Oxford Bulletin of Economics and Statistics

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Bayesian Inference for Joint Estimation Models Using Copulas to Handle Endogenous Regressors

Rouven E. Haschka[1,2]

[1]Chair of Data Analytics, Technical University of Kaiserslautern-Landau, Kaiserslautern, Germany | [2]Institute of Strategy and Management, Corvinus University, Budapest, Hungary

Correspondence: Rouven E. Haschka (rouven.haschka@rptu.de)

Received: 6 February 2025 | **Revised:** 9 September 2025 | **Accepted:** 17 September 2025

Keywords: Bayesian inference | copula function | endogeneity | Markov chain Monte Carlo

ABSTRACT

This study proposes a Bayesian approach for finite-sample inference of the Gaussian copula endogeneity correction. Extant studies use frequentist inference, build on a priori computed estimates of marginal distributions of explanatory variables, and use bootstrapping to obtain standard errors. The proposed Bayesian approach facilitates precise statistical inference through Markov chain Monte Carlo simulation techniques and requires neither asymptotics nor tuning. It is one-step, where regression coefficients, error variance, copula correlations, and probability masses of marginals are treated as random and sampled jointly, rather than fixed or pre-estimated. Simulation experiments illustrate finite-sample performance, complemented by an empirical application.

JEL Classification: C11, C14, C21, C51, C61, M31

1 | Introduction

The frequentist estimator proposed by Park and Gupta [1] is widely used to handle regressor-endogeneity without external, that is, instrumental information [2]. It is a two-stage approach that relies on Sklar's theorem [3] and copula functions to approximate the joint distribution of endogenous regressors and the structural error term. In the first step, data-driven cumulative distribution functions (cdfs) of endogenous regressors are obtained. These, along with an assumed distribution for structural errors, are used as plug-in estimates for the copula function in the second step, where the joint distribution is estimated using maximum likelihood (ML). Park and Gupta [1] showed that maximising the likelihood based on the joint distribution is equivalent to incorporating 'copula correction terms' as additional regressors in a linear regression model, allowing estimation by least squares (OLS) as well (see also [4]). Accordingly, pre-estimation of the marginal distributions of endogenous regressors makes it a two-step procedure. Hence, uncertainty is not completely characterised, as the uncertainties that arise within the first stage are typically overlooked in the second stage estimation. With a nonlinear regressor-error relation, model identification does not require any instrumental variables [5]. This can be established when (i) the endogenous regressors are non-normal and (ii) structural errors are assumed to be normally distributed. Using a Gaussian copula additionally requires assuming (iii) Gaussian dependence structures.

Against this background, this study proposes a one-step Bayesian approach for precise finite-sample inference in copula-based identification of endogenous regressors. In contrast to frequentist estimation (we refer to [1, 4, 5]), the proposed approach does not rely on fixed estimates for cdfs or correlations among regressors. Instead, marginal distributions are modelled as non-parametric functions using Dirichlet priors, and the full copula correlation matrix is treated as a random variable using an inverse Wishart prior. We use an MCMC sampling algorithm that combines Gibbs steps and adaptive Metropolis-Hastings to draw probability

masses for regressor distributions and the correlation matrix of the Gaussian copula jointly with regression coefficients and variance of structural errors. Weakly informative, but proper, priors enhance stability of the posterior, especially in regions where the likelihood is flat, without imposing dogmatic assumptions. The approach provides valid, credible intervals without reliance on asymptotic arguments. The intervals convey the uncertainty associated with the parameter estimates and avoid the simplicity of point null hypothesis tests, which marks a key strength of the Bayesian approach. Finally, numerically stable and adaptive proposal densities that do not require parameter tuning yield computationally efficient MCMC simulation techniques.

Simulation experiments demonstrate that the proposed Bayesian approach is a promising alternative to existing frequentist estimation, especially in small samples, because uncertainties associated with plug-in estimates can be avoided. We demonstrate that bootstrap-based frequentist standard errors are not valid when the model is not identified. Consequently, relying on standard errors as an indicator of model non-identification, as commonly done in the frequentist approach, is misleading. To illustrate the Bayesian approach in practice, we apply it to scanner data from Dominick's Finer Foods, a former major supermarket chain in the Chicago area. The aggregated dataset records weekly store-level information on sales, prices, and promotions for multiple orange juice brands. This allows us to examine brand-specific demand while accounting for cross-brand effects.

The paper begins with a brief review of recent developments in copula-based identification for endogenous regressors. Subsequently, Bayesian model estimation is discussed in Section 3. Monte Carlo studies are conducted in Section 4 to assess the inferential performance of the Bayesian approach. An empirical analysis is presented in Section 5 to address (cross-)price endogeneity in demand estimation, and Section 6 concludes. Full conditionals for Gibbs sampling, score vectors, and working weights for proposal densities are derived in the Web Appendices A–C. The MCMC sampling algorithm is described in detail in Web Appendix D, and Web Appendix E describes the implementation of the Bayesian approach in the empirical application.

2 | Recent Developments in Copula-Based Identification

In some fields, applied researchers increasingly rely on the copula approach introduced by Park and Gupta [1]—hereafter abbreviated as PG—to address endogenous regressors in empirical analyses. The method has gained particular prominence in marketing and management research (for a detailed overview of applications, see [6, 7]). For clarification, consider a stylised linear regression model:

$$y_i = \alpha + \beta x_i + \delta z_i + e_i, i = 1, \ldots, N \tag{1}$$

and the researcher is interested in estimating the marginal effects β and δ on the outcome y_i. Correlation between the continuous regressor z_i and the unobserved error e_i generates the endogeneity problem, and the regressor x_i is assumed uncorrelated with the error, that is, exogenous. Instead of deriving estimates from a likelihood based on the marginal distribution of e (assuming normality, this yields the (biased) least squares estimates), Park and Gupta [1] propose targeting the joint distribution of z and e. They assume normality for e and, by means of Sklar's theorem [3], use a Gaussian copula to link the marginals with the ecdf of z as plug-in estimates for the copula function. In addition, model identification requires non-normality of z because otherwise it is difficult to distinguish the variations as results of the endogenous regressor from the variation due to the error. Park and Gupta [1] demonstrated that estimates derived from the likelihood of this joint distribution are consistent.

A large volume of recent studies investigates the performance of PG under different scenarios, particularly to assess the adverse effects of model non-identification [8, 9] and to provide practical guidelines for empirical applications, such as the sufficient deviation from normality of endogenous regressors [6, 10]. Qian et al. [11] use simulations to examine PG's handling of higher-order terms, finding that for main effects, accounting for higher-order endogeneity is unnecessary. Monte Carlo studies [6, 9] show that PG's performance depends on sample size, the distribution of endogenous regressors, and the presence of an intercept: Small samples with an intercept require highly non-normal regressors, whereas large samples yield reliable results.

Eckert and Hohberger [8] highlight biases when identifying assumptions fail; skewed structural errors lead to bias, while symmetric errors are handled well (see also [1]). Despite potential bias from non-parametric regressor-error dependence [8], other studies demonstrate the Gaussian copula's robustness to various non-Gaussian dependencies [1, 5].

Recent studies propose generalisations of PG and often relax model assumptions. Haschka [5] shows that PG assumes independence between exogenous and endogenous regressors, which causes bias when these covariates are correlated, and he models their joint distribution with structural errors using ML estimation and bootstrapped standard errors. Yang et al. [4] demonstrate that Haschka's model can also be estimated via two-stage least squares (for a simpler setup that still uses bootstrap standard errors, see [12]). While Yang et al. [4] provide a proof of consistency, Breitung et al. [13] note that it assumes knowledge of the endogenous regressor distribution, rarely available empirically.

Tran and Tsionas [14] generalise PG further by estimating marginal distributions of endogenous regressors and structural errors simultaneously via one-step sieve ML, relaxing the normality assumption to allow differing distributions. Although asymptotic properties hold, Qian and Xie [15] point out scalability issues with multiple regressors and the need to preselect tuning parameters. They propose a semiparametric odds ratio model nesting PG via profile ML, retaining normal structural errors but allowing regressor-error dependencies within the exponential family. However, modelling all pairwise dependencies among many explanatory variables requires numerous candidate models. Recently, Haschka [16] extends PG to regression models with nonlinear effects and varying coefficients via P-splines, using the Bayesian framework introduced in this article.

Taken together, in reviewing the studies that have examined performance of the copula approach and proposed generalisations in different directions, as well as the extensive empirical literature

review by Becker et al. [6], the following considerations are worth remarking. Existing studies (both theoretical and methodological) either use ML or OLS. Accordingly, the approach has so far been carried out only from a frequentist perspective. Furthermore, since a priori-calculated marginal distributions of the model regressors are used as plug-in estimates for the copula function, inferential results such as standard errors, confidence intervals and test decisions require bootstrapping procedures (with the procedures by Tran and Tsionas [14] and Qian and Xie [15] as exceptions).

3 | Method

Consider the general linear regression model with multiple endogenous and exogenous variables:

$$y_i = \alpha + \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\delta} + e_i \tag{2}$$

where $i = 1, \ldots, N$ indexes either time or cross-sectional units, y_i is the dependent variable, α is an intercept, and coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are corresponding linear effects of a $(L \times 1)$ vector of exogenous regressors $\boldsymbol{x}_i$ and a $(K \times 1)$ vector of endogenous regressors $\boldsymbol{z}_i$. Note that while endogenous regressors $\boldsymbol{z}_i$ are at best continuous or at least discrete with multiple outcomes, $\boldsymbol{x}_i$ may also include dummies or discrete variables with few outcomes. Furthermore, if interactions or higher-order terms of the (endogenous) explanatory variables are considered, these are also collected in $\boldsymbol{x}_i$ [11].

Regressors in $\boldsymbol{z}_i$ and $\boldsymbol{x}_i$ are allowed to be correlated. Finally, e_i represents an unobserved normally distributed structural error term, that is, $e_i \sim \mathrm{N}(0, \sigma^2)$. Correlation between $\boldsymbol{z}_i$ and e_i generates the endogeneity problem such that estimates derived for both $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ will be inconsistent if the estimation procedure does not involve an endogeneity correction. Note that if $\boldsymbol{z}_i$ is uncorrelated with $\boldsymbol{x}_i$, $\boldsymbol{\beta}$ can be estimated consistently. While we restrict the model outline to the standard linear regression model, the proposed approach can be easily extended to random-coefficient linear regression models to account for slope heterogeneity.

3.1 | Deriving the Likelihood

To tackle the endogeneity problem introduced by regressor-error correlation, we follow Haschka [5] and determine the joint distribution of $\boldsymbol{z}_i$, $\boldsymbol{x}_i$, and e_i by means of a copula function. The approach is based on Park and Gupta [1], who target the joint distribution of $\boldsymbol{z}_i$ and e_i and thus assume independence between endogenous and exogenous regressors, but Haschka [5] highlights the risk of omitted-variable bias when this assumption is violated.

