

Modelling correlated proportion data: Bayesian and classical approaches to inference and diagnostics

Lucas Machado Moschen^{1,2} and Luiz Max Carvalho¹

¹School of Applied Mathematics, Getulio Vargas Foundation, Rio de Janeiro, Brazil

²Department of Mathematics, Imperial College London, London, United Kingdom

August 28, 2026

Abstract

Correlated proportions appear in many real-world applications and present a unique challenge in finding an appropriate probabilistic model and performing inference due to their constrained nature. The bivariate beta is a natural extension of the well-known beta distribution to the space of correlated quantities on $[0, 1]^2$, but its construction is not unique. Over the years, many bivariate beta distributions have been proposed, ranging from three to eight or more parameters, and for which the joint density and distribution moments vary in terms of mathematical tractability. As a consequence, the literature on statistical inference for correlated proportion data is scattered. In this paper, we address this gap by thoroughly studying two rather different bivariate models from their construction to inference and diagnostics. We investigate two bivariate constructions namely the four-parameter bivariate Beta proposed by [Olkin and Trikalinos \(2015\)](#) (OT) and a bivariate logistic normal distribution, which has one extra parameter. We provide a thorough account of inference, exploring both classical (frequentist) and Bayesian approaches to estimation, utilising the method of moments and latent variable augmentation coupled with Hamiltonian Monte Carlo, respectively. The elicitation of a bivariate beta distribution as a prior is also discussed. Further, we develop diagnostics for checking model fit and adequacy and test their performance with Monte Carlo experiments under well-specified and misspecified data-generating settings. We illustrate these methods using data on childhood vaccination coverage and on sensitivity/specificity from COVID-19 tests, comparing the OT bivariate beta model with a bivariate logit-normal model.

Keywords: Bivariate beta; correlated proportions; logistic normal; Diagnostics; method of moments; Bayesian estimation.

1 Introduction

Correlated proportions appear in many real-world applications, such as modelling pollen distributions in forests ([Nadarajah et al., 2017](#)), the relationship between drought frequency and duration ([Nadarajah et al., 2007](#)), and the sensitivity and specificity of imperfect detection systems (e.g., disease tests ([Dahabreh et al., 2012](#))). The beta distribution is widely used for variables on $[0, 1]$, and extensions to $[0, 1]^2$ are natural, yielding the class of bivariate beta distributions, i.e., bivariate distributions with beta-distributed marginals. Many constructions are possible and vary with respect to the number of parameters, the attainable correlation structure, and the mathematical and computational tractability of the joint probability density function (pdf). See [Balakrishnan and Lai \(2009\)](#) and [Trick et al. \(2023\)](#) for reviews of continuous bivariate distributions. Examples include gamma-ratio constructions ([Libby and Novick, 1982](#); [Olkin and Liu, 2003](#); [Sarabia and Castillo, 2006](#)), flexible but intractable likelihoods ([Arnold and Ng, 2011](#)), and families with only positive correlations ([Magnussen, 2004](#); [Trick et al., 2023](#)); see also [Ali et al. \(1978\)](#); [Nadarajah and Kotz \(2005\)](#); [Nadarajah et al. \(2007, 2017\)](#); [Barros \(2015\)](#); [Crackel and Flegal \(2017\)](#).

A common way to introduce correlation between beta marginals is through shared gamma-distributed random variables. The multivariate construction of [Libby and Novick \(1982\)](#) with generalised beta marginals yields, in the bivariate case, a six-parameter model with closed-form joint density and series expressions for moments. A three-parameter special case was proposed in [Olkin and Liu \(2003\)](#) and extended by [Sarabia and Castillo \(2006\)](#), who studied several bivariate distributions with generalised beta marginals and argued that three parameters are needed to control the marginals' mean, variance, and skewness. An alternative by [Nadarajah et al. \(2007\)](#) keeps generalised beta marginals but lacks closed-form moments (simulation via rejection sampling), though the joint density is available up to normalisation and both maximum likelihood (MLE) and method of moments (MME) estimators were considered. The five-parameter family of [Arnold and Ng \(2011\)](#), built from independent gammas with

a common scale, allows correlations across the full range $(-1, 1)$, unlike earlier constructions, but lacks closed-form density and product moments, prompting a modified MLE that combines marginal MLE with MME for the correlation parameter. This approach has two drawbacks: moment-based estimates can become negative, and arbitrary beta marginals are not guaranteed; see also (Olkin and Trikalinos, 2015) on difficulties extending to higher dimensions.

More recent developments include a six-parameter model with elementary pdfs (Nadarajah et al., 2017), for which product moments are given by double infinite sums and empirical evidence suggests correlations in $[0, 1)$, and another six-parameter family from Trick et al. (2023) that accommodates arbitrary beta marginals but only positive correlations and lacks closed-form product moments; the latter also notes practical identifiability issues and proposes a constrained five-parameter variant. On the estimation side, Barros (2015) exploits that products of independent beta variables are beta distributed to obtain closed-form moments for the model of Nadarajah and Kotz (2005) (but without joint pdf and strictly positive correlation), while Crackel and Flegal (2017) employ approximate Bayesian computation (ABC) for Arnold and Ng (2011) and an eight-parameter extension due to intractable likelihoods. Overall, these works illustrate a trade-off between flexibility (number of parameters, correlation range) and tractability (closed-form density/moments). In what follows we detail the construction of Olkin and Trikalinos (2015), which we argue strikes a good balance between tractability and parameter richness.

The bivariate distribution described by Olkin and Trikalinos (2015) has beta-distributed marginals, support $[0, 1]^2$, and correlation over the full range $(-1, 1)$. It is constructed as follows: let $\mathbf{U} \sim \text{Dirichlet}(\boldsymbol{\alpha})$, with $\mathbf{U} = (U_1, U_2, U_3, U_4)$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$, where $\alpha_i > 0$ for $i = 1, \dots, 4$ and $U_4 = 1 - U_1 - U_2 - U_3$. The joint density of $\mathbf{U}$ with respect to the Lebesgue measure on $\mathbb{R}^3$ is given by

$$f_U(u_1, u_2, u_3) = \frac{1}{B(\boldsymbol{\alpha})} u_1^{\alpha_1-1} u_2^{\alpha_2-1} u_3^{\alpha_3-1} (1 - u_1 - u_2 - u_3)^{\alpha_4-1}, \quad (1)$$

for $u_i \in [0, 1]$ ($i = 1, 2, 3$) with $u_1 + u_2 + u_3 \leq 1$, and $f_U(u_1, u_2, u_3) = 0$ otherwise. The normalising constant is the multivariate beta function,

$$B(\mathbf{v}) = \frac{\prod_{i=1}^n \Gamma(v_i)}{\Gamma(\sum_{i=1}^n v_i)},$$

for $\mathbf{v} \in \mathbb{R}^n$. Let $X = U_1 + U_2$ and $Y = U_1 + U_3$. In Section 2 we show that the marginal distributions of X and Y are beta, and thus (X, Y) has a bivariate beta distribution with parameter $\boldsymbol{\alpha}$. This yields a four-parameter construction with the correlation between the margins covering the full range and with a non-explicit integral expression for the joint density, but with closed-form moments. Moreover, simulation is straightforward via the Dirichlet representation, which facilitates inferential tasks such as prediction.

Another route for modelling correlated proportions is through copulas by specifying marginals with beta-type distributions, and then the dependence structure via copula models (Joe, 1997; Genest and Rivest, 1993; Nelsen, 2006; Genest et al., 2009). A comparison with six copula families for the two applications studied in the paper is reported in Appendix B. These models are flexible and well-studied, but they form a broad family whose properties depend on the chosen copula. In particular, Pearson correlations and product moments are usually not available in closed form, and model assessment involves a proper selection of the copula family. Thus, we take a parametric approach.

In this paper, we focus on the parameter estimation for the bivariate beta distribution of Olkin and Trikalinos (2015), following both frequentist (method of moments) and Bayesian approaches (latent variable representation coupled with Hamiltonian Monte Carlo). We also address an important but often neglected aspect of inference, which is the development of diagnostics for assessing model fit and adequacy. Given a sample $\mathbf{Z} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$, we (i) derive exact method of moments solutions when they exist, characterise regions where the moments do not correspond to a well-defined bivariate beta, and organise two classes of moment-based estimators; (ii) develop a Bayesian estimator by introducing a latent-variable formulation that circumvents the intractable likelihood and enables efficient Hamiltonian Monte Carlo; (iii) propose custom diagnostics to detect model-data mismatches under both classical and Bayesian inference; and (iv) illustrate over simulated and real-data scenarios.

2 Bivariate beta model specification

Let (X, Y) follow the bivariate beta distribution of Olkin and Trikalinos (2015) with parameter $\boldsymbol{\alpha}$ via $\mathbf{U} \sim \text{Dirichlet}(\boldsymbol{\alpha})$. By the aggregation property (Ng et al., 2011, Section 2.2), we have

$$X \sim \text{Beta}(\alpha_1 + \alpha_2, \alpha_3 + \alpha_4), \quad Y \sim \text{Beta}(\alpha_1 + \alpha_3, \alpha_2 + \alpha_4).$$

The joint density admits the one-dimensional integral representation

$$f_{X,Y}(x,y) = \frac{1}{B(\boldsymbol{\alpha})} \int_{\max(0,x+y-1)}^{\min(x,y)} u^{\alpha_1-1} (x-u)^{\alpha_2-1} (y-u)^{\alpha_3-1} (1-x-y+u)^{\alpha_4-1} du. \quad (2)$$

See also (Olkin and Trikalinos, 2015, Appendix A) for an expression using special functions. From here on we work with the moments.

Denote $s_{\boldsymbol{\alpha}} = \sum_{i=1}^4 \alpha_i$. Straightforward calculations give

$$\mathbb{E}[X] = \frac{\alpha_1 + \alpha_2}{s_{\boldsymbol{\alpha}}}, \quad \mathbb{E}[Y] = \frac{\alpha_1 + \alpha_3}{s_{\boldsymbol{\alpha}}}, \quad \text{Var}(X) = \frac{(\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4)}{s_{\boldsymbol{\alpha}}^2(s_{\boldsymbol{\alpha}} + 1)}, \quad \text{Var}(Y) = \frac{(\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4)}{s_{\boldsymbol{\alpha}}^2(s_{\boldsymbol{\alpha}} + 1)},$$

and, using (Ng et al., 2011, p. 39),

$$\text{Cov}(X, Y) = \frac{1}{s_{\boldsymbol{\alpha}}^2(s_{\boldsymbol{\alpha}} + 1)} (\alpha_1 \alpha_4 - \alpha_2 \alpha_3).$$

Proposition 1 (Correlation and attainable range). *The correlation between X and Y is*

$$\text{Cor}(X, Y) = \frac{\alpha_1 \alpha_4 - \alpha_2 \alpha_3}{\sqrt{(\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4)(\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4)}}, \quad (3)$$

and it spans the full range $(-1, 1)$.

See [proof](#) in [Appendix A](#).

Remark 1 (Gamma representation). *If $Y_i \sim \text{Gamma}(\alpha_i, 1)$ for $i = 1, \dots, 4$, then $X = (Y_1 + Y_2) / \sum_i Y_i$, $Y = (Y_1 + Y_3) / \sum_i Y_i$ follow the above bivariate beta; cf. (Ng et al., 2011, Theorem 2.1). This is a limit case of (Arnold and Ng, 2011, Section 6.1) when $\delta_i \rightarrow 0$ for $i = 1, \dots, 4$.*

Denote $m_1 = \mathbb{E}[X]$, $m_2 = \mathbb{E}[Y]$, $v_1 = \text{Var}(X)$, $v_2 = \text{Var}(Y)$ and $\rho = \text{Cor}(X, Y)$. Since $X \in (0, 1)$ almost surely, we have $\mathbb{E}[X^2] < \mathbb{E}[X]$ and hence $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 < \mathbb{E}[X](1 - \mathbb{E}[X])$, and similarly for Y . Thus

$$v_i < m_i(1 - m_i) \implies \frac{m_i(1 - m_i)}{v_i} - 1 > 0, \quad i = 1, 2. \quad (4)$$

The moments satisfy the system

$$\begin{cases} m_1 = \frac{\alpha_1 + \alpha_2}{s_{\boldsymbol{\alpha}}}, \\ m_2 = \frac{\alpha_1 + \alpha_3}{s_{\boldsymbol{\alpha}}}, \\ v_1 = \frac{(\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4)}{s_{\boldsymbol{\alpha}}^2(s_{\boldsymbol{\alpha}} + 1)}, \\ v_2 = \frac{(\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4)}{s_{\boldsymbol{\alpha}}^2(s_{\boldsymbol{\alpha}} + 1)}, \\ \rho = \frac{\alpha_1 \alpha_4 - \alpha_2 \alpha_3}{\sqrt{(\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4)(\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4)}}. \end{cases} \quad (5)$$

The following proposition clarifies when the moments' system in (5) has a solution for $\boldsymbol{\alpha}$. In particular, it delineates when we cannot define a bivariate beta following Olkin and Trikalinos (2015)'s approach from the moments, which is an important step for model identification.

Proposition 2 (Solution to the moments' system). *System (5) has no solution unless*

$$\frac{m_1(1 - m_1)}{v_1} = \frac{m_2(1 - m_2)}{v_2}. \quad (6)$$

When (6) holds, the unique solution is

$$\begin{aligned} \alpha_1 &= (m_1 + m_2 - 1)s_{\boldsymbol{\alpha}} + \alpha_4, \\ \alpha_2 &= (1 - m_2)s_{\boldsymbol{\alpha}} - \alpha_4, \\ \alpha_3 &= (1 - m_1)s_{\boldsymbol{\alpha}} - \alpha_4, \\ \alpha_4 &= s_{\boldsymbol{\alpha}} \left(\rho \sqrt{m_1 m_2 (1 - m_1)(1 - m_2)} + (1 - m_1)(1 - m_2) \right), \end{aligned} \quad (7)$$

where

$$s_{\alpha} = \frac{m_1(1-m_1)}{v_1} - 1. \quad (8)$$

See [proof](#) in [Appendix A](#).

Notice that the first two equations of the system in (5) imply the relations

$$\alpha_1 = \frac{m_1 + m_2 - 1}{1 - m_1} \alpha_3 + \frac{m_2}{1 - m_1} \alpha_4 \quad (9)$$

$$\alpha_2 = \frac{1 - m_2}{1 - m_1} \alpha_3 + \frac{m_1 - m_2}{1 - m_1} \alpha_4, \quad (10)$$

which are used throughout the text.

The constraint $\alpha_i > 0$ further restricts the feasible correlation for given (m_1, m_2) to

$$\rho \in \left(-\frac{\min(m_1 m_2, (1-m_1)(1-m_2))}{\sqrt{m_1 m_2 (1-m_1)(1-m_2)}}, \frac{\min(m_1, m_2) - m_1 m_2}{\sqrt{m_1 m_2 (1-m_1)(1-m_2)}} \right) \quad (11)$$

from the expression for α_4 in (7). For example, when $m_1 = m_2 = m$, the upper bound is $\rho_{\max} = 1$ and the lower bound is

$$\rho_{\min} = -\frac{\min(m, 1-m)}{\max(m, 1-m)}.$$

For fixed (v_1, ρ) , define the feasible mean region

$$\mathcal{R}(v_1, \rho) = \left\{ (m_1, m_2) \in (0, 1)^2 : s_{\alpha} = \frac{m_1(1-m_1)}{v_1} - 1 > 0 \text{ and } \alpha_i(m_1, m_2, v_1, \rho) > 0 \text{ for } i = 1, \dots, 4 \right\},$$

where $\alpha_i(\cdot)$ are given by (7) and v_2 is implied by the compatibility condition $v_2 = m_2(1-m_2)/(s_{\alpha} + 1)$. [Figure 1](#) illustrates $\mathcal{R}(v_1, \rho)$ for several values of v_1 and ρ . When $|\rho|$ approaches 1, the feasible set in (11) shrinks considerably, as the blue regions in [Figure 1](#) illustrate. Thus, extreme correlations can only coexist with a narrow range of mean configurations. Once the means are fixed, the shared concentration s_{α} further restricts the marginal variances, as discussed in [Remark 2](#).

