Nonlinear Panel Data Methods for Multivariate Fractional Outcomes

Santiago Montoya-Blandón*

University of Glasgow

Abstract

This paper expands the applied researcher's toolkit for dealing with nonlinear panel data models with unobserved heterogeneity using multivariate fractional outcomes. It presents a wide range of methods that include maximum likelihood estimation for identifying the structural parameters of models specified by a conditional mean, a simple and scalable GMM approach using the probit link that allows to identify average partial effects, and a Bayesian estimator from a latent dependent variable specification to account for censoring or structural zeros in the data. I then show how all these methods can be modified to handle continuous endogenous covariates using a control function approach. A range of simulation exercises showcase the comparative advantages of each method and how they might be used to approach different situations that arise in applied microeconomics.

Keywords: Multivariate fractional outcomes; Panel data; Nonlinear models; Bayesian analysis.

JEL classification: C35, C33, C51

1 Introduction

In many applied microeconomic settings one usually encounters outcome variables that are multivariate fractional. These can arise naturally from a demand estimation setting, where they represent a household's expenditure shares across several goods (Woodland, 1979; Barnett and Serletis, 2008); in micro-finance, to represent the portfolio shares held by individuals or firms across risky assets (Mullahy, 2015); in industrial organization, to represent the market shares of firms in a given industry (Morais et al., 2018); in health production, to represent time of use across different activities (Mullahy and Robert, 2010); in analyzing voting patterns for several candidates (Katz and King, 1999); among many others (Aitchison, 2003). As the name suggest, the unifying characteristic of multivariate fractional data (or compositional data in the statistics literature) is that they are a set of outcomes, each between zero and one, that satisfy a unit-sum restriction.

While there have been many developments in creating modeling strategies for this type of data in a cross-sectional context or for univariate fractions in a panel data setting (Papke and Wooldridge, 1996, 2008; Murteira and Ramalho, 2016), there are currently no comprehensive and flexible way of modeling multivariate fractions in a panel data setting. That is, a strategy that simultaneously takes into account the inherent nonlinearity in the partial effects from covariates, unobserved heterogeneity that is potentially correlated to these covariates, and that imposes the unit-sum restriction present across the multivariate outcomes. Additionally, we would expect that such a framework would allow to control for endogeneity issues that are not created by unobserved heterogeneity and should also allow for structural zeros in the data.[1]

*Adam Smith Business School, University of Glasgow, Room 664, University Avenue, Glasgow, United Kingdom, G12 8QQ. E-mail: santiago.montoya-blandon@glasgow.ac.uk. Website: smontoyablandon.com.

[1]For example, in the demand estimation setting it is standard to have endogeneity arising from prices or behavioural factors and to observe households not consuming particular categories of goods.

The main contribution of the paper is to expand the available toolkit for modeling multivariate fractional outcomes using panel data, where the provided methods address most or all the previously mentioned requirements that arise naturally in applied microeconomic settings. To this end, I extend currently available approaches for cross-sectional multivariate fractional outcomes to a panel data setting and bring panel data methods that operate on univariate fractions to the multivariate case. This is done in a way that emphasizes robustness and flexibility, while maintaining the advantages of each framework. Recognizing that different applications are conceived with different objectives in mind, the paper introduces a wide range of methods that are each suitable for various settings encountered in applied research.

The first method is maximum likelihood estimation that allows for identification of the parameters in a conditional mean model (Hartzel et al., 2001). This method will be particularly useful when an application requires consistent estimation of the parameters, not just the signs or average partial effects (APEs). This is the case in many applications, particularly those where the conditional mean specification comes from a structural model, such that the parameters have a direct economic interpretation.[2] Of course, given consistent estimation of the parameters, most other quantities of interest can be consistently estimated. This method also has the potential of being efficient, in contrast to the other techniques introduced in the paper. While many existing likelihood-based approaches allow the specification of a distribution on the multivariate fractional outcomes, they can be restrictive or not generalize well to allow for unobserved heterogeneity and endogenous covariates. For example, this occurs for transformation methods that take the multivariate fractions from the simplex to an unbounded space before imposing a distributional assumption on the transformed shares. If we wanted to impose a conditional mean model defined directly on the share components before the transformation, we would require strong and implausible independence assumptions to recover the parameters of that conditional mean (Papke and Wooldridge, 1996). This is the case for most non-linear transformations such as the additive log-ratio (Aitchison and Shen, 1980), centered log-ratio (Aitchison, 1983), isometric log-ratio (Egozcue et al., 2003), or α (Tsagris et al., 2011) transformations. One could also choose a distribution directly on the share components that respects a given conditional mean model (Lijoi et al., 2005; Hijazi and Jernigan, 2009; Scealy and Welsh, 2011), but this approach relies heavily on correct specification of the distribution and will not generally be robust to misspecification. The maximum likelihood methods considered in this paper will allow for direct specification of a conditional mean and at least some degree of robustness to distributional misspecification, if not full robustness. The treatment and computation of this method draws on the statistical literature on generalized (non)linear mixed models for multivariate responses (for a review, see for example Davidian and Giltinan, 1995).

The second method extends Papke and Wooldridge (2008) to the multivariate fractional setting. By specifying a probit link for each of the shares, we can analytically integrate out unobserved heterogeneity from the conditional mean specification at the cost of losing identification of the conditional mean parameters. However, the average partial effects remain identified as long as we assume the probit link to be correctly specified. The estimator proposed in the paper uses the generalized method of moments (GMM) to fit the system of non-linear regressions and consistently recover the average partial effects. If the functional form of the conditional mean for each share in terms of the covariates is not captured by the probit link, this approach still provides the best mean-squared error approximation to the APEs that is afforded by the probit link. Simulations exercises in Section 3 show that the probit approximation is quite good and tends to yield appropriate results for the APEs even when there is some misspecification in the link function. Additionally, this method is the simplest to estimate and extend, it is not impeded in any way by the presence of zeros in the multivariate fractions, inference

[2]This occurs in demand estimation, where structural demand systems usually produce an estimating equation in terms of a conditional mean (Deaton and Muellbauer, 1980; Banks et al., 1997; Barnett and Serletis, 2008).

for it can be made fully robust to the potential misspecification of the probit link, and it can be easily scaled to handle a large amount of shares without much additional computational burden.[3]

I then discuss a latent dependent variable formulation that accounts for large incidence of censoring or corner outcomes, given by structural zeros in the multivariate fractions. Using the simple transformation in Wales and Woodland (1983), I extend the Bayesian data-augmentation approach of Kasteridis et al. (2011) to account for the availability of panel data and correlated random effects. Accounting for unobserved heterogeneity in this method is then also a multivariate generalization to Loudermilk (2007).[4] The simplicity of this resulting approach is in line with previous literature where the Bayesian paradigm tends to be preferred to frequentist simulation-based approaches given their simplicity in dealing with the latent variables (McCulloch et al., 2000). Still, simulation methods such as the methods of simulated moments (McFadden, 1989) or simulated scores (Hajivassiliou and McFadden, 1998) would remain valid given this setting and their exploration in this context could be a potential avenue for further research. This approach also directly accounts for the presence of zeros in the multivariate fractions. Other methods that allow for zeros begin with transforming the zeros to a small but positive quantity that can be incorporated into standard techniques. In these cases, zeros are usually assumed to be caused by detection errors (not structural zeros) and thus imputation methods are used to transform the values in some optimal way to try and minimize the ad hoc nature of this operation (Fry et al., 2000; Martín-Fernández et al., 2003). While some transformations, likelihoods or regression-based approaches can also deal directly with structural zeros, their application is still subject to similar caveats as those mentioned before (Stewart and Field, 2011; Tsagris and Stewart, 2018).

The remaining of the paper proceeds as follows. Section 2 reviews the general assumptions and theory that supports the estimation methods that are then introduced. Special emphasis is made in implementation of the methods using fully robust inference. Section 3 presents several Monte Carlo exercises that showcase the comparative advantages of each of the methods, their possible weaknesses and robustness, as well as specific cases where they will be most useful. Finally, Section 4 presents the concluding remarks.

2 Methodology

I begin by stating the general assumptions that hold for all the methods considered in the paper. Let $\boldsymbol{Y}$ be a multivariate fractional outcome of d shares. The unifying characteristic of $\boldsymbol{Y}$ across different applications is that it is supported on the d-dimensional simplex

$$\mathcal{S}^d = \left\{ (y_1, \ldots, y_d) \in \mathbb{R}^d : 0 \leq y_j \leq 1, j = 1, \ldots, d; \sum_{j=1}^{d} y_j = 1 \right\}. \tag{1}$$

That is, each component of $\boldsymbol{Y}$ is a fraction, and taken together, these components add up to one. For each share Y_j, I also assume that we have a K_j-dimensional vector of covariates denoted by $\boldsymbol{X}_j$. Similarly, as is customary in panel data models, I allow for the presence of unobserved heterogeneity that is potentially correlated to the covariates, which is denoted by $\boldsymbol{C}$. The following assumption

[3] In the context of demand estimation, this means that using this method it is possible to find the average partial effects of covariates on expenditure shares for systems of hundreds or thousands of products, such as those encountered in online shopping applications.

[4] Loudermilk (2007) allows for dynamics by including the lagged outcome as a regressor. While the previous approaches require strict exogeneity, the Bayesian method only requires sequential exogeneity for its validity. Thus, while the previous methods cannot handle a lagged dependent variable as covariate, this is no issue for this Bayesian approach.

summarizes the type of panel data structures that are within the scope of this paper and which arise frequently in applied microeconomics.

Assumption 1 (Panel data)**.**

1. Let $(\boldsymbol{Y}', \boldsymbol{X}', \boldsymbol{C})'$ be a $(2d + K)$-dimensional random-vector with true distribution H, where $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$ takes values on $\mathcal{S}^d$, $\boldsymbol{X} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_j')'$ has support $\mathcal{X} \subset \mathbb{R}^{K_1+\cdots+K_D}$ with $K = K_1 + \cdots + K_d$, and $\boldsymbol{C} = (C_1, \ldots, C_d)'$.

2. There is access to a random sample of size n from H in the cross section, given by $\{\boldsymbol{Y}_i', \boldsymbol{X}_i'\}_{i=1}^n$, where $\boldsymbol{Y}_i \in \times_{t=1}^{T_i} \mathcal{S}^d$. That is, for each random draw i there are T_i time periods, and within each i and time period t, the outcomes are multivariate fractional.

The first part of Assumption 1 introduces unobserved heterogeneity as part of the true distribution that defines the population of interest. Emphasizing this true distribution will also allow us to discuss inference that takes into account possible misspecification in the maximum likelihood method that is presented shortly. From the second part, note that the paper is sufficiently general as to allow for unbalanced panels, but it does assume that the reason for the unbalance is completely at random. In this sense, the methods introduced in the paper will not remain valid under possible issues of attrition or other sample selection rules that are dependent on the covariates. Of course, since $\boldsymbol{C}$ is unobserved by definition, it does not show up in the information available to the econometrician for estimation and inference. Additionally, at this point I note that all the asymptotic results in the paper rely on short panels; i.e., where T_i is taken as fixed while the cross section n goes to infinity. The dimensionality of the simplex given by d is not restricted and we will introduce methods that allow for d to be large, which might occur, for example, in a demand estimation problem with many goods in consideration. With this in mind, I now consider the following estimation procedures that will contain some more specialized assumptions conditional on the inferencial goal of each method.

2.1 Maximum Likelihood Estimator

For this and the next subsection, we need to assume a conditional mean model that relates the multivariate fractional outcome $\boldsymbol{Y}$ to the covariates $\boldsymbol{X}$ and the unobserved heterogeneity $\boldsymbol{C}$. One possibility would be to assume for each $i = 1, \ldots, n$, $t = 1, \ldots, T_i$, and $j = 1, \ldots, d$,

$$\mathrm{E}[Y_{itj} | \boldsymbol{X}_{itj} = \boldsymbol{x}_{itj}, C_{ij} = c_{ij}] = m_j(\boldsymbol{x}_{itj}'\boldsymbol{\beta}_{0,j} + c_{ij}) \,,$$

for some $\boldsymbol{\beta}_{0,j} \in \mathcal{B}_j \subset \mathbb{R}^{K_j}$, where c_{ij} represents time-invariant unobserved heterogeneity for each individual i in outcome equation j, and the functions $m_j(\cdot)$ would satisfy $0 < m_j(z) < 1$ and $\sum_{j=1}^d m_j(z) = 1$ for all $z \in \mathbb{R}$, $j = 1, \ldots, d$. However, the unit-sum restriction on the link functions and the outcome shares creates an identification problem that prevents us from proceeding with this approach. As noted by Montoya-Blandón (2021), the fact that the outcome variables are supported on $\mathcal{S}^d$ prevents the recovery of one of the parameter vectors $\boldsymbol{\beta}_{0,j}, j = 1, \ldots, d$ as all information about one of the outcomes can be obtained from the distribution of the others. To address this issue, we will instead work with the $D \equiv d - 1$ dimensional system by setting a base category, assumed to be d hereafter. This conditional mean would also miss an interesting possibility that I use as the basis for the two special cases of a maximum likelihood estimator in this setting. Thus, I instead introduce the following assumption.

