Distribution Agreement

Flexible Estimation Methods for Multivariate Fractional Outcomes

By

Santiago Montoya Blandon

Doctor of Philosophy

Economics

David Jacho-Chavez
Advisor

Elena Pesavento
Committee Member

Christoph Breunig
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Flexible Estimation Methods for Multivariate Fractional Outcomes

By

Santiago Montoya Blandon

M.A., Emory University, 2020
M.Sc., Universidad EAFIT, 2015
B.A., Universidad EAFIT, 2013

Advisor: David Jacho-Chavez, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Economics

2021

Abstract

Flexible Estimation Methods for Multivariate Fractional Outcomes

By Santiago Montoya Blandon

Multivariate fractional outcomes are defined as vectors where each component is bounded to the unit interval and together they add up to 1. This dissertation expands the available toolkit for analyzing both univariate and multivariate fractional outcomes as well as their applications to economics and other fields. As these variables arise naturally in several areas of applied microeconomics, the focus is on cross-sectional and panel data. Emphasis is placed on providing methods that are flexible and robust while exploring several approaches to modeling of these outcomes in a variety of settings. In each chapter a different facet of multivariate fractional outcomes is studied. The first chapter presents a semiparametric extension of a quasi-likelihood estimator that is heavily used in applications with a univariate fractional outcome. As documented in the chapter, large biases can arise when the nonlinear link function is misspecified, which can be countered by the use of our extension. The second chapter provides a unified estimation methodology using copulas for multivariate fractional outcomes with a conditional mean specification. This methodology satisfies the fractional and unit-sum constraints of the outcomes, allows for cross-equation restrictions that are crucial in structural estimation, and can handle variable selection. The final chapter extends both the existing and newly proposed methods to a panel data setting, focusing on several robust alternatives and their numerical implementations. All chapters use simulation exercises and applications to showcase the performance of the proposed methods.

Flexible Estimation Methods for Multivariate Fractional Outcomes

By

Santiago Montoya Blandon

M.A., Emory University, 2020
M.Sc., Universidad EAFIT, 2015
B.A., Universidad EAFIT, 2013

Advisor: David Jacho-Chavez, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Economics

2021

Acknowledgments

The journey towards the completion of my PhD and my formation as a researcher has only been fruitful thanks to the incredible support I have received from many wonderful people along the way. Advisors, professors, colleagues, relatives, partners, friends, and many others were instrumental in this achievement and I want to give my sincerest appreciation for everyone that was part of it in any way. While this is by no means an exhaustive list, I hope to highlight several people whose contributions were invaluable.

Thank you David, my advisor and mentor, for helping me as he does many others in taking the first steps in the PhD program and the profession as a whole. You were my most constant source of support and I would not be where I am without your advice in all matters professional and personal. I will carry all your teachings with me throughout my career and hopefully I will pass them on to my own students in due time.

Thank you Elena for your insightful comments and honesty. Your help in all facets of research, and specially writing and presenting, has had a deep impact in the way I approach my work. You were also an integral part of my job market experience and I do not believe I could have gotten a job, let alone a great job like the one I was able to get, without everything you taught me. Thank you also to Christoph and Sara for their comments and constant guidance.

Thank you Stephanie, Renee, Marie, Elizabeth and the rest of the Economics department staff for making daily life simple and sorting with marvelous speed through my many administrative requests.

Thanks to my parents, Alonso and Marleny, for being my rocks and my first Econ professors. Everything I have accomplished is thanks to the foundation you laid for us. You taught me to trace my own life path, and this dissertation is but the culmination of one of those steps. I also want to thank Juliana, my sister, for showing me all the ways to enjoy life while being true to yourself. I could not have asked for a better hermanita.

Thank you Nicol for what seems like an infinite amount of love. We circled around each other during life but I am happy that we found each other, specially during this important and difficult last stage of my doctoral life. Thank you for putting up with the almost 3,200 kilometers that separated us. Te amo.

Thank you Daniela for being my anchor to Medellín, for our late-night conversations about the universe, and for teaching me so much about life. I hope we can keep supporting each other as we go on. Thank you Carla as well, I will always treasure our roller coaster rides. You and Dani were the core of my Atlanta family and I keep everything we went through in my heart.

Thank you Juan for all of our adventures and all the amazing work I know we will accomplish together. The passion you show for our profession is one of the main reasons I am excited to continue moving forward. Thank you to Kaylyn as well, the honorary Latina that supported us through so much. Thank you Diego, whom I always felt understood me and with whom I share many passions in common. Thank you Yiqing and Sasha, I am happy we survived the PhD together.

Thank you Andrés, Oliver, Lina, Fernando, Mónica and all other Colombians I met in Atlanta for being that small piece of home away from home. We shared amazing experiences together that I will always cherish and we also held the best dance parties Atlanta has ever seen. In that same spirit I want to take the members Latinx Graduate Student Association, specially Juan, Taína, Sandra and Ana María. I truly believe we provided a necessary place for Latinx students and Spanish speaking students in general to come together and practice our heritage. Within the incredibly diverse group of Emory students, I also want to thank Lenore, Marisol, Erick, Sofia, Massi, Alessandro, along many others.

Thank you to my dear friends in Colombia, Sara, Pipe, Davi, D, Mike, Elo, San, and many more for your support when I left and celebrating every time I came back. Despite the distance, I could not have made it out the other side with my sanity intact without you.

Contents

List of Figures

List of Tables

Introduction

The analysis of multivariate fractional outcomes $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$ is prevalent in several fields such as biology, chemistry, economics, geology, and others (Aitchison, 2003; Kieschnick and McCullough, 2003). The nature of the outcomes implies that they are both fractional (i.e., bounded between 0 and 1) and satisfy a unit-sum constraint across the d shares. These types of observations are known as compositional data in the statistics literature and are characterized as belonging to the d-dimensional simplex

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

Fractional outcomes arise naturally in economic applications when estimating a demand system in which the dependent variables are given as expenditure shares on d different categories of goods (Woodland, 1979; Barnett and Serletis, 2008). They are also central in other contexts such as in finance, where they can represent portfolio shares allocated to different stocks (Glassman and Riddick, 1994; Stavrunova and Yerokhin, 2012; Mullahy, 2015), in industrial organization and management when discussing market shares for different companies within a given industry (Morais et al., 2018), or in social choice when analyzing voting patterns in elections with several candidates (Katz and King, 1999). Other applications for these outcomes include time of use in health production functions (Mullahy and Robert, 2010), dividends and firm analysis (Loudermilk, 2007; Ramalho and Silva, 2009; Sosa, 2009; Sigrist and Stahel, 2011), psychology (Smithson and Verkuilen, 2006; Johnson and Mislin, 2011), among others.

This dissertation focuses on cross-sectional and panel data settings, as most analysis involving multivariate fractional outcomes rely on such data structures, leaving aside most time series concerns for future research.[1] The concepts are addressed in ascending level of complexity with respect to the outcome of interest. That is, the first chapter addresses a method for a univariate fractional outcome in a cross-sectional setting, the second chapter focuses on multivariate systems of fractions again within the cross-section, and the final chapter on multivariate fractional outcomes in panel data.

Across the three chapters several estimation methods are introduced and emphasis is placed on both flexibility and robustness. The first chapter presents a semiparametric extension of the robust estimator introduced by Papke and Wooldridge (1996). As documented within this chapter, when the link function for the conditional expectation is misspecified, a situation that can easily occur in practice, large biases in estimating the conditional mean parameters are bound to arise. In order to avoid such biases, a nonparametric kernel estimate of the link function is paired with a quasi-likelihood approach to obtain estimates of the conditional mean parameters. In essence, by consistently estimating a link function instead of assuming it known, we are able to avoid the biases associated to misspecification. While this creates a more computationally intensive method, we show that it produces sensible results in both simulations and an empirical application, while inference remains largely unaffected.

Computational considerations largely prevent a working version of this semiparametric estimator in a multivariate setting, although such extension would certainly be possible. Heading in another direction, the second chapter introduces a more general parametric framework using copulas to study models for fractional outcomes that arise in both structural and reduced form microeconometric approaches. Within this framework, the paper presents an estimation procedure that simultaneously accounts for the specific distributional concerns with multivariate fractional variables; the conditional mean structures that arise in many empirical models; and that allows for both variable selection and cross-equation restrictions that become necessary in certain structural scenarios. This approach yields several other features. First, the use of copulas allow for efficiency gains compared to other approaches while still accommodating a degree of robustness to dependence structure misspecification. Second, structural demand estimation models can create the need for variable selection, particularly in the presence of big data, which is taken into the ac-

[1] In Chapter 3 that assumes access to panel data, autocorrelation and other time series behavior is either accounted for by using standard errors robust to these possible patterns or directly modeled depending on the context.

count. Third, this variable selection is handled using a Bayesian approach using regularization that also guarantees correct inference. Finally, the paper presents a couple of technical contributions in parametric copula models that arise when proving the consistency and asymptotic normality of the resulting estimator.

The final chapter builds on the previous two and presents a comprehensive set of tools for the analysis of multivariate fractional outcomes in a panel data context, which requires dealing with unobserved heterogeneity in nonlinear models. It provides multivariate and panel extensions to methods that are previously available in the literature and to those introduced in this dissertation. Specifically, the paper presents several estimation procedures that should prove useful in different situations. First, a maximum likelihood method that in at least two special cases allows for identification and consistent estimation of conditional mean parameters and average partial effects. Second, a multivariate probit estimator that provides excellent approximations to the average partial effects, is computationally efficient, scales easily with the number of shares, and allows for endogeneity. Finally, to deal with censoring introduced by structural zeros in the data, this chapter introduces a Bayesian procedure using data augmentation. All these methods are tested in several numerical exercises that showcase their applicability and robustness in different scenarios.

Chapter 1

Semiparametric Quasi Maximum Likelihood Estimation of the Fractional Response Model

Note: The content in this chapter is reproduced from Montoya-Blandón, S., & Jacho-Chávez, D. T. (2020). "Semiparametric quasi maximum likelihood estimation of the fractional response model." *Economics Letters*, 186, 108769.

In the context of univariate fractional outcomes, this chapter proposes a kernel-based semiparametric quasi-maximum likelihood estimator (SPQMLE) which adapts Papke and Wooldridge's (1996) estimator to an unknown link function. The proposed adaptation inherits the nice properties of the original estimator, such as dealing with boundary values—where the response variable is allowed to take values exactly equal to 1 or 0—and it is robust to potential misspecification in the link function. Furthermore, the asymptotic properties are derived allowing for data-dependent smoothing parameters as well as possible random trimming. By deriving the exact formula of the asymptotic variance-covariance matrix for the proposed SPQMLE it is shown that there is no estimation effect from replacing the unknown link function by a consistent nonparametric kernel estimator.

A Monte Carlo experiment provides evidence that our method performs well in small-sample settings, and this performance is comparable to the performance achieved by a benchmark maximum likelihood estimation method (MLE) and a correctly specified quasi-likelihood method, but uniformly dominates methods with a misspecified link function. An empirical implementation of the proposed estimator utilizing data from Papke and Wooldridge (1996) is also included. Our point estimates are numerically smaller than those originally obtained in Papke and Wooldridge (1996) and closer to the baseline linear regression model.

The remainder of the paper is organized as follows: Section 1.1 introduces the estimator along with its asymptotic properties, Section 1.2 presents the results of our Monte Carlo simulation comparing our method with other suitable candidates, while Section 1.3 presents the results of our empirical application, and Section 1.4 concludes.

1.1 Estimator and Asymptotic Properties

1.1.1 Estimator

Assume one has access to an independent and identically distributed (i.i.d.) sample $\{\boldsymbol{y}_i', \boldsymbol{x}_i'\}_{i=1}^n$ from the joint distribution of $(\boldsymbol{Y}', \boldsymbol{X}')$ where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are k and d dimensional random vectors respectively. We will assume that $\boldsymbol{Y}$ takes values in $\mathcal{S}^2$. Note that in this case, one can focus the modeling strategy on one of the components of $\boldsymbol{Y}$ as the other will then be fully determined. Specifically, we will center our attention on $\boldsymbol{Y}^{(1)}$, which we will hereafter denote simply as Y. Given the characteristics of the data discussed before, we introduce the SPQMLE framework. Let the following index restriction holds almost surely (a.s.)

$$\mathrm{E}[Y_i|\boldsymbol{x}_i] = \mathrm{E}[Y_i|\boldsymbol{x}_i'\boldsymbol{\beta}_0] \equiv m(\boldsymbol{x}_i'\boldsymbol{\beta}_0) \tag{1.1}$$

for some $\boldsymbol{\beta}_0 \in \mathcal{B} \subset \mathbb{R}^p$ and $\boldsymbol{x}_i \in \mathcal{X} \subset \mathbb{R}^p$, where $\mathcal{X}$ represents the support of $\boldsymbol{X}$. We assume $f(\boldsymbol{x}|z)$ is the density of $\boldsymbol{X}$ conditional on $z = \boldsymbol{X}'\boldsymbol{\beta}$ with respect to a measure μ. Our estimator for $\boldsymbol{\beta}_0$ is based on the semiparametric quasi-likelihood function

$$\mathcal{L}_n(\boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{i=1}^{n}\{y_i \log[\widehat{m}(\boldsymbol{x}_i'\boldsymbol{\beta})] + (1-y_i)\log[1-\widehat{m}(\boldsymbol{x}_i'\boldsymbol{\beta})]\}\widehat{t}_{ni}\,, \tag{1.2}$$

where $\widehat{m}(\boldsymbol{x}_i'\boldsymbol{\beta})$ estimates the conditional mean $M(\boldsymbol{x}_i'\boldsymbol{\beta}) = \mathrm{E}[m(\boldsymbol{x}_i'\boldsymbol{\beta}_0)|\boldsymbol{x}_i'\boldsymbol{\beta}]$, using a (leave-one-out) Nadaraya-Watson estimator as $\widehat{m}(\boldsymbol{x}_i'\boldsymbol{\beta}) = \widehat{G}(\boldsymbol{x}_i'\boldsymbol{\beta})/\widehat{f}(\boldsymbol{x}_i'\boldsymbol{\beta})$, where $\widehat{G}(\boldsymbol{x}_i'\boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{j\neq i}^{n} y_j K_{\widehat{h}_n}(\boldsymbol{x}_j'\boldsymbol{\beta} - \boldsymbol{x}_i'\boldsymbol{\beta})$, $\widehat{f}(\boldsymbol{x}_i'\boldsymbol{\beta}) \equiv \frac{1}{n}\sum_{j\neq i}^{n} K_{\widehat{h}_n}(\boldsymbol{x}_j'\boldsymbol{\beta} - \boldsymbol{x}_i'\boldsymbol{\beta})$ with $K_h(v) = h^{-1}K(v/h)$, $K(\cdot)$ a kernel function, and $\widehat{h}_n$ a possibly data-dependent bandwidth. As the dependent variable in this setting is not binary but a fraction, the likelihood defined in (1.2) is inherently misspecified (even with a correctly specified fixed $m(\cdot)$ function), and thus consistent estimation is guaranteed by the index restriction in (1.1) and the conditions given in Theorem 1.1 (see Papke and Wooldridge, 1996, for possible optimality properties of this quasi-likelihood in the class of the linear exponential family). Let $\mathbb{I}\{\cdot\}$ be the indicator function that equals 1 when its argument is true, and 0 otherwise. Then, $\widehat{t}_{ni} \equiv \mathbb{I}\{\widehat{f}(\boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}}) \geq \tau_n\}$ is a trimming function based on a preliminary consistent estimator of $\boldsymbol{\beta}_0$, denoted by $\widetilde{\boldsymbol{\beta}}$, and $\tau_n \to 0$ as $n \to \infty$ at a rate satisfying Assumption 1.8 below. This estimator could be obtained, for example, by maximizing (1.2) using $\widehat{t}_{ni} = \mathbb{I}\{\boldsymbol{x}_i \in A\}$, where $A \in \mathcal{X}$ is a compact subset. The proposed estimator is then given by

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}\in\mathcal{B}} \mathcal{L}_n(\boldsymbol{\beta})\,. \tag{1.3}$$

1.1.2 Asymptotic Properties

We apply the results in Gourieroux et al. (1984) and Escanciano et al. (2014) to show that our estimator of $\boldsymbol{\beta}_0$ in (1.1) defined by (1.2)–(1.3) is consistent and asymptotically normal. We begin by listing the required assumptions, which set up the model and are needed to guarantee the properties of kernel estimated functions. Throughout, C will denote a generic positive constant.

Assumption 1.1. Identification of $\boldsymbol{\beta}_0$: (i) there are no constant elements in $\boldsymbol{x}$, (ii) the first element of $\boldsymbol{x}$, say x_1 is continuous and its associated component of $\boldsymbol{\beta}_0$, say $\beta_1 = 1$, and (iii) if $m(\boldsymbol{x}'\boldsymbol{\beta}_1) = m(\boldsymbol{x}'\boldsymbol{\beta}_2)$ a.s. (with respect to the measure μ) then $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ (these are standard in single index models, see for example Ichimura, 1993; Klein and Spady, 1993 and Li and Racine, 2007, pp. 251–253).

The following four assumptions are standard and limit the general set up (Assumptions 1.2–1.3), introduce a general rth-order kernel (Assumption 1.4) and control the bias present in the nonparametric estimations (Assumption 1.5).

Assumption 1.2. The observations $\{y_i, \boldsymbol{x}_i'\}_{i=1}^n$ are an i.i.d. sample from the joint distribution of $(Y, \boldsymbol{X}')$, satisfying $\mathrm{E}[|Y|^{2+\delta}|\boldsymbol{X} = \boldsymbol{x}] < \infty$ for almost all $x \in \mathcal{X}$ and some $\delta > 0$.

Assumption 1.3. $\mathcal{B}$ is a compact set, and $\boldsymbol{\beta}_0 \in int(\mathcal{B})$.

Assumption 1.4. The kernel function $K : \mathbb{R} \to \mathbb{R}$ is bounded, symmetric, twice continuously differentiable and satisfies: $\int K(v)dv = 1$, $\int v^l K(v)dv = 0$ for $0 < l < r$, and $\int |v^r K(v)|dv < \infty$ for some $r \geq 2$. Letting $d^{(j)}K(v)/dv^j$ denote the jth derivative of $K(\cdot)$, we further assume that for $j = 1, 2$, $|d^{(j)}K(v)/dv^j| \leq C$, and for some $s > 1$, $|d^{(j)}K(v)/dv^j| \leq C|v|^{-s}$ for $|v| > L_j$, $0 < L_j < \infty$.

Assumption 1.5. For all $\boldsymbol{\beta}$ and $\boldsymbol{x} \in \mathcal{X}$, $f(\boldsymbol{x}'\boldsymbol{\beta})$, $m(\boldsymbol{x}'\boldsymbol{\beta})$, and $f(\boldsymbol{x}|z)$ are r-times continuously differentiable in $z = \boldsymbol{x}'\boldsymbol{\beta}$, with all functions and derivatives being uniformly bounded.

Assumption 1.6. The possibly data-dependent bandwidth $\widehat{h}_n$ satisfies $P_n(a_n \leq \widehat{h}_n \leq b_n) \to 1$ as $n \to \infty$, for deterministic sequences of positive numbers a_n and b_n such that $b_n \to 0$, $b_n^{2r} n \to 0$ and $a_n^3 n / \log n \to \infty$, for r as given by Assumption 1.4.

The final assumptions adapt those in Escanciano et al. (2014) (specifically, see their assumptions 5, B.7, B.8, and C.1) to guarantee uniform convergence of the estimated functions and their derivatives while allowing for data-dependent bandwidths such as those obtained by plug-in rules and cross-validation (Andrews, 1995), as well as deal with random trimming. Let $\widehat{t}_{ni} \equiv \mathbb{I}\{\boldsymbol{x}_i \in \widehat{\mathcal{X}}_n\}$ represent a trimming function where $\widehat{\mathcal{X}}_n \subset \mathcal{X}$ could potentially be the result of an estimation procedure, such as a subset based on values of $\widehat{f}$. Let $\mathcal{X}_n$ represent a deterministic set and define $t_{ni} \equiv \mathbb{I}\{\boldsymbol{x}_i \in \mathcal{X}_n\}$, as well as the rate $d_n \equiv (\max\{\log 1/a_n, \log\log n\}/a_n n)^{1/2} + b_n^r$.

Assumption 1.7. The following two conditions are satisfied: (i) there is a sequence τ_n of positive numbers satisfying $\tau_n \leq \inf_{\boldsymbol{\beta} \in \mathcal{B}, \boldsymbol{x} \in \mathcal{X}_n} f(\boldsymbol{x}'\boldsymbol{\beta})$, $d_n^4 n/\tau_n^6 \to 0$ and $d_n/\tau_n \to 0$; and (ii) $P_n(\boldsymbol{X}_i \in \mathcal{X}_n) \to 1$ as $n \to \infty$ and $\mathrm{E}[|\widehat{t}_{ni} - t_{ni}|] = o(n^{-1/2})$.

Finally, in order to ensure that the estimated conditional mean asymptotically belongs to a sufficiently well-behaved class, we can further introduce $d_{mn} \equiv (\max\{\log 1/a_n, \log\log n\}/a_n^3 n)^{1/2}$.

Assumption 1.8. The rate d_{mn} is such that $d_{mn} = O(1)$.

The main result of the paper is summarized by the following theorem (a corresponding outline for the proof can be found in the supplemental material)

Theorem 1.1. *Given Assumptions 1.1–1.8,* $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$ *and* $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1})$*, where*

$$\boldsymbol{A} = \mathrm{E}\left\{ \frac{m'(\boldsymbol{X}'\boldsymbol{\beta}_0)^2}{m(\boldsymbol{X}'\boldsymbol{\beta}_0)[1 - m(\boldsymbol{X}'\boldsymbol{\beta}_0)]} (\boldsymbol{X} - \mathrm{E}[\boldsymbol{X}|\boldsymbol{X}'\boldsymbol{\beta}_0])(\boldsymbol{X} - \mathrm{E}[\boldsymbol{X}|\boldsymbol{X}'\boldsymbol{\beta}_0])' \right\}, \tag{1.4}$$

$$\boldsymbol{B} = \mathrm{E}\left\{ \left(\frac{[y_i - m(\boldsymbol{X}'\boldsymbol{\beta}_0)]m'(\boldsymbol{X}'\boldsymbol{\beta}_0)}{m(\boldsymbol{X}'\boldsymbol{\beta}_0)[1 - m(\boldsymbol{X}'\boldsymbol{\beta}_0)]} \right)^2 (\boldsymbol{X} - \mathrm{E}[\boldsymbol{X}|\boldsymbol{X}'\boldsymbol{\beta}_0])(\boldsymbol{X} - \mathrm{E}[\boldsymbol{X}|\boldsymbol{X}'\boldsymbol{\beta}_0])' \right\}. \tag{1.5}$$

Notice that, although semiparametric estimation introduces a correction term when compared to the parametric case, equations (1.4) and (1.5) show there is no estimation effect from replacing the unknown link function with a consistent estimator as in Ichimura (1993) and Klein and Spady (1993).

Following Theorem 1.1, one can estimate the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$ as follows: define $\widehat{y}_i = \widehat{m}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}})$, $\widehat{u}_i = y_i - \widehat{y}_i$, and $\widehat{g}_i = \widehat{m}'(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}})$. Obtain $\widehat{\boldsymbol{x}}_i = \widehat{\mathrm{E}}[\boldsymbol{X}_i|\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}]$ using a Nadaraya-Watson kernel estimator and define $\widetilde{\boldsymbol{x}}_i = \boldsymbol{x}_i - \widehat{\boldsymbol{x}}_i$. An estimate of the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$ is given by $\widehat{\mathrm{Asy.\ Var}}(\widehat{\boldsymbol{\beta}}) = \widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{A}}^{-1}$, where

$$\widehat{\boldsymbol{A}} = \sum_{i=1}^{n} \frac{\widehat{g}_i^2}{\widehat{y}_i(1 - \widehat{y}_i)} \widetilde{\boldsymbol{x}}_i \widetilde{\boldsymbol{x}}_i' \quad \text{and} \quad \widehat{\boldsymbol{B}} = \sum_{i=1}^{n} \left[\frac{\widehat{u}_i \widehat{g}_i}{\widehat{y}_i(1 - \widehat{y}_i)} \right]^2 \widetilde{\boldsymbol{x}}_i \widetilde{\boldsymbol{x}}_i'. \tag{1.6}$$

1.2 Monte Carlo Experiment

The following simulation study is conducted. The true values for the coefficients were set at $\boldsymbol{\beta}_0 = (1, \beta) = (1, -0.5)$ so as to satisfy the identification restrictions. Two covariates were generated from a $\mathcal{N}(0, 1)$ distribution using sample sizes of $n \in \{100, 200, 400, 800\}$. To generate fractional responses satisfying (1.1), the response variable is drawn as $y_i \sim \mathrm{Beta}(m(\boldsymbol{x}_i'\boldsymbol{\beta}_0)\phi, [1 - m(\boldsymbol{x}_i'\boldsymbol{\beta}_0)]\phi)$ for $i = 1, \ldots, n$, where $m(\cdot)$ is the Logit link (Ferrari and Cribari-Neto, 2004; Simas et al., 2010). We generate data for several variance configurations given by the precision parameter $\phi \in \{1, 5, 25, 50, 100\}$. Small values of ϕ allow us to introduce bimodality in the distribution of y_i, as well as instances where $y_i = 0$ or $y_i = 1$, for which standard methods can fail.

As a benchmark, we use the beta regression method of Simas et al. (2010); a correctly specified MLE. This methodology allows for analytically correct standard errors, as well as estimation of

ϕ (results are available upon request). We compare performance with our method and four other estimators, which implement the quasi-maximum likelihood (QMLE) methodology of Papke and Wooldridge (1996) with different link functions.

We simulate 1,000 data sets for each sample size and variance configuration. We focus our results on β as the only free parameter in our semiparametric estimation. The results of our simulation exercise are given in Table 1.1 and Figure 1.1. Table 1.1 presents the ratios of bias and standard errors of all estimators with respect to the MLE benchmark. At modest sample sizes and variance levels, as those with $\phi = 25$ and $n = 400$, our method comes within 20% of the benchmark in all performance measures. As expected, the correctly specified Logit link has remarkable performance. This is in contrast to misspecified methods which perform poorly and remain biased regardless of the sample size. We also note that inference is not greatly affected by our semiparametric method: in some cases, standard errors can get to within 7% of those produced by the benchmark method.

Figure 1.1 gives a representation of the asymptotic normality approximation for estimators of β. We observe how the estimator's distribution grows closer to its asymptotic limit as the sample size increases, for all variance configurations. Similar processes occur for the correctly specified link estimator and the benchmark. The same cannot be said for the misspecified models, which fail to correctly center and scale the distribution.

1.3 Empirical Application

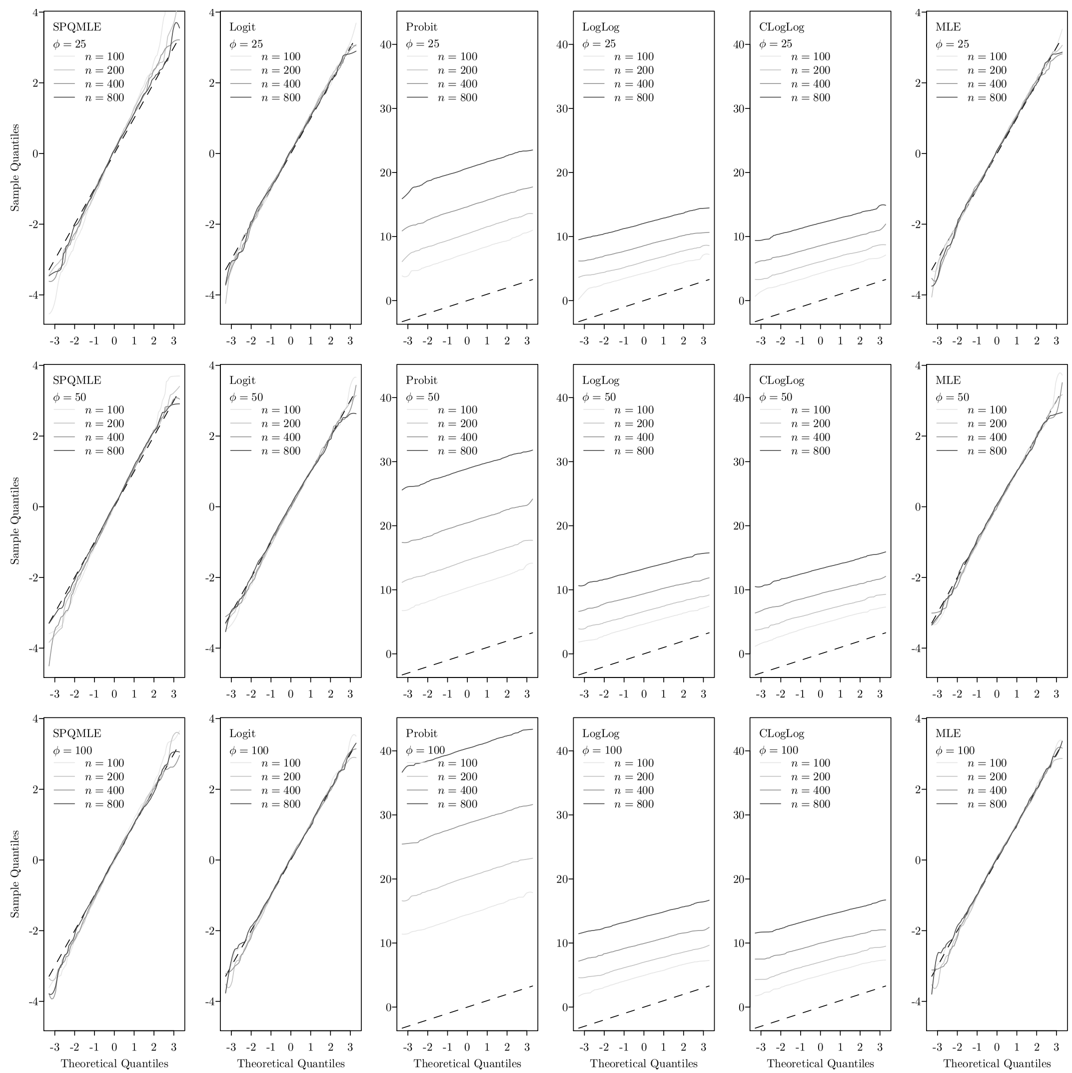
This section reassesses the model in Papke and Wooldridge (1996) using the new SPQMLE introduced in this paper. The authors use plan-level data on 401k accounts to estimate the effect of the match rate (percentage of employees' contributions matched by the firm) on the participation rate of each plan (ratio of eligible to enrolled employees). Due to institutional considerations, the match rate is not limited to 1 in the data set. The authors consider two separate estimations, either restricting the sample by match rate or keeping the full sample; we only present the latter here (restricted sample estimation can be found in the supplementary material subsection 1.C). To control for plan and firm characteristics, the authors include as covariates the log of total firm employment, age of the plan, their squares, and an indicator for whether the 401k was the sole plan offered by the firm. The authors also show that non-linearities in the match rate are important

Table 1.1: Ratios of Root Mean Squared Errors (RMSE) and Standard Errors for Estimators of β

		$\phi = 1$		$\phi = 5$		$\phi = 25$		$\phi = 50$		$\phi = 100$	
Estimator	n	RMSE	Std. Error	RMSE	Std. Error	RMSE	Std. Error	RMSE	Std. Error	RMSE	Std. Error
SPQMLE	100	2.142	1.194	1.709	1.072	1.410	1.074	1.245	1.106	1.261	1.126
	200	2.191	1.245	1.640	1.097	1.223	1.069	1.167	1.087	1.161	1.102
	400	1.915	1.323	1.351	1.118	1.190	1.075	1.179	1.080	1.131	1.082
	800	1.821	1.354	1.350	1.131	1.166	1.077	1.163	1.073	1.114	1.079
Logit	100	1.380	1.344	1.102	1.076	1.018	1.001	1.003	0.994	1.000	0.992
	200	1.409	1.352	1.094	1.082	1.022	1.008	1.007	0.997	0.999	1.000
	400	1.372	1.353	1.093	1.083	1.021	1.013	1.001	1.003	1.001	1.002
	800	1.352	1.355	1.102	1.087	1.011	1.015	0.997	1.005	1.004	1.003
Probit	100	1.773	0.792	2.225	0.638	4.365	0.595	5.978	0.592	8.456	0.594
	200	2.419	0.799	3.157	0.642	6.069	0.599	8.443	0.594	11.764	0.598
	400	3.319	0.802	4.506	0.643	8.841	0.603	12.102	0.598	16.951	0.600
	800	4.660	0.803	6.403	0.645	12.428	0.604	17.697	0.599	24.593	0.600
LogLog	100	1.935	0.944	2.487	0.861	4.921	1.147	6.775	1.468	9.580	1.947
	200	2.659	0.959	3.522	0.868	6.813	1.155	9.507	1.469	13.248	1.945
	400	3.687	0.969	5.057	0.874	9.956	1.162	13.631	1.465	19.098	1.940
	800	5.202	0.975	7.195	0.878	13.985	1.161	19.921	1.465	27.685	1.939
CLogLog	100	1.944	0.951	2.500	0.862	4.929	1.157	6.764	1.463	9.563	1.944
	200	2.702	0.969	3.559	0.872	6.845	1.158	9.535	1.461	13.291	1.939
	400	3.708	0.974	5.065	0.876	9.935	1.158	13.625	1.467	19.084	1.944
	800	5.233	0.981	7.205	0.879	13.980	1.157	19.916	1.465	27.679	1.940

Note: This table presents the ratio of several performance measures in relation to a correctly specified MLE of five estimators, our semiparametric proposal (SPQMLE) and four implementations of the Papke and Wooldridge's (1996) estimator for Logit, Probit, LogLog and CLogLog links. Standard Errors (Std. Error) are calculated using (1.6). n represents different sample sizes and ϕ the precision parameter used to generate the data across 1,000 simulations.

Figure 1.1: QQ plot for Estimators of β



Note: Quantile-Quantile plots with reference to the standard normal distribution of five estimators, our semiparametric proposal (SPQMLE) and four implementations of the Papke and Wooldridge (1996) estimator for Logit, Probit, LogLog and CLogLog links. n represents different sample sizes and ϕ the precision parameter used to generate the data across 1,000 simulations.

when dealing with the full sample, and therefore include this variable squared.

We compute both the linear regression (OLS) and QMLE (with Logit link) estimates for the preferred specification. To make results more directly comparable to those of our introduced method, we also estimate restricted QMLE models that mimic the identification conditions in Assumption 1.1: setting the intercept equal to 0 and the coefficient of a continuous variable, in this case age of the plan, equal to 1. Finally, we compute the SPQMLE following the computational considerations outlined in the supplemental material subsection 1.B.

Table 1.2 presents our results. We observe that both the OLS and unrestricted QMLE columns correspond exactly to the results in Papke and Wooldridge (1996) for the appropriate specifications. The restricted QMLE specification is not sensitive to the optimization method and resembles the unrestricted model. Strikingly, we see that using the semiparametric approach actually leads to results that are closer to OLS than to the QMLE proposed by the authors. Adding flexibility and robustness to the specification through our method results in a move towards the baseline estimates. This sheds light on the fact that assuming a specific link function in the QMLE approach might be too restrictive and could potentially create bias problems such as those illustrated in our simulation study.

To focus away from the coefficient estimates and into more intuitive and comparable results, the table also present an estimate for the average partial effect (APE) of match rate on participation rate. In general, we observe that the APEs remain fairly close to one another, with the QMLE one being the largest. Using the results from our SPQMLE method, we observe that a change in the match rate of 10 percentage points (10 cents for every dollar contributed by the employees) increases participation in the plan by approximately 0.9 percentage points.

1.4 Conclusions

We proposed a semiparametric extension of the parametric QML estimator in Papke and Wooldridge (1996) that allows for flexible estimation of fractional response models and is robust to potential misspecification of the link function. The main result in the paper proves the consistency and asymptotic normality of the estimator allowing for data-driven smoothing parameter and random trimming. We confirm through a Monte Carlo experiment that our estimator performs compar-

Table 1.2: Empirical Results with Additional Methods

Dependent variable: Participation Rate	OLS (1)	QMLE (2)	Rest. QMLE[a] (3)	Rest. QMLE[b] (4)	SPQMLE (5)
Match Rate	0.143	1.665	1.660	1.655	0.188
	(0.008)	(0.104)	(0.188)	(0.179)	(0.005)
Match Rate^2	−0.029	−0.332	−0.335	−0.334	−0.039
	(0.002)	(0.026)	(0.050)	(0.049)	(0.001)
log(Employment)	−0.099	−1.031	−1.079	−1.078	−0.100
	(0.012)	(0.110)	(0.205)	(0.031)	(0.003)
log(Employment)^2	0.0050	0.0536	0.0461	0.0460	0.0048
	(0.0008)	(0.0071)	(0.0117)	(0.0035)	(0.0002)
Age	0.0056	0.0548	1.000	1.000	1.000
	(0.0007)	(0.0077)	—	—	—
Age^2	−0.00007	−0.00063	−0.01931	−0.01931	−0.01246
	(0.00001)	(0.00018)	(0.00297)	(0.00035)	(0.00001)
Sole Plan	0.0066	0.0643	0.1552	0.1523	0.0162
	(0.0051)	(0.0498)	(0.0785)	(0.0785)	(0.0042)
Constant	1.170	5.105	—	—	—
	(0.042)	(0.416)			
Average Partial Effect of Match Rate	0.099	0.143	0.109	0.109	0.090
R^2	0.182	0.197	—	—	0.215
Log-likelihood	—	—	−2,571.0	−2,571.0	—

Note: Match rate is unrestricted, leaving 4,734 observations at the plan level. Heteroskedasticity-robust standard errors are in parenthesis. Restricted QMLE methods impose a 0 constant term and normalized the coefficient of Age to 1: [a] estimated using the augmented Lagrange optimization method and [b] estimated by iteratively re-weighted least squares with an offset given by the Age variable.

atively well with respect to the parametric maximum likelihood and correctly specified quasi-likelihood alternatives. As practitioners seldom know the correct form of the link function in practice, our method offers a robust alternative to existing parametric methods.

Appendices

1.A Proof

Proof of Theorem 1.1 (Outline). First, consistency follows from an application of the uniform consistency results for kernel estimators of Escanciano et al. (2014) as well as theorem 1 of Gourieroux et al. (1984). Note that our assumptions encompass those of Lemma B.4 of Escanciano et al. (2014) and thus guarantee the convergence of $\widehat{m}$ uniformly over $\boldsymbol{\beta}$ and the bandwidth, therefore that of the maximizing function in (1.2). Since we similarly satisfy conditions a of Gourieroux et al. (1984) for the resulting likelihood of the linear exponential family, and given our index restriction imposed in (1.1) as well as the identification assumptions, we guarantee consistency of $\widehat{\boldsymbol{\beta}}$ to $\boldsymbol{\beta}_0$.