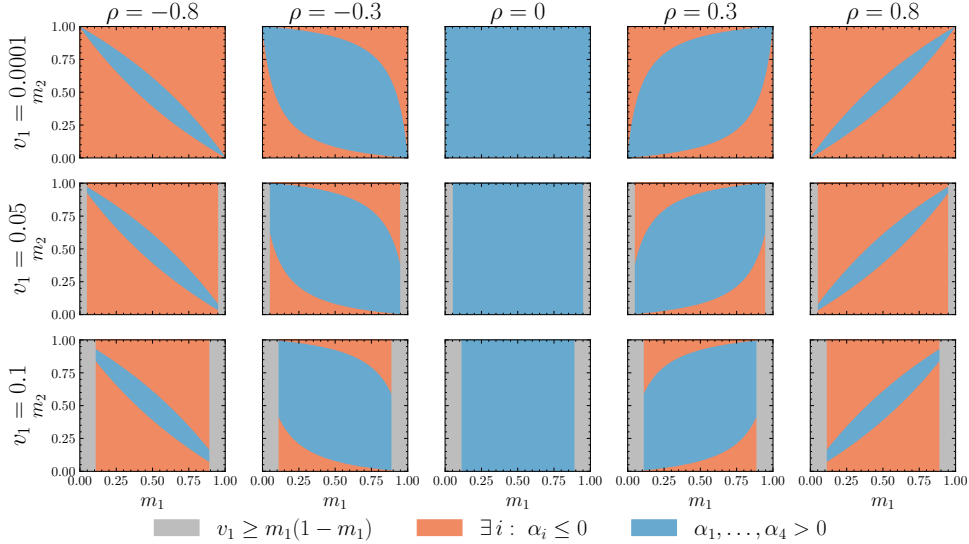


Figure 1: Positive solution for the moments' system: Representation of the sets $\mathcal{R}(v_1, \rho)$ in blue for different values of v_1 and ρ . In orange, we denote the regions where some α_i is negative in solution (7). In grey, we highlight the region where the variance and mean fail the beta-marginal constraint (4).

A slight modification of [Proposition 2](#) solves the moments' system without the equations related to v_1 and v_2 , culminating in the following proposition:

Proposition 3 (System of three moments). *Considering the system in (5) without the equations of v_1 and v_2 , the solution is*

$$\begin{aligned}\alpha_1 &= \alpha_4 \frac{m_1 m_2 + \rho \sqrt{m_1 m_2 (1 - m_1)(1 - m_2)}}{(1 - m_1)(1 - m_2) + \rho \sqrt{m_1 m_2 (1 - m_1)(1 - m_2)}}, \\ \alpha_2 &= \alpha_4 \frac{m_1(1 - m_2) - \rho \sqrt{m_1 m_2 (1 - m_1)(1 - m_2)}}{(1 - m_1)(1 - m_2) + \rho \sqrt{m_1 m_2 (1 - m_1)(1 - m_2)}}, \\ \alpha_3 &= \alpha_4 \frac{m_2(1 - m_1) - \rho \sqrt{m_1 m_2 (1 - m_1)(1 - m_2)}}{(1 - m_1)(1 - m_2) + \rho \sqrt{m_1 m_2 (1 - m_1)(1 - m_2)}},\end{aligned}\tag{12}$$

with $\alpha_4 > 0$ being a free parameter.

The proof of this proposition uses the relations (9) and (10).

Remark 2 (Link between the two marginal variances). *Under the OT bivariate beta model, both marginals share the same concentration s_α . Hence*

$$v_1 = \frac{m_1(1 - m_1)}{s_\alpha + 1}, \quad v_2 = \frac{m_2(1 - m_2)}{s_\alpha + 1},$$

so in particular

$$s_\alpha + 1 = \frac{m_1(1 - m_1)}{v_1} = \frac{m_2(1 - m_2)}{v_2}.$$

Condition (6) is therefore a structural feature of the OT bivariate beta distribution. Because both marginals share the same concentration s_α , once (m_1, m_2) are fixed, the two marginal variances cannot vary independently. This is reasonable when both proportions arise from comparable sample sizes or information levels, but less so when one margin is much more concentrated than the other. In such cases, the OT model might be too restrictive, and alternative bivariate beta constructions (e.g., using copulas) may be more appropriate. The gain in tractability thus comes at the cost of reduced flexibility, and this restriction should be verified in applications, which will be discussed in [Section 5.3](#).

3 Parameter estimation

We now develop estimators for the parameter α from data $\mathbf{Z} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$, with $\mathbf{Z} \stackrel{iid}{\sim} \text{BivariateBeta}(\alpha)$. We present a method of moments approach and a Bayesian approach.

3.1 Method of moments

Let

$$\begin{aligned}\hat{m}_1 &= \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{m}_2 = \frac{1}{n} \sum_{i=1}^n Y_i, \quad \hat{\rho} = \frac{\sum_{i=1}^n (X_i - \hat{m}_1)(Y_i - \hat{m}_2)}{n \sqrt{\hat{v}_1 \hat{v}_2}} \\ \hat{v}_1 &= \frac{1}{n} \sum_{i=1}^n (X_i - \hat{m}_1)^2 \quad \text{and} \quad \hat{v}_2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{m}_2)^2.\end{aligned}\tag{13}$$

Define $\hat{s}_\alpha = \hat{m}_1(1 - \hat{m}_1)/\hat{v}_1 - 1$ which is positive almost surely under continuous models over $[0, 1]$, since

$$\hat{m}_1(1 - \hat{m}_1) - \hat{v}_1 = \hat{m}_1 - \hat{m}_1^2 - \frac{1}{n} \sum_{i=1}^n X_i^2 + \hat{m}_1^2 = \frac{1}{n} \sum_{i=1}^n X_i(1 - X_i) \geq 0,\tag{14}$$

with equality only holding if $X_i \in \{0, 1\}$ for every i . The method of moments proceeds by matching the theoretical to the observed moments to find a solution for α . To quantify the uncertainty in our estimates, we employ a simple non-parametric bootstrap method ([Efron, 1979](#)), with $B = 500$ resamples over the pairs of observations.

MoM-Closed: closed-form inversion (four equations)

Let $L(\cdot)$ and $U(\cdot)$ be the feasible bounds in (11). Set the projected correlation

$$\tilde{\rho} = \min\{U(\hat{m}_1, \hat{m}_2) - \nu, \max\{L(\hat{m}_1, \hat{m}_2) + \nu, \hat{\rho}\}\}$$

for a fixed small $\nu > 0$. Using (7) from Section 2 with $(m_1, m_2, v_1, \rho) = (\hat{m}_1, \hat{m}_2, \hat{v}_1, \tilde{\rho})$ define

$$\begin{aligned}\hat{\alpha}_4 &= \hat{s}_\alpha \left(\tilde{\rho} \sqrt{\hat{m}_1 \hat{m}_2 (1 - \hat{m}_1)(1 - \hat{m}_2)} + (1 - \hat{m}_1)(1 - \hat{m}_2) \right) \\ \hat{\alpha}_1 &= (\hat{m}_1 + \hat{m}_2 - 1) \hat{s}_\alpha + \hat{\alpha}_4, \\ \hat{\alpha}_2 &= (1 - \hat{m}_2) \hat{s}_\alpha - \hat{\alpha}_4, \\ \hat{\alpha}_3 &= (1 - \hat{m}_1) \hat{s}_\alpha - \hat{\alpha}_4.\end{aligned}\tag{15}$$

These are strictly positive by the lemma below. In addition,

$$\widehat{\text{Var}}(Y) = \hat{v}_1 \cdot \frac{\hat{m}_2(1 - \hat{m}_2)}{\hat{m}_1(1 - \hat{m}_1)},$$

is the implied variance of Y since the estimate does not use its sample variance.

Lemma 1 (Positivity of MoM-Closed). *The solution to (15) yields $\hat{\alpha}_i > 0$ for $i = 1, \dots, 4$.*

Proof. From (14), $\hat{s}_\alpha > 0$ almost surely. Projecting $\hat{\rho}$ to $\tilde{\rho} \in (L, U)$ ensures each term in (15) is positive by the bounds derived in (11). Hence $\hat{\alpha}_i > 0$ for $i = 1, \dots, 4$. $\square$

MoM-NLS: joint moment matching

We also consider the nonlinear least squares estimator defined by Olkin and Trikalinos (2015), which seeks to minimise the following expression:

$$\begin{aligned}\left(\hat{m}_1 - \frac{\alpha_1 + \alpha_2}{s_\alpha} \right)^2 &+ \left(\hat{m}_2 - \frac{\alpha_1 + \alpha_3}{s_\alpha} \right)^2 + \left(\hat{\rho} - \frac{\alpha_1 \alpha_4 - \alpha_2 \alpha_3}{\sqrt{(\alpha_1 + \alpha_2)(\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4)(\alpha_3 + \alpha_4)}} \right)^2 \\ &+ \left(\hat{v}_1 - \frac{(\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4)}{s_\alpha^2(s_\alpha + 1)} \right)^2 + \left(\hat{v}_2 - \frac{(\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4)}{s_\alpha^2(s_\alpha + 1)} \right)^2,\end{aligned}\tag{16}$$

subject to $\alpha_1, \alpha_2, \alpha_3, \alpha_4 > 0$. Moreover, Olkin and Trikalinos (2015) consider the additional restriction

$$s_\alpha \leq \max \left(\frac{\hat{m}_1(1 - \hat{m}_1)}{\hat{v}_1}, \frac{\hat{m}_2(1 - \hat{m}_2)}{\hat{v}_2} \right) - 1,\tag{17}$$

based on Remark 2 to avoid over-concentrating the marginals. The solution $\hat{\alpha}$ to the minimisation problem (16) is the MoM-NLS estimator. As a warm initialisation for the optimisation problem, we consider the MoM-Closed estimate. We implicitly defined the quadratic loss for the minimisation strategy since it is differentiable and computationally inexpensive. Nevertheless, other losses can be used, such as the absolute, the absolute percentage, or the weighted quadratic, depending on the application.

Remark 3. *We experimented with two intermediate moment estimators:*

- (i) *a three-equation variant based on Proposition 3 where α_4 is chosen to approximate both variances using a nonlinear least square estimator; and*
- (ii) *a two-equation variant in which (α_1, α_2) are substituted out via (9) and (10), and (α_3, α_4) are then estimated by constrained nonlinear least squares matching the two marginal variances and the correlation.*

In the simulations, these two variants did not offer a consistent accuracy gain over MoM-Closed or MoM-NLS. They were both slower than MoM-Closed and faster than MoM-NLS. To keep the paper focused, we do not use them here.

Consistency of the method of moments estimators

Both moment estimators are consistent under the bivariate beta. We outline the argument. Let $\mathbf{Z} = \{(X_i, Y_i)\}_{i=1}^n$ be i.i.d. random variables with $(X_i, Y_i) \sim \text{BivariateBeta}(\boldsymbol{\alpha}_0)$ for some true parameter $\boldsymbol{\alpha}_0$. Denote by

$$\mu(\boldsymbol{\alpha}) = (m_1(\boldsymbol{\alpha}), m_2(\boldsymbol{\alpha}), v_1(\boldsymbol{\alpha}), v_2(\boldsymbol{\alpha}), \rho(\boldsymbol{\alpha}))$$

the vector of theoretical moments defined in (5), and by

$$\hat{\mu}_n = (\hat{m}_1, \hat{m}_2, \hat{v}_1, \hat{v}_2, \hat{\rho})$$

their empirical counterparts given in (13). The correlation projection is done with $\nu_n = \nu/n$ for some $\nu > 0$.

By the strong law of large numbers, as $n \rightarrow \infty$,

$$\hat{\mu}_n \xrightarrow{\text{a.s.}} \mu(\boldsymbol{\alpha}_0),$$

so that the empirical moment vector is a consistent estimator for the true moment vector.

Proposition 2 establishes that if the compatibility condition (6) holds, the moment equations have a unique solution $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$. In other words, the mapping

$$\psi : \boldsymbol{\alpha} \mapsto (m_1(\boldsymbol{\alpha}), m_2(\boldsymbol{\alpha}), v_1(\boldsymbol{\alpha}), \rho(\boldsymbol{\alpha}))$$

is injective on the admissible parameter space. This guarantees that the true parameter is identified by the moments. By the continuous mapping theorem, since $\nu_n \rightarrow 0$, for MoM-Closed we obtain

$$\hat{\boldsymbol{\alpha}}_{\text{MoM-Closed}} \xrightarrow{\text{a.s.}} \boldsymbol{\alpha}_0.$$

For MoM-NLS, standard M-estimation theory implies that, under regularity, the sample objective function converges uniformly to its population analogue, which has a unique minimiser at $\boldsymbol{\alpha}_0$. Consistency

$$\hat{\boldsymbol{\alpha}}_{\text{MoM-NLS}} \xrightarrow{\text{a.s.}} \boldsymbol{\alpha}_0$$

follows from the argmin continuous mapping theorem for M-estimators (see, e.g., Chapter 5 in [Van der Vaart \(2000\)](#)). The argument above is stated without the additional restriction (17), which we keep in the numerical implementation as a finite-sample regularisation. Under correct OT specification, this restriction does not change the consistency target because the empirical upper bound converges to $s_{\boldsymbol{\alpha}_0}$ and the true parameter remains feasible. A full asymptotic treatment of the constrained estimator, which depends on the sample, is beyond the scope of this paper.

Therefore, both proposed moment-based estimators are strongly consistent. This ensures that, as the sample size increases, the estimated parameters converge almost surely to the true parameters of the bivariate beta distribution. It also justifies the use of bootstrap procedures to approximate the sampling distribution in finite samples. Under misspecification, for instance if condition (6) fails, there is no true OT parameter $\boldsymbol{\alpha}_0$ to recover. However, the empirical moments still converge to the corresponding population moments, so, when the resulting coordinates are positive, MoM-Closed converges to the value obtained by applying the closed-form formula (7) to the limiting moments (m_1, m_2, v_1, ρ) of the true distribution. This value need not reproduce the second marginal variance. By contrast, MoM-NLS targets the minimiser of the corresponding population least-squares moment criterion. These two limiting targets need not coincide under misspecification.

3.2 Bayesian estimation

Given a prior $\pi(\boldsymbol{\alpha})$ and the likelihood in (2), Bayes' rule updates our knowledge about the parameter conditional on $\mathbf{Z}$ through the posterior distribution, $p(\boldsymbol{\alpha}|\mathbf{Z})$. After setting a loss function, we can derive the Bayes estimator $\hat{\boldsymbol{\alpha}}$ by minimising the expected posterior loss. Here we report the posterior mean (quadratic loss) and posterior median (absolute loss) as point estimation strategies, which we call BE-Mean and BE-Med, respectively. Because $p(\boldsymbol{\alpha}|\mathbf{Z})$ is intractable, we use dynamic Hamiltonian Monte Carlo (HMC; [Neal, 2011](#)) and its No-U-Turn Sampler variant (NUTS; [Hoffman and Gelman, 2014](#)) as implemented in Stan ([Stan Development Team, 2022](#)).