Assumption 2 (Conditional mean)**.** For each $i = 1, \ldots, n$, $t = 1, \ldots, T_i$, and $j = 1, \ldots, d$,

$$\mathrm{E}[Y_{itj} | \boldsymbol{X}_{it}, \boldsymbol{c}_i] = m_j(\boldsymbol{X}_{it}\boldsymbol{\beta}_0 + \boldsymbol{c}_i) \,, \tag{2}$$

for some $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_D')' \in \mathcal{B} \subset \mathbb{R}^K$, where $K = \sum_{j=1}^D K_j$, $\boldsymbol{c}_i = (c_{i1}, \ldots, c_{iD})'$, and the link functions are defined for all $j = 1, \ldots, d$ as $m_j : \mathbb{R}^D \to \mathbb{R}$ to satisfy $0 < m_j(\boldsymbol{z}) < 1$ and $\sum_{j=1}^d m_j(\boldsymbol{z}) = 1$ for all $\boldsymbol{z} \in \mathbb{R}^D$. Finally, $\boldsymbol{X}_{it}$ is a $D \times K$ matrix defined as

$$\boldsymbol{X}_{it} = \begin{bmatrix} \boldsymbol{x}_{it1}' & \cdots & \boldsymbol{0}_{1\times K_D} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0}_{1\times K_1} & \cdots & \boldsymbol{x}_{itD}' \end{bmatrix}$$

This assumption introduces a few key ideas. First, as is usual in panel data models, dealing with $\boldsymbol{c}_i$ will be one of the main challenges of obtaining reliable estimators (Wooldridge, 2010, section 10). Second, we have a family of link functions $m_j(\cdot)$ where each outcome can potentially depend on the covariates and unobserved heterogeneity of all other outcomes, allowing for very rich dependence between shares. Third, note that it is assumed there is a true $\boldsymbol{\beta}_0$ such that the conditional mean assumption holds for all outcomes. Finally, note that (2) is general enough to allow for outcome-specific intercepts, time effects and covariates, while allowing for the same covariates to enter different share equations and having possibly time-invariant covariates. It is also assumed that $\boldsymbol{x}_{itj}$ contains a 1 at the beginning of the vector for each $j = 1, \ldots, D$.

Throughout the paper, we will need stacked versions of (2) across outcomes and time. These are given by

$$\mathrm{E}[\boldsymbol{Y}_{it}|\boldsymbol{X}_{it}, \boldsymbol{c}_i] = \boldsymbol{m}(\boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{c}_i) \tag{3}$$

and

$$\mathrm{E}[\boldsymbol{Y}_i|\boldsymbol{X}_i, \boldsymbol{c}_i] = \boldsymbol{m}_{T_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{c}_i)\,, \tag{4}$$

where $\boldsymbol{Y}_{it} = (Y_{it1}, \ldots, Y_{itD})'$ and $\boldsymbol{m}(\boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{c}_i) = (m_1(\boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{c}_i), \ldots, m_D(\boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{c}_i))'$ are D-dimensional vectors, $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1}', \ldots, \boldsymbol{Y}_{iT_i}')'$ and $\boldsymbol{m}_{T_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{c}_i) = (\boldsymbol{m}(\boldsymbol{X}_{i1}\boldsymbol{\beta} + \boldsymbol{c}_i)', \ldots, \boldsymbol{m}(\boldsymbol{X}_{iT_i}\boldsymbol{\beta} + \boldsymbol{c}_i)')'$ are DT_i-dimensional vectors, and $\boldsymbol{X}_i = \begin{bmatrix} \boldsymbol{X}_{i1}' & \cdots & \boldsymbol{X}_{iT_i}' \end{bmatrix}'$ is a $DT_i \times K$ matrix.

As noted by Papke and Wooldridge (2008), assumptions 1 and 2 on their own are not enough to carry out estimation of the conditional mean parameters . To this end, I make two additional assumptions.

Assumption 3 (Strict exogeneity)**.** For all $i = 1, \ldots, n$, and $j = 1, \ldots, d$,

$$\mathrm{E}[Y_{itj}|\boldsymbol{X}_i, \boldsymbol{c}_i] \equiv \mathrm{E}[Y_{itj}|\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{iT_i}, \boldsymbol{c}_i] = \mathrm{E}[Y_{itj}|\boldsymbol{X}_{it}, \boldsymbol{c}_i]\,.$$

Assumption 4 (Mundlak device)**.** For all $i = 1, \ldots, n$,

$$\boldsymbol{c}_i|\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{iT_i} \sim \mathcal{N}(\bar{\boldsymbol{X}}_i\boldsymbol{\xi}, \boldsymbol{\Gamma})\,, \tag{5}$$

where $\bar{\boldsymbol{X}}_i = (1/T_i)\sum_{t=1}^{T_i} \boldsymbol{X}_{it}$ are the time averages for the time-varying covariates, $\boldsymbol{\xi}$ is a K-dimensional coefficient vector and $\boldsymbol{\Gamma}$ is a $D \times D$ covariance matrix.

Assumption 3 is standard and simply states that, conditional on unobserved heterogeneity, the covariates are uncorrelated to time-varying unobservables. It also rules out the use of lagged dependent variables as covariates or explanatory variables that correlate to paste values of the outcome variables (Papke and Wooldridge, 2008). Assumption 4 is a correlated random effect (CRE) assumption that uses Mundlak's (1978) device for specifying the relationship between covariates and unobserved heterogeneity. Note that under a pure random effects assumption, $\boldsymbol{\xi} = \boldsymbol{0}$ and there would be no need to

worry about correlation with unobserved heterogeneity. Of course, a more flexible model such as that by Chamberlain (1980) could be allowed, at the expense of slightly more complex models. The use of (5) is made for convenience and to allow for particularly simple estimation methods for $\boldsymbol{\beta}$. Other non or semiparametric alternatives that assume less structure on the distribution of $\boldsymbol{c}_i$ conditional on $\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{iT_i}$ are also available, again at the expense of more intensive computations (Hartzel et al., 2001). As the maximum likelihood method to be introduced shortly can already be computationally demanding, this paper maintains (5) for simplicity. Finally, the paper does not consider fixed effects transformations to eliminate $\boldsymbol{c}_i$, as these require correct specification (of both H and $\boldsymbol{m}$) and are only available for a handful of distributions with special forms and sufficient statistics (see, e.g., Magnac, 2004).

Note that, given (5), we can write $\boldsymbol{c}_i = \bar{\boldsymbol{X}}_i\boldsymbol{\xi} + \boldsymbol{b}_i$, where $\boldsymbol{b}_i | \boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{iT_i} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma})$. Replacing this into (3) and using Assumption 3 yields

$$\mathrm{E}[\boldsymbol{Y}_{it} | \boldsymbol{X}_i, \boldsymbol{c}_i] = \boldsymbol{m}(\boldsymbol{X}_{it}\boldsymbol{\beta} + \bar{\boldsymbol{X}}_i\boldsymbol{\xi} + \boldsymbol{b}_i)\,.$$

Writing $\tilde{\boldsymbol{X}}_{it} = \begin{bmatrix} \boldsymbol{X}_{it} & \bar{\boldsymbol{X}}_i \end{bmatrix}$ and $\boldsymbol{\alpha} = (\boldsymbol{\beta}', \boldsymbol{\xi}')'$, we can then find

$$\mathrm{E}[\boldsymbol{Y}_{it} | \boldsymbol{X}_i, \boldsymbol{c}_i] = \boldsymbol{m}(\tilde{\boldsymbol{X}}_{it}\boldsymbol{\alpha} + \boldsymbol{b}_i)\,, \tag{6}$$

with $\boldsymbol{b}_i$ independent of $\tilde{\boldsymbol{X}}_{it}$. This is of the same form as (3) but with $\boldsymbol{b}_i$ representing unobserved heterogeneity that is uncorrelated from the covariates. For notational simplicity, the remaining of the paper assumes that (3) (and thus 4) represents a random effects specification, so that $\boldsymbol{c}_i$ can be taken as independent from covariates $\boldsymbol{X}_{it}$. Keep in mind that this will only be true after the transformation given by (6) if the original covariates are thought to be correlated to unobserved heterogeneity, which is usually the case in most applications. A subtle point is that for the computation of average partial effects, or any derivation that follows from the original conditional mean model in (2), $\bar{\boldsymbol{X}}_i$ needs to be integrated out for each $t = 1, \ldots, T_i$ (Papke and Wooldridge, 2008).

Armed with Assumptions 1 through 4, I can now present the general maximum likelihood estimator for multivariate fractional outcomes and two interesting special cases. Let $F(\cdot; \boldsymbol{\beta})$ denote a D-dimensional distribution for $\boldsymbol{Y}_{it} | \boldsymbol{X}_{it}, \boldsymbol{c}_i$ that satisfies (3). As the random effects $\boldsymbol{c}_i$ (or $\boldsymbol{b}_i$ after the transformation in 6) are unobserved, we need to integrate over them in the definition of the likelihood. Assuming conditional independence across t, we can define the log-likelihood contribution for each i in this problem as

$$\ell_i^{(\mathrm{ind})}(\boldsymbol{\beta}, \boldsymbol{\Gamma}) = \log \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T_i} F(\boldsymbol{Y}_{it} | \boldsymbol{X}_{it}, \boldsymbol{c}_i; \boldsymbol{\beta}) \right] \phi_D(\boldsymbol{c}_i; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma})\, \mathrm{d}\boldsymbol{c}_i\,, \tag{7}$$

where $\phi_D(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density of a D-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. A second approach that does not impose conditional independence across time, is given by the pooled likelihood approach

$$\ell_i^{(\mathrm{pool})}(\boldsymbol{\beta}, \boldsymbol{\Gamma}) = \sum_{t=1}^{T_i} \log \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F(\boldsymbol{Y}_{it} | \boldsymbol{X}_{it}, \boldsymbol{c}_i; \boldsymbol{\beta}) \phi_D(\boldsymbol{c}_i; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma})\, \mathrm{d}\boldsymbol{c}_i\,. \tag{8}$$

Writing $\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathrm{vech}(\boldsymbol{\Gamma})')'$, where $\mathrm{vech}(\cdot)$ is the half-vectorization operator that selects the lower triangular portion of a square matrix, we have that a general maximum likelihood estimator based on either (7) or (8) is given by

$$\widehat{\boldsymbol{\theta}}_l \equiv \arg\max_{\boldsymbol{\theta}} \frac{1}{n} \sum_{i=1}^{n} \ell_i^{(l)}(\boldsymbol{\theta}), l \in \{\mathrm{ind}, \mathrm{pool}\}\,. \tag{9}$$

For $l \in \{\text{ind}, \text{pool}\}$, if we do not assume correct specification of F, general quasi-likelihood theory, such as that in White (1982), yields consistency of $\widehat{\boldsymbol{\theta}}_l$ to the minimizer of the Kullback-Leibler divergence between F and H, denoted as $\boldsymbol{\theta}_l^*$. Furthermore, if F is chosen to be a member of the linear exponential family, as long as the link function $\boldsymbol{m}$ is correctly specified, then the $\boldsymbol{\beta}^*$ component of $\boldsymbol{\theta}_l^*$ will equal the $\boldsymbol{\beta}_0$ specified in Assumption 2 (Gourieroux et al., 1984). This is the basis for one of the special cases introduced as Estimator 1. The second special case, Estimator 2, specifies F using a copula approach. Following the results in Montoya-Blandón (2021), we observe that as long as the marginals in F are correctly specified (which again requires correct specification of the link), even if the dependence structure is not, then $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$ also holds. In both of these cases, we can thus guarantee consistent estimation of the underlying conditional mean parameters $\boldsymbol{\beta}_0$.

Once consistency is established, the results in the previously mentioned literature can be used to obtain asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\theta}}_l - \boldsymbol{\theta}_l^*)$ with asymptotic variance given by

$$\text{Asy. Var}(\sqrt{n}(\widehat{\boldsymbol{\theta}}_l - \boldsymbol{\theta}_l^*)) = A_l^{-1} B_l A_l^{-1}\,, \tag{10}$$

where $A_l = \mathrm{E}_H[\partial^2 \ell_i^{(l)}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}']$ is the Hessian matrix of the log-likelihood contributions, $B_l = \mathrm{E}_H[\partial \ell_i^{(l)}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} \cdot \partial \ell_i^{(l)}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}']$ is the outer product of the scores, and the notation E_H emphasizes that the expectation is taken with respect to the true distribution. Inference that is fully robust to possible distributional misspecification (and to autocorrelation in the scores in the case of the pooled log-likelihood approach) follows from using

$$\widehat{A}_l = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \ell_i^{(l)}(\widehat{\boldsymbol{\theta}}_l)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \quad \text{and} \quad \widehat{B}_l = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \ell_i^{(l)}(\widehat{\boldsymbol{\theta}}_l)}{\partial\boldsymbol{\theta}} \cdot \frac{\partial \ell_i^{(l)}(\widehat{\boldsymbol{\theta}}_l)}{\partial\boldsymbol{\theta}'}\,, \tag{11}$$

to estimate the asymptotic variance in (10). The way this model is specified is similar to nonlinear mixed models (or generalized mixed models if F is assumed to be a distribution from the linear exponential family) used heavily in the statistics literature (Davidian and Giltinan, 1995). Pinheiro and Bates (1995) is a standard reference for computation of the integrals in (7) or (8). For adaptive (Liu and Pierce, 1994) or nonadaptive (Jäckel, 2005) quadrature, Appendix A presents some general formulas to compute these integrals. Whereas the literature tends to favor Laplace approximations to these integrals, quadrature or Monte Carlo methods should be used in this case, as we will usually want to assume a distribution that is not necessarily correctly specified. A Laplace approximation to an already misspecified distribution would likely introduce larger bias into the estimation process. Quadrature methods will also be reliable only for a small dimension D as the number of evaluations grows exponentially with D. For larger dimensions, one could use an expectation-maximization (EM) algorithm as outlined in Hartzel et al. (2001). When deciding between each method it is also important to keep in mind that the pooled approach requires more integral evaluations; (7) requires n integrals to be computed, while (8) requires $\sum_{i=1}^{n} T_i$ of them (or nT for a balanced panel).

