Florian Dost
Chair of Marketing
Brandenburg
University of Technology
Cottbus - Senftenberg

Addressing Endogeneity in Observational Data Models with Copula-based Methods

When Gaussian Copulas Work, When They Crack and a Novel Approach for Taking Up the Pieces

PDW @ AOM 25, Copenhagen, DK

27 July 2025

Rouven Haschka
rouven.haschka@rptu.de

Florian Dost
dost@b-tu.de
florian.dost@manchester.ac.uk

Florian Dost
Chair of Marketing
Brandenburg
University of Technology
Cottbus - Senftenberg

Agenda and Materials

Our plan for today:

1) Opening: *5min*
2) Introduction to the Gaussian Copula method: *25 min*
3) Using Gaussian Copula methods in R: *20 min*
4) Break: *5 min*
5) Revisiting the assumptions in a DAG; introducing new method: *15 min*
6) Using the new method in R: *10 min*
7) Discussion, Q&A, buffer: *10 min*

Materials: https://github.com/HashtagHaschka/AOM-Workshop

Notes/References:

Other stuff you might want or need:

- Papers on Copula:
  - Haschka, R. E. (2022). Handling endogenous regressors using copulas: a generalization to linear panel models with fixed effects and correlated regressors. Journal of Marketing Research, 59(4), 860-881.
  - Haschka, R. E. (2025). Robustness of copula-correction models in causal analysis: Exploiting between-regressor correlation. IMA Journal of Management Mathematics, 36(1), 161-180.
- Recent paper when Copula cracks:
  - Dost, Florian and Haschka, Rouven E., The Gaussian Copula Control Function Method Does Not Help Against Traditional Omitted Variable Bias (June 07, 2025). Available at SSRN: http://dx.doi.org/10.2139/ssrn.5285127
- New instrument-free method for omitted variables bias
  - Haschka, Rouven E. and Dost, Florian, ICA at the Cocktail Party: Casting Instrument-free Omitted Variable Bias Correction as a Blind Source Separation Problem (July 22, 2025). Available at SSRN: http://dx.doi.org/10.2139/ssrn.5361801

Florian Dost
Chair of Marketing
Brandenburg University of Technology Cottbus - Senftenberg

Explicit and implicit assumptions for Gaussian Copula Control Function (GCCF) methods

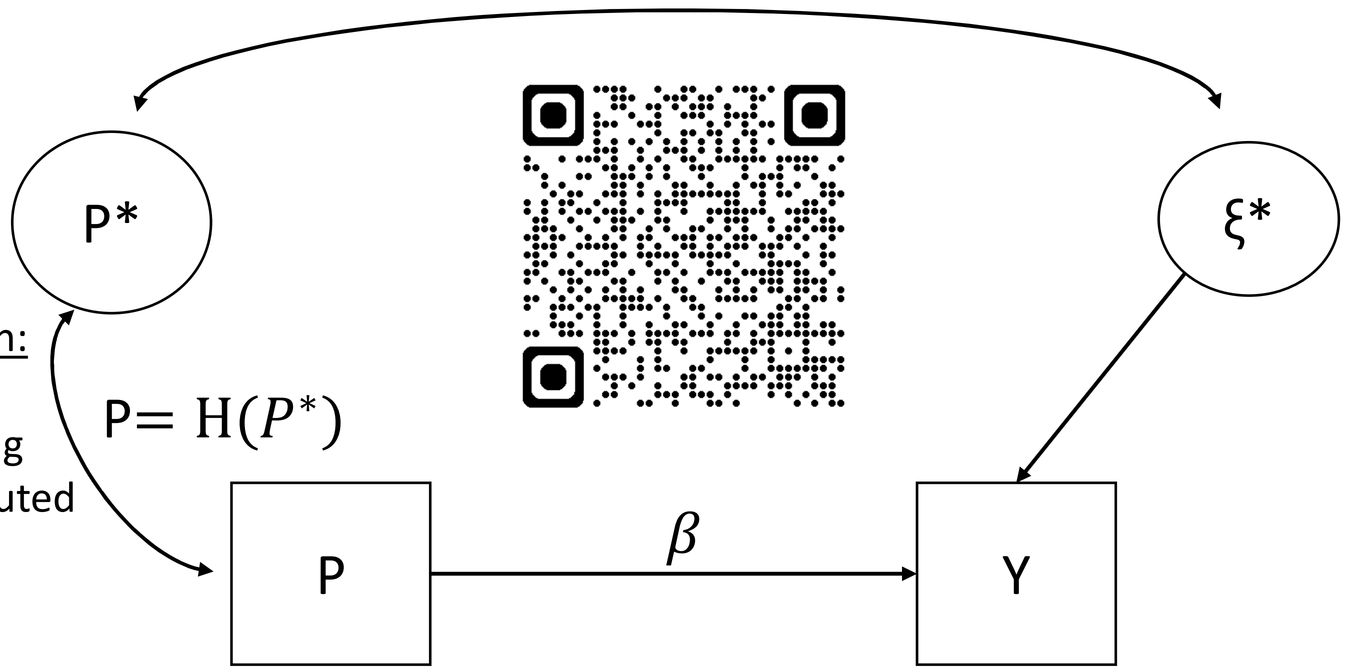
1st assumption: Joint dependence is bivariate-normal, with correlation ρ, such that a Gaussian copula captures the dependence

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

2nd assumption:
Error term ξ is normally distributed

$\xi^* \sim N$

$P^* \sim N$



4th, an implicit assumption:
H() is a bijective and non-linear one-to-one mapping between normally distributed P* and non-normally distributed P