Direct HMC on (2) is inconvenient due to the inner integral. We circumvent this by introducing a latent variable. Considering the density of the Dirichlet distribution given in (1) and setting $(w, x, y) = (U_1, U_1 + U_2, U_1 + U_3)$, by the Change of Variables formula,

$$f(w, x, y|\alpha) = \frac{1}{B(\alpha)} w^{\alpha_1-1} (x-w)^{\alpha_2-1} (y-w)^{\alpha_3-1} (1-x-y+w)^{\alpha_4-1} \mathbb{1}\{w \in \Omega(x, y)\},$$

where $\Omega(x, y) := (\max(0, x+y-1), \min(x, y))$. For n observations, introduce latent variables $w \in \mathbb{R}^n$ with $w_i \in \Omega(x_i, y_i)$. Writing the complete likelihood function $L(\alpha, w|x, y) = f(x, y|\alpha, w)$ and using the conditional density expression, we get the posterior

$$\begin{aligned} p(\alpha, w|x, y) &\propto \prod_{i=1}^n L(\alpha, w_i|x_i, y_i) f(w_i|\alpha) \pi(\alpha), \\ &\propto \prod_{i=1}^n \frac{f(w_i, x_i, y_i|\alpha)}{f(w_i|\alpha)} f(w_i|\alpha) \pi(\alpha), \\ &= \prod_{i=1}^n f(w_i, x_i, y_i|\alpha) \pi(\alpha). \end{aligned}$$

To stabilise HMC, we reparameterise the model as $\alpha = s\theta$ with $s > 0$ and $\theta \in \mathcal{S} := \{\theta = (\theta_1, \theta_2, \theta_3, \theta_4) \in (0, 1)^4 : \theta_1 + \theta_2 + \theta_3 + \theta_4 = 1\}$. This change of variables separates the concentration parameter s from the composition θ , reducing the geometry curvature when some α_j is small, and yielding more stable NUTS geometry while keeping the statistical model intact. We use simulation-based calibration (SBC) (Talbot et al., 2018) to validate this implementation. In this method, we use the fact that if $\tilde{\alpha} \sim \pi(\alpha)$, $\tilde{Z} \sim \text{BivariateBeta}(\tilde{\alpha})$ and $\{\alpha^1, \dots, \alpha^L\} \sim p(\alpha|\tilde{Z})$, the rank statistic

$$r(\{\alpha_i^1, \dots, \alpha_i^L\}, \tilde{\alpha}_i) = \sum_{l=1}^L \mathbb{1}[\alpha_i^l < \tilde{\alpha}_i]$$

has distribution $\text{Unif}\{0, \dots, L\}$ for each $i = 1, \dots, 4$. Using the priors $s \sim \text{Gamma}(4, 1)$ and $\theta \sim \text{Dirichlet}(1, 1, 1, 1)$ with $\alpha = s\theta$, which is exactly equivalent to $\alpha_i \stackrel{\text{iid}}{\sim} \text{Gamma}(1, 1)$, yields approximately uniform rank histograms as depicted in Figure 2. All supplementary material is provided in Online Resource 1. An analogous SBC check under the log-normal prior on s is reported in Supplementary Figure S2 and leads to the same conclusion.

Besides SBC, we monitor standard MCMC/HMC diagnostics, including $\hat{R}$ (Gelman and Rubin, 1992), divergent transitions, energy (BFMI) (Betancourt, 2016), and effective sample sizes (ESS) (Vehtari et al., 2021). All diagnostics and scripts are available at <https://github.com/lucamoschen/bivariate-beta>. With the per-observation latent variables, the complete-data likelihood factorises across the samples. We exploit this with vectorised Stan code, which is substantially faster than quadrature-based likelihood evaluation.

3.2.1 Prior distributions for α

A natural question to ask is which prior distribution should one place on α . We adopt the scale-simplex parameterisation $\alpha = s\theta$ with $s > 0$ and $\theta \in \mathcal{S}$, and place priors on (s, θ) . We choose $\theta \sim \text{Dirichlet}(a_1, \dots, a_4)$ with $a_i > 0$, and for s we consider:

- (a) $s \sim \text{Gamma}(\kappa, \lambda)$. Choosing $\kappa = a_1 + \dots + a_4$ and a common rate λ is equivalent to setting independent $\alpha_i \sim \text{Gamma}(a_i, \lambda)$.
- (b) $s \sim \text{log-normal}(\log s_0, \tau)$, where s_0 is the prior median of s and sets the typical concentration of the beta marginals, and τ is the log-scale standard deviation.

We use Gamma and log-normal priors for the concentration parameter s because both are supported on $(0, \infty)$, they are computationally tractable and easy to calibrate. A Gamma prior is convenient when prior information is expressed through a mean and variance, and combined with a Dirichlet on θ it recovers independent Gamma priors on the α_i . Alternatively, a log-normal prior is useful when prior information is given in terms of a median effective sample size and multiplicative uncertainty, and it also has a heavier upper tail than the Gamma prior. Prior predictive checking (see e.g. Gabry et al. (2019)) helps us understand the implications of each prior choice. Two empirical takeaways:

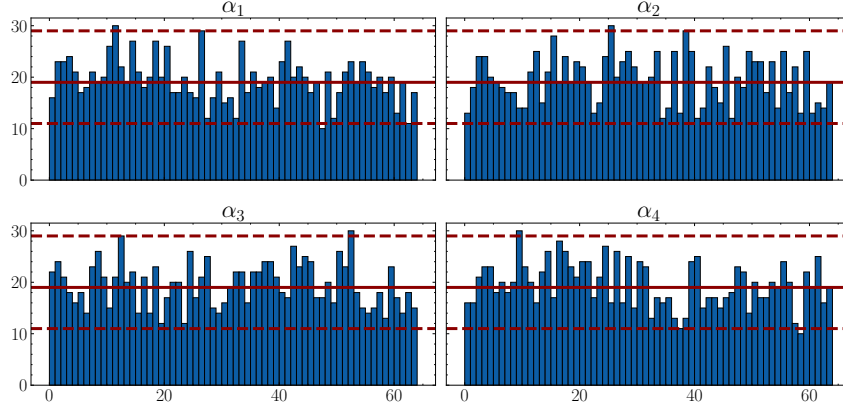


Figure 2: **Simulation-based calibration:** Histogram of the rank statistics r for each component of α . The dashed red lines mark the 95% interval under a discrete uniform distribution, and the solid red line indicates the median. We used $L = 63$ draws and $N = 1252$ experiments.

- (a) Under a symmetric prior on θ , the induced prior on ρ is centred and approximately symmetric about 0 (see Figure 3). To encode prior information on the sign of ρ , shift mass from (θ_2, θ_3) toward (θ_1, θ_4) (or vice versa). Such sign-informed priors are most reasonable when prior knowledge suggests a mechanism for concordance or discordance between the two proportions. Distinguishing components of θ also changes the marginal means.
- (b) The choice of prior for s primarily controls the overall concentration of the marginals, hence their variances. The larger s concentrates the beta marginals, while the smaller s produces more diffuse priors. The difference between the Gamma and the log-normal prior in s does not affect the induced priors in m_1 , m_2 , and ρ , but affects the variances v_1 and v_2 . The log-normal has a slightly heavier upper tail in s , thus yielding a modest shift toward smaller marginal variances.

For the corresponding supplementary simulation checks, see Supplementary Figures S2–S4. We do not pursue Jeffreys or Bernardo priors here because inference is implemented through a latent-variable augmentation rather than a closed-form likelihood, which makes their derivation and posterior propriety analysis non-trivial.

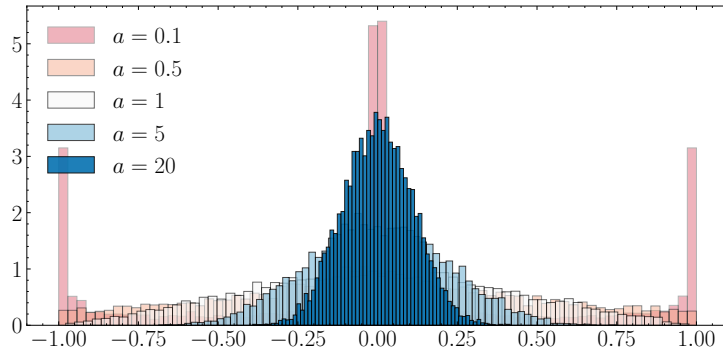


Figure 3: **Induced prior on correlation under Gamma priors:** Distribution of $\rho = \text{Cor}(X, Y)$ when $s \sim \text{Gamma}(4a, a)$ and $\theta \sim \text{Dirichlet}(a, a, a, a)$ for different values of a , leading to $\alpha_i \stackrel{\text{iid}}{\sim} \text{Gamma}(a, a)$.

4 Elicitation of a bivariate beta prior

A practical use of the bivariate beta is as a joint prior on two proportions when prior correlation is expected, such as the sensitivity and specificity of a diagnostic test. In this section we briefly discuss how to elicit a bivariate beta, i.e., specify α by reasoning in terms of marginal behaviour and correlation. As

noticed in [Remark 2](#), s_α is shared by both marginals and plays the role of a prior effective sample size (pseudo-trials).

Let $\varphi = (\varphi_1, \varphi_2) \in [0, 1]^2$ denote the parameter of interest and suppose that $\varphi \sim \text{BivariateBeta}(\alpha)$. We propose an elicitation approach for α . To identify the parameter α , we consider that the researcher wants to encode the values $m_1 = \mathbb{E}[\varphi_1]$, $m_2 = \mathbb{E}[\varphi_2]$, $v_1 = \text{Var}(\varphi_1)$, $v_2 = \text{Var}(\varphi_2)$ and $\rho = \text{Cor}(\varphi_1, \varphi_2)$ in the parameter α . By the expression in (6) from [Proposition 2](#), we already know whether one can define a bivariate beta distribution with these moments. Therefore, two practical elicitation modes are useful.

- (a) The practitioner specifies marginal means m_1 , m_2 , a prior strength $s > 0$ (effective number of pseudo-trials), and a target correlation ρ . In practice we take the projected correlation $\tilde{\rho}$ as in the MoM-Closed estimator. Define α by the closed-form inversion (7) with $s_\alpha = s$.
- (b) The practitioner specifies marginal means m_1 , m_2 , a target variance for one margin (e.g., v_1), and a target correlation ρ . Use the projected correlation and set

$$s = \frac{m_1(1 - m_1)}{v_1} - 1 \implies v_2^* = \frac{m_2(1 - m_2)}{s + 1}$$

for the other marginal variance. If both variances (v_1, v_2) are provided, feasibility requires (6). Otherwise, one must prioritise a margin (keep v_1 and accept v_2^* , or vice-versa), or adopt a compromise s , such as the arithmetic average of the two implied strengths. Set the shared strength as above, then apply (7) with $s_\alpha = s$.

Remark 4. *Although we describe information about marginal means and variances as inputs, any two marginal summaries that identify a beta distribution can be used (e.g. mean + quantile, two quantiles, tail probabilities). One first solves for the parameters (a_x, b_x) and (a_y, b_y) of the beta marginals and sets $s_\alpha = a_x + b_x = a_y + b_y$. If these totals disagree, use the average or prioritise one margin. Correlation is then chosen (and projected if necessary) exactly as above.*

5 Numerical experiments

We use 1,000 Monte Carlo simulations at sample sizes $n = 50$ and $n = 200$ to compare four estimation strategies for α (MoM-Closed, MoM-NLS, BE-Mean, and BE-Med) using bias, mean squared error (MSE), mean absolute percentage error (MAPE), runtime, and 95% interval coverage (bootstrap CIs for MoM and equal-tailed BCIs for Bayes). To contextualise the computational benchmarks reported throughout this section, all experiments were run on a standard personal computer (Apple M3 processor, 8 cores, 16 GB memory). Code and replication materials are available at <https://github.com/lucamoschen/bivariate-beta>.

5.1 Recovering parameters under the bivariate beta

We begin our investigation with the well-specified case. Let $(X, Y) \sim \text{BivariateBeta}(\alpha)$, where α is unknown. For this experiment, we construct five representative settings to consider symmetry, concentration, and correlation sign and magnitude: (a) $\alpha = (1, 1, 1, 1)$ ($m_1 = m_2 = 0.5$, $\rho = 0$); (b) $\alpha = (3, 1, 1, 3)$ (same means, smaller variances, $\rho \approx 0.5$); (c) $\alpha = (1, 7.4, 2.6, 1)$ (asymmetric means, strong negative correlation $\rho \approx -0.6$); (d) $\alpha = (0.4, 0.5, 1.4, 0.7)$ (low concentration, mild negative correlation $\rho \approx -0.21$); (e) $\alpha = (19, 1, 1, 19)$ ($m_1 = m_2 = 0.5$, $\rho = 0.9$). For method of moments estimators we report 95% bootstrap CIs ($B = 500$ resamples) and coverage. For the Bayesian estimates, we used symmetric weakly informative $s \sim \text{Gamma}(4, 1)$ and $\theta \sim \text{Dirichlet}(1, 1, 1, 1)$ priors for all five specifications, equivalently $\alpha_i \stackrel{\text{iid}}{\sim} \text{Gamma}(1, 1)$. This choice was made based on the induced prior on ρ in [Figure 3](#). It is centred near 0, reasonably spread, while also avoiding excessive prior mass near the boundaries ± 1 . Numerical tables are provided in Supplementary Tables S5–S10, including summaries for the induced estimates of ρ . Figures below summarise performance at $n = 50$ and $n = 200$ for the average MAPE and absolute bias across $\alpha_1, \dots, \alpha_4$.

In settings (a)–(d), [Figure 4](#) shows that BE-Mean and BE-Med typically attain the lowest MAPE, with BE-Med slightly more stable. Among frequentist methods, MoM-NLS generally reduces MAPE relative to MoM-Closed. This is expected since MoM-NLS jointly matches all five moments, whereas MoM-Closed does not rely on $\hat{v}_2$ and is more sensitive to sampling noise in $\hat{v}_1$ and $\hat{\rho}$, because it solves the

equations exactly. Turning to bias, Figure 5 indicates that MoM-NLS produces smaller absolute bias than MoM-Closed for most settings. The Bayesian estimators are competitive overall, but under the default prior they can show larger absolute bias than MoM-NLS when the true concentration is high. This is most apparent in settings (c) and (e) in Figure 5, where s_α is 12 and 40, respectively, and thus lies in the tail of the Gamma(4, 1) prior on s . For instance, prior to observing the data, we have $\mathbb{P}(\alpha_2 > 7) \lesssim 0.001$, which helps explain the larger Bayesian bias under the default prior. Setting (e) is even more extreme, where the moment estimators keep a reasonable MAPE at $\rho = 0.9$ but have the largest absolute bias of all five settings. More informative priors reduce this bias, as seen in setting (a) in Figure 5 and further detailed for setting (c) in Supplementary Figures S3–S4. However, we retain the same prior across all settings for a fair comparison. Thus, if low bias is the main objective and prior information is not available, the results often favour MoM-NLS, whereas the Bayesian procedure is preferable when prior information is available or posterior inference is desired. Finally, as the sample size increases, errors for both frequentist estimators decrease, consistent with their strong consistency.

