NBER WORKING PAPER SERIES

ADDRESSING ENDOGENEITY USING A TWO-STAGE COPULA GENERATED REGRESSOR APPROACH

Fan Yang
Yi Qian
Hui Xie

Working Paper 29708
http://www.nber.org/papers/w29708

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2022

We thank colleagues for helpful comments. We acknowledge the support by the Social Sciences and Humanities Research Council of Canada [grant 435-2018-0519], Natural Sciences and Engineering Research Council of Canada [grant RGPIN-2018-04313] and US National Institute of Health [grants R01CA178061] . All inferences, opinions, and conclusions drawn in this study are those of the authors, and do not reflect the opinions or policies of the funding agencies and data stewards. No personal identifying information was made available as part of this study. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Addressing Endogeneity Using a Two-stage Copula Generated Regressor Approach
Fan Yang, Yi Qian, and Hui Xie
NBER Working Paper No. 29708
January 2022
JEL No. C01,C1,C13,C18,C4

ABSTRACT

A prominent challenge when drawing causal inference using observational data is the ubiquitous presence of endogenous regressors. The classical econometric method to handle regressor endogeneity requires instrumental variables that must satisfy the stringent condition of exclusion restriction, making it infeasible to use in many settings. We propose new instrument-free methods using copulas to address the endogeneity problem. The existing copula correction method focuses only on the endogenous regressors and may yield biased estimates when exogenous and endogenous regressors are correlated. Furthermore, (nearly) normally distributed endogenous regressors cause model non-identification or finite-sample poor performance. Our proposed two-stage copula endogeneity correction (2sCOPE) method simultaneously overcomes the two key limitations and yields consistent causal-effect estimates with correlated endogenous and exogenous regressors as well as normally distributed endogenous regressors. 2sCOPE employs generated regressors derived from existing regressors to control for endogeneity, and is straightforward to use and broadly applicable. Moreover, we prove that exploiting correlated exogenous regressors can address the problem of insufficient regressor non-normality, relax identification requirements and improve estimation precision (by as much as 50% in empirical evaluation). Overall, 2sCOPE can greatly increase the ease of and broaden the applicability of instrument-free methods for dealing with regressor endogeneity. We demonstrate the performance of 2sCOPE via simulation studies and an empirical application.

Fan Yang
Sauder School of Business
2053 Main Mall
Vancouver, BC V6T 1Z2
Canada
fan.yang@sauder.ubc.ca

Yi Qian
Sauder School of Business
University of British Columbia
2053 Main Mall
Vancouver, BC V6T 1Z2
CANADA
and NBER
yi.qian@sauder.ubc.ca

Hui Xie
Department of Biostatistics
School of Public Health
University of Illinois at Chicago
huixie@uic.edu

1. Introduction

Causal inference is central to many problems faced by academics and practitioners, and becomes increasingly important as rapidly-available observational data in this digital era promise to offer real-world evidence on cause-and-effect relationships for better decision makings. However, a prominent challenge faced by empirical researchers to draw valid causal inferences from these data is the presence of endogenous regressors that are correlated with the structural error in the population regression model representing the causal relationship of interest. For example, omitted variables such as ability would cause endogeneity of schooling when examining schooling's effect on wages (Angrist and Krueger 1991).

Regressor endogeneity poses great empirical challenges to researchers and demands special handling of the issue in order to draw valid causal inferences. One classical method to deal with the endogeneity issue is using instrumental variables (IV). The ideal IV has to meet two requirements: it is correlated with the endogenous regressor via an explainable and validated relationship (i.e., relevance restriction), yet uncorrelated with the structural error (i.e., exclusion restriction). Although the theory of IVs is well-developed, researchers often face the challenge of finding good IVs satisfying these two requirements. Potential IVs often suffer from either weak correlation with endogenous regressors or challenging justification for exclusion restriction, which hampers using IVs to correct for the underlying endogeneity concerns (Rossi 2014).

To address the lack of suitable IVs, there has been a growing interest in developing and applying IV-free endogeneity-correction methods. Several instrument-free approaches have been developed, including identification via higher moments (Lewbel 1997, Erickson and Whited 2002), heteroscedasticity (Rigobon 2003, Hogan and Rigobon 2003), and latent instrumental variables (Ebbes et al. 2005). All three IV-free methods decompose the endogenous regressor into an exogenous part and an endogenous part. The assumption of the endogenous regressor containing an exogenous component is akin to the stringent condition of exclusion restriction for IVs, and thus can be difficult to justify.

Park and Gupta (2012) propose an alternative instrument-free method that uses the copula model to capture the regressor-error dependence.[1] Compared with the three IV-free methods above, their copula method does not impose the exogeneity assumption as it directly models the association between the structural error and the endogenous regressor via copula. Furthermore,

[1] In statistics, a copula is a multivariate cumulative distribution function where the marginal distribution of each variable is a uniform distribution on $[0, 1]$. Copulas permit modeling dependence without imposing assumptions on marginal distributions.

the copula method can handle discrete endogenous regressors better than other IV-free methods. These features considerably increase the feasibility of endogeneity correction, as evidenced by the rapidly increasing use of the copula correction method (see examples of recent applications in the next section on literature review). However, similar to other IV-free methods, the copula correction method also requires the distinctiveness between the distributions of the endogenous regressor and the structural error (Park and Gupta 2012). This means that the endogenous regressor is required to have a non-normal distribution for model identification with the commonly assumed normal structural error distribution. Furthermore, we show that the existing copula correction method implicitly requires all exogenous regressors to be uncorrelated with the linear combination of copula transformations of endogenous regressors (henceforth referred to as copula control function (CCF)) used to control for endogeneity, and may yield significant bias when there are noticeable correlations between the CCF and exogenous regressors.

In practice, we often encounter endogenous regressors or include transformations of endogenous variables as regressors that have close-to-normality distributions. Correlations between the CCF and exogenous regressors are quite common in practical applications, especially when the exogenous regressors are included to control for observed confounders. Although regressor normality or insufficient regressor non-normality leads to more severe identification issues, including model non-identifiability and poor finite sample performance (Table 1), correlations between CCF and exogenous regressors may occur more frequently than close-to-normality of endogenous regressors. Thus, we consider the two requirements of sufficient regressor non-normality and no correlation between CCF and exogenous regressors as being equally important, and either one can significantly limit the applicability of the copula correction method.

In this paper, we develop a generalized two-stage copula endogeneity correction method, denoted as 2sCOPE. Similar to the existing copula method (Park and Gupta 2012, denoted as CopulaP&G), 2sCOPE requires neither IVs nor the assumption of exclusion restriction. It corrects endogeneity by adding residuals obtained from regressing latent copula data for each endogenous regressor on the latent copula data for the exogenous regressors as generated regressors in the structural regression model. To demonstrate the benefits of 2sCOPE, we also consider as a benchmark another proposed method, called COPE, that corrects endogeneity by adding latent copula data themselves as generated regressors. Both COPEs methods (referring to COPE and 2sCOPE) are straightforward to use. However, only the preferred 2sCOPE overcomes the above two key limitations of CopulaP&G as shown in Table 1. CopulaP&G can be viewed as a special case of the 2sCOPE. Importantly, we prove that the 2sCOPE can identify causal effects under much broader settings than CopulaP&G, as summarized in Table 1.

The contributions of this work are several folds. *First*, we identify an implicit assumption required for CopulaP&G to yield consistent estimation, and provide conditions to verify this implicit assumption to ensure consistent causal-effect estimation. This helps improve the effectiveness of the rapidly adopted method for addressing the endogeneity issue. An important result is that the existence of the correlations between endogenous and exogenous regressors alone does not necessarily invalidate CopulaP&G. Instead, we show that the implicit assumption is the uncorrelatedness of the exogenous regressors with the CCF, the *linear combination* of copula transformations of endogenous regressors used to control for endogeneity. The difference between the implicit assumption and the condition of zero pairwise correlations between endogenous and exogenous regressors can be substantial, especially with multiple endogenous regressors.[2]

Second, we prove that the new 2sCOPE method yields consistent causal-effect estimates when the implicit assumption above is violated, which we show can cause biased causal effect estimates for CopulaP&G. *Third*, we relax the nonnormality assumption on the endogenous regressors. Specifically, we prove that the structural model with normally distributed endogenous regressors can be identified using the 2sCOPE method as long as one of the exogenous regressors correlated with endogenous ones is nonnormal, which is considerably more feasible in many practical applications. *Fourth*, we prove that when both COPE and 2sCOPE methods yield consistent estimates, 2sCOPE improves the efficiency (i.e., precision) of the structural model estimation by exploiting the correlations between the endogenous and exogenous regressors. The efficiency gain is substantial and can be up to ∼50% in our empirical evaluation, meaning that sample size can be reduced by ∼50% to achieve the same estimation efficiency as compared with the COPE method that does not exploit the correlations between endogenous and exogenous regressors.

Finally, 2sCOPE employs generated regressors to address endogeneity. There are a number of benefits associated with the generated-regressor approach. By including generated regressors

[2]Specifically, it is possible that with multiple endogenous regressors, the CCF is uncorrelated with exogenous regressors when pairwise correlations between endogenous and exogenous regressors are non-zeros. Even if there is only one endogenous regressor and CCF reduces to be proportional to the copula transformation of the endogenous regressor, the correlation coefficient is not invariant to nonlinear transformations and thus changes after the copula transformation of the endogenous regressor.

in the structural model to control endogeneity, 2sCOPE substantially reduces the burden to address the endogeneity issues. The vast majority of applications of the existing copula correction method have used the generated-regressor approach (Becker et al., 2021). We demonstrate that 2sCOPE retains this desirable property of simplicity for a range of commonly used models in marketing studies, as shown in Table 1, while relaxing the two key limitations of CopulaP&G. Furthermore, the generated-regressor approach facilitates studying theoretical properties of the proposed COPEs procedures and the comparison of these procedures. In this work, we provide theoretical proofs for the implicit assumption needed to ensure consistency of CopulaP&G, and the consistency and efficiency comparison for the proposed COPEs under correlated regressors and normally distributed regressors.

A novel finding of our theoretical investigation is as follows. Although the exogenous regressors being correlated with endogenous regressors require special handling for consistent causal-effect estimation (the first and second contributions above), they can be beneficial as well by providing additional information to help relax model identification requirements. They could help address the problem of insufficient regressor non-normality, and sharpen model estimates, as described for the third and fourth contributions above. Overall, the proposed 2sCOPE method can greatly broaden the applicability of the instrument-free methods for dealing with endogeneity issues in practice.

To our knowledge, 2sCOPE is the first copula-correction method that relaxes the non-normality assumption of endogenous regressors and handles correlated endogenous and exogenous regressors (Table 1). The theoretical results presented in this work contribute to better understanding of the properties of the copula correction methods and guiding their practical use, and are much needed because model identifiability is central to address the endogeneity issue and timely given recent progress made in this area (Table 1). Despite rapid adoption of the copula correction method, recent research based on simulation studies raise concerns about its performance in models with the intercept term and insufficient regressor non-normality, and calls for further studies of its properties (Becker et al. 2021). However, Becker et al. (2021) has not considered exogenous regressors. Haschka (2021)[3] generalizes Park and Gupta (2012) to linear panel models with correlated regressors using maximum likelihood estimation (MLE). However, as noted in Haschka (2021), Haschka's approach still requires non-normality of en-

[3]During the final stage of our writing, we became aware of Haschka (2021), which we add to Table 1.

dogenous regressors. In addition, the analysis restricts to linear panel models with common slope coefficients and heterogeneity only in individual-specific intercepts. First-difference or fix-effects transformation eliminates these individual-specific intercept parameters, resulting in simple likelihood containing no individual-specific parameters. [4] Consequently, the approach cannot estimate the intercept term, which is removed from the model prior to estimation using first-difference or fixed-effects transformation (Web Appendix A.8 of Haschka (2021)). Overall, the approach does not address the same question of having both intercept and insufficient regressor non-normality as raised in Becker et al. (2021). Finally, owing to the complex form of the estimation method, Haschka (2021) notes the lack of theoretical proofs of required model identification conditions and estimation consistency as one limitation, and thus has to rely solely on simulation studies to evaluate its empirical properties.

We study the general case of having intercept, regressor normality, and correlated regressors for a variety of types of structural models (Table 1). We precisely identify the implicit identification requirement for CopulaP&G.[5] We then develop 2sCOPE employing a copula control function approach with theoretical proofs of its capability to eliminate the identification requirement and to address the problem of regressor normality in the presence of the structural intercept term. We further provide proofs of required conditions for model identification and estimation consistency. Consequently, the 2sCOPE provides a theoretically-sound and widely applicable solution to the issue raised in Becker et al. (2021) besides the other noted issues.

The remainder of this paper unfolds as follows. Section 2 reviews the related literature on methods for causal inference with endogenous regressors. In Section 3, we show the implicit assumption of CopulaP&G. We propose COPE first for an easier transition followed by the recommended 2sCOPE procedure, providing theoretical proofs for the consistency of the proposed COPEs methods as well as for efficiency gain and model identifiability with normally distributed regressors under the 2sCOPE method. We also summarize the estimation procedure of the proposed methods. In Section 4, we evaluate the performance of our proposed 2sCOPE method using simulation studies and compare it with CopulaP&G under different scenarios.

[4] It is cumbersome to extend Haschka (2021) to models with heterogeneous slope coefficients, which cannot be eliminated by first-difference or fix-effects transformation and will yield complicated likelihood requiring numerical integration over random coefficients (see Section 3.5).

[5] Although Haschka (2021) explains why correlated regressors can cause potential bias for CopulaP&G, no condition of when bias can occur is given. As explained above, non-zero pairwise correlations between endogenous and exogenous regressors does not necessa6rily cause the bias problem for CopulaP&G.

In Section 5, we apply the proposed 2sCOPE method to estimate price elasticity using store purchase databases. We conclude the paper in Section 6.

Table 1: A Comparison of Copula Methods

Features	CopulaP&G	Haschka (2021)	2sCOPE
Nonnormality of Endogenous Regressors[1]	Required	Required	Not Required[2]
No Correlated Exogenous Regressors[3]	Required (implicit)	Not Required	Not Required
Intercept Included[4]	NO	NO[5]	YES
Theoretical Proof	YES	NO	YES
Estimation Method	Control Function & MLE	MLE	Control Function
Structural Model	Linear Regression RCL Slope Endogeneity	LPM-FE	Linear Regression LPM-FE, LPM-RE, LPM-ME RCL, Slope Endogeneity

Note: [1]: When required, normality of any endogenous regressor leads to non-identifiable models. Insufficient normality of endogenous regressors can also cause poor finite sample performance (finite sample bias and large standard errors) and require extremely large sample size to perform well.
[2]: Non-normality of endogenous regressors is not required as long as at least one correlated exogenous regressor is not normally distributed.
[3]: In our paper, correlated exogenous regressors refer to those exogenous regressors correlated with the CCF (copula control function) used to control for endogeneity.
[4]: Becker et al. (2021) shows the significance of including intercept in marketing applications, and the problem of adding intercept using the copula method CopulaP&G (Park and Gupta 2012).
[5]: The approach cannot estimate the intercept term, which is removed from the panel model prior to estimation using first-difference or fix-effects transformation (Web Appendix A.8 of Haschka (2021)).
LPM: Linear Panel Model; FE: Fixed Effects for individual-specific intercepts with common slope coefficients; RE: Random Effects; ME: Mixed-Effects (including both fixed-effects and random coefficients); RCL: Random Coefficient Logit

2. Literature Review

The marketing, economic and statistics literature develops a rich set of methods to draw causal inferences. The gold standard to estimate causal effects is randomized assignment such as controlled lab experiments and field experiments (Johnson et al. 2017, Anderson and Simester 2004, Godes and Mayzlin 2009). When controlled experiments are not feasible, quasi-experimental designs such as regression discontinuity and difference in differences are used to mimic randomized experiments and to enable the identification of causal effects with observational data (Hahn et al. 2001, Hartmann et al. 2011, Narayanan and Kalyanam 2015, Athey and Imbens

2006, Shi et al. 2017). However, these quasi-experimental designs have special data and design requirements, and cannot cope with the general issue of endogenous regressors when estimating causal effects using observational data.

There is a large literature focusing on approaches to addressing endogenous regressors when inferring causal effects. Rutz and Watson (2019), Papies et al. (2017) and Park and Gupta (2012) provide an overview of addressing endogeneity in marketing. Three broad classes of solutions are discussed. The most commonly used solution is to find observed instrumental variables to correct for endogeneity (Kleibergen and Zivot 2003, Qian 2008, Ataman et al. 2010, Van Heerde et al. 2013 and Novak and Stern 2009). Angrist and Krueger (2001) and Rossi (2014) provide a survey of literature that uses the instrumental variables approach. Rossi (2014) surveyed 10 years of publications in *Marketing Science* and *Quantitative Marketing and Economics*, which revealed that the most commonly used instrumental variables are lagged variables, costs, fixed effects and Hausman style variables from other markets. However, the survey found that the strength of the instruments is rarely measured and reported, which is needed to detect the weak instrument problem. Moreover, one generally cannot test the exclusion restriction condition and verify the validity of instruments. The survey also found that most papers lack a discussion of why the instruments used are valid. In a word, though the theory of instrumental variables is well-developed, good instruments are difficult to find, making the IV approach hard to implement in practice. Studies that identify good instruments are subsequently highly valued.