Let $\boldsymbol{u}_{z,i} = (F_{z1}(z_{1,i}), \ldots, F_{zK}(z_{K,i}))'$, $\boldsymbol{u}_{x,i} = (F_{x1}(x_{1,i}), \ldots, F_{xL}(x_{L,i}))'$, and $u_{e,i} = \Phi(e_i; \sigma^2)$ denote probability integral transformed explanatory variables and error, respectively, such that margins are $(\boldsymbol{u}_{z,i}, \boldsymbol{u}_{x,i}, u_{e,i})' \in [0,1]^{K+L+1}$. The F's denote respective marginal cdfs and $\Phi(e_i; \sigma^2)$ is the cdf of normal distribution with variance σ^2. Using a Gaussian copula, $\boldsymbol{\xi}_{z,i} = (\Phi^{-1}(u_{z,1i}), \ldots, \Phi^{-1}(u_{z,Ki}))'$, $\boldsymbol{\xi}_{x,i} = (\Phi^{-1}(u_{x,1i}), \ldots, \Phi^{-1}(u_{x,Li}))'$, and $\xi_{e,i} = \Phi^{-1}(u_{e,i}) = e_i/\sigma$ follow a standard multivariate normal distribution of dimension $(K + L + 1)$ with correlation matrix Σ.

According to Papadopoulos [10], the Gaussian copula is most flexible because it allows modelling multi-dimensional dependence structures separately and has many desirable properties (see also [17]). Specifically, it is able to model the full range of Pearson correlation, and specific consistency (i.e., robustness) theory for quasi-maximum likelihood estimation for Gaussian copula-based models exists [18]. Furthermore, if the true dependence is different from what the Gaussian copula assumes, the literature has demonstrated its robustness in capturing various non-Gaussian dependencies [1, 5, 6]; although the true dependence should not be non-parametric [8] or asymmetric [10], something we will elaborate on later.

By means of Sklar's theorem [3], the joint distribution can then be written as:

$$\begin{aligned} f(\boldsymbol{z}_i, \boldsymbol{x}_i, e_i) &= \frac{1}{\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2} \begin{pmatrix} \boldsymbol{\xi}_{z,i} \\ \boldsymbol{\xi}_{x,i} \\ \xi_{e,i} \end{pmatrix}' (\Sigma^{-1} - I) \begin{pmatrix} \boldsymbol{\xi}_{z,i} \\ \boldsymbol{\xi}_{x,i} \\ \xi_{e,i} \end{pmatrix} \right) \\ &\quad \times \phi(e_i; \sigma^2) \times \prod_{k=1}^{K} f_{zk}(z_{ki}) \times \prod_{l=1}^{L} f_{xl}(x_{li}) \end{aligned} \tag{3}$$

such that $\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\boldsymbol{z}_i, \boldsymbol{x}_i, e_i)\, dz_{1i} \ldots dz_{Ki} dx_{1i} \ldots dx_{Li}\, de_i = 1$, that is, (3) is a probability density function. The copula in the first line links the structural error and all explanatory variables to encode information about the entire dependence in the model. More precisely, the correlation matrix Σ in Equation (3) formalises the (linear) dependence among $(\boldsymbol{\xi}_{z,i}, \boldsymbol{\xi}_{x,i}, \xi_{e,i})'$ by means of pairwise Pearson correlations. Hence, the copula captures joint variation, whereas univariate densities ϕ and f in the second line describe marginal behaviour.

Before deriving the likelihood, we briefly recall model identification based on the joint distribution in Equation (3). First, the Gaussian copula requires linear dependence structures and continuous margins [17], although it has been shown that this restriction can be relaxed by allowing for discrete endogenous regressors with multiple outcomes [1]. Thus, the approach is restricted to endogenous regressors with many outcomes and linear forms of endogeneity under normal score transformation, that is, correlation between $\boldsymbol{\xi}_z$ and ξ_e should be of Pearson-type [19]. Second, model identification does not require any instrumental variables (IV) if the regressor-error relation is nonlinear, which can be established if structural errors are normal but endogenous regressors are not.[1] Then, $\mathbb{E}[e|\boldsymbol{z}]$ becomes a nonlinear function and this nonlinearity identifies linear coefficients $\boldsymbol{\delta}$ without IV information [5]. However, if endogenous regressors are normal, identification breaks down because the copula fails to distinguish the variations as a result of endogenous regressors from the variation due to the structural error. The reason for that is that $\mathbb{E}[e|\boldsymbol{z}]$ then becomes a linear function, and this linearity makes it impossible to separate linear coefficients $\boldsymbol{\delta}$ [5].

Extant studies compute cdfs of regressors a priori, most frequently by means of empirical cdfs, and then use these to generate $\boldsymbol{\xi}_{z,i}$ and $\boldsymbol{\xi}_{x,i}$.[2] Haschka [5] further computes empirical (Pearson) correlations between $\hat{\boldsymbol{\xi}}_z$ and $\hat{\boldsymbol{\xi}}_x$ as a preliminary step. He subsequently inserts these correlation estimates into

corresponding elements in matrix Σ (a similar approach is adopted by Haschka [12] and Yang et al. [4], who employ residuals from the first-stage regression of $\hat{\xi}_z$ on $\hat{\xi}_x$ as plug-in estimates). Tran and Tsionas [14] report that when mean correlations between endogenous and exogenous variables are estimated empirically and used as plug-in values in the second stage, this can lead to severe finite-sample problems.

By contrast, we neither use a priori computed CDFs nor first-stage estimation, but rather extract all information (including margins and correlations) in one single step within the Bayesian approach. For a generic regressor $\varpi \in (\boldsymbol{z}, \boldsymbol{x})'$, we treat univariate densities f_ϖ in Equation (3) as non-parametric functions that assign nonzero probability mass $\lambda_\varpi = (\lambda_{\varpi,1}, \ldots, \lambda_{\varpi,m})'$ on uniquely observed data values of ϖ (for a similar approach, see [20]); and let $\upsilon_\varpi = (1, \ldots, m_\varpi)'$ collect counts of m_ϖ unique values. The corresponding margins in the copula function are then obtained by taking cumulative sums of λ_ϖ of ascendingly ordered ϖ, that is, $u_{\varpi,j} = \sum_{j=1}^{m_\varpi} \lambda_{\varpi,j}$. Accordingly, elements in the copula function are now specified as $\xi_{\varpi,i} = \Phi^{-1}(u_{\varpi,i}(\lambda_\varpi))$ with $u_{\varpi,i}(\lambda_\varpi)$ assigning cumulative probability masses similar to a cdf. Through this perspective, note that we assume that υs exhibit categorical distributions, that is, $\upsilon_\varpi \sim \text{Cat}(m_\varpi, \lambda_\varpi)$.

Within the Bayesian approach, we treat these probability masses as random rather than fixed (as would be the case, for instance, when using an empirical cdf), and sample them simultaneously along with the full covariance matrix of the Gaussian copula and regression coefficients, which are also assumed random.[3] This allows for simultaneous inference, and our approach is not subject to uncertainties brought in by using plug-in estimates since we do not estimate anything a priori. Through the assumption that $\boldsymbol{x}$ is exogenous, entries in Σ which capture correlations between $\xi_{x,i}$ and $\xi_{e,i}$ are restricted to zero following Haschka [5]. Due to the zero restrictions, the number of unknown elements in Σ reduces by L, and model identification due to endogeneity is only subject to $z_1, \ldots, z_K$ but not to $x_1, \ldots, x_L$.

To derive the likelihood, note that in contrast to Haschka [5], functions f in Equation (3) cannot be dropped because, as pmfs of categorical distributions, they contain probability masses of unique values of explanatory variables that are also of interest:

$$
\begin{aligned}
&L(\boldsymbol{\theta}|y, \boldsymbol{z}, \boldsymbol{x}) \\
&\propto \prod_{i=1}^{N} \frac{1}{\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}\begin{pmatrix} \xi_{z,i} \\ \xi_{x,i} \\ e_i/\sigma \end{pmatrix}' \left(\Sigma^{-1} - I\right) \begin{pmatrix} \xi_{z,i} \\ \xi_{x,i} \\ e_i/\sigma \end{pmatrix}\right) \\
&\quad \times \phi(e_i; \sigma^2) \times \prod_{k=1}^{K} \left([\upsilon_{zk} = 1]\lambda_{zk,1} + \ldots + [\upsilon_{zk} = m_{zk}]\lambda_{zk,m}\right) \\
&\quad \times \prod_{l=1}^{L} \left([\upsilon_{xl} = 1]\lambda_{xl,1} + \ldots + [\upsilon_{xl} = m_{xl}]\lambda_{xl,m}\right) \qquad (4)
\end{aligned}
$$

where $[\cdot = j]$ is the Iverson bracket. In Equation (4), $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2, \Sigma, \lambda_{z1}, \ldots, \lambda_{zK}, \lambda_{x1}, \ldots, \lambda_{xL})'$ collects all unknown coefficients, which are treated as random variables, and $e_i = y_i - \alpha - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{z}_i'\boldsymbol{\delta}$. Subvector $(\alpha, \boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2, \Sigma)'$ is of dimension $((K + L + 1)^2 - (K + L + 1))/2 + K + 2$. Note that $\boldsymbol{\xi}_z$ and $\boldsymbol{\xi}_x$ can only be obtained given the λ's.

3.2 | Prior Specification

The likelihood combines a Gaussian copula with normal density for structural errors and categorical distributions for regressor values, yielding an always unimodal but possibly flat function. To ensure numerical stability of the posterior, we employ proper priors for all unknown coefficients. This prevents the posterior from becoming flat even if the likelihood is flat, as the combination with a non-flat prior ensures this. We choose independent and weakly informative priors to avoid imposing strong beliefs, prevent the prior from dominating and distorting the posterior, and mitigate potential identification issues that may arise from dependent priors.

3.2.1 | Linear Effects

Shrinkage priors are used for regression coefficients [21]. This regularisation technique not only improves the numerical stability of the posterior but also enhances the robustness of the estimates, particularly in situations with limited sample sizes. Following Carvalho et al. [22], we define the following hierarchical structure for every coefficient $\gamma \in (\alpha, \boldsymbol{\beta}, \boldsymbol{\delta})'$ and adhere to a global shrinkage (common to all γ's):

$$
\gamma|\varphi \sim \text{N}(0, \varphi^2), \quad \varphi|\tau \sim C^+(0, \tau), \quad \tau|\imath \sim C^+(0, \imath), \quad p(\imath) \propto 1/\imath \qquad (5)
$$

where C^+ is the half-Cauchy distribution. Treating variances $\imath$, τ, and φ as nuisance parameters, after marginalising them out, we obtain a 'horseshoe-type' prior given by:

$$
p(\gamma) = \int_0^\infty \int_0^\infty \int_0^\infty p(\gamma|\varphi, \tau, \imath)\, d\imath\, d\tau\, d\varphi \propto \sqrt{\frac{1}{\gamma^2}} \qquad (6)
$$

3.2.2 | Variance of Structural Error

As is common [23], we assign an inverse Gamma prior to the error variance, that is, $\sigma^2 \sim \text{IG}(a, b)$, with corresponding density

$$
p(\sigma^2|a, b) = \frac{b^a}{\Gamma(a)}(1/\sigma^2)^{a+1} \exp\{-b/\sigma^2\} \qquad (7)
$$

with common hyperparameters $a = b = 0.001$ [24, 25].