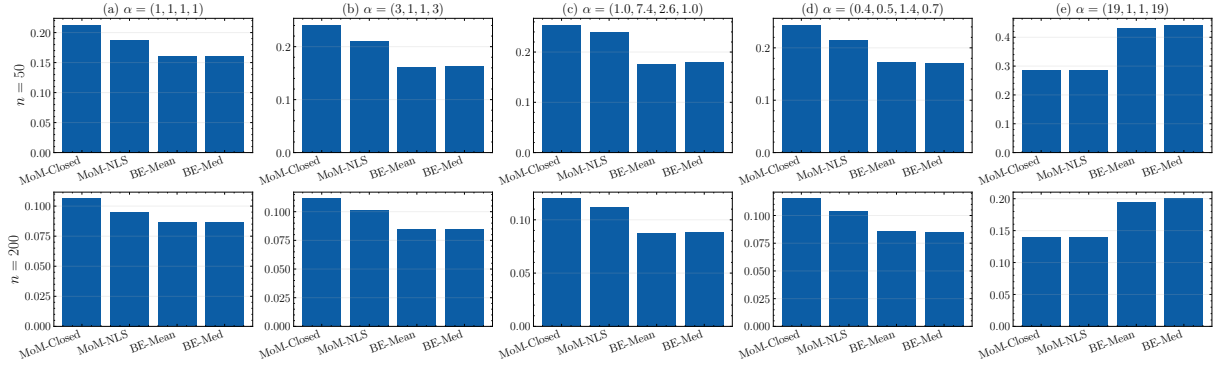


Figure 4: **MAPE across five designs.** Mean absolute percentage error for $\hat{\alpha}_1, \dots, \hat{\alpha}_4$ under MoM-Closed, MoM-NLS, BE-Mean, and BE-Med. Columns correspond to parameter settings (a)-(e) and rows to the number of samples.

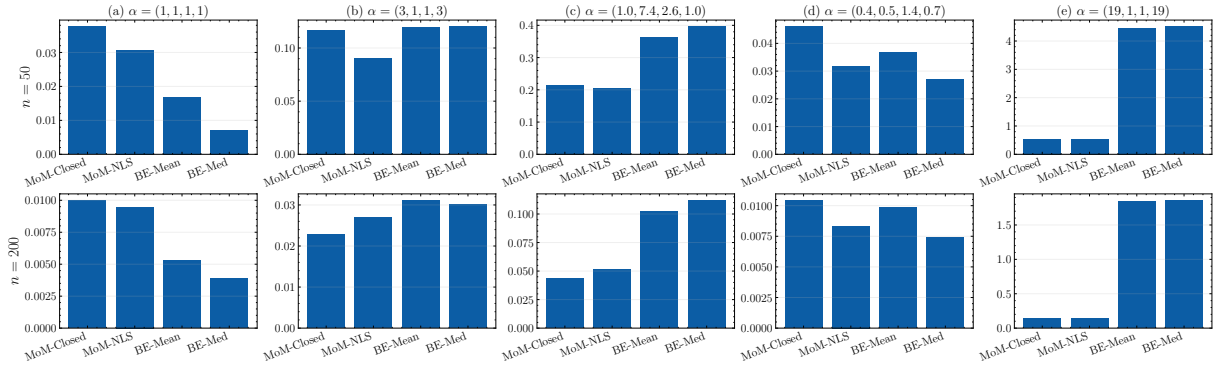


Figure 5: **Absolute bias across five designs.** Average absolute bias for $\hat{\alpha}_1, \dots, \hat{\alpha}_4$ under MoM-Closed, MoM-NLS, BE-Mean, and BE-Med. Columns correspond to parameter settings (a)-(e) and rows to the number of samples.

The observed coverage of the confidence intervals for MoM-Closed and MoM-NLS is generally close to the nominal 95%. In settings (a), (b), and (d), it ranges from 92.4% to 95.8%, while in the other two, it is slightly worse, but remains at least 89.8%. We thus conclude that the bootstrap method produces good interval estimates in these experiments. It is important to note that the Bayesian credible intervals (BCI) need not have good frequency characteristics, so their empirical coverage should be interpreted only as a frequentist diagnostic. For the BCI, the coverage is close to nominal in settings (a), (b), and (d), ranging from 93.0% to 97.5%, while in setting (c), the BCI underestimates the coverage for α_2 and α_3 , with coverage between 84.5% and 95.7% across the four parameters for $n = 50$ and 93.2% to 94.2% for $n = 200$. In setting (e), the Bayesian undercoverage is much stronger, with coverage below 50% for $n = 50$, and ranging from 20% to 61% for $n = 200$. Additionally, to understand the main drivers of

MoM-Closed variability, **Figure 6** reports a sensitivity analysis for $\hat{\alpha}_3$ and $\hat{\alpha}_4$ at $n = 50$. The scatterplots make clear that fluctuations in $\hat{v}_1$ predominantly impact $\hat{\alpha}_3$, whereas the projected correlation $\hat{\rho}$ largely drives $\hat{\alpha}_4$, which is consistent with the closed-form inversion in (7).



Figure 6: **Sensitivity of MoM-Closed to sample moments** ($n = 50$). Each cell reports $\Delta R^2(\%)$, the additional percentage of variance in $\hat{\alpha}_i$ explained by the moment j beyond all other moments. Larger values mean stronger unique influence on $\hat{\alpha}_i$. The sign marker indicates the coefficient sign of moment j in the full linear regression. Rows index the target parameter estimators for α_3 and α_4 . Columns index observed moments. Values are averaged over Monte Carlo replications.

Finally, we analyse the runtime. At $n = 50$, MoM-Closed is essentially instantaneous. MoM-NLS is one order of magnitude slower given the nonlinear optimisation, and the Bayesian estimators (BE-Mean/BE-Med, via NUTS with the latent-variable formulation) are a further three orders of magnitude slower than MoM-NLS. The ordering is the same at $n = 200$, but absolute times increase modestly for MoM-NLS and more noticeably for the Bayesian methods.

5.2 Model misspecification

We next study performance under misspecification, taking the true data-generating process to be a different distribution on $[0, 1]^2$. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. random variables with a distribution whose unknown density $h(x, y)$ has support over $[0, 1]^2$, and suppose we use the bivariate beta density $f_{\alpha}(x, y)$ to approximate h . For the method of moments, we know that the solution in (7) converges (almost surely) to $\hat{\alpha}$ such that $f_{\hat{\alpha}}$ has the same means, one marginal variance and correlation as h if $\hat{\alpha}$ has positive coordinates. Nonetheless, several issues may arise: (i) the empirical moment estimators given in equation (13) may have high variances and thus yield bad estimates; (ii) the system may yield negative solutions, even for large values of n , given the results presented in **Figure 1**; and (iii) matching moments may not capture other aspects of interest of h , such as probabilities.

Consider the joint density $h(x, y)$ of $X = 1/(1 + e^{-H_1})$ and $Y = 1/(1 + e^{-H_2})$, such that $H = (H_1, H_2)$ and $H \sim \text{Normal}(\mu, \Sigma)$, where $\mu \in \mathbb{R}^2$ and $\Sigma \in \mathbb{R}^{2 \times 2}$ is a covariance matrix¹. For this experiment, we set two representative sets of parameters: $\mu = (0, 0)$ and $\Sigma = [[1, 0.1], [0.1, 1]]$ and $\mu = -(1, 1)$ and $\Sigma = [[2.25, -1.2], [-1.2, 1]]$, and simulate $n = 50$ samples to estimate α . The Bayesian estimators use the same priors as in **Section 5.1**. We then use 1,000 Monte Carlo simulations to compare bias, MAPE and MSE between the true moments and the implied by $\hat{\alpha}$. Numerical tables can be found in Supplementary Tables S11–S12. In the first experiment, the true moments are, approximately $\mathbb{E}[X] = \mathbb{E}[Y] = 0.5$, $\text{Var}(X) = \text{Var}(Y) = 0.0433$ and $\text{Cor}(X, Y) = 0.098$. In this setting, MoM-Closed estimator is very precise because means and variances obey the relation (6). The other estimators yield similar results. Despite that, the MAPE value is greater than 100% for the correlation estimate, as we can observe in **Figure 7**. This is explained by the large variability of the sample correlation as an estimator of ρ . For instance, for $n = 50$ and true correlation $\rho \approx 0.098$, the sampling variability of $\hat{\rho}$ is large with $\mathbb{P}(|\hat{\rho} - \rho| > 0.1) \approx 0.49$. This sampling variability explains the large MAPE observed for correlation.

Remark 5. *Given the high variability of the empirical correlation in small samples, it is natural to ask whether covariance should be used instead in the moment equations. Our choice is not arbitrary: the closed-form system in (5) and (3) is naturally expressed in terms of correlation, not covariance.*

¹Since the moments of X and Y are not closed-form, we obtain them through Monte Carlo.

Moreover, correlation is scale-free and directly comparable across marginals, which avoids redundancy since the variances v_1, v_2 are already matched. Finally, our interest in eliciting bivariate beta priors also makes correlation the more natural quantity to work with.

The second experiment has the true moments $\mathbb{E}[X] = 0.33$, $\mathbb{E}[Y] = 0.3$, $\text{Var}(X) = 0.062$, $\text{Var}(Y) = 0.033$ and $\text{Cor}(X, Y) = -0.73$. These values do not yield a well-defined bivariate beta because the solution in (7) does not belong to the parameter space. Therefore, each method proposes approximations based on how it was defined. In Figure 7, we see that MoM-Closed incurs larger error in $\text{Var}(Y)$ since it ignores this variance in its closed-form inversion. By contrast, MoM-NLS reduces the error in $\text{Var}(Y)$ but at the cost of larger deviations in the means. The Bayesian estimators (BE-Mean and BE-Med) balance errors more evenly across the moments, though correlation remains difficult to approximate due to its high variability in small samples. In practice, the preferred estimator will depend on which features (means, variances, correlation) are most relevant for the application.

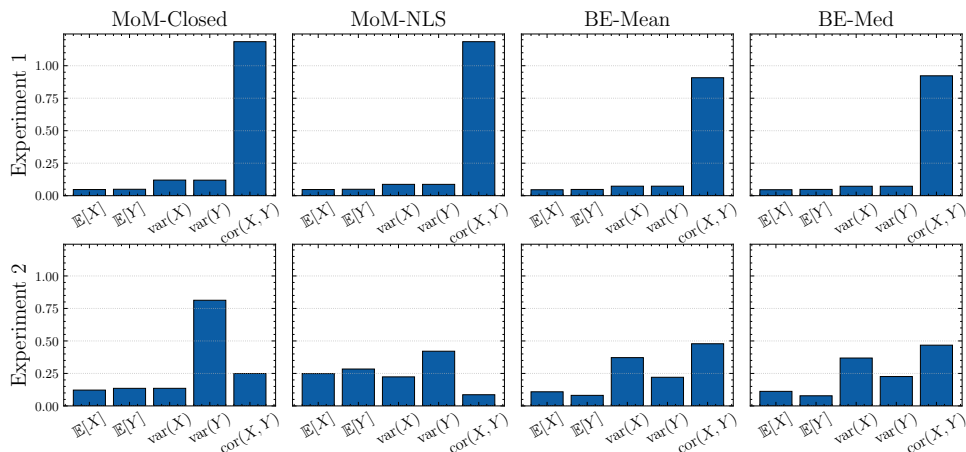


Figure 7: **MAPE values** for the experiments in the misspecified case considering the true moments against the estimated through the bivariate beta with parameter $\hat{\alpha}$ obtained from MoM-Closed and MoM-NLS.

Finally, we turn our attention to the whole density we want to approximate – see Figure 8. In the first experiment, the estimated moments are close to the true ones, especially when n is large. The densities are similar, but h seems more dispersed. If we compare the joint tail behaviour, we notice that the bivariate beta model produces a value of $\Pr(X > 0.9, Y > 0.9) \approx 1.3 \cdot 10^{-3}$ against $3.9 \cdot 10^{-4}$ of the true distribution. This illustrates that even when the matched moments are close, this distribution may not represent other desired aspects. In the same fashion, the bivariate beta in the second experiment represents the mode of the distribution poorly, driven by a poor method of moments estimate. The estimated distribution does however give small probabilities to regions where the true probability is also low.

5.3 Diagnostics

As we have discussed in Section 5.2, the bivariate beta distribution does not always approximate the unknown distribution of the data $\mathbf{Z}$ well. This is not a problem with the distribution itself but reveals its inflexibility. In this section, we propose a systematic way of diagnosing the compatibility of a given dataset with a bivariate beta distribution.

5.3.1 Asymptotic diagnostics

For the first diagnostic, we consider the necessary relation established by Proposition 2,

$$\frac{m_1(1 - m_1)}{v_1} = \frac{m_2(1 - m_2)}{v_2}.$$

This is a necessary condition for having a bivariate beta with marginal means m_1 and m_2 and marginal variances v_1 and v_2 . Equivalently, this diagnostic checks whether the restriction from Remark 2 holds in

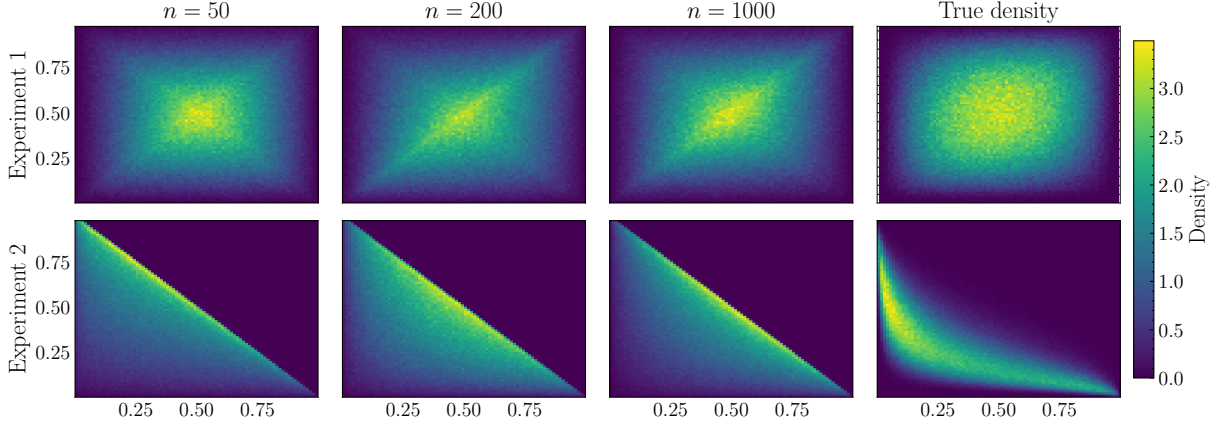


Figure 8: **Comparing the contour plots** of the bivariate beta density with parameter $\hat{\alpha}$ and the true density of h for both experiments in the misspecified case. MoM-NLS was used as the estimator based on different sample sizes: $n = 50, 200$, and $1,000$.

the data. Defining

$$g(x_1, x_2, x_3, x_4) = x_1(1 - x_1)x_4 - x_2(1 - x_3)x_3, \quad G_n = g(\hat{m}_1, \hat{v}_1, \hat{m}_2, \hat{v}_2).$$

If data comes from a bivariate beta distribution, $G_n \xrightarrow{p} 0$ by the consistency of the method of moments and, by [Proposition 4](#) in the Appendix,

$$\sqrt{n} \frac{G_n}{\hat{\sigma}_n} \rightarrow N(0, 1),$$

where $\hat{\sigma}_n^2$ is a consistent estimator for $\nabla g(m_1, v_1, m_2, v_2)^T \Sigma \nabla g(m_1, v_1, m_2, v_2)$. [Figure 9](#) illustrates this approximation for $n = 30$ across different parameter values.

