Copula Estimation and Model Selection with Multivariate Fractional Outcomes

Santiago Montoya-Blandón*

Emory University
Job Market Paper

Abstract
This paper introduces several estimation procedures using copulas for multivariate fractional outcomes with a conditional mean specification. All methods satisfy the fractional and unit-sum constraints on the outcomes while allowing for possible cross-equation restrictions (which are central in demand estimation). While ultimately Bayesian in nature, the paper rigorously examines the asymptotic properties of the arising frequentist estimators, as they are themselves an addition to the literature. The methodology can also be augmented to handle model selection using regularization in a Bayesian framework. A range of numerical exercises evaluate the properties of the estimators and showcase their flexibility under different contexts such as structural or reduced form models. An empirical application to transportation expenditures in Canada is also presented.

Keywords: Multivariate fractional data; Copula; Bayesian methods; Model selection; Demand estimation

JEL classification: C35, C51, C52, D12

1 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). These arise naturally in economics when estimating a demand system in which the dependent variables are given as expenditure shares on different categories of goods (Woodland, 1979, Barnett and Serletis, 2008). There are several important characteristics to keep in mind when dealing in this context. First, the nature of the outcomes implies that they are both fractional (i.e., bounded between 0 and 1) and satisfy a unit-sum constraint. This type of observations are known as compositional data in the statistics literature and are characterized as belonging to the d-dimensional simplex

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

*Department of Economics, Emory University, Rich Building 310, 1602 Fishburne Dr., Atlanta, GA 30322-2240, USA. E-mail: smonto2@emory.edu. Website: smontoyablandon.com

For example, in the demand estimation context, d represents the number of categories of goods. Second, most reduced form or structural models produce an estimating equation in the form of a conditional mean such as

$$\mathrm{E}[\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}]=\boldsymbol{m}(\boldsymbol{x},\boldsymbol{\beta})\,, \tag{1}$$

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 which 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). Structural models such as the Almost Ideal Demand (AID) system of Deaton and Muellbauer (1980) present one possible specification for $\boldsymbol{m}(\boldsymbol{x},\boldsymbol{\beta})$, for example. This paper starts from the conditional mean as primary object and introduces methods which impose this restriction while achieving flexibility.

Third, these models usually impose constraints on the parameter vector $\boldsymbol{\beta}$ to satisfy the economic regularity of the demand functions they produce. Furthermore, they are not only restrictions within each equation of (1) but include cross-equation restrictions (Barnett, 2002). The AID model, for example, imposes homogeneity in expenditure 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. 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, Chang and Serletis, 2014). In estimating these models, the second and third points are considered at length, but the simplex nature of the multivariate fractional outcomes is generally ignored by assuming an unrestricted distribution on the disturbances $\boldsymbol{u}\equiv\boldsymbol{Y}-\boldsymbol{m}(\boldsymbol{x},\boldsymbol{\beta})$ (Barnett and Serletis, 2008). This paper aims to correct that.

The main contribution of this paper is introducing Bayesian estimation procedures via copulas that incorporates all three points previously discussed. That is, they impose the fractional and unit-sum constraints of these limited support outcomes, satisfy a conditional mean structure, and can incorporate cross-equation restrictions. The use of copulas also broaden the possible dependence patterns between each share in the system, a general concern in the compositional data literature (Aitchison, 2003). The scope of the paper is realized as follows: first, two ways of constructing a likelihood using copulas in this setup are presented, where the marginal distributions satisfy the fractional restriction, and the joint captures the dependence structure and unit-sum constraint between shares. The frequentist maximum likelihood estimators (MLE) arising from these constructions are themselves contributions to the compositional data literature and are applicable in both the structural demand estimation context and reduced form models. Therefore, to introduce these estimators as stand-alone alternatives, their asymptotic properties in a standard frequentist context are presented.

To complete a Bayesian specification, one could place standard prior distributions on the marginal and dependence parameters to form a simple Bayesian estimator. However, a more general class of priors can be used to augment the base estimators to handle covariate selection through the use of regularization (Park and Casella, 2008, Hans, 2009). This is important in reduced form models for choosing relevant covariates and in the demand estimation context where certain functions can be approximated using polynomials whose degree would need to be chosen from the data. Such selection would be useful even in the case where the dimensionality of the covariates was large or grew with the sample size (i.e., high-dimensional settings, see Li and Lin, 2010). Finally, the use of a Bayesian framework guarantees that, even with a selection step, inference is simple not only for the estimated parameters (which would be useful in a structural context) but also for functions of interest computed from these parameters. These include quantities such as elasticities after estimation of a demand system or average partial effects (APE) in reduced form models. The empirical applications employing these methodologies show that the income elasticities of gasoline, local transportation and intercity transportation are all close to unity across a range of different household segments. In terms of price elasticity, it appears that gasoline is more elastic than in previous studies, particularly when including higher-order polynomials on deflated expenditure.

The paper 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 3 introduces the class of prior distributions considered for the coefficients of the conditional mean and outlines the Bayesian estimation algorithm. Numerical exercises in Section 4 showcase the properties and flexibility of these estimators, as well as their comparison with other method available in the literature. Section 5 presents an application of the proposed method to the demand of transportation services in Canada from a structural demand estimation perspective. Section 6 presents the concluding remarks.

2 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) consider 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 even when only (1) is imposed, while sacrificing efficiency.[1] Not having a correctly specified likelihood also precludes the use of the Bayesian approach and its advantages. This is why the paper combines copulas—expanding the possible dependence structure allowed between shares while adding robustness—with a full-likelihood approach to take advantage of Bayesian methods in estimation, selection and inference.

Starting from a model on $\boldsymbol{Y} \in \mathcal{S}^d$ with an additive mean-zero disturbance is usually the way in which (1) is justified. The disturbances $\boldsymbol{u} \equiv \boldsymbol{Y} - \boldsymbol{m}(\boldsymbol{x}, \boldsymbol{\beta})$ can be interpreted as random deviations from the theory given by preference shocks or can be tied to specific sources of heterogeneity of the agents (Lewbel and Pendakur, 2009). The additive disturbance reasoning remains valid as long as exogeneity is a plausible assumption in these models so as to satisfy $\mathrm{E}[\boldsymbol{u}|\boldsymbol{X} = \boldsymbol{x}] = 0$.

Example 1. (Demand Estimation) As noted before, the AID is one possibility for these conditional mean specifications, where $\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{\beta})$ is given by

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

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 the covariates represent total expenditure 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 (QAID, Banks et al., 1997), Minflex Laurent (ML, Barnett, 1983, Barnett and Lee, 1985), and recently the exact affine Stone index (EASI, 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 common 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,

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

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

2007). For example, a model could take the form of

$$m_j(\boldsymbol{x}, \boldsymbol{\beta}) = \begin{cases} \frac{\exp(\boldsymbol{x}'\boldsymbol{\beta}_j)}{1+\sum_{l=1}^{j-1}\exp(\boldsymbol{x}'\boldsymbol{\beta}_l)} & \text{for } j = 1, \ldots, d-1 \\ \frac{1}{1+\sum_{l=1}^{j-1}\exp(\boldsymbol{x}'\boldsymbol{\beta}_l)} & \text{for } j = d \end{cases} \tag{3}$$

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

Applications of these models are not important only to economics (Mullahy and Robert, 2010, Koch, 2015), but to other fields such as dividends and firm analysis (Loudermilk, 2007, Ramalho and Silva, 2009, Sosa, 2009, Sigrist and Stahel, 2011), portfolios (Glassman and Riddick, 1994, Stavrunova and Yerokhin, 2012, Mullahy, 2015), psychology (Smithson and Verkuilen, 2006, Johnson and Mislin, 2011), among others.

2.1 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, imposes a conditional mean as in (1), and allows for greater flexibility in the dependence structure between shares in the system. Let $(\boldsymbol{Y}', \boldsymbol{X}')'$ be a $(d+p)$-dimensional random-vector where $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$ 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 $j = 1, \ldots, d$. 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 $j = 1, \ldots, d$ and almost all $\boldsymbol{x} \in \mathcal{X}$. As mentioned previously, the following conditional mean specification is assumed to hold throughout:

Assumption 1. The joint distribution of $(\boldsymbol{Y}, \boldsymbol{X})$ satisfies

$$\mathrm{E}[Y_j|\boldsymbol{X} = \boldsymbol{x}] = m_j(\boldsymbol{x}, \boldsymbol{\beta}_0)\,, \tag{4}$$

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 $\boldsymbol{\beta}$, $j = 1, \ldots, d$.

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: $\sum_{j=1}^d m_j(\boldsymbol{x}, \boldsymbol{\beta}) = 1$. An application of Sklar's (1959) theorem allows for a representation using copulas as $H(y_1, \ldots, y_d|\boldsymbol{X} = \boldsymbol{x}) = C(H_1(y_1|\boldsymbol{X} = \boldsymbol{x}), \ldots, H_d(y_d|\boldsymbol{X} = \boldsymbol{x}))$, 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. The marginals $H_j, j = 1, \ldots, d$ and the copula C admit density functions conditional on $\boldsymbol{X} = \boldsymbol{x}$, and denoted by $h_j, j = 1, \ldots, d$ and c, respectively.

Given Assumption 2, the conditional joint density $h(y_1, \ldots, y_d | \boldsymbol{X} = \boldsymbol{x})$ is well-defined, and so is the unconditional density. In an estimation setting, a given copula C_Y and marginals F_j are selected from families $\mathcal{C}$ and $\mathcal{F}_j$ for each $j = 1, \ldots, d$, respectively. Taking a parametric stance on the definition of the copula, the conditional joint can be expressed as

$$F_{1,\ldots,d}(\boldsymbol{y}|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\psi}) = C_Y(F_1(y_1|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_1), \ldots, F_d(y_d|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_d), \boldsymbol{\psi}), \tag{5}$$

where $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \ldots, \boldsymbol{\delta}_d')' \in \Delta$ 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 $\Delta = \times_{j=1}^{d} \Delta_j \subset \mathbb{R}_j^D$, with D_j the dimensionality of each $\boldsymbol{\delta}_j, j = 1, \ldots, d$, and $\Psi \subset \mathbb{R}^S$. However, note that some issues arise when dealing directly with the object defined by (5) 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 be then given by

$$F_d(y_d|\boldsymbol{X} = \boldsymbol{x}) = 1 - F_W(1 - y_d|\boldsymbol{X} = \boldsymbol{x}), \tag{6}$$

where $F_W(w|\boldsymbol{X} = \boldsymbol{x}) = \lim_{v_j \to \infty, j=2,\ldots,d-1} \Pr(Y_1 + \cdots + Y_{d-1} \leq w, Y_2 \leq v_2, \ldots, Y_{d-1} \leq v_{d-1}|\boldsymbol{X} = \boldsymbol{x})$,[3] and this probability is taken over the joint distribution of $(Y_1, \ldots, Y_{d-1})'$, which could be obtained from a second application of Sklar's theorem. 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 have 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 which yields (6) 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.