The second class of solutions to mitigate endogeneity is to specify the economic structure that generates the observational data including endogenous regressors (e.g., a supply-side model for marketing-mix variables). Doing so allows researchers not only to recover parameters of interest and make causal inferences, but also to perform counterfactual analysis (Chintagunta et al. 2006). Some other examples of this approach in the marketing literature are Berry (1994), Sudhir (2001), Dubé et al. (2002), Yang et al. (2003), Sun (2005), Dotson and Allenby (2010) and Otter et al. (2011). The key concern with this approach is that the performance highly depends on model assumptions of supply side. Incorrect assumptions or insufficient information of the supply-side can lead to biased estimates (Chintagunta et al. 2006, Hartmann et al. 2011)

The third class of solutions in the domain of endogeneity correction is instrument-free methods. This is a more recent stream of methodological development. Three extant instrument-free approaches are discussed in Ebbes et al. (2009): the higher moments (HM) approach (Lewbel

1997, Erickson and Whited 2002), the identification through heteroscedasticity (IH) estimator (Rigobon 2003, Hogan and Rigobon 2003), the latent instrumental variables (LIV) method (Ebbes et al. 2005). Recently Wang and Blei (2019) proposed a deconfounder approach that has some flavor of the LIV approach. All these approaches divide the endogenous regressor P into an endogenous and an exogenous part, $P = f(Z) + v$, where $f(Z)$ is treated as an exogenous random variable with unique structures imposed for model identification in different methods. However, the assumption of $f(Z)$ being exogenous is hard to guarantee. Park and Gupta (2012) introduce another instrument-free method that doesn't require the exogeneity of $f(Z)$. It directly models the association between the structural error and the endogenous regressor via copula.

The copula method has been rapidly adopted by researchers to deal with the endogeneity problem because of its feasibility to use in that no instruments are needed. For example, the copula method has been used to study the effects of marketing activities such as promotion, advertising and loyalty programs (Burmester et al. 2015, Datta et al. 2015, Gruner et al. 2019, Keller et al. 2019, Bombaij and Dekimpe 2020, Guitart et al. 2018, Lamey et al. 2018); to study product design and brand equity (Wetzel et al. 2018, Heitmann et al. 2020); to study sales force training (Atefi et al. 2018); to study healthy food consumption (Elshiewy and Boztug 2018). Haschka (2021) develops an MLE method that extends Park and Gupta (2012) to linear panel models with fixed-effect intercepts and constant slope coefficients in the presence of correlated regressors. In our paper, we delineate the precise and verifiable condition for CopulaP&G to yield consistent estimates with correlated endogenous and exogenous regressors. For the case when this condition fails, we develop a new two-stage endogeneity correction method using copula control functions (2sCOPE) that relaxes two key assumptions imposed in Park and Gupta (2012): (1) all endogenous regressors must have non-normal distributions and (2) exogenous regressors must be uncorrelated with the CCF used to control for the endogeneity. We provide proofs of the theoretical properties of the proposed methods, including consistency and efficiency comparisons. We derive the new procedures for a variety of types of structural models, including the random coefficients models commonly used in marketing studies. As a result, the proposed 2sCOPE method is applicable in more general settings with the capability to exploit exogenous regressors to improve model identification and estimation.

3. Methods

In this section, we develop two copula-based instrument-free methods to handle endogenous regressors when there exist exogenous regressors that are correlated with endogenous regressors. We first review the CopulaP&G method in Park and Gupta (2012). We show that CopulaP&G implicitly assumes no correlations between the exogenous regressors and the CCF, as well as how the violation of the assumption can cause bias in the structural model parameter estimates for the current copula-based instrument-free method. Then we present two proposed methods to deal with the problem and the detailed estimation procedure. We also show how exogenous regressors correlated with endogenous regressors can sharpen structural model parameter estimates and permit the identification of the structural model containing normally distributed endogenous regressors, known to cause the model non-identifiability issue for CopulaP&G.

3.1 *Assumptions in the Existing Copula Endogeneity-Correlation Method (CopulaP&G)*

Consider the following linear structural regression model with an endogenous regressor and a vector of exogenous regressors[6]:

$$Y_t = \mu + P_t\alpha + W_t'\beta + \xi_t, \tag{1}$$

where $t = 1, 2, ..., T$ indexes either time or cross-sectional units, Y_t is a (1×1) dependent variable, P_t is a (1×1) endogenous regressor, W_t is a (k×1) vector of exogenous regressors, ξ_t is the structural error term, and (μ, α, β) are model parameters. P_t is correlated with ξ_t, and this correlation generates the endogeneity problem. W_t is exogenous, which means it is not correlated with ξ_t, but can be correlated with the endogenous variable P_t.

The key idea of the copula method (Park and Gupta 2012) is to use a copula to jointly model the correlation between the endogenous regressor P_t and the error term ξ_t. The advantage of using copula is that marginals are not restricted by the joint distribution. Using information contained in the observed data, marginals of the endogenous regressor and the error term are first obtained respectively. Then the copula model enables researchers to construct a flexible multivariate joint distribution that captures the correlation between the two variables.

Let $F(P, \xi)$ be the joint cumulative distribution function (CDF) of the endogenous regressor P_t and the structural error ξ_t with marginal CDFs $H(P)$ and $G(\xi)$, respectively. For notational simplicity, we may omit the index t in P_t and ξ_t below when appropriate. According to Sklar's

[6] Unlike Park and Gupta (2012), our model includes the intercept term. As shown in Becker et al. (2021), it is important to include the intercept term when evaluating the copula correction method.

theorem (Sklar 1959), there exists a copula function $C(\cdot,\cdot)$ such that for all P and ξ,

$$F(P,\xi) = C(H(P), G(\xi)) = C(U_p, U_\xi), \tag{2}$$

where $U_p = H(P)$ and $U_\xi = G(\xi)$, and they both follow uniform(0,1) distributions. Thus, the copula maps the marginal CDFs of the endogenous regressor and the structural error to their joint CDF, and makes it possible to separately model the marginals and correlations of these random variables.

To capture the association between the endogenous regressor P and the error ξ, Park and Gupta (2012) uses the following Gaussian copula for its many desirable properties (Danaher and Smith 2011):

$$\begin{aligned} F(P,\xi) &= C(U_p, U_\xi) = \Psi_\rho(\Phi^{-1}(U_p), \Phi^{-1}(U_\xi)) \\ &= \frac{1}{2\pi(1-\rho^2)^{1/2}} \int_{-\infty}^{\Phi^{-1}(U_p)} \int_{-\infty}^{\Phi^{-1}(U_\xi)} \exp\left[\frac{-(s^2 - 2\rho \cdot s \cdot t + t^2)}{2(1-\rho^2)}\right] dsdt, \end{aligned} \tag{3}$$

where $\Phi(\cdot)$ denotes the univariate standard normal distribution function and $\Psi_\rho(\cdot,\cdot)$ denotes the bivariate standard normal distribution with the correlation coefficient ρ. In the Gaussian copula model, ρ captures the endogeneity of the regressor P, and a non-zero value of ρ corresponds to P being endogenous.

Under the above copula model for (P_t, ξ_t) and the commonly-assumed normal distribution for the structural error ξ_t, Park and Gupta (2012) develop the following generated regressor procedure to correct for regressor endogeneity. Let $P_t^* = \Phi^{-1}(U_p)$ and $\xi_t^* = \Phi^{-1}(U_\xi)$, the above Gaussian copula assumes $[P_t^*, \xi_t^*]'$ follow the standard bivariate normal distribution with the correlation coefficient ρ as follows:

$$\begin{pmatrix} P_t^* \\ \xi_t^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \tag{4}$$

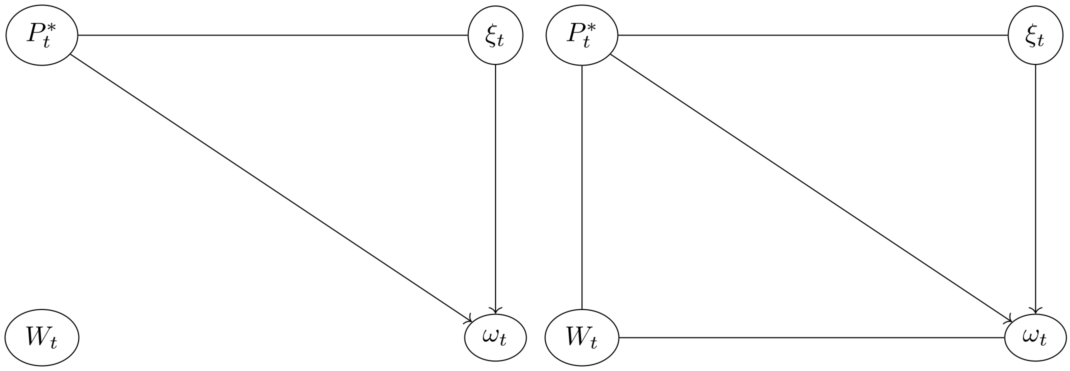
Under the assumption that the structural error ξ_t follows $N(0, \sigma_\xi^2)$, Park and Gupta (2012) show that the structural error can be split into two parts as follows:

$$\xi_t = \sigma_\xi \xi_t^* = \sigma_\xi \rho P_t^* + \sigma_\xi \sqrt{1-\rho^2}\omega_t, \tag{5}$$

where the first part $\sigma_\xi \rho P_t^*$ captures the correlation between ξ_t and the endogenous regressor, and the other part $\sigma_\xi \cdot \sqrt{1-\rho^2}\omega_t$ being an independent new error term. Equation (1) can be rewritten as follows:

$$Y_t = \mu + P_t\alpha + W_t\beta + \sigma_\xi \cdot \rho \cdot P_t^* + \sigma_\xi \cdot \sqrt{1-\rho^2} \cdot \omega_t. \tag{6}$$

Based on the above representation, Park and Gupta (2012) suggest the following generated regressor approach to correcting for the endogeneity of P_t: the ordinary least square (OLS)



(a) No Correlation Between P_t^* and W_t (b) Has Correlation Between P_t^* and W_t

Figure 1. Correlation Between W_t and New Error ω_t. Presence (absence) of a solid line between two variables means the two variables are correlated (uncorrelated). A line without an arrow represents stochastic association between two nodes. A line with an arrow represents a deterministic relationship. Specifically, ω_t is determined jointly by P_t^* and ξ_t.

estimation of Equation (6) with $P_t^* = \Phi^{-1}(U_p)$ included as an additional regressor will yield consistent model estimates. Park and Gupta (2012) also pointed out that in order for the above approach to work, P_t needs to have a non-normal distribution. Suppose P_t is normally distributed, $P_t = P_t^* \cdot \sigma_p$, resulting in perfect collinearity between P_t and P_t^* and violating the full rank assumption required for identifying the linear regression model in Equation (6). Thus, P_t should follow a different distribution from the normal error term so that the causal effect of P that is independent of all other regressors can be identified.

However, we show here that an additional and implicit assumption for the above generated regressor approach to yield consistent model estimates is the uncorrelatedness between P_t^* and W_t. For the OLS estimation to yield consistent estimation, the error term ω_t in Equation (6) is required to be uncorrelated with all the regressors on the right-hand side of the equation: P_t, W_t, P_t^*. Figure 1 shows how the correlation between W_t and the new error term ω_t changes when W_t becomes correlated with P_t^*. Absence of a line between two variables means that the two variables are not correlated. When W_t is not correlated with P_t^*, W_t should also be uncorrelated with ω_t, which is determined by ξ_t and P_t^*, because of the exogenous feature of W_t (Figure 1 (a)). However, when W_t is correlated with P_t^*, it would become correlated with ω_t because (1) ω_t is a linear combination of ξ_t and P_t^* (Equation 5), and (2) W_t is uncorrelated with ξ_t. The induced correlation between the exogenous regressor W_t and the new error term ω_t is intuitively shown in Figure 1 (b) and formally proved in Theorem 1 below. Thus, the correlation between the exogenous regressor W_t and the generated regressor P_t^* would cause biased OLS estimates of Equation (6) using CopulaP&G because of the induced correlation between the error term ω_t and W_t. That is, W_t becomes endogenous in Equation (6) when W_t and P_t^* are correlated.

Theorem 1. ***Assuming (1) the error term is normal, (2) a Gaussian Copula for the structural error term and P_t, and (3) P_t is endogenous:*** $\rho \neq 0$, $Cov(\omega_t, W_t) = -\frac{\rho}{\sqrt{1-\rho^2}} Cov(W_t, P_t^*) \neq 0$ ***if*** P_t^* ***and*** W_t ***are correlated.***

Proof: See Online Appendix A, Proof of Theorem 1.

To summarize, the generated regressor procedure based on Equation (6) makes the following set of assumptions.

Assumption 1. *The structural error follows a normal distribution;*

Assumption 2. *P_t and the structural error follow a Gaussian copula;*

Assumption 3. *Nonnormality of the endogenous regressor P_t;*

Assumption 4. *W_t and P_t^* are uncorrelated.*

In Park and Gupta (2012), all the above assumptions except Assumption (4) have been made explicit. Among the first three assumptions, Park and Gupta (2012) have shown reasonable robustness of their copula method to non-normal distributions of error term (Assumption 1) and alternative forms of copula functions (Assumption 2). By contrast, the assumption that the endogenous regressor P_t follows a non-normal distribution (Assumption 3) is critical. An endogenous regressor following a normal distribution can cause the structural model to be unidentifiable regardless of sample size; a nearly normally distributed endogenous regressor may require a very large sample size for the method to perform well and may cause the method to have poor performance for a finite sample size. Moreover, we have shown above that for their method to work, there should be no exogenous regressors that are correlated with P_t^* (Assumption 4). Both the Assumptions (3 and 4) can be too strong and substantially limit the applicability of the instrument-free copula method in practice. In the following two subsections, we will develop two new methods in order to relax the latter two critical assumptions.

3.2 *Proposed Method I: Copula Endogeneity-correction (COPE)*

The first proposed copula method jointly models the endogenous regressor, P_t, the correlated exogenous variable, W_t, and the structural error term, ξ_t. The copula model implies that

$[P_t^*, W_t^*, \xi_t^*]$ follows the multivariate normal distribution:

$$\begin{pmatrix} P_t^* \\ W_t^* \\ \xi_t^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{pw} & \rho_{p\xi} \\ \rho_{pw} & 1 & 0 \\ \rho_{p\xi} & 0 & 1 \end{bmatrix} \right), \tag{7}$$

where $P_t^* = \Phi^{-1}(H(P_t))$, $W_t^* = \Phi^{-1}(L(W_t))$, and $\xi_t^* = \Phi^{-1}(G(\xi_t))$, and $H(\cdot)$, $L(\cdot)$ and $G(\cdot)$ are marginal CDFs of P_t, W_t and ξ_t respectively. Since ξ_t is assumed to be normally distributed, $\xi_t^* = \Phi^{-1}(\Phi_{(0,\sigma_\xi)}(\xi_t)) = \sigma_\xi \cdot \xi_t$. The above multivariate distribution can be rewritten as follows:

$$\begin{pmatrix} P_t^* \\ W_t^* \\ \xi_t^* \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \rho_{pw} & \sqrt{1-\rho_{pw}^2} & 0 \\ \rho_{p\xi} & \frac{-\rho_{pw}\rho_{p\xi}}{\sqrt{1-\rho_{pw}^2}} & \sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2\rho_{p\xi}^2}{1-\rho_{pw}^2}} \end{pmatrix} \cdot \begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \\ \omega_{3,t} \end{pmatrix},$$

$$\begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \\ \omega_{3,t} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right). \tag{8}$$

Then, the structural error in Equation (5) can be re-expressed as

$$\xi_t = \sigma_\xi \cdot \xi_t^* = \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2} P_t^* + \frac{-\sigma_\xi \rho_{pw}\rho_{p\xi}}{1-\rho_{pw}^2} W_t^* + \sigma_\xi \sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2\rho_{p\xi}^2}{1-\rho_{pw}^2}}\,\omega_{3,t}. \tag{9}$$

In this way, the structural error term ξ_t is split into two parts: one part as a function of P_t^* and W_t^* that captures the endogeneity of P_t and the association of W_t with $\xi_t|P_t$[7], and the other part as an independent new error term. Then, we substitute Equation (9) into the main model in Equation (1), and obtain the following regression equation:

$$Y_t = \mu + P_t\alpha + W_t\beta + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2} P_t^* + \frac{-\sigma_\xi \rho_{pw}\rho_{p\xi}}{1-\rho_{pw}^2} W_t^* + \sigma_\xi \sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2\rho_{p\xi}^2}{1-\rho_{pw}^2}} \cdot \omega_{3,t}. \tag{10}$$

Given P_t^* and W_t^* as additional regressors, $\omega_{3,t}$ is not correlated with all regressors on the right-hand side of Equation (10) as proved in Theorem 2 below, and thus we can consistently estimate the model using the least squares estimator. The regressors P_t^* and W_t^* can be generated from the nonparametric distribution of P_t and W_t as $P_t^* = \Phi^{-1}(\widehat{H}(P_t))$ and $W_t^* = \Phi^{-1}(\widehat{L}(W_t))$, where $\widehat{H}(P_t)$ and $\widehat{L}(W_t)$ are the empirical CDFs of P_t and W_t, respectively.

Theorem 2. ***Estimation Consistency.*** *Assuming (1) the error term is normal, (2) the endogenous regressor P_t and exogenous regressors W_t are non-normally distributed, and (3) a Gaussian Copula for the error term, P_t and W_t, $Cov(\omega_{3,t}, W_t) = Cov(\omega_{3,t}, P_t) = Cov(\omega_{3,t}, W_t^*) = Cov(\omega_{3,t}, P_t^*) = 0$ and thus the OLS estimation of Equation (10) yields consistent estimates of*

[7] Although the exogenous regressor W_t and ξ_t are uncorrelated, W_t and $\xi_t|P_t$ (the error component in ξ_t remaining after removing the effect of the endogenous regressor P_t) can be correlated as seen by the correlation between W_t and ω_t in Figure 1 (b).

model parameters.

Proof: See Online Appendix A, Proof of Theorem 2.

As shown in Theorem 2, the proposed COPE method does not require the uncorrelatedness between P_t^* and W_t for consistent model estimation, an assumption needed for CopulaP&G. In fact, CopulaP&G can be obtained as a special case of the COPE: when W_t is uncorrelated with P_t (i.e. $\rho_{pw} = 0$) and also uncorrelated with P_t^* under the joint copula model, $\frac{-\sigma_\xi \rho_{pw} \rho_{p\xi}}{1-\rho_{pw}^2} W_t^*$ in Equation (10) vanishes and COPE based on Equation (10) reduces to CopulaP&G base on Equation (6). This broader applicability of COPE is a merit of COPE. However, similar to CopulaP&G, COPE requires the normality of the endogenous regressor P_t to fulfill the full-rank identification assumption. Moreover, a correlation between endogenous regressor P and the exogenous regressors W will cause CopulaP&G to transfer the endogeneity from P to W; the correction for the induced endogenous regressor W should have the same non-normality assumption for model identification as with P. In the next subsection, we will develop a novel two-stage COPE method that relaxes the regressor normality assumption. We further extend the model to incorporate multiple endogenous variables in the following subsection 3.4.