3rd assumption:
Regressor P is sufficiently non-normally distributed

$P \nsim N$

$$Y = \alpha + \beta P + \xi$$
$$\xi = \xi^* \sigma$$

Notes/References: Park & Gupta (2012)
Dost & Haschka (2025) The Gaussian Copula Control Function Method does not help against Traditional Omitted Variable Bias. https://ssrn.com/abstract=5285127

Florian Dost
Chair of Marketing
Brandenburg
University of Technology
Cottbus - Senftenberg

What researchers use GCCF for? Omitted variable bias…

Literature using GCCF:

- **82** Articles using GCCF for identification
- Published before 01/06/2024.
- Journals searched: *Journal of Marketing, Journal of Marketing Research, Marketing Science, Journal of Consumer Research, Journal of the Academy of Marketing Science, Journal of Retailing, International Journal of Research in Marketing, Journal of Consumer Psychology, International Journal of Information Management, Academy of Management Perspectives, Management Science, Journal of Interactive Marketing, Quantitative Marketing and Economics.*

Nature of endogeneity problem:

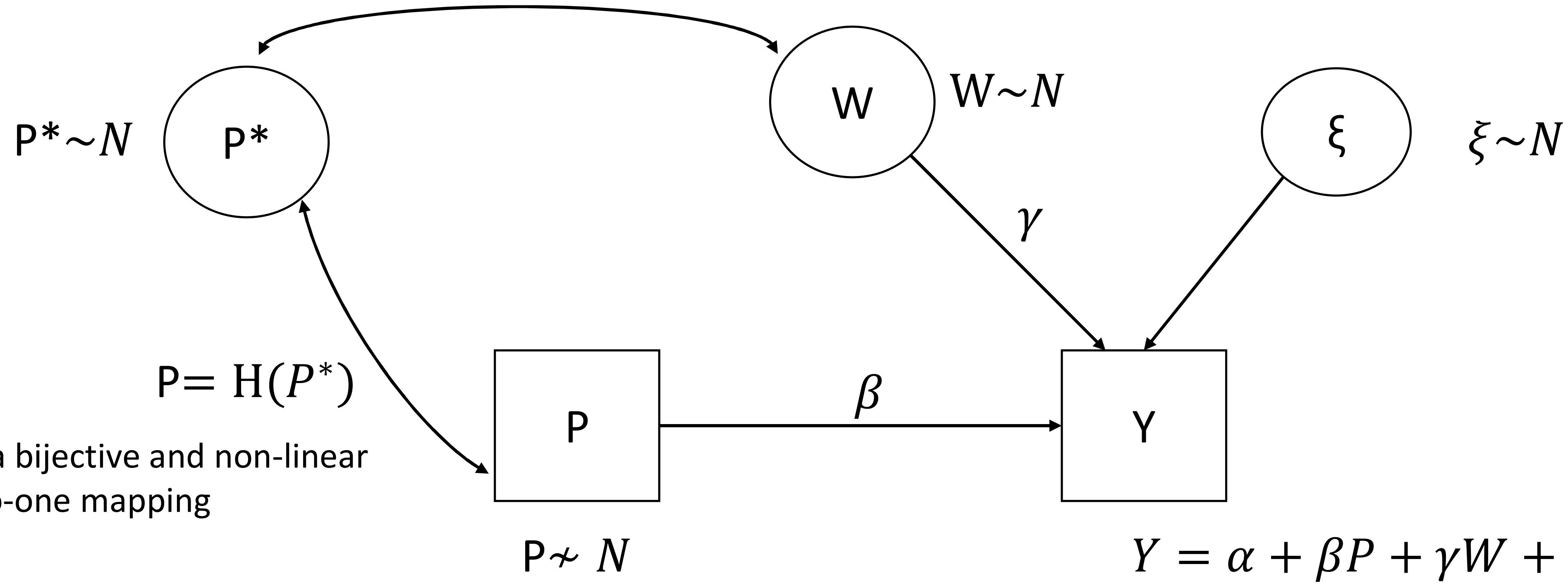
- Of 82 articles, 50 detail the assumed data-generating process.
- Of those, 46 articles **(92%) assume omitted variables bias**

Notes/References: Dost & Haschka (2025) The Gaussian Copula Control Function Method does not help against Traditional Omitted Variable Bias. https://ssrn.com/abstract=5285127

Extending Gaussian Copula Control Function (GCCF) assumptions to omitted variable DGP

Florian Dost
Chair of Marketing
Brandenburg University of Technology Cottbus - Senftenberg

$Cov[W, P^*] \neq 0$ e.g.,: $\binom{P^*}{W} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right)$

