MEMORANDUM

No 18/2011

Models of Truncation, Sample Selection, and Limited Dependent Variables: Suggestions for a Common Language

Erik Biørn
Knut R. Wangen

ISSN: 0809-8786

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.sv.uio.no/econ
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 17/11	Steinar Holden , Victoria Sparrman *Do Government Purchases Affect Unemployment?*
No 16/11	Ola Lotherington Vestad *Who pays for occupational pensions?*
No 15/11	Mads Greaker, Michael Hoel *Incentives for environmental R&D*
No 14/11	Halvor Mehlum , Karl Moene , Ragnar Torvik *Mineral Rents and Social Development in Norway*
No 13/11	André K. Anundsen, Eilev S. Jansen *Self-reinforcing effects between housing prices and credit: Evidence from Norway*
No 12/11	Nils Chr. Framstad *Portfolio Separation with α-symmetric and Psuedo-isotropic Distributions*
No 11/11	Jørgen Modalsli *Inequality and growth in the very long run: inferring inequality from data on social groups*
No 10/11	Eilev S. Jansen, Tord S. H. Krogh *Credit conditions indices: Controlling for regime shifts in the Norwegian credit market*
No 09/11	Karin Jacobsen , Kari H. Eika, Leif Helland, Jo Thori Lind, Karine Nyborg *Are nurses more altruistic than real estate brokers?*
No 08/11	Michael Hoel *The supply side of CO_2 with country heterogeneity*

Previous issues of the memo-series are available in a PDF® format at:
http://www.sv.uio.no/econ/forskning/publikasjoner/memorandum

MODELS OF TRUNCATION, SAMPLE SELECTION, AND LIMITED DEPENDENT VARIABLES: SUGGESTIONS FOR A COMMON LANGUAGE

ERIK BIØRN
Department of Economics, University of Oslo,
P.O. Box 1095 Blindern, 0317 Oslo, Norway.
E-mail: erik.biorn@econ.uio.no

KNUT R. WANGEN
Department of Health Management and Health Economics, University of Oslo,
P.O. Box 1089 Blindern, 0317 Oslo, Norway
E-mail: k.r.wangen@medisin.uio.no

ABSTRACT: The aim of this paper is two-fold: (a) to establish a 'map' for describing the wide class of Limited Dependent Variables (LDV) univariate and multivariate models in the econometric literature and (b) to localize typical models in this tradition within the structure, extending typologies of Heckman (1976) and Amemiya (1984). The classification system, or language, proposed, is given at different level of detail. Its scope is (1) that the latent variables can have any parametric distribution, (2) that a set of observation rules which include the observed, censored, missing status, is imposed, (3) that it should be possible to write a model combining (1) and (2) by means of a computer algorithm, also potentially applicable for generating samples and (4) that the models belonging to the system should have names to facilitate communication among researchers. The likelihood functions corresponding to the models' observed endogenous variables are discussed, but the paper is not concerned with describing the application of these functions for inference.

KEYWORDS: Micro-econometrics. Limited dependent variables. Latent variables. Discrete choice. Censoring. Truncation. Missing observations.

JEL CLASSIFICATION: C16, C24, C25, C34, C35, C51

ACKNOWLEDGEMENTS: Comments from seminar participants at Statistics Norway are gratefully acknowledged.

1. Introduction

In the first half of the twentieth century, pioneering contributions to univariate Limited Dependent Variables (LDV) models were made by Fisher (1946/37), Bliss (1937), Hald (1949) and other statisticians. The steadily decreasing price of computer power has allowed application of increasingly more demanding LDV models. Tobin (1958) introduced discrete-continuous LDV models with regressors and censoring to the econometric literature. For (elements of) recent historical reviews related to discrete LDV models in the logistic class, see *e.g.*, Agresti (1990, Appendix B) and Hilbe (2009, Section 1.3). This, together with vastly improved software for computations, spawned a whole branch of econometric modeling. The early contributions typically had a case-by-case approach — discussing LDV models one by one — motivated by one or a few particular applications. A few later studies developed a more extensive typology and showed that seemingly disparate LDV models share a common structure, notably Heckman (1976) and Amemiya (1984).

Heckman's and Amemiya's studies are limited in scope. Heckman (1976, p. 475) states that "To simplify the exposition, I consider a two equation model. Few new points arise in the multivariate case, and the multivariate extension is straightforward", while Amemiya (1984, p. 4), surveying the class of censored regression models often nicknamed 'Tobit models', states "My review of the empirical literature suggests that roughly 95 % of the econometric applications of Tobit models fall into one of ... five types". In subsequent literature, relatively few studies have considered models with more than two equations, which may suggest that extending to higher dimensions is not completely trivial. Although a typology based on previous *empirical* literature can be useful for review purposes, it may, from a theoretical perspective, have at least two disadvantages. First, the empirical literature existing at any point in time is limited by the currently available computing resources, in particular the publicly available software. Second, the number of possible types of LDV models is infinite, and thus, the empirical literature cannot cover all cases.

The LDV literature benefits from a classification system that is disconnected from limitations set by current computational resources. This is a primary motivation for writing the present paper. Several models in the existing econometric literature, with catchwords like censoring, truncation, selectivity, discrete choice, missing observations etc., emerge as variations on a common theme, being essentially special cases of a more general structure that has not, to our knowledge, yet been fully specified. Our intended contribution is to offer a generalized framework, a 'language', for classifying a wide set of uni- and multi-variate LDV models that have censored, truncated, missing observations, or all the three in combination. We introduce a notation which enables us to describe any model within the framework in a compact manner. We provide examples where models with different nicknames in econometric literature – indicating that they are essentially different – have in fact a close 'family likeness' in their structure. Thus, we believe the classification system may ease the communication between experts from fields with different jargons, say, in exchange of estimation software.

Our framework has three basic elements: a specification of a distribution of latent variables, a partition of its support into subsets, and an observation rule for each subset, defining whether each latent variable is observed, censored, or missing. To clarify the link with previous literature, we provide examples where the latent variables are normally distributed, and where the subsets are defined by linear restrictions. Since our focus is on classification and not on estimation *per se*, we take the discussion up to the point of demonstrating how likelihood functions can be constructed, but refrain from discussing typical inference issues, such as concavity of the likelihood or identification of parameters.

The scope of the classification system for models to be proposed and exemplified in this paper can be briefly described as follows: [1] The latent variables can have any parametric distribution. [2] A set of rules which include the observed, censored, missing status is imposed. [3] It should be possible, in principle, to write a model combining the variables and the rules by means of a computer algorithm, also potentially applicable for generating samples. [4] At a certain level of specification, the models should have names facilitating identification and communication among researchers. The likelihood functions of the models' observed endogenous variables (conditional on the exogenous variables when such occur) are discussed, and described to some level of precision, but it is beyond the scope of the paper to describe how these functions can be used for inference.

To draw a precise borderline between being missing, being censored, and being truncated is far from easy. A variable of interest for which observations are unavailable may be considered a particular status of a stochastic (potentially) observable variable: the analyst is unable to obtain a value regardless of the efforts paid. The 'missing data' field touches onto censoring, truncation, and latent variables (mismeasurement being a special case); see Little and Rubin (1987) for a survey of the 'missing data' field in statistics. We will not attempt to draw a precise borderline between missing and latent variables and between missing observations and qualitative variables.[1] However, the distinction between an observation being zero and being missing, which is, unfortunately, sometimes disguised by sloppy practice of statistical agencies and 'data organizers', is of the utmost importance.

The organization of the paper is as follows. In Section 2 we discuss, partly with examples, notation and definitions. Next, in Section 3 we describe the compilation of likelihood functions for univariate and bivariate models in some detail. A discussion of the multidimensional generalization, using still more compact notation, and a

[1] The missing status may be relative and temporary, as costs, efforts, institutional and legal arrangements etc. may be involved. To decide whether a variable 'unable of being observed' may be observable in other 'regimes' is difficult. We will pay attention to a 'globally missing' variable only when it is related to other variables of interest which are observable to some extent, and will classify a variable as missing only if it is *structural* according to some theory. Hence, measurement errors and disturbances are not missing variables by this usage. Neither are model parameters to be classified as missing variables; for the distinction between parameters and variables relative to the 'structure' concept, see Marschak (1953, Sections 2 and 3).

further generalization accounting for covariates follows in Section 4. Section 5 first presents a label system for models, next uses this for reference to models in the literature and for elaborating examples. Section 6 provides concluding remarks.

2. Definitions and notation

2.1. **Latent variables, subsets, and basic observation rules.** Our general framework has three *basic elements*. The *first* is a vector of latent stochastic variables, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)$, defined over the N-dimensional Euclidian space, $\mathbb{R}^N$. The *second* basic element is a partition of $\mathbb{R}^N$ into I subsets, denoted as α_i, so that

$$\bigcup_{i=1}^{I} \alpha_i = \mathbb{R}^N, \quad \alpha_i \cap \alpha_j = \emptyset \; \forall \; i \neq j. \tag{2.1}$$

The subsets are arranged in the tuple $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_I)$. The *third* basic element is a register of observation rules. Over each subset, each latent variable η_n, $n = 1, \ldots, N$, has one among three possible observational statuses: *observable*, *censored* or *missing*, indicated by the letters o, c, and m, respectively. The observation rule for subset i is denoted r_i, and is a 'word' with N letters indicating the observational statuses for all latent variables.[2] The observational rules are collected in the tuple $\boldsymbol{r} = (r_1, \ldots, r_I)$.

Example: Censored univariate normal distribution, $N = 1$, $I = 2$: Suppose the latent variable is univariate, normally distributed, and that the real line is divided in two parts by a threshold value, θ_1. In the interval below the threshold (α_1), the observation status is 'censored'. In the interval above the threshold (α_2), the observation status is 'observable'. Then the model can be described by

$$\begin{aligned} \alpha_1 &= \{\eta_1 \in \mathbb{R}^1; \eta_1 \leq \theta_1\}, \\ \alpha_2 &= \{\eta_1 \in \mathbb{R}^1; \eta_1 > \theta_1\}, \\ \boldsymbol{\alpha} &= (\alpha_1, \alpha_2), \\ \boldsymbol{r} &= (c, o). \end{aligned}$$

Example: Truncated univariate normal distribution $N = 1$, $I = 2$: We make the same assumptions as in the Censored univariate normal distribution example, except that the latent variable is missing below the threshold. The resulting truncated univariate model can be described by replacing the observation rules above with $\boldsymbol{r} = (m, o)$.