For the asymptotic normality part, we use a combination of standard Taylor expansion methods with the uniform convergence and uniform representation results. Consider the first order conditions

$$0 = \frac{\partial \mathcal{L}_n}{\partial \beta}(\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n}[y_i - \widehat{m}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}})]\widehat{\psi}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}})\widehat{t}_{ni}\,, \tag{1.7}$$

where $\widehat{\psi}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}) \equiv \{\widehat{m}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}})[1-\widehat{m}(\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}})]\}^{-1}\partial\widehat{m}(\boldsymbol{x}_i'\boldsymbol{\beta})/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}$, $\widehat{t}_{ni} = \mathbb{I}\{\widehat{f}(\boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}}) \geq \tau_n\}$, $\tau_n \to 0$ as $n \to \infty$ at a rate that satisfies Assumption 1.8, and $\widetilde{\boldsymbol{\beta}}$ is a preliminary consistent estimator for $\boldsymbol{\beta}_0$. Performing a Taylor expansion yields

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = \boldsymbol{H}_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[y_i - \widehat{m}(\boldsymbol{x}_i'\boldsymbol{\beta}_0)]\widehat{\psi}(\boldsymbol{x}_i'\boldsymbol{\beta}_0)\widehat{t}_{ni} + o_p(1)\,,$$

where

$$\boldsymbol{H}_n = -\left.\frac{\partial^2\mathcal{L}_n}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}} \quad \text{and} \quad |\bar{\boldsymbol{\beta}}-\boldsymbol{\beta}_0| \leq |\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0|\,.$$

Following the index restriction, consistency of $\widehat{\boldsymbol{\beta}}$, the uniform representation theorem and uniform consistency results of kernel estimators in Escanciano et al. (2014), the results in Gourieroux et al. (1984), as well as the continuous mapping theorem, it follows that $\boldsymbol{H}_n \xrightarrow{p} \boldsymbol{A}$ as previously defined and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[y_i - \widehat{m}(\boldsymbol{x}_i'\boldsymbol{\beta}_0)]\widehat{\psi}(\boldsymbol{x}_i'\boldsymbol{\beta}_0)\widehat{t}_{ni} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[y_i - m(\boldsymbol{x}_i'\boldsymbol{\beta}_0)]\psi(\boldsymbol{x}_i'\boldsymbol{\beta}_0) + o_p(1)\,, \tag{1.8}$$

where $\psi(\boldsymbol{x}_i'\boldsymbol{\beta}_0) = \{m(\boldsymbol{x}_i'\boldsymbol{\beta}_0)[1-m(\boldsymbol{x}_i'\boldsymbol{\beta}_0)]\}^{-1}\partial M(\boldsymbol{x}_i'\boldsymbol{\beta})/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$. Note that

$$\left.\frac{\partial M(\boldsymbol{x}_i'\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = m'(\boldsymbol{x}_i'\boldsymbol{\beta}_0)(\boldsymbol{x}_i - \mathrm{E}[\boldsymbol{X}_i|\boldsymbol{x}_i'\boldsymbol{\beta}_0])\,, \tag{1.9}$$

which can be found either by the chain rule (Newey, 1994) or by a couple of Taylor expansions. An application of the Lindeberg-Levy CLT yields $n^{-1/2}\sum_{i=1}^{n}[y_i - m(\boldsymbol{x}_i'\boldsymbol{\beta}_0)]\psi(\boldsymbol{x}_i'\boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}(\boldsymbol{0},\boldsymbol{B})$, so that finally,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1})\,. \qquad \square$$

1.B Computational Considerations

Since the SPQMLE has a similar structure to that of Klein and Spady (1993), for computation purposes we will leverage the capacities of the `np` package in the `R` software (Hayfield and Racine, 2008). In particular, we make use of the `npindex(..., method = ‘kleinspady’, ...)` routine. As a simplification, and to remain in line with the package's computational strategy, estimation of $\boldsymbol{\beta}_0$ will be performed jointly with the bandwidth $\widehat{h}_n$, which is allowed by our method as a data-dependent bandwidth, i.e., $(\widehat{\boldsymbol{\beta}}, \widehat{h}_n) = \arg\max_{\boldsymbol{\beta}\in\mathcal{B}, h_n\in\mathbb{R}_{++}} \mathcal{L}_n(\boldsymbol{\beta}, h_n)$ (Hardle et al., 1993; Escanciano et al., 2016). We modify the package to reflect the characteristics of our estimation method by eliminating the requirement for binary data and correcting the variance-covariance estimator formula in order to obtain valid statistical inference. All the numerical exercises in the paper make use of this implementation.

1.C Empirical Application Table

Table 1.C.1: Replication of Papke and Wooldridge (1996) with additional methods on restricted sample

Dependent variable: Participation Rate	OLS (1)	QMLE (2)	Rest. QMLE[a] (3)	Rest. QMLE[b] (4)	SPQMLE (5)
Match Rate	0.156	1.390	1.167	1.165	0.148
	(0.011)	(0.108)	(0.195)	(0.190)	(0.013)
log(Employment)	0.112	1.002	1.036	1.036	0.095
	(0.013)	(0.110)	(0.199)	(0.033)	(0.004)
log(Employment)^2	0.0057	0.0522	0.0434	0.0433	0.0048
	(0.0009)	(0.0071)	(0.0114)	(0.0037)	(0.0003)
Age	0.0060	0.0501	1.000	1.000	1.000
	(0.0009)	(0.0088)	—	—	—
Age^2	0.00007	0.00052	0.02011	0.02011	0.01221
	(0.00002)	(0.00021)	(0.00339)	(0.00061)	(0.00001)
Sole Plan	−0.0001	0.0079	0.1311	0.1302	0.0027
	(0.0060)	(0.0502)	(0.0814)	(0.0813)	(0.0043)
Constant	1.213	5.058	—	—	—
	(0.048)	(0.421)			
Average Partial Effect of Match Rate	0.156	0.173	0.111	0.111	0.126
R^2	0.143	0.152	—	—	0.179
Log-likelihood	—	—	−2,285.3	−2,285.3	—

Note: Match rate is limited to a maximum of 1, leaving 3,784 observations at the plan level. Heteroskedasticity-robust standard errors in parenthesis. Restricted QMLE methods impose a 0 constant term and the coefficient of Age being equal to 1: [a] estimated using the augmented Lagrange optimization method and [b] estimated by iteratively reweighted least squares with an offset given by the Age variable.

Chapter 2

Copula Estimation and Variable Selection with Multivariate Fractional Outcomes

In microeconomics, multivariate fractional outcomes are salient in two strands of the literature: structural microeconomics, specifically within demand system estimation, and reduced form regression analysis. In both contexts, there are similar key model characteristics that need to be taken into account.

First, most reduced form or structural models produce an estimating equation in the form of a conditional mean such as

$$\mathrm{E}[\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}]=\boldsymbol{m}(\boldsymbol{x},\boldsymbol{\beta}),$$

where $\boldsymbol{Y}$ represents the outcomes that take values in $\mathcal{S}^d$; $\boldsymbol{X}$ are some covariates such as price, expenditure, and functions of these and other variables; $\boldsymbol{\beta}$ represents the parameters of interest that may or may not have a structural interpretation; and $\boldsymbol{m}(\boldsymbol{x},\boldsymbol{\beta})=(m_1(\boldsymbol{x},\boldsymbol{\beta}),\ldots,m_d(\boldsymbol{x},\boldsymbol{\beta}))'$ is a vector of (possibly) nonlinear functions of covariates and parameters (Papke and Wooldridge, 1996, 2008). Example 1 in subsection 2.1.1 presents the conditional mean for the Almost Ideal Demand (AID) model of Deaton and Muellbauer (1980), a widely used structural demand system. Example 2 presents a multivariate fractional logit specification, which is a popular functional form for regression analysis with multivariate fractional outcomes (Mullahy, 2015; Murteira and Ramalho, 2016). This chapter starts from the conditional mean as the primary object and builds methods that impose such specification while maintaining flexibility.

A second key fact is that variable selection can be crucial. For example, when the dimensionality of the outcomes in structural demand systems is large or when many determinants of the allocations are considered, selecting which effects remain important for determining household consumption patterns is a variable selection issue. Additionally, there are meaningful ways in which the fit of structural demand systems can be improved by considering polynomials to approximate certain functions underlying the specification (Lewbel, 1991). The degrees of these polynomials would then need to be selected from the data (Lewbel and Pendakur, 2009). Similarly, covariate selection remains an important specification issue in reduced form models. It is thus necessary that the methods used to estimate these models can also handle variable selection. Inference would then need to be adjusted to account for the effect of selection, but this adjustment can be technically complex (Knight and Fu, 2000; Chernozhukov et al., 2018). To address this issue, this chapter employs Bayesian methods, which can incorporate selection via regularization in a similar way to LASSO and its alternatives while inference remains simple (Park and Casella, 2008; Li and Lin, 2010; Leng et al., 2014).

Third, structural demand models usually impose constraints on the parameter vector $\boldsymbol{\beta}$ to satisfy the economic regularity of the demand functions they produce. These are not only restrictions within each equation of the conditional mean but may also include cross-equation restrictions (Barnett, 2002). The AID model, for example, imposes homogeneity in expenditures and prices as well as symmetry of the Slutsky matrix via these cross-equation restrictions, both of which are important testable assumptions of the theory. Perhaps more important within this literature is the idea of curvature that is encoded in the negative semidefiniteness of the Slutsky matrix (Blundell et al., 2012; Chang and Serletis, 2014). Much of the research in demand estimation is thus dedicated to introducing and analyzing the properties of different models that can both expand the theoretical foundation of demand systems and capture important patterns in the data (Lewbel and Pendakur, 2009; Barnett and Serletis, 2008). In estimating these models, the first and third key facts are considered at length in the literature, but the second fact is not generally taken into account. The simplex nature of the multivariate fractional outcomes is also generally ignored by assuming an unrestricted distribution for $\boldsymbol{Y}$ centered at $\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{\beta})$ (Barnett and Serletis, 2008). This chapter aims to correct this gap.

The main contribution of this chapter is to introduce a unified estimation procedure via copulas that simultaneously incorporate all points discussed previously. That is, these methods impose the fractional and unit-sum constraints of multivariate fractional outcomes, satisfy a conditional mean regression structure, allow for variable selection with correct inference, and can incorporate cross-equation restrictions. The use of copulas also broaden the possible dependence patterns between each share in the system, which is a general concern in the compositional data literature (Aitchison, 2003). The chapter first presents two ways of constructing a likelihood using copulas. The marginal distributions impose the conditional mean specification and satisfy the fractional restriction, while the joint distribution captures the dependence structure and unit-sum constraint between shares. The generality in constructing the likelihood functions allows for a unified way to estimate both structural demand systems and reduced form models. As the maximum likelihood estimators (MLE) arising from this construction are themselves contributions to the literature on multivariate fractional outcome models, the chapter derives the asymptotic properties of these estimators in a standard frequentist context before diving into a full Bayesian solution.

In order to handle model selection, the chapter then uses a general class of priors in a Bayesian framework to augment the base estimators through the use of regularization (Park and Casella, 2008; Hans, 2009). This form of selection is also useful even in the case where the dimensionality of the covariates is large or grows with the sample size (i.e., high-dimensional settings, see Li and Lin, 2010). Finally, the use of Bayesian methods guarantees that, even with a selection step, inference is simple not only for the estimated parameters, but also for functions of interest computed from these parameters. These include quantities such as average partial effects (APE) in reduced form models or price and income elasticities after estimation of a demand system.

The chapter proceeds as follows. The next section introduces the specification of a parametric likelihood constructed using copulas in two different ways. The properties of the resulting maximum likelihood estimators are then analyzed. Section 2.2 introduces the class of prior distributions for the coefficients of the conditional mean and outlines the Bayesian estimation algorithm. Numerical exercises in Section 2.3 showcase the properties and flexibility of these estimators, as well as their comparison with other methods available in the literature. Section 2.4 presents an application of the proposed methods to the demand of transportation services in Canada from a structural demand system perspective. Section 2.5 presents the concluding remarks.

2.1 Methodological Framework

Existing methods for estimating models with compositional outcomes can be broadly categorized into transformation and (possibly quasi-) likelihood-based methods. The former operate by taking the shares in the simplex space $\mathcal{S}^d$ to an unrestricted domain and then fitting a regression on the transformed outcomes. Aitchison (1982, 1983) considers a multivariate normal distribution on the additive log-ratio transformation of the share system, resulting in a seemingly unrelated regression (SUR) framework with transformed outcomes (Zellner, 1962; Allenby and Lenk, 1994). More general transformations have been considered in the literature and include the centered log-ratio (Aitchison, 1983), isometric log-ratio (Egozcue et al., 2003), and α (Tsagris et al., 2011) transformations. The problem with using these methods in econometric modeling is that they induce properties that complicate the recovery of the conditional mean of $\boldsymbol{Y}$ on $\boldsymbol{X}$. As noted previously, this is the object of interest in a regression framework and cannot be obtained after these transformations unless implausibly strong assumptions are imposed, even in the simpler univariate case (see, e.g., Papke and Wooldridge, 1996).

The latter likelihood-based methods impose certain distributional assumptions — which may or may not need to be correctly specified (Montoya-Blandón and Jacho-Chávez, 2020) — to estimate the coefficients associated with the variables in a regression framework using link functions (see, e.g., Papke and Wooldridge, 1996, 2008). These include multivariate normal (Barten, 1969; Woodland, 1979), Dirichlet (Hijazi and Jernigan, 2009) and fractional multinomial (Mullahy, 2015; Murteira and Ramalho, 2016) regression models. The methods in this paper stand between full distributional assumptions and the quasi-likelihood approach. In particular, the few distributions that can fit data directly on $\mathcal{S}^d$ tend to have restrictive dependence structures between variables, such as having all pairwise correlations be negative in the case of the Dirichlet distribution. Additionally, while efficient if correctly specified, they are not guaranteed to be consistent if the distributional assumption fails. On the other hand, quasi-likelihood estimation remains consistent while sacrificing efficiency.[1] Not having a correctly-specified likelihood also precludes the use of the Bayesian

[1]Some efficiency could be recovered by imposing higher-order moment conditions (Gourieroux et al., 1984; Mullahy, 2015).

approach and its advantages. This is why this paper combines copulas — expanding the possible dependence structure allowed between shares while adding robustness — with a full-likelihood approach in order to take advantage of Bayesian methods in estimation, selection and inference.

2.1.1 Likelihood and Identification

The rest of this section outlines the construction of the likelihood function using marginal distributions on a bounded support, which are then combined via copulas. This is done in a way that respects the unit-sum constraint and imposes the conditional mean specification. Let $(\boldsymbol{Y}', \boldsymbol{X}')'$ be a $(d+p)$-dimensional random-vector, where $\boldsymbol{Y} = (Y_1, \dots, Y_d)'$ takes values on $\mathcal{S}^d$ and $\boldsymbol{X}$ has support $\mathcal{X} \subset \mathbb{R}^p$. Let H denote the true joint distribution of $(\boldsymbol{Y}', \boldsymbol{X}')'$ and P_X denote the marginal distribution of the covariates. Additionally, let $H_{Y|X}$ denote the true conditional joint distribution of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$ and $H_{Y_j|X}$ denote the associated conditional marginal distributions for $j = 1, \dots, d$. For notational convenience, these will be written as H and H_j, respectively, with their conditional nature made clear within their arguments. Each marginal distribution satisfies the fractional restriction; i.e., $H_j(y_j|\boldsymbol{X} = \boldsymbol{x}) = 0$ if $y_j < 0$ and $H_j(y_j|\boldsymbol{X} = \boldsymbol{x}) = 1$ if $y_j > 1$ for each $j = 1,\dots, d$ and almost all $\boldsymbol{x} \in \mathcal{X}$. As mentioned previously, the following conditional mean specification is assumed to hold throughout.

Assumption 2.1. The joint distribution of $(\boldsymbol{Y}, \boldsymbol{X})$ satisfies

$$\mathrm{E}[Y_j|\boldsymbol{X} = \boldsymbol{x}] = m_j(\boldsymbol{x}, \boldsymbol{\beta}_0)\,, \tag{2.1}$$

for almost all $\boldsymbol{x} \in \mathcal{X}$, some K-dimensional $\boldsymbol{\beta}_0 \in \mathcal{B} \subset \mathbb{R}^K$, and known functions $m_j : \mathbb{R}^p \times \mathbb{R}^k \to \mathbb{R}$, such that $0 < m_j(\boldsymbol{x}, \boldsymbol{\beta}) < 1$ for all $\boldsymbol{x}$ and $\boldsymbol{\beta}$, $j = 1, \dots, d$.

Note that this is a restriction on the family of conditional marginal distributions of $\boldsymbol{Y}$. In order to obtain sensible predictions, one should place an additional unit-sum constraint on the expectations: $\sum_{j=1}^{d} m_j(\boldsymbol{x}, \boldsymbol{\beta}) = 1$. The following examples present a couple of popular functional forms in both structural and reduced form models that satisfy Assumption 2.1.

Example 1. (Demand Estimation) As noted before, the almost ideal demand (AID) system is a popular model in demand estimation with a conditional mean specification $\boldsymbol{m}(\boldsymbol{x},\boldsymbol{\beta})$ given by

$$m_j(\boldsymbol{x},\boldsymbol{\beta}) = \alpha_j + \sum_{l=1}^{d}\gamma_{jl}\log p_l + \pi_j\left\{\log e - \alpha_0 - \sum_{l=1}^{d}\alpha_l\log p_l - \frac{1}{2}\sum_{k=1}^{d}\sum_{l=1}^{d}\gamma_{kl}\log p_k\log p_j\right\} \quad (2.2)$$

for all $j = 1,\ldots,d$, where $\boldsymbol{\beta} = (\alpha_0,\ldots,\alpha_d,\pi_1,\ldots,\pi_d,\gamma_{11},\ldots,\gamma_{dd})'$ are the structural parameters and $\boldsymbol{x} = (e,\boldsymbol{p}')'$, so that the covariates represent total expenditures and prices. Additionally, the following cross-equation restrictions are imposed to satisfy homogeneity of degree zero in prices and total expenditure, as well as a symmetric Slutsky matrix: $\sum_{j=1}^{d}\alpha_j = 1$, $\sum_{j=1}^{d}\pi_j = \sum_{j=1}^{d}\gamma_{jl} = \sum_{j=1}^{d}\gamma_{lj} = 0$ and $\gamma_{jl} = \gamma_{lj}$. Other demand systems exist, which extend the theoretical properties and provide a better fit to the data. The most popular in the literature are the quadratic AID (Banks et al., 1997), Minflex Laurent (Barnett, 1983; Barnett and Lee, 1985), and recently the exact affine Stone index (Lewbel and Pendakur, 2009). After estimating these models, price elasticities and other quantities of interest are computed for which standard errors are required. Demand systems also generally admit a fully linear approximation that reduces each component of $\boldsymbol{m}(\boldsymbol{x},\boldsymbol{\beta})$ to an identity link on a single-index. All of these models rely on imposing parameter restrictions to satisfy the unit-sum constraint, while not imposing the fractional constraint of the outcomes.[2]

Example 2. (Reduced Form) A model that specifies each component of $\boldsymbol{m}(\boldsymbol{x},\boldsymbol{\beta})$ as a link function on a single-index can also arise from several different contexts. It is commonly used when a researcher wants to explore the relationship between covariates and outcomes with no particular structural justification in mind. However, these specifications also arise from some structural frameworks when additional assumptions are imposed (Considine and Mount, 1984; Dubin, 2007). For example, a model could take the form of a multivariate fractional logit (Mullahy, 2015):

$$m_j(\boldsymbol{x},\boldsymbol{\beta}) = \begin{cases} \frac{\exp(\boldsymbol{x}'\boldsymbol{\beta}_j)}{1+\sum_{l=1}^{j-1}\exp(\boldsymbol{x}'\boldsymbol{\beta}_l)} & \text{for } j = 1,\ldots,d-1\,, \\ \frac{1}{1+\sum_{l=1}^{j-1}\exp(\boldsymbol{x}'\boldsymbol{\beta}_l)} & \text{for } j = d\,, \end{cases} \quad (2.3)$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1',\ldots,\boldsymbol{\beta}_{d-1}')'$. Perhaps more interesting in these types of nonlinear models is the average partial effect of variable k on outcome j, given by $\partial\,\mathrm{E}[Y_j|\boldsymbol{X}=\boldsymbol{x}]/\partial x_k$. Inference about this object is thus of great importance in an applied setting.

[2]The fractional constraint also guarantees positivity, a restriction that is generally ignored or checked only after estimating a particular demand system, and is not imposed in the estimation process.

An application of Sklar's (1959) theorem allows for a representation of H using copulas as $H(y_1,\dots,y_d|\boldsymbol{X}=\boldsymbol{x}) = C(H_1(y_1|\boldsymbol{X}=\boldsymbol{x}),\dots,H_d(y_d|\boldsymbol{X}=\boldsymbol{x}))$, where $C(\cdot)$ is a copula function linking together the conditional marginals with $\boldsymbol{x}$ common across all distributions. The following assumption on the underlying distributions will be important.

Assumption 2.2. The marginals $H_j, j=1,\dots,d$ and the copula C admit density functions conditional on $\boldsymbol{X}=\boldsymbol{x}$, which are denoted by $h_j, j=1,\dots,d$ and c, respectively.

Given Assumption 2.2, the conditional joint density $h(y_1,\dots,y_d|\boldsymbol{X}=\boldsymbol{x})$ is well-defined as is the unconditional density. Modeling can then take place in two steps. First, marginals F_j are selected for each outcome $y_j, j=1,\dots,d$ from the general class of distributions on the unit interval that satisfy Assumption 2.1 (denoted here as $\mathcal{F}$). Then, a copula C_Y can be chosen from class $\mathcal{C}$. Taking a parametric stance on the definition of the copula, the conditional joint can be expressed as

$$F_{1,\dots,d}(\boldsymbol{y}|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta},\boldsymbol{\psi}) = C_Y(F_1(y_1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_1),\dots,F_d(y_d|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_d);\boldsymbol{\psi})\,, \tag{2.4}$$

where $\boldsymbol{\delta}=(\boldsymbol{\delta}_1',\dots,\boldsymbol{\delta}_d')'\in\Delta$ are the parameters that govern the marginal distribution of each component and $\boldsymbol{\psi}\in\Psi$ defines the dependence structure between the variables in the copula. These parameters are defined on the spaces $\Delta=\times_{j=1}^{d}\Delta_j\subset\mathbb{R}_j^D$, where D_j is the dimensionality of each $\boldsymbol{\delta}_j, j=1,\dots,d$, and $\Psi\subset\mathbb{R}^S$. However, note that some issues arise when dealing directly with the object defined by (2.4) in this context. Due to the nature of the simplex, there is a redundancy in the sense that one of the variables can always be obtained from the others (Murteira and Ramalho, 2016; Elfadaly and Garthwaite, 2017). To illustrate this fact, take d as a base category and let $W=Y_1+\cdots+Y_{d-1}$. The distribution of Y_d will then be given by

$$F_d(y_d|\boldsymbol{X}=\boldsymbol{x}) = 1-F_W(1-y_d|\boldsymbol{X}=\boldsymbol{x})\,, \tag{2.5}$$

where

$$F_W(w|\boldsymbol{X}=\boldsymbol{x}) = \lim_{w_j\to\infty, j=2,\dots,d-1}\Pr(Y_1+\cdots+Y_{d-1}\le w, Y_2\le w_2,\dots,Y_{d-1}\le w_{d-1}|\boldsymbol{X}=\boldsymbol{x})\,.$$

This probability is taken over the joint distribution of $(Y_1,\dots,Y_{d-1})'$ conditional on $\boldsymbol{X}=\boldsymbol{x}$, which could be obtained from a second application of Sklar's theorem.[3] Thus, F_d is completely determined by the remaining components and a likelihood function based on this joint distribution would be constant with respect to $\boldsymbol{\delta}_d$. As identifiability is a property of the likelihood, this implies that $\boldsymbol{\delta}_d$ would not be identifiable separately from $(\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_{d-1}')'$. In a frequentist context, nothing else could be said about this remaining component. However, in a Bayesian framework, if there was some prior information linking $(\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_{d-1}')'$ and $\boldsymbol{\delta}_d$ together, it could be possible to achieve a posterior updating of $\boldsymbol{\delta}_d$ conditional on the data (Poirier, 1998).

As an example of this identification failure, consider specifying a Gaussian copula with Gaussian marginals (forgetting for a moment about the fractional restriction). The unit-sum constraint that yields (2.5) would imply a singular covariance matrix between the components of $\boldsymbol{Y}$. In a demand estimation context, Barten (1969) explores these effects, showing how to perform maximum likelihood estimation (MLE) of the parameters of the resulting demand system by eliminating one of the equations.

This paper considers two ways of imposing a copula on a D-dimensional object with $D \equiv d - 1$ in a way that both the unit-sum constraint from the simplex and the conditional mean specification in (2.1) are satisfied. For this reason and to simplify notation, some D-dimensional objects will be used interchangeably with their d-dimensional counterparts, but their distinctions will be made clear when necessary.

Copula Specification on Y

Consider placing a copula similar to (2.4) except that the object of interest is the D-dimensional vector $\boldsymbol{Y}_{-d} = (Y_1, \ldots, Y_D)'$, where the d-th component is taken as the base and is thus eliminated:

$$F(\boldsymbol{y}_{-d}|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\psi}) = C_Y(F_1(y_1|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_1), \ldots, F_D(y_D|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_D); \boldsymbol{\psi})\,. \tag{2.6}$$

Now, while identification is no longer an issue, there is still the fact that F has support on $[0, 1]^D$. That is, it places some probability outside of the set $\mathcal{T} = \{(y_1, \ldots, y_D) \in \mathbb{R}^D : 0 \leq y_j \leq 1, j = 1, \ldots, d; \sum_{j=1}^{D} y_j \leq 1\}$, so that it does not correspond to a valid distribution on $\mathcal{S}^d$ after marginal-

[3] This particular formula arises by considering the inverse transformation $Y_1 = W - Y_2 - \cdots - Y_{d-1}, Y_2 = V_2, \ldots, Y_{d-1} = V_{d-1}$ and obtaining the marginal for W. Similar formulas would set $Y_j = W - Y_1 - \cdots - Y_{j-1} - Y_{j+1} - \cdots - Y_{d-1}$ for some j in $1, \ldots, d-1$ and integrate over the remaining components.

izing the last component. Additionally, generating values from the distribution in (2.6) would yield draws that do not satisfy the unit-sum constraint with some probability. The amount of density placed outside of $\mathcal{T}$ depends on the distribution of W as previously defined. The following proposition gives the details of the general case from (2.5). All proofs can be found in Appendix 2.A.

Proposition 1. *The cdf of $W = Y_1 + \ldots + Y_D$ conditional on $\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\delta}$, and $\boldsymbol{\psi}$ is given by*

$$F_W(w|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\psi}) = \int_0^{w-D+l} \int_0^{w-D+l-y_D} \cdots \int_0^{w-D+l-\sum_{k=D-l+2}^{D} y_k} \int_0^1 \cdots \int_0^1 \mathrm{d}F(y_1, \ldots, y_{D-l}, y_{D-l+1}, \ldots, y_{D-1}, y_D|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\psi}) \,, \tag{2.7}$$

when $w \in (D-l, D-l+1]$ for $l = 1, \ldots, D$.

Based on this characterization, we can find $\Pr(\boldsymbol{Y}_{-d} \in \mathcal{T}|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\psi}) = F_W(1|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\psi})$. Under the following assumption, it is possible to obtain a density on $\boldsymbol{Y}_{-d}$ given by the truncation of the copula density to the set $\mathcal{T}$.

Assumption 2.3.A. The marginals $F_j, j = 1, \ldots, D$ and the copula C_Y admit density functions conditional on $\boldsymbol{X} = \boldsymbol{x}$, which are denoted by $f_j, j = 1, \ldots, D$ and c_Y, respectively.

Then, by Assumption 2.3.A,

$$\begin{aligned} f(\boldsymbol{y}_{-d}|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\psi}; \mathcal{T}) &= \begin{cases} \frac{f(\boldsymbol{y}_{-d}|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta},\boldsymbol{\psi})}{F_W(1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta},\boldsymbol{\psi})} & \text{if } \boldsymbol{y}_{-d} \in \mathcal{T}\,, \\ 0 & \text{if } \boldsymbol{y}_{-d} \notin \mathcal{T}\,, \end{cases} \\ &= \mathbb{I}(\boldsymbol{y}_{-d} \in \mathcal{T}) \frac{f(\boldsymbol{y}_{-d}|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\psi})}{F_W(1|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\psi})} \,, \end{aligned} \tag{2.8}$$

where $\mathbb{I}(\cdot)$ is the indicator function that takes the value of 1 if its argument is true and 0 otherwise. The nontruncated density is given by

$$f(\boldsymbol{y}_{-d}|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\psi}) = c_Y(F_1(y_1|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_1), \ldots, F_D(y_D|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_D); \boldsymbol{\psi}) \prod_{j=1}^{D} f_j(\boldsymbol{y}_j|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_j) \,.$$

While this method of constructing a likelihood function satisfies the conditional mean specification and unit-sum constraints, the possibly high-dimensional integral can be a complicated computation.

Some algorithms, such as the AEP of Arbenz et al. (2011), are devised for the specific purpose of approximating the integral in (2.7). This is used in the numerical implementation of the algorithm to drastically reduce the computational burden compared to general multivariate integration or Monte Carlo methods.

Copula Specification on Z

With the drawbacks outlined in the previous subsection, a second way of constructing a likelihood is considered here that does not suffer from such computational complexity. This is achieved by introducing a transformation step for the vector $\boldsymbol{Y}$ in order to impose more structure. Most transformations mapping $\mathcal{S}^d$ to $\mathbb{R}^d$ or $\mathbb{R}^{d-1}$ have an inverse mapping with a closure structure; i.e., they take each vector component and divide it by the sum of the whole vector. The resulting ratios make it so that recovering the conditional mean $\mathrm{E}[\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}]$ from the transformation is complicated and entails strong and implausible assumptions (Papke and Wooldridge, 1996). In contrast, this paper employs a transformation that has a multiplicative structure for the inverse mapping. That way, it is possible to obtain the conditional mean for $\boldsymbol{Y}$ on $\boldsymbol{X}$. Assuming that Y_d is selected as the base variable again, the so-called stick-breaking transformation (Connor and Mosimann, 1969) is used to produce new variables $Z_1, \ldots, Z_d$, such that

$$Z_1 = Y_1, \quad Z_j = \frac{Y_j}{1 - \sum_{l=1}^{j-1} Y_l} \quad \text{for } j = 2, \ldots, d-1, \quad \text{and} \quad Z_d = 1\,. \tag{2.9}$$

This mapping is denoted as $\boldsymbol{s}(\boldsymbol{Y}) = (s_1(\boldsymbol{Y}), \ldots, s_D(\boldsymbol{Y}))'$, where $Z_j = s_j(\boldsymbol{Y})$ for $j = 1, \ldots, D$. Note that after this transformation, Z_d becomes fixed, which once again highlights the redundancy problem in the original $\boldsymbol{Y}$ vector: it can be transformed into a lower-dimensional vector without sacrificing information. Here, it is important to note that although any category can be chosen as a base, subsequent analyses will depend on this base category. However, this failure to be permutation invariant is generally not viewed as an issue in most of the econometric literature as long as it is taken into consideration (Mullahy, 2015; Murteira and Ramalho, 2016).

Additionally, observe that $\boldsymbol{Z} = (Z_1, \ldots, Z_D)'$ takes values in $[0,1]^D$. Thus, placing a copula structure on $\boldsymbol{Z}$ analogous to (2.6) would not need to be truncated as it would always satisfy the unit-sum constraint of the original $\boldsymbol{Y}$ for any marginals and dependence structure. Therefore, the following distribution is considered:

$$G(z_1,\ldots,z_D|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\omega},\boldsymbol{\xi}) = C_Z(G_1(z_1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\omega}_1),\ldots,G_D(z_d|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\omega}_D);\boldsymbol{\xi})\,, \tag{2.10}$$

where $\boldsymbol{\omega} = (\boldsymbol{\omega}_1',\ldots,\boldsymbol{\omega}_D')' \in \Omega$ are the marginal parameters and $\boldsymbol{\xi} \in \Xi$ are the copula parameters. Here, similar to (2.6), $G_j, j = 1,\ldots,D$ are marginals respecting the fractional constraint, $\Omega = \times_{j=1}^{D}\Omega_j$ with each $\Omega_j \subset \mathbb{R}_j^O$, and $\Xi \subset \mathbb{R}^S$. In order to satisfy the conditional mean specification in (2.1), the restrictions given by the following proposition must be imposed on the conditional means of $\boldsymbol{Z}$.

Proposition 2. *There exist conditional mean functions* $\mathrm{E}[Z_j|\boldsymbol{X}=\boldsymbol{x}] \equiv \mu_j(\boldsymbol{x};\boldsymbol{\beta},\boldsymbol{\omega},\boldsymbol{\xi})$ *such that the conditional mean for* $\boldsymbol{Y}$ *on* $\boldsymbol{X}$ *satisfies Assumption 2.1. In particular, any such objects that are a solution to*

$$\mu_j(\boldsymbol{x};\boldsymbol{\beta},\boldsymbol{\omega},\boldsymbol{\xi}) + \frac{\mathrm{E}\left[\widetilde{Z}_j\prod_{l=1}^{j-1}\left(1-\widetilde{Z}_l-\mu_l(\boldsymbol{x};\boldsymbol{\omega},\boldsymbol{\xi})\right)\Big|\boldsymbol{X}=\boldsymbol{x}\right]}{1-\sum_{l=1}^{j-1}m_l(\boldsymbol{x},\boldsymbol{\beta})} = \frac{m_j(\boldsymbol{x},\boldsymbol{\beta})}{1-\sum_{l=1}^{j-1}m_l(\boldsymbol{x},\boldsymbol{\beta})} \tag{2.11}$$

will satisfy $\mathrm{E}[Y_j|\boldsymbol{X}=\boldsymbol{x}] = m_j(\boldsymbol{x},\boldsymbol{\beta})$*, where* $\widetilde{Z}_j \equiv Z_j - \mathrm{E}[Z_j|\boldsymbol{X}=\boldsymbol{x}]$.

Thus, by Proposition 2, we can sequentially find the conditional mean for $\boldsymbol{Z}$ in a way that imposes Assumption 2.1. This means that by setting up the moments of $\boldsymbol{Z}$ in a specific way, the copula would place a dependence structure on $\boldsymbol{Y}$ that is flexible and satisfies all the requirements for a multivariate fractional response model. This, of course, requires the existence of the necessary moments for a given copula C_Z. The challenging part of applying Proposition 2 comes from computing these cross-moments of $\boldsymbol{Z}$. However, in an important special case, given by the elliptical copulas with correlation matrix R, such as the Gaussian or t copulas, it is possible to show that all cross-moments depend only on the elements of R. This is due to Wick's theorem for elliptical distributions (Frahm et al., 2003) and the consequences are explored in the following example.

Example 3. (Gaussian Copula) Take a system with $d = 3$ shares and let C_Z be a Gaussian copula with correlation parameter ξ. Additionally, let both Z_1 and Z_2 have beta marginals in a mean-precision parameterization with precisions ϕ_1 and ϕ_2, respectively. Write $\mu_j \equiv \mu_j(\boldsymbol{x};\boldsymbol{\beta},\boldsymbol{\omega},\boldsymbol{\xi})$. Then, $\mathrm{E}[\widetilde{Z}_1\widetilde{Z}_2|\boldsymbol{X}=\boldsymbol{x}] = \xi\sqrt{\mathrm{Var}(Z_1|\boldsymbol{X}=\boldsymbol{x})\mathrm{Var}(Z_2|\boldsymbol{X}=\boldsymbol{x})}$ and the variance of a beta distribution in this parameterization is given by $\mathrm{Var}(Z_j|\boldsymbol{X}=\boldsymbol{x}) = \mu_j(1-\mu_j)/(1+\phi_j)$. Equation (2.11) would then take the form $\mu_1 = m_1(\boldsymbol{x},\boldsymbol{\beta})$ for $j=1$. For $j=2$, it reduces to $\mu_2 - b\sqrt{\mu_2(1-\mu_2)} = c$, where $b \equiv (\xi/\sqrt{(1+\phi_1)(1+\phi_2)})\sqrt{\mu_1/(1-\mu_1)}$ and $c \equiv m_2(\boldsymbol{x},\boldsymbol{\beta})/[1-m_1(\boldsymbol{x},\boldsymbol{\beta})]$. This has the solution

$$\mu_2 = \frac{b^2+2c \pm b\sqrt{b^2+4c(1-c)}}{2(b^2+1)},$$

which exists in the real unit interval as long as $c<1$, which in itself is guaranteed by the unit-sum constraint of the conditional mean functions $m_j(\cdot), j=1,\ldots,d$. In this setting, we have $\boldsymbol{\omega}_1 = (\mu_1,\phi_1)$ and $\boldsymbol{\omega}_2 = (\mu_2,\phi_1)$. This yields (2.1) for the $\boldsymbol{Y}$ transformed via the inverse transformation (2.A.1).

This way of introducing dependency from the underlying $\boldsymbol{Z}$ to $\boldsymbol{Y}$ is quite flexible. Proposition 2 acts in a similar way to a method of moments approach; i.e., given the copula structure in (2.10), the moments of $\boldsymbol{Z}$ are chosen to match those of $\boldsymbol{Y}$. Thus, it is also possible to have additional moments of each Y_j be matched by those of the underlying marginals. The parameters in this construction are then also written as $\boldsymbol{\delta}$. This implicit relationship depends on both the marginal and copula parameters and is denoted by $\boldsymbol{\delta} = \boldsymbol{v}(\boldsymbol{x};\boldsymbol{\beta},\boldsymbol{\omega},\boldsymbol{\xi})$. In a practical application, a researcher might only want to match the marginal moments of each Y_j and not impose a full copula structure. In this case, one could assume the $\boldsymbol{Z}$ to be independent of each other, reducing the conditional means to

$$\mu_j(\boldsymbol{x};\boldsymbol{\beta},\boldsymbol{\omega},\boldsymbol{\xi}) = \frac{m_j(\boldsymbol{x},\boldsymbol{\beta})}{1-\sum_{l=1}^{j-1} m_l(\boldsymbol{x},\boldsymbol{\beta})}.$$

The other marginal moments can be matched given the simplification of independence. Even by assuming this independence copula, the resulting $\boldsymbol{Y}$ are still correlated, although the patterns of this correlation are reduced. Consider again Example 3 but with $\boldsymbol{Z}$ assumed to be independent. If independent beta marginals are combined in this way, it is possible to recover the generalized Dirichlet distribution on $\boldsymbol{Y}$, which is a more flexible alternative to the Dirichlet used in practice (Connor and Mosimann, 1969).

As the Jacobian of the stick-breaking transformation is given by $\prod_{j=1}^{D} 1/(1-\sum_{l=1}^{j-1} Y_l)$, the next assumption, which mimics Assumption 2.3.A, yields a distribution for $\boldsymbol{Y}$.

Assumption 2.3.B. The marginals $G_j, j=1,\ldots,D$ and the copula C_Z admit density functions conditional on $\boldsymbol{X} = \boldsymbol{x}$, which are denoted by $g_j, j = 1, \ldots, D$ and c_Z, respectively.

Then, by Assumption 2.3.B and a change of variables from $\boldsymbol{Z}$ to $\boldsymbol{Y}$,

$$\begin{aligned} g(\boldsymbol{y}|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\xi}) &= g(\boldsymbol{s}(\boldsymbol{y})|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\xi}) \\ &= c_Z(G_1(s_1(\boldsymbol{y})|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_1), \ldots, G_D(s_D(\boldsymbol{y})|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_D), \boldsymbol{\xi}) \times \\ &\quad \prod_{j=1}^{D} \frac{g_j(\boldsymbol{y}_j|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_j)}{1 - \sum_{l=1}^{j-1} Y_l} . \end{aligned} \tag{2.12}$$

2.1.2 Frequentist Estimation and Asymptotic Properties

While the ultimate goal of this paper is to construct Bayesian estimators based on the joint distributions introduced in the previous subsection, to the best of my knowledge, the frequentist estimators have not been previously explored in the literature. Therefore, for completeness and to present an alternative to existing methods, the asymptotic properties of these estimators are derived in this subsection and prior specifications are postponed until the next section.

The following assumptions are introduced in order to construct a likelihood function from both (2.8) and (2.12).

Assumption 2.4. There is access to an independent and identically distributed (i.i.d.) sample of size n from the joint distribution of $(\boldsymbol{Y}', \boldsymbol{X}')'$, given by $\{(\boldsymbol{y}_i', \boldsymbol{x}_i')'\}_{i=1}^n$.

Define $\boldsymbol{\theta}_Y = (\boldsymbol{\delta}', \boldsymbol{\psi}')'$ and $\boldsymbol{\theta}_Z = (\boldsymbol{\delta}', \boldsymbol{\xi}')'$. The associated log-likelihoods are then given by

$$\begin{aligned} \ell_Y(\boldsymbol{\theta}_Y) = \frac{1}{n} \sum_{i=1}^{n} \Bigg\{ & \log c_Y(F_1(y_{1,i}|\boldsymbol{X} = \boldsymbol{x}_i; \boldsymbol{\delta}_1), \ldots, F_D(y_{D,i}|\boldsymbol{X} = \boldsymbol{x}_i; \boldsymbol{\delta}_D); \boldsymbol{\psi}) \\ & + \sum_{j=1}^{d} \log f_j(y_{j,i}|\boldsymbol{X} = \boldsymbol{x}_i; \boldsymbol{\delta}_j) - \log F_W(1|\boldsymbol{X} = \boldsymbol{x}_i; \boldsymbol{\delta}, \boldsymbol{\psi}) \Bigg\} \end{aligned} \tag{2.13}$$

and

$$\begin{aligned} \ell_Z(\boldsymbol{\theta}_Z) = \frac{1}{n} \sum_{i=1}^{n} \Bigg\{ & \log c_Z[G_1(s_1(\boldsymbol{y}_i)|\boldsymbol{X} = \boldsymbol{x}_i; \boldsymbol{\delta}_1), \ldots, G_D(s_D(\boldsymbol{y}_i)|\boldsymbol{X} = \boldsymbol{x}_i; \boldsymbol{\delta}_D); \boldsymbol{\xi}] \\ & + \sum_{j=1}^{d} \log g_j(s_j(\boldsymbol{y}_i)|\boldsymbol{X} = \boldsymbol{x}_i; \boldsymbol{\delta}_j) \Bigg\} , \end{aligned} \tag{2.14}$$

where the Jacobian term in (2.14) is not included as it does not depend on $\boldsymbol{\theta}_Z$. Once these likelihoods have been defined, a natural way to construct the estimators is

$$\widehat{\boldsymbol{\theta}}_Y \equiv \underset{\boldsymbol{\theta}_Y \in \Delta\times\Psi}{\arg\max}\, \ell_Y(\boldsymbol{\theta}_Y) \qquad \text{and} \qquad \widehat{\boldsymbol{\theta}}_Z \equiv \underset{\boldsymbol{\theta}_Z \in \Delta\times\Xi}{\arg\max}\, \ell_Z(\boldsymbol{\theta}_Z)\,. \tag{2.15}$$

The following assumptions guarantee identification and introduce correct specification of the marginals and copulas.