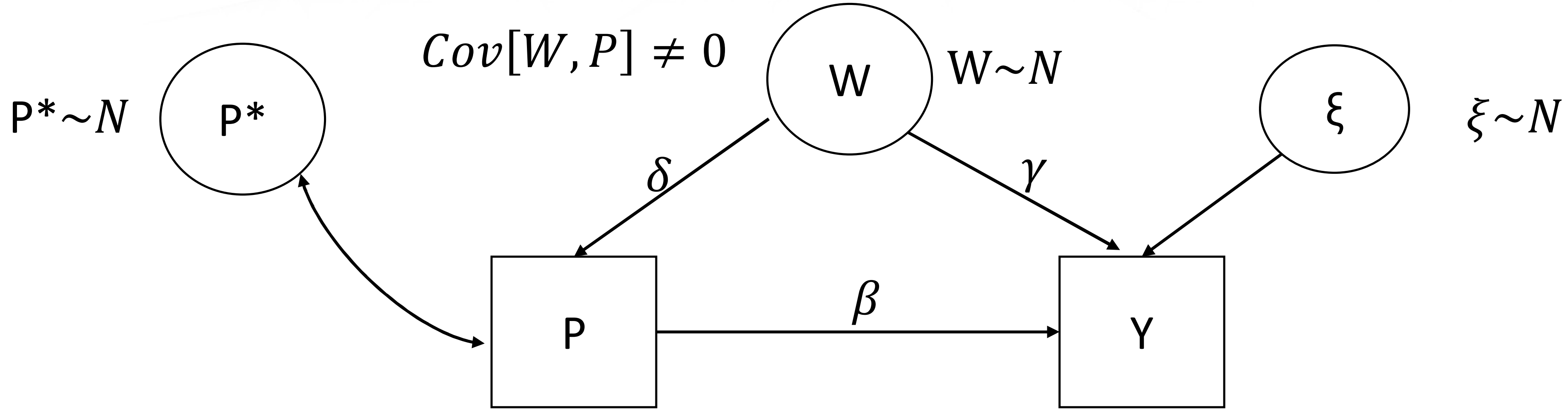


H() is a bijective and non-linear one-to-one mapping

Notes/References: Dost & Haschka (2025) The Gaussian Copula Control Function Method does not help against Traditional Omitted Variable Bias. https://ssrn.com/abstract=5285127

Omitted variable DGP with additive process and nonlinear transform

Florian Dost
Chair of Marketing
Brandenburg University of Technology Cottbus - Senftenberg



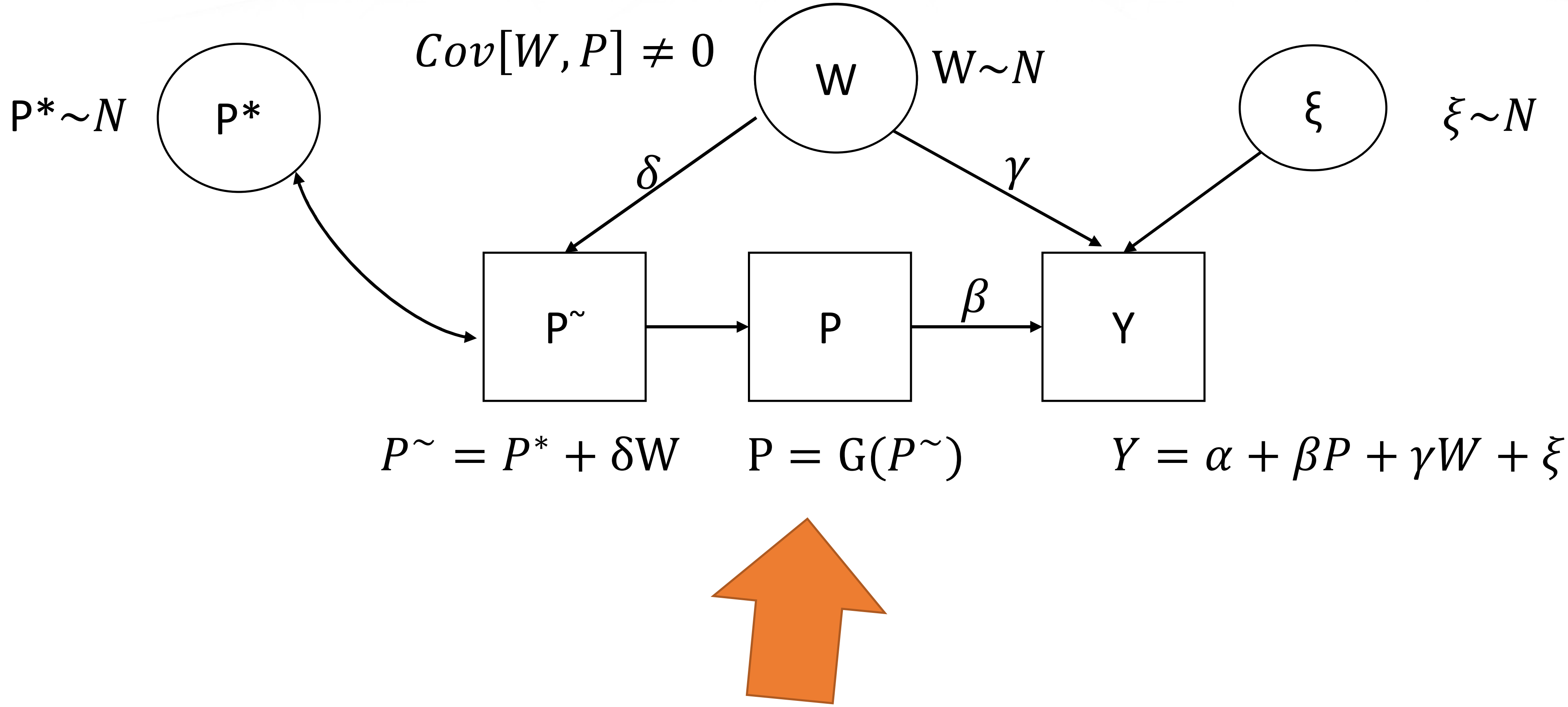
$$\mathrm{P} = \mathrm{G}(P^{*} + \delta \mathrm{W})$$

$$Y = \alpha + \beta P + \gamma W + \xi$$

G() is bijective and non-linear, such that: P$\nsim N$

Florian Dost
Chair of Marketing
Brandenburg University of Technology Cottbus - Senftenberg

