methods (Chib, 1992) and maximum likelihood (Amemiya, 1973). Maximizing the Tobit likelihood offers advantages, since it provides consistent and efficient estimates (cf. Amemiya, 1985, Ch. 10), while for some parameterizations the likelihood is globally concave in the parameters (Olsen, 1978). The Tobit likelihood needs to be maximized iteratively using, for example, the Newton-Raphson algorithm. However, for multivariate Tobit models the numerical evaluation of the likelihood has been difficult or even impossible, since it involves high dimensional integrals. This seems to have hampered the application of Tobit models to high dimensional multivariate behavioral data subject to censoring. Multivariate Tobit models have been previously limited in most cases to three variables, probably for those reasons of computational feasibility. Similar dimensionality problems have plagued the evaluation of other models for mixed type data (cf. Cox & Wermuth, 1992).

We propose a class of multivariate Tobit models that reduces the dimensionality problem and provides a parsimonious summary of high-dimensional censored survey data. The models are tailored to situations where there are J variables measured on N units, and are suited for exploratory analysis where the aim is to derive a number of latent variables that capture the multivariate dependencies among the J observed variables, as in standard factor models (Bartholomew & Knott, 1999; Bartholomew, 1987). Our work is in line with that of Sammel and Ryan (1996), Sammel, Ryan and Legler (1997), Arminger and Küsters (1988), and Lance, Cornwell and Mulaik (1988),who proposed latent variable models for mixed type data.

Maximum likelihood factor analysis was originally developed for continuous (normally distributed) observed variables, factor models for binary outcome variables were developed by Bartholomew (1987), while Muthén (1989) first proposed a factor model for censored data. We extend the pioneering work of Muthén (1989) and Waller and Muthén (1992), who developed a three-stage estimation procedure for confirmatory factor analysis models for censored data. We extend their work in several ways. First, we develop a framework for both exploratory and confirmatory tobit factor modeling. Second, we use a flexible type-2 tobit formulation. Third, we employ a simultaneous estimation procedure applying simulated maximum likelihood (Gouriéroux & Montfort, 1996), to solve the problem of

the evaluation of higher-order integrals involved in the estimation. The factor structure that we impose on the covariance matrix of the unobserved variables not only renders estimation simpler, but lends itself to interesting substantive interpretations and graphical representation.

Our purpose is to add to the factor analysis literature by developing exploratory factor models for outcome variables that are censored, which seems to be particularly useful in behavioral research. With our model one can deal with exploratory and confirmatory ML factor analysis of data with mixed type variables, in which a (large) number of observations equal to zero. We describe the model and estimation procedure next, and provide two applications to behavioral data, on the prescription of pharmaceutical products, and multi category buying behavior, respectively.

Multivariate Tobit Factor Models

Model Specification

Assume a rectangular data matrix Y classified by $n = 1, \ldots, N$ sampling units and $j = 1, \ldots, J$ variables. The observations are realizations of the random variable $y = (y_j)$ and may take on non-negative values. We consider a Type-2 Tobit model. These are more general than Type-1 Tobit models in that they allow for a different data-generating mechanism for the zero and the nonzero data. In particular, we allow for two types of partially observed variables y^1_{nj} and y^2_{nj} that drive the zero and non-zero observations, respectively. These partially observed variables differ in their mean values, η_j and μ_j, and are linked through a common set of latent factors, x_n, causing a common covariance structure among them. Thus we assume that there is a common set of latent factors underlying the zero and non-zero observed data (alternatively, the model could be seen as a set of seemingly unrelated regressions, with truncated outcome variables). A type-1 Tobit factor model can be obtained by constraining $\eta_j = \mu_j$ for all j. If the Type-1 model holds, then the data generation mechanisms for the zero and non-zero data are the same. The different means for the two types of partially observed variables allows the percentage of zero observations for each observed variable to vary independently of the mean. We define the model as follows:

$$
\begin{aligned}
(1) \qquad & y^1_{nj} = \eta_j + u_{nj}, \\
& y^2_{nj} = \mu_j + u_{nj} \\
& y_{nj} = y^2_{nj} \qquad \text{if } y^1_{nj} > 0, \\
& y_{nj} = 0 \qquad \text{if } y^1_{nj} \leq 0.
\end{aligned}
$$

and specify a factor structure on $u_n = (u_{nj})$:

$$u_n = x_n \mathbf{\Lambda}' + \varepsilon_n, \tag{2}$$

with x_n the n^{th} row of the $(N \times P)$ matrix $\mathbf{X}$ representing the scores of the subjects on P latent factors for some assumed value of P, $\mathbf{\Lambda}$ a $(J \times P)$ matrix of fixed parameters, $\mu = (\mu_j)$ and $\eta = (\eta_j)$ $(J \times 1)$ vector with an intercept for each observed variable, and $\varepsilon_n = (\varepsilon_{nj})$ a vector of independent error terms with $\varepsilon_{nj} \sim N(0,\psi^2_j)$. We specify the x_{nj} i.i.d standard normal, so that $E(u_n) = 0$ and $Cov(u_n) = \Psi + \Lambda'\Lambda$, with $\Psi = diag(\psi^2_j)$. The x_{nj} are the latent factors, the λ_{jp} are called the factor weights and ψ^2_j the unique variances for the J variables. We accommodate both exploratory factor models, in which $\mathbf{\Lambda}$ is free, and confirmatory factor models, in which according to prior theory, the number of factors P, as well as the structure of the weight matrix $\mathbf{\Lambda}$ are known. Since Muthén (1989) already dealt with confirmatory factor models, we predominantly restrict attention to the exploratory factor models.

Often in factor analysis one wants to interpret either standardized factor weights: , or the factor loadings, which are defined as the correlation of each partially observed variable with the latent factor in question (Bartholomew, 1987, p. 49):

$$\hat{\mathbf{\Gamma}} = diag\left(\hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}' + \hat{\Psi}\right)^{-\frac{1}{2}} \hat{\mathbf{\Lambda}}\,. \tag{3}$$

We have the conditional density:

$$f\left(y_{nj}|x_n;\mathbf{\Theta}\right) = [\phi(y^2_{nj}|x_n;\mathbf{\Theta})\mathbf{\Phi}(y^1_{nj}|x_n;\mathbf{\Theta})]^{I(y_{nj}>0)}[1-\mathbf{\Phi}(y^1_{nj}|\mathbf{\Theta})]^{I(y_{nj}=0)}, \tag{4}$$

in which Θ contains all parameters and $\Phi(\cdot)$ and $\phi(\cdot)$ are the normal distribution and density functions. The expectation of y_{nj}, given that it is positive, (cf. Amemiya 1985, p. 368) is:

$$E\left(y_{nj}|x_n;y_{nj}>0\right) = \mu_j + x_n\lambda'_j + \psi_j h\left(\frac{\eta_j + x_n\lambda'_j}{\psi_j}\right), \tag{5}$$

with $h(\cdot) = \phi(\cdot)/\Phi(\cdot)$ the hazard rate of the normal distribution. Thus the unconditional expectation equals:

$$E\left(y_{nj}|x_n\right) = \left(\mu_j + x_n\lambda'_j\right)\mathbf{\Phi}\left(\frac{\eta_j + x_n\lambda'_j}{\psi_j}\right) + \psi_j\phi\left(\frac{\eta_j + x_n\lambda'_j}{\psi_j}\right). \tag{6}$$

The likelihood contribution of subject *n* is:

$$(7)\quad L_n = \int \prod_1 \phi\left(y_{nj}^2 | x_n; \boldsymbol{\Theta}\right) \boldsymbol{\Phi}\left(y_{nj}^1 | x_n; \boldsymbol{\Theta}\right) \prod_0 [1 - \boldsymbol{\Phi}(y_{nj}^1 | x_n; \boldsymbol{\Theta})] \phi(x_n; 0,1) dx_n,$$

where $\prod_0$ and $\prod_1$ denote the product over the censored and the uncensored observations, respectively. Note that the proposed factor structure reduces the *J*-tuple normal integral in the full multivariate Tobit model to a *P*-tuple ($P < J$) integral. Our approach is in line with exploratory factor models that have been developed in the psychometrics literature for normal and binomial variables (Bartholomew, 1987; Krzanowski & Mariott, 1995, Ch. 12; Bartholomew & Knott, 1999).