The 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 conditional mean specification in (4) 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.

2.1.1 Copula Specification on Y

Consider placing a copula similar to (5) 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 base and thus eliminated:

$$F(\boldsymbol{y}_{-d}|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}, \boldsymbol{\psi}) = C_Y(F_1(y_1|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_1), \ldots, F_D(y_D|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_D), \boldsymbol{\psi}), \tag{7}$$

[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 marginalize over the remaining components.

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 $\mathcal{T} = \{(y_1,\ldots,y_D) \in \mathbb{R}^D : 0 \le y_j \le 1, j = 1,\ldots,d; \sum_{j=1}^{D} y_j \le 1\}$, so that it does not correspond to a valid distribution on $\mathcal{S}^d$ after marginalizing the last component. Additionally, generating values from the distribution in (7) 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 (6). All proofs can be found in appendix 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*

$$F_W(w|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta},\boldsymbol{\psi}) = \int_0^{w-D+l}\int_0^{w-D+l-y_D}\cdots\int_0^{w-D+l-\sum_{k=D-l+2}^{D} y_k}\int_0^1\cdots\int_0^1 \mathrm{d}F(y_1,\ldots,y_{D-l},y_{D-l+1},\ldots,y_{D-1},y_D|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta},\boldsymbol{\psi}) \tag{8}$$

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

Then, by Assumption 3.A,

$$\begin{aligned} f(\boldsymbol{y}_{-d}|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta},\boldsymbol{\psi};\mathcal{T}) &= \begin{cases} \frac{f(\boldsymbol{y}_{-d}|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta},\boldsymbol{\psi})}{F_W(1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta},\boldsymbol{\psi})} & ;\ \text{if } \boldsymbol{y}_{-d} \in \mathcal{T}\,, \\ 0 & ;\ \text{if } \boldsymbol{y}_{-d} \notin \mathcal{T}\,, \end{cases} \\ &= \mathbb{I}(\boldsymbol{y}_{-d} \in \mathcal{T})\frac{f(\boldsymbol{y}_{-d}|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta},\boldsymbol{\psi})}{F_W(1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta},\boldsymbol{\psi})}\,, \end{aligned} \tag{9}$$

where $\mathbb{I}(\cdot)$ is the indicator function that takes the value of 1 if its argument is true and 0 otherwise, and 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 complicated to compute. Some algorithms, such as the AEP of Arbenz et al. (2011), are devised for the specific purpose of approximating the integral in (8). 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.

2.1.2 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 $\boldsymbol{Y}$, to impose some 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

$$Z_1 = Y_1, \quad Z_j = \frac{Y_j}{1-\sum_{l=1}^{j-1} Y_l} \quad \text{for } j=2,\ldots,d-1, \quad \text{and} \quad Z_d = 1\,. \tag{10}$$

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 (7) 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:

$$G(z_1,\ldots,z_D|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\omega},\boldsymbol{\xi}) = C_Z(G_1(z_1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\omega}_1),\ldots,G_D(z_d|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\omega}_D);\boldsymbol{\xi})\,, \tag{11}$$

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 (7), $G_j, j=1,\ldots,D$ are marginals respecting the fractional constraint, $\Omega = \times_{j=1}^{D}\Omega_j$ with each $\Omega_j \subset \mathbb{R}^{O}_j$, and $\Xi \subset \mathbb{R}^S$. In order to satisfy the conditional mean specification in (4), 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 1. In particular, any such objects that are a solution to*

$$\mu_j(\boldsymbol{x};\boldsymbol{\beta},\boldsymbol{\omega},\boldsymbol{\xi}) + \frac{\mathrm{E}\Big[\widetilde{Z}_j \prod_{l=1}^{j-1}\Big(1-\widetilde{Z}_l - \mu_l(\boldsymbol{x};\boldsymbol{\omega},\boldsymbol{\xi})\Big)\,\Big|\boldsymbol{X}=\boldsymbol{x}\Big]}{1-\sum_{l=1}^{j-1} m_l(\boldsymbol{x},\boldsymbol{\beta})} = \frac{m_j(\boldsymbol{x},\boldsymbol{\beta})}{1-\sum_{l=1}^{j-1} m_l(\boldsymbol{x},\boldsymbol{\beta})}\,, \tag{12}$$

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 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. (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 parametrization 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 parametrization is given by $\mathrm{Var}(Z_j|\boldsymbol{X} = \boldsymbol{x}) = \mu_j(1-\mu_j)/(1+\phi_j)$. Equation (12) 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 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 $c < 1$, itself guaranteed by the unit-sum constraint of the conditional mean functions $m_j(\cdot), j = 1, \ldots, d$. In this setting, we have $\boldsymbol{\omega}_1 = (\mu_1, \phi_1)$, $\boldsymbol{\omega}_2 = (\mu_2, \phi_1)$. This yields (4) for the $\boldsymbol{Y}$ transformed via the inverse transformation (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 (11), 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, and thus the parameters in this construction are 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

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

and 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 the previous example but where the $\boldsymbol{Z}$ are assumed to be independent. If independent beta marginals are combined in this way, it is possible to recover the generalized Dirichlet distribution on $\boldsymbol{Y}$, 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, that mimics Assumption 3.A, yields a distribution for $\boldsymbol{Y}$.

Assumption 3.B. The marginals $G_j, j = 1, \ldots, D$ and the copula C_Z admit density functions conditional on $\boldsymbol{X}$ denoted by $g_j, j = 1, \ldots, D$ and c_Z respectively.

Then, by Assumption 3.B and a change of variables from $\boldsymbol{Z}$ to $\boldsymbol{Y}$,

$$g(\boldsymbol{y}|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta},\boldsymbol{\xi}) = g(\boldsymbol{s}(\boldsymbol{y})|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta},\boldsymbol{\xi}) = c_Z(G_1(s_1(\boldsymbol{y})|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_1),\ldots,G_D(s_D(\boldsymbol{y})|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_D),\boldsymbol{\xi}) \times \prod_{j=1}^{D} \frac{g_j(\boldsymbol{y}_j|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_j)}{1-\sum_{l=1}^{j-1} Y_l} \quad . \tag{13}$$

2.2 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 (9) and (13):

Assumption 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

$$\ell_Y(\boldsymbol{\theta}_Y) = \frac{1}{n}\sum_{i=1}^{n}\Bigg\{\log c_Y(F_1(y_{1,i}|\boldsymbol{X}=\boldsymbol{x}_i;\boldsymbol{\delta}_1),\ldots,F_D(y_{D,i}|\boldsymbol{X}=\boldsymbol{x}_i;\boldsymbol{\delta}_D);\boldsymbol{\psi}) + \sum_{j=1}^{d}\log f_j(y_{j,i}|\boldsymbol{X}=\boldsymbol{x}_i;\boldsymbol{\delta}_j) - \log F_W(1|\boldsymbol{X}=\boldsymbol{x}_i;\boldsymbol{\delta},\boldsymbol{\psi})\Bigg\}, \tag{14}$$

and

$$\ell_Z(\boldsymbol{\theta}_Z) = \frac{1}{n}\sum_{i=1}^{n}\Bigg\{\log c_Z[G_1(s_1(\boldsymbol{y}_i)|\boldsymbol{X}=\boldsymbol{x}_i;\boldsymbol{\delta}_1),\ldots,G_D(s_D(\boldsymbol{y}_i)|\boldsymbol{X}=\boldsymbol{x}_i;\boldsymbol{\delta}_D);\boldsymbol{\xi}] + \sum_{j=1}^{d}\log g_j(s_j(\boldsymbol{y}_i)|\boldsymbol{X}=\boldsymbol{x}_i;\boldsymbol{\delta}_j)\Bigg\}, \tag{15}$$

where the Jacobian term in (15) 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

$$\widehat{\boldsymbol{\theta}}_Y \equiv \underset{\boldsymbol{\theta}_Y\in\Delta\times\Psi}{\arg\max}\,\ell_Y(\boldsymbol{\theta}_Y) \qquad \text{and} \qquad \widehat{\boldsymbol{\theta}}_Z \equiv \underset{\boldsymbol{\theta}_Z\in\Delta\times\Xi}{\arg\max}\,\ell_Z(\boldsymbol{\theta}_Z)\,. \tag{16}$$

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

Assumption 5. (Identification)

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

2. *(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 $\text{Image}(m_j) = \text{Range}(m_j)$, both for $j = 1, \dots, D$.

Assumption 6.A. (Correct specification) (i) There exists $\boldsymbol{\psi}_0 \in \Psi$ and $\boldsymbol{\delta}_0 = (\boldsymbol{\delta}'_{0,1}, \dots, \boldsymbol{\delta}'_{0,D})' \in \Delta$ 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 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 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, e.g., White, 1982) leads to one of the main results of the paper:

Theorem 1. *Under assumptions 1, 2, 3.A or 3.B, 4, 5, 6.A, and regularity conditions, 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*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_e - \boldsymbol{\theta}_{e,0}) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_{e,0}))\,, \tag{17}$$

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

Now, as the focus of the paper is estimating the coefficients associated to the conditional mean, the full strength of Assumption 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 6.A is introduced next:

Assumption 6.B. (Possibly misspecified copula) There exists $\boldsymbol{\delta}_0 = (\boldsymbol{\delta}'_{0,1}, \dots, \boldsymbol{\delta}'_{0,D})' \in \Delta$ such that $H_j(\cdot|\boldsymbol{X} = \boldsymbol{x}) = F_j(\cdot|\boldsymbol{X} = \boldsymbol{x}; \boldsymbol{\delta}_{0,j})$ for all $j = 1, \dots, d$ 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 1 that uses assumption 6.B instead of 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 1. (KL divergence for copula likelihoods) *Under assumptions 1, 2 and 3.A, the KL divergence between the true distribution h, and f as defined by (9), is given by*

$$\begin{aligned}\mathrm{KL}(h, f; \boldsymbol{\theta}_Y) = \mathrm{E}_h\left[\log \frac{c(H_1(Y_1|\boldsymbol{X}=\boldsymbol{x}), \ldots, H_D(Y_D|\boldsymbol{X}=\boldsymbol{x}))}{c_Y(F_1(Y_1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_1), \ldots, F_D(Y_D|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\delta}_D);\boldsymbol{\psi})}\right] + \\ \sum_{j=1}^{D} \mathrm{KL}(h_j, f_j; \boldsymbol{\delta}_j) + \mathrm{E}_h\left[\log \frac{F_W(1|\boldsymbol{X}=\boldsymbol{x};\boldsymbol{\theta}_Y)}{\mathbb{I}(\boldsymbol{Y}\in\mathcal{T})}\right].\end{aligned} \tag{18}$$