Assumption 2.5. (Identification)

1. F_j and G_j are absolutely continuous and globally identified for $j = 1, \ldots, D$ and the same is true for C_Y and C_Z;

2. For $j = 1, \ldots, D$ *(i)* if $m_j(\boldsymbol{x}, \boldsymbol{\beta}_1) = m_j(\boldsymbol{x}, \boldsymbol{\beta}_2)$ for almost all $\boldsymbol{x} \in \mathcal{X}$ then $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, and *(ii)* $\mathcal{X}$ must be such that $\text{Image}(m_j) = \text{Range}(m_j)$.

Assumption 2.6.A. (Correct specification) (i) There exists $\boldsymbol{\psi}_0 \in \Psi$ and $\boldsymbol{\delta}_0 = (\boldsymbol{\delta}'_{0,1}, \ldots, \boldsymbol{\delta}'_{0,D})' \in \Delta$, such that $h(\cdot|\boldsymbol{X} = \boldsymbol{x}) = f(\cdot|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_0, \boldsymbol{\psi}_0)$ for almost all $\boldsymbol{x} \in \mathcal{X}$; (ii) Similarly, there exists $\boldsymbol{\xi}_0 \in \Xi$ and $\boldsymbol{\omega}_0 \in \Omega$, such that $h(\cdot|\boldsymbol{X} = \boldsymbol{x}) = g(\cdot|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_0, \boldsymbol{\xi}_0)$ for almost all $\boldsymbol{x} \in \mathcal{X}$, where $\boldsymbol{\delta}_0 = \boldsymbol{v}(\boldsymbol{x}; \boldsymbol{\beta}_0, \boldsymbol{\omega}_0, \boldsymbol{\xi}_0)$.

While identification of $\boldsymbol{\delta}$ depends solely on the marginals, the dependence structure parameter is more sensitive to discontinuities. In particular, this identification can be compromised when the covariates do not allow a wide range of the $[0,1]$-domain to be covered in the regression structures exploited in this paper (Genest and Nešlehová, 2007; Trivedi and Zimmer, 2017). Point masses on the marginal distributions could potentially be accommodated by robust correction techniques (Martín-Fernández et al., 2003) or in a Bayesian setting by data augmentation (Smith and Khaled, 2012). All link functions usually considered in the literature satisfy Assumption 2.5.2.(i). These include functions on a single-index or those including additional parameters in reduced form models, such as the nested logit or dogit models (Murteira and Ramalho, 2016). A simple way to guarantee 2.5.2.(ii) is to have a continuous regressor with unbounded support and a nonzero coefficient associated with it.

Combining all previous assumptions with the standard regularity conditions (see Appendix 2.B and White, 1982) leads to one of the main results of the paper.

Theorem 2.1. *Under Assumptions 2.1–2.6.A and regularity conditions R1–R6, the resulting estimators* $\widehat{\boldsymbol{\theta}}_Y$ *and* $\widehat{\boldsymbol{\theta}}_Z$ *are consistent and asymptotically normal; i.e., for* $e \in \{Y, Z\}$, $\widehat{\boldsymbol{\theta}}_e \xrightarrow{p} \boldsymbol{\theta}_{e,0}$, *and*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_e - \boldsymbol{\theta}_{e,0}) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_{e,0}))\,, \tag{2.16}$$

where $\mathcal{I}(\boldsymbol{\theta}_{e,0}) = -\mathrm{E}[\partial^2 \ell(\boldsymbol{\theta}_{e,0})/\partial\boldsymbol{\theta}_e\partial\boldsymbol{\theta}_e']$ *is the Fisher information matrix at the true parameter vector.*

Inference is easily obtained by plugging in $-\partial\ell(\widehat{\boldsymbol{\theta}}_e)/\partial\boldsymbol{\theta}_e\partial\boldsymbol{\theta}_e'$ as an estimator for $\mathcal{I}(\boldsymbol{\theta}_{e,0})$, where $e \in \{Y, Z\}$. Now, as the focus of the paper is estimating the coefficients associated to the conditional mean, the full strength of Assumption 2.6.A is not necessary to obtain consistency and asymptotic normality of the estimator from the copula on $\boldsymbol{Y}$. A modified version of Assumption 2.6.A is introduced next.

Assumption 2.6.B. (Possibly misspecified copula) There exists $\boldsymbol{\delta}_0 = (\boldsymbol{\delta}_{0,1}', \ldots, \boldsymbol{\delta}_{0,D}')' \in \Delta$ such that $H_j(\cdot|\boldsymbol{X} = \boldsymbol{x}) = F_j(\cdot|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_{0,j})$ for all $j = 1, \ldots, d$ and almost all $\boldsymbol{x} \in \mathcal{X}$. However, $C(\cdot) \neq C_Y(\cdot; \boldsymbol{\psi}_0)$ for all $\boldsymbol{\psi}_0 \in \Psi$.

The following lemma will be useful in proving an analog to Theorem 2.1 that uses Assumption 2.6.B instead of 2.6.A. It presents a decomposition of the Kullback-Leibler (KL) divergence when dealing with copula estimation, where the KL divergence between two distributions h and f, indexed by some parameter vector $\boldsymbol{\theta}$, is defined as follows: $\mathrm{KL}(h, f; \boldsymbol{\theta}) = \mathrm{E}_h[\log(h/f)]$, with E_h denoting that the expectation is taken with respect to distribution h.

Lemma 2.1. (KL divergence for copula likelihoods) *Under Assumptions 2.1–2.3.A and regularity conditions R1 and R2, the KL divergence between the true distribution* h*, when* f *is defined by* (2.8)*, is given by*

$$\begin{aligned}\mathrm{KL}(h, f; \boldsymbol{\theta}_Y) = \mathrm{E}_h\left[\log \frac{c(H_1(Y_1|\boldsymbol{X}=\boldsymbol{x}), \ldots, H_D(Y_D|\boldsymbol{X}=\boldsymbol{x}))}{c_Y(F_1(Y_1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_1), \ldots, F_D(Y_D|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_D);\boldsymbol{\psi})}\right] + \\ \sum_{j=1}^{D} \mathrm{KL}(h_j, f_j; \boldsymbol{\delta}_j) + \mathrm{E}_h\left[\log \frac{F_W(1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\theta}_Y)}{\mathbb{I}(\boldsymbol{Y} \in \mathcal{T})}\right].\end{aligned} \tag{2.17}$$

The main message from Lemma 2.1 is that the KL divergence can be decomposed into three parts: the first term represents a measure of the divergence between the true and the assumed copula;

the second are the actual KL divergences between the true and assumed marginals; and the third is the difference between the true and derived log-probability that $\boldsymbol{y}$ is in the set $\mathcal{T}$. Using this result, it is now possible to show that, as long as the marginals are correctly specified even if the copula is not, the coefficients $\boldsymbol{\theta}_Y$ can be consistently recovered. In such a case, the $\widehat{\boldsymbol{\delta}}$ parameters in the marginals converge to their true counterpart, while the dependence structure parameters $\widehat{\boldsymbol{\psi}}$ converge to the pseudo-true values that minimize the KL divergence along that dimension. In this sense, the proposed estimator is semiparametric with respect to the copula; i.e., robust to copula misspecification.

Theorem 2.2. *Under assumptions 2.1–2.3.A, 2.4–2.6.B and regularity conditions R1–R6, the resulting estimator* $\widehat{\boldsymbol{\theta}}_Y$ *is consistent and asymptotically normal. In particular,* $\widehat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}_0$ *and* $\widehat{\boldsymbol{\psi}} \xrightarrow{p} \boldsymbol{\psi}^*$*, where* $\boldsymbol{\psi}^*$ *is the value of* $\boldsymbol{\psi} \in \Psi$ *that minimizes the Kullback-Leibler divergence. Additionally,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_Y - \boldsymbol{\theta}_Y^*) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}_h^{-1}(\boldsymbol{\theta}_Y^*)\mathcal{J}_h(\boldsymbol{\theta}_Y^*)\mathcal{I}_h^{-1}(\boldsymbol{\theta}_Y^*))\,, \tag{2.18}$$

where $\boldsymbol{\theta}_Y^* = (\boldsymbol{\delta}_0', \boldsymbol{\psi}^{*\prime})'$ *is the pseudo-true value,* $\mathcal{I}_h(\boldsymbol{\theta}_Y^*) = \mathrm{E}_h[\partial^2 \log f(\boldsymbol{y}_i|\boldsymbol{X} = \boldsymbol{x}_i; \boldsymbol{\theta}_Y^*; \mathcal{T})/\partial\boldsymbol{\theta}_Y \partial\boldsymbol{\theta}_Y']$ *and* $\mathcal{J}_h(\boldsymbol{\theta}_Y^*) = \mathrm{E}_h[\partial \log f(\boldsymbol{y}_i|\boldsymbol{X} = \boldsymbol{x}_i; \boldsymbol{\theta}_Y^*; \mathcal{T})/\partial\boldsymbol{\theta}_Y \cdot \partial \log f(\boldsymbol{y}_i|\boldsymbol{X} = \boldsymbol{x}_i; \boldsymbol{\theta}_Y^*; \mathcal{T})/\partial\boldsymbol{\theta}_Y']$.

Theorem 2.2 is a specialization of the results in White (1982), tackling misspecified maximum likelihood estimation, and thus expected values are taken with respect to the true underlying joint distribution h. This result represents an additional advantage in this context, as some copulas have a truncation probability, $F_W(1|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\psi})$ in (2.13), which is easier to compute than others. Using these copulas will still recover the underlying marginal parameters while ensuring that the dependence parameters are consistent to a meaningful counterpart; the computational burden is therefore reduced. Furthermore, in the copula estimation context, it is not generally the case that $\mathcal{I}_h(\boldsymbol{\theta}_Y^*)$ has a block-diagonal structure, so that the full sandwich estimator is necessary to conduct inference regarding $\boldsymbol{\beta}$. Consistent estimators of these matrices can be computed in a standard fashion by using

$$\widehat{\mathcal{I}}_h(\widehat{\boldsymbol{\theta}}_Y) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log f(\boldsymbol{y}_i|\boldsymbol{X}=\boldsymbol{x}_i;\widehat{\boldsymbol{\theta}}_Y;\mathcal{T})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\,,$$
$$\widehat{\mathcal{J}}_h(\widehat{\boldsymbol{\theta}}_Y) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f(\boldsymbol{y}_i|\boldsymbol{X}=\boldsymbol{x}_i;\widehat{\boldsymbol{\theta}}_Y;\mathcal{T})}{\partial\boldsymbol{\theta}} \cdot \frac{\partial \log f(\boldsymbol{y}_i|\boldsymbol{X}=\boldsymbol{x}_i;\widehat{\boldsymbol{\theta}}_Y;\mathcal{T})}{\partial\boldsymbol{\theta}'}\,. \tag{2.19}$$

It is also simple to see why Theorem 2.2 does not apply to the estimator based on the copula on $\boldsymbol{Z}$. As Proposition 2 shows, the marginal parameters depend on the underlying copula parameters $\boldsymbol{\xi}$ via $\boldsymbol{\delta} = \boldsymbol{v}(\boldsymbol{x};\boldsymbol{\beta},\boldsymbol{\omega},\boldsymbol{\xi})$. If no $\boldsymbol{\xi} \in \Xi$ allows for a correct specification of the copula, the inferred relationship cannot reflect the correct marginal structure. The preceding theorems introduce a trade-off in the empirical analysis of copulas for demand estimation or reduced form models. While the estimator of the copula on $\boldsymbol{Y}$ is robust to copula misspecification, it is more expensive to compute. On the other hand, placing a copula on $\boldsymbol{Z}$, particularly an elliptical copula, creates an easier to compute model; however, it might be biased for computing the coefficients of interest. This trade-off is explored numerically in Section 2.3 using Monte Carlo simulations.

This theorem also presents a powerful result whose proof is generally applicable to copula estimation: correct marginals with misspecified dependence structure still leads to consistent and asymptotically normal estimators. The result is formally stated in the next corollary.

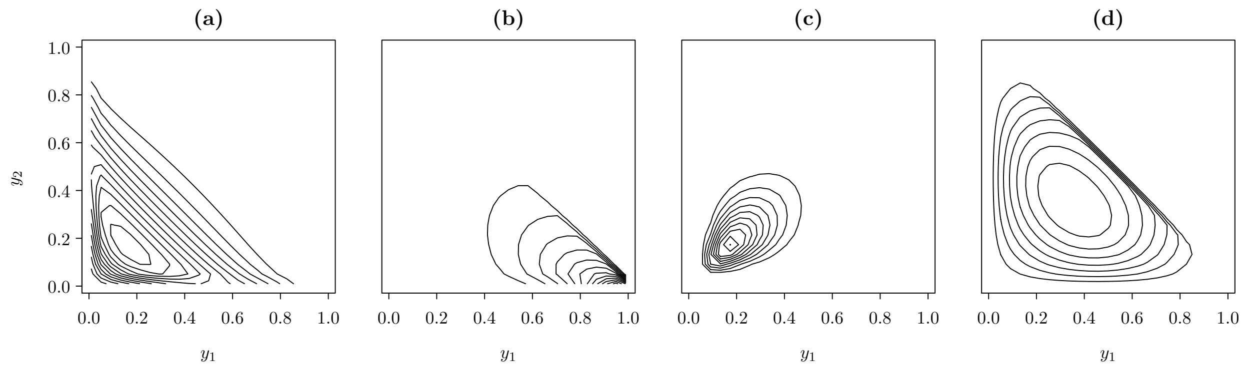
Corollary 2.1. *Let the support of* $\boldsymbol{Y}$ *be* $\mathbb{R}^D$ *instead of* $\mathcal{S}^d$. *Under Assumptions 2.2, 2.3.A, 2.4, 2.5.1, 2.6.B and regularity conditions R1–R6, an estimator* $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\delta}}', \widehat{\psi}')'$ *based on* (2.13) *(without the truncation probability) is consistent and has an asymptotically normal distribution as in* (2.18).

This is a potentially overlooked result in the copula estimation literature, as most attention is centered on correctly modeling the dependence structure without focusing on the marginals.[4] Corollary 2.1 presents a contrasting view: if the attention is shifted to the marginals, the copula specification parameters become nuisance parameters and the marginals can be recovered.

The estimators introduced in this paper cover several important cases in the literature. Several marginals can be chosen such that the regression structure given in (2.1) is preserved. Examples include the beta with a reparametrization (Ferrari and Cribari-Neto, 2004; Simas et al., 2010),

[4]This view is one usually found in most financial or actuarial applications, while the opposite tends to be true in economics and econometrics (Charpentier et al., 2007; Trivedi and Zimmer, 2007).

Figure 2.1: Dependence Patterns in Copulas



Note: (a) Beta marginals with $\boldsymbol{\delta}_1 = (0.5, 10), \boldsymbol{\delta}_2 = (0.5, 10)$ and a normal copula with $\psi = -0.5$; (b) Beta marginals with $\boldsymbol{\delta}_1 = (0.7, 10), \boldsymbol{\delta}_2 = (0.2, 10)$ and a normal copula with $\psi = -0.5$; (c) Simplex marginals with $\boldsymbol{\delta}_1 = (0.5, 1), \boldsymbol{\delta}_2 = (0.5, 1)$ and a normal copula with $\psi = 0.5$; and (d) Beta marginals with $\boldsymbol{\delta}_1 = (0.8, 10), \boldsymbol{\delta}_2 = (0.8, 10)$ and a FGM copula with $\psi = -0.5$.

simplex (Song and Tan, 2000; Liu et al., 2020), truncated normals, and skew-normals (Martínez-Flórez et al., 2020). Furthermore, there are many methods to create new distributions on the unit interval that satisfy this restriction (Rodrigues et al., 2020). Some distributions can even be made to handle point masses at the extremes to deal with boundary values that can occur in the data and that can be hard to introduce into a parametric analysis (Papke and Wooldridge, 1996; Martín-Fernández et al., 2003; Smithson and Shou, 2017). Once these marginals are selected, general copulas can be used to link them in a flexible way. As an example of this flexibility inherent to the copula approach, Figure 2.1 plots the densities under several configurations of marginals, copulas, and their parameters, obtaining a wide array of possible distributional shapes.

Example 1. (Continued) Now, as one of the objectives of the paper is to be able to deal with the type of cross-equation restrictions that arise in the estimation of demand systems, it will be useful to consider the more general estimator for $e \in \{Y, Z\}$ given by

$$\begin{aligned} \widetilde{\boldsymbol{\theta}} &\equiv \underset{\boldsymbol{\theta}_e \in \Theta_e}{\arg\max}\, \ell_e(\boldsymbol{\theta}) \\ &\quad \text{subject to } \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{a} \text{ and } \boldsymbol{B}\boldsymbol{\beta} \leq \boldsymbol{b}\,, \end{aligned} \tag{2.20}$$

where $\Theta_Y = \Delta \times \Psi$ and $\Theta_Z = \Delta \times \Xi$. Implementation of these types of (possible) cross-equation restrictions is simple in the full-likelihood estimation case. This is in contrast to the alternative two-step approach known in the literature as inference functions for margins (IFM), which first estimates $\boldsymbol{\delta}$ and then $\boldsymbol{\psi}$ or $\boldsymbol{\xi}$ (Joe and Xu, 1996). Imposition of cross-equation restrictions in this framework is complicated and usually leads to larger efficiency losses (Joe, 2014). However, an issue with the full estimator is numerical instability. The Bayesian approach can further aid in this issue, as the introduction of prior information usually leads to posteriors that are less flat than the likelihood in the regions of the parameter space that are of interest.[5]

2.2 Priors and Variable Selection

Armed with the likelihood function, prior distributions on the parameters can be imposed to carry out Bayesian estimation, which produces posterior distributions for $\boldsymbol{\theta}$. Inference then follows from a measure of uncertainty or from credible sets of these posterior distributions. Model selection in a traditional sense would follow from the same probability rules and yield posterior model probabilities that could be used for both selection and averaging. Instead, the objective of this paper is to further augment the proposed estimators to handle covariate selection by introducing regularization. This is done to leverage recent results on Bayesian analogs of the LASSO and related estimation methods (Tibshirani, 1996). Furthermore, the Bayesian framework allows the researcher to obtain statistical inference through simple numerical methods. Such a framework would be useful even in contexts where the dimensionality of the covariate space is large or grows with sample size, as occurs in high-dimensional settings (Li and Lin, 2010). In demand estimation, this could correspond to approximating the indirect utility or cost functions to an arbitrarily large degree of precision using polynomials and interaction terms, which can aid the performance and economic regularity of the resulting models (Chang and Serletis, 2014). Additionally, a researcher would need to obtain inference on functions of the parameters, such as the price elasticities in demand estimation or average partial effects in reduced form models. Frequentist methods rely on the Delta method or variants of bootstrapping to produce this inference, but they are either computationally complex or not supported theoretically.[6] On the other hand, Bayesian methods can produce inference for these objects at no real additional computational cost apart from the

[5] This property of Bayesian methods have made them very popular in macroeconomic modeling (see, e.g., Sims and Zha, 1998).

[6] For example, Koch (2015) and Mullahy (2015) deal with inference on the average partial effects for the multivariate fractional logit by using different kinds of bootstrap methods. However, the validity of these bootstrap methods is never assessed.

estimation itself.

The driving idea behind this framework is that regularization can be applied to any globally convex function, such as the negative of the log-likelihoods given in (2.13) and (2.14) (Zou and Hastie, 2005; Tibshirani et al., 2012). Thus, to automatically include a selection step, the objective function could be augmented to solve

$$\arg\min_{\boldsymbol{\theta}_e \in \Theta_e} \{-\ell_e(\boldsymbol{\theta}_e) + \rho_{\boldsymbol{\lambda}}(\boldsymbol{\beta})\}\,, \tag{2.21}$$

where the covariates are now assumed to be standardized and $\rho_{\boldsymbol{\lambda}}(\boldsymbol{\beta})$ is a penalization term of the regression coefficients that is indexed by a vector of regularization parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_M)'$. It is assumed that only the $\boldsymbol{\beta}$ or a subset of them are penalized, as these coefficients directly interact with the covariates to define the conditional mean.

Example 4. (LASSO and group LASSO) Useful forms of the penalty could be given by

$$\rho_{\boldsymbol{\lambda}}(\boldsymbol{\beta}) = \lambda||\boldsymbol{\beta}||_1 \qquad \text{or} \qquad \rho_{\boldsymbol{\lambda}}(\boldsymbol{\beta}) = \lambda \sum_{l=1}^{L} ||\boldsymbol{\beta}_l||_2\,, \tag{2.22}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_L')'$ so that there is a partition of the coefficient vector into L groups and $||\cdot||_1$ and $||\cdot||_2$ are the L^1 and L^2 norms in Euclidean spaces, respectively. The first penalty is the usual LASSO, whereas the second takes the form of the group LASSO (Yuan and Lin, 2006).

While frequentist methods can be used to solve (2.21), a Bayesian solution to this problem is still attractive. Frequentist penalization methods act such as LASSO act by simultaneously imposing shrinkage and selecting relevant features. The Bayesian framework can also naturally impose shrinkage into estimation by virtue of prior information. Recent literature shows how this pattern of Bayesian shrinkage can replicate those introduced by LASSO or its alternatives and how selection can be achieved (Park and Casella, 2008; Li and Lin, 2010; Leng et al., 2014). The connection between both methods was recognized at the onset of the penalized regression literature and the introduction of the LASSO, which can be obtained from a Bayesian interpretation (Tibshirani, 1996; Ročková and George, 2018).

However, the main consideration for adopting a Bayesian framework is its ability to obtain inference through simple probabilistic concepts (Kyung et al., 2010). Frequentist methods initially focused on fast coefficient estimation and tuning of the penalty parameters, but were generally unsuited for inference due to their nonstandard limiting distribution (Knight and Fu, 2000). Advancements in the literature have introduced different ways to circumvent this issue. These include approximations to the objective function (Tibshirani, 1996; Osborne et al., 2000; Wang and Leng, 2007), bootstrap (Knight and Fu, 2000; Hansen and Liao, 2019), use of nonconcave penalties (Fan and Li, 2001; Ning et al., 2017), inversion of Karush-Kuhn-Tucker conditions (also known as "desparsification", Javanmard and Montanari, 2014; van de Geer et al., 2014; Zhang and Zhang, 2014; Breunig et al., 2020), post-selection inference (Belloni et al., 2014, 2016; Lee et al., 2016), and double or debiased machine learning (Athey et al., 2018; Chernozhukov et al., 2018).[7] Most of these advancements involve linear regression and instrumental variable models, while some cover up to generalized linear models, which provide sufficient structure to the problem (Fan and Tang, 2013; Ning et al., 2017). The regression structure with the likelihood functions considered in this paper do not fall into these categories. Furthermore, the necessary technical conditions to adapt some of the previous methods that are sufficiently general to cover this setting are still unknown and left for future research. A Bayesian specification, on the other hand, is easy to establish without additional technical considerations and provides statistical inference as a by-product of the estimation algorithm. Additionally, the Bayesian framework can attach uncertainty to the estimates of nonselected variables — those estimated to be 0 — whereas this cannot be done satisfactorily under most methods in the frequentist approach. While this paper implements model selection by using the class of priors defined below in (2.23), several alternatives exist within the Bayesian literature (Chipman et al., 2001; Ishwaran and Rao, 2005; Yuan and Lin, 2006; Yen, 2011; Ročková and George, 2018).

To complete a Bayesian specification of the problem, this paper considers a general class of priors that implement regularization in an analog way to the usual frequentist solutions. For simplicity, it is assumed hereafter that the marginals can be entirely described, conditional on $\boldsymbol{X}$, by using the vector of coefficients $\boldsymbol{\beta}$ and precision parameters $\boldsymbol{\phi} = (\phi_1, \dots, \phi_D) \in \Phi \subset \mathbb{R}^D$. That is, we can write $\boldsymbol{\delta}_j = (\boldsymbol{\beta}', \phi_j)'$ for all $j = 1, \dots, d$, or $\boldsymbol{\delta} = (\boldsymbol{\beta}', \boldsymbol{\phi}')'$. The $\boldsymbol{\phi}$ are precision parameters such that for a fixed mean, larger $\boldsymbol{\phi}$ imply smaller variances and as $\boldsymbol{\phi} \to \infty$, the distribution degenerates

[7] Double machine learning methods are also connected to resampling ideas, which can be given a Bayesian interpretation (Smith and Gelfand, 1992).

to the mean value (Ferrari and Cribari-Neto, 2004). This is the case for all marginal distributions considered in the paper.

Most work on adapting the LASSO-type estimators to a Bayesian context shows that, essentially, different penalties are implemented by changing the priors in a systematic way (Park and Casella, 2008; Hans, 2009; Kyung et al., 2010). Furthermore, different representations of the Bayesian interpretation of the priors alters both the theoretical and computational properties of the solutions. This idea leads to the following general class of priors $\pi(\boldsymbol{\beta})$ to handle estimation and model selection in this framework:

$$\pi(\boldsymbol{\beta}) \propto \exp\left\{ -\frac{1}{2}\rho_{\boldsymbol{\lambda}}(\boldsymbol{\beta})\right\}. \tag{2.23}$$

Example 4. (Continued) For the penalties in (2.22), these priors can be implemented using a hierarchical Bayesian approach. For a LASSO penalty, the following hierarchy achieves the desired results:

$$\begin{aligned} \boldsymbol{\beta}|\tau_1, \dots, \tau_K &\sim \mathcal{N}_K(\mathbf{0}, D_\tau), D_\tau = \text{diag}(\tau_1, \dots, \tau_K)\,, \\ \tau_k|\lambda^2 &\sim \text{Exponential}\left(\frac{\lambda^2}{2}\right), k = 1, \dots, K\,, \end{aligned}$$

where $\mathcal{N}_K$ represents a multivariate K-dimensional normal distribution, $\tau_1, \dots, \tau_K$ are hierarchical parameters, and $\text{diag}(\tau_1, \dots, \tau_K)$ represents a $K \times K$ diagonal matrix with the diagonal given by its arguments. This hierarchical structure borrows from the linear regression framework, but its properties hold remarkably well in these nonlinear settings (Park and Casella, 2008). For the group-LASSO penalty, a similar structure can implement this prior distribution:

$$\begin{aligned} \boldsymbol{\beta}_l|\tau_l &\sim \mathcal{N}_{L_l}(\mathbf{0}, \tau_l I_{L_l}), l = 1, \dots, L\,, \\ \tau_l|\lambda^2 &\sim \text{Gamma}\left(\frac{L_l+1}{2}, \frac{\lambda^2}{2}\right), l = 1, \dots, L\,, \end{aligned}$$

where L_l is the number of elements of each group, there are a total of L groups, and I_{L_l} is the identity matrix of order L_l (Kyung et al., 2010; Leng et al., 2014).

Thus, the complete specification would yield $\pi(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\psi}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\phi})\pi(\boldsymbol{\psi})$. Priors on $\boldsymbol{\phi}$ can be placed in a standard fashion for each precision parameter; say, by choosing a flat Jeffrey's prior, a

Gamma distribution, or an adjusted Scaled-Beta2 distribution (Pérez et al., 2016; Ramírez-Hassan and Montoya-Blandón, 2020). The prior on $\boldsymbol{\xi}$, on the other hand, is dependent on the class of copula functions considered. For example, for a Gaussian copula whose dependent structure is characterized by a correlation matrix, a plausible prior could be given like the one in Lewandowski et al. (2009). If $d = 3$ so that only $D = 2$ shares need to be modeled, the dependence reduces to a single correlation parameter and flexible alternatives can be placed as priors, such as a diffuse uniform distribution on the support $[-1, 1]$ or (modified) beta distribution (LeSage, 2004; Smith and Khaled, 2012). Additionally, in the Bayesian framework, the tuning parameters $\boldsymbol{\lambda}$ can either be chosen by a suitable method such as the expectation-maximization (EM) algorithm or they can be given hierarchical priors to remain fully consistent with the paradigm. Given the complex nonlinear nature of the likelihood function constructed in this paper, it becomes simpler to tune a hyperprior for $\boldsymbol{\lambda}$. The most popular example sets a gamma prior on λ^2 for both LASSO and group-LASSO penalty parameters (Park and Casella, 2008; Kyung et al., 2010). Finally, although constraints can be implemented in a frequentist solution to (2.21) as in Gaines et al. (2018), Bayesian constraints are also consistently implemented as support restrictions on the prior distributions.[8]

Example 1. (Continued) There are meaningful ways in which sparsity and selection can play a role in the estimation of structural demand models. Consider the matrix form of the AID equations (2.2). Assuming that the expenditure and price variables are already defined in terms of their logarithms, we can write $\tilde{e} \equiv e - \alpha_0 - \boldsymbol{\alpha}'\boldsymbol{p} - (1/2)\boldsymbol{p}'\Gamma\boldsymbol{p}$ so that $\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{\alpha} + \Gamma\boldsymbol{p} + \boldsymbol{\pi}\tilde{e}$. One could allow further flexibility into the model by allowing polynomials on $\tilde{\boldsymbol{e}}$ of varying degrees, such as Blundell et al. (1993), which includes a second degree term, or Lewbel and Pendakur (2009), which empirically decide on including up to 5 terms.[9] Incorporating these ideas, one could in general write

$$\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{\alpha} + \Gamma\boldsymbol{p} + \sum_{r=1}^{R} \boldsymbol{\pi}_r \tilde{e}^r \,, \tag{2.24}$$

with $\boldsymbol{\beta} = (\alpha_0, \boldsymbol{\alpha}', \Gamma, \boldsymbol{\pi}_1', \ldots, \boldsymbol{\pi}_R')'$. It is then apparent that choosing R is a model selection issue

[8] For example, in the context of demand estimation, curvature can be imposed via support restrictions in the AID model (Geweke, 1989; Tiffin and Aguiar, 1995).

[9] While these models are derived from different structural assumptions compared to the AID system, this framework is kept for simplicity.

that could be undertaken using the penalties in (2.22). The group LASSO penalty is particularly suitable as one would naturally select or exclude together the d-dimensional vectors $\boldsymbol{\pi}_r$ from all equations.

Example 2. (Continued) In a similar fashion, the reduced form approach outlined in (2.3) could benefit from the feature selection accomplished by the class of priors considered in this paper. Letting the dimensionality p of the covariate vector $\boldsymbol{x}$ be large and assuming there are some redundant variables that should be excluded from the model, the penalized model will be more suitable. Furthermore, this setup also naturally lends itself to a grouped penalty structure, as the coefficients associated to the same variable in different equations can be placed together to form each group. Furthermore, if the goal is to introduce a correlation between the selected coefficients in a more structured manner, the fused-LASSO penalty of Tibshirani et al. (2005) could also be introduced. In all cases, λ controls the strength of the regularization imposed into each penalty.

Based on previous considerations, the following steps summarize a way to estimate and obtain inference for the Bayesian regularized copula regression model:

Step 1. Let $\mathcal{F}$ represent the class of marginal distributions satisfying the fractional and index restrictions (2.1). Choose $F_j, G_j \in \mathcal{F}$ for all $j = 1, \ldots, D$.

Step 2. Let $\mathcal{C}_D$ represent a class of copula functions of dimension D. Choose $C_Y, C_Z \in \mathcal{C}$. Together with the previous step, this allows us to find likelihood functions $f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\psi})$ and $g(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\xi})$ by (2.13) and (2.14).

Step 3. Choose a prior distribution $\pi(\boldsymbol{\theta}_Y)$ and $\pi(\boldsymbol{\theta}_Z)$ that belongs to the class outlined in (2.23). If constraints of the form $\boldsymbol{A\beta} = \boldsymbol{a}$ and $\boldsymbol{B\beta} \leq \boldsymbol{b}$ are present, the support of the prior distribution should be modified to the set $\mathcal{A}$ such that these constraints hold. Include a prior distribution for $\boldsymbol{\lambda}$.

Step 4. Combine the likelihood function and the prior distribution via Bayes's theorem to obtain the posterior distribution $\pi(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\psi}|\boldsymbol{Y}, \boldsymbol{X})$ and $\pi(\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\xi}|\boldsymbol{Y}, \boldsymbol{X})$. Point estimates $\check{\boldsymbol{\theta}}$ can be obtained as the mean, median, or mode from the posterior.[10] Inference can be obtained as

[10] The posterior mean is optimal in a decision-theoretic framework as it minimizes the squared loss. Similarly, the median minimizes the absolute value loss and the posterior mode does so with a zero-one loss. In particular, most Bayesian LASSO analogs target a mode interpretation to their frequentist counterparts but use the posterior mean and median for simplicity.

a credible set of the posterior; for example, using a highest posterior density interval of a given probability coverage.

A second way to implement a Bayesian solution is through the use of a least squares approximation (Wang and Leng, 2007; Leng et al., 2014). Given Assumptions 2.1–2.6.A, the likelihood function can be approximated by a Taylor expansion as

$$\ell_e(\boldsymbol{\theta}_e) \approx L(\widehat{\boldsymbol{\theta}}_e) + \frac{1}{2}(\boldsymbol{\theta}_e - \widehat{\boldsymbol{\theta}}_e)'\mathcal{I}(\widehat{\boldsymbol{\theta}}_e)(\boldsymbol{\theta}_e - \widehat{\boldsymbol{\theta}}_e)\,, \tag{2.25}$$

where $\widehat{\boldsymbol{\theta}}_e$ is the MLE in (2.15) for $e \in \{Y, Z\}$. Employing the same algorithm outlined previously with this expansion of the likelihood yields an approximate Bayesian solution for which closed form conditionals exist. Thus, this procedure could be implemented via a simpler Gibbs-sampling algorithm for which theoretical properties are readily available.

Furthermore, by virtue of Lemma 2.1 and standard results for parametric Bayesian estimators, Bayes estimates $\check{\boldsymbol{\theta}}$ found from this algorithm are also consistent (Strasser, 1981; Bunke and Milhaud, 1998). For convenience, this is stated in the following theorem.

Theorem 2.3. *(i) Under assumptions 2.1–2.6.A and regularity conditions R1–R3 and R7–R9, then $\check{\boldsymbol{\theta}}_e$, defined as a mean, median, or mode of the posterior distribution $\pi(\boldsymbol{\theta}_e|\boldsymbol{Y}, \boldsymbol{X})$, is consistent; i.e., $\check{\boldsymbol{\theta}}_e \xrightarrow{p} \boldsymbol{\theta}_{e,0}$, for $e \in \{Y, Z\}$.*

(ii) Under Assumptions 2.1–2.3.A, 2.4–2.6.B and regularity conditions R1–R3 and R7–R9, then $\check{\boldsymbol{\theta}}_Y$ as defined above, is consistent to the minimizer of the Kullback-Leibler divergence; i.e., $\check{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^*_Y$, where $\boldsymbol{\theta}^*_Y = (\boldsymbol{\delta}'_0, \boldsymbol{\psi}^{*\prime})'$.

2.3 Monte Carlo Study

To test the performance of the estimator defined by (2.15) as well as the theoretical properties found in the previous two sections, a range of numerical exercises is conducted. These follow the structure of Examples 1 and 2, and change the form of the conditional mean function. Data are simulated from several scenarios that maintain the conditional mean as correctly specified; link function misspecification would be a source of bias distinct to likelihood misspecification (Montoya-Blandón and Jacho-Chávez, 2020). Numerical optimization of the log-likelihoods (2.13) and (2.14) produce estimates $\widehat{\boldsymbol{\theta}}_e$ for $e \in \{Y, Z\}$. To simplify the exposition of the results, the main estimation method used is one that assumes a Gaussian copula and beta marginals. That is, the copula density $c_e(\cdot)$ takes the form

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

where $u_j, j = 1, \ldots, D$ are the pseudo-observations found by transforming the variables through a distribution function, R is a $D \times D$ correlation matrix with elements in the lower triangular block given by the vector of copula parameters $\boldsymbol{\psi}$, and $\Phi^{-1}(\cdot)$ is the quantile function for the standard normal distribution. The pseudo-observations are computed using the marginal distributions; in this case, a beta in a mean-precision parameterization so that for each j in $1, \ldots, D$, u_j is given by

$$u_j \equiv \int_0^{y_j} \frac{\Gamma(\phi_j)}{\Gamma[m_j(\boldsymbol{x};\boldsymbol{\beta}\phi_j)]\Gamma[[1 - m_j(\boldsymbol{x};\boldsymbol{\beta})]\phi_j]} t^{m_j(\boldsymbol{x};\boldsymbol{\beta})\phi_j} (1-t)^{[1-m_j(\boldsymbol{x};\boldsymbol{\beta})]\phi_j} \, \mathrm{d}t \ ,$$

where $\Gamma(\cdot)$ is the gamma function. Additional combinations using different marginals and copulas, along with other extensions, can be found in Appendix 2.C.

2.3.1 Reduced Form

Due to the ease of simulating from a reduced form setup, the paper focuses on this example first. A multivariate fractional logit structure as in (2.3) is imposed for $d = 3$ shares; i.e.,

$$\mathrm{E}[Y_1 | \boldsymbol{X} = \boldsymbol{x}] = \frac{\exp(\boldsymbol{x}'\boldsymbol{\beta}_1)}{1 + \exp(\boldsymbol{x}'\boldsymbol{\beta}_1) + \exp(\boldsymbol{x}'\boldsymbol{\beta}_2)} ,$$
$$\mathrm{E}[Y_2 | \boldsymbol{X} = \boldsymbol{x}] = \frac{\exp(\boldsymbol{x}'\boldsymbol{\beta}_2)}{1 + \exp(\boldsymbol{x}'\boldsymbol{\beta}_1) + \exp(\boldsymbol{x}'\boldsymbol{\beta}_2)} ,$$

and $\mathrm{E}[Y_3|\boldsymbol{X} = \boldsymbol{x}] = 1 - \mathrm{E}[Y_1|\boldsymbol{X} = \boldsymbol{x}] - \mathrm{E}[Y_2|\boldsymbol{X} = \boldsymbol{x}]$. True coefficient values are set at $\boldsymbol{\beta}_1 = (-1, 0.5, 0)$ and $\boldsymbol{\beta}_2 = (-1.5, 0, 0.5)$. Two covariates, x_1 and x_2, are generated independently from a standard normal distribution. For the first exercise, beta marginals with a mean-precision parameterization are used, setting $\phi_1 = \phi_2 = 10$. A Gaussian copula with a correlation parameter of $\psi = 0.5$ links the two free marginals together. Values for $\boldsymbol{y}$ are generated via rejection sampling for sample sizes $n \in \{100, 200, 400, 800\}$ and 1,000 simulations under this setting. No constraints are set on $\boldsymbol{\beta}$ but the natural nonnegativity constraints on $\boldsymbol{\phi}$ and ψ belonging to $(-1, 1)$ are imposed to guarantee numerical stability. Aside from the copula estimators introduced in this paper, several competing estimation methods are implemented. First, the multivariate fractional quasi-likelihood method (Mullahy, 2015; Murteira and Ramalho, 2016) is estimated as a flexible alternative and multivariate generalization of the popular estimator proposed by Papke and Wooldridge (1996). This estimator should remain consistent regardless of the generating distribution as it only relies on a correctly specified conditional mean. The next method is a Dirichlet distribution using a parameterization similar to the beta (Hijazi and Jernigan, 2009; Murteira and Ramalho, 2016). As a Dirichlet distribution is a special case of the beta marginals with a copula on Z, their performance should be similar. Finally, the additive log-ratio transformation regression of Aitchison (1982) is used as a simple alternative that requires no real modeling choice. This procedure is equivalent to a SUR model on the transformed outcomes; given the assumption of common covariates across shares, it further simplifies to estimating D equations by ordinary least squares (OLS). However, as previously noted, this procedure will not recover the true conditional mean.

Results from this first exercise are presented in Table 2.1 in terms of the root mean squared error (RMSE) across 1,000 simulations. We can observe the consistency of the proposed methods as the RMSE shrinks at an expected rate. In general, the copula estimators outperform the other likelihood-based methods and are chosen as preferable by the Akaike and Bayesian information criteria (AIC and BIC, respectively). The logistic normal distribution remains inconsistent and performs poorly in comparison to the other methods.