3.2.3 | Correlation Matrix

Under the copula representation in Equation (3), it is $\boldsymbol{\xi}|\Sigma \sim \text{MN}(\boldsymbol{0}, \Sigma)$, $\boldsymbol{\xi} = (\boldsymbol{\xi}_z, \boldsymbol{\xi}_x, \boldsymbol{\xi}_e)'$. Since a proper prior distribution for a correlation matrix is hard to find, we follow Zhang et al. [26] and do not assign a prior directly to Σ, but rather to the corresponding covariance matrix W (for an application in a similar context, see [25]), which is particularly useful in Bayesian copula modelling [27]. Due to its conjugacy in multivariate normal distributions, a common choice of prior distribution for covariance matrices is the inverse Wishart distribution, that is, $W \sim \mathcal{W}^{-1}(\Psi, \nu)$. The density is given by

$$
p(W|\Psi, \nu) = \frac{|\Psi|^\nu |W|^{-(\nu+K+L+1)/2} \exp\{-\text{tr}(\Psi W^{-1})/2\}}{2^{\nu(K+L+1)/2}\Gamma(\nu/2)} \qquad (8)
$$

with $\Psi = \boldsymbol{I}$ and $\nu = K + L + 1$.

Although entries in Σ which capture correlations between $\boldsymbol{\xi}_x$ and ξ_e will be restricted to zero to ensure exogeneity of $\boldsymbol{x}$, the prior is assigned on the full matrix W; and the full conditional of W can be derived and used for Gibbs sampling (see Web Appendix A). We use Cholesky factorisation $\Sigma = \text{diag}(W)^{-\frac{1}{2}} W \text{diag}(W)^{-\frac{1}{2}}$ to transform samples for W into samples for Σ [26].

3.2.4 | Probability Masses formulating CDFs

Because of its conjugacy with the categorical distribution, we assign a Dirichlet prior on $\lambda_\varpi = (\lambda_{\varpi,1}, \ldots, \lambda_{\varpi,m})'$, that is, $\lambda_\varpi \sim \text{Dir}(\boldsymbol{\omega}_\varpi)$, with density

$$p(\lambda_{\varpi,1}, \ldots, \lambda_{\varpi,m} | \omega_{\varpi,1}, \ldots, \omega_{\varpi,m}) = \frac{1}{\text{B}(\boldsymbol{\omega}_\varpi)} \prod_{j=1}^{m_\varpi} \left(\lambda_{\varpi,j}\right)^{\omega_{\varpi,j}-1} \quad (9)$$

with $\text{B}(\boldsymbol{\omega}) = \prod_{j=1}^{m_\varpi} \Gamma\left(\omega_{\varpi,j}\right) / \Gamma(\sum_{j=1}^{m_\varpi} \omega_{\varpi,j})$, and $\sum_{j=1}^{m_\varpi} \lambda_{\varpi,j} = 1$ and $\lambda_{\varpi,j} \in (0,1)$ for all $j \in \{1, \ldots, m_\varpi\}$ unique values of ϖ for $i = 1, \ldots, N$. In Equation (9), $\boldsymbol{\omega}_\varpi = (\omega_{\varpi,1}, \ldots, \omega_{\varpi,m})'$ are hyperparameters also of dimension m_ϖ that can be interpreted as weights of individual observation points. With $\boldsymbol{\omega}_\varpi = \mathbf{1}$, the prior is assigned to each regressor $\varpi \in (z_1, \ldots, z_K, x_1, \ldots, x_L)'$, subject to the number of corresponding unique data values m_ϖ. The full conditional of $\lambda = (\lambda_{z1}, \ldots, \lambda_{zK}, \lambda_{x1}, \ldots, \lambda_{xL})'$ can be derived and used for Gibbs sampling (see Web Appendix B).[4]

3.3 | Bayesian Inference

Bayesian inference is based on the posterior of the model given by:

$$p(\boldsymbol{\theta} | y, \boldsymbol{z}, \boldsymbol{x}) \propto L(\boldsymbol{\theta} | y, \boldsymbol{z}, \boldsymbol{x})\ p(\boldsymbol{\theta}) \quad (10)$$

where the likelihood is given in Equation (4) and $p(\boldsymbol{\theta}) \propto p(\alpha)p(\boldsymbol{\beta})p(\boldsymbol{\delta})p(\sigma^2)p(W)p(\lambda)$. We use MCMC simulation techniques due to the high dimensionality of $\boldsymbol{\theta}$ and the fact that analytical derivation of the joint posterior distribution is intractable. Combining likelihood with priors, we have the following four sets of conditionals:

$$\begin{aligned} &(\alpha, \boldsymbol{\beta}, \boldsymbol{\delta}) | \sigma^2, W, \lambda \quad \sigma^2 | (\alpha, \boldsymbol{\beta}, \boldsymbol{\delta}), W, \lambda \\ &W | (\alpha, \boldsymbol{\beta}, \boldsymbol{\delta}), \sigma^2, \lambda \quad \lambda | (\alpha, \boldsymbol{\beta}, \boldsymbol{\delta}), \sigma^2, W \end{aligned} \quad (11)$$

While we can derive full conditional distributions for W and λ as shown in Web Appendices A and B, respectively, full conditionals of α, $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, and σ^2 are also intractable. Consequently, we use a sampling algorithm that is a combination of Gibbs steps (updating W and λ by drawing from corresponding full conditional distributions) and a generalised Metropolis-Hastings step, where proposal distributions are adaptive and based on approximations to the full conditionals (which are derived in Web Appendix C). A detailed description of the MCMC sampler is given in Web Appendix D.

4 | Simulation Studies

In this section, we assess the performance of the proposed Bayesian MCMC through simulation studies.[5] We examine scenarios of both model identification and non-identification to evaluate the reliability and exactness of the MCMC sampling procedure. The simulated model structure is similar to previous studies [4–6, 8]. The goal is to assess the effect of the priors, the accuracy of posterior means as point estimates, and the validity of credible intervals.[6]

The simulations compare the proposed Bayesian approach with frequentist copula estimation (2sCOPE) and stylised OLS estimation to highlight the impact of neglected endogeneity. The simulations do not include an IV estimator, because this has been frequently done in the literature (e.g., [1, 4, 5]). Nevertheless, we will use an IV estimator in the empirical application later to assess the validity of the new Bayesian approach.

4.1 | Inference Under Model Identification

The data generating processes (DGPs) condition on one endogenous regressor (z_i) and one exogenous variable (x_i), being correlated with each other, and an intercept:

$$y_i = \alpha + x_i\beta + z_i\delta + e_i;\ i = 1, \ldots, N \quad (12)$$

with $\alpha = 2$, $\beta = 6$, $\delta = -4$, and $N = \{50, 100, 500\}$. To begin, we draw underlying components as:

$$\begin{pmatrix} \zeta_{e,i} \\ \zeta_{z,i} \\ \zeta_{x,i} \end{pmatrix} \sim \text{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.7 & 0 \\ 0.7 & 1 & 0.3 \\ 0 & 0.3 & 1 \end{bmatrix} \right) \quad (13)$$

$$x_i = \Phi^{-1}\left(\Phi\left(\zeta_{x,i}\right)\right) \quad (14)$$

$$e_i = \Phi^{-1}\left(\Phi\left(\zeta_{e,i}; \sigma^2 = 5\right)\right) \quad (15)$$

$$z_i = F^{-1}_{\text{lognormal}(0,1)}\left(\Phi\left(\zeta_{z,i}\right)\right) \quad (16)$$

where $F^{-1}_{\text{lognormal}(0,1)}$ is the inverse cumulative distribution function of the lognormal(0,1) distribution such that the endogenous regressor is non-normal, that is, $z_i \sim \text{lognormal}(0,1)$. Structural errors and the exogenous regressor are normal, that is, $e_i \sim \text{N}(0,5)$ and $x_i \sim \text{N}(0,1)$. With two continuous explanatory variables, both of which have only unique values in the observations, there are $6 + 2N$ unknown coefficients in the posterior.

We consider two specifications of the proposed Bayesian approach, which differ only with regard to the priors assigned on $\gamma \in (\alpha, \boldsymbol{\beta}, \boldsymbol{\delta})'$. Specifically, we compare the horseshoe prior outlined in Equation (5) to using flat priors, that is, $p(\gamma) \propto \text{const}$. The proposed horseshoe-type is known for its robustness- and stability-enhancing effects, especially in situations where likelihood information is sparse [33], while the consideration of flat priors enables the isolation of the combined impact arising from the joint estimation of cdfs and the full copula correlation matrix. This isolation facilitates the direct assessment of the effects attributed to avoiding a priori calculated CDFs and circumventing plug-in estimates.

Simulation results for the DGP in Equations (12–16) are given in Table 1, and show average posterior means, standard deviations of the means, and coverage frequencies based on 95% credible intervals, along with their frequentist counterparts for OLS and

TABLE 1 | Monte Carlo results for simulations of DGP in Equations (12–16) using OLS, 2sCOPE, and Bayesian approaches with different prior specifications.

