NBER WORKING PAPER SERIES

A PRACTICAL GUIDE TO ENDOGENEITY CORRECTION USING COPULAS

Yi Qian
Anthony Koschmann
Hui Xie

Working Paper 32231
http://www.nber.org/papers/w32231

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2024

We acknowledge the support by Social Sciences and Humanities Research Council of Canada [grants 435-2018-0519 and 435-2023-0306], Natural Sciences and Engineering Research Council of Canada [grant RGPIN-2018-04313 and 2023-04348] and US National Institute of Health [grant 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. Procedures used were in compliance with British Columbia's Freedom in Information and Privacy Protection Act. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

A Practical Guide to Endogeneity Correction Using Copulas
Yi Qian, Anthony Koschmann, and Hui Xie
NBER Working Paper No. 32231
March 2024
JEL No. C01,C10,C5

ABSTRACT

Causal inference is of central interests in many empirical applications yet often challenging because of the presence of endogenous regressors. The classical approach to the problem requires using instrumental variables that must satisfy the stringent condition of exclusion restriction. At the forefront of recent research, instrument-free copula methods have been increasingly used to handle endogenous regressors. This article aims to provide a practical guide for how to handle endogeneity using copulas. The authors give an overview of copula endogeneity correction and its usage in marketing research, discuss recent advances that broaden the understanding, applicability, and robustness of copula correction, and examine implementation challenges of copula correction such as construction of copula control functions and handling of higher-order terms of endogenous regressors. To facilitate the appropriate usage of copula correction, the authors detail a process of checking data requirements and identification assumptions to determine when and how to use copula correction methods, and illustrate its usage using empirical examples.

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

Anthony Koschmann
Eastern Michigan University
121 Hill Hall
College of Business
Ypsilanti, MI 48178
akoschma@emich.edu

Many research questions in marketing, economics, and health sciences are interested in matters of causality rather than simply questions of association. Frequently, these questions are tackled by using relevant data to estimate structural regression models representing causal relationships. A pervasive issue in these empirical investigations is the presence of endogenous regressors, which can arise when the regressors representing the causes (e.g., an economic program to be evaluated, marketing mix variables, etc.) are not randomly assigned in the data; the regressors thus correlate with unobservables (e.g., unobserved product characteristics or common market shocks) in the structural error term (Villas-Boas and Winer 1999). Estimation methods that ignore the presence of regressor-error dependence, such as the ordinary least squares (OLS) method, can lead to severe bias in the estimates of structural model parameters (i.e., endogeneity bias).

Given the ubiquity of endogenous regressors and the importance of addressing endogeneity bias, a large body of literature is devoted to developing appropriate methods to solve or mitigate the endogeneity issue. The instrumental variable (IV) method is the classical econometric approach to correct for endogeneity bias (Wooldridge 2010). This method relies on the existence of valid and strong IVs to satisfy the stringent requirement of exclusion restriction, which makes IVs difficult to find and justify in practice (Ebbes et al. 2005; Ebbes, Wedel, and Böckenholt 2009; Park and Gupta 2012). When there exists theory or knowledge about the underlying mechanism of endogeneity, an alternative approach is to model the exact process of yielding the observed values of the endogenous regressors, which is then estimated jointly with the structural model of primary interest. For instance, in estimating a consumer demand model, a supply-side model reflecting researchers' beliefs about the managerial decisions determining the supply-side marketing mix variables (such as price and promotions) can be specified and jointly estimated with the demand model (e.g., Sudhir 2001; Yang, Chen, and Allenby 2003; Manchanda, Rossi, and Chintagunta 2004; Luan and Sudhir 2010). When the supply-side model is specified correctly, this approach can successfully correct for endogeneity bias in parameter estimates of the demand model.

Recently, there have been growing interests in developing endogeneity correction methods that require neither observed IVs nor knowledge to correctly specify an auxiliary supply model. These instrument-free methods exploit higher moments (Lewbel 1997), heteroscedastic error structures (Rigobon 2003), latent IVs (Ebbes et al. 2005), and copulas[1] (Park and Gupta 2012; Becker, Proksch, and Ringle 2021; Christopoulos, McAdam, and Tzavalis 2021; Tran and Tsionas 2021; Eckert and Hohberger 2022; Haschka 2022; Yang, Qian, and Xie 2022) to control for endogeneity bias. Ebbes, Wedel, and Böckenholt (2009), Park and Gupta (2012), Papies, Ebbes, and Heerde (2017), and Rutz and Watson (2019) provide detailed comparisons of these IV-free methods with alternative methods.

Copula correction methods provide substantial advantages for addressing the prevalent and thorny issue of endogenous regressors. These methods directly address the regressor-error dependence using copulas, a widely used multivariate dependence model applicable in many practical applications (Joe 2015; Danaher 2007; Danaher and Smith 2011). Unlike the traditional IV approach and other IV-free methods, the copula correction methods do not require the endogenous regressor contain an (observed or latent) exogenous component satisfying the stringent exclusion restriction condition that can be hard to justify in practical applications. Thus, copula correction methods are feasible to use in many situations under appropriate conditions. Although copula correction originally required sufficient nonnormality of endogenous regressors (Park and Gupta 2012), limiting its applicability, the recent two-stage copula correction method by Yang, Qian, and Xie (2022) relaxes this condition as long as one of the correlated exogenous regressors is nonnormally distributed, which is a considerably weaker requirement and feasible in many applications.

Furthermore, one can implement copula correction by including copula control functions derived from existing regressors as additional regressors in the structural regression model to control for endogeneity. Thus, copula correction using the control function is straightforward to apply in a wide array of settings, including both linear and nonlinear models (e.g., discrete

[1] "Copula" was introduced by Sklar (1959) from the Latin "to link", as a function linking two variables. Copulas encompass different forms, but we use 'copulas' here to speak synonymously with Gaussian copulas.

choice models) and the challenging slope endogeneity problem.

Focusing on copula correction methods, the objectives of this article are: (a) to raise awareness of the importance to address endogenous regressors in marketing studies; (b) to provide practical guidance to empirical researchers employing copula endogeneity correction; and (c) to demonstrate use of copula endogeneity correction in practical applications.

With these objectives in mind, this article makes the following contributions. One, we provide a comprehensive overview of how the copula procedures have been used in marketing research to correct for endogeneity. Over the past ten years, the copula approach has been adopted in a range of substantive areas to establish causality. We review the substantive areas for which copula methods are useful, wide variances among empirical researchers on copula use and implementations, and recent advances in copula correction methodology. We synthesize the literature to provide a theoretical and empirical foundation for appropriate use of copula correction methods.

Two, building upon recent advances and our evaluations on variations of copula implementation, we provide an updated guidance on when and how to use copula correction, accessible to academics and practitioners alike. Despite the advantages of copula correction methods and growing popularity, the effectiveness of these methods depends on whether important data requirements are met and whether the analysis is implemented appropriately. Indeed, recent research points to pitfalls resulting from misuse of copula methods. Furthermore, significant methodological advances have been made since the Park and Gupta's 2012 study, such that clear guidelines regarding the use of expanded copula correction toolbox are lacking. In this article, we create a 'cookbook' for how copulas should be applied based on the latest research, in a flowchart with checkpoints and data requirements that characterize the settings where copula correction methods are useful and where they may fail.

Three, we address the lack of clear guidelines regarding two implementation variations that are less studied but have substantial effects on the performance of copula correction. The first issue regards estimation bias in models with an intercept, discovered by Becker,

Proksch, and Ringle (2021); we show that an alternative implementation of copula transformation with theoretical support solves this bias issue and informs better copula implementation. The second issue is the handling of moderated endogenous regressors (i.e., higher-order effects like squared terms or interactions between two endogenous regressors). To the best of our knowledge, no study has been conducted to compare different copula approaches, let alone establishing an optimal approach to addressing endogenous higher-order regressors, where a clear guideline is needed. By making an analogy to the control function using IVs (Papies, Ebbes, and Heerde 2017), interactions between an endogenous and exogenous regressor need no additional copula term: only the copula term for the main effect of the endogenous regressor is needed. However, no theoretical optimality nor magnitude of empirical difference for different copula approaches are found in the existing literature. This may explain the variations in copula handling of higher-order endogenous regressors. Researchers may deem including copula terms for higher-order endogenous regressors as having comparable performance with little harm[2]; researchers may even believe including these copula terms is a good practice to control for endogeneity of higher-order regressors, or at the request of gatekeepers (e.g., journal reviewers). This study establishes not only the sufficiency but also theoretical and empirical optimality of excluding copula higher-order correction terms from endogeneity correction. We highlight large adverse effects - significant finite sample bias and greatly inflated estimation variability - when such higher-order copula terms are included, both in simulations and real-life data.

In the next section, we survey substantive marketing areas where copula correction has been used, as well as important variations in the use and implementation of copula correction. Next, we present relevant methodological background: how the copula handles endogeneity, how the copula is generated, how to generalize copula correction for correlated exogenous regressors or close-to-normal endogenous regressors, and how copulas should be

[2]This may hold in control functions using IVs. Depending on the strength of IVs, the control functions for main and higher-order endogenous terms may cause much less severe multicollinearity issues than the counterpart copula control functions.

used for moderated endogenous regressors. Then we discuss the boundary conditions and data requirements for applying copula correction, presenting a flowchart of checkpoints evaluating these conditions. We provide two empirical examples to walk through this process of applying copula correction. Finally, we close with conclusions and implications for both academics and practitioners.

IMPACTS OF COPULA ENDOGENEITY CORRECTION

Largely due to the aforementioned advantages, copula correction has gained increasing popularity in empirical research since Park and Gupta's 2012 study for addressing endogeneity (Rutz and Watson 2019; Becker, Proksch, and Ringle 2021; Haschka 2022; Eckert and Hohberger 2022). Table 1, and the pie chart in Figure W1 of Web Appendix A, break down by substantive area copula correction publications that appeared in leading marketing journals [3] from 2013 to 2022.

Table 1: Examples of Substantive Areas in Marketing with Applications of Copula Endogeneity Correction.

Study	Product	Price	Place	Prom.	SF[a] & CRM	Other[a]
Schwedel and Knox (2013)					X	
Burmester et al (2015)				X		
Datta, Foubert, and van Heerde (2015)				X		
Glady, Lemmens, and Croux (2015)						X
Mathys, Burmester, and Clement (2016)	X			X		
Datta, Ailawadi, and van Heerde (2017)		X	X	X		
Lenz, Wetzel, and Hammerschmidt (2017)						X
Atefi et al (2018)					X	
Gielens et al (2018)	X			X		
Gijsbrechts, Campo, and Vroegrijk (2018)						X
Guitart, Gonzalez, and Stremersch (2018)		X		X		
Lamey et al (2018)		X		X		
Lim, Tuli, and Dekimpe (2018)		X				

continued ...

[3] This list includes *Journal of Marketing, Journal of Marketing Research, Marketing Science, Journal of Consumer Research, Journal of the Academy of Marketing Science, Journal of Retailing, International Journal of Research in Marketing,* and *Journal of Consumer Psychology.*

Study	Product	Price	Place	Prom	SF[a] & CRM	Other[a]
Ter Braak and Deleersnyder (2018)	X	X				X
Wetzel et al (2018)					X	
Zhao et al (2018)	X					
Carson and Ghosh (2019)					X	
Keller, Deleersnyder, and Gedenk (2019)		X				
Nath et al (2019)						X
Schulz, Shehu, and Clement (2019)						X
Vieira et al (2019)				X		X
Aydinli et al (2020)		X				X
Bombaij and Dekimpe (2020)						X
Bornemann, Hattula, and Hattula (2020)	X					
Campo et al (2020)	X	X				
De Jong, Zacharias, and Nijssen (2020)						X
Garrido-Morgado et al (2020)	X	X				
Guitart and Stremersch (2020)		X		X		X
Guitart, Hervet, and Gelper (2020)				X		
Heitmann et al (2020)						
Homburg, Vomberg, Muehlhaeuser (2020)	X	X		X		X
Liu et al (2020)		X				
Magnotta et al (2020)					X	
Shehu, Papies, and Neslin (2020)		X				
Van Ewijk et al (2020)		X		X		
Vomberg, Homburg, and Gwinner (2020)					X	
Bachmann, Meierer, and Näf (2021)					X	
Cron et al (2021)					X	
Dhaoui and Webster (2021)						X
Fossen and Bleier (2021)						X
Hoskins et al (2021)						X
Kidwell et al (2021)						X
Lamey, Breugelmans, and ter Braak (2021)						X
Sawant, Hada, and Blanchard (2021)						X
Bhattacharaya, Morgan, and Rego (2022)						X
Borah et al (2022)	X			X		X
Cao (2022)	X					X
Cao et al (2022)						X
Danaher (2022)		X				
Datta et al (2022)	X	X	X			
Gielens et al (2022)	X	X				
Janani et al (2022)					X	
Krämer et al (2022)					X	X

continued ...

Study	Product	Price	Place	Prom	SF[a] & CRM	Other[a]
Ludwig et al (2022)					X	
Maesen et al (2022)	X	X				
Moon, Tuli, and Mukherjee (2022)						X
Nahm et al (2022)		X				
Rajavi, Kushwaha, and Steenkamp (2022)	X	X	X	X		
Scholdra et al (2022)	X	X	X	X		
Umashankar, Kim, and Reutterer (2022)						X
Van Ewijk, Gijsbrechts, Steenkamp (2022a)	X	X	X	X		
Van Ewijk, Gijsbrechts, Steenkamp (2022b)	X	X	X	X		
Widdecke et al (2022)		X		X		
Zhang et al (2022)		X				

[a]: "SF" denotes Salesforce;"Other" includes word-of-mouth, warranty claims, and store visits, etc. The detailed list of the publications appears in Web Appendix A.

A common use for copula correction stems from applications of the marketing mix (price, product, place, and promotion) of goods and services. A primary reason for this is such regressors are often correlated with the error term in a regression model because of uncaptured managerial knowledge in decision-making (i.e., setting prices is often related to the cost of production or anticipating consumer demand; advertising budgets are often set as a percentage of sales). For instance, Park and Gupta (2012) initially use copulas for prices, noting "there are unmeasured product characteristics, or demand shocks, that influence not only consumer decisions but also retailer pricing decisions" (p.582). Danaher (2023) uses copulas for price when looking at optimal advertising targeting of consumers. The concern for pricing is that managers may set prices relative to the cost of production, as a percentage of sales, or anticipating consumer demand. In their study of electronics and appliance sales, Datta et al. (2022) use copulas for line length, price, and distribution; retailers may stock more models of brands that sell better, which of course may get increased sales from greater distribution reach. Besides line length, product features can encompass elements like R&D spending, such as Walmart's sustainability mandate for its suppliers (Gielens et al. 2018), or even movies where the brand equity of actors may be endogenous due to the number of movie appearances, award nominations, or award wins (Mathys, Burmester, and Clement 2016). Advertising also commonly uses copulas, since managers often set advertising budgets

as a percentage of sales or relative to a competitor or industry benchmark. In modeling the conversion of customers to contact an insurance agent, Guitart, Hervet, and Gelper (2020) use copulas for the focal brand's advertising, particularly in its relation to when and where the brand's primary competitor is advertising.

Another area using copula correction is salesforce and customer relationship management (CRM) (Table 1). Endogeneity can arise in this area because allocating particular sales personnel to particular clients, or incentivizing sales personnel may be correlated with unobserved variables like the motivation and/or ability of the sales personnel or the value of clients. Atefi et al. (2018) use copulas for salesforce training, and Burchett, Murtha, and Kohli (2023) use copulas for salesperson's interactions with secondary items (either other people or objects like computers) when talking with customers. CRM endogeneity may occur in efforts to connect with customers, such as donation frequency and amounts (Schweidel and Knox 2013), or communications with buyers (Ludwig et al. 2022).

Copula correction can also be found in areas other than traditional marketing mix and sales efforts. A recurring explanation for the use of copula correction in the studies noted in Table 1 is where reverse causality or common shocks could affect the endogenous variable. In retail research, for instance, Gijsbrechts, Campo, and Vroegrijk (2018) examine household grocery spending, in particular with using copulas for visiting hard discounters (i.e., stores with very low prices), since this becomes habit reinforcing for consumers to then spend their budget there. With social media, Fossen and Bleier (2021) use copulas to examine endogeneity when studying if online program engagement of television shows (word-of-mouth volume and deviation) affects audience size. The testing is warranted since increasing audience size may reversely cause an increase in word-of-mouth activities.

In these cases, copula correction provides a feasible approach to controlling for the thorny regressor endogeneity issue and offers opportunities for optimal managerial decision makings, as further illustrated in the following running example.

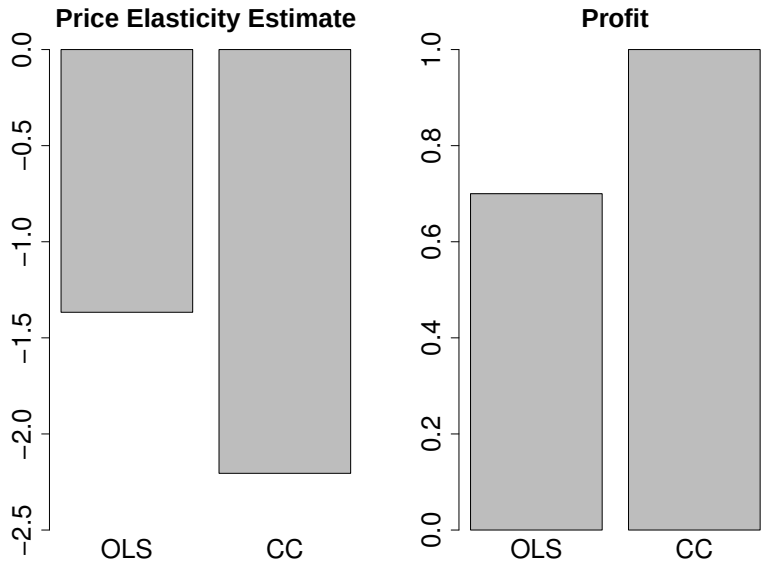

Figure 1: Example 1: Impact of copula correction on price sensitivity estimation. OLS: ordinary least squares; CC:copula correction.

Example 1: Price Sensitivity Estimation. Store managers and policy-makers are often interested in learning price sensitivity for category demand growth. This example estimates price sensitivity for the diapers' category using store scanner purchase data from the IRI Academic data set for the years 2002-2006 (261 weeks) for one focal store in the Buffalo, NY market. In this instance, price was typically treated as endogenous because of unobserved variables (e.g., product characteristics, retailer pricing decisions, number of shelf facings) that, when omitted from a model, become part of the structural error. It is expected that these unobserved characteristics induce positive correlation between price and the error term, thereby causing the OLS estimate of price sensitivity biased toward zero (i.e., less negative). As shown in a later section, the OLS price estimate in this data set is -1.367, which is significantly less than the price estimate of -2.205 from copula endogeneity correction (Figure 1). Using the OLS price estimate, the manager will underestimate consumer price sensitivity and mistakenly set the price too high, resulting in lost revenue and profit. The analysis in the later section shows that using the OLS price estimate will yield 30% less profit compared to using the copula corrected price sensitivity estimate (Figure 1).

We will return and speak more to this later in Example 1, but it directly indicates the impact of a "wrong" estimate: without correcting for endogeneity, OLS yields a price elasticity of -1.367, but using a copula to correct for endogeneity shows a price elasticity of -2.205, a 61% difference. Meta-analyses of studies that compare estimates after endogeneity correction to uncorrected estimates also find similar differences. Bijmolt, Van Heerde, and Pieters (2005) found price elasticity was -2.47 without endogeneity correction, but -3.74 when corrected. Sethuraman, Tellis, and Briesch (2011) found "Advertising elasticity is lower when endogeneity in advertising is not incorporated in the model" (p.470). With personal selling (i.e., salesforce), models that account for endogeneity have lower elasticity (.282) than models without endogeneity correction (.373), a significant difference of .091 that importantly represents an over-estimation of 32% (Albers, Mantrala, and Sridhar 2010). The importance of endogeneity correction should be apparent: without its correction, managers and academics are likely experiencing under-estimated effects of pricing and advertising and over-estimated effects of salesforce.

VARIATIONS IN THE USE OF COPULA CORRECTION

Given the importance of endogeneity correction and the growing popularity of copula correction, several questions arise for best practices. How should researchers utilize copula endogeneity correction? Under what conditions can copula correction be used or not used? Are there concerns when higher-order terms – like interactions and squared terms of endogenous regressors – when using copula correction? Per Table 2, there exist appreciable variations in the use of copula endogeneity corrections among researchers and practitioners.