Based on the previous formulas, the paper proposes two special cases that will be of particular interest in applications. Both start from a multinomial logit conditional mean as it satisfies the unit-sum restriction given in Assumption 2. That is, these estimators take $\boldsymbol{m}(\cdot)$ as

$$\boldsymbol{m}(\boldsymbol{X}_{it}'\boldsymbol{\beta} + \boldsymbol{c}_i) = \begin{cases} \dfrac{\exp\left(\boldsymbol{x}_{itj}'\boldsymbol{\beta}_j + c_{ij}\right)}{1 + \sum_{p=1}^{D} \exp\left(\boldsymbol{x}_{itp}'\boldsymbol{\beta}_p + c_{ip}\right)} & \text{for } j = 1, \ldots, D\,, \\ \dfrac{1}{1 + \sum_{p=1}^{D} \exp\left(\boldsymbol{x}_{itp}'\boldsymbol{\beta}_p + c_{ip}\right)} & \text{for } j = d\,. \end{cases} \tag{12}$$

Estimator 1 (Multinomial Logit QMLE)**.**

1. Use

$$F(\boldsymbol{Y}_{it}|\boldsymbol{X}_{it},\boldsymbol{c}_i;\boldsymbol{\beta}) = \prod_{j=1}^{d} m_{ijt}^{y_{ijt}}\,,$$

in either (7) or (8) with $m_{itj} \equiv m_j(\boldsymbol{X}_{it}'\boldsymbol{\beta} + \boldsymbol{c}_i)$ according to the multinomial logit link.

2. Estimate $\widehat{\boldsymbol{\theta}}$ as in (9) computing the integrals as in Appendix A.

3. As the multinomial likelihood is inherently misspecified, use the fully robust estimators given in (11).

Appendix B contains a formula for the score $\partial \ell_i^{(l)}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ that can be used to motivate a quasi-Newton algorithm as in Hartzel et al. (2001) and also to obtain the fully robust variance estimator. As in Papke and Wooldridge (1996), this estimator, while being inherently misspecified, should achieve some optimality properties in the class of linear exponential families for this problem. Another possible approach would be to specify a population-averaged estimator that uses general estimating equations (GEE) to gain efficiency (Liang and Zeger, 1986). These would start by specifying $\mathrm{E}[Y_{itj}|\boldsymbol{X}_i t]$ directly as in (2), perhaps using a multinomial logit link. Note that no model would actually correspond to this link after integration of the random effects. Additionally, given that the multinomial distribution is inherently misspecified, it might not be worthwhile to attempt to gain more efficiency by correctly specifying other features of the distribution. Thus, I recommend the use of the fully robust approach as noted Estimator 1.

If efficiency is a concern, there is another route. As shown in Montoya-Blandón (2021), copulas can be used to model multivariate fractional outcomes in a way that achieves flexibility in the dependence patterns between shares, while retaining some robustness to distributional misspecification. Furthermore, if the copula and marginals are correctly specified, this leads to an efficient maximum likelihood approach. This is summarized in the following procedure.

Estimator 2 (Multinomial Logit Copula)**.**

1. Choose marginals $G_j(\cdot;\boldsymbol{\beta},\phi_j), j = 1,\ldots,D$ that satisfy (12), such as beta distributions, and copula $G(\cdot;\boldsymbol{\psi})$, for example a Gaussian copula. Then, use

$$\begin{aligned} F(\boldsymbol{Y}_{it}|\boldsymbol{X}_{it},\boldsymbol{c}_i;\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\psi}) = g(G_1(y_{it1}|\boldsymbol{X}_{it};\boldsymbol{\beta},\phi_1),\ldots,G_D(y_{itD}|\boldsymbol{X}_{it};\boldsymbol{\beta},\phi_D);\boldsymbol{\psi}) \\ \times \prod_{j=1}^{D} g_j(y_{itj}|\boldsymbol{X}_{it};\boldsymbol{\beta},\phi_j)\,, \end{aligned}$$

in either (7) or (8). The copula approach adds some additional precision parameters for the marginals and dependence parameters for the copula (which can be misspecified). Compute the integrals as in Appendix A.

2. Estimate $(\boldsymbol{\theta}',\boldsymbol{\phi}',\boldsymbol{\psi}')$ as in (9).

3. If the copula is potentially correctly specified, use $\widehat{A}_l^{-1}$ as the estimator for the asymptotic variance in (10). Otherwise, use the fully robust (11).

Estimator 2 also encompasses the use of a Dirichlet joint distribution with a multinomial logit link, as this can be expressed using an independent copula with beta marginals after a transformation (Connor and Mosimann, 1969; Hijazi and Jernigan, 2009). If there is no reason to believe that the

copula might be correctly specified, then by using the fully robust asymptotic variance estimator in both the multinomial logit and copula models, we would usually expect Estimator 1 to actually be more efficient, as it has to estimate less parameters to arrive at a solution. This is studied numerically in Section 3.

As a final consideration, recall that these estimators can recover the conditional mean parameters (and random effects variance) that can then be used to estimate the average partial effects by estimating the derivatives of covariates with respect to (12). However, if our only goal was to consistently estimate these partial effects, you could simply estimate a multinomial logit link via quasi-maximum likelihood and obtain average partial effects as noted in Wooldridge (2005), which requires no integration. While this is a perfectly valid approach, this method would not generalize well to the inclusion of possible endogenous covariates. Thus, we instead consider the probit link version of this issue in the next subsection, that does allow for simple inclusion of endogeneity.

2.2 Probit Estimator

With the notation and assumptions outlined in the previous subsection, it becomes easy to define a very simple estimator that parallels that in Papke and Wooldridge (2008). This time, instead of a multinomial logit link, assume a probit link for each share:

$$\mathrm{E}[\boldsymbol{Y}_{it}|\boldsymbol{X}_{it},\boldsymbol{c}_i] = \boldsymbol{m}(\boldsymbol{X}_{it}\boldsymbol{\beta}+\boldsymbol{c}_i) = \begin{bmatrix} \Phi(\boldsymbol{x}_{it1}'\boldsymbol{\beta}_1 + c_{i1}) \\ \vdots \\ \Phi(\boldsymbol{x}_{itD}'\boldsymbol{\beta}_D + c_{iD}) \end{bmatrix},$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function (CDF). Using the properties of the normal CDF, we can readily integrate the unobserved heterogeneity from the conditional mean function to arrive at

$$\mathrm{E}[\boldsymbol{Y}_{it}|\boldsymbol{X}_{it}] = \begin{bmatrix} \Phi\left(\boldsymbol{x}_{it1}'\left(\frac{\boldsymbol{\beta}_1}{\sqrt{1+\gamma_1^2}}\right)\right) \\ \vdots \\ \Phi\left(\boldsymbol{x}_{it1}'\left(\frac{\boldsymbol{\beta}_D}{\sqrt{1+\gamma_D^2}}\right)\right) \end{bmatrix} = \begin{bmatrix} \Phi(\boldsymbol{x}_{it1}'\boldsymbol{\beta}_{1c}) \\ \vdots \\ \Phi(\boldsymbol{x}_{itD}'\boldsymbol{\beta}_{Dc}) \end{bmatrix}, \tag{13}$$

where for $j = 1,\ldots,D$, $\boldsymbol{\beta}_{cj} = \boldsymbol{\beta}_j/(1+\gamma_j^2)^{1/2}$ and γ_j^2 is the j-th diagonal element of $\boldsymbol{\Gamma}$. Thus, similarly to Papke and Wooldridge (2008), identification of the conditional mean parameters is no longer possible (and the same is true for $\boldsymbol{\Gamma}$) but the average partial effects are still identified. Indeed, as shown by Wooldridge (2005), the average partial effect of covariate x_{itjk} on outcome y_{itj} is given as the derivative or difference (if it is categorical) of

$$\mathrm{E}_{\bar{x}_{ij}}[\Phi(\boldsymbol{x}_{itj}'\boldsymbol{\beta}_{cj} + \bar{\boldsymbol{x}}_{ij}'\boldsymbol{\xi}_{cj})] \tag{14}$$

where $\boldsymbol{\xi}_{cj} = \boldsymbol{\xi}_j/(1+\gamma_j^2)^{1/2}$ and I explicitly include $\bar{\boldsymbol{x}}_{ij}$ to emphasize that it is being integrated out of this unconditional expectation. Then, given a consistent estimator of the scaled parameters of the probit link, the average partial effects can be identified. In obtaining this consistent estimator, however, we run into an important issue: the probit link itself does not necessarily satisfy Assumption 2. Specifically, define $m_d(\boldsymbol{X}_{it}\boldsymbol{\beta}+\boldsymbol{c}_i) = 1 - \sum_{j=1}^{D}\Phi(\boldsymbol{x}_{itj}'\boldsymbol{\beta}_j + c_{ij})$. Then it is not necessarily the case that $m_d(\boldsymbol{X}_{it}\boldsymbol{\beta}+\boldsymbol{c}_i) > 0$, as the probit link does not collectively impose $\sum_{j=1}^{D}\Phi(\boldsymbol{x}_{itj}'\boldsymbol{\beta}_j + c_{ij}) < 1$ as is done by the multinomial logit link. This would imply that the conditional mean might not be correctly specified, and thus estimating $\boldsymbol{\beta}_c$ from (13) might not consistently estimate $\boldsymbol{\beta}_{0c}$.

However, it is important to note that this method would still provide the best probit link approximation to each of the conditional mean functions for each fraction separately. By also taking into account the correlation between each share in the system, it operates in a way similar to a seemingly unrelated regressions (SUR) approach. That is, imagine fitting a probit link conditional expectation to each fractional outcome Y_{itj} using panel methods, where the base category is taken to be $1 - Y_{itj}$. If we expect this to be a correctly specified model, then we would be consistently estimating $\boldsymbol{\beta}_{0,j}$. If we repeat this thought experiment for each $j = 1, \ldots, D$, and accept the probit link as a correctly specified link at each step, then the multivariate solution that approximates each of the conditional means while taking into account the correlation between shares should be a good approximation to the system as a whole. Finally, the method provides this approximation for the coefficients and partial effects in a way that is simple, computationally fast, and can incorporate continuous endogenous covariates using standard control function arguments (Papke and Wooldridge, 2008). We can also proceed with estimation by multivariate nonlinear least squares and adjust inference for the use of a potentially misspecified conditional mean function.

Formally, writing $\boldsymbol{\alpha}_c = (\boldsymbol{\beta}_c, \boldsymbol{\xi}_c)$ and given the objective function contribution

$$q_i(\boldsymbol{\alpha}_c) \equiv q(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\alpha}_c) = \frac{1}{2}[\boldsymbol{Y}_i - \boldsymbol{m}_{T_i}(\tilde{\boldsymbol{X}}_i\boldsymbol{\alpha}_c)]'[\boldsymbol{Y}_i - \boldsymbol{m}_{T_i}(\tilde{\boldsymbol{X}}_i\boldsymbol{\alpha}_c)] \tag{15}$$

the pooled multivariate nonlinear least squares estimator of $\boldsymbol{\alpha}_c = (\boldsymbol{\beta}_c, \boldsymbol{\xi}_c)$ with the probit link is found as

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_c &\equiv \arg\min_{\boldsymbol{\alpha_c}} \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}[\boldsymbol{Y}_{it} - \boldsymbol{m}(\tilde{\boldsymbol{X}}_{it}\boldsymbol{\alpha}_c)]'[\boldsymbol{Y}_{it} - \boldsymbol{m}(\tilde{\boldsymbol{X}}_{it}\boldsymbol{\alpha}_c)] \\ &= \arg\min_{\boldsymbol{\alpha_c}} \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\sum_{j=1}^{D}[y_{itj} - \Phi(\tilde{\boldsymbol{x}}'_{itj}\boldsymbol{\alpha}_{cj})]^2 \end{aligned} \tag{16}$$

where the definitions of $\tilde{\boldsymbol{x}}$ and $\boldsymbol{\alpha}$ come from (6). Thus, as outlined in White (1981) and section 12.3 of Wooldridge (2010), even if the probit link is potentially misspecified as a conditional mean for the multivariate fractions, $\widehat{\boldsymbol{\alpha}}_c$ is consistent to the value $\boldsymbol{\alpha}_c^*$ that creates the best probit link approximation, in a mean squared error sense, to the true conditional mean $\mathrm{E}[\boldsymbol{Y}_{it}|\boldsymbol{X}_{it}]$. Furthermore, if $\sum_{j=1}^{D}\Phi(\tilde{\boldsymbol{x}}'_{itj}\widehat{\boldsymbol{\alpha}}_{cj}) < 1$ for all i and t, we have no reason to expect that the probit link approximation would be a poor one.