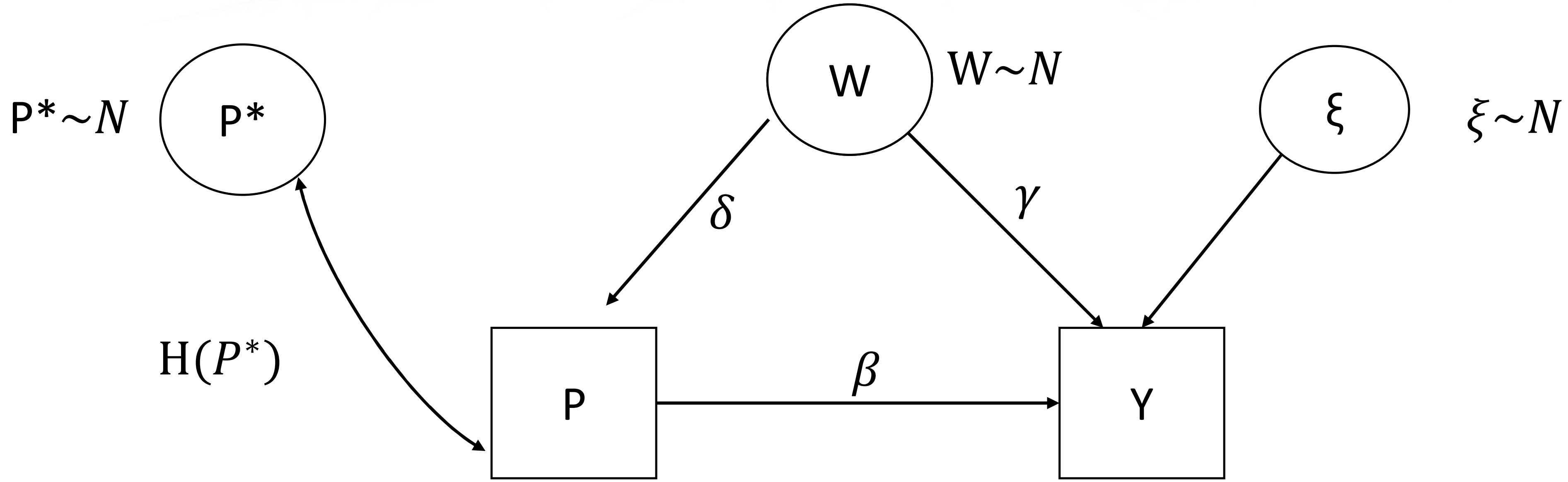
Equivalent notation with intermediate step:



Necessary!
G() is bijective and nonlinear

Notes/References: Dost & Haschka (2025) The Gaussian Copula Control Function Method does not help against Traditional Omitted Variable Bias. https://ssrn.com/abstract=5285127

Traditional omitted variable DGP with non-Gaussian exogenous component



$P \nsim N$

$Y = \alpha + \beta P + \gamma W + \xi$

$P = H(P^{*}) + \delta W$

Prove that GCCF estimation creates a bias:

$$\text{Bias} = \gamma \cdot \frac{\delta \sigma_W^2 - \text{Cov}\left(\mathbb{E}[P_t \mid \tilde{P}^*], \mathbb{E}[W_t \mid \tilde{P}^*]\right)}{\mathbb{E}[\text{Var}(P_t \mid \tilde{P}^*)]}$$

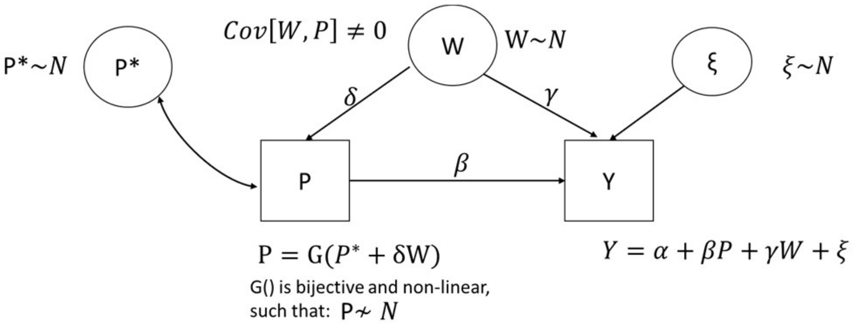
Notes/References: Dost & Haschka (2025) The Gaussian Copula Control Function Method does not help against Traditional Omitted Variable Bias. https://ssrn.com/abstract=5285127

Florian Dost
Chair of Marketing
Brandenburg University of Technology Cottbus - Senftenberg

Two simulation studies for illustration

Scenario A:

Omitted variable with bijective transform

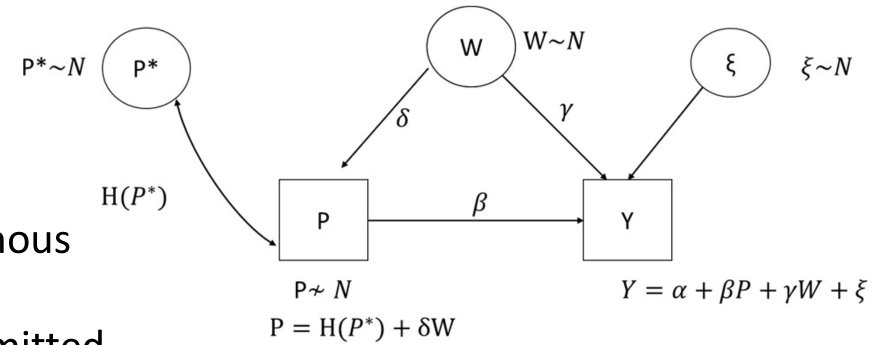


$P_t = G(P^*_t + \delta W_t)$ is the observed variable

- $\mathbf{Y_t} = \alpha + \beta P_t + \gamma W_t + \xi_t$,
- $\boldsymbol{\xi_t}$ is structural error with $E[\xi_t \mid P_t, W_t] = 0$, and $\xi_t \sim N(0, 1)$,
- $\alpha = 0, \beta = 1, \gamma = 1$,
- n = 1000 sample size, r = 500 repetitions
- $\mathbf{P^*_t} \sim N(0, 1)$is exogenous variance,
- $\mathbf{W_t} \sim N(0, 1)$ is the omitted variable,
- **δ** is randomly drawn from [0, 0.4] to induce endogeneity.
- $G() = H() = \Phi^{-1}()$ the canonical nonlinear transform to uniform.
- **Bias** = $\beta_{true} - \beta_{OLS/GCCF}$

Scenario B:

Traditional omitted variable



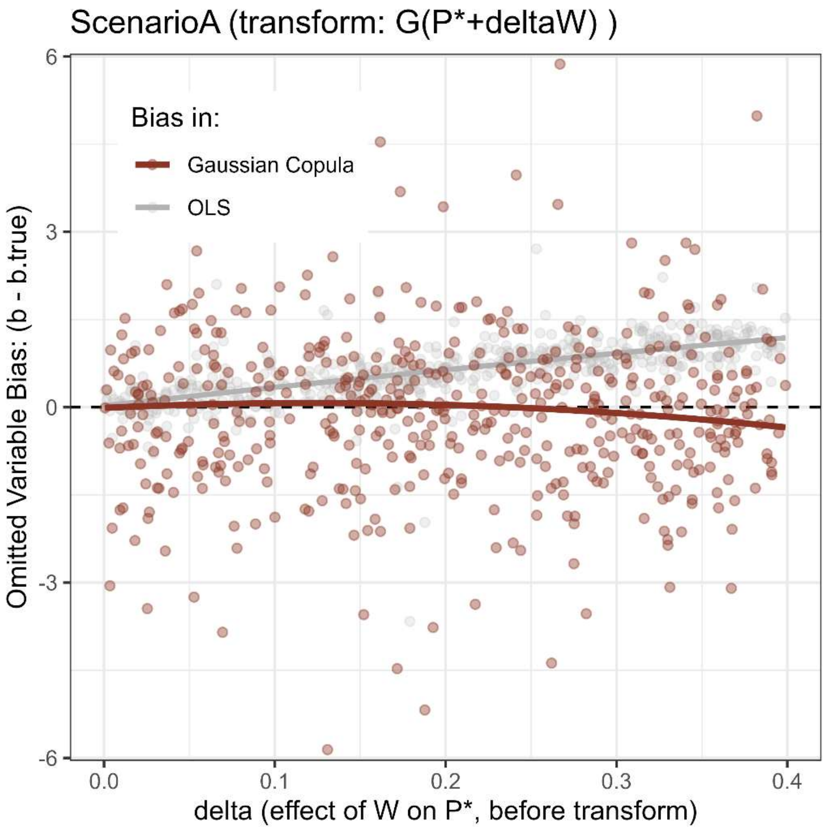
$P_t = H(P^*_t) + \delta W_t$ is the observed variable

Notes/References: Dost & Haschka (2025) The Gaussian Copula Control Function Method does not help against Traditional Omitted Variable Bias. https://ssrn.com/abstract=5285127

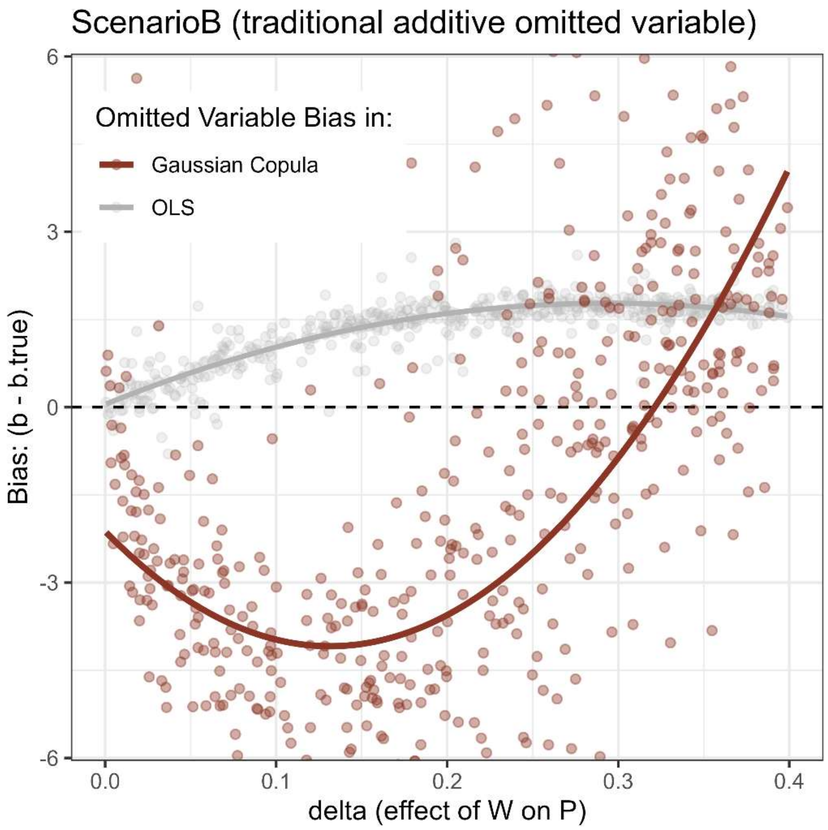
Florian Dost
Chair of Marketing
Brandenburg University of Technology Cottbus - Senftenberg