The main message from Lemma 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 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. *Under assumptions 1, 2, 3.A, 4, 5, 6.B, and regularity conditions, 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,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_Y - \boldsymbol{\theta}_Y^*) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}_h^{-1}(\boldsymbol{\theta}_Y^*)\mathcal{J}_h(\boldsymbol{\theta}_Y^*)\mathcal{I}_h^{-1}(\boldsymbol{\theta}_Y^*)), \tag{19}$$

where $\boldsymbol{\theta}_Y^* = (\boldsymbol{\delta}_0', \boldsymbol{\psi}^{*\prime})'$ is the pseudo-true parameter vector, $\mathcal{I}_h(\boldsymbol{\theta}_Y^*) = \mathrm{E}_h[\partial^2\ell(\boldsymbol{\theta}_Y^*)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}']$, and $\mathcal{J}_h(\boldsymbol{\theta}_Y^*) = \mathrm{E}_h[\partial\ell(\boldsymbol{\theta}_Y^*)/\partial\boldsymbol{\theta}\,(\partial\ell(\boldsymbol{\theta}_Y^*)/\partial\boldsymbol{\theta})']$.

Theorem 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. It 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 (14), that 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 and that the computational burden is 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 standard fashion.

It is also simple to see why Theorem 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 correct specification of the copula, the inferred relationship cannot reflect the correct marginal structure. The preceding theorems introduce a trade-off in 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, but it might be biased for computing the coefficients of interest. This trade-off is explored numerically in Section 4 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 1. *Let the support of $\boldsymbol{Y}$ be $\mathbb{R}^D$ instead of $\mathcal{S}^d$. Under assumptions 2, 3.A, 4, 5.1, 6.B, and regularity conditions, an estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\delta}}', \widehat{\boldsymbol{\psi}}')'$ based on (14) (without the truncation probability) is consistent and has an asymptotically normal distribution as in (19).*

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 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 (4) is preserved. For example, beta with a reparametrization (Ferrari and Cribari-Neto, 2004, Simas et al., 2010), 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 by following for example 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 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

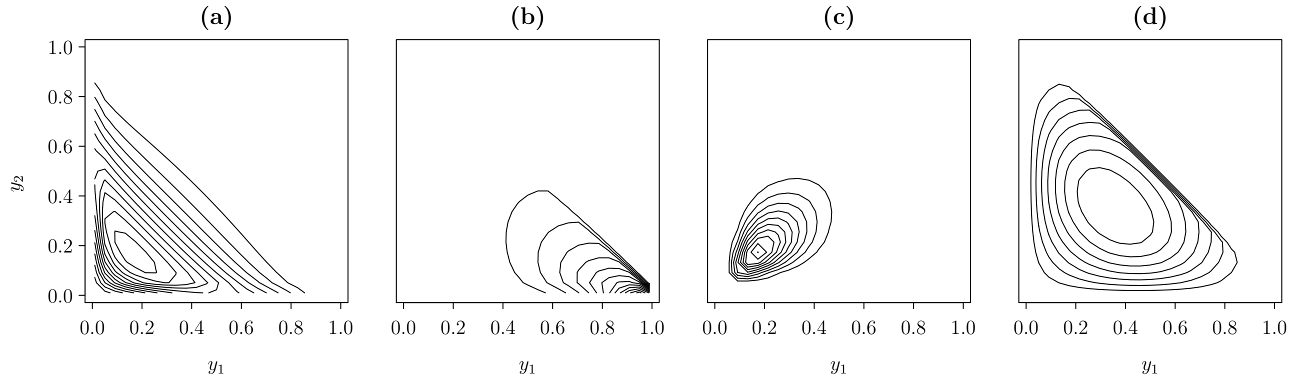
$$\begin{aligned} \widetilde{\boldsymbol{\theta}} \equiv & \arg\max_{\boldsymbol{\theta}_e \in \Theta_e} \ell_e(\boldsymbol{\theta}) \\ & \text{subject to } \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{a} \text{ and } \boldsymbol{B}\boldsymbol{\beta} \leq \boldsymbol{b} \end{aligned} \tag{20}$$

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]

[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).

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

Figure 1: Dependence patterns in copulas



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

3 Priors and Model 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). An important distinction is that, in contrast to frequentist alternatives that are usually used for prediction, Bayesian analysis allows to obtain statistical inference through simple numerical methods. Such a framework would be useful even in contexts where the dimensionality of the covariate space was large or grew 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 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 bootstrap to produce this inference, but they are either computationally complex or not supported theoretically.[6] However, in Bayesian methods, inference follows as a by-product of the algorithms used in Bayesian methods.

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 (14) and (14) (Zou and

[6] To the best of my knowledge, only Mullahy (2015) deals with inference on the average partial effects for multivariate fractional response models. The author does not provide the asymptotic behavior of these estimates or the validity of the bootstrap procedure they used for inference.

Hastie, 2005, Tibshirani et al., 2012). Thus, to automatically include a selection step, the objective function could be augmented to solve

$$\underset{\boldsymbol{\theta}_e \in \Theta_e}{\arg\min}\{-\ell_e(\boldsymbol{\theta}_e) + \rho_{\boldsymbol{\lambda}}(\boldsymbol{\beta})\} \tag{21}$$

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. Useful forms of the penalty could be given by

$$\rho_{\boldsymbol{\lambda}}(\boldsymbol{\beta}) = \lambda ||\boldsymbol{\beta}||_1 \qquad \text{or} \qquad \rho_{\boldsymbol{\lambda}}(\boldsymbol{\beta}) = \lambda \sum_{l=1}^{L} ||\boldsymbol{\beta}_l||_2 \tag{22}$$

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, while the second takes the form of the group LASSO (Yuan and Lin, 2006).

A Bayesian solution to the problem presented in (21) is attractive due to several considerations. First, as penalization methods essentially act as shrinkage, which is a salient feature of the Bayesian paradigm, these estimators usually have a Bayesian interpretation. This connection was recognized at the onset of the penalized regression literature and the introduction of the LASSO (Tibshirani, 1996). They are also connected to resampling ideas (Smith and Gelfand, 1992), which are at the core of recent double or debiased machine learning in high-dimensional settings (Athey et al., 2018, Chernozhukov et al., 2018). However, the latter apply to a semi-parametric framework in which the object of interest is a small-dimensional set of coefficients that depend on an infinite dimensional object estimated using machine learning techniques, and a connection to the concepts considered in this paper is still currently unknown. Additionally, one of the main consideration for adopting a Bayesian framework is its ability to obtain inference through simple probabilistic concepts. While frequentist methods allow for fast coefficient estimation and tuning of the penalty parameters, they are usually unsuited for inference (Kyung et al., 2010). In particular, approximation by ridge regression (Tibshirani, 1996) or asymptotic and bootstrap methods (Knight and Fu, 2000, Fan and Li, 2001) work well to obtain standard errors for the coefficients of selected variables but are unsatisfactory for those of nonselected variables, i.e., those estimated to be 0. Osborne et al. (2000) considers a different and suitable approach, but their results rely on the estimates being approximately linear transformations. This condition does not hold when penalization is substantial (say, a large λ in 22), precisely when these methods would be most attractive.

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 $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_D) \in \Phi \subset \mathbb{R}^D$. That is, we can write $\boldsymbol{\delta}_j = (\boldsymbol{\beta}', \phi_j)'$ for all $j = 1, \ldots, d$, 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 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 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:

$$\pi(\boldsymbol{\beta}) \propto \exp\Big\{ -\frac{1}{2}\rho_{\boldsymbol{\lambda}}(\boldsymbol{\beta})\Big\} . \tag{23}$$

Example. For the penalties in (22), these priors can be implemented using a hierarchical Bayesian approach. For a LASSO penalty, the following hierarchy achieves the desired results:

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

where $\mathcal{N}_K$ represents a multivariate K-dimensional normal distribution, $\tau_1, \ldots, \tau_K$ are hierarchical parameters, and $\mathrm{diag}(\tau_1, \ldots, \tau_K)$ represents a $K \times K$ diagonal matrix with 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:

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

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 by that of Lewandowski et al. (2009). In the context of three shares, so that $D = 2$ need to be modeled and the dependence reduces to a single correlation parameter, flexible alternatives such as a diffuse uniform distribution on the support $[-1, 1]$ or (modified) beta distribution can be placed (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 (21) as in Gaines et al. (2018), Bayesian constrains are also consistently implemented as support restrictions on the prior distributions.[7]

Example 1. (Continued) There are meaningful ways in which sparsity and selection can play a role in demand estimation of structural models. Consider the matrix form of the AID equations (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) that includes a second degree term, or Lewbel and Pendakur (2009) that empirically decide on including up to 5 terms.[8] Incorporating these ideas, one could in general write

$$\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{\alpha} + \Gamma\boldsymbol{p} + \sum_{r=1}^{R} \boldsymbol{\pi}_r \tilde{e}^r \,, \tag{24}$$

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 that could be undertaken using the penalties in (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 (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 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 (4). 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 to define 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 (14) and (15).

Step 3. Choose a prior distribution $\pi(\boldsymbol{\theta}_Y)$ and $\pi(\boldsymbol{\theta}_Z)$ that belongs to the class outlined in (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 $\boldsymbol{\mathcal{A}}$ such that these constraints hold. Include a prior distribution for $\boldsymbol{\lambda}$.

[7] 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).

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

Step 4. Combine the likelihood function and the prior distribution via Bayes' 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.[9] Inference can be obtained as a credible set of the posterior, for example using highest posterior density calculations of a given probability.