Asymptotic normality centered around $\boldsymbol{\alpha}_c^*$ also holds, so that $\sqrt{n}(\widehat{\boldsymbol{\alpha}}_c - \boldsymbol{\alpha}_c^*)$ is asymptotically normal with asymptotic variance given by

$$\text{Asy. Var}(\sqrt{n}(\widehat{\boldsymbol{\alpha}}_c - \boldsymbol{\alpha}_c^*)) = A^{-1}BA^{-1}\,, \tag{17}$$

where, similar to the previous subsection, $A = \mathrm{E}_H[\partial^2 q_i(\boldsymbol{\alpha}_c)/\partial\boldsymbol{\alpha}_c\partial\boldsymbol{\alpha}_c']$ is the Hessian matrix of the objective contributions and $B = \mathrm{E}_H[\partial q_i(\boldsymbol{\alpha}_c)/\partial\boldsymbol{\alpha}_c \cdot \partial q_i(\boldsymbol{\alpha}_c)/\partial\boldsymbol{\alpha}_c']$ is the outer product of the scores. By using the full Hessian that does not assume $\mathrm{E}_H[\boldsymbol{Y}_i - \boldsymbol{m}_{T_i}(\tilde{\boldsymbol{X}}_i\boldsymbol{\alpha}_c)] = 0$, inference is made robust to the possible misspecification of the probit link, as well as autocorrelation in the scores. Estimation of the asymptotic variance in (17) follows as

$$\widehat{A} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 q_i(\widehat{\boldsymbol{\alpha}}_c)}{\partial\boldsymbol{\alpha}_c\partial\boldsymbol{\alpha}_c'} \quad \text{and} \quad \widehat{B} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial q_i(\widehat{\boldsymbol{\alpha}}_c)}{\partial\boldsymbol{\alpha}_c}\cdot\frac{\partial q_i(\widehat{\boldsymbol{\alpha}}_c)}{\partial\boldsymbol{\alpha}_c}'\,. \tag{18}$$

Given that the probit link is a simple special case, formulas for both the scores and Hessian are available; these are given in Appendix B. This procedure is summarized as follows.

Estimator 3 (Probit pooled multivariate NLS)**.**

1. Estimate $\boldsymbol{\alpha}_c$ from (16) by pooling across observations, time and outcome equations with the probit link.

2. For fully robust inference, estimate the covariance matrix for $\widehat{\boldsymbol{\alpha}}_c$ from (18) using the formulas in Appendix B.

If the probit link is deemed to be a good approximation, a possible next step to gain efficiency is to use a two-step estimator that specifies a weighted adjustment to the objective function in (15). As the estimator defined in (16) is also a generalized method of moments (GMM) estimator with an identity weighting matrix, the two-step choice could be implemented by using a different weight matrix choice. While the identity choice does not incorporate the correlation structure between the shares, this correlation is accounted for in the inference step when using the estimators (18). Furthermore, both consistency and asymptotic normality is unaffected; the choice of weighting matrix should only affect efficiency concerns. Given that there is a potential misspecification problem, once again it does not seem worthwhile to pursue larger efficiency gains if a crucial part of the distribution might not be correct. For more details, see, e.g., section 12.4 in Wooldridge (2010).

2.3 Bayesian Latent Variable Estimator

While the previous methods are able to handle zeros in the data naturally, they do not account for the possibly large probability that might accumulate at 0 for some fractions (Liu et al., 2020). There is now abundant research in ways to deal with these zeros in multivariate fractional outcomes. However, to account for non-trivial probability at zero; i.e., censoring for corner outcomes, the literature usually focuses on limited dependent variable approaches. To this end, I maintain Assumptions 1, 3 and 4.[5] I will assume the following limited dependent variable (LDV) model holds for all i, t, and j:

$$y^*_{itj} = \boldsymbol{x}'_{itj}\boldsymbol{\beta}_j + c_{ij} + \varepsilon_{itj}\,.$$

Here, y^*_{itj} is an unobservable latent variable. We can stack the previous model as before, to obtain

$$\boldsymbol{Y}^*_i = \boldsymbol{X}'_i\boldsymbol{\beta} + \boldsymbol{W}_i\boldsymbol{c}_i + \boldsymbol{\varepsilon}_i\,,$$

where the definitions mimic those in (4) with the addition of $\boldsymbol{W}_i = \iota_{T_i} \otimes I_D$, a $DT_i \times D$ matrix, where ι_{T_i} is a T_i-dimensional vector of ones and I_{T_i} is a $T_i \times T_i$ identity matrix. To allow for possible autocorrelation and contemporaneous correlation between outcomes, I assume $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_{DT_i}(\mathbf{0}_{DT_i\times 1}, \lambda_i^{-1}(\Omega_i \otimes \Sigma))$. In this specification Σ is a $D \times D$ contemporaneous covariance matrix that is left unrestricted, Ω_i is assumed to be known or to be the result of a specific VARMA process whose parameters need to be estimated, and λ_i^{-1} is a precision parameter. As outlined by Chib (2008), if λ_i is given a gamma $\mathcal{G}(\nu/2, \nu/2)$ prior and integrated out, then ε_i would have a marginal multivariate t distribution with ν degrees of freedom and scale matrix $\Omega_i \otimes \Sigma$. That is, we can allow for robust non-normal errors by giving the precision parameter an appropriate prior.

Now, in contrast to a usual probit or Tobit LDVs, there is no unified way to map the latent variables $\boldsymbol{Y}^*_{it}$ to the simplex $\mathcal{S}^d$ and obtain its inverse transformation. Even when focusing to those that allow for zeros, there have been several proposals in the literature, such as re-scaling the sum of the positive $\boldsymbol{Y}^*_{it}$ (Wales and Woodland, 1983), via Box-Cox transformations of ratios of variables (Fry et al., 2000; Tsagris et al., 2011), by minimizing the Euclidean distance from $\boldsymbol{Y}^*_{it}$ to $\mathcal{S}^d$ (Butler and Glasbey, 2008), among others. Due to the computational simplicity of the resulting simulation scheme, I focus on

[5] As noted by Chib (2008), Bayesian estimation can usually relax the strict exogeneity assumption for one of sequential exogeneity, given the distributional assumptions and dynamic completeness of the resulting likelihoods.

the scaling transformation given by (Wales and Woodland, 1983) and described as part of a Bayesian cross-sectional approach in Kasteridis et al. (2011).

This approach fixes the sum of the underlying latent variables to 1, and transforms to observable variables supported on $\mathcal{S}^d$ by using

$$y_{itj} = \frac{\max\{y^*_{itj}, 0\}}{1 - \sum_{(t,p)\in E_i} y^*_{itp}}, \tag{19}$$

for all i, t, and j, where $E_i = \{1 \leq t \leq T_i, 1 \leq j \leq D : y^*_{itj} \leq 0\}$. The censored set is defined in this way given that ε_i is not necessarily independent over time and thus both temporal and contemporaneous correlations will influence whether a particular latent observation falls into the censoring set or not. It will also be necessary for the simulation algorithm to be introduced shortly. Note that fixing the sum is related to the identification issue mentioned previously, as not constraining the support of $\boldsymbol{Y}^*_{it}$ results in infinitely many solutions to the inverse problem of finding the $\boldsymbol{Y}^*_{it}$ that generate a particular observable $\boldsymbol{Y}_{it}$.

The Bayesian paradigm recognizes that Assumption 4 is simply a prior distribution on the correlated random effects. For simplicity, I once again assume that $\boldsymbol{c}_i$ directly represents a random effect, as would occur after employing the Mundlak device. By assigning prior distributions to the remaining parameters over which there is uncertainty, we can combine them with the likelihood implied by the normality assumption on ε_i to produce a posterior distribution. I assume the following normal and inverse Wishart conjugate prior distributions on the remaining model parameters:

$$\begin{aligned} \boldsymbol{\beta} &\sim \mathcal{N}(\boldsymbol{\beta}_0, \boldsymbol{B}_0)\,, \\ \boldsymbol{\Gamma} &\sim \mathcal{IW}(\nu_\Gamma, \boldsymbol{R}_\Gamma)\,, \\ \boldsymbol{\Sigma} &\sim \mathcal{IW}(\nu_\Sigma, \boldsymbol{R}_\Sigma)\,. \end{aligned} \tag{20}$$

The data augmentation approach due to Albert and Chib (1993) that is common in Bayesian estimation of LDVs includes the $\boldsymbol{Y}^*_i$ as parameters (McCulloch et al., 2000). Thus, with these prior distributions in place, the posterior for all the parameters $\boldsymbol{\beta}$, $\boldsymbol{Y} = (\boldsymbol{Y}^{*'}_1, \ldots, \boldsymbol{Y}^{*'}_n)'$, $\boldsymbol{c} = (\boldsymbol{c}'_1, \ldots, \boldsymbol{c}'_n)'$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ conditional on data $\boldsymbol{Y} = (\boldsymbol{Y}'_1, \ldots, \boldsymbol{Y}'_n)'$, $\boldsymbol{X} = (\boldsymbol{X}'_1, \ldots, \boldsymbol{X}'_n)'$, and $\boldsymbol{W} = (\boldsymbol{W}'_1, \ldots, \boldsymbol{W}'_n)'$, denoted by $\pi(\cdot|\cdot)$ yields

$$\begin{aligned} \pi(\boldsymbol{\beta}, \boldsymbol{Y}^*, \boldsymbol{c}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}, \boldsymbol{\lambda} | \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{W}) \propto \prod_{i=1}^{n} \Bigg\{ \Bigg[& \prod_{t=1}^{T_i} \prod_{j=1}^{D} I(y_{itj} = 0) I(y^*_{itj} \leq 0) \\ & + I(y_{itj} > 0) I\left(y_{itj} = \frac{y^*_{itj}}{1 - \sum_{(t,p)\in E_i} y^*_{itp}} \right) \Bigg] \\ & \times \phi_{DT_i}(\boldsymbol{Y}^*_i; \boldsymbol{X}'_i\boldsymbol{\beta} + \boldsymbol{W}_i\boldsymbol{c}_i; \lambda_i^{-1}(\boldsymbol{\Omega}_i \otimes \boldsymbol{\Sigma})) \Bigg\} \\ & \times \pi(\boldsymbol{\beta})\pi(\boldsymbol{c})\pi(\boldsymbol{\lambda})\pi(\boldsymbol{\Gamma})\pi(\boldsymbol{\Sigma})\,. \end{aligned} \tag{21}$$

In this equation, $\pi(\cdot)$ for each parameter refers to their assumed prior distribution and $I(\cdot)$ denotes an indicator function that is equal to 1 when its argument is true and 0 otherwise. Note that for all i, t and j such that $y_{itj} = 0$, the posterior implies a normal distribution for y^*_{itj} truncated to $(-\infty, 0]$. For all positive parameters, the distribution is singular and puts all mass at the inversely transformed values given by

$$y^*_{itj} = y_{itj}\left(1 - \sum_{(t,p)\in E_i} y^*_{itp}\right). \tag{22}$$

From (21), we can obtain the conditional distribution of each parameter on all other model parameters and the data to propose a Gibbs sampling scheme to simulate from the posterior. This is summarized in the following procedure and uses the usual Bayesian updates with conjugate priors under normality (see, e.g., Chib, 2008).

Estimator 4 (Bayesian LDV estimator)**.** For simplicity, this assumes that $\boldsymbol{\lambda} = \iota_n$ and $\boldsymbol{\Omega}_i = I_{T_i}$ but incorporating other structures is simple. At the s-th simulation step:

1. For each i, draw $y_{itj}^{*(s)}$ for all those $(t,j) \in E_i$ from

$$\mathcal{TN}_{(-\infty,0]}(\mu_{itj|-(tj)}, \sigma^2_{itj|-(tj)}),$$

where $\mathcal{TN}$ represents a truncated normal distribution with mean given by $\mu_{itj|-(tj)} = \mathrm{E}[y^*_{itj}|\boldsymbol{Y}^{*(s-1)}_{i,-(tj)}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}]$, variance $\sigma^2_{itj|-(tj)} = \mathrm{Var}(y^*_{itj}|\boldsymbol{Y}^{*(s-1)}_{i,-(tj)}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$, and where $\boldsymbol{Y}^*_{i,-(tj)}$ denotes the vector $\boldsymbol{Y}^*_i$ excluding the tj component. Calculate the remaining components of $\boldsymbol{Y}_i^{*(s)}$ with $(t,j) \notin E_i$ via (22).