3.3 *Proposed Method II: Two-stage Copula Endogeneity-correction (2sCOPE)*

In this subsection, we further propose a two-stage COPE (2sCOPE) method and will show that this method can relax both the uncorrelatedness assumption between the copula-transformed endogenous regressor and the exogenous regressors (Assumption 4) and the key identification assumption of normality on the endogenous regressors (Assumption 3).

Under the Gaussian copula assumption in Equation (7), we have a linear relationship between P_t^* and W_t^*, and we take advantage of this information to construct the two-stage COPE estimation. We have the following system of equations that are similar to two-stage least square method. However, we do not require any variable that satisfies the exclusion restriction.

$$Y_t = \mu + P_t \alpha + W_t \beta + \xi_t \tag{11}$$

$$P_t^* = W_t^* \gamma + \epsilon_t, \tag{12}$$

where P_t^*, W_t^* are the Gaussian copula transformations of P_t and W_t, respectively. ϵ_t and ξ_t follow a bivariate joint normal distribution, since they are a linear combination of tri-normal variate (ξ_t^*, P_t^*, W_t^*) under the Gaussian copula assumption. ϵ_t and ξ_t are correlated because of the endogeneity of P_t. We can derive the relationship between P_t^* and W_t^*, and between ϵ_t and

ξ_t directly from Equation (7). Specifically we have

$$P_t^* \quad = \quad \rho_{pw} W_t^* + \sqrt{(1-\rho_{pw}^2)} \cdot \omega_{2,t} = \rho_{pw} W_t^* + \epsilon_t, \tag{13}$$

which shows γ in Equation (12) is ρ_{pw}. Then we substitute ϵ_t in Equation (13) into the outcome regression in Equation (10) and rewrite the model below

$$\begin{aligned} Y_t &= \mu + P_t\alpha + W_t\beta + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2}(P_t^* - \rho_{pw} W_t^*) + \sigma_\xi \sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2 \rho_{p\xi}^2}{1-\rho_{pw}^2}} \cdot \omega_{3,t}, \\ &= \mu + P_t\alpha + W_t\beta + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2}\epsilon_t + \sigma_\xi \sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2 \rho_{p\xi}^2}{1-\rho_{pw}^2}} \cdot \omega_{3,t}. \end{aligned} \tag{14}$$

Equation (14) suggests adding the estimate of the error term ϵ_t from the first stage regression as a generated regressor to the outcome regression instead of using P_t^* and W_t^*. As shown in Theorem 3, the new error term $\omega_{3,t}$ is uncorrelated with all the regressors in Equation (14), ensuring the consistency of model estimates. This two-step procedure, named as 2sCOPE, adds the first-stage residual term $\widehat{\epsilon}_t$ to control for endogeneity and in this aspect is similar to the control function approach of Petrin and Train (2010). However, unlike Petrin and Train (2010), 2sCOPE requires no use of instrumental variables.

Theorem 3. ***Estimation Consistency.*** *Assuming (1) the error term is normal, (2) the endogenous variable P_t or correlated regressors W_t is nonnormal, and (3) a Gaussian Copula for the error term, P_t and W_t, $Cov(\omega_{3,t}, W_t) = Cov(\omega_{3,t}, P_t) = Cov(\omega_{3,t}, \epsilon_t) = 0$ in Equation (14).*

Proof: See Online Appendix A, Proof of Theorem 3.

According to Theorems 2 and 3, both the proposed COPE and 2sCOPE can yield consistent estimates when assumptions are met. However, compared with the COPE method above, the 2sCOPE method uses additional information, the correlation between P_t^* and W_t^*, for model identification. Thus, intuitively, we expect 2sCOPE to have greater estimation efficiency. Theorem 4 shows that the estimates of 2sCOPE indeed are more precise, with smaller standard errors than those of COPE.

Theorem 4. ***Variance Reduction.*** *Assuming (1) the error term is normal, (2) the endogenous variable P_t and correlated regressors W_t are nonnormal, and (3) a Gaussian Copula for the error term, P_t and W_t, $\mathbf{Var}(\widehat{\theta}_\mathbf{2}) \leq \mathbf{Var}(\widehat{\theta}_\mathbf{1})$, where $\widehat{\theta}_1$ and $\widehat{\theta}_2$ denote parameter estimates from COPE and 2sCOPE, respectively.*

Proof: See Online Appendix A, Proof of Theorem 4.

Theorem 4 shows that under the assumptions when both COPE and 2sCOPE yield consistent estimates, 2sCOPE further reduces the variance of the estimates and improves estimation efficiency. Besides estimation efficiency, we show in Theorem 5 below that 2sCOPE also outperforms COPE in dealing with normal endogenous regressors.

Theorem 5. ***Nonnormality Assumption Relaxed.*** *Assuming (1) the error term is normal, (2) one of the exogenous regressors W_t is nonnormal, and (3) a Gaussian Copula for the error term, P_t and W_t, 2sCOPE estimator $\widehat{\theta}_{\mathbf{2}}$ is consistent when P_t follows a normal distribution while the COPE estimator $\widehat{\theta}_{\mathbf{1}}$ is not consistent.*

Proof: See Online Appendix A, Proof of Theorem 5.

Theorem 5 shows that as long as one of the exogenous regressors that are correlated with the endogenous regressor P_t is nonnormally distributed, 2sCOPE can correct for endogeneity for normal P_t while COPE cannot. Intuitively, when one of P_t and W_t is normal, P_t^* (or W_t^*) in Equation (10) becomes a linear function of P_t (or W_t) under the Gaussian copula assumption, rendering the COPE model to fail the full rank assumption and become unidentified. Thus, our first proposed method COPE cannot deal with normal endogenous regressors. For the proposed 2sCOPE method in Equation (14), adding the first stage helps model identification with extra information, the correlation between P_t^* and W_t^*. As long as not all W_t are normal, ϵ_t would not be a linear function of P_t and W_t and thus would satisfy the full rank assumption for model identification. Thus, our proposed method 2sCOPE can relax the nonnormality assumption on the endogenous regressor required in Park and Gupta (2012) as long as one of W_t is nonnormally distributed.

To sum up, we have proved the consistency of both COPE and 2sCOPE methods (Theorems 2, 3). Theorem 4 and 5 further show that the 2sCOPE method outperforms the COPE method in terms of estimation efficiency gain and relaxing the nonnormality assumption on the endogenous regressors required in CopulaP&G by satisfying a very loose condition.

3.4 *Multiple Endogenous Regressors*

In the above three subsections, we focus on the case of one endogenous regressor, study the explicit and implicit assumptions of the existing copula correction method CopulaP&G, and propose two new COPE procedures to relax some key assumptions of CopulaP&G. In this subsection, we extend these results to the general case of multiple endogenous regressors. Consider the following structural linear regression model with two endogenous regressors ($P_{1,t}$

and $P_{2,t}$) that are potentially correlated with the exogenous regressor W_t:

$$Y_t = \mu + P_{1,t} \cdot \alpha_1 + P_{2,t} \cdot \alpha_2 + W_t\beta + \xi_t. \tag{15}$$

Assumptions in CopulaP&G We first examine CopulaP&G for multiple endogenous regressors. According to Park and Gupta (2012), under a Gaussian copula model for $(P_{1,t}, P_{2,t}, \xi_t)$, the structural regression model in Equation (15) can be re-expressed as

$$\begin{aligned} Y_t =& \mu + P_{1,t}\alpha_1 + P_{2,t}\alpha_2 + W_t\beta + \sigma_\xi \frac{\rho_{\xi 1} - \rho_{12}\rho_{\xi 2}}{1-\rho_{12}^2} \cdot P^*_{1,t} + \sigma_\xi \frac{\rho_{\xi 2} - \rho_{12}\rho_{\xi 1}}{1-\rho_{12}^2} \cdot P^*_{2,t} \\ &+ \sigma_\xi \cdot \sqrt{1 - \rho_{\xi 1}^2 - \frac{(\rho_{\xi 2} - \rho_{12}\rho_{\xi 1})^2}{1-\rho_{12}^2}} \cdot \omega_t. \end{aligned} \tag{16}$$

where $P^*_{1,t} = \Phi^{-1}(H_1(P_{1,t}))$, $P^*_{2,t} = \Phi^{-1}(H_2(P_{2,t}))$, and $H_1(\cdot)$ and $H_2(\cdot)$ are CDFs of $P_{1,t}$ and $P_{1,t}$, respectively, ρ_{12} is the correlation between $P^*_{1,t}$ and $P^*_{2,t}$, $\rho_{\xi 1}$ is the correlation between ξ and $P^*_{1,t}$, $\rho_{\xi 2}$ is the correlation between ξ and $P^*_{2,t}$, and ω_t is a standard normal random variable that is independent of $P^*_{1,t}$ and $P^*_{2,t}$. Park and Gupta (2012) suggest the OLS estimation of Equation (16) yields consistent estimates of model parameters. Their method simply adds $P^*_{1,t}$ and $P^*_{2,t}$ as two "generated regressors" into the original structural regression model to control for endogeneity bias. The derivation provided in Park and Gupta (2012) makes explicit the following three assumptions: the error term ξ_t follows a normal distribution, ($P_{1,t}$ $P_{2,t}$, ξ_t) follow Gaussian copula, and all endogenous regressors $(P_{1,t}, P_{2,t})$ are nonnormally distributed (similar to Assumptions 1, 2, 3 noted in Subsection 3.1 for one endogenous regressor). However, Assumptions 1, 2, 3 are insufficient to guarantee the consistency of CopulaP&G. For the OLS estimation of Equation (16) to yield consistent estimates, W_t need also be uncorrelated with ω_t, which requires that $Cov(W_t, \frac{\rho_{\xi 1}-\rho_{12}\rho_{\xi 2}}{1-\rho_{12}^2} \cdot P^*_{1,t} + \frac{\rho_{\xi 2}-\rho_{12}\rho_{\xi 1}}{1-\rho_{12}^2} \cdot P^*_{2,t}) = 0$ (Assumption 4(b) below) where $\frac{\rho_{\xi 1}-\rho_{12}\rho_{\xi 2}}{1-\rho_{12}^2} \cdot P^*_{1,t} + \frac{\rho_{\xi 2}-\rho_{12}\rho_{\xi 1}}{1-\rho_{12}^2} \cdot P^*_{2,t}$ is the CCF used to control for endogeneity in CopulaP&G.

Assumption 4(b). *When there are multiple endogenous regressors, W_t is uncorrelated with the CCF, i.e., the linear combination of P^*_t that is used to control for endogenous regressors. Specifically, $Cov(W_t, \frac{\rho_{\xi 1}-\rho_{12}\rho_{\xi 2}}{1-\rho_{12}^2} \cdot P^*_{1,t} + \frac{\rho_{\xi 2}-\rho_{12}\rho_{\xi 1}}{1-\rho_{12}^2} \cdot P^*_{2,t}) = 0$ is required in the 2-endogenous regressors case.*[8]

Assumption 4 and 4(b) are verifiable and provide users with criteria to check whether CopulaP&G would fail to work when there exist exogenous regressors that may be correlated with the CCF. With only one endogenous regressor, one can simply check the correlations between

[8]It is clear that this requirement is not the same as either $Cov(W_t, P^*_{1,t}) = 0, Cov(W_t, P^*_{2,t}) = 0$ or $Cov(W_t, P_{1,t}) = 0, Cov(W_t, P_{2,t}) = 0$.

the copula transformation of this endogenous regressor with each exogenous regressor. For multiple endogenous regressors, one should check the correlations between the CCF (i.e., the linear combination of copula transformations of these endogenous regressors used to control for endogeneity) in CopulaP&G with each exogenous regressor. If there exists one exogenous regressor in W_t that fails the Assumption 4 or 4(b), CopulaP&G yields biased estimates, and our proposed COPE or 2sCOPE method should be used, which are derived below.

COPE Under the Gaussian Copula assumption that $[P^*_{1,t}, P^*_{2,t}, W^*_t, \xi^*_t]$ follows a multivariate normal distribution:

$$\begin{pmatrix} P^*_{1,t} \\ P^*_{2,t} \\ W^*_t \\ \xi^*_t \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_p & \rho_{wp1} & \rho_{\xi p1} \\ \rho_p & 1 & \rho_{wp2} & \rho_{\xi p2} \\ \rho_{wp1} & \rho_{wp2} & 1 & 0 \\ \rho_{\xi p1} & \rho_{\xi p2} & 0 & 1 \end{bmatrix}\right),$$

we have:

$$\begin{pmatrix} P^*_{1,t} \\ P^*_{2,t} \\ W^*_t \\ \xi^*_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \rho_p & \sqrt{1-\rho_p^2} & 0 & 0 \\ \rho_{wp1} & \frac{\rho_{wp2}-\rho_p\rho wp1}{\sqrt{1-\rho_p^2}} & \sqrt{1-\rho_{wp1}^2 - \frac{(\rho_{wp2}-\rho_p\rho_{wp1})^2}{1-\rho_p^2}} & 0 \\ \rho_{\xi p1} & \frac{\rho_{\xi p2}-\rho_p\rho_{\xi p1}}{\sqrt{1-\rho_p^2}} & \frac{-\rho_{wp1}\rho_{\xi p1} - \frac{(\rho_{wp2}-\rho_p\rho_{wp1})(\rho_{\xi p2}-\rho_p\rho_{\xi p1})}{1-\rho_p^2}}{\sqrt{1-\rho_{wp1}^2 - \frac{(\rho_{wp2}-\rho_p\rho_{wp1})^2}{1-\rho_p^2}}} & m \end{pmatrix} \cdot \begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \\ \omega_{3,t} \\ \omega_{4,t} \end{pmatrix},$$

$$\begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \\ \omega_{3,t} \\ \omega_{4,t} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\right), \tag{17}$$

where m is a function of all the ρs. Under the Gaussian Copula assumption above, we can derive ξ^*_t as a function of P_t and W_t. After simplification, the structural error in Equation (15) can be decomposed as

$$\xi_t = \sigma_\xi \xi^*_t = \eta_1 P^*_{1,t} + \eta_2 P^*_{2,t} - (\eta_1\rho_{wp1} + \eta_2\rho_{wp2})W^*_t + \sigma_\xi \cdot m \cdot \omega_{4,t}. \tag{18}$$

where

$$\eta_1 = \frac{\sigma_\xi\rho_{\xi p1}(1-\rho_{wp2}^2) - \sigma_\xi\rho_{\xi p2}(\rho_p - \rho_{wp1}\rho_{wp2})}{1-\rho_p^2-\rho_{wp1}^2-\rho_{wp2}^2+2\rho_p\rho_{wp1}\rho_{wp2}}, \quad \eta_2 = \frac{\sigma_\xi(\rho_{wp1}\rho_{wp2}\rho_{\xi p1} + \rho_{\xi p2} - \rho_p\rho_{\xi p1} - \rho_{wp1}^2\rho_{\xi p2})}{1-\rho_p^2-\rho_{wp1}^2-\rho_{wp2}^2+2\rho_p\rho_{wp1}\rho_{wp2}}.$$

The COPE method with one endogenous regressor in Equation (10) is then extended to

$$Y_t = \mu + P_{1,t}\alpha_1 + P_{2,t}\alpha_2 + W_t\beta + \eta_1 P^*_{1,t} + \eta_2 P^*_{2,t} - (\eta_1\rho_{wp1} + \eta_2\rho_{wp2})W^*_t + \sigma_\xi \cdot m \cdot \omega_{4,t}. \tag{19}$$

In Equation (19), the new error term $\omega_{4,t}$ is uncorrelated with all the regressors on the right-hand side of Equation (19). Thus, the OLS estimation of Equation (19) provides consistent estimates of structural regression model parameters $(\mu, \alpha_1, \alpha_2, \beta)$.

2sCOPE Under the multivariate Gaussian distribution assumption on $(\xi_t, P^*_{1,t}, P^*_{2,t}, W^*_t)$, the system equations of 2sCOPE method in Equation (11, 12) are easily extended to the case with two endogenous regressors as

$$Y_t = \mu + P_{1,t}\alpha_1 + P_{2,t}\alpha_2 + W_t\beta + \xi_t, \tag{20}$$

$$P^*_{1,t} = \rho_{wp1}W^*_t + \epsilon_{1,t}, \tag{21}$$

$$P^*_{2,t} = \rho_{wp2}W^*_t + \epsilon_{2,t}, \tag{22}$$

where Equations (21) and (22) can be directly derived from the Gaussian copula assumption; $(\xi_t, \epsilon_{1,t}.\epsilon_{2,t})$ are a linear transformation of $(\xi_t, P^*_{1,t}, P^*_{2,t}, W^*_t)$, and thus also follow a multivariate Gaussian distribution. As a result, we can decompose the structural error ξ_t as additive terms for $\epsilon_{1,t}$, $\epsilon_{2,t}$ and a remaining independent error term $\omega_{4,t}$ as follows

$$Y_t \quad = \mu + P_{1,t}\alpha_1 + P_{2,t}\alpha_2 + W_t\beta + \eta_1\epsilon_{1,t} + \eta_2\epsilon_{2,t} + \sigma_\xi \cdot m \cdot \omega_{4,t}. \tag{23}$$

Note that Equation (23) can also be obtained from Equation (19) by noting that $\epsilon_{1,t} = P^*_{1,t} - \rho_{wp1}W^*_t$ and $\epsilon_{2,t} = P^*_{2,t} - \rho_{wp2}W^*_t$, and thus OLS estimation of Equation (23) also yields consistent estimation of structural model parameters. However, the 2sCOPE procedure adds only two residual terms $(\epsilon_{1,t}, \epsilon_{2,t})$ as generated regressors instead of three copula transformations of regressors $(P^*_{1,t}, P^*_{2,t}, W^*_t)$ as generated regressors. Thus, 2sCOPE adds a smaller number of generated regressors than COPE, and thus provides higher estimation efficiency. In addition, by adding residual terms as the generated regressors, 2sCOPE relaxes the assumption of regressor non-normality required in COPE. The proof for the estimation consistency of COPE and 2sCOPE, estimation efficiency gain and relaxation of the regressor-nonnormality assumption for 2sCOPE can be found in Appendix under the related Theorems 2, 3, 4, 5.