Simulation studies results:

Scenario A:
Omitted variable with bijective transform



Scenario B:
Traditional omitted variable



Notes/References: Dost & Haschka (2025) The Gaussian Copula Control Function Method does not help against Traditional Omitted Variable Bias. https://ssrn.com/abstract=5285127

Florian Dost
Chair of Marketing
Brandenburg
University of Technology
Cottbus - Senftenberg

Recommendations for GCCF use:

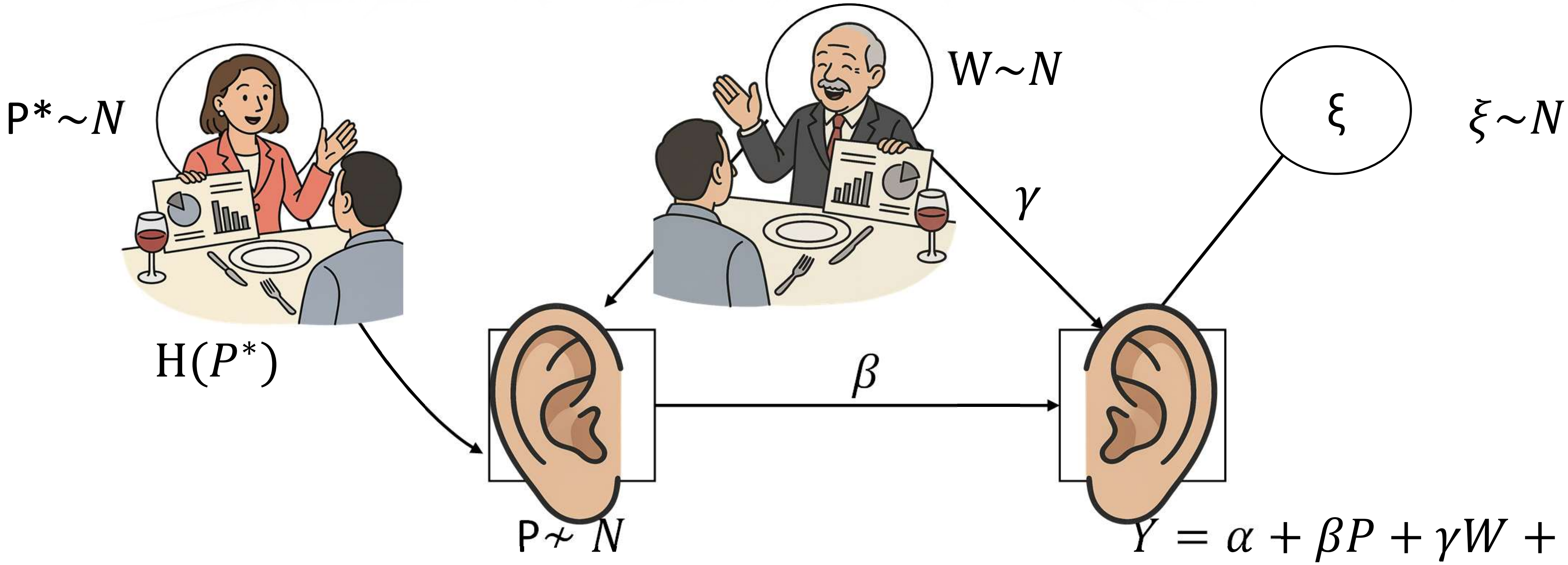
TLDR: Explicitly answer: Why is P **endogenous? AND:** Why is P **non-normal?**

- **1. Diagnose the DGP first, estimate second.** Begin every empirical project by writing down a verbal or formal data-generating process. If the story is "unobserved trait W adds to both P and the error," copula correction is contraindicated.
- **2. Treat marginal non-normality as necessary, never sufficient.** Confirming that P is non-normal merely clears only one of several hurdles; it does not establish that a monotone transform of the endogenous component drives P .
- **3. Document the monotone-transform rationale.** When using any copula method, spell out why the underlying mechanism should be nonlinear and strictly monotone (e.g., saturation, diminishing returns, ranking processes).
- **4. Document the order of additions and transforms in the DGP.** Importantly, the transform needs to happen after the omitted variable additively affected the exogenous variance in P.
- **5. Triangulate with IVs when possible.** If a credible instrument is available, compare IV and GCCF estimates. Divergence is an immediate red flag that the monotone-transform assumption may be violated.