2. Draw $\boldsymbol{\beta}^{(s)}|\boldsymbol{Y}^{*(s)}, \boldsymbol{\Gamma}^{(s-1)}, \boldsymbol{\Sigma}^{(s-1)} \sim \mathcal{N}(\bar{\boldsymbol{\beta}}^{(s)}, \bar{\boldsymbol{B}}^{(s)})$ where

$$\begin{aligned}
\bar{\boldsymbol{B}}^{(s)} &= \left(\boldsymbol{B}_0^{-1} + \sum_{i=1}^n \boldsymbol{X}_i' \boldsymbol{V}_i^{-1(s-1)} \boldsymbol{X}_i\right)^{-1}, \\
\bar{\boldsymbol{\beta}}^{(s)} &= \bar{\boldsymbol{B}}^{(s)} \left(\boldsymbol{\beta}_0^{-1} + \sum_{i=1}^n \boldsymbol{X}_i' \boldsymbol{V}_i^{-1(s-1)} \boldsymbol{Y}_i^{*(s)}\right)^{-1}, \\
\boldsymbol{V}_i^{(s-1)} &= (I_{T_i} \otimes \boldsymbol{\Sigma}^{(s-1)}) + \boldsymbol{W}_i \boldsymbol{\Gamma}^{(s-1)} \boldsymbol{W}_i'.
\end{aligned}$$

3. For each i, draw $\boldsymbol{c}_i^{(s)}|\boldsymbol{Y}^{*(s)}, \boldsymbol{\beta}^{(s)}, \boldsymbol{\Gamma}^{(s-1)}, \boldsymbol{\Sigma}^{(s-1)} \sim \mathcal{N}(\bar{\boldsymbol{c}}_i^{(s)}, \bar{\boldsymbol{\Gamma}}_i^{(s)})$ where

$$\begin{aligned}
\bar{\boldsymbol{\Gamma}}_i &= \left[\boldsymbol{\Gamma}^{-1(s-1)} + \boldsymbol{W}_i'(I_{T_i} \otimes \boldsymbol{\Sigma}^{-1(s-1)})\boldsymbol{W}_i\right]^{-1}, \\
\bar{\boldsymbol{c}}_i &= \bar{\boldsymbol{\Gamma}}_i \boldsymbol{W}_i'(I_{T_i} \otimes \boldsymbol{\Sigma}^{-1(s-1)})(\boldsymbol{Y}_i^{*(s)} - \boldsymbol{X}_i \boldsymbol{\beta}^{(s)}).
\end{aligned}$$

4. Draw $\bar{\boldsymbol{\Gamma}}^{(s)}|\boldsymbol{c}_i^{(s)} \sim \mathcal{IW}(\bar{\nu}, \bar{\boldsymbol{R}}_\Gamma^{(s)})$ where

$$\begin{aligned}
\bar{\nu}_\Gamma &= \nu_\Gamma + n, \\
\bar{\boldsymbol{R}}_\Gamma^{(s)} &= \boldsymbol{R}_\Gamma + \sum_{i=1}^n \boldsymbol{c}_i^{(s)} \boldsymbol{c}_i^{\prime(s)}.
\end{aligned}$$

5. Draw $\bar{\boldsymbol{\Sigma}}^{(s)}|\boldsymbol{c}_i^{(s)} \sim \mathcal{IW}(\bar{\nu}, \bar{\boldsymbol{R}}_\Sigma^{(s)})$ where

$$\begin{aligned}
\bar{\nu}_\Sigma &= \nu_\Sigma + \sum_{i=1}^n T_i, \\
\bar{\boldsymbol{R}}_\Gamma^{(s)} &= \boldsymbol{R}_\Gamma + \sum_{i=1}^n \boldsymbol{e}_i^{\prime(s)} \boldsymbol{e}_i^{(s)},
\end{aligned}$$

and $\boldsymbol{\epsilon}_i^{(s)}$ is a $T_i \times D$ matrix such that $\mathrm{vec}(\boldsymbol{e}_i^{\prime(s)}) = \boldsymbol{Y}_i^{*(s)} - \boldsymbol{X}_i \boldsymbol{\beta}^{(s)} - \boldsymbol{W}_i \boldsymbol{c}_i^{(s)}$; i.e., the i-th residuals in matrix form. This is perhaps the only nonstandard update that arises from the connection

between the vector representation of the distribution for $\boldsymbol{\varepsilon}_i$ with the matricvariate representation (see section A.1.12 of Greenberg, 2012). That is, given that $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_{DT_i}(\mathbf{0}_{DT_i\times 1}, I_{T_i} \otimes \Sigma)$, then define the $T_i \times D$ random matrix $\boldsymbol{\epsilon}_i$ such that $\text{vec}(\boldsymbol{\epsilon}_i') = \boldsymbol{\varepsilon}_i$. Then $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{T_i\times D}(\mathbf{0}_{T_i\times D}, \Omega_i, \Sigma)$ is matricvariate normal.

An important final observation is that, just as the LDV approach recognizes the use of Assumption 4 as a prior distribution, the same could be done for the maximum likelihood approach in Section 2.1. While the main deterrent from using Bayesian analysis for this class of generalized or nonlinear mixed effects models has been computational, there are now many available tools that allow for simulating the posterior of a system using priors (20) along with the likelihoods given in (7) or (8). Furthermore, as Fong et al. (2010) point out, the use of priors for the covariance matrix of the random effects allows for a more realistic inclusion of the uncertainty of these estimates in contrast to the use of a single estimate. This would be reflected as more believable standard errors for the estimated panel coefficients.

3 Numerical Exercises

To test the performance and comparative advantages of each method, I present several Monte Carlo exercises. To ensure that each method satisfies the assumptions laid out in the previous section and to test them under distinct conditions that might be found in practice, I use several data-generating processes to test each estimator. Some of these should be well-suited to the specifics of each method while others will test their robustness to possible misspecification. To keep matters concise, I will be focusing specifically on the procedures outline in Estimators 1 through 4.

3.1 Copula Data-Generating Process

Given that the multinomial logit is a misspecified distribution by construction, it does not allow for the generation of data that could be used to test the behavior of Estimators 1 and 2 under correct specification. Therefore, the first Monte Carlo exercise draws variables from a copula model as that in Montoya-Blandón (2021). Specifically, I will use a Gaussian copula with beta marginals and a multinomial logit link, which was found to be one of the most numerically stable and robust methods both for generation and estimation. To this end, I draw pseudo observations $u_1, \dots u_D$ from the Gaussian copula density

$$c(u_1, \dots, u_D) = \frac{1}{\sqrt{\det R}} \exp\left(-\frac{1}{2} \begin{bmatrix} \Phi^{-1}(u_1) & \cdots & \Phi^{-1}(u_D) \end{bmatrix} \cdot (R^{-1} - I_D) \cdot \begin{bmatrix} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_D) \end{bmatrix} \right),$$

with $D \times D$ correlation matrix R, where $\Phi^{-1}(\cdot)$ is the quantile function for the standard normal distribution. I then use the probability integral transform to guarantee that the draws are from beta marginals in a mean-precision parameterization. Thus, for each j in $1, \dots, D$, u_j is transform by the inverse of the cumulative distribution function of the beta density with mean m_j and precision ϕ_j, which is given as

$$\frac{\Gamma(\phi_j)}{\Gamma(m_j)\Gamma[(1-m_j)\phi_j]} y_j^{m_j\phi_j} (1-y_j)^{[1-m_j]\phi_j},$$

for $0 < y_j < 1$. In this first scenario, I draw $D = 2$ shares (y_{it1}, y_{it2}) for $i = 1, \dots, n$ individuals with $n \in \{100, 200\}$ and $t = 1, 2$ time periods for a total of 200 or 400 observations on each share. The third share y_{it3} is set to $1 - y_{it1} - y_{it2}$ for all i and t. I set $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ with $\boldsymbol{\beta}_1 = (-1, 0.5, 0)'$ and

$\boldsymbol{\beta}_2 = (-1.5, 0, 0.5)'$. Two covariates x_{it1} and x_{it2} are drawn from standard normal distributions and unobserved heterogeneity is added in the form of a random effect $\boldsymbol{c}_i$ drawn from a multivariate normal distribution with zero mean and covariance matrix $\boldsymbol{\Gamma}$ with $\Gamma_{11} = \Gamma_{22} = 1$ and $\Gamma_{12} = \Gamma_{21} = 0.5$. I assume a multinomial logit link as that given in (12) for the means m_{it1} and m_{it2} of each beta distribution. The precision parameters are set to $\phi_1 = \phi_2 = 10$ and a correlation of $\rho = 0.5$ is used to form matrix R for use in the Gaussian copula density.

Across 500 Monte Carlo simulations with the previous baseline scenarios, the multinomial quasi-maximum likelihood (QMLE) and the copula maximum likelihood estimators were calculated using the conditionally independent version of the likelihood, as in (7) and use nonadaptive quadrature with 10 evaluation points in each dimension. For a given application, I would recommend using the nonadaptive version with a larger number of evaluation points as a starting point to then use the adaptive version with relatively fewer until the differences are not noticeable between successive estimates. The probit pooled multivariate nonlinear least squares (PMNLS) is by far the most efficient method, as it has no need for evaluating integrals and the availability of scores and Hessian greatly simplify the computation of robust inference.

Table 1: RMSE for Coefficients in a from a Gaussian Copula with Beta Marginals and Multinomial Logit Link

Method	$\beta_{1,0}$	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{2,0}$	$\beta_{2,1}$	$\beta_{2,2}$
$nT = 200$						
Multinomial QMLE	0.113	0.095	0.084	0.114	0.088	0.094
Copula MLE	0.187	0.080	0.082	0.190	0.086	0.088
Probit PMNLS	0.277	0.248	0.084	0.452	0.101	0.258
$nT = 400$						
Multinomial QMLE	0.079	0.068	0.061	0.098	0.077	0.064
Copula MLE	0.153	0.057	0.059	0.161	0.065	0.059
Probit PMNLS	0.277	0.250	0.077	0.451	0.093	0.258

Note: RMSE across 500 simulations for each estimation procedure when data are generated from a Gaussian copula with beta marginals.

The results from using Estimators 1 through 3 are given in Table 1 in the forms of root mean squared errors (RMSE) from the true parameters. The analysis focuses on the conditional mean coefficients $\boldsymbol{\beta}$.[6] As expected, given a correctly specified link function, the estimates remain consistent to the true parameters, as evidenced by the declining RMSE at an expected rate. Both the multinomial QMLE and copula estimators compete in terms of RMSE but it is not surprising that the copula estimator tends to be slightly better, given that it is a correctly specified MLE. The probit estimator, on the other hand, remains inconsistent, which is to be expected given the incorrect link. As observed by Montoya-Blandón and Jacho-Chávez (2020), link misspecification can cause large biases even when two relatively similar links such as the logit and probit are used in one specification. However, the RMSE information hides an important point. We know from the theory in the previous section that when unobserved heterogeneity is involved, the probit would not even identify the correct coefficients, so its inconsistency for the true $\boldsymbol{\beta}_0$ is not surprising.

A more complete depiction is given in the following set of results, found in Table 2. This table presents the mean coefficients and standard errors across the 500 Monte Carlo simulations. First, note that once again the multinomial QMLE and copula MLE are quite close in their performance, both

[6]The results for the complete parameters are available upon request.

in terms of mean coefficients and standard errors. This is interesting given that the copula standard errors rely on the correctly specified variance covariance matrix, while the multinomial QMLE uses the fully robust formulas (see 10). Thus, as expected, the fact that the copula model estimates a larger number of parameters likely diminishes the possible efficiency gains from correctly specifying the distribution. Now, as mentioned before, while the probit PMNLS is not correctly capturing the underlying conditional mean coefficients, it should provide the best probit link approximation to the scaled coefficients. Since we know that both true unobserved heterogeneity variances equal 1, this will mean that the probit will identify and consistently estimate $\boldsymbol{\beta}^*/\sqrt{2}$. We note this value under the true conditional mean coefficients in Table 2. As can be observed, the probit PMNLS approach is indeed quite close to these values. The remaining bias is likely explained by the link misspecification and small sample sizes. Still, this implies that the average partial effects recovered from using these scaled coefficients will likely be close to the true effects, or at least as close as the marginal effects from a multinomial and probit specification can be. As an example, the true average partial effect of x_{it1} on y_{it1} evaluated at $x_{it1} = x_{it2} = 0$ using the multinomial logit link is 0.088. Averaging across the Monte Carlo simulations, I find that this effect is estimated to be 0.084 on average from the multinomial logit link, and 0.077 from the probit approximations, where both examples use the full 400 observations.

Table 2: Coefficients from a Multinomial Logit Link in a Gaussian Copula with Beta Marginals

Method	$\beta_{1,0}$	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{2,0}$	$\beta_{2,1}$	$\beta_{2,2}$
			$nT = 200$			
Multinomial QMLE	−1.033	0.459	−0.021	−1.524	−0.020	0.462
	(0.107)	(0.084)	(0.084)	(0.118)	(0.093)	(0.094)
Copula MLE	−1.122	0.500	−0.022	−1.628	−0.013	0.494
	(0.115)	(0.074)	(0.074)	(0.124)	(0.082)	(0.082)
Probit PMNLS	−0.729	0.257	−0.071	−1.052	−0.088	0.247
	(0.054)	(0.046)	(0.044)	(0.056)	(0.046)	(0.050)
$\boldsymbol{\beta}_0/\sqrt{2}$	−0.707	0.354	0.000	−1.061	0.000	0.354
			$nT = 400$			
Multinomial QMLE	−1.027	0.444	−0.036	−1.555	−0.051	0.461
	(0.073)	(0.059)	(0.059)	(0.084)	(0.067)	(0.068)
Copula MLE	−1.113	0.495	−0.020	−1.627	−0.017	0.490
	(0.083)	(0.053)	(0.052)	(0.089)	(0.058)	(0.058)
Probit PMNLS	−0.726	0.252	−0.071	−1.051	−0.086	0.245
	(0.038)	(0.033)	(0.031)	(0.040)	(0.033)	(0.036)
$\boldsymbol{\beta}_0/\sqrt{2}$	−0.707	0.354	0.000	−1.061	0.000	0.354

Note: Average coefficients and standard errors across 500 simulations for each estimation procedure when data are generated from a Gaussian copula with beta marginals. Standard errors are in parenthesis. For multinomial QMLE and probit PMNLS these are robust to distributional misspecification in each iteration.

3.2 Probit Data-Generating Process

To test an opposing situation to the one in the previous subsection, I now generate values from the probit PMNLS model. To this end, I generate values of $y_{itj}, j = 1, 2$ according to

$$y_{itj} = \Phi\left(\boldsymbol{x}'_{itj}\frac{\boldsymbol{\beta}_j}{\sqrt{2}}\right) + r_{itj}$$

where $r_{itj} \sim \mathcal{N}(0, 0.01)$ is an additional error term that is independent across units, time and shares. The variance is set low enough so that the multivariate fractions stay within the unit interval with sufficiently large probability after generation. This generation scheme assumes the probit link has already integrated out the underlying unobserved heterogeneity and so it generates directly from the conditional mean of Y_{itj} given $\boldsymbol{x}_{itj}$. All remaining values stay the same as in the previous scenario. Using this data-generating process, the values for RMSE can be found in Table 3 and the coefficients with associated standard errors in Table 4.