			$N = 50$						$N = 100$						$N = 500$					
		Parameter	α	β	δ	σ^2	$\rho_{x,z}$	$\rho_{z,e}$	α	β	δ	σ^2	$\rho_{x,z}$	$\rho_{z,e}$	α	β	δ	σ^2	$\rho_{x,z}$	$\rho_{z,e}$
		True value	**2**	**6**	**−4**	**5**	**0.3**	**0.7**	**2**	**6**	**−4**	**5**	**0.3**	**0.7**	**2**	**6**	**−4**	**5**	**0.3**	**0.7**
OLS		Mean	1.88	5.86	−2.48	3.43			2.05	5.65	−2.49	3.25			2.01	5.68	−2.62	3.39		
		Sd	0.309	0.323	0.111	0.867			0.188	0.207	0.475	0.544			0.098	0.105	0.130	0.411		
		Coverage	0.935	0.927	0.042	0.926			0.943	0.586	0	0.189			0.955	0.085	0	0.006		
2sCOPE		Mean	1.95	5.98	−3.92	4.80	0.302	0.673	2.00	5.97	−3.94	4.81	0.296	0.677	2.00	5.99	−3.97	4.94	0.296	0.696
		Sd	0.317	0.326	0.655	1.47	0.123	0.098	0.273	0.287	0.510	1.29	0.110	0.080	0.099	0.105	0.136	0.425	0.041	0.030
		Coverage	0.966	0.955	0.964	0.957	0.948	0.971	0.944	0.942	0.939	0.913	0.944	0.944	0.945	0.955	0.942	0.930	0.946	0.943
		AMSE			Q1: 44.1; Q3: 49.7						Q1: 30.4; Q3: 35.9						Q1: 11.2; Q3: 13.0			
Bayes	Horseshoe	P. Mean	1.87	6.03	−3.99	5.26	0.313	0.680	2.00	5.98	−3.96	4.96	0.294	0.693	2.00	5.97	−3.95	4.98	0.298	0.696
		Sd	0.281	0.308	0.629	1.24	0.108	0.081	0.243	0.259	0.453	1.20	0.099	0.069	0.088	0.103	0.128	0.413	0.041	0.030
		P. Coverage	0.946	0.932	0.946	0.952	0.946	0.956	0.952	0.951	0.956	0.954	0.951	0.952	0.953	0.944	0.950	0.955	0.945	0.957
		AMSE			Q1: 39.8; Q3: 45.5						Q1: 27.0; Q3: 28.8						Q1: 11.0; Q3: 12.7			
		Gelman-Rubin	1.31	1.05	1.04	1.18	1.07	1.05	1.05	1.07	1.08	1.09	1.01	1.11	1.04	1.06	1.09	1.04	1.01	1.08
	Flat	P. Mean	1.97	6.02	−3.99	4.92	0.302	0.684	1.95	6.03	−4.01	5.06	0.304	0.685	2.01	6.05	−4.02	5.07	0.297	0.701
		Sd	0.304	0.318	0.647	1.33	0.117	0.098	0.266	0.274	0.489	1.26	0.109	0.078	0.095	0.105	0.133	0.418	0.042	0.030
		P. Coverage	0.951	0.937	0.948	0.957	0.944	0.960	0.953	0.948	0.946	0.934	0.948	0.964	0.947	0.955	0.951	0.960	0.952	0.944
		AMSE			Q1: 41.7; Q3: 49.1						Q1: 29.4; Q3: 33.0						Q1: 11.3; Q3: 13.0			
		Gelman-Rubin	1.04	1.07	1.02	1.10	1.08	1.05	1.05	1.07	1.03	1.09	1.04	1.04	1.01	1.01	1.03	1.02	1.01	1.04

Note: The table displays average posterior mean estimates (P. Mean), standard deviation of posterior means (Sd), and posterior coverage (P. Coverage) of 95% credible intervals. For frequentist estimators (OLS, 2sCOPE), mean estimates (Mean), Sd of means, and coverage frequencies based on 95% confidence intervals are reported. $\rho_{x,z}$ and $\rho_{z,e}$ denote entries in the Gaussian copula correlation matrix. Posterior quantities for probability masses of non-parametric functions defining marginal distributions of z and x are omitted for brevity.

2sCOPE. Further, unreported acceptance rates are in the optimal range, so we can assume that (iteratively updating) proposal distributions for (α, β, δ) and σ^2 are adequate choices. Unsurprisingly, OLS exhibits significant biases and incorrect confidence intervals for slope coefficients and residual variance, even when the sample size increases. The only exception is the intercept, which shows no bias and valid confidence intervals. Both specifications of the proposed Bayesian approach and 2sCOPE are largely unbiased.

In the smallest sample scenario examined ($N = 50$), the Bayesian approach employing flat priors on regression coefficients exhibits posterior means that are more tightly distributed around the true values than estimates obtained by 2sCOPE. Consequently, avoiding a priori calculated CDFs and circumventing plug-in estimates significantly reduces uncertainty in small samples. Interestingly, the Bayesian approach with the horseshoe prior produces slightly biased estimates for the intercept and the variance of the structural errors but exhibits even smaller uncertainty compared to the flat prior approach. As the sample size increases, the estimators converge more closely to the true coefficients, and the differences among them diminish. For $N = 500$, they perform nearly identically.

Taken together, the proposed approach demonstrates superiority over 2sCOPE in small samples and allows for exact finite sample inference. Through simultaneous estimation of margins, correlations, and regression coefficients, wherein priors enhance the numerical stability of the posterior, the Bayesian approach more effectively accounts for uncertainty and exhibits increased robustness, particularly in scenarios with small sample sizes. Posterior credible intervals cover all true coefficients with correct probability, regardless of the sample size. Gelman-Rubin statistics are calculated to evaluate the convergence of Markov chains, with all statistics consistently below 1.1, indicating convergence. Visual inspection of individual Markov chains (not shown) further confirms their stationarity and minimal remaining autocorrelation. Due to the superior performance in terms of estimation accuracy, we proceed with the specification that employs the horseshoe prior for the subsequent analysis.

4.2 | Inference When Identifying Assumptions Are Not Met

The literature identifies three scenarios of model non-identification: (i) normality of the endogenous regressor, (ii) violation of the Gaussian regressor-error dependence assumption, and (iii) misspecification of the structural error distribution (for an overview, see [17]). While it is well known that frequentist copula estimates are biased in these scenarios [5, 8], the impact on the performance of the Bayesian approach remains uncertain. Additionally, there is a lack of research examining the validity of frequentist inference, such as bootstrap standard errors, in these scenarios.

4.2.1 | Near Normality

It has been well documented in the literature that when the endogenous regressor approaches normality, estimates obtained from the frequentist copula approach drift towards OLS estimates while standard errors increase sharply [1, 5, 6]. Against this background, we reassess the simulations from above and use the DGP given in Equations (12–15). To introduce model non-identification, we follow Park and Gupta [1] and generate the endogenous regressor from a Student t distribution with various degrees of freedom, that is, we replace (16) by

$$z_i = F^{-1}_{t(df)}\left(\Phi\left(\zeta_{z,i}\right)\right) \tag{17}$$

where $F^{-1}_{t(df)}$ is the inverse cumulative distribution function of the t distribution with degrees of freedom given by df. Then, the endogenous regressor exhibits a $t(df)$ distribution and we distinguish scenarios with $df = \{5, 15, 30\}$. When degrees of freedom are as small as $df = 5$, the t distribution has thicker tails in comparison with the normal distribution, which is expected to provide model identification. However, both distributions become increasingly similar as df increases, such that identification breaks down if e_i and z_i are both normal.

The estimation results in Table 2 reveal the impact of endogenous regressor normality on parameter estimates. While OLS performance remains consistently bad regardless of the regressor distribution, both 2sCOPE and Bayes suffer from breakdowns in model identification. Slightly biased estimates arise when the regressor follows a $t(5)$ distribution. Confidence intervals (2sCOPE) and posterior credible intervals (Bayes) fail to cover the true values accurately. Hence, the non-normality of a $t(5)$ distribution is insufficient for model identification. Increasing the degrees of freedom to $df = 15$ and $df = 30$ leads to similar undesirable properties in Bayesian estimation and 2sCOPE. Both coefficients δ and β drift towards the OLS estimate, and the corresponding correlation estimate approaches zero, even though no identification problem is associated with x. This bias is likely due to the identification problem with δ and adverse effects transmitted through the correlation between x and z, which has not been documented in the previous literature. Furthermore, as the difference between the distribution of the endogenous regressor and the structural error diminishes, the coverage probabilities of all parameters, except α and $\rho_{x,z}$, shrink towards zero. While the mean performance of 2sCOPE and the proposed Bayesian approach is generally similar, the empirical coverage frequencies differ significantly, indicating incorrect inference in both cases.

4.2.2 | 'Non-Parametric' Dependence Structure

In empirical contexts, endogeneity arises from economic mechanisms that induce dependencies between regressors and random shocks. Thus, the actual dependence may differ from the assumed Gaussian copula. Previous research has demonstrated the Gaussian copula's ability to flexibly capture various parametric dependence structures [34, 35]. Nevertheless, Eckert and Hohberger [8] find that the Gaussian copula fails to model dependence structures that do not originate in parametric dependence generators, leading to biased estimates.

The DGP is the same as in Equation (12), and coefficients and distributions are kept as in the beginning of Section 4. However, we first draw model components independently from the

TABLE 2 | Monte Carlo results for simulations using DGP in Equations (12–15) and (17).

		$df = 5$						$df = 15$						$df = 30$					
	Parameter	α	β	δ	σ^2	$\rho_{x,z}$	$\rho_{z,e}$	α	β	δ	σ^2	$\rho_{x,z}$	$\rho_{z,e}$	α	β	δ	σ^2	$\rho_{x,z}$	$\rho_{z,e}$
	True value	**2**	**6**	**−4**	**5**	**0.3**	**0.7**	**2**	**6**	**−4**	**5**	**0.3**	**0.7**	**2**	**6**	**−4**	**5**	**0.3**	**0.7**
OLS	Mean	2.00	5.50	−1.82	2.39			2.00	5.48	−2.16	2.31			2.00	5.49	−2.22	2.31		
	Sd	0.040	0.075	0.231	0.301			0.048	0.051	0.056	0.104			0.049	0.052	0.053	0.106		
	log(MAE)	−3.60	−0.700	0.779	0.958			−3.73	−0.666	0.612	0.990			−3.68	−0.669	0.576	0.990		
	Coverage	0.945	0	0	0			0.953	0	0	0			0.945	0	0	0		
2sCOPE	Mean	2.00	5.96	−3.80	4.56	0.302	0.699	2.00	5.75	−3.11	3.61	0.300	0.396	2.00	5.62	−2.68	3.36	0.300	0.194
	Sd	0.034	0.067	0.090	0.268	0.028	0.022	0.061	0.251	0.873	1.64	0.029	0.310	0.059	0.306	1.02	1.62	0.029	0.409
	log(MAE)	−3.99	−3.22	−1.74	−1.00	−4.17	−3.73	−3.54	−1.57	−0.311	0.434	−4.16	−1.68	−3.50	−1.18	0.041	0.560	4.19	−1.15
	Coverage	1	0.949	0.706	0.844	0.956	0.856	0.971	0.724	0.724	0.839	0.948	0.729	0.980	0.731	0.725	0.940	0.941	0.725
Bayes	P. Mean	2.00	5.95	−3.77	4.52	0.302	0.662	2.00	5.58	−2.52	2.71	0.300	0.176	2.00	5.52	−2.34	2.63	0.300	0.069
	Sd	0.033	0.064	0.093	0.269	0.027	0.023	0.050	0.113	0.370	0.305	0.029	0.167	0.050	0.105	0.306	0.198	0.029	0.150
	log(MAE)	−2.97	−2.56	−1.54	−0.732	−3.49	−3.27	−2.91	−0.913	0.362	0.829	−3.47	−0.701	−2.91	−0.765	0.481	0.866	−3.47	−0.490
	P. Coverage	0.944	0.899	0.352	0.665	0.970	0.771	0.961	0.161	0.125	0.136	0.954	0.120	0.957	0.084	0.049	0.072	0.942	0.054
	Gelman–Rubin	1.02	1.22	5.84	7.72	1.03	5.98	1.03	8.67	11.4	13.2	1.01	11.9	1.04	13.7	18.2	22.5	1.04	19.0

Note: Endogenous regressor approaches normality in a 'stepwise' manner; $N = 1000$. For further notes, see Table 1.

following distributions:

$$e_i \sim \mathrm{N}(0, 5) \tag{18}$$

$$x_i \sim \mathrm{N}(0, 1) \tag{19}$$

$$z_i \sim \mathrm{lognormal}(0, 1) \tag{20}$$

Non-parametric dependence is then introduced using the algorithm by Ruscio and Kaczetow [36], such that (empirical) correlations are

$$\mathrm{Corr}[e, x] = 0, \qquad \mathrm{Corr}[e, z] = 0.3, \qquad \mathrm{Corr}[x, z] = -0.1 \tag{21}$$

measured either by Pearson or Spearman coefficients. The algorithm rotates the model component matrix (e, x, z) until empirical correlations in Equation (21) are reached while keeping marginal distributions specified in Equations (18–20) unchanged. Hence, data are simulated and correlations are introduced without assuming parametric dependence (for similar setup, see [8]).