Identification

The standard exploratory factor model suffers from location, scale and rotation indeterminacy (Bartholomew, 1987, p. 97; Bekker, Merckens, Wansbeek, 1994, pp. 84-90). For an extensive discussion about indeterminacy in the factor analysis model, see the debate in this Journal, initiated by Mauran (1996). Here, we expand upon location, scale and rotation invariance in the factor Tobit model described above. The model is invariant under an arbitrary translation of the means of the distribution of x_n. Scale invariance occurs since the scale of Λ and the variance of x_n are not separately identified. Location and scale invariance are alleviated by fixing the mean and variance of the distribution of the latent factors: $x_n \sim \phi(0,1)$. If we assume $\Lambda^* = \Lambda R$ with **R** an orthogonal rotation matrix, we have $f(y_{nj}|x_n;\Theta) = f(y_{nj}|x^*_n;\Theta^*)$, for $x^*_n = x_n R^{-1}$. The distribution of x^*_n equals: $f_x(x^*_n R)$, since the Jacobean is 1. Since the distribution of $x_n R^{-1}$ is the same as that of x_n for *R* orthogonal and x_n normal (Lancaster, 1954), the unconditional distribution of y_{nj} under the new parameterization, $f(y_{nj};\Theta^*)$, is the same as that under the original parameterization, $f(y_{nj};\Theta)$. Thus, the model is rotation invariant. Rather than imposing constraints on the parameters to alleviate this invariance, which provides confirmatory models, we adhere to the convention in the exploratory factor analysis literature to choose from among the set of possible solutions the one that has most substantive meaning (cf. Bartholomew, 1987, p. 96; Krzanowski & Mariott, 1995, p. 138). In reporting the number of parameters we take the identification constraints into account. Note that several sets of $P(P - 1)/2$ constraints imposed on the matrix $\boldsymbol{\Lambda}$ alleviate the rotation indeterminacy and renders the model a confirmatory factor analysis model, in line with Muthén (1989) and Waller and Muthén (1992), which greatly facilitates in obtaining

identified models. Bekker, Mercens and Wansbeek (1994, p.87-88) elaborately discuss linear identifying restrictions for the factor model.

SML Estimation

We estimate the Tobit factor model by maximizing the likelihood functions defined as the product of the individual likelihood functions in Equation 7 across *n*. We use simulation to evaluate the integrals (Gouriéroux & Montfort, 1996). In simulated maximum likelihood (SML) estimation, the likelihood contributions in Equation 7 are approximated by:

$$(8) \qquad L_n \approx \sum_{t=1}^{T} \prod_{1} \phi\left(y_{nj}^2 | z^t; \boldsymbol{\Theta}\right) \boldsymbol{\Phi}\left(y_{nj}^1 | z^t; \boldsymbol{\Theta}\right) \prod_{0} [1 - \boldsymbol{\Phi}(y_{nj}^1 | z^t; \boldsymbol{\Theta})],$$

where z^t is drawn T times from $\phi(0,1)$. An appealing aspect of SML estimation is that the simulated likelihood function in Equation 8 is twice differentiable, simplifying optimization with Newton-type algorithms (details on SML and the first order derivatives of our model are provided in the appendix).

The large-sample properties of ML estimators in factor analysis have been investigated by Anderson and Rubin (1956), and by Gill (1977), who demonstrate consistency of the estimates. Ignoring the censoring mechanism in formulating the likelihood for the model defined by Equations 1 and 2, however, leads to inconsistent estimates. This can be seen from Equation 5, using $x_n \sim \phi(0,1)$, so that the results of Greene (1981) apply. Consistency of the MLE in the factor Tobit model, given an arbitrary restriction on Λ to obtain uniqueness (Bekker, Merckens & Wansbeek, 1994, p. 87), follows from Amemiya (1973), who proved that the ML estimator in the Tobit model is consistent [if the parameter space is compact, (x_n) are uniformly bounded and $\lim_{N\mapsto} X'X/N > 0$].

Gouriéroux and Montfort (1996) describe the application of SML to the estimation of a Tobit model. Their model includes an individual random effect that follows a (standard) normal distribution and arises as a special case of our model for $P = 1$ and $\Lambda = \lambda_0$, a scalar. They show that the SML estimator is consistent if $\sqrt{N} / T \to 0$ as $N \mapsto \infty$ and asymptotically equivalent to the MLE. The bias is of order $1/T$. Simulation studies by Keane (1993) and (Lee, 1995, 1997), show that SML has excellent properties for finite values of T. We use $T = 200$ in the application below (for example, Harris & Keane, 1999, use a similar number in a choice model).

Determination of P

In some cases, prior theory may be available to guide the choice of the number of latent factors, P. In cases where the aim is only to take heterogeneity into account often $P = 1$ is chosen. If the aim is graphical display of the dependency structure of the J *parially observed* variables as in the factor analysis literature (cf. Bartholomew, 1987), $P = 2$ or $P = 3$ is a convenient choice. However, in many cases the value of P needs to be determined from the data. Akaike (1987) argues for the AIC statistic to compare models with various values of P. A limitation of the AIC statistic, however, is that it is not dimension consistent: it does not asymptotically indicate the true model among a set of candidate models (Bozdogan, 1987). In response, several authors have proposed dimension consistent criteria, such as the Bayesian Information Criterion, $BIC = -2 \ln L + [J(P + 2) - P(P - 1)]\ln(JN)$ (Schwartz, 1978), and the very similar Consistent Akaike Information Criterion, *CAIC* statistic (Bozdogan, 1987). Based on the assumptions that model dimensionality is fixed as $N \mapsto \infty$ and that the true model is among the set of candidate models, these statistics indicate the true model with probability one, asymptotically. Dimension consistent criteria have been criticized because the assumption that the true model is among the set of models considered is unlikely to hold in most applications, while in addition high probabilities of selecting the true model accrue only at large sample sizes (Burnham & Anderson, 1998, p. 69). However, in our application below the number of observations is large, a situation where empirical work supports the use of BIC (Rust, Simester, Brodie & Nilikant, 1995). We therefore base model selection on the BIC statistic (reporting AIC as well), which tends to result in more parsimonious models than the AIC statistic. This parsimony is preferable because of the easier interpretability of the factor solution in low-dimensional spaces.

Illustration to the Analysis of Drug Prescription Behavior

Prescription-drug therapy is highly cost-effective as compared to other medical interventions, such as hospitalization or surgery, while it often eliminates the need for those types of treatment. Pharmaceutical manufacturers' sell prescription drugs to wholesalers, who again distribute the products to retail HMOs, hospitals, and clinics. Industry sales reached $124.6 billion in 1998 (PhRMA, 1999). The main drivers of growth in this sector of the pharmaceutical market have been non-price factors such as the increased volume and the changing mix of prescriptions, accounting for 80 percent of growth in 1998 (PhRMA, 1999). Even though prescription drugs

are sold to patients, most of the manufacturers' marketing effort is focused on physicians, who often decide the particular drug to be used by the patient. Prescription drug sales are hardly affected by pricing strategies of the pharmaceutical companies since neither the prescribing physicians nor the patients bear the costs. The American Medical Association is concerned over the increased competition from "over-the-counter" (OTC) drugs, because of the potential of mistreatment of illnesses. Thus, there is great interest in the volume and pattern of prescriptions by physicians, from pharmaceutical companies and policy makers in health care alike. In order to gain insights into the tendencies of drug prescription among a heterogeneous population of physicians, we analyze U.S. data on prescriptions for 33 different pharmaceutical drugs written by a sample of 500 physicians during a period of one year. The drugs fall into six classes: convulsion, Parkinson, psychotherapeutic, anti-depressants, analgesics, and arthritic drugs.

We apply our Tobit factor model to identify latent dimensions underlying physician prescription behavior. Prescription behavior is assumed to derive from physicians latent tendencies to prescribe drugs, which may be affected by their specialization, the type and numbers of patients they treat as well as marketing activities of pharmaceutical companies. We take the prescribed volumes of drugs as indicators of those latent tendencies, where we do not have a-priori hypotheses on those prescription tendencies. Once the parameters for our multivariate Tobit factor model are obtained, we apply them to compute factor scores for a hold-out sample of 4,361 physicians. We estimated the model for $P = 1$ through $P = 4$ latent factors, obtaining the statistics in Table 1. The BIC criterion is minimal for the 3-factor solution.

Parameter estimates for the $P = 3$ solution are shown in Table 2 (pp. 63-64; the factor weights are standardized and rotated to achieve better interpretability using the Varimax method Kaiser, 1958). Relatively high factor weights are underlined. This table also shows two measures of fit at the item (drug) level, the correlation between the observed and fitted volume

Table 1
Fit Statistics for the Tobit Factor Model for the Drug Prescription Data

Number of Factors	No. of Parameters	Log-Likelihood	BIC
1	132	-41,486.1	83,792.6
2	164	-40,481.0	81,987.5
3	195	-40,353.7	81,937.8
4	227	-40,343.7	82,123.1

Table 2
Parameter Estimates for the $P = 3$ Tobit Factor Model, Drug Prescription Data[a]