A second way to implement a Bayesian solution is through the use of a least squares approximation (Leng et al., 2014). Given assumptions 1–6.A, the likelihood function can be approximated by a Taylor expansion as

$$\ell_e(\boldsymbol{\theta}_e) \approx L(\widehat{\boldsymbol{\theta}}_e) + \frac{1}{2}(\boldsymbol{\theta}_e - \widehat{\boldsymbol{\theta}}_e)'\mathcal{I}(\widehat{\boldsymbol{\theta}}_e)(\boldsymbol{\theta}_e - \widehat{\boldsymbol{\theta}}_e)\,, \tag{25}$$

where $\widehat{\boldsymbol{\theta}}_e$ is the MLE in (16) for $e \in \{Y, Z\}$. Employing the same algorithm outlined previously with this expansions 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 1 and standard results for parametric Bayesian estimators, Bayes estimates $\check{\boldsymbol{\theta}}$ found from this algorithm are also consistent (Strasser, 1981, Bunke et al., 1998). For convenience, this is stated in the following theorem:

Theorem 3. *(i) Under assumptions 1, 2, 3.A or 3.B, 4, 5, 6.A, and regularity conditions,* $\check{\boldsymbol{\theta}}_e$*, 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 1, 2, 3.A or 3.B, 4, 5, 6.B, and regularity conditions, $\check{\boldsymbol{\theta}}_Y$ *as defined above, is consistent to the minimizer of the Kullback-Leibler divergence, i.e.* $\check{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^*_Y$*, where* $\boldsymbol{\theta}^*_Y = (\boldsymbol{\delta}'_0, \boldsymbol{\psi}^{*\prime})'$*.*

4 Monte Carlo Study

To test the performance of the estimator defined by (16) 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, changing the form of the conditional mean function. Data are simulated from several scenarios that maintain the conditional mean as correctly specified, as link function misspecification would be a source of bias distinct to likelihood misspecification (Montoya-Blandón and Jacho-Chávez, 2020). Numerical optimization on the log-likelihoods (14) and (15) 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 which assumes a Gaussian copula and beta marginals. That is,

[9] 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 analogues target a mode interpretation to their frequentist counterparts but use the posterior mean and median for simplicity.

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 parametrization, 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 B.

4.1 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 (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 $\boldsymbol{\beta}_2 = (-1.5, 0, 0.5)$, for $p = 2$ regressors, x_1 and x_2, generated independently from a standard normal distribution. For the first exercise, beta marginals with a mean-precision parametrization are used, setting $\phi_1 = \phi_2 = 10$. A Gaussian copula with correlation parameter of $\psi = 0.5$ links together the two free marginals. Values for $\boldsymbol{y}$ are generated by rejection sampling for sample sizes $n \in \{100, 200, 400, 800\}$ and 1000 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 parametrization 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, and given the assumption of common covariates across shares, 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 1 in terms of the root mean squared error (RMSE) across 1000 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 0.3 correlation in a Gaussian distribution. The results are presented in Table 2. Now, as expected from Theorem 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 B.2, the estimated copula 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. 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 B.2 in the appendix).

Moving away from sampling directly from a correctly specified copula likelihood, the next exercise in Table 3 draws observations from a Dirichlet distribution. As it is possible to maintain the conditional mean intact under this parametrization, 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 B.3 in the appendix, it is noticeable that the model does capture 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.

To produce a Bayesian estimator into this setting, the following setup is used: first, 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 1. A

Table 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	$\psi\|\xi$	AIC	BIC
					$n = 100$						
Copula Y	9.102	8.059	8.075	10.878	9.688	9.281	15.720	17.051	20.311	-403.53	-380.08
Copula Z	9.147	8.176	8.093	11.636	11.132	9.230	15.785	67.007	41.837	-338.24	-314.79
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	—	—	-346.14	-327.90
Logistic Norm.	18.874	17.116	11.402	38.984	17.394	29.083	—	—	—	592.84	608.47
					$n = 200$						
Copula Y	6.548	5.558	5.487	7.755	6.857	6.361	11.414	12.151	14.661	-816.65	-786.96
Copula Z	6.436	5.770	5.487	8.056	8.907	6.358	11.430	67.996	38.208	-684.58	-654.89
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	—	—	-699.10	-676.01
Logistic Norm.	17.289	14.892	7.840	37.735	12.972	26.862	—	—	—	1188.21	1208.00
					$n = 400$						
Copula Y	5.090	4.014	4.013	6.086	4.849	4.630	8.561	9.437	10.787	-1643.86	-1607.94
Copula Z	4.715	4.326	4.016	5.741	7.508	4.700	8.579	68.130	36.581	-1379.98	-1344.06
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	—	—	-1406.40	-1378.46
Logistic Norm.	16.612	14.004	5.827	37.208	10.065	25.642	—	—	—	2378.82	2402.77
					$n = 800$						
Copula Y	3.997	2.785	2.936	4.874	3.451	3.184	6.690	7.375	8.691	-3291.71	-3249.55
Copula Z	3.449	3.248	3.010	4.415	6.591	3.493	6.772	68.559	35.274	-2761.51	-2719.35
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	—	—	-2815.58	-2782.79
Logistic Norm.	16.108	13.297	4.343	36.877	8.306	24.878	—	—	—	4762.20	4790.30

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 different amount of parameters to be estimated. "—" implies the parameter is not part of the model and information criteria were not computed for the quasi-likelihood method.

Table 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	$\psi\|\xi$	AIC	BIC
					$n = 100$						
Copula Y	8.416	8.137	7.792	10.443	9.109	8.925	15.598	15.897	237.074	−380.02	−356.58
Copula Z	9.324	9.183	9.045	12.386	10.562	10.284	16.585	59.563	193.189	−314.77	−291.32
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	—	—	−351.85	−333.61
Logistic Norm.	18.548	17.507	10.874	38.331	15.572	29.477	—	—	—	604.69	620.33
					$n = 200$						
Copula Y	5.934	5.447	5.535	7.210	6.126	6.156	10.689	10.848	237.323	−768.82	−739.14
Copula Z	10.363	9.534	10.871	13.366	9.995	12.092	15.204	62.439	189.734	−626.38	−596.69
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	—	—	−710.24	−687.15
Logistic Norm.	17.103	15.503	7.875	37.330	11.396	26.971	—	—	—	1211.31	1231.10
					$n = 400$						
Copula Y	4.586	3.909	4.035	5.505	4.503	4.457	7.336	7.575	237.161	−1545.68	−1509.75
Copula Z	10.984	10.574	12.394	15.332	11.219	12.809	15.932	63.952	187.821	−1241.65	−1205.72
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	—	—	−1424.71	−1396.77
Logistic Norm.	16.377	14.317	5.529	36.753	9.137	25.857	—	—	—	2429.21	2453.16
					$n = 800$						
Copula Y	3.114	2.772	2.877	3.790	3.023	3.147	5.440	5.403	237.675	−3099.65	−3057.49
Copula Z	10.849	10.296	12.022	15.373	10.865	12.350	15.625	63.269	189.708	−2486.79	−2444.63
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	—	—	−2857.27	−2824.47
Logistic Norm.	15.952	13.854	4.033	36.597	7.408	25.327	—	—	—	4861.60	4889.70

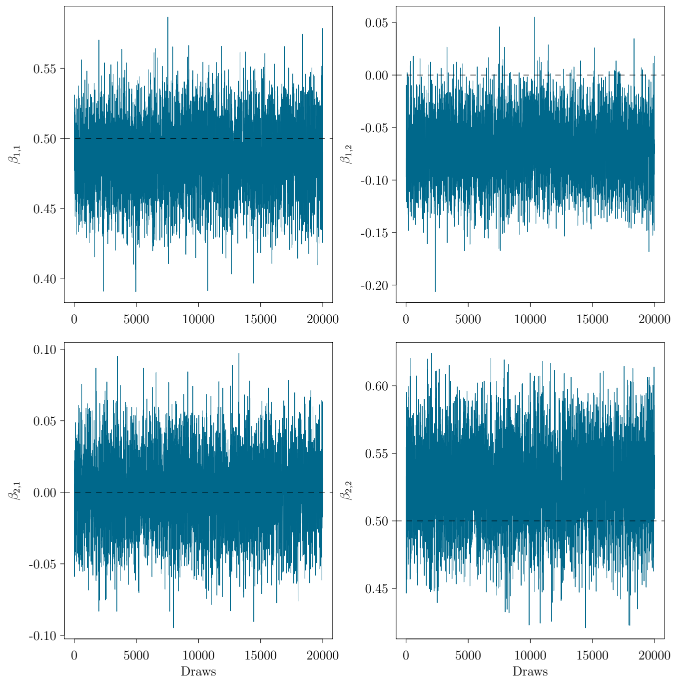
Note: 100 times RMSE for each estimation procedure when data are generated from a Farlie–Gumbel–Morgenstern copula with beta marginals. Akaike and Bayesian information criteria (AIC and BIC, respectively) computed as models have different amount of parameters to be estimated. "—" implies the parameter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

Table 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	−371.59	−348.14
Copula Z	7.662	7.722	7.296	9.158	8.645	8.372	14.459	−313.59	−290.15
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	−375.86	−357.63
Logistic Norm.	20.193	16.133	9.747	40.454	14.379	28.222	—	591.78	607.41
$n = 200$									
Copula Y	5.283	5.342	5.160	6.451	5.812	5.777	9.454	−753.55	−723.87
Copula Z	5.286	5.319	5.088	6.529	6.658	5.733	9.457	−637.18	−607.50
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	−760.38	−737.29
Logistic Norm.	19.236	14.433	7.067	39.945	11.069	25.979	—	1185.60	1205.39
$n = 400$									
Copula Y	3.685	3.741	3.608	4.680	4.209	4.059	7.011	−1517.52	−1481.59
Copula Z	3.684	3.761	3.569	4.738	5.283	4.055	7.012	−1284.79	−1248.86
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	−1528.83	−1500.89
Logistic Norm.	18.709	13.422	5.095	39.269	8.719	24.879	—	2370.38	2394.33
$n = 800$									
Copula Y	2.616	2.615	2.526	3.339	2.996	2.919	4.935	−3042.68	−3000.52
Copula Z	2.616	2.627	2.496	3.376	4.416	2.911	4.932	−2575.64	−2533.48
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	−3063.26	−3030.47
Logistic Norm.	18.254	13.065	3.736	38.840	7.459	24.328	—	4740.63	4768.74

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 different amount of parameters to be estimated. "—" implies the parameter is not part of the model and information criteria were not computed for the quasi-likelihood method.