Table 3 presents the simulation results. OLS is slightly biased for β, while parameters δ and σ^2 are more biased. Interestingly, the confidence intervals never meet the coverage probability of 95%, and this is also the case for α, although OLS is unbiased here. The performance of both 2sCOPE and the Bayesian approach varies depending on the scenario; both estimators consistently exhibit bias and reveal incorrect coverage frequencies. The limitations of 2sCOPE and the Bayesian approach can be attributed to the use of a Gaussian copula, which fails to capture non-parametric dependencies between the model regressors and structural error term accurately. For Pearson correlations, posterior means of δ and σ^2 exhibit substantial overestimation (in absolute terms), and their corresponding credible intervals fail to contain the true values in any replication. In the case of Spearman correlations, posterior means of δ and σ^2 display biased estimates in opposite directions, and the credible intervals are also invalid. However, both α and β remain unbiased, with small biases observed for $\rho_{x,z}$. It is worth noting that, unlike the simulations about non-normal errors, where the credible intervals of α and $\rho_{x,z}$ were valid despite the lack of model identification, none of these intervals now accurately cover the true coefficients.

4.2.3 | Non-Normal Structural Errors

While the assumption of normally distributed structural errors is crucial in copula-based identification and widely used, the true error distribution is unobserved and may deviate from normality. Recent frequentist studies have shown that the copula approach is robust against error kurtosis but not against skewness [5, 8]. In light of this, we revisit the first set of simulations in Section 4, where the DGP given in Equations (12–16) is used. However, we now generate structural errors from the skew generalised t distribution, replacing (15) with the new specification:

$$e_i = F^{-1}_{\mathrm{sgt}(\mu,\sigma,\lambda,p,q)}\left(\Phi\left(\zeta_{e,i}\right)\right) \tag{22}$$

where $F^{-1}_{\mathrm{sgt}(\mu,\sigma,\lambda,p,q)}$ is the inverse cumulative distribution function of the skew generalised t distribution with parameters μ, σ, λ, p, and q. As in Reference Eckert and Hohberger [8], we fix $\mu = 0$, $\sigma = 1$, and $p = 2$ and regard combinations of $\lambda = \{0, 0.8\}$ and $q = \{2, \infty\}$ such that the error distribution exhibits either moderate degrees of skewness (determined by λ) or kurtosis (determined by q). With $\lambda = 0$ and $q = 2$, the structural errors are symmetric but exhibit leptokurtosis, while combinations of $\lambda = 0.8$ and $q = \infty$ generate skewed errors with normal kurtosis. Finally, $\lambda = 0.8$ and $q = 2$ induce both skewness and leptokurtosis. In all scenarios, the structural errors are substantially different from normality.

Table 4 presents the estimation results. OLS shows severe biases regardless of the error distribution, while the performance of 2sCOPE and the Bayesian approach depends on the structural

TABLE 3 | Monte Carlo results for simulations using DGP in Equations (18–21).

		Pearson						Spearman					
	Parameter	α	β	δ	σ^2	$\rho_{x,z}$	$\rho_{z,e}$	α	β	δ	σ^2	$\rho_{x,z}$	$\rho_{z,e}$
	True value	**2**	**6**	**−4**	**5**	**−0.1**	**0.3**	**2**	**6**	**−4**	**5**	**−0.1**	**0.3**
OLS	Mean	2.01	6.07	−3.29	4.54			2.00	6.05	−3.28	4.54		
	Sd	0.104	0.003	0.107	0.290			0.107	0.012	0.129	0.299		
	log(MAE)	−2.92	−2.70	−0.352	−0.987			−2.89	−3.03	−0.344	−0.970		
	Coverage	0.792	1	0	0.451			0.799	1	0	0.448		
2sCOPE	Mean	2.00	5.90	−5.04	7.26	−0.082	0.781	2.00	6.00	−3.78	4.76	−0.131	0.279
	Sd	0.110	0.018	0.287	0.699	0.009	0.025	0.102	0.006	0.039	0.313	2.6E−4	0.010
	log(MAE)	−2.81	−2.33	−0.003	0.756	−4.12	−0.734	−2.89	−5.62	−1.54	−1.49	−3.47	−3.96
	Coverage	0.870	1	0	0.011	0.999	0	0.820	1	0.433	0.736	1	1
Bayes	P. Mean	2.00	5.98	−5.01	7.21	−0.082	0.775	2.00	6.00	−3.77	4.78	−0.131	0.274
	Sd	0.110	0.018	0.275	0.677	0.009	0.026	0.102	0.007	0.041	0.314	0.001	0.010
	log(MAE)	−2.39	−2.79	−0.033	0.737	−3.52	−0.746	−2.42	−2.89	−1.48	−1.10	−3.32	−3.16
	P. Coverage	0.763	0.737	0	0.003	0.838	0	0.804	0.604	0.601	0.660	0.972	0.529
	Gelman−Rubin	1.83	1.24	15.4	13.8	1.09	14.0	1.94	1.22	15.7	14.2	1.11	15.1

Note: Simulations with non-parametric dependence between endogenous regressor and structural error; $N = 1000$ For further notes, see Table 1.

TABLE 4 | Monte Carlo results for simulations using DGP in Equations (12–14), (16), and (22).

		$\lambda = 0,\ q = 2$						$\lambda = 0.8,\ q = \infty$						$\lambda = 0.8,\ q = 2$					
		(no skewness, moderate kurtosis)						(moderate skewness, no kurtosis)						(moderate skewness and kurtosis)					
	Parameter	α	β	δ	σ^2	$\rho_{x,z}$	$\rho_{z,e}$	α	β	δ	σ^2	$\rho_{x,z}$	$\rho_{z,e}$	α	β	δ	σ^2	$\rho_{x,z}$	$\rho_{z,e}$
	True value	**2**	**6**	**−4**	**1**	**0.3**	**0.7**	**2**	**6**	**−4**	**1**	**0.3**	**0.7**	**2**	**6**	**−4**	**1**	**0.3**	**0.7**
OLS	Mean	2.00	5.87	−3.40	0.676			2.00	5.86	−3.37	0.640			2.00	5.85	−3.34	0.603		
	Sd	0.018	0.031	0.092	0.046			0.020	0.031	0.105	0.044			0.021	0.028	0.091	0.054		
	log(MAE)	−4.66	−2.03	−0.522	−1.14			−4.56	−1.98	−0.475	−1.03			−4.53	−1.92	−0.419	−0.932		
	Coverage	0.990	0.006	0	0			0.983	0.005	0	0			0.981	0.002	0	0		
2scope	Mean	2.00	5.99	−3.94	0.933	0.299	0.661	2.00	5.97	−3.84	0.842	0.301	0.619	2.00	5.92	−3.67	0.705	0.300	0.497
	Sd	0.003	0.036	0.086	0.087	0.029	0.032	0.005	0.030	0.051	0.047	0.029	0.024	0.011	0.032	0.101	0.082	0.028	0.052
	log(MAE)	−6.73	−3.88	−2.92	−2.90	−4.19	−3.55	−5.98	−3.61	−1.91	−1.90	−4.16	−2.56	−5.24	−2.73	−1.15	−1.26	−4.25	−1.62
	Coverage	1	0.930	0.877	0	0.939	0.728	1	0.771	0.051	0	0.943	0.017	1	0.303	0.001	0	0.954	0
Bayes	P. Mean	2.00	5.98	−3.92	0.930	0.299	0.646	2.00	5.96	−3.83	0.838	0.301	0.614	2.00	5.92	−3.66	0.707	0.299	0.475
	Sd	0.005	0.039	0.108	0.086	0.029	0.056	0.006	0.030	0.051	0.047	0.029	0.024	0.012	0.034	0.105	0.079	0.027	0.068
	log(MAE)	−3.94	−3.27	−2.56	−2.50	−3.48	−3.09	−3.94	−3.15	−1.84	−1.87	−3.47	−2.50	−3.88	−2.59	−1.12	−1.27	−3.51	−1.53
	P. Coverage	1	0.810	0.505	0	0.942	0.519	1	0.686	0.017	0	0.943	0.037	1	0.215	0	0	0.956	0
	Gelman−Rubin	1.04	1.08	1.03	1.08	1.02	1.12	1.03	1.11	1.06	1.14	1.07	1.01	1.03	1.09	1.03	1.15	1.10	1.02

Note: Simulations with non-normal structural errors; $N = 1,000$. For further notes, see Table 1.

error distribution. When the errors are symmetric with moderate kurtosis, the posterior means and the mean estimates from 2sCOPE are unbiased. These findings confirm the robustness of copula-based identification in the presence of symmetric errors [5, 8]. However, when the errors are skewed, regardless of the presence of kurtosis, model identification breaks down. In these cases, the estimates of δ, σ^2, and $\rho_{z,e}$ are biased, while α, β, and $\rho_{x,z}$ remain unbiased. Nevertheless, the confidence intervals (2sCOPE) and the posterior credible intervals (Bayes) do not provide correct coverage of the true coefficients, even though the empirical coverage frequencies show less variation compared to previous simulations. This observation is particularly interesting in scenarios with symmetric errors, where the estimates are unbiased, but the intervals are invalid.

5 | Application

We illustrate the applicability of the proposed Bayesian approach to handle (cross-)price endogeneity in demand estimation using store-level scanner data of orange juices from Dominick's Finer Foods. The data include unit sales, retail price, and a dummy indicating the use of an in-store display for eight brands of orange juice (64 oz). The sample covers individual brand sales in 83 stores ($s = 1, \ldots, 83$) over a time span of 121 weeks ($t = 1, \ldots, 121$). Dominick's data are widely used in marketing research [37–42]. According to these studies, brand-level data and weekly observations offer advantages over aggregate data, such as a more detailed understanding of consumer preferences because of cross-brand effects, and capture short-term variations.