Until now, we have shown the derivation of our proposed COPE and 2sCOPE methods, the difference of these two proposed methods from the existing copula method CopulaP&G, and the assumptions for each method to work. Table 2 summarizes the assumptions for the three methods. Our proposed methods can deal with the case when there are exogenous regressors that are correlated with the endogenous regressors. Moreover, 2sCOPE can further relax the regressor-nonnormality assumption. Table 3 summarizes the estimation procedure of the two methods.

Table 2: Summary of Assumptions for the Three Methods

CopulaP&G	COPE	2sCOPE
• The structural error follows a normal distribution (Asm. 1); • P_t and the structural error follow a Gaussian copula (Asm. 2); • All regressors in P_t are nonnormally distributed (Asm. 3); • W_t is uncorrelated with the CCF (copula control function which is the linear combination of all P_t^* used to control for endogeneity) (Asm. 4, 4(b)).	• The structural error follows a normal distribution; • P_t, W_t and the structural error follow a Gaussian copula; • All regressors in P_t and W_t are nonnormally distributed.	• The structural error follows a normal distribution; • P_t, W_t and the structural error follow a Gaussian copula; • P_t can be normally distributed as long as one of W_t is nonnormal.

Table 3: Estimation Procedure

COPE	2sCOPE
	Stage 1:
• Obtain empirical CDFs for each regressor in P_t and W_t, denoted as $\widehat{H}(P_t)$ and $\widehat{L}(W_t)$; • Compute $P_t^* = \Phi^{-1}(\widehat{H}(P_t))$ and $W_t^* = \Phi^{-1}(\widehat{L}(W_t))$; • Add P_t^* and W_t^* to the outcome structural regression model as generated regressors.	• Obtain empirical CDFs for each regressor in P_t and W_t, $\widehat{H}(P_t)$ and $\widehat{L}(W_t)$; • Compute $P_t^* = \Phi^{-1}(\widehat{H}(P_t))$ and $W_t^* = \Phi^{-1}(\widehat{L}(W_t))$; • Regress each endogenous regressor in P_t^* separately on W_t^* and obtain residual $\widehat{\epsilon}_t$;
	Stage 2:
	• Add $\widehat{\epsilon}_t$ to the outcome structural regression model as generated regressors.
• Standard errors of parameter estiamtes are estimated using bootstrap in both methods.	

3.5 *COPEs for Random Coefficient Linear Panel Model with Correlated and Normally Distributed Regressors*

We consider the following random coefficient model for linear panel data

$$Y_{it} = \bar{\mu} + \mu_i + P_{it}'\alpha_i + W_{it}'\beta_i + \xi_{it}, \tag{24}$$

where $i = 1, \cdots, N$ indexes cross-sectional units and $t = 1, \cdots, T$ indexes occasions. P_{it} (W_{it}) denotes a vector of endogenous (exogenous) regressors. P_{it} and W_{it} can be correlated. The error term $\xi_{it} \overset{iid}{\sim} N(0, \sigma_\xi^2)$, which is correlated with P_{it} due to the endogeneity of P_{it} but is uncorrelated with the exogenous regressors in W_{it}. The individual-specific intercept μ_i and individual-specific slope coefficients (α_i, β_i) permit heterogeneity in both intercepts and regressor effects across cross-sectional units. Extant marketing studies have shown the ubiquitous presence of heterogeneous consumers' responses to marketing mix variables (e.g., price sensitivity) and substantial bias associated with ignoring such heterogeneity in slope coefficients. Thus, it is important to permit individual-specific slope coefficients, especially in marketing studies.

The linear panel data model as specified in Equation (24) is general and includes the linear panel model with only individual-specific intercepts considered in Haschka (2021) as a special case. Specifically, Haschka (2021) fixes (α_i, β_i) to be the same value (α, β) across all units, assuming all cross-sectional units have the same slope coefficients. In contrast, the model in Equation (24) relaxes this strong assumption and can generate unit-specific slope parameters, which can be used for targeting purposes.

A fully random coefficient model typically assumes $(\mu_i, \alpha_i, \beta_i)$ follows a multivariate normal distribution. Estimation algorithms for such random coefficient models are well-established when all regressors are exogenous. Alternatively, one can assume a mixed-effect model where μ_i is a fix-effect parameter with μ_i's allowed to be correlated with the regressors P_{it} and W_{it}. To avoid potential incidental parameter problem associated with these fix-effect parameters, one often uses the first-difference or fixed-effects transformation to eliminate the incidental intercept parameters as follows

$$\tilde{y}_{it} = \tilde{P}_{it}'\alpha_i + \tilde{W}_{it}'\beta_i + \tilde{\xi}_{it}, \tag{25}$$

where $\tilde{y}_{it}$, $\tilde{P}_{it}$, $\tilde{W}_{it}$ and $\tilde{\xi}_{it}$ denote new variables obtained from the first-difference or fixed-effect transformation. Haschka (2021) considered a special case of Equation (25) by fixing (α_i, β_i) to be constants. As shown above, the fixed-effect approach eliminates the intercepts from the model and can make the intercept parameter in-estimable (Web Appendix A8 in Haschka (2021)).

It is straightforward to apply COPEs to address regressor endogeneity in the general random coefficient model for linear panel data in Equation (24) and the transformed one without intercepts in Equation (25).[9] Assuming $(P_{it}, W_{it}, \xi_{it})$ follow a Gaussian copula, COPE adds the generated regressor $P_{it}^* = \Phi^{-1}(\widehat{H}(P_{it}))$ and $W_{it}^* = \Phi^{-1}(\widehat{L}(W_{it}))$ into Equation (24) to control for regressor endogeneity. The 2sCOPE procedure adds the residuals obtained from regressing P_{it}^* on W_{it}^*. Both COPEs methods can be implemented using standard methods for random coefficient linear panel models. By contrast, maximum likelihood approach for copula correction for the random coefficients model would require constructing complicated joint likelihood on the error term, P_t and W_t, which involves newly appearing numerical integrals with respect to random effects and cannot be maximized by standard estimation algorithms for random coefficient models. Finally, current applications applying CopulaP&G do not consider the role of exogenous regressors. Our analysis shows that this may yield bias if any exogenous regressor is correlated with the CCF added to control endogeneity, for which COPEs should be used to address regressor endogeneity.

3.6 *COPEs for Slope Endogeneity and Random Coefficient Logit Model*

In Online Appendix B and C, we derive the COPEs methods to tackle the slope endogeneity problem and address endogeneity bias in random coefficient logit models with correlated and normally distributed regressors. In these two cases, we show how to apply COPEs to correct for the endogenous bias, which can avoid the potential bias of CopulaP&G due to the potential correlations between the exogenous regressors and CCF, as well as make use of the correlated exogenous regressors to relax the non-normality assumption of endogenous regressors, improve model identification and sharpen model estimates. As shown there, both COPE and 2sCOPE can be implemented using standard estimation methods by adding generated regressors to control for endogenous regressors. By contrast, the maximum likelihood approach can require constructing complicated joint likelihood that is not what the standard estimation method uses and thus requires separate development and significantly more computation involving numerical integration.

[9]Similar to Haschka (2021), a GLS transformation can be applied to both sides of Equation (25), resulting in a pooled regression for which COPEs can be directly applied.

4. Simulation Study

In this section, we conduct Monte Carlo simulation studies for the following goals: (a) to assess the performance of the proposed methods for correlated regressors, (b) to assess the performance of the proposed methods under regressor normality, (c) to assess generalizability and restrictions of the distributional assumptions about the endogenous and exogenous regressors, and (d) to compare the performance of the proposed methods with existing methods. Following Park and Gupta (2012), we measure the estimation bias using t_{bias} calculated as the ratio of the absolute difference between the mean of the sampling distribution and the true parameter value to the standard error of the parameter estimate. As defined above, t_{bias} represents the size of bias relative to the sampling error. Online Appendix D provides additional simulation results on the robustness of COPEs to the mis-specifications of the structural error distribution and the copula dependence structure.

4.1 *Case 1: Non-normal Regressors*

We first examine the case when P and W are correlated. The specific data-generating process (DGP) is summarized below:

$$\begin{pmatrix} P_t^* \\ W_t^* \\ \xi_t^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{pw} & \rho_{p\xi} \\ \rho_{pw} & 1 & 0 \\ \rho_{p\xi} & 0 & 1 \end{bmatrix} \right) = N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix} \right), \tag{26}$$

$$\xi_t = G^{-1}(U_{\xi,t}) = G^{-1}(\Phi(\xi_t^*)) = \Phi^{-1}(\Phi(\xi^*)) = 1 \cdot \xi_t^*, \tag{27}$$

$$P_t = H^{-1}(U_{P,t}) = H^{-1}(\Phi(P_t^*)), \quad W_t = L^{-1}(U_{W,t}) = L^{-1}(\Phi(W_t^*)), \tag{28}$$

$$Y_t = \mu + \alpha \cdot P_t + \beta \cdot W_t + \xi_t = 1 + 1 \cdot P_t + (-1) \cdot W_t + \xi_t. \tag{29}$$

where ξ_t^* and P_t^* are correlated with the correlation coefficient $\rho_{p\xi} = 0.5$, and thus ξ_t and P_t are correlated, generating the endogeneity problem. W_t^* is exogenous and is not correlated with ξ_t^*. But W_t^* and P_t^* are correlated with the correlation coefficient $\rho_{pw} = 0.5$, and thus W_t and P_t are correlated. We consider four different estimation methods: (i) OLS, (ii) CopulaP&G in the form of Equation (6), (iii) the proposed method COPE in the form of Equation (10), and the proposed method 2sCOPE in the form of Equation (14). We set the sample size T = 1000, and generate 1000 data sets as replicates using the DGP above. In the simulation, we use the gamma distribution $Gamma(1, 1)$ with shape and rate equal to 1 for P_t and the exponential distribution $Exp(1)$ with rate 1 for W_t. Models are estimated on all generated data sets, providing the empirical distributions of the parameter estimates.

Table 4: Results of the Simulation Study Case 1

			OLS			CopulaP&G			COPE			2sCOPE		
ρ_{pw}	Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
0.5	μ	1	0.689	0.045	6.964	1.231	0.081	2.849	1.012	0.093	0.129	1.009	0.059	0.157
	α	1	1.571	0.036	15.75	1.055	0.069	0.791	0.985	0.072	0.213	0.986	0.070	0.197
	β	-1	-1.259	0.031	8.236	-1.289	0.031	9.169	-0.997	0.067	0.038	-0.995	0.042	0.123
	$\rho_{p\xi}$	0.5	-	-	-	0.570	0.047	1.504	0.505	0.055	0.090	0.504	0.038	0.097
	σ_ξ	1	0.862	0.020	6.902	1.011	0.043	0.244	1.008	0.041	0.206	1.006	0.040	0.143
	D-error			-			-			0.002613			0.001614	
0.7	μ	1	0.730	0.041	6.629	1.307	0.076	4.037	1.011	0.085	0.124	1.005	0.053	0.088
	α	1	1.800	0.041	19.67	1.260	0.068	3.838	0.988	0.078	0.148	0.991	0.075	0.118
	β	-1	-1.529	0.037	14.21	-1.567	0.037	15.36	-0.997	0.071	0.041	-0.994	0.056	0.110
	$\rho_{p\xi}$	0.5	-	-	-	0.633	0.043	3.130	0.503	0.057	0.048	0.500	0.026	0.000
	σ_ξ	1	0.799	0.018	11.18	0.980	0.044	0.468	1.007	0.041	0.160	1.003	0.040	0.084
	D-error			-			-			0.002902			0.001760	

Note: Mean and SE denote the average and standard deviation of parameter estimates over all the 1,000 simulated samples.

Table 4 reports estimation results. As expected, OLS estimates of both α and β are biased (t_{bias} = 15.75/8.24) as a result of the regressor endogeneity. The estimation result of CopulaP&G reduces the bias, but still shows significant bias for both the coefficient estimates of P_t and W_t. The bias of CopulaP&G depends on the strength of the correlation between W and P. Stronger correlations between P^* and W^* can cause a larger bias of CopulaP&G estimates. For example, when the correlation between W^* and P^* increases from 0.5 to 0.7, the bias of estimated α increases by around five times (from 0.055 to 0.260 in Table 4 under the column "CopulaP&G"). The bias confirms our derivation in the model section, demonstrating that using the existing copula method may not solve the endogeneity problem completely with correlated regressors.

We next examine our proposed methods. Both methods (COPE and 2sCOPE) provide consistent estimates without the use of instruments. The average estimates of $\rho_{p\xi}$ is close to the true value 0.5 and is significantly different from 0, implying significant correlation between the endogeneity regressor and the error term. Moreover, the proposed method 2sCOPE shows larger efficiency. The standard error of $\alpha(\beta)$ in 2sCOPE is 0.070 (0.042), which is 2.78% (37.31%) smaller than the corresponding standard errors using COPE. We further calculate the estimation precision of COPE and 2sCOPE using the D-error measure $|\Sigma|^{1/K}$ (Arora and Huber 2001, Qian and Xie 2021), where Σ is the covariance matrix of the parameter estimates in the regression mean function, and K is the number of these parameters. A smaller value of D-error means greater estimation efficiency and improved estimation precision. When $\rho_{pw} = 0.5$, the D-error measure is 0.002613 for COPE and is 0.001614 for 2sCOPE (Table 4), and thus 2sCOPE increases estimation precision by 38.2%, meaning that for 2sCOPE to achieve the same precision with COPE, sample size can be reduced by 38.2%. A 39.3% of efficiency gain for 2sCOPE is found for $\rho_{pw} = 0.7$ in Table 4.

We perform a further simulation study for a small sample size. Specifically, we use the same DGP as described above to generate synthetic data, except with the sample size T=200. Online Appendix D Table 11 reports the results and shows that OLS estimates have endogeneity bias and CopulaP&G reduces the endogeneity bias but significant bias remains. Both our proposed methods, COPE and 2sCOPE, perform well and have unbiased estimates for the small sample size T=200. The efficiency gain of 2sCOPE relative to COPE appears to be greater when sample size becomes smaller. When the correlation between P^* and W^* is 0.5, the D-error measures are 0.0166 and 0.0091 for COPE and 2sCOPE (Online Appendix Table 11), respectively, meaning that 2sCOPE increases estimation precision by 1-0.0091/.0166=46% compared with COPE, and thus sample size can be reduced by almost a half (∼50%) for 2sCOPE to achieve the same estimation precision as that achieved by COPE. A similar magnitude of efficiency gain for 2sCOPE relative to COPE (∼50%) is observed when the correlation between P^* and W^* is 0.7 (Online Appendix Table 11).

4.2 *Case 2: Normal Regressors*

Next, we examine the case when the endogenous regressor and (or) the correlated exogenous regressor are normally distributed. We pay special attention to this case because normality is not allowed for endogenous regressors in Park and Gupta (2012). We use the Gaussian copula as described in Equations (26) to Equations (29) for DGP to generate the data, except that the marginal CDFs for regressors ($H(\cdot)$ and $L(\cdot)$) are chosen according to the distributions listed in the first two columns in Table 5.

Table 5 summarizes the estimation results. As expected, OLS estimates are biased. CopulaP&G produces biased estimates whenever the endogenous regressor P follows a normal distribution. The estimates of CopulaP&G are biased when P follows a gamma distribution (first row of Table 5) for a different reason: P and W are correlated. Similar to CopulaP&G, the results of proposed COPE are biased in all the three scenarios when at least one of P_t and W_t is normal. When W_t is normal, β is 0.323 away from the true value -1; when P_t is normally distributed, α is 0.684 away from the true value; when both P_t and W_t are normal, α is 0.663 away from the true value 1 and β is 0.324 away from the true value -1. Thus, similar to CopulaP&G, COPE also requires the assumption of regressor non-normality for estimation consistency. This is expected because COPE adds P_t^* and W_t^*, the copula transformation of regressors, as additional regressors, and will cause perfect co-linearity and model non-identification problem whenever at least one of these regressors is normally distributed.

Table 5: Results of Case 2: Normal Regressors

Distribution				OLS			CopulaP&G			COPE			2sCOPE		
P	W	Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
Gamma	Normal	μ	1	0.431	0.045	12.63	1.018	0.078	0.227	1.017	0.080	0.217	1.015	0.077	0.190
		α	1	1.569	0.037	15.40	0.979	0.070	0.302	0.979	0.070	0.296	0.985	0.070	0.212
		β	-1	-1.259	0.030	8.619	-1.333	0.028	11.78	-1.323	0.433	0.746	-0.997	0.045	0.067
		$\rho_{p\xi}$	0.5	-	-	-	0.640	0.039	3.556	0.589	0.141	0.631	0.506	0.036	0.151
		σ_ξ	1	0.861	0.019	7.240	1.064	0.046	1.394	1.135	0.162	0.837	1.005	0.038	0.134
Normal	Exp	μ	1	1.286	0.042	6.777	1.286	0.045	6.374	0.994	0.073	0.081	1.023	0.070	0.334
		α	1	1.628	0.031	20.36	1.532	0.462	1.152	1.684	0.437	1.568	1.048	0.126	0.381
		β	-1	-1.286	0.032	8.956	-1.287	0.032	8.960	-0.992	0.066	0.127	-1.024	0.062	0.383
		$\rho_{p\xi}$	0.5	-	-	-	0.089	0.419	0.980	-0.167	0.384	1.738	0.465	0.074	0.473
		σ_ξ	1	0.829	0.018	9.492	0.940	0.151	0.394	0.981	0.151	0.129	0.980	0.063	0.318
Normal	Normal	μ	1	1.001	0.026	0.046	1.002	0.030	0.052	1.001	0.033	0.024	1.002	0.028	0.057
		α	1	1.668	0.030	22.38	1.663	0.450	1.474	1.663	0.460	1.441	1.655	0.395	1.657
		β	-1	-1.335	0.029	11.44	-1.335	0.029	11.42	-1.324	0.438	0.740	-1.328	0.197	1.668
		$\rho_{p\xi}$	0.5	-	-	-	0.006	0.412	1.198	0.001	0.412	2.426	0.010	0.303	1.616
		σ_ξ	1	0.816	0.019	9.687	0.917	0.155	0.534	1.003	0.211	0.016	0.879	0.092	1.317

By contrast, the proposed 2sCOPE method provides consistent estimates as long as P_t and W_t are not both normally distributed. Both α and β are tightly distributed near the true value whenever P_t or W_t is nonnormally distributed. Unlike CopulaP&G and COPE, 2sCOPE adds the residual term obtained from regressing P_t^* on W_t^* as the generated regressor. Thus, as long as P_t and W_t are not both normally distributed, the residual term is not perfectly co-linear with the original regressors, permitting model identification. Only when both P_t and W_t are normally distributed (the last scenario in Table 5), the residual term added into the structural regression model becomes a linear combination of P_t and W_t, causing perfect co-linearity and model non-identification. Overall, this simulation study demonstrates the advantage of the proposed 2sCOPE to relax the nonnormality assumption in CopulaP&G as long as one of P_t and W_t is nonnormally distributed.