Example: Amemiya's 'Type 2 Tobit model' $N = 2$, $I = 2$ (Amemiya, 1984): Suppose the there are two latent variables ($N = 2$) that follow a bivariate normal distribution. There are two subsets, defined by the value of the first latent variable, η_1. The variable η_1 is censored in both subsets (Amemiya, 1984, p. 31, assumes that only the sign can be observed). The second variable, η_2 is censored in one subset (if η_1 is negative in Amemiya's setting) and observed in the other (if η_1 is positive in Amemiya's setting). This model can be described by

$$\begin{aligned} \alpha_1 &= \{(\eta_1, \eta_2) \in \mathbb{R}^2; \eta_1 \leq \theta_1\}, \\ \alpha_2 &= \{(\eta_1, \eta_2) \in \mathbb{R}^2; \eta_1 > \theta_1\}, \\ \boldsymbol{\alpha} &= (\alpha_1, \alpha_2), \\ \boldsymbol{r} &= (cc, co). \end{aligned}$$

[2] In this definition, and in the next section, we borrow elements from formal languages as used in computational theory, confer for instance Hein (2002). This link to formal languages is only chosen because it enables us to make use of pieces of an already well-established notation.

2.2. **Coding of observations.** We will now offer a way of representing data generated within the framework described above. Broadly speaking, data are assumed to be generated in three steps: first, realizations of the latent variables are drawn. Second, each realization is assigned to a subset α_i with the observation rule r_i. Third, depending on the observation rule for each realization an observation is recorded — say, written to a computer readable file. Since each latent variable can be observable, censored, or missing, and the number of latent variables may be arbitrarily large, this can be complicated. Thus, we think the following, slightly formalistic description is warranted.

Let a realization, t, of the vector of latent variables be denoted as $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{Nt})$. Each realization, $\boldsymbol{\eta}_t$, belongs to a specific subset (α_i) with a corresponding observation rule (r_i). For conceptual purposes we define a vector of *observable stochastic variables*, $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})$, corresponding to realization t of $\boldsymbol{\eta}_t$ and with the same dimension, regardless the observation status. Let the observation rule for realization t be denoted R_t. We can then define an observation as a pair, $(\boldsymbol{y}_t, R_t)$.

Consider first the univariate case: If the latent variable is observable in the subset, then $y_{1t} = \eta_{1t}$; if the latent variable is censored in the subset, then $y_{1t} = i$, which is the subset number; if the latent variable is missing in the subset, then the realization is disregarded – which we denote as $(\boldsymbol{y}_t, R_t) = \Lambda$, Λ representing an empty string. This coding can be illustrated by the following extensions of the three examples above:

Example: Censored univariate normal distribution ($N=1$, $I=2$), continued:

$$(\boldsymbol{y}_t, R_t) = \begin{cases} (\eta_{1t}, o) & \text{if } \eta_{1t} \in \alpha_1, \text{ when } r_1 = o, \\ (2, c) & \text{if } \eta_{1t} \in \alpha_2, \text{ when } r_2 = c. \end{cases}$$

Example: Truncated univariate normal distribution ($N=1$, $I=2$), continued:

$$(\boldsymbol{y}_t, R_t) = \begin{cases} (\eta_{1t}, o) & \text{if } \eta_{1t} \in \alpha_1, \text{ when } r_1 = o, \\ \Lambda & \text{if } \eta_{1t} \in \alpha_2, \text{ when } r_2 = m. \end{cases}$$

Example: Amemiya's 'Type 2 Tobit model' ($N=2$, $I=2$), continued:

$$(\boldsymbol{y}_t, R_t) = \begin{cases} \big((1,1), cc\big) & \text{if } (\eta_{1t}, \eta_{2t}) \in \alpha_1, \text{ when } r_1 = cc, \\ \big((2, \eta_{2t}), co\big) & \text{if } (\eta_{1t}, \eta_{2t}) \in \alpha_2, \text{ when } r_2 = co. \end{cases}$$

For multivariate models an important distinction goes between cases where all latent variables are missing and cases where some are observed or censored. When all are missing we will define $(\boldsymbol{y}_t, R_t) = \Lambda$ when $r_i = m, mm, mmm, \ldots$, as an extension of the definition in the univariate case. When only a subset of the latent variables are missing we can choose to use the same coding as if they were censored. Suppose, for a model with $N=2$, that the observational status in subset i is om, meaning that η_1 is observable and η_2 is missing. Then, assuming we know the subset number, we may, without loss of generality, treat η_2 as censored and code $R_t = oc$ and $y_{2t} = i$, as usual. This type of coding reduces complexity since among the observation rules for the 3^N subsets only (2^N+1) rules are used in the coding:

Example: Bivariate case, $N = 2$. There are $3^N = 9$ possible observation rules for *subset* i, $r_i \in \{mm, mc, mo, cm, cc, co, om, oc, oo\}$. The $(2^N+1) = 5$ *observation rules* used in coding observation t are $R_t \in \{oo, oc, co, cc, mm\}$ as we do not need to distinguish mc from cc, mo from co, cm from cc, and om from oc.

Example: Trivariate case, $N = 3$. There are $3^N = 27$ possible observation rules for *subset* i, $r_i \in \{mmm, mmc, mmo, mcm, mcc, mco, mom, moc, moo, cmm, cmc, cmo, ccm, ccc, cco, com, coc, coo, omm, omc, omo, ocm, occ, oco, oom, ooc, ooo\}$. The $(2^N + 1) = 9$ observation rules used in coding *observation* t are $R_t \in \{ooo, ooc, oco, occ, coo, coc, cco, ccc, mmm\}$.

Above we have implicitly assumed that the subset number is known. This may not always be the case. The coding can be adjusted to accommodate such situations, but some simplicity is lost, confer the following example:

Example: inspired by Cohen (1950, case III), $N = 1, I = 3$: A univariate normal distribution where the real line is divided in three parts by, θ_1, θ_2 ($\theta_1 < \theta_2$) and where censoring occurs in both the upper and lower tails, can be represented as

$$\begin{aligned} \alpha_1 &= \{\eta_1 \in \mathbb{R}^1 : \eta_1 \le \theta_1\}, \\ \alpha_2 &= \{\eta_1 \in \mathbb{R}^1 : \theta_1 < \eta_1 \le \theta_2\}, \\ \alpha_3 &= \{\eta_1 \in \mathbb{R}^1 : \theta_2 < \eta_1\}, \\ \boldsymbol{\alpha} &= (\alpha_1, \alpha_2, \alpha_3), \\ \boldsymbol{r} &= (c, o, c). \end{aligned}$$

If the subsets are known, we would suggest the standard coding:

$$(\boldsymbol{y}_t, R_t) = \begin{cases} (1, c) & \text{if } \eta_{1t} \in \alpha_1, \\ (\eta_{1t}, o) & \text{if } \eta_{1t} \in \alpha_2, \\ (3, c) & \text{if } \eta_{1t} \in \alpha_3. \end{cases}$$

If we are unable to distinguish realizations in the upper tail from those in the lower tail, the realizations can for instance be coded

$$(\boldsymbol{y}_t, R_t) = \begin{cases} (\eta_{1t}, o) & \text{if } \eta_{1t} \in \alpha_2, \\ (2, c) & \text{if } \eta_{1t} \in \alpha_1 \cup \alpha_3. \end{cases}$$

3. Compilation of likelihood functions, uni- and bivariate models

3.1. Univariate models. The density function of $\boldsymbol{\eta} = \eta_1$ is

$$f(\boldsymbol{\eta}, \boldsymbol{\gamma}) = f(\eta_1, \boldsymbol{\gamma}), \tag{3.1}$$

where $\boldsymbol{\gamma}$ is a vector of parameters. We restrict attention to cases where the subsets are defined as continuous intervals, such that each of the I intervals are limited by pairs of thresholds, collected in the vectors $\boldsymbol{\theta}_i = (\underline{\theta}_i, \overline{\theta}_i)$. Specifically

$$\alpha_i = \{\eta_1 \in \mathbb{R} : \underline{\theta}_i \le \eta_1 < \overline{\theta}_i\}, \qquad i = 1, \ldots, I. \tag{3.2}$$

The probability that the latent variable belongs to subset i is denoted

$$\mathcal{F}(\boldsymbol{\theta}_i, \boldsymbol{\gamma}) = \int_{\underline{\theta}_i}^{\overline{\theta}_i} f(\eta_1, \boldsymbol{\gamma}) d\eta, \qquad i = 1, \ldots, I. \tag{3.3}$$

Suppose we have a set of observations which we will refer to as a *sample* and denote T. The likelihood for observation t ($t \in T$) takes different forms depending on the value of R_t. Let us first consider models where η_1 is either observed or censored in all subsets. Then the likelihood for observation t is defined as

$$\mathcal{L}_t(y_t, R_t) = \begin{cases} f(\eta_{1t}, \boldsymbol{\gamma}) & \text{if } R_t = o, \\ \mathcal{F}(\boldsymbol{\theta}_i, \boldsymbol{\gamma}) & \text{if } R_t = c. \end{cases} \tag{3.4}$$

Let $T_i \subseteq T$ denote the subset of observations that falls in subset α_i. We use the notation $|.|$ to denote the number of elements in a set, so that the number of observation in the full sample is $|T| = \sum_{i=1}^{I} T_i$. The likelihood for the full set of observations can be written as

$$\mathcal{L} = \prod_{i=1}^{I} \prod_{t \in T_i} \mathcal{L}_t(y_t, R_t). \tag{3.5}$$

Example: Double censoring (Case II in Cohen, 1950), $N=1$, $I=3$: Suppose there are three pairs of thresholds, so that the real line is split in three parts:

$$\boldsymbol{\theta}_1 \equiv (-\infty, \overline{\theta}_1), \quad \boldsymbol{\theta}_2 \equiv (\overline{\theta}_1, \overline{\theta}_2), \quad \boldsymbol{\theta}_3 \equiv (\overline{\theta}_2, \infty).$$

If the latent variable is censored in α_1 and α_3 while it is observed in α_2, then a sample likelihood for the $|T|$ observations can be represented as

$$\begin{aligned}\mathcal{L} &= \prod_{t \in T_1} \mathcal{L}_t(1, c) \prod_{t \in T_2} \mathcal{L}_t(\eta_{1t}, o) \prod_{t \in T_3} \mathcal{L}_t(3, c) = \prod_{t \in T_1} \mathcal{F}(\boldsymbol{\theta}_1, \boldsymbol{\gamma}) \prod_{t \in T_2} f(\eta_{1t}, \boldsymbol{\gamma}) \prod_{t \in T_3} \mathcal{F}(\boldsymbol{\theta}_3, \boldsymbol{\gamma}) \\ &= \mathcal{F}(\boldsymbol{\theta}_1, \boldsymbol{\gamma})^{|T_1|} \mathcal{F}(\boldsymbol{\theta}_3, \boldsymbol{\gamma})^{|T_3|} \prod_{t \in T_2} f(\eta_{1t}, \boldsymbol{\gamma}).\end{aligned}$$