In sum, the (unbalanced) panel consists of 9649 observations. Table 5 provides summary statistics pooled across stores of average weekly prices, market shares, and unit sales of the brands. As Table 5 reveals, the brands can be classified into three price-quality tiers: Premium brands, national brands, and the store brand. Differences in quality across the tiers are well represented by higher (lower) average prices for higher (lower) quality tier brands. Average weekly prices and market shares of all brands vary considerably, reflecting frequent use of promotions.

Empirical challenges arise when considering the potential endogeneity of (cross-)prices [43]. Firm decisions on prices and marketing factors are often endogenous due to considerations of product attributes [44]. Additionally, while datasets may include information on sales promotions, there may be unmeasured brand-specific characteristics like shelf location, shelf-space allocation, or regional advertising [45]. These unmeasured variables could introduce correlation with the prices of included brands, leading to endogeneity problems [45, 46]. Research suggests that both own and cross-prices should be treated as endogenous within demand models [47, 48].

To handle endogeneity, commonly used methods involve IV estimation, with wholesale prices frequently used as instruments for retail prices [45, 49, 50]. If wholesalers consider marketplace heterogeneity, it leads to increased profitability from category-level markups [51]. According to Song and Chintagunta [45], wholesale prices are typically assumed to be uncorrelated with store-specific factors influencing retail prices, thus serving as valid instruments for controlling endogeneity) see also [49]).

We estimate demand models for the national brand with the highest market share, Minute Maid, considering both own and cross effects of price and in-store display. Unlike previous Gaussian copula studies that focus on a single endogenous regressor [1, 4, 5], our model incorporates multiple endogenous regressors. The model follows Brezger and Steiner [37] and reads as:

$$\begin{aligned}\log(sales_{st}) &= \alpha + \delta_1 \log(price_{st}) + \delta_2 \log(price\_premium_{st}) \\ &\quad + \delta_3 \log(price\_national_{st}) \\ &\quad + \delta_4 \log(price\_dominicks_{st}) \\ &\quad + \beta_1 display_{st} + \beta_2 display\_premium_{st} \\ &\quad + \beta_3 display\_national_{st} + \beta_4 display\_dominicks_{st} \\ &\quad + \sum_{t=1}^{T=120} \beta_{t+4} week_t + e_{st} \end{aligned} \tag{23}$$

TABLE 5 | Descriptive statistics for brand prices, market shares, and unit sales averaged over stores and weeks.

	Retail price			Market share			Unit sales	
Brand	**Range ($)**	**Mean ($)**	**Std Dev**	**Range (%)**	**Mean (%)**	**Std Dev (%)**	**Min**	**Max**
Premium brands								
Tropicana Pure	[1.69, 3.55]	2.95	0.53	[3, 73]	15	15	6388	101,712
Florida Natural	[1.57, 3.16]	2.86	0.33	[1, 53]	5	7	1138	56,037
National brands								
Citrus Hill	[1.09, 2.82]	2.31	0.31	[1, 78]	8	12	2006	151,570
Minute Maid	[1.29, 2.92]	2.33	0.40	[3, 87]	21	22	4805	243,711
Tropicana	[1.49, 2.75]	2.20	0.35	[2, 75]	21	23	3041	102,629
Florida Gold	[0.990, 2.83]	2.17	0.39	[1, 63]	4	8	325	150,945
Tree Fresh	[1.07, 2.48]	2.15	0.27	[1, 42]	4	6	916	39,401
Store brand								
Dominick's	[0.990, 2.47]	1.75	0.40	[1, 83]	22	22	2170	189,462

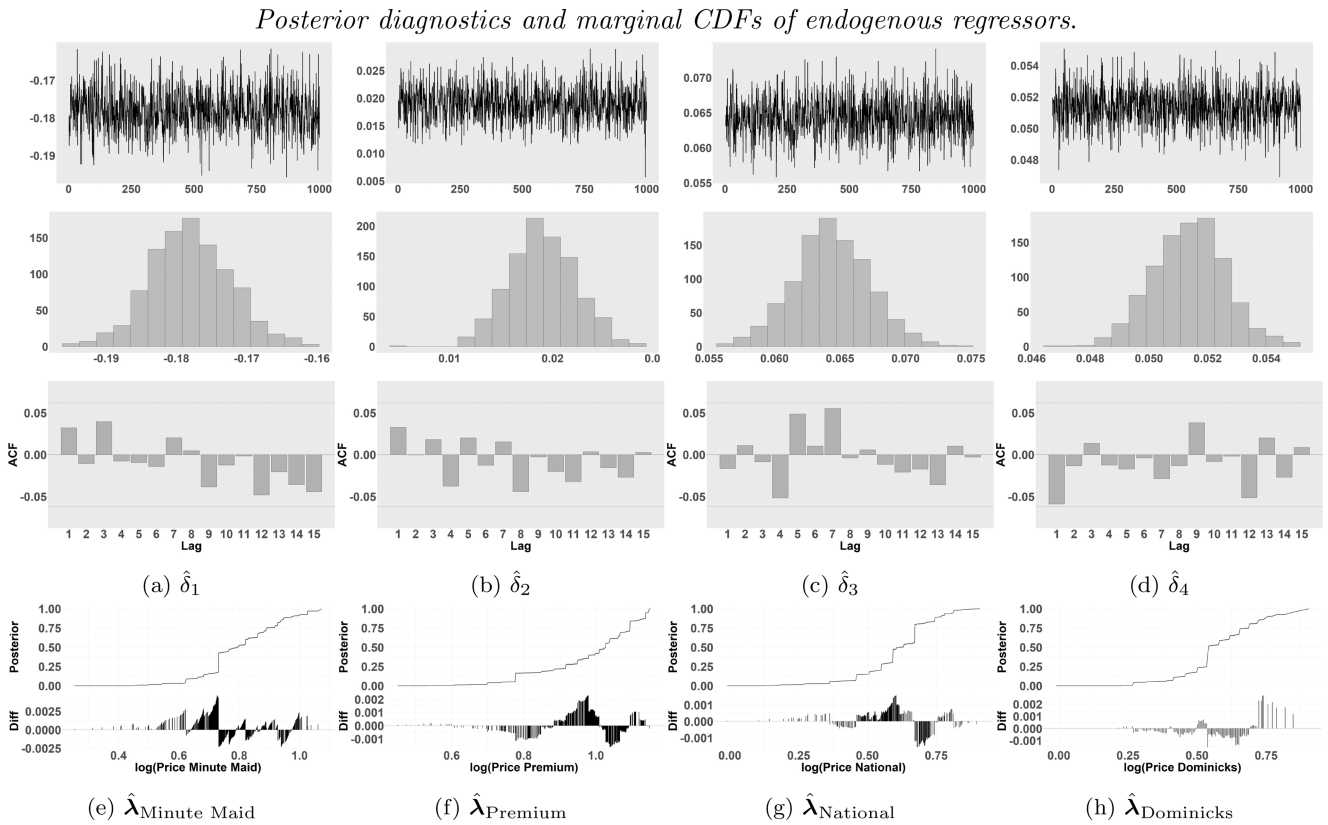

FIGURE 1 | *Posterior diagnostics and marginal CDFs of endogenous regressors.* (a) $[\hat{\delta}_1]$, (b) $[\hat{\delta}_2]$, (c) $[\hat{\delta}_3]$, (d) $[\hat{\delta}_4]$, (e) $[\hat{\lambda}_{\text{Minute Maid}}]$, (f) $[\hat{\lambda}_{\text{Premium}}]$, (g) $[\hat{\lambda}_{\text{National}}]$, and (h) $[\hat{\lambda}_{\text{Dominicks}}]$. Time series plot (first row), histograms (second row), and autocorrelation functions (third row) of coefficients attached to endogenous regressors. The fourth row shows the estimated CDFs of the price variables, based on the posterior means of the sampled λ values, together with 95% credible intervals. Each x-axis point corresponds to a unique value for which posterior draws of the corresponding λ were obtained. Directly below, the difference between the posterior mean CDF and the ecdf is displayed. For further notes, see Table 6.

where $sales_{st}$ denotes unit sales of the brand Minute Maid in store s and week t, *price* is the brand's price, and *display* is a dummy variable (= 1) if an in-store display for Minute Maid is used. Following Blattberg and George [52] and Brezger and Steiner [37], we capture cross-price effects in a parsimonious way through the use of competitive variables at the tier level rather than the individual brand level: *price_premium* and *price_national* indicate the minimum price for competing brands within the premium brand and the national brand tier, respectively, without own-price; and *price_dominicks* is the actual price of Dominick's private label brand. The dummy variables *display_premium* and *display_national* indicate the usage of in-store display for at least one brand within the respective tier, and *display_dominicks* is the display dummy for the private label. Weekly dummies account for time variation, and structural errors are assumed normally distributed, that is, $e_{st} \sim \text{N}(0, \sigma^2)$, as the use of a log-normal model is common when relating brand sales to (cross-) prices and promotional instruments [4, 5, 53].

While price endogeneity is a concern, in-store display may also be endogenous if managers adjust promotional activities based on performance indicators such as sales and awareness [54, 55]. However, endogeneity is less likely with weekly data, as promotional activities do not change rapidly in response to demand shocks [56]. Furthermore, there is a large body of literature that treats in-store display variables as exogenous (e.g., [57–59]). Therefore, we allow for endogeneity of (cross-)price variables while assuming exogeneity of display variables. We compare the proposed Bayesian approach to the frequentist 2sCOPE estimator (non-normality of the price variables is checked a priori), and use IV estimation to assess the validity of the two IV-free Gaussian copula models. Following Song and Chintagunta [45], brand-specific wholesale prices of the eight brands are used as instruments for endogenous price variables (see also [49]).

Details on the Bayesian implementation are provided in Web Appendix E. Figure 1 shows posterior samples of the coefficients for the endogenous variables. The time series appears stationary with negligible autocorrelation, and the histograms are approximately normal, indicating no convergence issues. Given that simultaneous estimation of the marginals is one of the most attractive features of the approach, the figure also reports posterior means with 95% credible intervals. Differences between the posterior mean cdfs and the ecdf are minor, mostly due to the large sample and weak priors. Importantly, the Bayesian approach provides simultaneous inference, with marginals drawn conditional on all other coefficients, unlike the independent treatment in frequentist ecdfs.

Estimation results for the empirical model are given in Table 6. Cross-price effects exhibit variation across tiers, suggesting asymmetric competitive price reactions (for similar findings,

TABLE 6 | Estimation results for model (23).