4.3 *Case 3: Multiple Endogenous Regressors*

In this case, we examine the performance of our proposed methods when the model has multiple endogenous regressors. We use the data-generating process (DGP) with two endogenous regressors and one exogenous regressor that is correlated with the endogenous regressor below:

$$\begin{pmatrix} P^*_{1,t} \\ P^*_{2,t} \\ W^*_t \\ \xi^*_t \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_p & \rho_{wp1} & \rho_{\xi p1} \\ \rho_p & 1 & \rho_{wp2} & \rho_{\xi p2} \\ \rho_{wp1} & \rho_{wp2} & 1 & 0 \\ \rho_{\xi p1} & \rho_{\xi p2} & 0 & 1 \end{bmatrix}\right) = N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.3 & 0.4 & 0.5 \\ 0.3 & 1 & 0.4 & 0.5 \\ 0.4 & 0.4 & 1 & 0 \\ 0.5 & 0.5 & 0 & 1 \end{bmatrix}\right) \quad (30)$$

$$\xi_t = G^{-1}(U_{\xi,t}) = G^{-1}(\Phi(\xi^*_t)) = \Phi^{-1}(\Phi(\xi^*)) = 1 \cdot \xi^*_t, \quad (31)$$

$$P_{1,t} = H_1^{-1}(U_{p1}) = H_1^{-1}(\Phi(P^*_{1,t})),\ \ P_{2,t} = H_2^{-1}(\Phi(P^*_{2,t})),\ \ W_t = L^{-1}(U_{W,t}) = L^{-1}(\Phi(W^*_t)), \quad (32)$$

$$Y_t = \mu + \alpha \cdot P_t + \beta \cdot W_t + \xi_t = 1 + 1 \cdot P_{1,t} + 1 \cdot P_{2,t} + (-1) \cdot W_t + \xi_t, \quad (33)$$

where $H_1^{-1}(\cdot)$ ($H_2^{-1}(\cdot)$) and $L^{-1}(\cdot)$ are the inverse distribution functions of the gamma and exponential distributions used to generate these regressors. Sample size T = 1000. We generate 1000 data sets, and use existing methods and our proposed methods to estimate the model. Table 6 shows the estimation results. Both the OLS and CopulaP&G estimates are biased, while our proposed methods provide unbiased estimates for all parameters, indicating that our proposed methods perform well with multiple endogenous regressors.

4.4 *Case 4: Multiple Exogenous Control Covariates*

We investigate the performance of our proposed methods when there exist multiple exogenous regressors consisting of both continuous and discrete variables. We generate the data using

Table 6: Results of the Simulation Study Case 3: Multiple Endogenous Regressors

		OLS			CopulaP&G			COPE			2sCOPE		
Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
μ	1	0.419	0.045	13.02	1.267	0.090	2.949	1.012	0.097	0.125	1.008	0.069	0.120
α_1	1	1.450	0.029	15.46	1.040	0.060	0.665	0.990	0.060	0.166	0.991	0.059	0.153
α_2	1	1.450	0.031	14.72	1.040	0.059	0.673	0.990	0.058	0.177	0.991	0.056	0.167
β	-1	-1.320	0.029	11.04	-1.353	0.028	12.56	-0.997	0.057	0.061	-0.995	0.040	0.134
$\rho_{\xi p1}$	0.5	-	-	-	0.567	0.043	1.545	0.503	0.049	0.052	0.502	0.040	0.048
$\rho_{\xi p2}$	0.5	-	-	-	0.568	0.042	1.625	0.503	0.047	0.073	0.503	0.038	0.075
σ_ξ	1	0.772	0.018	12.58	1.019	0.048	0.402	1.012	0.044	0.283	1.010	0.042	0.233

the following DGP:

$$\begin{pmatrix} P_t^* \\ W_{1,t}^* \\ W_{2,t}^* \\ \xi_t^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{pw1} & \rho_{pw2} & \rho_{\xi p} \\ \rho_{pw1} & 1 & \rho_w & 0 \\ \rho_{pw2} & \rho_w & 1 & 0 \\ \rho_{\xi p} & 0 & 0 & 1 \end{bmatrix}\right) = N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.3 & 0 \\ 0.5 & 0.3 & 1 & 0 \\ 0.5 & 0 & 0 & 1 \end{bmatrix}\right) \tag{34}$$

$$\xi_t = G^{-1}(\Phi(\xi_t^*)) = \Phi^{-1}(\Phi(\xi^*)) = 1 \cdot \xi_t^*, \tag{35}$$

$$P_t = H^{-1}(\Phi(P_t^*)), \quad W_{1,t} = L^{-1}(\Phi(W_{1,t}^*)), \tag{36}$$

$$W_{2,t} = \begin{cases} 1, & \text{if} \quad \Phi(W_{2,t}^*) \geq 0.5 \\ 0. & \text{if} \quad \Phi(W_{2,t}^*) < 0.5 \end{cases}, \tag{37}$$

$$Y_t = \mu + \alpha \cdot P_t + \beta_1 \cdot W_{1,t} + \beta_2 \cdot W_{2,t} + \xi_t = 1 + 1 \cdot P_t + (-1) \cdot W_{1,t} + (-1) \cdot W_{2,t} + \xi_t, \tag{38}$$

where $H^{-1}(\cdot)$ and $L^{-1}(\cdot)$ are the inverse distribution functions of the gamma and exponential distributions. $W_{2,t}$ is a binary variable that follows a Bernoulli distribution. We set sample size T = 1000 and generate 1000 data sets to estimate parameters using OLS and copula methods. We follow the approach of Park and Gupta (2012) to generate latent copula data for discrete variables. Specifically, for a discrete regressor W_t, such as the binary exogenous regressor $W_{2,t}$, we define $U_{W,t}$, uniformly distributed on [0,1], as the CDF for a latent variable W_t^* that determines the discrete value of W_t. We then relate $U_{W,t}$ to W_t through the following inequality: $K(W_t - 1) < U_{W,t} < K(W_t)$, where $K(\cdot)$ is the CDF of W_t and can be directly estimated from the frequencies of the observed data. The above inequality implies the following relationship between $W_t^* = \Phi^{-1}(U_{W,t})$ and $K_{W,t}$: $\Phi^{-1}(K(W_t - 1)) < W_t^* < \Phi^{-1}(K(W_t))$.

The estimation results for the multiple-exogenous-regressor case with both discrete and continuous ones are summarized in Table 7. The OLS and CopulaP&G estimates are biased because of endogeneity and correlated exogenous regressors, respectively. The proposed COPEs methods perform well and provide consistent estimates for all parameters. This indicates that our proposed methods perform well with multiple exogenous correlated regressors. Moreover, correcting for endogeneity using our proposed methods does not require every exogenous correlated regressor to be informative (i.e., continuously distributed).

Table 7: Results of the Simulation Study Case 4: Multiple Exogenous Control Covariates

		OLS			CopulaP&G			COPE			2sCOPE		
Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
μ	1	0.701	0.046	6.452	1.281	0.083	3.394	1.007	0.115	0.057	1.005	0.061	0.085
α	1	1.573	0.038	15.10	1.037	0.071	0.532	0.985	0.073	0.208	0.987	0.072	0.180
β_1	-1	-1.225	0.041	5.523	-1.220	0.039	5.584	-0.990	0.069	0.140	-0.992	0.048	0.161
β_2	-1	-1.096	0.075	1.273	-1.202	0.073	2.758	-1.006	0.115	0.051	-1.003	0.080	0.042
$\rho_{p\xi}$	0.5	-	-	-	0.589	0.045	1.976	0.503	0.061	0.053	0.504	0.038	0.097
σ_ξ	1	0.862	0.020	7.066	1.023	0.044	0.532	1.011	0.040	0.264	1.006	0.040	0.115

4.5 *Case 5: Random Coefficient Linear Panel Model*

We investigate the performance of our proposed COPEs methods in random coefficient linear panel model. We use the copula and marginal distributions for $[P_{it}, W_{it}, \xi_{it}]$ as specified in Case 1 (Equations 26-28). We assign $\rho_{pw} = 0.7$ as an example. We then generate the outcome Y_{it} using the following standard random coefficient linear panel model:

$$Y_{it} = \bar{\mu} + \mu_i + P_{it}(\bar{\alpha} + a_i) + W_{it}(\bar{\beta} + b_i) + \xi_{it} = 1 + \mu_i + P_{it}(1 + a_i) + W_{it}(-1 + b_i) + \xi_{it},$$

where $[\mu_i, a_i, b_i] \sim N(0, I_3)$, $t = 1, ..., 50$ indexes occasions for repeated measurements, and $i = 1, ..., 500$ indexes the individual units. The above random coefficients model permits individual units to have heterogeneous baseline preferences (μ_i) and heterogeneous responses to regressors (a_i, b_i). Such random coefficients models are frequently used in marketing studies to capture individual heterogeneity and to profile and target individuals. The correlation between ξ_{it} and P_{it} creates the regressor endogeneity problem in the random coefficient model, which can cause biased estimates for standard linear random coefficient estimation methods ignoring the regressor-error correlation. We generate individual-level panel data as described above for

1000 times and use the data for estimation. Estimation results are in Table 8. LME is the standard estimation method for linear mixed models assuming all regressors are exogenous, as implemented in the R function $lme()$. LME and CopulaP&G are biased because of endogeneity and correlated exogenous regressors, respectively. Our proposed methods provide unbiased estimates that are tightly distributed around the true values for all parameters.

Table 8: Results of the Simulation Study Case 5: Random Coefficient Linear Panel Model

		LME			CopulaP&G			COPE			2sCOPE		
Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
$\bar{\mu}$	1	0.722	0.046	6.052	1.314	0.049	6.399	1.001	0.054	0.016	1.004	0.048	0.091
$\bar{\alpha}$	1	1.853	0.045	18.83	1.293	0.045	6.469	1.000	0.045	0.009	1.000	0.046	0.008
$\bar{\beta}$	-1	-1.557	0.045	12.39	-1.598	0.044	13.56	-0.996	0.048	0.079	-1.000	0.044	0.005
σ_μ	1	0.985	0.033	0.459	0.982	0.033	0.547	0.985	0.033	0.463	0.984	0.031	0.522
σ_α	1	0.988	0.036	0.326	0.987	0.034	0.397	0.986	0.035	0.403	0.989	0.035	0.316
σ_β	1	0.993	0.031	0.235	0.992	0.033	0.249	0.992	0.031	0.264	0.992	0.033	0.248
$\rho_{p\xi}$	0.5	-	-	-	0.646	0.009	16.33	0.509	0.012	0.757	0.507	0.005	1.365
σ_ξ	1	0.794	0.004	57.71	0.957	0.010	4.439	0.985	0.009	1.689	0.985	0.009	1.640

Note: $\sigma_\mu, \sigma_\alpha, \sigma_\beta$ are standard deviations of μ_i, a_i, b_i.

5. Empirical Application

In this section, we apply our methods to a real marketing application. We illustrate the proposed methods to address the price endogeneity issue using store-level sales data of toothpaste category in Chicago over 373 weeks from 1989 to 1997[10]. To control for product size, we select toothpaste with the most common size, which is 6.4 oz. Retail price is usually considered endogenous. The endogeneity of retail price can come from unmeasured product characteristics or demand shocks that can influence both consumers' and retailers' decisions. Since these variables are unobserved by researchers, they are absorbed into the structural error, leading to the endogeneity problem. Prices of different stores are correlated and often used as an IV for each other. This allows us to test the performance of the proposed COPEs methods in an empirical setting where a good IV exists. Besides the endogenous price, two promotion-related variables, bonus promotion and direct price reduction, would also affect demand. Following Park and Gupta (2012), we

[10] We obtained the data from https://www.chicagobooth.edu/research/kilts/datasets/dominicks.

treat the promotion variables as exogenous regressors. We focus on category sales in two large stores in Chicago (referred to as Stores 1 and 2). We convert retail price, in-store promotion and sales from UPC level to aggregate category level. They are computed as weekly market share-weighted averages of UPC-level variables. The correlation between log retail price and

Table 9: Summary Statistics

	Store 1				Store 2			
Variables	Mean	SD	Max	Min	Mean	SD	Max	Min
Sales (Unit)	115	52.8	720	35	165.7	93.7	1334	26
Price ($)	2.06	0.20	2.48	1.46	2.10	0.21	2.48	1.47
Bonus	0.18	0.20	0.80	0.00	0.16	0.19	0.79	0.00
PriceRedu	0.10	0.19	0.72	0.00	0.10	0.19	0.73	0.00

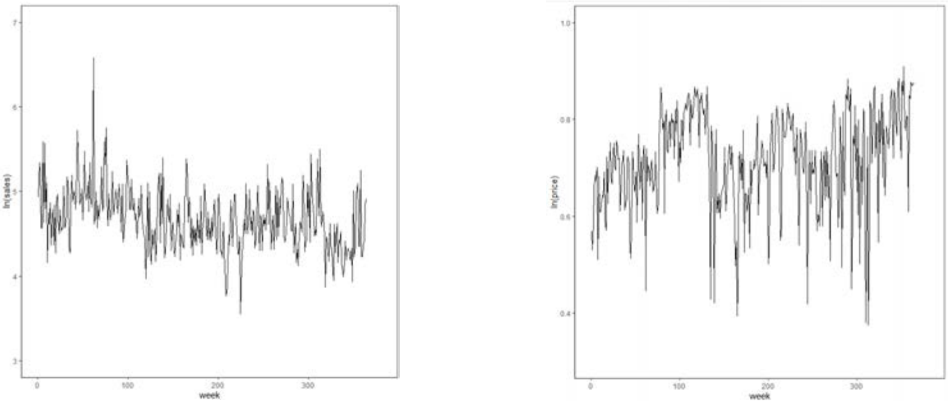
bonus promotion in Store 1 (Store 2) is -0.30 (-0.15), and the correlation between log retail price and price reduction promotion in Store 1 (Store 2) is -0.23 (-0.35). Both the correlations are significantly different from zero. The appreciable correlations between price and promotion variables actually provide a good setting for testing our methods and examining the impact that our proposed methods can make in the setting of correlated endogenous and exogenous regressors. Summary statistics of key variables are summarized in Table 9.

We estimate the following linear regression model:

$$\log(\text{Sales}_t) = \beta_0 + \log(\text{Retail Price}_t) \cdot \beta_1 + W_t' \beta_2 + \xi_t,$$

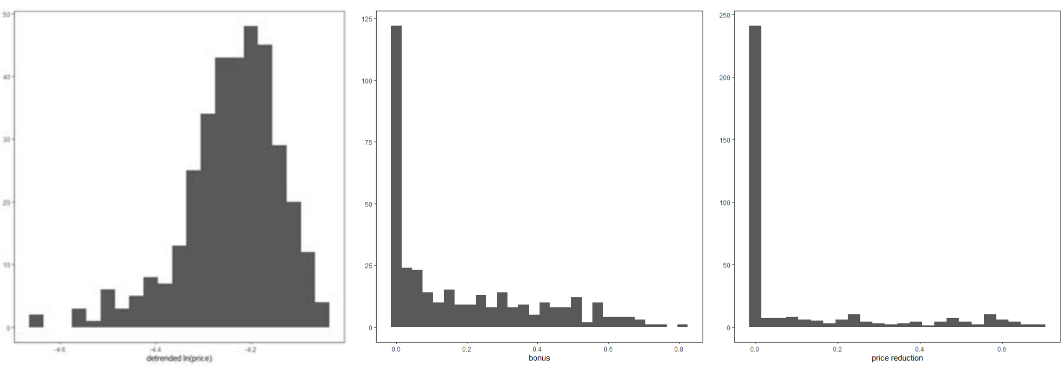
where $t = 1, 2, ..., T$ indexes week. The vector W_t includes all exogenous regressors, which are two promotion variables, bonus promotion and price reduction, in this application. Figure 2 shows log sales and log retail prices of toothpaste at store 1 over time (store 2 is very similar). To control for the possible trend of retail price over time, we use de-trended log retail prices (as instrumental variables as well) for estimation below.

Figure 3 shows the histograms of detrended log retail prices and the two promotion variables. All the three variables are continuous variables. Moreover, except log retail price, which is a bit close to normal distribution, the other two regressors, bonus and price reduction, are both nonnormally distributed. Therefore, we expect that the proposed 2sCOPE method can exploit these additional features of exogenous regressors correlated with the endogenous regressor for model identification and estimation even if the endogenous regressor is close to normal distri-



(a) Store 1 log sales (b) Store 1 log retail price

Figure 2. Log Sales and Log Retail Price of Toothpaste in Store 1.



(a) detrended log price (b) bonus (c) price reduction

Figure 3. Histogram of Log Retail Price, Bonus and Price Reduction in Store 1

bution. We estimate the model using the OLS, two-stage least squares (TSLS), CopulaP&G and our two proposed COPEs methods.