Example: Quadruple censoring, $N=1$, $I=5$: Suppose there are five pairs of thresholds, so that the real line is split in five parts:

$$\boldsymbol{\theta}_1 \equiv (-\infty, \overline{\theta}_1), \quad \boldsymbol{\theta}_2 \equiv (\underline{\theta}_1, \overline{\theta}_2), \quad \boldsymbol{\theta}_3 \equiv (\underline{\theta}_2, \overline{\theta}_3), \quad \boldsymbol{\theta}_4 \equiv (\underline{\theta}_3, \overline{\theta}_4), \quad \boldsymbol{\theta}_5 \equiv (\underline{\theta}_4, \infty).$$

If the latent variable is censored in α_1, α_3 and α_5 while it is observed in α_2 and α_4, then a sample likelihood for the $|T|$ observations can be represented as

$$\begin{aligned}\mathcal{L} &= \prod_{t \in T_1} \mathcal{L}_t(1, c) \prod_{t \in T_2} \mathcal{L}_t(\eta_{1t}, o) \prod_{t \in T_3} \mathcal{L}_t(3, c) \prod_{t \in T_4} \mathcal{L}_t(\eta_{1t}, o) \prod_{t \in T_5} \mathcal{L}_t(5, c) \\ &= \mathcal{F}(\boldsymbol{\theta}_1, \boldsymbol{\gamma})^{|T_1|} \mathcal{F}(\boldsymbol{\theta}_3, \boldsymbol{\gamma})^{|T_3|} \mathcal{F}(\boldsymbol{\theta}_5, \boldsymbol{\gamma})^{|T_5|} \left[\prod_{t \in T_2} f(\eta_{1t}, \boldsymbol{\gamma})\right]\left[\prod_{t \in T_4} f(\eta_{1t}, \boldsymbol{\gamma})\right].\end{aligned}$$

If we assume η_1 is missing in at least one subset, it convenient to define

$$\mathcal{F}_z = \sum_{i: Y_i = z} \mathcal{F}(\boldsymbol{\theta}_i, \boldsymbol{\gamma}), \qquad z = o, c, m,$$

and note that $\mathcal{F}_o + \mathcal{F}_c + \mathcal{F}_m = 1$. The likelihood for observation t can now be defined

$$\mathcal{L}_t(y_t, R_t) = \begin{cases} \dfrac{f(y_t, \boldsymbol{\gamma})}{\mathcal{F}_o + \mathcal{F}_c} & \text{if } R_t = o, \\ \displaystyle\int_{\underline{\theta}_i}^{\overline{\theta}_i} \frac{f(\eta, \boldsymbol{\gamma})}{\mathcal{F}_o + \mathcal{F}_c} d\eta = \frac{\mathcal{F}(\boldsymbol{\theta}_i, \boldsymbol{\gamma})}{\mathcal{F}_o + \mathcal{F}_c} \quad (y_t = i) & \text{if } R_t = c. \end{cases} \tag{3.6}$$

Otherwise the setup is similar to that considered above, with η_1 missing nowhere.

Example: Double truncation (Case I in Cohen, 1950), $N=1$, $I=3$: Suppose that the thresholds are as in the Double censoring example. Now the latent variable is missing in α_1 and α_3 and still observed in α_2. Then the likelihood function for the $|T_2|$ observations can be represented as

$$\mathcal{L} = \prod_{t \in T_2} \mathcal{L}_t(y_t, o) = \prod_{t \in T_2} \frac{f(\eta_{1t}, \boldsymbol{\gamma})}{F_o + F_c} = \prod_{t \in T_2} \frac{f(\eta_{1t}, \boldsymbol{\gamma})}{\mathcal{F}(\boldsymbol{\theta}_2, \boldsymbol{\gamma})}.$$

EXAMPLE: QUADRUPLE TRUNCATION, $N=1$, $I=5$: Suppose that the thresholds are as in the Quadruple censoring example. Now, the latent variable is assumed to be missing in α_1, α_3 and α_5 while it is observed in α_2 and α_4. Then a sample likelihood for the $|T_2|+|T_4|$ observations can be represented as

$$\mathcal{L} = \prod_{t\in T_2} \mathcal{L}_t(y_t, o) \prod_{t\in T_4} \mathcal{L}_t(y_t, o) = \Big[\prod_{t\in T_2} \frac{f(\eta_{1t},\boldsymbol{\gamma})}{\mathcal{F}(\boldsymbol{\theta}_2,\boldsymbol{\gamma})}\Big]\Big[\prod_{t\in T_4} \frac{f(\eta_{1t},\boldsymbol{\gamma})}{\mathcal{F}(\boldsymbol{\theta}_4,\boldsymbol{\gamma})}\Big]$$
$$= [\mathcal{F}(\boldsymbol{\theta}_2,\boldsymbol{\gamma})]^{-|T_2|}[\mathcal{F}(\boldsymbol{\theta}_4,\boldsymbol{\gamma})]^{-|T_4|}\Big[\prod_{t\in T_2} f(\eta_{1t},\boldsymbol{\gamma})\Big]\Big[\prod_{t\in T_4} f(\eta_{1t},\boldsymbol{\gamma})\Big].$$

3.2. **Bivariate models.** The density function of $\boldsymbol{\eta}$ is

$$f(\boldsymbol{\eta},\boldsymbol{\gamma}) = f(\eta_1,\eta_2,\boldsymbol{\gamma}), \tag{3.7}$$

where again $\boldsymbol{\gamma}$ is a vector of parameters known to the 'compiler'. There are I subsets shaped as rectangles. Each subset is limited by two pairs of known[3] thresholds, $\boldsymbol{\theta}_{1i} = (\underline{\theta}_{1i}, \overline{\theta}_{1i})$ and $\boldsymbol{\theta}_{2i} = (\underline{\theta}_{2i}, \overline{\theta}_{2i})$, so that

$$\alpha_i = \{(\eta_1,\eta_2) \in \mathbb{R}^2 : \quad \underline{\theta}_{1i} \le \eta_1 < \overline{\theta}_{1i}, \underline{\theta}_{2i} \le \eta_2 < \overline{\theta}_{2i}\}. \tag{3.8}$$

Define

$$\mathcal{F}[\boldsymbol{\theta}_i,\boldsymbol{\gamma}] = \int_{\underline{\theta}_{1i}}^{\overline{\theta}_{1i}} \int_{\underline{\theta}_{2i}}^{\overline{\theta}_{2i}} f(\eta_1,\eta_2,\boldsymbol{\gamma})d\eta_1 d\eta_2, \tag{3.9}$$

$$\begin{aligned} F_1[\eta_2,\boldsymbol{\theta}_{1i},\boldsymbol{\gamma}] &= \int_{\underline{\theta}_{1i}}^{\overline{\theta}_{1i}} f(\eta_1,\eta_2,\boldsymbol{\gamma})d\eta_1, \\ F_2[\eta_1,\boldsymbol{\theta}_{2i},\boldsymbol{\gamma}] &= \int_{\underline{\theta}_{2i}}^{\overline{\theta}_{2i}} f(\eta_1,\eta_2,\boldsymbol{\gamma})d\eta_2, \end{aligned} \qquad i = 1,\dots,I, \tag{3.10}$$

where $\boldsymbol{\theta}_i = (\boldsymbol{\theta}_{1i},\boldsymbol{\theta}_{2i})$ and subscripts 1 and 2 on the F functions indicate that they are obtained by integrating the density f across the intervals of η_1 and of η_2, respectively. The probability that $\boldsymbol{\eta}$ belongs to subset (rectangle) i is $\mathcal{F}[\boldsymbol{\theta}_i,\boldsymbol{\gamma}]$. Furthermore, let

$$\begin{aligned} \mathcal{F}_z &= \textstyle\sum_{i:Y_i=z} \mathcal{F}(\boldsymbol{\theta}_i,\boldsymbol{\gamma}), \qquad z = oo, oc, om, co, cc, cm, mo, mc, mm, \\ \textstyle\sum_z \mathcal{F}_z &= \textstyle\sum_z \sum_{i:Y_i=z} \mathcal{F}(\boldsymbol{\theta}_i,\boldsymbol{\gamma}) = 1. \end{aligned}$$

Using (3.9) we can formalize how the form of the likelihood for observation t, $\mathcal{L}_t(\boldsymbol{y}_t, R_t)$, depends on R_t. If neither of (η_{1t},η_{2t}) is missing in any rectangle, *i.e.*, $F_{mm}=0$, then (3.13) takes the simpler form, generalizing (3.4):

$$\mathcal{L}_t(\boldsymbol{y}_t, R_t) = \begin{cases} f(y_{1t}, y_{2t}, \boldsymbol{\gamma}) & \text{if } \boldsymbol{\eta}_t \in \alpha_i \text{ and } R_t = oo, \\ F_1(y_{2t}, \boldsymbol{\theta}_{1i}, \boldsymbol{\gamma}) & \text{if } \boldsymbol{\eta}_t \in \alpha_i \text{ and } R_t = co, \\ F_2(y_{1t}, \boldsymbol{\theta}_{2i}, \boldsymbol{\gamma}) & \text{if } \boldsymbol{\eta}_t \in \alpha_i \text{ and } R_t = oc, \\ \mathcal{F}(\boldsymbol{\theta}_i,\boldsymbol{\gamma}) & \text{if } \boldsymbol{\eta}_t \in \alpha_i \text{ and } R_t = cc. \end{cases} \tag{3.11}$$

The likelihood for all observations, generalizing (3.5), becomes

$$\mathcal{L} = \prod_{i=1}^{I} \prod_{t\in T_i} \mathcal{L}_t(\boldsymbol{y}_t, R_t). \tag{3.12}$$

[3]If some thresholds are unknown to the analyst – reflecting, *e.g.*, that the variables are only ordinally measurable – they will become unknown parameters in the (elements of the) likelihood function.