These variations in copula correction methods and implementation can substantially affect the performance of copula correction. Becker, Proksch, and Ringle (2021) discovered substantial bias of Park & Gupta's copula corrected parameter estimates if the structural model contains the intercept, and cautioned the use of copula correction in such models with small to moderate sample sizes. We study this issue and evaluate an alternative im-

Table 2: Variations in Copula Endogeneity Correction Methods

Items	Approach
Copula transformation of the largest value	• Assigned a fixed value (Gui et al. 2023; Becker, Proksch, and Ringle 2021) • Assigned the same value as that of the second largest value (Papies, Ebbes, and Heerde 2017) • Assigned as $\Phi^{-1}(\frac{n}{n+1})$ + (Yang, Qian, and Xie 2022)
Endogenous regressors with insufficient nonnormality	• Not allowed in Park and Gupta (2012); Eckert and Hohberger (2022) Becker, Proksch, and Ringle (2021); Haschka (2022) • Allowed in Yang, Qian, and Xie (2022)
Exogenous Regressors	• Do not account for correlated exogenous regressors Park and Gupta (2012); Eckert and Hohberger (2022) Becker, Proksch, and Ringle (2021) • Account for correlated exogenous regressors Haschka (2022); Yang, Qian, and Xie (2022)
Higher-order Endogenous regressors	• Include corresponding copula correction terms (see Table 3) • Exclude corresponding copula correction terms (see Table 3)

+: A justification of this formula is provided in the note under Table 4.

plementation of copula transformation that has strong theoretical support and avoids such bias. Recent research also shows that failure to account for exogenous regressors correlated with endogenous regressors can adversely affect copula correction effectiveness in eliminating endogeneity bias (Haschka 2022; Yang, Qian, and Xie 2022). Originally, copula correction required sufficient nonnormality of endogenous regressors, but a recent two-stage copula correction method relaxes this requirement, and can handle endogenous regressors that are normally distributed or close-to-normal (Yang, Qian, and Xie 2022).

Another important and unaddressed issue arises regarding the best way to address endogeneity bias for models containing higher-order terms of endogenous regressors (Table 3). Many applications in different fields are interested in estimating structural models with higher-order terms of endogenous regressors. Polynomial regressions are employed to study non-monotonic causal relationships, such as an inverted-U relationship, often to determine optimal policy and managerial intervention (Aghion et al. 2005; Qian 2007). Interaction terms are included in models to study relevant moderators of causal relationships.

Table 3: Examples of Applications Involving Higher-order Endogenous Terms.

Study	Higher-Order Endogenous Regressors	CHI*
Burmester et al. (2015)	Ad Stock * Publicity Stock	Yes
Blauw and Franses (2016)	Mobile Phone Ownership^2	Yes
Lenz, Wetzel, and Hammerschmidt (2017)	Corporate Social Responsibility^2	No
Lamey et al. (2018)	Promotion Intensity * Store context	No
Gielens et al. (2018)	R& D * Retailer Power	No
Yoon et al. (2018)	Knowledge * Government Activity	Yes
Atefi et al. (2018)	Trained Percentage^2	Yes
	Trained Percentage *Performance Diversity	
Guitart, Gonzalez, and Stremersch (2018)	Advertising * Price	No
Wetzel et al. (2018)	Recruitment Spend * Brand Age	No
Keller, Deleersnyder, and Gedenk (2019)	Price Index * Price Premium	No
Heitmann et al. (2020)	Complexity *Segment Typicality	No
Vomberg, Homburg, and Gwinner (2020)	Failure Culture* Reacquisition Policies	No
Guitart and Stremersch (2021)	Ad Stock^2, Price^2, Informational^2	Yes
Magnotta, Murtha, and Challagalla (2020)	Salesperson Training*Salesperson Incentive	No
Homburg, Vomberg, and Muehlhaeuser (2020)	Direct Channel Usage*Formalization	No
Liu et al. (2021)	Price Discount^2, order Coupon^2	Yes
Krämer et al. (2022)	Industrial Service Share^2	Yes

CHI: copula correction terms for high-order terms of endogenous regressors included.

As shown in Table 3, there exists inconsistencies in the literature regarding how to handle higher-order terms of endogenous regressors with the copula correction procedure. While some studies did not include copula generated regressors for endogenous higher-order terms (often without stating the reason), other applications argued for including these generated regressors to control for endogeneity. For instance, Atefi et al. (2018) (p.730) note "we added the squared term of the training percentage (TPS) to the regression model to determine whether we could replicate the nonlinear pattern obtained in Panels A and B of Figure 1. Following the suggestions in the literature (Wooldridge 2010), we treated TPS as a second endogenous variable and thus added a second Copula endogeneity-correction term to the regression model", while Yoon et al. (2018) (p.249) state, "The interaction terms of knowledge with areas of government activities are also subject to endogeneity. Therefore, we constructed additional variables [copula correction terms]".

To illustrate the impact of variations in using copula correction, consider the following

running example.

Example 2: Moderator of Price Sensitivity Of interest here is that price and a retail store's feature advertising likely work together to achieve interactive, synergistic effects on sales. This can be tested by estimating the interaction term between price and feature advertisement in a sales model, with feature advertisement as a potential moderator of price. Blattberg and Neslin (1990) note that feature advertising "may interact with price discounts. If the consumer is not informed that a price discount is offered, the price elasticity is likely to be small" (p.347). This suggests a negative sign for the interaction term between price and feature advertisement.

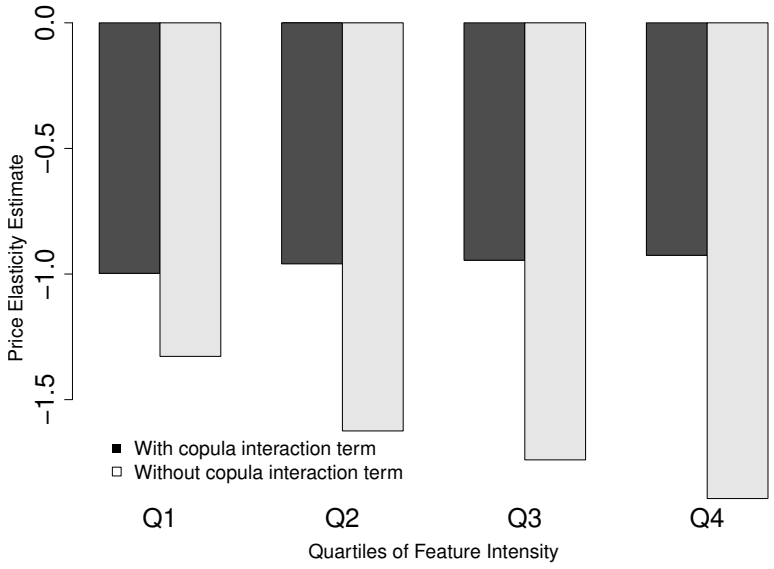

Figure 2: Mean price sensitivity estimates per quartile of feature intensity.

Figure 2 presents the mean price sensitivity estimates per quartile of feature intensity for the peanut butter category, predicted from a sales demand model with an interaction term between price and feature, estimated using the IRI academic data for a store in New York city. The black (white) bars are price sensitivity estimates estimated with (without) a copula term for the interaction term. Including the copula term for the interaction term yields price sensitivity estimates that are the same across different feature intensity (meaning lack of interactive effect); excluding the copula term yields a greater magnitude of price sensitivity,

and the price sensitivity estimates increase with greater feature advertisement. As shown later, adding the copula term for the interaction term can induce bias and greatly increase variability of parameter estimates.

METHODOLOGICAL BACKGROUND

In the section, we discuss the methodological aspects of the copula endogeneity correction. Our discussion aims to acquaint readers with the concepts and procedures of copula correction, to address the inconsistencies in the use of copula correction, and to inform the decision process guiding the proper use of copula correction.

Accounting for regressor-error dependence using copula

A primer on the copula joint estimation approach

We first review the copula endogeneity correction approach of Park and Gupta (2012), to account for the dependence between endogenous regressors and the error term. Consider the following linear structural model:

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

where $t = 1, \cdots, T$ indexes time, market, or cross-sectional units; Y_t is a scalar response variable (e.g., log-transformed volume of diapers sold in week t in Example 1), P_t contains the endogenous regressor (log-transformed price in Example 1), and W_t contains a vector of exogenous control variables. The regression coefficient α is the structural model parameter capturing the causal or independent effects of P_t.

A noted previously, the association between the endogenous regressor P_t and the structural error E_t can lead to biased estimates of model parameters (i.e., endogeneity bias) when using estimation methods (e.g., OLS) that ignore the regressor-error dependence. One approach to addressing this endogeneity bias is to directly model and incorporate the regressor-error dependence into inference. Park and Gupta (2012) proposed a novel approach, henceforth denoted as P&G, positing a Gaussian copula (GC) to link the marginal distributions of P_t and E_t together to obtain the joint distribution of (E_t, P_t). The GC model has desir-

able properties, making it frequently used in management to robustly capture multivariate dependence (Danaher 2007; Danaher and Smith 2011). In particular, the GC model with nonparametric empirical marginals depends on the rank-order of raw data only, and is invariant to strictly monotonic transformations of variables in (P_t, E_t).

Park and Gupta (2012) propose two estimation methods based on the GC model under the assumption of a normal structural error, $E_t \sim N(0, \sigma^2)$. The first maximizes the likelihood function derived from the joint distribution of (E_t, P_t). See Park and Gupta (2012) and recent extensions of the maximum likelihood estimation by Tran and Tsionas (2021) and Haschka (2022) for more details. The second uses a generated regressor approach that is straightforward to apply and has been used in the majority of applications using copula correction (Becker, Proksch, and Ringle 2021; Eckert and Hohberger 2022; Yang, Qian, and Xie 2022). Thus, our discussion hereafter focuses on the generated regressor approach that estimates the following augmented regression model

$$Y_t = \mu + \alpha P_t + + \beta' W_t + \gamma P_t^* + \epsilon_t \tag{2}$$

$$\text{where} \quad P_t^* = \Phi^{-1}(F_P(P_t)); \tag{3}$$

$F_P(\cdot)$ denotes the marginal cumulative distribution function (CDF) of P; $\Phi^{-1}(\cdot)$ denotes the inverse CDF of the standard normal distribution; γ is the coefficient parameter for P^*.

Under the GC model for (P_t, E_t), the added term P_t^* in Equation 2 captures the correlation between the endogenous regressor P and the error term E, and consequently the new error term ϵ_t in Equation 2 is independent of P_t given P_t^* in the model. Based on this result, the P&G procedure includes the copula term P_t^* as an additional control variable in the structural model to correct for the endogeneity of P. The computation of the generated regressor $P_t^* = \Phi^{-1}(F_P(P_t))$ requires an estimate of $F_P(\cdot)$, the unknown marginal CDF of the endogenous regressor P_t. The popular approach is to estimate $F_P(\cdot)$ with the empirical CDF, $\widehat{F}_P(\cdot)$, which assigns probability mass to the uniquely observed values of P_t in the sample according to their sample frequencies. To account for the additional uncertainty introduced during the estimation of $F_P(\cdot)$, standard errors of the model estimates are obtained using

bootstrap resampling (Park and Gupta 2012).

The P&G procedure can handle multiple endogenous regressors. For K continuous endogenous regressors $(P_1, \cdots, P_K)$, the generated regressor approach estimates the following augmented regression model:

$$Y_t = \mu + \sum_{k=1}^{K} P_{t,k}\alpha_k + +\beta' W_t + \sum_{k=1}^{K} P^*_{t,k}\gamma_k + \epsilon_t, \tag{4}$$

$$\text{where} \quad P^*_{t,k} = \Phi^{-1}(\widehat{F}_{P_k}(P_{t,k})); \tag{5}$$

γ_k is the coefficient parameter for P^*_k; $\sum_{k=1}^{K} P^*_{t,k}\gamma_k$ is the linear combination of the K copula terms $\{P^*_{t,k}\}$ used to control for the endogenous regressors and thus is denoted as the copula control function (CCF).

Assumptions of the P&G procedure

For proper use of the P&G procedure, it is important to understand the assumptions behind the method. The P&G procedure makes the following assumptions.

- Assumption 1. The structural error follows a normal distribution.
- Assumption 2. P_t and the structural error follow a Gaussian copula.
- Assumption 3. Full rank of all regressors and $Cov(W_t, E_t) = 0$.
- Assumption 4. P_t is nonnormally-distributed.
- Assumption 5: The linear combination of $P^*_{t,k}$, $\sum_{k=1}^{K} P^*_{t,k}\gamma_k$, is uncorrelated with W_t.

Assumptions 1 and 2 are used in the conversion from the GC model for (P, E) to the augmented regression models in Equations 2 and 4. However, the P&G procedure exhibits reasonable robustness to nonnormal error distributions and alternative non-Gaussian copulas (Park and Gupta 2012), but might not withstand gross departures from the two assumptions, such as highly skewed error distributions or arbitrary dependence structures (Becker, Proksch, and Ringle 2021; Eckert and Hohberger 2022). Eckert and Hohberger (2022) also show that the P&G method performs on par with or better than the alternative IV estimation with a moderately skewed error distribution. If a highly skewed error distribution is suspected, it is advisable to consider alternative model specifications (e.g., transforming variables). As noted in Danaher and Smith (2011) and Eckert and Hohberger (2022), the GC

model can capture dependence among variables in most applications. Specifically, the structural error can often be expressed as the summed term for the combined effect of unmeasured confounders and a white noise term; in many settings the combined effect of unmeasured confounders and the endogenous regressor jointly follow a GC model, leading to a GC model for the endogenous regressor and the error term (i.e., Assumption 2).

Assumptions 3 to 5 are needed for ensuring the consistency of augmented OLS regression in Equations 2 and 4. Two important conditions are required for consistency of the augmented OLS estimates: full column rank condition of the regressor matrix, and zero correlation between regressors and the new error term ϵ (Wooldridge 2010). Assumption 3 is essential for all common econometric methods, such as OLS and IV regression. Assumption 4 is important and established in the literature: almost all the applications of the P&G method checked for this condition. If P approaches the normal distribution and consequently is close to a linear function of P^*, the resulting collinearity between P and P^* can lead to large standard errors; this renders the precise evaluation of the independent effect of P impossible with a finite sample size (Park and Gupta 2012). In the extreme case when P is normally distributed, the augmented OLS regression fails by violating the full rank condition of the regressor matrix. In contrast, Assumption 5 was implicit until recently[4]. When Assumption 5 is violated, the new error term ϵ_t in the augmented OLS regression becomes correlated with the exogenous regressors W_t, which subsequently may bias estimates of all model parameters. Thus, Assumptions 4 and 5 limit the applicability of the P&G procedure. We describe in a later subsection a recent development that relaxes Assumptions 4 and 5, the two-stage copula endogeneity correction method. Before then, the next subsection discusses the algorithm to produce generated regressor P^*, which can substantially affect copula correction performance.

[4]As shown in Yang, Qian, and Xie (2022), this assumption is weaker than the assumption that exogenous and endogenous regressors are uncorrelated as suggested in Haschka (2022).

Proper construction of nonparametric rank-based copula transformation

As noted above, almost all applications of copula endogeneity correction employ the nonparametric rank-based copula transformation based on the empirical marginal distributions of regressors (Equation 5). Although convenient and immune to misspecifications of these nuisance marginal distributions, the empirical copula transformation requires special handling of mapping from ranks to latent copula data. To demonstrate how the empirical rank-based copula transformation is constructed, consider the example of the selling price of twenty goods from a small retailer, as shown in Table 4. The construction of the empirical rank-based copula follows two steps, per Equation 5. First, the observations are ordered and mapped to a ranked percentile according to the empirical cumulative distribution, $F(\cdot)$. For example, the first observation (of twenty) is $\frac{1}{20}$, or 5% of the cumulative observations; the second observation is $\frac{2}{20}$, or 10%, and so on. The second step computes the inverse normal CDF of that ranked percentile as shown in the column "Price*": an observation in the bottom 5% (or fifth percentile) maps onto the far left end of a standard normal distribution, in this case about -1.6449 standard deviations below 0.

One item from Table 4 is of particular importance: the last observation is technically the 100th percentile, however, the inverse normal CDF of the 100th percentile is undefined. This is because the probability (reflected as F) must be between 0 and 1. The latent copula data, Price*, for the 20th observation here reflects an adjustment, where $F(\cdot)$ becomes the observation count divided by the observation count plus one (i.e., $\frac{n}{n+1} = \frac{20}{21}$) for the reason given in the note under Table 4. That is, we compute the copula transformation as

$$P_t^* = \Phi^{-1}(F_P(P_t)) \quad = \quad \begin{cases} \Phi^{-1}(\texttt{Rank}(P_t)/n) & \text{if } P_t < \max(P) \\ \Phi^{-1}(n/(n+1)) & \text{if } P_t = \max(P). \end{cases} \tag{6}$$

Besides ensuring that the copula transformed values maintain the same rank order as the original regressor values for any sample size[5], the percentile adjustment for the maximum

[5]By contrast, in their example of 100 observations, Papies, Ebbes, and Heerde (2017) set the percentile for the last observation to 0.99, which is the same as the second to last observation even though these two raw data points do not have the same rank order.

Table 4: Example Creation of the Rank-based Gaussian Copula

Obs	Price	F(Price)	Price*	Obs	Price	F(Price)	Price*
1	\$14.00	0.05	-1.6449	11	\$32.10	0.55	0.1257
2	\$15.20	0.10	-1.2816	12	\$33.00	0.60	0.2533
3	\$16.30	0.15	-1.0364	13	\$34.60	0.65	0.3853
4	\$16.50	0.20	-1.0364	14	\$34.90	0.70	0.3853
5	\$21.00	0.25	-0.6745	15	\$37.00	0.75	0.6745
6	\$24.20	0.30	-0.5244	16	\$42.00	0.80	0.8416
7	\$27.00	0.35	-0.3853	17	\$43.50	0.85	1.0364
8	\$29.00	0.40	-0.2533	18	\$44.10	0.90	1.2816
9	\$29.50	0.45	-0.2533	19	\$45.00	0.95	1.6449
10	\$30.00	0.50	0.0000	20	\$47.80	0.9524^{+}	1.6684

+: To avoid generating undefined latent copula data, the rank for the maximum value of Price is changed from 1 to n/(n+1), which is 20/21=0.9524 for the sample size $n = 20$ here. A justification of this formula is that the expected value of the maximum of a standard normal sample of size n can be approximated by $\Phi^{-1}(\frac{n-\alpha}{n+1-2\alpha})$ with a recommended value for α as $\alpha = 0.375$ (Royston 1982). The use of $\Phi^{-1}(\frac{n}{n+1})$ can be viewed as setting $\alpha = 0$ in the formula, which is simpler to use and leads to almost identical result as setting $\alpha = 0.375$ for typical sample size (i.e., $n >> \alpha$) seen in practical studies. $\Phi^{-1}(\cdot)$ is the standard normal inverse cumulative distribution function.

value yields a theoretically valid maximum value of the underlying copula data, and stabilizes the copula transformation without producing an extremely transformed value.

To demonstrate the importance of the empirical copula transformation, consider an alternative empirical copula construction as implemented in R package `REndo` (Gui et al. 2023), which is considered in Becker, Proksch, and Ringle (2021) to set the percentile for the last observation to a fixed value of 0.9999999:

$$P^*_{t,Fix} = \Phi^{-1}(F_P(P_t)) \quad = \quad \begin{cases} \Phi^{-1}(\texttt{Rank}(P_t)/n) & \text{if } P_t < \max(P) \\ \Phi^{-1}(0.9999999) = 5.1999 & \text{if } P_t = \max(P), \end{cases} \tag{7}$$

where P^*_{Fix} means a fixed percentile value is used for the largest rank. The fixed value is chosen to be 0.9999999 (close to 1) in order to maintain the same rank order after copula transformation unless sample size is extremely large (i.e, n >1,000,000). However, when sample size is small or moderate, copula transformation of the maximum can differ substantially from the theoretically predicted value; this becomes an outlier in the augmented OLS

regression, which can adversely impact the performance of copula correction.

To assess the impact of empirical copula construction on the performance of copula correction, we compare the algorithm in Equation 6 with the algorithm in Equation 7 using simulation studies[6] in which parameter estimates are compared to the known true values. Consistent with prior literature, the Monte Carlo study employed the same set up as described in Becker, Proksch, and Ringle (2021) and in Web Appendix B. Data is simulated from structural model $Y_t = \mu + \alpha P_t + E_t$, with a Gaussian copula model between the error term and the endogenous regressor P_t that follows a uniform distribution on (0,1). For each simulated data, we apply both our algorithm in Equation 6 and the algorithm in Equation 7 to obtain P^*. For both algorithms, P_t^* is added as a generated regressor in the augmented OLS regression to obtain the corrected estimate of α.