As a second exercise, consider what happens when, under a similar setting to before, the copula function is changed from a Gaussian to a Farlie–Gumbel–Morgenstern (FGM) copula. As the FGM copula generates relatively low amounts of dependence, its parameter is set to 0.9, which translates to about a 0.3 correlation in a Gaussian distribution. The results are presented in Table 2.2. Now, as expected from Theorem 2.2, the copula on Y remains a consistent estimator, while the copula on Z (and similarly the Dirichlet distribution) are inconsistent and have a reduced performance. Also as expected from the theoretical results, the copula parameter is not recovered in its original scale and thus its RMSE remains high. However, as noted in Table 2.C.2, the estimated copula

Table 2.1: RMSE for Coefficients in a Reduced Form Model from a Gaussian Copula with Beta Marginals

Method	$\beta_{0,1}$	$\beta_{1,1}$	$\beta_{2,1}$	$\beta_{0,2}$	$\beta_{1,2}$	$\beta_{2,2}$	ϕ_1	ϕ_2	$\psi\vert\xi$	AIC	BIC
$n = 100$											
Copula Y	9.102	8.059	8.075	10.878	9.688	9.281	15.720	17.051	20.311	−403.5	−380.1
Copula Z	9.147	8.176	8.093	11.636	11.132	9.230	15.785	67.007	41.837	−338.2	−314.8
MF Logit	9.213	8.718	8.524	11.104	10.822	10.378	—	—	—	—	—
Dirichlet	10.928	8.807	8.485	13.405	9.785	9.895	22.126	—	—	−346.1	−327.9
Logistic Norm.	18.874	17.116	11.402	38.984	17.394	29.083	—	—	—	592.8	608.5
$n = 200$											
Copula Y	6.548	5.558	5.487	7.755	6.857	6.361	11.414	12.151	14.661	−816.7	−787.0
Copula Z	6.436	5.770	5.487	8.056	8.907	6.358	11.430	67.996	38.208	−684.6	−654.9
MF Logit	6.550	6.077	5.835	7.767	7.706	7.286	—	—	—	—	—
Dirichlet	8.525	6.167	5.805	10.789	6.672	6.866	21.307	—	—	−699.1	−676.0
Logistic Norm.	17.289	14.892	7.840	37.735	12.972	26.862	—	—	—	1,188.2	1,208.0
$n = 400$											
Copula Y	5.090	4.014	4.013	6.086	4.849	4.630	8.561	9.437	10.787	−1,643.9	−1,607.9
Copula Z	4.715	4.326	4.016	5.741	7.508	4.700	8.579	68.130	36.581	−1,380.0	−1,344.1
MF Logit	5.071	4.377	4.343	6.057	5.630	5.356	—	—	—	—	—
Dirichlet	7.064	4.597	4.220	9.301	4.741	5.215	20.700	—	—	−1,406.4	−1,378.5
Logistic Norm.	16.612	14.004	5.827	37.208	10.065	25.642	—	—	—	2,378.8	2,402.8
$n = 800$											
Copula Y	3.997	2.785	2.936	4.874	3.451	3.184	6.690	7.375	8.691	−3,291.7	−3,249.6
Copula Z	3.449	3.248	3.010	4.415	6.591	3.493	6.772	68.559	35.274	−2,761.5	−2,719.4
MF Logit	3.896	3.167	3.263	4.776	4.167	3.781	—	—	—	—	—
Dirichlet	6.230	3.430	3.053	8.501	3.301	3.941	20.634	—	—	−2,815.6	−2,782.8
Logistic Norm.	16.108	13.297	4.343	36.877	8.306	24.878	—	—	—	4,762.2	4,790.3

Note: 100 times RMSE for each estimation procedure when data are generated from a Gaussian copula with beta marginals. Akaike and Bayesian information criteria (AIC and BIC, respectively) computed as models have a different amount of parameters to be estimated. For coefficients, "—" implies that the parameter is not part of the model. Information criteria are not computed for the quasi-likelihood method.

parameter is around 0.3, which is the true dependence within the range allowed by the Gaussian copula. It is still the case that the copula model is selected by both information criteria regardless of sample size. In this example, it becomes necessary to adjust inference to control for misspecification, which is readily implemented in the numerical optimization routine used for the paper using (2.19). Inference is not compromised using the estimation method introduced in the paper as standard errors remain close or below those of comparable consistent methods (results on inference for this exercise can be found in Table 2.C.2 in the Appendix).

Moving away from sampling directly from a correctly specified copula likelihood, the next exercise in Table 2.3 draws observations from a Dirichlet distribution. As it is possible to maintain the conditional mean intact under this parameterization, all methods should remain consistent. One of the drawbacks from the Dirichlet distribution is that no pairwise correlation can be positive, something that the previous examples allowed and that could in general occur in an applied setting. This table does not present results for the correlation parameter or second precision parameters as these have no true counterpart. However, in Table 2.C.3 in the Appendix, it is noticeable that the model captures the negative correlation present in the data-generating process with a mean of around -0.4 across the simulations. Once again, this is a manifestation of the theoretical properties derived in Section 2.1.

To produce a Bayesian estimator into this setting, the following setup is used. To streamline the results, only the copula on Y estimator is considered. As the Bayesian estimates are conditional on data, a sample of $n = 800$ is drawn from the setting used in Table 2.1. A Gaussian copula with beta marginals is given as a likelihood and the priors are of the form

$$\begin{aligned}
\beta_{0,j} &\sim \text{Uniform}(-\infty, \infty), j = 1, 2\,, \\
\beta_{k,j} &\sim \mathcal{N}(0, 5) \text{ for } k = 1, 2 \text{ and } j = 1, 2\,, \\
\phi_j &\sim \text{Gamma}(1, 1), j = 1, 2\,, \\
\psi &\sim \text{Uniform}(-1, 1)\,.
\end{aligned}$$

The use of improper prior distributions for the constants is standard in Bayesian analysis and results remain unchanged if a proper prior similar to the other coefficients is assigned. The estimation uses the Hamiltonian Monte Carlo algorithm to sample from the posterior distribution in four chains

Table 2.2: RMSE for Coefficients in a Reduced Form Model from a FGM Copula with Beta Marginals

Method	$\beta_{0,1}$	$\beta_{1,1}$	$\beta_{2,1}$	$\beta_{0,2}$	$\beta_{1,2}$	$\beta_{2,2}$	ϕ_1	ϕ_2	$\psi \vert \xi$	AIC	BIC
$n = 100$											
Copula Y	8.416	8.137	7.792	10.443	9.109	8.925	15.598	15.897	237.074	−380.0	−356.6
Copula Z	9.324	9.183	9.045	12.386	10.562	10.284	16.585	59.563	193.189	−314.8	−291.3
MF Logit	8.620	8.607	8.276	10.834	9.806	10.004	—	—	—	—	—
Dirichlet	9.923	8.342	8.151	12.345	9.238	9.094	17.605	—	—	−351.9	−333.6
Logistic Norm.	18.548	17.507	10.874	38.331	15.572	29.477	—	—	—	604.7	620.3
$n = 200$											
Copula Y	5.934	5.447	5.535	7.210	6.126	6.156	10.689	10.848	237.323	−768.8	−739.1
Copula Z	10.363	9.534	10.871	13.366	9.995	12.092	15.204	62.439	189.734	−626.4	−596.7
MF Logit	6.090	5.942	5.868	7.450	6.875	7.082	—	—	—	—	—
Dirichlet	7.650	5.732	5.878	9.699	6.331	6.384	16.758	—	—	−710.2	−687.2
Logistic Norm.	17.103	15.503	7.875	37.330	11.396	26.971	—	—	—	1,211.3	1,231.1
$n = 400$											
Copula Y	4.586	3.909	4.035	5.505	4.503	4.457	7.336	7.575	237.161	−1,545.7	−1,509.8
Copula Z	10.984	10.574	12.394	15.332	11.219	12.809	15.932	63.952	187.821	−1,241.7	−1,205.7
MF Logit	4.442	4.267	4.258	5.456	5.019	5.005	—	—	—	—	—
Dirichlet	6.535	4.118	4.173	8.545	4.586	4.629	16.839	—	—	−1,424.7	−1,396.8
Logistic Norm.	16.377	14.317	5.529	36.753	9.137	25.857	—	—	—	2,429.2	2,453.2
$n = 800$											
Copula Y	3.114	2.772	2.877	3.790	3.023	3.147	5.440	5.403	237.675	−3,099.7	−3,057.5
Copula Z	10.849	10.296	12.022	15.373	10.865	12.350	15.625	63.269	189.708	−2,486.8	−2,444.6
MF Logit	3.147	3.051	3.086	3.863	3.492	3.616	—	—	—	—	—
Dirichlet	5.656	2.954	3.055	7.560	3.025	3.439	16.703	—	—	−2,857.3	−2,824.5
Logistic Norm.	15.952	13.854	4.033	36.597	7.408	25.327	—	—	—	4,861.6	4,889.7

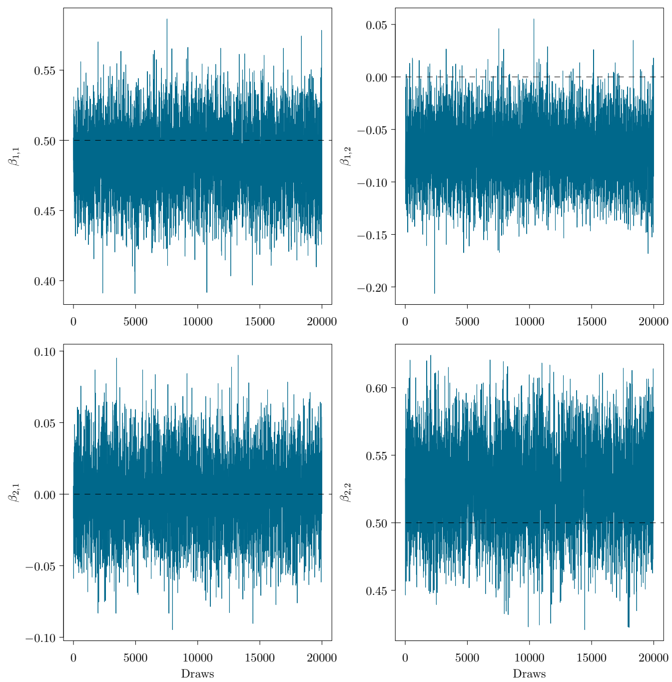
Note: 100 times RMSE for each estimation procedure when data are generated from a Farlie–Gumbel–Morgenstern copula with beta marginals. Akaike and Bayesian information criteria (AIC and BIC, respectively) computed as models have a different amount of parameters to be estimated. For coefficients, "—" implies that the parameter is not part of the model. Information criteria are not computed for the quasi-likelihood method.

Table 2.3: RMSE for Coefficients in a Reduced Form Model from a Dirichlet

Method	$\beta_{0,1}$	$\beta_{1,1}$	$\beta_{2,1}$	$\beta_{0,2}$	$\beta_{1,2}$	$\beta_{2,2}$	ϕ_1	AIC	BIC
$n = 100$									
Copula Y	7.664	7.798	7.409	9.167	8.203	8.386	14.448	−371.6	−348.1
Copula Z	7.662	7.722	7.296	9.158	8.645	8.372	14.459	−313.6	−290.2
MF Logit	7.722	8.001	7.790	9.352	9.277	9.392	—	—	—
Dirichlet	7.434	7.592	7.341	8.523	8.039	8.235	10.157	−375.9	−357.6
Logistic Norm.	20.193	16.133	9.747	40.454	14.379	28.222	—	591.8	607.4
$n = 200$									
Copula Y	5.283	5.342	5.160	6.451	5.812	5.777	9.454	−753.6	−723.9
Copula Z	5.286	5.319	5.088	6.529	6.658	5.733	9.457	−637.2	−607.5
MF Logit	5.339	5.581	5.411	6.598	6.463	6.399	—	—	—
Dirichlet	5.158	5.245	5.119	6.060	5.731	5.650	6.893	−760.4	−737.3
Logistic Norm.	19.236	14.433	7.067	39.945	11.069	25.979	—	1,185.6	1,205.4
$n = 400$									
Copula Y	3.685	3.741	3.608	4.680	4.209	4.059	7.011	−1,517.5	−1,481.6
Copula Z	3.684	3.761	3.569	4.738	5.283	4.055	7.012	−1,284.8	−1,248.9
MF Logit	3.736	3.934	3.773	4.833	4.742	4.538	—	—	—
Dirichlet	3.565	3.661	3.575	4.428	4.160	3.959	4.890	−1,528.8	−1,500.9
Logistic Norm.	18.709	13.422	5.095	39.269	8.719	24.879	—	2,370.4	2,394.3
$n = 800$									
Copula Y	2.616	2.615	2.526	3.339	2.996	2.919	4.935	−3,042.7	−3,000.5
Copula Z	2.616	2.627	2.496	3.376	4.416	2.911	4.932	−2,575.6	−2,533.5
MF Logit	2.670	2.742	2.615	3.427	3.372	3.241	—	—	—
Dirichlet	2.522	2.555	2.496	3.157	2.965	2.838	3.440	−3,063.3	−3,030.5
Logistic Norm.	18.254	13.065	3.736	38.840	7.459	24.328	—	4,740.6	4,768.7

Note: 100 times RMSE for each estimation procedure when data are generated from a Dirichlet distribution. Akaike and Bayesian information criteria (AIC and BIC, respectively) computed as models have a different amount of parameters to be estimated. For coefficients, "—" implies that the parameter is not part of the model. Information criteria are not computed for the quasi-likelihood method.

Figure 2.2: Trace Plot of Bayesian Chains in a Reduced Form Model

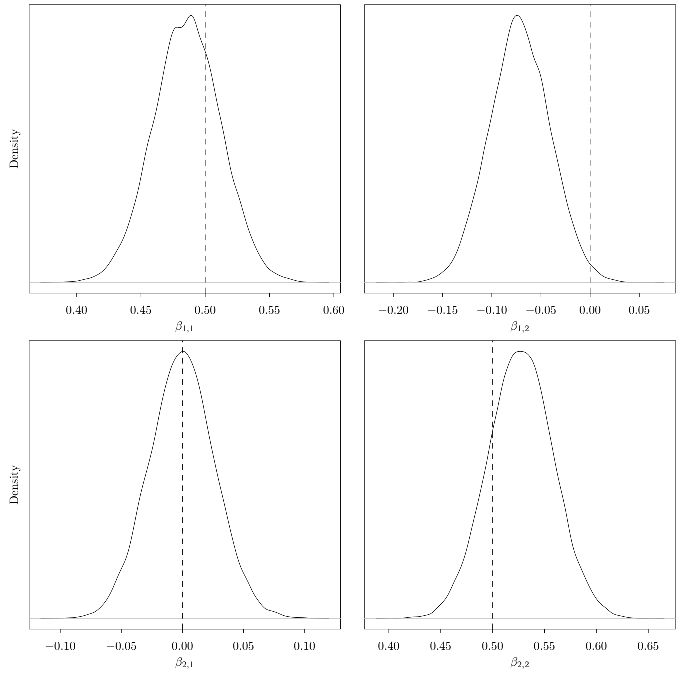


Note: Combination of 4 chains, each of 5,000 draws. The dotted line shows the true value.

from random starting values (Carpenter et al., 2017). The chains pass all of the usual diagnostics for assessing convergence to the target distribution (Brooks and Gelman, 1998; Vehtari et al., 2020). The results, along with the corresponding MLE output on the same data, are presented in Table 2.4. As expected, both approaches capture the correct values closely and have small standard errors that imply significant variables when they have a nonzero coefficient. However, note that for $\beta_{1,2}$ in this data set, the MLE estimates would imply that it is significantly different from 0 even when this is not the case in the population model. This is not the case for the Bayesian estimates that correctly single out the statistically insignificant coefficients. For further visual assessment, Figures 2.2 and 2.3 present the trace and density plots of the chains, respectively, for the main slope coefficients in $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. These combine the output from all four chains. We can see that the draws tend to gather close to the true values and thus most of the density is concentrated around these values as well.

In an applied setting, an important quantity of interest is the average partial effect (APE) of variable x_k on outcome y_j, which can be computed as an estimate of $\partial \operatorname{E}[Y_j|\boldsymbol{X} = \boldsymbol{x}]/\partial x_k$ (see, e.g., Appendix 1 in Mullahy, 2015). For notational convenience, this is written simply as $\text{APE}_{k,j}$.

Figure 2.3: Density Plot of Bayesian Chains in a Reduced Form Model



Note: Combination of 4 chains, each of 5,000 draws. The dotted line shows the true value.

While in frequentist methods you would need to use the Delta method or bootstrap for inference on this object, in the Bayesian framework it comes as a by-product of the estimation process. By simple probability arguments, calculating this quantity for each draw of the chain and obtaining the resulting mean (or median) and standard deviation yields appropriate estimation and inference. These results are presented in Table 2.5. The computed APEs are similar between all chains in terms of both point estimate and standard error. They also approximate the true effect quite well, where this true effect is simply the APE under the true coefficient vector. Figures 2.4 and 2.5 present the trace and density plots for the estimated APEs, showcasing the simplicity of the Bayesian approach in obtaining point estimates and inference of these complicated functions.

Selection using a LASSO penalty and estimating a Gaussian copula with beta marginals solves

Table 2.4: Bayesian and Frequentist Estimates for a Reduced Form Model

Parameter	Chain 1	Chain 2	Chain 3	Chain 4	MLE
$\beta_{0,1}$	−1.0603	−1.0598	−1.0620	−1.0611	−1.0614
	(0.0299)	(0.0293)	(0.0295)	(0.0298)	(0.0293)
$\beta_{1,1}$	0.4855	0.4859	0.4860	0.4866	0.4860
	(0.0258)	(0.0262)	(0.0263)	(0.0265)	(0.0262)
$\beta_{2,1}$	0.0001	0.0006	−0.0016	−0.0005	−0.0005
	(0.0268)	(0.0266)	(0.0268)	(0.0267)	(0.0264)
$\beta_{0,2}$	−1.5678	−1.5669	−1.5692	−1.5683	−1.5692
	(0.0352)	(0.0355)	(0.0355)	(0.0351)	(0.0352)
$\beta_{1,2}$	−0.0721	−0.0713	−0.0716	−0.0710	−0.0720
	(0.0307)	(0.0310)	(0.0308)	(0.0311)	(0.0310)
$\beta_{2,2}$	0.5276	0.5280	0.5258	0.5271	0.5276
	(0.0314)	(0.0310)	(0.0312)	(0.0314)	(0.0312)

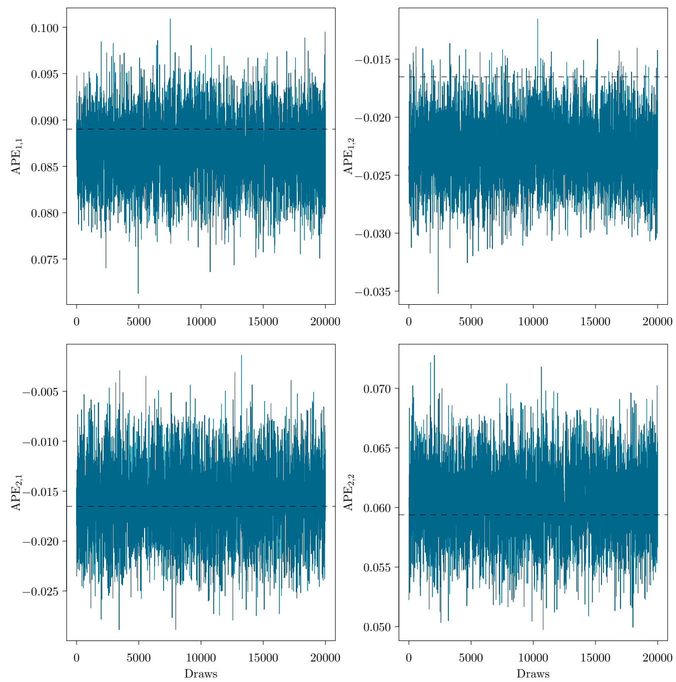
Note: Bayesian and MLE estimates from a Gaussian copula with beta marginals specification. Standard errors are in parentheses (standard deviations in each chain for Bayesian and asymptotic for MLE).

Table 2.5: Bayesian Estimates and Inference of APEs for a Reduced Form Model

Parameter	Chain 1	Chain 2	Chain 3	Chain 4	True
$\text{APE}_{1,1}$	0.0866	0.0866	0.0866	0.0867	0.0890
	(0.0037)	(0.0038)	(0.0038)	(0.0038)	
$\text{APE}_{2,1}$	−0.0159	−0.0158	−0.0161	−0.0160	−0.0165
	(0.0039)	(0.0039)	(0.0039)	(0.0039)	
$\text{APE}_{1,2}$	−0.0229	−0.0229	−0.0229	−0.0228	−0.0165
	(0.0030)	(0.0030)	(0.0029)	(0.0030)	
$\text{APE}_{2,2}$	0.0606	0.0607	0.0604	0.0606	0.0594
	(0.0032)	(0.0032)	(0.0032)	(0.0032)	

Note: Bayesian estimates from a Gaussian copula with beta marginals specification. Standard errors (standard deviation of each chain) are in parentheses.

Figure 2.4: Trace Plot of APE Chains in a Reduced Form Model



Note: Combination of 4 chains, each of 5,000 draws. The dotted line shows the true value.

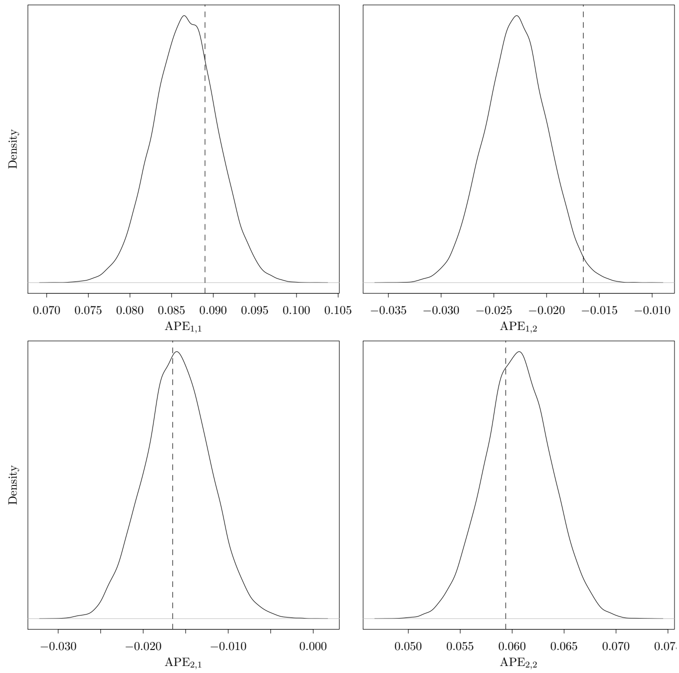
the following optimization problem:

$$\underset{(\boldsymbol{\beta},\boldsymbol{\phi},\psi)\in\mathcal{B}\times\Phi\times\Psi}{\arg\min}\Bigg\{-\log c_Y(F_1(y_1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\beta},\phi_1),\ldots,F_D(y_D|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\beta},\phi_D);\boldsymbol{\psi}) \\ -\sum_{j=1}^{d}\log f_j(y_j|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\beta},\phi_j)+\log F_W(1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\beta},\boldsymbol{\phi},\boldsymbol{\psi})+\lambda||\boldsymbol{\beta}||_1\Bigg\}.$$

Obtaining solutions for different values of λ using the simulated data set shows the effect of regularization. In the frequentist case, it operates as shown in Figure 2.6, where the parameters are moved towards 0 in absolute value and eventually set to 0 given a large enough penalty parameter λ. The coefficient $\beta_{2,1}$ does not appear in the picture as it is already estimated to be close to 0 even without regularization.

From a Bayesian perspective, to get a sense of the selection effect that the class of priors discussed in (2.23) can possess, the previous simulation is extended to a setting with 10 variables. The variables $x_1,\ldots,x_{10}$ are drawn independently from a standard normal distribution and are assigned coefficients as $\boldsymbol{\beta}_1=\boldsymbol{\beta}_2=(-2,1,-1,1,-1,1,0,0,0,0,0)$, so that the last five variables are

Figure 2.5: Density Plot of APE Chains in a Reduced Form Model



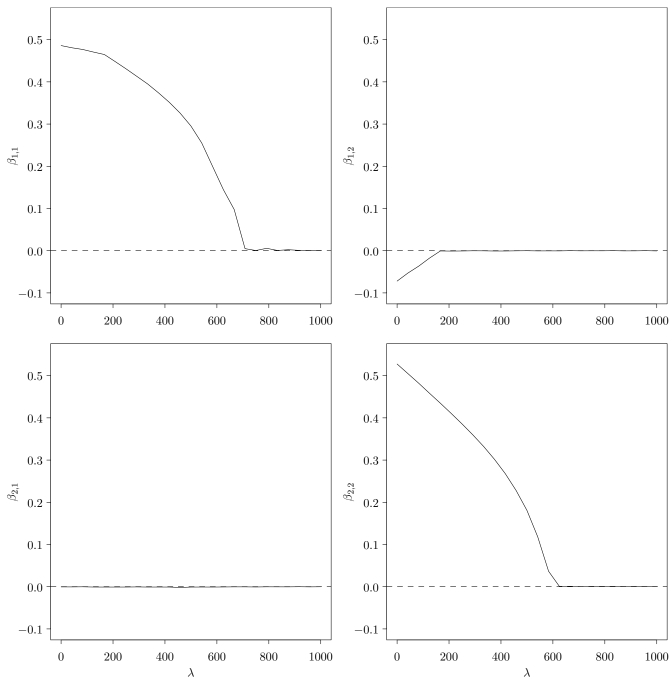
Note: Combination of 4 chains, each of 5,000 draws. The dotted line shows the true value.

redundant in the model. The following setup for priors allows for the implementation of a Bayesian LASSO penalty on this simulated data set (which due to the symmetry of the setup, will also mimic the behavior of the group-LASSO penalty):

$$\begin{aligned}
\beta_{0,j} &\sim \text{Uniform}(-\infty, \infty), j = 1, 2\,, \\
\beta_{k,j} &\sim \mathcal{N}(0, \tau_{k,j}^2) \text{ for } k = 1, \ldots, 10 \text{ and } j = 1, 2\,, \\
\tau_{k,j}^2 &\sim \text{Exponential}(\lambda^2/2) \text{ for } k = 1, \ldots, 10 \text{ and } j = 1, 2\,, \\
\lambda^2 &\sim \text{Exponential}(1)\,, \\
\phi_j &\sim \text{Gamma}(1, 1), j = 1, 2\,, \\
\psi &\sim \text{Uniform}(-1, 1)\,.
\end{aligned}$$

The resulting point estimates and inference can be found in Table 2.C.8. As expected, these are shrunk towards 0, which is a consequence of the LASSO penalty encoded in the prior distributions. Table 2.6 shows the relevant selection aspects for these coefficients and APEs for each variable. While Bayesian selection is in general not sharp, other methods such as the credible interval or scaled

Figure 2.6: Frequentist LASSO in a Reduced Form Model with a Gaussian Copula and Beta Marginals



Note: Dotted line at 0. Optimization of the Gaussian copula with beta marginals likelihood over 25 equally spaced values of λ from 0 to 1,000.

neighborhood criteria can be used to select variables based on estimates from this specification (Li and Lin, 2010).[11] The credible interval method sets a coefficient $\beta_{k,j}$ to 0 if its credible interval at a given level $\bar{l}$ (computed here as the highest posterior density interval) contains 0. On the other hand, the scaled neighborhood method takes a dual approach by computing the posterior probability within the interval defined by the standard errors (given by the standard deviation of the chains) and excludes the variable if it surpasses a given threshold; i.e., $\Pr[(-\operatorname{sd}(\beta_{k,j}), \operatorname{sd}(\beta_{k,j}))] > \bar{p}$ for some $\bar{p} \in (0, 1)$.

As can be seen in Table 2.6, the APEs are still precisely estimated. The very fact that it is simple to obtain inference for this quantity after undertaking a selection step is one of the virtues of regularization in the Bayesian framework. Additionally, the employed selection methods seem to capture the effects for the significant variables, while dropping the irrelevant ones. The scaled neighborhood method gets all of the variables right using a $\bar{p} = 0.5$, while there are some

[11] Other attractive methods exist, which combine the frequentist and Bayesian properties of selection. See, for example, the method in Leng et al. (2014) that performs a frequentist penalized regression with each λ sample in the chain and selects those variables which appear in 50 percent or more of the models.

issues if $\bar{l} = 0.5$ is used for the credible interval approach. If the level is increased slightly, say to $\bar{l} = 0.55$, then the method also successfully selects the correct model in this context. Importantly, by including a prior distribution for λ, the mean or median posterior value for this quantity can be used as a guidance for selecting the amount of regularization. In this example, both the mean and median value for λ is around 1.79, indicating that only a slight amount of penalization is necessary to exclude the redundant variables of this system.

Table 2.6: Bayesian APEs and Selection for an Extended Reduced Form Model

Variable	True $\text{APE}_{k,1}$	True $\text{APE}_{k,2}$	$\text{APE}_{k,1}$	$\text{APE}_{k,2}$	CI y_1	CI y_2	SN y_1	SN y_2
x_1	0.091	0.091	0.080 (0.004)	0.080 (0.004)	✓	✓	✓	✓
x_2	−0.091	−0.091	−0.082 (0.004)	−0.076 (0.004)	✓	✓	✓	✓
x_3	0.091	0.091	0.083 (0.004)	0.081 (0.004)	✓	✓	✓	✓
x_4	−0.091	−0.091	−0.082 (0.004)	−0.084 (0.004)	✓	✓	✓	✓
x_5	0.091	0.091	0.081 (0.004)	0.081 (0.004)	✓	✓	✓	✓
x_6	0.000	0.000	−0.002 (0.003)	−0.003 (0.003)	✓	✓	×	×
x_7	0.000	0.000	−0.004 (0.003)	0.004 (0.003)	×	×	×	×
x_8	0.000	0.000	−0.002 (0.003)	0.000 (0.003)	×	×	×	×
x_9	0.000	0.000	−0.004 (0.003)	0.001 (0.003)	✓	×	×	×
x_{10}	0.000	0.000	−0.001 (0.003)	−0.003 (0.003)	×	✓	×	×

Note: Bayesian estimates from a Gaussian copula with beta marginals specification. $\text{APE}_{k,j}$ denotes the average partial effect for a variable on outcome $j = 1, 2$. Standard errors (the standard deviation of each chain) are in parentheses. CI y_j represents credible interval selection with $\bar{l} = 0.5$ and SN y_j represents the scaled neighborhood method with $\bar{p} = 0.5$; both regarding outcome $j = 1, 2$. "✓" indicates that a variable is present in that outcome's equation and "×" denotes its absence. The Bayesian algorithm chooses a regularization parameter $\lambda = 1.79$.

2.3.2 Demand Estimation

To mimic some of the properties present in the empirical application of the next section, an almost ideal demand system with $d = 3$ shares is simulated from (2.2) by choosing the following population values for the parameters:

$$\alpha_0 = 0.675\,, \qquad \boldsymbol{\alpha} = \begin{bmatrix} 0.929 \\ 0.297 \\ -0.226 \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} 0.062 & -0.033 & -0.029 \\ -0.033 & -0.058 & 0.091 \\ -0.029 & 0.091 & -0.062 \end{bmatrix}, \qquad \boldsymbol{\pi} = \begin{bmatrix} -0.064 \\ -0.029 \\ 0.093 \end{bmatrix}.$$

These values satisfy the constraints of an AID system for homogeneity of degree one in prices and expenditures, as well as the symmetry of the Slutsky matrix. In order to generate values from this model, the following exercises use either a Gaussian copula with beta marginals or generate from a multivariate normal distribution directly, while restricting the values to lie on $\mathcal{S}^d$. Prices are generated from a uniform distribution between 1.2 and 1.5 for all three simulated goods. Expenditures were drawn from a log-normal distribution with a mean of 6 and a standard deviation of 0.25 in the log scale. For each generating exercise, there are 1,000 simulations. For now, the paper examines the maximum likelihood estimation results, leaving the Bayesian results for the empirical application, which will be conditional on the examined data.

For estimation purposes in the standard AID framework, there are only $(d^2 + 3d - 1)/2$ free parameters to estimate as the constraints allow us to eliminate one parameter each from $\boldsymbol{\alpha}$ and $\boldsymbol{\pi}$ and all but $d(d-1)/2$ parameters from the Γ matrix. These can be recovered in each iteration of the estimation algorithm, ensuring that the constraints are always satisfied. Furthermore, the use of marginals that respect the fractional restriction encourages positivity on the system (all predicted shares being greater than 0), as the likelihood is undefined if the underlying values lead to predictions outside of this range.

The flexibility and robustness of the methodology introduced in the paper even in this context is showcased in Tables 2.7 and 2.8. The main difference is in the generating marginal distributions. In the first table, betas with mean-precision parameterization are used, whereas the second table uses normal distributions. The tables estimate four of the same models as before: a copula on Y, a copula on Z, a multivariate fractional quasi-likelihood (it is no longer a logit as the conditional mean specification changes), and a Dirichlet. The final method is a regular multivariate normal distribution, where the ϕ parameters take on a precision interpretation for each marginal, and ψ or ξ represents the correlation parameter. As a Gaussian copula with Gaussian marginals is equivalent to a multivariate normal distribution, this second exercise is closer to what is usually used in practice, where no appropriate restriction on the estimating functional form is imposed.

The main features from the previous simulations are maintained in this setting as well. Both the copula on Y and Z estimators are consistent due to their correctly-specified nature in Table 2.7. Both AIC and BIC select the copula on Y as the preferable estimator at all sample sizes, with the regular AID coming in at a close second place in terms of performance. This is also to be expected, as part of the attractive features of the normal distribution are that the normal distribution is consistent under the same conditions as the multivariate fractional quasi-likelihood, even under misspecification (Gourieroux et al., 1984). While this multivariate fractional distribution is generally only used in conjunction with a logit link, this exercise also confirms its ability to remain consistent only under correct conditional mean specifications. Table 2.8 presents a similar view; however, the copula on Z estimator becomes less reliable. This is to be expected due to its failure to be consistent under more general conditions than the copula on Y estimator. Surprisingly, the normal AID system does not become much more dominant in this setting, which could be related to the positivity argument discussed before, as the current configuration could try to pull the parameters toward violating the fractional restriction on the outcomes.

To examine the role of a more flexible alternative to the AID system, the next two simulations implement a setting similar to the previous one, except that polynomials on the deflated expenditures are added as outlined in (2.24). Two extra terms are added to the generating process, where the new population coefficients are just $\boldsymbol{\pi}_2 = \boldsymbol{\pi}_1^2$ and $\boldsymbol{\pi}_3 = \boldsymbol{\pi}_1^3$, with $\boldsymbol{\pi}_1$ being the original coefficients in the first two simulation exercises. Tables 2.9 and 2.10 present the results for this configuration. In general, the patterns observed in this iteration track the previous results very closely. It is worth noting that the copula on Z estimator becomes even more erratic with the inclusion of extra parameters, so that the copula on Y estimator remains a preferred choice. We have seen throughout this Monte Carlo study, even in a Bayesian setting, that it has strong a performance compared to the methods previously available in the literature.

Table 2.7: RMSE for Coefficients in a Structural Demand Model from a Gaussian Copula with Beta Marginals

Method	α_0	α_1	α_2	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{2,2}$	π_1	π_2	ϕ_1	ϕ_2	$\psi\|\xi$	AIC	BIC
						$n = 100$							
Copula Y	2.613	2.895	1.749	1.467	0.823	0.853	0.553	0.323	3.409	4.490	4.784	−366.5	−337.8
Copula Z	46.744	7.030	2.838	1.757	0.873	0.907	0.539	0.334	3.404	12.660	4.058	−188.9	−160.3
Multi. Frac.	1.975	2.929	1.887	1.598	0.857	1.070	0.555	0.347	—	—	—	—	—
Dirichlet	0.923	3.099	1.696	1.677	0.845	1.078	0.599	0.330	1.631	—	—	−312.9	−289.5
AID	18.587	3.683	9.054	1.573	0.955	88.715	0.556	3.565	1.860	2.379	5.664	−337.9	−309.2
						$n = 200$							
Copula Y	3.056	2.231	1.400	1.043	0.573	0.613	0.413	0.246	3.006	4.174	4.562	−744.4	−708.1
Copula Z	6.650	2.216	1.515	1.105	0.590	0.641	0.403	0.269	3.001	12.808	3.768	−387.3	−351.0
Multi. Frac.	0.670	2.238	1.522	1.125	0.591	0.758	0.415	0.267	—	—	—	—	—
Dirichlet	2.576	2.416	1.316	1.176	0.603	0.769	0.452	0.249	1.605	—	—	−634.5	−604.8
AID	9.416	2.496	30.154	1.258	3.842	84.369	0.423	6.542	1.840	2.371	5.621	−686.8	−650.5
						$n = 400$							
Copula Y	3.731	1.746	1.184	0.732	0.406	0.456	0.313	0.196	2.854	3.981	4.443	−1,502.9	−1,459.0
Copula Z	10.029	1.858	1.349	0.827	0.429	0.517	0.322	0.235	2.907	12.837	3.657	−784.6	−740.7
Multi. Frac.	4.870	1.744	1.331	0.782	0.418	0.580	0.314	0.221	—	—	—	—	—
Dirichlet	1.213	1.911	1.065	0.819	0.421	0.584	0.348	0.197	1.603	—	—	−1,279.3	−1,243.4
AID	7.991	1.886	10.664	0.847	0.757	46.232	0.324	1.847	1.840	2.366	5.517	−1,386.9	−1,343.0
						$n = 800$							
Copula Y	3.137	1.480	1.053	0.523	0.286	0.326	0.251	0.164	2.775	3.873	4.373	−3,016.7	−2,965.1
Copula Z	8.271	1.592	1.309	0.713	0.351	0.428	0.263	0.222	2.874	12.827	3.542	−1,564.9	−1,513.3
Multi. Frac.	1.998	1.535	1.209	0.558	0.294	0.438	0.252	0.190	—	—	—	—	—
Dirichlet	4.610	1.613	0.897	0.577	0.293	0.429	0.288	0.164	1.616	—	—	−2,564.3	−2,522.2
AID	7.356	1.619	2.128	0.736	0.348	32.587	0.266	1.908	1.833	2.367	5.493	−2,790.5	−2,739.0

Note: 10 times RMSE for each estimation procedure when data are generated from a Gaussian copula with beta marginals. Akaike and Bayesian information criteria (AIC and BIC, respectively) computed as models have a different amount of parameters to be estimated. For coefficients, "—" implies that the parameter is not part of the model. Information criteria are not computed for the quasi-likelihood method.

Table 2.8: RMSE for Coefficients in a Structural Demand Model from a Gaussian Distribution

Method	α_0	α_1	α_2	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{2,2}$	π_1	π_2	ϕ_1	ϕ_2	$\psi\|\xi$	AIC	BIC
						$n = 100$							
Copula Y	54.963	11.151	4.388	3.262	1.567	1.705	0.857	0.619	6.240	3.354	14.511	−184.0	−155.4
Copula Z	70.756	12.002	6.708	3.471	1.795	1.903	0.879	0.612	6.277	14.642	4.607	−49.5	−20.9
Multi. Frac.	2.106	4.760	2.960	2.486	1.499	1.789	0.833	0.626	—	—	—	—	—
Dirichlet	34.743	7.554	3.161	2.768	1.514	1.715	0.862	0.621	7.096	—	—	−199.4	−175.9
AID	31.751	7.744	3.333	2.779	1.528	1.745	0.834	0.625	1.252	1.881	15.166	−157.5	−128.8
						$n = 200$							
Copula Y	21.665	5.580	2.353	1.982	1.050	1.278	0.689	0.463	6.367	3.458	14.368	−379.3	−343.0
Copula Z	29.222	6.706	2.268	2.517	1.251	1.428	0.714	0.472	6.416	14.726	4.496	−103.7	−67.4
Multi. Frac.	3.642	4.085	2.194	1.796	1.056	1.337	0.667	0.482	—	—	—	—	—
Dirichlet	4.849	4.239	2.063	1.820	1.059	1.312	0.694	0.472	7.267	—	—	−408.4	−378.7
AID	9.861	4.129	2.030	1.762	1.046	1.304	0.668	0.483	1.234	1.868	15.106	−328.0	−291.7
						$n = 400$							
Copula Y	10.187	3.638	1.856	1.241	0.731	0.948	0.558	0.401	6.402	3.480	14.354	−771.0	−727.1
Copula Z	9.147	3.703	1.764	1.997	1.010	1.224	0.594	0.409	6.446	14.491	4.663	−182.7	−138.8
Multi. Frac.	1.446	3.571	1.802	1.231	0.759	1.001	0.545	0.413	—	—	—	—	—
Dirichlet	3.134	3.699	1.655	1.253	0.746	0.989	0.571	0.410	7.317	—	—	−827.6	−791.7
AID	11.977	3.749	1.940	1.208	0.757	0.970	0.546	0.413	1.226	1.861	15.119	−670.4	−626.5
						$n = 800$							
Copula Y	12.375	3.518	1.436	0.845	0.565	0.759	0.476	0.356	6.455	3.509	14.311	−1,550.6	−1,499.1
Copula Z	9.477	3.391	1.616	1.627	0.911	1.014	0.530	0.382	6.491	14.426	4.680	−350.5	−299.9
Multi. Frac.	2.999	3.290	1.554	0.861	0.571	0.798	0.471	0.366	—	—	—	—	—
Dirichlet	5.043	3.373	1.388	0.884	0.577	0.797	0.489	0.367	7.378	—	—	−1,662.0	−1,619.9
AID	5.521	3.262	1.606	0.836	0.564	0.764	0.472	0.366	1.220	1.857	15.083	−1,350.6	−1,299.1

Note: 10 times RMSE for each estimation procedure when data are generated from a multivariate Gaussian distribution. Akaike and Bayesian information criteria (AIC and BIC, respectively) computed as models have a different amount of parameters to be estimated. For coefficients, "—" implies that the parameter is not part of the model. Information criteria are not computed for the quasi-likelihood method.