Figure 2: Trace plot of Bayesian chains in reduced form model



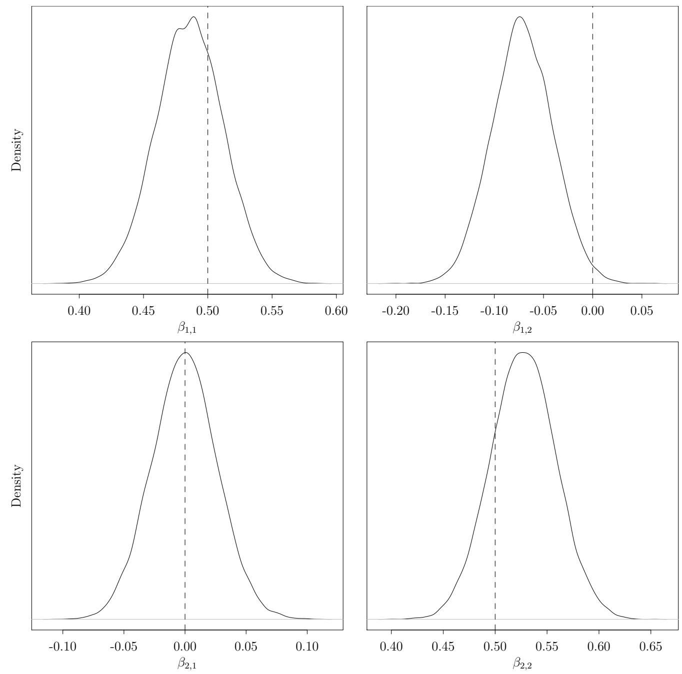
Note: Combination of 4 chains, each of 5000 draws. True value in dotted line.

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. Estimation uses the Hamiltonian Monte Carlo algorithm to sample from the posterior distribution in four chains 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 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 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 and 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.

Figure 3: Density plot of Bayesian chains in reduced form model



Note: Combination of 4 chains, each of 5000 draws. True value in dotted line.

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 $\partial \mathrm{E}[Y_j|\boldsymbol{X}=\boldsymbol{x}]/\partial x_k$ (see, e.g., Appendix 1 in Mullahy, 2015). For notational convenience this is written simply as $\mathrm{APE}_{k,j}$. 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 5. The computed APEs can be seen to be 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 4 and 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 the following optimization problem:

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

Obtaining solutions for different values of λ using the simulated data set, shows the effect of

Table 4: Bayesian and frequentist estimates for 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)

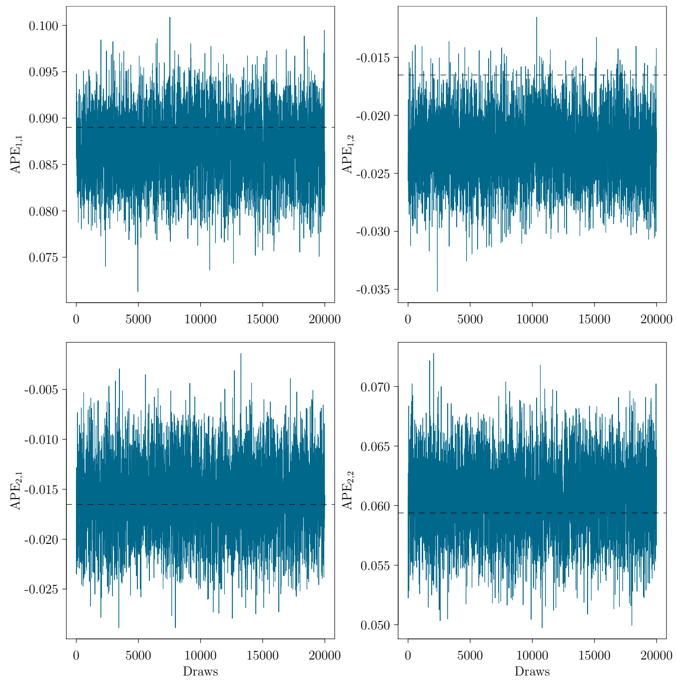
Note: Bayesian and MLE estimates from a Gaussian copula with beta marginals specification. Standard errors are in parentheses (standard deviations in each chain for Bayesian and asymptotic for MLE).

Table 5: Bayesian estimates and inference of APEs for reduced form model

Parameter	Chain 1	Chain 2	Chain 3	Chain 4	True
$\mathrm{APE}_{1,1}$	0.0866	0.0866	0.0866	0.0867	0.0890
	(0.0037)	(0.0038)	(0.0038)	(0.0038)	
$\mathrm{APE}_{2,1}$	-0.0159	-0.0158	-0.0161	-0.0160	-0.0165
	(0.0039)	(0.0039)	(0.0039)	(0.0039)	
$\mathrm{APE}_{1,2}$	-0.0229	-0.0229	-0.0229	-0.0228	-0.0165
	(0.0030)	(0.0030)	(0.0029)	(0.0030)	
$\mathrm{APE}_{2,2}$	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 4: Trace plot of APE chains in reduced form model



Note: Combination of 4 chains, each of 5000 draws. True value in dotted line.

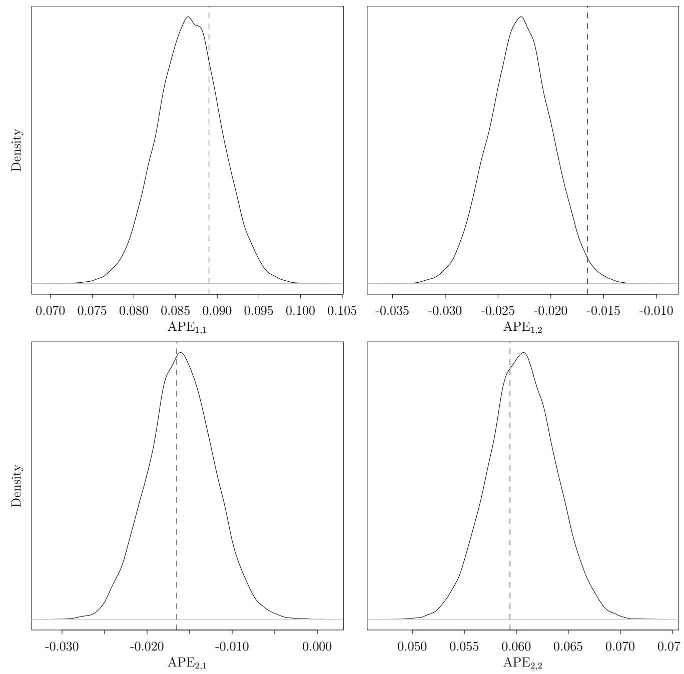
regularization. In the frequentist case, it operates as shown in Figure 6, where the parameters are moved towards 0 in absolute value and eventually set to 0 given a large enough penalty parameter λ. The coefficient $\beta_{2,1}$ does not appear in the picture as it is already estimated to be close to 0 even without regularization.

From a Bayesian perspective, to get a sense of the selection effect that the class of priors discussed in (23) can have, 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 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 redundant in the model. The following setup for priors allows for 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 B.8. As expected, these are shrunk towards 0, a consequence of the LASSO penalty encoded in the prior distributions. Table 6 shows the relevant selection aspects for these coefficients and APEs for each variable.

Figure 5: Density plot of APE chains in reduced form model



Note: Combination of 4 chains, each of 5000 draws. True value in dotted line.

While Bayesian selection is in general not sharp, other methods such as the credible interval or scaled neighborhood criteria can be used to select variables based on estimates from this specification (Li and Lin, 2010).[10] The credible interval method sets a coefficient $\beta_{k,j}$ to 0 if its credible interval 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. It computes 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., $\Pr[(-\mathrm{sd}(\beta_{k,j}), \mathrm{sd}(\beta_{k,j}))] > \bar{p}$ for some $\bar{p} \in (0, 1)$.

As can be seen in Table 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 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.

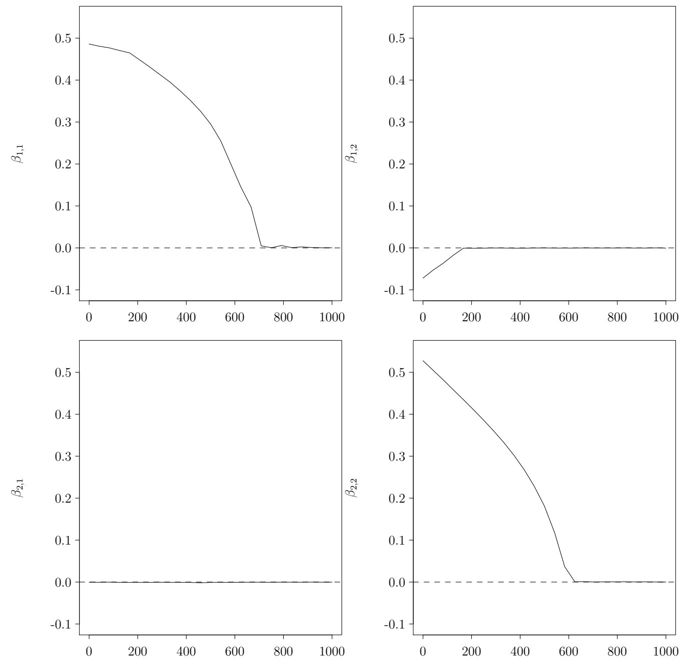
[10]Other attractive methods exist, which combine the frequentist and Bayesian properties of selection. See, e.g., the method in Leng et al. (2014) that does frequentist penalized regression with each λ sample in the chain and selects those variables which appear in 50% or more of the models.

Table 6: Bayesian APEs and selection for 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	0.080	✓	✓	✓	✓
			(0.004)	(0.004)				
x_2	−0.091	−0.091	−0.082	−0.076	✓	✓	✓	✓
			(0.004)	(0.004)				
x_3	0.091	0.091	0.083	0.081	✓	✓	✓	✓
			(0.004)	(0.004)				
x_4	−0.091	−0.091	−0.082	−0.084	✓	✓	✓	✓
			(0.004)	(0.004)				
x_5	0.091	0.091	0.081	0.081	✓	✓	✓	✓
			(0.004)	(0.004)				
x_6	0.000	0.000	−0.002	−0.003	✓	✓	×	×
			(0.003)	(0.003)				
x_7	0.000	0.000	−0.004	0.004	×	×	×	×
			(0.003)	(0.003)				
x_8	0.000	0.000	−0.002	0.000	×	×	×	×
			(0.003)	(0.003)				
x_9	0.000	0.000	−0.004	0.001	✓	×	×	×
			(0.003)	(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 (standard deviation of each chain) are in parentheses. CI y_j represents credible interval selection with $\bar{l} = 0.5$ and SN y_j the scaled neighborhood method with $\bar{p} = 0.5$, both for outcome $j = 1, 2$. "✓" indicates a variable is present in that outcome's equation, and "×" denotes its absence. Bayesian algorithm chooses a regularization parameter $\lambda = 1.79$.