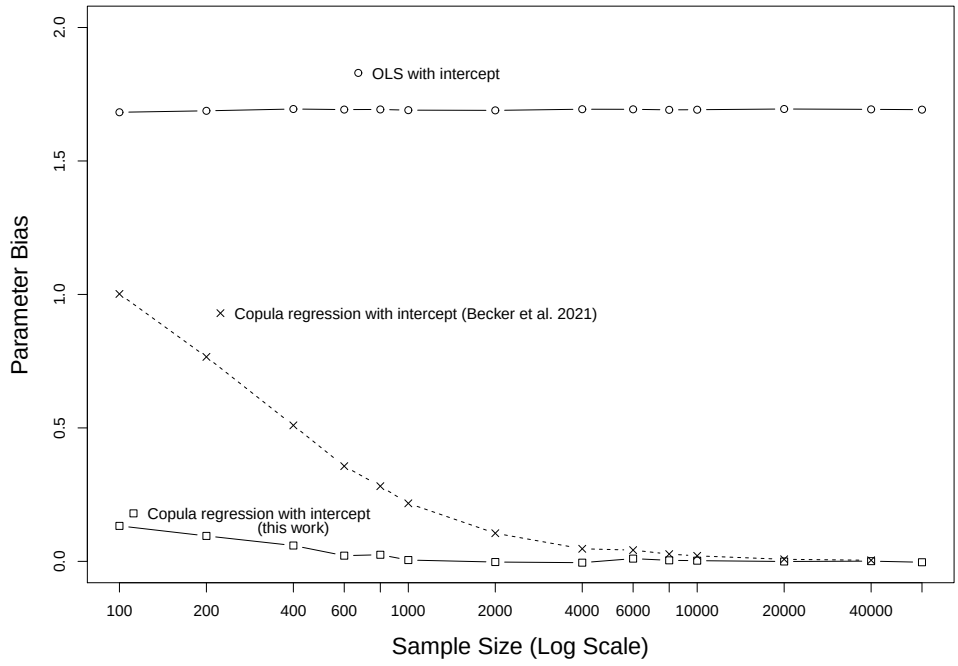

Figure 3: Bias of the endogenous regressor.

Figure 3 shows the bias of α, evaluated as the difference between the mean parameter estimate averaged over 1,000 simulated data sets and its true value, for different estimation

[6]The R codes for simulation studies and empirical examples are available at https://osf.io/by2ge/?view_only=27cc862a9c02446abbafd3a745722603.

methods at sample sizes ranging from 100 to 60,000 (Figure 3 x-axis). OLS, as the curve with circles in Figure 3, exhibits substantial bias (> 1.5) in the coefficient estimate α for endogenous regressor P. Furthermore, this bias remains the same regardless of sample size. Consistent with Becker, Proksch, and Ringle (2021), the P&G method using Equation 7 (the curve with cross marks in Figure 3) substantially reduces the bias in the OLS estimates, but does not resolve the endogeneity in many situations: substantial bias remains after copula correction in small to moderate sample sizes. The endogenous regressor's coefficient estimation bias only becomes negligible for sample sizes larger than 4,000. The finite sample bias for P&G copula regression with intercept discovered in Becker, Proksch, and Ringle (2021) is a significant problem that needs addressing, so as to ensure appropriate use of copula correction. This is relevant because prior to Becker, Proksch, and Ringle (2021), users of copula correction were unaware of such surprisingly severe bias concerns.

We shed light on this issue and reveal that more principled handling of nonparametric rank-based copula transformation is crucial for the performance of copula correction. A key finding of this study is that the substantial bias of the P&G copula correction method for models with intercept[7], discovered in Becker, Proksch, and Ringle (2021), is largely solved by adjusting the largest rank using Equation 6 (the curve with squares in Figure 3). The algorithm in Equation 6 results in considerably improved performance of the P&G copula correction method; the endogenous regressor's coefficient estimate bias now becomes negligible when sample size reaches 400 rather than 4,000. Furthermore, even sample sizes as small as 100 exhibit a bias of about 0.15 for our algorithm, which is quite smaller than 1.0 using the algorithm in Equation 7. The theoretical reason is that constructing the empirical copula using the fixed-value percentile for the largest rank can substantially distort the distribution of generated regressor P^*, resulting in suboptimal performance of the P&G copula correction method and substantial finite sample bias. In conclusion, we recommend

[7]Interestingly, models without intercept are robust to the algorithms to handle largest ranked value when constructing empirical copula. Both algorithms (Equations 6 and 7) yield unbiased estimates for models without intercept (results not shown here as regression models often include the intercept (Becker, Proksch, and Ringle 2021)).

against assigning a fixed percentile value for the largest rank, instead favoring the algorithm in Equation 6 to produce valid empirical copula construction regardless of sample size.

Handling endogenous regressors with insufficient nonnormality and correlated exogenous regressors

The 2sCOPE procedure

Although Figure 3 showed that the algorithm in Equation 6 eliminates the majority of bias in copula regression, noticeable bias remains for the P&G method when sample size is small (e.g., n=100). This is not surprising because the copula correction method, like instrumental variables and other IV-free methods, is a large sample procedure requiring sufficient information for satisfactory performance. More importantly, regardless of the algorithms used to construct empirical copula, the P&G method cannot solve the following two problems that can limit its applicability.

First, as shown in Equation 5 and the paragraphs under Equation 5, the P&G method requires sufficient nonnormality of the endogenous regressor P: a normally or close-to-normally distributed endogenous regressor P can lead to model nonidentification or significant finite sample bias (Becker, Proksch, and Ringle 2021; Eckert and Hohberger 2022). Second, Assumption 5 (uncorrelatedness between CCF and exogenous regressors) of the P&G method may not hold when correlated exogenous regressors are included in the model. Importantly, these two problems - bias due to insufficient regressor nonnormality and the correlation between CCF and exogenous regressors - cannot be solved by employing the algorithm in Equation 6 to construct the empirical copula.

To overcome these limitations of the P&G method, Yang, Qian, and Xie (2022) propose a two-stage copula endogeneity correction (2sCOPE) method that does not require regressor nonnormality or presume uncorrelatedness between endogenous and exogenous regressors; the method leverages correlated exogenous regressors to sharpen structural model parameter estimates. The 2sCOPE method includes the P&G method as a special case and reduces to

the P&G method when no correlated exogenous regressors exist in the model.

For the augmented OLS regression in Equation 2, the generated regressor P^* does not use exogenous regressors in W; this can produce biased estimates when the generated regressor P^* is correlated with the exogenous regressors W. The idea of 2sCOPE is to remove from P^* the component that is correlated with the exogenous regressors, and use the remaining cleaned part of P^* to control for endogeneity. Under the assumption of Gaussian copula for the regressors (P, W) and the error term E, we have:

$$P_t^* = \delta' W_t^* + V_t. \tag{8}$$

where δ contains coefficient parameters, W_t^* is copula transformation of W_t, and V_t is the component of P_t^* that is unrelated to the exogenous regressors but is correlated with the structure error term E_t. With a normal error term E_t, the two error terms V_t and E_t follow a bivariate normal distribution: the correlation coefficient captures the endogeneity of P. For instance, both E_t and V_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. One can then run the following two-stage augmented OLS regression, denoted as 2sCOPE, to correct endogeneity:

1. Regress P_t^* on W_t^* as in Equation 8 and obtain the first-stage residual that removes from P^* the component related to exogenous regressors: $V_t = P_t^* - \widehat{\delta}' W_t^*$.
2. Include the first-stage residual V_t as an additional regressor in the structural model in Equation 1 and perform the following augmented OLS regression:

$$Y_t = \mu + \alpha P_t + \beta' W_t + \gamma V_t + \omega_t. \tag{9}$$

By conditioning on the first-stage residual V_t (the component in P that causes endogeneity but uncorrelated with exogenous regressors), the structural error E_t becomes independent of both P_t and W_t, thereby ensuring the consistency of standard estimation methods.

For K continuous endogenous regressors $(P_1, \cdots, P_K)$, 2sCOPE estimates the following

augmented regression model:

$$Y_t = \mu + \sum_{k=1}^{K} P_{t,k}\alpha_k + +\beta' W_t + \sum_{k=1}^{K} V_{t,k}\gamma_k + \omega_t, \tag{10}$$

$$\text{where} \quad V_{t,k} = P^*_{t,k} - \widehat{\delta}'_k W^*_t; \tag{11}$$

thus, $\sum_{k=1}^{K} V_{t,k}\gamma_k$ is the linear combination of the K residual terms $\{V_{t,k}\}$ used to control for the endogenous regressors (i.e., CCF).

This two-step procedure (2sCOPE) first regresses each $P^*_{t,k}$ on W^*_t and then adds these first-stage residual terms $\{V_{t,k}\}$ to control for endogeneity. In this aspect, $\sum_{k=1}^{K} V_{t,k}\gamma_k$ serves as a control function to correct for endogeneity bias in a similar manner to the control function approach of Petrin and Train (2010). Unlike Petrin and Train (2010), 2sCOPE requires no IVs that must satisfy the stringent condition of exclusion restriction, a much stronger requirement than exogeneity. Furthermore, no arguments for the nature and direction of correlation between W and P are needed: empirical association is sufficient when using 2sCOPE. These gains by 2sCOPE greatly increase the practicality of endogeneity correction.

The 2sCOPE method extends the P&G method in three important aspects. First, unlike P&G, 2sCOPE adds the first-stage residual terms as the control function instead of P^*. As a result, the control function in 2sCOPE accounts for the correlated exogenous regressors. Second, 2sCOPE does not require endogenous regressors to have a nonnormal distribution. Even if the endogenous regressor is normally distributed, 2sCOPE can identify the model as long as one of the correlated W is nonnormally distributed, which is feasible in many empirical applications. Third, while exogenous regressors are not used for generating the CCF in P&G, 2sCOPE can leverage these exogenous regressors to sharpen the structural model estimates. If a powerful regressor is available and included in the model to generate the CCF, 2sCOPE can eliminate the finite sample bias of the P&G method in small samples (Yang, Qian, and Xie 2022). This demonstrates the power of leveraging relevant exogenous regressors to increase the accuracy of the parameter estimates.

Assumptions of the 2sCOPE procedure

The 2sCOPE method makes the following assumptions:

- Assumption 1. The structural error follows a normal distribution.
- Assumption 2. P_t, W_t and the structural error follow a Gaussian copula.
- Assumption 3. Full rank of all regressors and $Cov(W_t, E_t) = 0$.
- Assumption 4. Either P_t or one correlated regressor in W_t is nonnormally-distributed.

As shown in Yang, Qian, and Xie (2022), 2sCOPE increases modeling robustness and reduces dependence on model assumptions as compared with the P&G method. As a result, 2sCOPE has increased robustness to small sample size, nonnormal error distributions, and violations of Gaussian copula dependence. Assumption 3 is not specific to 2sCOPE, but a standard assumption invoked in other commonly used econometric method, such as OLS, two-stage least squares estimation using IVs, and the P&G method. Assumption 3 should be evaluated when specifying the econometric model before deciding on particular estimation strategies. Finally, Assumption 4 is less stringent than P&G's Assumption 4 (nonnormal distribution of P), while 2sCOPE eliminates Assumption 5 in P&G.

Optimal Copula Estimation of Endogenous Moderating and Nonlinear Effects

Many practical applications in different fields are interested in estimating structural models with higher-order terms of endogenous regressors to gain deeper understanding of causal mechanisms. However, considerable variability exists in how to handle these higher-order endogenous regressors. In this section we consider the best approach to handling these higher-order terms when using copula correction.

Theoretical results on optimality

Consider the following general structural model containing higher-order terms of endogenous regressors:

$$Y_t = \mu + \alpha_1' P_t + \alpha_2' f_1(P_t) + \alpha_3' f_2(P_t, W_t) + \beta' W_t + E_t, \tag{12}$$

where P_t is a vector of K continuous and endogenous regressors (i.e., associated with the error term E_t), and W_t is a vector of exogenous regressors. The structural model in Equation

12 expands the model in Equation 1 to include higher-order endogenous terms, namely $f_1(P_t)$ and $f_2(P_t, W_t)$. Below are examples of these higher-order terms:

- Polynomial functions of a scalar P_t: $\alpha_2' f_1(P_t) = \alpha_2 P_t^2$
- Interaction of two endogenous regressors $P_t = (P_{1t}, P_{2t})$: $\alpha_2' f_1(P_t) = \alpha_2 P_{1t} P_{2t}$
- Interaction of endogenous and exogenous regressors: $\alpha_3'(P_t, W_t) = \alpha_3 P_t W_t$

These higher-order terms of endogenous regressors are added into the structural model to capture non-additive interactive effects and more complex nonlinear effects (such as an inverted-U relationship captured by the squared term $\alpha_2 P_t^2$ in the above). For instance, our Example 2 is interested in estimating the synergistic interaction effect between price and feature advertisement. Because these higher-order terms of endogenous regressors are also endogenous and correlated with the structural error, questions arise regarding the optimal approach to handling these higher-order terms. Since both $f_1(P_t)$ and $f_2(P_t, W_t)$ are endogenous, it is tempting to control their endogeneity by adding separate copula correction terms for them. However, the point of not needing these copula correction terms for these higher-order terms is clearly shown in the following augmented OLS regression, including only copula correction terms for the first-order endogenous terms (i.e., main effect):

$$Y_t = \mu + \alpha_1' P_t + \alpha_2' f_1(P_t) + \alpha_3' f_2(P_t, W_t) + \beta' W_t + \gamma' C_{t,main} + \epsilon_t, \tag{13}$$

where $C_{t,main} = (C_{t,1}, \cdots, C_{t,K})$ contains copula correction terms for main terms P_t only, and $C_{t,k} = V_{t,k}, k = 1, \cdots K$, are the first-stage residual terms defined in Equation 11 when 2sCOPE is used, and reduce to $P^*_{t,k}, k = 1, \ldots, K$, in Equation 5 when P&G is used (e.g., no correlated W in the model). Because the new error term ϵ is independent of P and W under the GC model, ϵ is also independent of $f_1(P)$ and $f_2(P, W)$, both of which are deterministic functions of P and W. Thus, once the copula correction terms for main effects C_{main} are included as control variables into the structural model of Equation 13, the new error term ϵ is already independent of (and uncorrelated with) $f_1(P)$ and $f_2(P, W)$, so extra correction terms for $f_1(P)$ and $f_2(P, W)$ are not needed. This simplicity of handling higher-order endogenous regressors is a merit of the copula correction approach.

Although it is unnecessary to add the copula correction terms for higher-order terms, i.e., $C_{f_1(P_t)}$ and $C_{f_2(P_t,W_t)}$, a further question is what will happen if the additional copula generated regressors for the higher-order terms is included. Will doing this lead to better or worse performance of the copula correction?

The issue with adding unnecessary regressors $C_{f_1(P_t)}$ and $C_{f_2(P_t,W_t)}$ is the significant collinearity between these higher-order copula terms and their co-varying constituents (P, $f_1(P)$, $f_2(P,W)$, and C_{main}), making it harder to distinguish the independent effects of the first-order and higher-order terms involving the endogenous regressors. As a result, this substantially decreases precision of the regression coefficient estimates, and makes copula correction methods perform worse than otherwise, shown formally by Theorem 1 in Web Appendix C.

Empirical Assessment

In addition to the above theoretical results, we present empirical evidences using simulated data to demonstrate (1) that there is no need to add correction terms for higher-order terms of endogenous regressors to control for their endogeneity, and more importantly, (2) harmful effects occur if correction terms for higher-order terms are added to control for their endogeneity. These effects include potential finite sample bias and inflated variability of structural model parameter estimates, as predicted by the theoretical results in the previous section. The simulations further highlight the magnitude of such harmful effects.

Case I: Interaction of endogenous regressors Data were simulated from the following model (subscript t omitted for simplicity):

$$Y = \mu + \alpha_1 P_1 + \alpha_2 P_2 + \alpha_3 P_1 * P_2 + E, \tag{14}$$

where the endogenous regressors (P_1, P_2) and the error term E were generated from a Gaussian copula (Web Appendix D). For each simulated data set, the following three estimation

procedures were applied regressing Y on the following sets of regressors:

$$\begin{aligned} \text{OLS:} \quad & P_1, P_2 \\ \text{Copula-Main:} \quad & P_1, P_2, C_{P_1}, C_{P_2} \\ \text{Copula-All:} \quad & P_1, P_2, C_{P_1}, C_{P_2}, C_{P_1 * P_2} \end{aligned}$$

where $C_{P_1} = \Phi^{-1}(\widehat{F}_{P_1}(P_1))$, $C_{P_2} = \Phi^{-1}(\widehat{F}_{P_2}(P_2))$, and $C_{P_1 * P_2} = \Phi^{-1}(\widehat{F}_{P_1 * P_2}(P_1 * P_2))$ are the copula correction terms; Copula-Main indicates including copula correction terms for the main effect only, while Copula-All signifies including copula correction for all terms involving endogenous regressors, including the interaction term.

Table 5: Results from Case I: Interaction of Endogenous Regressors.

N	Method	$\mu(=0)$	$\alpha_1(=1)$	$\alpha_2(=-1)$	$\alpha_3(=1)$	$\sigma(=1)$	D-error
500	OLS	**-7.624**	**2.281**	**-1.546**	**1.432**	**0.297**	
		(0.290)	(0.058)	(0.312)	(0.066)	(0.019)	—
	Copula-Main	**-0.119**	1.019	**-1.104**	1.024	0.99	
		(0.899)	(0.179)	(0.254)	(0.047)	(0.076)	0.0117
	Copula-All	**0.176**	0.974	**-0.702**	**0.923**	1.051	
		(0.902)	(0.178)	(0.331)	(0.077)	(0.086)	0.0165
5,000	OLS	**-7.623**	**2.281**	**-1.549**	**1.432**	**0.298**	
		(0.092)	(0.018)	(0.099)	(0.021)	(0.006)	—
	Copula-Main	-0.012	1.002	-1.017	1.003	1.000	
		(0.291)	(0.058)	(0.080)	(0.015)	(0.024)	0.0011
	Copula-All	**0.202**	0.968	**-0.713**	**0.929**	1.044	
		(0.318)	(0.061)	(0.240)	(0.058)	(0.041)	0.0031

Table presents the averages of the estimates and standard errors (in parenthesis) over the repeated samples. Bold numbers highlight estimates with bias of at least 0.05. Underlined numbers highlight where the standard errors of the Copula-All estimates are inflated by at least 50% compared with the corresponding ones from Copula-Main. Results for sample size=200 and 50,000 are in Table W1 in Web Appendix D.

Across simulations, sample sizes (N) of 200, 500, 5,000, and 50,000 are examined. For each sample size N, we generate 5,000 data sets as replicates to systematically evaluate average performance (estimation bias and variability) for the three estimation methods. For each parameter, Table 5 reports the average of the estimates and standard errors (in parenthesis) computed across replicates for sample size=500 and 5,000. Results for sample size=200 and

50,000 are reported in Table W1 (Web Appendix D). As expected, OLS estimates have significant bias for all model parameters at all sample sizes. For example, even for a large sample size of N=5,000, the OLS regression (without any correction terms) yields large bias for the regression parameter estimates ($\widehat{\alpha}_1$: 2.281 [0.018]; $\widehat{\alpha}_2$: -1.549 [0.099]; $\widehat{\alpha}_3$: 1.432 [0.021]) and the error standard deviation ($\widehat{\sigma}$: 0.298 [0.006]). Copula-Main corrects for the endogenous bias ($\widehat{\alpha}_1$: 1.002 [0.058]; $\widehat{\alpha}_2$: -1.017 [0.080]; $\widehat{\alpha}_3$: 1.003 [0.015]), demonstrating that there is no need to additionally include the copula correction term, $C_{P_1*P_2}$. Furthermore, Copula-Main performs substantially better in both estimation bias and variability for all parameter estimates than Copula-All which includes $C_{P_1*P_2}$. In fact, Copula-All yields significantly biased parameter estimates, even at the large sample size of N=5,000 ($\widehat{\alpha}_0$: 0.202 [0.318]; $\widehat{\alpha}_2$: -0.713 [0.240]; $\widehat{\alpha}_3$: 0.929 [0.058]); bias decreases as sample size increases, but remains apparent even for a large sample size of 50,000 (Table W1 in Web Appendix D).

We further compare the efficiency of Copula-All and Copula-Main using the D-error measure (Arora and Huber 2001; Qian and Xie 2022). The D-error measure is defined as $|\Sigma|^{1/K}$ where Σ is the variance-covariance matrix of the regression coefficient estimates, and K is the number of explanatory variables in the structural regression model. A larger D-error value means lower efficiency, with a $\Delta\%$ increase in D-error corresponding to a $\Delta\%$ larger sample size required to achieve the same level of estimation precision. As shown in Table 5, the D-error inflation for Copula-All is about 3-times at N=5,000. In this case, Copula-All requires about 3-times the sample size in order to achieve approximately the same accuracy for estimating α_1, α_2 and α_3 jointly as Copula-Main. The variance inflation for the Copula-All estimate of α_3, the coefficient for the interaction term, is much larger and equals $(\frac{0.058}{0.015})^2 \approx 15$ when N=5,000. This means 15-times the sample size is required for Copula-All to achieve the same estimation accuracy of the interaction term as Copula-Main.

Case II and III We also consider the cases of an interaction between an endogenous regressor and an exogenous regressor (Case II in Web Appendix E) and a square term of

an endogenous regressor (Case III in Web Appendix F). The overall conclusion remains the same as that from Case I, in that Copula-Main outperforms Copula-All for correcting the endogeneity bias of OLS estimates. Compared with Copula-Main, Copula-All yields substantially less estimate precision (up to 4-times larger standard errors) and significant finite sample bias (up to 30% bias).