Table 2.9: RMSE for Coefficients in an Extended Structural Demand Model from a Gaussian Copula with Beta Marginals

Method	α_0	α_1	α_2	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{2,2}$	$\pi_{1,1}$	$\pi_{2,1}$	$\pi_{1,2}$	$\pi_{2,2}$	$\pi_{1,3}$	$\pi_{2,3}$	ϕ_1	ϕ_2	$\psi\|\xi$	AIC	BIC
								$n = 100$									
Copula Y	1.893	3.257	3.107	0.221	0.111	0.129	1.919	1.713	0.581	0.468	0.075	0.060	0.471	0.624	0.639	−381.7	−342.6
Copula Z	2.079	3.623	3.279	0.242	0.117	0.137	2.023	1.746	0.609	0.483	0.090	0.065	0.464	1.263	0.360	−175.8	−136.8
Multi. Frac.	1.272	2.936	2.867	0.233	0.110	0.144	1.809	1.677	0.516	0.455	0.078	0.069	—	—	—	—	—
Dirichlet	2.084	3.521	3.258	0.281	0.121	0.148	1.889	1.686	0.559	0.454	0.084	0.062	0.123	—	—	−339.2	−305.3
AID	1.845	3.591	3.100	0.236	0.117	0.129	2.017	1.767	0.579	0.492	0.075	0.070	0.197	0.245	0.724	−360.4	−321.3
								$n = 200$									
Copula Y	1.595	3.093	2.763	0.166	0.081	0.084	1.744	1.462	0.457	0.345	0.056	0.038	0.417	0.575	0.622	−776.0	−726.5
Copula Z	1.908	3.342	2.799	0.174	0.077	0.085	1.824	1.443	0.494	0.351	0.066	0.041	0.411	1.302	0.321	−362.7	−313.2
Multi. Frac.	0.970	2.672	2.576	0.165	0.076	0.091	1.596	1.486	0.385	0.335	0.041	0.035	—	—	—	—	—
Dirichlet	1.866	3.339	2.990	0.193	0.084	0.101	1.751	1.553	0.449	0.378	0.057	0.047	0.089	—	—	−687.3	−644.4
AID	1.659	3.213	2.814	0.171	0.082	0.088	1.846	1.552	0.508	0.391	0.066	0.046	0.194	0.244	0.717	−733.9	−684.4
								$n = 400$									
Copula Y	1.249	2.715	2.343	0.108	0.055	0.055	1.523	1.282	0.347	0.274	0.034	0.024	0.391	0.547	0.618	−1,562.9	−1,503.1
Copula Z	1.619	2.865	2.357	0.112	0.055	0.059	1.618	1.315	0.420	0.318	0.055	0.035	0.382	1.315	0.301	−728.3	−668.4
Multi. Frac.	0.708	2.225	2.223	0.104	0.053	0.066	1.383	1.354	0.324	0.301	0.032	0.027	—	—	—	—	—
Dirichlet	1.547	2.874	2.609	0.118	0.054	0.072	1.531	1.375	0.360	0.301	0.042	0.031	0.082	—	—	−1,381.9	−1,330.0
AID	1.214	2.781	2.389	0.117	0.056	0.212	1.624	1.372	0.397	0.313	0.048	0.030	0.193	0.243	0.716	−1,477.5	−1,417.7
								$n = 800$									
Copula Y	1.016	2.360	2.052	0.081	0.037	0.041	1.354	1.165	0.293	0.250	0.027	0.023	0.378	0.536	0.618	−3,140.5	−3,070.2
Copula Z	1.380	2.483	1.952	0.091	0.035	0.042	1.409	1.127	0.345	0.262	0.041	0.026	0.370	1.314	0.287	−1,448.3	−1,378.1
Multi. Frac.	0.491	2.139	1.925	0.085	0.036	0.047	1.297	1.188	0.282	0.251	0.023	0.019	—	—	—	—	—
Dirichlet	1.109	2.603	2.262	0.086	0.040	0.050	1.503	1.315	0.346	0.287	0.037	0.026	0.078	—	—	−2,774.5	−2,713.6
AID	1.058	2.612	1.977	0.094	0.040	1.725	1.494	1.168	0.335	0.321	0.032	0.057	0.192	0.242	0.718	−2,970.4	−2,900.1

Note: RMSE for each estimation procedure when data are generated from a Gaussian copula with beta marginals. Akaike and Bayesian information criteria (AIC and BIC, respectively) computed as models have a different amount of parameters to be estimated. For coefficients, "—" implies that the parameter is not part of the model. Information criteria are not computed for the quasi-likelihood method.

Table 2.10: RMSE for Coefficients in an Extended Structural Demand Model from a Gaussian Distribution

Method	α_0	α_1	α_2	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{2,2}$	$\pi_{1,1}$	$\pi_{2,1}$	$\pi_{1,2}$	$\pi_{2,2}$	$\pi_{1,3}$	$\pi_{2,3}$	ϕ_1	ϕ_2	$\psi\|\xi$	AIC	BIC
$n = 100$																	
Copula Y	2.523	4.358	4.215	0.421	0.225	0.278	2.312	2.095	0.723	0.574	0.106	0.077	0.543	0.268	1.548	−197.2	−158.1
Copula Z	2.453	4.439	4.060	0.406	0.225	0.265	2.379	2.025	0.746	0.559	0.119	0.078	0.513	1.376	0.464	−16.6	22.5
Multi. Frac.	1.415	3.468	3.107	0.384	0.198	0.241	2.217	1.869	0.705	0.534	0.112	0.075	—	—	—	—	—
Dirichlet	2.313	4.206	3.830	0.387	0.212	0.263	2.293	1.985	0.727	0.571	0.107	0.082	0.618	—	—	−216.6	−182.7
AID	2.117	4.191	3.730	0.386	0.218	0.267	2.380	2.001	0.804	0.623	0.128	0.098	0.132	0.195	1.601	−169.0	−129.9
$n = 200$																	
Copula Y	2.013	3.811	3.604	0.267	0.145	0.174	2.043	1.882	0.571	0.488	0.079	0.062	0.570	0.290	1.530	−405.6	−356.1
Copula Z	2.264	3.716	3.417	0.259	0.146	0.170	1.917	1.740	0.561	0.478	0.082	0.064	0.528	1.388	0.445	−27.2	22.3
Multi. Frac.	0.982	2.967	2.847	0.241	0.138	0.162	1.846	1.664	0.500	0.396	0.062	0.042	—	—	—	—	—
Dirichlet	2.249	3.949	3.730	0.276	0.150	0.173	2.099	1.879	0.625	0.489	0.107	0.068	0.654	—	—	−442.8	−399.9
AID	1.836	3.580	3.503	0.252	0.145	0.180	1.974	1.868	0.557	0.506	0.074	0.069	0.128	0.192	1.593	−350.5	−301.0
$n = 400$																	
Copula Y	2.037	3.732	3.375	0.190	0.110	0.133	1.903	1.682	0.481	0.411	0.065	0.055	0.588	0.303	1.515	−822.3	−762.5
Copula Z	2.116	3.647	3.359	0.193	0.108	0.134	1.931	1.714	0.545	0.463	0.082	0.068	0.577	1.478	0.456	−195.4	−135.5
Multi. Frac.	0.903	2.810	2.516	0.175	0.096	0.118	1.709	1.506	0.417	0.348	0.042	0.035	—	—	—	—	—
Dirichlet	1.902	3.427	3.138	0.188	0.106	0.134	1.826	1.656	0.484	0.422	0.067	0.057	0.675	—	—	−894.6	−842.7
AID	1.674	3.664	2.900	0.176	0.097	0.121	1.912	1.597	0.467	0.393	0.052	0.047	0.127	0.191	1.583	−714.2	−654.3
$n = 800$																	
Copula Y	1.363	3.144	2.777	0.122	0.069	0.094	1.722	1.540	0.397	0.350	0.043	0.037	0.596	0.309	1.510	−1,655.0	−1,584.7
Copula Z	1.708	2.935	2.599	0.132	0.084	0.111	1.522	1.324	0.367	0.307	0.043	0.035	0.540	1.386	0.422	−120.8	−50.5
Multi. Frac.	0.596	2.454	2.256	0.114	0.066	0.088	1.551	1.405	0.366	0.308	0.035	0.025	—	—	—	—	—
Dirichlet	1.601	3.228	2.864	0.120	0.071	0.099	1.628	1.483	0.367	0.312	0.042	0.029	0.683	—	—	−1,799.3	−1,738.4
AID	1.278	2.863	2.646	0.113	0.068	0.092	1.595	1.508	0.375	0.349	0.041	0.037	0.126	0.190	1.580	−1,444.0	−1,373.7

Note: RMSE for each estimation procedure when data are generated from a multivariate Gaussian distribution. Akaike and Bayesian information criteria (AIC and BIC, respectively) computed as models have a different amount of parameters to be estimated. For coefficients, "—" implies that the parameter is not part of the model. Information criteria are not computed for the quasi-likelihood method.

2.4 Empirical Application

As a complement and extension to the numerical study undertaken in the previous section, this section puts into action the methods introduced in the paper. This empirical application uses the data set in Chang and Serletis (2014) (hereafter referred to as CS), which collects information on household transportation expenditures in Canada from the Canadian *Survey of Household Spending* between the years of 1997 and 2009. Using these observations, CS fit an almost ideal demand system, as well as its quadratic extension, and the Minflex Laurent model (Deaton and Muellbauer, 1980; Barnett, 1983; Barnett and Lee, 1985; Banks et al., 1997). Focusing on the AID system, in the language of this paper's Example 1, it translates to fitting the following model for household i in $1, \ldots, n$:

$$\mathrm{E}[\boldsymbol{Y}_i | e_i, \boldsymbol{p}_i] = \boldsymbol{\alpha} + \Gamma \boldsymbol{p}_i + \boldsymbol{\pi}[e_i - \alpha_0 - \boldsymbol{\alpha}' \boldsymbol{p}_i - (1/2) \boldsymbol{p}_i' \Gamma \boldsymbol{p}_i] \,. \tag{2.26}$$

Using the notation developed thus far, there are expenditure shares for $d = 3$ goods, where y_1 represents gasoline, y_2 is local transportation, and y_3 is intercity transportation. The base category of analysis will be the same as used in CS, given by the third good. Prices of these goods are normalized with 2002 serving as the base. To rule out the effect of possible unobserved heterogeneity, CS assumes that households with similar demographic characteristics share similar consumption patterns. Thus, instead of including these characteristics to complicate the structural model, CS focus only on households between 25 and 64 years old, living in urban areas with a population of at least 30,000 in English Canada. The authors also restrict the sample to households with a larger than 0 expenditure on all three goods, to avoid the issue of boundary values. Furthermore, the sample is split between three types of households: single-member households, married couples without children, and married couples with one child. Summary statistics for the variables are presented in Table 2.11. While this table uses the data in levels, prices and expenditures are understood to have been transformed to natural logarithms for estimation purposes in (2.26).

For modeling purposes, CS assume that all observations are independent and identically distributed, which is a reasonable assumption as data is collected as repeated cross-sections at the household level. The authors also acknowledge possible endogeneity issues, but given the use of

Table 2.11: Summary Statistics for Data in Chang and Serletis (2014)

Variable	Good	Mean	Std. Dev.	Minimum	Maximum
Single member households, 2,218 observations					
Budget shares	Gasoline	0.499	0.237	0.002	0.986
	Local transportation	0.095	0.128	0.001	0.856
	Intercity transportation	0.406	0.228	0.003	0.985
Prices	Gasoline	1.157	0.269	0.726	1.751
	Local transportation	1.038	0.131	0.801	1.307
	Intercity transportation	1.011	0.132	0.755	1.233
Expenditures		2,430.7	1,703.0	161	24,620
Married couples without children, 3,326 observations					
Budget shares	Gasoline	0.524	0.234	0.005	0.990
	Local transportation	0.083	0.114	0.000	0.866
	Intercity transportation	0.392	0.224	0.003	0.985
Prices	Gasoline	1.170	0.268	0.726	1.751
	Local transportation	1.046	0.131	0.801	1.307
	Intercity transportation	1.017	0.132	0.755	1.233
Expenditures		3,920.5	2,396.7	170	26,230
Married couples with one child, 6,141 observations					
Budget shares	Gasoline	0.575	0.237	0.002	0.997
	Local transportation	0.092	0.117	0.000	0.886
	Intercity transportation	0.333	0.229	0.002	0.980
Prices	Gasoline	1.146	0.261	0.726	1.751
	Local transportation	1.035	0.127	0.801	1.307
	Intercity transportation	1.005	0.130	0.755	1.233
Expenditures		4,858.4	3,021.8	259	37,490

Note: Sample covers the period from 1997 to 2009. Intercity transportation is taken as the base category.

individual-level consumption instead of an aggregated level, it is likely that there is no simultaneity bias in the determination of household consumption and yearly aggregate prices. Furthermore, even when endogeneity is addressed by means of the generalized method of moments (GMM) or iterative three-stage least squares (3SLS), estimates tend to be similar to the baseline ones. Therefore, the conditional mean assumption in (2.1) is likely to be satisfied.

As seen in the Monte Carlo evidence from the previous section, the copula on Y estimator stands out as a flexible alternative to model structural estimation in demand models. Table 2.12 presents the estimation results using beta marginals with Gaussian or FGM copulas. The two represent widely-used copulas in applied research and belong to the two most important classes of copulas: elliptical and Archimedean. The resulting estimates are quite similar within each of the three population segments regardless of the copula — a consequence of Theorem 2.2 in action. The only main differences for the parameters of the AID system are in α_0, but this parameter is known to be identified only up to a scale factor so that it tends to vary with any estimation procedure (Deaton and Muellbauer, 1980). The estimates also align closely with those obtained in Table II of CS and mimic other replications of their results (Velásquez-Giraldo et al., 2018). Interestingly, the negative correlation between the two outcomes is reflected as a correlation coefficient in the Gaussian distribution of about -0.4. As the FGM copula cannot produce as much negative dependence, the estimates tend to be close to the lower bound of 1. Inference also remains quite similar between both specifications.[12] Standard errors are consistent with the magnitude and role of each parameter and also closely resemble those previously found in the literature.

As a second exercise, an estimation can be done in the Bayesian framework, using similar techniques as before. However, one of the issues with using Bayesian directly on the AID conditional mean (2.26) is the scale of all parameters except for $\boldsymbol{\pi}$. In the original scales, the Hamiltonian Monte Carlo algorithm used to explore the parameter space and draw from the posterior can get stuck and over-reject as many combinations of parameter values do not satisfy the positivity constraints. To this end, a reparameterization similar to that in Lewbel and Pendakur (2009) becomes necessary. The authors use the natural logarithm of the expenditure variable after having subtracted the median of the log-transformed value; i.e., they define $e_{\text{new}} = e - \text{median}(e)$. In the AID system,

[12] As numerical optimization is done in an unrestricted domain, the standard errors for the precision and correlation parameters are Delta method transformations.

Table 2.12: MLE Estimates of AID System using the Copula Y Estimator with Different Copulas and Beta Marginals

Parameter	Single households			Married couples			Married with one child		
	Gaussian	FGM	Reparam.	Gaussian	FGM	Reparam.	Gaussian	FGM	Reparam.
α_0	0.871	0.358	1.282	0.379	−0.401	0.216	0.655	1.599	0.961
	(0.126)	(0.083)	(0.028)	(0.507)	(0.120)	(0.073)	(0.034)	(0.461)	(0.012)
α_1	0.889	0.884	0.403	1.086	1.121	0.494	1.149	1.048	0.491
	(0.071)	(0.074)	(0.016)	(0.037)	(0.054)	(0.007)	(0.038)	(0.049)	(0.007)
α_2	0.247	0.273	0.073	0.259	0.286	0.080	0.246	0.239	0.075
	(0.016)	(0.017)	(0.004)	(0.018)	(0.017)	(0.002)	(0.012)	(0.014)	(0.002)
$\gamma_{1,1}$	0.057	0.056	0.086	0.002	0.007	0.045	−0.043	−0.028	0.007
	(0.042)	(0.043)	(0.041)	(0.034)	(0.034)	(0.031)	(0.025)	(0.025)	(0.024)
$\gamma_{2,1}$	−0.019	−0.014	−0.008	−0.023	−0.024	−0.010	−0.031	−0.031	−0.018
	(0.012)	(0.012)	(0.012)	(0.008)	(0.009)	(0.008)	(0.007)	(0.007)	(0.007)
$\gamma_{2,2}$	−0.032	−0.041	−0.028	0.053	0.052	0.057	0.052	0.042	0.056
	(0.033)	(0.032)	(0.033)	(0.025)	(0.025)	(0.025)	(0.021)	(0.021)	(0.021)
π_1	−0.060	−0.056	−0.060	−0.074	−0.072	−0.074	−0.076	−0.072	−0.076
	(0.010)	(0.010)	(0.010)	(0.008)	(0.007)	(0.007)	(0.005)	(0.005)	(0.005)
π_2	−0.022	−0.024	−0.022	−0.023	−0.024	−0.023	−0.020	−0.022	−0.020
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
ϕ_1	3.551	3.589	3.551	3.718	3.769	3.718	3.498	3.505	3.498
	(0.102)	(0.099)	(0.102)	(0.083)	(0.081)	(0.082)	(0.059)	(0.058)	(0.059)
ϕ_2	7.313	7.367	7.313	7.881	7.987	7.881	7.382	7.357	7.382
	(0.359)	(0.353)	(0.361)	(0.297)	(0.292)	(0.297)	(0.189)	(0.183)	(0.188)
ψ	−0.390	−0.999	−0.390	−0.400	−1.000	−0.400	−0.363	−0.995	−0.363
	(0.026)	(0.002)	(0.026)	(0.021)	(0.001)	(0.021)	(0.017)	(0.021)	(0.017)
Log-lik.	3,352.7	3,330.1	3,352.7	5,660.6	5,635.6	5,660.6	9,734.5	9,677.5	9,734.4
Obs.		2,218			3,326			6,141	

Note: Sample covers the period from 1997 to 2009. Intercity transportation is taken as the base category. Standard errors robust to copula misspecification are in parentheses. The third column of each data set includes a reparameterized model with a Gaussian copula.

this reparameterization keeps $\boldsymbol{\pi}$ intact, while ensuring that α_0, $\boldsymbol{\alpha}$, and Γ take on scales that are more likely to respect the fractional restriction for the conditional mean. Table 2.12 includes a third column for each data set where the AID system is estimated using e_{new} instead of e. As expected, the slope estimates $\widehat{\boldsymbol{\pi}}$ remain the same, while other estimated parameters change in scale. Note, for example, how the $\widehat{\boldsymbol{\alpha}}$ are now closer to the mean expenditure of each good.

With this reparameterization, the Bayesian algorithm becomes more accurate and can produce results without needing many iterations. In particular, after around 300 tuning iterations, the algorithm rarely produces rejections based on violations of positivity constraints. This is also due to the beta marginals that — similar to the frequentist case — encourage parameter values that satisfy the fractional restrictions of multivariate fractional outcomes. Within this new parameterization, the following priors are imposed:

$$
\begin{aligned}
\alpha_0 &\sim \mathcal{N}(0,5)\,,\\
\alpha_j &\sim \mathcal{N}(0,1), j=1,2\,,\\
\gamma_{j,l} &\sim \mathcal{N}(0,1), j=1,2, l\leq j,\,,\\
\pi_j &\sim \mathcal{N}(0,1), j=1,2\,,\\
\phi_j &\sim \text{Gamma}(1,1), j=1,2\,,\\
\psi &\sim \text{Uniform}(-1,1)\,.
\end{aligned}
$$

The slightly tighter priors are useful in avoiding many proposal rejections in the posterior exploration algorithm, as it is clear that larger values of the parameters are generally incompatible with the fractional restriction. Table 2.13 presents the estimation results from a Bayesian perspective. Estimates are the mean of the chains, where there are five chains, each providing 700 draws (after the 300 tuning period). Similar to before, the chains are checked and pass the usual convergence diagnostics. As can be observed, the results remain similar to the maximum likelihood ones, when the reparameterization is considered. The Bayesian standard errors tend to be more narrow for the $\boldsymbol{\alpha}$ and Γ parameters, but slightly larger for the slopes $\boldsymbol{\pi}$, which become statistically insignificant in the first model. Figures 2.7 and 2.8 present the trace and density plots for the core AID parameters in the data set for married couples with one child. As expected, the most variability is given in the chain for α_0. There appears to be some possible auto-correlation in the other $\boldsymbol{\alpha}$ parameter chains, which can be solved by thinning the chain before computing estimates; this is done for the results presented in Table 2.13.

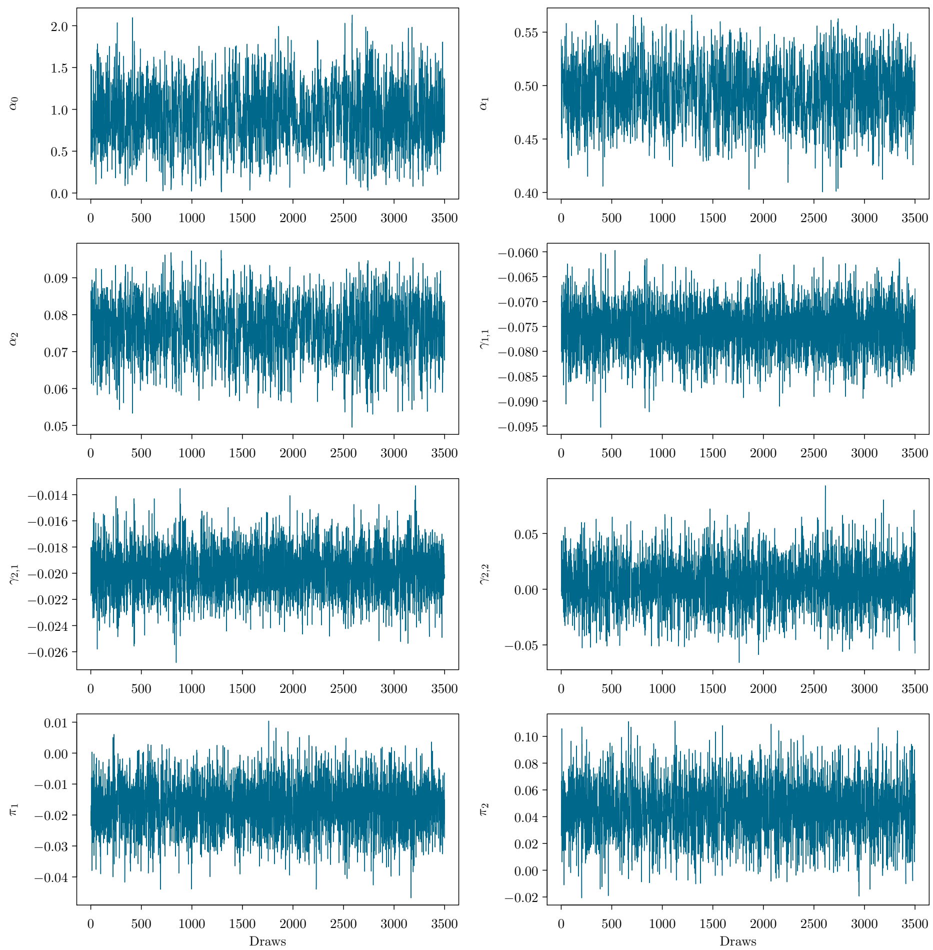
Table 2.13: Bayesian Estimates of a Reparameterized AID System using the Copula Y Estimator with a Gaussian Copula and Beta Marginals

Parameter	Single households	Married couples	Married with one child
α_0	0.651 (0.354)	0.697 (0.368)	0.928 (0.369)
α_1	0.446 (0.021)	0.461 (0.027)	0.494 (0.028)
α_2	0.086 (0.009)	0.069 (0.009)	0.076 (0.008)
$\gamma_{1,1}$	−0.058 (0.008)	−0.073 (0.007)	−0.076 (0.005)
$\gamma_{2,1}$	−0.022 (0.003)	−0.023 (0.002)	−0.020 (0.002)
$\gamma_{2,2}$	0.050 (0.031)	0.034 (0.027)	0.005 (0.022)
π_1	−0.004 (0.014)	−0.007 (0.010)	−0.017 (0.008)
π_2	−0.017 (0.032)	0.045 (0.025)	0.045 (0.021)
ϕ_1	3.563 (0.093)	3.725 (0.081)	3.503 (0.056)
ϕ_2	7.339 (0.244)	7.890 (0.207)	7.386 (0.149)
ψ	−0.388 (0.018)	−0.399 (0.015)	−0.362 (0.011)
Obs.	2,218	3,326	6,141

Note: Sample covers the period from 1997 to 2009. Intercity transportation is taken as the base category. Standard deviations for the chains are in parentheses.

Looking beyond the parameter estimates in the AID system, it is important to be able to provide price and income elasticities, as well as inference with respect to these parameters. As previously stated, this inference is simple in the Bayesian context. While these functions can be complicated and highly nonlinear with respect to the parameters so as to make the application of the Delta method challenging, computing them for a given set of estimates is simple. Table 2.16 presents the income and uncompensated price elasticities for the AID. Following CS, these are the elasticities evaluated at the average prices and, given the parameterization necessary for a Bayesian estimation, are at the average median-centered expenditure. These elasticities are slightly larger

Figure 2.7: Trace Plot of Coefficient Chains in a Reparameterized Bayesian AID System



Note: Results for the data set on married couples with one child. Combination of 5 chains with 700 draws each for a total of 3,500 draws.

than those in CS, but are for the most part consistent with economic theory. Note, however, the large standard errors for elasticities associated to local transportation (Good 2). This phenomenon most likely occurs because of a few outliers in the chains, combined with the generally small share of the budget allocated to this good. As the predicted shares get closer to the lower bound of 0, the computed elasticities can suffer from numerical issues. The fact that the mean remains close to the expected values, however, is a sign this occurs only a few times throughout the chain.

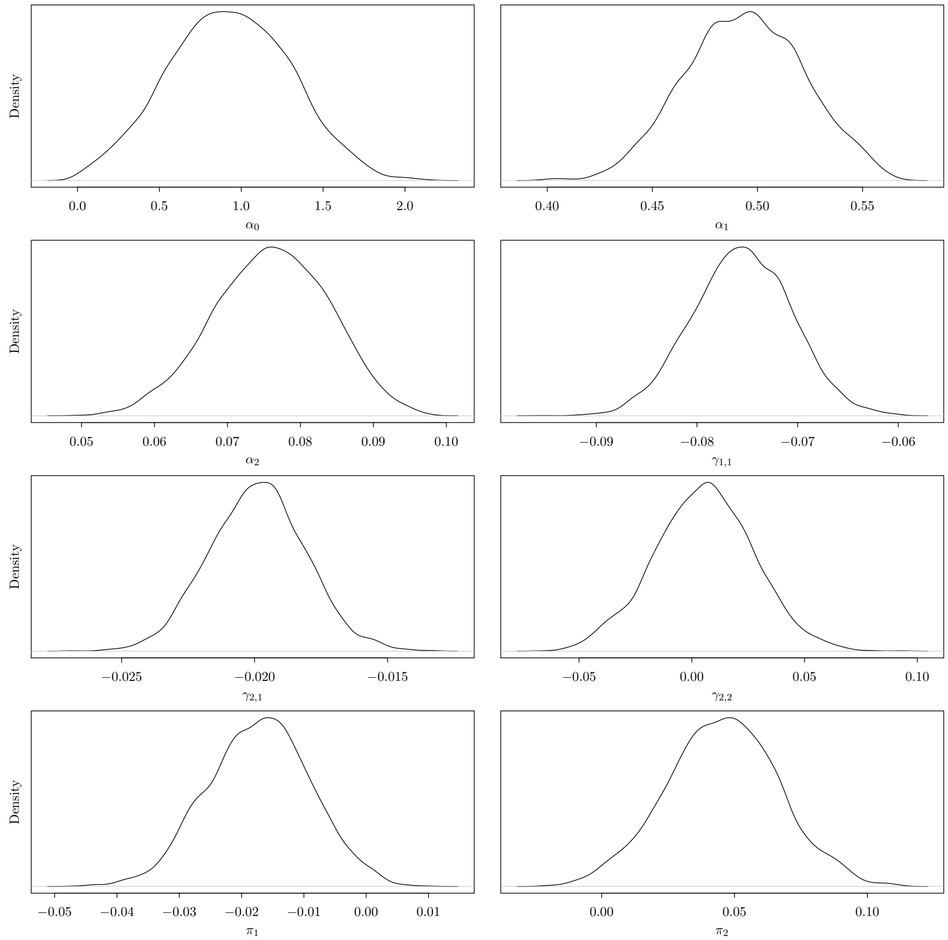
In order to resolve some of these issues and improve the fit, the paper now considers an extension of the AID system to account for polynomials on deflated real expenditures $\tilde{e}$. In particular, the

Table 2.14: Elasticity Estimates and Inference from a Bayesian AID System

Good	Elasticities			
	Income	Price (1)	Price (2)	Price (3)
Single member households, 2,218 observations				
(1)	0.991	−1.129	−0.049	0.188
	(0.031)	(0.027)	(0.008)	(0.023)
(2)	0.914	−0.221	−0.402	−0.291
	(0.674)	(0.507)	(0.723)	(0.828)
(3)	1.048	0.152	−0.065	−1.135
	(0.076)	(0.031)	(0.068)	(0.076)
Married couples without children, 3,326 observations				
(1)	0.986	−1.154	−0.049	0.218
	(0.021)	(0.023)	(0.006)	(0.022)
(2)	−0.420	0.931	−1.218	0.708
	(104.842)	(85.301)	(42.383)	(61.464)
(3)	0.926	0.224	−0.017	−1.133
	(0.051)	(0.031)	(0.055)	(0.063)
Married couples with one child, 6,141 observations				
(1)	0.966	−1.136	−0.038	0.207
	(0.016)	(0.017)	(0.005)	(0.016)
(2)	1.539	−0.531	−1.174	0.166
	(52.174)	(42.687)	(16.385)	(19.273)
(3)	0.941	0.235	0.036	−1.212
	(0.046)	(0.029)	(0.049)	(0.061)

Note: Elasticities are computed at the average median-normed expenditures and average prices for each chain. Point estimates are given by the mean of the chains. Standard deviations for the chains are in parentheses.

Figure 2.8: Density Plot of Coefficient Chains in a Reparameterized Bayesian AID System



Note: Results for the data set on married couples with one child. Combination of 5 chains with 700 draws each for a total of 3,500 draws.

following conditional mean obtained in one of the examples is used:

$$\tilde{e}_{\text{new},i} \equiv e_{\text{new},i} - \alpha_0 - \boldsymbol{\alpha}'\boldsymbol{p}_i - (1/2)\boldsymbol{p}_i'\Gamma\boldsymbol{p}_i \,,$$

$$\mathrm{E}[\boldsymbol{Y}_i | e_{\text{new},i}, \boldsymbol{p}_i] = \boldsymbol{\alpha} + \Gamma\boldsymbol{p}_i + \sum_{r=1}^{R} \boldsymbol{\pi}_r \tilde{e}_{\text{new},i}^{r} \,.$$

The reparameterization of the model in terms of the median-centered expenditure also plays a crucial role in this setting as it makes the magnitudes of the coefficients $\boldsymbol{\pi}_r, r = 1, \ldots, R$, directly comparable (Blundell et al., 1993; Lewbel and Pendakur, 2009). Having this standardized measure of the covariates allows for selection to be both accurate and more meaningful. For simplicity, R is set equal to 3, so that there is a third-degree polynomial on the conditional mean equation for each share. To implement the estimation and shrinkage of the coefficients using the LASSO penalty, the following priors are assumed:

$$
\begin{aligned}
\alpha_0 &\sim \mathcal{N}(0,5)\,, \\
\alpha_j &\sim \mathcal{N}(0,1), j=1,2\,, \\
\gamma_{j,l} &\sim \mathcal{N}(0,1), j=1,2, l \leq j,\,, \\
\pi_{r,j}|\tau_{r,j} &\sim \mathcal{N}(0,\tau_{r,j}), j=1,2, r=1,2,3\,, \\
\tau_{r,j}|\lambda^2 &\sim \text{Exponential}\left(\frac{\lambda^2}{2}\right)\,, \\
\lambda^2 &\sim \text{Exponential}(1)\,, \\
\phi_j &\sim \text{Gamma}(1,1), j=1,2\,, \\
\psi &\sim \text{Uniform}(-1,1)\,.
\end{aligned}
$$

The results for selection performance are given in Table 2.15. Using the credible interval and scaled neighborhood approaches to selection in the Bayesian framework, it appears that a third-degree polynomial on deflated expenditures is relevant for modeling the demand for gasoline. It does not seem to be the case for local transportation, where the methods are dependent on the demographic characteristics of the consumers. For example, while the second-order term is significant in the single-member households, no polynomial is selected for the married without children households. In the final population segment, both measures are inconclusive and this is the only instance in which the methods disagree with one another.

Simultaneous to the selection step, the estimation of the extended AID coefficients is straightforward. Table 2.14 presents the results for the income and price elasticities in this model, which are simple to obtain due to the Bayesian approach. Furthermore, it appears that the inclusion of the polynomial terms not only makes the model more flexible, but it also stabilizes the values and inference for these elasticities. The signs are in concordance with economic theory: all of the goods are normal with a relatively large income elasticity that is close to unity. The own-price elasticities are all negative and suggest that gasoline and intercity transportation are slightly elastic, whereas local transport is somewhat inelastic. The magnitudes also vary across the demographic groups, with married couples with one child having the largest price reactions. As these elasticities are

Table 2.15: Selection of Polynomial Terms in an Extended Bayesian AID System

Polynomial	CI (1)	CI (2)	SN (1)	SN (2)
Single member households, 2,218 observations				
$\tilde{e}$	✓	✓	✓	✓
$\tilde{e}^2$	×	✓	×	✓
$\tilde{e}^3$	✓	×	✓	×
Married couples without children, 3,326 observations				
$\tilde{e}$	✓	✓	✓	✓
$\tilde{e}^2$	✓	×	✓	×
$\tilde{e}^3$	✓	×	✓	×
Married couples with one child, 6,141 observations				
$\tilde{e}$	✓	✓	✓	✓
$\tilde{e}^2$	✓	✓	✓	×
$\tilde{e}^3$	×	✓	×	✓

Note: CI (1) and CI (2) represents credible interval selection with $\bar{l} = 0.5$ for each good's equation. SN (1) and SN (2) uses the scaled neighborhood method with $\bar{p} = 0.5$; "✓" indicates a variable is present in that outcome's equation; and "×" denotes its absence. The Bayesian algorithm chooses a regularization parameter $\lambda = 1.97$ for the first sample; $\lambda = 1.95$ for the second and third.

uncompensated, the possibility of these households reacting to price variations might bear some correlation with income or other socioeconomic variables. These interactions might not be fully accounted for by the use of different estimation samples. The cross-price elasticities are slightly more erratic, as they suggest some substitution effect between gasoline and intercity transportation, but the complementary nature of gasoline and local transport is maintained (as is seen in CS). Figures 2.9 and 2.10 present the trace and density plots for these elasticities, respectively.

2.5 Conclusion

The paper introduces several estimation procedures for multivariate fractional outcomes, which are useful in both structural and reduced form contexts. A likelihood function is constructed using copulas in two ways, one of which is found to be robust to deviations from the model assumptions. These likelihoods also allow for more flexibility in the dependence structure between shares than the usual joint distributions assumed on outcomes in the unit-simplex. Both of the introduced methods allow the researcher to satisfy the main characteristic that comes with multivariate fractional responses — a conditional mean specification and the fractional and unit-sum restrictions in the outcomes — and allows for the inclusion of cross-equation restrictions. The latter point is

Table 2.16: Elasticity Estimates and Inference from an Extended Bayesian AID System

Good	Elasticities			
	Income	Price (1)	Price (2)	Price (3)
Single member households, 2,218 observations				
(1)	0.966	−1.226	−0.009	0.270
	(0.012)	(0.053)	(0.050)	(0.062)
(2)	1.056	−0.094	−0.804	−0.158
	(0.053)	(0.252)	(0.057)	(0.272)
(3)	1.023	0.228	−0.023	−1.227
	(0.016)	(0.065)	(0.052)	(0.112)
Married couples without children, 3,326 observations				
(1)	0.958	−1.247	−0.041	0.331
	(0.010)	(0.067)	(0.040)	(0.082)
(2)	1.049	−0.323	−0.890	0.164
	(0.083)	(0.294)	(0.083)	(0.333)
(3)	1.035	0.278	0.025	−1.338
	(0.019)	(0.060)	(0.048)	(0.099)
Married couples with one child, 6,141 observations				
(1)	0.956	−1.321	−0.101	0.466
	(0.013)	(0.090)	(0.033)	(0.119)
(2)	0.943	−0.614	−1.020	0.692
	(0.057)	(0.221)	(0.059)	(0.258)
(3)	1.057	0.438	0.110	−1.605
	(0.018)	(0.057)	(0.040)	(0.086)

Note: Elasticities are computed at the average median-normed expenditures and average prices for each chain. Point estimates are given by the mean of the chains. Standard deviations for the chains are in parentheses.

of particular importance in structural demand estimation models where these restrictions are at the heart of guaranteeing economic regularity of the underlying demand functions. The paper also shows how Bayesian methods can be crucial in this setting by showing how the methods can be augmented to handle covariate selection using a Bayesian analog of regularization. Inference is still simple in this framework, even after performing a selection step, which can be hard to accomplish in frequentist settings. As the objects of interest in applied research are complicated functions of the parameters, the Bayesian approach allows for a natural way to handle inference of these quantities as well. Numerical exercises and an empirical application of a structural demand system to transportation expenditures in Canada showcase the flexibility of the proposed methods and their usefulness in an applied setting.

As a matter of future research, it would be interesting to extend this kind of Bayesian copula estimation to broader settings apart from the multivariate fractional outcome context. While Bayesian methods, regularization, and copulas are popular topics in econometrics and statistics, the combination of all of these elements could prove to be valuable in adding flexibility while preserving structure in different modeling problems. Additionally, it would be interesting to bring these tools to more applications in which multivariate fractional outcomes naturally arise. Examples include data for market shares on a given industry, portfolio shares in financial econometrics, industrial organization and firm analysis, among many others.