DRUG	Therapy	σ_j	μ_j	η_j	λ_{j1}	λ_{j2}	λ_{j3}	γ_{i1}	γ_{i2}	γ_{i3}	R	%C
ADDERALL	convulsion	93.6	12.0	-87.9	-0.042	-0.030	0.992	-0.042	-0.030	0.999	0.53	86.4
AMBIEN	Psychoter	41.2	53.5	99.2	0.074	0.206	1.144	0.064	0.177	0.982	0.65	94.6
ARICEPT	alzheimer	14.8	13.9	2.0	0.505	0.202	0.526	0.668	0.267	0.695	0.51	42.8
CATAFLAM	arthritic	13.1	-0.2	-8.5	0.168	0.791	0.039	0.208	0.977	0.048	0.19	30.6
DAYPRO	arthritic	33.0	8.2	13.1	0.224	0.847	-0.073	0.255	0.963	-0.083	0.50	64.0
DURACT	analgesic	14.3	-0.5	-9.4	0.168	0.956	-0.087	0.173	0.981	-0.089	0.47	66.4
DURAGESIC	analgesic	18.4	8.4	-7.6	0.133	0.317	-0.104	0.370	0.883	-0.288	0.16	65.2
EFFEXOR	antidepress	16.4	18.0	13.1	0.146	-0.058	1.702	0.086	-0.034	0.996	0.56	59.0
EFFEXOR XR	antidepress	32.7	17.3	-7.7	0.025	0.032	1.231	0.020	0.026	0.999	0.33	29.0
IMITREX (INJ)	analgesic	4.7	1.2	-1.2	4.255	1.902	0.184	0.912	0.408	0.039	0.97	90.6
IMITREX (TAB)	analgesic	24.2	14.6	23.0	1.153	0.656	0.139	0.865	0.491	0.104	0.64	65.0
IMITREX NASL	analgesic	8.8	2.6	-2.7	0.904	0.530	0.148	0.854	0.501	0.140	0.57	73.6
LODINE	arthritic	7.8	3.0	-5.5	0.165	0.757	-0.078	0.212	0.972	-0.100	0.36	70.2
LODINE XL	arthritic	24.3	0.8	-11.9	0.084	0.800	0.232	0.101	0.956	0.277	0.42	65.6
LUVOX	antidepress	19.3	8.6	-3.9	0.055	-0.121	1.861	0.029	-0.065	0.997	0.63	73.8
NAPRELAN	arthritic	13.1	1.0	-8.0	0.243	0.874	0.066	0.267	0.961	0.073	0.33	68.2
NEURONTIN	convulsion	77.5	31.4	17.8	0.221	0.156	0.486	0.398	0.281	0.873	0.17	49.8
ORUVAIL	arthritic	9.4	2.0	-7.2	0.190	0.848	-0.073	0.218	0.972	-0.083	0.43	71.0
PAXIL	antidepress	53.8	95.9	139.9	0.318	-0.159	2.045	0.153	-0.076	0.985	0.80	88.8
PROSOM	Psychoter	6.3	5.7	-6.3	0.032	0.203	0.213	0.107	0.687	0.719	0.15	85.2
RELAFEN	arthritic	37.1	15.5	24.4	0.280	0.823	-0.090	0.320	0.942	-0.103	0.52	70.0

Table 2 (cont.)

DRUG	Therapy	σ_j	μ_j	η_j	λ_{j1}	λ_{j2}	λ_{j3}	γ_{i1}	γ_{i2}	γ_{i3}	R	%C
REMERON	antidepress	56.6	21.0	-15.7	0.072	-0.141	1.346	0.053	-0.104	0.993	0.71	74.0
RISPERDAL	Psychoter	34.6	28.5	22.4	0.157	-0.261	2.219	0.070	-0.117	0.991	0.84	48.4
SERZONE	antidepress	23.0	25.7	23.9	0.298	-0.140	2.215	0.133	-0.063	0.989	0.80	54.6
SINEMET CR	parkinson	19.8	10.5	-10.3	0.503	0.321	0.134	0.823	0.525	0.219	0.48	31.0
STADOL NS	analgesic	9.4	7.1	-4.7	0.431	0.440	0.392	0.590	0.603	0.537	0.28	73.8
TEGRETOL XR	convulsion	40.4	17.3	-39.5	0.507	0.053	0.572	0.662	0.069	0.746	0.49	87.8
TORADOL ORL	analgesic	2.1	2.0	-2.9	0.068	0.502	-0.303	0.115	0.850	-0.514	0.17	87.2
ULTRAM	analgesic	29.9	21.8	32.0	0.236	0.975	0.065	0.235	0.970	0.064	0.64	77.4
VICOPROFEN	analgesic	14.7	5.1	-12.0	0.067	0.523	0.144	0.123	0.957	0.264	0.24	77.4
VOLTAREN-XR	arthritic	17.9	0.4	-9.5	0.150	0.879	-0.109	0.167	0.979	-0.121	0.43	63.4
ZOLOFT	antidepress	77.5	110.0	205.2	0.246	-0.105	1.829	0.133	-0.057	0.989	0.73	91.6
ZYPREXA	Psychoter	51.4	15.6	-8.2	0.093	-0.356	1.910	0.048	-0.183	0.982	0.72	25.8

[a] Factor weights are standardized. *R* = correlation between actual and fitted prescription volume among prescribing physicians. %C = percentage of correct predictions of observed or censored data

of prescriptions among prescribing physicians (R), and the percentage of correct predictions of whether each data point is observed or censored at zero (%C). Based on these two measures, one may conclude that the model fits reasonably well to the discrete portion of the Tobit model, correctly predicting between 26% (for Zyprexa) and 95% (for Ambien) of the censored/non-censored observations. For the continuous portion of the model the correlations between actual and fitted prescription volumes are between 0.15 (for Prosom) and 0.97 (for Imitrex Inj.).

The censoring intercepts η_j indicate that several drugs have a particularly high probability of being prescribed: *Zoloft, Serzone* and *Paxil* (Anti-depressives), *Imitrex Tablets* (Migraine), *Ultram* (Analgesic), *Risperdal* (Psychotherapeutic), and *Relafen* and *Dyapro* (Arthritis). Note that most of the drugs with large censoring intercepts (i.e., high market penetration) also have relatively large intercepts for the continuous portion of our model (μ_j) (the correlation between the two sets of intercepts is 0.89). This indicates that physicians who prescribe these drugs tend to prescribe a large volume compared to other drugs. In contrast, the parameters for some of the low-penetration drugs such as *Aderall* and *Tegretol* indicate that while a small proportion of physicians ever prescribe those drugs (small $\hat{\eta}_j$), those who do so, tend to prescribe them in relatively high volumes (i.e., $\hat{\mu}_j > \hat{\eta}_j$).

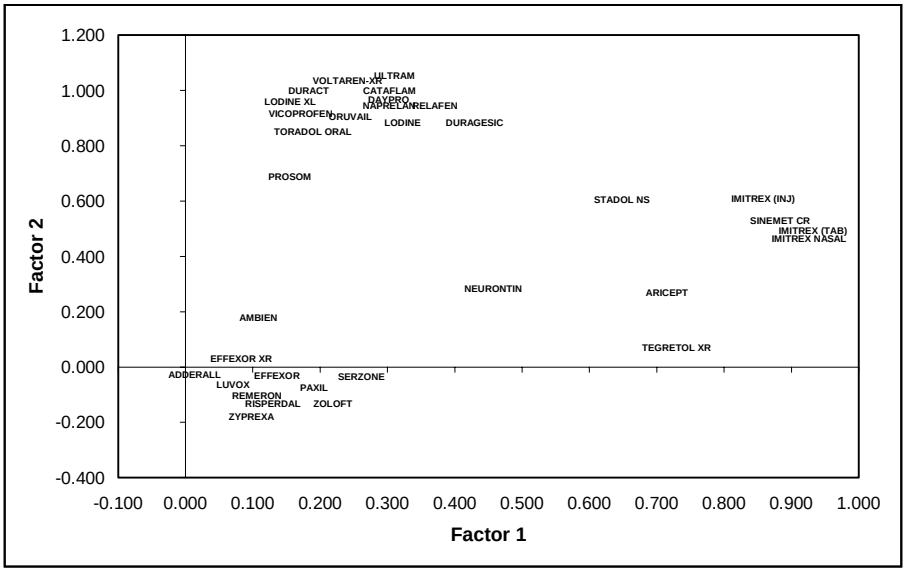
The high correlation of the two sets of intercepts may indicate that a type-1 Tobit model is more appropriate. In order to test for the differences in means among the discrete and continuous portions of our multivariate Tobit model, we compare the Type 2 factor model against a Type 1 model, where $\mu_j = \eta_j$ for all j. The LR test for the 3-factor solution (also chosen for the Type 1 model on the basis of BIC) yields a value of 1639.2 on 33 *df*, which is highly significant, showing that the Type 1 model provides a worse representation of the data and that both types of means are required. This shows that the proportion of zeros in the data is independent of the mean of the distribution of the observed variables, but that the positive and zero data each obey specific data generation processes. Whereas the means capture the aggregate prescription behavior across the population, the factor structure captures heterogeneity among physicians through the underlying distribution of the latent factors.

Table 2 presents the estimated factor loadings, $\hat{\mathbf{\Gamma}}$ cf. Equation 3, with the largest loadings underlined. The loadings are displayed graphically in Figures 1 and 2. These figures show the loadings for the drugs (top panel) and related therapies (bottom panel) on factors 1-2 and 1-3, respectively. Each point on these plots represents a vector terminus for a particular drug (cf. Bartholomew, 1984).