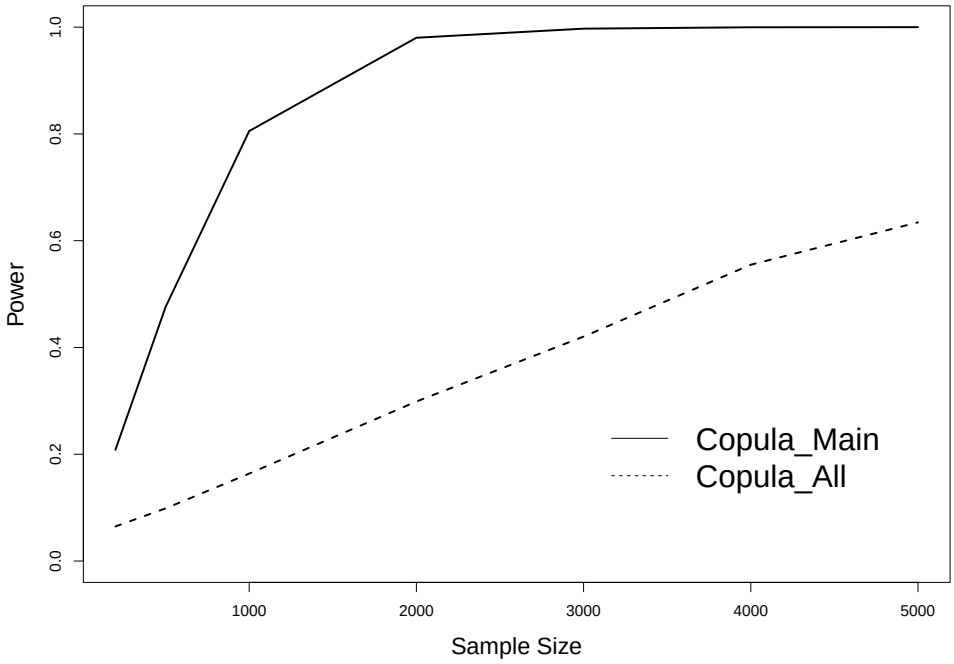

Figure 4: Statistical Power to detect the squared term P^2 with the copula squared term (Copula-All) and without the copula squared term (Copula-Main).

Such a large magnitude of variance inflation has important inferential consequences and managerial implications. Figure 4 shows substantial loss of power of Copula-All to detect the presence of the squared term (P^2) for sample size up to 5,000. For example, when sample size is 1,000, the statistical power to detect the squared effect is about 8-fold for Copula-Main ($\approx$ 80% power) of that for Copula-All ($\approx$ 10% power).

Mean-centering regressors Lastly, we examine whether mean-centering resolve the under-performance of Copula-All. One may suspect that mean-centering might reduce the multi-collinearity issue and improve the performance of Copula-All. However, as shown in Web

Appendix G, mean-centering regressors does not overturn the sub-optimal performance of adding the unnecessary copula correction for higher-order terms, demonstrating again that these unnecessary copula correction terms should be omitted from empirical models.

GUIDANCE FOR PRACTICAL USE

As described in the preceding sections, considerable advances have been made since Park and Gupta's 2012 study with more flexible and general copula correction methods becoming available. We also show that variations in implementing copula correction have substantial impacts on the effectiveness to correct endogeneity. Informed by these findings and advances, this section describes a procedure guiding practical usage of copula correction methods.

Figure 5 presents a step-by-step flowchart for the steps and checkpoints in using copula correction. When conditions are met, the P&G method can be followed, but more recent research relaxes these conditions and presents the path to perform copula corrections when these conditions are not met. Before entering the flowchart, one should ensure the structural model is appropriately specified and theoretically supported, with pertinent exogenous control variables included in W and the regressor matrix being full rank. Revise model specifications (e.g., transform variables) if the error distribution is suspected to be highly skewed, and assess the plausibility of GC dependence in the focal application. As shown in the prior section, copula correction only needs to include CCFs corresponding to the first-order terms P_{main} of endogenous regressors, even when the structural model contains higher-order terms of endogenous regressors. Thus, the flowchart in Figure 5 only needs to consider P_{main}. Furthermore, when the structural model includes an intercept, the copula transformation should use the algorithm in Equation 6 to avoid the estimation bias discovered in Becker, Proksch, and Ringle (2021).

Step 1. This step checks whether the endogenous regressor P_{main} has sufficient support. The copula procedures can handle continuous and discrete endogenous regressors with sufficient support (> 4 ordinal levels), but should not be applied for a binary endogenous

regressor or nominal endogenous regressors whose levels have no natural ordering (Park and Gupta 2012; Eckert and Hohberger 2022; Haschka 2022).

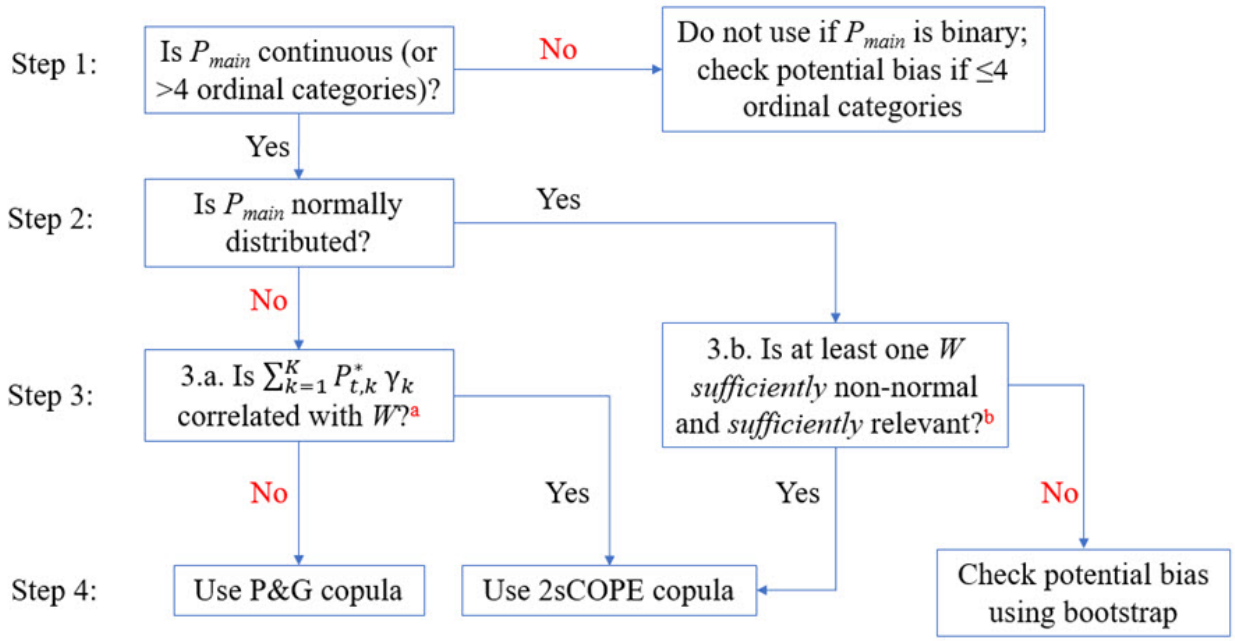

Figure 5: Flowchart for Copula Procedure.

Note: P_{main} denotes the first-order terms of endogenous regressors. W denotes exogenous control variables.

[a]: For multiple endogenous regressors $(P_{main,1}, \cdots, P_{main,K})$, a less stringent condition for using P&G is no correlation between $\sum_{k=1}^{K} P^*_{main,k}\gamma_k$ (the linear combination of copula transformations of all the first-order endogenous regressor terms) and each W. Use the stabilized copula transformation formula in Equation 6 especially when the model includes the intercept.

[b]: W is sufficiently nonnormal when normality test $p < 0.001$ and is sufficiently relevant to P_{main} when F statistic > 10.

Step 2. This step checks whether P_{main} is normally distributed or not. Previously, if P_{main} is normally distributed, Gaussian copulas could not be used (Park and Gupta 2012; Becker, Proksch, and Ringle 2021; Eckert and Hohberger 2022; Haschka 2022) because the model is unidentified. However, the 2sCOPE procedure shows even if P_{main} is normally distributed, it can still be a candidate for copula correction through the 2sCOPE procedure. Yet, this route follows a different path, as seen in Figure 5 and discussed more below in Step 3.b. The literature notes that more powerful tests for normality, such as the Shapiro-Wilk test or Anderson-Darling test, might not fully rule out nonidentification, because these tests can detect small departures from normality that are insufficient for copula correction

(Becker, Proksch, and Ringle 2021; Eckert and Hohberger 2022). The Kolmogorov-Smirnov (KS) test is relatively conservative among the most commonly used normality tests; a p-value less than 0.05 from the KS normality test has been shown to perform well for ruling out finite sample bias due to insufficient regressor nonnormality (Yang, Qian, and Xie 2022). The KS test compares the focal empirical CDF distribution - a quantity linked to copula transformation - with the reference CDF, and is an overall and comprehensive measure to quantify nonnormality.

Step 3. This step marks one of the biggest shifts in copula usage since Park and Gupta (2012), consisting of two disjoint steps (3.a and 3.b), depending on the outcome of Step 2.

3.a. If the endogenous regressor P_{main} is found to have sufficient nonnormality in Step 2 above, Step 3 will check an additional condition to determine if the P&G method or the 2sCOPE method should be used. As noted in the preceding section, the P&G method requires the condition of its control function (i.e., $\sum_{k=1}^{K} P^*_{main,k}\gamma_k$ as the linear combination of the copula transformations of endogenous regressors when P_{main} contains K endogenous regressors) be uncorrelated with exogenous regressors. The correlation between P&G's control function and each exogenous regressor can be checked using Fisher's Z test for correlation. When this condition is met, the P&G method is preferred to 2sCOPE because a simpler and valid model outperforms a more general method. Otherwise, one should use 2sCOPE to handle correlated exogenous regressors. Since P_{main} already has sufficient nonnormality, there is no need for correlated exogenous regressors to be nonnormally distributed.

3.b. If the endogenous regressor P_{main} is found to have insufficient nonnormality in Step 2, then one cannot use the P&G method, but can use 2sCOPE to leverage correlated exogenous regressors to achieve model identification. In order to compensate for the lack of nonnormality of endogenous regressor P, at least one exogenous regressor W needs to satisfy the following two conditions: (1) sufficient nonnormality, and (2) sufficient association with the endogenous regressor P. A conservative rule of thumb for such a W is the p-value from the KS test on W being < 0.001 and a strong association with P (F statistic for the effect of

W^* on $P^*_{main} > 10$ in the first-stage regression). When these conditions are met, even when P_{main} is normally distributed, 2sCOPE is expected to yield estimates with negligible bias. When these conditions are not met, Yang, Qian, and Xie (2022) suggest gauging potential bias of 2sCOPE for data at hand via a bootstrap procedure described there, and using 2sCOPE only if the potential bias is small.

As seen above, only one of 3.a or 3.b is taken in Step 3. Importantly, if P already has sufficient nonnormality that leads to 3.a, there is no need to do 3.b to check if any W has sufficient nonnormality and is associated with P. These conditions are only checked if we need to find a useful W to compensate for the lack of nonnormality of P. In 3.b, 2sCOPE uses W to tease out an exogenous and nonnormally distributed part of the endogenous regressor for model identification. A good starting place to find such W is in the exogenous control variables pre-existing in the OLS or IV regressions. Unlike IVs, these exogenous control variables (e.g., exogenous demand shocks) do not need to satisfy the stringent exclusion restriction condition. That is, these Ws do not have to be excluded from the structural model (e.g., Equation 1), and can affect the outcome directly and not through the endogenous regressors. Such Ws are more readily available than IVs, and because empirical association between the candidate W and P is sufficient, researchers using copula correction do not need to argue for the causal pathways between W and P like in the case of IVs.

Step 4— The final step is to apply the appropriate copula procedure by including in the structural model the generated regressor, which is P^*_{main} if the P&G method is used or the residual term V_{main} from the first-stage regression if 2sCOPE is used. If the generated regressor (i.e., copula correction term) is not statistically significant, this suggests the endogenous regressor P_{main} is not sufficiently correlated with the error term, and endogeneity is unlikely. Thus, non-significant generated regressors should be dropped and the model re-estimated. Marketing studies have dropped copula correction terms at the $p < .10$ level (e.g., Datta et al. 2022), suggesting even marginally significant copula correction terms are still worth retaining. If none of the generated regressors is significant, then the model can be estimated

in a more traditional manner (i.e., OLS).

COPULA IMPLEMENTATION EXAMPLES

In this section, we illustrate use of the flowchart to guide the implementation of copula correction via two examples using weekly store sales data from the IRI Academic data set (Bronnenberg, Kruger, and Mela 2008). To correct for price endogeneity, the first example examines the main effect of price, while the second example examines higher-order moderating effects captured by the interaction between price and store feature (i.e., weekly store flyer promoting products).

Example 1: Main Effects Application of Copula Correction

Returning to our running Example 1, the outcome of interest is the weekly sale volume in the diaper category for one focal store in the Buffalo, NY market in the years 2002-2006, where volume is measured in diaper counts. Price is defined on an equitable volume across UPCs, since pack sizes vary in diapers per pack. IRI additionally collected information on whether UPCs were featured in the store's weekly flyer that week. Category price and feature are evaluated as market-share weighted averages of UPC-level price and feature, respectively.

In this instance, price was treated as endogenous because of unobserved variables (e.g., retailer pricing decisions, number of shelf facings) that, when omitted from a model, become part of the structural error. For brevity, we use "Price" and "Volume" hereafter to refer to the log-transformations of category price and sale volume, respectively. The expected impacts of price and feature advertising appear in the following model:

$$\text{Volume}_t = \mu + \alpha P_t + \beta' W_t + E_t. \tag{15}$$

In the model, P_t is the endogenous regressor as log-transformed price. W_t is a vector of control variables including feature, week, and binary variables for quarters 2, 3, and 4. We treat feature as exogenous because decisions to promote items in the store flyer are made on quarterly basis and require weeks of implementation times, and thus are unlikely

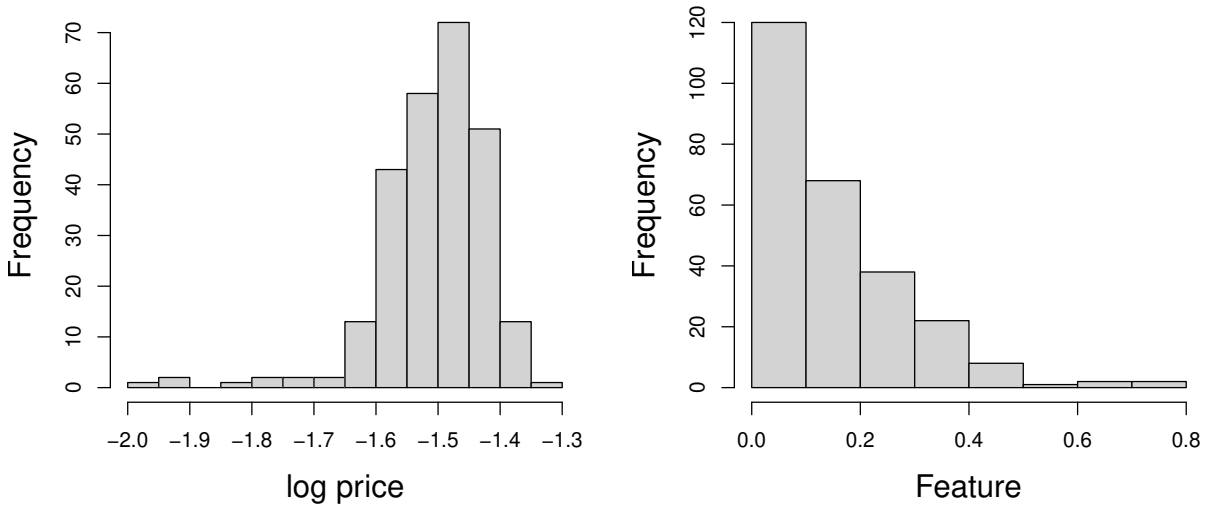

Figure 6: Distributions of Price and Feature in Example 1.

to be correlated with weekly unobservables (Chintagunta 2002; Sriram, Balachander, and Kalwani 2007). The week variable is included as a control variable to account for a small but significant trend in price increases over time. Before we present the results, below we walk through the steps of flowchart in Figure 5.

Step 1 – is P_{main} continuous (or >4 ordinal categories)? The endogenous regressor, Price, is a continuous measure, ranging from \$0.140 to \$0.262 per diaper, with a mean of \$0.221, median of \$0.224, and standard deviation of \$0.018.

Step 2 – is P_{main} normally distributed? Figure 6 shows somewhat skewness to the left for the price variable. However, the skewness is not strong enough to reject the the KS test for normality ($D = 0.08$, $p = 0.06$) at the 0.05 level of significance. Thus, we take a conservative stance and conclude insufficient nonnormality of the price. This means that the endogenous regressor may not have sufficient nonnormality needed for the P&G method to perform well for data at hand. One solution is to leverage related exogenous regressors with sufficient nonnormality via the 2sCOPE method as described next in Step 3.b.

Step 3.b – Is at least one W sufficiently nonnormal and correlated with P_{main}? The first-stage regression shows only one exogenous regressor is sufficiently correlated with the price (F-stat > 10): feature ($F = 16.8$). The regressor, feature, is highly skewed (Figure 6)

and nonnormally distributed based on the KS test ($D = 0.14$, $p < 0.0001$).

Step 4 – Perform 2sCOPE estimation. The above steps show that conditions have been verified such that the 2sCOPE method can be used to handle the price endogeneity. The standard errors are obtained using 500 bootstrap samples.

The estimation results appear in Table 6, which compares 2sCOPE to OLS and two-stage least-squares (2SLS), an instrumental variable approach, where the diaper price of another store in the same market was used as an IV. Prices are correlated for both stores, with the belief that wholesale prices are similar for products sold by the two stores (relevance), but uncaptured product characteristics (including retailer decisions like shelf facings and shelf location) are unlikely related to wholesale prices (exclusion restriction). The 2sCOPE estimation results in Table 6 show that the copula correction term C_{price} (i.e., the first-stage residual) is significant (Est. = .077, SD = .038, $p < .05$), indicating the presence of price endogeneity, so we retain the CCF in the model to control for price endogeneity.

Table 6: Estimation Results for Example 1

Parameters	OLS	2SLS	2sCOPE
Intercept	6.005 (0.205)***	4.371 (0.978)***	4.763 (0.695)***
Price	-1.367 (0.137)***	-2.470 (0.661)***	-2.205 (0.465)***
Feature	0.298 (0.095)**	0.059 (0.178)	0.124 (0.134)
Week	-0.002 (0.000)***	-0.002 (0.000)***	-0.002 (0.000)***
Q_2	-0.019 (0.031)	-0.014 (0.035)	-0.018 (0.033)
Q_3	-0.018 (0.032)	-0.034 (0.036)	-0.029 (0.034)
Q_4	-0.018 (0032)	-0.061 (0.041)	-0.044 (0.037)
C_{price}			0.077 (0.038)**
ρ			0.366 (0.162)**

Note: Table presents estimates and bootstrapped standard errors in the parentheses. * is $p < .10$, ** is $p < .05$, *** is $p < .01$

The results show that while price has the smallest absolute effect in the OLS model (Est. = -1.367, SE = .137, $p < .01$), the effect is greatest in the 2SLS model (Est. = -2.470, SE = .661, $p < .01$); the 2sCOPE price estimate falls in between and is much closer to the 2SLS price estimate (Est. = -2.205, SE = .465, $p < .01$). Compared to 2SLS using IV, the 2sCOPE

results are not unlike that of 2SLS, within one SD of the 2SLS price estimates. The 2SLS price estimate differs somewhat from the 2sCOPE price estimate by 12.0%. Although the correlation in prices between the two stores is significant and passes the weak instruments test ($F = 13.89$, $p < .01$), the correlation is not especially strong ($r = 0.218$). Thus, the difference between 2sCOPE and 2SLS seen here could be because the other store's price as an IV is not particularly strong, and a strong IV is not always readily available. The 2sCOPE shows that price is positively correlated with the error term (Est. = 0.366, SE = 0.162, p <0.05, Table 6), indicating the presence of price endogeneity. This finding is consistent with the result of the Wu-Hausman test ($H = 3.56$, $p < .07$) from 2SLS, which also suggests endogeneity was likely present. Overall, the comparison with 2sCOPE shows that without endogeneity correction, managers would severely under-estimate consumer price elasticity based on the OLS findings for this store, by 38.0%.

Example 2: Copula estimation of endogenous interactions

While Example 1 detailed how to correct for price endogeneity, we now examine what to do when an endogenous regressor has a higher-order effect, such as a squared term or interaction (moderation) with another variable. For brevity, we speak to these higher-order effects simply as interactions. The "METHODOLOGICAL BACKGROUND" section provides studies based on simulated data showing that including a copula for the interaction term may induce bias and inflated estimation variability, and that the best course is to only include copula correction terms for the main effects.

