

# Workshop

## Addressing endogeneity in observational data models with Copula-based methods

Florian Dost and Rouven E. Haschka

AoM Meeting

2025 – 27 – 07

# Defining Endogeneity

$$y_i = \beta x_i + e_i$$

Gauß-Markov:  $\mathbb{E}[e_i|x_i] = 0 \rightarrow \mathbb{E}[\hat{\beta}_{\text{OLS}}] = \beta$

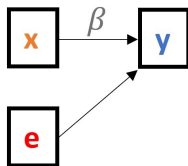
Violation  $\mathbb{E}[e_i|x_i] \neq 0$ :

- ▶ Omitted variables
- ▶ Measurement error (in the independent variable)
- ▶ Reverse causality

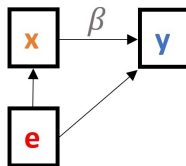
→ Endogeneity:  $\mathbb{E}[\hat{\beta}_{\text{OLS}}] \neq \beta$  and  $\text{plim}_{n \rightarrow \infty} \hat{\beta}_{\text{OLS}} \neq \beta$

# Defining Endogeneity

$$y_i = \beta x_i + e_i$$



(a) Typical regression



(b) Endogeneity

# Understanding Endogeneity

$$y_i = \beta x_i + e_i$$

Endogeneity:

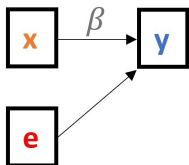
- (i)  $x_i \uparrow \rightarrow y_i \uparrow$
  - (ii)  $x_i \uparrow \rightarrow e_i \uparrow$
  - (iii)  $e_i \uparrow \rightarrow y_i \uparrow$
- $\rightarrow y_i \uparrow \uparrow$

If  $\mathbb{E}[e|x]$  is not  $= 0$ , but may be  $\mathbb{E}[e|x] = \gamma x$

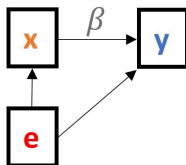
$$\begin{aligned}\mathbb{E}[y|x] &= \beta x + \mathbb{E}[e|x] \\ &= \beta x + \gamma x = (\beta + \gamma)x\end{aligned}$$

# What are Instrumental Variables?

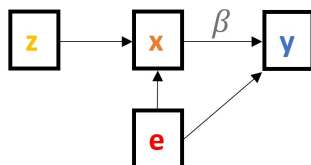
$$y_i = \beta x_i + e_i$$



(c) Typical regression



(d) Endogeneity

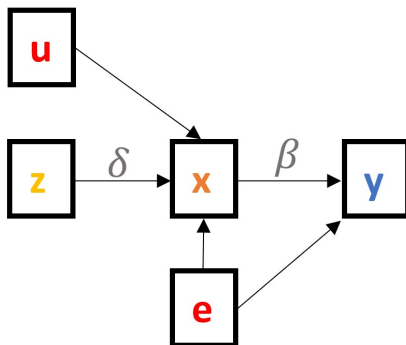


(e) IV estimation

# Two stage least squares

1st stage:  $x_i = \delta z_i + u_i$

2nd stage:  $y_i = \beta x_i + e_i$



# Implementation

$$y_i = \beta x_i + e_i$$

Implementation:

1. Regress  $x_i$  on  $z_i$ .
2. Obtain the fitted values  $\hat{x}_i = \hat{\alpha} + \hat{\delta}z_i$ .
3. Replace  $x_i$  by  $\hat{x}_i$ .
4. Estimate  $y_i = \beta\hat{x}_i + e_i$ .

Why it works:

- ▶ Within the first stage, we tease out the part of  $x_i$  that is not correlated with  $e_i$ ; this is  $\hat{x}_i$ .
- ▶ Within the second stage, since  $\hat{x}_i$  is uncorrelated with  $e_i$ , it's marginal effect gives the causal effect on  $y_i$ .

# Violation of assumptions

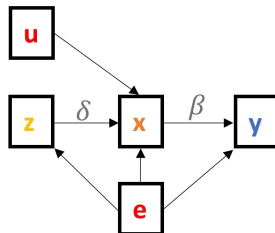
$$y_i = \beta x_i + e_i$$

$$x_i = \delta z_i + u_i$$

Crucial requirements:

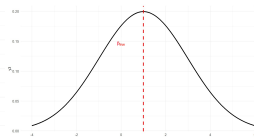
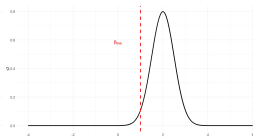
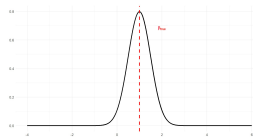
- ▶ The instrument is relevant:  $z_i$  is correlated with  $x_i$ , i.e.,  $\text{Corr}[z, x] > 0$
- ▶ The instrument is exogenous:  $z_i$  is not correlated with  $e_i$ , i.e.,  $\text{Cov}[z, e] = 0$

Violation of the exclusion restriction:





# Violation of assumptions



(f) Exogenous & strong (g) Endogenous & strong (h) Exogenous & weak

# The Gaussian copula endogeneity correction

It is very hard to convince reviewers that the employed instruments are valid!

Consider the regression model:

$$Y_t = \mu + X_t\beta + P_t\alpha + \xi_t, \quad t = 1, \dots, T$$

Least-squares estimation is equivalent to Maximum Likelihood estimation based on  $f_\xi(\xi)$ , assuming  $\xi \sim N(0, \sigma^2)$

Now the idea is to first derive the joint distribution of  $\xi$  and  $P$ ,  $f(\xi, P)$ , and then obtaining estimates based on it  
→ The joint distribution should account for the dependence between  $\xi$  and  $P$  and solve the endogeneity problem!

# The Gaussian copula endogeneity correction

$$Y_t = \mu + X_t\beta + P_t\alpha + \xi_t, \quad t = 1, \dots, T$$

- ▶ Goal: Approximate the joint distribution of  $\xi_t$  and  $P_t$  using a copula function:  $f(\xi, P) = c(F_\xi(\xi), F_P(P)) \times f_\xi(\xi) \times f_P(P)$
- ▶ Transform variables to Gaussian margins
  - ▶  $P_t^* = \Phi^{-1}[\hat{F}_P(P_t)]$
  - ▶  $\xi_t^* = \Phi^{-1}[\Phi(\xi_t, \sigma^2)]$ , assuming  $\xi_t \sim N(0, \sigma^2)$
- ▶ Resulting marginal distributions
  - ▶  $P_t^* \sim \mathcal{N}(0, 1)$
  - ▶  $\xi_t^* \sim \mathcal{N}(0, 1)$
- ▶  $(P_t^*, \xi_t^*)'$  follows a bivariate normal distribution with correlation  $\rho$

In general, we could use any copula function and distributional assumption for  $\xi$ !

# The approach by Park and Gupta (2012)

$$Y_t = \mu + X_t\beta + P_t\alpha + \xi_t, \quad t = 1, \dots, T$$

$$\begin{pmatrix} P_t^* \\ \xi_t^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right),$$

$$\begin{pmatrix} P_t^* \\ \xi_t^* \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} \varpi_{1,t} \\ \varpi_{2,t} \end{pmatrix},$$

with  $(\varpi_{1,t}, \varpi_{2,t})' \sim N(\mathbf{0}_2, \mathbf{I}_2)$ . Then,

$\xi_t = \sigma \xi_t^* = \sigma \rho P_t^* + \sigma \sqrt{1-\rho^2} \varpi_{2,t}$  and the model can be rewritten as:

$$Y_t = \mu + X_t\beta + P_t\alpha + \sigma \rho P_t^* + \sigma \sqrt{1-\rho^2} \varpi_{2,t}.$$

# The approach by Park and Gupta (2012)

$$\begin{pmatrix} P_t^* \\ \xi_t^* \end{pmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$

$$\begin{pmatrix} P_t^* \\ \xi_t^* \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} \varpi_{1,t} \\ \varpi_{2,t} \end{pmatrix},$$

with  $(\varpi_{1,t}, \varpi_{2,t})' \sim N(\mathbf{0}_2, \mathbf{I}_2)$ .  $\xi_t = \sigma \xi_t^* = \sigma \rho P_t^* + \sigma \sqrt{1-\rho^2} \varpi_{2,t}$   
and the model can be rewritten as:

$$Y_t = \mu + X_t \beta + P_t \alpha + \sigma \rho P_t^* + \sigma \sqrt{1-\rho^2} \varpi_{2,t}.$$

Model is identified because:

- ▶  $\varpi_{1,t}, \varpi_{2,t}$  are independent
- ▶  $P_t^*$  is a linear functions of  $\varpi_{1,i}$ , normal, and independent of  $\varpi_{2,i}$
- ▶  $P_t$  as a (nonlinear) function of  $P_t^*$  is uncorrelated with  $\varpi_{2,i}$   
 $\rightarrow \varpi_{2,t}$  is not correlated with  $P_t$ , and  $P_t^*$

# The approach by Park and Gupta (2012)

Starting with the regression model

$$Y_t = \mu + X_t\beta + P_t\alpha + \xi_t, \quad t = 1, \dots, T$$

- ▶ Generate  $\hat{P}_t = \Phi^{-1}(\hat{F}_P(P))$
- ▶ Include  $\hat{P}_t$  as additional regressor to model: Regress  $Y_t$  on  $X_t$ ,  $P_t$ , and  $\hat{P}_t$
- ▶ The additional regressor  $\hat{P}_t$  absorbs the endogeneity bias

Identifying assumptions:

- ▶  $P$  is continuous
- ▶ Normality of error:  $\xi \sim N(0, \sigma^2)$
- ▶ Gaussian copula dependence
- ▶ Nonlinear regressor error relation (as  $\xi$  is normal,  $P$  must be non-normal)  $\rightarrow P$  is not normally distributed
- ▶  $P$  is independent of  $X$

# The approach by Park and Gupta (2012) - DGP

$$\begin{pmatrix} P_t^* \\ \xi_t^* \end{pmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$

$$\xi_t = \xi_t^* \sigma,$$

$$P_t = F^{-1}[\Phi(P_t^*)],$$

$$Y_t = \mu + X_t \beta + P_t \alpha + \xi_t,$$

- ▶ In general,  $P_t^*$  is (standard) normal and  $F^{-1}[\Phi(\cdot)]$  is a nonlinear transformation function such that  $P_t$  will be non-normal
- ▶ In practice,  $F^{-1}$  is the inverse cdf (quantile function) of a continuous non-normal distribution, such that  $P_t$  will be follow this distribution
- ▶ Because it is assumed that  $P$  and  $X$  are independent, it does not matter how it is generated

# The approach by Haschka (2024)

$$Y_t = \mu + X_t\beta + P_t\alpha + \xi_t, \quad t = 1, \dots, T$$

- ▶ Allow  $X_t$  and  $P_t$  to be correlated
- ▶ Approximate the joint distribution of  $\xi_t$ ,  $P_t$ , and  $X_t$  using Gaussian copula function
- ▶  $(P^* \ X^* \ \xi^*)'$  follow a three-dimensional standard normal distribution

Steps:

- ▶ Generate  $\hat{P}^* = \Phi^{-1}(\hat{F}_P(P))$  and  $\hat{X}^* = \Phi^{-1}(\hat{F}_X(X))$
- ▶ Regress  $\hat{P}^*$  on  $\hat{X}^*$  (without constant) and obtain the residuals  $\widehat{res}_t = \hat{P}_t^* - \hat{r}\hat{X}_t^*$
- ▶ Include  $\widehat{res}_t$  as additional regressor to model: Regress  $Y_t$  on  $P_t$ ,  $X_t$ , and  $\widehat{res}_t$
- ▶ The additional regressor  $\widehat{res}_t$  absorbs the endogeneity bias



# The approach by Haschka (2024) - DGP

$$\begin{pmatrix} P_t^* \\ X_t^* \\ \xi_t^* \end{pmatrix} \sim \mathcal{N} \left[ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & r & \rho \\ r & 1 & 0 \\ \rho & 0 & 1 \end{pmatrix} \right],$$

$$\xi_t = \xi_t^* \sigma,$$

$$P_t = F_{\cdot}^{-1} [\Phi(P_t^*)],$$

$$X_t = G_{\cdot}^{-1} [\Phi(X_t^*)],$$

$$Y_t = \mu + X_t \beta + P_t \alpha + \xi_t,$$

- ▶ In practice,  $F_{\cdot}^{-1}$  is the inverse cdf (quantile function) of a continuous non-normal distribution, such that  $P_t$  will be follow this distribution
- ▶  $G$  can also be the inverse cdf of the normal distribution
- ▶ Linear relationship between  $P^*$  and  $X^*$

# The approach by Yang et al. (2025) - 2sCOPE

$$Y_t = \mu + X_t\beta + P_t\alpha + \xi_t, \quad t = 1, \dots, T$$

Steps:

- ▶ Generate  $\hat{P}^* = \Phi^{-1}(\hat{F}_P(P))$  and  $\hat{X}^* = \Phi^{-1}(\hat{F}_X(X))$
- ▶ Regress  $\hat{P}^*$  on  $\hat{X}^*$  (with constant) and obtain the residuals  $\widehat{res}_t = \hat{P}_t^* - \hat{\tau} - \hat{r}\hat{X}_t^*$
- ▶ Include  $\widehat{res}_t$  as additional regressor to model: Regress  $Y_t$  on  $P_t$ ,  $X_t$ , and  $\widehat{res}_t$
- ▶ The additional regressor  $\widehat{res}_t$  absorbs the endogeneity bias

# The approach by Yang et al. (2025) - DGP

$$\begin{pmatrix} P_t^* \\ X_t^* \\ \xi_t^* \end{pmatrix} \sim \mathcal{N} \left[ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & r & \rho \\ r & 1 & 0 \\ \rho & 0 & 1 \end{pmatrix} \right],$$

$$\xi_t = \xi_t^* \sigma,$$

$$P_t = F_{\cdot}^{-1} [\Phi(P_t^*)],$$

$$X_t = G_{\cdot}^{-1} [\Phi(X_t^*)],$$

$$Y_t = \mu + X_t \beta + P_t \alpha + \xi_t,$$

- ▶ Either  $F_{\cdot}^{-1}[\Phi(\cdot)]$  or  $G_{\cdot}^{-1}[\Phi(\cdot)]$  should be nonlinear function
- ▶ Put differently: Either  $F^{-1}$  or  $G^{-1}$  must be the inverse cdf of a non-normal distribution (the other can be normal)
- ▶ Linear relationship between  $P^*$  and  $X^*$

# The approach by Liengard et al. (2025)

$$Y_t = \mu + X_t\beta + P_t\alpha + \xi_t, \quad t = 1, \dots, T$$

- ▶ Exogenous regressor  $X_t$  is binary (or categorical)
- ▶ The correlation between  $P_t$  and  $\xi_t$  is moderated by  $X_t$
- ▶ The amount of endogeneity depends on the category of  $X_t$

Steps:

- ▶ Generate a copula correction term only using observations for  $X = 0$ ,  $P_t^{*[X_t=0]}$ , and a second term only using observations for  $X = 1$ ,  $P_t^{*[X_t=1]}$
- ▶ These two correction terms are then interacted with the binary variable and included to the regression model

$$Y_t = \mu + \beta X_t + \alpha P_t + \gamma_1 P_t^{*[X_t=1]} 1_{[X_t=1]} + \gamma_2 P_t^{*[X_t=0]} 1_{[X_t=0]} + error_t$$

# The approach by Liengard et al. (2025) - DGP

$$\begin{pmatrix} P_t^* \\ \xi_t^* \end{pmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{bmatrix} \right) \text{ if } X_t = 0,$$

$$\begin{pmatrix} P_t^* \\ \xi_t^* \end{pmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \right) \text{ if } X_t = 1,$$

$$\xi_t = \xi_t^* \sigma,$$

$$P_t = F^{-1} [\Phi (P_t^*)],$$

$$X_t = \text{Binom}(\pi),$$

$$Y_t = \mu + X_t \beta + P_t \alpha + \xi_t,$$

- ▶  $X_t$  is categorical (not necessarily binary); endogeneity regimes
- ▶ If there is an additional continuous exogenous regressor, the approaches by Haschka (2024) or Yang et al. (2025) are employed taking only observations for each category of  $X_t$

# The approach by Hu et al. (2025)

$$Y_t = \mu + X_t\beta + P_t\alpha + \xi_t, \quad t = 1, \dots, T$$

- ▶ All first-stage regressions inherently assume a specific dependency structure between  $P_t$  on  $X_t$
- ▶ Estimate the conditional distribution of  $P_t$  given  $X_t$
- ▶ Avoids first-stage regression
- ▶ Generalisation of the approach by Lienggaard et al. (2025)
- ▶ Applicable to binary endogenous regressors

Steps:

- ▶ Estimate  $\hat{F}_{P|X}$ , the conditional cdf of  $P_t$  given  $X_t$

$$Y_t = \mu + X_t\beta + P_t\alpha + \gamma\Phi^{-1}[\hat{F}_{P|X}(P_t)] + error_t,$$

# The approach by Hu et al. (2025) - DGP

$$\begin{pmatrix} P_t^* \\ \xi_t^* \end{pmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$

$$\xi_t = \xi_t^* \sigma,$$

$$P_t = F_{P|X}^{-1} [\Phi(P_t^*)],$$

$$Y_t = \mu + X_t \beta + P_t \alpha + \xi_t,$$

- As  $P|X$  is used, the dependence between  $P$  and  $X$  does not need to be specified a priori, i.e.,  $P_t = f(X_t) + \dots$

Different perspective:

$$\begin{pmatrix} P_t^* \\ \xi_t^* \end{pmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho = f(X) \\ \rho = f(X) & 1 \end{bmatrix} \right),$$

$$P_t = F^{-1} [\Phi(P_t^*)]$$

# The approach by Breitung et al. (2024)

$$Y_t = \mu + X_t\beta + P_t\alpha + \xi_t, \quad t = 1, \dots, T$$

- ▶ First-stage regression other than in Yang et al. (2025)
- ▶ Assume a linear relation between  $P_t$  and  $X_t$  instead of between  $\hat{P}^*$  and  $\hat{X}^*$

Steps:

- ▶ Regress  $P_t$  on  $X_t$  (with constant) and obtain the residuals  $\widehat{res}_t = P_t - \hat{\tau} - \hat{r}X_t$
- ▶ Generate  $\widehat{res}^* = \Phi^{-1}(\hat{F}_{res}(res))$
- ▶ Include  $\widehat{res}_t^*$  as additional regressor to model: Regress  $Y_t$  on  $P_t$ ,  $X_t$ , and  $\widehat{res}_t^*$
- ▶ The additional regressor  $\widehat{res}_t^*$  absorbs the endogeneity bias



# The approach by Breitung et al. (2024) - DGP

$$\begin{aligned}Y_t &= \mu + \beta X_t + \alpha P_t + \xi_t, \\P_t &= \delta X_t + e_t, \quad (X_t, e_t) \not\sim N, \\ \xi_t &= \rho \eta_t + \epsilon_t, \quad \epsilon_t \sim N, \\ \eta_t &= \Phi^{-1}[F_e(e)]\end{aligned}$$

- ▶ Linear dependence between  $P$  and  $X$  is assumed (not between  $P^*$  and  $X^*$ !)
- ▶ Although the error can be non-normal, endogeneity must fully run through its normal part

# The approach by Haschka (2022)

$$Y_{it} = \mu_i + X_{it}\alpha + P_{it}\beta + \xi_{it},$$

- ▶ First employ fixed-effects transformation to eliminate  $\mu_i$
- ▶ Apply the common GLS-transformation for further pre-whitening
- ▶ Set up the joint distribution of the transformed  $X_{it}$ ,  $P_{it}$ , and  $P_{it}$  using Gaussian copula
- ▶ Derive the likelihood function

# Key take-aways

- ▶ A nonlinear bijective transformation function is always required
  - ▶ Either a nonlinear regressor-error relation (between  $P$  and  $\xi$ )
    - Then,  $P$  must not be additively decomposable
  - ▶ Or a nonlinear endogenous-exogenous regressor relation (between  $P$  and  $X$ )
    - Then,  $X$  must not be additively decomposable
- In any case, the endogeneity must fully run through the normal part in  $\xi$
- ▶ A Gaussian copula generates the latent data  $P^*$ ,  $X^*$ , which are then (nonlinearly) transformed into the observables  $P$ ,  $X$ , and then used to generate  $Y$

# Methods to obtain the cdf's

## ► Kernel-based estimation

- Park and Gupta (2012) use kernel density estimation with Epanechnikov kernel and Silverman's rule of thumb
- Haschka (2022) uses a different bandwidth rule

→ The cdf is then computed via numerical integration of the estimated pdf

## ► Empirical CDF (ecdf)

- Rescaling required due to  $\hat{F}(\max(x)) = a \rightarrow 1 \Rightarrow \Phi^{-1}(a) \rightarrow \infty$
- Qian and Xie (2024): Multiply ecdf by  $T/(T+1)$ , also used by Haschka (2024) and Yang et al. (2025)
- Becker et al. (2022) and Eckert and Hohberger (2022): Replace  $\hat{F}(x) = 0$  with  $10^{-7}$  and  $\hat{F}(x) = 1$  with  $1 - 10^{-7}$
- Liengaard et al. (2025) propose:  

$$\hat{F}(x) = \frac{1}{2T} + \frac{T-1}{T^2} \sum_{i=1}^T I(X_i \leq x)$$

# Park and Gupta (2024) blog post

<https://www.ama.org/marketing-news/a-review-of-copula-correction-methods-to-address-regressor-error-correlation>

## A Review of Copula Correction Methods to Address Regressor–Error Correlation

5.15.2024 • Sungho Park and Sachin Gupta

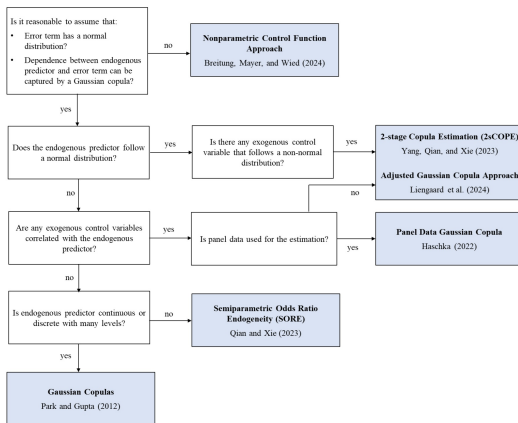
The omnipresent error term in regression models does not always receive careful attention by model builders. What factors are included in this error? Naturally, it would be ideal if the error were entirely due to random shocks. However, sometimes factors that should be explicitly incorporated in the model but cannot be observed or are unavailable to be used as explanatory variables are also present in the error. Worse, often our accumulated knowledge and theories indicate that the variables seeping into the error term are systematically related to the explanatory variables included in the model. This results in regressor–error correlation, which, if ignored, leads to biased estimates.

# Recommendations by Herhausen et al. (2025)

https:

//papers.ssrn.com/sol3/papers.cfm?abstract\_id=5246466

**FIGURE 5: Decision Tree for Endogeneity Correction through Gaussian Copulas**



# Bibliography

- Becker, J.-M., D. Proksch, and C. M. Ringle (2022). Revisiting Gaussian copulas to handle endogenous regressors. *Journal of the Academy of Marketing Science* 50(1), 46–66.
- Breitung, J., A. Mayer, and D. Wied (2024). Asymptotic properties of endogeneity corrections using nonlinear transformations. *The Econometrics Journal*, utae002. DOI: 10.1093/ectj/utae002.
- Eckert, C. and J. Hohberger (2022). Addressing endogeneity without instrumental variables: An evaluation of the Gaussian copula approach for management research. *Journal of Management*, 1–36.
- Haschka, R. E. (2022). Handling endogenous regressors using copulas: A generalisation to linear panel models with fixed effects and correlated regressors. *Journal of Marketing Research* 59(4), 860–881. DOI: 0.1177/00222437211070820.
- Haschka, R. E. (2024). Robustness of copula-correction models in causal analysis: Exploiting between-regressor correlation. *IMA Journal of Management Mathematics* 36(1), 161–180. DOI: 10.1093/imaman/dpae018.
- Herhausen, D., H. Van Heerde, S. Ludwig, and D. Grewal (2025). Has the pendulum swung too far? Reassessing endogeneity concerns in marketing research. *Reassessing Endogeneity Concerns in Marketing Research (May 08, 2025)*.
- Hu, X., Y. Qian, and H. Xie (2025). Correcting endogeneity via instrument-free two-stage nonparametric copula control functions. *NBER Working Paper*. DOI: 10.3386/w33607.
- Lienggaard, B. D., J.-M. Becker, M. Bennedsen, P. Heiler, L. N. Taylor, and C. M. Ringle (2025). Dealing with regression models' endogeneity by means of an adjusted estimator for the Gaussian copula approach. *Journal of the Academy of Marketing Science* 53, 279—299. DOI: 10.1007/s11747-024-01055-4.
- Park, S. and S. Gupta (2012). Handling endogenous regressors by joint estimation using copulas. *Marketing Science* 31(4), 567–586.
- Park, S. and S. Gupta (2024). A review of copula correction methods to address regressor–error correlation. *Impact at JMR*.
- Qian, Y. and H. Xie (2024). Correcting regressor-endogeneity bias via instrument-free joint estimation using semiparametric odds ratio models. *Journal of Marketing Research* 61(5), 914–936. DOI: 10.1177/00222437231195577.
- Yang, F., Y. Qian, and H. Xie (2025). Addressing endogeneity using a two-stage copula generated regressor approach. *Journal of Marketing Research* 62(4), 601–623. DOI: 10.1177/00222437241296453.