Visual inspection of the pattern of factor weights and factor loadings inTable 2 and the top panel of Figure 1 and discussion with pharmacists, leads

to the interpretation of factor 1 as the propensity to prescribe *neurological drugs*, in particular *Imitrex* for migraines, and drugs for Parkinson's disease and seizures. Note, however, that specific drugs for Altzheimer (*Aricept*) and analgesics (*Stadol* and *Tegretol*) also load relatively high on this dimension, indicating that there is a tendency to prescribe these drugs jointly with these neurological drugs. The second factor is related to *analgesic drugs* such as *Daypro, Naprelan* and *Lodine*. Therefore, physicians with a high score on this dimension are more likely to prescribe and to be heavy prescribers of analgesics and drugs against arthritis. However, note that several of the drugs for migraine (*Imitrex*), psychotherapy (*Prosom*) and Parkinson (*Sinemet*) also have somewhat higher loadings, indicating a propensity to prescribe those drugs in conjunction with the analgesics. Based on Figure 2 and Table 2, Factor 3 can be interpreted as the propensity to prescribe drugs against depression, and psychosis. Physicians with a high score on this dimension are more likely to prescribe this type of drugs, and more likely to prescribe a higher volume of these drugs than other physicians. Note that again a Parkinson drug (*Aricept*) loads high on this dimension, while the general purpose analgesics, *Stadol* and in particular *Tegretol* also load relatively high. Note that another analgesic, *Toradol*, tends not to be prescribed with drugs against psychosis.

In order to ascertain the validity of our results, we use the parameter estimates from Table 2 to compute the factor scores for the hold-out sample of 4,361 physicians and verify whether their scores are consistent with their area of specialization. Figure 3 shows the average score of all physicians by area of specialization. On its top panel, one can see that *Neurologists* have the highest scores on Factor 1, which was identified, as the propensity to prescribe neurological drugs, a quite intuitive finding. Similarly, *Orthopedists* have the highest average scores on Factor 2, and therefore have the highest propensity to prescribe anti-arthritic drugs and analgesics. Note that physicians in *Family*, *Internal* and *Preventive* medicine also have high scores for Factor 2, while their scores on Factor 1 are higher than for the Orthopedists. These less specialized physicians tend to prescribe analgesics among a wider range of drugs. Some drugs load high on Factors 1 and 3 also load on Factor 2 may because they are often first prescribed to patients by neurological and psychiatric specialists, but later prescriptions are taken over by family doctors (this is known to occur for e.g. *Imitrex*, *Sinemet*, *Stadol*, *Prosom*). The bottom panel of Figure 3 shows that Psychiatrists have the highest scores on Factor 3, the propensity to prescribe psychiatric drugs, again a quite intuitive finding that lends the results face validity. Figure 3 reveals a number of meaningful clusters of specialties based on prescriptions. For example, the related specializations *Orthopedy*,



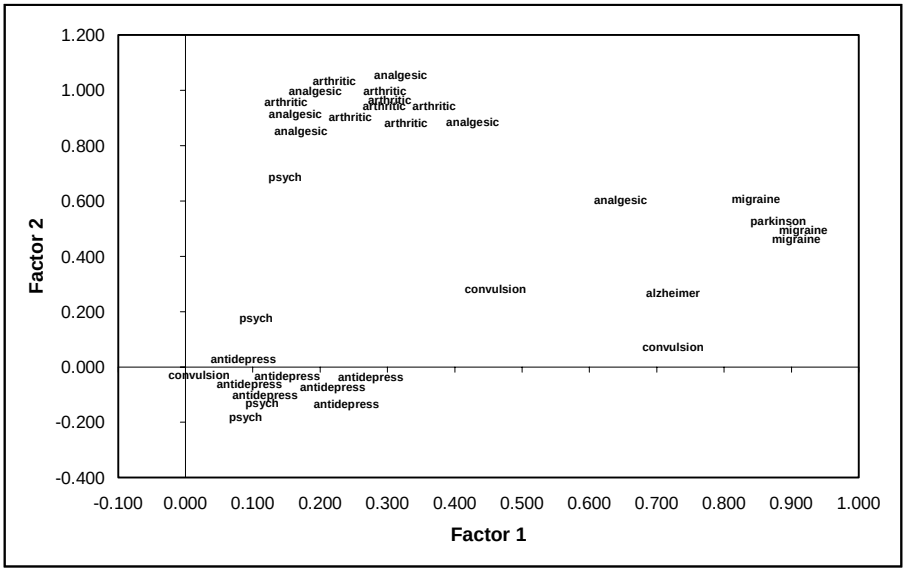
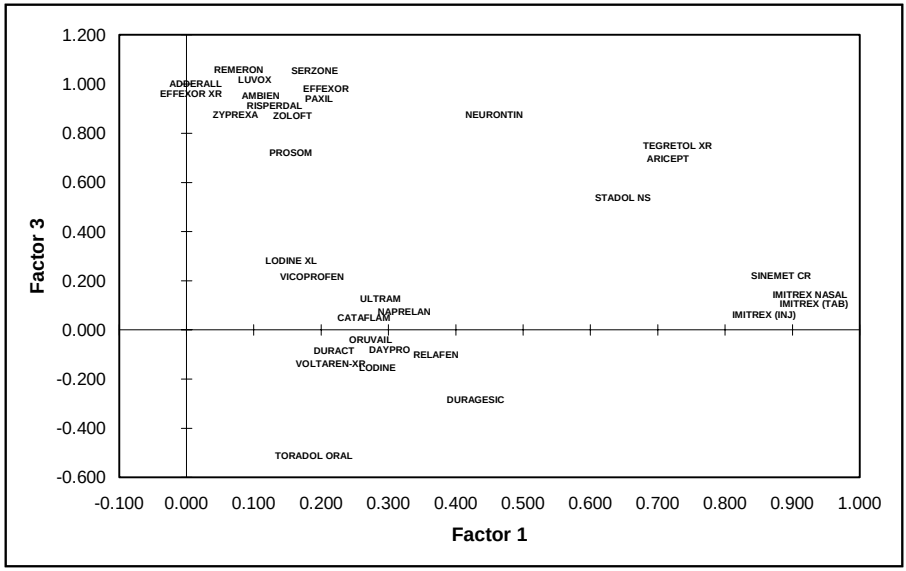


Figure 1
Drug Prescription Factors 1 and 2



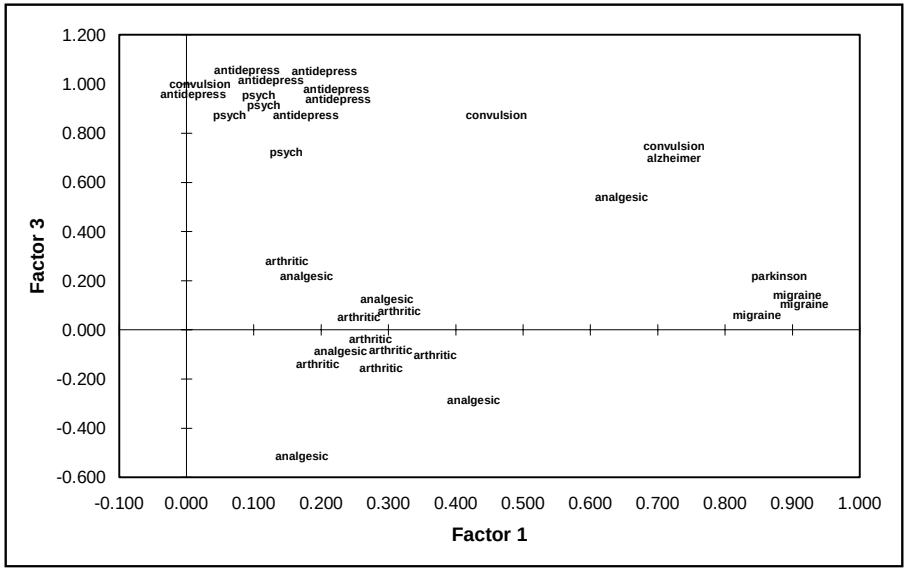
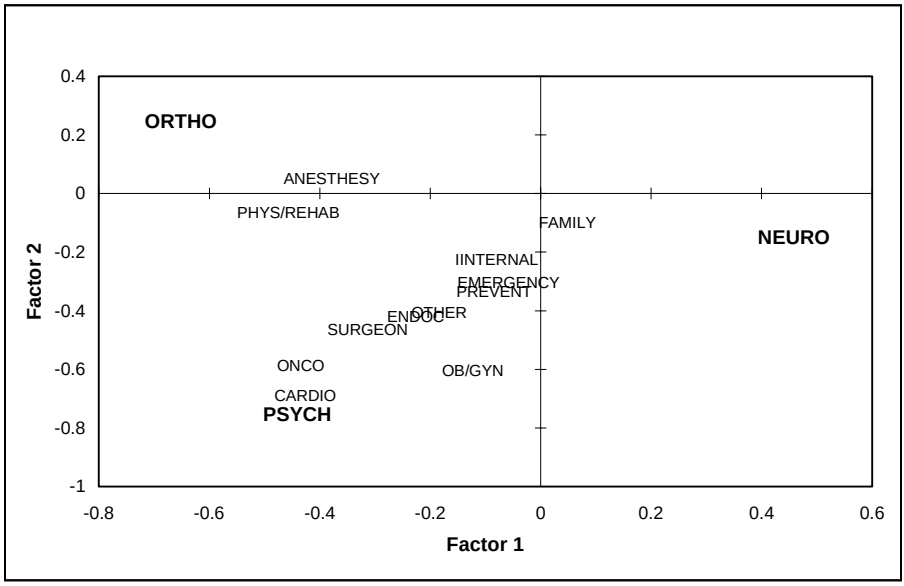