To show how copula correction is applied with interactions of endogenous regressors and examine the adverse effects of including high-order copula correction terms in an empirical application, we extend the sales response model in Equation 15 to include an interaction term ($P_t * F_t$) between price and feature as follows:

$$\text{Volume}_t = \mu + \alpha * P_t + \beta' W_t + \phi P_t * F_t + E_t, \tag{16}$$

where P_t and F_t are category price and feature, respectively, and W_t includes F_t, week,

and binary variables for quarters 2, 3, and 4. We use the IRI academic data set for a new store and product category, a New York City store and its peanut butter sales for the years 2001-2003 (156 weeks), allowing for price and feature to work together as an interaction. Such interactions are common to both academics and managers, as marketing efforts often work together. Of interest here is that price and feature advertising likely work together to achieve interactive, synergistic effects on sales. This can be tested by estimating the interaction term between price and feature advertisement in the above sales model, with feature advertisement as a potential moderator of price. Like Example 1, we follow the same steps in Figure 5 to guide the selection of the appropriate copula method.

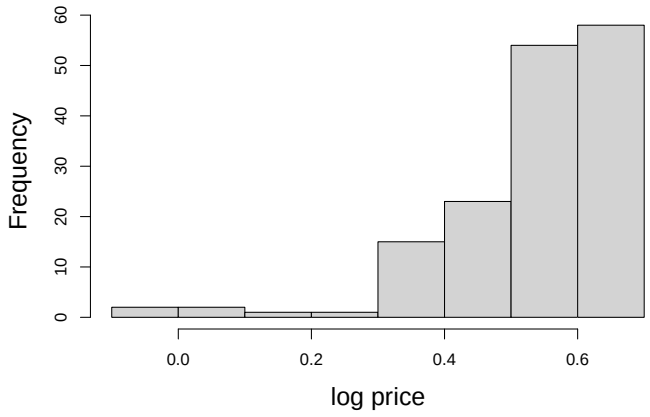

Figure 7: Price Distribution in Example 2.

Step 1 – is P_{main} continuous (or >4 ordinal categories)? Price is a continuous measure here, ranging from \$0.957 to \$1.963 per pound, with a mean of \$1.714, median of \$1.798, and standard deviation of \$0.195.

Step 2 – is P_{main} normally distributed? Unlike Example 1, the price variable in Example 2 is highly skewed (Figure 7) and rejects the KS test for normality ($D = 0.18$, $p < 0.001$) at the 0.05 level of significance. The flowchart in Figure 5 show that what is needed is either P_{main} or one related W is nonnormally distributed. There is no need for both P_{main} and W to be nonnormally distributed. This means that when the endogenous regressor already has sufficient nonnormality, we do not need to check any exogenous regressor W for sufficient

nonnormality and sufficient association with P, like what was needed in Figure 6 of Example 1. To determine if we should use P&G or 2sCOPE, we next check the uncorrelatedness between the linear combination of copula transformations of P_{main} with each W. When P_{main} is a scalar, this condition reduces to check the uncorrelatedness between P^*_{main} and each W.

Step 3.a – is P^*_{main} correlated with W? The copula transformation of endogenous regressor price, P^*, is correlated with the following exogenous regressors at the 0.10 level of significance: week (r = .21, $p < .05$), feature (r = -.76, $p < .01$), Q3 (r = -.16, $p < .06$), and Q4 (r = .16, $p < .04$). This indicates we should use 2sCOPE for endogeneity correction.

Step 4 – Perform 2sCOPE estimation. Until now, the steps had been met to indicate price was a candidate to use the 2sCOPE method. Table 7 presents the 2sCOPE result that includes the copula correction term (i.e., the first-stage residual) for price only. The results show the price copula correction term (i.e., the first-stage residual) is significant (Est. = .069, SE = .033, $p < .05$), indicating the presence of endogeneity. Like Example 1, we also compare the results to OLS and 2SLS, as well as to when a copula correction term (the first-stage residual) for the interaction term is also included (2sCOPE W/Int).

Similar to Example 1, price has the smallest absolute effect in the OLS model (Est. = -.453, SE = .274, $p < .10$) and the greatest absolute effect in the 2SLS model (Est. = -1.554, SE = .606, $p < .05$). The 2sCOPE estimate falls in between, closer to 2SLS in both effect and SE (Est. = -1.314, SE = .580, $p < .10$). The closeness to 2SLS is more expected here since the usage of another store's price is a strong instrument (r = .90, $p < .01$), as 2SLS rejects the test for weak instrument (F = 21.567, $p < .01$); the Wu-Hausman test also suggests endogeneity (W= 4.863, $p < .03$). Without correcting for endogeneity in this example, managers would under-estimate the price elasticity by 65.5% in OLS.

Importantly, the 2sCOPE results point to a contrast with 2sCOPE when a copula correction term $C_{Price*Feature}$ is included for the interaction between price and feature. Here, the price estimate is substantially smaller and becomes insignificant (Est. = -.999, SE =

Table 7: Estimation Results for Example 2

Parameters	OLS	2SLS	2sCOPE	2sCOPE W/Int
Intercept	6.038 (0.165)***	6.688 (0.356)***	6.544 (0.346)***	6.344 (0.394) ***
Price	-0.453 (0.274)***	-1.554 (0.606)***	-1.314 (0.580)**	-0.999 (0.665)
Feature	1.513 (0.234)**	0.646 (0.487)	0.837 (0.491)*	0.619 (0.568)
Price*Feature	-2.125 (0.379)**	-0.950 (0.694)	-1.176 (0.671)*	0.148 (0.999)
Week	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***	0.001 (0.000)***
Q_2	-0.028 (0.034)	-0.020 (0.036)	-0.022 (0.034)	-0.038 (0.041)
Q_3	-0.083 (0.035)	-0.099 (0.038)	-0.096 (0.034)***	-0.089 (0.047)*
Q_4	-0.090 (0.036)	-0.081 (0.038)	-0.080 (0.034)***	-0.066 (0.040)*
C_{price}			0.069 (0.033)**	0.058 (0.038)
$C_{Price*Feature}$				-0.168 (0.088)
ρ_1			0.185 (0.096)*	0.128 (0.101)
ρ_2				-0.456 (0.218)**

Note: Table presents estimates and bootstrapped standard errors in the parentheses. * is $p < .10$, ** is $p < .05$, *** is $p < .01$

.665, $p > .10$ under column "2sCOPE W/Int" in Table 7), which can lead the academic or practitioner to incorrectly conclude price had no significant effect on sales. A more striking difference regards the estimate of the interaction term Price*Feature. The Price*Feature estimates from 2SLS and 2sCOPE (excluding the copula interaction term) are both negative and close: the 2SLS Est. = -0.950 (SE = 0.694, $p > .10$) and 2sCOPE Est. = -1.176 (SE = 0.671, $p < .10$). By contrast, 2sCOPE including the copula term for Price*Feature yields an interaction estimate with the opposite sign and larger SE (Est. = 0.148, SE = 0.999, $p > 0.10$). These results mark an important point: when adding copula correction terms, only copula terms for the main effects should be included, and no copula terms for higher-order terms should be included. Adding the unnecessary higher-order copula terms can lead to substantially varied and biased estimates, including estimates with the opposite sign.

Managerial and Academic Implications

The two examples highlight both how copulas can correct for endogeneity to remove bias in estimation, as well as how copulas should be correctly specified in models with interactions.

Example 1 showed that without the copula, the OLS estimate for price elasticity was severely under-estimated (Est. = -1.367) compared to both 2SLS (Est. = -2.470) and 2sCOPE (Est. = -2.205). The result was price elasticity in OLS was 38% lower than 2sCOPE. We also noted that the instrument was significant but not particularly strong, attributing to the difference between 2SLS and 2sCOPE estimates.

Controlling for endogeneity in price elasticity estimates can have important managerial implications. Price elasticity estimates are often a crucial piece of information for managers to set the optimal pricing that maximizes profit. Let the profit function $p(Price) = V * (Price - Cost)$, where V is the sale volume and *cost* is the marginal cost. The maximum profit is then the value of $Price$ that satisfies the condition $\frac{\partial \ln p(Price)}{\partial Price} = 0$. Following the Amoroso-Robinson relation, the profit-maximizing price is $Price_{optim} = \frac{\alpha}{1+\alpha} Cost$, where α is the price elasticity. In Example 1, we find the optimal pricing is $Price_{ols} = \frac{-1.367}{-1.367+1} Cost = 3.72 * Cost$ if the OLS price elasticity estimate is used, and $Price_{cop} = \frac{-2.205}{-2.205+1} Cost = 1.83 * Cost$ if the 2sCOPE price elasticity estimate is used. Because of the price endogeneity problem associated with the scanner panel data, the biased OLS estimate underestimates the size of price elasticity, meaning that OLS considers consumers less price sensitive than they actually are. Thus, the manager will set the price more aggressively; in Example 1, using the OLS price elasticity estimate means the manager will set price at approximately 100% higher than the actual optimal price.

This considerable difference in optimal pricing based on the OLS and 2sCOPE price elasticity estimates results in a substantial profit difference as well. It can be shown that the profits achieved at the different prices has the following relationship: $\ln \frac{p_{cop}}{p_{ols}} = \alpha \ln[Price_{cop}/Price_{ols}] + \ln[(Price_{cop} - Cost)/(Price_{ols} - Cost)]$, where p_{cop} and p_{ols} refer to the profit achieved when using the 2sCOPE and OLS price elasticity estimates, respectively. From Example 1, the calculation shows $\frac{p_{cop}}{p_{ols}} = 1.46$, which corresponds to a loss of 31% in profit when using the incorrect OLS price elasticity estimate, compared to using the correct 2sCOPE price elasticity estimate (Figure 1).

Example 2 presented the case of the interaction between an endogenous and exogenous regressor. Like Example 1, price elasticity in the absence of feature was substantially underestimated in OLS (Est. = -0.453) than 2SLS (Est. = -1.554) or 2sCOPE (-1.314). The OLS price elasticity estimate was nearly a third that of 2sCOPE.

Furthermore, 2sCOPE including a copula term for the interaction term biased the price elasticity estimate downwards (Est. = -0.999), about 30% lower as compared with the estimate of -1.314 from 2sCOPE excluding the copula term for the interaction term. This bias in the price elasticity estimate becomes even larger as feature intensity increases. Including the copula term for the endogenous interaction term —Price*Feature— yields a severely biased interaction effect estimate; while 2sCOPE without this unnecessary copula term had a negative estimate of -1.176, 2sCOPE including this term (2sCOPE W/Int) produced a positive estimate of 0.148 (Table 7). As shown in Figure 2, including the unnecessary copula term for Price*Feature yields price sensitivity estimates that are the same across different feature intensity (meaning lack of interactive effect); excluding the copula term yields much greater magnitude of price sensitivity that increases with greater feature advertisement. Such drastic differences in price elasticity estimates can have substantive managerial implications, including the optimal price setting and profit maximization, as demonstrated in Example 1. Thus, when the endogenous regressor has a higher-order term (e.g., a squared term or interaction), no copula for the interaction term should be included – only the copulas for the main effects should be included to avoid biasing the estimates.

CONCLUSION

Estimation bias due to the presence of endogenous regressors is a prevalent and important issue to address in business and many other fields. The instrument-free copula correction approach has been increasingly used to address endogeneity bias given its practical advantages and feasible implementation. Yet, like all other causal estimation procedures designed for use with nonexperimental data, the validity of the copula correction requires correct im-

plementation of the method and demands boundary conditions and data requirements to be met in its empirical application.

This study contributes to the marketing research field in three areas. One, we provide a review for how the copula procedure has been used in marketing to correct for endogeneity, across substantive areas, and how it has been applied (and misapplied). Two, we build on recent advances to provide an updated best practices "cookbook" for both managers and academics to follow in implementing the copula procedure. Three, we speak to implementation variations (such as including an intercept and higher-order effects of moderation), showing theoretically and with real-world data best practices for copula correction usage.

We demonstrate that existing variations in the implementation of copula correction have substantial impacts on its performance. Our discussions on the methodological aspects of the copula method informs optimal and theoretically sound implementation for copula correction. We present a theoretically sound way of constructing copula transformation that avoids potential finite sample bias problem and substantially improves the performance of copula correction. We show that excluding the copula terms for higher order endogenous regressors is optimal and substantially outperforms including these copula terms. A theoretical proof shows that copula terms for higher-order effects are not only unnecessary, but also substantially inflate estimation variability: the higher the correlations between the extra higher-order copula term and other regressors, the greater the estimation variance inflation. Our empirical evaluation shows consequential adverse effects of taking alternative suboptimal approaches: larger standard errors (by up to 5-times as shown in our simulation studies), substantial estimation bias (about 30% of parameter values), and significant loss of statistical power to detect moderating and nonlinear effects (e.g., a reduction of power from 80% to 10% in Figure 4). The empirical application of peanut butter sales further demonstrates this adverse bias: omitting the higher-order copula term yields model estimates closest to that of two-stage least squares using instrumental variables; including the copula interaction term produces the opposite sign for the coefficient estimate of the endogenous interaction

term, and greater estimation variability.

We also discuss the latest extensions that expand the applicability, flexibility and robustness of copula correction, highlighting endogeneity correction when the conditions and requirements of the prior copula correction approach are not met by the data at hand. For cases where the endogenous regressors have insufficient nonnormality, and the traditional method (Park and Gupta 2012) fails to work, we describe how a two-stage copula correction (2sCOPE) can still work by leveraging related and nonnormally distributed exogenous regressors. We demonstrate that applications of traditional methods need to check the uncorrelatedness between the copula control functions and exogenous regressors; if this condition fails, alternative copula correction methods need to be used.

We synthesize the above discussions into a flowchart with easy-to-follow checkpoints and boundary conditions. This guide is practical for researchers - in both academia and industry - to employ copula correction methods. In addition to making the copula code available, we illustrate its usage in two empirical examples for two different product categories.

While this study was intended to integrate the latest research on copula usage, and present a helpful guide for users to address endogeneity, future avenues of related research remain. One research direction is to relax the assumption of the Gaussian copula correlation structure. Excluding copula correction terms for higher-order endogenous regressors assumes a Gaussian copula model for the first-order endogenous regressors only. This should be robust to Gaussian copula correlation structure violations, at least relative to including the copula terms for higher-order endogenous regressors. Nonetheless, exploring methods to further relax the Gaussian copula correlation assumption will increase the robustness of copula correction. Another area may be a Bayesian approach, which is frequently used in marketing research. Extending copula correction to Bayesian inference can expand its applicability.

REFERENCES

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

Aiken, Leona S and Stephen G West (1991), *Multiple Regression: Testing and Interpreting Interactions* Newbury Park: Sage Publications.

Albers, Sönke, Murali K Mantrala, and Shrihari Sridhar (2010), "Personal selling elasticities: a meta-analysis," *Journal of Marketing Research*, 47 (5), 840–853.

Arora, Neeraj and Joel Huber (2001), "Improving Parameter Estimates and Model Prediction by Aggregate Customization in Choice Experiments.," *Journal of Consumer Research*, 28, 273–83.

Atefi, Yashar, Michael Ahearne, James G Maxham III, Todd D Donavan, and Brad D Carlson (2018), "Does Selective Sales Force Training Work?," *Journal of Marketing Research*, 55 (5), 722–737.

Becker, Jan-Michael, Dorian Proksch, and Christian M Ringle (2021), "Revisiting Gaussian Copulas to Handle Endogenous Regressors," *Journal of the Academy of Marketing Science*, pages 1–21.

Bijmolt, Tammo HA, Harald J Van Heerde, and Rik GM Pieters (2005), "New empirical generalizations on the determinants of price elasticity," *Journal of marketing research*, 42 (2), 141–156.

Blattberg, Robert C. and Scott A. Neslin (1990), *Sales Promotion - Concepts, Methods, and Strategies* Englewood Cliffs, NJ: Prentice-Hall.

Blauw, Sanne and Philip Hans Franses (2016), "Off the Hook: Measuring the Impact of Mobile Telephone Use on Economic Development of Households in Uganda using Copulas," *Journal of Development Studies*, 52(3), 315–330.

Bronnenberg, Bart J., Michael W. Kruger, and Carl F. Mela (2008), "Database paper - The IRI marketing data set," *Marketing Science*, 27(4), 745–748.

Burchett, Molly R, Brian Murtha, and Ajay K Kohli (2023), "Secondary Selling: Beyond the Salesperson–Customer Dyad," *Journal of Marketing*, page 00222429221138302.

Burmester, Alexa B, Jan U Becker, Harald J van Heerde, and Michel Clement (2015), "The Impact of Pre-and Post-launch Publicity and Advertising on New Product Sales," *International Journal of Research in Marketing*, 32 (4), 408–417.

Chintagunta, 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, 422–437.

Danaher, Peter J (2023), "Optimal microtargeting of advertising," *Journal of Marketing Research*, 60 (3), 564–584.

Danaher, Peter J. and Michael Smith (2011), "Modeling Multivariate Distributions Using Copulas: Applications in Marketing," *Marketing Science*, 30, 4–21.

Datta, Hannes, Harald J van Heerde, Marnik G Dekimpe, and Jan-Benedict EM Steenkamp (2022), "Cross-national differences in market response: line-length, price, and distribution elasticities in 14 Indo-Pacific Rim economies," *Journal of Marketing Research*, 59 (2), 251–270.

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.

Echambadi, Raj and James D Hess (2007), "Mean-Centering Does Not Alleviate Collinearity Problems in Moderated Multiple Regression Models," *Marketing Science*, 26(3), 438–445.

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*.

Fossen, Beth L and Alexander Bleier (2021), "Online program engagement and audience size during television ads," *Journal of the Academy of Marketing Science*, 49, 743–761.

Gielens, Katrijn, Inge Geyskens, Barbara Deleersnyder, and Max Nohe (2018), "The New Regulator in Town: The Effect of Walmart's Sustainability Mandate on Supplier Shareholder Value," *Journal of Marketing*, 82, 124–141.

Gijsbrechts, Els, Katia Campo, and Mark Vroegrijk (2018), "Save or (over-) spend? The impact of hard-discounter shopping on consumers' grocery outlay," *International Journal of Research in Marketing*, 35 (2), 270–288.

Gui, Raluca, Markus Meierer, Patrik Schilter, and René Algesheimer (2023), "REndo: Internal Instrumental Variables to Address Endogeneity," *Journal of Statistical Software*, 107, 1–43.

Guitart, Ivan A, Jorge Gonzalez, and Stefan Stremersch (2018), "Advertising Non-premium Products as if They Were Premium: The Impact of Advertising Up on Advertising Elasticity and Brand Equity," *International Journal of Research in Marketing*, 35 (3), 471–489.

Guitart, Ivan A, Guillaume Hervet, and Sarah Gelper (2020), "Competitive advertising strategies for programmatic television," *Journal of the Academy of Marketing Science*, 48, 753–775.

Guitart, Ivan A and Stefan Stremersch (2021), "The Impact of Informational and Emotional Television Ad Content on Online Search and Sales," *Journal of Marketing Research*, 58 (2), 299–320.

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, https://doi.org/10.1177/00222437211070820*.

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, 257–277.

Homburg, Christian, Arnd Vomberg, and Stephan Muehlhaeuser (2020), "Design and Governance of Multichannel Sales Systems: Financial Performance Consequences in Business-to-business Markets," *Journal of Marketing Research*, 57 (6), 1113–1134.

Joe, H. (2015), *Dependence Modeling with Copulas* Boca Raton, FL: CRC Press.

Keller, Wiebke IY, Barbara Deleersnyder, and Karen Gedenk (2019), "Price Promotions and Popular Events," *Journal of Marketing*, 83 (1), 73–88.

Kopalle, Praveen K. and Donald R. Lehmann (2006), "Setting Quality Expectations When Entering a Market: What Should the Promise Be?," *Marketing Science*, 25, 8–24.

Krämer, Martin, Christina Desernot, Sascha Alavi, Christian Schmitz, Felix Brüggemann, and Jan Wieseke (2022), "The Role of Salespeople in Industrial Servitization: How to Manage Diminishing Profit Returns From Salespeople's Increasing Industrial Service Shares," *International Journal of Research in Marketing*.

Lamey, Lien, Barbara Deleersnyder, Jan-Benedict EM Steenkamp, and Marnik G Dekimpe (2018), "New Product Success in the Consumer Packaged Goods Industry: A Shopper Marketing Approach," *International Journal of Research in Marketing*, 35 (3), 432–452.

Lenz, Isabell, Hauke A Wetzel, and Maik Hammerschmidt (2017), "Can Doing Good Lead to Doing Poorly? Firm Value Implications of CSR in the Face of CSI," *Journal of the Academy of Marketing Science*, 45 (5), 677–697.

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*, 65, 1201–1214.

Liu, Huan, Lara Lobschat, Peter C Verhoef, and Hong Zhao (2021), "The Effect of Permanent Product Discounts and Order Coupons on Purchase Incidence, Purchase Quantity, and Spending," *Journal of Retailing*, 97 (3), 377–393.

Luan, Y Jackie and K Sudhir (2010), "Forecasting Marketing-mix Responsiveness for New Products," *Journal of Marketing Research*, 47 (3), 444–457.

Ludwig, Stephan, Dennis Herhausen, Dhruv Grewal, Liliana Bove, Sabine Benoit, Ko De Ruyter, and Peter Urwin (2022), "Communication in the gig economy: Buying and selling in online freelance marketplaces," *Journal of Marketing*, 86 (4), 141–161.