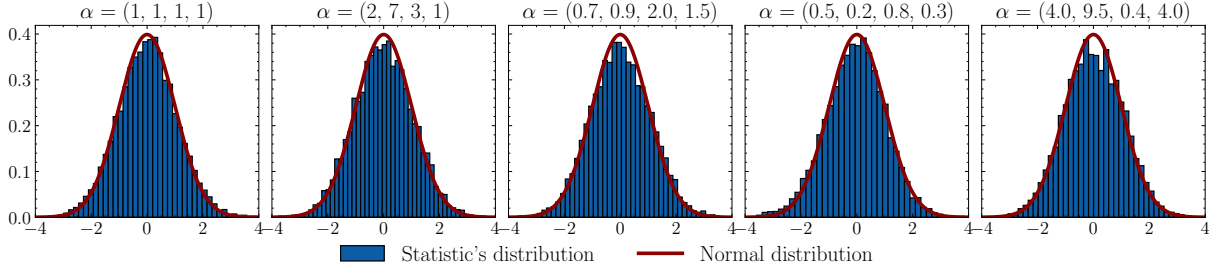


Figure 9: **Sampling distribution of the diagnostic.** Histogram of the standardized statistic $\sqrt{n}G_n/\hat{\sigma}_n$ over Monte Carlo replications overlaid with the standard normal density (red), for $n = 30$ and five specifications of α .

The testing procedure has the null hypotheses that the data are compatible with a bivariate beta model, i.e. $H_0 : g(m_1, v_1, m_2, v_2) = 0$, against $H_1 : g(m_1, v_1, m_2, v_2) \neq 0$. Given an observed sample, compute $s = \sqrt{n}G_n/\hat{\sigma}_n$. The (asymptotic) two-sided p -value is $p = 2\Phi(-|s|)$, where Φ is the standard normal cumulative distribution function (CDF). Supplementary Figure S5 illustrates the distribution of p -values under both the null and alternative specifications.

The second diagnostic targets the correlation aspect of [Proposition 2](#), namely relation (7). If the data come from a bivariate beta, then the closed-form inversion in (7) yields a positive solution. Let $\hat{\alpha}$ be the solution from (7) built from $(\hat{m}_1, \hat{m}_2, \hat{v}_1, \hat{\rho})$. Define $\hat{\beta}_i = \hat{\alpha}_i/\hat{s}_\alpha$ and the summary statistic

$$M = \min\{\hat{\beta}_1, \dots, \hat{\beta}_4\}.$$

This choice is convenient because $M > 0$ if and only if all four positivity constraints are satisfied, so M records the signed difference to the boundary. Moreover, one always has $M \geq -1/8$, so the statistic cannot

become arbitrarily negative. Other summaries, such as the total negative mass $\sum_i (-\hat{\beta}_i)_+$, could also be considered, but they are less directly interpretable and do not have tractable asymptotic distributions either. Accordingly, the proposed procedure rejects compatibility with the bivariate beta whenever $M \leq c$ for a fixed calibrated threshold $c < 0$. Because MoM-Closed is consistent, $M \rightarrow \min_i \{\alpha_i / s_{\alpha}\} > 0$ under H_0 and $c = 0$ would give a conservative large-sample test. For finite n , small true components can drive negative estimates. The distribution of M is analytically intractable, even asymptotically, because it is the minimum of correlated statistics. We therefore calibrate c by nonparametric bootstrap under H_0 . Supplementary Figure S6 shows the $n = 50$ quantile distribution, and Supplementary Figure S7 displays the distribution of the 5th percentile of M for $n = 10, 20, 50, 100, 200$ over uniformly sampled $\alpha \in [0, 0.5]^4$, a region that mostly causes negative estimates. These experiments show that the threshold depends on the sample size. For $n = 10$, the calibrated lower tail moves to -0.1 . By $n = 20$, the distribution is already shifted to the right of $c = -0.05$, and for $n = 100$ and $n = 200$ this shift is even more pronounced due to the consistency of the estimator. For the sample sizes emphasised in this paper, the choice $c = -0.05$ keeps the null rejection probability below 0.05 and is slightly conservative. In practice, however, c should be recalibrated by the same bootstrap scheme for the sample size at hand. Results for the 1st and 10th percentiles are similar.

5.3.2 Bayesian diagnostics

Bayesian methods need a different suite of diagnostics owing to the fact that now a full posterior distribution is being generated. Posterior predictive checks (PPC, [Gabry et al. \(2019\)](#)) use the parameters' posterior samples to look into the distribution of induced quantities, comparing them to the observed in the sample. In our case, we can compare the observed moments $\hat{m}_1, \hat{m}_2, \hat{v}_1, \hat{v}_2$ and $\hat{\rho}$ to the calculated through the samples of the posterior $p(\alpha|\mathbf{Z})$. [Figure 10](#) shows an example of PPC. The observed moments fall inside the 95% posterior predictive intervals. Informative priors can further sharpen these checks when justified by domain knowledge.

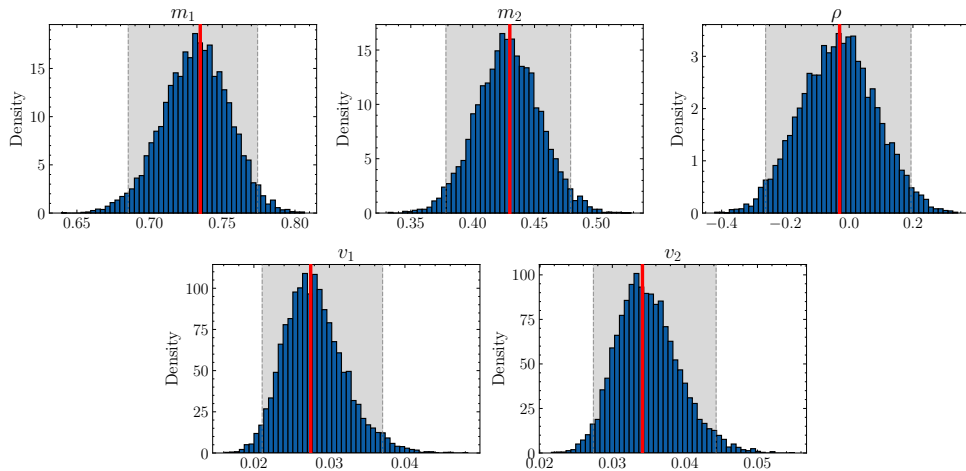


Figure 10: **Posterior predictive checks (moments)**. Posterior distributions of sample moments and the observed values in red. The shaded area represents 95% equal-tailed interval. Data generated with $\alpha = (2, 3, 0.7, 1)$ and $n = 50$ under independent $\text{Gamma}(1, 1)$ priors.

6 Applications to real data

We illustrate the proposed estimators and diagnostics in two very distinct real datasets comprising pairs of proportions on $[0, 1]^2$. In both cases, we (i) briefly describe the data and pre-processing; (ii) fit the model using MoM-Closed, MoM-NLS, and compute the posterior distribution; (iii) report the asymptotic diagnostic in [Proposition 4](#) and the positivity diagnostic induced by (7); and (iv) compare with a baseline model we introduce later. Posterior predictive checking (PPC, cf. [\(Gabry et al., 2019\)](#)) complements the analysis. All estimators use the empirical moments in (13). MoM-NLS is initialised at MoM-Closed, Bayes uses the latent-variable formulation from [Section 3.2](#) with 4 chains, NUTS, and weakly informative priors

as noted below. Uncertainties for the method of moments use the nonparametric bootstrap ($B = 500$ over pairs).

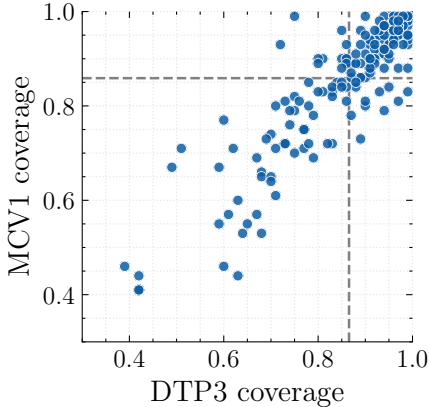
To complement the bivariate Beta (BB), we fit a bivariate logit-normal (BLN) baseline (Ali et al., 1978; Atchison and Shen, 1980), as already used in Section 5.2. The BLN has one additional free parameter, which gives it slightly more flexibility and thus an advantage in predictive comparisons. We report both maximum likelihood and Bayesian estimates. For Bayes, we take $\mu \sim N(0, \tau^2 I)$, $\sigma_j \sim N^+(0, \kappa_j^2)$, $\rho \sim \text{Unif}(-1, 1)$, and

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

where N^+ denotes a half-normal prior, and $\tau > 0$, and $\kappa_j > 0$ are spread hyperparameters. We compare the fitted models using three complementary criteria: (i) overall predictive accuracy: for Bayesian fits we report Pareto smoothed importance sampling leave-one-out cross-validation (PSIS-LOO, an estimate of expected log predictive density computed by importance-sampling leave-one-out posteriors cf. Vehtari et al. (2017)), while for MoM and MLE fits, we report the corresponding leave-one-out CV log-score; (ii) tail event prediction on the region $\{X > 0.9, Y > 0.9\}$ with Brier score and log score; and (iii) moment-based fit diagnostics via posterior or bootstrap distributions of m_1 , m_2 , v_1 , v_2 , and ρ . Additional visual comparisons between OT and BLN, including marginal Q-Q plots and contour overlays for both case studies, are reported in Supplementary Figures S8 and S10. Supplementary Tables S3–S4 also report an energy-based goodness-of-fit analysis, with full numerical tables. For each fitted model, we compare the observed data with a large sample simulated from that fitted model using the two-sample energy statistic with permutation calibration, and we report the resulting p -values as fitted-model discrepancy summaries.

6.1 Case A: Childhood vaccination coverage in the world

We consider the WHO-UNICEF vaccination coverage series as compiled by Our World in Data (OWID) from WUENIC (WHO & UNICEF, 2025), focusing on the share of one-year-olds vaccinated against diphtheria-tetanus-pertussis (DTP) and measles. For each country j , we set $X_j = \text{DTP3}$ (third dose DTP) coverage and $Y_j = \text{MCV1}$ (first-dose measles) coverage in 2024, expressed as proportions. The final sample size is $n = 195$. The two series have similar means and comparable variances with positive correlation, so the identity (6) holds closely, making this a good example for the Olkin-Trikalinos bivariate Beta. Figure 11a shows the resulting scatter plot.



(a) DTP3 (X) vs. MCV1 (Y) coverage in one-year-olds per country. Dashed lines mark sample means.

	m_1	v_1	m_2	v_2	ρ
Observed	0.865	0.017	0.859	0.020	0.855
MoM-Closed	0.865	0.017	0.859	0.018	0.855
MoM-NLS	0.866	0.018	0.859	0.019	0.855
BE-Mean	0.854	0.018	0.852	0.018	0.816
BE-Med	0.855	0.018	0.853	0.019	0.816
$\Delta = \frac{m_1(1-m_1)}{v_1} - \frac{m_2(1-m_2)}{v_2} \approx 0.81$, $M \approx 0.014$					
$\sqrt{n}G_n/\hat{\sigma}_n \approx 1.14$, (p -value ≈ 0.26).					

(b) Observed and fitted moments for the OT estimators.

Figure 11: **Case A: Childhood vaccination.** Panel (a) scatter plot by country; panel (b) observed and fitted moments with diagnostics.

MoM-Closed applies the closed-form inversion with no projection for ρ . MoM-NLS matches all five moments, thus slightly increasing the implied m_1 in comparison to $\hat{m}_1$ to improve the approximation of $\hat{v}_2$. For the Bayesian fit, we use weakly informative priors $\theta \sim \text{Dirichlet}(1, 1, 1, 1)$ and $s \sim \text{log-normal}(\log(4), 1)$. A prior-sensitivity analysis with eight different prior settings reported in Supplementary Figure S9 leads to the same qualitative conclusions and very similar PSIS-LOO values.

The sampling algorithm does not produce any divergences with $\hat{R} \approx 1.00$, bulk/tail ESS $\gg 1000$, and BFMI > 0.7 . The asymptotic test based on G_n does not reject the compatibility of the marginals with p -value $p = 0.26$. In addition, the positivity statistic satisfies $M > -0.05$, so at level 0.05, the test does not reject the positivity of the estimated parameter. MoM-Closed, MoM-NLS, BE-Mean, and BE-Med agree closely, with Bayes showing a small shrinkage of ρ toward 0, as expected under a symmetric prior on θ .

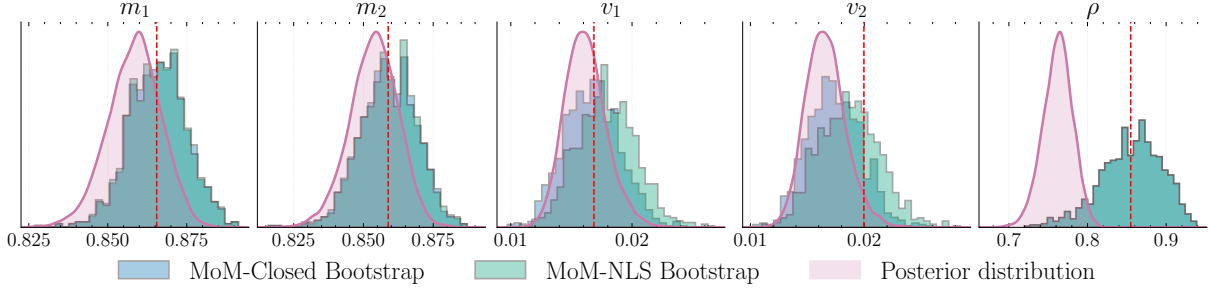


Figure 12: **Distributions of population moments across estimators.** MoM-Closed and MoM-NLS bootstrap samples and posterior distribution of moments m_1 , m_2 , v_1 , v_2 , and ρ with observed values in red.

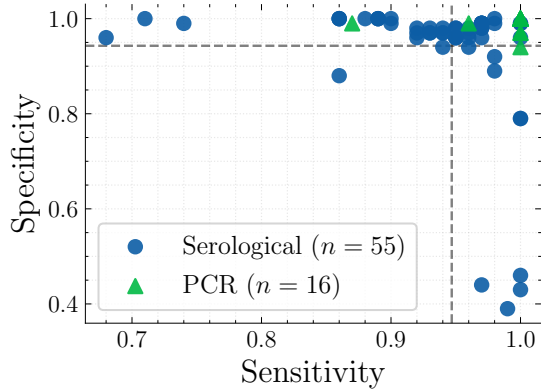
Figure 12 shows the bootstrap distributions of the implied population moments for MoM-Closed and MoM-NLS, and the posterior distributions of the population moments under the Bayesian fit. They largely overlap, and the observed sample summaries (red) appear within high-density regions for m_1 , m_2 , v_1 , and v_2 . For ρ , the Bayesian curve sits left of the sample $\hat{\rho}$, which is expected under a weak, symmetric prior on θ that pulls the population correlation toward zero. Furthermore, a single $\hat{\rho}$ is a noisy statistic whose sampling distribution is skewed near the boundary (Fisher- z 95% range at $N = 195$ is about 0.81-0.89), so the posterior remains within the plausible population range.