Appendices

2.A Proof of Main Results

Proof of Proposition 1. This is a specialized version of the formulas in Gijbels and Herrmann (2014). As

$$F_W(w|\boldsymbol{X};\boldsymbol{\delta},\boldsymbol{\eta}) = \int_{\mathcal{T}_w} dF_{1,\ldots,D}(y_1,\ldots,y_D|\boldsymbol{X};\boldsymbol{\delta},\boldsymbol{\eta}),$$

where $\mathcal{T}_w = \{(y_1,\ldots,y_D) \in \mathbb{R}^D : 0 \le y_j \le 1, j = 1,\ldots,d; \sum_{j=1}^D y_j \le w\}$, then the set $\mathcal{T}_w$ can be expressed using multiple integrals corresponding to (2.7). □

Proof of Proposition 2. The existence of a solution is guaranteed if $\sum_{j=1}^{d} m_j(\boldsymbol{x}, \boldsymbol{\beta}) = 1$ is imposed, as the right-hand term of (2.11) will always be less than 1. To obtain a solution, first note that the inverse mapping for the stick-breaking transformation (2.9), $\boldsymbol{Y} = \boldsymbol{s}^{-1}(\boldsymbol{Z})$, is given by

$$Y_1 = Z_1, \quad Y_j = Z_j \prod_{l=1}^{j-1}(1 - Z_l) \quad \text{for } j = 2, \ldots, d\,. \tag{2.A.1}$$

Additionally, this mapping satisfies the following property:

$$\prod_{l=1}^{j}(1 - Z_l) = 1 - \sum_{l=1}^{j} Y_l\,, \tag{2.A.2}$$

for $j = 1, \ldots, D$. First, set $\mu_1(\boldsymbol{x}; \boldsymbol{\gamma}, \boldsymbol{\psi}) = m_1(\boldsymbol{x}, \boldsymbol{\beta})$. For $j = 2, \ldots, D$, take the definition of Y_j in (2.A.1), replace $Z_j = \widetilde{Z}_j + m_j(\boldsymbol{x}, \boldsymbol{\beta}_j)$, and take conditional expectations on both sides. This results in

$$m_j(\boldsymbol{x}, \boldsymbol{\beta}) = \mathrm{E}\left[\widetilde{Z}_j \prod_{l=1}^{j-1}\Big(1 - \widetilde{Z}_l - \mu_l(\boldsymbol{x}; \boldsymbol{\gamma}, \boldsymbol{\psi})\Big) \middle| \boldsymbol{X} = \boldsymbol{x}\right] + \mu_j(\boldsymbol{x}; \boldsymbol{\gamma}, \boldsymbol{\psi}) \cdot \mathrm{E}\left[\prod_{l=1}^{j-1}(1 - Z_l) \middle| \boldsymbol{X} = \boldsymbol{x}\right]$$

While the first expectation cannot be reduced, the second can be replaced by taking conditional expectations of (2.A.2) for $j - 1$. Dividing by this term gives the desired result. □

Proof of Theorem 2.1. For $\widehat{\boldsymbol{\theta}}_Y$, the only non-standard part of the likelihood is the integral corresponding to the probability of set $\mathcal{T}$, given by $\Pr_f(\boldsymbol{Y}_{-d} \in \mathcal{T} | \boldsymbol{X} = \boldsymbol{x}_i; \boldsymbol{\theta}_Y)$, where the subscript emphasizes that the probability is taken with respect to the assumed joint distribution. However, since $\boldsymbol{\theta}_{Y,0}$ satisfies $H(\cdot|\boldsymbol{X}) = F(\cdot|\boldsymbol{X}; \boldsymbol{\theta}_{Y,0})$ by Assumption 2.6.A, the relevant probability becomes $\Pr_h(\boldsymbol{Y}_{-d} \in \mathcal{T} | \boldsymbol{X} = \boldsymbol{x}_i)$, where the notation emphasizes that it is taken with respect to the true H. This probability equals 1, as it is assumed that H is a joint distribution with support in $\mathcal{S}^d$. Thus, the log of this probability equals 0 and the term is irrelevant in the population. The usual argument would then guarantee consistency in light of Assumption 2.5; the same is true for $\widehat{\boldsymbol{\theta}}_Z$. The rest of the argument for asymptotic normally is standard as outlined; e.g., in Joe (2014), pp. 227. □

Proof of Lemma 2.1. First, note that since P_X (the marginal distribution of $\boldsymbol{X}$) is given, we have

$$\mathrm{KL}(h, f; \boldsymbol{\theta}_Y) = \mathrm{E}_P[\mathrm{KL}(h_{Y|X}, f_{Y|X}; \boldsymbol{\theta}_Y)]\,, \tag{2.A.3}$$

where E_P means that the expectation is taken with respect to $\boldsymbol{X} \sim P_X$ and $\mathrm{KL}(h_{Y|X}, f_{Y|X}; \boldsymbol{\theta}_Y)$ is the KL divergence between the conditional distributions $h(\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x})$ and $f(\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\theta}_Y)$. Thus, we only need to focus on the conditional KL divergence. This can be derived as follows:

$$\begin{aligned}
\log\left[\frac{h(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x})}{f(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\theta}_Y)}\right] &= \log\Bigg[\frac{c(H_1(Y_1|\boldsymbol{X}=\boldsymbol{x}),\ldots,H_D(Y_D|\boldsymbol{X}=\boldsymbol{x}))}{c_Y(F_1(Y_1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_1),\ldots,F_D(Y_D|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_D);\boldsymbol{\eta})} \times \\
&\qquad \prod_{j=1}^{D}\frac{h_j(Y_j|\boldsymbol{X}=\boldsymbol{x})}{f_j(Y_D|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_j)} \times \frac{F_W(1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\theta}_Y)}{\mathbb{I}(\boldsymbol{Y}\in\mathcal{T})}\Bigg] \\
&= \log\left[\frac{c(H_1(Y_1|\boldsymbol{X}=\boldsymbol{x}),\ldots,H_D(Y_D|\boldsymbol{X}=\boldsymbol{x}))}{c_Y(F_1(Y_1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_1),\ldots,F_D(Y_D|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_D);\boldsymbol{\eta})}\right] + \\
&\qquad \sum_{j=1}^{D}\log\left[\frac{h_j(Y_j|\boldsymbol{X}=\boldsymbol{x})}{f_j(Y_D|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_j)}\right] + \log\left[\frac{F_W(1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\theta}_Y)}{\mathbb{I}(\boldsymbol{Y}\in\mathcal{T})}\right].
\end{aligned}$$

Taking conditional expectations with respect to $h(\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x})$ yields $\mathrm{KL}(h_{Y|X}, f_{Y|X}; \boldsymbol{\theta}_Y)$. Due to (2.A.3), another expectation — this time with respect to P_X — gives the desired result. □

Proof of Theorem 2.2. From Lemma 2.1, we can write the KL divergence as

$$\begin{aligned}
\mathrm{KL}(h, f; \boldsymbol{\theta}_Y) = \underbrace{\mathrm{E}_h\left[\log\frac{c(H_1(Y_1|\boldsymbol{X}=\boldsymbol{x}),\ldots,H_D(Y_D|\boldsymbol{X}=\boldsymbol{x}))}{c_Y(F_1(Y_1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_1),\ldots,F_D(Y_D|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_D);\boldsymbol{\psi})}\right]}_{T_1} + \\
\underbrace{\sum_{j=1}^{D}\mathrm{KL}(h_j, f_j; \boldsymbol{\delta}_j)}_{T_2} + \underbrace{\mathrm{E}_h\left[\log\frac{F_W(1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\theta}_Y)}{\mathbb{I}(\boldsymbol{Y}\in\mathcal{T})}\right]}_{T_3},
\end{aligned}$$

where there are three terms, T_1, T_2, and T_3, each representing a divergence measure between either the copulas, marginals, or truncation probability. Similar to the proof of Theorem 2.1, $\mathrm{E}_h[\log \mathbb{I}(\boldsymbol{Y} \in \mathcal{T})] = 0$ under the true density. Furthermore, as long as $f(\cdot)$ places a positive amount of density in $\mathcal{T}$, the numerator of the T_3 term will be well-defined.

Now, based on Assumptions 2.5 and 2.6.B, there exists a true $\boldsymbol{\delta}_0$ that correctly specifies all the marginals, but no $\boldsymbol{\eta}$ that does so for the copula. Evaluating T_2 at $\boldsymbol{\delta}_0$ shows that $\mathrm{KL}(h_j, f_j; \boldsymbol{\delta}_{j,0}) = \mathrm{KL}(h_j, h_j) = 0, j = 1, \ldots, D$. Similarly, evaluating T_1 at $\boldsymbol{\delta}_0$ yields

$$\mathrm{E}_h \left[\log \frac{c(H_1(Y_1|\boldsymbol{X}=\boldsymbol{x}), \ldots, H_D(Y_D|\boldsymbol{X}=\boldsymbol{x}))}{c_Y(H_1(Y_1|\boldsymbol{X}=\boldsymbol{x}), \ldots, F_D(Y_D|\boldsymbol{X}=\boldsymbol{x}); \boldsymbol{\psi})} \right],$$

so that T_1 reduces to the KL divergence based solely on the dependence structure. Thus, consistency of the subvector $\widehat{\boldsymbol{\delta}}$ in $\widehat{\boldsymbol{\theta}}_Y$ to $\boldsymbol{\delta}_0$ is guaranteed by Theorem 2.2 in White (1982). Consistency of $\widehat{\boldsymbol{\eta}}$ is guaranteed to $\boldsymbol{\eta}^*$, which is the minimizer of T_1 and the maximizer of T_3 given $\boldsymbol{\delta}_0$. Asymptotic normality follows from Theorem 3.2 in White (1982) and requires the full sandwich covariance matrix as there is no diagonality in either $\mathcal{I}_h$ or $\mathcal{J}_h$ to exploit in the copula estimation (see Joe, 2014, pp. 228). □

Proof of Corollary 2.1. In this setting, similar to Theorem 2.2, the KL divergence can be split into two terms:

$$\mathrm{KL}(h, f; \boldsymbol{\theta}_Y) = \underbrace{\mathrm{E}_h \left[\log \frac{c(H_1(Y_1|\boldsymbol{X}=\boldsymbol{x}), \ldots, H_D(Y_D|\boldsymbol{X}=\boldsymbol{x}))}{c_Y(F_1(Y_1|\boldsymbol{X}=\boldsymbol{x}; \boldsymbol{\delta}_1), \ldots, F_D(Y_D|\boldsymbol{X}=\boldsymbol{x}; \boldsymbol{\delta}_D); \boldsymbol{\psi})} \right]}_{T_1} + \underbrace{\sum_{j=1}^{D} \mathrm{KL}(h_j, f_j; \boldsymbol{\delta}_j)}_{T_2}.$$

As T_2 vanishes when evaluated at $\boldsymbol{\delta}_0$ and T_1 becomes the KL divergence between the copula dependence structures, the proof can follow the same steps as that of Theorem 2.2 to show consistency and asymptotic normality. □

Proof of Theorem 2.3. (i) Note that the assumptions plus the additional regularity conditions are stronger than those needed for correctly specified Bayesian posteriors (see, e.g., Theorem 2.3 in Strasser, 1981). This guarantees consistency of the posterior distribution as a whole in neighborhoods around $\boldsymbol{\theta}_{e,0}$ for $e \in \{Y, Z\}$. That is, for any open set $\mathcal{U}$ containing $\boldsymbol{\theta}_{e,0}$,

$$\lim_{n\to\infty} \pi(\mathcal{U}|\boldsymbol{Y}, \boldsymbol{X}) = 1\,, \tag{2.A.4}$$

where $\pi(\mathcal{U}|\boldsymbol{Y},\boldsymbol{X})$ is defined as the posterior probability in set $\mathcal{U}$; i.e.,

$$\pi(\mathcal{U}|\boldsymbol{Y},\boldsymbol{X}) = \int_{\mathcal{U}} \pi(\theta_e|\boldsymbol{Y},\boldsymbol{X})\ \mathrm{d}\theta_e = \int_{\mathcal{U}} \frac{\ell_e(\theta_e)\pi(\theta_e)}{\int_{\Theta_e} \ell_e(\theta_e)}\ \mathrm{d}\theta_e\ .$$

(ii) Similarly, under the established assumptions and regularity conditions, the Bayesian posterior are consistent in a KL divergence sense. Formally, this implies that consistency is not to $\boldsymbol{\theta}_{Y,0}$, but to the KL pseudo-true values (minimizers of the KL divergence). Thus, (2.A.4) holds for open sets $\mathcal{U}$ containing $\boldsymbol{\theta}_Y^*$ (see, e.g., Theorem 2.1 in Bunke and Milhaud, 1998).

Establishing posterior consistency yields mean and mode consistency of the posteriors, so that (i) $\check{\boldsymbol{\theta}}_e \xrightarrow{p} \boldsymbol{\theta}_{e,0}$ for $e \in \{Y, Z\}$ and (ii) $\check{\boldsymbol{\theta}}_Y \xrightarrow{p} \boldsymbol{\theta}_Y^*$. The median can also be shown to hold this property (see Remarks 3, 4, and 5 in Bunke and Milhaud, 1998). □

2.B Regularity Conditions

This is a list of the necessary regularity conditions required for the paper's proofs. It essentially reproduces the assumptions in White (1982) and Bunke and Milhaud (1998) that are not implied by Assumptions 2.1–2.6.B. To simplify notation, let $\boldsymbol{U} = (\boldsymbol{Y}', \boldsymbol{X}')' \subset \mathcal{S}^d \times \mathcal{X} = \Upsilon$. Then, for $u \in \Upsilon$ write $F(\boldsymbol{u}, \boldsymbol{\theta}_Y) = F(\boldsymbol{y}|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\theta}_Y)P_X(\boldsymbol{x})$ and let $f(\boldsymbol{u}, \boldsymbol{\theta}_Y)$ be its associated density. The density $g(\boldsymbol{u}, \boldsymbol{\theta}_Z)$ is defined analogously. Both of these densities are assumed to be obtained with respect to a measure ν.

Assumption R1. The densities $f(\boldsymbol{u}, \boldsymbol{\theta}_Y)$ and $g(\boldsymbol{u}, \boldsymbol{\theta}_Z)$ are measurable in $\boldsymbol{u}$ for all $\boldsymbol{\theta}_Y \in \Theta_Y$ and $\boldsymbol{\theta}_Z \in \Theta_Z$, as well as continuous in $\boldsymbol{\theta}_Y$ and $\boldsymbol{\theta}_Z$ for all $u \in \Upsilon$. Θ_Y and Θ_Z are also assumed to be compact.

Assumption R2. (i) The expectation $\mathrm{E}[\log h(\boldsymbol{U})]$ exists and both $\log f(\boldsymbol{u}, \boldsymbol{\theta}_Y)$ and $\log g(\boldsymbol{u}, \boldsymbol{\theta}_Z)$ are dominated by functions integrable with respect to H. (ii) $\mathrm{KL}(h, f; \boldsymbol{\theta}_Y)$ has a unique minimum at $\boldsymbol{\psi}^* \in \Psi$ given $\boldsymbol{\delta}_0$.

Assumption R3. The gradients $\partial \log f(\boldsymbol{u}, \boldsymbol{\theta}_Y)/\partial \boldsymbol{\theta}_Y$ and $\partial \log g(\boldsymbol{u}, \boldsymbol{\theta}_Z)/\partial \boldsymbol{\theta}_Z$ are measurable functions of $\boldsymbol{u}$ for each $\boldsymbol{\theta}_e \in \Theta_e$ and continuously differentiable functions of $\boldsymbol{\theta}_e$ for each $\boldsymbol{u} \in \Upsilon$, where $e \in \{Y, Z\}$.

Assumption R4. These derivatives $\left\| \partial^2 \log f(\boldsymbol{u}, \boldsymbol{\theta}_Y)/\partial\boldsymbol{\theta}_Y\partial\boldsymbol{\theta}_Y' \right\|_2$, $\left\| \partial^2 \log g(\boldsymbol{u}, \boldsymbol{\theta}_Z)/\partial\boldsymbol{\theta}_Z\partial\boldsymbol{\theta}_Z' \right\|_2$, $\| \partial \log f(\boldsymbol{u}, \boldsymbol{\theta}_Y)/\partial\boldsymbol{\theta}_Y \cdot \partial \log f(\boldsymbol{u}, \boldsymbol{\theta}_Y)/\partial\boldsymbol{\theta}_Y' \|_2$ and $\| \partial \log g(\boldsymbol{u}, \boldsymbol{\theta}_Z)/\partial\boldsymbol{\theta}_Z \cdot \partial \log g(\boldsymbol{u}, \boldsymbol{\theta}_Z)/\partial\boldsymbol{\theta}' \|_2$ are dominated by functions integrable with respect to H for all $\boldsymbol{u} \in \Upsilon$, $\boldsymbol{\theta}_Y \in \Theta_Y$ and $\boldsymbol{\theta}_Z \in \Theta_Z$.

Assumption R5. For the information equality, $\| \partial[\partial \log f(\boldsymbol{u}, \boldsymbol{\theta}_Y)/\partial\boldsymbol{\theta}_Y \cdot f(\boldsymbol{u}, \boldsymbol{\theta}_Y)]/\partial\boldsymbol{\theta}_Y \|_2$ and $\| \partial[\partial \log g(\boldsymbol{u}, \boldsymbol{\theta}_Z)/\partial\boldsymbol{\theta}_Z \cdot g(\boldsymbol{u}, \boldsymbol{\theta}_Z)]/\partial\boldsymbol{\theta}_Z \|_2$ are dominated by functions integrable with respect to ν for all $\boldsymbol{\theta}_Y \in \Theta_Y$ and $\boldsymbol{\theta}_Z \in \Theta_Z$.

Assumption R6. (i) $\boldsymbol{\theta}_{Y,0}, \boldsymbol{\theta}_Y^* \in \text{int}(\Theta_Y)$ and $\boldsymbol{\theta}_{Z,0} \in \text{int}(\Theta_Z)$; (ii) $\mathcal{I}(\boldsymbol{\theta}_{Y,0})$, $\mathcal{I}(\boldsymbol{\theta}_{Z,0})$ and $\mathcal{I}(\boldsymbol{\theta}_Y^*)$ have constant rank in a neighborhood of their arguments; (iii) $\mathcal{J}_h(\boldsymbol{\theta}_Y^*)$ is nonsingular.

Assumption R7. There are positive constants c, b_0 such that for all $\boldsymbol{\theta}_Y \in \Theta_Y$

$$\int \left\| \frac{\partial \log f(\boldsymbol{u}, \boldsymbol{\theta}_Y)}{\partial\boldsymbol{\theta}_Y} \right\|_2^{4(|\Theta_Y|+1)} f(\boldsymbol{u}, \boldsymbol{\theta}_Y)\nu(\mathrm{d}\boldsymbol{u}) < c(1 + \|\boldsymbol{\theta}_Y\|^{b_0}),$$

where $|\Theta_Y|$ is the dimensionality of Θ_Y. The same condition holds for $g(\boldsymbol{u}, \boldsymbol{\theta}_Z)$.

Assumption R8. For some positive constant b_1, $\int [f(\boldsymbol{u}, \boldsymbol{\theta}_Y)h(\boldsymbol{u})]^{1/2}\nu(\mathrm{d}\boldsymbol{u}) < c\|\boldsymbol{\theta}_Y\|^{-b_1}$ and $\int [g(\boldsymbol{u}, \boldsymbol{\theta}_Z)h(\boldsymbol{u})]^{1/2}\nu(\mathrm{d}\boldsymbol{u}) < c\|\boldsymbol{\theta}_Z\|^{-b_1}$, for all $\boldsymbol{\theta}_Y \in \Theta_Y$ and $\boldsymbol{\theta}_Z \in \Theta_Z$.

Assumption R9. Take $e \in \{Y, Z\}$ and let $S(\boldsymbol{\theta}_e, r)$ represent a ball centered at $\boldsymbol{\theta}_e$ with radius r. Then, $\pi(\boldsymbol{\theta}_e)$ assigns probability $\pi(S(\boldsymbol{\theta}_e, r)) > 0$ for all $\boldsymbol{\theta}_e \in \Theta_e$ and $r > 0$, and there are positive constants b_2 and b_3 so that for all $\boldsymbol{\theta}_e \in \Theta_e$ and $r > 0$ it holds that

$$\pi(S(\boldsymbol{\theta}_e, r)) \leq c \cdot r^{b_2}[1 + (\|\boldsymbol{\theta}_e\| + r)^{b_3}].$$

2.C Additional Numerical Exercises

Table 2.C.1: Estimates and Standard Errors in a Reduced Form Model from a Gaussian Copula with Beta Marginals

Method	$\beta_{0,1}$	$\beta_{1,1}$	$\beta_{2,1}$	$\beta_{0,2}$	$\beta_{1,2}$	$\beta_{2,2}$	ϕ_1	ϕ_2	$\psi\|\xi$
				$n = 100$					
Copula Y	−1.027	0.492	−0.013	−1.538	−0.018	0.495	10.809	10.947	0.486
	(0.085)	(0.079)	(0.079)	(0.103)	(0.090)	(0.089)	(1.503)	(1.585)	(0.124)
Copula Z	−1.015	0.481	−0.014	−1.490	−0.062	0.483	10.802	5.268	0.625
	(0.084)	(0.080)	(0.080)	(0.098)	(0.088)	(0.090)	(1.515)	(0.744)	(0.111)
MF Logit	−1.024	0.487	−0.017	−1.536	−0.026	0.490	—	—	—
	(0.085)	(0.084)	(0.083)	(0.103)	(0.098)	(0.1)			
Dirichlet	−0.950	0.480	0.000	−1.430	−0.003	0.476	8.473	—	—
	(0.079)	(0.078)	(0.078)	(0.091)	(0.086)	(0.086)	(0.825)		
Logistic Norm.	−1.153	0.621	−0.017	−1.862	−0.068	0.736	—	—	—
	(0.108)	(0.108)	(0.108)	(0.141)	(0.142)	(0.142)			
				$n = 200$					
Copula Y	−1.026	0.493	−0.009	−1.535	−0.018	0.497	10.614	10.711	0.484
	(0.060)	(0.056)	(0.056)	(0.073)	(0.063)	(0.063)	(1.042)	(1.097)	(0.088)
Copula Z	−1.014	0.480	−0.010	−1.486	−0.064	0.484	10.610	5.138	0.621
	(0.059)	(0.056)	(0.056)	(0.070)	(0.062)	(0.063)	(1.044)	(0.506)	(0.078)
MF Logit	−1.023	0.487	−0.015	−1.532	−0.026	0.491	—	—	—
	(0.060)	(0.060)	(0.059)	(0.073)	(0.070)	(0.071)			
Dirichlet	−0.949	0.481	0.003	−1.427	−0.003	0.478	8.304	—	—
	(0.056)	(0.055)	(0.055)	(0.064)	(0.060)	(0.061)	(0.571)		
Logistic Norm.	−1.155	0.623	−0.014	−1.864	−0.069	0.740	—	—	—
	(0.076)	(0.077)	(0.077)	(0.101)	(0.101)	(0.101)			
				$n = 400$					
Copula Y	−1.026	0.494	−0.009	−1.535	−0.015	0.498	10.522	10.637	0.483
	(0.042)	(0.039)	(0.039)	(0.051)	(0.045)	(0.044)	(0.730)	(0.770)	(0.062)
Copula Z	−1.015	0.482	−0.010	−1.485	−0.061	0.485	10.520	5.095	0.620
	(0.042)	(0.040)	(0.039)	(0.050)	(0.044)	(0.045)	(0.739)	(0.361)	(0.056)
MF Logit	−1.023	0.489	−0.014	−1.532	−0.023	0.492	—	—	—
	(0.043)	(0.042)	(0.042)	(0.052)	(0.049)	(0.050)			
Dirichlet	−0.949	0.482	0.004	−1.426	0.000	0.479	8.243	—	—
	(0.039)	(0.039)	(0.038)	(0.045)	(0.043)	(0.043)	(0.401)		
Logistic Norm.	−1.157	0.626	−0.014	−1.865	−0.065	0.742	—	—	—
	(0.054)	(0.054)	(0.054)	(0.071)	(0.071)	(0.071)			
				$n = 800$					
Copula Y	−1.026	0.494	−0.009	−1.534	−0.013	0.498	10.465	10.566	0.480
	(0.030)	(0.028)	(0.028)	(0.036)	(0.032)	(0.031)	(0.514)	(0.541)	(0.044)
Copula Z	−1.012	0.483	−0.009	−1.482	−0.058	0.485	10.469	5.056	0.618
	(0.032)	(0.029)	(0.029)	(0.039)	(0.031)	(0.032)	(0.560)	(0.257)	(0.041)
MF Logit	−1.023	0.489	−0.014	−1.531	−0.022	0.491	—	—	—
	(0.030)	(0.030)	(0.030)	(0.037)	(0.035)	(0.035)			
Dirichlet	−0.948	0.482	0.003	−1.425	0.001	0.479	8.190	—	—
	(0.028)	(0.028)	(0.027)	(0.032)	(0.030)	(0.030)	(0.281)		
Logistic Norm.	−1.156	0.626	−0.015	−1.865	−0.063	0.741	—	—	—
	(0.038)	(0.038)	(0.038)	(0.051)	(0.050)	(0.051)			

Note: MLE estimates and (copula misspecification robust) asymptotic standard errors for each estimation procedure. Data are generated from a Gaussian copula with beta marginals. "—" implies the parameter is not part of the model.

Table 2.C.2: Estimates and Standard Errors in a Reduced Form Model from a FGM Copula with Beta Marginals

Method	$\beta_{0,1}$	$\beta_{1,1}$	$\beta_{2,1}$	$\beta_{0,2}$	$\beta_{1,2}$	$\beta_{2,2}$	ϕ_1	ϕ_2	$\psi\|\xi$
				$n = 100$					
Copula Y	−1.014	0.498	−0.004	−1.518	−0.007	0.501	10.646	10.626	0.283
	(0.082)	(0.077)	(0.077)	(0.099)	(0.087)	(0.087)	(1.437)	(1.491)	(0.126)
Copula Z	−1.000	0.496	0.008	−1.475	−0.036	0.505	10.629	5.686	0.472
	(0.084)	(0.079)	(0.078)	(0.105)	(0.087)	(0.087)	(1.465)	(0.953)	(0.121)
MF Logit	−1.013	0.499	−0.006	−1.517	−0.010	0.499	—	—	—
	(0.084)	(0.083)	(0.081)	(0.102)	(0.097)	(0.098)			
Dirichlet	−0.957	0.493	0.008	−1.441	0.006	0.490	8.848	—	—
	(0.078)	(0.077)	(0.076)	(0.089)	(0.085)	(0.085)	(0.863)		
Logistic Norm.	−1.153	0.631	−0.006	−1.857	−0.052	0.742	—	—	—
	(0.103)	(0.104)	(0.104)	(0.137)	(0.137)	(0.137)			
				$n = 200$					
Copula Y	−1.013	0.498	−0.004	−1.515	−0.008	0.500	10.413	10.394	0.280
	(0.058)	(0.055)	(0.054)	(0.070)	(0.062)	(0.062)	(1.007)	(1.040)	(0.090)
Copula Z	−0.973	0.518	0.040	−1.441	−0.008	0.534	10.264	5.487	0.485
	(0.066)	(0.061)	(0.064)	(0.084)	(0.069)	(0.068)	(1.076)	(0.698)	(0.091)
MF Logit	−1.012	0.498	−0.006	−1.514	−0.011	0.498	—	—	—
	(0.059)	(0.059)	(0.058)	(0.072)	(0.069)	(0.070)			
Dirichlet	−0.956	0.493	0.007	−1.438	0.006	0.489	8.666	—	—
	(0.055)	(0.054)	(0.054)	(0.063)	(0.060)	(0.060)	(0.597)		
Logistic Norm.	−1.154	0.632	−0.008	−1.860	−0.052	0.744	—	—	—
	(0.073)	(0.073)	(0.074)	(0.097)	(0.097)	(0.098)			
				$n = 400$					
Copula Y	−1.011	0.497	−0.005	−1.513	−0.007	0.499	10.272	10.275	0.280
	(0.045)	(0.040)	(0.041)	(0.054)	(0.046)	(0.045)	(0.740)	(0.776)	(0.067)
Copula Z	−0.958	0.536	0.065	−1.421	0.016	0.561	10.170	5.392	0.493
	(0.050)	(0.048)	(0.050)	(0.062)	(0.054)	(0.055)	(0.799)	(0.544)	(0.067)
MF Logit	−1.011	0.495	−0.007	−1.512	−0.011	0.496	—	—	—
	(0.042)	(0.042)	(0.041)	(0.051)	(0.049)	(0.049)			
Dirichlet	−0.954	0.491	0.007	−1.434	0.007	0.488	8.547	—	—
	(0.039)	(0.039)	(0.038)	(0.045)	(0.042)	(0.042)	(0.416)		
Logistic Norm.	−1.154	0.631	−0.009	−1.861	−0.054	0.745	—	—	—
	(0.052)	(0.052)	(0.052)	(0.069)	(0.069)	(0.069)			
				$n = 800$					
Copula Y	−1.011	0.497	−0.005	−1.514	−0.007	0.499	10.224	10.219	0.277
	(0.029)	(0.028)	(0.027)	(0.035)	(0.031)	(0.031)	(0.497)	(0.515)	(0.045)
Copula Z	−0.951	0.540	0.068	−1.408	0.021	0.561	10.176	5.411	0.485
	(0.036)	(0.036)	(0.034)	(0.046)	(0.039)	(0.038)	(0.595)	(0.397)	(0.051)
MF Logit	−1.010	0.495	−0.007	−1.512	−0.012	0.496	—	—	—
	(0.030)	(0.030)	(0.029)	(0.036)	(0.035)	(0.035)			
Dirichlet	−0.953	0.492	0.008	−1.435	0.006	0.488	8.512	—	—
	(0.028)	(0.027)	(0.027)	(0.032)	(0.030)	(0.030)	(0.293)		
Logistic Norm.	−1.155	0.632	−0.009	−1.863	−0.055	0.746	—	—	—
	(0.037)	(0.037)	(0.037)	(0.049)	(0.049)	(0.049)			

Note: MLE estimates and (copula misspecification robust) asymptotic standard errors for each estimation procedure. Data are generated from a Farlie–Gumbel–Morgenstern copula with beta marginals. "—" implies the parameter is not part of the model.

Table 2.C.3: Estimates and Standard Errors in a Reduced Form Model from a Dirichlet

Method	$\beta_{0,1}$	$\beta_{1,1}$	$\beta_{2,1}$	$\beta_{0,2}$	$\beta_{1,2}$	$\beta_{2,2}$	ϕ_1
			$n = 100$				
Copula Y	−1.004	0.498	0.000	−1.508	0.006	0.500	10.368
	(0.075)	(0.072)	(0.072)	(0.091)	(0.080)	(0.080)	(1.409)
Copula Z	−1.004	0.492	−0.001	−1.510	−0.025	0.504	10.366
	(0.075)	(0.072)	(0.071)	(0.091)	(0.080)	(0.080)	(1.410)
MF Logit	−1.004	0.498	−0.001	−1.508	0.003	0.501	—
	(0.076)	(0.076)	(0.075)	(0.093)	(0.089)	(0.090)	
Dirichlet	−1.004	0.497	−0.001	−1.505	0.005	0.498	10.319
	(0.073)	(0.073)	(0.072)	(0.085)	(0.081)	(0.081)	(1.011)
Logistic Norm.	−1.180	0.620	−0.017	−1.885	−0.048	0.734	—
	(0.091)	(0.091)	(0.092)	(0.123)	(0.124)	(0.124)	
			$n = 200$				
Copula Y	−1.003	0.499	0.000	−1.508	0.003	0.500	10.168
	(0.053)	(0.052)	(0.051)	(0.064)	(0.057)	(0.057)	(0.987)
Copula Z	−1.003	0.493	0.000	−1.510	−0.030	0.504	10.166
	(0.053)	(0.051)	(0.050)	(0.065)	(0.057)	(0.057)	(0.987)
MF Logit	−1.003	0.499	0.000	−1.508	0.000	0.500	—
	(0.054)	(0.054)	(0.053)	(0.066)	(0.063)	(0.064)	
Dirichlet	−1.003	0.498	−0.001	−1.505	0.002	0.498	10.156
	(0.052)	(0.051)	(0.051)	(0.060)	(0.057)	(0.057)	(0.703)
Logistic Norm.	−1.181	0.623	−0.018	−1.890	−0.053	0.736	—
	(0.065)	(0.065)	(0.065)	(0.088)	(0.088)	(0.088)	
			$n = 400$				
Copula Y	−1.003	0.499	0.000	−1.505	0.002	0.500	10.100
	(0.038)	(0.037)	(0.036)	(0.046)	(0.040)	(0.041)	(0.698)
Copula Z	−1.003	0.493	0.000	−1.507	−0.031	0.505	10.098
	(0.038)	(0.036)	(0.036)	(0.046)	(0.041)	(0.040)	(0.698)
MF Logit	−1.003	0.499	−0.001	−1.505	0.000	0.500	—
	(0.038)	(0.039)	(0.038)	(0.047)	(0.045)	(0.045)	
Dirichlet	−1.003	0.498	−0.001	−1.503	0.001	0.499	10.092
	(0.037)	(0.036)	(0.036)	(0.043)	(0.040)	(0.040)	(0.494)
Logistic Norm.	−1.182	0.623	−0.018	−1.887	−0.055	0.737	—
	(0.046)	(0.046)	(0.046)	(0.062)	(0.062)	(0.062)	
			$n = 800$				
Copula Y	−1.001	0.501	0.000	−1.502	0.001	0.501	10.066
	(0.027)	(0.026)	(0.025)	(0.032)	(0.029)	(0.029)	(0.493)
Copula Z	−1.001	0.494	0.000	−1.505	−0.032	0.505	10.062
	(0.027)	(0.026)	(0.025)	(0.032)	(0.029)	(0.029)	(0.493)
MF Logit	−1.001	0.501	−0.001	−1.501	0.001	0.499	—
	(0.027)	(0.027)	(0.027)	(0.033)	(0.032)	(0.032)	
Dirichlet	−1.001	0.501	−0.001	−1.501	0.001	0.499	10.054
	(0.026)	(0.026)	(0.025)	(0.030)	(0.028)	(0.028)	(0.348)
Logistic Norm.	−1.180	0.625	−0.018	−1.886	−0.056	0.737	—
	(0.032)	(0.032)	(0.032)	(0.044)	(0.044)	(0.044)	

Note: MLE estimates and (copula misspecification robust) asymptotic standard errors for each estimation procedure. Data are generated from a Dirichlet distribution. "—" implies the parameter is not part of the model.

Table 2.C.4: Estimates and Standard Errors in a Structural Demand Model from a Gaussian Copula with Beta Marginals

Method	α_0	α_1	α_2	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{2,2}$	π_1	π_2	ϕ_1	ϕ_2	$\psi\|\xi$
					$n = 100$						
Copula Y	0.665	0.806	0.205	0.069	−0.028	−0.051	−0.046	−0.017	13.685	15.320	0.322
	(7.458)	(0.435)	(0.221)	(0.147)	(0.079)	(0.084)	(0.052)	(0.030)	(1.873)	(2.203)	(0.134)
Copula Z	0.900	0.804	0.196	0.072	−0.030	−0.048	−0.047	−0.012	13.667	2.848	0.621
	(7.172)	(0.439)	(0.218)	(0.157)	(0.083)	(0.089)	(0.051)	(0.029)	(1.876)	(0.361)	(0.112)
MF Logit	0.626	0.816	0.197	0.063	−0.031	−0.046	−0.046	−0.014	—	—	—
	(1.775)	(0.307)	(0.194)	(0.161)	(0.084)	(0.106)	(0.052)	(0.031)			
Dirichlet	0.677	0.795	0.225	0.056	−0.030	−0.046	−0.042	−0.017	8.947	—	—
	(8.472)	(0.495)	(0.272)	(0.196)	(0.115)	(0.123)	(0.060)	(0.039)	(0.861)		
AID	0.839	0.790	0.150	0.069	−0.027	0.121	−0.046	−0.052	59.029	164.150	0.280
	(2.711)	(0.378)	(0.238)	(0.167)	(0.089)	(0.123)	(0.052)	(0.053)	(7.756)	(24.731)	(0.128)
					$n = 200$						
Copula Y	0.632	0.812	0.204	0.074	−0.027	−0.048	−0.047	−0.017	13.299	15.016	0.320
	(5.918)	(0.315)	(0.155)	(0.103)	(0.056)	(0.059)	(0.037)	(0.021)	(1.285)	(1.523)	(0.095)
Copula Z	0.513	0.822	0.192	0.075	−0.030	−0.044	−0.047	−0.012	13.270	2.804	0.618
	(5.514)	(0.293)	(0.138)	(0.106)	(0.058)	(0.063)	(0.036)	(0.021)	(1.286)	(0.250)	(0.079)
MF Logit	0.697	0.812	0.192	0.070	−0.029	−0.042	−0.047	−0.014	—	—	—
	(1.842)	(0.290)	(0.126)	(0.117)	(0.059)	(0.075)	(0.038)	(0.023)			
Dirichlet	0.714	0.805	0.227	0.065	−0.028	−0.044	−0.044	−0.017	8.724	—	—
	(6.598)	(0.345)	(0.181)	(0.139)	(0.082)	(0.087)	(0.042)	(0.027)	(0.593)		
AID	0.772	0.804	0.287	0.069	−0.042	−0.462	−0.046	−0.064	57.271	160.672	0.276
	(2.3)	(0.262)	(0.177)	(0.108)	(0.063)	(0.085)	(0.037)	(0.032)	(5.414)	(16.808)	(0.091)
					$n = 400$						
Copula Y	0.626	0.817	0.207	0.074	−0.027	−0.046	−0.048	−0.017	13.200	14.802	0.321
	(4.904)	(0.237)	(0.108)	(0.072)	(0.039)	(0.041)	(0.026)	(0.015)	(0.901)	(1.061)	(0.067)
Copula Z	0.808	0.820	0.195	0.076	−0.029	−0.042	−0.049	−0.012	13.217	2.798	0.616
	(3.599)	(0.177)	(0.081)	(0.074)	(0.040)	(0.044)	(0.025)	(0.015)	(0.9)	(0.176)	(0.056)
MF Logit	0.774	0.807	0.187	0.069	−0.028	−0.039	−0.048	−0.014	—	—	—
	(2.687)	(0.145)	(0.121)	(0.082)	(0.041)	(0.055)	(0.027)	(0.016)			
Dirichlet	0.726	0.804	0.226	0.065	−0.028	−0.041	−0.044	−0.017	8.628	—	—
	(5.437)	(0.252)	(0.127)	(0.097)	(0.058)	(0.062)	(0.030)	(0.019)	(0.415)		
AID	0.751	0.809	0.141	0.072	−0.027	0.097	−0.047	−0.028	57.251	158.636	0.274
	(1.043)	(0.162)	(0.103)	(0.074)	(0.043)	(0.079)	(0.026)	(0.020)	(3.785)	(11.754)	(0.064)
					$n = 800$						
Copula Y	0.582	0.817	0.206	0.076	−0.027	−0.044	−0.047	−0.016	13.141	14.684	0.322
	(3.671)	(0.173)	(0.069)	(0.050)	(0.027)	(0.029)	(0.018)	(0.010)	(0.635)	(0.744)	(0.047)
Copula Z	0.732	0.817	0.186	0.076	−0.028	−0.040	−0.048	−0.011	13.208	2.818	0.612
	(2.451)	(0.122)	(0.056)	(0.051)	(0.028)	(0.031)	(0.018)	(0.010)	(0.631)	(0.124)	(0.039)
MF Logit	0.769	0.811	0.190	0.070	−0.028	−0.036	−0.047	−0.013	—	—	—
	(1.490)	(0.180)	(0.063)	(0.066)	(0.031)	(0.038)	(0.021)	(0.012)			
Dirichlet	0.549	0.806	0.225	0.066	−0.028	−0.038	−0.044	−0.017	8.558	—	—
	(3.885)	(0.178)	(0.085)	(0.069)	(0.041)	(0.044)	(0.021)	(0.014)	(0.291)		
AID	0.746	0.803	0.192	0.070	−0.028	0.064	−0.046	−0.030	56.618	158.493	0.275
	(2.384)	(0.180)	(0.112)	(0.055)	(0.032)	(0.040)	(0.021)	(0.014)	(2.761)	(8.499)	(0.046)

Note: MLE estimates and (copula misspecification robust) asymptotic standard errors for each estimation procedure. Data are generated from a Gaussian copula with beta marginals. "—" implies the parameter is not part of the model.