Notes/References:

Reframe the omitted variable as “the cocktail party problem” from signal processing

Florian Dost
Chair of Marketing
Brandenburg University of Technology Cottbus - Senftenberg



$$Y = \alpha + \beta P + \gamma W + \xi$$

$$P = H(P^*) + \delta W$$

→ Implementation of an estimate for W as control function (R code)
→ Extensive simulation evidence for robustness in wide array of DGPs
→ canonical empirical cases

Notes/References: Haschka, Rouven E. and Dost, Florian, ICA at the Cocktail Party: Casting Instrument-free Omitted Variable Bias Correction as a Blind Source Separation Problem (July 22, 2025). Available at SSRN: https://ssrn.com/abstract=5361801

Using Independent Component Analysis (ICA) to disentangle exogeneous from omitted signal

Florian Dost
Chair of Marketing
Brandenburg University of Technology Cottbus - Senftenberg

- ICA:
  - Blind source separation technique
  - Based on whitening and rotation
  - Several variants (e.g., JADE)

- ICA assumptions:
  - As many (or more) measured signals as independent latent sources
  - Each signal is a linear combination of sources
  - Ideally, all sources are uncorrelated and non-normal
  - At most one source can be normal

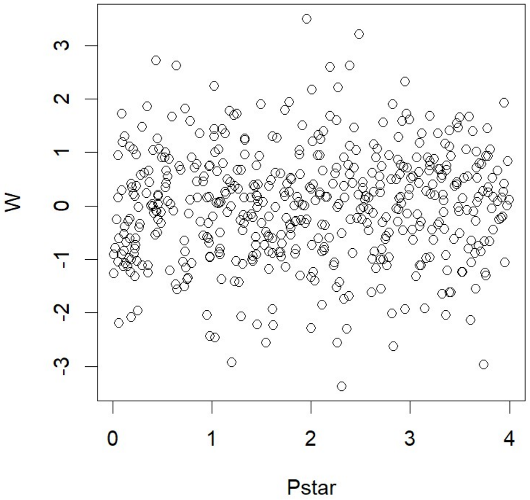
- Additional assumptions for our use case:
  - The exogeneous signal is non normal; the omitted signal (source of endogeneity) is normal
    → ICA is blind, we need the assumption to decide which component is exogenous

Notes/References: Haschka, Rouven E. and Dost, Florian, ICA at the Cocktail Party: Casting Instrument-free Omitted Variable Bias Correction as a Blind Source Separation Problem (July 22, 2025). Available at SSRN: https://ssrn.com/abstract=5361801

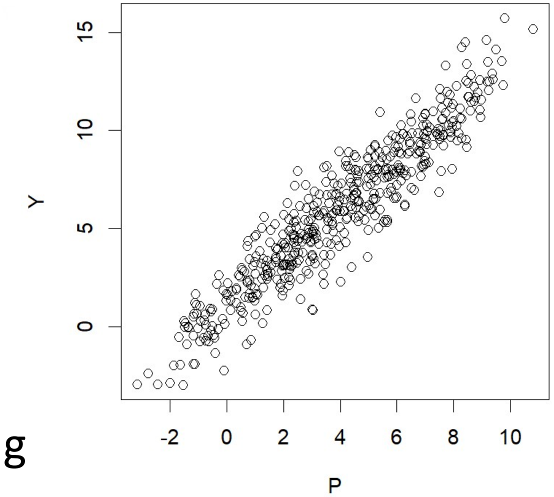
Florian Dost
Chair of Marketing
Brandenburg
University of Technology
Cottbus - Senftenberg

Using Independent Component Analysis (ICA) to disentangle exogeneous from omitted signal

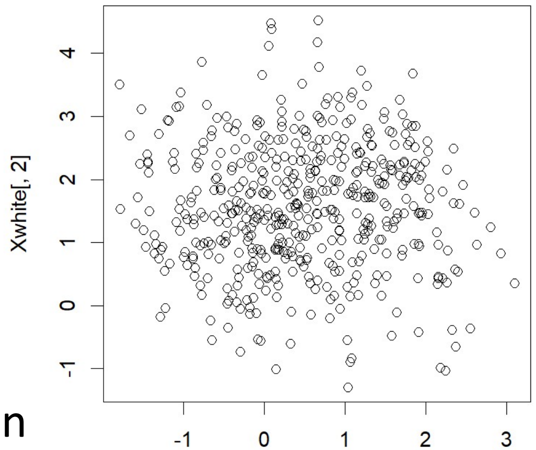
- Example:
  - $\mathbf{Y_t} = 2 + 1P_t + 1W_t + \xi_t$,
  - $\boldsymbol{\xi_t}$ is structural error with $E[\xi_t \mid P_t, W_t] = 0$, and $\xi_t \sim N\,(0, 1)$,
  - $\mathbf{P_t} = 2P^*_t + 1.5W_t$
  - $\mathbf{P^*_t} \sim U\,[0, 4]$is exogenous variance,
  - $\mathbf{W_t} \sim N\,(0, 1)$ is the omitted variable,
- Latent sources:



- Observed variables:



- ICA step 1: whitening



- ICA next step: rotation

Notes/References: Haschka, Rouven E. and Dost, Florian, ICA at the Cocktail Party: Casting Instrument-free Omitted Variable Bias Correction as a Blind Source Separation Problem (July 22, 2025). Available at SSRN: https://ssrn.com/abstract=5361801

Florian Dost
Chair of Marketing
Brandenburg University of Technology Cottbus - Senftenberg

Code and paper (with identification proof)