Regarding the predictive comparison with the baseline model, the fitted procedures have comparable PSIS-LOO / CV log-scores, with overlapping one-SE bars and a slight edge to MoM-Closed. Posterior-predictive moment checks are acceptable for all fitted models (Z-scores within ± 2), with the BLN estimators closest to zero overall. The bivariate beta shows small positive misfit in m_1 , m_2 and v_2 , and a mild positive shift for ρ . For the joint tail event $\{X > 0.9, Y > 0.9\}$, apart from MoM-NLS, all estimators have the empirical value within the confidence or credible intervals. When comparing the frequentist estimators, MoM-Closed estimator stands out with a slightly higher predictive fit and good predictive moments checks, except for v_2 . Overall, the bivariate beta and BLN fits give similar predictive performance, while BLN provides slightly better calibration of moments. See these results in Figure 14a. Supplementary Tables S3–S4 show somewhat larger discrepancies for the OT Bayesian summaries (posterior mean and median) than for BLN-Bayes, BLN-MLE, MoM-Closed, and MoM-NLS, although the overall differences are small and the latter four fits are very similar on this metric.

6.2 Case B: COVID-19 diagnostic tests from ANVISA registrations

We compiled $n = 71$ reports of accuracy from various coronavirus diagnostic tests submitted to the Brazilian regulator (ANVISA) at the time of device registration, up to June 23, 2020. Each observation consists of a pair (X_i, Y_i) , where the first component is the sensitivity point estimate and the second component is the specificity point estimate for the diagnostic test. The tests cover multiple platforms and types of specimens. Figure 13a displays the resulting scatter plot that separates them into (i) PCR and (ii) serological. Some observations lie on the boundary of the square $[0, 1]^2$. For numerical stability, we replace boundary values (0 or 1) by $0 + \varepsilon$ or $1 - \varepsilon$ with $\varepsilon \stackrel{\text{iid}}{\sim} \text{Unif}(10^{-6}, 5 \times 10^{-4})$. We selected 5×10^{-4} as the smallest tested upper bound that yielded stable HMC sampling. Sensitivity checks with upper bounds 10^{-4} and 10^{-3} gave essentially the same fitted moments and posterior conclusions (Supplementary Table S2). Because the observed sample correlation is modest, we also include as a benchmark the fit of each marginal separately using the MLE of beta distribution parameters.

Figure 13a already signals a difficult modelling problem given the high density close to the boundary, especially at $Y \approx 1$. This example illustrates how the proposed diagnostics behave when the data drive the OT model into an extreme regime. For this dataset, the observed moments are $\hat{m}_1 \approx 0.947$, $\hat{v}_1 \approx 0.0042$,



	m_1	v_1	m_2	v_2	ρ
Observed	0.947	0.004	0.943	0.017	-0.187
MoM-Closed	0.947	0.004	0.943	0.005	-0.058
MoM-NLS	0.901	0.011	0.897	0.011	-0.112
BE-Mean	0.921	0.013	0.936	0.010	0.382
BE-Med	0.922	0.012	0.937	0.010	0.381

$\Delta = \frac{m_1(1-m_1)}{v_1} - \frac{m_2(1-m_2)}{v_2} \approx 8.81$, $M \approx -6.7 \cdot 10^{-3}$
 $\sqrt{n}G_n/\hat{\sigma}_n \approx 2.14$, (p -value $\approx 3.2 \cdot 10^{-2}$).

(b) Observed and fitted moments for the OT estimators.

(a) Sensitivity (X) vs. specificity (Y) by diagnostic test. Triangles: PCR. Circles: serological. Dashed lines mark the sample means.

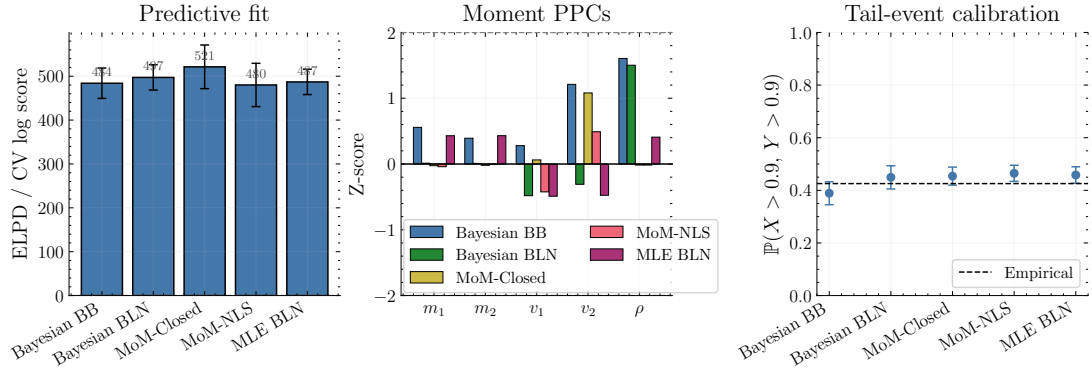
Figure 13: **Case B: ANVISA registrations.** Panel (a) scatter plot by diagnostic test; panel (b) observed and fitted moments with diagnostics.

$\hat{m}_2 \approx 0.943$, $\hat{v}_2 \approx 0.0173$, and $\hat{\rho} \approx -0.187$. Thus, while the marginal means are similar, the variances differ by a factor of about 4. Numerically, $m_1(1-m_1)/v_1 - m_2(1-m_2)/v_2 \approx 8.81$ (see Table 13b). Furthermore, $\hat{\rho}$ falls outside the feasible interval in (11) implied by $(\hat{m}_1, \hat{m}_2)$. These features jointly violate the identity in (6) and the positivity constraints on α that underpin the Dirichlet construction. Consequently, any four-parameter bivariate beta fit must either (i) equalise variances by moving the means or (ii) keep the observed means at the cost of forcing the correlation closer to 0.

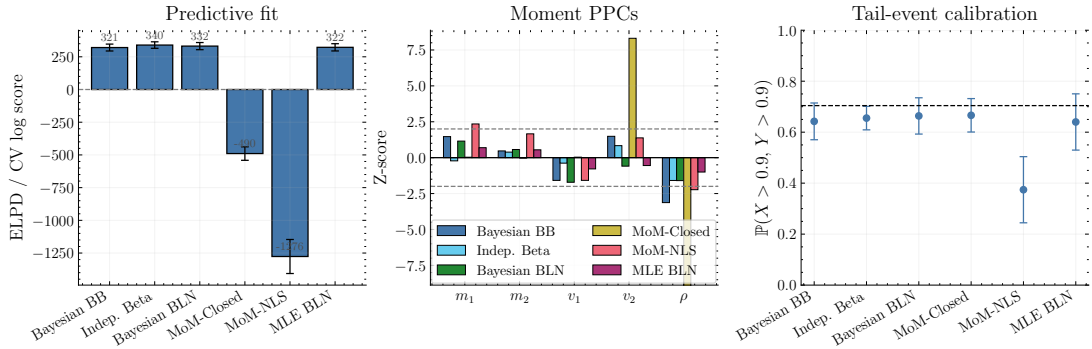
In terms of estimation, MoM-Closed yields a feasible solution only after correlation projection. As expected, it matches m_1, m_2, v_1 exactly and pulls v_2 down to the implied value $\hat{v}_2^{\text{MoM-Closed}} \approx 0.005$, with a milder correlation $\hat{\rho} \approx -0.058$. The positivity statistic is negative ($M \approx -6.7 \cdot 10^{-3}$), but above the calibrated rejection threshold -0.05 . MoM-NLS balances all five moments at the expense of the means, moving both margins to $(\hat{m}_1, \hat{m}_2) \approx (0.901, 0.897)$ while increasing the second variance to $\hat{v}_2^{\text{MoM-NLS}} \approx 0.011$, and retaining a modest negative correlation $\hat{\rho}^{\text{MoM-NLS}} \approx -0.112$. This trade-off mirrors the behaviour seen in Section 5.2. To perform Bayesian inference, we set priors $s \sim \log\text{-normal}(\log(4), 1)$ and $\theta \sim \text{Dirichlet}(1, 1, 1, 1)$, reflecting weakly informative priors. Supplementary prior-sensitivity analyses, as carried out in Case A, lead to the same conclusions (Supplementary Figure S11). In particular, posterior ρ remains positive, indicating that the positive posterior correlation is driven by the model restrictions rather than by the prior choice. Posterior summaries reproduce m_1 and m_2 but tend to underestimate v_2 and assign most mass to $\rho > 0$ even though the observed sample correlation is negative and the weak prior is centred at 0.

The moment compatibility in (6) fails decisively: $\Delta = \frac{m_1(1-m_1)}{v_1} - \frac{m_2(1-m_2)}{v_2} \approx 8.81$ is large, and the asymptotic test based on G_n rejects at 5% level. The positivity diagnostic $\hat{M}$ is close to zero, indicating a near-boundary fit under MoM-Closed even after correlation projection, but it does not reject the null hypothesis since $M > -0.05$. Bayesian diagnostics also clarify the problem of using the bivariate beta distribution of Olkin and Trikalinos (2015) to fit this dataset. Both BE-Mean and BE-Med produce a positive posterior correlation, indicating a model-data mismatch rather than a sampling failure, given good sampler diagnostics. For all monitored parameters, under these priors, we observe $\hat{R} \approx 1.0$, with no divergences, no `max_treedepth` saturations, and no energy pathologies (BFMI > 0.75). However, the posterior for ρ concentrates on positive values with $\mathbb{P}(\rho < 0.2 | \mathbf{Z}) < 0.001$, reflecting the model's constraint that both marginals share the same concentration s_α (Remark 2).

For this dataset, the added independent marginals beta benchmark attains the largest CV log-score, followed closely by the Bayesian BLN fit, while the Bayesian bivariate beta and BLN MLE are slightly lower but still within one-SE bars; the numerical values are given in Supplementary Table S1. Posterior-predictive diagnostics indicate a mismatch. The two BLN estimators match the sample moments more closely for some quantities, as measured by the Z-score criterion. In particular, the bivariate beta shows noticeable misfit for v_2 and ρ . The same pattern appears in the joint tail event $\{X > 0.9, Y > 0.9\}$. Under the BLN model fit, the predicted tail probabilities from both estimators are closer to the empirical rate and achieve lower (which is better) Brier and log scores than the bivariate beta posterior. The



(a) Case A: predictive fit (ELPD), moment PPC Z-scores, and tail-event calibration.



(b) Case B: predictive fit (ELPD), moment PPC Z-scores, and tail-event calibration.

Figure 14: Model comparison on two datasets. Each subfigure shows (left) ELPD / CV log-score with one-SE error bars, (middle) posterior-predictive Z-scores for $\{m_1, m_2, v_1, v_2, \rho\}$, and (right) upper-tail probability $\mathbb{P}(X > 0.9, Y > 0.9)$ with posterior mean, one-SD error bars, and empirical value with dashed line.

independent marginals beta benchmark results indicate that the modest negative sample correlation has limited impact on the practical conclusions for sensitivity and specificity. The two method-of-moments fits perform poorly, with large negative ELPD and extreme moment Z-scores. See these results in [Figure 14b](#). Supplementary Tables S3–S4 show that Case B is more difficult for all fitted models. On this metric, the independent-beta benchmark has the smallest discrepancy, the OT Bayesian summaries improve on the BLN fits, and MoM-NLS has by far the largest energy discrepancy. Additional visual comparisons based on marginal Q-Q plots and contour overlays in Supplementary Figures S8 and S10 suggest similar marginal fits for OT and BLN, with OT tracking the joint geometry more closely, especially in Case A. By contrast, the logit-normal produces a broader and smoother 95% probability region than is suggested by the observed data.

7 Discussion and conclusions

In this paper, we have dealt with a particular construction of the bivariate distribution ([Olkin and Trikalinos, 2015](#)), which strikes a balance between parameter-richness and tractability. We leverage a detailed mathematical analysis of the moments to propose a class of moment-based estimators for the parameter of interest, $\alpha \in \mathbb{R}_+^4$. We also exploit a latent variable construction to propose an efficient representation of the posterior distribution in order to obtain Bayesian estimates. Finally, we have discussed the construction of simple yet powerful diagnostics to evaluate model fit, from both frequentist and Bayesian perspectives. Importantly, we have also compared its performance under both Bayesian and classical approaches to a logistic normal formulation. From the Bayesian side, we focus on proper weakly informative priors with sensible prior-predictive distributions and provide tools for prior specification. Objective priors such as Jeffreys or Bernardo are left for future work because our computation relies on a latent-variable augmentation rather than a closed-form likelihood.

We find that no single method performs best under all situations, but that there is an inherent trade-off between statistical and computational performance: moments estimators that are based on solving fewer equations numerically tend to have lower runtimes but also perform worse in terms of bias or MAPE. When low bias and speed are the priorities, MoM-NLS is often the more reliable choice in our experiments. Alternatively, the Bayesian approach is more useful when full posterior or predictive inference is needed and prior information is available. The applications illustrate two aspects of the OT model: in Case A, where the OT restrictions are approximately compatible with the data, OT and BLN have similar predictive performance, while in Case B, the diagnostics clearly reveal model mismatch and simpler or more flexible baselines perform as well or better. We therefore view the main contribution as a practical framework for fitting the OT model, understanding when it works, and detecting when it does not. In general, the bivariate beta provides a simple model that can be quickly fitted to data, but our experiments in the misspecified case show that it can sometimes be a rather inflexible distribution and fail to correctly capture correlation between the data coordinates, which is a major feature of any bivariate model.

Future research will focus on testing a plethora of bivariate models beyond the ones considered here on a larger diversity of real datasets. In particular, a systematic empirical comparison of the OT distribution with other bivariate beta constructions, such as those of [Crackel and Flegal \(2017\)](#) and [Trick et al. \(2023\)](#), is planned. Moreover, an extensive comparison with copula models under Beta marginals is warranted. A preliminary comparison of six such models is reported in [Appendix B](#), in which for the analysed applications, it found little statistical gain. The analysis of sensitivity/specificity data, which are useful when analysing imperfect test data under a Bayesian paradigm ([Gelman et al., 2020](#)), deserves special attention.

In summary, we hope to have provided applied researchers with the tools to quickly fit the bivariate beta model to data and diagnose problems, while at the same time furnishing a modern discussion of a complete suite of tools to guide model development and assessment in the case of correlated proportions.

Declarations

Data availability

All code and data are available from <https://github.com/lucamoschen/bivariate-beta>.

Supplementary Information

Online Resource 1 contains the supplementary simulations, diagnostics, figures, and numerical tables.

Competing interests

None.

Funding

For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising. L. M. Moschen acknowledges financial support from a master's scholarship awarded by the School of Applied Mathematics at Fundação Getulio Vargas, where most of this work was carried out, and from a Roth Scholarship at Imperial College London.

Authors' contributions

LMC conceived the study. LMM carried out calculations and implementation. LMM and LMC wrote the paper.

Acknowledgements

We thank Rodrigo Targino and Eduardo Mendes for insightful discussions.