Table 3: RMSE for Coefficients from a Multivariate Nonlinear Least Squares with Probit Link

Method	$\beta_{1,0}$	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{2,0}$	$\beta_{2,1}$	$\beta_{2,2}$
$nT = 200$						
Multinomial QMLE	0.207	0.270	0.157	0.266	0.208	0.193
Copula MLE	0.501	0.224	0.138	0.444	0.199	0.140
Probit PMNLS	0.033	0.038	0.021	0.100	0.026	0.087
$nT = 400$						
Multinomial QMLE	0.168	0.298	0.162	0.230	0.265	0.206
Copula MLE	0.504	0.217	0.130	0.442	0.193	0.130
Probit PMNLS	0.029	0.034	0.016	0.098	0.018	0.083

Note: RMSE across 500 simulations for each estimation procedure when data are generated from a multivariate nonlinear least squares conditional mean with additive error.

As expected, the situation has reversed in comparison to the previous scenario. In this setting, the likelihood-based methods no longer remain consistent to the new true value of the parameters $\boldsymbol{\beta}_0/\sqrt{2}$. Their RMSE is erratic and their coefficients remain biased regardless of the sample size. The standard errors for all approaches are also lower than in the previous scenarios, likely due to the reduced variation introduced by the r_{itj} additive errors in comparison to that from the copula generating mechanism. On the other hand, the probit estimator now appears to be consistent with RMSE decreasing with larger sample size. The estimates remain much closer to the true value in comparison to before, reflecting the correct specification assumption. Interestingly, using a similar example as before, it appears that the probit link approximates the average partial effects much better even when misspecified. In the previous example, the approximation was fairly close to the averaged estimates from the multinomial QMLE APEs. This does not seem to occur in this reverse scenario. Now, the true average partial effect of x_{it1} on y_{it1} evaluated at $x_{it1} = x_{it2} = 0$ using the probit link is 0.109. The average of the estimated APEs from the correctly specified probit is 0.102, but the approximation by the multinomial logit is 0.084, which remains essentially unchanged from the previous scenario. Thus, while it seems that the probit link adapts quite well when it is misspecified, this does not seem to be the case for the multinomial logit QMLE.

3.3 Censored Data-Generating Process

Finally, consider a scenario that takes into account the possibility of having corner solutions expressed as structural zeros within the data:

$$y^*_{itj} = \boldsymbol{x}'_{itj}\boldsymbol{\beta}_j + c_{ij} + \varepsilon_{itj} \,. \tag{23}$$

This creates the need to adjust the values previously used for generation, as the underlying latent variable model (23) tends to yield too many zeros if the linear index induces a lot of variance on

Table 4: Coefficients from a Multivariate Nonlinear Least Squares with Probit Link

Method	$\beta_{1,0}$	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{2,0}$	$\beta_{2,1}$	$\beta_{2,2}$
			$nT = 200$			
Multinomial QMLE	−0.907	0.619	0.149	−1.319	0.198	0.538
	(0.047)	(0.049)	(0.048)	(0.057)	(0.057)	(0.058)
Copula MLE	−1.199	0.569	0.128	−1.497	0.188	0.477
	(0.081)	(0.056)	(0.055)	(0.094)	(0.063)	(0.062)
Probit PMNLS	−0.683	0.323	0.000	−0.964	−0.004	0.272
	(0.023)	(0.023)	(0.022)	(0.026)	(0.024)	(0.028)
$\boldsymbol{\beta}_0/\sqrt{2}$	−0.707	0.354	0.000	−1.061	0.000	0.354
			$nT = 400$			
Multinomial QMLE	−0.868	0.635	0.126	−1.273	0.236	0.550
	(0.033)	(0.035)	(0.034)	(0.04)	(0.041)	(0.040)
Copula MLE	−1.209	0.568	0.124	−1.500	0.188	0.476
	(0.056)	(0.039)	(0.038)	(0.066)	(0.044)	(0.044)
Probit PMNLS	−0.682	0.324	−0.001	−0.964	−0.004	0.273
	(0.016)	(0.017)	(0.015)	(0.019)	(0.017)	(0.020)
$\boldsymbol{\beta}_0/\sqrt{2}$	−0.707	0.354	0.000	−1.061	0.000	0.354

Note: Average coefficients and standard errors across 500 simulations for each estimation procedure when data are generated from a multivariate nonlinear least squares conditional mean with additive error. Standard errors are in parenthesis. Maximum likelihood methods use the fully robust standard errors.

$\boldsymbol{Y}^*$. Thus, I adjust the population values of the coefficients to $\boldsymbol{\beta}_1 = (-0.2, 0.15, -0.2)'$ and $\boldsymbol{\beta}_2 = (-0.15, -0.2, 0.15)'$ and it is now assumed that the variances for both the unobserved heterogeneity and the additive errors ε_{itj} are given by $\boldsymbol{\Gamma} = \boldsymbol{\Sigma}$ with the diagonal components equal to 0.02 and covariance 0.01. Furthermore, the covariates are generated from normal distributions with mean equal to 3.5 and standard deviation equal to 0.25. Generating (y^*_{it1}, y^*_{it2}) and mapping to observable multivariate fractions via (19) was found to produce approximately 20% censoring in the data. This large proportion of zeros can be taken into account by using the Bayesian alternative given in Estimator 4.

For estimation purposes, given the conjugate priors outlined for the Bayesian estimator in Section 2.3, all that remains is to specify the hyperparameters of these distributions. I choose standard uninformative priors for the coefficients by setting $\boldsymbol{\beta}_0 = \mathbf{0}_{K\times 1}$, $\boldsymbol{B}_0 = 1000 I_K$, $\nu_\Gamma = \nu_\Sigma = D + 1$ and $\boldsymbol{R}_\Gamma = \boldsymbol{R}_\Sigma = I_D$. With these values, I executed the Gibbs sampling algorithm outlined in Estimator 4 to find the posterior mean and median across from 5000 simulations after a burn-in period of 1000. The results for the mean of these Bayesian estimates across 500 Monte Carlo simulations can be found in Table 5. The parameter values can be seen to be close to the appropriate starting values and get better with a larger sample size. Furthermore, the standard errors, as measured by the standard deviation across the simulation chains is seen to also decrease with sample size, as expected. These simulations showcase the simplicity of dealing with censoring using a Bayesian perspective with a data augmentation scheme.

Finally, Figures 1 and 2 give a graphical depiction of the posterior chains for the coefficients in a single Monte Carlo draw. One of the major advantages of the Bayesian approach is its ability to produce a complete distribution for each parameter of interest from which all proceeding information is derived. As observed in the figures, the distribution of the coefficients centers around their true values and most sampling steps are taken close to the median. Using the usual diagnostics, I also confirmed

Table 5: Coefficients from a Bayesian Latent Dependent Variable Model

Estimate	$\beta_{1,0}$	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{2,0}$	$\beta_{2,1}$	$\beta_{2,2}$
			$nT = 200$			
Mean	−0.178	0.139	−0.191	−0.137	−0.222	0.175
Median	−0.177	0.138	−0.190	−0.136	−0.221	0.174
Std. Dev.	(0.058)	(0.041)	(0.046)	(0.051)	(0.04)	(0.043)
			$nT = 400$			
Mean	−0.197	0.141	−0.215	−0.121	−0.217	0.165
Median	−0.194	0.140	−0.213	−0.120	−0.215	0.164
Std. Dev.	(0.038)	(0.029)	(0.030)	(0.035)	(0.028)	(0.029)

Note: Average posterior mean and medians across 500 simulations from a latent dependent variable model. Standard errors are given as the standard deviation of the chains.

that the chains satisfy the criteria for convergence to their stationary distribution.

4 Conclusion

Multivariate fractional outcomes can arise from many interesting applied economic problems. As the literature has expanded to cover many interesting use of this data in statistics and econometrics, there have not been many developments that are useful in a panel data context. This paper attempts to fill that gap by introducing a wide range of methods for dealing with multivariate fractions in a way that deals with the specific issues surrounding these limited dependent variables, while also remaining flexible and robust enough to be widely applicable. First, a general maximum likelihood estimator that allows for correlated random effects was introduced, and noted that it remains robust to distributional misspecification. A second approach, and perhaps the one that will be most useful, is a multivariate nonlinear least squares estimator with a probit link that allows for identification of average partial effects and can incorporate endogeneity, arguably some of the most interesting challenges in any particular application. A final approach that allows for directly incorporating the zeros and accounting for this censoring was presented. In line with the literature of limited dependent variable models, a Bayesian solution is found to be flexible and computationally feasible comparative to other simulation-based alternatives.

As avenues for future research, it would be interesting to push the limits of these methods, particularly for applications with many shares, such as budget share allocations across many goods. Furthermore, it would be interesting to take these method to richer data sets that would allow to explore additional possibilities for estimation and inference, while providing important answers to problems where multivariate fractional outcomes can arise.

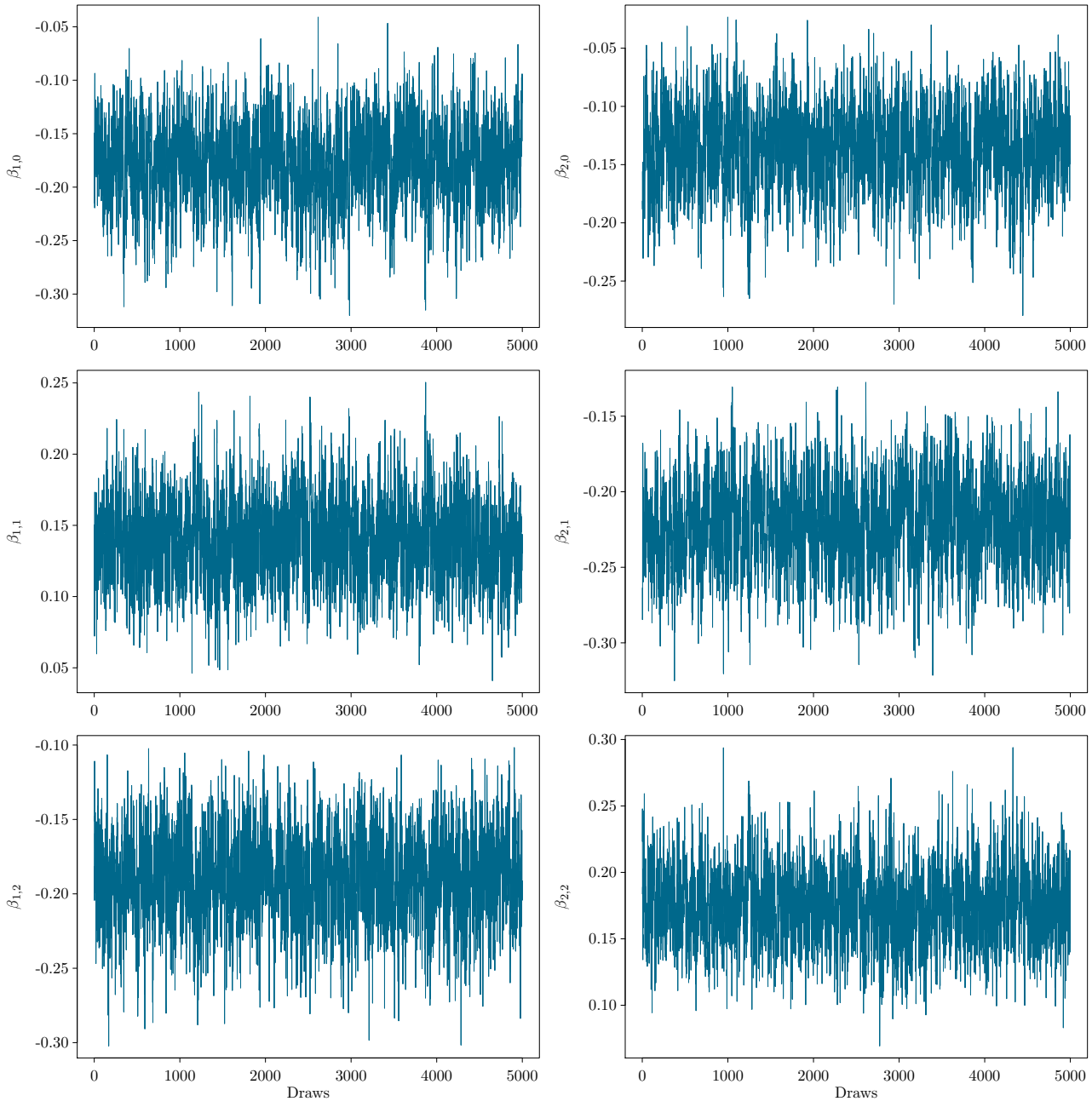
References

Aitchison, J. (1983). Principal Component Analysis of Compositional Data. *Biometrika*, 70(1):57–65.

Aitchison, J. (2003). *The Statistical Analysis of Compositional Data*. Blackburn Press.