The generalization of (3.6), accounting for missing observations, is

$$
\mathcal{L}_t(\boldsymbol{y}_t, R_t) = \begin{cases} \dfrac{f(y_{1t}, y_{2t}, \boldsymbol{\gamma})}{\sum_{z \neq mm} \mathcal{F}_z} & \text{if } \boldsymbol{\eta}_t \in \alpha_i \text{ and } R_t = oo, \\ \dfrac{F_1(y_{2t}, \boldsymbol{\theta}_{1i}, \boldsymbol{\gamma})}{\sum_{z \neq mm} \mathcal{F}_z} & \text{if } \boldsymbol{\eta}_t \in \alpha_i \text{ and } R_t \in \{co, mo\}, \\ \dfrac{F_2(y_{1t}, \boldsymbol{\theta}_{2i}, \boldsymbol{\gamma})}{\sum_{z \neq mm} \mathcal{F}_z} & \text{if } \boldsymbol{\eta}_t \in \alpha_i \text{ and } R_t \in \{oc, om\}, \\ \dfrac{\mathcal{F}(\boldsymbol{\theta}_i, \boldsymbol{\gamma})}{\sum_{z \neq mm} \mathcal{F}_z} & \text{if } \boldsymbol{\eta}_t \in \alpha_i \text{ and } R_t \in \{cc, cm, mc\}. \end{cases} \tag{3.13}
$$

The analyst may, for convenience or some other reason, choose to curtail the sample, by including *only observations for which no observation is missing*, *i.e.*, include only the $|T_{oo}|+|T_{co}|+|T_{oc}|+|T_{cc}|$ observations for which $R_t = oo, co, oc, cc$. Then (3.13) should be replaced by

$$
\mathcal{L}_t(\boldsymbol{y}_t, R_t) = \begin{cases} \dfrac{f(y_{1t}, y_{2t}, \boldsymbol{\gamma})}{\mathcal{F}_{oo}+\mathcal{F}_{co}+\mathcal{F}_{oc}+\mathcal{F}_{cc}} & \text{if } \boldsymbol{\eta}_t \in \alpha_i \text{ and } R_t = oo, \\ \dfrac{F_1(y_{2t}, \boldsymbol{\theta}_{1i}, \boldsymbol{\gamma})}{\mathcal{F}_{oo}+\mathcal{F}_{co}+\mathcal{F}_{oc}+\mathcal{F}_{cc}} & \text{if } \boldsymbol{\eta}_t \in \alpha_i \text{ and } R_t = co, \\ \dfrac{F_2(y_{1t}, \boldsymbol{\theta}_{2i}, \boldsymbol{\gamma})}{\mathcal{F}_{oo}+\mathcal{F}_{co}+\mathcal{F}_{oc}+\mathcal{F}_{cc}} & \text{if } \boldsymbol{\eta}_t \in \alpha_i \text{ and } R_t = oc, \\ \dfrac{\mathcal{F}(\boldsymbol{\theta}_i, \boldsymbol{\gamma})}{\mathcal{F}_{oo}+\mathcal{F}_{co}+\mathcal{F}_{oc}+\mathcal{F}_{cc}} & \text{if } \boldsymbol{\eta}_t \in \alpha_i \text{ and } R_t = cc, \end{cases} \tag{3.14}
$$

and

$$
\begin{aligned} \mathcal{L} = {} & [\mathcal{F}_{oo}+\mathcal{F}_{co}+\mathcal{F}_{oc}+\mathcal{F}_{cc}]^{-(|T_{oo}|+|T_{co}|+|T_{oc}|+|T_{cc}|)} \\ & \times \prod_{t \in T_{oo}} \mathcal{L}_t(\boldsymbol{y}_t, R_t) \prod_{t \in T_{oc}} \mathcal{L}_t(\boldsymbol{y}_t, R_t) \prod_{t \in T_{co}} \mathcal{L}_t(\boldsymbol{y}_t, R_t) \prod_{t \in T_{cc}} \mathcal{L}_t(\boldsymbol{y}_t, R_t). \end{aligned} \tag{3.15}
$$

Example: Bivariate ordered response: Assume that $I = JK$ and let subset α_i ($i = 1, \dots, I$) is defined as follows:

$$
\boldsymbol{\eta} = (\eta_1, \eta_2) \in \alpha_i \quad \text{means:} \quad \begin{cases} \eta_1 \in (\underline{\theta}_{1,j}, \overline{\theta}_{1,j}], & j = 1, 2, \dots, J, \\ \eta_2 \in (\underline{\theta}_{2,k}, \overline{\theta}_{2,k}], & k = 1, 2, \dots, K, \end{cases} \quad i = j + (k-1)J.
$$

where $[\underline{\theta}_{1,j}, \overline{\theta}_{1,j}]_{j=1}^{j=J}$ and $[\underline{\theta}_{2,k}, \overline{\theta}_{2,k}]_{k=1}^{k=K}$ denote the threshold values, for each variable arranged in ascending order, $\underline{\theta}_{1,1} = \underline{\theta}_{2,1} = -\infty$, $\overline{\theta}_{1,J} = \overline{\theta}_{2,K} = \infty$, and $\underline{\theta}_{1,j} = \overline{\theta}_{1,j-1}$ ($j = 2, \dots, J$), $\underline{\theta}_{2,k} = \overline{\theta}_{2,k-1}$ ($k = 2, \dots, K$). The observed variables, defined as the interval indexes, are

$$
\begin{aligned} y_1 &= \textstyle\sum_{j=1}^{J} j \mathbf{1}\{\eta_1 \in (\underline{\theta}_{1,j}, \overline{\theta}_{1,j}]\}, \\ y_2 &= \textstyle\sum_{k=1}^{K} k \mathbf{1}\{\eta_2 \in (\underline{\theta}_{2,k}, \overline{\theta}_{2,k}]\}. \end{aligned}
$$

Let $\mathcal{F}(\boldsymbol{\theta}_i, \boldsymbol{\gamma})$ be defined as in (3.9). The likelihood function for observation t, $\boldsymbol{y}_t = (y_{1t}, y_{2t})$ can then be written as

$$
\mathcal{L}_t(\boldsymbol{y}_t, R_t) = \mathcal{F}(\boldsymbol{\theta}_i; \boldsymbol{\gamma}) \quad \text{if} \quad \boldsymbol{\eta} \in \alpha_i, \qquad i = 1, \dots, I.
$$

Let T_i denote the set of observations which belongs to subset i, let $|T_i|$ be the number of such observations, and let $T = \cup_{i=1}^{I} T_i$ and $|T| = \sum_{i=1}^{I} |T_i|$. The sample likelihood for the $|T|$ observations can then be written as

$$\mathcal{L} = \prod_{t \in T} \mathcal{L}_t(\boldsymbol{y}_t, R_t) = \prod_{i=1}^{I} \prod_{t \in T_i} \mathcal{L}_t(\boldsymbol{y}_t, R_t) = \prod_{i=1}^{I} \prod_{t \in T_i} \mathcal{F}(\boldsymbol{\theta}_i, \boldsymbol{\gamma}) = \prod_{i=1}^{I} \prod_{t : \boldsymbol{\eta}_t \in \alpha_i} \mathcal{F}(\boldsymbol{\theta}_i, \boldsymbol{\gamma}).$$

Although $\mathcal{L}$ is formally a function of $\boldsymbol{y}_1, \dots, \boldsymbol{y}_{|T|}$, the expression after the last equality sign does not involve the particular choice of $\boldsymbol{y}$ 'metric' adopted. The pair of interval indexes $y_1 = \sum_{j=1}^{J} j\mathbf{1}\{\eta_1 \in (\underline{\theta}_{1,j}, \overline{\theta}_{1,j}]\}$ and $y_2 = \sum_{k=1}^{K} k\mathbf{1}\{\eta_2 \in (\underline{\theta}_{2,k}, \overline{\theta}_{2,k}]\}$ do not enter the function. We might have used for example the set of $I = JK$ subset (rectangle) dummies instead.

4. Extensions

We now generalize the setup in Section 3 in two respects. First, we generalize from $N=1$ and 2 to an arbitrary dimension (Section 4.1), second, we introduce covariates which have so far been suppressed (Section 4.2). The first generalization necessitates some more compact notation.

4.1. **Extending to N-variate case.** Assume that $\boldsymbol{\eta}$ follows an N-variate distribution with density $f(\boldsymbol{\eta}, \boldsymbol{\gamma})$. Subset i is defined by

$$\boldsymbol{\alpha}_i = \{(\eta_1, \dots, \eta_N) \in \mathbb{R}^N : \underline{\theta}_{ni} \leq \eta_n < \overline{\theta}_{ni}, \quad n = 1, \dots, N\}. \tag{4.1}$$

Since N and I are finite, the number of interval bounds is at most $2NI$, which is a finite number. Some bounds may be unknown to the analyst and therefore unknown parameters in the likelihood function for the model. We also define interval and variable specific vectors of bounds,

$$\boldsymbol{\theta}_{ni} \equiv \left(\underline{\theta}_{ni}, \overline{\theta}_{ni}\right), \qquad n = 1, \dots, N;\ i = 1, \dots, I.$$

Define the index set

$$\mathcal{N} \equiv \{1, \dots, N\}.$$

and let $\mathcal{A}_i$ and it complement $\mathcal{A}_i^*$ be any set containing the indices of the variables which, *in subset* i, are observed and non-observed (*i.e.*, either censored or missing), respectively.[4] Let the elements in $\boldsymbol{\eta}$ which, in subset i, are observed and non-observed, be collected in, respectively,

$$\boldsymbol{\eta}_{\mathcal{A}i} \equiv \{\eta_n : n \in \mathcal{A}_i\}, \qquad \boldsymbol{\eta}_{\mathcal{A}*i} \equiv \{\eta_n : n \in \mathcal{A}_i^*\},$$

and let $\boldsymbol{\theta}_{\mathcal{A}i}$ and $\boldsymbol{\theta}_{\mathcal{A}*i}$ be the set of interval bounds relating to observed and unobserved variables, respectively, in subset i:

$$\boldsymbol{\theta}_{\mathcal{A}i} \equiv \{\boldsymbol{\theta}_{ni} : n \in \mathcal{A}_i\}, \qquad \boldsymbol{\theta}_{\mathcal{A}*i} \equiv \{\boldsymbol{\theta}_{ni} : n \in \mathcal{A}_i^*\}.$$

In particular,

$$\mathcal{A}_i = \mathcal{N} \implies \boldsymbol{\eta}_{\mathcal{A}i} = \boldsymbol{\eta}, \quad \boldsymbol{\alpha}_{\mathcal{A}i} = \boldsymbol{\alpha}_i, \quad \boldsymbol{\theta}_{\mathcal{A}i} = \boldsymbol{\theta}_i.$$

[4] For any i: $\mathcal{A}_i \subseteq \mathcal{N}, \quad \mathcal{A}_i^* \subseteq \mathcal{N}, \quad \mathcal{A}_i \cup \mathcal{A}_i^* = \mathcal{N}, \quad \mathcal{A}_i \cap \mathcal{A}_i^* = \emptyset.$

Amemiya's 'Type 2 Tobit model' $N=2$, $I=2$ reconsidered. In the compact notation, the 'Tobit 2 type' example, considered in Section 2.1, reads

For α_1: $\mathcal{A}_i=\emptyset,\ \mathcal{A}_i^*=\{1,2\}$; $\boldsymbol{\eta}_{\mathcal{A}i}=\emptyset,\ \boldsymbol{\eta}_{\mathcal{A}_i^*}=(\eta_1,\eta_2)$;
$\boldsymbol{\theta}_{\mathcal{A}}=\emptyset,\ \boldsymbol{\theta}_{\mathcal{A}^*}=(\boldsymbol{\theta}_{11},\boldsymbol{\theta}_{21})=[(-\infty,\theta_1),(-\infty,\infty)]$.