Magnotta, Sarah, Brian Murtha, and Goutam Challagalla (2020), "The Joint and Multilevel Effects of Training and Incentives From Upstream Manufacturers on Downstream Salespeople's Efforts," *Journal of Marketing Research*, 57 (4), 695–716.

Manchanda, Puneet, Peter E Rossi, and Pradeep K Chintagunta (2004), "Response Modeling with Nonrandom Marketing-mix Variables," *Journal of Marketing Research*, 41 (4), 467–478.

Mathys, Juliane, Alexa B Burmester, and Michel Clement (2016), "What drives the market popularity of celebrities? A longitudinal analysis of consumer interest in film stars," *International Journal of Research in Marketing*, 33 (2), 428–448.

Papies, Dominik, Peter Ebbes, and Harald J van Heerde "Addressing Endogeneity in Marketing Models," T. Bijmolt P. Leeflang, J. Wieringa and K. Pauwels, editors, "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, 567–586.

Petrin, A and K Train (2010), "A control function approach to endogeneity in consumer choice models," *Journal of Marketing Research*, 47, 3–13.

Qian, Yi (2007), "Do National Patent Laws Stimulate Domestic Innovation in a Global Patenting Environment? A Cross-Country Analysis of Pharmaceutical Patent Protection, 1978-2002," *The Review of Economics and Statistics*, 89, 436–453.

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.

Rigobon, Roberto (2003), "Identification Through Heteroskedasticity," *Review of Economics and Statistics*, 85, 777–792.

Royston, J. Patrick (1982), "Algorithm AS 177: Expected normal order statistics (exact and approximate)," *Journal of the Royal Statistical Society. Series C (Applied statistics)*, 31 (2), 161–165.

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.

Schweidel, David A and George Knox (2013), "Incorporating direct marketing activity into latent attrition models," *Marketing Science*, 32 (3), 471–487.

Sethuraman, Raj, Gerard J Tellis, and Richard A Briesch (2011), "How well does advertising work? Generalizations from meta-analysis of brand advertising elasticities," *Journal of Marketing Research*, 48 (3), 457–471.

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, 42–60.

Tran, Kien C. and Mike G. Tsionas (2021), "Efficient Semiparametric Copula Estimation of Regression Models with Endogeneity," *Econometric Reviews*, 41(5), 1–28.

Villas-Boas, J. Miguel and Russell S. Winer (1999), "Endogeneity in Brand Choice Models," *Management Science*, 45, 1324–1338.

Vomberg, Arnd, Christian Homburg, and Olivia Gwinner (2020), "Tolerating and Managing Failure: An Organizational Perspective on Customer Reacquisition Management," *Journal of Marketing*, 84 (5), 117–136.

Wetzel, Hauke A, Stefan Hattula, Maik Hammerschmidt, and Harald J van Heerde (2018), "Building and Leveraging Sports Brands: Evidence From 50 Years of German Professional Soccer," *Journal of the Academy of Marketing Science*, 46 (4), 591–611.

Wooldridge, Jeffrey M (2010), *Econometric Analysis of Cross Section and Panel Data* Cambridge, MA: MIT Press.

Yang, Fan, Yi Qian, and Hui Xie (2022), "Addressing Endogeneity Using a Two-stage Copula Generated Regressor Approach," *NBER Working Paper, https://www.nber.org/papers/w29708.*

Yang, Sha, Yuxin Chen, and Greg Allenby (2003), "Bayesian Analysis of Simultaneous Demand and Supply," *Quantitative Marketing and Economics*, 1, 251–275.

Yoon, Hyungseok David, Namil Kim, Bernard Buisson, and Fred Phillips (2018), "A Cross-national Study of Knowledge, Government Intervention, and Innovative Nascent Entrepreneurship," *Journal of Business Research*, 84, 243–252.

A Practical Guide to Endogeneity Correction Using Copulas

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: SUBSTANTIVE AREAS IN MARKETING WITH APPLICATIONS OF COPULA CORRECTION

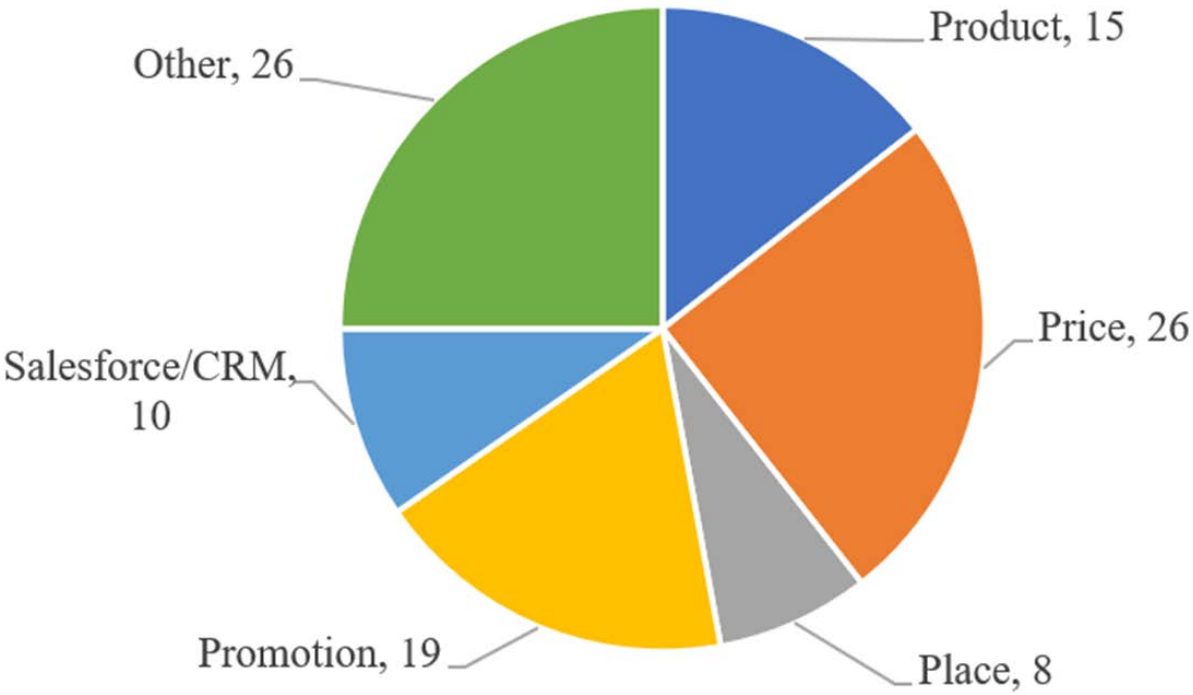

Figure W1: Number of papers that use copula endogeneity correction by substantive areas.

List of Publications in Table 1 and Figure W1

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.

Aydinli, Aylin, Lien Lamey, Kobe Millet, Anne ter Braak, and Maya Vuegen (2021), "How do customers alter their basket composition when they perceive the retail store to be crowded? An empirical study," *Journal of Retailing*, 97 (2), 207–216.

Bachmann, Patrick, Markus Meierer, and Jeffrey Näf (2021), "The role of time-varying contextual factors in latent attrition models for customer base analysis," *Marketing Science*, 40 (4), 783–809.

Bhattacharya, Abhi, Neil A Morgan, and Lopo L Rego (2022), "Examining why and when market share drives firm profit," *Journal of Marketing*, 86 (4), 73–94.

Bombaij, Nick JF and Marnik G Dekimpe (2020), "When do loyalty programs work? The moderating role of design, retailer-strategy, and country characteristics," *International Journal of Research in Marketing*, 37 (1), 175–195.

Borah, Abhishek, S Cem Bahadir, Anatoli Colicev, and Gerard J Tellis (2022), "It pays to pay attention: How firm's and competitor's marketing levers affect investor attention and firm value," *International Journal of Research in Marketing*, 39 (1), 227–246.

Bornemann, Torsten, Cornelia Hattula, and Stefan Hattula (2020), "Successive product generations: Financial implications of industry release rhythm alignment," *Journal of the Academy of Marketing Science*, 48, 1174–1191.

Burchett, Molly R, Brian Murtha, and Ajay K Kohli (2023), "Secondary Selling: Beyond the Salesperson–Customer Dyad," *Journal of Marketing*, page 00222429221138302.

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

Campo, Katia, Lien Lamey, Els Breugelmans, and Kristina Melis (2021), "Going online for groceries: Drivers of category-level share of wallet expansion," *Journal of Retailing*, 97 (2), 154–172.

Cao, Zixia (2022), "Brand equity, warranty costs, and firm value," *International Journal of Research in Marketing*, 39 (4), 1166–1185.

Cao, Zixia, Reo Song, Alina Sorescu, and Ansley Chua (2023), "Innovation Potential, Insider Sales, and IPO Performance: How Firms Can Mitigate the Negative Effect of Insider Selling," *Journal of Marketing*, 87 (4), 550–574.

Carson, Stephen J and Mrinal Ghosh (2019), "An integrated power and efficiency model of contractual channel governance: Theory and empirical evidence," *Journal of Marketing*, 83 (4), 101–120.

Cron, William L, Sascha Alavi, Johannes Habel, Jan Wieseke, and Hanaa Ryari (2021), "No conversion, no conversation: consequences of retail salespeople disengaging from unpromising prospects," *Journal of the Academy of Marketing Science*, 49, 502–520.

Dall?Olio, Filippo and Demetrios Vakratsas (2023), "The impact of advertising creative strategy on advertising elasticity," *Journal of Marketing*, 87 (1), 26–44.

Danaher, Peter J (2023), "Optimal microtargeting of advertising," *Journal of Marketing Research*, 60 (3), 564–584.

Datta, Hannes, Kusum L Ailawadi, and Harald J Van Heerde (2017), "How well does consumer-based brand equity align with sales-based brand equity and marketing-mix response?," *Journal of Marketing*, 81 (3), 1–20.

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.

Datta, Hannes, Harald J van Heerde, Marnik G Dekimpe, and Jan-Benedict EM Steenkamp (2022), "Cross-national differences in market response: line-length, price, and distribution elasticities in 14 Indo-Pacific Rim economies," *Journal of Marketing Research*, 59 (2), 251–270.

de Jong, Ad, Nicolas A Zacharias, and Edwin J Nijssen (2021), "How young companies can effectively manage their slack resources over time to ensure sales growth: the contingent role of value-based selling," *Journal of the Academy of Marketing Science*, 49, 304–326.

Dhaoui, Chedia and Cynthia M Webster (2021), "Brand and consumer engagement behaviors on Facebook brand pages: Let's have a (positive) conversation," *International Journal of Research in Marketing*, 38 (1), 155–175.

Fossen, Beth L and Alexander Bleier (2021), "Online program engagement and audience size during television ads," *Journal of the Academy of Marketing Science*, 49, 743–761.

Garrido-Morgado, Álvaro, Óscar González-Benito, Mercedes Martos-Partal, and Katia Campo (2021), "Which products are more responsive to in-store displays: utilitarian or hedonic?," *Journal of Retailing*, 97 (3), 477–491.

Gielens, Katrijn, Marnik G Dekimpe, Anirban Mukherjee, and Kapil Tuli (2023), "The future of private-label markets: A global convergence approach," *International Journal of Research in Marketing*, 40 (1), 248–267.

Gielens, Katrijn, Inge Geyskens, Barbara Deleersnyder, and Max Nohe (2018), "The new regulator in town: The effect of Walmart's sustainability mandate on supplier shareholder value," *Journal of Marketing*, 82 (2), 124–141.

Gijsbrechts, Els, Katia Campo, and Mark Vroegrijk (2018), "Save or (over-) spend? The impact of hard-discounter shopping on consumers' grocery outlay," *International Journal of Research in Marketing*, 35 (2), 270–288.

Glady, Nicolas, Aurélie Lemmens, and Christophe Croux (2015), "Unveiling the relationship between the transaction timing, spending and dropout behavior of customers," *International Journal of Research in Marketing*, 32 (1), 78–93.

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

Guitart, Ivan A, Guillaume Hervet, and Sarah Gelper (2020), "Competitive advertising strategies for programmatic television," *Journal of the Academy of Marketing Science*, 48, 753–775.

Guitart, Ivan A and Stefan Stremersch (2021), "The impact of informational and emotional television ad content on online search and sales," *Journal of Marketing Research*, 58 (2), 299–320.

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.

Homburg, Christian, Arnd Vomberg, and Stephan Muehlhaeuser (2020), "Design and governance of multichannel sales systems: Financial performance consequences in business-to-business markets," *Journal of Marketing Research*, 57 (6), 1113–1134.

Hoskins, Jake, Shyam Gopinath, J Cameron Verhaal, and Elham Yazdani (2021), "The influence of the online community, professional critics, and location similarity on review ratings for niche and mainstream brands," *Journal of the Academy of Marketing Science*, 49, 1065–1087.

Janani, Saeed, Ranjit M Christopher, Atanas Nik Nikolov, and Michael A Wiles (2022), "Marketing experience of CEOs and corporate social performance," *Journal of the Academy of Marketing Science*, pages 1–22.

Keller, Wiebke IY, Barbara Deleersnyder, and Karen Gedenk (2019), "Price promotions and popular events," *Journal of Marketing*, 83 (1), 73–88.

Kidwell, Blair, Jonathan Hasford, Broderick Turner, David M Hardesty, and Alex Ricardo Zablah (2021), "Emotional calibration and salesperson performance," *Journal of Marketing*, 85 (6), 141–161.

Krämer, Martin, Christina Desernot, Sascha Alavi, Christian Schmitz, Felix Brüggemann, and Jan Wieseke (2022), "The role of salespeople in industrial servitization: how to manage diminishing profit returns from salespeople?s increasing industrial service shares," *International Journal of Research in Marketing*, 39 (4), 1235–1252.

Lamey, Lien, Els Breugelmans, Maya Vuegen, and Anne ter Braak (2021), "Retail service innovations and their impact on retailer shareholder value: Evidence from an event study," *Journal of the Academy of Marketing Science*, 49, 811–833.

Lamey, Lien, Barbara Deleersnyder, Jan-Benedict EM Steenkamp, and Marnik G Dekimpe (2018), "New product success in the consumer packaged goods industry: A shopper marketing approach," *International Journal of Research in Marketing*, 35 (3), 432–452.

Lenz, Isabell, Hauke A Wetzel, and Maik Hammerschmidt (2017), "Can doing good lead to doing poorly? Firm value implications of CSR in the face of CSI," *Journal of the Academy of Marketing Science*, 45, 677–697.

Lim, Leon Gim, Kapil R Tuli, and Marnik G Dekimpe (2018), "Investors' evaluations of price-increase preannouncements," *International Journal of Research in Marketing*, 35 (3), 359–377.

Liu, Huan, Lara Lobschat, Peter C Verhoef, and Hong Zhao (2021), "The effect of permanent product discounts and order coupons on purchase incidence, purchase quantity, and spending," *Journal of Retailing*, 97 (3), 377–393.

Ludwig, Stephan, Dennis Herhausen, Dhruv Grewal, Liliana Bove, Sabine Benoit, Ko De Ruyter, and Peter Urwin (2022), "Communication in the gig economy: Buying and selling in online freelance marketplaces," *Journal of Marketing*, 86 (4), 141–161.

Maesen, Stijn and Lien Lamey (2023), "The impact of organic specialist store entry on category performance at incumbent stores," *Journal of Marketing*, 87 (1), 97–113.

Maesen, Stijn, Lien Lamey, Anne ter Braak, and Léon Jansen (2022), "Going healthy: how product characteristics influence the sales impact of front-of-pack health symbols," *Journal of the Academy of Marketing Science*, 50 (1), 108–130.

Magnotta, Sarah, Brian Murtha, and Goutam Challagalla (2020), "The joint and multilevel effects of training and incentives from upstream manufacturers on downstream salespeople?s efforts," *Journal of Marketing Research*, 57 (4), 695–716.

Mathys, Juliane, Alexa B Burmester, and Michel Clement (2016), "What drives the market popularity of celebrities? A longitudinal analysis of consumer interest in film stars," *International Journal of Research in Marketing*, 33 (2), 428–448.

Moon, Sungkyun, Kapil R Tuli, and Anirban Mukherjee (2023), "Does disclosure of advertising spending help investors and analysts?," *Journal of Marketing*, 87 (3), 359–382.

Nahm, Irene Y, Michael J Ahearne, Nick Lee, and Seshadri Tirunillai (2022), "Managing Positive and Negative Trends in Sales Call Outcomes: The Role of Momentum," *Journal of Marketing Research*, 59 (6), 1120–1140.

Nath, Pravin, Ahmet H Kirca, Saejoon Kim, and Trina Larsen Andras (2019), "The effects of retail banner standardization on the performance of global retailers," *Journal of Retailing*, 95 (3), 30–46.

Rajavi, Koushyar, Tarun Kushwaha, and Jan-Benedict EM Steenkamp (2023), "Brand Equity in Good and Bad Times: What Distinguishes Winners from Losers in Consumer Packaged Goods Industries?," *Journal of Marketing*, 87 (3), 472–489.

Sawant, Rajeev J, Mahima Hada, and Simon J Blanchard (2021), "Contractual discrimination in franchise relationships," *Journal of Retailing*, 97 (3), 405–423.

Scholdra, Thomas P, Julian RK Wichmann, Maik Eisenbeiss, and Werner J Reinartz (2022), "Households under economic change: How micro-and macroeconomic conditions shape grocery shopping behavior," *Journal of Marketing*, 86 (4), 95–117.

Schulz, Petra, Edlira Shehu, and Michel Clement (2019), "When consumers can return digital products: Influence of firm-and consumer-induced communication on the returns and profitability of news articles," *International Journal of Research in Marketing*, 36 (3), 454–470.

Schweidel, David A and George Knox (2013), "Incorporating direct marketing activity into latent attrition models," *Marketing Science*, 32 (3), 471–487.

Shehu, Edlira, Dominik Papies, and Scott A Neslin (2020), "Free shipping promotions and product returns," *Journal of Marketing Research*, 57 (4), 640–658.

ter Braak, Anne and Barbara Deleersnyder (2018), "Innovation cloning: The introduction and performance of private label innovation copycats," *Journal of Retailing*, 94 (3), 312–327.

Umashankar, Nita, Kihyun Hannah Kim, and Thomas Reutterer (2023), "Understanding Customer Participation Dynamics: The Case of the Subscription Box," *Journal of Marketing*, 87 (5), 719–735.

Van Ewijk, Bernadette J, Els Gijsbrechts, and Jan-Benedict EM Steenkamp (2022a), "The dark side of innovation: How new SKUs affect brand choice in the presence of consumer uncertainty and learning," *International Journal of Research in Marketing*, 39 (4), 967–987.

Van Ewijk, Bernadette J, Els Gijsbrechts, and Jan-Benedict EM Steenkamp (2022b), "What drives brands? price response metrics? An empirical examination of the Chinese packaged goods industry," *International Journal of Research in Marketing*, 39 (1), 288–312.

Van Ewijk, Bernadette J, Astrid Stubbe, Els Gijsbrechts, and Marnik G Dekimpe (2021), "Online display advertising for CPG brands:(When) does it work?," *International Journal of Research in Marketing*, 38 (2), 271–289.

Vieira, Valter Afonso, Marcos Inácio Severo de Almeida, Raj Agnihotri, Nôga Simões De Arruda Corrêa da Silva, and S Arunachalam (2019), "In pursuit of an effective B2B digital marketing strategy in an emerging market," *Journal of the Academy of Marketing Science*, 47, 1085–1108.

Vomberg, Arnd, Christian Homburg, and Olivia Gwinner (2020), "Tolerating and managing failure: An organizational perspective on customer reacquisition management," *Journal of Marketing*, 84 (5), 117–136.

Wetzel, Hauke A, Stefan Hattula, Maik Hammerschmidt, and Harald J van Heerde (2018), "Building and leveraging sports brands: evidence from 50 years of German professional soccer," *Journal of the Academy of Marketing Science*, 46, 591–611.

Widdecke, Kai A, Wiebke IY Keller, Karen Gedenk, and Barbara Deleersnyder (2023), "Drivers of the synergy between price cuts and store flyer advertising at supermarkets and discounters," *International Journal of Research in Marketing*, 40 (2), 455–474.

Zhang, Junzhou and Yuping Liu-Thompkins (2023), "Personalized email marketing in loyalty programs: The role of multidimensional construal levels," *Journal of the Academy of Marketing Science*, pages 1–21.

Zhang, Yufei, Clay M Voorhees, Chen Lin, Jeongwen Chiang, G Tomas M Hult, and Roger J Calantone (2022), "Information search and product returns across mobile and traditional online channels," *Journal of Retailing*, 98 (2), 260–276.

Zhao, Yanhui, Yufei Zhang, Joyce Wang, Wyatt A Schrock, and Roger J Calantone (2020), "Brand relevance and the effects of product proliferation across product categories," *Journal of the Academy of Marketing Science*, 48, 1192–1210.