Table 2.C.5: Estimates and Standard Errors in a Structural Demand Model from a Gaussian Distribution

Method	α_0	α_1	α_2	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{2,2}$	π_1	π_2	ϕ_1	ϕ_2	$\psi\|\xi$
					$n = 100$						
Copula Y	0.370	0.655	0.157	0.021	−0.010	−0.022	−0.025	0.003	5.473	7.388	−0.166
	(2.678)	(0.562)	(0.384)	(0.272)	(0.155)	(0.167)	(0.078)	(0.050)	(0.826)	(0.989)	(0.142)
Copula Z	0.573	0.682	0.158	0.033	−0.016	−0.013	−0.028	0.003	5.462	2.336	0.331
	(5.345)	(0.850)	(0.473)	(0.293)	(0.171)	(0.179)	(0.092)	(0.054)	(0.846)	(0.291)	(0.129)
MF Logit	0.708	0.626	0.163	0.022	−0.013	−0.015	−0.025	0.004	—	—	—
	(2.209)	(0.451)	(0.316)	(0.248)	(0.149)	(0.170)	(0.076)	(0.051)			
Dirichlet	0.799	0.597	0.184	0.019	−0.006	−0.016	−0.022	0.004	4.963	—	—
	(13.099)	(0.788)	(0.549)	(0.266)	(0.174)	(0.194)	(0.077)	(0.058)	(0.455)		
AID	0.556	0.622	0.165	0.026	−0.014	−0.018	−0.025	0.005	27.346	61.059	−0.200
	(13.008)	(0.730)	(0.524)	(0.250)	(0.152)	(0.170)	(0.077)	(0.052)	(3.876)	(8.652)	(0.139)
					$n = 200$						
Copula Y	1.154	0.592	0.177	0.038	−0.011	−0.012	−0.023	0.002	5.345	7.183	−0.164
	(2.237)	(0.369)	(0.208)	(0.183)	(0.103)	(0.116)	(0.056)	(0.036)	(0.592)	(0.689)	(0.103)
Copula Z	0.433	0.651	0.184	0.035	−0.016	−0.009	−0.027	0.001	5.329	2.331	0.324
	(5.505)	(0.579)	(0.340)	(0.207)	(0.117)	(0.125)	(0.057)	(0.038)	(0.603)	(0.206)	(0.091)
MF Logit	0.759	0.614	0.171	0.033	−0.012	−0.009	−0.023	0.003	—	—	—
	(1.274)	(0.304)	(0.196)	(0.174)	(0.104)	(0.120)	(0.054)	(0.037)			
Dirichlet	0.532	0.615	0.196	0.025	−0.010	−0.008	−0.020	0.003	4.854	—	—
	(11.215)	(0.495)	(0.320)	(0.179)	(0.119)	(0.133)	(0.054)	(0.041)	(0.314)		
AID	1.458	0.588	0.170	0.041	−0.014	−0.011	−0.023	0.003	26.854	59.740	−0.200
	(10.167)	(0.474)	(0.271)	(0.174)	(0.104)	(0.118)	(0.054)	(0.037)	(2.689)	(5.981)	(0.098)
					$n = 400$						
Copula Y	0.098	0.643	0.167	0.044	−0.011	−0.012	−0.025	0.002	5.299	7.111	−0.165
	(3.932)	(0.405)	(0.170)	(0.133)	(0.072)	(0.082)	(0.040)	(0.026)	(0.425)	(0.489)	(0.073)
Copula Z	0.837	0.635	0.191	0.061	−0.024	−0.008	−0.029	−0.001	5.293	2.439	0.315
	(2.038)	(0.242)	(0.140)	(0.139)	(0.076)	(0.101)	(0.038)	(0.025)	(0.496)	(0.199)	(0.081)
MF Logit	0.686	0.627	0.170	0.041	−0.012	−0.008	−0.025	0.003	—	—	—
	(7.470)	(0.649)	(0.363)	(0.141)	(0.083)	(0.099)	(0.041)	(0.031)			
Dirichlet	0.615	0.619	0.190	0.034	−0.010	−0.007	−0.023	0.003	4.821	—	—
	(8.601)	(0.309)	(0.195)	(0.125)	(0.084)	(0.094)	(0.039)	(0.029)	(0.220)		
AID	0.495	0.629	0.177	0.046	−0.014	−0.012	−0.025	0.004	26.662	59.123	−0.202
	(7.506)	(0.295)	(0.183)	(0.120)	(0.073)	(0.083)	(0.038)	(0.026)	(1.887)	(4.182)	(0.069)
					$n = 800$						
Copula Y	1.705	0.596	0.176	0.052	−0.012	−0.011	−0.025	0.002	5.258	7.064	−0.164
	(1.620)	(0.138)	(0.094)	(0.085)	(0.051)	(0.058)	(0.028)	(0.018)	(0.3)	(0.343)	(0.052)
Copula Z	0.706	0.648	0.195	0.054	−0.024	−0.009	−0.031	−0.001	5.260	2.480	0.313
	(1.471)	(0.162)	(0.098)	(0.089)	(0.053)	(0.058)	(0.026)	(0.018)	(0.303)	(0.107)	(0.046)
MF Logit	0.587	0.627	0.172	0.046	−0.012	−0.008	−0.025	0.003	—	—	—
	(12.234)	(0.413)	(0.236)	(0.109)	(0.059)	(0.063)	(0.030)	(0.019)			
Dirichlet	0.560	0.624	0.193	0.041	−0.012	−0.007	−0.023	0.003	4.786	—	—
	(5.204)	(0.176)	(0.119)	(0.088)	(0.059)	(0.066)	(0.027)	(0.021)	(0.154)		
AID	0.416	0.632	0.168	0.051	−0.014	−0.012	−0.025	0.003	26.487	58.691	−0.201
	(4.932)	(0.182)	(0.108)	(0.084)	(0.051)	(0.059)	(0.027)	(0.018)	(1.325)	(2.938)	(0.049)

Note: MLE estimates and (copula misspecification robust) asymptotic standard errors for each estimation procedure. Data are generated from a multivariate Gaussian distribution. "—" implies the parameter is not part of the model.

Table 2.C.6: Estimates and Standard Errors in an Extended Structural Demand Model from a Gaussian Copula with Beta Marginals

Method	α_0	α_1	α_2	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{2,2}$	$\pi_{1,1}$	$\pi_{2,1}$	$\pi_{1,2}$	$\pi_{2,2}$	$\pi_{1,3}$	$\pi_{2,3}$	ϕ_1	ϕ_2	$\psi\|\xi$
							$n = 100$								
Copula Y	0.258	0.917	0.468	0.081	−0.029	−0.037	−0.169	−0.125	0.046	0.015	−0.005	0.000	15.607	18.359	0.244
	(0.482)	(1.458)	(1.202)	(0.190)	(0.103)	(0.113)	(1.028)	(0.684)	(0.307)	(0.177)	(0.035)	(0.021)	(2.231)	(2.736)	(0.138)
Copula Z	0.087	0.595	0.335	0.090	−0.025	−0.032	0.200	−0.122	−0.072	0.030	0.008	−0.003	15.465	2.902	0.603
	(0.521)	(1.387)	(1.182)	(0.193)	(0.106)	(0.117)	(0.992)	(0.675)	(0.308)	(0.191)	(0.039)	(0.026)	(2.183)	(0.358)	(0.115)
MF Logit	0.560	1.007	0.490	0.062	−0.031	−0.032	−0.182	−0.180	0.040	0.038	−0.005	−0.005	—	—	—
	(1.630)	(6.088)	(6.506)	(0.368)	(0.188)	(0.237)	(4.633)	(4.231)	(1.258)	(0.970)	(0.119)	(0.082)			
Dirichlet	0.073	0.943	0.256	0.075	−0.026	−0.033	−0.315	−0.055	0.100	0.013	−0.011	−0.001	10.121	—	—
	(0.681)	(1.509)	(1.285)	(0.245)	(0.146)	(0.160)	(1.050)	(0.752)	(0.303)	(0.196)	(0.034)	(0.022)	(0.982)		
AID	0.292	0.507	0.470	0.084	−0.028	−0.038	0.034	−0.185	0.006	0.034	−0.002	−0.001	69.944	196.805	0.199
	(1.159)	(4.039)	(4.311)	(0.339)	(0.176)	(0.242)	(2.684)	(2.184)	(0.697)	(0.434)	(0.066)	(0.037)	(16.174)	(49.798)	(0.154)
							$n = 200$								
Copula Y	0.278	0.867	0.452	0.075	−0.031	−0.039	−0.125	−0.155	0.029	0.029	−0.003	−0.002	14.984	17.613	0.242
	(0.369)	(1.356)	(1.091)	(0.133)	(0.070)	(0.076)	(0.895)	(0.587)	(0.230)	(0.126)	(0.023)	(0.012)	(1.494)	(1.825)	(0.097)
Copula Z	0.111	0.471	0.333	0.085	−0.031	−0.034	0.152	−0.089	−0.042	0.012	0.004	−0.001	14.855	2.782	0.600
	(0.483)	(1.457)	(1.070)	(0.137)	(0.073)	(0.078)	(0.936)	(0.590)	(0.252)	(0.138)	(0.027)	(0.015)	(1.485)	(0.239)	(0.082)
MF Logit	0.495	0.954	0.548	0.066	−0.032	−0.036	−0.143	−0.230	0.026	0.046	−0.002	−0.004	—	—	—
	(1.372)	(6.017)	(5.394)	(0.275)	(0.153)	(0.177)	(3.857)	(3.050)	(0.880)	(0.603)	(0.072)	(0.042)			
Dirichlet	0.195	0.855	0.436	0.068	−0.028	−0.035	−0.070	−0.103	0.011	0.013	−0.001	−0.001	9.677	—	—
	(0.478)	(1.359)	(1.085)	(0.175)	(0.101)	(0.110)	(0.880)	(0.612)	(0.233)	(0.145)	(0.025)	(0.015)	(0.662)		
AID	0.452	0.565	0.483	0.087	−0.031	−0.037	0.065	−0.249	−0.013	0.065	0.000	−0.006	67.189	189.514	0.195
	(0.910)	(3.641)	(2.855)	(0.199)	(0.103)	(0.108)	(2.311)	(1.577)	(0.561)	(0.327)	(0.051)	(0.027)	(9.618)	(28.424)	(0.098)
							$n = 400$								
Copula Y	0.350	0.521	0.570	0.078	−0.027	−0.043	0.082	−0.222	−0.017	0.040	0.001	−0.003	14.682	17.205	0.240
	(0.223)	(1.338)	(0.932)	(0.093)	(0.048)	(0.051)	(0.839)	(0.488)	(0.192)	(0.093)	(0.016)	(0.007)	(1.020)	(1.254)	(0.069)
Copula Z	0.484	0.376	0.491	0.088	−0.028	−0.036	0.205	−0.146	−0.049	0.021	0.004	−0.001	14.510	2.761	0.597
	(0.297)	(1.203)	(0.859)	(0.094)	(0.049)	(0.053)	(0.800)	(0.490)	(0.202)	(0.113)	(0.020)	(0.012)	(1.010)	(0.167)	(0.057)
MF Logit	0.775	0.631	0.574	0.069	−0.027	−0.035	0.044	−0.253	−0.013	0.049	0.001	−0.003	—	—	—
	(1.203)	(7.511)	(5.775)	(0.312)	(0.140)	(0.130)	(4.637)	(3.367)	(0.978)	(0.663)	(0.074)	(0.045)			
Dirichlet	0.239	0.678	0.510	0.065	−0.025	−0.035	−0.053	−0.201	0.022	0.040	−0.003	−0.003	9.446	—	—
	(0.393)	(1.380)	(1.073)	(0.121)	(0.070)	(0.077)	(0.850)	(0.588)	(0.204)	(0.127)	(0.019)	(0.011)	(0.457)		
AID	0.634	0.115	0.550	0.091	−0.028	−0.042	0.344	−0.246	−0.078	0.052	0.006	−0.004	65.817	184.504	0.192
	(0.827)	(4.389)	(3.138)	(0.179)	(0.083)	(0.088)	(2.769)	(1.778)	(0.628)	(0.347)	(0.051)	(0.024)	(6.467)	(19.972)	(0.069)
							$n = 800$								
Copula Y	0.447	0.738	0.542	0.074	−0.027	−0.041	−0.020	−0.221	0.001	0.045	0.000	−0.004	14.550	17.057	0.238
	(0.151)	(1.327)	(0.824)	(0.066)	(0.033)	(0.035)	(0.818)	(0.425)	(0.177)	(0.079)	(0.014)	(0.006)	(0.730)	(0.899)	(0.049)
Copula Z	0.601	0.354	0.498	0.086	−0.028	−0.035	0.211	−0.188	−0.050	0.034	0.004	−0.002	14.365	2.780	0.595
	(0.184)	(1.216)	(0.734)	(0.065)	(0.033)	(0.036)	(0.777)	(0.415)	(0.182)	(0.088)	(0.017)	(0.008)	(0.702)	(0.119)	(0.041)
MF Logit	0.732	0.818	0.594	0.068	−0.029	−0.034	−0.066	−0.268	0.009	0.052	−0.001	−0.004	—	—	—
	(0.880)	(5.469)	(4.022)	(0.177)	(0.083)	(0.089)	(3.343)	(2.256)	(0.689)	(0.425)	(0.048)	(0.027)			
Dirichlet	0.689	0.341	0.678	0.078	−0.024	−0.038	0.220	−0.273	−0.052	0.048	0.004	−0.003	9.355	—	—
	(0.253)	(1.264)	(0.870)	(0.085)	(0.048)	(0.053)	(0.814)	(0.510)	(0.190)	(0.110)	(0.017)	(0.009)	(0.320)		
AID	0.617	0.010	0.507	0.093	−0.024	−0.089	0.410	−0.227	−0.090	0.043	0.007	−0.002	65.239	183.159	0.190
	(0.623)	(8.194)	(4.365)	(0.225)	(0.081)	(0.059)	(4.951)	(2.447)	(1.042)	(0.469)	(0.077)	(0.031)	(6.737)	(19.920)	(0.051)

Note: MLE estimates and (copula misspecification robust) asymptotic standard errors for each estimation procedure. Data are generated from a Gaussian copula with beta marginals. "—" implies the parameter is not part of the model.

Table 2.C.7: Estimates and Standard Errors in an Extended Structural Demand Model from a Gaussian Distribution

Method	α_0	α_1	α_2	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{2,2}$	$\pi_{1,1}$	$\pi_{2,1}$	$\pi_{1,2}$	$\pi_{2,2}$	$\pi_{1,3}$	$\pi_{2,3}$	ϕ_1	ϕ_2	$\psi\|\xi$
							$n = 100$								
Copula Y	−0.282	0.605	0.285	0.046	−0.004	−0.007	0.027	−0.158	−0.011	0.056	0.001	−0.007	5.962	7.987	−0.213
	(2.940)	(13.726)	(15.480)	(1.999)	(0.809)	(0.717)	(7.401)	(6.045)	(1.592)	(0.968)	(0.133)	(0.071)	(1.735)	(4.024)	(0.212)
Copula Z	−0.178	0.554	0.314	0.081	−0.012	0.005	0.074	−0.017	−0.036	−0.006	0.006	0.001	6.315	2.686	0.334
	(1.417)	(5.694)	(4.673)	(0.765)	(0.472)	(0.428)	(3.363)	(2.259)	(0.862)	(0.588)	(0.090)	(0.068)	(1.563)	(0.412)	(0.268)
MF Logit	0.493	1.026	0.146	0.042	−0.010	−0.001	−0.321	−0.007	0.082	0.008	−0.009	−0.001	—	—	—
	(3.321)	(5.305)	(5.420)	(0.648)	(0.379)	(0.423)	(4.379)	(3.785)	(1.153)	(1.009)	(0.129)	(0.100)			
Dirichlet	0.051	0.786	0.157	0.056	−0.013	0.002	−0.332	0.002	0.117	0.000	−0.013	0.001	5.447	—	—
	(0.965)	(1.678)	(1.588)	(0.344)	(0.225)	(0.256)	(1.275)	(0.991)	(0.417)	(0.295)	(0.052)	(0.037)	(0.505)		
AID	0.103	0.561	0.096	0.059	−0.010	−0.001	−0.077	−0.053	0.032	0.038	−0.004	−0.006	29.410	67.693	−0.240
	(0.835)	(1.540)	(1.421)	(0.358)	(0.204)	(0.226)	(1.246)	(0.937)	(0.414)	(0.283)	(0.052)	(0.038)	(4.253)	(9.762)	(0.138)
							$n = 200$								
Copula Y	0.083	0.636	0.163	0.065	−0.013	−0.005	−0.098	0.108	0.038	−0.039	−0.005	0.004	5.719	7.620	−0.208
	(5.512)	(6.114)	(7.698)	(0.603)	(0.229)	(0.330)	(7.691)	(2.836)	(2.582)	(0.716)	(0.259)	(0.054)	(0.985)	(1.074)	(0.145)
Copula Z	−0.028	0.769	0.276	0.075	−0.005	0.002	0.028	−0.050	−0.029	0.007	0.004	0.000	6.212	2.697	0.337
	(1.408)	(7.925)	(7.372)	(0.626)	(0.419)	(0.379)	(5.331)	(3.719)	(1.338)	(0.745)	(0.129)	(0.067)	(1.181)	(0.357)	(0.151)
MF Logit	0.506	0.906	0.283	0.051	−0.006	−0.004	−0.201	−0.073	0.036	0.017	−0.002	−0.001	—	—	—
	(3.256)	(6.192)	(8.674)	(0.532)	(0.359)	(0.366)	(3.233)	(4.730)	(1.093)	(0.847)	(0.121)	(0.051)			
Dirichlet	0.068	0.346	0.221	0.060	−0.005	−0.002	0.162	0.027	−0.038	−0.009	0.003	0.000	5.227	—	—
	(0.829)	(1.593)	(1.382)	(0.244)	(0.158)	(0.181)	(1.071)	(0.851)	(0.324)	(0.243)	(0.043)	(0.031)	(0.344)		
AID	0.254	0.674	0.224	0.062	−0.014	0.002	−0.135	−0.019	0.046	0.001	−0.006	0.001	28.330	64.736	−0.239
	(0.663)	(1.507)	(1.377)	(0.228)	(0.136)	(0.158)	(1.091)	(0.819)	(0.322)	(0.212)	(0.036)	(0.024)	(2.862)	(6.538)	(0.097)
							$n = 400$								
Copula Y	−0.045	0.308	0.359	0.058	−0.010	−0.005	0.201	−0.162	−0.055	0.053	0.005	−0.007	5.585	7.444	−0.203
	(0.960)	(4.143)	(3.467)	(0.258)	(0.144)	(0.159)	(2.408)	(1.843)	(0.554)	(0.373)	(0.054)	(0.033)	(0.549)	(0.713)	(0.077)
Copula Z	0.200	0.507	0.359	0.062	−0.013	−0.002	0.004	−0.018	0.004	−0.013	−0.001	0.002	5.722	2.329	0.318
	(1.372)	(3.604)	(3.300)	(0.244)	(0.158)	(0.168)	(2.154)	(1.653)	(0.689)	(0.371)	(0.097)	(0.048)	(0.617)	(0.163)	(0.098)
MF Logit	0.631	0.729	0.427	0.057	−0.011	−0.001	−0.063	−0.140	0.007	0.027	0.000	−0.002	—	—	—
	(2.043)	(8.389)	(4.864)	(0.307)	(0.160)	(0.220)	(5.407)	(2.536)	(1.187)	(0.610)	(0.091)	(0.058)			
Dirichlet	0.282	0.821	0.298	0.047	−0.007	0.001	−0.144	−0.020	0.036	0.001	−0.005	0.000	5.107	—	—
	(0.634)	(1.487)	(1.294)	(0.168)	(0.107)	(0.119)	(0.992)	(0.776)	(0.270)	(0.190)	(0.031)	(0.022)	(0.236)		
AID	0.245	−0.249	0.163	0.061	−0.009	−0.001	0.444	−0.020	−0.092	0.007	0.007	0.000	27.774	63.597	−0.236
	(0.529)	(1.391)	(1.159)	(0.164)	(0.095)	(0.105)	(0.900)	(0.645)	(0.236)	(0.148)	(0.024)	(0.015)	(1.975)	(4.522)	(0.069)
							$n = 800$								
Copula Y	0.427	0.174	0.345	0.073	−0.014	−0.002	0.206	−0.060	−0.037	0.005	0.002	0.000	5.528	7.372	−0.202
	(0.589)	(4.091)	(2.912)	(0.170)	(0.094)	(0.102)	(2.422)	(1.551)	(0.536)	(0.297)	(0.048)	(0.023)	(0.376)	(0.467)	(0.053)
Copula Z	0.171	0.452	0.391	0.080	−0.001	0.012	0.083	−0.081	−0.019	0.005	0.001	0.001	6.238	2.693	0.344
	(0.603)	(17.191)	(10.617)	(0.456)	(0.217)	(0.115)	(7.727)	(4.322)	(1.264)	(0.636)	(0.081)	(0.038)	(0.959)	(0.153)	(0.168)
MF Logit	0.784	0.967	0.517	0.051	−0.013	−0.002	−0.250	−0.200	0.053	0.039	−0.004	−0.002	—	—	—
	(1.402)	(5.672)	(6.069)	(0.243)	(0.105)	(0.133)	(3.435)	(3.474)	(0.721)	(0.670)	(0.056)	(0.045)			
Dirichlet	−0.014	0.078	0.561	0.057	−0.015	0.001	0.178	−0.212	−0.018	0.043	0.000	−0.003	5.058	—	—
	(0.320)	(1.442)	(1.145)	(0.116)	(0.076)	(0.083)	(0.849)	(0.611)	(0.191)	(0.128)	(0.016)	(0.011)	(0.165)		
AID	0.499	0.547	0.517	0.059	−0.017	−0.003	−0.043	−0.204	0.024	0.042	−0.004	−0.003	27.534	63.084	−0.236
	(0.391)	(1.439)	(1.055)	(0.110)	(0.065)	(0.072)	(0.918)	(0.592)	(0.216)	(0.125)	(0.020)	(0.011)	(1.390)	(3.183)	(0.049)

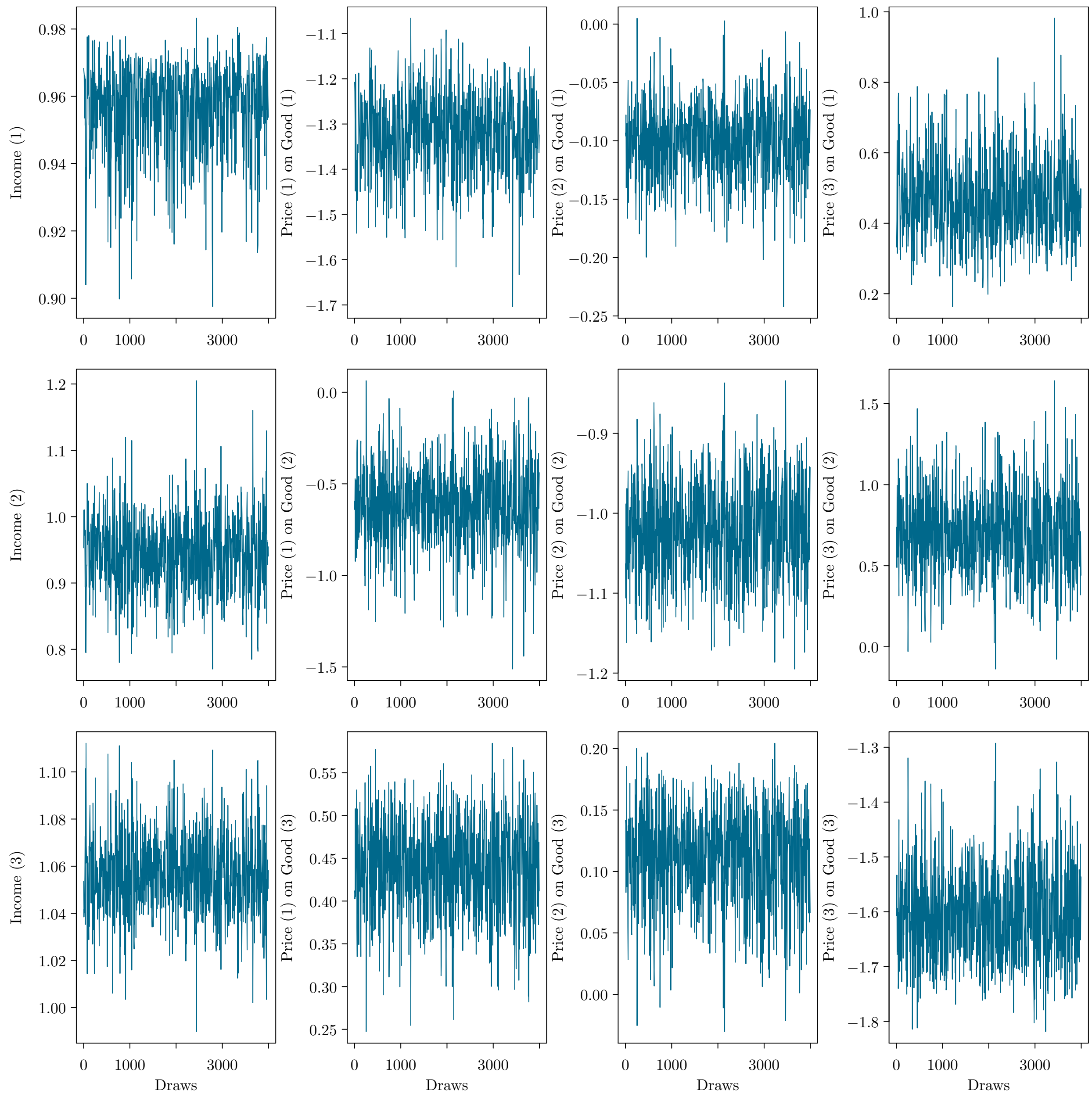
Note: MLE estimates and (copula misspecification robust) asymptotic standard errors for each estimation procedure. Data are generated from a multivariate Gaussian distribution. "—" implies the parameter is not part of the model.

Table 2.C.8: Bayesian Point Estimates and Inference for an Extended Reduced Form Model

Variable	Outcome 1	Outcome 2
Constant	−2.002	−2.033
	(0.041)	(0.043)
x_1	0.841	0.848
	(0.042)	(0.043)
x_2	−0.846	−0.828
	(0.041)	(0.042)
x_3	0.869	0.871
	(0.042)	(0.043)
x_4	−0.867	−0.892
	(0.042)	(0.042)
x_5	0.849	0.861
	(0.042)	(0.043)
x_6	−0.023	−0.026
	(0.030)	(0.031)
x_7	−0.020	0.023
	(0.030)	(0.031)
x_8	−0.015	−0.006
	(0.029)	(0.030)
x_9	−0.026	−0.001
	(0.031)	(0.031)
x_{10}	−0.018	−0.023
	(0.030)	(0.030)

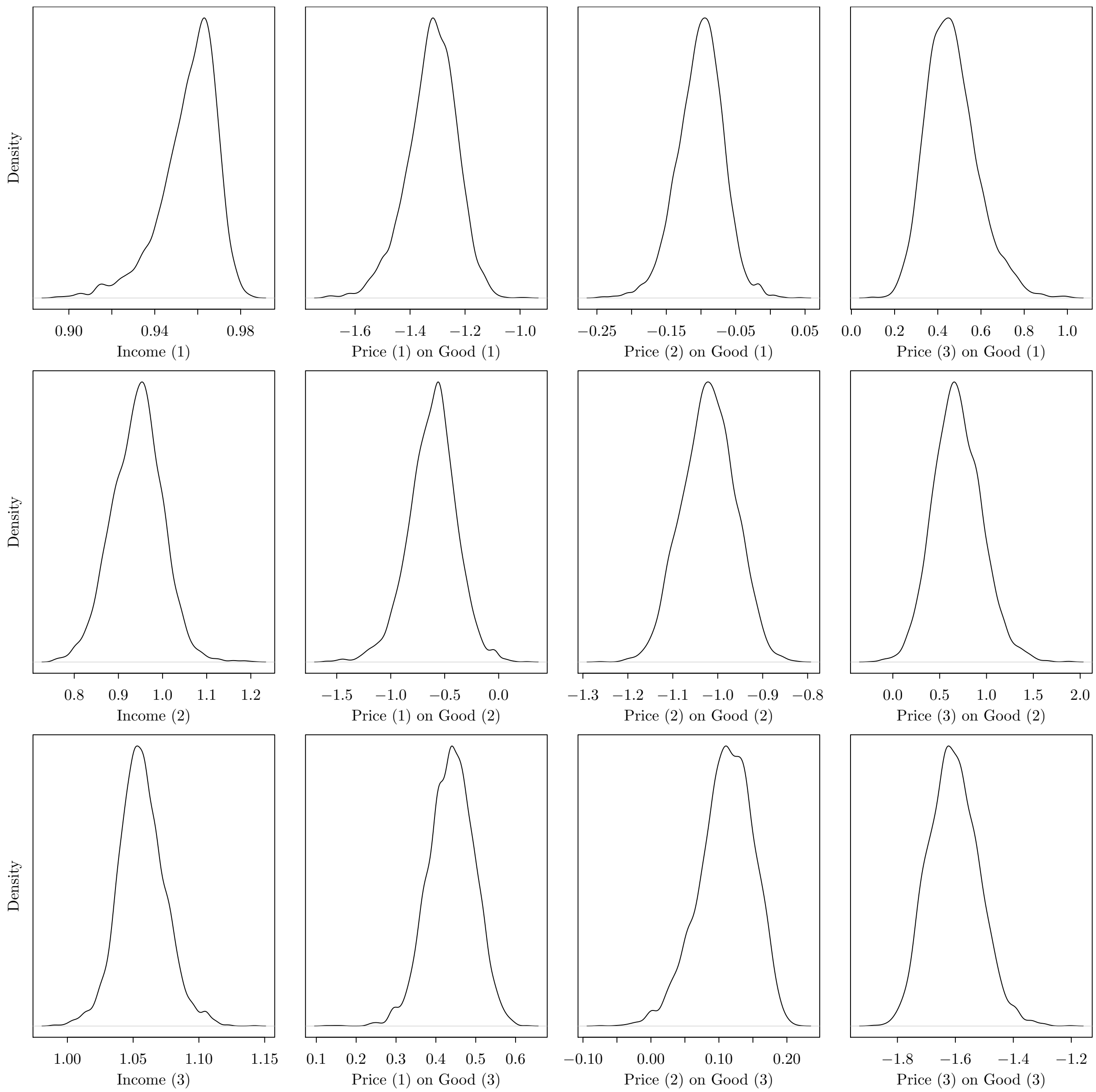
Note: Bayesian estimates from a Gaussian copula with beta marginals specification. Entries denote coefficient of the associated variable in each of the outcome equations. Standard errors (standard deviation of the chains) in parentheses.

Figure 2.9: Trace Plot of Elasticity Chains in an Extended Bayesian AID System



Note: Results for the data set on married couples with one child. Combination of 5 chains with 800 draws each for a total of 4,000 draws.

Figure 2.10: Density Plot of Elasticity Chains in an Extended Bayesian AID System



Note: Results for the data set on married couples with one child. Combination of 5 chains with 800 draws each for a total of 4,000 draws.

Chapter 3

Multivariate Fractional Panel Data Methods

While there have been many developments in creating modeling strategies for multivariate fractional outcomes 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 ways of modeling multivariate fractions in a panel data setting. That is, strategies that simultaneously take into account the inherent nonlinearity in the partial effects from covariates, unobserved heterogeneity that is potentially correlated to these covariates, and that impose the unit-sum restriction present across the multivariate outcomes. Additionally, we would expect that such a framework would allow to control for further endogeneity issues that are not captured by unobserved heterogeneity and also allow for structural zeros in the data.[1]

The main contribution of the chapter is then to expand the available toolkit for modeling multivariate fractional outcomes using panel data in applied microeconomic settings. Recognizing that different applications are conceived with different objectives in mind, the chapter introduces a wide range of methods that are suitable in a variety of 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

[1] For example, in the demand estimation setting by allowing some households to spend none of their income on a particular good.

each framework.

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 draws on the statistical literature on generalized (non)linear mixed models for multivariate responses (for a review, see for example Davidian and Giltinan, 1995). This method will be particularly useful when an application requires consistent estimation of the parameters, not just the signs or average partial effects. Of course, given consistent estimation of the parameters, these other quantities can be consistently estimated. It also has the potential of being efficient in comparison to the other methods introduced in the chapter. While many available 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. For example, transformation methods that take the multivariate fractions to an unbounded space before imposing a distributional assumption, such as the additive log-ratio (Aitchison and Shen, 1980), centered log-ratio (Aitchison, 1983), centered log-ratio (Egozcue et al., 2003), or α (Tsagris et al., 2011) transformations require strong independence assumptions to recover the parameters of a conditional mean model defined directly on the share components (Papke and Wooldridge, 1996). Other distributions might allow for a regression structure but will generally not be robust to misspecification (Hijazi and Jernigan, 2009; Scealy and Welsh, 2011). The maximum likelihood methods considered in this chapter will allow for direct specification of a conditional mean and at least some degree of robustness to distributional misspecification, if not full robustness.

The second method extends Papke and Wooldridge (2008) to a multivariate fractional setting by using pooled multivariate nonlinear least squares with a probit link. While this approach might be potentially misspecified and thus not consistently estimate the parameters of the conditional mean (up to a scale factor), it provides the best mean squared error approximation to these quantities that is afforded by the probit link. Furthermore, if these approximations are believed to be accurate (and numerical simulation results in Section 3.2 show that this tends to be the case), this approach would allow for the identification and estimation of average partial effects, the inclusion of continuous endogenous covariates, and inference can be made fully robust to the potential misspecification of the conditional mean. Additionally, this method is not impeded by zeros in the underlying multivariate fractions and can be scaled to handle a large amount of shares without much additional computational burden.

I then discuss a latent dependent variable formulation that accounts for censoring, given by structural zeros in the multivariate fractions. Using the simple transformation in Wales and Woodland (1983), I extend the Bayesian approach of Kasteridis et al. (2011) to account for panel data and correlated random effects using a data augmentation algorithm that accounts for censoring (Albert and Chib, 1993). Accounting for unobserved heterogeneity in this method is then also a multivariate generalization to Loudermilk (2007). 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. Additionally, it is important to note that the Bayesian estimator allows for potential endogeneity that is not captured by the unobserved heterogeneity, similar to the probit method (Ramírez-Hassan, 2021). This approach also directly accounts for the presence of zeros in the multivariate fractions. Other methods that allow for zeros usually take these as possible detection errors, and thus create imputation methods in some optimal way to minimize the ad hoc nature of this operation (Fry et al., 2000; Martín-Fernández et al., 2003). Furthermore, some transformation and likelihood-based approaches can also deal with zeros, but they can suffer from similar caveats as those mentioned before (Stewart and Field, 2011; Tsagris and Stewart, 2018).

The remaining of the chapter proceeds as follows. Section 3.1 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 3.2 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 3.3 presents the concluding remarks.

3.1 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. For each share Y_j, I 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 summarizes the type of panel data structures that are within the scope of this paper and which arise frequently in applied microeconomics.

Assumption 3.1 (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 3.1 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.

3.1.1 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 3.2 (Conditional mean)**.** For each $i = 1, \ldots, n$, $t = 1, \ldots, T_i$, and $j = 1, \ldots, d$,

$$\mathrm{E}[Y_{itj} | \boldsymbol{X}_{it}, \boldsymbol{c}_i] = m_j(\boldsymbol{X}_{it}\boldsymbol{\beta}_0 + \boldsymbol{c}_i), \tag{3.1.1}$$

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 (3.1.1) 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 (3.1.1) across outcomes and time. These are given by

$$\mathrm{E}[\boldsymbol{Y}_{it}|\boldsymbol{X}_{it}, \boldsymbol{c}_i] = \boldsymbol{m}(\boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{c}_i) \tag{3.1.2}$$

and

$$\mathrm{E}[\boldsymbol{Y}_i|\boldsymbol{X}_i, \boldsymbol{c}_i] = \boldsymbol{m}_{T_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{c}_i)\,, \tag{3.1.3}$$

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 $\boldsymbol{X}_i = [\boldsymbol{X}_{i1}' \cdots \boldsymbol{X}_{iT_i}']'$ is a $DT_i \times K$ matrix.

As noted by Papke and Wooldridge (2008), assumptions 3.1 and 3.2 on their own are not enough to carry out estimation of the conditional mean parameters . To this end, I make two additional assumptions.

Assumption 3.3 (Strict exogeneity)**.** For all $i = 1, \dots, n$, and $j = 1, \dots, d$,

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

Assumption 3.4 (Mundlak device)**.** For all $i = 1, \dots, n$,

$$\boldsymbol{c}_i|\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{iT_i} \sim \mathcal{N}(\bar{\boldsymbol{X}}_i\boldsymbol{\xi}, \boldsymbol{\Gamma})\,, \tag{3.1.4}$$

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 3.3 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 3.4 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 (3.1.4) 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 $\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{iT_i}$ 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 (3.1.4) 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 (3.1.4), we can write $\boldsymbol{c}_i = \bar{\boldsymbol{X}}_i\boldsymbol{\xi} + \boldsymbol{b}_i$, where $\boldsymbol{b}_i|\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{iT_i} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma})$. Replacing this into (3.1.2) and using Assumption 3.3 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 $\tilde{\boldsymbol{X}}_{it} = [\boldsymbol{X}_{it}\, \bar{\boldsymbol{X}}_i]$ and $\boldsymbol{\alpha} = (\boldsymbol{\beta}', \boldsymbol{\xi}')'$, we can then find

$$\mathrm{E}[\boldsymbol{Y}_{it}|\boldsymbol{X}_i, \boldsymbol{c}_i] = \boldsymbol{m}(\tilde{\boldsymbol{X}}_{it}\boldsymbol{\alpha} + \boldsymbol{b}_i)\,, \tag{3.1.5}$$

with $\boldsymbol{b}_i$ independent of $\tilde{\boldsymbol{X}}_{it}$. This is of the same form as (3.1.2) but with $\boldsymbol{b}_i$ representing unobserved heterogeneity that is uncorrelated from the covariates. For notational simplicity, the remaining of the paper assumes that (3.1.2) (and thus 3.1.3) 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 (3.1.5) 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 (3.1.1), $\bar{\boldsymbol{X}}_i$ needs to be integrated out for each $t = 1, \ldots, T_i$ (Papke and Wooldridge, 2008).

Armed with Assumptions 3.1 through 3.4, 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 (3.1.2). As the random effects $\boldsymbol{c}_i$ (or $\boldsymbol{b}_i$ after the transformation in 3.1.5) 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

$$\ell_i^{(\mathrm{ind})}(\boldsymbol{\beta}, \boldsymbol{\Gamma}) = \log \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T_i} F(\boldsymbol{Y}_{it}|\boldsymbol{X}_{it}, \boldsymbol{c}_i; \boldsymbol{\beta}) \right] \phi_D(\boldsymbol{c}_i; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma})\, \mathrm{d}\boldsymbol{c}_i\ , \tag{3.1.6}$$

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

$$\ell_i^{(\mathrm{pool})}(\boldsymbol{\beta}, \boldsymbol{\Gamma}) = \sum_{t=1}^{T_i} \log \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F(\boldsymbol{Y}_{it}|\boldsymbol{X}_{it}, \boldsymbol{c}_i; \boldsymbol{\beta}) \phi_D(\boldsymbol{c}_i; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma})\, \mathrm{d}\boldsymbol{c}_i\ . \tag{3.1.7}$$

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 (3.1.6) or (3.1.7) is given by

$$\widehat{\boldsymbol{\theta}}_l \equiv \arg\max_{\boldsymbol{\theta}} \frac{1}{n}\sum_{i=1}^{n} \ell_i^{(l)}(\boldsymbol{\theta}), l \in \{\text{ind}, \text{pool}\}\,. \tag{3.1.8}$$

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 3.2 (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

$$\text{Asy. Var}(\sqrt{n}(\widehat{\boldsymbol{\theta}}_l - \boldsymbol{\theta}_l^*)) = A_l^{-1} B_l A_l^{-1}\,, \tag{3.1.9}$$

where $A_l = \text{E}_H[\partial^2 \ell_i^{(l)}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}']$ is the Hessian matrix of the log-likelihood contributions, $B_l = \text{E}_H[\partial \ell_i^{(l)}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} \cdot \partial \ell_i^{(l)}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}']$ 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

$$\widehat{A}_l = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \ell_i^{(l)}(\widehat{\boldsymbol{\theta}}_l)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \quad \text{and} \quad \widehat{B}_l = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \ell_i^{(l)}(\widehat{\boldsymbol{\theta}}_l)}{\partial\boldsymbol{\theta}} \cdot \frac{\partial \ell_i^{(l)}(\widehat{\boldsymbol{\theta}}_l)}{\partial\boldsymbol{\theta}'}\,, \tag{3.1.10}$$

to estimate the asymptotic variance in (3.1.9). 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 (3.1.6) or (3.1.7). For adaptive (Liu and Pierce, 1994) or nonadaptive (Jäckel, 2005) quadrature, Appendix 3.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; (3.1.6) requires n integrals to be computed, while (3.1.7) 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 3.2. That is, these estimators take $\boldsymbol{m}(\cdot)$ as

$$\boldsymbol{m}(\boldsymbol{X}_{it}'\boldsymbol{\beta} + \boldsymbol{c}_i) = \begin{cases} \frac{\exp\left(\boldsymbol{x}_{itj}'\boldsymbol{\beta}_j + c_{ij}\right)}{1+\sum_{p=1}^{D}\exp\left(\boldsymbol{x}_{itp}'\boldsymbol{\beta}_p + c_{ip}\right)} & \text{for } j = 1,\ldots,D\,, \\ \frac{1}{1+\sum_{p=1}^{D}\exp\left(\boldsymbol{x}_{itp}'\boldsymbol{\beta}_p + c_{ip}\right)} & \text{for } j = d\,. \end{cases} \tag{3.1.11}$$

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 (3.1.6) or (3.1.7) 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 (3.1.8) computing the integrals as in Appendix 3.A.

3. As the multinomial likelihood is inherently misspecified, use the fully robust estimators given in (3.1.10).

Appendix 3.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 (3.1.1), 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 (3.1.11), such as beta distributions, and copula $G(\cdot;\boldsymbol{\psi})$, for example a Gaussian copula. Then, use

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

   in either (3.1.6) or (3.1.7). 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 3.A.

2. Estimate $(\boldsymbol{\theta}',\boldsymbol{\phi}',\boldsymbol{\psi}')$ as in (3.1.8).

3. If the copula is potentially correctly specified, use $\widehat{A}_l^{-1}$ as the estimator for the asymptotic variance in (3.1.9). Otherwise, use the fully robust (3.1.10).

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 3.2.

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 (3.1.11). 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.