Figure 2
Drug Prescription Factors 1 and 3



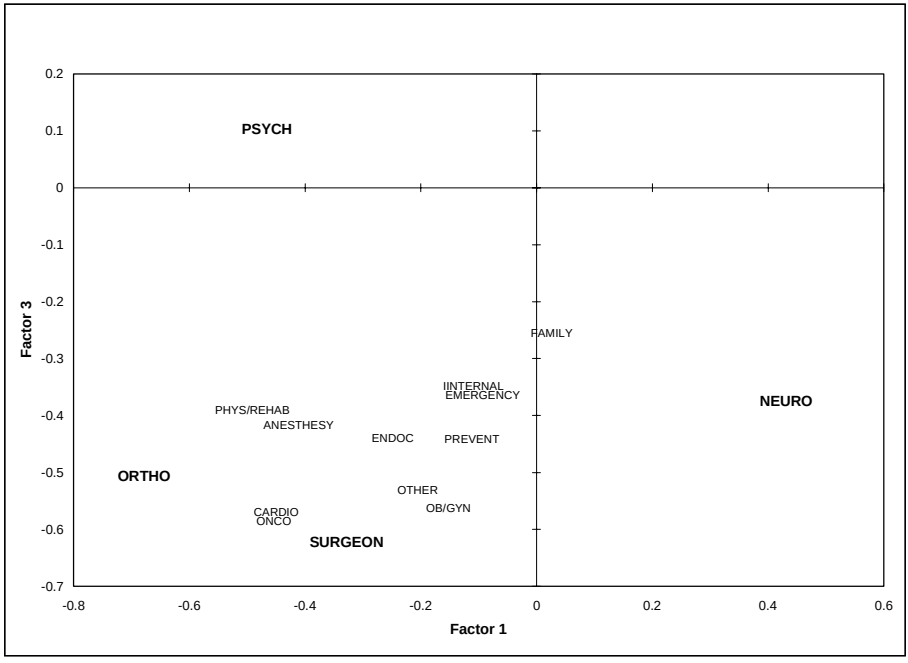


Figure 3
Drug Prescription Average Factor Scores for Physicians by Specialty

Anesthesy, and *Physiotherapy* cluster together on the three factors, and there is a general medicine cluster consisting of family, internal, preventive end emergency medicine.

Illustration to the Analysis of Cross-Category Buying

The number of brands in the global marketplace has rapidly expanded in the eighties and nineties. In the face of competition, manufacturers push the boundaries of product use, create new usage situations for existing brands, extend product lines to transfer brand equity beyond the original category and pursue bundling and cross-category selling. For retailers, category management has emerged as an effective marketing strategy. In category management, the firm markets brands in a way to exploit the pattern in which consumers assemble baskets of products, based on an understanding of their behavior in question. For purposes of manufacturer and retailer marketing strategies alike, knowledge of preferences across and within categories is essential. Knowledge of multiple-category preference patterns allows the retailer to predict the likely composition of market baskets and allows manufacturers and retailers to assess the profitability of product bundling strategies. Therefore, the academic marketing literature has recently seen an upsurge in the analysis of cross-category buying behavior of consumers (e.g. Russell & Kamakura, 1997, Ainsly & Rossi, 1998, Seetharaman, Ainsly & Chintagunta, 1999). In line with that stream of research, we apply the factor Tobit model to analyze cross-category purchasing data and reveal consumers propensities to buy products from different categories to address the central role of cross-category preferences in the design and implementation of these marketing strategies.

The data analyzed here are taken from a panel of 626 Canadian households in one market area. The data consist of the total volume (in equivalent units) purchased of brands in four paper goods categories (Toilet Paper, TP, 10 brands; Paper Towel, PT, 10 brands; Facial Tissue, FT, 9 brands; Table Napkins, TN, 10 brands) for a one year period. The data are collected by scanning equipment at checkout counters of retailers, among the panel of consumers that are identified through special cards. The data set was analyzed previously by Russell and Kamakura (1997) using a latent-class model. The analysis is based upon the purchase volumes of twelve brands with a (volume based) market share greater than 0.5% in at least one product category, and the remaining brands combined into an "Other" brand group. Seven of the brands compete in all product categories. For reasons of confidentiality,

we cannot disclose the brand names in these categories, but indicate them with letters. We use the factor models to identify the latent propensities to buy products across those four categories and to address the question of whether brand-specific (complementary) or category specific (competitive) factors can be identified. This facilitates the development of cross-category marketing strategies, including bundling, cross category promotional programs, cross category loyalty programs and cross-category positioning.

We apply the factor models for $P = 1$ to $P = 5$ factors. Table 3 shows the statistics for the models. BIC is minimal for $P = 3$. The factor weights and loadings for the $P = 3$ model are reported in Table 4 (the factor weights are standardized and rotated using Varimax; high weights and loadings are underlined). This table also shows the correlation between the observed and fitted volume of purchases among buyers (R) and the percentage of correct predictions of whether each product is bought (%C). Based on these two measures, one concludes that the model fits reasonably well to the discrete portion of the Tobit model, correctly predicting between 34% (for PT-SCT) and 97% (for TN-RTA) of the censored/non-censored observations. For the continuous portion of the model the correlations between actual and fitted prescription volumes are between 0.18 (for TP-FAC) and 0.81 (for PT-GRN).

The censoring intercepts η_j indicate that some brands have a high probability of being purchased, in particular KLN in the facial tissue category and to a lesser extent FAC in the toilet paper category and RTB is the table napkin category. Note that the intercepts reflect differences in purchase incidence of the categories as a whole, where the table napkin category has a low purchase incidence and the toilet paper and paper towel categories have higher incidence. As indicated by high values of μ_j, products that are bought in large volumes are KLN in the facial tissue category and RTB in the table napkin category. Here, the correlation between the censoring

Table 3
Fit Statistics for the Tobit factor Model Paper Goods Data

Number of Factors	No. of Parameters	Log-Likelihood	BIC
1	156	-24585.9	503836
2	195	-24585.9	50383.6
3	234	-24446.8	50347.9
4	273	-24338.3	50373.1

intercepts (indicating high incidence) and the intercepts of the continuous part of the model (μ_j) (indicating high volumes bought) is lower (0.47) than in the drug application. This indicates that the Type 1 Tobit factor model would not provide a good representation of the category data and shows that purchase incidence is independent of the mean of the distribution of the purchase quantities: the positive and zero data each obey specific data generation processes. Whereas the means capture the aggregate purchase behavior, the factor structure captures heterogeneity among consumers through the distribution of the latent factors.

The estimates in Table 4 display an interesting pattern of strong cross-category purchase tendencies for a few dominant brands, with a mixed pattern of within and between category competition. We interpret the patterns of factor weights and factor loadings in Table 4 (we have a slight preference for the factor weights since the solution is more simple to interpret, while the varimax rotation tends to produce a solution where many brands load high on the first factor). The factor weights and loadings show that the first factor is clearly a brand-specific factor capturing the unique purchase predisposition of consumers towards the (national) "green" brand GRN. Next to GRN, there is a cluster of brands with high loadings on this factor (FAC, SCT, and WHI). For example, in the paper towels (PT) category, high purchase rates of GRN tend to co-occur with the category specific national brands MAJ and WHI. This indicates category specific competition between those two brands. The second factor clearly represents a store-brand dimension, with strong factor weights on the two retail brands (RTA and RTB), and low weights on all other brands for all categories.

The third factor captures the position of the MFM brand across categories, since we find highly negative weights for MFM in all four categories. This particular brand appears to be uniquely positioned in all categories, showing a strong pattern of joint purchases across categories, with little direct competition with other brands. The single exception is the paper towels category, where competition comes from towels manufactured by HID (as well as SCT and FAC), which show strongly negative weights on this factor. This third factor also shows higher positive loadings for the "green" brand GRN, indicating an almost diametrically opposite positioning for the MFM and GRN brands; consumers who buy one brand is highly unlikely to buy the other.

Figure 4 plots the three factors to display the competitive structure graphically. Once one considers the three-dimensional directions of the vectors representing each category-brand combination, one can see strong cross-category competitive positions for the MFM brand, and for the "green" brand GRN. These two national brands display very strong and

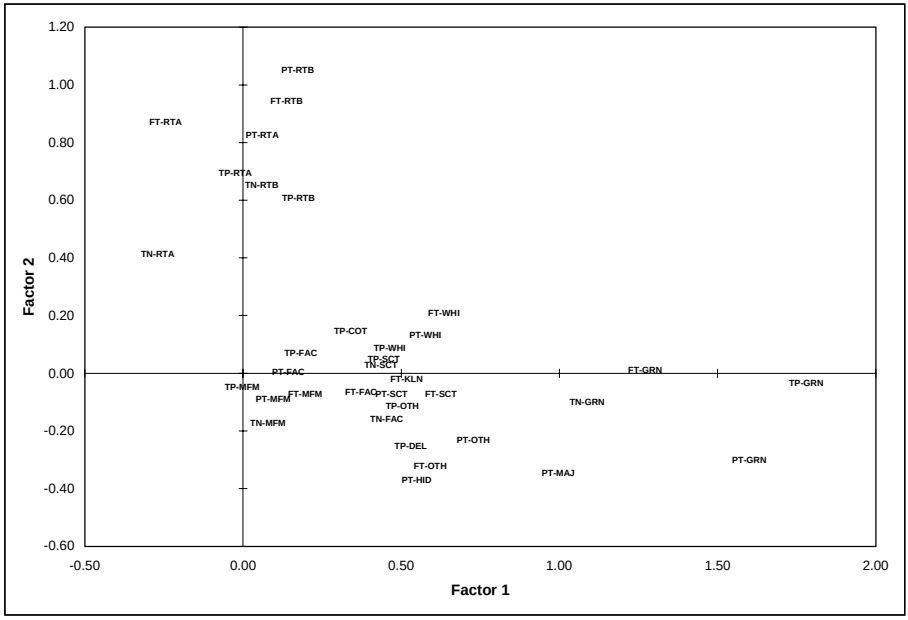
Table 4
Parameter Estimates for the $P = 3$ Tobit Factor Model, Paper Goods Data[a]

Cat.-Brand	σ_j	μ_j	η_j	λ_{j1}	λ_{j2}	λ_{j3}	γ_{i1}	γ_{i2}	γ_{i3}	R	%C
TP-MFM	2.57	0.94	-3.83	-0.09	-0.05	-0.75	-0.11	-0.06	-0.99	0.45	0.88
TP-RTA	4.50	0.96	-5.51	-0.07	0.69	0.07	-0.10	0.99	0.10	0.55	0.84
TP-RTB	1.45	0.32	-2.90	0.09	0.61	0.44	0.13	0.80	0.58	0.49	0.94
TP-FAC	4.63	5.66	5.28	0.10	0.07	-0.13	0.59	0.39	-0.71	0.18	0.54
TP-GRN	2.00	-1.50	-4.32	1.70	-0.03	0.90	0.88	-0.02	0.47	0.78	0.85
TP-SCT	0.81	1.40	-0.68	0.37	0.05	0.03	0.99	0.13	0.09	0.27	0.79
TP-WHI	2.00	2.49	-0.36	0.39	0.09	-0.16	0.91	0.20	-0.37	0.38	0.66
TP-COT	3.32	4.03	2.35	0.29	0.09	-0.25	0.74	0.23	-0.63	0.42	0.76
TP-DEL	2.72	3.20	1.18	0.45	-0.25	-0.33	0.73	-0.41	-0.54	0.46	0.72
TP-OTH	2.62	3.03	-0.06	0.42	-0.12	-0.17	0.90	-0.25	-0.36	0.50	0.59
PT-MFM	1.58	0.40	-2.20	0.01	-0.09	-1.33	0.01	-0.07	-1.00	0.82	0.86
PT-RTA	5.82	-2.36	-12.53	-0.02	0.82	-0.09	-0.02	0.99	-0.11	0.41	0.94
PT-RTB	2.16	0.04	-3.05	0.09	1.05	0.04	0.09	1.00	0.04	0.74	0.82
PT-FAC	3.87	3.52	-0.81	0.06	0.00	-0.36	0.17	0.01	-0.98	0.31	0.58
PT-GRN	2.39	-1.20	-4.24	1.52	-0.30	0.61	0.91	-0.18	0.37	0.81	0.89
PT-SCT	4.83	4.30	2.57	0.39	-0.07	-0.51	0.60	-0.12	-0.79	0.51	0.34
PT-WHI	4.27	4.35	1.89	0.50	0.13	-0.32	0.82	0.22	-0.53	0.52	0.62
PT-HID	4.22	3.20	0.88	0.47	-0.37	-0.90	0.44	-0.34	-0.83	0.71	0.77
PT-MAJ	1.46	1.32	-1.62	0.92	-0.35	-0.13	0.93	-0.35	-0.13	0.63	0.82
PT-OTH	2.73	2.46	-0.32	0.65	-0.23	-0.44	0.79	-0.29	-0.54	0.57	0.65
FT-MFM	2.61	3.13	-2.93	0.11	-0.07	-0.71	0.16	-0.10	-0.98	0.56	0.83

Table 4 (cont.)

Cat.-Brand	σ_j	μ_j	η_j	λ_{j1}	λ_{j2}	λ_{j3}	γ_{i1}	γ_{i2}	γ_{i3}	R	$\%C$
FT-RTA	1.05	0.20	-2.38	-0.32	0.87	-0.04	-0.35	0.94	-0.04	0.40	0.93
FT-RTB	3.35	-0.30	-5.24	0.06	0.94	0.38	0.06	0.93	0.38	0.70	0.88
FT-FAC	3.40	3.64	1.08	0.29	-0.07	-0.19	0.83	-0.18	-0.53	0.28	0.38
FT-GRN	1.11	-0.37	-2.92	1.19	0.01	0.91	0.79	0.01	0.61	0.56	0.91
FT-SCT	3.19	3.37	-0.89	0.55	-0.07	-0.31	0.86	-0.11	-0.49	0.48	0.60
FT-WHI	2.99	2.60	-1.11	0.56	0.21	-0.08	0.93	0.35	-0.14	0.45	0.68
FT-KLN	9.32	9.87	11.92	0.44	-0.02	-0.44	0.71	-0.04	-0.71	0.52	0.87
FT-OTH	3.60	1.68	-4.21	0.51	-0.32	-0.17	0.81	-0.52	-0.27	0.46	0.84
TN-MFM	8.25	1.98	-12.34	0.00	-0.18	-0.77	-0.01	-0.22	-0.97	0.62	0.89
TN-RTA	6.78	4.33	-15.59	-0.35	0.41	0.01	-0.65	0.76	0.03	0.58	0.97
TN-RTB	7.09	5.90	-11.55	-0.02	0.65	0.16	-0.03	0.97	0.24	0.43	0.90
TN-FAC	3.00	3.13	-2.98	0.38	-0.13	-0.21	0.83	-0.28	-0.47	0.18	0.83
TN-GRN	3.42	0.91	-8.11	1.00	-0.10	0.26	0.96	-0.10	0.25	0.49	0.95
TN-SCT	2.80	3.92	-2.55	0.36	0.05	-0.19	0.87	0.13	-0.47	0.37	0.81
TN-WHI	3.06	2.96	-3.47	0.29	0.22	-0.18	0.71	0.54	-0.45	0.40	0.85
TN-HID	6.18	3.43	-5.99	0.36	-0.32	-0.41	0.56	-0.50	-0.65	0.30	0.82
TN-KLN	1.37	1.70	-1.14	0.03	-0.14	-0.26	0.10	-0.48	-0.87	0.37	0.80
TN-OTH	2.25	2.64	-1.85	0.48	-0.13	-0.03	0.96	-0.27	-0.06	0.47	0.79

[a] Factor weights are standardized. R = correlation between actual and fitted prescription volume among prescribing physicians. $\%C$ = percentage of correct predictions of observed or censored data.



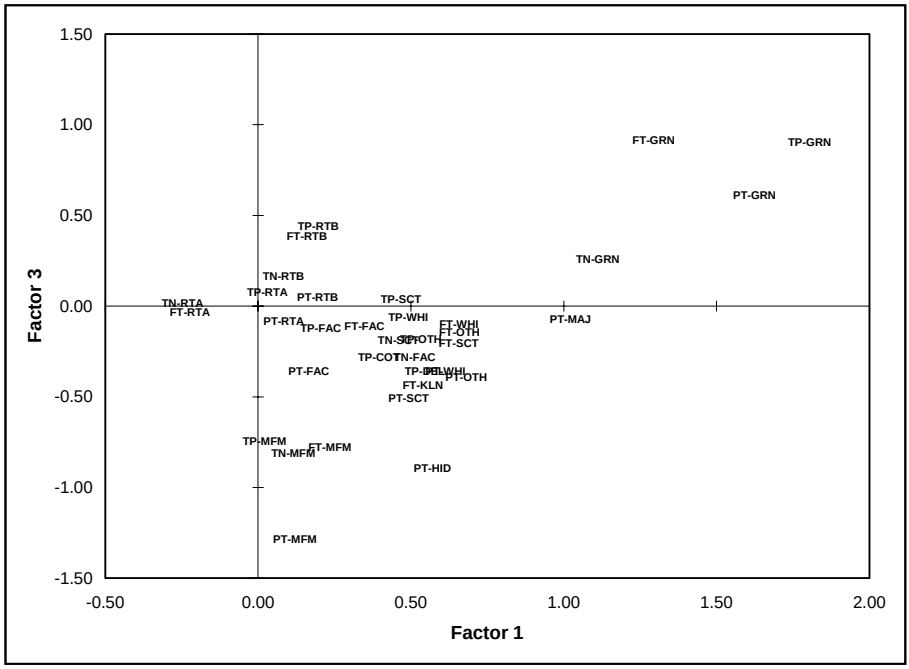


Figure 4
Factor Plots for the Paper Goods Application

unique brand franchises across all four categories. Consumers who buy the brand in one category are highly likely to also buy the brand in all other category. Most importantly, they are also unlikely to buy any other brand in these categories. The only exceptions are the paper towels by HID, which tend to be bought by buyers of the MFM brand, and paper towels by MAJ, which are bought by buyers of the GRN brand.