	OLS Est.	IV Est.	Proposed Bayes P.mean	GR
Endogenous regressors				
$\hat{\delta}_1$ (log price)	−0.1240	−0.1773	−0.1780	1.09
	(0.0025)	(0.0031)	(−0.1895 −0.1660)	
$\hat{\delta}_2$ (log price_premium)	0.0346	0.0138	0.0190	1.02
	(0.0024)	(0.0033)	(0.0128 0.0254)	
$\hat{\delta}_3$ (log price_national)	0.0162	0.0597	0.0643	1.07
	(0.0021)	(0.0030)	(0.0584 0.0701)	
$\hat{\delta}_4$ (log price_dominicks)	0.0270	0.0526	0.0514	1.07
	(0.0019)	(0.0027)	(0.0489 0.0539)	
Exogenous regressors				
$\hat{\alpha}$ (intercept)	0.6430	0.6607	0.4688	1.05
	(0.0122)	(0.0125)	(0.4441 0.4958)	
$\hat{\beta}_1$ (display)	0.0457	0.0450	0.0486	1.09
	(0.0011)	(0.0012)	(0.0462 0.0518)	
$\hat{\beta}_2$ (display_premium)	−0.0092	−0.0113	−0.0096	1.05
	(0.0009)	(0.0010)	(−0.0122 −0.0078)	
$\hat{\beta}_3$ (display_national)	−0.0088	−0.0088	−0.0071	1.05
	(0.0008)	(0.0008)	(−0.0111 −0.0009)	
$\hat{\beta}_4$ (display_dominicks)	−0.0026	−0.0028	−0.0025	1.04
	(0.0007)	(0.0007)	(−0.0044 −0.0007)	
weekly dummies	Yes	Yes	Yes	
Correlations				
$\hat{\rho}$(error, log price)			0.1917	1.15
			(0.1840 0.1972)	
$\hat{\rho}$(error, log price_premium)			−0.2499	1.08
			(−0.2561 −0.2387)	
$\hat{\rho}$(error, log price_national)			0.2937	1.09
			(0.2710 0.3125)	
$\hat{\rho}$(error, log price_dominicks)			0.2082	1.01
			(0.1997 0.2135)	

Note: Estimates obtained by the proposed Bayesian approach (posterior means (P.mean) and 95% credible intervals in parentheses), and by the frequentist estimators (estimated coefficients (Est.) and (bootstrap) standard errors in parentheses). Bayesian estimation is based on 1,002,000 full MCMC iterations. We discard the first 2000 iterations (burn-in) and keep each 10,000th iterate (thinning). Inference is thus based on 1000 draws from posteriors. The average acceptance rate is 0.671.

see [48, 60]). Own-price coefficients obtained through OLS are significantly higher than those obtained from IV estimation, indicating potential endogeneity issues. Neglecting endogeneity leads to underestimation of effects from other national brands and Dominick's private label, while overestimating the effects of premium brands. In terms of in-store display coefficients, there are no systematic differences among the estimators. In line with findings by Allenby and Rossi [61], promoting premium brands has the strongest negative impact on demand, while the effects diminish for lower quality tiers, indicating asymmetric cross-promotional effects.

Assuming the IV estimator provides a valid benchmark, we can use it to gauge the reliability of the proposed Bayesian approach by comparing the estimates. For all coefficients, posterior means are close to the IV estimates. However, IV estimation does not deliver quantitative assessments of the amount or direction of endogeneity, whereas copula approaches can. By estimating correlations between endogenous regressors and structural errors, valuable insights into the underlying economic mechanisms can be gleaned. A positive correlation is found between own-price and structural errors, indicating that unmeasured product characteristics influence both demand and price [62]. This holds true for the price variables of other national brands and Dominick's private label in relation to demand for Minute Maid. Interestingly, a negative correlation is observed between prices of premium brands and structural errors, possibly due to the strong distinction between premium brands and Minute

Maid [63]. The different signs may also imply that national brands are perceived as more similar to Dominick's private label compared to premium brands, supporting the presence of neighbourhood effects [64]. This observation gains additional support through the demonstrated underestimation of effects from other national brands and Dominick's private label when endogeneity is neglected, along with the concurrent overestimation of effects attributed to premium brands. The endogeneity-robust neighbourhood effects suggest a customer perception that aligns other national brands and Dominick's private label more closely with Minute Maid, while relatively distancing premium brands from Minute Maid, something which is relevant for store managers and manufacturers.

6 | Conclusion

Empirical applications of the frequentist two-stage Gaussian copula method are increasingly common, making it the most popular IV-free method in marketing and management research [6, 8]. Against this background, this study proposes a flexible one-step Bayesian approach to copula-based identification for handling endogenous regressors that enables precise finite-sample inference. In contrast to Haschka [5] and Yang et al. [4], we neither estimate cdfs of explanatory variables a priori nor use empirical correlation coefficients as plug-in estimates for the correlation matrix in the Gaussian copula. Within the Bayesian approach, we instead treat both the probability masses formalising non-parametric marginal distributions of explanatory variables and the full correlation matrix as random variables. For these, we derive the full conditional distributions by assigning Dirichlet and inverse Wishart priors, respectively, for Gibbs sampling. Sampling these jointly with the remaining unknown coefficients (intercept, slopes, structural error variance) allows for simultaneous inference. Regularisation priors on regression coefficients enhance numerical stability, while the simultaneous estimation of regressors' marginal distributions enables a more effective consideration of uncertainties compared to using plug-in estimates.

We conduct simulation experiments to assess the performance of our proposed Bayesian approach and validate Bayesian inference, with a particular focus on scenarios where crucial identifying assumptions are violated. The results show that drawing margins of explanatory variables and correlations, rather than keeping them fixed as in frequentist estimation, is more effective in accounting for uncertainties. When the model is identified, posterior means provide unbiased estimates of the true coefficients, and the corresponding credible intervals correctly cover the true values with the appropriate probability, even in small samples ($N = 50$) and with an intercept. However, if the model lacks identification, Bayesian inference becomes invalid. In such cases, we also find frequentist inference to be invalid due to incorrect bootstrap standard errors—a problem that, to our knowledge, has not yet been documented in the literature.

Acknowledgments

Data were provided by the James M. Kilts Centre, GSB, University of Chicago. Helpful comments by the participants of the ISMS 2023 Marketing Science conference are gratefully acknowledged. We also thank the Editor, Jonathan Temple, and an anonymous reviewer for valuable feedback. Open Access funding enabled and organized by Projekt DEAL.

Endnotes

[1] For a proof which shows that meeting these assumptions leads to model identification, see Web Appendix B in Reference Yang et al. [4]. Their results also hold for our Bayesian model because we make the same identifying assumptions.

[2] When using ecdf (see footnote 3 in [1]) for a generic regressor $\varpi \in (\boldsymbol{z}, \boldsymbol{x})'$, $\hat{u}_{\varpi} = (u_1/u_{m_{\varpi}}, \ldots, u_{m_{\varpi}}/u_{m_{\varpi}})'$, where u_j, $j = 1, \ldots, m_{\varpi}$ are unique values of ϖ for $i = 1, \ldots, N$. Then, $\hat{u}_{\varpi}$ is usually rescaled by $N/(N+1)$ to ensure that $\hat{\xi}_{\varpi} = \Phi^{-1}(\hat{u}_{\varpi})$ is not undefined.

[3] Note that although our approach is fully Bayesian, we stick to parametric modelling. While we acknowledge that the proposed approach may introduce more uncertainty in the margin estimation compared to a fully non-parametric Bayesian approach, we show by means of simulations that uncertainty in the margin estimation is smaller compared to the frequentist plug-in approach.

[4] For applications of Dirichlet processes to flexibly modelling the joint distribution and permitting heterogeneity of various forms in the context of instrumental-based identification, we refer the reader to Conley et al. [28] and Hu et al. [29].

[5] Codes for implementing the proposed approach are available on GitHub: https://github.com/HashtagHaschka.

[6] We conduct 1000 replications of each experiment, obtaining unique samples from the posterior distributions of model parameters. Each estimation run consists of 102,000 MCMC iterations, with a burn-in of 2000 iterations and keeping every 100th iterate (thinning). The reason for running such long chains is that we want to rule out convergence problems caused by chains that are too short. This results in 1000 posterior draws for inference. Gelman-Rubin convergence checks are performed using 100 parallel chains to check convergence. According to Gelman et al. [30], convergence can be assumed when $GR \leq 1.1$. This threshold has been adopted widely by practitioners [31]. Nevertheless, the reliability of this criterion is a subject of debate, particularly concerning its sensitivity to Type I and II errors [32]. Questions arise regarding whether this criterion would exclusively identify severe non-identification cases or also capture minor cases. There is a possibility that it might incorrectly flag models as non-identified, even when they are, in fact, identified. As a precaution, we recommend supplementing convergence checks with additional visual inspection.

References

1. S. Park and S. Gupta, "Handling Endogenous Regressors by Joint Estimation Using Copulas," *Marketing Science* 31, no. 4 (2012): 567–586.

2. D. Papies, P. Ebbes, and E. M. Feit, "Endogeneity and Causal Inference in Marketing," SSRN Working Paper (2022), https://doi.org/10.2139/ssrn.4091717.

3. M. Sklar, "Fonctions de repartition an dimensions et leurs marges," *Publications De L'institut De Statistique De L'université De Paris* 8 (1959): 229–231.

4. F. Yang, Y. Qian, and H. Xie, "Addressing Endogeneity Using a Two-Stage Copula Generated Regressor Approach," *Journal of Marketing Research* 62, no. 4 (2024): 601–623.

5. R. E. Haschka, "Handling Endogenous Regressors Using Copulas: A Generalisation to Linear Panel Models With Fixed Effects and Correlated Regressors," *Journal of Marketing Research* 59, no. 4 (2022): 860–881.

6. J.-M. Becker, D. Proksch, and C. M. Ringle, "Revisiting Gaussian Copulas to Handle Endogenous Regressors," *Journal of the Academy of Marketing Science* 50 (2021): 46–66.

7. F. Dost and R. E. Haschka, "The Gaussian Copula Control Function Method Does not Help Against Traditional Omitted Variable Bias," SSRN Working Paper (2025), https://doi.org/10.2139/ssrn.5285127.

8. C. Eckert and J. Hohberger, "Addressing Endogeneity Without Instrumental Variables: An Evaluation of the Gaussian Copula Approach for Management Research," *Journal of Management* 49, no. 4 (2023): 1460–1495.

9. F. Falkenström, S. Park, and C. N. McIntosh, "Using Copulas to Enable Causal Inference From Nonexperimental Data: Tutorial and Simulation Studies," *Psychological Methods* 28, no. 2 (2023): 301–321.

10. A. Papadopoulos, "Accounting for Endogeneity in Regression Models Using Copulas: A Step-By-Step Guide for Empirical Studies," *Journal of Econometric Methods* 11, no. 1 (2022): 127–154.