References

- Ali, M. M., Mikhail, N., and Haq, M. S. (1978). A class of bivariate distributions including the bivariate logistic. Journal of multivariate analysis, 8(3):405–412.
- Arellano-Valle, R. B., Harnik, S. B., and Genton, M. G. (2021). On the asymptotic joint distribution of multivariate sample moments. In Advances in Statistics-Theory and Applications, pages 181–206. Springer.
- Arnold, B. C. and Ng, H. K. T. (2011). Flexible bivariate beta distributions. Journal of Multivariate Analysis, 102(8):1194–1202.
- Atchison, J. and Shen, S. M. (1980). Logistic-normal distributions: Some properties and uses. Biometrika, 67(2):261–272.
- Balakrishnan, N. and Lai, C. D. (2009). Continuous bivariate distributions. Springer Science & Business Media.
- Barros, O. A. d. (2015). Estimação dos parâmetros da distribuição beta bivariada: aplicações em severidade de doenças em plantas. PhD thesis, Universidade de São Paulo.
- Betancourt, M. (2016). Diagnosing suboptimal cotangent disintegrations in Hamiltonian Monte Carlo. arXiv preprint arXiv:1604.00695.
- Crackel, R. and Flegal, J. (2017). Bayesian inference for a flexible class of bivariate beta distributions. Journal of Statistical Computation and Simulation, 87(2):295–312.
- Dahabreh, I. J., Trikalinos, T. A., Lau, J., and Schmid, C. (2012). An empirical assessment of bivariate methods for meta-analysis of test accuracy. AHRQ Methods for Effective Health Care 12(13)-EHC136-EF, Agency for Healthcare Research and Quality (US), Rockville, MD.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. The Annals of Statistics, 7(1):1–26.
- Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019). Visualization in bayesian workflow. Journal of the Royal Statistical Society: Series A (Statistics in Society), 182:389–402.
- Gelman, A., Carpenter, B., et al. (2020). Bayesian analysis of tests with unknown specificity and sensitivity. Journal of the Royal Statistical Society Series C, 69(5):1269–1283.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. Statistical science, 7(4):457–472.
- Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics, 44(2):199–213.
- Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association, 88(423):1034–1043.
- Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research, 15(1):1593–1623.
- Joe, H. (1997). Multivariate Models and Dependence Concepts. Number 73 in Monographs on Statistics and Applied Probability. Chapman & Hall, London.
- Libby, D. L. and Novick, M. R. (1982). Multivariate generalized beta distributions with applications to utility assessment. Journal of Educational Statistics, 7(4):271–294.
- Magnussen, S. (2004). An algorithm for generating positively correlated beta-distributed random variables with known marginal distributions and a specified correlation. Computational statistics & data analysis, 46(2):397–406.
- Nadarajah, S. et al. (2007). A new bivariate beta distribution with application to drought data. Metron-International Journal of Statistics, 65(2):153–174.

- Nadarajah, S. and Kotz, S. (2005). Some bivariate beta distributions. *Statistics*, 39(5):457–466.
- Nadarajah, S., Shih, S. H., and Nagar, D. K. (2017). A new bivariate beta distribution. *Statistics*, 51(2):455–474. Publisher: Taylor & Francis eprint: <https://doi.org/10.1080/02331888.2016.1240681>.
- Neal, R. M. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11):2.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York, 2nd edition.
- Ng, K. W., Tian, G.-L., and Tang, M.-L. (2011). *Dirichlet and related distributions: Theory, methods and applications*. John Wiley & Sons.
- Olkin, I. and Liu, R. (2003). A bivariate beta distribution. *Statistics & Probability Letters*, 62(4):407–412.
- Olkin, I. and Trikalinos, T. A. (2015). Constructions for a bivariate beta distribution. *Statistics & Probability Letters*, 96:54–60.
- Sarabia, J. M. and Castillo, E. (2006). Bivariate distributions based on the generalized three-parameter beta distribution. In *Advances in distribution theory, order statistics, and inference*, pages 85–110. Springer.
- Stan Development Team (2022). Stan modeling language users guide and reference manual, 2.29. <https://mc-stan.org>.
- Talts, S., Betancourt, M., Simpson, D., Vehtari, A., and Gelman, A. (2018). Validating bayesian inference algorithms with simulation-based calibration. *arXiv preprint arXiv:1804.06788*.
- Trick, S., Rothkopf, C. A., and Jäkel, F. (2023). Parameter estimation for a bivariate beta distribution with arbitrary beta marginals and positive correlation. *METRON*, 81:163–180.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.
- Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432.
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved R-hat for assessing convergence of MCMC. *Bayesian Analysis*, 16(2):667–718.
- WHO & UNICEF (2025). Share of children vaccinated, by vaccine. [dataset]. Processed by Our World in Data. Based on WHO & UNICEF, *WHO Immunization Data - Vaccination coverage*; United Nations, *World Population Prospects* (2024).

A Proofs

Proof of Proposition 1. Combine the variance and covariance formulas to obtain the expression of the correlation. Fix $\rho \in (-1, 1)$, $a > 0$, and set $\boldsymbol{\alpha} = \frac{a}{2}(1 + \rho, 1 - \rho, 1 - \rho, 1 + \rho)$, which yields $X, Y \sim \text{Beta}(a, a)$ with correlation ρ . $\square$

Proof of Proposition 2. Notice that we can simplify the third and fourth equations since

$$\frac{\alpha_3 + \alpha_4}{s_{\boldsymbol{\alpha}}} = \frac{s_{\boldsymbol{\alpha}} - (\alpha_1 + \alpha_2)}{s_{\boldsymbol{\alpha}}} = 1 - m_1, \quad \frac{\alpha_2 + \alpha_4}{s_{\boldsymbol{\alpha}}} = 1 - m_2.$$

Therefore,

$$v_1 = \frac{m_1(1 - m_1)}{s_{\boldsymbol{\alpha}} + 1} \quad \text{and} \quad v_2 = \frac{m_2(1 - m_2)}{s_{\boldsymbol{\alpha}} + 1}.$$

These relations tell us that the system does not have a solution if

$$\frac{m_1(1 - m_1)}{v_1} \neq \frac{m_2(1 - m_2)}{v_2}.$$

The first two equations of system (5) can be rewritten as a linear system:

$$\begin{aligned} (m_1 - 1)\alpha_1 + (m_1 - 1)\alpha_2 + m_1\alpha_3 + m_1\alpha_4 &= 0, \\ (m_2 - 1)\alpha_1 + m_2\alpha_2 + (m_2 - 1)\alpha_3 + m_2\alpha_4 &= 0, \end{aligned} \iff \begin{aligned} \alpha_1 + \alpha_2 + \frac{m_1}{m_1 - 1}\alpha_3 + \frac{m_1}{m_1 - 1}\alpha_4 &= 0, \\ \alpha_2 + \frac{1 - m_2}{m_1 - 1}\alpha_3 + \frac{m_1 - m_2}{m_1 - 1}\alpha_4 &= 0, \end{aligned}$$

so that we can write α_1 and α_2 as functions of α_3 and α_4 as in (9) and (10).

Based on the expressions above, denote $\alpha_1 = a_3\alpha_3 + a_4\alpha_4$, $\alpha_2 = b_3\alpha_3 + b_4\alpha_4$, $c_3 = a_3 + b_3 + 1$, and $c_4 = a_4 + b_4 + 1$. Then, the third equation can be rewritten as

$$s_\alpha = a_3\alpha_3 + a_4\alpha_4 + b_3\alpha_3 + b_4\alpha_4 + \alpha_3 + \alpha_4 = c_3\alpha_3 + c_4\alpha_4 = \frac{m_1(1 - m_1)}{v_1} - 1,$$

which implies that $\alpha_3 = (m_1(1 - m_1) - v_1 - c_4v_1\alpha_4)/(c_3v_1)$. By substituting this formula into the expressions for α_1 and α_4 , we obtain the first three lines of (7).

Now rewrite the fifth equation using the first two equations from system (5) as follows

$$\begin{aligned} \rho &= \frac{\alpha_1\alpha_4 - \alpha_2\alpha_3}{\sqrt{(\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4)(\alpha_1 + \alpha_3)(\alpha_2 + \alpha_4)}}, \\ &= \frac{\alpha_1\alpha_4 - \alpha_2\alpha_3}{s_\alpha^2 \sqrt{m_1m_2(1 - m_1)(1 - m_2)}}, \\ &= \frac{(m_1 + m_2 - 1)s_\alpha\alpha_4 + \alpha_4^2 - ((1 - m_2)s_\alpha - \alpha_4)((1 - m_1)s_\alpha - \alpha_4)}{s_\alpha^2 \sqrt{m_1m_2(1 - m_1)(1 - m_2)}}, \\ &= \frac{\alpha_4 - (1 - m_1)(1 - m_2)s_\alpha}{s_\alpha \sqrt{m_1m_2(1 - m_1)(1 - m_2)}}, \end{aligned} \tag{18}$$

and the solution is, therefore, $\alpha_4 = s_\alpha \left(\rho \sqrt{m_1m_2(1 - m_1)(1 - m_2)} + (1 - m_1)(1 - m_2) \right)$, which concludes the proof. $\square$

Proposition 4. Let $(X_i, Y_i)_{i=1}^n$ be i.i.d. samples from a distribution on $[0, 1]^2$ with marginal means (m_1, m_2) variances (v_1, v_2) , and finite fourth moments. Define

$$G_n = g(\bar{X}_n, S_{X,n}^2, \bar{Y}_n, S_{Y,n}^2), \quad g(x_1, x_2, x_3, x_4) = x_1(1 - x_1)x_4 - x_3(1 - x_3)x_2,$$

where $\bar{X}_n, \bar{Y}_n$ are sample means and $S_{X,n}^2, S_{Y,n}^2$ the sample variances. If the data come from a bivariate beta distribution, then $g(m_1, v_1, m_2, v_2) = 0$ and

$$\sqrt{n}G_n \xrightarrow{d} N(0, \nabla g(m_1, v_1, m_2, v_2)^\top \Sigma \nabla g(m_1, v_1, m_2, v_2)),$$

where Σ is the asymptotic covariance matrix of $(\bar{X}_n, S_{X,n}^2, \bar{Y}_n, S_{Y,n}^2)$. Moreover, letting $\hat{\sigma}_n^2$ be a consistent estimator of the asymptotic variance,

$$\sqrt{n} \frac{G_n}{\hat{\sigma}_n} \xrightarrow{d} N(0, 1).$$

Proof. By (Arellano-Valle et al., 2021, Proposition 2),

$$\sqrt{n}(\bar{X}_n - m_1, S_{X,n}^2 - v_1, \bar{Y}_n - m_2, S_{Y,n}^2 - v_2) \xrightarrow{d} N(0, \Sigma),$$

for a covariance matrix Σ depending on second and fourth moments of (X, Y) . Since g is continuously differentiable on $\mathbb{R}^4$, the multivariate Delta method yields

$$\sqrt{n}(g(\bar{X}_n, S_{X,n}^2, \bar{Y}_n, S_{Y,n}^2) - g(m_1, v_1, m_2, v_2)) \xrightarrow{d} N(0, \nabla g(m_1, v_1, m_2, v_2)^\top \Sigma \nabla g(m_1, v_1, m_2, v_2)).$$

If the data come from a bivariate beta distribution, then $g(m_1, v_1, m_2, v_2) = 0$, which gives the stated limit distribution of $\sqrt{n}G_n$. Finally, by continuity of ∇g and consistency of the sample moments, $\nabla g(\bar{X}_n, S_{X,n}^2, \bar{Y}_n, S_{Y,n}^2)$ is a consistent estimator of $\nabla g(m_1, v_1, m_2, v_2)$. This provides a consistent variance estimator $\hat{\sigma}_n^2$, and Slutsky's theorem then implies

$$\sqrt{n} \frac{G_n}{\hat{\sigma}_n} \xrightarrow{d} N(0, 1).$$

$\square$

B Copula models on two datasets

We performed a brief analysis of the fit of some popular copula models to the two datasets addressed in the main text: worldwide childhood vaccination coverage (Case A) and ANVISA registrations of COVID-19 diagnostic tests (Case B).

B.1 Copula models considered

We considered six copula families: two elliptical families and four Archimedean families. Recall that a copula model is said to be Archimedean if it admits the representation

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)),$$

where the generator $\varphi : (0, 1] \rightarrow [0, \infty)$ is a continuous, strictly decreasing, convex function that satisfies $\varphi(1) = 0$. The six families are defined as follows.

- **Gaussian copula:**

$$C_{\rho}^{\text{Ga}}(u, v) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \rho),$$

where Φ is the standard normal distribution function, $\Phi_2(\cdot, \cdot; \rho)$ is the bivariate normal distribution function with correlation ρ , and $\rho \in (-1, 1)$. This family exhibits neither upper nor lower tail dependence.

- **Student-t copula:**

$$C_{\rho, \nu}^t(u, v) = T_2(T_{\nu}^{-1}(u), T_{\nu}^{-1}(v); \rho, \nu),$$

where $\nu > 0$ is the parameter controlling the degrees of freedom, T_{ν} is the univariate t distribution function, and T_2 is the bivariate t distribution function with parameters ρ and ν . It exhibits symmetric tail dependence, with upper and lower tail dependence coefficients

$$\lambda_U = \lambda_L = 2T_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right) > 0.$$

- **Clayton copula:** Archimedean with generator $\varphi(t) = (t^{-\theta} - 1)/\theta$ and $\theta > 0$, which gives

$$C_{\theta}^{\text{Cl}}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}.$$

It exhibits lower tail dependence $\lambda_L = 2^{-1/\theta}$ and no upper tail dependence.

- **Gumbel–Hougaard copula:** Archimedean with generator $\varphi(t) = (-\log t)^{\theta}$ and $\theta \geq 1$, which gives

$$C_{\theta}^{\text{Gu}}(u, v) = \exp\left\{-\left[(-\log u)^{\theta} + (-\log v)^{\theta}\right]^{1/\theta}\right\}.$$

It exhibits upper tail dependence $\lambda_U = 2 - 2^{1/\theta}$ and no lower tail dependence.

- **Frank copula:** Archimedean with generator $\varphi(t) = -\log\{(e^{-\theta t} - 1)/(e^{-\theta} - 1)\}$ and $\theta \in \mathbb{R} \setminus \{0\}$, which gives

$$C_{\theta}^{\text{Fr}}(u, v) = -\frac{1}{\theta} \log\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right).$$

The Frank copula is radially symmetric and has no tail dependence.

- **Joe copula:** Archimedean with generator $\varphi(t) = -\log\{1 - (1 - t)^{\theta}\}$ and $\theta \geq 1$, which gives

$$C_{\theta}^{\text{Jo}}(u, v) = 1 - \left[(1 - u)^{\theta} + (1 - v)^{\theta} - (1 - u)^{\theta}(1 - v)^{\theta}\right]^{1/\theta}.$$

It exhibits upper tail dependence $\lambda_U = 2 - 2^{1/\theta}$ and no lower tail dependence.

B.2 Fitting, model comparison, and diagnostic

Instead of doing full maximum likelihood, we adopt a two-stage approach, commonly called inference functions for margins (IFM). First, the marginal parameters are estimated separately:

$$(\hat{\alpha}_j, \hat{\beta}_j) = \arg \max_{\alpha_j, \beta_j} \sum_{i=1}^n \log f_j(\cdot; \alpha_j, \beta_j),$$

for $j = 1, 2$. In the second stage, the estimated marginals are used to form pseudo-observations $\hat{u}_i = F(x_i; \hat{\alpha}_1, \hat{\beta}_1)$ and $\hat{v}_i = G(y_i; \hat{\alpha}_2, \hat{\beta}_2)$, and the copula parameter is estimated by

$$\hat{\theta}_{\text{IFM}} = \arg \max_{\theta} \sum_{i=1}^n \log c(\hat{u}_i, \hat{v}_i; \theta).$$

The IFM estimator is consistent and asymptotically normal under standard regularity conditions, though it is generally less efficient than maximum likelihood estimation for the full model.