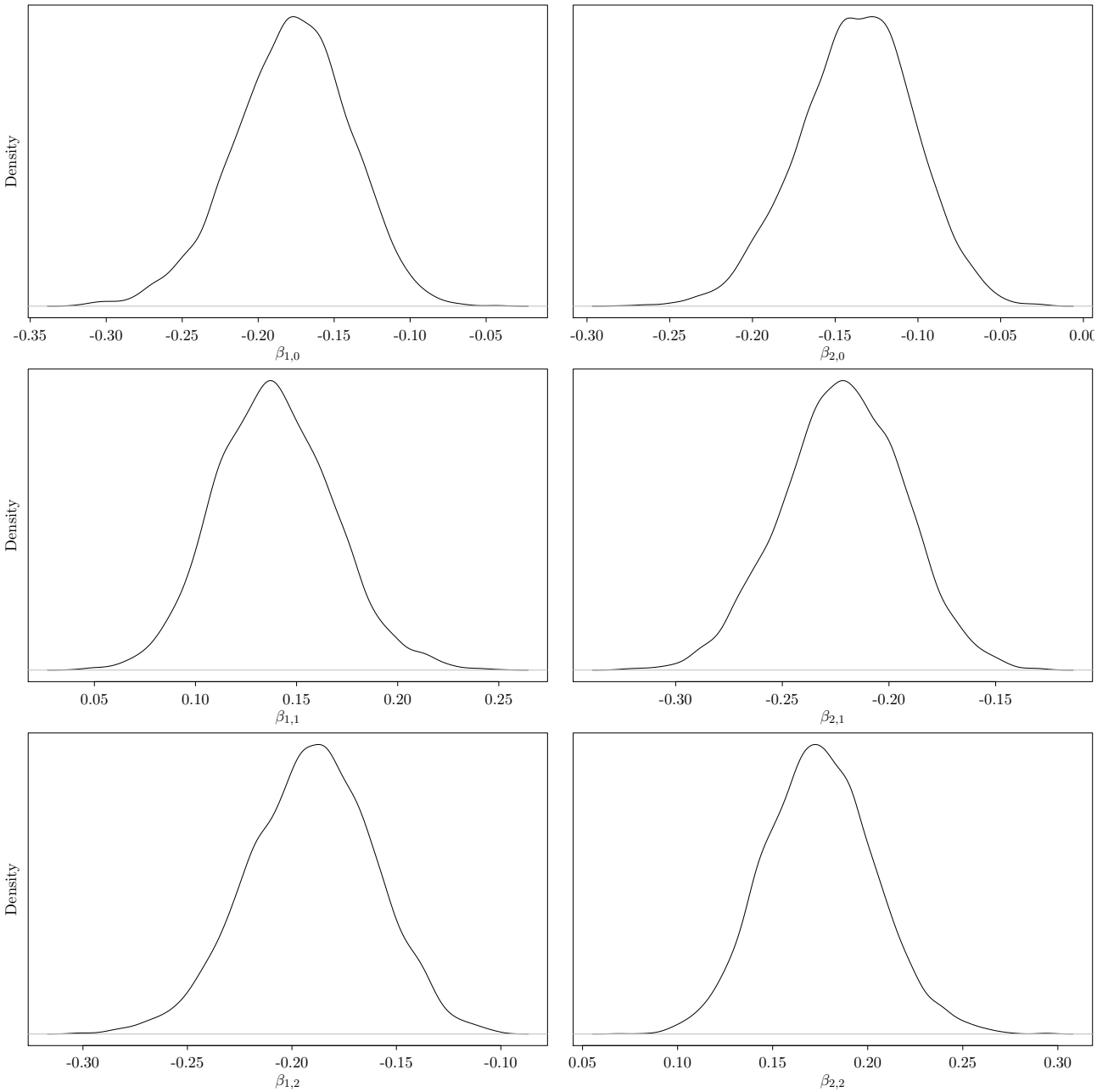
Aitchison, J. and Shen, S. M. (1980). Logistic-Normal Distributions: Some Properties and Uses. *Biometrika*, 67(2):261–272.

Figure 1: Trace Plot of Coefficients for Latent Dependent Variable Model



Note: Results from 5000 simulations after a burn-in period of 1000. The draws on the coefficients integrate out the unobserved heterogeneity.

Figure 2: Density Plot of Coefficients for Latent Dependent Variable Model



Note: Results from 5000 simulations after a burn-in period of 1000. The draws on the coefficients integrate out the unobserved heterogeneity.

Albert, J. H. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, 88(422):669–679.

Banks, J., Blundell, R., and Lewbel, A. (1997). Quadratic engel curves and consumer demand. *Review of Economics and statistics*, 79(4):527–539.

Barnett, W. A. and Serletis, A. (2008). Consumer Preferences and Demand Systems. *Journal of Econometrics*, 147(2):210–224.

Butler, A. and Glasbey, C. (2008). A Latent Gaussian Model for Compositional Data with Zeros. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57(5):505–520.

Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *The Review of Economic Studies*, 47(1):225–238.

Chib, S. (2008). Panel Data Modeling and Inference: A Bayesian Primer. In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, volume 46 of *Advanced Studies in Theoretical and Applied Econometrics*, pages 479–515. Springer.

Connor, R. J. and Mosimann, J. E. (1969). Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution. *Journal of the American Statistical Association*, 64(325):194–206.

Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*, volume 62 of *Monographs on Statistics and Applied Probability*. CRC Press.

Deaton, A. and Muellbauer, J. (1980). An Almost Ideal Demand System. *American Economic Review*, 70(3):312–326.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barcelo-Vidal, C. (2003). Isometric Logratio Transformations for Compositional Data Analysis. *Mathematical Geology*, 35(3):279–300.

Fong, Y., Rue, H., and Wakefield, J. (2010). Bayesian Inference for Generalized Linear Mixed Models. *Biostatistics*, 11(3):397–412.

Fry, J. M., Fry, T. R. L., and McLaren, K. R. (2000). Compositional Data Analysis and Zeros in Micro Data. *Applied Economics*, 32(8):953–959.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo Maximum Likelihood Methods: Theory. *Econometrica*, 52(3):681–700.

Greenberg, E. (2012). *Introduction to Bayesian Econometrics*. Cambridge University Press.

Hajivassiliou, V. A. and McFadden, D. L. (1998). The Method of Simulated Scores for the Estimation of LDV Models. *Econometrica*, 66(4):863–896.

Hartzel, J., Agresti, A., and Caffo, B. (2001). Multinomial Logit Random Effects Models. *Statistical Modelling*, 1(2):81–102.

Hijazi, R. H. and Jernigan, R. W. (2009). Modeling Compositional Data Using Dirichlet Regression Models. *Journal of Applied Probability & Statistics*, 4(1):77–91.

Jäckel, P. (2005). A Note on Multivariate Gauss-Hermite Quadrature. Unpublished Manuscript.

Kasteridis, P., Yen, S. T., and Fang, C. (2011). Bayesian Estimation of a Censored Linear Almost Ideal Demand System: Food Demand in Pakistan. *American Journal of Agricultural Economics*, 93(5):1374–1390.

Katz, J. N. and King, G. (1999). A Statistical Model for Multiparty Electoral Data. *American Political Science Review*, 93(1):15–32.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal Data Analysis using Generalized Linear Models. *Biometrika*, 73(1):13–22.

Lijoi, A., Mena, R. H., and Prünster, I. (2005). Hierarchical mixture modeling with normalized inverse-gaussian priors. *Journal of the American Statistical Association*, 100(472):1278–1291.

Liu, P., Yuen, K. C., Wu, L.-C., Tian, G.-L., and Li, T. (2020). Zero-One-Inflated Simplex Regression Models for the Analysis of Continuous Proportion Data. *Statistics and Its Interface*, 13(2):193–208.

Liu, Q. and Pierce, D. A. (1994). A Note on Gauss-Hermite Quadrature. *Biometrika*, 81(3):624–629.

Loudermilk, M. S. (2007). Estimation of Fractional Dependent Variables in Dynamic Panel Data Models With an Application to Firm Dividend Policy. *Journal of Business & Economic Statistics*, 25(4):462–472.

Magnac, T. (2004). Panel Binary Variables and Sufficiency: Generalizing Conditional Logit. *Econometrica*, 72(6):1859–1876.

Martín-Fernández, J. A., Barceló-Vidal, C., and Pawlowsky-Glahn, V. (2003). Dealing with Zeros and Missing Values in Compositional Data Sets using Nonparametric Imputation. *Mathematical Geology*, 35(3):253–278.

McCulloch, R. E., Polson, N. G., and Rossi, P. E. (2000). A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters. *Journal of Econometrics*, 99(1):173–193.

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica*, pages 995–1026.

Montoya-Blandón, S. (2021). Copula Estimation and Model Selection with Multivariate Fractional Outcomes. Working Paper.

Montoya-Blandón, S. and Jacho-Chávez, D. T. (2020). Semiparametric Quasi Maximum Likelihood Estimation of the Fractional Response Model. *Economics Letters*, 186:108769.

Morais, J., Thomas-Agnan, C., and Simioni, M. (2018). Using Compositional and Dirichlet Models for Market Share Regression. *Journal of Applied Statistics*, 45(9):1670–1689.

Mullahy, J. (2015). Multivariate Fractional Regression Estimation of Econometric Share Models. *Journal of Econometric Methods*, 4(1):71–100.

Mullahy, J. and Robert, S. A. (2010). No Time to Lose: Time Constraints and Physical Activity in the Production of Health. *Review of Economics of the Household*, 8(4):409–432.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46(1):69–85.

Murteira, J. M. R. and Ramalho, J. J. S. (2016). Regression Analysis of Multivariate Fractional Data. *Econometric Reviews*, 35(4):515–552.

Papke, L. E. and Wooldridge, J. M. (1996). Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. *Journal of Applied Econometrics*, 11(6):619–632.

Papke, L. E. and Wooldridge, J. M. (2008). Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates. *Journal of Econometrics*, 145(1-2):121–133.

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the Log-likelihood Function in the Nonlinear Mixed-effects Model. *Journal of Computational and Graphical Statistics*, 4(1):12–35.

Scealy, J. L. and Welsh, A. H. (2011). Regression for Compositional Data by Using Distributions Defined on the Hypersphere. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3):351–375.

Stewart, C. and Field, C. (2011). Managing the Essential Zeros in Quantitative Fatty Acid Signature Analysis. *Journal of Agricultural, Biological, and Environmental Statistics*, 16(1):45–69.

Tsagris, M., Preston, S., and Wood, A. (2011). A Data-based Power Transformation for Compositional Data. In *Proceedings of the 4rth Compositional Data Analysis Workshop, Girona, Spain.*

Tsagris, M. and Stewart, C. (2018). A Dirichlet Regression Model for Compositional Data with Zeros. *Lobachevskii Journal of Mathematics*, 39(3):398–412.

Wales, T. J. and Woodland, A. D. (1983). Estimation of Consumer Demand Systems with Binding Non-negativity Constraints. *Journal of Econometrics*, 21(3):263–285.

White, H. (1981). Consequences and Detection of Misspecified Nonlinear Regression Models. *Journal of the American Statistical Association*, 76(374):419–433.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 50(1):1–25.

Woodland, A. D. (1979). Stochastic Specification and the Estimation of Share Equations. *Journal of Econometrics*, 10(3):361–383.

Wooldridge, J. M. (2005). Unobserved Heterogeneity and Estimation of Average Partial Effects. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 27–55. Cambridge University Press.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, second edition edition.

A Details on Integration Methods for MLE

The integrals given by the conditionally independent (7) and pooled (8) likelihoods can be cast in a general way as the problem of numerically evaluating the following integral for some function $f : \mathbb{R}^D \times \mathbb{R}^p \to \mathbb{R}$:

$$V \equiv \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\boldsymbol{c}, \boldsymbol{z}) \phi_D(\boldsymbol{c}; \mathbf{0}_{D\times 1}, \boldsymbol{\Gamma}) \, \mathrm{d}\boldsymbol{c} \, , \tag{24}$$

where $\boldsymbol{z} \in \mathbb{R}^p$ represents other possible arguments to the function. From Liu and Pierce (1994), recall that the Gauss-Hermite quadrature allows one to evaluate the one-dimensional integral

$$\int_{-\infty}^{\infty} g(c, \boldsymbol{z}) \exp\left(-c^2\right) \mathrm{d}c \approx \sum_{s=1}^{S} w_s g(a_s, \boldsymbol{z}) \,, \tag{25}$$

where $g : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}$, the abscissas a_s denote the zeros of the S-th order Hermite polynomial and w_s are their corresponding weights.

A.1 Adaptive Quadrature

The adaptive approach to evaluate the multidimensional integral in (24) begins by transforming the integrand as

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\frac{f(\boldsymbol{c}, \boldsymbol{z}) \phi_D(\boldsymbol{c}; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma})}{\phi_D(\boldsymbol{c}; \boldsymbol{\omega}, \boldsymbol{Q})} \right] \phi_D(\boldsymbol{c}; \boldsymbol{\omega}, \boldsymbol{Q}) \, \mathrm{d}\boldsymbol{c} \,,$$

By a substitution $\boldsymbol{u} = (2\boldsymbol{Q})^{-1/2}(\boldsymbol{c} - \boldsymbol{\omega})$, this integral becomes

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} 2^{\frac{D}{2}} |\boldsymbol{Q}|^{\frac{1}{2}} \exp\left(\boldsymbol{u}'\boldsymbol{u}\right) f(\boldsymbol{c}(\boldsymbol{u}), \boldsymbol{z}) \phi_D(\boldsymbol{c}(\boldsymbol{u}); \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma}) \exp\left(-\boldsymbol{u}'\boldsymbol{u}\right) \mathrm{d}\boldsymbol{u} \,,$$

where $\boldsymbol{c}(\boldsymbol{u}) = \boldsymbol{\omega} + \sqrt{2}\boldsymbol{Q}^{1/2}\boldsymbol{u}$, $\boldsymbol{Q}^{1/2}$ is the matrix resulting from a Cholesky decomposition of $\boldsymbol{Q}$ and $|\boldsymbol{Q}|$ is the determinant of $\boldsymbol{Q}$. Defining the function $h(\boldsymbol{c}) = \log f(\boldsymbol{c}, \boldsymbol{z}) + \log \phi_D(\boldsymbol{c}; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma})$, the adaptive approach estimates $\boldsymbol{\omega}$ and $\boldsymbol{Q}$ as the mode and curvature at the mode, respectively, of $h(\boldsymbol{c})$; i.e.,

$$\begin{aligned} \widehat{\boldsymbol{\omega}} &= \arg\max_{\boldsymbol{c}} h(\boldsymbol{c}) \,, \\ \widehat{\boldsymbol{Q}} &= \left. \frac{\partial^2 h(\boldsymbol{c})}{\partial \boldsymbol{c} \partial \boldsymbol{c}'} \right|_{\boldsymbol{c}=\widehat{\boldsymbol{\omega}}} . \end{aligned}$$

Given that $f(\cdot)$ is taken to be a (potentially misspecified) distribution for the multivariate fractions $\boldsymbol{Y}$, then $\widehat{\boldsymbol{\omega}}$ can be interpreted as the posterior mode of $\boldsymbol{c}$ using likelihood f and a Gaussian prior centered at 0. As noted by Liu and Pierce (1994), these estimators ensure that the log of the chosen Gaussian density has the same scores and Hessian as $f(\boldsymbol{c}, \boldsymbol{z})\phi_D(\boldsymbol{c}; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma})$. It is in this sense that this method is adaptive to the specific integrand.