11. Y. Qian, H. Xie, and A. Koschmann, "Should Copula Endogeneity Correction Include Generated Regressors for Higher-Order Terms? No, It Hurts," NBER Working Paper 29978 (2022), https://doi.org/10.3386/w29978.

12. R. E. Haschka, "Robustness of Copula-Correction Models in Causal Analysis: Exploiting Between-Regressor Correlation," *IMA Journal of Management Mathematics* 33, no. 1 (2014): 161–180.

13. J. Breitung, A. Mayer, and D. Wied, "Asymptotic Properties of Endogeneity Corrections Using Nonlinear Transformations," *Econometrics Journal* 27, no. 3 (2024): 362–383.

14. K. C. Tran and M. G. Tsionas, "Efficient Semiparametric Copula Estimation of Regression Models With Endogeneity," *Econometric Reviews* 41, no. 5 (2021): 485–504.

15. Y. Qian and H. Xie, "Correcting Regressor-Endogeneity Bias via Instrument-Free Joint Estimation Using Semiparametric Odds Ratio Models," *Journal of Marketing Research* 61, no. 5 (2024): 914–936.

16. R. E. Haschka, "Endogeneity-Robust Estimation of Nonlinear Price-Response Curves in Demand Models: A Bayesian Approach Through Joint Estimation Using Copulas," (2023), https://doi.org/10.2139/ssrn.4451591.

17. R. B. Nelsen, *An Introduction to Copulas* (Springer, 2006).

18. A. Prokhorov and P. Schmidt, "Likelihood-Based Estimation in a Panel Setting: Robustness, Redundancy and Validity of Copulas," *Journal of Econometrics* 153, no. 1 (2009): 93–104.

19. R. E. Haschka and H. Herwartz, "Endogeneity in Pharmaceutical Knowledge Generation: An Instrument-Free Copula Approach for Poisson Frontier Models," *Journal of Economics and Management Strategy* 31, no. 4 (2022): 942–960.

20. Y. Qian and H. Xie, "No Customer Left Behind: A Distribution-Free Bayesian Approach to Accounting for Missing Xs in Marketing Models," *Marketing Science* 30, no. 4 (2011): 717–736.

21. S. Kim, C. Lee, and S. Gupta, "Bayesian Synthetic Control Methods," *Journal of Marketing Research* 57, no. 5 (2020): 831–852.

22. C. M. Carvalho, N. G. Polson, and J. G. Scott, "The Horseshoe Estimator for Sparse Signals," *Biometrika* 97, no. 2 (2010): 465–480.

23. L. Fahrmeir and S. Lang, "Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data," *Annals of the Institute of Statistical Mathematics* 53, no. 1 (2001): 11–30.

24. L. Fahrmeir, T. Kneib, S. Lang, and B. D. Marx, *Regression: Models, Methods and Applications* (Springer, 2021).

25. M. Wiesenfarth, C. M. Hisgen, T. Kneib, and C. Cadarso-Suarez, "Bayesian Nonparametric Instrumental Variables Regression Based on Penalized Splines and Dirichlet Process Mixtures," *Journal of Business & Economic Statistics* 32, no. 3 (2014): 468–482.

26. X. Zhang, W. J. Boscardin, and T. R. Belin, "Sampling Correlation Matrices in Bayesian Models With Correlated Latent Variables," *Journal of Computational and Graphical Statistics* 15, no. 4 (2006): 880–896.

27. M. S. Smith, "Bayesian Approaches to Copula Modelling," in *Bayesian Theory and Applications Bayesian Theory and Applications*, eds. P. Damien, P. Dellaportas, N. G. Polson, and D. A. Stephens (Oxford University Press, 2013), 336–358.

28. T. G. Conley, C. B. Hansen, R. E. McCulloch, and P. E. Rossi, "A Semi-Parametric Bayesian Approach to the Instrumental Variable Problem," *Journal of Econometrics* 144, no. 1 (2008): 276–305.

29. X. Hu, M. K. Munkin, and P. K. Trivedi, "Estimating Incentive and Selection Effects in the Medigap Insurance Market: An Application With Dirichlet Process Mixture Model," *Journal of Applied Econometrics* 30, no. 7 (2015): 1115–1143.

30. A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis* (Chapman and Hall/CRC, 2004).

31. D. Vats and C. Knudson, "Revisiting the Gelman–Rubin Diagnostic," *Statistical Science* 36, no. 4 (2021): 518–529.

32. H. Du, Z. Ke, G. Jiang, and S. Huang, "The Performances of Gelman-Rubin and Geweke's Convergence Diagnostics of Monte Carlo Markov Chains in Bayesian Analysis," *Journal of Behavioral Data Science* 2, no. 2 (2022): 47–72.

33. N. G. Polson and V. Sokolov, "Bayesian Regularization: From Tikhonov to Horseshoe," *Wiley Interdisciplinary Reviews: Computational Statistics* 11 (2019): e1463.

34. P. J. Danaher and M. S. Smith, "Modeling Multivariate Distributions Using Copulas: Applications in Marketing," *Marketing Science* 30, no. 1 (2011): 4–21.

35. X.-K. Song and P. X.-K. Song, "Multivariate Dispersion Models Generated From Gaussian Copula," *Scandinavian Journal of Statistics* 27, no. 2 (2000): 305–320.

36. J. Ruscio and W. Kaczetow, "Simulating Multivariate Nonnormal Data Using an Iterative Algorithm," *Multivariate Behavioral Research* 43, no. 3 (2008): 355–381.

37. A. Brezger and W. J. Steiner, "Monotonic Regression Based on Bayesian P–Splines: An Application to Estimating Price Response Functions From Store-Level Scanner Data," *Journal of Business & Economic Statistics* 26, no. 1 (2008): 90–104.

38. Y. Chen and S. Yang, "Estimating Disaggregate Models Using Aggregate Data Through Augmentation of Individual Choice," *Journal of Marketing Research* 44, no. 4 (2007): 613–621.

39. A. L. Montgomery, "Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data," *Marketing Science* 16, no. 4 (1997): 315–337.

40. R. J. Slotegraaf and K. Pauwels, "The Impact of Brand Equity and Innovation on the Long-Term Effectiveness of Promotions," *Journal of Marketing Research* 45, no. 3 (2008): 293–306.

41. S. Sriram, S. Balachander, and M. U. Kalwani, "Monitoring the Dynamics of Brand Equity Using Store-Level Data," *Journal of Marketing* 71, no. 2 (2007): 61–78.

42. M. Wedel and J. Zhang, "Analyzing Brand Competition Across Subcategories," *Journal of Marketing Research* 41, no. 4 (2004): 448–456.

43. H. Hruschka, "Functional Flexibility, Latent Heterogeneity and Endogeneity in Aggregate Market Response Models," *Marketing ZFP* 39, no. 3 (2017): 17–31.

44. J. M. Villas-Boas and R. S. Winer, "Endogeneity in Brand Choice Models," *Management Science* 45, no. 10 (1999): 1324–1338.

45. I. Song and P. K. Chintagunta, "Measuring Cross-Category Price Effects With Aggregate Store Data," *Management Science* 52, no. 10 (2006): 1594–1609.

46. Y. Li and A. Ansari, "A Bayesian Semiparametric Approach for Endogeneity and Heterogeneity in Choice Models," *Management Science* 60, no. 5 (2014): 1161–1179.

47. D. Besanko, S. Gupta, and D. Jain, "Logit Demand Estimation Under Competitive Pricing Behavior: An Equilibrium Framework," *Management Science* 44, no. 11 (1998): 1533–1547.

48. M. Wedel, J. Zhang, and F. Feinberg, "Implementing Retail Category Management: A Model-Based Approach to Setting Optimal Markups," *Customer Needs and Solutions* 2, no. 2 (2015): 165–176.

49. P. K. Chintagunta, J.-P. Dubé, and V. Singh, "Balancing Profitability and Customer Welfare in a Supermarket Chain," *Quantitative Marketing and Economics* 1, no. 1 (2003): 111–147.

50. N. M. Fong, D. I. Simester, and E. T. Anderson, "Private Label vs. National Brand Price Sensitivity: Evaluating Non-Experimental Identification Strategies," MIT Working Paper (2011).

51. S. Yang, Y. Chen, and G. M. Allenby, "Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics* 1, no. 3 (2003): 251–275.

52. R. C. Blattberg and E. I. George, "Shrinkage Estimation of Price and Promotional Elasticities: Seemingly Unrelated Equations," *Journal of the American Statistical Association* 86, no. 414 (1991): 304–315.

53. A. Brezger and S. Lang, "Generalized Structured Additive Regression Based on Bayesian P-Splines," *Computational Statistics & Data Analysis* 50, no. 4 (2006): 967–991.

54. P. E. Rossi, "Even the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications," *Marketing Science* 33, no. 5 (2014): 655–672.

55. R. Sethuraman, G. J. Tellis, and R. A. Briesch, "How Well Does Advertising Work? Generalizations From Meta-Analysis of Brand Advertising Elasticities," *Journal of Marketing Research* 48, no. 3 (2011): 457–471.

56. W. I. Keller, B. Deleersnyder, and K. Gedenk, "Price Promotions and Popular Events," *Journal of Marketing* 83, no. 1 (2019): 73–88.

57. V. R. Nijs, S. Srinivasan, and K. Pauwels, "Retail-Price Drivers and Retailer Profits," *Marketing Science* 26, no. 4 (2007): 473–487.

58. K. Pauwels, "How Dynamic Consumer Response, Competitor Response, Company Support, and Company Inertia Shape Long-Term Marketing Effectiveness," *Marketing Science* 23, no. 4 (2004): 596–610.

59. S. Srinivasan, K. Pauwels, and V. Nijs, "Demand-Based Pricing Versus Past-Price Dependence: A Cost–Benefit Analysis," *Journal of Marketing* 72, no. 2 (2008): 15–27.

60. F. J. Mulhern and R. P. Leone, "Implicit Price Bundling of Retail Products: A Multiproduct Approach to Maximizing Store Profitability," *Journal of Marketing* 55, no. 4 (1991): 63–76.

61. G. M. Allenby and P. E. Rossi, "Quality Perceptions and Asymmetric Switching Between Brands," *Marketing Science* 10, no. 3 (1991): 185–204.

62. T. H. Bijmolt, H. J. Van Heerde, and R. G. Pieters, "New Empirical Generalizations on the Determinants of Price Elasticity," *Journal of Marketing Research* 42, no. 2 (2005): 141–156.

63. D. Ackerberg and M. Rysman, "Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects," NBER Working Paper 8798 (2002), https://doi.org/10.3386/w8798.

64. R. Sethuraman, V. Srinivasan, and D. Kim, "Asymmetric and Neighborhood Cross-Price Effects: Some Empirical Generalizations," *Marketing Science* 18, no. 1 (1999): 23–41.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Web Appendix**: Supporting Information.