3.1.2 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

$$\mathrm{E}[\boldsymbol{Y}_{it}|\boldsymbol{X}_{it}] = \begin{bmatrix} \Phi\left(\boldsymbol{x}'_{it1}\left(\frac{\boldsymbol{\beta}_1}{\sqrt{1+\gamma_1^2}}\right)\right) \\ \vdots \\ \Phi\left(\boldsymbol{x}'_{it1}\left(\frac{\boldsymbol{\beta}_D}{\sqrt{1+\gamma_D^2}}\right)\right) \end{bmatrix} = \begin{bmatrix} \Phi(\boldsymbol{x}'_{it1}\boldsymbol{\beta}_{1c}) \\ \vdots \\ \Phi(\boldsymbol{x}'_{itD}\boldsymbol{\beta}_{Dc}) \end{bmatrix}, \tag{3.1.12}$$

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

$$\mathrm{E}_{\bar{x}_{ij}}[\Phi(\boldsymbol{x}'_{itj}\boldsymbol{\beta}_{cj} + \bar{\boldsymbol{x}}'_{ij}\boldsymbol{\xi}_{cj})] \tag{3.1.13}$$

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 3.2. 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 (3.1.12) 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

$$q_i(\boldsymbol{\alpha}_c) \equiv q(\boldsymbol{Y}_i, \boldsymbol{X}_i; \boldsymbol{\alpha}_c) = \frac{1}{2}[\boldsymbol{Y}_i - \boldsymbol{m}_{T_i}(\tilde{\boldsymbol{X}}_i\boldsymbol{\alpha}_c)]'[\boldsymbol{Y}_i - \boldsymbol{m}_{T_i}(\tilde{\boldsymbol{X}}_i\boldsymbol{\alpha}_c)] \tag{3.1.14}$$

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

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_c &\equiv \arg\min_{\boldsymbol{\alpha_c}} \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}[\boldsymbol{Y}_{it} - \boldsymbol{m}(\tilde{\boldsymbol{X}}_{it}\boldsymbol{\alpha}_c)]'[\boldsymbol{Y}_{it} - \boldsymbol{m}(\tilde{\boldsymbol{X}}_{it}\boldsymbol{\alpha}_c)] \\ &= \arg\min_{\boldsymbol{\alpha_c}} \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\sum_{j=1}^{D}[y_{itj} - \Phi(\tilde{\boldsymbol{x}}_{itj}'\boldsymbol{\alpha}_{cj})]^2 \end{aligned} \tag{3.1.15}$$

where the definitions of $\tilde{\boldsymbol{x}}$ and $\boldsymbol{\alpha}$ come from (3.1.5). 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

$$\mathrm{Asy.\,Var}(\sqrt{n}(\widehat{\boldsymbol{\alpha}}_c - \boldsymbol{\alpha}_c^*)) = A^{-1}BA^{-1}\,, \tag{3.1.16}$$

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 (3.1.16) follows as

$$\widehat{A} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 q_i(\widehat{\boldsymbol{\alpha}}_c)}{\partial \boldsymbol{\alpha}_c \partial \boldsymbol{\alpha}_c'} \quad \text{and} \quad \widehat{B} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial q_i(\widehat{\boldsymbol{\alpha}}_c)}{\partial \boldsymbol{\alpha}_c} \cdot \frac{\partial q_i(\widehat{\boldsymbol{\alpha}}_c)}{\partial \boldsymbol{\alpha}_c}' . \tag{3.1.17}$$

Given that the probit link is a simple special case, formulas for both the scores and Hessian are available; these are given in Appendix 3.B. This procedure is summarized as follows.

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

1. Estimate $\boldsymbol{\alpha}_c$ from (3.1.15) 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 (3.1.17) using the formulas in Appendix 3.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 (3.1.14). As the estimator defined in (3.1.15) 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 (3.1.17). 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).

3.1.3 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 3.1, 3.3 and 3.4.[2] 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 (3.1.3) 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 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

[2]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.

variables supported on $\mathcal{S}^d$ by using

$$y_{itj} = \frac{\max\{y^*_{itj}, 0\}}{1 - \sum_{(t,p)\in E_i} y^*_{itp}}\,, \tag{3.1.18}$$

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 3.4 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:

$$\begin{aligned} \boldsymbol{\beta} &\sim \mathcal{N}(\boldsymbol{\beta}_0, \boldsymbol{B}_0)\,, \\ \boldsymbol{\Gamma} &\sim \mathcal{IW}(\nu_\Gamma, \boldsymbol{R}_\Gamma)\,, \\ \boldsymbol{\Sigma} &\sim \mathcal{IW}(\nu_\Sigma, \boldsymbol{R}_\Sigma)\,. \end{aligned} \tag{3.1.19}$$

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

$$\begin{aligned}\pi(\boldsymbol{\beta}, \boldsymbol{Y}^*, \boldsymbol{c}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{W}) \propto \prod_{i=1}^{n} \Bigg\{ \Bigg[& \prod_{t=1}^{T_i}\prod_{j=1}^{D} I(y_{itj}=0) I(y_{itj}^* \leq 0) \\ & + I(y_{itj} > 0) I\left(y_{itj} = \frac{y_{itj}^*}{1 - \sum_{(t,p)\in E_i} y_{itp}^*} \right) \Bigg] \\ & \times \phi_{DT_i}(\boldsymbol{Y}_i^*; \boldsymbol{X}_i'\boldsymbol{\beta} + \boldsymbol{W}_i \boldsymbol{c}_i; \lambda_i^{-1}(\boldsymbol{\Omega}_i \otimes \boldsymbol{\Sigma})) \Bigg\} \\ & \times \pi(\boldsymbol{\beta})\pi(\boldsymbol{c})\pi(\boldsymbol{\lambda})\pi(\boldsymbol{\Gamma})\pi(\boldsymbol{\Sigma})\,. \end{aligned} \tag{3.1.20}$$

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

$$y_{itj}^* = y_{itj}\left(1 - \sum_{(t,p)\in E_i} y_{itp}^*\right). \tag{3.1.21}$$

From (3.1.20), 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 $\mu_{itj|-(tj)} = \mathrm{E}\left[y_{itj}^* | \boldsymbol{Y}_{i,-(tj)}^{*(s-1)}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma}\right]$, variance $\sigma^2_{itj|-(tj)} = \mathrm{Var}(y_{itj}^*|\boldsymbol{Y}_{i,-(tj)}^{*(s-1)}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\Sigma})$, 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 (3.1.21).

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

$$\bar{\boldsymbol{B}}^{(s)} = \left(\boldsymbol{B}_0^{-1} + \sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{V}_i^{-1(s-1)} \boldsymbol{X}_i \right)^{-1},$$

$$\bar{\boldsymbol{\beta}}^{(s)} = \bar{\boldsymbol{B}}^{(s)} \left(\boldsymbol{\beta}_0^{-1} + \sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{V}_i^{-1(s-1)} \boldsymbol{Y}_i^{*(s)} \right)^{-1},$$

$$\boldsymbol{V}_i^{(s-1)} = (I_{T_i} \otimes \boldsymbol{\Sigma}^{(s-1)}) + \boldsymbol{W}_i \boldsymbol{\Gamma}^{(s-1)} \boldsymbol{W}_i' .$$

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

$$\bar{\boldsymbol{\Gamma}}_i = \left[\boldsymbol{\Gamma}^{-1(s-1)} + \boldsymbol{W}_i' (I_{T_i} \otimes \boldsymbol{\Sigma}^{-1(s-1)}) \boldsymbol{W}_i \right]^{-1},$$

$$\bar{\boldsymbol{c}}_i = \bar{\boldsymbol{\Gamma}}_i \boldsymbol{W}_i' (I_{T_i} \otimes \boldsymbol{\Sigma}^{-1(s-1)})(\boldsymbol{Y}_i^{*(s)} - \boldsymbol{X}_i \boldsymbol{\beta}^{(s)}) .$$

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

$$\bar{\nu}_\Gamma = \nu_\Gamma + n ,$$

$$\bar{\boldsymbol{R}}_\Gamma^{(s)} = \boldsymbol{R}_\Gamma + \sum_{i=1}^{n} \boldsymbol{c}_i^{(s)} \boldsymbol{c}_i^{\prime(s)} .$$

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

$$\bar{\nu}_\Sigma = \nu_\Sigma + \sum_{i=1}^{n} T_i ,$$

$$\bar{\boldsymbol{R}}_\Gamma^{(s)} = \boldsymbol{R}_\Gamma + \sum_{i=1}^{n} \boldsymbol{e}_i^{\prime(s)} \boldsymbol{e}_i^{(s)} ,$$

and $\boldsymbol{\epsilon}_i^{(s)}$ is a $T_i \times D$ matrix such that $\text{vec}(\boldsymbol{e}_i^{\prime(s)}) = \boldsymbol{Y}_i^{*(s)} - \boldsymbol{X}_i \boldsymbol{\beta}^{(s)} - \boldsymbol{W}_i \boldsymbol{c}_i^{(s)}$; 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 3.4 as a prior distribution, the same could be done for the maximum likelihood approach in Section 3.1.1. 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 (3.1.19) along with the likelihoods given in (3.1.6) or (3.1.7). 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.

3.2 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.

3.2.1 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

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

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 $1,\ldots,D$, 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 $i = 1,\ldots,n$ 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 (3.1.11) 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 (3.1.6) 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 3.1: 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 3.1 in the forms of root mean squared errors (RMSE) from the true parameters. The analysis focuses on the conditional mean coefficients $\boldsymbol{\beta}$.[3] 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 3.2. 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 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 3.1.9). Thus, as expected, the fact that the copula

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

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 3.2. 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 3.2: 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.

3.2.2 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 3.3 and the coefficients with associated standard errors in Table 3.4.

Table 3.3: 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.

Table 3.4: 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.

3.2.3 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:

$$y^*_{itj} = \boldsymbol{x}'_{itj}\boldsymbol{\beta}_j + c_{ij} + \varepsilon_{itj}\,. \tag{3.2.1}$$

This creates the need to adjust the values previously used for generation, as the underlying latent variable model (3.2.1) tends to yield too many zeros if the linear index induces a lot of variance on $\boldsymbol{Y}^*$. 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 (3.1.18) 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.

Table 3.5: Coefficients from a Bayesian Latent Dependent Variable Model

Method	$\beta_{1,0}$	$\beta_{1,1}$	$\beta_{1,2}$	$\beta_{2,0}$	$\beta_{2,1}$	$\beta_{2,2}$
			$nT = 200$			
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.

For estimation purposes, given the conjugate priors outlined for the Bayesian estimator in Section 3.1.3, all that remains is to specify the hyperparameters of these distributions. I choose standard uninformative priors for the coefficients by setting $\boldsymbol{\beta}_0 = \mathbf{0}_{K\times 1}$, $\boldsymbol{B}_0 = 1{,}000 I_K$, $\nu_\Gamma = \nu_\Sigma = D + 1$ 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 3.5. 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.

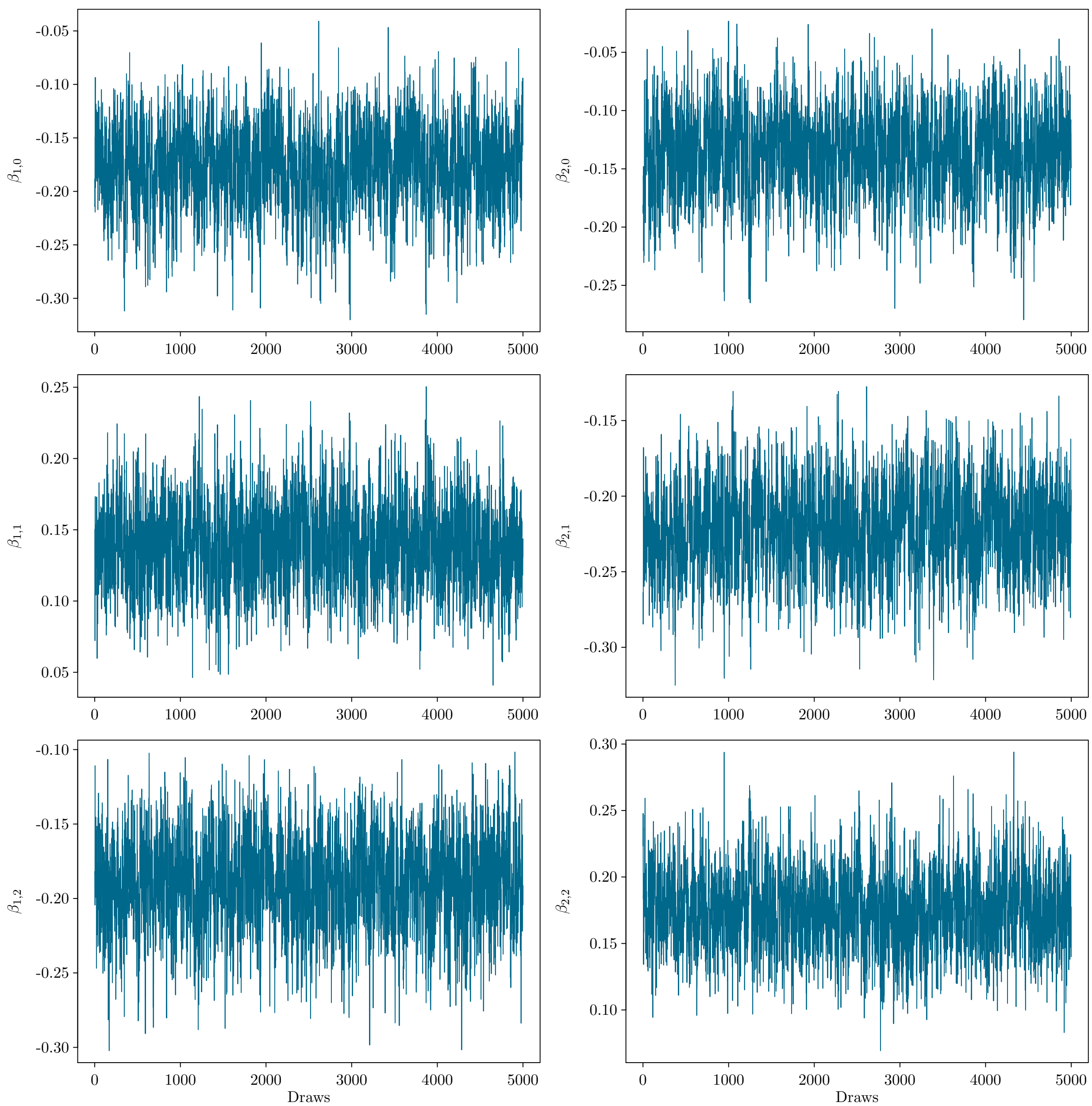
Finally, Figures 3.1 and 3.2 give a graphical depiction of the posterior chains for the coefficients in a single Monte Carlo draw. One of the major advantages of the Bayesian approach is its ability to produce a complete distribution for each parameter of interest from which all proceeding information 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 that the chains satisfy the criteria for convergence to their stationary distribution.

3.3 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.

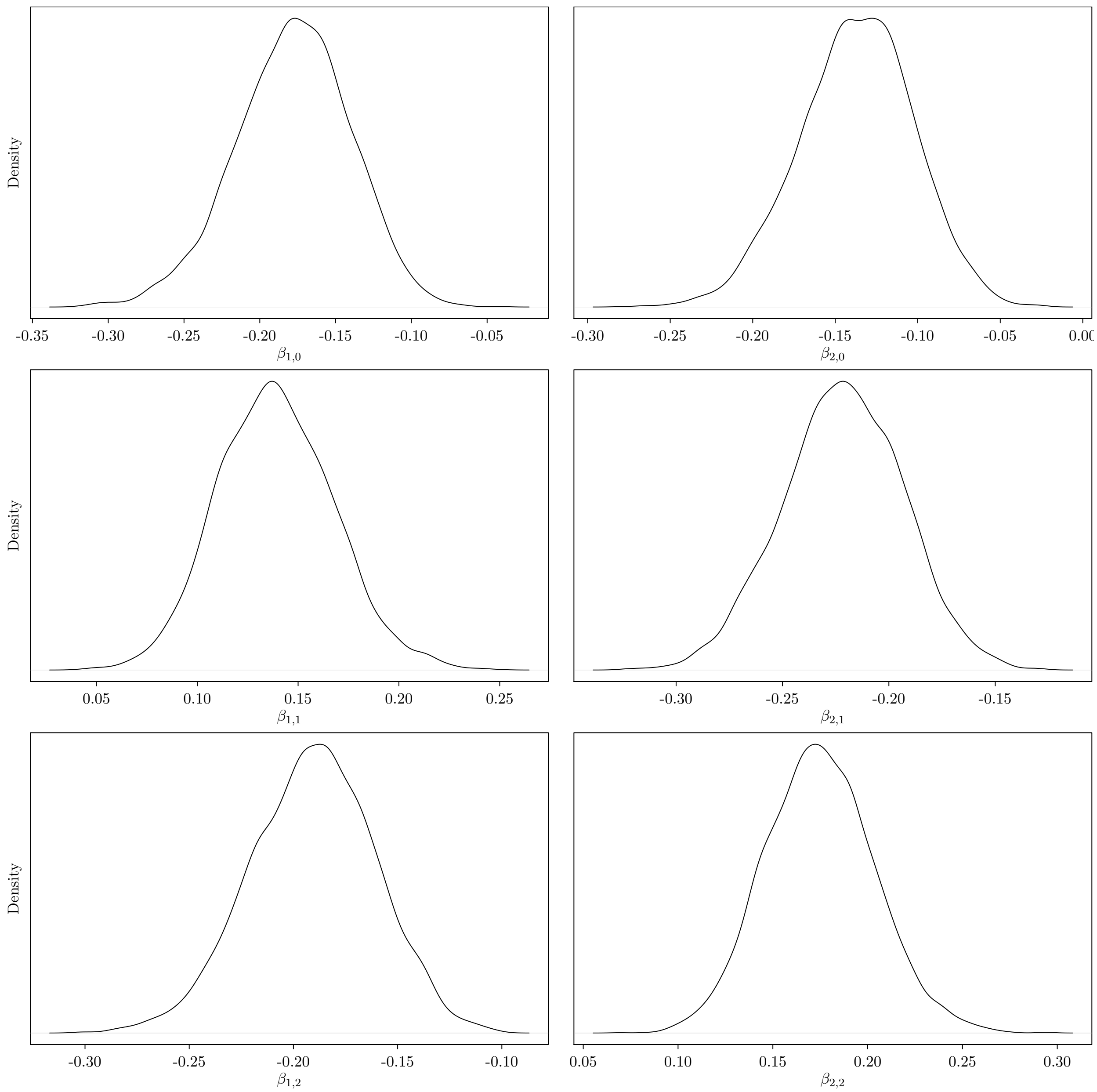
Figure 3.1: Trace Plot of Coefficients for Latent Dependent Variable Model



Note: Results from 5,000 simulations after a burn-in period of 1,000. The draws on the coefficients integrate out the unobserved heterogeneity.

Note:

Figure 3.2: Density Plot of Coefficients for Latent Dependent Variable Model



Note: Results from 5,000 simulations after a burn-in period of 1,000. The draws on the coefficients integrate out the unobserved heterogeneity.

Appendices

3.A Details on Integration Methods for MLE

The integrals given by the conditionally independent (3.1.6) and pooled (3.1.7) likelihoods can be cast in a general way as the problem of numerically evaluating the following integral for some function $f : \mathbb{R}^D \times \mathbb{R}^p \to \mathbb{R}$:

$$V \equiv \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\boldsymbol{c}, \boldsymbol{z}) \phi_D(\boldsymbol{c}; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma}) \, \mathrm{d}\boldsymbol{c} \; , \tag{3.A.1}$$

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

$$\int_{-\infty}^{\infty} g(c, \boldsymbol{z}) \exp\left(-c^2\right) \mathrm{d}c \approx \sum_{s=1}^{S} w_s g(a_s, \boldsymbol{z}) \, , \tag{3.A.2}$$

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.

3.A.1 Adaptive Quadrature

The adaptive approach to evaluate the multidimensional integral in (3.A.1) 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.,

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

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

$$V_{\text{adaptive}} \approx 2^{\frac{D}{2}} |\widehat{\boldsymbol{Q}}|^{\frac{1}{2}} \sum_{s_1=1}^{S} \cdots \sum_{s_D=1}^{S} \prod_{j=1}^{D} w_{s_j} \exp\left(\boldsymbol{a}_s'\boldsymbol{a}_s\right) f(\boldsymbol{a}_s^*, \boldsymbol{z}) \phi_D(\boldsymbol{a}_s^*; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma}) \tag{3.A.3}$$

3.A.2 Nonadaptive Quadrature

This method operates by noting that, since we are already starting from a function times a Gaussian density in (3.A.1), 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, (3.A.1) 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

$$V_{\text{nonadaptive}} \approx \pi^{-\frac{D}{2}} \sum_{s_1=1}^{S} \cdots \sum_{s_D=1}^{S} \prod_{j=1}^{D} w_{s_j} f(\boldsymbol{a}_s^*, \boldsymbol{z})\,. \tag{3.A.4}$$

3.A.3 Pruning

One final issue that is of interest for the computation of both (3.A.3) and (3.A.4) 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.

3.B Derivatives for MLE and Probit Estimators

3.B.1 Scores for Independent and Pooled MLE

Starting from (3.1.6) or (3.1.7), replace the multinomial logit link (3.1.11) 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

$$\begin{aligned}\frac{\partial \ell_i^{(\mathrm{ind})}(\boldsymbol{\beta}, \boldsymbol{\Gamma})}{\partial \boldsymbol{\beta}_k} = L_i^{(\mathrm{ind})}(\boldsymbol{\beta}, \boldsymbol{\Gamma}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Bigg\{ \left[\prod_{t=1}^{T_i} \prod_{j=1}^{d} m_{ijt}^{y_{ijt}} \right] \left[\sum_{t=1}^{T_i} (y_{itk} - m_{itk})\boldsymbol{x}_{itk} \right] \\ \times \phi_D(\boldsymbol{c}_i; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma}) \Bigg\} \mathrm{d}\boldsymbol{c}_i \,,\end{aligned} \tag{3.B.1}$$

for the likelihood assuming conditional independence and

$$\begin{aligned}\frac{\partial \ell_i^{(\mathrm{pool})}(\boldsymbol{\beta}, \boldsymbol{\Gamma})}{\partial \boldsymbol{\Gamma}} = \sum_{t=1}^{T_i} L_{it}^{(\mathrm{pool})}(\boldsymbol{\beta}, \boldsymbol{\Gamma}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Bigg\{ \left[\prod_{j=1}^{d} m_{ijt}^{y_{ijt}} \right] (y_{itk} - m_{itk})\boldsymbol{x}_{itk} \\ \times \phi_D(\boldsymbol{c}_i; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma}) \Bigg\} \mathrm{d}\boldsymbol{c}_i \,,\end{aligned} \tag{3.B.2}$$

for the pooled likelihood. The terms $L_i^{(\mathrm{ind})}(\boldsymbol{\beta}, \boldsymbol{\Gamma})$ and $L_{it}^{(\mathrm{pool})}(\boldsymbol{\beta}, \boldsymbol{\Gamma})$ 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

$$\begin{aligned}\frac{\partial \ell_i^{(\mathrm{ind})}(\boldsymbol{\beta}, \boldsymbol{\Gamma})}{\partial \boldsymbol{\Gamma}} = L_i^{(\mathrm{ind})}(\boldsymbol{\beta}, \boldsymbol{\Gamma}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Bigg\{ \left[\prod_{t=1}^{T_i} \prod_{j=1}^{d} m_{ijt}^{y_{ijt}} \right] \boldsymbol{\Gamma}^{-1}(I_D - \boldsymbol{c}_i \boldsymbol{c}_i' \boldsymbol{\Gamma}^{-1}) \\ \times \phi_D(\boldsymbol{c}_i; \boldsymbol{0}_{D\times 1}, \boldsymbol{\Gamma}) \Bigg\} \mathrm{d}\boldsymbol{c}_i \,,\end{aligned} \tag{3.B.3}$$

for the likelihood assuming conditional independence and

$$\frac{\partial \ell_i^{(\mathrm{pool})}(\boldsymbol{\beta}, \boldsymbol{\Gamma})}{\partial \boldsymbol{\beta}_k} = \sum_{t=1}^{T_i} L_{it}^{(\mathrm{pool})}(\boldsymbol{\beta}, \boldsymbol{\Gamma}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \left[\prod_{j=1}^{d} m_{ijt}^{y_{ijt}} \right] \boldsymbol{\Gamma}^{-1} (I_D - \boldsymbol{c}_i \boldsymbol{c}_i' \boldsymbol{\Gamma}^{-1}) \times \phi_D(\boldsymbol{c}_i; \mathbf{0}_{D\times 1}, \boldsymbol{\Gamma}) \right\} \mathrm{d}\boldsymbol{c}_i \ , \tag{3.B.4}$$

for the pooled likelihood.

3.B.2 Score and Hessian for Probit NLS

Starting from the objective function (3.1.14), 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

$$\frac{\partial q_i(\boldsymbol{\alpha}_c)}{\partial \boldsymbol{\alpha}_c} = -\sum_{t=1}^{T_i} \begin{bmatrix} \phi(\tilde{\boldsymbol{x}}_{it1}' \boldsymbol{\alpha}_{1c}) [y_{it1} - \Phi(\tilde{\boldsymbol{x}}_{it1}' \boldsymbol{\alpha}_{1c})] \tilde{\boldsymbol{x}}_{it1} \\ \vdots \\ \phi(\tilde{\boldsymbol{x}}_{itD}' \boldsymbol{\alpha}_{Dc}) [y_{itD} - \Phi(\tilde{\boldsymbol{x}}_{itD}' \boldsymbol{\alpha}_{Dc})] \tilde{\boldsymbol{x}}_{itD} \end{bmatrix} . \tag{3.B.5}$$

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

$$\frac{\partial^2 q_i(\boldsymbol{\alpha}_c)}{\partial \boldsymbol{\alpha}_{kc} \partial \boldsymbol{\alpha}_{kc}} = \sum_{t=1}^{T_i} \phi(\tilde{\boldsymbol{x}}_{itk}' \boldsymbol{\alpha}_{kc}) \{ \phi(\tilde{\boldsymbol{x}}_{itk}' \boldsymbol{\alpha}_{kc}) + \tilde{\boldsymbol{x}}_{itk}' \boldsymbol{\alpha}_{kc} [y_{itk} - \Phi(\tilde{\boldsymbol{x}}_{itk}' \boldsymbol{\alpha}_{kc})] \} \tilde{\boldsymbol{x}}_{itk} \tilde{\boldsymbol{x}}_{itk}' \, , \tag{3.B.6}$$

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

Bibliography

Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 44(2):139–177.

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.

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.

Allenby, G. M. and Lenk, P. J. (1994). Modeling Household Purchase Behavior with Logistic Normal Regression. *Journal of the American Statistical Association*, 89(428):1218–1231.

Andrews, D. W. K. (1995). Nonparametric Kernel Estimation for Semiparametric Models. *Econometric Theory*, 11(3):560–586.

Arbenz, P., Embrechts, P., and Puccetti, G. (2011). The AEP Algorithm for the Fast Computation of the Distribution of the Sum of Dependent Random Variables. *Bernoulli*, 17(2):562–591.

Athey, S., Imbens, G. W., and Wager, S. (2018). Approximate Residual Balancing: Debiased Inference of Average Treatment Effects in High Dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4):597–623.

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. (1983). New Indices of Money Supply and the Flexible Laurent Demand System. *Journal of Business & Economic Statistics*, 1(1):7–23.

Barnett, W. A. (2002). Tastes and Technology: Curvature is not Sufficient for Regularity. *Journal of Econometrics*, 108(1):199–202.

Barnett, W. A. and Lee, Y. W. (1985). The Global Properties of the Minflex Laurent, Generalized Leontief, and Translog Flexible Functional Forms. *Econometrica*, 53(6):1421–1437.

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

Barten, A. P. (1969). Maximum Likelihood Estimation of a Complete System of Demand Equations. *European Economic Review*, 1(1):7–73.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on Treatment Effects after Selection Among High-Dimensional Controls. *Review of Economic Studies*, 81(2):608–650.

Belloni, A., Chernozhukov, V., and Wei, Y. (2016). Post-Selection Inference for Generalized Linear Models with Many Controls. *Journal of Business & Economic Statistics*, 34(4):606–619.

Blundell, R., Horowitz, J. L., and Parey, M. (2012). Measuring the Price Responsiveness of Gasoline Demand: Economic Shape Restrictions and Nonparametric Demand Estimation. *Quantitative Economics*, 3(1):29–51.

Blundell, R., Pashardes, P., and Weber, G. (1993). What do we learn about Consumer Demand Patterns from Micro Data? *American Economic Review*, 83(3):570–597.

Breunig, C., Mammen, E., and Simoni, A. (2020). Ill-Posed Estimation in High-Dimensional Models with Instrumental Variables. *Journal of Econometrics*, 219(1):171–200.

Brooks, S. P. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.

Bunke, O. and Milhaud, X. (1998). Asymptotic Behavior of Bayes Estimates under Possibly Incorrect Models. *Annals of Statistics*, 26(2):617–644.

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.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1):1–32.

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

Chang, D. and Serletis, A. (2014). The Demand for Gasoline: Evidence from Household Survey Data. *Journal of Applied Econometrics*, 29(2):291–313.

Charpentier, A., Fermanian, J.-D., and Scaillet, O. (2007). The Estimation of Copulas: Theory and Practice. In Rank, J., editor, *Copulas: From Theory to Application in Finance*, pages 35–64. Risk Books.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/Debiased Machine Learning for Treatment and Structural Parameters. *Econometrics Journal*, 21(1):C1–C68.

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.

Chipman, H., George, E. I., McCulloch, R. E., Clyde, M., Foster, D. P., and Stine, R. A. (2001). The Practical Implementation of Bayesian Model Selection. In Lahiri, P., editor, *Model Selection*, volume 38 of *Lecture Notes-Monograph Series*, pages 65–134. Institute of Mathematical Statistics.

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.

Considine, T. J. and Mount, T. D. (1984). The Use of Linear Logit Models for Dynamic Input Demand Systems. *Review of Economics and Statistics*, pages 434–443.

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.

Dubin, J. A. (2007). Valuing Intangible Assets with a Nested Logit Market Share Model. *Journal of Econometrics*, 139(2):285–302.

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.

Elfadaly, F. G. and Garthwaite, P. H. (2017). Eliciting Dirichlet and Gaussian Copula Prior Distributions for Multinomial Models. *Statistics and Computing*, 27(2):449–467.

Escanciano, J. C., Jacho-Chávez, D. T., and Lewbel, A. (2014). Uniform Convergence of Weighted Sums of Non and Semiparametric Residuals for Estimation and Testing. *Journal of Econometrics*, 178:426–443.

Escanciano, J. C., Jacho-Chávez, D. T., and Lewbel, A. (2016). Identification and Estimation of Semiparametric Two-Step Models. *Quantitative Economics*, 7(2):561–589.

Fan, J. and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Fan, Y. and Tang, C. Y. (2013). Tuning Parameter Selection in High Dimensional Penalized Likelihood. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pages 531–552.

Ferrari, S. and Cribari-Neto, F. (2004). Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, 31(7):799–815.

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

Frahm, G., Junker, M., and Szimayer, A. (2003). Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*, 63(3):275–286.

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.

Gaines, B. R., Kim, J., and Zhou, H. (2018). Algorithms for Fitting the Constrained Lasso. *Journal of Computational and Graphical Statistics*, 27(4):861–871.

Genest, C. and Nešlehová, J. (2007). A Primer on Copulas for Count Data. *ASTIN Bulletin: The Journal of the IAA*, 37(2):475–515.

Geweke, J. (1989). Bayesian Inference in Econometric Models using Monte Carlo Integration. *Econometrica*, 57(6):1317–1339.

Gijbels, I. and Herrmann, K. (2014). On the Distribution of Sums of Random Variables with Copula-induced Dependence. *Insurance: Mathematics and Economics*, 59:27–44.

Glassman, D. A. and Riddick, L. A. (1994). A New Method of Testing Models of Portfolio Diversification: An Application to International Portfolio Choice. *Journal of International Financial Markets, Institutions & Money*, 4:27–47.

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.

Hans, C. (2009). Bayesian Lasso Regression. *Biometrika*, 96(4):835–845.

Hansen, C. and Liao, Y. (2019). The Factor-Lasso and K-Step Bootstrap Approach for Inference in High-Dimensional Economic Applications. *Econometric Theory*, 35(3):465–509.

Hardle, W., Hall, P., and Ichimura, H. (1993). Optimal Smoothing in Single-Index Models. *Annals of Statistics*, 21(1):157–178.

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

Hayfield, T. and Racine, J. S. (2008). Nonparametric Econometrics: The np Package. *Journal of Statistical Software*, 27(5):1–32.

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

Ichimura, H. (1993). Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models. *Journal of Econometrics*, 58(1–2):71–120.

Ishwaran, H. and Rao, J. S. (2005). Spike and Slab Variable Selection: Frequentist and Bayesian Strategies. *Annals of Statistics*, 33(2):730–773.

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

Javanmard, A. and Montanari, A. (2014). Confidence Intervals and Hypothesis Testing for High-Dimensional Regression. *Journal of Machine Learning Research*, 15(1):2869–2909.

Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman and Hall/CRC.

Joe, H. and Xu, J. J. (1996). The Estimation Method of Inference Functions for Margins for Multivariate Models. Technical Report 166, University of British Columbia, Department of Statistics.

Johnson, N. D. and Mislin, A. A. (2011). Trust Games: A Meta-Analysis. *Journal of Economic Psychology*, 32(5):865–889.

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.

Kieschnick, R. and McCullough, B. D. (2003). Regression Analysis of Variates Observed on (0, 1): Percentages, Proportions and Fractions. *Statistical Modelling*, 3(3):193–213.

Klein, R. W. and Spady, R. H. (1993). An Efficient Semiparametric Estimator for Binary Response Models. *Econometrica*, 61(2):387–421.

Knight, K. and Fu, W. (2000). Asymptotics for Lasso-type Estimators. *Annals of Statistics*, pages 1356–1378.

Koch, S. F. (2015). On the Performance of Fractional Multinomial Response Models for Estimating Engel Curves. *Agrekon*, 54(1):28–52.

Kyung, M., Gill, J., Ghosh, M., and Casella, G. (2010). Penalized Regression, Standard Errors, and Bayesian Lassos. *Bayesian Analysis*, 5(2):369–411.

Lee, J. D., Sun, D. L., Sun, Y., and Taylor, J. E. (2016). Exact Post-Selection Inference, with Application to the Lasso. *Annals of Statistics*, 44(3):907–927.

Leng, C., Tran, M.-N., and Nott, D. (2014). Bayesian Adaptive Lasso. *Annals of the Institute of Statistical Mathematics*, 66(2):221–244.

LeSage, J. P. (2004). Introduction to Spatial and Spatiotemporal Econometrics. In *Spatial and Spatiotemporal Econometrics*, volume 18 of *Advances in Econometrics*. Emerald Group Publishing Ltd.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating Random Correlation Matrices based on Vines and Extended Onion Method. *Journal of Multivariate Analysis*, 100(9):1989–2001.

Lewbel, A. (1991). The Rank of Demand Systems: Theory and Nonparametric Estimation. *Econometrica*, pages 711–730.

Lewbel, A. and Pendakur, K. (2009). Tricks with Hicks: The EASI Demand System. *American Economic Review*, 99(3):827–63.

Li, Q. and Lin, N. (2010). The Bayesian Elastic Net. *Bayesian Analysis*, 5(1):151–170.

Li, Q. and Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

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

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.

Martínez-Flórez, G., Leiva, V., Gómez-Déniz, E., and Marchant, C. (2020). A Family of Skew-Normal Distributions for Modeling Proportions and Rates with Zeros/Ones Excess. *Symmetry*, 12(9):1439.

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

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

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.

Newey, W. K. (1994). The Asymptotic Variance of Semiparametric Estimators. *Econometrica*, 62(6):1349–1382.

Ning, Y., Zhao, T., and Liu, H. (2017). A Likelihood Ratio Framework for High-Dimensional Semiparametric Regression. *Annals of Statistics*, 45(6):2299–2327.

Osborne, M. R., Presnell, B., and Turlach, B. A. (2000). On the LASSO and its Dual. *Journal of Computational and Graphical Statistics*, 9(2):319–337.

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.

Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482):681–686.

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.

Poirier, D. J. (1998). Revising Beliefs in Nonidentified Models. *Econometric Theory*, 14(4):483–509.

Pérez, M.-E., Pericchi, L. R., and Ramírez, I. C. (2016). The scaled beta2 distribution as a robust prior for scales. *Bayesian Analysis*.

Ramalho, J. J. S. and Silva, J. V. (2009). A Two-Part Fractional Regression Model for the Financial Leverage Decisions of Micro, Small, Medium and Large Firms. *Quantitative Finance*, 9(5):621–636.

Ramírez-Hassan, A. (2021). Bayesian Estimation of the EASI Demand System: Replicating the Lewbel and Pendakur (2009) Results. *Journal of Applied Econometrics*.

Ramírez-Hassan, A. and Montoya-Blandón, S. (2020). Forecasting from others' experience: Bayesian estimation of the generalized bass model. *International Journal of Forecasting*, 36(2):442–465.

Rodrigues, J., Bazán, J. L., and Suzuki, A. K. (2020). A Flexible Procedure for Formulating Probability Distributions on the Unit Interval with Applications. *Communications in Statistics - Theory and Methods*, 49(3):738–754.

Ročková, V. and George, E. I. (2018). The Spike-and-Slab LASSO. *Journal of the American Statistical Association*, 113(521):431–444.

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.

Sigrist, F. and Stahel, W. A. (2011). Using the Censored Gamma Distribution for Modeling Fractional Response Variables with an Application to Loss Given Default. *ASTIN Bulletin: The Journal of the IAA*, 41(2):673–710.

Simas, A. B., Barreto-Souza, W., and Rocha, A. V. (2010). Improved Estimators for a General Class of Beta Regression Models. *Computational Statistics & Data Analysis*, 54(2):348–366.

Sims, C. A. and Zha, T. (1998). Bayesian Methods for Dynamic Multivariate Models. *International Economic Review*, 39(4):949–968.

Sklar, M. (1959). Fonctions de Repartition an Dimensions et Leurs Marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian Statistics without Tears: A Sampling-Resampling Perspective. *The American Statistician*, 46(2):84–88.

Smith, M. S. and Khaled, M. A. (2012). Estimation of Copula Models with Discrete Margins via Bayesian Data Augmentation. *Journal of the American Statistical Association*, 107(497):290–303.

Smithson, M. and Shou, Y. (2017). CDF-Quantile Distributions for Modelling Random Variables on the Unit Interval. *British Journal of Mathematical and Statistical Psychology*, 70(3):412–438.

Smithson, M. and Verkuilen, J. (2006). A Better Lemon Squeezer? Maximum-Likelihood Regression with Beta-Distributed Dependent Variables. *Psychological Methods*, 11(1):54.

Song, P. X.-K. and Tan, M. (2000). Marginal Models for Longitudinal Continuous Proportional Data. *Biometrics*, 56(2):496–502.

Sosa, M. L. (2009). Application-Specific R&D Capabilities and the Advantage of Incumbents: Evidence from the Anticancer Drug Market. *Management Science*, 55(8):1409–1422.

Stavrunova, O. and Yerokhin, O. (2012). Two-Part Fractional Regression Model for the Demand for Risky Assets. *Applied Economics*, 44(1):21–26.

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.

Strasser, H. (1981). Consistency of Maximum Likelihood and Bayes Estimates. *Annals of Statistics*, 9(5):1107–1113.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1):267–288.

Tibshirani, R., Bien, J., Friedman, J., Hastie, T., Simon, N., Taylor, J., and Tibshirani, R. J. (2012). Strong Rules for Discarding Predictors in Lasso-type Problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(2):245–266.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and Smoothness via the Fused Lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108.

Tiffin, R. and Aguiar, M. (1995). Bayesian Estimation of an Almost Ideal Demand System for fresh fruit in Portugal. *European Review of Agricultural Economics*, 22(4):469–480.

Trivedi, P. and Zimmer, D. (2017). A Note on Identification of Bivariate Copulas for Discrete Count Data. *Econometrics*, 5(1):10.

Trivedi, P. K. and Zimmer, D. M. (2007). Copula Modeling: An Introduction for Practitioners. *Foundations and Trends® in Econometrics*, 1(1):1–111.

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.

van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On Asymptotically Optimal Confidence Regions and Tests For High-Dimensional Models. *Annals of Statistics*, 42(3):1166–1202.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2020). Rank-normalization, Folding, and Localization: An Improved $\widehat{R}$ for Assessing Convergence of MCMC. *Bayesian Analysis*.

Velásquez-Giraldo, M., Canavire-Bacarreza, G., Huynh, K., and Jacho-Chávez, D. (2018). Flexible Estimation of Demand Systems: A Copula Approach. *Journal of Applied Econometrics*, 33(7):1109–1116.

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.

Wang, H. and Leng, C. (2007). Unified LASSO estimation by Least Squares Approximation. *Journal of the American Statistical Association*, 102(479):1039–1048.

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.

Yen, T.-J. (2011). A Majorization-Minimization Approach to Variable Selection using Spike and Slab Priors. *The Annals of Statistics*, 39(3):1748–1775.

Yuan, M. and Lin, Y. (2006). Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association*, 57(298):348–368.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence Intervals for Low Dimensional Parameters in High Dimensional Linear Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pages 217–242.

Zou, H. and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.