Let $\boldsymbol{a}_s = (a_{s_1}, \ldots, a_{s_D})$ and compute $\boldsymbol{a}_s^* = \widehat{\boldsymbol{\omega}} + \sqrt{2}\widehat{\boldsymbol{Q}}^{(1/2)}\boldsymbol{a}_s$. As $\exp(-\boldsymbol{u}'\boldsymbol{u}) = \exp\left(-u_1^2\right) \times \cdots \times \exp\left(-u_D^2\right)$, we can apply the univariate Gauss-Hermite quadrature process D times to solve the multivariate integral yielding

$$V_{\text{adaptive}} \approx 2^{\frac{D}{2}} |\widehat{\boldsymbol{Q}}|^{\frac{1}{2}} \sum_{s_1=1}^{S} \cdots \sum_{s_D=1}^{S} \prod_{j=1}^{D} w_{s_j} \exp\left(\boldsymbol{a}_s'\boldsymbol{a}_s\right) f(\boldsymbol{a}_s^*, \boldsymbol{z}) \phi_D(\boldsymbol{a}_s^*; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma}) \tag{26}$$

A.2 Nonadaptive Quadrature

This method operates by noting that, since we are already starting from a function times a Gaussian density in (24), we only need to deal with the correlation between unobserved heterogeneity values before using Gauss-Hermite quadrature in each dimension. While there is no generally best way of incorporating this correlation structure into the Gauss-Hermite procedure, Jäckel (2005) describes one of the most numerically robust methods as follows. Using a singular value decomposition, find $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ such that of $\boldsymbol{\Gamma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'$. By a similar substitution to before, define $\boldsymbol{u} = \boldsymbol{R}'(2\boldsymbol{\Lambda})^{-1/2}\boldsymbol{U}'\boldsymbol{c}$, where $\boldsymbol{R}$

is the resulting matrix from multiplying together $(D-1)$ planar rotation matrices of $45°$ degrees each. Then, (24) becomes

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi^{-\frac{D}{2}} f(\boldsymbol{c}(\boldsymbol{u}), \boldsymbol{z}) \exp\left(-\boldsymbol{u}'\boldsymbol{u}\right) \mathrm{d}\boldsymbol{u} \ ,$$

with $\boldsymbol{c}(\boldsymbol{u}) = \sqrt{2}\boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{R}\boldsymbol{u}$. This time, compute $\boldsymbol{a}_s^* = \sqrt{2}\boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{R}\boldsymbol{a}_s$. Thus, the desired approximation is given by

$$V_{\text{nonadaptive}} \approx \pi^{-\frac{D}{2}} \sum_{s_1=1}^{S} \cdots \sum_{s_D=1}^{S} \prod_{j=1}^{D} w_{s_j} f(\boldsymbol{a}_s^*, \boldsymbol{z}) \, . \tag{27}$$

A.3 Pruning

One final issue that is of interest for the computation of both (26) and (27) is the use of pruning. Since some of the evaluation points will be given very small weights that might not contribute much to the value of the integral, one can set these to 0 and decrease the amount of function evaluations needed without sacrificing much precision. As the individual weights are always multiplied together for any approximation, set $\boldsymbol{w}_s = \prod_{j=1}^{D} w_{s_j}$. Given a threshold τ_S, the idea of pruning is to use weights

$$\boldsymbol{w}_s^* = \boldsymbol{w}_s I(\boldsymbol{w}_s > \tau_S) \, ,$$

in each evaluation. While τ_S can be chosen to be any arbitrary constant designed to reduce computational intensity without sacrificing numerical precision, Jäckel (2005) recommends using

$$\tau_S = \min_s\{\boldsymbol{w}_s\}^{D-1} \cdot \max_s\{\boldsymbol{w}_s\} \, .$$

This is the value that I use throughout the paper for all integral evaluations.

B Derivatives for MLE and Probit Estimators

B.1 Scores for Independent and Pooled MLE

Starting from (7) or (8), replace the multinomial logit link (12) into $F(\boldsymbol{Y}_{it}|\boldsymbol{X}_{it}, \boldsymbol{c}_i; \boldsymbol{\beta})$ and take logs to obtain

$$\log F(\boldsymbol{Y}_{it}|\boldsymbol{X}_{it}, \boldsymbol{c}_i; \boldsymbol{\beta}) = \sum_{j=1}^{d} y_{tij} \left[\boldsymbol{x}_{itj}'\boldsymbol{\beta}_j + c_{ij} - \log\left(1 + \sum_{p=1}^{D} \exp\left(\boldsymbol{x}_{itp}'\boldsymbol{\beta}_p + c_{ip}\right)\right)\right] \, .$$

Differentiating this equation with respect to some $\boldsymbol{\beta}_k$ yields the usual multinomial score

$$\begin{aligned} \frac{\partial \log F(\boldsymbol{Y}_{it}|\boldsymbol{X}_{it}, \boldsymbol{c}_i; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k} &= \sum_{j=1}^{d} y_{tij}[I(j=k) - m_{itk}]\boldsymbol{x}_{itk} \, , \\ &= (y_{itk} - m_{itk})\boldsymbol{x}_{itk} \, , \end{aligned}$$

where the last step follows from $\boldsymbol{Y}_{it} \in \mathcal{S}^d$. We now have the derivative that would apply to the logarithm of the integrand. Exchanging differentiation and integration, we then have

$$\begin{aligned} \frac{\partial \ell_i^{(\text{ind})}(\boldsymbol{\beta}, \boldsymbol{\Gamma})}{\partial \boldsymbol{\beta}_k} = L_i^{(\text{ind})}(\boldsymbol{\beta}, \boldsymbol{\Gamma}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Bigg\{ \left[\prod_{t=1}^{T_i} \prod_{j=1}^{d} m_{ijt}^{y_{ijt}}\right] \left[\sum_{t=1}^{T_i} (y_{itk} - m_{itk})\boldsymbol{x}_{itk}\right] \\ \times \phi_D(\boldsymbol{c}_i; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma}) \Bigg\} \mathrm{d}\boldsymbol{c}_i \ , \end{aligned} \tag{28}$$

for the likelihood assuming conditional independence and

$$\frac{\partial \ell_i^{(\text{pool})}(\boldsymbol{\beta}, \boldsymbol{\Gamma})}{\partial \boldsymbol{\Gamma}} = \sum_{t=1}^{T_i} L_{it}^{(\text{pool})}(\boldsymbol{\beta}, \boldsymbol{\Gamma}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \left[\prod_{j=1}^{d} m_{ijt}^{y_{ijt}} \right] (y_{itk} - m_{itk}) \boldsymbol{x}_{itk} \times \phi_D(\boldsymbol{c}_i; \mathbf{0}_{D\times 1}, \boldsymbol{\Gamma}) \right\} \mathrm{d}\boldsymbol{c}_i \,, \tag{29}$$

for the pooled likelihood. The terms $L_i^{(\text{ind})}(\boldsymbol{\beta}, \boldsymbol{\Gamma})$ and $L_{it}^{(\text{pool})}(\boldsymbol{\beta}, \boldsymbol{\Gamma})$ represent the likelihood before taking logarithms; i.e., the complete integrals. Stacking across all $k = 1, \ldots, D$ yields the total score. The scores for $\boldsymbol{\Gamma}$ are similar and rely on the score for the normal distribution and the matrix derivatives of $\boldsymbol{\Gamma}$. They are given as

$$\frac{\partial \ell_i^{(\text{ind})}(\boldsymbol{\beta}, \boldsymbol{\Gamma})}{\partial \boldsymbol{\Gamma}} = L_i^{(\text{ind})}(\boldsymbol{\beta}, \boldsymbol{\Gamma}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \left[\prod_{t=1}^{T_i} \prod_{j=1}^{d} m_{ijt}^{y_{ijt}} \right] \boldsymbol{\Gamma}^{-1}(I_D - \boldsymbol{c}_i \boldsymbol{c}_i' \boldsymbol{\Gamma}^{-1}) \times \phi_D(\boldsymbol{c}_i; \mathbf{0}_{D\times 1}, \boldsymbol{\Gamma}) \right\} \mathrm{d}\boldsymbol{c}_i \,, \tag{30}$$

for the likelihood assuming conditional independence and

$$\frac{\partial \ell_i^{(\text{pool})}(\boldsymbol{\beta}, \boldsymbol{\Gamma})}{\partial \boldsymbol{\beta}_k} = \sum_{t=1}^{T_i} L_{it}^{(\text{pool})}(\boldsymbol{\beta}, \boldsymbol{\Gamma}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \left[\prod_{j=1}^{d} m_{ijt}^{y_{ijt}} \right] \boldsymbol{\Gamma}^{-1}(I_D - \boldsymbol{c}_i \boldsymbol{c}_i' \boldsymbol{\Gamma}^{-1}) \times \phi_D(\boldsymbol{c}_i; \mathbf{0}_{D\times 1}, \boldsymbol{\Gamma}) \right\} \mathrm{d}\boldsymbol{c}_i \,, \tag{31}$$

for the pooled likelihood.

B.2 Score and Hessian for Probit NLS

Starting from the objective function (15), we see that it can be written as a summation across both t and j, such that

$$q_i(\boldsymbol{\alpha}_c) = \frac{1}{2} \sum_{t=1}^{T_i} \sum_{j=1}^{D} [y_{itj} - \Phi(\tilde{\boldsymbol{x}}_{itj}' \boldsymbol{\alpha}_{jc})]^2 \,.$$

Taking the derivative with respect to some $\boldsymbol{\alpha}_{kc}$ yields

$$\frac{\partial q_i(\boldsymbol{\alpha}_c)}{\partial \boldsymbol{\alpha}_{kc}} = -\sum_{t=1}^{T_i} \phi(\tilde{\boldsymbol{x}}_{itk}' \boldsymbol{\alpha}_{kc}) [y_{itk} - \Phi(\tilde{\boldsymbol{x}}_{itk}' \boldsymbol{\alpha}_{kc})] \tilde{\boldsymbol{x}}_{itk} \,.$$

Stacking across $k = 1, \ldots, D$ gives the score as

$$\frac{\partial q_i(\boldsymbol{\alpha}_c)}{\partial \boldsymbol{\alpha}_c} = -\sum_{t=1}^{T_i} \begin{bmatrix} \phi(\tilde{\boldsymbol{x}}_{it1}' \boldsymbol{\alpha}_{1c}) [y_{it1} - \Phi(\tilde{\boldsymbol{x}}_{it1}' \boldsymbol{\alpha}_{1c})] \tilde{\boldsymbol{x}}_{it1} \\ \vdots \\ \phi(\tilde{\boldsymbol{x}}_{itD}' \boldsymbol{\alpha}_{Dc}) [y_{itD} - \Phi(\tilde{\boldsymbol{x}}_{itD}' \boldsymbol{\alpha}_{Dc})] \tilde{\boldsymbol{x}}_{itD} \end{bmatrix} . \tag{32}$$

Note that each element depends only on its respective coefficient and so $\partial^2 q_i(\boldsymbol{\alpha}_c) / \partial \boldsymbol{\alpha}_{kc} \partial \boldsymbol{\alpha}_{jc} = 0$ for $j \neq k$. This then implies that the Hessian will be a diagonal matrix. Taking another derivative with

respect to some $\boldsymbol{\alpha}_{kc}$ and using $\mathrm{d}\phi(z)/\mathrm{d}z = -z\phi(z)$ for any $z \in \mathbb{R}$, we have that each diagonal term will be of the form

$$\frac{\partial^2 q_i(\boldsymbol{\alpha}_c)}{\partial \boldsymbol{\alpha}_{kc} \partial \boldsymbol{\alpha}_{kc}} = \sum_{t=1}^{T_i} \phi(\tilde{\boldsymbol{x}}'_{itk}\boldsymbol{\alpha}_{kc})\{\phi(\tilde{\boldsymbol{x}}'_{itk}\boldsymbol{\alpha}_{kc}) + \tilde{\boldsymbol{x}}'_{itk}\boldsymbol{\alpha}_{kc}[y_{itk} - \Phi(\tilde{\boldsymbol{x}}'_{itk}\boldsymbol{\alpha}_{kc})]\}\tilde{\boldsymbol{x}}_{itk}\tilde{\boldsymbol{x}}'_{itk}\,, \tag{33}$$

for all $k = 1, \ldots, D$.