Figure 6: Frequentist LASSO in a reduced form model with 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 1000.

4.2 Demand Estimation

To mimic some of the properties present in the empirical application of the next section, in this part of the paper an almost ideal demand system with $d = 3$ shares is simulated from (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 expenditure, as well as symmetry of the Slutsky matrix. In order to generate 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 were 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 1000 simulations and for now, the paper examines maximum likelihood estimation, 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 one 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 which 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 7 and 8. The main difference is in the generating marginal distributions; in the first betas with mean-precision parametrization are used, whereas the second set are normal distributions. The tables estimate four of the same models as before: copula on Y, copula on Z, multivariate fractional quasi-likelihood (it is no longer a logit as the conditional mean specification changes), and 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 copula on Y and Z estimators are consistent due to the correctly specified nature in Table 7. Both AIC and BIC select the copula on Y as estimator as preferable 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 is that it 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 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 towards 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 (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 9 and 10 present the results for this configuration. In general, the patterns observed in this iteration track the previous results very closely. Worth noting is 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, and even in a Bayesian setting, that it has strong performance comparative to the methods previously available in the literature.

5 Empirical Application

As a complement and extension to the numerical study undertaken in the previous section, this section puts in 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*

Table 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	−366.48	−337.82
Copula Z	46.744	7.030	2.838	1.757	0.873	0.907	0.539	0.334	3.404	12.660	4.058	−188.90	−160.25
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	—	—	−312.90	−289.46
AID	18.587	3.683	9.054	1.573	0.955	88.715	0.556	3.565	1.860	2.379	5.664	−337.88	−309.22
						$n = 200$							
Copula Y	3.056	2.231	1.400	1.043	0.573	0.613	0.413	0.246	3.006	4.174	4.562	−744.39	−708.11
Copula Z	6.650	2.216	1.515	1.105	0.590	0.641	0.403	0.269	3.001	12.808	3.768	−387.28	−351.00
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	—	—	−634.49	−604.80
AID	9.416	2.496	30.154	1.258	3.842	84.369	0.423	6.542	1.840	2.371	5.621	−686.82	−650.54
						$n = 400$							
Copula Y	3.731	1.746	1.184	0.732	0.406	0.456	0.313	0.196	2.854	3.981	4.443	−1502.88	−1458.98
Copula Z	10.029	1.858	1.349	0.827	0.429	0.517	0.322	0.235	2.907	12.837	3.657	−784.56	−740.65
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	—	—	−1279.32	−1243.39
AID	7.991	1.886	10.664	0.847	0.757	46.232	0.324	1.847	1.840	2.366	5.517	−1386.89	−1342.99
						$n = 800$							
Copula Y	3.137	1.480	1.053	0.523	0.286	0.326	0.251	0.164	2.775	3.873	4.373	−3016.67	−2965.14
Copula Z	8.271	1.592	1.309	0.713	0.351	0.428	0.263	0.222	2.874	12.827	3.542	−1564.87	−1513.34
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	—	—	−2564.31	−2522.15
AID	7.356	1.619	2.128	0.736	0.348	32.587	0.266	1.908	1.833	2.367	5.493	−2790.49	−2738.96

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 different amount of parameters to be estimated. For coefficients "—" implies the parameter is not part of the model. Information criteria not computed for the quasi-likelihood method.

Table 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	−184.02	−155.36
Copula Z	70.756	12.002	6.708	3.471	1.795	1.903	0.879	0.612	6.277	14.642	4.607	−49.54	−20.88
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	—	—	−199.36	−175.91
AID	31.751	7.744	3.333	2.779	1.528	1.745	0.834	0.625	1.252	1.881	15.166	−157.48	−128.83
						$n = 200$							
Copula Y	21.665	5.580	2.353	1.982	1.050	1.278	0.689	0.463	6.367	3.458	14.368	−379.25	−342.97
Copula Z	29.222	6.706	2.268	2.517	1.251	1.428	0.714	0.472	6.416	14.726	4.496	−103.71	−67.43
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	—	—	−408.39	−378.71
AID	9.861	4.129	2.030	1.762	1.046	1.304	0.668	0.483	1.234	1.868	15.106	−327.98	−291.70
						$n = 400$							
Copula Y	10.187	3.638	1.856	1.241	0.731	0.948	0.558	0.401	6.402	3.480	14.354	−771.00	−727.09
Copula Z	9.147	3.703	1.764	1.997	1.010	1.224	0.594	0.409	6.446	14.491	4.663	−182.70	−138.79
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	—	—	−827.60	−791.68
AID	11.977	3.749	1.940	1.208	0.757	0.970	0.546	0.413	1.226	1.861	15.119	−670.44	−626.53
						$n = 800$							
Copula Y	12.375	3.518	1.436	0.845	0.565	0.759	0.476	0.356	6.455	3.509	14.311	−1550.58	−1499.05
Copula Z	9.477	3.391	1.616	1.627	0.911	1.014	0.530	0.382	6.491	14.426	4.680	−350.51	−298.98
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	—	—	−1662.04	−1619.88
AID	5.521	3.262	1.606	0.836	0.564	0.764	0.472	0.366	1.220	1.857	15.083	−1350.60	−1299.07

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 different amount of parameters to be estimated. For coefficients "—" implies the parameter is not part of the model. Information criteria not computed for the quasi-likelihood method.

Table 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	−381.65	−342.57
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	−175.84	−136.76
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	—	—	−339.16	−305.29
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	−360.39	−321.32
									$n = 200$								
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	−775.97	−726.50
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	−362.70	−313.23
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	—	—	−687.30	−644.42
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	−733.85	−684.38
									$n = 400$								
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	−1562.92	−1503.05
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	−728.29	−668.42
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	—	—	−1381.89	−1330.00
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	−1477.53	−1417.65
									$n = 800$								
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	−3140.51	−3070.24
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	−1448.33	−1378.06
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	—	—	−2774.52	−2713.62
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	−2970.36	−2900.10

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 different amount of parameters to be estimated. For coefficients "—" implies the parameter is not part of the model. Information criteria not computed for the quasi-likelihood method.

Table 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	−197.19	−158.12
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	−16.60	22.48
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	—	—	−216.57	−182.70
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	−168.99	−129.91
$n = 200$																	
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	−405.58	−356.11
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	−27.22	22.26
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	—	—	−442.76	−399.88
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	−350.46	−300.98
$n = 400$																	
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	−822.34	−762.47
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	−195.38	−135.50
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	—	—	−894.60	−842.71
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	−714.21	−654.34
$n = 800$																	
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	−1655	−1584.73
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	−120.78	−50.51
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	—	—	−1799.33	−1738.43
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	−1443.92	−1373.65

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 different amount of parameters to be estimated. For coefficients "—" implies the parameter is not part of the model. Information criteria not computed for the quasi-likelihood method.

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 $1, \ldots, n$:

$$\mathrm{E}[\boldsymbol{Y}_i | e_i, \boldsymbol{p}_i] = \boldsymbol{\alpha} + \Gamma \boldsymbol{p}_i + \boldsymbol{\pi}[e_i - \alpha_0 - \boldsymbol{\alpha}' \boldsymbol{p}_i - (1/2) \boldsymbol{p}_i' \Gamma \boldsymbol{p}_i] \,. \tag{26}$$

Using the notation developed thus far, there are expenditure shares for $d = 3$ goods, where y_1 represents gasoline, y_2 local transportation, and y_3 is intercity transportation. The base category of analysis will be the same as used in CS, given by good 3. Prices of these goods are normalized with 2002 serving as the base. To rule out the effect of possible unobserved heterogeneity, CS assume 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 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 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 (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 again relying on the use of 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 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 (4) 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 even structural estimation in demand models. Table 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 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.[11] Standard errors are consistent with

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

Table 11: Summary statistics for data in Chang and Serletis (2014)

Variable	Good	Mean	Std. Dev.	Minimum	Maximum
Single member households, 2218 observations					
	Gasoline	0.499	0.237	0.002	0.986
Budget shares	Local transportation	0.095	0.128	0.001	0.856
	Intercity transportation	0.406	0.228	0.003	0.985
	Gasoline	1.157	0.269	0.726	1.751
Prices	Local transportation	1.038	0.131	0.801	1.307
	Intercity transportation	1.011	0.132	0.755	1.233
Expenditure		2,430.7	1,703.0	161	24,620
Married couples without children, 3326 observations					
	Gasoline	0.524	0.234	0.005	0.990
Budget shares	Local transportation	0.083	0.114	0.000	0.866
	Intercity transportation	0.392	0.224	0.003	0.985
	Gasoline	1.170	0.268	0.726	1.751
Prices	Local transportation	1.046	0.131	0.801	1.307
	Intercity transportation	1.017	0.132	0.755	1.233
Expenditure		3,920.5	2,396.7	170	26,230
Married couples with one child, 6141 observations					
	Gasoline	0.575	0.237	0.002	0.997
Budget shares	Local transportation	0.092	0.117	0.000	0.886
	Intercity transportation	0.333	0.229	0.002	0.980
	Gasoline	1.146	0.261	0.726	1.751
Prices	Local transportation	1.035	0.127	0.801	1.307
	Intercity transportation	1.005	0.130	0.755	1.233
Expenditure		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.

the magnitude and role of each parameter, and also closely resemble those previously found in the literature.

Table 12: MLE estimates of AID system using 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.	3352.683	3330.05	3352.679	5660.563	5635.622	5660.556	9734.453	9677.537	9734.422
Obs.		2218			3326			6141	

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

As a second exercise, 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 (26), is the scale of all parameters except $\boldsymbol{\pi}$. 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 since many combination of parameter values do not satisfy the positivity constraints. To this end, a re-parametrization 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 $e_{\text{new}} = \log e - \text{median}(\log e)$. In the AID system, this re-parametrization 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 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 re-parametrization, 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, which, similar to the frequentist case, encourage parameter values which always satisfy the fractional restrictions of multivariate fractional outcomes. Within this new parametrization, 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 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 were checked and pass the usual convergence diagnostics. As can be observed, the results remain similar to the maximum likelihood ones, when the re-parametrization is considered. The Bayesian standard errors tend to be narrower for the $\boldsymbol{\alpha}$ and Γ parameters, but slightly larger for the slopes $\boldsymbol{\pi}$, which become statistically insignificant in the first model. Figures 7 and 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 13.