WEB APPENDIX B: A SIMULATION STUDY FOR COPULA CORRECTION IN MODELS WITH INTERCEPT

In this study, we use the following data generating process (DGP) that is the same as specified in Equations 1-4 in Becker, Proksch, and Ringle (2021):

$$\begin{bmatrix} E_t^* \\ \\ P_t^* \end{bmatrix} = N\left(\begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.50 \\ \\ 0.50 & 1 \end{bmatrix}\right) \tag{W1}$$

$$E_t = \Phi^{-1}(\Phi(E_t^*)) \tag{W2}$$

$$P_t = \Phi(P_t^*) \tag{W3}$$

$$Y_t = -1P_t + E_t, \tag{W4}$$

where Y_t, P_t, and E_t represent the dependent variable, endogenous regressor, and the error term, respectively. The DGP specifies a linear model with the endogenous regressor P following a uniform distribution, and a correlation coefficient of 0.50 between P_t^* and the error term E_t. The simulation study varies sample size N from 100 to 60,000 (100, 200, 400, 600, 800, 1,000, 2,000, 4,000, 6,000, 8,000, 10,000, 20,000, 40,000, 60,000). For each sample size, we generate 1,000 datasets from the above DGP.

For each generated data set, we apply OLS, the Park and Gupta (P&G) method using the algorithm in Equation 7 to obtain generated regressor, and the P&G method using the algorithm in Equation 6 to obtain the generated regressor in estimating the structural model. While the intercept term $\mu = 0$ in the DGP, the estimation does not assume this a-priori but instead estimates the intercept parameter jointly with other model parameters. The difference between the average of the estimates across 1,000 simulated datasets and its true value is the bias of an estimator, which is plotted in Figure 3 for α.

WEB APPENDIX C: PROOF OF THEOREM 1

Theorem 1. ***Optimality of excluding higher-order copula terms.*** *Let* $(\widehat{\theta}_k^{Main}), k = 1, \cdots, K$, *denote the structural model parameter estimates when only the copula terms for the main endogenous effects are included to correct for endogeneity, and* $(\widehat{\theta}_k^{All}), k = 1, \cdots, K$, *denote the corresponding estimates when copula terms for both the main effects and higher-order endogenous regressors are included. This yields:*

$$\mathit{Var}(\widehat{\theta}_k^{All}) \geq \mathit{Var}(\widehat{\theta}_k^{Main}) \quad \textit{for } k = 1, \cdots, K.$$

Thus, $\widehat{\theta}_k^{Main}$ *yields optimal copula estimation of structural model parameters with less variance and mean squared errors than* $\widehat{\theta}_k^{All}$*, for all* k*.*

Proof: Consider the OLS regression of the model when only the copula main terms are included to correct for endogeneity:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \;\; V(\boldsymbol{\epsilon}) = \sigma_c^2 \mathbf{I}_n, \tag{W5}$$

where $\mathbf{X}$ includes the intercept, the regressors in the structural model, and $\mathbf{C}_{main}$ (the copula generated regressors for the main effects); $\boldsymbol{\theta}$ collects all the coefficients of these regressors. Math symbols in bold represent matrices and vectors. The variance of the estimates using copula terms for main effects only is:

$$V(\widehat{\boldsymbol{\theta}}^{Main}) = \sigma_c^2 (\mathbf{X}'\mathbf{X})^{-1}. \tag{W6}$$

Then after introducing additional copula terms $\mathbf{C}$ for higher-order terms into the model in

Equation (W5), we have:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{C}\boldsymbol{\phi} + \boldsymbol{\epsilon_1}, \;\; V(\boldsymbol{\epsilon_1}) = \sigma_c^{\prime 2}\mathbf{I}_n, \tag{W7}$$

According to linear regression theory, the new estimates after entering the copula higher-order terms $\mathbf{C}$ in the model become:

$$\widehat{\boldsymbol{\theta}}^{All} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{Y} - \mathbf{C}\widehat{\phi}), \qquad \widehat{\phi} = (\mathbf{C}'\mathbf{R}\mathbf{C})^{-1}\mathbf{C}'\mathbf{R}\mathbf{Y}, \tag{W8}$$

$$V(\widehat{\boldsymbol{\theta}}^{All}) = \sigma_c^{\prime 2}\big[(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} + \mathbf{M}(\mathbf{C}'\mathbf{R}\mathbf{C})^{-\mathbf{1}}\mathbf{M}'\big], \tag{W9}$$

where $\mathbf{M} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{C}$, $\mathbf{R} = \mathbf{I_n} - \mathbf{P}$, and $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$. Note that $\mathbf{P}$ is the projection matrix representing the orthogonal projection that maps the responses to the fitted values, and $\mathbf{R} = \mathbf{I_n} - \mathbf{P}$ represents the orthogonal projection that maps the responses to the residuals. Given that the newly added higher-order copula terms in $\mathbf{C}$ are highly correlated with the higher-order terms in the structural model (as well as other copula terms already included in the model), the extra variability in $\mathbf{Y}$ explained by adding $\mathbf{C}$ is small. Thus, $\sigma_c^{\prime 2} \approx \sigma_c^2$ and:

$$V(\widehat{\boldsymbol{\theta}})^{All} - V(\widehat{\boldsymbol{\theta}})^{Main} \approx \sigma_c^2\big[(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} + \mathbf{M}(\mathbf{C}'\mathbf{R}\mathbf{C})^{-\mathbf{1}}\mathbf{M}' - (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\big] \tag{W10}$$

$$= \sigma_c^2\big[\mathbf{M}(\mathbf{C}'\mathbf{R}\mathbf{C})^{-\mathbf{1}}\mathbf{M}'\big]. \tag{W11}$$

Since the matrix $\mathbf{M}(\mathbf{C}'\mathbf{R}\mathbf{C})^{-\mathbf{1}}\mathbf{M}'$ is positive semi-definite, all the diagonal elements are greater than or equal to zero. For each of the K structural model parameters:

$$\mathtt{Var}(\widehat{\theta}_k^{All}) \geq \mathtt{Var}(\widehat{\theta}_k^{Main}) \quad \text{for } k = 1, \cdots, K. \tag{W12}$$

The magnitude of variance inflation is inversely related to $\mathbf{C}'\mathbf{R}\mathbf{C}$, which represents the

matrix of sum of squared residuals, obtained from regressing $\mathbf{C}$ on $\mathbf{X}$. Thus, the higher the correlation between the extra higher-order term $\mathbf{C}$ and existing regressors in $\mathbf{X}$, the smaller the sum of squares, which leads to greater variance inflation of $\texttt{Var}(\widehat{\theta}_k^{All})$. Q.E.D.

WEB APPENDIX D: A SIMULATION STUDY FOR MODELS WITH AN INTERACTION TERM BETWEEN TWO ENDOGENOUS REGRESSORS

Data were simulated from the following structural regression model with an interaction between two endogenous regressors, P_1 and P_2:

$$\begin{aligned} Y &= \alpha_0 + \alpha_1 P_1 + \alpha_2 P_2 + \alpha_3 P_1 * P2 + E \\ \begin{pmatrix} E^* \\ P_1^* \\ P_2^* \end{pmatrix} &= N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho_{E1} & \rho_{E2} \\ \rho_{E1} & 1 & \rho_{12} \\ \rho_{E2} & \rho_{12} & 1 \end{bmatrix}\right) \\ E = H_E^{-1}(\Phi(E^*)) &= \Phi^{-1}(\Phi(E^*)), \qquad P_1 = H_{P_1}^{-1}(\Phi(P_1^*)), \quad P_2 = H_{P_2}^{-1}(\Phi(P_2^*)). \end{aligned} \tag{W13}$$

In this simulation, we set $H_{P_1}(\cdot)$ as the CDF of the uniform distribution on $[4, 6]$, $H_{P_2}(\cdot)$ as the CDF of the truncated standard normal with a lower bound of 0, and parameters $\alpha_0 = 0, \alpha_1 = 1, \alpha_2 = -1, \alpha_3 = 1, \rho_{E1} = \rho_{E2} = 0.5, \rho_{12} = -0.5$. The OLS version regresses Y on P_1, P_2 and $P_1 * P_2$ without any correction for the endogeneity of these regressors. Copula-Main adds two copula correction terms, C_{P_1} and C_{P_2}, to control for the endogeneity of these three regressors, where:

$$C_{P_1} = \Phi^{-1}(\hat{H}_{P_1}(P_1)), \qquad C_{P_2} = \Phi^{-1}(\hat{H}_{P_2}(P_2)). \tag{W14}$$

In addition to C_{P_1} and C_{P_2}, Copula-All adds the copula correction term $C_{P_1 * P_2}$, where:

$$C_{P_1 * P_2} = \Phi^{-1}(\hat{H}_{P_1 * P_2}(P_1 * P_2)) \tag{W15}$$

and $\hat{H}_{P_1}$, $\hat{H}_{P_2}$ and $\hat{H}_{P_1 * P_2}$ denote the empirical marginal distribution functions of P_1, P_2 and $P_1 * P_2$ in the observed sample, respectively.

Bias and SEs of parameter estimates The simulation results appear in Table W1. As expected, OLS regression yields significant bias for all model parameters at all sample sizes. For example, even for a large sample size of N=5,000, the OLS regression without any correction terms yields large bias for the regression parameter estimates ($\widehat{\alpha}_1$: 2.281 [0.018]; $\widehat{\alpha}_2$: -1.549 [0.099]; $\widehat{\alpha}_3$: 1.432 [0.021]) and the error standard deviation ($\widehat{\sigma}$: 0.298 [0.006]). Copula-Main corrects for the endogenous bias ($\widehat{\alpha}_1$: 1.002 [0.058]; $\widehat{\alpha}_2$: -1.017 [0.080]; $\widehat{\alpha}_3$: 1.003 [0.015]), demonstrating that there is no need to additionally include the copula correction term, $C_{P_1*P_2}$. Furthermore, Copula-Main performs substantially better in both estimation bias and variability for all parameter estimates than Copula-All which includes $C_{P_1*P_2}$. In fact, Copula-All yields significantly biased parameter estimates, even at the large sample size of N=5,000 ($\widehat{\alpha}_0$: 0.202 [0.318]; $\widehat{\alpha}_2$: -0.713 [0.240]; $\widehat{\alpha}_3$: 0.929 [0.058]); bias decreases as sample size increases, but remains apparent even for a sample size as large as 50,000, as including the copula term for the interaction $P_1 * P_2$ causes significant estimation bias.

The same conclusion - that Copula-Main performs substantially better than Copula-All in terms of both estimation bias and variability for all parameter estimates - applies to all other sample sizes, except for the intercept parameter (α_0) at small sample size N=200. The exception likely results from both a small sample size and strong multicollinearity induced by the interaction term; however, the bias in the intercept estimate bears less practical implication, since the intercept parameter is often of less interest.

Copula-All also yields less precise estimates (larger standard errors) than Copula-Main; underlined standard errors in Table W1 highlight much larger SE for Copula-All versus Copula-Main. This imprecision includes an SE 3.00-times that for α_2 and 3.86-times that for α_3 compared to Copula-Main at a sample size of 5,000.

Overall Estimation Efficiency and Accuracy Regarding overall estimation efficiency, the D-error ratios for Copula-All to Copula-Main increase as sample size increases, from 1.26-times (N=200) to 1.41-times (N=500) to 2.82-times (N=5,000) to 4.64-times (N=50,000).

We also compute the ratio of mean squared error (MSE) of the structural estimate $\widehat{\alpha}_k$, comparing Copula-All to Copula-Main (where $\text{MSE}(\widehat{\alpha}_k) = \text{Bias}^2(\widehat{\alpha}_k)+\text{Var}(\widehat{\alpha}_k)$, measuring overall estimation accuracy). Notably, Copula-All increases MSEs for all model parameter estimates, with the harmful effects being largest for the interaction parameter estimate $\widehat{\alpha}_3$, whose MSE is more than 80-times that of Copula-Main when sample size N=50,000 (Figure W2).

Table W1: Results from Case I: Interaction of Endogenous Regressors.

N	Method	$\alpha_0(=0)$	$\alpha_1(=1)$	$\alpha_2(=-1)$	$\alpha_3(=1)$	$\sigma(=1)$	D-error
200	OLS	**-7.627**	**2.282**	**-1.546**	**1.433**	**0.294**	
		(0.464)	(0.093)	(0.501)	(0.106)	(0.031)	—
	Copula-Main	**-0.358**	1.046	**-1.187**	1.043	0.963	
		(1.363)	(0.271)	(0.417)	(0.079)	(0.121)	0.0293
	Copula-All	**-0.058**	1.012	**-0.794**	**0.930**	1.028	
		(1.364)	(0.270)	(0.468)	(0.107)	(0.134)	0.0368
500	OLS	**-7.624**	**2.281**	**-1.546**	**1.432**	**0.297**	
		(0.290)	(0.058)	(0.312)	(0.066)	(0.019)	—
	Copula-Main	**-0.119**	1.019	**-1.104**	1.024	0.99	
		(0.899)	(0.179)	(0.254)	(0.047)	(0.076)	0.0117
	Copula-All	**0.176**	0.974	**-0.702**	**0.923**	1.051	
		(0.902)	(0.178)	(0.331)	(0.077)	(0.086)	0.0165
5000	OLS	**-7.623**	**2.281**	**-1.549**	**1.432**	**0.298**	
		(0.092)	(0.018)	(0.099)	(0.021)	(0.006)	—
	Copula-Main	-0.012	1.002	-1.017	1.003	1.000	
		(0.291)	(0.058)	(0.080)	(0.015)	(0.024)	0.0011
	Copula-All	**0.202**	0.968	**-0.713**	**0.929**	1.044	
		(0.318)	(0.061)	(0.240)	(0.058)	(0.041)	0.0031
50000	OLS	**-7.621**	**2.281**	**-1.551**	**1.433**	**0.298**	
		(0.029)	(0.006)	(0.031)	(0.007)	(0.002)	—
	Copula-Main	0.001	1.000	-1.003	1.000	1.000	
		(0.092)	(0.018)	(0.025)	(0.005)	(0.008)	0.00011
	Copula-All	**0.064**	0.990	**-0.912**	0.978	1.013	
		(0.133)	(0.023)	(0.158)	(0.038)	(0.023)	0.00051

See the same note under Table 5.

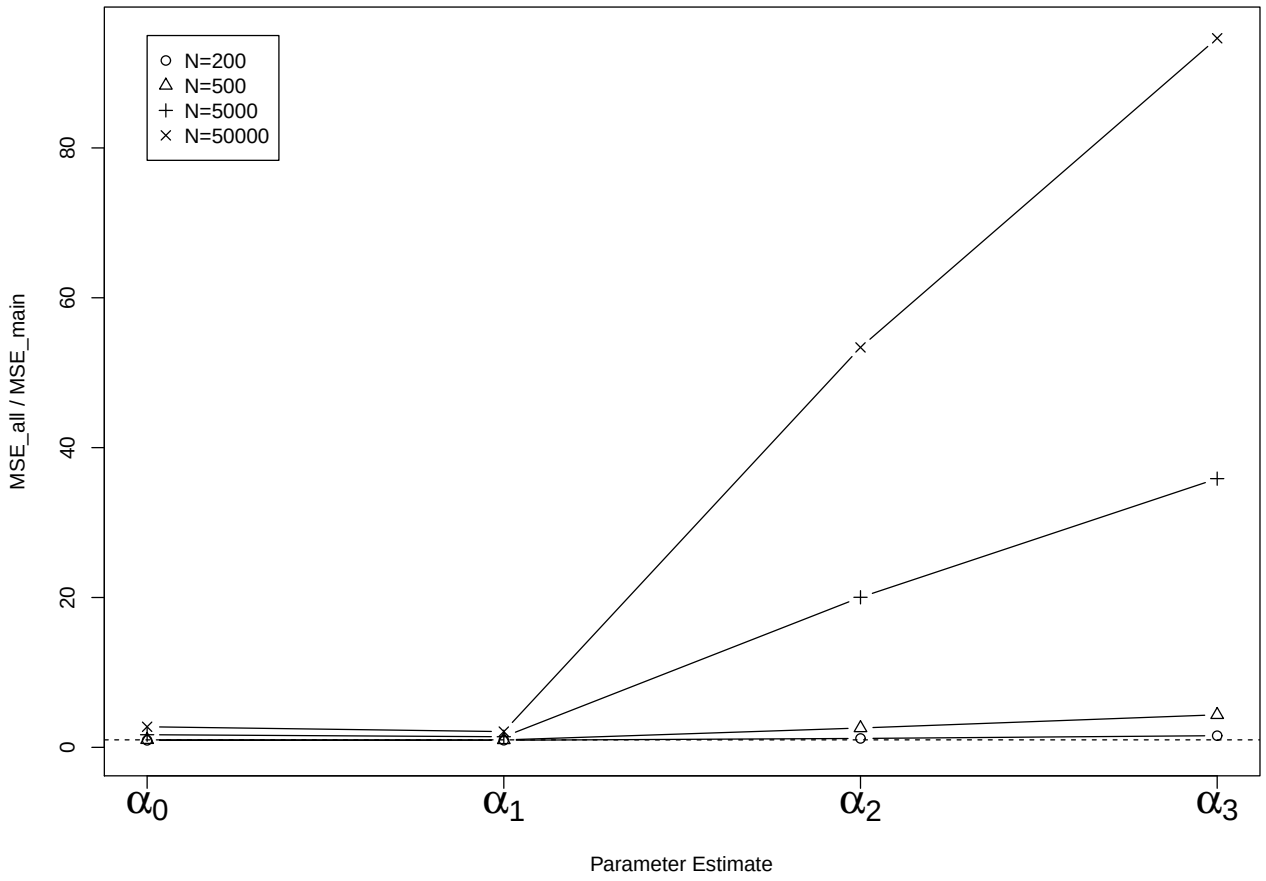

Figure W2: Ratio of mean squared errors of structural model estimates, with using the copula interaction term (Copula-All) to those without using the copula interaction term (Copula-Main).

WEB APPENDIX E: A SIMULATION STUDY FOR MODELS WITH AN INTERACTION TERM BETWEEN AN ENDOGENOUS REGRESSOR AND AN EXOGENOUS REGRESSOR

We simulated data from the following structural regression model with an interaction term between an exogenous regressor X and an endogenous regressor P:

$$\begin{aligned} Y &= \alpha_0 + \beta_1 X + \alpha_1 P + \alpha_2 X * P + E \\ \begin{pmatrix} P^* \\ E^* \end{pmatrix} &= N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \\ E = H_E^{-1}(\Phi(E^*)) &= \Phi^{-1}(\Phi(E^*)), \qquad P = H_P^{-1}(\Phi(P^*)), \end{aligned} \tag{W16}$$

where $H_P(\cdot)$ is the CDF of the truncated standard normal on $[0, \infty]$. We simulated X from a uniform distribution on $[4, 6]$, and set $\alpha_0 = 0, \beta_1 = 1, \alpha_1 = -1, \alpha_2 = 1$ and $\rho = 0.5$ with sample sizes of 200, 500, 5,000 and 50,000. For each sample size, we generated 5,000 repeated samples. For each generated sample, we then apply three estimation procedures: OLS, Copula-Main and Copula-All. The OLS regresses Y on P, X and $X * P$ without any correction for the endogeneity of P and $X * P$. Copula-Main adds one copula correction term, $C_P = \Phi^{-1}(\hat{H}_P(P))$, to control for endogeneity of P and $X * P$. In addition to C_P, Copula-All adds the copula correction term $C_{X*P} = \Phi^{-1}(\hat{H}_{X*P}(X * P))$. $\hat{H}_P(\cdot)$ and $\hat{H}_{X*P}(\cdot)$ denote the empirical marginal distribution functions of P and $X * P$ in the observed sample, respectively. Results over 5,000 simulated samples are summarized in Table W2.

Table W2: Results from Case II: Interaction between Endogenous and Exogenous Regressors

N	Method	$\alpha_0(=0)$	$\beta_1(=1)$	$\alpha_1(=-1)$	$\alpha_2(=1)$	$\sigma(=1)$	D-error
200	OLS	**-0.650**	1.003	**-0.194**	0.999	**0.876**	
		(0.913)	(0.181)	(0.929)	(0.185)	(0.044)	—
	Copula-Main	-0.032	1.002	-0.976	0.999	1.008	
		(0.952)	(0.178)	(0.993)	(0.182)	(0.127)	0.0406
	Copula-All	**0.073**	0.981	**-0.789**	0.962	1.02	
		(1.097)	(0.210)	(1.315)	(0.250)	(0.129)	0.0792
500	OLS	**-0.639**	1.000	**-0.199**	1.000	**0.876**	
		(0.573)	(0.114)	(0.581)	(0.115)	(0.028)	—
	Copula-Main	-0.002	1.000	-1.003	1.000	1.006	
		(0.594)	(0.111)	(0.620)	(0.113)	(0.082)	0.0156
	Copula-All	**0.103**	0.978	**-0.799**	0.961	1.013	
		(0.747)	(0.148)	(1.015)	(0.190)	(0.082)	0.0375
5000	OLS	**-0.643**	1.001	**-0.198**	0.999	**0.877**	
		(0.186)	(0.037)	(0.185)	(0.037)	(0.009)	—-
	Copula-Main	-0.003	1.001	-1.000	1.000	1.001	
		(0.192)	(0.036)	(0.195)	(0.036)	(0.025)	0.0016
	Copula-All	**0.075**	0.983	**-0.836**	0.969	1.003	
		(0.361)	(0.077)	(0.654)	(0.121)	(0.028)	0.0064
50000	OLS	**-0.637**	1.000	**-0.202**	1.000	**0.877**	
		(0.056)	(0.011)	(0.056)	(0.011)	(0.003)	—
	Copula-Main	0.000	1.000	-1.000	1.000	1.000	
		(0.059)	(0.011)	(0.060)	(0.011)	(0.008)	0.0002
	Copula-All	0.028	0.994	**-0.942**	0.989	1.000	
		(0.169)	(0.037)	(0.329)	(0.061)	(0.010)	0.0009

See the same note under Table 5.