A similar pattern of cross-category preferences is found for the retail brands. Consumers who buy a retail brand in one product category are highly likely to also buy that same retail brand in other product categories. In contrast to the MFM and GRN brands, which occupied unique competitive positions, our Tobit factor map indicates that the two retail brands RTA and RTB clearly compete directly against each other.

Conclusions

Making distribution assumptions on observed variables has essentially moved factor analysis from the realm of descriptive methods to the area of statistical modeling. This practice dates back to Lawley (1940), who first developed maximum likelihood estimation methods for factor analysis. Later, factor analysis was developed for binary variables (Bartholomew, 1980), and for truncated variables (Muthén, 1989). Along the lines of these studies, we propose a factor model that is based on mixed discrete-continuous data. Rather than specifying different distributions for different variables, we have employed the Tobit framework that has gained great popularity for the analysis of consumption behavior. In doing that, we build directly on the pioneering work of Muthén (1989). Our model allows for the exploration of high dimensional data by displaying the latent structure underlying it, being based on assumptions on the type of censoring mechanism. Much of our enterprise has been made possible by the advent of simulation methods. The Tobit modeling framework has the advantages over previously proposed models for mixed data that it is feasible for large numbers of variables (cf. Cox & Wermuth, 1992) and directly restricts the outcomes of the continuous data to be positive (cf. Sammel, Ryan & Legler, 1997; Bartholomew & Knott, 1999, p. 173).

Our approach, enables full (S)ML estimation of exploratory factor models, but also of confirmatory factor models when appropriate restrictions have been identified. Muthén (1989) proposes a confirmatory tobit-type factor analysis, which was subsequently applied by Waller and Muthén (1992) to behavioral genetics. They propose a three-step procedure. First, univariate Tobit models are employed to estimate the mean and variance of the latent censored variable, using ML. Then, the bivariate correlation is

estimated from the bivariate distributions by maximum likelihood, fixing the mean and variance parameters at the estimated values in the first step. Thirdly, generalized least squares is applied to these estimated correlations to estimate a confirmatory factor model. A consistent estimator of the asymptotic covariance matrix of the correlations estimated from the two previous steps is used as a weight matrix in the GLS estimation procedure. Muthéns approach overcomes the problem of high dimensional integration by reducing the P-variate normal integral to $[J(J + 1)/2]$ two-dimensional integrals. Thus, the procedure requires running $[J(J + 1)/2]$ Tobit models and a GLS confirmatory factor model. Although being a multi-stage procedure, Muthén's method provides consistent –but not efficient- estimates of the factor model parameters. Our procedure extends that of Muthén in several ways, building upon his developments. First, Muthén's approach deals with confirmatory models while we deal with both confirmatory and exploratory factor analysis, focussing on the latter. However, Muthén's method may be relatively easily modified to deal exploratory models as well. Secondly, Muthén presumably accommodates a Type-1 Tobit model, while we deal with a type-2 factor model. The latter offers the advantage of providing a more flexible model of the censored and non-censored data. In the empirical applications on drug prescription and multi-category purchasing, we showed that the fit of the type-2 model is significantly better than that of a type-1 model. Further, rather than the three stage estimation approach of Muthén, we estimate all parameters simultaneously with Simulated Likelihood. This gives us consistent and asymptotically efficient estimates of the model parameters. The application of SML to Tobit factor models has not been previously described. A current limitation of the proposed SML procedure is its computational cost, which is a curse shared by most models estimated using simulation. We expect this limitation to become less and less of a problem in the future with the increasing speed of computers. Note that Muthéns method would require the estimation of 561 and 780 Tobit models for our two applications, respectively, with associated programming and data handling costs.

Due to our simultaneous estimation procedure, our approach easily lends itself to extensions in various directions, that we however consider beyond the purpose of the present paper, including different distributions for the observed variable and the inclusion of predictor variables. Given the assumption of a normal distribution of the latent factors, our factor model may be seen as a way to include heterogeneity in Tobit models, along the lines of Gouriéroux and Montfort (1996). Instead of including a single random term with fixed variance to capture misspecification, our Tobit factor model includes P random terms, which are weighted differently for the observed

variables, creating covariance among them. Gouriéroux and Montfort (1996) include predictors, which we have not done (except for the intercepts) since none were available in our applications, but the model can be extended to include such predictor effects.

References

Ainslie, A. & Rossi, P. (1998). Similarities in choice behavior across product categories. *Marketing Science, 17*, 91-106.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52* (*1*), 317-332.

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica, 41*, 723-732.

Amemiya, T. (1985). *Advanced econometrics*. Cambridge: Harvard University Press.

Anderson, T. W. & Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability, 5*, 111-150.

Arminger, G. & Küsters, U. (1988). Latent trait models with indicators of mixed measurement level. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models.* New York: Plenum.

Bartholomew, D. J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society, B, 42*, 293-321.

Bartholomew, D. J. (1987). *Latent variable models and factor analysis.* New York: Oxford University Press.

Bartholomew, D. J. & Knott, M. (1999). *Latent variable models and factor analysis.* London: Edward Arnold.

Bekker, P., Merckens, A., & Wansbeek T. J. (1994). *Identification, equivalent models and computer algebra*. New York: Academic Press Inc.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika 52*, 345-370.

Burkett, J. P. (1998). Bureaucratic behavior modeled by reduced-rank regression: The case of expenditures from the Soviet state budget. *Journal of Economic Behavior and Organization, 34*.

Burnham, K. P. & Anderson, D. R. (1998). *Model selection and inference*. New York: Springer Verlag.

Chib, S. (1992). Bayes inference in the Tobit Censoring Regression Model. *Journal of Econometrics, 51*, 79-99.

Cox, D. R. & Wermuth, N. (1992). Response models for mixed binary and quantitative variables. *Biometrika, 79*, 441-461.

DeSarbo W. S. & Choi J. (1999). A latent structure double hurdle regression model for exploring heterogeneity in consumer search patterns. *Journal of Econometrics, 89*, 423-456.

Donovan, J. E. (1993). Young adult drinking-driving: Behavioral and psychosocial correlates. *Journal of Studies on Alcohol*, 54 600-614.

Fitzmaurice, G. M. & Laird, N. M. (1995). Regression models for bivariate discrete and continuous outcome with clustering. *Journal of the American Statistical Association, 90*, 845-852.

Gill, R. D. (1977). Consistency of maximum likelihood estimator of the factor analysis model when the observations are not multivariate normal. In J. R. Bara, F. Brodeau, G. Romier & B. van Cutsem (Eds.), *Recent developments in statistics* (pp. 437-440). Amsterdam: North Holland.

Gourieroux, C. & Montfort, A. (1996). *Simulation based econometric methods*. Oxford: Oxford University Press.

Greene, W. H. (1981). On the asymptotic bias of the ordinary least-squares estimator of the Tobit model. *Econometrica, 49*, 505-513.

Harris, K. M. & Keane, M. P. (1999). A model of health plan choice: inferring preferences and perceptions from a combination of revealed preference and attitudinal data. *Journal of Econometrics, 89*, 131-158.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement, 5*, 475-492.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995). *Continuous univariate distributions, Volume 2*. New York: Wiley.

Jones, A. & Posnett, J. (1991). Charitable donations by U.S. households: Evidence from family expenditure survey. *Applied Economics, 23*, 343-351.

Kaiser, H. F. (1958). The Varimax criterion for analytical rotation in factor analysis. *Psychometrika, 23*, 187-200.

Keane, M. P. (1993). Simulation estimation for panel data models with limited dependent variables. In G. S. Maddala, C. R. Rao, H. D. Vinod (Eds.), *Handbook of statistics*. Amsterdam: Elsevier.

Krzanowski, W. J. & Mariott, F. H. C. (1995). *Multivariate analysis, Kendall Library of Statistics 2*. London: Arnold.

Lancaster, H. O. (1954). Traces and cumulants of quadratic forms in normal variables. *Journal of the Royal Statistical Society, B, 16*, 247-254.

Lance, C. E., Cornwell, J. M. & Mulaik, S. A. (1988). Limited information parameter estimates for latent of mixed manifest and latent variable models. *Multivariate Behavioral Research, 23*, 171-187.

Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh, 61*, 176-185.

Lee, L. F. (1995). Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. *Econometric Theory*, 437-483.

Lee, L. F. (1997). Simulated maximum likelihood estimation of dynamic discrete choice statistical models: Some Monte Carlo results. *Journal of Econometrics, 82*, 1-35.

Mauran, M. D. (1996). Metaphor taken as math: Indeterminancy in the factor analysis model. *Multivariate Behavioral Research, 31*(*4*), 517-538.

Mingshan, L. (1999). Separating the true effect from gaming in incentive-based contracts in health care. *Journal of Economics & Management Strategy, 8*, 383-423.