We use the IV-based TSLS estimator as a benchmark to test the validity of our proposed methods. Following Park and Gupta (2012), we use retail price at the other store as an instrument for price. This variable can be a valid instrument as it satisfies the two key requirements. First, retail prices across stores in a same market can be highly correlated because wholesale prices are usually offered the same (or very close). The Pearson correlation between the detrended log retail prices at Stores 1 and 2 is 0.79, providing strong explanatory power on the endogenous price. The correlation is comparable to that in Park and Gupta (2012). Second, some unmeasured product characteristics such as shelf-space allocation, shelf location and category location are determined by retailers and are usually not systematically related to wholesale prices (exclusion restriction). For the three copula-based methods, we make use of information from the existing endogenous and exogenous regressors and no extra IVs are needed. In CopulaP&G, we add the copula transformation of the detrended log price, $\mathrm{logP}^* = \Phi^{-1}(\widehat{H}(\mathrm{logP}))$, as a "generated regressor" to the outcome regression. For the COPE method, we add another two "generated regressors", copula transformation of bonus and price reduction ($\mathrm{Bonus}^* = \Phi^{-1}(\widehat{L}_1(\mathrm{Bonus}))$, $\mathrm{PriceRedu}^* = \Phi^{-1}(\widehat{L}_2(\mathrm{PriceRedu}))$). For the 2sCOPE method, we first regress logP^* on Bonus^* and $\mathrm{PriceRedu}^*$, and then add the residual as the only "generated regressor" to the outcome regression. $\widehat{H}(\cdot), \widehat{L}_1(\cdot), \widehat{L}_2(\cdot)$ are all estimated CDFs using the univariate empirical distribution for each regressor. Standard errors of parameter estimates are obtained using bootstrap.

Table 10 reports the estimation results. Beginning with the results from Store 1, OLS estimates are significantly different from TSLS estimates, indicating that the price endogeneity issue occurs. Instrumenting for retail price changes the price coefficient estimate from -0.767 to -1.797, implying that there is a positive correlation between unobserved product characteristics and the price. The estimates of ρ in the three IV-free copula-based methods, representing the correlation between the endogenous regressor P_t and the error term, are all significantly positive, further confirming our previous conclusion. This direction of correlation is consistent with previous empirical findings (e.g., Villas-Boas and Winer 1999, Chintagunta et al. 2005). The price elasticity estimates from the CopulaP&G, the proposed COPE and 2sCOPE are -3.082, -3.111 and -2.014, respectively. Among the three estimates, the estimate of -2.014 from the proposed 2sCOPE is close to the estimate of -1.797 from the TSLS method, whereas the

Table 10: Estimation Results: Toothpaste Sales

		OLS			TSLS			CopulaP&G			COPE			2sCOPE		
Store	Parameters	Est	SE	t-value	Est	SE	t-value	Est	SE	t-value	Est	SE	t-value	Est	SE	t-value
Store 1	Constant	1.301	1.197	0.25	-2.993	1.646	1.82	-8.526	2.619	3.26	-8.569	2.820	3.04	-3.908	2.314	1.69
	Price	-0.767	0.288	2.66	-1.797	0.396	4.54	-3.082	0.620	4.97	-3.111	0.664	4.69	-2.014	0.555	3.63
	Bonus	0.371	0.122	3.31	0.104	0.141	0.74	0.415	0.115	3.61	0.522	0.288	1.81	0.064	0.171	0.37
	PriceRedu	0.498	0.115	4.33	0.285	0.125	2.28	0.544	0.111	4.90	1.033	0.211	4.90	0.275	0.143	1.92
	ρ	-	-	-	-	-	-	0.521	0.098	5.32	0.662	0.117	5.66	0.297	0.089	3.34
Store 2	Constant	-3.898	1.246	3.13	0.763	1.943	0.39	1.107	3.404	0.33	1.324	3.430	0.39	0.001	2.702	0.00
	Price	-1.982	0.300	6.61	-0.864	0.467	1.85	-0.799	0.807	0.99	-0.783	0.811	0.96	-1.048	0.648	1.62
	Bonus	0.062	0.116	0.53	0.286	0.148	1.93	0.032	0.117	0.27	-0.819	0.426	1.92	0.239	0.151	1.58
	PriceRedu	0.283	0.111	2.55	0.540	0.137	3.94	0.275	0.110	2.5	0.540	0.194	2.78	0.467	0.152	3.07
	ρ	-	-	-	-	-	-	-0.319	0.177	1.80	-0.358	0.164	2.18	-0.188	0.109	1.72

existing copula and the proposed COPE yield substantially smaller price elasticity estimates. We confirm in the literature that the TSLS and 2sCOPE estimates are reasonable because the price elasticity of toothpaste category in the literature is around -2.0 (Hoch et al. 1995, Mackiewicz and Falkowski 2015). Comparing the estimates of ρ from the three IV-free copula-based methods, our proposed 2sCOPE provides a much smaller estimate of ρ (0.297 for 2sCOPE vs 0.521 for CopulaP&G and 0.662 for COPE in Table 10), consistent with the over-correction in both CopulaP&G and COPE.

Reasons for the substantial difference in the estimates from the CopulaP&G include (1) its ignoring correlated endogenous and exogenous regressors which can lead to inconsistent estimates, and (2) the unimodal close-to-normality distribution for the logarithm of price variable leading to potentially poor finite sample performance. In fact, the correlations between logP* and the exogenous regressors are -0.44 for Bonus and -0.26 for PriceRedu, both of which are substantially larger than the corresponding correlations (-0.30 and -0.15, respectively) between logP and the exogenous regressors. The p-value for the null hypothesis of these correlations being zeros are significantly less than 0.05 (< 0.001), indicating a violation of Assumption 4 required for CopulaP&G to yield consistent estimates.

Reasons for the substantial difference in the estimates from the proposed COPE method include (1) a uni-modal close-to-normality distribution for the price variable leading to potentially poor finite sample performance of COPE, and (2) loss of estimation precision manifested due to a larger standard error of estimates as compared with those from 2sCOPE. By contrast, the proposed 2sCOPE can relax the non-normality assumption of the endogenous regressor, and yield consistent and efficient estimates even if the endogenous regressor follows a normal or nearly normal distribution. Moreover, 2sCOPE provides estimates with smaller standard error than COPE, which confirms Theorem 4 showing that using two-stage copula estimation reduces estimation variance.

Unlike Store 1, the results from Store 2 indicate that the retail price is not endogenous. First, The estimates of ρ, which is the correlation between price and the error term, are not significantly different from 0 for both CopulaP&G and 2sCOPE (t-value ≤ 1.96 under columns "CopulaP&G" and "2sCOPE" for Store 2 in Table 10), and only slightly significantly different from 0 for COPE (a t-value of 2.18, slightly larger than 1.96 under Column "COPE" in Table 10). The estimate of ρ for the COPE, however, is questionable because of the limitations of COPE mentioned in the paragraph above. Second, the estimated price coefficient of OLS is -1.982, which is very close to the estimates of TSLS and 2sCOPE in store 1 and further confirming no endogeneity of price in store 2. Overall, the price elasticity estimates from TSLS and the three IV-free copulas-based methods are close to each other for Store 2, and the observed differences between them and the OLS estimate can be attributed to estimation variability incurred from using more complicated models instead of the presence of endogeneity.

In sum, the convergence of results between TSLS and the proposed method 2sCOPE in both stores supports the validity of the proposed methods in addressing the endogeneity issue. Moreover, the difference between the estimates in the two proposed methods (COPE vs 2sCOPE) in store 1 shows the advantages of 2sCOPE in terms of relaxing the non-normality assumption of the endogenous regressor and estimation efficiency gain by exploiting additional information from correlated exogenous regressors.

6. Conclusion

Causal inference lies at the center of social science research, and observational studies often beg rigorous post-study designs and methodologies to overcome endogeneity concerns. In this paper, we focus on the instrument-free copula method to handle the problem of endogenous regressors. We propose a generalized two-stage copula endogeneity correction (2sCOPE) method that overcomes two key limitations of the existing copula-based method in Park and Gupta (2012) (CopulaP&G), and extend CopulaP&G to more general settings. Specifically, 2sCOPE allows exogenous regressors to be correlated with endogenous regressors and relaxes the non-normality assumption on the endogenous regressors. To demonstrate the benefits of 2sCOPE, we compare it with the other proposed method, called COPE. Similar to CopulaP&G, both COPEs (COPE and 2sCOPE) methods correct endogeneity by adding "generated regressors" derived from the existing regressors and are straightforward to use. COPE is a direct extension to CopulaP&G by adding latent copula transformation of existing regressors, while 2sCOPE has two stages and adds the residuals from regressing latent copula data for the endogenous regressor on the latent copula data for the exogenous regressors as a "generated regressor" in the structural regression model. We theoretically prove that both proposed COPEs methods can yield consistent causal-effect estimates when exogenous regressors are correlated with the endogenous regressors, which can cause biased estimates in the method of Park and Gupta (2012). Moreover, the 2sCOPE method can further relax the nonnormality assumption on the endogenous regressors and improve estimation efficiency.

We conduct simulation studies and use an empirical marketing application to empirically verify the performance of our proposed methods. The simulation results show that both methods yield consistent estimates under relaxed assumptions. Moreover, 2sCOPE method outperforms COPE in terms of dealing with normal endogenous regressors and improving estimation efficiency. Endogenous regressors are allowed to be normally distributed as long as one of the exogenous regressors is nonnormally distributed, which is a very weak assumption. The efficiency gain is substantial and can be up to ∼50%, meaning that sample size can be reduced by ∼50% to achieve the same estimation efficiency as compared with COPE method that does not exploit the correlations between endogenous and exogenous regressors. Last but not least, our robustness checks show that the proposed methods are reasonably robust to the structural error distributional assumption and non-Gaussian copula correlation structure (Online Appendix D). We further apply our methods to a commonly used public dataset in marketing. When dealing with endogenous price, we find that the estimated price coefficient using our proposed 2sCOPE is very close to the TSLS estimate, while OLS and CopulaP&G show large biases. Moreover, results of the two proposed methods demonstrate the advantage of 2sCOPE in dealing with (nearly) normal endogenous regressors and improving estimation efficiency.

These findings have rich implications for guiding the practical use of copula-based instrument-free methods to handle endogeneity. A known critical assumption for CopulaP&G is the non-normality of endogenous regressors. The users of the method in the literature have all been practicing the check and verification of this assumption. However, our work shows that this is insufficient: one also needs to check Assumption 4 for the one-endogenous-regressor case, and Assumption 4(b) for the multiple-endogenous-regressors case. Note that neither assumption is the same as checking the pairwise correlations between the endogenous and exogenous regressors. Assumption 4 evaluates pairwise correlations involving copula transformation of the endogenous regressor, which, as shown in our empirical application, can be substantially different from the pairwise correlations using the regressor itself (Danaher and Smith 2011). Assumption 4(b) evaluates the correlations between exogenous regressors and the linear combination of generated regressors, which are even more different from checking pairwise correlations on the regressors themselves. When the above assumptions are satisfied, CopulaP&G is preferred to our proposed COPEs methods, since the simpler and valid model outperforms more general but more complex models.

If any endogenous regressor fails to have sufficient departure from being normally distributed, or any exogenous regressor violates the Assumptions 4 or 4(b), our proposed COPEs methods should be used instead of CopulaP&G. Then the next step is to decide which of COPE and 2sCOPE to use. Both methods employ the generated regressor approach and are straightforward to use. However, if any endogenous regressor is normally distributed or is close to be normally distributed, only 2sCOPE can perform well. 2sCOPE also performs better by reducing estimation variance. Overall, 2sCOPE has much to recommend, and COPE is considered here mostly for demonstrating the benefits of 2sCOPE.

The 2sCOPE is straightforward to extend to many other settings, and we have derived 2sCOPE for a range of commonly used marketing models, including linear regression models, linear panel models with mixed-effects, random coefficient logit models and slope endogeneity. The 2sCOPE method proposed here can be applied to these cases and many other cases not studied here while accounting for correlations between exogenous and endogenous regressors and exploiting the correlations for model identification in the presence of insufficient non-normality of endogenous regressors.

Although the proposed 2sCOPE contributes to the literature by relaxing key assumptions of the existing copula correction method CopulaP&G and extending it to more general settings, it is not without limitations. For the 2sCOPE to work best, the distributions of the endogenous regressors need to contain adequate information. The condition is violated when the endogenous regressors follow Bernoulli distributions or discrete distributions with small support, as noted in Park and Gupta (2012). The proposed 2sCOPE method does not address this limitation.

Developing instrument-free methods to handle such inadequately distributed endogenous regressors is an important topic for future research. The simplicity of 2sCOPE hinges on the normal structural error and Gaussian copula dependence structure. Although 2sCOPE demonstrates reasonable robustness to departures from these assumptions as shown in Online Appendix D, future research is needed for more flexible methods testing and relaxing these assumptions. Despite these limitations, we expect that the proposed 2sCOPE will provide a useful alternative to a broad range of empirical problems when instruments are not available.

References

Anderson, E. T. and Simester, D. I. (2004). Long-run effects of promotion depth on new versus established customers: three field studies. *Marketing Science*, 23(1):4–20.

Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014.

Angrist, J. D. and Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic perspectives*, 15(4):69–85.

Arora, N. and Huber, J. (2001). Improving parameter estimates and model prediction by aggregate customization in choice experiments. *Journal of Consumer Research*, 28(2):273–283.

Ataman, M. B., Van Heerde, H. J., and Mela, C. F. (2010). The long-term effect of marketing strategy on brand sales. *Journal of Marketing Research*, 47(5):866–882.

Atefi, Y., Ahearne, M., Maxham III, J. G., Donavan, D. T., and Carlson, B. D. (2018). Does selective sales force training work? *Journal of Marketing Research*, 55(5):722–737.

Athey, S. and Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497.

Becker, J.-M., Proksch, D., and Ringle, C. M. (2021). Revisiting gaussian copulas to handle endogenous regressors. *Journal of the Academy of Marketing Science*, pages 1–21.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262.

Bombaij, N. J. and Dekimpe, M. G. (2020). When do loyalty programs work? the moderating role of design, retailer-strategy, and country characteristics. *International Journal of Research in Marketing*, 37(1):175–195.

Burmester, A. B., Becker, J. U., van Heerde, H. J., and Clement, M. (2015). The impact of pre-and post-launch publicity and advertising on new product sales. *International Journal of Research in Marketing*, 32(4):408–417.

Chintagunta, P., Dubé, J.-P., and Goh, K. Y. (2005). Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models. *Management Science*, 51(5):832–849.

Chintagunta, P., Erdem, T., Rossi, P. E., and Wedel, M. (2006). Structural modeling in marketing: review and assessment. *Marketing Science*, 25(6):604–616.

Danaher, P. and Smith, M. (2011). Modeling multivariate distributions using copulas: Applications in marketing (with discussion and rejoinder). *Marketing Science*, 30:4–21.

Datta, H., Foubert, B., and Van Heerde, H. J. (2015). The challenge of retaining customers acquired with free trials. *Journal of Marketing Research*, 52(2):217–234.

Dotson, J. P. and Allenby, G. M. (2010). Investigating the strategic influence of customer and employee satisfaction on firm financial performance. *Marketing Science*, 29(5):895–908.

Dubé, J.-P., Chintagunta, P., Petrin, A., Bronnenberg, B., Goettler, R., Seetharaman, P. S., Sudhir, K., Thomadsen, R., and Zhao, Y. (2002). Structural applications of the discrete choice model. *Marketing letters*, pages 207–220.

Ebbes, P., Wedel, M., and Böckenholt, U. (2009). Frugal iv alternatives to identify the parameter for an endogenous regressor. *Journal of Applied Econometrics*, 24(3):446–468.

Ebbes, P., Wedel, M., Böckenholt, U., and Steerneman, T. (2005). Solving and testing for regressor-error (in) dependence when no instrumental variables are available: With new evidence for the effect of education on income. *Quantitative Marketing and Economics*, 3(4):365–392.

Elshiewy, O. and Boztug, Y. (2018). When back of pack meets front of pack: How salient and simplified nutrition labels affect food sales in supermarkets. *Journal of Public Policy & Marketing*, 37(1):55–67.

Erickson, T. and Whited, T. M. (2002). Two-step gmm estimation of the errors-in-variables model using high-order moments. *Econometric Theory*, 18(3):776–799.

Godes, D. and Mayzlin, D. (2009). Firm-created word-of-mouth communication: Evidence from a field test. *Marketing science*, 28(4):721–739.

Gruner, R. L., Vomberg, A., Homburg, C., and Lukas, B. A. (2019). Supporting new product launches with social media communication and online advertising: sales volume and profit implications. *Journal of Product Innovation Management*, 36(2):172–195.

Guitart, I. A., Gonzalez, J., and Stremersch, S. (2018). Advertising non-premium products as if they were premium: The impact of advertising up on advertising elasticity and brand equity. *International journal of research in marketing*, 35(3):471–489.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.

Hartmann, W., Nair, H. S., and Narayanan, S. (2011). Identifying causal marketing mix effects using a regression discontinuity design. *Marketing Science*, 30(6):1079–1097.

Haschka, R. E. (2021). Express: Handling endogenous regressors using copulas: A generalization to linear panel models with fixed effects and correlated regressors. *Journal of Marketing Research, First appeared online on Dec 18, 2021.*

Heitmann, M., Landwehr, J. R., Schreiner, T. F., and van Heerde, H. J. (2020). Leveraging brand equity for effective visual product design. *Journal of Marketing Research*, 57(2):257–277.

Hoch, S. J., Kim, B.-D., Montgomery, A. L., and Rossi, P. E. (1995). Determinants of store-level price elasticity. *Journal of marketing Research*, 32(1):17–29.

Hogan, V. and Rigobon, R. (2003). Using unobserved supply shocks to estimate the returns to education. *Unpublished manuscript.*

Johnson, G. A., Lewis, R. A., and Nubbemeyer, E. I. (2017). Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research*, 54(6):867–884.

Keller, W. I., Deleersnyder, B., and Gedenk, K. (2019). Price promotions and popular events. *Journal of Marketing*, 83(1):73–88.

Kleibergen, F. and Zivot, E. (2003). Bayesian and classical approaches to instrumental variable regression. *Journal of Econometrics*, 114(1):29–72.