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 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 parametrization necessary for Bayesian estimation, at the average median-centered expenditure. These elasticities are slightly larger 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

Table 13: Bayesian estimates of reparametrized AID system using Copula Y estimator with Gaussian copulas and beta marginals

Parameter	Single households	Married couples	Married with one child
α_0	0.651	0.697	0.928
	(0.354)	(0.368)	(0.369)
α_1	0.446	0.461	0.494
	(0.021)	(0.027)	(0.028)
α_2	0.086	0.069	0.076
	(0.009)	(0.009)	(0.008)
$\gamma_{1,1}$	−0.058	−0.073	−0.076
	(0.008)	(0.007)	(0.005)
$\gamma_{2,1}$	−0.022	−0.023	−0.020
	(0.003)	(0.002)	(0.002)
$\gamma_{2,2}$	0.050	0.034	0.005
	(0.031)	(0.027)	(0.022)
π_1	−0.004	−0.007	−0.017
	(0.014)	(0.010)	(0.008)
π_2	−0.017	0.045	0.045
	(0.032)	(0.025)	(0.021)
ϕ_1	3.563	3.725	3.503
	(0.093)	(0.081)	(0.056)
ϕ_2	7.339	7.890	7.386
	(0.244)	(0.207)	(0.149)
ψ	−0.388	−0.399	−0.362
	(0.018)	(0.015)	(0.011)
Obs.	2218	3326	6141

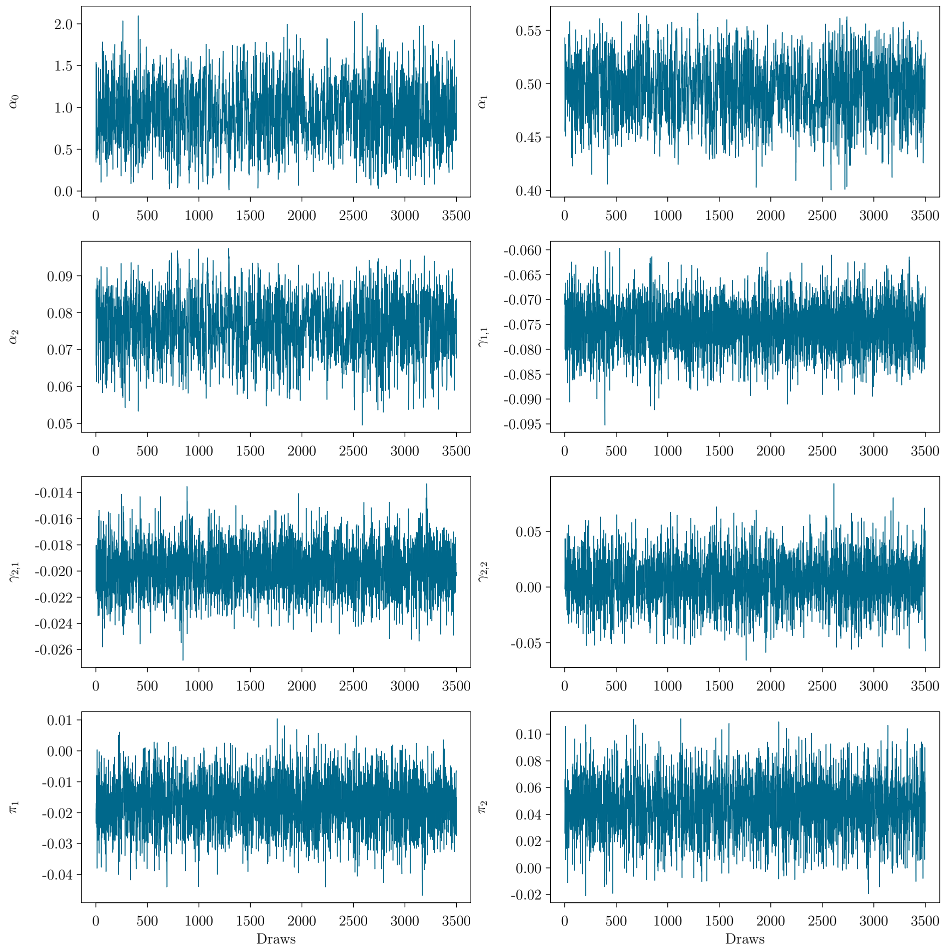
Note: Sample covers the period from 1997 to 2009. Intercity transportation is taken as the base category. Standard errors (standard deviation of the chains) in parentheses.

Table 14: Elasticity estimates and inference from extended Bayesian AID system

Good	Elasticities			
	Income	Price (1)	Price (2)	Price (3)
Single member households, 2218 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, 3326 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, 6141 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 computed at the average median-normed income and average prices for each chain. Point estimates are given by the mean. Standard deviation for the chains in parentheses.

Figure 7: Trace plot of coefficient chains in reparametrized AID system



Note: Results for data set on married couples with one children. Combination of 5 chains with 700 draws for a total of 3500.

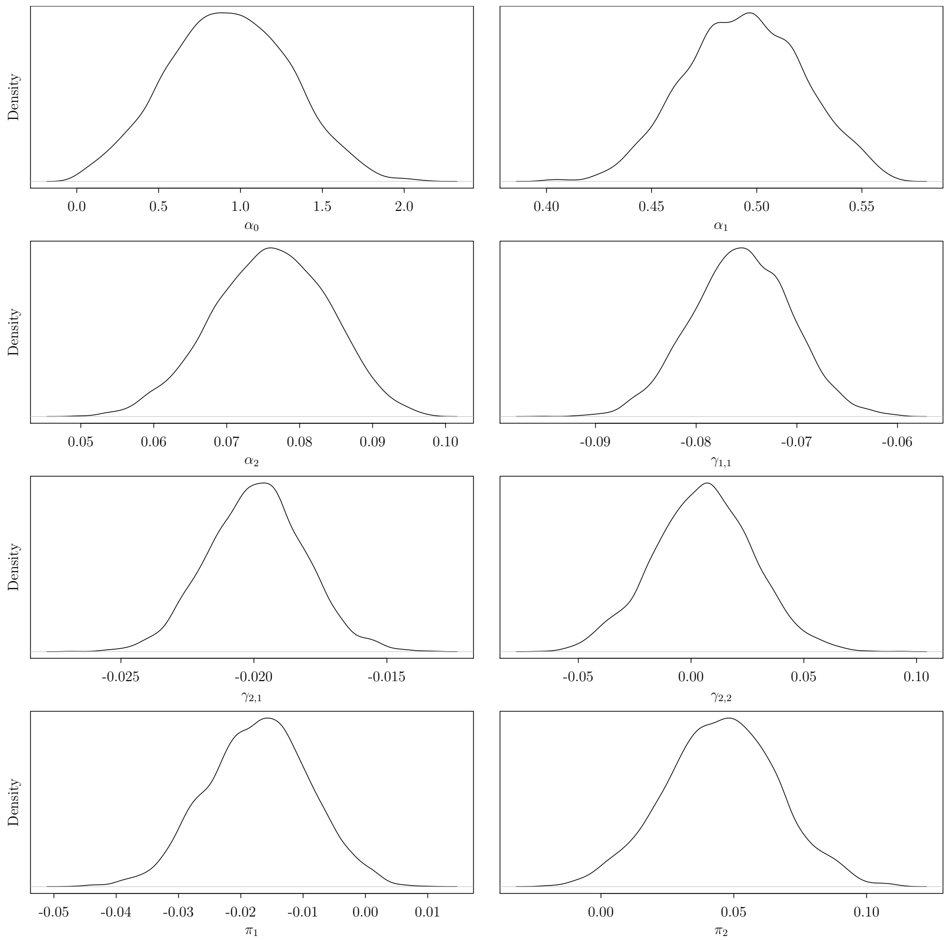
particular, the following conditional mean obtained in one of the examples is used:

$$\tilde{e}_i \equiv e_i - \alpha_0 - \boldsymbol{\alpha}' \boldsymbol{p}_i - (1/2)\boldsymbol{p}_i' \Gamma \boldsymbol{p}_i \,,$$

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

The re-parametrization 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 (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 degree 3 polynomial on the conditional mean equation for each share. To implement estimation and shrinkage of the coefficients using the LASSO penalty, the

Figure 8: Density plot of coefficient chains in reparametrized AID system



Note: Results for data set on married couples with one children. Combination of 5 chains with 700 draws for a total of 3500.

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 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 is the only instance in which the methods disagree with one another.

Table 15: Selection of polynomial terms in an extended Bayesian AID system

Polynomial	CI (1)	CI (2)	SN (1)	SN (2)
Single member households, 2218 observations				
$\tilde{e}$	✓	✓	✓	✓
$\tilde{e}^2$	×	✓	×	✓
$\tilde{e}^3$	✓	×	✓	×
Married couples without children, 3326 observations				
$\tilde{e}$	✓	✓	✓	✓
$\tilde{e}^2$	✓	×	✓	×
$\tilde{e}^3$	✓	×	✓	×
Married couples with one child, 6141 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. Bayesian algorithm chooses a regularization parameter $\lambda = 1.97$ for the first sample, and $\lambda = 1.95$ for the second and third.

Simultaneous to the selection step, estimation of the extended AID coefficients is straightforward. Table 14 presents the results for the income and price elasticities in this model, and 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 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 in 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 uncompensated, the possibility of these households to react to price variations might bear some correlation with income or other socioeconomic variables. 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 in CS. Figures 9 and 10 present the trace and density plots for these elasticities.

Table 16: Elasticity estimates and inference from Bayesian AID system

Good	Elasticities			
	Income	Price (1)	Price (2)	Price (3)
Single member households, 2218 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, 3326 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, 6141 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 computed at the average median-normed income and average prices for each chain. Point estimates are given by the mean. Standard deviation for the chains in parentheses.

6 Conclusions

The paper introduced several estimation procedures for multivariate fractional outcomes, that is 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 much more 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, the fractional and unit-sum restrictions in the outcomes, and allows for the inclusion of cross-equation restrictions. The latter point is 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 an 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 estimators 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 these elements could prove to be valuable in adding flexibility while preserving structure into different modeling problems. Additionally, it would be interesting to bring these tools to more applications in which multivariate fractional outcomes naturally arise. For example, in data for market shares on a given industry, portfolio shares in financial econometrics, industrial organization and firm analysis, among many others.