Mulaik, S. A. (1972). *The foundations of factor analysis.* New York: McGraw Hill.

Muthén, B. O. (1989). Tobit factor analysis. *British Journal of Mathematical and Statistical Psychology, 42*, 241-250.

Olsen, R. J. (1978). A note on the uniqueness of the maximum likelihood estimator for the Tobit model, *Econometrica, 46*, 1211-1215.

PhRMA (1999). *Industry profile 1998*. PhRMA, USA.

Russell, G. J. & Kamakura, W. A. (1997). Modeling multiple category brand preference with household basket data. *Journal of Retailing, 73*, 439-462.

Rust, R. T., Simester, D., Brodie, R., & Nilikant, V. (1995). Model selection criteria: an investigation of relative accuracy, posterior probabilities and combinations of criteria. *Management Science 41*, 322-333.

Sammel, M. D. & Ryan, L. M. (1996). Latent variable models with fixed effects. *Biometrics, 52*, 220-243.

Sammel, M. D., Ryan, L. M. & Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, B, 59* (*3*), 667-678.

Seetharaman, P. B. Ainslie, A. & Chintagunta, P. K. (1999). Investigating household state dependence across categories. *Journal of Marketing Research, 36*, 488-500.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-464.

Sher, K. J. & Wood, M. D. (1996). Alcohol outcome expectancies and alcohol use: A latent variable cross-lagged panel study. *Journal of Abnormal Psychology, 105*, 561, 575.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica, 26*, 24-36.

Vandenberg, J. & Richardson, R. (1999). The impact of high involvement work processes on organizational effectiveness. *Group and Organization Management, 24* (*3*), 300-340.

Wales, T. J. & Woodland, A. D. (1980). Sample selectivity and the estimation of labor supply functions. *International Economic Review, 21*, 437-468.

Waller, G. N. & Muthén, B. O. (1992). Genetic Tobit factor analysis: Quantitative genetic modeling with censored data. *Behavior Genetics, 22*, 265-292.

Accepted July, 2000.

Appendix

Simulated Maximum Likelihood (Gouriéroux & Montfort, 1996)

The generic problem in which simulation can be applied is to evaluate a log-likelihood equation of the form:

$$l(\boldsymbol{\Theta}|y) = \sum_n \ln \int L(y_n|x;\boldsymbol{\Theta})\phi(x)dx, \tag{A1}$$

where x is a P-dimensional multivariate random variable with a normal density $\phi(x)$, and y_n is a J-dimensional observation vector. The estimator $\hat{\boldsymbol{\Theta}}$ obtained by maximizing Equation A1, which is often done numerically, is consistent, efficient and asymptotically normal for a large class of models. If the dimensionality P of x is larger than three, standard numerical integration cannot be used to evaluate the log-likelihood. The idea of simulation is to draw T random variables z^t from $\phi(\cdot)$ and use

(A2) $$\tilde{l}(\boldsymbol{\Theta}|y) = \sum_n \ln \sum_t \tilde{L}\left(y_n|z^t;\boldsymbol{\Theta}\right)\Big/T .$$

Now, $\tilde{l}(\boldsymbol{\Theta}|y) \to l(\boldsymbol{\Theta}|y)$ as $T \mapsto$ from the strong law of large numbers, so that the simulated likelihood function is a consistent simulator of the likelihood function. The value of Θ that maximizes Equation A2 is the SML estimator. SML provides consistent estimators only if $T \mapsto \infty$ as $N \mapsto \infty$. This can be seen as follows:

$$\lim_{N,T\to\infty} \sum_n \ln \sum_t \tilde{L}\left(y_n|z^t;\boldsymbol{\Theta}\right)\Big/T = \lim_{N\to\infty} \sum_n \ln \int \tilde{L}\left(y_{nj}|z;\boldsymbol{\Theta}\right)dz ,$$

since the mean over t converges to the integral function for $T \mapsto \infty$. Because $\tilde{L}(\cdot)$ is a consistent simulator of $L(\cdot)$, this equals

$$E\left[\sum_n \ln \int L(y_{nj}|z;\boldsymbol{\Theta})dz\right],$$

so that the estimator is consistent and asymptotically equivalent to the ML estimator.

The procedure for maximizing Equation A2 works as follows.

1.Assume $z \sim \phi(z|\mu,\Sigma)$, multivariate normal. Fix a value of T. To draw z from the multivariate normal distribution, first draw NPT values of u, with $u_n \sim \phi(0,I_P)$ independent normal with the same dimensionality as z.

2.Compute the Choleski decomposition $CC' = \Sigma$. Then $z^t_n = \mu + Cu^t_n \sim \phi(\mu,\Sigma)$. Store the NPT values of z computed in this way. These values will remain the same throughout the optimization procedure.

3.Compute the simulated likelihood function in Equation A2 based on the stored values z_n.

4.Maximize Equation A2 numerically over Θ using a Newton type algorithm to find the SML estimator. For that purpose one needs the first order derivatives of the simulated log-likelihood function:

$$\sum_n \frac{\sum_t \partial\tilde{L}\left(y_n|z^t;\boldsymbol{\Theta}\right)\Big/\partial\boldsymbol{\Theta}}{\sum_t \tilde{L}\left(y_n|z^t;\boldsymbol{\Theta}\right)}$$

Details on Estimation of the Proposed Model

In order to speed-up estimation, we replace the standard normal ogive by a logistic approximation, so that the individual log-likelihood contribution is computed as:

$$\text{(A3)} \qquad l_n \approx \ln\left\{\sum_{t=1}^{T}\prod_{1}\phi(y_{nj}|z_n^t;\Theta)\frac{e^{1.7(\eta_j+z_n^t\Lambda')}}{1+e^{1.7(\eta_j+z_n^t\Lambda')}}\prod_{0}\left[\frac{1}{1+e^{1.7(\eta_j+z_n^t\Lambda')}}\right]\right\}$$

This approximation of the cumulative normal by the cumulative logistic distribution function is accurate, since there is a close similarity on shape between the normal and logistic distributions, while the difference, attributed to the longer tails of the logistic, has hardly any effect on the cumulative distribution function (Johnson, Kotz & Balakrishnan, 1995, p. 119). Alternatively, we could have formulated the model in terms of the logistic distribution rather than the normal, but since that seems counter to the current practice of Tobit modeling, we will not do so. The gradients needed for a Newton-Raphson search are:

$$\text{(A4)} \qquad \frac{\partial \ell_n}{\partial \mu_j} = \sum_{t}^{T}(P_{nt}A_{jnt}/\sigma_j)^{I(y_{jn}>0)},$$

$$\text{(A5)} \qquad \frac{\partial \ell_n}{\partial \sigma_j} = \sum_{t}^{T}\left[P_{nt}\left(\frac{A_{jnt}y_{jn}-\sigma_j}{\sigma_j^2}\right)\right]^{I(y_{jn}>0)},$$

$$\text{(A6)} \quad \frac{\partial \ell_n}{\partial \lambda_{jk}} = \sum_{t}^{T}\left[P_{nt}\left(A_{jnt}+1.7Q_{jnt}\right)z_{kt}\right]^{I(y_{jn}>0)} + \sum_{t}^{T}\left[1.7P_{nt}(Q_{jnt}-1)z_{kt}\right]^{I(y_{jn}=0)},$$

and

$$\text{(A7)} \quad \frac{\partial \ell_n}{\partial \eta_j} = \sum_{t}^{T}(1.7P_{nt}Q_{jnt}/\sigma_j)^{I(y_{jn}>0)} + \sum_{t}^{T}\left[1.7P_{nt}(Q_{jnt}-1)/\sigma_j\right]^{I(y_{jn}=0)},$$

where,

(A8) $$A_{jnt} = \frac{y_{jn} - \mu_j}{\sigma_j} - z_n^t \Lambda',$$

(A9) $$Q_{jnt} = \left[1 + e^{1.7\left(\frac{\eta_j}{\sigma_j} + z_n^t \Lambda' \right)} \right]^{-1}$$

and

(A10) $$P_{nt} = \frac{\prod_{1} \phi\left(y_{nj} | z_n^t; \Theta\right)\left(1 - Q_{jnt}\right) \prod_{0} Q_{jnt}}{\sum_{t=1}^{T} \prod_{1} \phi\left(y_{nj} | z_n^t; \Theta\right)\left(1 - Q_{jnt}\right) \prod_{0} Q_{jnt}}$$

Once the parameters of the model are estimated, each subject *n* can be evaluated along the latent dimensions, by solving the non-linear equations below for the factor scores, x_n:

(A11) $$\partial \ell_n / \partial x_n = \sum_{o} \lambda_j \left(\frac{y_{jn} - \mu_j}{\sigma_j} - x_n \lambda'_j + 1 \right) - \sum_{j=1}^{J} \lambda_j \left[\frac{e^{1.7\left(\frac{\eta_j}{\sigma_j} + x_n \lambda'_j \right)}}{1 + e^{1.7\left(\frac{\eta_j}{\sigma_j} + x_n \lambda'_j \right)}} \right] = 0$$

where $\sum_{0}$ indicates a sum over the non-censored data.