For evaluation of (relative) goodness-of-fit, we used commonly employed information criteria in AIC and BIC:

$$\text{AIC} = -2\hat{\ell} + 2k, \quad \text{BIC} = -2\hat{\ell} + k \log n,$$

where $\hat{\ell}$ is the maximum log-likelihood and k is the number of estimated parameters. Regarding the diagnostic, let C_n denote the empirical copula, defined for $\mathbf{u} = (u, v) \in [0, 1]^2$ as

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\hat{u}_i \leq u, \hat{v}_i \leq v).$$

The Cramér–von Mises test statistic of [Genest et al. \(2009\)](#) is

$$S_n = n \int_{[0,1]^2} \left[C_n(u, v) - C(u, v; \hat{\theta}) \right]^2 dC_n(u, v) = \sum_{i=1}^n \left[C_n(\hat{u}_i, \hat{v}_i) - C(\hat{u}_i, \hat{v}_i; \hat{\theta}) \right]^2$$

and p -values are obtained by parametric bootstrap ([Genest et al., 2009](#)), which we perform with $B = 200$ bootstrap replications.

B.3 Results

We begin with the data on the coverage of the diphtheria-measles vaccine in Section 6.1 of the main text (Case A). [Figure B.15](#) shows the data and the PIT transformation after the first stage of fitting the marginals. From this, it is clear that the Beta marginals fit somewhat well. Then, in [Figure B.16](#) we show the contours for the observed data and the fitted copulas. Finally, the goodness-of-fit and bootstrap diagnostic results are shown in [Figure B.17](#). The results show that the Clayton copula fits well and achieves both the lowest AIC/BIC and also leads to an S_n statistic compatible with the observed data (p -value > 0.05). In conclusion, for this dataset it is possible to find a well-fitting copula model in the Clayton model. However, as we show in the main text, the OT model also fits well.

For Case B, the results of the same analyses are presented in [Figure B.18](#), [B.19](#) and [B.20](#). We can clearly see that for this dataset no model really fits well, as reflected by the bootstrap p -values, for instance. As mentioned, our custom diagnostic for the OT model also detects problems. Interestingly, from the left panel of [Figure B.18](#), it seems that part of the problem might lie in specifying Beta marginals.

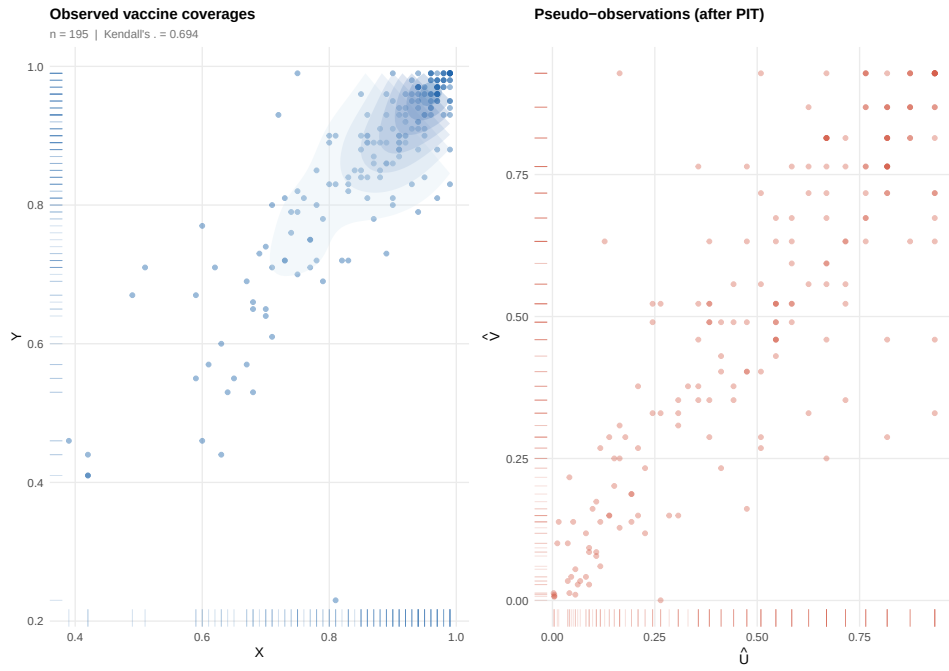


Figure B.15: **Scatterplot and transformed percentiles (Case A)**. We show the data on the left and the probability integral transform (PIT) values from the maximum likelihood marginals on the right panel.

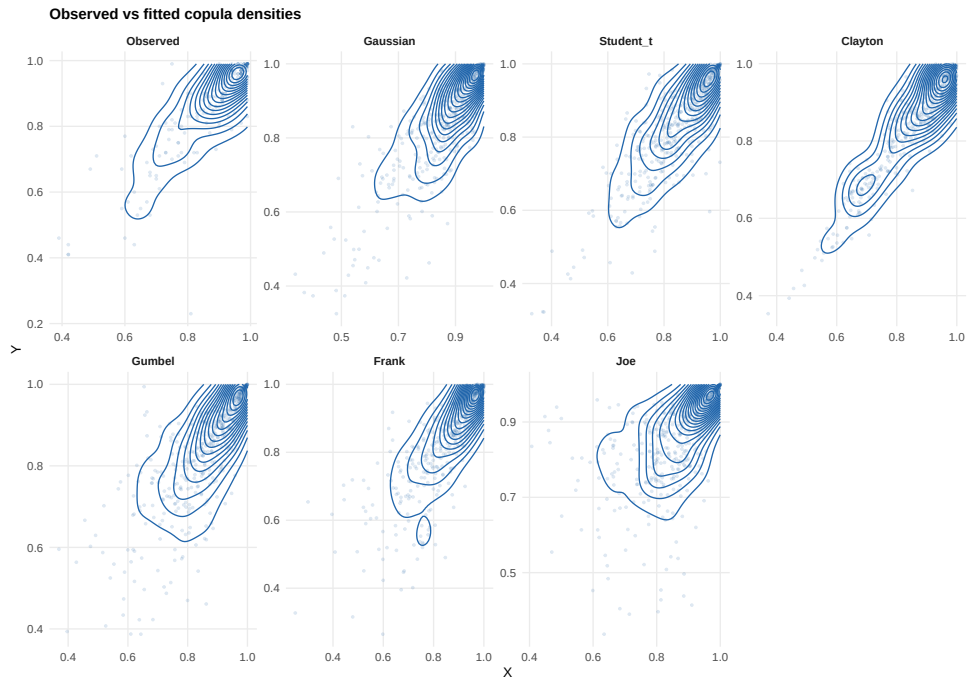


Figure B.16: **Contour plots for the fitted formulas (Case A)**. We show the contour plot for the observed data and the six copula families considered here.

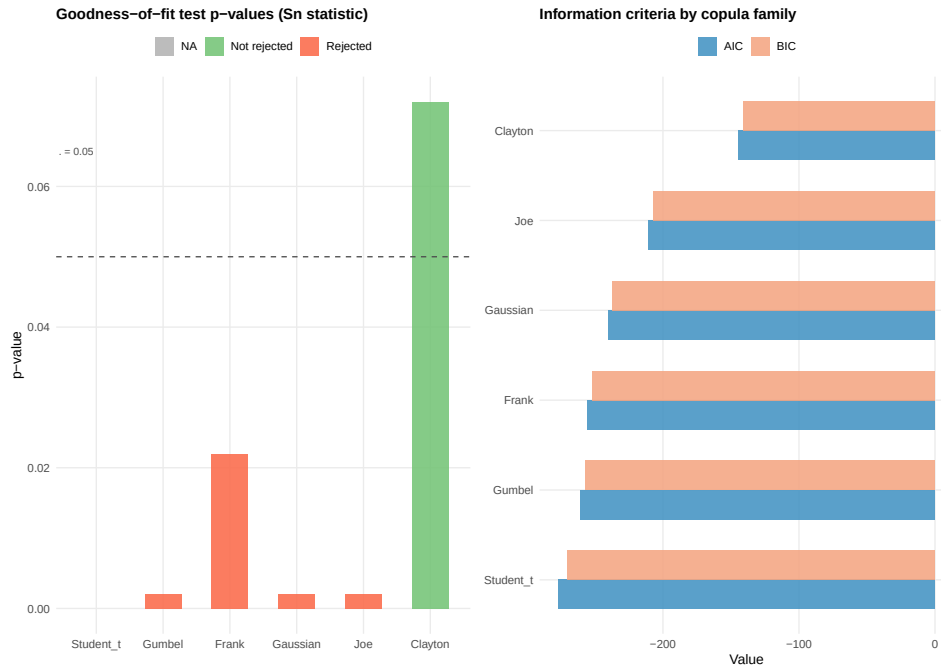


Figure B.17: **Bootstrap adequacy and (relative) goodness-of-fit metrics via information criteria (Case A).** On the left, we show the bootstrap p -values obtained for the S_n statistic under each parametric copula family. The right panel shows the AIC and BIC values for each model, with lower being better – in this case the Clayton model is to be selected.

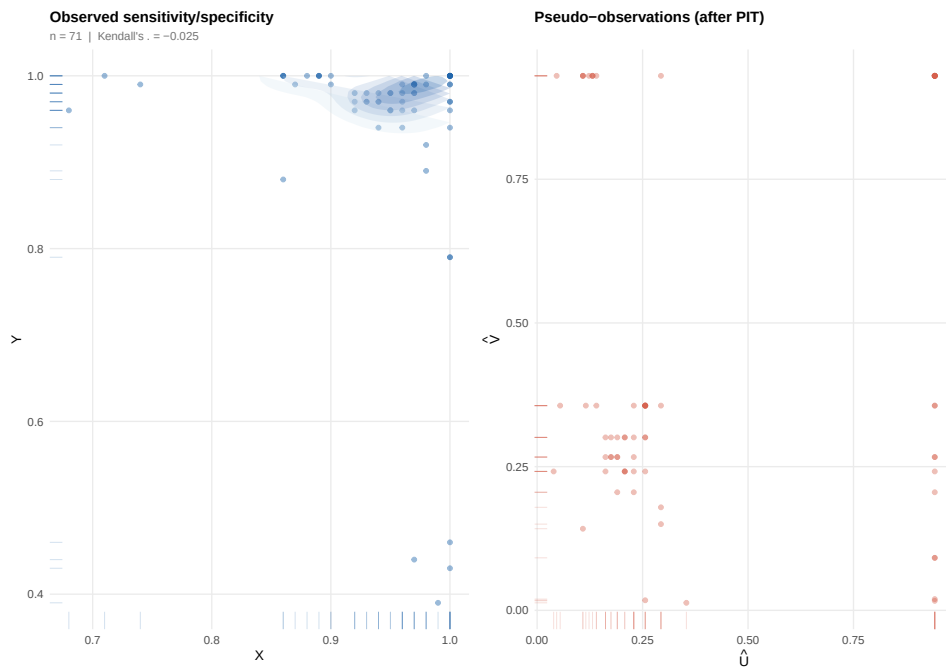


Figure B.18: **Scatterplot and transformed percentiles (Case B).** Same setting as in [Figure B.15](#).

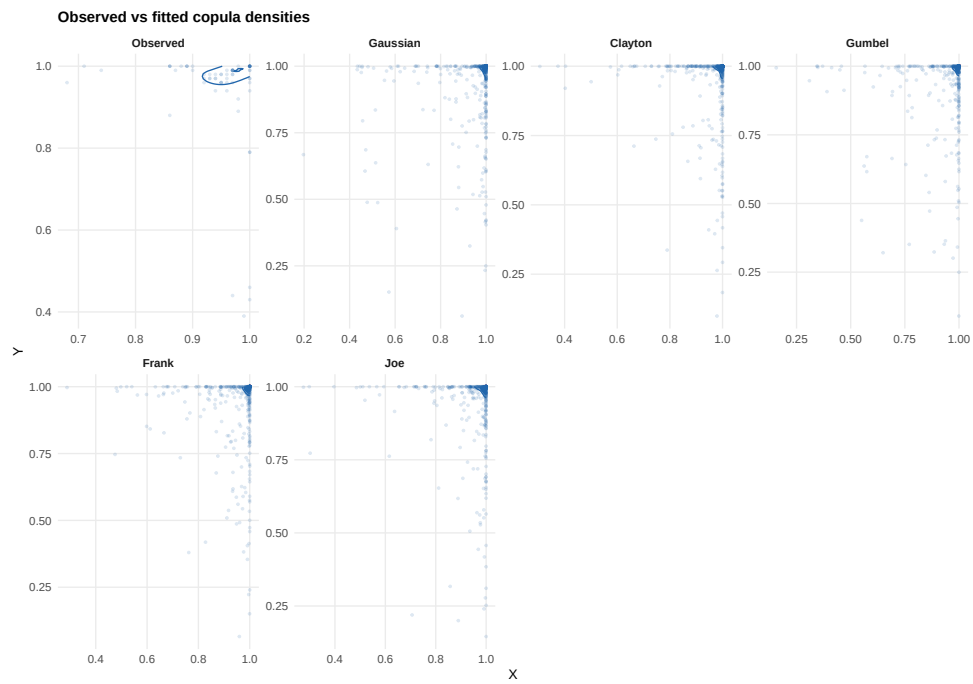


Figure B.19: Contour plots for the fitted formulas (Case B). Same setting as in Figure B.16.

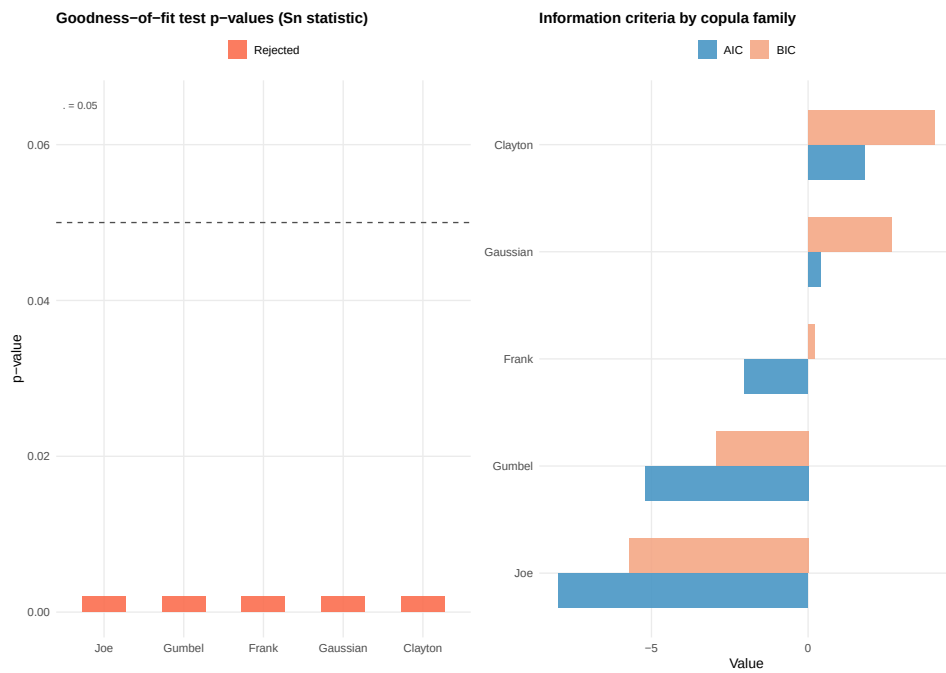


Figure B.20: Bootstrap adequacy and (relative) goodness-of-fit metrics via information criteria (Case B). Same setting as in Figure B.17.