References

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.

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.

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.

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.

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

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.

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.

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.

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.

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.

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

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

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.

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

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

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.

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.

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.

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. 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.

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.

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.

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.

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

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.

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

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.

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. 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.

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.

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*.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., Bürkner, P.-C., et al. (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.

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.

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.

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.

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 (8). □

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 (12) will always be less than 1. To obtain a solution, first note that the inverse mapping for the stick-breaking transformation (10), $\boldsymbol{Y} = \boldsymbol{s}^{-1}(\boldsymbol{Z})$, is given by

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

Additionally, this mapping satisfies the following property:

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

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 (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

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

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

Proof of Theorem 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 taking 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 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 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 5, and 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 1. First, note that since P_X (the marginal distribution of $\boldsymbol{X}$) is given, we have

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

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:

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

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 (A.3), another expectation, this time with respect to P_X, gives the desired result. $\square$

Proof of Theorem 2. From Lemma 1, we can write the KL divergence as

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

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 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 T_3 term will be well-defined.

Now, based on assumptions 5 and 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 $\boldsymbol{\eta}^*$, 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 copula estimation (see also Joe, 2014, pp. 228). $\square$

Proof of Corollary 1. In this setting, similar to Theorem 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 to show consistency and asymptotic normality. $\square$

Proof of Theorem 3. (i) Note that the assumptions and regularity conditions for proving consistency and asymptotic normality in the previous theorems (as well as the prior specifications considered in the paper) 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 U containing $\boldsymbol{\theta}_{e,0}$,

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

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

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

(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, (A.4) holds for open sets U containing $\boldsymbol{\theta}^*_Y$ (see, e.g., Theorem 2.1 in Bunke et al., 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 et al., 1998). □

B Additional Numerical Exercises

Table B.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 B.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 B.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 B.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 B.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 B.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)	(0.8)	(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)	(19.92)	(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 B.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)	(15.48)	(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)	(0.1)			
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)	(3.3)	(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)	(0.9)	(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)

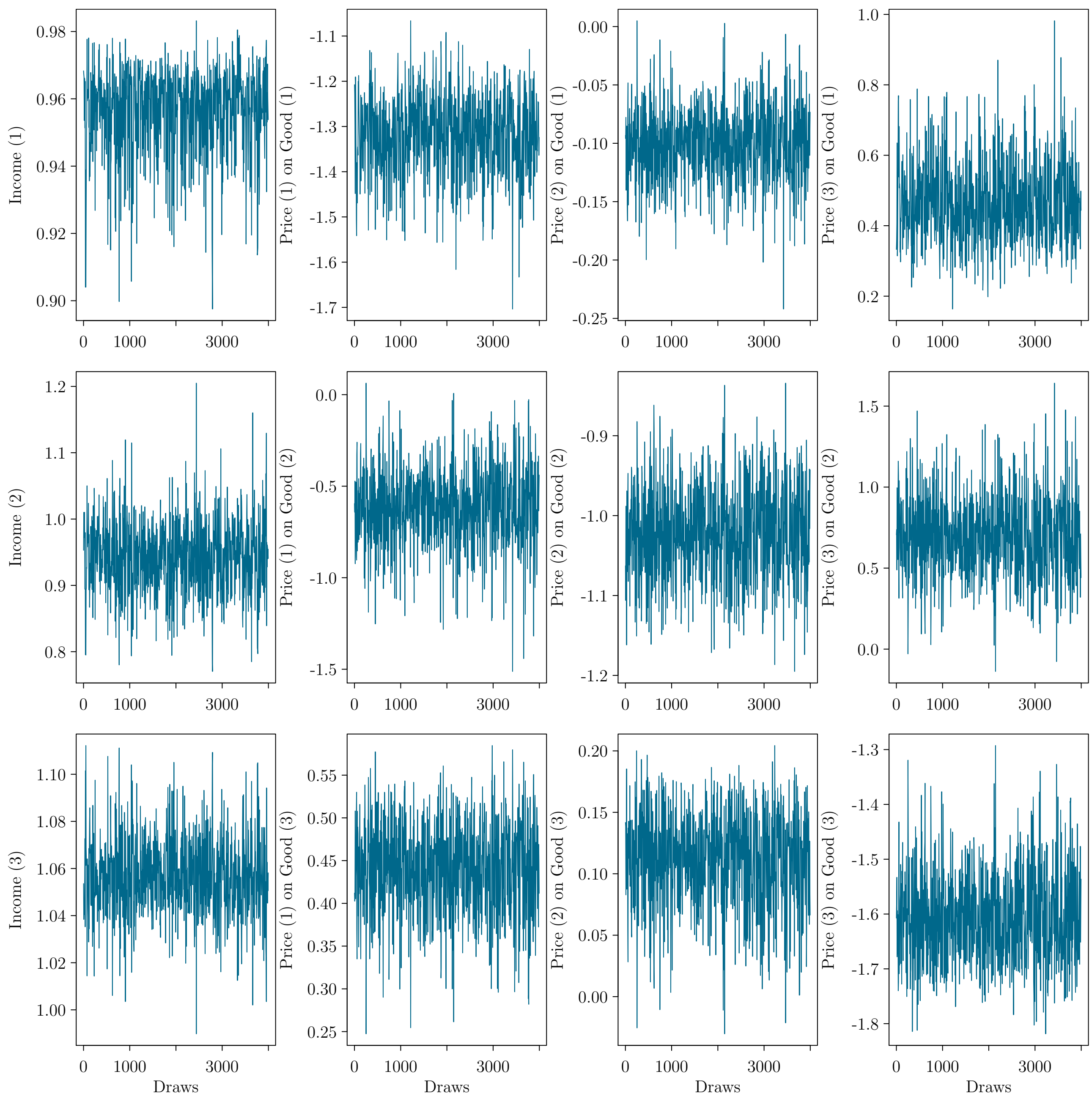
Note: MLE estimates and (copula misspecification robust) asymptotic standard errors for each estimation procedure. Data are generated from a multivariate Gaussian distribution. "—" implies the parameter is not part of the model.

Table B.8: Bayesian point estimates and inference for 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)

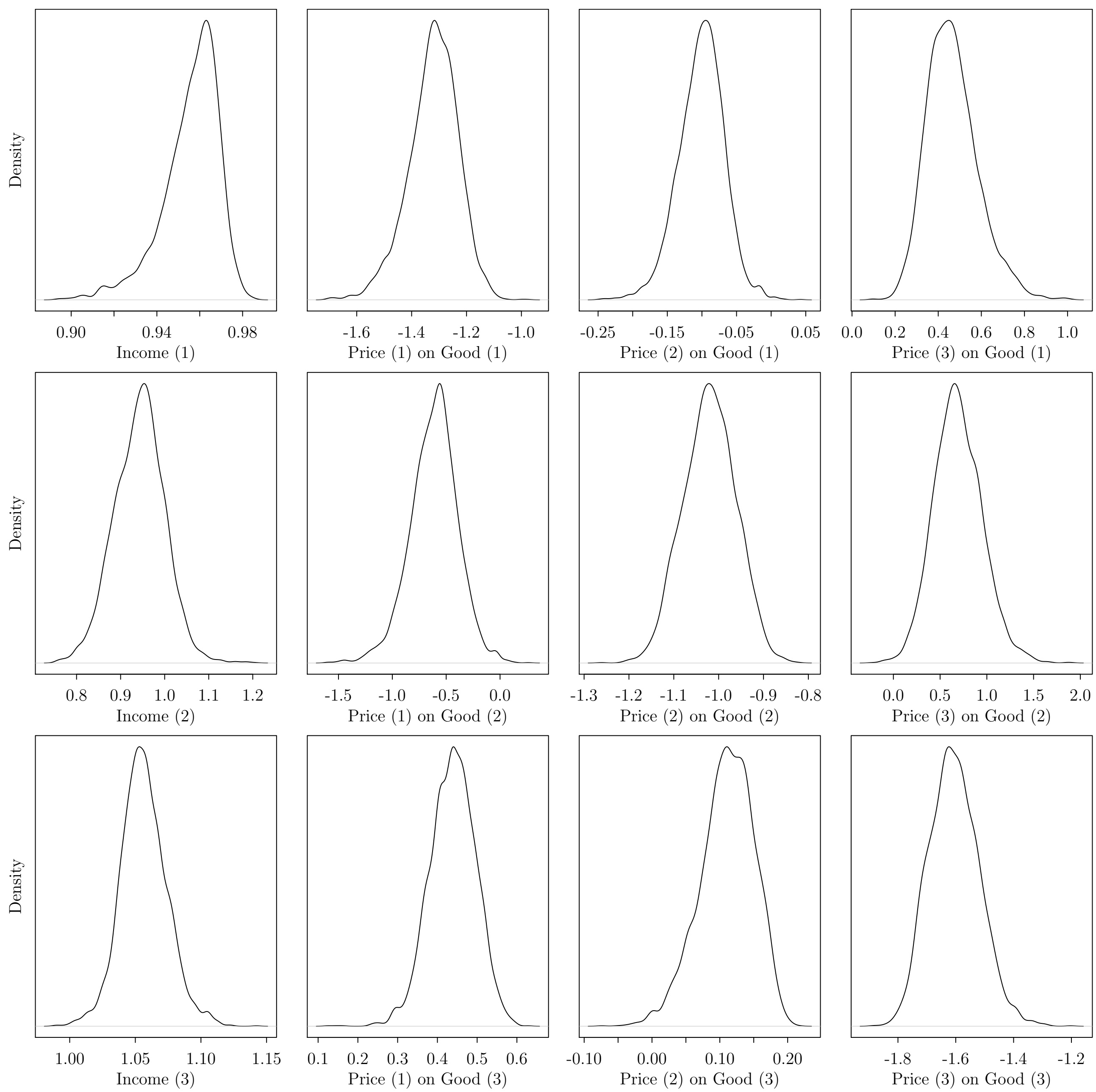
Note: Bayesian estimates from a Gaussian copula with beta marginals specification. Entries denote coefficient of the associated variable in each of the outcome equations. Standard errors (standard deviation of the chains) in parentheses.

Figure 9: Trace plot of elasticity chains in extended Bayesian AID system



Note: Results for data set on married couples with one children. Combination of 5 chains with 800 draws for a total of 4000.

Figure 10: Density plot of elasticity chains in extended Bayesian AID system



Note: Results for data set on married couples with one children. Combination of 5 chains with 800 draws for a total of 4000.