Lamey, L., Deleersnyder, B., Steenkamp, J.-B. E., and Dekimpe, M. G. (2018). New product success in the consumer packaged goods industry: A shopper marketing approach. *International Journal of Research in Marketing*, 35(3):432–452.

Lewbel, A. (1997). Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and r&d. *Econometrica: journal of the econometric society*, pages 1201–1213.

Mackiewicz, R. and Falkowski, A. (2015). The use of weber fraction as a tool to measure price sensitivity: a gain and loss perspective. *Advances in Consumer Research*, 43.

Narayanan, S. and Kalyanam, K. (2015). Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Science*, 34(3):388–407.

Novak, S. and Stern, S. (2009). Complementarity among vertical integration decisions: Evidence from automobile product development. *Management Science*, 55(2):311–332.

Otter, T., Gilbride, T. J., and Allenby, G. M. (2011). Testing models of strategic behavior characterized by conditional likelihoods. *Marketing Science*, 30(4):686–701.

Papies, D., Ebbes, P., and Van Heerde, H. J. (2017). Addressing endogeneity in marketing models. In *Advanced methods for modeling markets*, pages 581–627. Springer.

Park, S. and Gupta, S. (2012). Handling endogenous regressors by joint estimation using copulas. *Marketing Science*, 31(4):567–586.

Petrin, A. and Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of marketing research*, 47(1):3–13.

Qian, Y. (2008). Impacts of entry by counterfeiters. *Quarterly Journal of Economics*, 123:1577–1609.

Qian, Y. and Xie, H. (2021). Simplifying bias correction for selective sampling: A unified distribution-free approach to handling endogenously selected samples. *Marketing Science, Forthcoming, available at https://www.nber.org/papers/w28801.*

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics*, 85(4):777–792.

Rossi, P. E. (2014). Even the rich can make themselves poor: A critical examination of iv methods in marketing applications. *Marketing Science*, 33(5):655–672.

Rutz, O. J. and Watson, G. F. (2019). Endogeneity and marketing strategy research: An overview. *Journal of the Academy of Marketing Science*, 47(3):479–498.

Shi, H., Sridhar, S., Grewal, R., and Lilien, G. (2017). Sales representative departures and customer reassignment strategies in business-to-business markets. *Journal of Marketing*, 81(2):25–44.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

Sudhir, K. (2001). Competitive pricing behavior in the auto market: A structural analysis. *Marketing Science*, 20(1):42–60.

Sun, B. (2005). Promotion effect on endogenous consumption. *Marketing science*, 24(3):430–443.

Van Heerde, H. J., Gijsenberg, M. J., Dekimpe, M. G., and Steenkamp, J.-B. E. (2013). Price and advertising effectiveness over the business cycle. *Journal of Marketing Research*, 50(2):177–193.

Villas-Boas, J. M. and Winer, R. S. (1999). Endogeneity in brand choice models. *Management science*, 45(10):1324–1338.

Wang, Y. and Blei, D. (2019). The blessings of multiple causes. *Journal of the American Statistical Association*, 114(528):1574–1596.

Wetzel, H. A., Hattula, S., Hammerschmidt, M., and van Heerde, H. J. (2018). Building and leveraging sports brands: evidence from 50 years of german professional soccer. *Journal of the Academy of Marketing Science*, 46(4):591–611.

Yang, S., Chen, Y., and Allenby, G. M. (2003). Bayesian analysis of simultaneous demand and supply. *Quantitative marketing and economics*, 1(3):251–275.

Online Appendix A: Proofs

Proof of Theorem 1

Under the Gaussian copula assumption for structural error term ξ_t and the endogenous regressor P_t, and the normality assumption of ξ_t, the outcome regression becomes (Equation 6)

$$Y_t = \mu + P_t\alpha + W_t\beta + \sigma_\xi \cdot \rho \cdot P_t^* + \sigma_\xi \cdot \sqrt{1-\rho^2} \cdot \omega_t$$

Because of the exogeneity assumption of W_t in linear model (Equation 1), $Cov(W_t, \xi_t) = 0$,

$$\begin{aligned} Cov(W_t, \xi_t) &= Cov(W_t, \sigma_\xi \cdot \rho \cdot P_t^* + \sigma_\xi \cdot \sqrt{1-\rho^2} \cdot \omega_t) \\ &= \sigma_\xi \cdot \rho \cdot Cov(W_t, P_t^*) + \sigma_\xi \cdot \sqrt{1-\rho^2} \cdot Cov(W_t, \omega_t) = 0. \end{aligned}$$

Thus, whenever W_t and P_t^* is correlated, the covariance between W_t and P_t^* is

$$Cov(W_t, \omega_t) = -\frac{\rho}{\sqrt{1-\rho^2}} Cov(W_t, P_t^*) \neq 0,$$

and W_t would be correlated with the new error term ω_t. **Theorem proved.**

Proof of Theorem 2: Consistency of COPE

Under the Gaussian copula model for (P_t, ξ_t) and the normality assumption of the error term ξ_t, we can divide ξ_t into an endogenous and exogenous part and our proposed COPE method is based on the OLS estimation of the regression below (Equation 10) by adding P_t^* and W_t^* as generated regressors.

$$Y_t = \mu + P_t\alpha + W_t\beta + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2} P_t^* + \frac{-\sigma_\xi \rho_{pw} \rho_{p\xi}}{1-\rho_{pw}^2} W_t^* + \sigma_\xi \sqrt{1 - \rho_{p\xi}^2 - \frac{\rho_{pw}^2 \rho_{p\xi}^2}{1-\rho_{pw}^2}} \cdot \omega_{3,t}$$

We want to prove that the new error term $\omega_{3,t}$ is uncorrelated with all terms of the right-hand side. Since $\omega_{1,t}$, $\omega_{2,t}$ and $\omega_{3,t}$ follow a standard multivariate Gaussian distribution (Equation 8), they are independent. According to the same equation, W_t^* and P_t^* are linear functions of $\omega_{1,t}$ and $\omega_{2,t}$. Thus, P_t^* and W_t^* are normally distributed and are independent of $\omega_{3,t}$. Since functions of independent variables are still independent, P_t (W_t), as a function of P_t^* (W_t^*), would be uncorrelated with $\omega_{3,t}$ and thus $\omega_{3,t}$ is not correlated with P_t, P_t^*, W_t and W_t^* on the right-hand side of Equation (10). Since P_t and W_t are nonnormal distributed, the full rank assumption is satisfied and thus COPE yields consistent estimates. **Theorem proved**.

Next we show that this result can be readily extended to the multi-dimension W_t case. We first derive the regression of the COPE method. Here we take 2-dimension W_t as an example. When there are one endogenous regressor P_t and two exogenous regressors W_t, the linear regression is:

$$Y_t = \beta_0 + \beta_1 P_t + \beta_2 W_{1,t} + \beta_3 W_{2,t} + \xi_t \tag{39}$$

Under the Gaussian Copula assumption,

$$\begin{pmatrix} P_t^* \\ W_{1,t}^* \\ W_{2,t}^* \\ \xi_t^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_\xi \\ \rho_1 & 1 & \rho_w & 0 \\ \rho_2 & \rho_w & 1 & 0 \\ \rho_\xi & 0 & 0 & 1 \end{bmatrix} \right) \tag{40}$$

The multivariate normal distribution can be written as follows:

$$\begin{pmatrix} P_t^* \\ W_{1,t}^* \\ W_{2,t}^* \\ \xi_t^* \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \rho_1 & \sqrt{1-\rho_1^2} & 0 & 0 \\ \rho_2 & \frac{\rho_w - \rho_1\rho_2}{\sqrt{1-\rho_1^2}} & \sqrt{1-\rho_2^2 - \frac{(\rho_w-\rho_1\rho_1)^2}{1-\rho_1^2}} & 0 \\ \rho_\xi & \frac{-\rho_1\rho_\xi}{\sqrt{1-\rho_1^2}} & \frac{\frac{(\rho_w-\rho_1\rho_2)\rho_1\rho_\xi}{1-\rho_1^2} - \rho_2\rho_\xi}{\sqrt{1-\rho_2^2-\frac{(\rho_w-\rho_1\rho_2)^2}{1-\rho_1^2}}} & \gamma \end{pmatrix} \cdot \begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \\ \omega_{3,t} \\ \omega_{4,t} \end{pmatrix},$$

where $\omega_{k,t} \sim N(0,1), k = 1,2,3,4,$ $\gamma = \sqrt{1-\rho_\xi^2 - \frac{\rho_1^2\rho_\xi^2}{1-\rho_1^2} - \left(\frac{\frac{(\rho_w-\rho_1\rho_2)\rho_1\rho_\xi}{1-\rho_1^2} - \rho_2\rho_\xi}{\sqrt{1-\rho_2^2-\frac{(\rho_w-\rho_1\rho_2)^2}{1-\rho_1^2}}} \right)^2}$. Structural error ξ_t can then be written as a function of P_t^* and W_t^*,

$$\xi_t = \sigma_\xi \xi_t^* = \frac{\sigma_\xi \rho_\xi (1-\rho_w^2)}{1-\rho_1^2-\rho_2^2+2\rho_1\rho_2\rho_w+\rho_w^2} \left(P_t^* - \frac{\rho_1-\rho_2\rho_w}{1-\rho_w^2} W_{1,t}^* - \frac{\rho_2-\rho_1\rho_w}{1-\rho_w^2} W_{2,t}^* \right) + \sigma_\xi \gamma \cdot \omega_{4,t}. \tag{41}$$

Thus, our COPE method in 2-W case becomes:

$$Y_t = \beta_0 + \beta_1 P_t + \beta_2 W_{1,t} + \beta_3 W_{2,t} + \beta_4 P_t^* + \beta_5 W_{1,t}^* + \beta_6 W_{2,t}^* + \sigma_\xi \gamma \cdot \omega_{4,t} \tag{42}$$

where

$$\beta_4 = \frac{\sigma_\xi \rho_\xi (1-\rho_w^2)}{1-\rho_1^2-\rho_2^2+2\rho_1\rho_2\rho_w+\rho_w^2}$$
$$\beta_5 = \frac{-\sigma_\xi \rho_\xi (1-\rho_w^2)}{1-\rho_1^2-\rho_2^2+2\rho_1\rho_2\rho_w+\rho_w^2} \cdot \frac{\rho_1-\rho_2\rho_w}{1-\rho_w^2}$$
$$\beta_6 = \frac{-\sigma_\xi \rho_\xi (1-\rho_w^2)}{1-\rho_1^2-\rho_2^2+2\rho_1\rho_2\rho_w+\rho_w^2} \cdot \frac{\rho_2-\rho_1\rho_w}{1-\rho_w^2}.$$

Since $\omega_{4,t}$ is independent of P_t^*, $W_{1,t}^*$ and $W_{2,t}^*$, it would also be uncorrelated with any functional form of P_t^*, $W_{1,t}^*$ and $W_{2,t}^*$, and thus $\omega_{4,t}$ is uncorrelated with any other terms in Equation (42). The COPE method can easily be extended to the case with multiple endogenous regressors by adding copula transformation of each regressor as generated regressors into the outcome regression, and the proof of estimation consistency is similar.

Proof of Theorem 3: Consistency of 2sCOPE

We have shown the derivation of 2sCOPE method in Section 3. The system of equations used in 2sCOPE method (Equations 11, 12) leads to the following equations

$$Y_t = \mu + P_t\alpha + W_t\beta + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2}\epsilon_t + \sigma_\xi\sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2\rho_{p\xi}^2}{1-\rho_{pw}^2}} \cdot \omega_{3,t},$$

$$P_t^* = \rho_{pw}W_t^* + \epsilon_t.$$

The proof of consistency is similar to the proof of Theorem 2. Since $\omega_{3,t}$ is independent of P_t^* and W_t^*, it would also be uncorrelated with any functional form of P_t^* and W_t^*, and thus $\omega_{3,t}$ is uncorrelated with P_t, W_t and ϵ_t. Once P_t or W_t is nonnormal, ϵ_t is not a linear function of P_t and W_t, satisfying the full rank condition required for model identification using the 2sCOPE method. **Theorem proved**.

Next we show that this result can be easily extended to the multi-dimension W_t case. We first derive the system of equations of the 2sCOPE method. Here we take 2-dimension W_t as an example. Because of the Gaussian relationship among P_t^* and W_t^* we assumed in Equation (40), the first stage regression becomes

$$\begin{aligned} P_t^* &= \frac{\rho_1 - \rho_2\rho_w}{1-\rho_w^2}W_{1,t}^* + \frac{\rho_2 - \rho_1\rho_w}{1-\rho_w^2}W_{2,t}^* + \sqrt{1-\rho_1^2 - \frac{(\rho_2-\rho_1\rho_w)^2}{1-\rho_w^2}}\omega_{3,t} \\ &= \frac{\rho_1 - \rho_2\rho_w}{1-\rho_w^2}W_{1,t}^* + \frac{\rho_2 - \rho_1\rho_w}{1-\rho_w^2}W_{2,t}^* + \epsilon_{2,t} \\ &= \gamma_1 W_{1,t}^* + \gamma_2 W_{2,t}^* + \epsilon_{2,t}. \end{aligned} \tag{43}$$

The structural error ξ_t in Equation (11) and the first-stage error term $\epsilon_{2,t}$ are linear transformations of the Gaussian data $(\xi_t, P_t^*, W_{1,t}^*, W_{2_t}^*)$ and thus follow a bivariate normal distribution. Thus, ξ_t can be decomposed to a sum of one term containing $\epsilon_{2,t}$ and an independent new error term, resulting in the following regression equation:

$$Y_t = \beta_0 + \beta_1 P_t + \beta_2 W_{1,t} + \beta_3 W_{2,t} + \beta_4\epsilon_{2,t} + \sigma_\xi\gamma\cdot\omega_{4,t}. \tag{44}$$

where

$$\beta_4 = \frac{\sigma_\xi\rho_\xi(1-\rho_w^2)}{1-\rho_1^2-\rho_2^2+2\rho_1\rho_2\rho_w+\rho_w^2}.$$

Since $\omega_{4,t}$ is independent of P_t^*, $W_{1,t}^*$ and $W_{2,t}^*$, it is uncorrelated with any functional form of P_t^*, $W_{1,t}^*$ and $W_{2,t}^*$, and thus $\omega_{4,t}$ is uncorrelated with P_t, $W_{1,t}$, $W_{2,t}$ and $\epsilon_{2,t}$ in Equation (44). Thus, 2sCOPE that performs OLS regression of Equation (44) yields consistent model estimates. Without loss of generality, the result can be extended to cases with any dimension of W_t.

Proof of Theorem 4: Variance Reduction

According to the COPE method in Equation (10),

$$Y_t = \mu + P_t\alpha + W_t\beta + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2} P_t^* + \frac{-\sigma_\xi \rho_{pw}\rho_{p\xi}}{1-\rho_{pw}^2} W_t^* + \sigma_\xi \sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2 \rho_{p\xi}^2}{1-\rho_{pw}^2}} \cdot \omega_{3,t}.$$

The coefficients of P_t^* and W_t^* follows a linear relationship. Denote δ_3 and δ_4 the coefficients of P_t^* and W_t^* respectively. Then,

$$\delta_4 + \rho_{pw}\delta_3 = 0.$$

With the two-stage estimation in 2sCOPE (Equation 14), ρ_{pw} is estimated in the first stage and is thus treated as a known parameter in the main regression. That is, 2sCOPE can be viewed as the COPE method with a linear restriction. The linear restriction is,

$$\delta_4 + \hat{\rho}_{pw}\delta_3 = 0. \tag{45}$$

In this case, the two-stage copula method (2sCOPE) can be viewed as one kind of restricted least square estimation based on COPE. We next prove that restricted least square can achieve reductions in standard errors. Suppose we simplify the regression expression in Equation (10) as

$$y = X\theta + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2 I)$, $X \equiv (1, P_t, W_t, P_t^*,\ W_t^*)$, and $\theta = (\mu, \alpha, \beta, \delta_3, \delta_4)$. The restriction in Equation (45) becomes

$$R\theta = 0, \text{where} R = (0, 0, 0, \hat{\rho}_{pw}, 1).$$

Thus, the 2sCOPE yields the least square estimates $\hat{\theta}_2$ of Equation (10) subject to the above restriction, whereas COPE yields the unrestricted least square estimates, $\hat{\theta}_1$, as follows.

$$\hat{\theta}_1 \sim N(\theta, \sigma^2 (X'X)^{-1}),$$
$$\hat{\theta}_2 \sim N(\theta, \sigma^2 M(X'X)^{-1}M').$$

where according to restricted least square theory, $M = I - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R$. Let us compare the variance of $\hat{\theta}_1$ and $\hat{\theta}_2$. Note that,

$$\begin{aligned}
&M(X'X)^{-1}M' \\
=&(I - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R)(X'X)^{-1}(I - R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}) \\
=&(X'X)^{-1} - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}.
\end{aligned}$$

Therefore,

$$\begin{aligned} Var(\hat{\theta}_1) - Var(\hat{\theta}_2) &= \sigma^2\{(X'X)^{-1} - M(X'X)^{-1}M'\} \\ &= \sigma^2(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1} \geq 0. \end{aligned}$$

Since the matrix $Var(\hat{\theta}_1) - Var(\hat{\theta}_2)$ is positive semi-definite, all the diagonal elements should be greater than or equal to zero. Thus, the imposition of the linear restriction brings about a variance reduction. **Theorem proved.**

We have proved that there would be variance reduction when there exist restriction of parameters. When the exogenous variable W_t is a scalar, the linear restriction is shown in Equation (45). We next show that when W_t is extended to a multi-dimension vector, there are still linear restrictions and variance reduction of 2sCOPE. We take a 2-dimension W_t as an example below. According to the 2sCOPE method with 2-dimension W_t in Equations (43, 44), 2sCOPE is equivalent to adding two restrictions to COPE in Equation (42). The two restrictions are:

$$\beta_5 + \hat{\gamma}_1\beta_4 = 0$$

$$\beta_6 + \hat{\gamma}_2\beta_4 = 0$$

where $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are estimated and obtained in the first stage in Equation (43). Thus, compared with COPE, we still have variance reduction using 2sCOPE in the 2-W case. Without loss of generality, this result can be extended to cases with any dimension of W_t.

Proof of Theorem 5: Nonnormality Assumption Relaxed

In this section, we prove that our proposed 2sCOPE method can relax the nonnormality assumption on the endogenous regressors imposed in CopulaP&G, while COPE does not.

We first examine the COPE method in Equation (10),

$$Y_t = \mu + P_t\alpha + W_t\beta + \frac{\sigma_\xi\rho_{p\xi}}{1-\rho_{pw}^2}P_t^* + \frac{-\sigma_\xi\rho_{pw}\rho_{p\xi}}{1-\rho_{pw}^2}W_t^* + \sigma_\xi\sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2\rho_{p\xi}^2}{1-\rho_{pw}^2}} \cdot \omega_{3,t}.$$

If the endogenous regressor P_t is normally distributed, $P_t = \Phi_{\sigma_p}^{-1}(\Phi(P_t^*)) = \sigma_p P_t^*$ and thus P_t^* and P_t would be fully collinear, violating the full rank assumption and making the model unidentified.

We then examine the 2sCOPE method in Equation (14).

$$\begin{aligned} Y_t &= \mu + P_t\alpha + W_t\beta + \frac{\sigma_\xi\rho_{p\xi}}{1-\rho_{pw}^2}\epsilon_t + \sigma_\xi\sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2\rho_{p\xi}^2}{1-\rho_{pw}^2}} \cdot \omega_{3,t}, \\ \epsilon_t &= P_t^* - \rho_{pw}W_t^*. \end{aligned}$$

When the endogenous regressor P_t is normally distributed, $P_t = \Phi_{\sigma_p}^{-1}(\Phi(P_t^*)) = \sigma_p P_t^*$. Since we add the residual ϵ_t from the first stage to the outcome regression instead of adding each P_t^* and W_t^*, ϵ_t would not be perfectly collinear with P_t and W_t as long as one of the Ws correlated with P_t is not normally distributed. **Theorem proved.**

Online Appendix B: COPEs for Slope Endogeneity with Correlated and Normally Distributed Regressors

In this section, we describe the COPEs approaches to addressing slope endogeneity with correlated regressors in the following model:

$$Y_t \quad = \quad \mu + P_t\alpha_t + W_t'\beta_t + \eta_t, \qquad \text{where } \alpha_t = \bar{\alpha} + \xi_t, \tag{46}$$

α_t, β_t are individual-specific regression coefficients and $\bar{\alpha}$ is the mean of α_i, $\xi_t \sim N(0, \sigma_\xi^2)$. The normal error term η_i is uncorrelated with the regressors P_t and W_t and thus causes no endogeneity concern. However, the random coefficient ξ_t can be correlated with the regressor P_t, causing the problem of "slope endogeneity". P_t and W_t can be correlated. Assuming that (P_t, W_t, α_t) follows a Gaussian copula model, the COPE approach to addressing the slope endogeneity problem is derived as follows.

$$\begin{aligned} Y_t &= \mu + P_t(\bar{\alpha} + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2} P_t^* + \frac{-\sigma_\xi \rho_{pw}\rho_{p\xi}}{1-\rho_{pw}^2} W_t^* + \sigma_\xi \sqrt{1 - \rho_{p\xi}^2 - \frac{\rho_{pw}^2 \rho_{p\xi}^2}{1-\rho_{pw}^2}} \omega_{3,t}) + W_t'\beta_t + \eta_t \\ &= \mu + P_t\bar{\alpha} + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2} P_t \times P_t^* + \frac{-\sigma_\xi \rho_{pw}\rho_{p\xi}}{1-\rho_{pw}^2} P_t \times W_t^* + W_t'\beta_t + \\ &\quad \sigma_\xi \sqrt{1 - \rho_{p\xi}^2 - \frac{\rho_{pw}^2 \rho_{p\xi}^2}{1-\rho_{pw}^2}} P_t \times \omega_{3,t} + \eta_t. \end{aligned} \tag{47}$$

Given both $P_t \times P_t^*$ and $P_t \times W_t^*$ in Equation (47), the unobserved variable $w_{3,t}$ is independent of all regressors (P_t, W_t, P_t^*, W_t^*) and uncorrelated with functions of these regressors. Thus, Equation (47) can be estimated using standard methods for random-effects models with $\omega_{3,t}$ as the random effect and $(P_t \times P_t^*,\ P_t \times W_t^*)$ as generated regressors. The method of Park and Gupta (2012) adds only $P_t \times P_t^*$ as a generated regressor, and may fail to yield consistent estimates when P_t and W_t are correlated, resulting in the correlation between the random effect in their method and the regressor W_t.

The 2sCOPE for addressing the slope endogeneity problem with correlated regressors is derived as follows

$$\begin{aligned} Y_t &= \mu + P_t(\bar{\alpha} + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2} \epsilon_t + \sigma_\xi \sqrt{1 - \rho_{p\xi}^2 - \frac{\rho_{pw}^2 \rho_{p\xi}^2}{1-\rho_{pw}^2}} \cdot \omega_{3,t}) + W_t'\beta_t + \eta_t \\ &= \mu + P_t\bar{\alpha} + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2} P_t^* \times \epsilon_t + W_t'\beta_t + \sigma_\xi \sqrt{1 - \rho_{p\xi}^2 - \frac{\rho_{pw}^2 \rho_{p\xi}^2}{1-\rho_{pw}^2}} P_t \times \omega_{3,t} + \eta_t, \end{aligned} \tag{48}$$

where only one generated regressor, $P_t^* \times \epsilon_t$, is needed, given which the random effect $\omega_{3,t}$ is independent of all regressors in Equation (48).

Both COPE and 2sCOPE can be implemented using the standard methods for random effects models by simply adding generated regressors to control for endogenous regressors. By contrast, the maximum likelihood approach requires constructing a complicated joint likelihood of $(\xi_t, \eta_t, P_t^*, W_t^*)$, which is not what the standard random effects method uses and thus requires separate development and significantly more computation involving numerical integration.

Online Appendix C: COPEs for Random Coefficient Logit Model with Correlated and Normally Distributed Regressors

We next consider endogeneity bias in the following random utility model with correlated endogenous and exogenous regressors:

$$\begin{aligned} u_{hjt} &= \psi_{hj} + P'_{jt}\alpha_h + W'_{jt}\beta_h + \xi_{jt} + \epsilon_{hjt}, \qquad j = 1, \cdots, J, \\ u_{h0t} &= \epsilon_{h0t}, \qquad j = 0 \text{ if no purchase,} \end{aligned}$$

where u_{hjt} denotes the utility for household $h = 1, \cdots, n_h$ at occasion $t = 1, \cdots, T$ with $j = 1, \cdots, J$ alternatives and $j = 0$ denotes the option of no purchase. In the utility function, ψ_{hj} is the individual-specific preference for choice j with ψ_{hJ} normalized to be zero for identification purpose, (P_{jt}, W_{jt}) include the choice characteristics, and (α_h, β_h) denote the individual-specific random coefficients. These individual-specific coefficients $(\psi_{hj}, \alpha_h, \beta_h)$ permit heterogeneity in both intercepts and regressor effects across cross-sectional units, such as consumers or households. In this model, the association between regressors in P_{jt} and the unobserved common shock ξ_{jt} causes endogeneity bias. We further allow P_{jt} and W_{jt} to be correlated. The term ϵ_{hjt} is the idiosyncratic error uncorrelated with all regressors. An individual at any occasion chose the alternative with the largest utility, i.e., $Y_{hjt} = 1$ iff $u_{hjt} > u_{hj't}\ \forall j' \neq j$. When ϵ_{hjt} follows an *i.i.d* Type I extreme value distribution, the choice probability follows the random-coefficient multinomial logit model.

The COPEs approach can be used to address the endogeneity issue using the following two-step procedure. In the first step, we estimate the model

$$u_{hjt} = \delta_{jt} + \widetilde{\psi}_{hj} + P'_{jt}a_h + W'_{jt}b_h + \epsilon_{hjt},$$

where $\delta_{jt} = \mu_j + P'_{jt}\bar{\alpha} + W'_{jt}\bar{\beta} + \xi_{jt}$, $(\mu_j, \bar{\alpha}, \bar{\beta})$ is the mean of random effects $(\psi_{hj}, \alpha_h, \beta_h)$, $\widetilde{\psi}_{hj} = \psi_{hj} - \mu_j, a_h = \alpha_h - \bar{\alpha}$ and $b_h = \beta_h - \bar{\beta}$. δ_{jt} is treated as occasion- and choice-specific fixed-effect parameters in this model. Since the regressors are uncorrelated with the error term ϵ_{hij}, there is no endogeneity bias in the model. In the second step, we estimate the equation below.

$$\widehat{\delta}_{jt} = \mu_j + P'_{jt}\bar{\alpha} + W'_{jt}\bar{\beta} + \xi_{jt} + \eta_{jt}, \tag{49}$$

where $\widehat{\delta}_{jt}$ denotes the estimate of the fix-effect δ_{jt}; η_{jt} denotes the estimation error of $\widehat{\delta}_{ij}$ and is approximately normally distributed. In the second-step model, the structural error is correlated with P_{jt}, leading to endogenous bias. We then apply COPEs to correct for the endogenous bias, which can avoid the potential bias of CopulaP&G due to the potential correlations between P and W, as well as make use of this correlation to relax the non-normality assumption of P_{it}, improve model identification and sharpen model estimates. The above development is for individual-level data. Park and Gupta (2012) also derived their copula method for addressing endogeneity bias in random coefficient logit models using aggregate-level data. It is straightforward to extend the COPEs to the setting with correlated regressors and (nearly) normal regressor distributions.

Online Appendix D: Additional Results

Case 1: Additional Results

Table 11: Results of the Simulation Study for Case 1 with Sample Size of 200

			OLS			CopulaP&G			COPE			2sCOPE		
ρ_{pw}	Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
0.5	μ	1	0.683	0.097	3.264	1.228	0.191	1.194	1.020	0.223	0.091	0.999	0.137	0.005
	α	1	1.583	0.079	7.388	1.048	0.178	0.271	0.990	0.184	0.056	0.996	0.175	0.023
	β	-1	-1.265	0.068	3.902	-1.291	0.068	4.293	-1.019	0.166	0.116	-1.004	0.101	0.044
	$\rho_{p\xi}$	0.5	-	-	-	0.559	0.122	0.489	0.493	0.139	0.048	0.489	0.097	0.109
	σ_ξ	1	0.857	0.044	3.224	1.016	0.107	0.148	1.018	0.100	0.176	1.001	0.094	0.013
	D-error			-			-			0.016598			0.009069	
0.7	μ	1	0.723	0.091	3.050	1.304	0.175	1.740	1.006	0.197	0.031	0.983	0.114	0.153
	α	1	1.817	0.095	8.583	1.255	0.161	1.584	1.032	0.182	0.175	1.044	0.174	0.253
	β	-1	-1.539	0.084	6.388	-1.574	0.086	6.686	-1.045	0.180	0.250	-1.033	0.131	0.251
	$\rho_{p\xi}$	0.5	-	-	-	0.624	0.103	1.200	0.490	0.135	0.077	0.480	0.067	0.297
	σ_ξ	1	0.796	0.039	5.156	0.988	0.105	0.116	0.999	0.096	0.011	0.982	0.090	0.205
	D-error			-			-			0.016245			0.008867	

Case D1: Misspecification of ξ_t

Similar to CopulaP&G, we assume the structural error to be normally distributed. Though the normality of ξ_t is a reasonable and commonly used assumption in marketing and economics literature, the true distribution of ξ_t is often unknown, resulting in possible misspecifications. In this simulation study, we examine the robustness of the proposed methods to the departures from the normality of ξ_t. We generate 1,000 data sets using the same multivariate normal distribution as in Equation (26). The rest of DGP is:

$$\xi_t = G^{-1}(U_{\xi,t}) \quad = \quad G^{-1}(\Phi(\xi_t^*)) = \Phi(\xi_t^*) - 0.5, \tag{50}$$

$$P_t = H^{-1}(U_{p,t}) = H^{-1}(\Phi(P_t^*)), \qquad W_t = L^{-1}(U_{w,t}) = L^{-1}(\Phi(W_t^*)), \tag{51}$$

$$Y_t = \mu + \alpha \cdot P_t + \beta \cdot W_t \quad + \quad \xi_t = 1 + 1 \cdot P_t + (-1) \cdot W_t + \xi_t. \tag{52}$$

where we set $P_t \sim Gamma(1,1)$ and $W_t \sim Exp(1)$ in the simulation. Note that the structural error ξ_t now follows a uniform distribution instead of a normal distribution. For estimation, we assume normality of ξ_t and use the OLS estimator and the proposed methods.

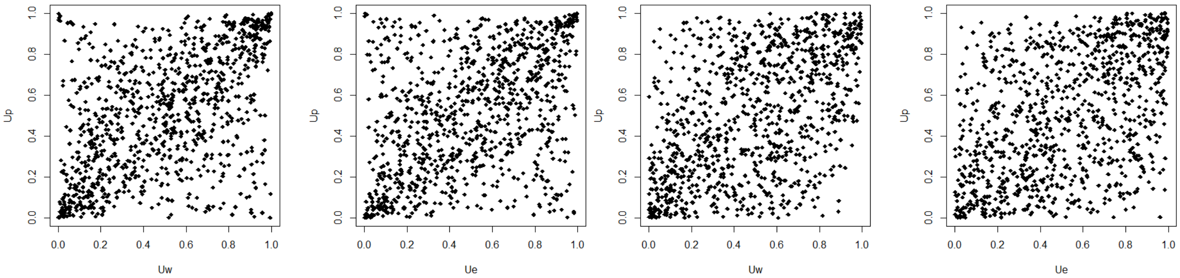
Table 12: Results of the Simulation Study Case D1: Misspecification of ξ_t

		OLS			COPE			2sCOPE		
Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
μ	1	0.912	0.013	6.807	1.005	0.026	0.203	1.004	0.016	0.232
α	1	1.161	0.010	16.30	0.992	0.017	0.506	0.992	0.016	0.481
β	-1	-1.073	0.009	8.180	-0.998	0.020	0.116	-0.996	0.011	0.321
$\rho_{p\xi}$	0.5	-	-	-	0.502	0.051	0.043	0.499	0.034	0.015
σ_ξ	0.289	0.251	0.004	8.624	0.293	0.009	0.476	0.292	0.009	0.385

Table 12 reports estimation results. As the same in Case 1, OLS estimates are biased. COPE and 2sCOPE can still recover the true parameter values despite the misspecification of ξ_t, demonstrating the robustness of the proposed COPEs methods to the normal error assumption.

D2: Misspecification of Copula

In the proposed methods, we use the Gaussian copula to capture the dependence structure among the regressors and error term (U_p, U_w and U_ξ). In practice, the dependence might come from an economic mechanism (such as marketing strategic decisions) and thus might be different from what the Gaussian copula generates. In this section, we examine the robustness of the Gaussian copula in simulated data. Specifically, we generate the dependence among U_p,



(a) T Copula ($\rho = 0.5$)

(b) Gaussian Copula ($\rho = 0.5$)

Figure 4. Scatter plots of Randomly Generated Pairs U_p, U_w (U_p,U_ξ) for Considered Copulas.

U_w and U_ξ using copula models other than the Gaussian copula. Specifically, we consider the following T copula models which provide flexible random general generation from arbitrary and heterogeneous correlation structures among more than two variables:

$$C(U_p, U_w, U_\xi) = \int_{-\infty}^{t_\nu^{-1}(U_p)} \int_{-\infty}^{t_\nu^{-1}(U_w)} \int_{-\infty}^{t_\nu^{-1}(U_\xi)} \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\pi\nu)^d|\Sigma|}} \left(1 + \frac{x'\Sigma^{-1}x}{\nu}\right) dx, \tag{53}$$

where t_ν^{-1} denotes the quantile function of a standard univariate t_ν distribution. We set the degree of freedom ν=2, and the dimension of the copula d=3 in this example. Σ is covariance matrix capturing correlations among variables. The data-generating process (DGP) of t copula is summarized below:

$$\begin{pmatrix} P_t^* \\ W_t^* \\ \xi_t^* \end{pmatrix} \sim t_\nu^d \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{pw} & \rho_{p\xi} \\ \rho_{pw} & 1 & 0 \\ \rho_{p\xi} & 0 & 1 \end{bmatrix} \right) = t_\nu^d \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix} \right). \tag{54}$$

Figure 4 shows the scatter plots of randomly generated (U_p, U_w, U_ξ) pairs from the above copulas, as well as the Gaussian copula with the same correlation of 0.5. The figure shows disparate dependence structures between U_p and ξ_t (U_p and U_w) for these two copulas.

We then use the following process to generate P_t, W_t and ξ_t:

$$\xi_t = G^{-1}(U_\xi) = \Phi^{-1}(U_\xi), \tag{55}$$

$$P_t = H^{-1}(U_p), W_t = L^{-1}(U_w), \tag{56}$$

$$Y_t = 1 + 1 \cdot P_t + (-1) \cdot W_t + \xi_t. \tag{57}$$

where $H(\cdot)$ is a gamma distribution and $L(\cdot)$ is an exponential distribution. We set $T = 1000$, generate 1000 data sets and estimate the parameters using the OLS estimator and the proposed COPEs methods.

Table 13 summarizes the estimation results. OLS estimates are still biased for all parameters. By contrast, estimates from the proposed COPE and 2sCOPE methods are centered closely around the true values. Therefore, the proposed methods based on the Gaussian copula are reasonably robust to the misspecification of the copula dependence structure among the regressors and the structural error.

Table 13: Results of the Simulation Study Case D2: Misspecification of Copula

		OLS			COPE			2sCOPE		
Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
μ	1	0.710	0.530	5.463	1.002	0.127	0.016	0.988	0.077	0.156
α	1	1.580	0.044	13.13	1.030	0.115	0.257	1.029	0.116	0.250
β	-1	-1.289	0.047	6.142	-1.033	0.127	0.262	-1.017	0.070	0.248
$\rho_{p\xi}$	0.5	-	-	-	0.463	0.085	0.435	0.458	0.067	0.622
σ_ξ	1	0.864	0.026	5.236	0.993	0.054	0.133	0.988	0.054	0.230