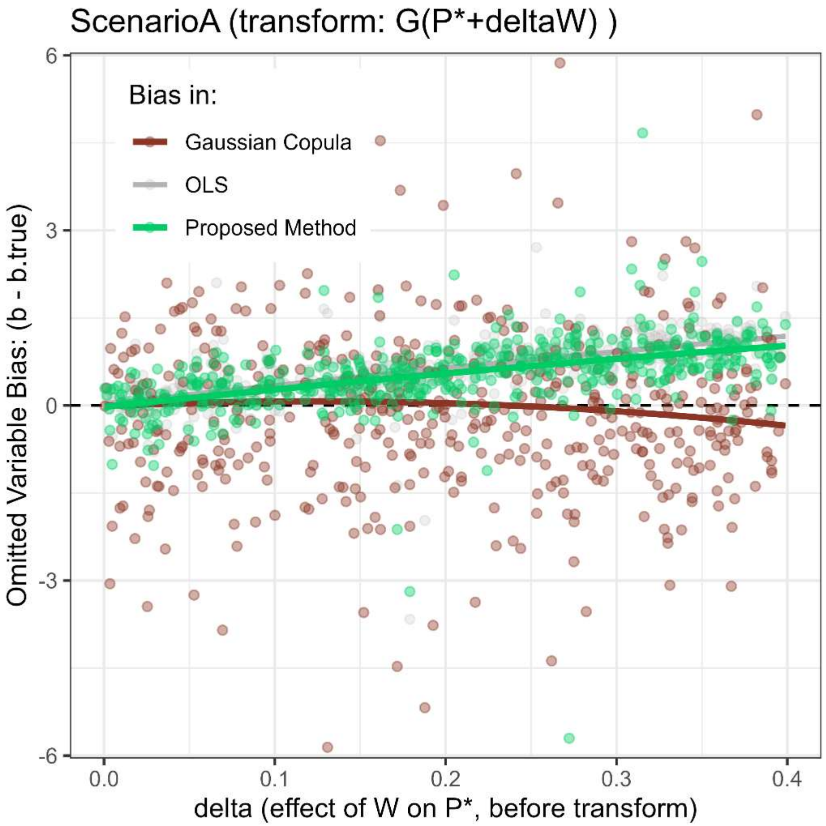
```r
# Step 3: Apply Independent Component Analysis (ICA)
data_matrix <- cbind(Y, P)
ica_result <- ica(X = data_matrix, nc = 2, method = "jade")
# Step 4: Find the ICA component that looks most normally distributed
hist(ica_result$S[, 1])
hist(ica_result$S[, 2])
# Step 5: Estimate the regression model, controlling for the normal component
model <- lm(Y ~ P + ica_result$S[, 1])
# View the results
summary(model)
```

Notes/References: Haschka, Rouven E. and Dost, Florian, ICA at the Cocktail Party: Casting Instrument-free Omitted Variable Bias Correction as a Blind Source Separation Problem (July 22, 2025). Available at SSRN: https://ssrn.com/abstract=5361801

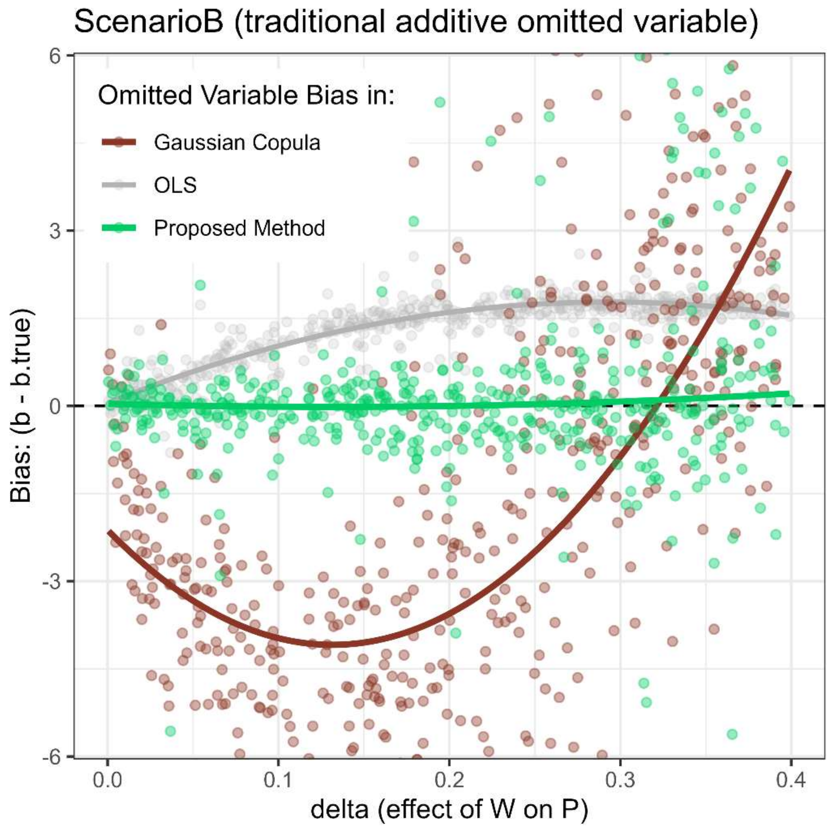
Florian Dost
Chair of Marketing
Brandenburg
University of Technology
Cottbus - Senftenberg

Revisit simulation studies:

Scenario A:
Omitted variable with bijective transform



Scenario B:
Traditional omitted variable



Notes/References:

Let‘s have a discussion!

- What did we miss that would make this method fail (spectacularly…)?
- What would you like to see the method do? (cases, scenarios, types of data, etc.)?

Brandenburgische Technische Universität Cottbus - Senftenberg

Dr. Florian Dost
Professor, Chair of Marketing

BTU Cottbus
Brandenburg University of Technology
Erich-Weinert-Strasse 1, 03046 Cottbus, Germany

E florian.dost@b-tu.de
T +49 (0) 355 69 2923

Honorary Professor at
Alliance Manchester Business School
Booth St West, Manchester M15 6PB, UK
E florian.dost@manchester.ac.uk

RPTU Faculty of Business and Economics

Dr. Rouven Haschka
Professor for Data Analytics

RPTU School of Business and Economics
Gottlieb-Daimler-Straße, 67663 Kaiserslautern, Germany

E rouven.haschka@rptu.de

Notes/References: