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, Revised June 2024

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, Revised June 2024
JEL No. C01,C1,C13,C18,C4

ABSTRACT

The ubiquitous presence of endogenous regressors presents a significant challenge when drawing causal inference using observational data. The classical econometric method used to handle regressor endogeneity requires IVs that must satisfy the stringent condition of exclusion restriction, rendering it unfeasible in many settings. Herein, we propose a new IV-free method that uses copulas to address the endogeneity problem. Existing copula correction methods require nonnormal endogenous regressors: normally or nearly normally distributed endogenous regressors cause model non-identification or significant finite-sample bias. Furthermore, existing copula control function methods presume the independence of exogenous regressors and the copula control function. Our proposed two-stage copula endogeneity correction (2sCOPE) method simultaneously relaxes the two key identification requirements. Assuming a Gaussian copula dependence structure for all regressors and a normally distributed structural error, we prove that 2sCOPE yields consistent causal-effect estimates with correlated endogenous and exogenous regressors as well as normally distributed endogenous regressors. In addition to relaxing the identification requirements, 2sCOPE has superior finite-sample performance and addresses the significant finite-sample bias problem due to insufficient regressor nonnormality. Moreover, 2sCOPE employs generated regressors derived from existing regressors to control for endogeneity, and can thus considerably increase the ease and broaden the applicability of IV-free methods for handling regressor endogeneity. We further demonstrate 2sCOPE's performance using simulation studies and illustrate its use in an empirical application.

Fan Yang
NEOMA Business School
59 Rue Pierre Taittinger
Reims, 51100
France
fan.yang@neoma-bs.fr

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

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

Causal inference is central to many problems that both academics and practitioners face. It has become increasingly important as rapidly available observational data in the current digital era promise to offer real-world evidence on cause-and-effect relationships for better decision makings. However, empirical researchers attempting to draw valid causal inferences from these data often encounter the presence of endogenous regressors correlated with the structural error in the population regression model representing the causal relationship of interest. Notably, omitted variables, such as unobserved demand shocks (e.g., product attributes, taste changes), may cause price endogeneity in observational scanner data. Ignoring this endogeneity can yield severely biased estimates of pricing effects on consumer demand (Villas-Boas and Winer 1999).

Instrumental variables (IV) have traditionally been used to address the endogeneity issue. The ideal IV must satisfy two requirements: first, it must be correlated with the endogenous regressor via an explainable and validated relationship (i.e., relevance restriction) and second, it must be uncorrelated with the structural error and must not directly affect the outcome (i.e., exclusion restriction). Although the theory of IVs is well-developed, researchers often struggle to find good IVs that meet both criteria. Potential IVs are often deficient by virtue of either weak relevance or challenging justification for exclusion restriction, hindering their use for addressing underlying endogeneity concerns (Rossi 2014).

The development and application of IV-free endogeneity-correction methods to address the lack of suitable IVs have gained traction in recent decades (Ebbes, Wedel, and Böckenholt 2009). Park and Gupta (2012) proposed an IV-free method that uses the copula model (Danaher 2007; Danaher and Smith 2011; Christopoulos, McAdam, and Tzavalis 2021) to directly model the regressor-error dependence.[1] In addition to requiring no IVs, the copula approach is straightforward to use: one can simply add the latent copula data for the endogenous regressors as control variables to correct for endogeneity. These features

[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.

considerably enhance the feasibility of endogeneity correction, as evidenced by the prolific use of the copula correction method (see examples of recent applications noted in the next section on literature review). Like other IV-free methods, however, the copula correction methods also require distinctiveness between the distributions of the endogenous regressor and the structural error. As such, the endogenous regressor is required to have a nonnormal distribution for model identification with the commonly assumed normal structural error distribution (Park and Gupta 2012; Papies, Ebbes, and Van Heerde 2017; Becker, Proksch, and Ringle 2021; Haschka 2022; Eckert and Hohberger 2022; Qian, Xie, and Koschmann 2022). Furthermore, we demonstrate that the existing copula control function correction methods implicitly require that all exogenous regressors 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 noticeable correlations are present between the CCF and exogenous regressors.

In practical applications, the requirements of sufficient regressor nonnormality and absence of correlations between CCF and exogenous regressors can be excessive and pose significant challenges to the application of the copula correction method. We often encounter endogenous regressors or include transformations of endogenous variables as regressors that have close-to-normal distributions. Examples of such regressors in economics and marketing management studies include stock market returns (Sorescu, Warren, and Ertekin 2017), corporate social responsibility (Eckert and Hohberger 2022), the organizational intelligent quotient (Mendelson 2000), and the logarithm of price (see Figure 4 in the Application). Theoretically, the endogenous regressor and the structural error can contain a common set of unobservables that collectively have a normal distribution, which can lead to a close-to-normal distribution of the endogenous regressor. In such scenarios, even if the model is identified asymptotically, the close-to-normality of endogenous regressors can cause estimation bias even in moderate sample sizes and require large sample sizes to mitigate the finite-sample bias (Becker, Proksch, and Ringle 2021). Correlations between the CCF and exogenous regressors are also common in practical applications, particularly when the exogenous regressors are included to control for potential confounders. Examples of such exogenous control variables abound in marketing and management studies, such as customer-specific variables (location, age, household size, income, past purchase behaviors, etc.) when estimating the returns of consumer targeting strategies on product sales (Papies, Ebbes, and Van Heerde 2017) and firms' similarity when estimating the effect of competition on innovation (Aghion et al. 2005).

The purpose of this paper is to provide a generalized copula control function procedure that relaxes the stringent requirements of both sufficient regressor nonnormality and the absence of correlations between CCF and exogenous regressors. Like the existing copula methods, 2sCOPE requires neither IVs nor the assumption of exclusion restriction. The 2sCOPE method corrects for endogeneity by adding residuals, obtained by regressing latent copula data for each endogenous regressor on the latent copula data for exogenous regressors, as generated regressors in the structural model. Unlike the original copula method (Park and Gupta 2012; Becker, Proksch, and Ringle 2021; Eckert and Hohberger 2022; henceforth $\text{Copula}_{\text{Origin}}$), 2sCOPE can account for the dependence between endogenous and exogenous regressors. $\text{Copula}_{\text{Origin}}$ thus constitutes a special case of 2sCOPE. Assuming that a Gaussian copula correlation model captures dependence among the endogenous regressors, correlated exogenous regressors, and the structural error, we prove that 2sCOPE can identify causal effects under weaker assumptions than $\text{Copula}_{\text{Origin}}$ and overcome the two key limitations menntioned above.

The remainder of this paper begins with a review of relevant literature on methods for causal inference with endogenous regressors and an overview of this work's contributions. We then propose 2sCOPE and prove its consistency with normally distributed and correlated regressors. Next, we evaluate 2sCOPE's performance using simulation studies under different scenarios and provide a flowchart to guide its use in practical applications. We further apply 2sCOPE to estimate price elasticity using store purchase databases.

LITERATURE REVIEW AND CONTRIBUTIONS

Marketing, economics, and statistics research has developed a rich array of methods for drawing causal inferences. Randomized experiments, such as controlled lab experiments and field experiments (Johnson, Lewis, and Nubbemeyer 2017; Anderson and Simester 2004; Godes and Mayzlin 2009) have long been the gold standard for estimating causal effects. When controlled experiments are not feasible, quasi-experimental designs, such as regression discontinuity, difference-in-difference, and synthetic control, are used to mimic randomized experiments and enable the identification of causal effects with observational data (Hartmann, Nair, and Narayanan 2011, Narayanan and Kalyanam 2015, Athey and Imbens 2006, Shi et al. 2017, Kim, Lee, and Gupta 2020). However, these quasi-experimental designs have special data and design requirements, and are not designed to cope with the general issue of endogenous regressors when estimating causal effects using observational data.

There exists a large literature on various approaches to addressing endogenous regressors when inferring causal effects. Papies, Ebbes, and Van Heerde (2017), Rutz and Watson (2019), and Park and Gupta (2012) provide overviews of approaches to addressing endogeneity in marketing. Among the three broad classes of solutions that they discuss, the most widely employed is the IV approach (Aghion et al. 2005, Qian 2008, Novak and Stern 2009, Ataman, Van Heerde, and Mela 2010, Van Heerde et al. 2013, Li and Ansari 2014). Rossi (2014) surveyed a decade of publications in *Marketing Science* and *Quantitative Marketing and Economics* and revealed that the most commonly used IVs are lagged variables, costs, fixed effects, and Hausman-style variables from other markets. However, the survey found that the IVs' strengths are rarely measured/reported, despite the necessity of doing so to detect weak IVs. Moreover, the exclusion restriction condition cannot generally be tested to verify the IVs' validity. The survey also found that most papers lack a discussion of why the IVs used are valid. In short, although IVs have a sound theoretical grounding, good ones are difficult to find, making the IV approach difficult to implement in practice.

The second class of solutions for mitigating endogeneity involves specifying the economic structure that generates the observed data on endogenous regressors (e.g., a supply-side model for marketing-mix variables) (Chintagunta et al. 2006, Sudhir 2001, Yang, Chen, and Allenby 2003, Sun 2005, Dotson and Allenby 2010 and Otter, Gilbride, and Allenby 2011). A key concern with this approach is that incorrect assumptions or insufficient information on the supply side may generate biased estimates (Chintagunta et al. 2006).

The third class of solutions to endogeneity correction involves IV-free methods, a more recent methodological development. Ebbes, Wedel, and Böckenholt (2009) discussed three extant IV-free approaches: the higher moments (HM) approach (Lewbel 1997), the identification through heteroscedasticity (IH) estimator (Rigobon 2003), and the latent instrumental variables (LIV) method (Ebbes et al. 2005). Wang and Blei (2019) recently proposed a deconfounder approach that has some flavor of the LIV approach. All these methods decompose an endogenous regressor into an exogenous and endogenous part. The assumption that the endogenous regressor contains an exogenous component that does not directly affect the outcome is akin to the stringent condition of exclusion restriction for observed IVs and thus can be difficult to justify. Park and Gupta (2012) introduced another IV-free method that does not require the stringent condition of exclusion restriction but directly models the dependence between the structural error and the endogenous regressor via copula. Researchers have enthusiastically adopted the copula method owing to its feasibility without the need for instruments (Becker, Proksch, and Ringle 2021; Haschka 2022; Eckert and Hohberger 2022; Qian, Xie, and Koschmann 2022; Datta, Foubert, and Van Heerde 2015; Heitmann et al. 2020; Atefi et al. 2018; Elshiewy and Boztug 2018). The 2sCOPE method contributes to the field by overcoming significant limitations of existing copula correction methods and by virtue of its broader applicability (Table 1).

The contributions of 2sCOPE are threefold. *First*, to our knowledge, this work is novel in providing formal proofs for copula correction methods' theoretical properties. These theoretical results are necessary given that model identifiability is central to addressing the endogeneity issue. Recent work has noted the lack of rigorous proofs of required model identification conditions and estimation properties (consistency and efficiency) for copula correction as a major area requiring further research (Becker, Proksch, and Ringle 2021; Haschka 2022)[2]. The theoretical results presented herein contribute to bridging this gap and facilitate a better understanding of the properties of the copula correction methods and their optimal use.

Table 1: A Comparison of Copula Correction Methods

Features	Park and Gupta (2012)	Haschka (2022)	2sCOPE
Nonnormality of endogenous regressors[1]	Required	Required	Not required[2]
Handle correlated exogenous regressors	NO	YES	YES
Intercept included	YES	NO[3]	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 nonnormality of endogenous regressors can also cause poor finite-sample performance (finite-sample bias and large standard errors) and require extremely large sample sizes to perform well.
[2]: Nonnormality of endogenous regressors is not required provided at least one correlated exogenous regressor is not normally distributed.
[3]: The approach cannot estimate the intercept term, which is removed from the panel model prior to estimation using first-difference or fixed effects transformation (Web Appendix A.8 of Haschka (2022)). Becker, Proksch, and Ringle (2021) shows the importance of including intercept in marketing applications.
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

A useful theoretical outcome is that the existence of the correlations between endogenous and exogenous regressors alone does not automatically introduce bias to $\text{Copula}_{\text{Origin}}$. Rather, we demonstrate that the implicit assumption for $\text{Copula}_{\text{Origin}}$ is the exogenous re-

[2]For instance, owing to the complex form of the estimation method, Haschka (2022) notes the lack of theoretical proofs of required model identification conditions and estimation consistency as one limitation of the copula correction method developed there, and thus must rely solely on simulation studies to evaluate its empirical properties.

gressors' uncorrelatedness 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, particularly with multiple endogenous regressors.[3] We prove that the proposed 2sCOPE yields consistent causal effect estimates when the above implicit assumption is violated, which can cause biased causal effect estimates for $\text{Copula}_{\text{Origin}}$. Our other novel finding is that although the exogenous regressors that are correlated with the CCF require special handling for consistent causal effect estimation, they can be leveraged efficiently by 2sCOPE to substantially improve the identification and finite-sample performance of copula correction. Significantly, we prove that the structural model with normally distributed endogenous regressors can be identified using 2sCOPE provided one of the exogenous regressors correlated with endogenous ones is nonnormal, which is considerably more feasible in many practical applications.

Second, the proposed 2sCOPE method is the first copula-correction method that simultaneously relaxes the nonnormality assumption of endogenous regressors and handles correlated endogenous and exogenous regressors (Table 1). Existing copula correction methods do not account for correlated endogenous and exogenous regressors. An exception is Haschka (2022), who generalizes Park and Gupta (2012) to fixed effects linear panel models with correlated regressors by jointly modeling the structural error and endogenous and exogenous regressors using copulas and maximum likelihood estimation (MLE). However, as noted in Haschka (2022), Haschka's approach still requires the nonnormality of endogenous regressors. Thus, all existing copula correction methods require sufficient nonnormality assumption of endogenous regressors for model identification and acceptable finite sample performance (Haschka

[3]Although Haschka (2022) explains why correlated regressors can cause potential bias for $\text{Copula}_{\text{Origin}}$, no condition for the occurrence of bias is specified. 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 is reduced 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 (Danaher and Smith 2011).

2022; Becker, Proksch, and Ringle 2021; Eckert and Hohberger 2022). Becker, Proksch, and Ringle (2021) suggest a minimum absolute skewness of 2 for an endogenous regressor to ensure good performance of Gaussian copula correction methods in a sample under 1000 (Figure 8 in Becker, Proksch, and Ringle 2021). These requirements can significantly limit the practical deployment of copula correction methods.

Our proposed 2sCOPE method overcomes these important restrictions of existing copula correction methods. Consistent with our theoretical results, the evaluation in the simulation Cases 2 and 3 demonstrates the superior finite-sample performance of 2sCOPE and shows that 2sCOPE eliminates or substantially reduces the significant problem of finite-sample bias associated with insufficient regressor nonnormality raised in Becker, Proksch, and Ringle (2021) and Eckert and Hohberger (2022). Even when the endogenous regressor is normal or close-to-normal with a skewness of 0, the estimation bias of 2sCOPE remains negligible for a sample size as small as 200 (Figure 1). We further conduct systematic simulation studies and provide an actionable guideline for using 2sCOPE (Figure 2), establishing sufficient conditions regarding exogenous regressors that are verifiable using tests of nonnormality and relevance to endogenous regressors, to effectively handle endogenous regressors with insufficient nonnormality using data at hand. Overall, 2sCOPE can substantially broaden IV-free methods' practical applicability for handling endogeneity issues.

Third, the proposed 2sCOPE provides a versatile and feasible copula control function method for handling regressor endogeneity. Although the vast majority of applications using the copula correction method have employed the generated-regressor approach (Becker, Proksch, and Ringle 2021; Eckert and Hohberger 2022), no existing copula control function method can handle endogenous regressors that have insufficient nonnormality or correlated with exogenous regressors. 2sCOPE addresses this need and enjoys several benefits associated with the control function versus the alternative MLE approach. These include— but are not limited to—little extra computational and modeling burdens to be integrated with complex outcome models commonly used in marketing studies (Table 1), broader applicability with weaker assumptions, and increased robustness to model mis-specifications.[4]

In many such models, the MLE approach becomes considerably more difficult or computationally infeasible while 2sCOPE is straightforward. Footnote 8 offers an example demonstrating that extension of Haschka (2022)'s MLE approach to random coefficient linear panel models (RC-LPMs) with correlated endogenous and exogenous regressors requires the numerical evaluation of potentially high-dimensional integrals of complicated functions containing the product of copula density functions, evaluated at repeated measurement occasions. However, 2sCOPE involves none of these integrals and can be implemented using standard software programs for RC-LPMs, assuming all regressors are exogenous. Furthermore, although 2sCOPE assumes a normal error distribution, we show its robustness to symmetric nonnormal error distributions (Web Appendix E.4), in contrast to existing methods' sensitivity to such error mis-specifications (Becker, Proksch, and Ringle 2021). Thus, the 2sCOPE control function approach leveraging correlated exogenous regressors can enhance robustness to model mis-specifications.

METHODS

In this section, we develop a copula-based IV-free method for handling endogenous regressors with insufficient nonnormality or correlated with exogenous regressors. We first review $\text{Copula}_{\text{Origin}}$ and demonstrate that it implicitly assumes no correlations between exogenous regressors and the CCF as well as the bias in the structural model parameter estimates that may arise from the violation of this assumption. We then propose a new approach to the problem and the detailed estimation procedure. We also demonstrate how exogenous regressors correlated with endogenous regressors can sharpen structural model parameter estimates and enable the identification of the structural model containing normally distributed endogenous regressors, which are known to cause model non-identifiability for $\text{Copula}_{\text{Origin}}$.

[4] As Becker, Proksch, and Ringle (2021) demonstrate, the Gaussian copula control function approach is more robust against error term mis-specifications than the Gaussian copula MLE approach.

Assumptions of Existing Copula Endogeneity-Correction Method (Copula$_{Origin}$)

Consider the following linear structural regression model [5]:

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

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) continuous endogenous regressor (e.g., price); W_t is a $(k_W\times1)$ 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 causes the endogeneity problem. W_t is exogenous—i.e., it is not correlated with ξ_t, but can be correlated with the endogenous variable P_t. An example of W_t is weather when estimating the pricing effect on consumer demand of ice cream using historical sales data. Weather is an exogenous factor that can affect both the pricing and sales of ice cream: the inclusion of weather in the sales model for ice cream mitigates the concern of price endogeneity and improves the demand prediction accuracy.

The key idea of Copula$_{\text{Origin}}$ (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. This method has the advantage that marginals are not restricted by the joint distribution. Thus, the copula model allows researchers to construct a flexible multivariate joint distribution that captures the correlation among these 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 theorem (Sklar 1959), there exists a copula function $C(\cdot, \cdot)$ such that

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

where $U_p = H(P)$ and $U_\xi = G(\xi)$, and 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 these random variables' marginals and

[5] As shown in Becker, Proksch, and Ringle (2021), it is important to include the intercept term when evaluating the copula correction method.

correlations. To capture the association between the endogenous regressor P and the error ξ, Park and Gupta (2012) use the following Gaussian copula for its desirable properties (Danaher 2007; 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. With empirical marginal CDFs, the above Gaussian copula model depends on the rank order of raw data only, and is invariant to strictly monotonic transformations of variables in (P,ξ). Thus, the above Gaussian copula model is regarded as general and robust for most marketing applications (Danaher and Smith 2011). 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.

Let $P_t^* = \Phi^{-1}(U_p)$ and $\xi_t^* = \Phi^{-1}(U_\xi)$, the above Gaussian copula means $[P_t^*, \xi_t^*]'$ follow the standard bivariate normal distribution with the correlation coefficient ρ as follows:

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

Based on the assumption that the structural error ξ_t follows $N(0,\sigma_\xi^2)$, Park and Gupta (2012) demonstrated that the structural error can be divided 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 $\sigma_\xi \cdot \sqrt{1-\rho^2}\omega_t$ is a new independent error term. Equation (1) can then 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) suggested augmented ordinary least squares (OLS) estimation of Equation (6) with $P_t^* = \Phi^{-1}(U_p)$ included as an additional regressor to correct for the endogeneity of P_t. Park and Gupta (2012) further noted that for the above approach to work, P_t must have a nonnormal distribution. If 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 to identify the linear regression model in Equation (6).

Below, we demonstrate that an implicit assumption for the above generated regressor approach to yield consistent model estimates is the uncorrelatedness between P_t^* and W_t. A non-zero correlation between the exogenous regressor W_t and the generated regressor P_t^* would cause biased OLS estimates of Equation (6) using Copula$_{\text{Origin}}$ owing to the induced correlation between the error term ω_t and W_t, which Theorem 1 formally proves.

Theorem 1. *Assuming (1) $(1, P, W)$ is full rank and W is exogenous, (2) the error term is normal, (3) a Gaussian Copula for the structural error term and P_t, (4) P_t is endogenous: $\rho \neq 0$, and (5) P_t^* and W_t are correlated, $Cov(\omega_t, W_t) = -\frac{\rho}{\sqrt{1-\rho^2}} Cov(W_t, P_t^*) \neq 0$ and consequently, the OLS estimates of Equation (6) are inconsistent.*

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

To summarize, Copula$_{\text{Origin}}$ based on Equation (6) makes the set of assumptions listed in Table 2. Assumption 5 has been discovered by Haschka (2022). However, as Web Appendix A.2 details, Assumption 5 should be replaced with the more general Assumption 5(b) for multiple endogenous regressors.[6] Assumptions 5 and 5(b) are verifiable and provide users with criteria to determine whether Copula$_{\text{Origin}}$ will provide consistent estimation when exogenous regressors exist. With only one endogenous regressor, one can simply check the correlations between the copula transformation of this endogenous regressor and 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 Copula$_{\text{Origin}}$ and each exogenous regressor, using Fisher's Z test, as described in Web Appendix E.7. Where at least one exogenous regressor exists in W_t that fails Assumption 5 or 5(b), Copula$_{\text{Origin}}$ yields biased estimates, and our proposed 2sCOPE can be used, as derived in the next subsection.

[6] For instance, in the 2-endogenous regressors case, Assumption 5(b) means $Cov(W_t, \frac{\rho_{\xi 1} - \rho_{12}\rho_{\xi 2}}{1-\rho_{12}^2} \cdot P_{1,t}^* + \frac{\rho_{\xi 2} - \rho_{12}\rho_{\xi 1}}{1-\rho_{12}^2} \cdot P_{2,t}^*) = 0$ (Web Appendix A.2), which is not the same as either $Cov(W_t, P_{1,t}^*) = 0$, $Cov(W_t, P_{2,t}^*) = 0$ or $Cov(W_t, P_{1,t}) = 0, Cov(W_t, P_{2,t}) = 0$.

Table 2: Assumptions in Copula$_{\text{Origin}}$

Assumption 1.	*Full rank* [1] *of all regressors and exogeneity of* W [2]*.*
Assumption 2.	*The structural error follows a normal distribution.*
Assumption 3.	P_t *and the structural error follow a Gaussian copula.*
Assumption 4.	*Nonnormality of the endogenous regressor* P_t*.*
Assumption 5.	*For a scalar endogenous regressor* P_t*,* W_t *and* P_t^* *are uncorrelated.*
Assumption 5(b).	*For multiple endogenous regressors,* W_t *and the CCF* [3] *are uncorrelated.*

[1]: Full rank means $\text{rank}(X^{'}X) = k$, in which $X = (1, P, W)$ with column dimension of k;
[2]: Exogeneity of W means $Cov(W, \xi) = 0$. When P and W are uncorrelated, the exogeneity assumption of W can be relaxed if the interest is only on α, the coefficient of P (Web Appendix E.11), but P still must be nonnormally distributed to achieve identification.
[3]: CCF (copula control function) is the linear combination of P_t^* used to control for endogenous regressors.

The full rank of all regressors and exogeneity of W_t in Assumption 1 of Table 2 are assumptions that are made in several other commonly used econometric methods, including OLS and IV methods, to ensure estimation consistency. For Assumptions 2 to 4, Park and Gupta (2012) have demonstrated their copula method's reasonable robustness to nonnormal distributions of the structural error (Assumption 2) and alternative forms of copula functions (Assumption 3), although it is not surprising to observe Copula$_{\text{Origin}}$'s sensitivity to gross violations of these assumptions, such as highly skewed error distributions (Becker, Proksch, and Ringle 2021; Eckert and Hohberger 2022). By contrast, the assumption that the endogenous regressor P_t follows a nonnormal distribution (Assumption 4) is critical. An endogenous regressor following a normal distribution violates the full-rank condition in Equation (6) and causes model unidentification regardless of the 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 perform poorly for a finite sample size. Both Assumptions 4 and 5 (or 5(b)) can be too strong and substantially limit the copula method's applicability.

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

Here, we propose a two-stage copula endogeneity-correction (2sCOPE) method and demonstrate its ability to relax both the uncorrelatedness assumption between CCF and the exogenous regressors (Assumption 5(b)) and the key identification assumption of nonnormal endogenous regressors (Assumption 4). 2sCOPE jointly models the endogenous regressor, P_t, the correlated exogenous variable, W_t, and the structural error term, ξ_t, using the Gaussian copula model, which 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.

Under the above Gaussian copula model in Equation (7), one may develop a direct extension of $\text{Copula}_{\text{Origin}}$, that adds the generated regressors P_t^* and W_t^* into the structural regression model to correct for endogeneity bias. The resulting method —COPE (COPula Endogeneity correction)— is shown to yield consistent causal effect estimates without requiring the exogeneity of W and Assumption 5 (or Assumption 5(b)) that are needed for $\text{Copula}_{\text{Origin}}$ (Web Appendix A.1). However, COPE requires that endogenous regressors P_t and exogenous regressors W_t both have sufficient nonnormality and yields substantial bias when regressors have insufficient nonnormality (see simulation results in Table 6 and Figure 1). Furthermore, the addition of numerous generated regressors for control variables W can cause severe multicollinearity issues and exert significantly adverse impacts on causal effect estimation efficiency and stability (the simulation results in Web Appendix E.3 demonstrate that COPE can require a sample size five times greater than our proposed method to achieve the same estimation precision). Overall, COPE suffers from the low face-validity problem— i.e., it requires generated regressors for exogenous regressors even though, conceptually, only generated regressors for endogenous regressors are needed to control for the endogeneity problem. To overcome these limitations, we have derived the 2sCOPE method, which relaxes both Assumptions 4 and 5(b) of $\text{Copula}_{\text{Origin}}$ below.

Under the Gaussian copula model in Equation (7), we have the following system of equations:

$$Y_t = \mu + P_t\alpha + W_t\beta + \xi_t \tag{8}$$

$$P_t^* = W_t^*\gamma + \epsilon_t. \tag{9}$$

Assuming that ξ_t follows a normal distribution, ϵ_t and ξ_t follow a bivariate normal distribution, since they are a linear combination of tri-normal variate (ξ_t^*, P_t^*, W_t^*) under the Gaussian copula assumption. Equation (9) expresses the copula transformation of the endogenous regressor, determined by the rank order of P_t, as a linear combination of observed and unobserved variables. The two error terms ϵ_t and ξ_t are correlated owing to the endogeneity of P_t. For example, both ξ_t and ϵ_t may contain an additive component corresponding to a common omitted variable. The above model is then obtained when the omitted variable and regressors follow a Gaussian copula model.

The main idea of 2sCOPE is to make use of the fact that, by conditioning on ϵ_t, the structural error ξ_t becomes independent of both P_t and W_t. That is, by conditioning on the component of P_t that causes its endogeneity (here, ϵ_t), the structural error is not correlated with either P_t or W_t, thereby ensuring the consistency of standard estimation methods. In this sense, ϵ_t serves as a (scaled) control function to address the endogeneity bias. To demonstrate this point, we rewrite the Gaussian copula model in Equation (7) as

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

$$\begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \\ \omega_{3,t} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right). \tag{10}$$

Given the above joint normal distribution for (P_t^*, W_t^*, ξ_t^*) and $\xi_t = \sigma_\xi \xi_t^*$, we have

$$P_t^* \quad = \quad \rho_{pw} W_t^* + \epsilon_t, \tag{11}$$

and

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

$$= \mu + P_t\alpha + W_t\beta + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2}\epsilon_t + \sigma_\xi\sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2\rho_{p\xi}^2}{1-\rho_{pw}^2}} \cdot \omega_{3,t}. \tag{12}$$

Equation (12) suggests adding the estimate of the error term ϵ_t from the first stage regression as a generated regressor instead of adding P_t^* and W_t^*. As Theorem 2 below illustrates, the linear model in Equation (12) satisfies both the full column rank condition of the regressor matrix and zero correlation between the new error term $\omega_{3,t}$ and each regressor in Equation (12), ensuring the consistency of OLS estimates (Wooldridge 2010, Ch. 4). This two-step procedure— termed as 2sCOPE— adds the first-stage residual $\widehat{\epsilon}_t$ to control for endogeneity and, in this respect, is similar to the control function approach adopted by Petrin and Train (2010). However, unlike Petrin and Train (2010), 2sCOPE requires no use of IVs. When P and W are uncorrelated (i.e., $\rho_{pw} = 0$), the generated regressor ϵ_t becomes P_t^* in Equation (12) and, consequently, 2sCOPE is reduced to $\text{Copula}_{\text{Origin}}$. 2sCOPE is more general than $\text{Copula}_{\text{Origin}}$ by permitting nonzero values of ρ_{pw}.

Theorem 2. ***Consistency of the 2sCOPE Estimator.*** *Assuming (1)* $(1, P, W)$ *is full rank and* W *is exogenous, (2) the error is normal, (3) either the endogenous regressor* P_t *or one correlated regressor in* W_t *is nonnormal, and (4) a Gaussian Copula for* (ξ_t, P_t, W_t)*, 2sCOPE estimator is consistent.*

Proof: See Web Appendix B.1, Proof of Theorem 2.

According to Theorem 2, the proposed method 2sCOPE can yield consistent estimates when assumptions are met. Specifically, Assumptions 5 and 5(b) are relaxed because 2sCOPE can handle the case when the model includes exogenous regressors correlated with the CCF. Theorem 3 further demonstrates that 2sCOPE relaxes Assumption 4 (the nonnormality assumption on endogenous regressors), a critical model identification condition required in all other copula correction methods.

Theorem 3. ***Nonnormality Assumption Relaxed.*** *Assuming (1)* $(1, P, W)$ *is full rank and* W *is exogenous, (2) the error term is normal, (3) one of the correlated exogenous regressors* W_t *is nonnormal, and (4) a Gaussian Copula for the error term,* P_t *and* W_t*, 2sCOPE estimator is consistent when* P_t *follows a normal distribution.*

Proof: See Web Appendix B.2, Proof of Theorem 3.

Theorem 3 shows that as long as one exogenous regressor correlated with the endogenous regressor P_t is nonnormally distributed, 2sCOPE can correct for endogeneity for a normal regressor P_t while COPE cannot. Intuitively, when P_t (or W_t) is normal, P_t^* (or W_t^*) becomes a linear function of P_t (or W_t) under the Gaussian copula assumption, causing COPE to fail the full rank assumption and become unidentified. As such, COPE cannot deal with normal endogenous/exogenous regressors. For 2sCOPE in Equation (12), adding the first stage residual $\widehat{\epsilon}_t$ as the generated regressor improves model identification. Provided not all W_t are normal, $\widehat{\epsilon}_t$ would not be a linear function of P_t and W_t and thus the second stage model (Equation 12) in 2sCOPE would satisfy the full rank requirement for model identification. Thus, 2sCOPE can relax the nonnormality assumption on the endogenous regressor required by the method of Park and Gupta (2012) provided one of the W_t is nonnormally distributed.

Table 3: Assumptions in 2sCOPE

Assumption 1. *Full rank* [1] *of all regressors and exogeneity of* W [2].
Assumption 2. *The structural error follows a normal distribution.*
Assumption 3. P_t, W_t *and the structural error follow a Gaussian copula.*
Assumption 4. *Either* P_t *or one related regressor in* W_t *is nonnormally distributed.*

[1]: Full rank means rank$(X^{'}X) = k$, in which $X = (1, P, W)$ with column dimension of k;
[2]: Exogeneity of W means $Cov(W, \xi) = 0$. When P and W are uncorrelated, the exogeneity assumption of W can be relaxed if the interest is only on α, the coefficient of P (Web Appendix E.11), but P must be nonnormally distributed to achieve identification as, in this special case, 2sCOPE reverts to Copula$_{\text{Origin}}$ (see Note 2 under Table 2).

Table 3 summarizes the assumptions used in the proof of the properties of 2sCOPE. Among these assumptions, Assumption 4 and the full rank of the regressor matrix in Assumption 1 are required and verifiable. We highlight the remaining assumptions that may be difficult to verify given that they involve the unobserved error term. W's exogeneity is required when P and W are correlated, and this assumption should be evaluated based on institutional knowledge or economic theory. Although Assumptions 2 and 3 are used in the proof, our simulation study demonstrates that 2sCOPE is robust to a range of non-normal error distributions and reasonable departures from the Gaussian copula dependence model (Web Appendix E.4, E.5, and E.6). Furthermore, it is often reasonable to assume that the error term ξ may be expressed as $\xi = U + V$, where the normally distributed U stands for the joint effect of confounders (a linear combination of confounders) and V represents an independent Gaussian noise term. When U and regressors jointly follow the Gaussian copula model, Assumption 3 holds. The assumption that U and regressors follow a Gaussian copula dependence model appears plausible given that the Gaussian copula is a widely used dependence model that is deemed sufficiently flexible to adequately capture multivariate dependence in many practical applications (Danaher and Smith 2011; Eckert and Hohberger 2022). Meanwhile, we emphasize that the Gaussian copula assumption is generally untestable and warrants attention from those who employ copula methods.

In sum, we have demonstrated the consistency of 2sCOPE (Theorem 2). Theorem 3 and Proposition 1 (Web Appendix B.3) further establish that 2sCOPE outperforms COPE, the extended $\text{Copula}_{\text{Origin}}$, with respect to estimation efficiency and relaxation of the nonnormality assumption on endogenous regressors in $\text{Copula}_{\text{Origin}}$ by satisfying a looser condition.

Multiple Endogenous Regressors

Herein, we extend 2sCOPE 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:

$$Y_t = \mu + P_{1,t} \cdot \alpha_1 + P_{2,t} \cdot \alpha_2 + W_t\beta + \xi_t. \tag{13}$$

Under the multivariate Gaussian distribution assumption on $(\xi_t, P^*_{1,t}, P^*_{2,t}, W^*_t)$, the system of equations for the 2sCOPE method in Equations (8, 9) are readily extended to

$$Y_t = \mu + P_{1,t}\alpha_1 + P_{2,t}\alpha_2 + W_t\beta + \xi_t, \tag{14}$$

$$P^*_{1,t} = \rho_{wp1} W^*_t + \epsilon_{1,t}, \tag{15}$$

$$P^*_{2,t} = \rho_{wp2} W^*_t + \epsilon_{2,t}, \tag{16}$$

where $(\xi_t, \epsilon_{1,t} . \epsilon_{2,t})$ are linear transformations of $(\xi_t, P^*_{1,t}, P^*_{2,t}, W^*_t)$, and thus also follow a multivariate Gaussian distribution. Consequently, we can decompose the structural error ξ_t

as additive terms for $\epsilon_{1,t}$, $\epsilon_{2,t}$ and the remaining independent error term $\omega_{4,t}$ as follows

$$Y_t = \mu + P_{1,t}\alpha_1 + P_{2,t}\alpha_2 + W_t\beta + \eta_1\epsilon_{1,t} + \eta_2\epsilon_{2,t} + \sigma_\xi \cdot m \cdot \omega_{4,t}, \tag{17}$$

where $\epsilon_{1,t} = P^*_{1,t} - \rho_{wp1}W^*_t$ and $\epsilon_{2,t} = P^*_{2,t} - \rho_{wp2}W^*_t$; m is a constant that depends solely on the correlation coefficients in the Gaussian copula (Web Appendix B.1). The new (scaled) error term $\omega_{4,t}$ is independent of latent copula data $(P^*_{1,t}, P^*_{2,t}, W^*_t)$ and all functions of these latent data including $P_{1,t}, P_{2,t}, W_t, \epsilon_{1,t}, \epsilon_{2,t}$. Because $\omega_{4,t}$ is independent of all regressors in Equation (17), the OLS estimation of Equation (17) yields consistent estimates of structural model parameters provided the regressor matrix $(1, P_1, P_2, W, \epsilon_1, \epsilon_2)$ is of full column rank.

Web Appendix B presents the proof for the estimation consistency, relaxation of the regressor-nonnormality assumption, and estimation efficiency gain for 2sCOPE with multiple endogenous regressors under the related Theorems 2, 3, and Proposition 1. Table 4 summarizes the estimation procedure of 2sCOPE.

Table 4: Estimation Procedure for 2sCOPE

Stage 1:

- 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.

Note: Standard errors of parameter estimates are estimated using Bootstrap (Web Appendix F).

2sCOPE for Random Coefficient Linear Panel Models

We consider the following random coefficient model for linear panel data

$$Y_{it}|\mu_i, \alpha_i, \beta_i = \bar{\mu} + \mu_i + P'_{it}\alpha_i + W'_{it}\beta_i + \xi_{it}, \tag{18}$$

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} may be correlated. The error term $\xi_{it} \overset{iid}{\sim} N(0, \sigma^2_\xi)$, which is correlated with P_{it} owing 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 attested the ubiquitous presence of heterogeneous consumers' responses to marketing mix variables (e.g., price sensitivity) and substantial bias associated with ignoring such slope heterogeneity. Thus, it is important to allow individual-specific slope coefficients in marketing studies.

The linear panel data model as specified in Equation (18) is general and includes the linear panel model with only the individual-specific intercepts considered by Haschka (2022) as a special case. Specifically, Haschka (2022) fixes (α_i, β_i) to be the same value (α, β) across all units, assuming that all cross-sectional units have the same slope coefficients. By contrast, the model in Equation (18) relaxes this strong assumption and can generate unit-specific slope parameters that may be used for targeting purposes.

A random coefficient model typically assumes that $(\mu_i, \alpha_i, \beta_i)$ follows a multivariate normal distribution. When all regressors are exogenous, estimation algorithms for such random coefficient models are well-established and computationally feasible, even for a high-dimensional vector of random effects $(\mu_i, \alpha_i, \beta_i)$. With the normal conditional distribution for $Y_{it}|(\mu_i, \alpha_i, \beta_i)$ in Equation (18) and the multivariate normal prior distribution for random effects $(\mu_i, \alpha_i, \beta_i)$, Y_{it} marginally follows a normal distribution with a closed-form expression that contains no integrals with respect to random effects $(\mu_i, \alpha_i, \beta_i)$, leading to an easy-to-evaluate likelihood function (Wooldridge 2010). Alternatively, one may assume a mixed-effect model where μ_i is a fixed-effect parameter and can be correlated with the regressors P_{it} and W_{it}. In this case, the first-difference or fixed-effects transformation is typically used to eliminate the incidental intercept parameters as follows

$$\tilde{y}_{it}|\alpha_i, \beta_i \quad = \quad \tilde{P}'_{it}\alpha_i + \tilde{W}'_{it}\beta_i + \tilde{\xi}_{it}, \tag{19}$$

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 (2022) considered a special case of Equation (19) by fixing (α_i, β_i) as homogeneous (i.e., not varied by i).

It is straightforward to apply 2sCOPE to address regressor endogeneity in the general random coefficient model for linear panel data in Equation (18) and the transformed model without intercepts in Equation (19).[7] The 2sCOPE procedure adds the residuals obtained from regressing P_{it}^* on W_{it}^*. 2sCOPE may thus be implemented using standard software programs for random coefficient linear panel models assuming that all regressors are exogenous (see Simulation Study Case 4 for an illustration using the R function `lme`()). By contrast, the MLE approach for copula correction in the random coefficients model accounting for correlated endogenous and exogenous regressors has yet to be developed and would require the construction of 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.[8]

2sCOPE for Slope Endogeneity and Random Coefficient Logit Model

In Web Appendices C and D, we derive 2sCOPE to tackle slope endogeneity and address endogeneity bias in random coefficient logit models with correlated and normally distributed regressors. Unlike the MLE approach, 2sCOPE can be implemented using standard estimation methods through the addition of generated regressors to control for endogeneity.

SIMULATION STUDY

Here, we conduct Monte Carlo studies: (a) to assess the the proposed method's performance for correlated regressors, (b) to assess the proposed method's performance under regressor normality and near normality, (c) to assess the proposed method's performance under various types of structural models, and (d) to assess the proposed method's robustness to violations of model assumptions. We measure the estimation bias using t_{bias} calculated as the ratio of

[7]Similar to Haschka (2022), a GLS transformation can be applied to both sides of Equation (19), resulting in a pooled regression for which 2sCOPE can be directly applied.

[8]With endogenous regressors, the individual random effects parameters enter into both the density function for the outcome $Y_{it}|(\mu_i, \alpha_i, \beta_i)$ and the density of copula function $C(U_{\xi,it}, U_{P,it}, U_{W,it})$ via $U_{\xi,it}$, and thus cannot be integrated out in closed-form any more from the likelihood function even with the normal structural error term and normal random effects. Therefore, numerical integration is required for obtaining MLEs in random coefficient models with endogenous regressors, which cannot be performed with standard software programs for random coefficient model estimation.

the absolute difference between the mean of the sampling distribution and the true parameter value to the standard error of the parameter estimate (Park and Gupta 2012). Thus, t_{bias} represents the size of bias relative to the sampling error.

Case 1: Nonnormal Regressors

In the first case, P and W are correlated. The data generating process (DGP) is summarized below:

$$\begin{pmatrix} P_t^* \\ W_t^* \\ \xi_t^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{pw} & \rho_{p\xi} \\ \rho_{pw} & 1 & 0 \\ \rho_{p\xi} & 0 & 1 \end{bmatrix}\right) = N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix}\right), \tag{20}$$

$$\xi_t = G^{-1}(U_{\xi,t}) = G^{-1}(\Phi(\xi_t^*)) = \Phi^{-1}(\Phi(\xi^*)) = 1 \cdot \xi_t^*, \tag{21}$$

$$P_t = H^{-1}(U_{P,t}) = H^{-1}(\Phi(P_t^*)), \quad W_t = L^{-1}(U_{W,t}) = L^{-1}(\Phi(W_t^*)), \tag{22}$$

$$Y_t = \mu + \alpha \cdot P_t + \beta \cdot W_t + \xi_t = 1 + 1 \cdot P_t + (-1) \cdot W_t + \xi_t, \tag{23}$$

where ξ_t^* and P_t^* are correlated ($\rho_{p\xi} = 0.5$), generating the endogeneity problem; W_t^* is exogenous and uncorrelated with ξ_t^*; W_t^* and P_t^* are correlated ($\rho_{pw} = 0.5$), and therefore W_t and P_t are correlated. We consider four different estimation methods: (1) OLS regression of Equation (23); (2) $\text{Copula}_{\text{Origin}}$, which is the OLS regression of Equation (6); (3) the extended method COPE, which is the OLS regression of Equation $Y_t = \mu + P_t\alpha + W_t\beta + P_t^*\eta_1 + W_t^*\eta_2 + e_t$; and (4) the proposed 2sCOPE, which is the OLS regression of Equation (12). The generated regressors P_t^*, W_t^* and ϵ_t are estimated according to Table 4. We set the sample size $T =$ 1000 and generate 1000 datasets as replicates using the above DGP. 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 datasets, providing the empirical distributions of parameter estimates.

Table 5 reports estimation results. As expected, OLS estimates of both α and β are biased ($t_{bias} = 15.75$ and 8.24) due to the regressor endogeneity. $\text{Copula}_{\text{Origin}}$ reduces the bias, but still shows significant bias for the coefficient estimates of P_t and W_t. The bias of $\text{Copula}_{\text{Origin}}$ depends on the strength of the correlation between W and P. Stronger correlations between

Table 5: Results of the Simulation Study Case 1: Nonnormal Regressors

			OLS			$\text{Copula}_{\text{Origin}}$			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 across all 1,000 simulated samples.

P^* and W^* may cause the $\text{Copula}_{\text{Origin}}$ estimates to show a larger bias. For example, when the correlation between W^* and P^* increases from 0.5 to 0.7, the mean α estimate changes from 1.055 to 1.260 (Table 5 under the column "$\text{Copula}_{\text{Origin}}$") and consequently, the bias of estimated α increases by around five times (from 0.055 to 0.260). The bias confirms our derivation in the model section, demonstrating that using the existing copula method may not wholly resolve the endogeneity problem with correlated regressors.

The proposed 2sCOPE method provides consistent estimates without using instruments. The average estimate of $\rho_{p\xi}$ is close to the true value of 0.5 and differs significantly from 0, implying regressor endogeneity detected correctly using 2sCOPE. Moreover, 2sCOPE shows greater estimation 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 COPE's and 2sCOPE's estimation precision using the D-error measure $|\Sigma|^{1/K}$ (Arora and Huber 2001; Qian and Xie 2022), where Σ is the covariance matrix of the regression coefficient estimates and K is the number of explanatory variables in the structural model. A smaller D-error indicates greater estimation efficiency and improved estimation precision. When $\rho_{pw} = 0.5$, the D-error measure is 0.002613 for COPE and 0.001614 for 2sCOPE (Table 5). As such, 2sCOPE increases estimation precision by 38.2%, meaning that for 2sCOPE to achieve the same precision as COPE, the sample size can be reduced by 38.2%. A 39.3% efficiency gain for 2sCOPE is observed for $\rho_{pw} = 0.7$ (Table 5).

We perform a further simulation study for a small sample size using the same DGP as that described above but with the sample size T=200. The results presented in Web Appendix E Table W1 demonstrate that OLS estimates have endogeneity bias and Copula$_{\text{Origin}}$ reduces the endogeneity bias but that significant bias remains nonetheless. The proposed 2sCOPE performs well, yielding unbiased estimates for the small sample size T=200. The efficiency gain of 2sCOPE relative to COPE appears to be greater for a smaller sample size. For example, when the correlation between P^* and W^* is 0.5, the D-error measures are 0.0166 and 0.0091 for COPE and 2sCOPE (Web Appendix Table W1), respectively, meaning that 2sCOPE increases estimation precision by 1-0.0091/.0166=46% compared with COPE. Thus, sample size can be reduced by almost half (∼50%) for 2sCOPE to achieve the same estimation precision as that achieved by COPE.

Case 2: Normal Regressors

Next, we examine a case in which the endogenous regressor and (or) the correlated exogenous regressor are normally distributed. This case is of particular interest given that normality is not allowed for endogenous regressors in Park and Gupta (2012). We use the same DGP as that described in Equations (20) to (23), with the exception that the marginal CDFs for regressors, $H(\cdot)$ and $L(\cdot)$, are selected based on the distributions listed in the first two columns of Table 6, which summarizes the estimation results. As expected, the OLS estimates are biased. Copula$_{\text{Origin}}$ produces biased estimates whenever the endogenous regressor P follows a normal distribution. Its estimates are also biased when P follows a gamma distribution (first row of Table 6) for a different reason: P and W are correlated. Similar to Copula$_{\text{Origin}}$, the COPE estimators in all three scenarios are biased when either P_t or 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. 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 the model non-identification problem whenever at least one of these regressors is normally distributed.

Table 6: Results of Case 2: Normal Regressors

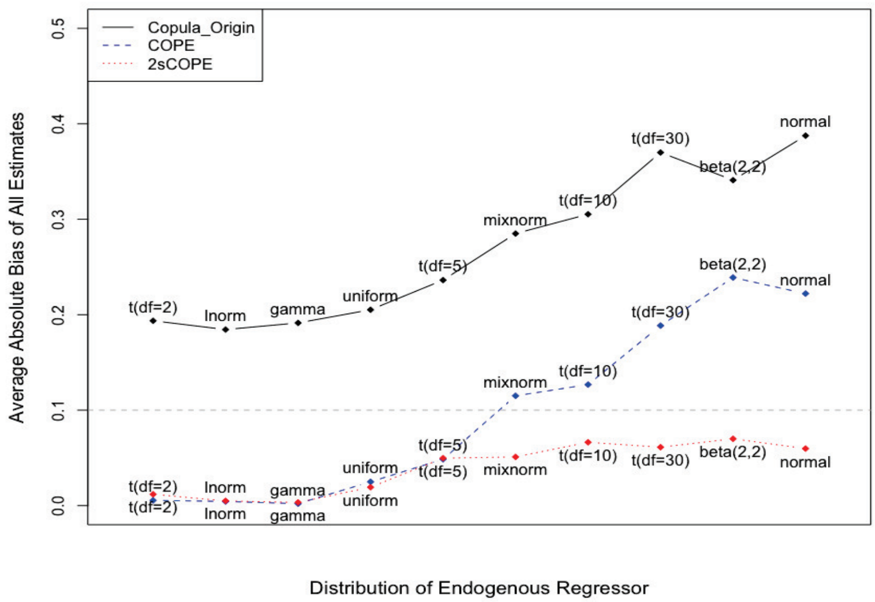
Distribution				OLS			Copula$_{\text{Origin}}$			COPE			2sCOPE		
P	W	Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
Gamma	Normal	μ	1	0.431	0.045	12.63	1.018	0.078	0.227	1.017	0.080	0.217	1.015	0.077	0.190
		α	1	1.569	0.037	15.40	0.979	0.070	0.302	0.979	0.070	0.296	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 Copula$_{\text{Origin}}$ and COPE, 2sCOPE adds the residual term obtained from regressing P_t^* on W_t^* as the generated regressor. Therefore, provided 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 6), 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 proposed 2sCOPE's ability to relax the nonnormality assumption in $\text{Copula}_{\text{Origin}}$ provided either P_t or W_t is nonnormally distributed.

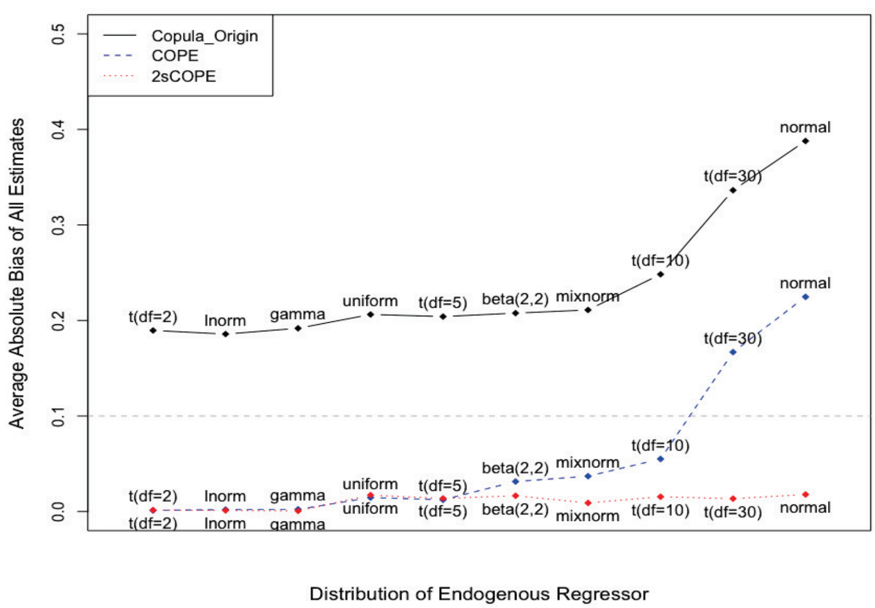
Case 3: Insufficient Nonnormality of Endogenous Regressors

The above case illustrates that the proposed 2sCOPE's capability to deal with normal endogenous regressors while $\text{Copula}_{\text{Origin}}$ and COPE cannot. Here, we examine the more frequent scenario of close-to-normal regressors. Although models are identified asymptotically (i.e., infinite sample size), appreciable finite-sample bias may occur with the realistic sample sizes commonly seen in marketing studies, if the endogenous regressor is too close to a normal distribution (Becker, Proksch, and Ringle 2021; Haschka 2022; Eckert and Hohberger 2022). Becker, Proksch, and Ringle (2021) suggest a minimum absolute skewness of 2 for an endogenous regressor for $\text{Copula}_{\text{Origin}}$ to have good performance in sample sizes less than 1000. This requirement can significantly limit the use of copula correction methods in practical applications. Given that 2sCOPE can handle normal endogenous regressors, we expect that 2sCOPE will be better able to handle the finite-sample bias caused by insufficient regressor nonnormality than the existing copula correction methods. We use the DGP described in Equations (20) to (23) to generate data, with the exception that the marginal CDF for the endogenous regressor ($H(\cdot)$) varies across some common distributions with varying closeness to normality. Specifically, we consider uniform, log normal, t, mixture normal, gamma, beta and normal distributions, and use the average absolute estimation bias of all the regression parameters (μ, α, β) in the structural model to measure the performance.

Figure 1 plots the estimation bias with different distributions of the endogenous regressor P. The results reveal that the $\text{Copula}_{\text{Origin}}$ estimates are biased with correlated endogenous and exogenous regressors, consistent with our theoretical proof (Theorem 1). COPE performs well when P has sufficient nonnormality ($t(2)$, log normal, gamma) and exhibits no bias even for a sample size as small as 200. However, COPE cannot handle a normal endogenous



(a) Sample Size N=200



(b) Sample Size N=1000

Figure 1: Average absolute estimation bias of all regression parameters (μ, α, β) in the structural model for different endogenous regressor distributions.

Note: 'lnorm' is lognormal(0,1), 'mixnorm' is N(-1,1) with the probability 0.5 and N(1,1) with the probability 0.5, 'uniform' is U[0,1], and 'gamma' is $Gamma(1, 1)$.

regressor and yields a large estimation bias that remains unchanged as the sample size increases, consistent with our theoretical proof in Theorem 3 (Web Appendix B.2) and the simulation result in Case 2. Furthermore, COPE suffers from finite-sample bias when the endogenous regressor P has distributions with insufficient nonnormality (e.g., beta(2,2), $t(\mathrm{df} = 30)$). Moreover, COPE's estimation bias is larger when the sample size is smaller or the distribution of the endogenous regressor P is closer to normal. For instance, t distribution with a degree of freedom of 30 is closer to normal than the t distribution with 10, 5, and 2 degrees of freedom, yielding a larger estimation bias. For t (df $= 30$) which is very close to normal, increasing the sample size from T=200 to 1000 scarcely changes the size of the estimation bias. By contrast, our proposed 2sCOPE method yields consistent estimates for all normal and close-to-normal regressor distributions and has negligible finite-sample bias even for a sample size as small as 200 (bias $<$ 5% of parameter values).

Case 4: Random Coefficient Linear Panel Model

We investigate the performance of 2sCOPE in the random coefficient linear panel model. We use the copula function and marginal distributions of $[P_{it}, W_{it}, \xi_{it}]$ as specified in Case 1 (Equations 20-22). 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). Marketing studies frequently use such random coefficient models to capture individual heterogeneity and to profile and target individuals. The correlation between ξ_{it} and P_{it} creates the regressor endogeneity problem, which can cause biased estimates for standard linear random coefficient estimation methods that ignore the regressor-error correlation. We generate individual-level panel data as described above 1000 times and use the data for estimation. Table 7 presents the estimation results. LME is the standard estimation method for linear mixed models assuming all regressors are exogenous, as implemented in the R function `lme()`. LME and $\text{Copula}_{\text{Origin}}$ are biased due to endogeneity and correlated exogenous regressors, respectively. Our proposed method, 2sCOPE, provides unbiased estimates that are tightly distributed around all parameters' true values.

Table 7: Results of the Simulation Study Case 4: Random Coefficient Linear Panel Model

		LME			$\text{Copula}_{\text{Origin}}$			2sCOPE		
Parameters	True	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	1.004	0.048	0.091
$\bar{\alpha}$	1	1.853	0.045	18.83	1.293	0.045	6.469	1.000	0.046	0.008
$\bar{\beta}$	-1	-1.557	0.045	12.39	-1.598	0.044	13.56	-1.000	0.044	0.005
σ_μ	1	0.985	0.033	0.459	0.982	0.033	0.547	0.984	0.031	0.522
σ_α	1	0.988	0.036	0.326	0.987	0.034	0.397	0.989	0.035	0.316
σ_β	1	0.993	0.031	0.235	0.992	0.033	0.249	0.992	0.033	0.248
$\rho_{p\xi}$	0.5	-	-	-	0.646	0.009	16.33	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	1.640

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

Additional Simulation Results and Robustness Checks

Web Appendix E provides additional simulation results on a small sample size (E.1); model estimation with multiple endogenous regressors (E.2); estimation with multiple exogenous control covariates including binary and close-to-normal control covariates (E.3); 2sCOPE's robustness to mis-specifications of the structural error distribution (E.4) and to mis-specifications of the copula dependence structure (E.5 & E.6); a test of Assumption 5(b) (E.7); experimental studies to obtain practical recommendations for using 2sCOPE (E.8); the performance of 2sCOPE with one 'strongly nonnormal' exogenous regressor vs. multiple 'weakly nonnormal' exogenous regressors for handling an endogenous regressor with insufficient nonnormality (E.9); the random coefficient logit model using 2sCOPE (E.10); 2sCOPE's performance when W is endogenous (E.11); and 2sCOPE's ability to leverage the empirical correlation between P and W(E.12). Overall, these results verify that 2sCOPE is robust to small sample sizes and reasonable violations of normal error and Gaussian copula assumptions, flexible to leverage control covariates and handle nonlinear models for choice outcomes, and provide guidance for 2sCOPE's use to obtain good performance, as summarized in the next section. Interestingly, the results reported in Web Appendix E.9 indicate that a 'strongly nonnormal' W is considerably more effective than multiple 'weakly nonnormal' Ws in helping the identification of the causal effect for an endogenous regressor with insufficient nonnormality.

GUIDELINES FOR USING 2SCOPE

To summarize, we have established theoretical results that guarantee 2sCOPE's desirable large sample properties where correlated exogenous regressors exist (Theorem 2) and endogenous regressors have nonnormal distributions (Theorem 3). As expected, simulation studies demonstrate that 2sCOPE performs well when the sample size is sufficiently large. Meanwhile, simulation studies also reveal that, for 2sCOPE to perform well in finite samples, it may require sufficient regressor nonnormality and relevance between P and W (e.g., Figure 1 (a)). To provide actionable guidelines for 2sCOPE's application to the data at hand, we conduct systematic simulation studies to establish the boundary conditions for using 2sCOPE. Specifically, the studies employ a factorial experimental design, which systematically varies the distributions of P and W, sample sizes, the level of endogeneity, and the strength of correlation between P and W. We evaluate 2sCOPE's performance using structural model parameter estimates' relative bias. Web Appendix E.8 provides details of the experimental design and results.

Figure 2 presents the decision tree for 2sCOPE's use based on the simulation studies' findings. The decision tree comprises three steps. In step 1, we test Assumption 5 (or 5(b) for multiple endogenous regressors) to choose between 2sCOPE and $\text{Copula}_{\text{Origin}}$. When Assumption 5 (5(b)) is satisfied, $\text{Copula}_{\text{Origin}}$ is preferred to 2sCOPE because, although both methods provide consistent estimates, $\text{Copula}_{\text{Origin}}$ is more efficient (Web Appendix E.7). Becker, Proksch, and Ringle (2021) provided a flowchart for the use of $\text{Copula}_{\text{Origin}}$.

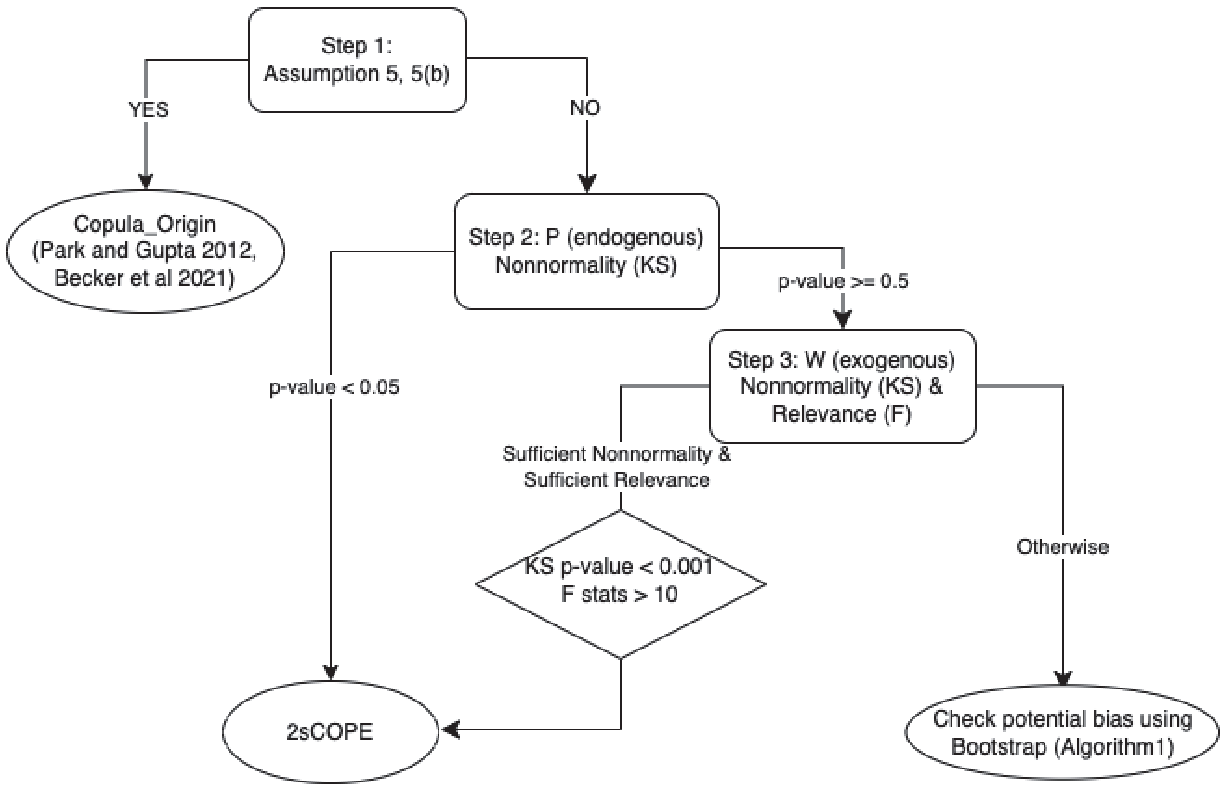


Figure 2: Decision Tree for Using 2sCOPE.
Note: P and W stand for endogenous and exogenous regressors, respectively; for both $\text{Copula}_{\text{Origin}}$ and 2sCOPE to work properly, W should satisfy the exogeneity assumption and be uncorrelated with the structural error; KS stands for the Kolmogorov-Smirnov test. An application of the decision tree to real-world data results in a high frequency (98% cases) of 2sCOPE usage (see more details in Footnote 15 and Table W11 in Web Appendix E.8)

Violation of Assumption 5 (5(b)) suggests the presence of relevant exogenous regressors that 2sCOPE may leverage to better handle endogeneity. In step 2, we test the nonnormality of the endogenous regressor P using the Kolmogorov-Smirnov (KS) test (see Web Appendix E.8 for the rationale of using the test of normality). If the KS test rejects the null at the 0.05 level[9], P possesses sufficient nonnormality, and 2sCOPE has a high probability of success in correcting endogeneity bias based on the results detailed in Web Appendix E.8. Otherwise, P has a close-to-normal distribution, which requires related exogenous regressors with sufficient nonnormality and relevance to help identification. Thus, in step 3, we check the W's nonnormality and relevance to P. Results in Web Appendix E.8 show: If the

[9]The choice of 0.05 level is conservative, as we find that a cut-off value of 0.10 can also ensure that 2sCOPE performs well, suggesting that one may still use 2sCOPE when $0.05 <$ p-value < 0.1. The cut-off points proposed here and elsewhere in the decision tree are indicative in light of the ongoing discussion surrounding strict p-value thresholds (McShane et al. 2024).

p-value of the KS test of an exogenous regressor W is smaller than 0.001 (i.e., sufficient nonnormality of W) and the relevance is sufficient (F statistic for the effect of W^* on $P^* > 10$ in the first-stage regression), 2sCOPE will have a high probability of success.

We have provided sufficient conditions of endogenous and exogenous regressors in steps 2 and 3 for 2sCOPE to have good finite-sample performance. These conditions are not necessary but are conservative. In particular, we consider extreme cases in which either the exogenous regressor in step 2 or the endogenous regressor in step 3 follows the normal distribution. In practice, regressors are more likely to have close-to-normal than exact normal distributions. The failure of W's sufficient condition tests does not necessarily preclude the use of 2sCOPE. For example, the estimation result of scenario 1 in Web Appendix Table W12 (P and W are close-to-normal and weakly nonnormal, respectively) demonstrates that 2sCOPE may still show acceptable finite-sample performance when the above (conservative) sufficient conditions are not satisfied. In this scenario (the rightmost branch in Figure 2), one can employ our proposed bootstrap resampling Algorithm 1 to evaluate 2sCOPE's finite-sample performance on a case-by-case basis.

Algorithm 1 A Bootstrap Algorithm for Evaluating Finite-sample Bias of 2sCOPE

Series Input: data Y, P, W, sample size N, $\widehat{\theta}(Y, P, W)-$ 2sCOPE estimates of the structural model parameters, $(\widehat{H}, \widehat{L})-$ empirical CDFs of P and W, and $\widehat{\Sigma}-$ Gaussian copula correlation structure estimate. If the $\hat{\rho}_{P\xi}$ is small and not significantly different from zero, set $\hat{\rho}_{P\xi} = \pm 0.5$ in $\widehat{\Sigma}$.
for $b = 1$ to B **do**
  Simulate P_b^*, W_b^*, ξ_b^* from Gaussian Copula $\Psi_{\widehat{\Sigma}}(\Phi^{-1}(U_P), \Phi^{-1}(U_w), \Phi^{-1}(U_\xi))$, sample size=N;
  Obtain $P_b = \widehat{H}^{-1}(\Phi(P_b^*)), W_b = \widehat{L}^{-1}(\Phi(W_b^*))$ and $\xi_b = \widehat{\sigma}_\xi \cdot \xi_b^*$, where $\widehat{\sigma}_\xi$ is the 2sCOPE estimate of the standard deviation of structural error term;
  Obtain $Y_b = f(P_b, W_b, \xi_b, \widehat{\theta}(Y, P, W))$, where f is the linear regression in this setting;
  Obtain the 2sCOPE estimate $\widehat{\theta}_b = \widehat{\theta}(Y_b, P_b, W_b)$ using the bth bootstrap sample.
end for
Calculate potential bias of the 2sCOPE estimator: $\frac{1}{B}\sum_{b=1}^{B} \widehat{\theta}_b - \widehat{\theta}(Y, P, W)$.

Bootstrap simulations can be used to evaluate the bias size in parameter estimates when sample size is small to moderate (Efron and Tibshirani 1994, Ch. 10; Hooker and Mentch

2018)[10], even if the estimation performs well for large samples. Specifically, Algorithm 1 randomly draws the same number of observations from the underlying copula model and the structural model estimated using the original sample,[11] and then performs the 2sCOPE estimation on the bootstrap sample as in the original sample. We repeat this simulation B times and obtain a distribution for each model coefficient estimate. We then compare the mean of each coefficient estimate's distribution with the corresponding coefficient estimate using the original data, which is the true parameter value in our model-based bootstrap re-sampling. The small-sample bias of a coefficient estimate is the difference between the average coefficient estimate from bootstrap samples and the original sample's coefficient estimate.

EMPIRICAL APPLICATION

In this section, we illustrate the proposed method to address the price endogeneity issue using store-level toothpaste category sales data for Chicago over 373 weeks from 1989 to 1997 [12]. To control for product size, we select the most common weight, which is 6.4 oz. Specifically, we estimate the following sales model:

$$\log(\text{Sales}_t) = \beta_0 + \log(\text{Retail Price}_t) \cdot \beta_1 + W_t'\beta_2 + \xi_t, \tag{24}$$

where $t = 1, 2, ..., T$ indexes week. Store/category managers and policymakers are often interested in the effect of price on the category demand (e.g., Nijs et al. 2001; Li, Linn, and Muehlegger 2014). Empirical estimates of category price elasticity (i.e., β_1) are key to optimal pricing, allowing retailers/manufacturers to expand category demands (the first source of profitable growth), and critical to designing effective interventions (e.g., gasoline or soda tax) for policymakers to achieve policy goals. Meanwhile, retail price is typically considered endogenous in the category sales demand model (Nijs et al. 2001; Li, Linn, and Muehlegger

[10]This bootstrap simulation is distinct from and should not be confused with the bootstrap method mentioned in Table 4 and used to obtain the standard errors of 2sCOPE estimates.

[11]When $\rho_{P\xi}$ is set at zero (i.e., no endogeneity), 2sCOPE is expected to have no finite-sample bias since in this case 2sCOPE reduces to OLS which is unaffected by regressor normality. Thus, if $\hat{\rho}_{P\xi}$ is small and not significantly different from zero, we recommend setting $\widehat{\rho}_{P\xi}$ to a plausible non-zero value (e.g., ± 0.5 as suggested in Algorithm 1).

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

2014; Park and Gupta 2012; Haschka 2022). Retail price endogeneity may derive from unmeasured product characteristics or demand shocks that can influence both consumers' and retailers' decisions. Unobserved by researchers, these variables are absorbed into the structural error, creating the endogeneity problem. Prices of different stores are correlated and often used as an IV for each other. This allows us to test the proposed 2sCOPE method's performance in an empirical setting with a good IV. Besides the endogenous price, two promotion-related variables—bonus promotion and direct price reduction—would also affect demand. The promotion decisions during the study period were typically made on a quarterly basis or less frequently, plus a long lead time (e.g., several weeks) for implementation; thus, they were unlikely to be correlated with the weekly unobserved demand shock and can be considered exogenous (Chintagunta 2002; Sriram, Balachander, and Kalwani 2007).[13] We focus on category sales in two large stores in Chicago (Stores 1 and 2). We convert retail price, in-store promotion and sales from UPC level to the aggregate category level, computed as weekly market share-weighted averages of UPC-level variables.

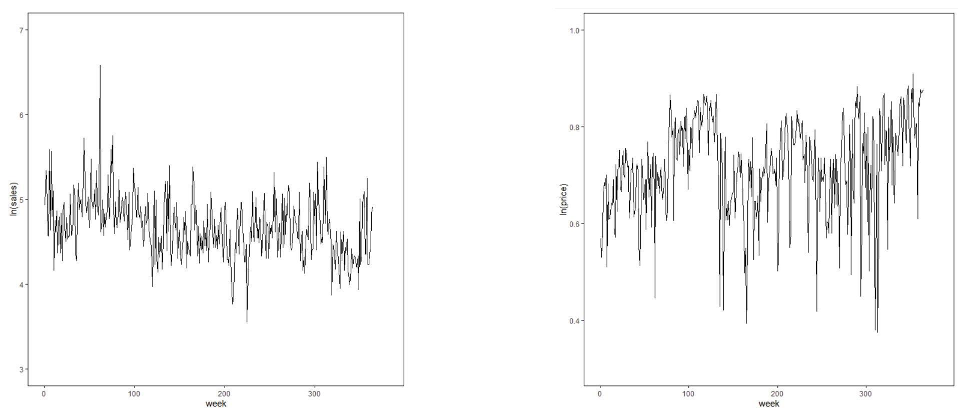
The correlation between log retail price and bonus promotion in Store 1 (Store 2) is -0.30 (-0.15), and the correlation between log retail price and price reduction promotion in Store 1 (Store 2) is -0.23 (-0.35). The appreciable correlations between price and promotion variables provide a suitable context for testing our method with correlated endogenous and exogenous regressors. The moderate sample size (T=373) also allows us to evaluate the 2sCOPE's finite-sample performance in the presence of potentially insufficient regressor nonnormality in real data. Table 8 reports summary statistics of the key variables.

Figure 3 plots the log sales and log retail prices of toothpaste at Store 1 over time (Store 2 is very similar). To control for the possible time trend of retail price, we use detrended log retail prices (and for IVs as well) for estimation. Figure 4 shows the histograms of detrended log retail prices and the two promotion variables. All three are continuous variables.

[13] We also checked the endogeneity of the bonus and price reduction promotion variables using the Hausman test employing IVs (promotions in the other store). The Hausman test's p-values are 0.30 for bonus promotion and 0.14 for price reduction in store 1 (store 2 is similar), thus yielding no evidence to suggest that the two promotion variables are endogenous, consistent with prior literature.

Table 8: 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



(a) Store 1 log sales

(b) Store 1 log retail price

Figure 3: Log sales and log retail price of toothpaste in Store 1.

The flowchart in Figure 2 guides our use of 2sCOPE. In Store 1, the correlations between logP^* and the exogenous regressors are -0.44[14] for bonus promotion and -0.26 for price reduction promotion, both of which differ substantially from zero with p-value $< 2.2 \times e^{-16}$ and $7.542 \times e^{-08}$ respectively, indicating a violation of Assumption 5, which $\text{Copula}_{\text{Origin}}$ requires to yield consistent estimates. Next, we check the endogenous regressor's sufficient nonnormality. The KS test of the endogenous price yields a p-value of $0.063 > 0.05$, concluding insufficient nonnormality of the endogenous price that $\text{Copula}_{\text{Origin}}$ (or COPE) cannot handle. We then proceed to check the exogenous regressors' nonnormality and their relevance with the endogenous regressor. Bonus variable is strongly nonnormal (p-value of KS test=$3.159 \times e^{-12}$) and sufficiently relevant (F statistic $= 89.5 > 10$). Price reduction is

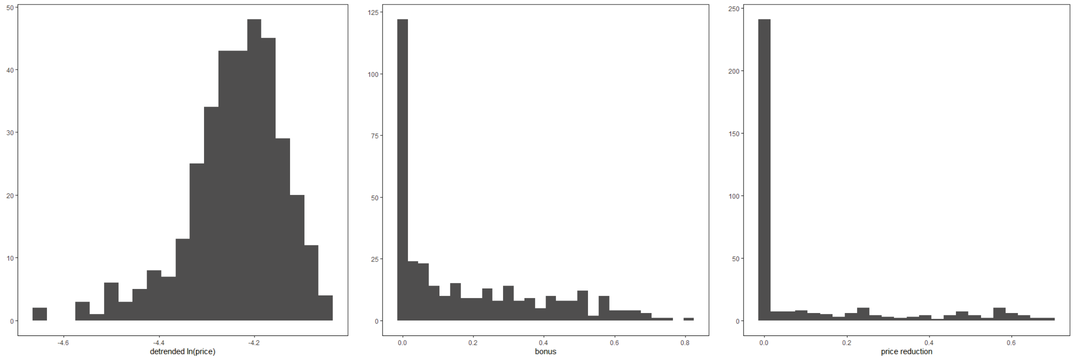
[14]These correlations differ from the correlations between the original endogenous and exogenous regressors without copula transformation—for example, cor(logP,bonus) = -0.30.

also strongly nonnormal (p-value of KS test $< 2.2 \times e^{-16}$) and is sufficiently relevant (F statistic $= 27.3 > 10$). Thus, according to Figure 2, the Store 1 dataset is appropriate for using 2sCOPE to correct endogeneity, and 2sCOPE is expected to have a high probability to achieve good finite-sample performance.

We next go through the flowchart for Store 2. First, the correlations between logP* and the exogenous regressors are -0.32 for bonus promotion and -0.36 for price reduction promotion, both of which are substantially different from zero with p-value $1.54 \times e^{-10}$ and $4.42 \times e^{-13}$ respectively, indicating a violation of Assumption 5 required for $\text{Copula}_{\text{Origin}}$ to yield consistent estimates. Next, we check the nonnormality of the endogenous regressor. The KS test of the endogenous price yields a p-value of $0.0053 < 0.05$, concluding sufficient nonnormality of the endogenous price. Thus, according to the decision tree in Figure 2, the Store 2 dataset is also appropriate for using 2sCOPE to correct endogeneity[15].

We use the IV-based two-stage least-squares (TSLS) estimator to cross-validate 2sCOPE's performance and the IV used. We use retail price at the other store as an instrument for price, which is commonly used in the literature (Park and Gupta 2012; Rossi 2014). This variable can be a valid instrument as it satisfies the two key requirements. First, retail prices across stores in a market can be highly correlated because wholesale prices are typically the same (or very similar). The Pearson correlation between the detrended log retail prices in 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, 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). Meanwhile, unobserved national advertisement is not expected to affect production cost and wholesale price on a weekly basis and is thus expected to exert only a small effect on the variance of weekly wholesale price. Given that national advertisement occurs only in a few in-

[15]We further tested all the 86 stores in the toothpaste category of the Dominick dataset. The result shows that all stores violate Assumption 5 which is required for $\text{Copula}_{\text{Origin}}$, and around 98% of stores satisfy the conditions of using 2sCOPE in the decision tree in Figure 2 (Table W11 in Web Appendix E.8).



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

Figure 4: Histogram of Log Retail Price, Bonus and Price Reduction in Store 1

stances in any given planning time horizon (e.g., a quarter or a year), one would expect these demand shocks to be highly correlated and have a small variance over time at the weekly frequency (Rossi 2014). These considerations suggest that the exclusion restriction condition is reasonably satisfied in the presence of unobserved national advertisement. However, like other IVs, the validity claim cannot be fully verified, and is debatable. We therefore perform both 2sCOPE and TSLS to cross-validate one another. Congruent results from the two methods increase our confidence in endogeneity correction. Like TSLS, 2sCOPE includes (and uses) the existing exogenous regressors in the first-stage regression; however, unlike TSLS, 2sCOPE neither includes nor needs IVs. We first regress $\text{logP}^* = \Phi^{-1}(\widehat{H}(\text{logP}))$ on $\text{Bonus}^* = \Phi^{-1}(\widehat{L}_1(\text{Bonus}))$ and $\text{PriceRedu}^* = \Phi^{-1}(\widehat{L}_2(\text{PriceRedu}))$, 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 (Web Appendix F).

Table 9 reports the estimation results. The OLS estimates in Store 1 differ significantly from TSLS estimates, suggesting the price endogeneity issue. Instrumenting for retail price changes the price coefficient estimate from -0.767 to -1.797, implying a positive correlation between unobserved product characteristics and price. The 2sCOPE estimate of ρ, representing the correlation between the endogenous regressor P_t and the error term, is 0.297 (t-value=3.34) and significantly positive, further confirming our conclusion and consistent with previous empirical findings (e.g., Villas-Boas and Winer 1999; Chintagunta, Dubé, and

Goh 2005).

Table 9: Estimation Results: Toothpaste Sales

		OLS			TSLS			2sCOPE		
Store	Parameters	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	-3.908	2.314	1.69
	Price	-0.767	0.288	2.66	-1.797	0.396	4.54	-2.014	0.555	3.63
	Bonus	0.371	0.122	3.31	0.104	0.141	0.74	0.064	0.171	0.37
	PriceRedu	0.498	0.115	4.33	0.285	0.125	2.28	0.275	0.143	1.92
	ρ	-	-	-	-	-	-	0.297	0.089	3.34
Store 2	Constant	-3.898	1.246	3.13	0.763	1.943	0.39	0.001	2.702	0.00
	Price	-1.982	0.300	6.61	-0.864	0.467	1.85	-1.048	0.648	1.62
	Bonus	0.062	0.116	0.53	0.286	0.148	1.93	0.239	0.151	1.58
	PriceRedu	0.283	0.111	2.55	0.540	0.137	3.94	0.467	0.152	3.07
	ρ	-	-	-	-	-	-	-0.188	0.109	1.72

This positive price-error correlation causes upward bias (i.e., less price sensitivity) in the OLS price estimate. By directly accounting for this price-error dependence and controlling the first-stage residual, which captures unobserved product characteristics causing the positive correlation between the endogenous price and the error term, 2sCOPE corrects the classic upward endogeneity bias of price elasticity from -0.767 to -2.014. The 2sCOPE price elasticity estimate of -2.014 is close to that of -1.797 from the TSLS method. Both 2sCOPE and TSLS price estimates show greater price sensitivity, suggesting that both correct the price endogeneity problem in the right direction. The TSLS and 2sCOPE estimates are reasonable because the price elasticity of the toothpaste category is approximately -2.0 (Hoch et al. 1995; Mackiewicz and Falkowski 2015; Chen and Lim 2022).

Unlike Store 1, Store 2's results indicate that the retail price is not endogenous. The estimate of ρ (the correlation between price and the error term) does not differ significantly from 0 for 2sCOPE (t-value $\leq$ 1.96 under column "2sCOPE" for Store 2 in Table 9). The OLS price estimate is -1.982, which is very close to the estimates of TSLS and 2sCOPE in Store 1, further confirming Store 2's lack of price endogeneity. Overall, the price elasticity estimates from TSLS and 2sCOPE are close for Store 2, and the observed differences between them and the OLS estimate may be attributed to the estimation variability incurred from the use of more complicated models in the absence of endogeneity.

Evaluating Finite-Sample Performance of Copula Correction Using Bootstrap

In the above, the convergence of results between TSLS and the proposed 2sCOPE in both stores supports the validity of the proposed method in addressing the endogeneity issue. The flowchart in Figure 2 also suggests our empirical data satisfy the boundary conditions under which 2sCOPE is expected to have good finite-sample performance. Although here it is unnecessary to empirically evaluate the finite-sample performance using the bootstrap resampling in Algorithm 1, we apply the algorithm to illustrate its usage in the empirical application. Specifically, we apply the bootstrap algorithm (Algorithm 1) to our empirical application with the true parameter values set as Store 1's 2sCOPE estimates (Table 9) rounded to the first non-zero number when generating bootstrap samples. We also consider the case in which ρ is set at 0.5, somewhat larger than the estimated value of ρ (=0.3), to assess the robustness of the bootstrap findings. The steps used to generate these bootstrap samples are detailed in Web Appendix G.

Table 10 summarizes the means and standard deviations of parameter estimates for OLS and 2sCOPE over the 1000 bootstrap samples, unlike the estimation result on a single observed dataset reported in Table 9. The estimation results are broadly consistent with those in Table 9. In both cases ($\rho = 0.3$ and 0.5), the estimates of 2sCOPE are distributed closely to the true values, demonstrating that 2sCOPE corrects the bias of OLS estimates and performs well with little finite-sample bias in our empirical application.

CONCLUSION

Observational studies often require rigorous study designs and methodologies to overcome endogeneity concerns. While it is preferable to bring exogeneity via good instruments for identification, this is not always possible. In this paper, we focus on the IV-free copula

Table 10: Finite-Sample Performance of Copula Correction.

		OLS			2sCOPE				OLS			2sCOPE		
Parameters	True	Est	SE	t_{bias}	Est	SE	t_{bias}	True	Est	SE	t_{bias}	Est	SE	t_{bias}
Constant	-4	1.514	0.777	7.098	-3.782	1.619	0.135	-4	5.256	0.635	14.57	-3.601	1.393	0.287
Price	-2	-0.678	0.186	7.099	-1.946	0.388	0.139	-2	0.220	0.152	14.59	-1.904	0.334	0.287
Bonus	0.1	0.458	0.088	4.046	0.113	0.128	0.103	0.1	0.706	0.073	8.290	0.126	0.112	0.236
PriceRedu	0.3	0.571	0.089	3.058	0.309	0.112	0.079	0.3	0.764	0.075	6.160	0.323	0.095	0.240
ρ	0.3	-	-	-	0.284	0.071	0.222	0.5	-	-	-	0.483	0.048	0.360

Note: "Est" and "SE" denote the mean and standard deviation of the estimates over 1000 bootstrap samples of Store 1 data.

method to handle endogenous regressors, proposing a generalized two-stage copula endogeneity correction (2sCOPE) method that extends the existing copula correction methods (Park and Gupta 2012; Becker, Proksch, and Ringle 2021; Haschka 2022; Eckert and Hohberger 2022) to more general settings. Specifically, 2sCOPE permits the correlation of exogenous regressors with endogenous regressors and relaxes the nonnormality assumption on the endogenous regressors. Similar to $\text{Copula}_{\text{Origin}}$, 2sCOPE corrects endogeneity by adding "generated regressors" derived from existing regressors and is straightforward to use. However, unlike $\text{Copula}_{\text{Origin}}$ that adds the latent copula transformations of endogenous regressors directly to the model, 2sCOPE has two stages. The first obtains the residuals from regressing latent copula data for the endogenous regressor on the latent copula data for the exogenous regressors, while the second uses the first-stage residual as a "generated regressor" in the structural regression model. We prove that 2sCOPE yields consistent causal effect estimates when the normally distributed structural error and all regressors follow a Gaussian copula correlation structure. 2sCOPE can also relax the nonnormality assumption on endogenous regressors and substantially improve copula correction's finite-sample performance.

We evaluate 2sCOPE's performance via simulation studies and demonstrate its use in an empirical application. The simulation results show that 2sCOPE yields consistent estimates under relaxed assumptions. Moreover, 2sCOPE outperforms $\text{Copula}_{\text{Origin}}$ (and COPE) in terms of dealing with close-to-normal or normal endogenous regressors and improving estimation efficiency. Endogenous regressors are allowed to have close-to-normal or even normal distributions with the help of exogenous regressors (Figure 2). The efficiency gain relative to COPE is substantial and may reach ∼80% in simulation studies (Web Appendix E.3), implying that 2sCOPE can reduce the sample size by ∼80% needed to achieve the same estimation efficiency as COPE that does not exploit the correlations between endogenous and exogenous regressors. Finally, robustness checks indicate that 2sCOPE is reasonably robust to the structural error distributional assumption and non-Gaussian copula correlation structure (Web Appendix E.4, E.5 & E.6). We further apply 2sCOPE to a public dataset in marketing. Regarding endogenous price, we find that the estimated price coefficient using our proposed 2sCOPE is very close to the TSLS estimate and the price coefficient reported in the literature, while the OLS estimator indicates substantial biases.

These findings have rich implications regarding the practical use of the copula-based IV-free methods for handling endogeneity. A known critical assumption for $\text{Copula}_{\text{Origin}}$ is the nonnormality of endogenous regressors. The users of the method have largely checked and verified this assumption. However, our findings indicate that this is insufficient: it is also necessary to check Assumption 5 for the single endogenous regressor case and Assumption 5(b) for the multiple endogenous regressors case. Neither assumption is the same as checking the pairwise correlations between the endogenous and exogenous regressors. Assumption 5 evaluates pairwise correlations involving copula transformation of the endogenous regressor, which, as shown in the literature (Danaher and Smith 2011) and in our specific empirical application, may differ substantially from the pairwise correlations using the original variables. Assumption 5(b) evaluates the correlations between exogenous regressors and the linear combination of generated regressors, which differ even more considerably from pairwise correlations on the regressors themselves. We expect that Assumptions 5 and 5(b) will be violated in the majority of practical applications (see the result of real-data application in Footnote 15 and Web Appendix E.8), for which 2sCOPE should be used. Even in exceptional cases that satisfy the above assumptions[16], 2sCOPE can still be used successfully (Web Appendix E.7). Yet in these exceptional cases, $\text{Copula}_{\text{Origin}}$ may be considered (Figure 2, Step 1) since the simpler and valid model gains estimation/prediction efficiency and is easier to understand, use, and communicate with as a decision calculus tool (Little 2004).

For endogenous regressors with insufficient nonnormality or exogenous regressors that violate Assumptions 5 or 5(b), 2sCOPE outperforms $\text{Copula}_{\text{Origin}}$ and is recommended. When all endogenous regressors have sufficient nonnormality, 2sCOPE is expected to perform well. If any endogenous regressor has insufficient nonnormality, 2sCOPE exploits exogenous regressors with sufficient relevance and nonnormality levels (Figure 2 details sufficient conditions) for satisfactory model identification in finite samples. One can empirically check and verify whether these conditions are satisfied for data at hand, using normality and relevance tests. For cases in which these conditions are not met, we also propose a novel bootstrap resampling method (Algorithm 1) to directly gauge and validate 2sCOPE's finite-sample performance in real applications on a case-by-case basis, complementing the above rules of thumb using tests of normality and relevance. According to the decision-tree check using real-world data, 98% of cases result in using 2sCOPE in the decision tree (Figure 2), demonstrating the overwhelming need for 2sCOPE (Table W11 in Web Appendix E.8).

Unlike the TSLS method, 2sCOPE requires no IVs that must satisfy exclusion restriction (ER). ER is considerably more stringent than the exogeneity condition in that the IV is not only exogenous but also does not appear in the outcome model, meaning that the IV cannot affect the outcome Y by any means other than the endogenous regressor.[17] It is typically impossible to test ER; one must rely on institutional knowledge and theoretical arguments to establish ER's credibility, which is often the most challenging aspect of IV applications. By contrast, our approach eliminates the requirement that any variable satisfy the ER assumption, which constitutes a significant gain. Using 2sCOPE, one needs not

[16] This can occur, *e.g.,* when the control variables include prognostic factors of only the outcome to improve prediction accuracy—i.e., W is uncorrelated with P.

[17] For example, when estimating the price elasticity for ice cream, weather satisfies exogeneity but does not satisfy ER because weather can affect the demand for ice cream directly rather than exclusively via pricing.

argue for ER.

Meanwhile, 2sCOPE is capable of leveraging relevant exogenous variables in W in the outcome model (e.g., in Equation (8)) for model identification. Marketing models rarely contain endogenous regressors exclusively; the majority of the outcome models estimated (e.g. using OLS or IVs) in marketing include exogenous variables for various reasons, such as the inclusion of exogenous regressors as control variables to mitigate the endogeneity concern of the primary explanatory variables, to improve model estimation and forecasting accuracy, to make the outcome models substantively complete and relevant, or to render the ER assumption of IVs more plausible. These exogenous control variables influence other variables in the marketing model but are determined outside the model and are unaffected by the model outcomes. Common examples include environmental factors (e.g., weather), macroeconomic indicators (e.g., interest rate, GDP growth, inflation rates), government policies and legal rules, natural disasters and events, customer characteristics used by firms for targeting, marketing activities (e.g., promotions) pre-arranged on an annual basis and independent of daily/weekly demand shocks.

These exogenous regressors are not used to generate the CCF in $\text{Copula}_{\text{Origin}}$. By contrast, 2sCOPE can leverage these exogenous variables to improve model identification and estimation. Such exogenous control variables are much more widely available than the IVs because 2sCOPE does not require any of these exogenous variables to satisfy the stringent ER condition. Furthermore, no theoretical arguments are required for the direction and intuition of correlation between W and P. An empirical correlation is sufficient (Web Appendix E.12).[18] Finally, when endogenous regressors have insufficient nonnormality, 2sCOPE can leverage exogenous regressors with certain nonnormality and relevance levels (Figure 2), which are feasible in many applications, for identification.

To fully benefit from leveraging relevant control covariates in W for handling endogeneity,

[18] An example is location-based targeted pricing. Researchers do not observe location but can reasonably assume that location sale-effects can be captured by a rich set of observed demographic variables W. In this case, W and P (price) are spuriously correlated. Price may be endogenous due to unobserved shocks affecting all locations.

these control covariates need to be exogenous. In practice, choosing a good set of exogenous control variables requires care in model specification and may pose empirical challenges for data analysts. Similar to OLS, TSLS, and $\text{Copula}_{\text{Origin}}$, the addition of endogenous variables to W can yield inconsistent model estimates for 2sCOPE.[19] As such, certain types of variables that violate the exogeneity condition, such as mediators or colliders, should be excluded from W. For example, consumer perceptions of price levels and emotions toward stores are found to lie in the causal pathway between pricing and demand (Cakici and Tekeli 2022). Such mediators are endogenous (affected by the endogenous pricing variable) and, consequently, including them (e.g., measured using periodic consumer surveys) in W and treating them as being exogenous when estimating a demand model will introduce overcontrol bias (Hernan and Robins 2023). Additionally, if these mediators are also influenced by the unobserved determinants of the sales outcome (e.g., store service quality), then the two mediators could also be colliders influenced *by* both the primary explanatory variables and the outcome. Including colliders in the regression and treating them as exogenous control variables is known to cause collider bias (Hernan and Robins 2023). The remedy is to exclude mediators and colliders from W. Like other econometric methods, the reasonableness of the exogenous W assumption should be evaluated and justified in the context of the objectives and scopes of analysis. In this respect, substantive or institutional knowledge is useful for guiding or justifying the selection of appropriate exogenous control covariates. To avoid violating the exogeneity assumption about W, we recommend practicing clean adjustment employing only exogenous control variables necessary to improve causal effect estimation. Control variables strongly believed to be endogenous should be treated as endogenous regressors in the model or removed from the model.[20] In this respect, 2sCOPE can offer ways of tackling some

[19]Theoretically, neither the OLS nor the IV method requires the presence of W. Practically, however, the structural model typically includes relevant W variables, for the reasons noted above. In the considerably more common cases in which the structural model includes W, these other methods (OLS, IV and $\text{Copula}_{\text{Origin}}$) all require the exogeneity of W, as does 2sCOPE (Web Appendix E.11).

[20]An exception is endogenous control variables that are the prognostic factors of only the outcome variable. Unlike TSLS using IVs, copula correction methods yield consistent causal-effect estimation between the focal endogenous regressor and the outcome even if such endogenous variables are included in W and treated as exogenous (Web Appendix E.11).

major challenges encountered in model specifications: e.g., certain control variables must be included to mitigate endogeneity concerns in OLS estimation or make the ER assumption of IVs plausible but are nonetheless believed to be endogenous and thus pose a dilemma as to whether they should be included in the model for the OLS or TSLS estimation.

Although 2sCOPE contributes to solving regressor endogeneity by relaxing key assumptions of the existing copula correction methods and extending them to more general settings, it is not without its limitations. For 2sCOPE to function optimally, the endogenous regressors' distributions must contain adequate information. The condition is violated when the endogenous regressors follow Bernoulli distributions or discrete distributions with small support, as Park and Gupta (2012) noted. The proposed 2sCOPE method does not address this limitation. The simplicity of 2sCOPE assumes the normal structural error and Gaussian copula dependence structure. Our evaluation demonstrates that 2sCOPE is robust to symmetric nonnormal error distributions, linear dependence among endogenous and exogenous regressors, and certain non-Gaussian copula structure (Web Appendix E.4, E.5 & E.6). Such robustness may not hold for asymmetric nonnormal error distributions or other forms of dependence or copula structure. Future research is needed to test and relax these assumptions using more flexible copula methods. Despite these limitations, we expect that 2sCOPE will provide a useful alternative to a broad range of empirical problems when instruments are unavailable. Although our empirical application only considered linear sales models, potential empirical applications of 2sCOPE abound. We have derived 2sCOPE and evaluated its performance via simulated data for a range of other commonly used marketing models, including linear panel models with mixed effects, random coefficient logit models, and slope endogeneity. Future empirical studies may apply 2sCOPE to these and many other cases not studied here.

REFERENCES

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt (2005), "Competition and innovation: An inverted-U relationship," *The Quarterly Journal of Economics*, 120 (2), 701–728.

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

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

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

Atefi, Yashar, Michael Ahearne, James G Maxham III, D Todd Donavan, and Brad D Carlson (2018), "Does selective sales force training work?," *Journal of Marketing Research*, 55 (5), 722–737.

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

Becker, Jan-Michael, Dorian Proksch, and Christian M Ringle (2021), "Revisiting Gaussian copulas to handle endogenous regressors," *Journal of the Academy of Marketing Science*, 50 (1), 1–21.

Cakici, A Celil and Sena Tekeli (2022), "The mediating effect of consumers' price level perception and emotions towards supermarkets," *European Journal of Management and Business Economics*, 31 (1), 57–76.

Chen, Hao and Alvin Lim "Were consumers less price sensitive to life necessities during the COVID-19 pandemic? An empirical study on Dutch consumers," "Proceedings of SAI Intelligent Systems Conference," pages 79–100, Springer (2022).

Chintagunta, Pradeep, Jean-Pierre Dubé, and Khim Yong Goh (2005), "Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models," *Management Science*, 51 (5), 832–849.

Chintagunta, Pradeep, Tülin Erdem, Peter E Rossi, and Michel Wedel (2006), "Structural modeling in marketing: review and assessment," *Marketing Science*, 25 (6), 604–616.

Chintagunta, Pradeep K (2002), "Investigating category pricing behavior at a retail chain," *Journal of Marketing Research*, 39 (2), 141–154.

Christopoulos, Dimitris, Peter McAdam, and Elias Tzavalis (2021), "Dealing with endogeneity in threshold models using copulas," *Journal of Business & Economic Statistics*, 39 (1), 166–178.

Danaher, Peter J (2007), "Modeling page views across multiple websites with an application to internet reach and frequency prediction," *Marketing Science*, 26 (3), 422–437.

Danaher, Peter J and Michael S Smith (2011), "Modeling multivariate distributions using copulas: Applications in marketing," *Marketing science*, 30 (1), 4–21.

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

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

Ebbes, Peter, Michel Wedel, and Ulf Böckenholt (2009), "Frugal IV alternatives to identify the parameter for an endogenous regressor," *Journal of Applied Econometrics*, 24 (3), 446–468.

Ebbes, Peter, Michel Wedel, Ulf Böckenholt, and Ton Steerneman (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.

Eckert, Christine and Jan Hohberger (2022), "Addressing Endogeneity Without Instrumental Variables: An Evaluation of the Gaussian Copula Approach for Management Research," *Journal of Management*, DOI: 10.1177/01492063221085913.

Efron, Bradley and Robert J Tibshirani (1994), *An introduction to the bootstrap* CRC press.

Elshiewy, Ossama and Yasemin Boztug (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.

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

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

Haschka, Rouven E (2022), "Handling Endogenous Regressors using Copulas: A Generalization to Linear Panel Models with Fixed Effects and Correlated Regressors," *Journal of Marketing Research*, 59 (4), 860–881.

Heitmann, Mark, Jan R Landwehr, Thomas F Schreiner, and Harald J van Heerde (2020), "Leveraging brand equity for effective visual product design," *Journal of Marketing Research*, 57 (2), 257–277.

Hernan, Miguel A. and James M. Robins (2023), *Causal Inference: What If.* Chapman & Hall/CRC Monographs on Statistics & Applied Probab, CRC Press.

Hoch, Stephen J, Byung-Do Kim, Alan L Montgomery, and Peter E Rossi (1995), "Determinants of store-level price elasticity," *Journal of Marketing Research*, 32 (1), 17–29.

Hooker, Giles and Lucas Mentch (2018), "Bootstrap bias corrections for ensemble methods," *Statistics and Computing*, 28 (1), 77–86.

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

Kim, Sungjin, Clarence Lee, and Sachin Gupta (2020), "Bayesian Synthetic Control Methods," *Journal of Marketing Research*, 57(5), 831–852.

Lewbel, Arthur (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.

Li, Shanjun, Joshua Linn, and Erich Muehlegger (2014), “Gasoline taxes and consumer behavior,” *American Economic Journal: Economic Policy*, 6 (4), 302–342.

Li, Yang and Asim Ansari (2014), “A Bayesian Semiparametric Approach for Endogeneity and Heterogeneity in Choice Models,” *Management Science*, 60, 1161–1179.

Little, John D. C. (2004), “Models and Managers: The Concept of a Decision Calculus,” *Management Science*, 50 (12_supplement), 1841–1853 https://doi.org/10.1287/mnsc.1040.0267.

Mackiewicz, Robert and Andrzej Falkowski (2015), “The Use of Weber Fraction as a Tool to Measure Price Sensitivity: a Gain and Loss Perspective.,” *Advances in Consumer Research*, 43.

McShane, Blakeley B, Eric T Bradlow, John G Lynch Jr, and Robert J Meyer (2024), ““Statistical Significance” and Statistical Reporting: Moving Beyond Binary,” *Journal of Marketing*, page 00222429231216910.

Mendelson, Haim (2000), “Organizational Architecture and Success in the Information Technology.,” *Management Science*, 46, 513–529.

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

Nijs, Vincent R, Marnik G Dekimpe, Jan-Benedict EM Steenkamps, and Dominique M Hanssens (2001), “The category-demand effects of price promotions,” *Marketing science*, 20 (1), 1–22.

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

Otter, Thomas, Timothy J Gilbride, and Greg M Allenby (2011), “Testing models of strategic behavior characterized by conditional likelihoods,” *Marketing Science*, 30 (4), 686–701.

Papies, Dominik, Peter Ebbes, and Harald J Van Heerde “Addressing endogeneity in marketing models,” “Advanced methods for modeling markets,” pages 581–627, Springer (2017).

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

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

Qian, Yi (2008), “Impacts of entry by counterfeiters,” *Quarterly Journal of Economics*, 123, 1577–1609.

Qian, Yi and Hui Xie (2022), “Simplifying Bias Correction for Selective Sampling: A Unified Distribution-Free Approach to Handling Endogenously Selected Samples,” *Marketing Science*, 41(2), 336–360.

Qian, Yi, Hui Xie, and Anthony Koschmann (2022), "Should Copula Endogeneity Correction Include Generated Regressors for Higher-order Terms? No, It Hurts," *NBER Working Paper, https://www.nber.org/papers/w29978.*

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

Rossi, Peter 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, Oliver J and George F Watson (2019), "Endogeneity and marketing strategy research: An overview," *Journal of the Academy of Marketing Science*, 47 (3), 479–498.

Shi, Huanhuan, Shrihari Sridhar, Rajdeep Grewal, and Gary Lilien (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.

Sorescu, Alina, Nooshin L. Warren, and Larisa Ertekin (2017), "Event Study Methodology in the Marketing Literature: An overview," *Journal of the Academy of Marketing Science*, 45, 186–207.

Sriram, Srinivasaraghavan, Subramanian Balachander, and Manohar U Kalwani (2007), "Monitoring the dynamics of brand equity using store-level data," *Journal of Marketing*, 71 (2), 61–78.

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

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

Van Heerde, Harald J, Maarten J Gijsenberg, Marnik G Dekimpe, and Jan-Benedict EM Steenkamp (2013), "Price and advertising effectiveness over the business cycle," *Journal of Marketing Research*, 50 (2), 177–193.

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

Wang, Yixin and David M Blei (2019), "The blessings of multiple causes," *Journal of the American Statistical Association*, 114 (528), 1574–1596.

Wooldridge, Jeffrey M (2010), *Econometric analysis of cross section and panel data* MIT press.

Yang, Sha, Yuxin Chen, and Greg M Allenby (2003), "Bayesian analysis of simultaneous demand and supply," *Quantitative marketing and economics*, 1 (3), 251–275.

Addressing Endogeneity using a Two-stage Copula Generated Regressor Approach

WEB APPENDIX

These materials have been supplied by the authors to aid in the understanding of their paper. The AMA is sharing these materials at the request of the authors.

TABLE OF CONTENTS

WEB APPENDIX A: PROOFS RELATED TO COPULA$_{\textbf{Origin}}$

Web Appendix A.1: Proof of Theorem 1

Under the Gaussian copula assumption for structural error 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 (Assumption 1 in Table 2), $Cov(W_t, \xi_t) = 0$,

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

Thus, whenever W_t and P_t^* is correlated, the covariance between W_t and ω_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. Consequently, this regressor-error dependence will cause inconsistent OLS estimates of β as well as α using Equation (6) when P and W are correlated (Footnote 19). **Theorem proved.**

COPE Method: A Direct Extension of Copula$_{\textbf{Origin}}$

Under the Gaussian copula model for the endogenous regressor, P_t, the correlated exogenous regressor, W_t, and the structural error term, ξ_t in Equations (7,10), the structural error in the main model (Equation 1) can be rewritten as

$$\xi_t = \sigma_\xi \cdot \xi_t^* = \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2} P_t^* + \frac{-\sigma_\xi \rho_{pw} \rho_{p\xi}}{1-\rho_{pw}^2} W_t^* + \sigma_\xi \sqrt{1 - \rho_{p\xi}^2 - \frac{\rho_{pw}^2 \rho_{p\xi}^2}{1-\rho_{pw}^2}} \omega_{3,t}. \qquad \text{(W1)}$$

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 $\xi_t|P_t$ [21], and the other part as an independent new error term. Then, we substitute Equation (W1) into the main model in Equation (1), and obtain the following regression equation:

$$Y_t = \mu + P_t\alpha + W_t\beta + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2} P_t^* + \frac{-\sigma_\xi \rho_{pw}\rho_{p\xi}}{1-\rho_{pw}^2} W_t^* + \sigma_\xi \sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2\rho_{p\xi}^2}{1-\rho_{pw}^2}} \cdot \omega_{3,t}. \quad \text{(W2)}$$

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 (W2) as proven in Theorem 2 in Web Appendix B.1, 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.

COPE method, the direct extension of $\text{Copula}_{\text{Origin}}$, does not require the uncorrelatedness between P_t^* and W_t for consistent model estimation, an assumption needed for $\text{Copula}_{\text{Origin}}$. However, similar to $\text{Copula}_{\text{Origin}}$, COPE requires the nonnormality of the endogenous regressor P_t to fulfill the full-rank identification assumption. In addition, COPE requires the nonnormality of all the exogenous regressors $W_t s$ to fulfill the full-rank identification assumption, while our proposed 2sCOPE method relaxes these assumptions.

[21] 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.

Web Appendix A.2: Assumption 5(b) in Copula$_{\text{Origin}}$

According to Park and Gupta (2012), under a Gaussian copula model for $(P_{1,t}, P_{2,t}, \xi_t)$, the structural model in Equation (13) with two endogenous regressors can be re-expressed as

$$\begin{aligned} Y_t =& \mu + P_{1,t}\alpha_1 + P_{2,t}\alpha_2 + W_t\beta + \sigma_\xi \frac{\rho_{\xi 1} - \rho_{12}\rho_{\xi 2}}{1-\rho_{12}^2} \cdot P_{1,t}^* + \sigma_\xi \frac{\rho_{\xi 2} - \rho_{12}\rho_{\xi 1}}{1-\rho_{12}^2} \cdot P_{2,t}^* \\ &+ \sigma_\xi \cdot \sqrt{1-\rho_{\xi 1}^2 - \frac{(\rho_{\xi 2} - \rho_{12}\rho_{\xi 1})^2}{1-\rho_{12}^2}} \cdot \omega_t. \end{aligned} \tag{W3}$$

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 $P_{2,t}^*$. For the OLS estimation of Equation (W3) to yield consistent estimates, W_t need also be uncorrelated with ω_t, which requires that $Cov(W_t, \sigma_\xi \cdot \sqrt{1-\rho_{\xi 1}^2 - \frac{(\rho_{\xi 2}-\rho_{12}\rho_{\xi 1})^2}{1-\rho_{12}^2}} \cdot \omega_t) = -Cov(W_t, \frac{\rho_{\xi 1}-\rho_{12}\rho_{\xi 2}}{1-\rho_{12}^2} \cdot P_{1,t}^* + \frac{\rho_{\xi 2}-\rho_{12}\rho_{\xi 1}}{1-\rho_{12}^2} \cdot P_{2,t}^*) = 0$ (Assumption 5(b) in the main text) 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 Copula$_{\text{Origin}}$.

WEB APPENDIX B: PROOFS FOR 2SCOPE

Web Appendix B.1: Proof of Theorem 2 Consistency of 2sCOPE

We have shown the derivation of 2sCOPE method in the main text. The system of equations used in the 2sCOPE method (Equations 8, 9) leads to the following equations

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

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

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, which satisfies the population orthogonality condition required for consistency of OLS (OLS.1 assumption in Wooldridge 2010). 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 consistency of OLS (OLS.2 assumption in Wooldridge 2010). **Theorem proved**.

2sCOPE in Multiple Exogenous Regressors Case 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. When there are one endogenous regressor P_t and two exogenous regressors W_t, the linear regression becomes:

$$Y_t = \beta_0 + \beta_1 P_t + \beta_2 W_{1,t} + \beta_3 W_{2,t} + \xi_t \tag{W4}$$

Under the Gaussian Copula assumption,

$$\begin{pmatrix} P_t^* \\ W_{1,t}^* \\ W_{2,t}^* \\ \xi_t^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_\xi \\ \rho_1 & 1 & \rho_w & 0 \\ \rho_2 & \rho_w & 1 & 0 \\ \rho_\xi & 0 & 0 & 1 \end{bmatrix} \right) \tag{W5}$$

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^*,

$$\xi_t = \sigma_\xi \xi_t^* = \frac{\sigma_\xi\rho_\xi(1-\rho_w^2)}{1-\rho_1^2-\rho_2^2+2\rho_1\rho_2\rho_w+\rho_w^2}\left(P_t^* - \frac{\rho_1-\rho_2\rho_w}{1-\rho_w^2}W_{1,t}^* - \frac{\rho_2-\rho_1\rho_w}{1-\rho_w^2}W_{2,t}^* \right) + \sigma_\xi\gamma\cdot\omega_{4,t}. \tag{W6}$$

Then we derive the first-stage regression of 2sCOPE

$$\begin{aligned} P_t^* &= \frac{\rho_1-\rho_2\rho_w}{1-\rho_w^2}W_{1,t}^* + \frac{\rho_2-\rho_1\rho_w}{1-\rho_w^2}W_{2,t}^* + \sqrt{1-\rho_1^2-\frac{(\rho_2-\rho_1\rho_w)^2}{1-\rho_w^2}}\omega_{3,t} \\ &= \frac{\rho_1-\rho_2\rho_w}{1-\rho_w^2}W_{1,t}^* + \frac{\rho_2-\rho_1\rho_w}{1-\rho_w^2}W_{2,t}^* + \epsilon_{2,t} \\ &= \gamma_1 W_{1,t}^* + \gamma_2 W_{2,t}^* + \epsilon_{2,t}. \end{aligned} \tag{W7}$$

The structural error ξ_t in Equation (W4) 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:

$$Y_t = \beta_0 + \beta_1 P_t + \beta_2 W_{1,t} + \beta_3 W_{2,t} + \beta_4 \epsilon_{2,t} + \sigma_\xi \gamma \cdot \omega_{4,t}. \tag{W8}$$

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 (W8). Thus, 2sCOPE that performs OLS regression of Equation (W8) yields consistent model estimates. Without loss of generality, the result can be extended to cases with any dimension of W_t.

2sCOPE in Multiple Endogenous Regressors Case

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:

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

$$\begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \\ \omega_{3,t} \\ \omega_{4,t} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right), \tag{W9}$$

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 (13) can be decomposed as

$$\xi_t = \sigma_\xi \xi^*_t = \eta_1 P^*_{1,t} + \eta_2 P^*_{2,t} - (\eta_1\rho_{wp1} + \eta_2\rho_{wp2})W^*_t + \sigma_\xi \cdot m \cdot \omega_{4,t}. \tag{W10}$$

where

$$\begin{aligned} \eta_1 &= \frac{\sigma_\xi\rho_{\xi p1}(1-\rho_{wp2}^2) - \sigma_\xi\rho_{\xi p2}(\rho_p - \rho_{wp1}\rho_{wp2})}{1-\rho_p^2-\rho_{wp1}^2-\rho_{wp2}^2+2\rho_p\rho_{wp1}\rho_{wp2}}, \\ \eta_2 &= \frac{\sigma_\xi(\rho_{wp1}\rho_{wp2}\rho_{\xi p1} + \rho_{\xi p2} - \rho_p\rho_{\xi p1} - \rho_{wp1}^2\rho_{\xi p2})}{1-\rho_p^2-\rho_{wp1}^2-\rho_{wp2}^2+2\rho_p\rho_{wp1}\rho_{wp2}}. \end{aligned} \tag{W11}$$

The 2sCOPE method with one endogenous regressor in Equation (12) is then extended to

$$
\begin{aligned}
Y_t &= \mu + P_{1,t}\alpha_1 + P_{2,t}\alpha_2 + W_t\beta + \eta_1\epsilon^*_{1,t} + \eta_2\epsilon^*_{2,t} + \sigma_\xi \cdot m \cdot \omega_{4,t},\\
\epsilon_{1,t} &= P^*_{1,t} - \rho_{wp1}W^*_t,\\
\epsilon_{2,t} &= P^*_{2,t} - \rho_{wp2}W^*_t.
\end{aligned}
$$

The main model in the first equation above is the same as Equation (17). The new error term $\omega_{4,t}$ is uncorrelated with all the regressors on the right-hand side of Equation (17). Thus, the OLS estimation of Equation (17) provides consistent estimates of structural regression model parameters $(\mu, \alpha_1, \alpha_2, \beta)$.

Web Appendix B.2: Proof of Theorem 3 Nonnormality Assumption Relaxed

In this section, we prove that our proposed 2sCOPE method can relax the nonnormality assumption on the endogenous regressors imposed in $\text{Copula}_{\text{Origin}}$, while COPE does not.

We first examine the COPE method in Equation (W2),

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

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 (12).

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

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

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.**

Web Appendix B.3: Variance Reduction Proposition of 2sCOPE

Proposition 1. ***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.*

According to the COPE method in Equation (W2),

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

The coefficients of P_t^* and W_t^* follow 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 12), 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,

$$\delta_4 + \hat{\rho}_{pw}\delta_3 = 0. \tag{W12}$$

In this case, the two-stage copula method (2sCOPE) can be viewed as one kind of restricted least squares estimation based on COPE. We next prove that restricted least squares can achieve reductions in standard errors. Suppose we simplify the regression expression in Equation (W2) 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 (W12) becomes

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

Thus, the 2sCOPE yields the least squares estimates $\hat{\theta}_2$ of Equation (W2) subject to the above restriction, whereas COPE yields the unrestricted least squares 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 squares 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.**

WEB APPENDIX C: 2SCOPE FOR SLOPE ENDOGENEITY

In this section, we describe the 2sCOPE approach to addressing slope endogeneity with correlated regressors in the following model:

$$Y_t \;=\; \mu + P_t\alpha_t + W_t'\beta_t + \eta_t, \qquad \text{where } \alpha_t = \bar{\alpha} + \xi_t, \tag{W13}$$

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.

$$\begin{aligned}
Y_t &= \mu + P_t\Big(\bar{\alpha} + \frac{\sigma_\xi\rho_{p\xi}}{1-\rho_{pw}^2}P_t^* + \frac{-\sigma_\xi\rho_{pw}\rho_{p\xi}}{1-\rho_{pw}^2}W_t^* + \sigma_\xi\sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2\rho_{p\xi}^2}{1-\rho_{pw}^2}}\,\omega_{3,t}\Big) + W_t'\beta_t + \eta_t \\
&= \mu + P_t\bar{\alpha} + \frac{\sigma_\xi\rho_{p\xi}}{1-\rho_{pw}^2}P_t \times P_t^* + \frac{-\sigma_\xi\rho_{pw}\rho_{p\xi}}{1-\rho_{pw}^2}P_t \times W_t^* + W_t'\beta_t + \\
&\quad \sigma_\xi\sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2\rho_{p\xi}^2}{1-\rho_{pw}^2}}\,P_t \times \omega_{3,t} + \eta_t.
\end{aligned} \tag{W14}$$

Given both $P_t \times P_t^*$ and $P_t \times W_t^*$ in Equation (W14), 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 (W14) 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

$$\begin{aligned} Y_t &= \mu + P_t(\bar{\alpha} + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2}\epsilon_t + \sigma_\xi \sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2\rho_{p\xi}^2}{1-\rho_{pw}^2}} \cdot \omega_{3,t}) + W_t'\beta_t + \eta_t \\ &= \mu + P_t\bar{\alpha} + \frac{\sigma_\xi \rho_{p\xi}}{1-\rho_{pw}^2} P_t^* \times \epsilon_t + W_t'\beta_t + \sigma_\xi \sqrt{1-\rho_{p\xi}^2 - \frac{\rho_{pw}^2\rho_{p\xi}^2}{1-\rho_{pw}^2}} P_t \times \omega_{3,t} + \eta_t, \end{aligned} \tag{W15}$$

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 (W15).

The 2sCOPE estimation 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.

WEB APPENDIX D: 2SCOPE FOR RANDOM COEFFICIENT LOGIT MODEL

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 2sCOPE 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.

$$\widehat{\delta}_{jt} = \mu_j + P'_{jt}\bar{\alpha} + W'_{jt}\bar{\beta} + \xi_{jt} + \eta_{jt}, \tag{W16}$$

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 2sCOPE to correct for the endogenous bias, which can avoid the potential bias of $\text{Copula}_{\text{Origin}}$ due to the potential correlations between P and W, as well as make use of this correlation to relax the nonnormality 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 2sCOPE to the setting with correlated regressors and (nearly) normal regressor distributions.

WEB APPENDIX E: ADDITIONAL SIMULATION RESULTS

Web Appendix E.1: Additional Results for Smaller Sample Size for Case 1

In the simulation study case 1, we use the sample size T=1000. Here we further check the robustness of results with respect to a smaller sample size. We simulate 1000 datasets, each of which has the sample size T=200, and use the same DGP as described in Case 1. Table W1 shows that 2sCOPE has unbiased estimates for a small sample size T=200. Hence, our proposed method is robust and can be applied to small sample sizes.

Table W1: Results of the Simulation Study for Case 1 with Sample Size of 200

			OLS			Copula$_{\text{Origin}}$			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	

Web Appendix E.2: Multiple Endogenous Regressors

In this case, we examine the performance of our proposed 2sCOPE when the model has multiple endogenous regressors. Specifically, we use the DGP with two endogenous regressors and one exogenous regressor that is correlated with the endogenous regressors below:

$$\begin{pmatrix} P^*_{1,t} \\ P^*_{2,t} \\ W^*_t \\ \xi^*_t \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.3 & 0.4 & 0.5 \\ 0.3 & 1 & 0.4 & 0.5 \\ 0.4 & 0.4 & 1 & 0 \\ 0.5 & 0.5 & 0 & 1 \end{bmatrix} \right), \tag{W17}$$

$$\xi_t = G^{-1}(U_{\xi,t}) = G^{-1}(\Phi(\xi^*_t)) = \Phi^{-1}(\Phi(\xi^*)) = 1 \cdot \xi^*_t, \tag{W18}$$

$$P_{1,t} = H_1^{-1}(U_{p1}) = H_1^{-1}(\Phi(P^*_{1,t})), \;\; P_{2,t} = H_2^{-1}(\Phi(P^*_{2,t})), \tag{W19}$$

$$W_t = L^{-1}(U_{W,t}) = L^{-1}(\Phi(W^*_t)), \tag{W20}$$

$$Y_t = \mu + \alpha \cdot P_t + \beta \cdot W_t + \xi_t = 1 + 1 \cdot P_{1,t} + 1 \cdot P_{2,t} + (-1) \cdot W_t + \xi_t, \tag{W21}$$

where $H_1^{-1}(\cdot)$, $H_2^{-1}(\cdot)$ and $L^{-1}(\cdot)$ are the inverse distribution functions of the Gamma(1,1), t(30) and Exp(1) distributions used to generate these regressors. We generate 1000 datasets, each of which has a sample size T=1000.

Table W2 shows the estimation results. First, the OLS estimates are biased. The COPE, the extended $\text{Copula}_{\text{Origin}}$, estimates are biased as well because of the close-to-normal endogenous regressor, t(30). However, our proposed 2sCOPE method provides unbiased estimates for all parameters, indicating that 2sCOPE performs well with multiple endogenous regressors, even for close-to-normal endogenous regressors. Moreover, 2sCOPE provides a much smaller d-error (0.002695) compared with COPE (0.006943), indicating that 2sCOPE can largely increase the estimation efficiency. The efficiency gain is 61.2% in this case.

Table W2: Results of the Simulation Study: Multiple Endogenous Regressors.

		OLS			COPE			2sCOPE		
Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
μ	1	0.903	0.040	2.426	1.002	0.079	0.019	1.016	0.076	0.213
α_1	1	1.436	0.029	14.88	0.998	0.058	0.032	0.995	0.058	0.079
α_2	1	1.487	0.025	19.23	1.367	0.479	0.765	1.028	0.141	0.197
β	-1	-1.338	0.029	11.76	-1.000	0.055	0.007	-1.010	0.054	0.196
$\rho_{\xi p1}$	0.5	-	-	-	0.394	0.136	0.781	0.501	0.042	0.029
$\rho_{\xi p2}$	0.5	-	-	-	0.110	0.422	0.923	0.472	0.095	0.295
σ_ξ	1	0.742	0.017	15.51	0.992	0.166	0.050	0.993	0.073	0.093
D-error						0.006943			0.002695	

Web Appendix E.3: Multiple Exogenous Control Covariates

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

$$\begin{pmatrix} \xi_t^* \\ P_{1,t}^* \\ W_{1,t}^* \\ P_{2,t}^* \\ W_{2,t}^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho & 0 & \rho & 0 \\ \rho & 1 & q & 0 & 0 \\ 0 & q & 1 & 0 & 0 \\ \rho & 0 & 0 & 1 & q \\ 0 & 0 & 0 & q & 1 \end{bmatrix} \right), \tag{W22}$$

$$\xi_t = G^{-1}(\Phi(\xi_t^*)) = \Phi^{-1}(\Phi(\xi^*)) = 1 \cdot \xi_t^*, \tag{W23}$$

$$P_{1,t} = \chi^2(2)^{-1}(\Phi(P_{1,t}^*)), \quad P_{2,t} = \chi^2(2)^{-1}(\Phi(P_{2,t}^*)), \tag{W24}$$

$$W_{1,t} = \Phi^{-1}(\Phi(W_{1,t}^*)), \tag{W25}$$

$$W_{2,t} = \begin{cases} 1, & \text{if} \quad \Phi(W_{2,t}^*) \geq 0.5 \\ 0. & \text{if} \quad \Phi(W_{2,t}^*) < 0.5 \end{cases}, \tag{W26}$$

$$Y_t = \mu + \alpha_1 \cdot P_{1,t} + \alpha_2 \cdot P_{2,t} + \beta_1 \cdot W_{1,t} + \beta_2 \cdot W_{2,t} + \xi_t \tag{W27}$$

$$= 1 + P_{1,t} + P_{2,t} + (-1) \cdot W_{1,t} + (-1) \cdot W_{2,t} + \xi_t, \tag{W28}$$

where $W_{1,t}$ is normally distributed and $W_{2,t}$ is a binary variable that follows a Bernoulli distribution. ρ is set to 0.4, and q is set to two cases, $\{0.3, 0.6\}$. We set the sample size $T =$ 1000 and generate 1000 datasets to estimate parameters using OLS and copula methods. For binary W, we compute W^* in the same way as for the continuous case, $W^* = \Phi^{-1}(F(W_t))$, where F is the cdf function of W.

The estimation results for the multiple-exogenous-regressor case with both discrete and continuous ones are summarized in Table W3. The OLS and Copula$_{\text{Origin}}$ estimates are biased because of endogeneity and correlated exogenous regressors, respectively. The proposed 2sCOPE method performs well and provides consistent estimates for all parameters. This indicates that our proposed method performs well with multiple exogenous correlated regressors. Moreover, correcting for endogeneity using our proposed method does not require every exogenous correlated regressor to be informative (i.e., continuously distributed) and nonnormally distributed.

Table W3: Results of the Simulation Study: Multiple Exogenous Control Covariates.

			OLS			Copula$_{\text{Origin}}$			COPE			2sCOPE		
q	Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
0.3	μ	1	0.312	0.055	12.54	1.100	0.098	1.012	1.100	0.101	0.983	1.000	0.091	0.001
	α_1	1	1.195	0.015	12.86	1.000	0.031	0.006	1.000	0.031	0.008	1.000	0.031	0.014
	α_2	1	1.191	0.015	12.57	1.000	0.032	0.002	1.000	0.033	0.002	1.000	0.032	0.001
	β_1	-1	-1.104	0.028	3.726	-1.130	0.027	4.897	-1.145	0.694	0.209	-0.999	0.036	0.015
	β_2	-1	-1.167	0.055	3.052	-1.206	0.053	3.882	-1.206	0.053	3.884	-1.003	0.070	0.042
	$\rho_{P_1,\xi}$	0.4	-	-	-	0.430	0.053	0.561	0.371	0.152	0.188	0.396	0.050	0.084
	$\rho_{P_2,\xi}$	0.4	-	-	-	0.417	0.055	0.307	0.364	0.078	0.468	0.398	0.052	0.042
0.6	μ	1	0.236	0.052	14.63	1.256	0.096	2.667	1.256	0.098	2.609	1.005	0.083	0.056
	α_1	1	1.256	0.017	14.68	0.999	0.032	0.037	0.999	0.032	0.033	1.000	0.031	0.016
	α_2	1	1.222	0.016	13.67	0.996	0.029	0.136	0.996	0.029	0.137	0.996	0.029	0.137
	β_1	-1	-1.277	0.032	8.620	-1.373	0.031	12.14	-1.367	0.629	0.584	-1.002	0.047	0.039
	β_2	-1	-1.379	0.057	6.621	-1.497	0.053	9.306	-1.497	0.053	9.314	-0.993	0.082	0.082
	$\rho_{P_1,\xi}$	0.4	-	-	-	0.566	0.045	3.659	0.487	0.232	0.376	0.396	0.036	0.110
	$\rho_{P_2,\xi}$	0.4	-	-	-	0.477	0.047	1.630	0.437	0.082	0.451	0.403	0.040	0.071

We further show the advantage of using 2sCOPE, compared with the direct extension of Copula$_{\text{Origin}}$ (COPE), in estimation efficiency and consistency in high-dimensional W using simulation. In particular, we use some commonly used distributions for the exogenous regressors Ws. The data-generating process (DGP) is summarized below:

$$
\begin{pmatrix} P_t^* \\ W_t^* \\ \xi_t^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0_{10} \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{pw} & \rho_{p\xi} \\ \rho_{pw} & \Sigma_w & 0 \\ \rho_{p\xi} & 0 & 1 \end{bmatrix} \right)
$$

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

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

$$
Y_t = \mu + \alpha \cdot P_t + \beta \cdot W_t + \xi_t = 1 + 1 \cdot P_t + (-1) \cdot W_t + \xi_t,
$$

where ξ_t^* and P_t^* are correlated ($\rho_{p\xi} = 0.5$), generating the endogeneity problem; W_t^* is a 8-dimensional exogenous regressors uncorrelated with ξ_t^*; Each exogenous regressor in W_t^* is correlated with P_t^* with $\rho_{pw} = 0.5$; Σ_w denotes the covariance matrix of W_t^* with all the diagonal items equal to one and all non-diagonal items equal to $\rho_w = 0.3$. We set the sample size $T = 1000$, and generate 1000 datasets as replicates using the DGP above. In the simulation, we use normal distribution $N(0,1)$ for P_t, and the eight distributions, Exp(1), t(2), binary, mixnorm, Gamma(1,1), truncated-normal, lognorm(0,1), Cauchy(0,0.5), in sequence for the 8-dimentional W_t.

Table W4 summarizes the estimation results, and confirms that 2sCOPE outperforms COPE in several dimensions. First, the estimated coefficient of the endogenous regressor for COPE is 2.360, which is far away from the true value, indicating that COPE cannot handle normally-distributed endogenous regressor, while 2sCOPE can provide unbiased estimate. Second, COPE estimates of some exogenous regressors with certain distributions are biased (31.9% bias for binary W, 23.5% bias for mix-normal W and 25.7% bias for truncated normal W), indicating that COPE is sensitive to the distributions of exogenous regressors included in the model, making all the estimates vulnerable. Third, the D-error of COPE and 2sCOPE estimates is 0.002531 and 0.000534 respectively, indicating that 2sCOPE is much more efficient than COPE and increases the efficiency by 78.9%. Adding too many generated regressors in the model, as COPE does, will significantly decrease the estimation efficiency. In this section, we illustrate using the exactly normally distributed endogenous regressor as an example. Please refer to Table W12 for weakly-nonnormal P case.

Table W4: The Performance of 2sCOPE with large-dimension of W

			OLS			COPE			2sCOPE		
Distribution	Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
	μ	1	1.848	0.051	16.73	1.281	0.228	1.229	1.040	0.089	0.453
N(0,1)	α	1	2.170	0.036	32.75	2.360	0.496	2.741	1.057	0.101	0.565
Exp(1)	β_1	-1	-1.210	0.024	8.862	-0.999	0.044	0.029	-1.009	0.035	0.245
t(2)	β_2	-1	-1.066	0.025	2.608	-1.001	0.013	0.107	-1.004	0.012	0.313
binary	β_3	-1	-1.363	0.045	8.119	-1.319	0.039	8.193	-1.014	0.073	0.196
mix-norm	β_4	-1	-1.233	0.022	10.76	-1.235	0.432	0.544	-1.012	0.039	0.311
Gamma(1,1)	β_5	-1	-1.212	0.024	8.812	-1.001	0.043	0.017	-1.011	0.035	0.313
truncated-N(0,1)	β_6	-1	-1.261	0.024	10.81	-1.257	0.470	0.547	-1.012	0.043	0.270
lognorm(0,1)	β_7	-1	-1.078	0.017	4.425	-1.000	0.015	0.025	-1.004	0.014	0.286
cauchy(0,0.5)	β_8	-1	-1.005	0.006	0.883	-1.000	0.002	0.023	-1.000	0.002	0.146
	ρ	0.5	-	-	-	-0.392	0.394	2.262	0.487	0.022	0.580
	σ_ξ	1	0.644	0.016	22.33	1.191	0.341	0.559	0.968	0.055	0.580
Bias				0.345			0.246			0.016	
RMSE				0.347			0.327			0.047	
D-error				0.000424			0.002531			0.000534	

Web Appendix E.4: Misspecification of ξ_t

Similar to Park and Gupta (2012), we assume the structural error ξ_t to be normally distributed, a reasonable and commonly used assumption in marketing and economics literature. However, the true distribution of ξ_t is often unknown. Thus, in this simulation study, we examine the robustness of 2sCOPE to the departures from the normality of ξ_t. We generate 1,000 datasets using the same multivariate normal distribution as in Equation (20). The rest of DGP is:

$$\xi_t = G^{-1}(U_{\xi,t}) \quad = \quad G^{-1}(\Phi(\xi_t^*)), \tag{W29}$$

$$P_t = H^{-1}(U_{p,t}) = H^{-1}(\Phi(P_t^*)), \qquad W_t = L^{-1}(U_{w,t}) = L^{-1}(\Phi(W_t^*)), \tag{W30}$$

$$Y_t = \mu + \alpha \cdot P_t + \beta \cdot W_t \quad + \quad \xi_t = 1 + 1 \cdot P_t + (-1) \cdot W_t + \xi_t, \tag{W31}$$

where we set $P_t \sim Gamma(1, 1)$ and $W_t \sim Exp(1)$ in the simulation. We check the robustness of the structural error ξ_t using different distributions (e.g., a uniform distribution, beta distribution and t distribution) instead of a normal distribution. For estimation, we assume normality of ξ_t and use the OLS estimator, $\text{Copula}_{\text{Origin}}$ and the proposed 2sCOPE method.

Table W5 reports estimation results. As shown in Table W5, 2sCOPE can recover the true parameter values despite the misspecification of ξ_t, demonstrating the robustness of the proposed 2sCOPE method to the normal error assumption.

Table W5: Results of the Simulation Study: Misspecification of ξ_t

			OLS			2sCOPE		
Distribution of ξ_t	Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}
U[-0.5,0.5]	μ	1	0.912	0.013	6.808	1.002	0.017	0.105
	α	1	1.160	0.010	16.41	0.996	0.017	0.233
	β	-1	-1.072	0.009	8.033	-0.998	0.011	0.147
	$\rho_{p\xi}$	0.5	-	-	-	0.495	0.035	0.155
	σ_ξ	0.289	0.251	0.004	9.018	0.290	0.008	0.197
Beta(0.5,0.5)	μ	1	0.896	0.016	6.461	1.003	0.020	0.145
	α	1	1.190	0.012	15.72	0.994	0.018	0.318
	β	-1	-1.086	0.011	7.763	-0.998	0.014	0.183
	$\rho_{p\xi}$	0.5	-	-	-	0.481	0.033	0.593
	σ_ξ	0.354	0.311	0.005	9.046	0.356	0.009	0.258
Beta(4,4)	μ	1	0.948	0.008	6.928	1.000	0.010	0.009
	α	1	1.095	0.006	16.61	1.000	0.010	0.044
	β	-1	-1.043	0.005	8.149	-1.000	0.007	0.030
	$\rho_{p\xi}$	0.5	-	-	-	0.499	0.037	0.025
	σ_ξ	0.167	0.144	0.003	7.969	0.167	0.006	0.011
t (df=3)	μ	1	0.504	0.082	6.071	0.983	0.127	0.135
	α	1	1.903	0.089	10.13	1.024	0.217	0.110
	β	-1	-1.410	0.064	6.448	-1.012	0.109	0.111
	$\rho_{p\xi}$	0.5	-	-	-	0.454	0.069	0.676
	σ_ξ	1.732	1.503	0.231	0.992	1.698	0.244	0.141
t (df=5)	μ	1	0.603	0.059	6.723	0.997	0.080	0.039
	α	1	1.727	0.053	13.65	1.006	0.113	0.057
	β	-1	-1.328	0.043	7.642	-1.002	0.067	0.037
	$\rho_{p\xi}$	0.5	-	-	-	0.486	0.047	0.292
	σ_ξ	1.291	1.118	0.049	3.506	1.289	0.070	0.032

We further examine the performance of 2sCOPE with misspecification of ξ under normally distributed endogenous regressor case. The estimation result in Table W6 shows that 2sCOPE can even work for normal endogenous regressor cases under misspecification of ξ.

Table W6: Results of the Simulation Study: Misspecification of ξ with Normal Endogenous Regressor

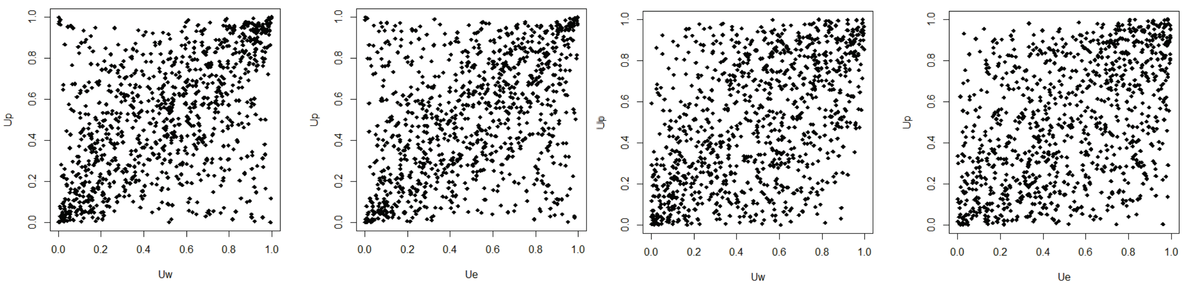
			OLS			2sCOPE		
Distribution of ξ	Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}
U[-0.5,0.5]	μ	1	1.080	0.012	6.778	1.002	0.020	0.086
	α	1	1.178	0.008	23.28	1.004	0.037	0.116
	β	-1	-1.080	0.009	9.070	-1.002	0.018	0.113
	$\rho_{p\xi}$	0.5	-	-	-	0.475	0.072	0.344
	σ_ξ	0.289	0.241	0.004	10.64	0.288	0.018	0.054
Beta(0.5,0.5)	μ	1	1.095	0.015	6.483	1.003	0.025	0.116
	α	1	1.211	0.009	22.85	1.007	0.045	0.157
	β	-1	-1.095	0.011	8.560	-1.003	0.022	0.130
	$\rho_{p\xi}$	0.5	-	-	-	0.457	0.075	0.573
	σ_ξ	0.354	0.299	0.005	10.51	0.352	0.021	0.088
Beta(4,4)	μ	1	1.047	0.007	6.962	1.002	0.012	0.146
	α	1	1.105	0.005	22.30	1.004	0.021	0.194
	β	-1	-1.047	0.005	9.219	-1.002	0.011	0.157
	$\rho_{p\xi}$	0.5	-	-	-	0.479	0.072	0.294
	σ_ξ	0.167	0.138	0.003	9.969	0.165	0.011	0.159
T(df=3)	μ	1	1.443	0.075	5.884	1.022	0.133	0.165
	α	1	1.988	0.077	12.82	1.052	0.260	0.201
	β	-1	-1.443	0.059	7.508	-1.021	0.125	0.168
	$\rho_{p\xi}$	0.5	-	-	-	0.438	0.089	0.695
	σ_ξ	1.732	1.461	0.221	1.226	1.692	0.247	0.161
T(df=5)	μ	1	1.358	0.052	6.932	1.012	0.090	0.135
	α	1	1.795	0.045	17.66	1.030	0.168	0.176
	β	-1	-1.359	0.041	8.795	-1.012	0.082	0.151
	$\rho_{p\xi}$	0.5	-	-	-	0.472	0.075	0.376
	σ_ξ	1.291	1.073	0.046	4.726	1.277	0.096	0.145

Web Appendix E.5: Misspecification of Copula

In the proposed method, 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 assumption in capturing the dependence among the endogenous regressors, exogenous regressors and the error term using simulated data. Specifically, we generate the dependence among U_p, U_w and U_ξ using copula models other than the Gaussian copula. Our simulation setting requires the availability of a random number generation routine from a tri-variate copula model other than Gaussian copula with non-homogeneous correlations among the three variables. Among copula models other than Gaussian copula, we find only T copula has this flexibility of providing flexible random number generation from arbitrary and heterogeneous correlation structures among more than two variables. We thus consider using the following T copula models in which

$$C(U_p, U_w, U_\xi) = \int_{-\infty}^{t_\nu^{-1}(U_p)} \int_{-\infty}^{t_\nu^{-1}(U_w)} \int_{-\infty}^{t_\nu^{-1}(U_\xi)} \frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\pi\nu)^d|\Sigma|}} \left(1 + \frac{x'\Sigma^{-1}x}{\nu}\right) dx, \qquad \text{(W32)}$$

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 the covariance matrix capturing correlations among variables. The data-generating process



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

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

Figure W1: Scatter plots of Randomly Generated Pairs U_p, U_w (U_p,U_ξ) for Considered Copulas.

(DGP) of t copula is summarized below:

$$\begin{pmatrix} P_t^* \\ W_t^* \\ \xi_t^* \end{pmatrix} \sim t_\nu^d \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{pw} & \rho_{p\xi} \\ \rho_{pw} & 1 & 0 \\ \rho_{p\xi} & 0 & 1 \end{bmatrix} \right) = t_\nu^d \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix} \right). \quad \text{(W33)}$$

Figure W1 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:

$$\xi_t = G^{-1}(U_\xi) = \Phi^{-1}(U_\xi), \quad \text{(W34)}$$

$$P_t = H^{-1}(U_p), W_t = L^{-1}(U_w), \quad \text{(W35)}$$

$$Y_t = 1 + 1 \cdot P_t + (-1) \cdot W_t + \xi_t. \quad \text{(W36)}$$

where $H(\cdot)$ is a gamma distribution and $L(\cdot)$ is an exponential distribution. We set $T = 1000$, generate 1000 datasets and estimate the parameters using the OLS estimator and the proposed 2sCOPE method.

Table W7 summarizes the estimation results. OLS and $\text{Copula}_{\text{Origin}}$ estimates are still biased for all parameters. By contrast, estimates from the proposed 2sCOPE method are centered closely around the true values. Therefore, the proposed method based on the Gaussian copula is reasonably robust to the mis-specifications of the assumed copula dependence structure, while recognizing that a limitation of the simulation study is the restriction to the multivariate T copula that has the flexibility of generating multivariate data from arbitrary and heterogeneous correlation structures.

Table W7: Results of the Simulation Study: Misspecification of Copula

		OLS			2sCOPE		
Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}
μ	1	0.710	0.530	5.463	0.988	0.077	0.156
α	1	1.580	0.044	13.13	1.029	0.116	0.250
β	-1	-1.289	0.047	6.142	-1.017	0.070	0.248
$\rho_{p\xi}$	0.5	-	-	-	0.458	0.067	0.622
σ_{ξ}	1	0.864	0.026	5.236	0.988	0.054	0.230

Web Appendix E.6: Linear Dependence Among Regressors

When using 2sCOPE with Gaussian Copula, the implied relation among regressors is restricted to a specific non-linear form. In this section, we further check the performance of the 2sCOPE estimator when the regressors are linearly related to each other. Specifically, we consider the following DGP.

$$\begin{pmatrix} X_t^* \\ \\ \xi_t^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{p\xi} \\ \\ \rho_{p\xi} & 1 \end{bmatrix} \right) = N\left(\begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ \\ 0.5 & 1 \end{bmatrix} \right), \tag{W37}$$

$$\xi_t = G^{-1}(\Phi(\xi_t^*)) = \Phi^{-1}(\Phi(\xi^*)) = 1 \cdot \xi_t^*, \tag{W38}$$

$$W_t \sim N(0, 1), \tag{W39}$$

$$P_t = \gamma W_t + F^{-1}_{\chi^2(1)}(\Phi(X_t^*)), \tag{W40}$$

$$Y_t = \mu + \alpha \cdot P_t + \beta \cdot W_t + \xi_t = 1 + 1 \cdot P_t + (-1) \cdot W_t + \xi_t \tag{W41}$$

We set the sample size $T = 1000$, and generate 1000 datasets as replicates using the DGP above. Equation (W40) shows the linear dependence among P and W, which violates the assumption of the 2sCOPE estimator. However, the estimation result in Table W8 shows that 2sCOPE estimation method can still get consistent estimates. Therefore, the results demonstrate the robustness of the proposed 2sCOPE method based on Gaussian copula, which implicitly requires a non-linear relationship among regressors, to the linear dependence among regressors.

Table W8: Results of the Simulation Study: Linear Dependence Among Regressors

			OLS			2sCOPE		
γ	Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}
0.6	μ	1	0.545	0.041	11.07	0.967	0.084	0.389
	α	1	1.226	0.016	14.48	1.015	0.039	0.383
	β	-1	-1.136	0.030	4.607	-1.010	0.038	0.270
	σ_ξ	1	0.891	0.021	5.176	0.987	0.041	0.305
1.2	μ	1	0.546	0.042	10.74	0.969	0.104	0.294
	α	1	1.227	0.016	14.37	1.015	0.050	0.294
	β	-1	-1.273	0.036	7.653	-1.019	0.069	0.274
	σ_ξ	1	0.892	0.020	5.554	0.990	0.048	0.217

Web Appendix E.7: Test Assumption 5(b)

As shown in the METHODS section, when P contains only one endogenous regressor, Assumption 5 (W and P^* are uncorrelated) has to be satisfied for Copula$_{\text{Origin}}$ to yield consistent estimates. In the multiple-endogenous-regressors case, W should be uncorrelated with the CCF term, the linear combination of P^*s (Assumption 5(b)). Assumption 5 for a single endogenous regressor is easy to check, while Assumption 5(b) is not that obvious. In this subsection, we describe how to test Assumption 5(b). Specifically, we consider two simulation scenarios: one satisfies Assumption 5(b) while the other doesn't. In addition, we will show that Copula$_{\text{Origin}}$ performs better if Assumption 5(b) is satisfied, and 2sCOPE is preferred if the assumption is violated. The data-generating process is summarized below.

$$\begin{pmatrix} P^*_{1,t} \\ P^*_{2,t} \\ W^*_{1,t} \\ W^*_{2,t} \\ \xi^*_t \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & p & q_1 & q_1 & \rho_1 \\ p & 1 & q_2 & q_2 & \rho_2 \\ q_1 & q_2 & 1 & q_{ww} & 0 \\ q_1 & q_2 & q_{ww} & 1 & 0 \\ \rho_1 & \rho_2 & 0 & 0 & 1 \end{bmatrix} \right), \tag{W42}$$

$$\xi_t = \Phi^{-1}(\Phi(\xi^*)) = 1 \cdot \xi^*_t, \quad W_t = L^{-1}(\Phi(W^*_t)), \tag{W43}$$

$$P_{1,t} = H^{-1}(\Phi(P^*_{1,t})), \quad P_{2,t} = H^{-1}(\Phi(P^*_{2,t})), \tag{W44}$$

$$Y_t = \mu + \alpha_1 \cdot P_{1,t} + \alpha_2 \cdot P_{2,t} + \beta \cdot W_t + \xi_t = 1 + P_{1,t} + P_{2,t} + (-1) \cdot W_t + \xi_t, \tag{W45}$$

where $P_t \sim Gamma(1,1)$ and $W_t \sim Exp(1)$ in both scenarios. The two scenarios differ in the covariance matrix in W42.

In Scenario 1, we set $p = 0$, $q_1 = q_2 = 0.4$, $q_{ww} = 0.2$, $\rho_1 = 0.5$ and $\rho_2 =$ -0.5;

In Scenario 2, we set $p = 0$, $q_1 = q_2 = 0.4$, $q_{ww} = 0.2$, $\rho_1 = 0.5$ and $\rho_2 = 0.5$.

We set $T = 1000$, generate 1000 datasets and estimate the parameters using the $\text{Copula}_{\text{Origin}}$ and 2sCOPE methods. To test the assumption 5(b) for $\text{Copula}_{\text{Origin}}$, we first estimate the coefficients of P_1^* and P_2^* ($\widehat{\gamma}_1$ and $\widehat{\gamma}_2$) using $\text{Copula}_{\text{Origin}}$, and obtain the CCF by calculating $\widehat{\gamma}_1 P_1^* + \widehat{\gamma}_2 P_2^*$. Then we check $\text{Cor}(W, \text{CCF})$, the correlation between W and the CCF=$\hat{\gamma}_1 \hat{P}_1^* + \hat{\gamma}_2 \hat{P}_2^*$ for each W. We use Fisher's Z test to test the null hypothesis of $\text{Cor}(W, \text{CCF}) = 0$. Assumption 5(b) is violated when the null hypothesis is rejected using the Fisher's Z test.

Table W9 summarizes the estimation results for the two scenarios with sample size N=1000. In Scenario 1, the average correlation between W_1 (W_2) and the CCF term across 1000 simulated datasets is -0.001316 (0.000369) with an average p-value of 0.493 (0.508), which means that the correlation is not significantly different from 0 and Assumption 5(b) holds. Correspondingly $\text{Copula}_{\text{Origin}}$ performs well with all the estimates centered closely around the true values. By contrast, estimates from $\text{Copula}_{\text{Origin}}$ are biased in Scenario 2, and the average correlation between W_1 (W_2) and the CCF term across 1000 simulated datasets is 0.504 (0.503) with the average p-value $< 2.2e^{-16}(< 2.2e^{-16})$, violating Assumption 5(b). In both scenarios, 2sCOPE provides unbiased estimates. However, when Assumption 5(b) holds, $\text{Copula}_{\text{Origin}}$ is more efficient than 2sCOPE with a smaller D-error ($0.001420 < 0.001611$) (Table W9), which means $\text{Copula}_{\text{Origin}}$ increases the estimation efficiency by 13.45%. When the dimension of W_t increases, we expect the efficiency gain of $\text{Copula}_{\text{Origin}}$ to be greater than that in this example.

To summarize, this subsection provides an example of how to test Assumption 5(b) with multiple endogenous regressors. When Assumption 5(b) holds, 2sCOPE can still be used with good performance but $\text{Copula}_{\text{Origin}}$ may be preferred over 2sCOPE, as the simpler $\text{Copula}_{\text{Origin}}$ procedure yields more efficient estimates with smaller D-error and $\text{Copula}_{\text{Origin}}$ is easier to understand, control, and communicate with (Little 2004). Otherwise, the proposed 2sCOPE method is preferred.

Table W9: Results of the Simulation Study: Testing Assumption 5(b)

			OLS			$\text{Copula}_{\text{Origin}}$			2sCOPE		
Simulation	Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
Scenario 1	μ	1	0.999	0.050	0.014	0.999	0.089	0.017	0.998	0.063	0.038
	α_1	1	1.454	0.033	13.94	0.998	0.060	0.026	1.000	0.059	0.003
	α_2	1	0.546	0.032	14.19	1.004	0.058	0.074	1.005	0.057	0.087
	β_1	-1	-1.000	0.029	0.010	-1.000	0.027	0.008	-1.002	0.039	0.050
	β_2	-1	-1.001	0.029	0.036	-1.001	0.026	0.040	-1.001	0.038	0.021
	ρ_1	0.5	-	-	-	0.499	0.049	0.014	0.499	0.034	0.041
	ρ_2	-0.5	-	-	-	-0.500	0.046	0.010	-0.501	0.032	0.019
	σ_ξ	1	0.769	0.017	13.86	1.004	0.043	0.084	1.002	0.043	0.049
	D-error					0.001420			0.001611		
	Cor(W_1, CCF), p value					-0.001316, 0.493					
	Cor(W_2, CCF), p value					0.000369, 0.508					
Scenario 2	μ	1	0.384	0.040	15.32	1.753	0.078	9.699	0.995	0.060	0.085
	α_1	1	1.763	0.031	24.26	1.161	0.047	3.426	1.001	0.040	0.038
	α_2	1	1.764	0.034	22.32	1.164	0.044	3.699	1.003	0.041	0.072
	β_1	-1	-1.456	0.025	18.58	-1.542	0.023	23.54	-1.000	0.031	0.013
	β_2	-1	-1.456	0.024	19.30	-1.541	0.023	23.78	-1.000	0.032	0.010
	ρ_1	0.5	-	-	-	0.666	0.032	5.239	0.491	0.034	0.260
	ρ_2	0.5	-	-	-	0.665	0.033	5.047	0.490	0.035	0.274
	σ_ξ	1	0.569	0.017	25.01	1.099	0.044	2.237	1.001	0.033	0.043
	Cor(W_1, CCF), p value					0.504, <2.2e-16					
	Cor(W_2, CCF), p value					0.503, <2.2e-16					

Web Appendix E.8: Simulation Experiments to Inform the Decision Tree of Using 2sCOPE

As noted in the main paper, as long as the sample size is sufficiently large and the assumptions of Theorems 2 and 3 are satisfied, 2sCOPE yields unbiased structural model estimates. However, in practical datasets with finite sample sizes, good performance of 2sCOPE may require sufficient nonnormality of regressors and sufficient relevance between P and W. Thus, in this section, we conduct systematic simulation studies to assess boundary conditions for using 2sCOPE in finite samples, and to inform the decision tree in Figure 2 of the main text.

To achieve this goal, in the simulation studies we systematically vary distributions of P and W, sample size, the endogeneity level, and the relevance level between P and W, to obtain empirically verifiable boundary conditions under which we can expect a good performance of 2sCOPE with a high probability. In the simulation studies, we use the Kolmogorov-Smirnov (KS) test to evaluate the regressor nonnormality for the following reasons. The KS test is one of the most commonly used tests for normality. The KS test statistic compares the empirical cumulative distribution of the standardized regressor with the CDF of the standard normal distribution, and is an overall and comprehensive measure to quantify nonnormality. Furthermore, more powerful normality tests, such as the Shapiro-Wilk test or Anderson-Darling test, may detect small departures from normality that are insufficient for the purpose of copula endogeneity correction (Cortina and Dunlap 1997, Eckert and Hohberger 2022, Ahad et al. 2011). Thus, among these most commonly used normality tests, we choose the relatively conservative KS test to be on the safe side. Last, because the performance of 2sCOPE improves with sample size when everything else is fixed, the measures for the sufficient nonnormality of regressors should also change with sample size: a minor departure from normality that is considered as insufficient nonnormality for a small sample can become sufficient for 2sCOPE to have good performance when the sample size is large. The p-value from the KS normality test satisfies this condition. Thus, we use the p-value from the KS test to inform sufficient nonnormality of regressors. Finally, we consider cases in which Assumptions 5 and 5(b) are violated because otherwise copula$_{\text{Origin}}$ should be used instead of 2sCOPE (Web Appendix E.7).

We first consider the sufficient condition of P in step 2 under the scenario when W is normally distributed, in which case the relevant W is expected to provide less help in identifying the causal effect of a close-to-normal endogenous regressor P than if W is sufficiently nonnormal. To identify the boundary condition of sufficient nonnormality of P for using 2sCOPE in this scenario, we conduct a factorial experiment using the data generating process from Equations (20-23) with a wide range of variations in sample size, endogeneity level, relevance, and distributions of regressors. Specifically, we simulate the endogenous regressor P using the nine nonnormal distributions in Figure 1, 14 levels of sample size {100, 200, 400, 600, 800, 1000, 2000, 4000, 6000, 8000, 10000, 20000, 40000, 60000}, eight endogeneity level $\rho_{p\xi}$ $\{0.1, 0.2, ..., 0.8\}$, and eight relevance level between P^* and W^* $\{0.1, 0.2, ..., 0.8\}$. This results in a total of $9 \times 14 \times 8 \times 8 = 8064$ cases. We generate 1000 datasets for each case, estimate the parameters using 2sCOPE, and calculate the average relative bias of $[\mu, \alpha, \beta]$ to measure the performance in each of the 8064 cases. In the end, we obtain 8064 observations in total. By examining the performance of 2sCOPE across all these cases, we evaluate the boundary condition of the nonnormality level of P for good performance of 2sCOPE. Table

W10 shows that in 5377 out of 8064 cases considered, the average p-value of the KS test of the normality of P over 1000 simulated datasets is less than 0.05 (i.e., rejecting the normality of P). As long as the p-value of the KS test of P is smaller than 0.05, 2sCOPE yields estimates with minor bias (relative bias less than 10%) with high probability ($\frac{5250}{5377}$ =97.6%) (Table W10). A relative bias of less than 10% is a commonly used criterion that indicates satisfactory performance of a method in terms of estimation bias (Forero, Maydeu-Olivares, and Gallardo-Pujol 2009, Holtmann et al. 2016, McNeish 2016, Wang and Kim 2017, Enders, Keller, and Levy 2018). Among the 5250 cases with an average relative bias less than 10%, the overall mean relative bias is small (1.0%) with a standard deviation (SD) of 1.8% over these 5250 cases (Table W10). Furthermore, among the 127 cases in which the average relative bias of the 2sCOPE estimates exceeds 10%, the bias is not large with a mean relative bias of 17% and a standard deviation of 7% (Table W10). Overall, we can conclude from this simulation experiment that 2sCOPE is expected to perform well in finite samples with high probabilities when the p-value from the KS test of the endogenous regressor P is less than 0.05.

Table W10: Condition for Sufficient Nonnormality of P to Use 2sCOPE in Step 2 of the Decision Tree.

KS Test P-value of Endogenous Regressor P	Number of Cases	Number of Cases Bias $\leq$ 10%	Number of Cases Bias > 10%	Percentage of Good Performance
< 0.05	5377	5250 (mean=1.0%, SD=1.8%)*	127 (mean=17%, SD=7.0%)*	97.6%

Note: *: Mean and standard deviation of the relative bias across cases are reported in the parenthesis.

Next, we consider the scenario when P fails the nonnormality test in step 2. Specifically, we consider the extreme case when P is normally distributed, and examine the sufficient condition of W for 2sCOPE to have good finite-sample performance. Similarly, to identify the sufficient condition of W, we simulate the exogenous regressor W using the nine nonnormal distributions in Figure 1, 14 levels of sample size {100, 200, 400, 600, 800, 1000, 2000, 4000, 6000, 8000, 10000, 20000, 40000, 60000}, eight endogeneity level $\rho_{p\xi}$ $\{0.1, 0.2, ..., 0.8\}$ and eight relevance level between P^* and W^* $\{0.1, 0.2, ..., 0.8\}$. This results in a total of $9 \times 14 \times 8 \times 8 = 8064$ cases. We generate 1000 datasets for each case, estimate the parameters using 2sCOPE, and calculate the average relative bias of $[\mu, \alpha, \beta]$ to measure the performance. In the end, we obtained 8064 observations in total. By examining the performance of 2sCOPE in all those simulation studies, we evaluate the sufficient condition of W for 2sCOPE to have good performance when P is normally distributed. Figure W2 shows the simulation results. When W is sufficiently nonnormal with the average p-value of KS test smaller than e^{-3} over 1000 simulated datasets, we only require a moderate relevance between P^* and W^* (average F stats > 10 for the effect of W^* on P^* in the first stage regression of 2sCOPE over 1000 simulated datasets) to have good performance of 2sCOPE (relative bias $\leq 10\%$) with a high probability ($\frac{4075}{4075+416} = 90.7\%$ in Figure W2): in 4075 out of 4491 (4075+416) cases in which W satisfies the sufficient nonnormality and sufficient relevance requirements above, 2sCOPE performs well with relative bias $\leq 10\%$. In other cases, we observe a considerably lower probability ($\frac{540}{3033+540}$ =15.1% [22]) of good finite-sample performance. The simulation experiment informs the sufficient condition of W for 2sCOPE to have good performance when the endogenous regressor P is normally distributed. Overall, this simulation experiment demonstrates that a combination of certain levels of nonnormality and relevance of W are needed to identify the normally distributed endogenous regressor with good finite-sample

[22]The 15.1% is the worst case in extreme scenario (P is set to be normally distributed). In practice when P is close-to-normal instead of exactly normal, 2sCOPE can have a greater probability of achieving good finite-sample performance than 15.1% when W does not meet the requirements of sufficient nonnormality and relevance.

performance.

We have provided sufficient conditions of endogenous and exogenous regressors above for 2sCOPE to have good finite-sample performance. These are not necessary conditions but are conservative ones to be on the safe side. In particular, to obtain sufficient conditions, we consider the extreme cases in which either the exogenous regressor in step 2 or the endogenous regressor in step 3 follows the normal distribution. However, in practice, regressors are likely to have close-to-normal rather than exact normal distributions. The failure of the sufficient condition tests of W in practice does not mean 2sCOPE cannot be used. For instance, the estimation result of scenario 1 in Table W12 (P and W are close-to-normal and weakly nonnormal, respectively) demonstrates that 2sCOPE may still have acceptable finite-sample performance when the above (conservative) sufficient conditions are not satisfied. In this situation, one can rely on our proposed bootstrap resampling Algorithm 1 to evaluate the finite-sample performance of 2sCOPE on a case-by-case basis.

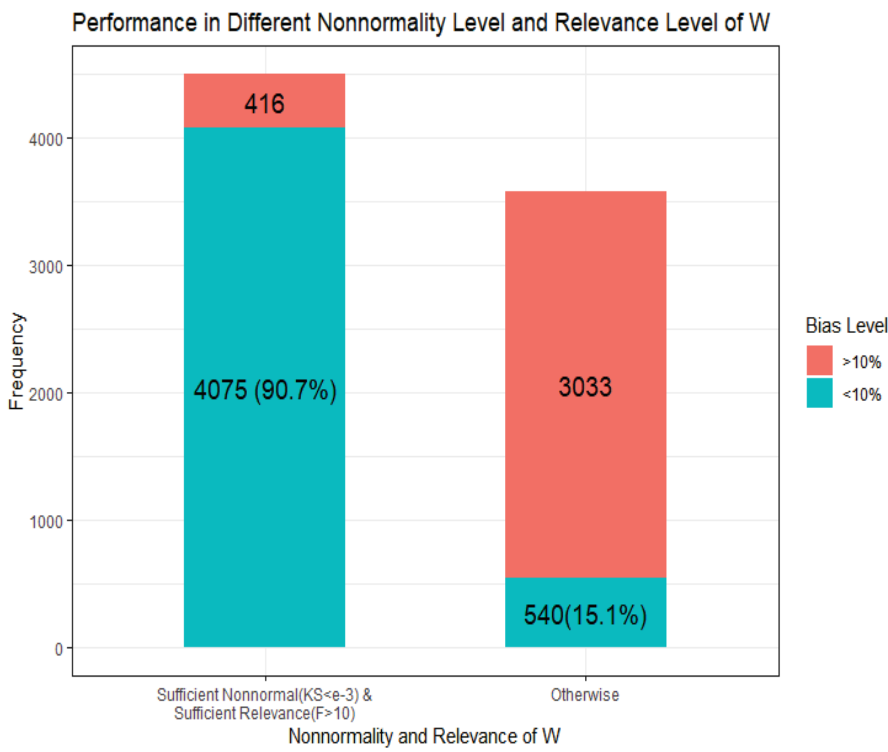


Figure W2: Sufficient Condition of W for Using 2sCOPE in Step 3 of the Decision Tree.

We further test the decision tree using Dominicks store-level toothpaste category sales data from all stores in Chicago (Table W11 under column "Total") and examine frequencies of satisfying various branches of the decision tree when using the real-life data to estimate the demand model in Equation 24 of our empirical application. We first check Step 1, Assumption 5. The result shows that all the 86 stores[23] reject Assumption 5 (Table W11 under column "Step 1"), which means that there exist exogenous regressors (Bonus and PriceRedu) that are correlated with the endogenous regressor (Price), and thus Copula$_{\text{origin}}$ should not be used for model estimation. Then we further check how many stores satisfy the conditions of Step 2 and Step 3 for using 2sCOPE. The result shows that among the 86 stores, 73 stores (84.9%) satisfy the nonnormality condition of the endogenous regressor in

[23] A total of 93 stores have data on toothpaste category sales but 7 stores had only about 30 weeks of sales data and thus were excluded from the analysis.

Step 2. Among the remaining $86 - 73 = 13$ stores with close-to-normal endogenous regressor, 11 stores (11/86=12.8%) have nonnormally distributed and sufficiently relevant exogenous regressors that satisfy conditions in Step 3 to help the model identification with the close-to-normal endogenous regressor. In total, 84 out of 86 stores (97.7%) meet the rule-of-thumb conditions established in the decision tree, demonstrating a high frequency of 2sCOPE usage for correcting endogeneity. The second part of Table W11 shows the average estimates of price coefficient using both OLS and 2sCOPE among the 84 stores that satisfy the conditions of using 2sCOPE. The result shows that the average estimate of the OLS price coefficient is -0.85 (SD=0.22 over the estimates from the 84 stores), which is significantly smaller in size than -2.0, the price elasticity of the toothpaste category reported in the literature (Hoch et al. 1995; Mackiewicz and Falkowski 2015; Chen and Lim 2022). This demonstrates strong price endogeneity that causes attenuation bias of OLS estimates. By contrast, 2sCOPE corrects the endogeneity bias in the right direction, and yields an average price elasticity estimate of -2.47 (Table W11). Moreover, in 71.4% of stores the 95% confidence intervals from 2sCOPE cover -2.0, the price coefficient of toothpaste in the literature. In comparison, in only 13.1% of stores the 95% confidence intervals from OLS cover -2.0 (Table W11).

Table W11: Decision Tree Check Using Real-world Data

	Decision Tree			
	Total	Step 1 (Accept Assumption 5)	Step 2 (P) (nonnormal)	Step 3 (W) (nonnormal and relevant)
Number of Stores	86	0	73	11
Percentage of Stores		0% (0/86)	84.9% (73/86)	12.8% (11/86)
Percentage using Copula$_{\text{origin}}$		0%		
Percentage using 2sCOPE			97.7%	
	Estimates of Price Coefficient			
	OLS Est.	95% CI covers -2.0	2sCOPE Est.	95% CI covers -2.0
	-0.85 (0.22)	11/84 (13.1%)	-2.47 (0.59)	60/84 (71.4%)

Web Appendix E.9: Multiple 'Weakly-nonnormal' Exogenous Covariates vs. One 'Strongly-nonnormal' Exogenous Covariate

According to the decision tree in Figure 2, 2sCOPE with one relevant exogenous regressor having sufficiently strong nonnormality can achieve good finite-sample performance when the endogenous regressor is normally distributed. In this section, we conduct further simulation studies to examine whether several 'weakly-nonnormal' exogenous covariates can add up to achieve the same performance as one 'strongly-nonnormal' exogenous covariate when the endogenous regressor is close to normal. We set the sample size $T = 1000$, and use the same data-generating process as in Equations (20 - 23) except for the dimension of W and the distributions of W and P. Specifically, we set the distribution of P to a close-to-normal distribution, t(30), and use three different scenarios of W to examine the capability of W to help identify the causal effect of the endogenous regressor. In scenario 1, we have one W following the $t(4)$ distribution, with the average p-value of the KS test over 1000 simulated datasets being 0.0054$>$ 0.001 and thus is a 'weakly-nonnormal' exogenous covariate defined in Figure 2. In scenario 2, we increase the number of 'weakly-nonnormal' Ws from 1 to 3, with a 0.2 correlation between different Ws. In scenario 3, we use one 'strongly-nonnormal' W following the $t(2)$ distribution, with the average p-value of KS test over 1000 datasets $1.88e^{-11} < e^{-10}$.

Table W12 shows the estimation results of the three scenarios. In both scenarios 1 (one weakly-nonnormal W) and 2 (three weakly-nonnormal Ws), the estimates of α of 2sCOPE have similar minor but noticeable finite-sample bias. Adding multiple 'weakly-nonnormal' exogenous regressors does not improve the estimation (the estimate of 1.137 for α in scenario

2 is farther away from the true value than 1.101 in scenario 1). However, the 2sCOPE's performance becomes better when using a 'strongly-nonnormal' W in scenario 3 (the estimate of 1.018 for α is closer to the true value than 1.101 in scenario 1 and 1.137 in scenario 2, Table W12). The D-error for 2sCOPE is also smallest when using a 'strongly-nonnormal' W in scenario 3 (Table W12). Thus, a 'strongly-nonnormal' W is better than multiple 'weakly-nonnormal' Ws in helping the identification for close-to-normal endogenous regressors. Adding multiple 'weakly-nonnormal' exogenous covariates will not help the identification of a normal (close-to-normal) endogenous regressor as effectively as one 'strongly-nonnormal' exogenous regressor. Moreover, the estimation results further confirm that our proposed 2sCOPE can largely improve the performance, compared with COPE (the extended $\text{Copula}_{\text{Origin}}$). For instance, in scenario 3, 2sCOPE has large improvement over COPE in both estimation consistency (the estimate of α is improved from 1.494 in COPE to 1.018 in 2sCOPE) and efficiency (the D-error is improved from 0.004830 in COPE to 0.001031 in 2sCOPE, increased by 78.7%), when both the endogenous and exogenous regressors are close-to-normally distributed.

Table W12: Multiple 'Weakly-nonnormal' Exogenous W vs. One 'Strongly-nonnormal' W

			OLS			COPE			2sCOPE		
Scenario	Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
1 W, $t(4)$	μ	1	1.001	0.025	0.026	1.001	0.034	0.030	1.000	0.030	0.016
Weakly-nonnormal	α	1	1.634	0.030	21.25	1.516	0.556	0.928	1.101	0.209	0.486
	β	-1	-1.227	0.026	8.684	-1.010	0.086	0.119	-1.038	0.078	0.482
	ρ	0.5	-	-	-	-0.018	0.454	1.142	0.425	0.132	0.569
	σ_ξ	1	0.822	0.019	9.438	1.016	0.225	0.072	0.965	0.095	0.368
	D-error						0.013740			0.002724	
3 Ws, $t(4)$	μ	1	1.000	0.021	0.016	1.002	0.031	0.080	0.999	0.029	0.019
Weakly-nonnormal	α	1	1.994	0.034	29.40	1.814	0.467	1.742	1.137	0.146	0.941
	β_1	-1	-1.258	0.027	9.677	-1.018	0.071	0.252	-1.036	0.043	0.848
	β_2	-1	-1.259	0.025	10.38	-1.022	0.073	0.297	-1.037	0.042	0.892
	β_3	-1	-1.258	0.024	10.89	-1.020	0.073	0.279	-1.035	0.042	0.851
	ρ	0.5	-	-	-	-0.258	0.394	8.345	0.459	0.046	0.903
	σ_ξ	1	0.697	0.017	17.93	1.005	0.206	1.923	0.933	0.072	0.928
	D-error						0.007816			0.001095	
1 W, $t(2)$	μ	1	1.001	0.026	0.040	1.002	0.033	0.054	1.001	0.031	0.037
Strongly-nonnormal	α	1	1.574	0.040	14.36	1.494	0.577	0.855	1.018	0.094	0.188
	β	-1	-1.086	0.033	2.617	-1.002	0.018	0.110	-1.003	0.018	0.194
	ρ	0.5	-	-	-	-0.005	0.467	1.081	0.486	0.057	0.249
	σ_ξ	1	0.839	0.019	8.273	1.025	0.229	0.108	0.993	0.052	0.133
	D-error						0.004830			0.001031	

Web Appendix E.10: Random Coefficient Logit Model with Regressor Endogeneity

In this section, we examine the performance of the proposed 2sCOPE in random coefficient logit (RCL) models using simulated individual-level data. The specific DGP is as follows:

$$u_{h0t} = \epsilon_{h0t} \tag{W46}$$

$$u_{h1t} = \bar{\beta}_1 + \bar{\beta}_3 \cdot W_{1t} + (\bar{\alpha} + a_h) \cdot P_{1t} + \xi_{1t} + \epsilon_{h1t}, \tag{W47}$$

$$u_{h2t} = \bar{\beta}_2 + \bar{\beta}_3 \cdot W_{2t} + (\bar{\alpha} + a_h) \cdot P_{2t} + \xi_{2t} + \epsilon_{h2t}, \tag{W48}$$

$$\alpha_h \sim N(0, \sigma_a^2) = N(0, 0.5^2), \tag{W49}$$

$$\begin{pmatrix} P_{jt}^* \\ W_{jt}^* \\ \xi_{jt}^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix}\right), \tag{W50}$$

$$\xi_t = G^{-1}(U_{\xi,t}) = G^{-1}(\Phi(\xi_t^*)) = \Phi^{-1}_{(0,0.2^2)}(\Phi(\xi^*)) = 0.2 \cdot \xi_t^*, \tag{W51}$$

$$P_{jt} = H^{-1}(\Phi(P_{jt}^*)) = 1.5 + 0.2 \cdot ln(-ln(1 - \Phi(P_{jt}^*))) \text{ for } j = 1, 2, \tag{W52}$$

$$W_{jt} = L^{-1}(\Phi(W_{jt}^*)) = I(\Phi(W_{jt}^*) > 0.7) \text{ for } j = 1, 2 \tag{W53}$$

where the purchase occasion t = 1,...,200, the brand $j = 1, 2$, and the consumer h = 1,...,100. The independent error term ϵ_{hjt} follows Type one extreme value distribution. P_{jt} is the endogenous regressor following an extreme value distribution, and W_{jt} is a binary exogenous regressor that is correlated with P_{jt}. Consumers have heterogeneous preferences with respect to P_{jt}. We generate individual-level choices as described above and apply the 2sCOPE method described in Web Appendix D to the simulated data to estimate the model parameters. We also use a benchmark method called OLS, which ignores the endogeneity by applying OLS to the first-step estimates $\widehat{\delta}_{jt}$ in Equation W16. Estimation results over 1000 simulated datasets are presented in Table W13. Results show that the OLS estimates are biased, while the estimates from 2sCOPE are tightly distributed around the true values.

Table W13: Results of the Simulation Study: RCL Model Using Individual-level Data.

		OLS			2sCOPE		
Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}
$\bar{\beta}_1$	0.7	0.114	0.098	6.009	0.683	0.306	0.056
$\bar{\beta}_2$	0.7	0.121	0.099	5.854	0.693	0.310	0.024
$\bar{\beta}_3$	1	0.928	0.037	1.974	1.012	0.058	0.205
$\bar{\alpha}$	-1	-0.563	0.090	4.880	-0.994	0.237	0.028
σ_a	0.5	0.515	0.046	0.322	0.515	0.046	0.322

Note: σ_a is estimated in the first step, and thus OLS and 2sCOPE methods have the same estimates.

Web Appendix E.11: Endogenous W

For the proposed 2sCOPE method to work properly, we require the regressors in W be exogenous, which is also a requirement for the OLS, copula$_{Origin}$, and IV approaches (Wooldridge 2010). In this section, we further discuss the case when W is endogenous, and examine the performance of 2sCOPE and alternative approaches when W is endogenous. The specific DGP is as follows.

$$\begin{pmatrix} P_t^* \\ W_t^* \\ Z_t^* \\ \xi_t^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & p & q & \rho_{p\xi} \\ p & 1 & q_2 & \rho_{w\xi} \\ q & q_2 & 1 & 0 \\ \rho_{p\xi} & \rho_{w\xi} & 0 & 1 \end{bmatrix} \right) = N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.5 & 0.5 & 0.5 \\ -0.5 & 1 & -0.3 & \rho_{w\xi} \\ 0.5 & -0.3 & 1 & 0 \\ 0.5 & \rho_{w\xi} & 0 & 1 \end{bmatrix} \right),$$

$$\xi_t = \Phi^{-1}(\Phi(\xi^*)) = 1 \cdot \xi_t^*, \quad W_t = L^{-1}(\Phi(W_t^*)),$$

$$P_t = H^{-1}(\Phi(P_t^*)), \quad Z_t = L^{-1}(\Phi(Z_t^*)),$$

$$Y_t = \mu + \alpha \cdot P_t + \beta \cdot W_t + \xi_t = 1 + P_t + (-1) \cdot W_t + \xi_t, \tag{W54}$$

where $P_t \sim Gamma(1,1)$, $W_t \sim Exp(1)$ and $Z_t \sim Exp(1)$. Z is a valid instrument variable of the endogenous regressor P because it satisfies the two conditions in page 89 of Wooldridge (2010): (1) Z is uncorrelated with the error term ξ and is excluded from the equation W54 for Y (i.e, exclusion restriction), and (2) Z is correlated with P (i.e., relevant). We set the sample size $N = 1000$, and estimate the model using OLS, Copula$_{\text{Origin}}$, 2sCOPE and TSLS methods.

Table W14 shows the estimation results over 1000 simulated datasets for four endogeneity levels of W, $\rho_{w\xi}$ ={0, 0.1, 0.2, 0.3}. As expected, OLS estimates are biased in all situations.

Copula$_{\text{Origin}}$ substantially reduces the bias of the OLS estimates but still has notable bias because of ignoring the correlation between P and W. As expected, when $\rho_{w\xi} = 0$ (i.e., W is exogenous), both 2sCOPE and TSLS yield consistent estimates that properly correct for endogeneity bias. However, when W becomes endogenous ($\rho_{w\xi}$ ={0.1, 0.2, 0.3}), all methods suffer from potential biases. For the model in Eqn (1), OLS estimate of $\hat{\alpha} = (P'P)^{-1}P'Y - (P'P)^{-1}P'[1, W][\hat{\mu}, \hat{\beta}']'$. Thus, when $P'W \neq 0$ (i.e., P and W are correlated), $\hat{\alpha}$ depends on $\hat{\beta}$, and the inconsistency of $\hat{\beta}$ will make $\hat{\alpha}$ biased even if P is exogenous. Similarly, the endogeneity of W can cause bias in $\hat{\beta}$, which, when P and W are correlated, can lead to bias in $\hat{\alpha}$ for both Copula$_{\text{Origin}}$ and 2sCOPE. In TSLS, only exogenous regressors and IVs can enter the first-stage regression in TSLS and so including endogenous W in the first-stage regression can lead to biased estimates for TSLS (Wooldridge 2010). Thus, all these methods require the exogeneity of W to yield consistent estimates when W and P are correlated. Moreover, according to the estimation results, the bias of 2sCOPE is relatively smaller than OLS, Copula$_{\text{Origin}}$ and IV approach.

Table W14: Simulation Results When P and W are correlated

			OLS			Copula$_{\text{Origin}}$			2sCOPE			TSLS		
Simulation	Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
$\rho_{w\xi} = 0$	μ	1	0.287	0.061	11.68	0.780	0.077	2.846	1.002	0.083	0.024	0.998	0.122	0.012
	α	1	1.523	0.037	14.30	0.930	0.068	1.027	0.997	0.055	0.054	0.999	0.083	0.009
	β	-1	-0.808	0.033	5.746	-0.712	0.033	8.790	-1.000	0.036	0.014	-0.998	0.044	0.040
	Avg Abs Bias			0.476			0.193			0.0018			0.0013	
	D-error			0.00096			0.00175			0.00140			0.00189	
$\rho_{w\xi} = 0.1$	μ	1	0.146	0.062	13.74	0.696	0.083	3.672	0.984	0.089	0.181	0.837	0.120	1.354
	α	1	1.560	0.036	15.56	0.899	0.071	1.423	0.945	0.057	0.965	1.053	0.083	0.639
	β	-1	-0.704	0.033	8.970	-0.597	0.031	13.00	-0.929	0.037	1.919	-0.889	0.043	2.349
	Avg Abs Bias			0.570			0.269			0.047			0.109	
	D-error			0.00092			0.00182			0.00150			0.00186	
$\rho_{w\xi} = 0.2$	μ	1	0.004	0.062	16.07	0.603	0.084	4.701	0.955	0.091	0.498	0.672	0.114	2.869
	α	1	1.597	0.038	15.73	0.878	0.071	1.723	0.903	0.057	1.701	1.108	0.077	1.405
	β	-1	-0.599	0.036	11.27	-0.482	0.032	16.22	-0.855	0.037	3.896	-0.779	0.043	5.201
	Avg Abs Bias			0.665			0.345			0.096			0.219	
	D-error			0.00101			0.00185			0.00156			0.00176	
$\rho_{w\xi} = 0.3$	μ	1	-0.143	0.060	18.98	0.513	0.079	6.149	0.927	0.086	0.849	0.502	0.110	4.514
	α	1	1.638	0.037	17.04	0.851	0.070	2.126	0.853	0.059	2.488	1.165	0.075	2.213
	β	-1	-0.495	0.035	14.42	-0.367	0.032	19.61	-0.781	0.034	6.399	-0.667	0.042	8.003
	Avg Abs Bias			0.762			0.423			0.146			0.332	
	D-error			0.00092			0.00185			0.00152			0.00155	

Table W15 further shows the estimation results using the same DGP as above except that W is uncorrelated with P. OLS estimates continue to have a large bias. One interesting finding is that the endogeneity of W will not cause bias in the estimate of the endogenous regressor P using 2sCOPE and Copula$_{\text{Origin}}$ when P and W are uncorrelated. By contrast, since both the endogenous W and the IV enter the first stage of TSLS, bias will arise in the estimate of α using TSLS (Wooldridge 2010). Thus, the copula methods (Copula$_{\text{Origin}}$ and 2sCOPE) are more robust than the IV method to endogenous W that is unrelated to P. This means that when using copula endogeneity correction methods, one does not need to argue for the exogeneity of the control variables that are prognostic factors of only the outcome variable; the purpose of including such prognostic variables in the model is to improve the accuracy of model estimates and predictions rather than to adjust for confounding.

These simulation studies demonstrate the importance of the exogeneity assumption for the control variables W and potential bias for different estimation methods when the exogeneity assumption of W is violated. Thus, we require the exogeneity of W for 2sCOPE to work properly when W and P are correlated.

Table W15: Simulation Results When P and W are uncorrelated

			OLS			$\text{Copula}_{\text{Origin}}$			2sCOPE			TSLS		
Simulation	Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
$\rho_{w\xi} = 0$	μ	1	0.548	0.048	9.322	0.999	0.077	0.020	0.999	0.077	0.009	1.001	0.080	0.008
	α	1	1.453	0.031	14.71	1.000	0.066	0.005	1.001	0.066	0.011	1.001	0.067	0.017
	β	-1	-1.000	0.028	0.008	-1.000	0.028	0.008	-1.000	0.032	0.017	-0.999	0.032	0.028
	Avg Abs Bias			0.3017			0.0007			0.0006			0.0009	
	D-error			0.00082			0.00150			0.00164			0.00169	
$\rho_{w\xi} = 0.1$	μ	1	0.460	0.050	10.83	0.910	0.080	1.124	0.911	0.082	1.089	0.867	0.078	1.715
	α	1	1.451	0.031	14.38	1.000	0.068	0.002	1.000	0.068	0.007	1.044	0.066	0.666
	β	-1	-0.909	0.028	3.221	-0.909	0.028	3.280	-0.909	0.031	2.935	-0.908	0.031	2.956
	Avg Abs Bias			0.361			0.060			0.060			0.089	
	D-error			0.00085			0.00152			0.00164			0.00158	
$\rho_{w\xi} = 0.2$	μ	1	0.366	0.050	12.70	0.817	0.077	2.370	0.818	0.080	2.268	0.729	0.077	3.523
	α	1	1.453	0.032	13.96	0.999	0.064	0.013	1.000	0.064	0.005	1.089	0.064	1.385
	β	-1	-0.820	0.027	6.630	-0.820	0.027	6.803	-0.820	0.030	5.930	-0.819	0.030	6.084
	Avg Abs Bias			0.423			0.121			0.121			0.180	
	D-error			0.00084			0.00141			0.00155			0.00148	
$\rho_{w\xi} = 0.3$	μ	1	0.272	0.048	15.06	0.726	0.076	3.630	0.727	0.080	3.421	0.592	0.075	5.415
	α	1	1.454	0.030	15.25	0.999	0.065	0.022	0.999	0.065	0.019	1.134	0.061	2.176
	β	-1	-0.728	0.029	9.344	-0.728	0.028	9.651	-0.728	0.033	8.301	-0.728	0.032	8.617
	Avg Abs Bias			0.485			0.183			0.182			0.271	
	D-error			0.00080			0.00146			0.00162			0.00145	

Web Appendix E.12: Leveraging Empirical Correlation between W and P

As noted in the main paper, 2sCOPE can leverage the exogenous regressors W pre-existing in the OLS, IV or Copula$_{\text{Origin}}$ estimation of the outcome model to improve model estimation. The correlation between W and the endogenous regressor P does not need to have a causal interpretation: association is sufficient. Thus, using 2sCOPE, one does not need to argue for a causal relationship between P and W. To demonstrate this point, we conduct the following simulation study. Specifically, the simulation study considers cases of correlations resulting from either a causal relationship or mere empirical spurious correlation.

First, we consider the case of a spurious correlation between W and P. A spurious correlation is an association in which W and P are associated but not causally related. A common reason for spurious correlations is the presence of a third, unobserved common factor (named as O here) that affects W and P simultaneously. Although W and P are empirically correlated, they have no direct causal relationship: given the common factor O, P and W are unrelated. The specific DGP is as follows.

$$\begin{pmatrix} e_{p,t} \\ \xi_t \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho \cdot \sigma_1 \cdot \sigma_2 \\ \rho \cdot \sigma_1 \cdot \sigma_2 & \sigma_2^2 \end{bmatrix}\right) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.5 & 0.5 \cdot \sqrt{0.5} \\ 0.5 \cdot \sqrt{0.5} & 1 \end{bmatrix}\right),$$

$$O_t \sim N(0, \sigma_o^2) = N(0,1), \quad e_{w,t} \sim N(0, \sigma_w^2) = N(0, 0.5),$$

$$P_t^* = \sqrt{0.5} \cdot O_t + e_{p,t}, \quad W_t^* = \sqrt{0.5} \cdot O_t + e_{w,t},$$

$$P_t = H^{-1}(\Phi(P_t^*)), \quad W_t = L^{-1}(\Phi(W_t^*)),$$

$$Y_t = \mu + \alpha \cdot P_t + \beta \cdot W_t + \xi_t = 1 + P_t + (-1) \cdot W_t + \xi_t,$$

where $P_t \sim Gamma(1,1)$ and $W_t \sim Exp(1)$. We set the sample size $N = 1000$, and estimate the model using OLS, $\text{Copula}_{\text{Origin}}$ and 2sCOPE methods. Table W16 shows the estimation results over 1000 simulated datasets. We can see that the OLS estimates are severely biased; estimates from $\text{Copula}_{\text{Origin}}$ improve upon the OLS estimates but still have a notable bias because of the spurious correlation between P and W, while the 2sCOPE method yields consistent estimates even if the association between P and W are spurious.

		OLS			$\text{Copula}_{\text{Origin}}$			2sCOPE		
Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
μ	1	0.781	0.047	4.654	1.155	0.082	1.885	1.003	0.060	0.045
α	1	1.404	0.037	10.94	1.065	0.072	0.682	1.000	0.070	0.006
β	-1	-1.184	0.034	5.368	-1.205	0.034	6.039	-1.001	0.044	0.031
σ_ξ	1	0.933	0.021	3.217	1.002	0.034	0.066	1.000	0.031	0.005

Table W16: Results of the Simulation Study: Spurious Correlation between P and W.

Next, we examine the case of a causal relationship between W and P. In this case, W directly affects P. The DGP is as follows.

$$\begin{pmatrix} e_{p,t} \\ \xi_t \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho \cdot \sigma_1 \cdot \sigma_2 \\ \rho \cdot \sigma_1 \cdot \sigma_2 & \sigma_2^2 \end{bmatrix}\right) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.5 & 0.5 \cdot \sqrt{0.5} \\ 0.5 \cdot \sqrt{0.5} & 1 \end{bmatrix}\right),$$

$$W_t^* \sim N(0,1), P_t^* = \sqrt{0.5} \cdot W_t^* + e_{p,t},$$

$$P_t = H^{-1}(\Phi(P_t^*)), \quad W_t = L^{-1}(\Phi(W_t^*)),$$

$$Y_t = \mu + \alpha \cdot P_t + \beta \cdot W_t + \xi_t = 1 + P_t + (-1) \cdot W_t + \xi_t,$$

where $P_t \sim Gamma(1,1)$ and $W_t \sim Exp(1)$. As shown above, W directly affects both the endogenous regressor P and the outcome variable. We set the sample size $N = 1000$, and estimate the model using OLS, $\text{Copula}_{\text{Origin}}$ and 2sCOPE methods. Table W17 shows the estimation results over 1000 simulated datasets. OLS and $\text{Copula}_{\text{Origin}}$ estimates are biased, while 2sCOPE yields consistent estimates when P and W are causally related.

Table W17: Results of the Simulation Study: P and W are causally related.

		OLS			$\text{Copula}_{\text{Origin}}$			2sCOPE		
Parameters	True	Mean	SE	t_{bias}	Mean	SE	t_{bias}	Mean	SE	t_{bias}
μ	1	0.809	0.043	4.415	1.213	0.082	2.582	1.001	0.052	0.019
α	1	1.580	0.044	13.04	1.201	0.076	2.653	1.000	0.079	0.005
β	-1	-1.388	0.041	9.449	-1.415	0.041	10.15	-1.001	0.062	0.016
σ_ξ	1	0.903	0.021	4.631	0.986	0.038	0.356	1.000	0.034	0.013

The simulation studies in this section demonstrate that 2sCOPE can leverage the association between P and W for model estimation, regardless of whether the association is causal or not.

WEB APPENDIX F: OBTAINING STANDARD ERRORS USING BOOTSTRAP

We generate B bootstrap datasets by randomly resampling the original dataset with replacement, and re-estimate the structural model parameters using 2sCOPE for each dataset. Then calculate the standard errors by calculating the standard deviation of the estimates obtained from these datasets. Algorithm 2 summarizes the detailed steps of how to obtain the standard errors of estimates using bootstrap.

Algorithm 2 Bootstrap Algorithm for Calculating Standard Error of 2sCOPE Estimates

Series Input: data Y, P, W, sample size N, and number of bootstrap B.

for $b = 1$ to B **do**

Randomly resample (Y_b, P_b, W_b) from the original data (Y, P, W) with replacement, sample size $= N$;

Obtain $P_b^* = \Phi^{-1}(\widehat{H}(P_b)), W_b^* = \Phi^{-1}(\widehat{L}(W_b))$, where $\hat{H}(\cdot)$ and $\hat{L}(\cdot)$ are estimated CDFs of P_b and W_b;

Obtain the 2sCOPE estimate $\widehat{\theta}_b = \widehat{\theta}(Y_b, P_b, W_b, P_b^*, W_b^*)$ using the bth bootstrap sample.

end for

Calculate standard error of the estimator: $\sqrt{\frac{\sum_{b=1}^{B}(\widehat{\theta}_b - \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}_b)^2}{B-1}}$.

WEB APPENDIX G: IMPLEMENTING THE BOOTSTRAP METHOD TO EVALUATE FINITE-SAMPLE BIAS IN EMPIRICAL APPLICATION

To gauge and validate the finite-sample performance of 2sCOPE, we apply the bootstrap algorithm described in Algorithm 1 to our empirical application and conduct a bootstrap re-sampling study by drawing repeated samples of the same size as the observed data from the underlying copula model and the structural model estimated from the original sample using data from store 1 in the application, and perform estimation on each bootstrap sample. Specifically, we generate data using the following DGP:

$$\begin{pmatrix} \text{Price}^* \\ \text{Bonus}^* \\ \text{PriceRedu}^* \\ \xi_t^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & -0.5 & -0.3 & 0.3 \\ -0.5 & 1 & -0.3 & 0 \\ -0.3 & -0.3 & 1 & 0 \\ 0.3 & 0 & 0 & 1 \end{bmatrix} \right), \tag{W55}$$

$$\xi_t = G^{-1}(\Phi(\xi_t^*)) = \Phi_\sigma^{-1}(\Phi(\xi^*)) = 0.4 \cdot \xi_t^*, \tag{W56}$$

$$\text{Price} = \hat{H}^{-1}(\Phi(\text{Price}^*)), \quad \text{Bonus} = \hat{L}_1^{-1}(\Phi(\text{Bonus}^*)), \tag{W57}$$

$$\text{PriceRedu} = \hat{L}_2^{-1}(\Phi(\text{PriceRedu}^*)), \tag{W58}$$

$$Y_t = -4 + (-2) \cdot \text{Price} + 0.1 \cdot \text{Bonus} + 0.3 \cdot \text{PriceRedu} + \xi_t, \tag{W59}$$

where $\hat{H}(\cdot), \hat{L}_1(\cdot), \hat{L}_2(\cdot)$ are all estimated CDFs using the univariate empirical distribution in the application for regressors Price, Bonus and PriceRedu, respectively. The correlation matrix of the copula transformation of variables (i.e., Price*, Bonus*, PriceRedu*, ξ^*) in Equation (W55) and the standard deviation of the error term (i.e., σ_ξ) are set according to the estimated parameter values using real data. After generating the regressors and the structural error, we set the coefficients using the 2sCOPE estimates of original data to generate Y in Equation (W59). We set the sample size $T = 373$, which is the same as the sample size in the application, and generate $B = 1000$ bootstrap datasets in each of which we estimate the structural model parameters using OLS and 2sCOPE.

WEB APPENDIX H: REFERENCES

Ahad, Nor Aishah, Teh Sin Yin, Abdul Rahman Othman, and Che Rohani Yaacob (2011), "Sensitivity of normality tests to non-normal data," *Sains Malaysiana,* **40 (6)**, 637–641.

Cortina, Jose M and William P Dunlap (1997), "On the logic and purpose of significance testing," *Psychological Methods,* **2 (2)**, 161.

Eckert, Christine and Jan Hohberger (2022), "Addressing Endogeneity Without Instrumental Variables: An Evaluation of the Gaussian Copula Approach for Management Research," *Journal of Management,* DOI: 10.1177/01492063221085913

Enders, Craig K, Brian T Keller, and Roy Levy (2018), "A fully conditional specification approach to multilevel imputation of categorical and continuous variable," *Psychological methods,* **23 (2)**, 298.

Forero, Carlos G, Alberto Maydeu-Olivares, and David Gallardo-Pujol (2019) "Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation," *Structural equation modeling,* **16 (4)**, 625–641.

Holtmann, Jana, Tobias Koch, Katharina Lochner, and Michael Eid (2016) "A comparison of ML, WLSMV, and Bayesian methods for multilevel structural equation models in small samples: A simulation study," *Multivariate behavioral research,* **51 (5)**, 661–680.

McNeish, Daniel (2016) "Estimation methods for mixed logistic models with few clusters," *Multivariate Behavioral Research,* **51(6)**, 790–804.

Wang, Yan and Eun Sook Kim (2017) "Evaluating model fit and structural coefficient bias: A Bayesian approach to multilevel bifactor model misspecification," *Structural Equation Modeling,* **24 (5)**, 699–713.