For α_2: $\mathcal{A}_i=\{2\},\ \mathcal{A}_i^*=\{1\}$; $\boldsymbol{\eta}_{\mathcal{A}i}=\eta_2,\ \boldsymbol{\eta}_{\mathcal{A}_i^*}=\eta_1$;
$\boldsymbol{\theta}_{\mathcal{A}}=\boldsymbol{\theta}_{22}=(-\infty,\infty),\ \boldsymbol{\theta}_{\mathcal{A}^*}=\boldsymbol{\theta}_{12}=(\theta_1,\infty)$.

The total number of $\mathcal{A}_i$ sets is, in principle, $2^N\ \forall i$, of which $N_p\equiv\binom{N}{p}$ $(p=0,\dots,N)$ sets contain p observed and $N-p$ non-observed variables. A set with all variables and a set with no variables observed (*i.e.*, either fully censored or fully missing) correspond to $N_N=\binom{N}{N}=1$ and $N_0=\binom{N}{0}=1$, respectively. This concurs with the binomial formula $\sum_{p=0}^{N}\binom{N}{p}=2^N$.[5] Let (p,r) index selection no. r $(r=1,\dots,N_p)$ among those sets which have p observable variables. Let $\mathcal{A}(p,r)_i$ symbolize that i belongs to the r'th selection among the sets containing p observed variables r $(r=1,\dots,N_p)$.

From the density $f(\boldsymbol{\eta},\boldsymbol{\gamma})$ we can define the prototype element in the likelihood function for any observability status in subset i characterized by the set $\mathcal{A}_i$ as follows:

$$F_{\mathcal{A}*i}(\boldsymbol{\eta}_{\mathcal{A}i},\boldsymbol{\theta}_{\mathcal{A}*i},\boldsymbol{\gamma})\equiv\int_{\boldsymbol{\eta}_{\mathcal{A}*i}\in\boldsymbol{\theta}_{\mathcal{A}*i}}f(\boldsymbol{\eta},\boldsymbol{\gamma})d\boldsymbol{\eta}_{\mathcal{A}*i},\qquad i=1,\dots,I. \tag{4.2}$$

Its arguments are: (i) the *observable* elements of the $\boldsymbol{\eta}$ vector in subset i, (ii) the interval bounds of the *non-observable* elements in subset i, and (iii) the parameter vector $\boldsymbol{\gamma}$ of the distribution of $\boldsymbol{\eta}$. This equation generalizes (3.10). The integration in constructing $F_{\mathcal{A}*i}(\cdot)$ goes across the *non-observable variables*, making the result a function of their known or (to the analyst) unknown interval bounds. For subsets with all, respectively no, variables observed, we have in particular:

$$F_{\mathcal{A}*i}(\boldsymbol{\eta}_{\mathcal{A}i},\boldsymbol{\theta}_{\mathcal{A}*i},\boldsymbol{\gamma})=f(\boldsymbol{\eta},\boldsymbol{\gamma}),\qquad \text{for } \mathcal{A}_i=\mathcal{N},\quad \mathcal{A}_i^*=\emptyset,$$
$$F_{\mathcal{A}*i}(\boldsymbol{\eta}_{\mathcal{A}i},\boldsymbol{\theta}_{\mathcal{A}*i},\boldsymbol{\gamma})=\int_{\boldsymbol{\eta}\in\boldsymbol{\theta}_i}f(\boldsymbol{\eta},\boldsymbol{\gamma})d\boldsymbol{\eta}\equiv\mathcal{F}(\boldsymbol{\theta}_i,\boldsymbol{\gamma}),\qquad \text{for } \mathcal{A}_i=\emptyset,\quad \mathcal{A}_i^*=\mathcal{N},$$

where $\mathcal{F}(\boldsymbol{\theta}_i,\boldsymbol{\gamma})$ is the subset probability for subset i; $\sum_{i=1}^{N}\mathcal{F}(\boldsymbol{\theta}_i,\boldsymbol{\gamma})=1$.

Example: $N=4$. The likelihood elements $F_{\mathcal{A}*i}(\boldsymbol{\eta}_{\mathcal{A}i},\boldsymbol{\theta}_{\mathcal{A}*i},\boldsymbol{\gamma})$ in $F_{\mathcal{A}(p,r)}(\cdot)$ notation. The likelihood function elements below form a *recursive structure*, with the density $f(\eta_1,\eta_2,\eta_3,\eta_4,\boldsymbol{\gamma})$ at the bottom. The recursion can be exemplified as follows. (Recursive computation may not, however, be a recommendable *numerical* procedure, owing to cumulative approximation errors.):

[5] The two middle alternatives specified for the bivariate case in, say, (3.11), *i.e.* $F_1(\cdot)$ and $F_2(\cdot)$, relate to the $\binom{2}{1}=2$ subsets with either η_1 or η_2 observed.

$$
\begin{array}{ll}
p=4,\ N_p=1: & F_{\mathcal{A}(4,1)}[\eta_1,\eta_2,\eta_3,\eta_4,\boldsymbol{\theta},\boldsymbol{\gamma}] = f(\eta_1,\eta_2,\eta_3,\eta_4,\boldsymbol{\gamma}), \\[1ex]
p=3,\ N_p=4: & \begin{array}{l}
F_{\mathcal{A}(3,1)}[\eta_1,\eta_2,\eta_3,\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{4i}}^{\overline{\theta}_{4i}} f(\eta_1,\eta_2,\eta_3,\eta_4,\boldsymbol{\gamma})d\eta_4, \\
F_{\mathcal{A}(3,2)}[\eta_1,\eta_2,\eta_4,\boldsymbol{\theta}_{3i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{3i}}^{\overline{\theta}_{3i}} f(\eta_1,\eta_2,\eta_3,\eta_4,\boldsymbol{\gamma})d\eta_3, \\
F_{\mathcal{A}(3,3)}[\eta_1,\eta_3,\eta_4,\boldsymbol{\theta}_{2i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{2i}}^{\overline{\theta}_{2i}} f(\eta_1,\eta_2,\eta_3,\eta_4,\boldsymbol{\gamma})d\eta_2, \\
F_{\mathcal{A}(3,4)}[\eta_2,\eta_3,\eta_4,\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{1i}}^{\overline{\theta}_{1i}} f(\eta_1,\eta_2,\eta_3,\eta_4,\boldsymbol{\gamma})d\eta_1,
\end{array} \\[1ex]
p=2,\ N_p=6: & \begin{array}{l}
F_{\mathcal{A}(2,1)}[\eta_1,\eta_2,\boldsymbol{\theta}_{3i},\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{3i}}^{\overline{\theta}_{3i}} F_{\mathcal{A}(3,1)}[\eta_1,\eta_2,\eta_3,\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}]d\eta_3, \\
F_{\mathcal{A}(2,2)}[\eta_1,\eta_3,\boldsymbol{\theta}_{2i},\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{2i}}^{\overline{\theta}_{2i}} F_{\mathcal{A}(3,1)}[\eta_1,\eta_2,\eta_3,\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}]d\eta_2, \\
F_{\mathcal{A}(2,3)}[\eta_1,\eta_4,\boldsymbol{\theta}_{2i},\boldsymbol{\theta}_{3i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{2i}}^{\overline{\theta}_{2i}} F_{\mathcal{A}(3,2)}[\eta_1,\eta_2,\eta_4,\boldsymbol{\theta}_{3i},\boldsymbol{\gamma}]d\eta_2, \\
F_{\mathcal{A}(2,4)}[\eta_2,\eta_3,\boldsymbol{\theta}_{1i},\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{1i}}^{\overline{\theta}_{1i}} F_{\mathcal{A}(3,1)}[\eta_1,\eta_2,\eta_3,\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}]d\eta_1, \\
F_{\mathcal{A}(2,5)}[\eta_2,\eta_4,\boldsymbol{\theta}_{1i},\boldsymbol{\theta}_{3i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{1i}}^{\overline{\theta}_{1i}} F_{\mathcal{A}(3,2)}[\eta_1,\eta_2,\eta_4,\boldsymbol{\theta}_{3i},\boldsymbol{\gamma}]d\eta_1, \\
F_{\mathcal{A}(2,6)}[\eta_3,\eta_4,\boldsymbol{\theta}_{1i},\boldsymbol{\theta}_{2i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{1i}}^{\overline{\theta}_{1i}} F_{\mathcal{A}(3,3)}[\eta_1,\eta_3,\eta_4,\boldsymbol{\theta}_{2i},\boldsymbol{\gamma}]d\eta_1,
\end{array} \\[1ex]
p=1,\ N_p=4: & \begin{array}{l}
F_{\mathcal{A}(1,1)}[\eta_1,\boldsymbol{\theta}_{2i},\boldsymbol{\theta}_{3i},\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{2i}}^{\overline{\theta}_{2i}} F_{\mathcal{A}(2,1)}[\eta_1,\eta_2,\boldsymbol{\theta}_{3i},\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}]d\eta_2, \\
F_{\mathcal{A}(1,2)}[\eta_2,\boldsymbol{\theta}_{1i},\boldsymbol{\theta}_{3i},\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{1i}}^{\overline{\theta}_{1i}} F_{\mathcal{A}(2,1)}[\eta_1,\eta_2,\boldsymbol{\theta}_{3i},\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}]d\eta_1, \\
F_{\mathcal{A}(1,3)}[\eta_3,\boldsymbol{\theta}_{1i},\boldsymbol{\theta}_{2i},\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{1i}}^{\overline{\theta}_{1i}} F_{\mathcal{A}(2,2)}[\eta_1,\eta_3,\boldsymbol{\theta}_{2i},\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}]d\eta_1, \\
F_{\mathcal{A}(1,4)}[\eta_4,\boldsymbol{\theta}_{1i},\boldsymbol{\theta}_{2i},\boldsymbol{\theta}_{3i},\boldsymbol{\gamma}] = \int_{\underline{\theta}_{1i}}^{\overline{\theta}_{1i}} F_{\mathcal{A}(2,3)}[\eta_1,\eta_4,\boldsymbol{\theta}_{2i},\boldsymbol{\theta}_{3i},\boldsymbol{\gamma}]d\eta_1,
\end{array} \\[1ex]
p=0,\ N_p=1: & \begin{array}{ll}
F_{\mathcal{A}(0,1)}[\boldsymbol{\theta}_{1i},\boldsymbol{\theta}_{2i},\boldsymbol{\theta}_{3i},\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}] & = \mathcal{F}[\boldsymbol{\theta}_i,\boldsymbol{\gamma}] \\
& = \int_{\underline{\theta}_{1i}}^{\overline{\theta}_{1i}}\int_{\underline{\theta}_{2i}}^{\overline{\theta}_{2i}}\int_{\underline{\theta}_{3i}}^{\overline{\theta}_{3i}}\int_{\underline{\theta}_{4i}}^{\overline{\theta}_{4i}} f(\eta_1,\eta_2,\eta_3,\eta_4,\boldsymbol{\gamma})d\eta_1\,d\eta_2\,d\eta_3\,d\eta_4 \\
& = \int_{\underline{\theta}_{1i}}^{\overline{\theta}_{1i}} F_{\mathcal{A}(1,1)}[\eta_1,\boldsymbol{\theta}_{2i},\boldsymbol{\theta}_{3i},\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}]d\eta_{1i} \\
& = \int_{\underline{\theta}_{2i}}^{\overline{\theta}_{2i}} F_{\mathcal{A}(1,2)}[\eta_2,\boldsymbol{\theta}_{1i},\boldsymbol{\theta}_{3i},\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}]d\eta_{2i} \\
& = \int_{\underline{\theta}_{3i}}^{\overline{\theta}_{3i}} F_{\mathcal{A}(1,3)}[\eta_3,\boldsymbol{\theta}_{1i},\boldsymbol{\theta}_{2i},\boldsymbol{\theta}_{4i},\boldsymbol{\gamma}]d\eta_{3i} \\
& = \int_{\underline{\theta}_{4i}}^{\overline{\theta}_{4i}} F_{\mathcal{A}(1,4)}[\eta_4,\boldsymbol{\theta}_{1i},\boldsymbol{\theta}_{2i},\boldsymbol{\theta}_{3i},\boldsymbol{\gamma}]d\eta_{4i}
\end{array}
\end{array}
$$

If no variable is missing in any rectangle – so that all $\mathcal{A}_i^*$ contain *censored variables only* – we can then, letting t index observation and $\boldsymbol{y}_{\mathcal{A}it}=\boldsymbol{\eta}_{\mathcal{A}it}$, generalize (3.11) as

$$
\mathcal{L}_t(\boldsymbol{y}_t) = \begin{cases} f(\boldsymbol{y}_t,\boldsymbol{\gamma}) & \text{if } \boldsymbol{\eta}_t \in \alpha_i,\ \mathcal{A}_i = \mathcal{N},\ \mathcal{A}_i^* = \emptyset, \\ F_{\mathcal{A}*i}(\boldsymbol{y}_{\mathcal{A}it},\boldsymbol{\theta}_{\mathcal{A}*i},\boldsymbol{\gamma}) & \text{if } \boldsymbol{\eta}_t \in \alpha_i,\ \mathcal{A}_i \subset \mathcal{N},\ \mathcal{A}_i^* \subset \mathcal{N}, \\ \mathcal{F}(\boldsymbol{\theta}_i,\boldsymbol{\gamma}) & \text{if } \boldsymbol{\eta}_t \in \alpha_i,\ \mathcal{A}_i = \emptyset,\ \mathcal{A}_i^* = \mathcal{N}. \end{cases} \tag{4.3}
$$

Let, in general, $\mathcal{F}_{NM}(\boldsymbol{\theta},\boldsymbol{\gamma})$, $\mathcal{F}_{SM}(\boldsymbol{\theta},\boldsymbol{\gamma})$, $\mathcal{F}_{AM}(\boldsymbol{\theta},\boldsymbol{\gamma})$ denote the total subset probabilities for the subsets, where, respectively, no variable, some variables, and all variables are missing:

$$
\begin{aligned}
\mathcal{F}_{NM}(\boldsymbol{\theta},\boldsymbol{\gamma}) &= \textstyle\sum_{i:\text{No}\,\eta\text{-var. missing}} \mathcal{F}(\boldsymbol{\theta}_i,\boldsymbol{\gamma}), \\
\mathcal{F}_{SM}(\boldsymbol{\theta},\boldsymbol{\gamma}) &= \textstyle\sum_{i:\text{Some}\,\eta\text{-var. missing}} \mathcal{F}(\boldsymbol{\theta}_i,\boldsymbol{\gamma}), \\
\mathcal{F}_{AM}(\boldsymbol{\theta},\boldsymbol{\gamma}) &= \textstyle\sum_{i:\text{All}\,\eta\text{-var. missing}} \mathcal{F}(\boldsymbol{\theta}_i,\boldsymbol{\gamma}), \\
\mathcal{F}_{NM}(\boldsymbol{\theta},\boldsymbol{\gamma}) &+ \mathcal{F}_{SM}(\boldsymbol{\theta},\boldsymbol{\gamma}) + \mathcal{F}_{AM}(\boldsymbol{\theta},\boldsymbol{\gamma}) = 1
\end{aligned}
$$

Likewise if missing variables are allowed for – provided that all $\mathcal{A}_i^*$ contain *at least some censored variables* – generalize (3.13) to

$$
\mathcal{L}_t(\boldsymbol{y}_t) = \begin{cases} \dfrac{f(\boldsymbol{y}_t, \boldsymbol{\gamma})}{\mathcal{F}_{NM}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + \mathcal{F}_{SM}(\boldsymbol{\theta}, \boldsymbol{\gamma})} & \text{if } \boldsymbol{\eta}_t \in \alpha_i,\ \mathcal{A}_i = \mathcal{N},\ \mathcal{A}_i^* = \emptyset, \\ \dfrac{F_{\mathcal{A}*i}(\boldsymbol{y}_{\mathcal{A}it}, \boldsymbol{\theta}_{\mathcal{A}*i}, \boldsymbol{\gamma})}{\mathcal{F}_{NM}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + \mathcal{F}_{SM}(\boldsymbol{\theta}, \boldsymbol{\gamma})} & \text{if } \boldsymbol{\eta}_t \in \alpha_i,\ \mathcal{A}_i \subset \mathcal{N},\ \mathcal{A}_i^* \subset \mathcal{N}, \\ \dfrac{\mathcal{F}(\boldsymbol{\theta}_i, \boldsymbol{\gamma})}{\mathcal{F}_{NM}(\boldsymbol{\theta}, \boldsymbol{\gamma}) + \mathcal{F}_{SM}(\boldsymbol{\theta}, \boldsymbol{\gamma})} & \text{if } \boldsymbol{\eta}_t \in \alpha_i,\ \mathcal{A}_i = \emptyset,\ \mathcal{A}_i^* = \mathcal{N}. \end{cases} \tag{4.4}
$$

Finally, if the analyst actively chooses to curtail the sample by omitting only observation sets for which some observations are missing (when $\mathcal{F}_{SM}(\boldsymbol{\theta}, \boldsymbol{\gamma}) > 0$ and $\mathcal{F}_{AM}(\boldsymbol{\theta}, \boldsymbol{\gamma}) > 0$ in the original data set) which is often done, the generalization of (3.14) – so that all $\mathcal{A}_i^*$ contain *censored variables only* – becomes

$$
\mathcal{L}_t(\boldsymbol{y}_t) = \begin{cases} \dfrac{f(\boldsymbol{y}_t, \boldsymbol{\gamma})}{\mathcal{F}_{NM}(\boldsymbol{\theta}, \boldsymbol{\gamma})} & \text{if } \boldsymbol{\eta}_t \in \alpha_i,\ \mathcal{A}_i = \mathcal{N},\ \mathcal{A}_i^* = \emptyset, \\ \dfrac{F_{\mathcal{A}*i}(\boldsymbol{y}_{\mathcal{A}it}, \boldsymbol{\theta}_{\mathcal{A}*i}, \boldsymbol{\gamma})}{\mathcal{F}_{NM}(\boldsymbol{\theta}, \boldsymbol{\gamma})} & \text{if } \boldsymbol{\eta}_t \in \alpha_i,\ \mathcal{A}_i \subset \mathcal{N},\ \mathcal{A}_i^* \subset \mathcal{N}, \\ \dfrac{\mathcal{F}(\boldsymbol{\theta}_i, \boldsymbol{\gamma})}{\mathcal{F}_{NM}(\boldsymbol{\theta}, \boldsymbol{\gamma})} & \text{if } \boldsymbol{\eta}_t \in \alpha_i,\ \mathcal{A}_i = \emptyset,\ \mathcal{A}_i^* = \mathcal{N}. \end{cases} \tag{4.5}
$$

Using the $t \in \mathcal{A}(p,r)_i$ notation to symbolize that observation t in subset i belongs to the r'th selection among those set having p observed variables r ($r = 1, \ldots, N_p$). The prototype expression for the likelihood function, generalizing all those above can then be written as

$$
\mathcal{L} = \prod_{i=1}^{I} \prod_{p=0}^{N} \prod_{r=1}^{N_p} \prod_{t \in \mathcal{A}(p,r)_i} \mathcal{L}_t(\boldsymbol{y}_t), \tag{4.6}
$$

where we can insert from (4.3), (4.4) or (4.5), as appropriate.

4.2. **Introducing covariates.**

We next fill the gap represented by the lack of covariates, letting the covariate column vector, $\boldsymbol{x}$, including a one attached to the intercept, be arbitrary, and let $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)'$ be considered the model's endogenous latent variables and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_N)'$ be the vector of disturbances.[6] Assume that the relationship is specified as N linear regressions,

$$
\begin{aligned} \mathsf{E}(\boldsymbol{\eta}|\boldsymbol{x}) &= \boldsymbol{A}\boldsymbol{x} & &\iff \\ \boldsymbol{\eta} &= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon}, & \boldsymbol{\epsilon} &= \boldsymbol{\eta} - \mathsf{E}(\boldsymbol{\eta}|\boldsymbol{x}), \end{aligned} \tag{4.7}
$$

where $\boldsymbol{A}$ is a coefficient matrix with N rows, including the intercept column. This equation formally defines a transformation from $(\boldsymbol{\eta}, \boldsymbol{x})$ to $(\boldsymbol{\epsilon}, \boldsymbol{x})$. The mapping from $\boldsymbol{\eta}$ to $\boldsymbol{y} = (y_1, \ldots, y_N)'$, as described in (2.1) is assumed to prevail, and we again,

[6] We will not address problems related to censored or missing values of $\boldsymbol{x}$, except that in truncated regression models, both η and $\boldsymbol{x}$ are assumed to be missing.

according to custom, introduce distributional properties for the disturbances, so that we will have to consider the corresponding partition of $\boldsymbol{\epsilon}$.

Let $f_\epsilon(\boldsymbol{\epsilon};\boldsymbol{\gamma})$ denote the density function of $\boldsymbol{\epsilon}$, which implies that $\boldsymbol{\eta}$ has density

$$f(\boldsymbol{\eta};\boldsymbol{\gamma},\boldsymbol{A},\boldsymbol{x}) \equiv f_\epsilon(\boldsymbol{\eta}-\boldsymbol{A}\boldsymbol{x};\boldsymbol{\gamma}). \tag{4.8}$$

The extension of the expression for the prototype element in the likelihood function for any observability status described by the set $\mathcal{A}_i$, (4.2), then becomes

$$\begin{aligned} F_{\mathcal{A}*i}(\boldsymbol{\eta}_{\mathcal{A}},\boldsymbol{\theta}_{\mathcal{A}*i};\boldsymbol{\gamma},\boldsymbol{A},\boldsymbol{x}) &= \int_{\boldsymbol{\eta}_{\mathcal{A}*i}\in\boldsymbol{\theta}_{\mathcal{A}*i}} f(\boldsymbol{\eta};\boldsymbol{\gamma},\boldsymbol{A},\boldsymbol{x})d\boldsymbol{\eta}_{\mathcal{A}*i} \\ &\equiv \int_{\boldsymbol{\eta}_{\mathcal{A}*i}\in\boldsymbol{\theta}_{\mathcal{A}*i}} f_\epsilon(\boldsymbol{\eta}-\boldsymbol{A}\boldsymbol{x};\gamma)d\boldsymbol{\eta}_{\mathcal{A}^*i}. \end{aligned} \tag{4.9}$$

Below follow three examples for the bivariate case ($N=2$), where we, for convenience stick to the, less compact notation in Section 3.2.

Example: Censored binormal regression:
Let $\boldsymbol{\eta}=(\eta_1,\eta_2)'$ and $\boldsymbol{A}=(\boldsymbol{a}_1',\boldsymbol{a}_2')'$ and let $f_\epsilon(\boldsymbol{\epsilon};\boldsymbol{\gamma})$ denote the density function of ϵ, which implies that $\boldsymbol{\eta}$ has density

$$\begin{gathered} f(\boldsymbol{\eta};\boldsymbol{\gamma},\boldsymbol{A},\boldsymbol{x}) \equiv f_\epsilon(\boldsymbol{\eta}-\boldsymbol{A}\boldsymbol{x};\boldsymbol{\gamma}) \iff \\ f(\eta_1,\eta_2;\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x}) \equiv f_\epsilon(\eta_1-\boldsymbol{a}_1\boldsymbol{x},\eta_2-\boldsymbol{a}_2\boldsymbol{x};\boldsymbol{\gamma}). \end{gathered}$$

Assume in addition that f_ϵ represents the binormal density function, $\boldsymbol{\gamma}$ containing the two variances and the correlation coefficient. Accordingly, extend (3.9), while using (3.8), to

$$\begin{aligned} \mathcal{F}[\boldsymbol{\theta}_i,\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x}] &= \int_{\underline{\theta}_{1i}}^{\overline{\theta}_{1i}}\int_{\underline{\theta}_{2i}}^{\overline{\theta}_{2i}} f(\eta_1,\eta_2;\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x})d\eta_1 d\eta_2 \\ &= \int_{\underline{\theta}_{1i}-\boldsymbol{a}_1\boldsymbol{x}}^{\overline{\theta}_{1i}-\boldsymbol{a}_1\boldsymbol{x}}\int_{\underline{\theta}_{2i}-\boldsymbol{a}_2\boldsymbol{x}}^{\overline{\theta}_{2i}-\boldsymbol{a}_2\boldsymbol{x}} f_\epsilon(\epsilon_1,\epsilon_2;\boldsymbol{\gamma})d\epsilon_1 d\epsilon_2, \\ F_1[\eta_2,\boldsymbol{\theta}_{1i},\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x}] &= \int_{\underline{\theta}_{1i}}^{\overline{\theta}_{1i}} f(\eta_1,\eta_2;\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x})d\eta_1 \\ &= \int_{\underline{\theta}_{1i}-\boldsymbol{a}_1\boldsymbol{x}}^{\overline{\theta}_{1i}-\boldsymbol{a}_1\boldsymbol{x}} f_\epsilon(\epsilon_1,\eta_2-\boldsymbol{a}_2\boldsymbol{x};\boldsymbol{\gamma})d\epsilon_1, \\ F_2[\eta_1,\boldsymbol{\theta}_{2i},\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x}] &= \int_{\underline{\theta}_{2i}}^{\overline{\theta}_{2i}} f(\eta_1,\eta_2;\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x})d\eta_2 \\ &= \int_{\underline{\theta}_{2i}-\boldsymbol{a}_2\boldsymbol{x}}^{\overline{\theta}_{2i}-\boldsymbol{a}_2\boldsymbol{x}} f_\epsilon(\eta_1-\boldsymbol{a}_1\boldsymbol{x},\epsilon_2;\boldsymbol{\gamma})d\epsilon_2, \qquad i=1,\dots,I. \end{aligned}$$

We can then compile the likelihood function by replacing in (3.13) [or in (3.11) or (3.14)] and in (3.12) [or in (3.15)]

$$\left.\begin{matrix} f(y_{1t},y_{2t},\boldsymbol{\gamma}) \\ F_1(y_{2t},\boldsymbol{\theta}_{1i},\boldsymbol{\gamma}) \\ F_2(y_{1t},\boldsymbol{\theta}_{2i},\boldsymbol{\gamma}) \\ \mathcal{F}(\boldsymbol{\theta}_i,\boldsymbol{\gamma}) \end{matrix}\right\} \text{ by } \left\{\begin{matrix} f(y_{1t},y_{2t};\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x}) \\ F_1[y_{2t},\boldsymbol{\theta}_{1i},\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x}] \\ F_2[y_{1t},\boldsymbol{\theta}_{2i},\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x}] \\ \mathcal{F}[\boldsymbol{\theta}_i,\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x}]. \end{matrix}\right.$$

The above functions $F_1[y_{2t},\boldsymbol{\theta}_{1i},\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x}]$ and $F_2[y_{1t},\boldsymbol{\theta}_{2i},\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x}]$ are obtained for $(\mathcal{A}_i,\mathcal{A}_i^*)=(2,1)$ and $(\mathcal{A}_i,\mathcal{A}_i^*)=(1,2)$, respectively, when expressed in the general notation (4.9).

Example: Truncated binormal regression:
This is similar to the censored binormal regression example, except that the likelihood function for a single observation, $\mathcal{L}(\boldsymbol{y}_t,R_t)$, in (3.11) is extended to

$$\mathcal{L}(\boldsymbol{y}_t,R_t) = \frac{f(y_{1t},y_{2t};\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x})}{\sum_{i:Y_i=oo}\mathcal{F}[\boldsymbol{\theta}_i,\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x}]}$$

so that the full likelihood function (3.12) is extended to

$$\mathcal{L} = \prod_{i=1}^{I}\prod_{t\in T_i}\frac{f(y_{1t},y_{2t};\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x})}{\sum_{i:Y_i=oo}\mathcal{F}[\boldsymbol{\theta}_i,\boldsymbol{\gamma},\boldsymbol{a}_1,\boldsymbol{a}_2,\boldsymbol{x}]}.$$

Example: Ordered bivariate probit:
Let $\mathcal{F}(\alpha_i;\boldsymbol{\gamma},\boldsymbol{A},\boldsymbol{x})$ be defined as in the Censored Binormal Regression example above and define, while recalling (3.8), the observable variable such that its value equals the rectangle number i:

$$y = \sum_{i=1}^{I} i\mathbf{1}\{\boldsymbol{\eta} \in \alpha_i\}.$$

The likelihood function for observation t can then be written as

$$\mathcal{L}_t(y_t, R_t) = \mathcal{F}(\boldsymbol{\theta}_i;\boldsymbol{\gamma},\boldsymbol{A},\boldsymbol{x}_t) \quad \text{if} \quad y_t = i, \qquad i = 1,\dots,I.$$

Let T_i denote the set of observations which belongs to interval i, and $|T_i|$ the number of such observations. The full sample likelihood for the $\sum_{i=1}^{I}|T_i|$ observations can then be written as

$$\mathcal{L} = \prod_{i=1}^{I}\prod_{t\in T_i}\mathcal{L}_t(y_t,R_t) = \prod_{i=1}^{I}\prod_{t\in T_i}\mathcal{F}(\boldsymbol{\theta}_i;\boldsymbol{\gamma},\boldsymbol{A},\boldsymbol{x}_t) = \prod_{i=1}^{I}\prod_{t:y_t=i}\mathcal{F}(\boldsymbol{\theta}_i;\boldsymbol{\gamma},\boldsymbol{A},\boldsymbol{x}_t).$$

Although $\mathcal{L}$ is formally a function of the rectangle counter $y_1,\dots,y_{|T|}$, the expression after the last equality sign does not involve the particular choice of 'metric' adopted: the rectangle index $y_t = \sum_{i=1}^{I} i\mathbf{1}\{\boldsymbol{\eta}_t \in \alpha_i\}$. The parameters apart, $\mathcal{L}$ only depends on the vectors of the exogenous variables $\boldsymbol{x}_1,\dots,\boldsymbol{x}_{|T|}$. We might have used for example I rectangle dummies instead.

5. Describing model types

We will now suggest short hand notations for labeling model types. Our discussion above and the basic notation from Section 2.1 will serve as a backcloth, and as we have seen, this notation can be supplemented, leading to formulations of models and likelihood functions with or without explanatory variables.

In such a label system, we will have to sacrifice some details unless it shall be unduly complicated. With this in mind, we will present two approaches with different attractions and disadvantages. The first is directly linked to the definitions in Section 2.1, having subsets as their core elements (Section 5.1), the second takes a list of variables as basis, describing the observational status for each variable (Section 5.2).

The econometric literature has traditionally given different LDV models pet names, maybe 'probit' and 'tobit' are the most prominent examples. Sometimes, such informal assignment of names may blur relationships between models with similar structure – or may lead us to overlook key differences. For instance, three 'friction models' presented by Maddala (1983 Section 6.8) can be represented as $\boldsymbol{r} = (c,c,c,m)$, $\boldsymbol{r} = (c,c,c)$, and $\boldsymbol{r} = (o,c,o)$, the second having a clear link to the standard probit model, and to ordered probit models in general. So does the first, but, as we have seen, models with missing latent variables have distinctly different likelihood functions.

Amemiya's *classification of censored regression, 'Tobit' models* (Amemiya, 1984) provides a notable contribution in providing a systematic labeling of related models. However, as remarked in the introduction, his suggested typology has a limited scope, and before presenting our own suggestions, we will point out some of its limitations. Amemiya defines five types, all in a setting where the latent variables are normally distributed; Type I is the familiar univariate censored regression 'tobit' (Tobin, 1958), Types II and III are based on the bivariate normal distribution, and

Types IV and V on the trivariate normal distribution. A virtue of Amemiya's typology is that it organizes an array of seemingly different models into a limited number of types. For instance it may be less than obvious for the uninitiated how the study "Application of a threshold regression model to household purchase of automobiles" (Dagenais, 1975) relates to Nelson's "Censored regression model with unobserved stochastic censoring thresholds" (Nelson, 1977). According to Amemiya's typology they are closely related — both are in fact Type II tobit models.

For *univariate* cases, Amemiya's classification is unsuited for distinguishing between models with different number of subsets. The standard 'tobit' has, in our notation, $I = 2$ and is characterized by $\boldsymbol{r} = (c, o)$ (or, by symmetry, $\boldsymbol{r} = (o, c)$), and models with more than two subsets can easily be handled. Examples include a "friction model" $\boldsymbol{r} = (o, c, o)$ (Maddala, 1983, Section 6.8) or a double censored distribution $\boldsymbol{r} = (c, o, c)$.

Amemiya's *bivariate* types, Type II and Type III, can be represented by $\boldsymbol{r} = (cc, co)$ and $\boldsymbol{r} = (cc, oo)$, respectively. As pet names for special cases with $I = 2$, this works well. However, if we increase the number of subsets to $I = 3$ and consider a case with $\boldsymbol{r} = (cc, co, oo)$, it is not clear whether the result should be labeled Type II, Type III, or both. Moreover, when $I = 2$ a case with $\boldsymbol{r} = (oo, co)$ is clearly a 'tobit-like model' (as at least one variable is observed in one subset and censored in another), but it is neither a Type II nor a Type III model.

Amemiya's *trivariate* models, Type IV and Type V, can be described as $\boldsymbol{r} = (cco, ooc)$ and $\boldsymbol{r} = (cco, coc)$, respectively. When $I = 2$, all of the configurations $\boldsymbol{r} = (ooo, ooc)$, (ooo, occ), (ooo, ccc), (ooc, ccc), (occ, ccc) represent 'tobit-like models', but they are unclassified in Amemiya's typology. Nor is it clear how his typology should be extended to account for more subsets. For instance, if $I = 3$ it is not clear whether $\boldsymbol{r} = (cco, ooc, coc)$ is a Type IV or a Type V model or both.

5.1. **A subset-oriented description.** The observation rules as elaborated in Section 2.1 can be used as a relatively accurate way of describing model types. As another extreme, it is possible to take a bird's eye view and only consider two key model characteristics; the dimension of the latent distribution, and the number of subsets. In some contexts it may seem excessive to list the observation rules for all subsets, or insufficiently accurate to take the bird's eye view, so we will suggest descriptions with intermediate levels of detail.

Our least detailed notation specifies only the dimension of the distribution of the latent variables and the number of subsets. In general, we suggest the notation $OCM(N, I)$, where OCM indicates that we are dealing with observed, censored and missing variables, and where N is the dimension and I is the number of subsets, as above. The standard 'tobit' is thus a $OCM(1, 2)$ model, and so are the standard probit and the truncated normal distribution. Since cases with missing variables have structure distinctly different from that in cases with no missing variables, the labels for the latter case can be modified to $OC(N, I)$. Similarly, an ordered probit with $N=1$ and an arbitrary number of alternatives, i.e. a $OCM(1,I)$ model, can simply be labeled $C(I)$.

A description with an intermediate level of detail can be obtained by counting the number of subsets with the same observation rule. All univariate models can be labeled in the format $o(\cdot)c(\cdot)m(\cdot)$, where the letters indicate observation rules and the arguments in parentheses express the number of subsets with a specific observation rule. For the standard tobit, the standard probit and the truncated normal can be denoted as $o(1)c(1)m(0)$, $o(0)c(2)m(0)$, and $o(1)c(0)m(1)$, respectively. By omitting from the labels observation rules which are not represented, these three models can be labeled simply $o(1)c(1)$, $c(2)$, and $o(1)m(1)$, respectively.

As we have seen in previous sections, the number of possible observation rules increase rapidly in the number of dimensions, N. A way to simplify the description is to ignore the order of letters in the observations rules and regard the string of letters as a product: we can then write $cc = c^2$, $oo = o^2$, $cco = c^2o$, $coc = c^2o$ and so on.

This allows us to represent Amemiya's tobit typology on four alternative levels of detail:

Tobit type I:	$\boldsymbol{r} = (c, o)$	$o(1)c(1)$	$o(1)c(1)$	$OC(1,2)$
Tobit type II:	$\boldsymbol{r} = (cc, co)$	$oc(1)cc(1)$	$oc(1)c^2(1)$	$OC(2,2)$
Tobit type III:	$\boldsymbol{r} = (cc, oo)$	$oo(1)cc(1)$	$o^2(1)c^2(1)$	$OC(2,2)$
Tobit type IV:	$\boldsymbol{r} = (cco, ooc)$	$ooc(1)cco(1)$	$o^2c(1)c^3(1)$	$OC(3,2)$
Tobit type V:	$\boldsymbol{r} = (cco, coc)$	$cco(1)coc(1)$	$oc^2(2)$	$OC(3,2)$

5.2. **A variable-oriented description.** In some cases it can be more convenient to choose a variable-oriented description of a model rather than the subset-oriented description introduced above. Since the subset-oriented and the variable-oriented descriptions are closely linked for the model class considered in Sections 4.1 and 4.2, which is characterized by the particular way of constructing the subsets, we will introduce the variable-oriented description with an example from this model class:

Example: Expenditure functions for $N=6$ commodities:
The latent variables are determined by

$$\eta_{nt} = \boldsymbol{x}_{nt}\boldsymbol{\beta}_n + \epsilon_{nt}, \qquad n=1,\ldots,N,$$

which exemplifies (4.7), where $(\epsilon_{1t},\ldots,\epsilon_{Nt})$ has density function with a 'full', t-invariant covariance matrix exemplified by (4.8). The pattern of observability of the y_{nt} variables is:

y_{1t} : observed over entire range,
y_{2t} : one threshold; censored twice both below and above threshold,
y_{3t} : one threshold; censored below, observed above threshold,
y_{4t} : one threshold; missing below, censored above threshold,
y_{5t} : one threshold; missing below, observed above threshold,
y_{6t} : missing over entire range.

Here y_{1t} may be a commodity used (and purchased) frequently by most households and fairly precisely recorded by the 'data producer'; $y_{2t},\ldots,y_{5t}$ may be commodities not universally used or purchased more or less infrequently by several households, say alcohols, tobacco, medical and holiday services, and consumer durables, while latent consumption of good 6, η_{6t}, according to the theory being determined jointly with $\eta_{1t},\ldots,\eta_{5t}$, and correlated with them via the disturbances $(\epsilon_{1t},\ldots,\epsilon_{6t})$, but because of limitations in the recording procedures, occurs as a completely missing variable in the data set.

To obtain the number of subsets, I, in this specific example we first note the number of intervals for each variable. Here y_{1t} and y_{6t} have one interval (the whole real line), the others have two intervals (below and above the threshold value). It follows that the number of subsets is the product of the number of intervals for each variable, $I = 1 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 1 = 16$, so following the subset-oriented labeling, this is an $OCM(6, 16)$ model.

Descriptions on intermediate levels of detail can be constructed by the variables separately. For each variable we define a sequence of letters (o, c, m), to describe its observations status. The length of the sequence is defined by the number of intervals, and the order of the letters corresponds to the way the intervals are distributed on the real line. Thus, in the above example, we describe the variable y_{1t} by o, y_{2t} by cc, y_{3t} by co, and so on. A description of the full model can be obtained by listing the description for each variable, say

$$o \cdot cc \cdot co \cdot mc \cdot mo \cdot m,$$

where the delimiter '$\cdot$' is chosen partly as a reminder that the number of subsets equals the product of the number of intervals for the intervals, and partly to avoid confusion with the alternative label system in Section 5.1. This description can be simplified by merging variables having the same observation status, by letting for instance $oc \cdot oc = (oc)^2$. Less detailed descriptions can be obtained by disregarding the order of the letters for a given variable, so that $oc = co$, and by using superscripts to simplify sequences of equal letters, say $cc = c^2$ or $commom = co^2m^3$. In this manner a consumption system with $N = 10$ commodities, all censored due to a non-negativity constraint, can be described alternatively as

$$\begin{aligned} &co \cdot co \cdot co \cdot co \cdot co \cdot co \cdot co \cdot co \cdot co \cdot co, \\ &(co)^{10}, \\ &OC(10, 1024). \end{aligned}$$

Amemiya (1984, p. 31, Table 2) also describes models based on the observational status of individual variables. However, since Amemiya's Types II through V do no belong to the model class in Sections 4.1 and 4.2 the link between our subset-oriented and the variable-oriented descriptions is lost: Amemiya's Type V tobit could still be given the variable-oriented description "$cc \cdot oc \cdot oc$", but it would be fallacious to infer from this that $I = 2 \cdot 2 \cdot 2 = 8$.

6. Concluding remarks

The typology suggested in this paper applies to a frequently used class of econometric models. Albeit it has been recognized for decades that members of this class have common features, previous attempts to describe the class have been implicit and deliberately incomplete. Our classification system is complete, in the sense of being applicable to models of any dimension of the latent variables, containing any number of subsets, and any combinations of observation rules. And depending on the amount of details one want the label system to provide, the names may be more or less concise. We have suggested two ways of describing the models. The first, and most generally applicable is the one called the subset-oriented description, while the second, the variable-oriented description, can sometimes be simpler to apply. However, for the N-variate case as presented in Section 4.1 the two descriptions coincide.

We believe that our proposed classification system, in addition to facilitating communication between masters of the LDV model field, may also benefit students and newcomers. In contemporary econometric textbooks terms like 'censoring', 'selection', 'incomplete observation', 'defective data' and 'incidental truncation' occurs frequently. And although the meaning usually is sufficiently clear within the context of a single book, it is not always obvious how these terms should be generalized to other models. With our classification system at hand the whole class of model can be presented with a few simple examples and straightforward induction.

References

Agresti, A. (1990): *Categorical Data Analysis*. New York: Wiley.

Amemiya, T. (1984): Tobit models: A survey. *Journal of Econometrics* 24, 3–61.

Bliss, C.I. (1937): The calculation of the time-mortality curve. Appendix by W.L. Stevens. *Annals of Applied Biology* 24, 815–852.

Cohen, A.C. (1950): Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. *The Annals of Mathematical Statistics* 21, 557–569.

Dagenais, M.G. (1975): Application of a threshold regression model to household purchases of automobiles. *Review of Economics and Statistics* 57, 275–285.

Fisher, R.A. (1946/1934): Tables VII–VIII. Exponential and sine and cosine integrals. In *Mathematical tables volume I*, pp. viii–ix. Cambridge: The British Association for the Advancement of Science.

Hald, A. (1949): Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Skandinavisk Aktuarietidsskrift* 32, 119–132.

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.

Hein, J.L. (2002): *Discrete Structures, Logic, and Computability, Second Edition*. Sudbury, MA: Jones and Bartlett Publishers.

Hilbe, J.M. (2009): *Logistic Regression Models*. New York: Chapman & Hall/CRC.

Little, R.D.A. and Rubin, D.B. (1987): *Statistical Analysis with Missing Data*. New York: Wiley.

Maddala, G.S. (1983): *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Marschak, J. (1953): Economic Measurements for Policy and Prediction. In Hood, W.C. and Koopmans, T.C. (eds.): *Studies in Econometric Methods*. New York: Wiley.

Nelson, F.D. (1977): Censored regression models with unobserved, stochastic censoring thresholds. *Journal of Econometrics* 6, 309–327.

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