WEB APPENDIX F: A SIMULATION STUDY FOR MODELS WITH A SQUARED TERM OF AN ENDOGENOUS REGRESSOR

Data were simulated from the following model (subscript t omitted for simplicity):

$$\begin{aligned} Y &= \alpha_0 + \alpha_1 P + \alpha_2 P^2 + E, \\ \begin{pmatrix} E^* \\ P^* \end{pmatrix} &= N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \\ E = H_E^{-1}(\Phi(E^*)) &= \Phi^{-1}(\Phi(E^*)), \qquad P = H_P^{-1}(\Phi(P^*)), \end{aligned} \tag{W17}$$

where $H_P(\cdot)$ is the CDF for the marginal distribution of P, $\alpha_0 = 0, \alpha_1 = -1, \alpha_2 = 1$ and $\rho = 0.7$. We set $H_P(\cdot)$ as the CDF of the truncated standard normal distribution on $[-0.5, 0.5]$. For each simulated data set, the following three estimation procedures were applied using OLS regression of Y on the following sets of regressors:

OLS:	P, P^2
Copula-Main:	P, P^2, C_P
Copula-All:	P, P^2, C_P, C_{P^2}

where $C_P = \Phi^{-1}(\hat{H}_P(P))$ and $C_{P^2} = \Phi^{-1}(\hat{H}_{P^2}(P^2))$ are the copula correction terms for endogenous regressors P and P^2, respectively; $\hat{H}_P$ and $\hat{H}_{P^2}$ denote the empirical marginal distribution functions of P and P^2 in the generated sample, respectively. Copula-Main indicates including copula correction terms for the main effect only, while Copula-All signifies including copula correction for all terms involving endogenous regressor P (i.e., higher-order terms).

Across simulations, sample sizes (N) of 200, 500, 5,000 and 50,000 are examined. For each sample size N, we generate 5,000 data sets as replicates to systematically evaluate average performance (estimation bias and variability) of different estimation methods. Averages and standard deviations (SD) of parameter estimates over these 5,000 data sets are computed for each method. The difference between the average of the estimates and its true value is the bias of one estimator; the SD of the parameter estimates over these 5,000 repeated samples is the standard error (SE) of the parameter estimate, capturing estimation variability.

Table W3 presents the simulation results. For each parameter, we report the average of the estimates and SE in the parenthesis computed using 5,000 generated data sets. As expected, OLS yields significant estimation bias at all values of N. For example, when N=200, the OLS regression yields large bias in the parameter estimates ($\widehat{\alpha}_1$: 1.413 [0.188]) and the error standard deviation ($\widehat{\sigma}$: 0.726 [0.037]) in the structural regression model. Copula-Main corrects for the endogenous bias ($\widehat{\alpha}_1$: -0.964 [1.049]; $\widehat{\sigma}$: 1.013 [0.202]), demonstrating that there is no need to additionally include C_{P^2}. Meanwhile, Copula-All yields substantial bias for the coefficient parameter of P^2 ($\widehat{\alpha}_2$: 0.771 [2.214]) because adding unnecessary generated regressor C_{P^2} leads to the finite sample bias problem. In contrast, Copula-Main eliminates the majority of the bias and performs much better in this small sample size with only small bias and the SE reduced by approximately 70% ($\widehat{\alpha}_2$: 0.922 [0.797]). In a large sample size (n=5,000), the finite sample bias in Copula-All is reduced. Yet, Copula-All continues to yield less precise estimates (i.e. larger standard errors) than Copula-Main.

Table W3: Results from Case III: Endogenous Squared Terms.

N	Method	$\alpha_0(=0)$	$\alpha_1(=-1)$	$\alpha_2(=1)$	$\sigma(=1)$	D-error
200	OLS	0.000	**1.413**	0.986	**0.726**	
		(0.078)	(0.188)	(0.742)	(0.037)	—
	Copula-Main	-0.001	-0.964	**0.922**	1.013	
		(0.099)	(1.049)	(0.797)	(0.202)	0.835
	Copula-All	0.009	-0.957	**0.771**	1.020	
		(0.190)	(1.057)	(2.214)	(0.203)	2.338
500	OLS	0.001	**1.410**	0.982	**0.728**	
		(0.048)	(0.118)	(0.472)	(0.024)	—
	Copula-Main	0.001	-0.978	0.951	1.005	
		(0.057)	(0.640)	(0.483)	(0.126)	0.309
	Copula-All	0.004	-0.974	**0.889**	1.008	
		(0.120)	(0.641)	(1.393)	(0.126)	0.891
5000	OLS	0.000	**1.413**	1.003	**0.728**	
		(0.015)	(0.036)	(0.146)	(0.007)	—
	Copula-Main	0.000	-1.000	0.994	1.001	
		(0.019)	(0.192)	(0.157)	(0.038)	0.030
	Copula-All	0.000	-1.000	0.997	1.001	
		(0.037)	(0.192)	(0.427)	(0.038)	0.082
50000	OLS	0.000	**1.415**	1.001	**0.728**	
		(0.005)	(0.012)	(0.047)	(0.002)	—
	Copula-Main	0.000	-1.004	1.000	1.001	
		(0.006)	(0.060)	(0.050)	(0.012)	0.003
	Copula-All	0.000	-1.004	0.999	1.001	
		(0.012)	(0.060)	(0.137)	(0.012)	0.008

Table presents the averages of the estimates and standard errors in the parenthesis over the repeated samples. Bold numbers highlight the estimates with bias of at least 0.05. Underlined numbers highlight the cases where the standard errors of the estimates from Copula-All are inflated by at least 50% compared with the corresponding ones from Copula-Main.

We also compute the ratio of mean squared error (MSE) of the structural estimate $\widehat{\alpha}_k$, comparing Copula-All to Copula-Main (where $\text{MSE}(\widehat{\alpha}_k) = \text{Bias}^2(\widehat{\alpha}_k) + \text{Var}(\widehat{\alpha}_k)$, measuring overall estimation accuracy). Notably, Copula-All increases MSEs for all model parameter estimates, with the harmful effects being greatest for the squared term estimate $\widehat{\alpha}_2$, whose MSE is more than 6-times that of Copula-Main for all sample sizes (Figure W3).

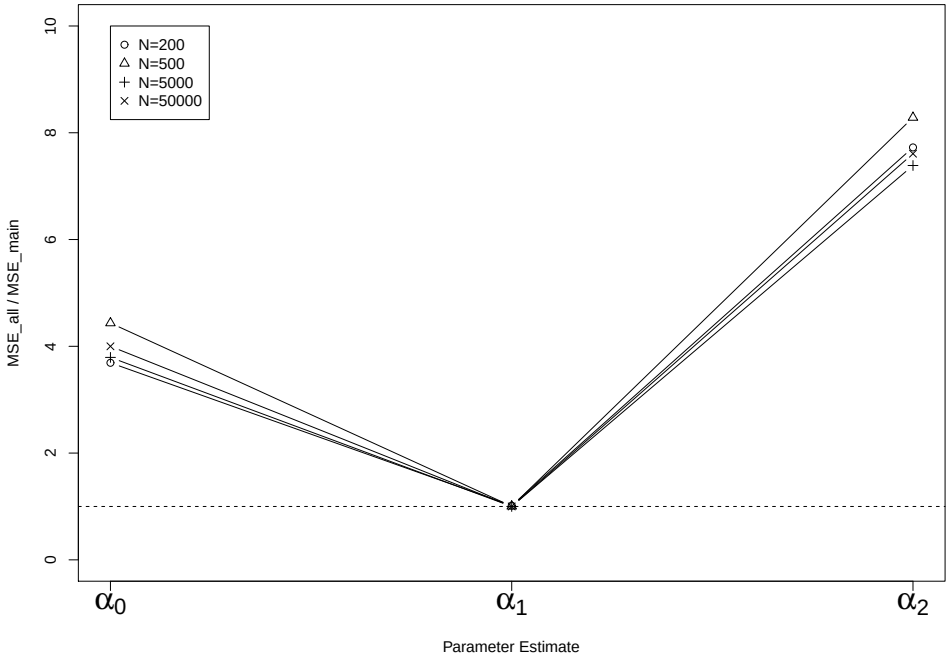

Figure W3: Ratio of mean squared errors of structural model estimates, with using the copula square term (Copula-All) to those without using the copula square term (Copula-Main).

WEB APPENDIX G: MEAN-CENTERING REGRESSORS

This section examines whether mean-centering helps improve the performance of Copula-All. A common practice for researchers in economics, management, and other fields is to mean-center the regressors before estimating models with higher-order terms. One argument for this practice is that by mean-centering the regressors, the correlation - and resulting collinearity problem - between the linear and higher-order terms (e.g., quadratic terms or interaction terms) is reduced (Aiken and West 1991; Kopalle and Lehmann 2006). However, Echambadi and Hess (2007) showed that mean-centering regressors does not alleviate collinearity problems in moderated regression models. Namely, none of the parameter estimates and sampling accuracy of main effects, simple effects, interactions, or R^2 is changed by mean-centering. By main effect and simple effect, we refer to the regression coefficient for a first-order term with and without mean-centering, representing the effect of a regressor when its moderators are set at their mean values and at zero (or absence of the attributed quantified by these moderators), respectively.

To illustrate this point, consider the following structural regression model with an interaction term:

$$Y = \alpha_0 + \alpha_1 P_1 + \alpha_2 P_2 + \alpha_3 P_1 * P2 + E$$

For the purposes of ease in interpretation or reducing the correlation between the linear and interaction terms, mean-centering regressors is often employed, which leads to the following equivalent model with parameter transformation:

$$Y = \alpha_0^c + \alpha_1^c (P_1 - \bar{P}_1) + \alpha_2^c (P_2 - \bar{P}_2) + \alpha_3^c (P_1 - \bar{P}_1) * (P_2 - \bar{P}_2) + E, \quad \text{(W18)}$$

where the parameters for the models before and after mean-centering have the following one-to-one relationship:

$$\begin{aligned}
\alpha_0^c &= \alpha_0 + \alpha_1\bar{P}_1 + \alpha_2\bar{P}_2 + \alpha_3\bar{P}_1\bar{P}_2 \\
\alpha_1^c &= \alpha_1 + \alpha_3\bar{P}_2 \\
\alpha_2^c &= \alpha_2 + \alpha_3\bar{P}_1 \\
\alpha_3^c &= \alpha_3. \qquad \text{(W19)}
\end{aligned}$$

As shown above, the regression coefficient α_1^c for the centered linear term $P_1 - \bar{P}_1$ represents the effect of P_1 when P_2 is equal to its mean value $\bar{P}_2$. Thus, α_1^c represents the main effect: the effect of P_1 when the other variables are at their mean values. In contrast, the coefficient using uncentered data, α_1, represents the simple effect: the effect of P_1 when the other variables are at zero (or absence of the attribute quantified by these other variables). The differences in estimates and standard errors between α_1 and α_1^c are due to the two coefficients having different substantive meanings, and both effects can be of substantive interest (Echambadi and Hess 2007). Quadratic terms can be considered a special case of the above model because a quadratic term can be considered as the interaction term of a regressor with itself. The relationship between parameters for models with quadratic terms before and after mean-centering can be derived similarly. Echambadi and Hess (2007) showed that the relationships in Equation W19 also holds for the OLS estimates of these model parameters.

However, our setting differs from the case of moderated regression models considered in Echambadi and Hess (2007), since we consider the more general case of endogeneity bias correction of structural regression models with endogenous higher-order regressors. Although

the relationships in Equation W19 hold exactly for OLS estimates (Echambadi and Hess 2007) for all data sets, such relationships only hold approximately for copula corrected estimates because copula generated regressors involve probability integral transformations. Specifically, we use the same data generating process for Cases I, II and III to generate data. When estimating models, we first mean-center all the first-order terms of the regressors, and then construct the higher-order terms using these mean-centered first-order terms. Copula correction terms are then constructed using these new regressors based on centered versions of the first-order terms of regressors. Because these copula correction terms involve probability integral transformation, the estimates and sampling accuracy of main effects, simple effects and interactions can change after mean centering, which differs from the case of Echambadi and Hess (2007) in which all regressors are exogenous.

For the models giving results in Tables W1, W2, and W3, we apply the OLS (without any correction), Copula-Main, and Copula-All to estimate the corresponding mean-centered structural regression models, with results summarized in Tables W4, W5, and W6, respectively. The true values for the parameters in the models after mean-centering are also listed in Tables W4 to W6. The mean values of the regressors $(\bar{P}_1, \bar{P}_2)$ used to compute these true parameter values are: $\frac{\phi(a)-\phi(b)}{\Phi(b)-\Phi(a)}$, where $\phi(\cdot)$ denotes the density function of the standard normal; when the marginal distribution of the regressor is the truncated standard normal on $[a, b]$, and $\frac{a+b}{2}$ when it is the uniform distribution on $[a, b]$.

Because copula correction terms for higher-order terms are not invariant to mean-centering, the ratios of the D-error for Copula-All to that of Copula-Main using mean-centered data will not be the same as those in Tables W1, W2, and W3, using uncentered data. Still, the same conclusion of inflated variability of estimates for Copula-All is apparent, and the

D-error measure ratios are all above 2. This finding is consistent with that of Echambadi and Hess (2007) in that mean-centering regressors does not alleviate collinearity problems in moderated regression models. Furthermore, mean-centering seemingly shifts the variance inflation from the regression coefficient estimates of first-order terms to those of the higher-order terms, and may hurt the estimation of the higher-order terms in some cases.

It is important to note, however, that this does not imply that mean-centering affects the estimation of the *same* first-order effects. As explained above, the regression coefficients for a first-order term (with and without mean-centering) represent different effects of one regressor evaluated at different values of its moderator: these regression coefficients represent the main effects when mean-centering regressors and the simple effects when using uncentered data. As such, regression coefficients for a first-order term with and without mean-centering are not directly comparable, although both main and simple effects can be of substantive interest (Echambadi and Hess 2007). When using the parameter estimates based on the centered data to compute the simple effects, we again find finite sample bias and inflated standard errors for the estimates of simple effects (results not shown here), as occurred when using uncentered data. In sum, we conclude that mean-centering does not overturn the under-performance of Copula-All relative to Copula-Main.

Table W4: Results from Case I: Interaction of Endogenous Regressors With Mean-Centering

N	Method	$\alpha_0^c(=8.192)$	$\alpha_1^c(=1.798)$	$\alpha_2^c(=4)$	$\alpha_3^c(=1)$	$\sigma(=1)$	D-error
200	OLS	**8.259**	**3.425**	**5.619**	**1.432**	**0.294**	
		(0.208)	(0.071)	(0.084)	(0.105)	(0.031)	—
	Copula-Main	8.172	1.897	4.072	1.041	0.967	
		(0.208)	(0.279)	(0.257)	(0.080)	(0.124)	0.0316
	Copula-All	8.180	1.896	4.069	**1.101**	0.972	
		(0.215)	(0.279)	(0.266)	(0.281)	(0.124)	0.0734
500	OLS	**8.262**	**3.425**	**5.615**	**1.431**	**0.297**	
		(0.134)	(0.045)	(0.051)	(0.065)	(0.02)	—
	Copula- M	8.184	1.838	4.018	1.025	0.990	
		(0.133)	(0.179)	(0.166)	(0.047)	(0.077)	0.0123
	Copula-All	8.189	1.838	4.020	**1.057**	0.992	
		(0.137)	(0.178)	(0.174)	(0.173)	(0.078)	0.0293
5000	OLS	**8.263**	**3.424**	**5.612**	**1.433**	**0.298**	
		(0.042)	(0.014)	(0.017)	(0.021)	(0.006)	—
	Copula-Main	8.191	1.803	3.999	1.003	1.000	
		(0.042)	(0.057)	(0.051)	(0.015)	(0.024)	0.0011
	Copula-All	8.192	1.803	3.999	1.009	1.000	
		(0.043)	(0.057)	(0.054)	(0.052)	(0.024)	0.0028
50000	OLS	**8.263**	**3.424**	**5.613**	**1.433**	**0.298**	
		(0.013)	(0.004)	(0.005)	(0.007)	(0.002)	—
	Copula-Main	8.192	1.799	3.999	1.000	1.000	
		(0.013)	(0.018)	(0.017)	(0.005)	(0.008)	0.0001
	Copula-All	8.192	1.799	3.999	1.002	1.000	
		(0.014)	(0.018)	(0.017)	(0.017)	(0.008)	0.0003

See the same note under Table W3.

Table W5: Results from Case II: Interaction between Endogenous and Exogenous Regressors With Mean-centering.

N	Method	$\alpha_0^c(=8.192)$	$\beta_1^c(=1.798)$	$\alpha_1^c(=4)$	$\alpha_2^c(=1)$	$\sigma(=1)$	D-error
200	OLS	8.190	1.800	**4.801**	0.999	**0.874**	
		(0.225)	(0.114)	(0.115)	(0.187)	(0.044)	—
	Copula-Main	8.183	1.800	4.010	1.000	1.009	
		(0.225)	(0.114)	(0.418)	(0.185)	(0.128)	0.0426
	Copula-All	8.182	1.800	4.004	0.999	1.051	
		(0.226)	(0.118)	0.(426)	(0.895)	(0.143)	0.1243
500	OLS	8.191	1.797	**4.801**	1.000	**0.875**	
		(0.143)	(0.075)	(0.072)	(0.116)	(0.028)	—
	Copula-Main	8.188	1.797	4.003	1.000	1.004	
		(0.143)	(0.074)	(0.259)	(0.113)	(0.081)	0.0167
	Copula-All	8.188	1.797	4.001	1.005	1.022	
		(0.143)	(0.076)	(0.262)	(0.558)	(0.084)	0.0489
5000	OLS	8.191	1.798	**4.799**	1.001	**0.876**	
		(0.045)	(0.023)	(0.023)	(0.036)	(0.009)	—
	Copula-Main	8.191	1.797	3.998	1.001	1.001	
		(0.045)	(0.023)	(0.082)	(0.036)	(0.026)	0.0016
	Copula-All	8.191	1.798	3.998	1.000	1.003	
		(0.045)	(0.023)	(0.082)	(0.170)	(0.026)	0.0046
50000	OLS	8.192	1.798	**4.799**	1.000	**0.877**	
		(0.015)	(0.007)	(0.007)	(0.011)	(0.003)	—
	Copula-Main	8.191	1.798	4.000	1.000	1.000	
		(0.015)	(0.007)	(0.025)	(0.011)	(0.008)	0.00015
	Copula-All	8.191	1.798	4.000	1.000	1.000	
		(0.015)	(0.008)	(0.025)	(0.053)	(0.008)	0.00045

See the same note under Table W3.

Table W6: Results from Case III: Endogenous Squared Terms With Mean-Centering

N	Method	$\alpha_0^c(=0)$	$\alpha_1^c(=-1)$	$\alpha_2^c(=1)$	$\sigma(=1)$	D-error
200	OLS	0.000	**1.414**	0.993	**0.727**	
		(0.080)	(0.188)	(0.737)	(0.037)	—
	Copula-Main	-0.001	-0.967	**0.912**	1.007	
		(0.085)	(1.008)	(0.785)	(0.193)	0.790
	Copula-All	0.000	-0.959	**0.857**	1.022	
		(0.196)	(1.019)	(2.353)	(0.194)	2.396
500	OLS	0.000	**1.414**	0.995	**0.729**	
		(0.049)	(0.117)	(0.458)	(0.024)	—
	Copula-Main	0.000	-0.993	0.949	1.005	
		(0.052)	(0.628)	(0.495)	(0.125)	0.311
	Copula-All	0.001	-0.999	**0.936**	1.011	
		(0.116)	(0.631)	(1.380)	(0.125)	0.871
5000	OLS	-0.001	**1.413**	1.002	**0.728**	
		(0.016)	(0.038)	(0.151)	(0.007)	—
	Copula-Main	-0.001	-0.993	0.995	0.999	
		(0.017)	(0.201)	(0.159)	(0.040)	0.031
	Copula-All	-0.002	-0.993	1.008	0.999	
		(0.036)	(0.202)	(0.417)	(0.040)	0.085
50000	OLS	-0.001	**1.415**	1.000	**0.728**	
		(0.005)	(0.013)	(0.045)	(0.002)	—
	Copula-Main	0.000	-1.003	1.000	1.001	
		(0.005)	(0.062)	(0.048)	(0.012)	0.003
	Copula-All	0.000	-1.003	0.998	1.001	
		(0.012)	(0.062)	(0.137)	(0.012)	0.009

See the same note under Table W3.