NBER WORKING PAPER SERIES

SHOULD COPULA ENDOGENEITY CORRECTION INCLUDE GENERATED REGRESSORS FOR HIGHER-ORDER TERMS? NO, IT HURTS

Yi Qian
Hui Xie
Anthony Koschmann

Working Paper 29978
http://www.nber.org/papers/w29978

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

We thank colleagues in the field for valuable comments. Yi Qian would like to acknowledge the SSHRC grant. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Should Copula Endogeneity Correction Include Generated Regressors for Higher-order Terms? No, It Hurts
Yi Qian, Hui Xie, and Anthony Koschmann
NBER Working Paper No. 29978
April 2022
JEL No. C01,C1

ABSTRACT

Causal inference in empirical studies is 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. A forefront of recent research is a new paradigm of handling endogenous regressors without using instrumental variables. Park and Gupta (Marketing Science, 2012) proposed instrument-free estimation using copulas that has been increasingly used in practical applications to address endogeneity bias. A relevant issue not studied is how to handle the higher-order terms (e.g., interaction and quadratic terms) of endogenous regressors using the copula approach. Recent applications of the approach have used disparate ways of handling these higher-order endogenous terms with unclear consequences. We show that once copula correction terms for the main effects of endogenous regressors are included as generated regressors, there is no need to include additional correction terms for the higher-order terms. This simplicity in handling higher-order endogenous regression terms is a merit of the instrument-free copula bias correction approach. More importantly, adding these unnecessary correction terms has harmful effects and leads to sub-optimal solutions of endogeneity bias, including finite-sample estimation bias and substantially inflated variability in estimates.

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

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

Anthony Koschmann
Eastern Michigan University
akoschma@emich.edu

1. INTRODUCTION

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

Because of 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 correcting for endogeneity bias (Wooldridge, 2010). Estimation can use either two-stage least squares or a control function approach (Petrin and Train, 2010). The method requires the existence of valid and strong IVs satisfying the stringent condition of exclusion restriction, which makes IVs hard to find and justify in practice (Ebbes *et al.*, 2005, 2009, Park and Gupta, 2012). When there exists theory or knowledge about the underlying mechanism of endogeneity, an alternative approach is to specify a model describing the exact process of yielding the observed values of the endogenous regressors, and estimate it jointly with the structural model of primary interest. For estimating a consumer demand model, a supply-side model reflecting researchers' belief about the managerial decisions and competitions 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 *et al.* 2003, Manchanda *et al.* 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 has been growing interest in developing endogeneity-correction methods that require neither observed IVs nor knowledge to correctly specify a supply model. These instrument-free methods exploit higher moments (Lewbel, 1997), heteroscedastic structural error structures (Rigobon, 2003), latent IVs (Ebbes *et al.*, 2005), and copula models (Park and Gupta, 2012) to correct for endogeneity bias. Ebbes *et al.* (2009), Park and Gupta (2012), Papies *et al.* (2017) and Rutz and Watson (2019) provided detailed comparisons of these IV-free methods with alternative methods. Unlike other IV-free methods, the copula correction method of Park and Gupta (2012) does not require the condition of each endogenous regressor containing an exogenous component, which is akin to the exclusion restriction condition and can be hard to justify in practical applications. Furthermore, one can implement the copula correction method by including copula transformation of endogenous regressors as additional regressors in the structural regression model to control for endogeneity. This copula correction method has been increasingly used in the marketing and related fields (see Table 1 for some recent applications of the method).

The main contribution of this work is to investigate the optimal copula correction approach to addressing endogeneity bias for models containing higher-order terms of endogenous regressors, an issue that has not been studied previously but is important to address given that such models are widely used in practice. Many practical applications in different fields are interested in estimating structural models with higher-order terms of endogenous regressors. Polynomial regressions are often employed to study non-monotonic relationships between causes and effects, such as an inverted-U relationship, in order to determine optimal policy and managerial intervention (Aghion *et al.*, 2005, Qian, 2007). Interaction terms are often included in the model to study relevant moderators of the relationships between causes and effects (Table 1).

As shown in Table 1, there exists apparent inconsistency in the literature regarding how to handle higher-order terms of endogenous regressors when applying the copula correction procedure. Specifically, the table presents applications for which we can ascertain whether copula generated regressors were included for higher-order terms. While some applications did not include copula generated regressors for endogenous higher-order terms without stating the reason, other applications argued for including them for higher-order endogenous terms to control for their endogeneity, as illustrated by the following two quotes.

"*In addition, 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 and followed the procedure for multiple endogenous covariates (Park and Gupta 2012).*" —- (Atefi *et al.*, 2018)

"*In our empirical model, the main possible endogenous variables are Scientific Knowledge*$_{it-1}$*, Technological Knowledge*$_{it-1}$*, and Areas of Government Activity*$_{it-1}$*. The interaction terms of knowledge with areas of government activities are also subject to endogeneity. Therefore, we constructed additional variables [copula correction terms]...*" —(Yoon *et al.*, 2018)

This research addresses the ambiguity regarding whether it is good practice to include copula generated regressors for endogenous higher-order regressors. Given the fast-growing adoption of the copula correction method, coupled with the importance and relevance of handling endogenous higher-order terms in practical applications, there is a pressing need to have such guidelines at hand. To clarify the issue, this work shows that under the copula endogeneity modeling framework, there is no need to include separate copula correction terms for endogenous higher-order terms: including copula correction terms only for the first-order terms of endogenous regressors is sufficient to obtain consistent estimates of structural regression model parameters. This simplicity can be considered as a merit of the copula approach to handling endogeneity. More importantly, we show that including copula generated regressors for higher-order regression terms is sub-optimal and can substantially hurt the performance of endogeneity correction, including finite-sample estimation bias and inflated standard errors of parameter estimates.

Section 2 next reviews the copula method to handle endogenous regressors and provides theoretical justifications of not including the generated regressors for higher-order terms, and expected harmful effects if one does. Section 3 describes a set of simulation studies supporting the theoretical justifications and demonstrating the harmful effects of including copula generated regressors for higher-order terms. We also investigate whether mean-centering regressors affects the conclusions. Section 4 ends with a discussion.

2. METHODOLOGY

2.1 *Review of the copula endogeneity-correction approach*

For completeness, we review briefly the copula endogeneity-correction approach; see Park and Gupta (2012) for more details. Consider the following structural linear regression model

$$Y_i \quad = \quad \alpha_0 + P_i\alpha + X_i^T\beta + E_i, \tag{1}$$

where $i = 1, \cdots N$ indexes independent units; Y_i is a scalar response variable; X_i contains a vector of exogenous regressors independent of the structural error E_i; P_i is a scalar continuous endogenous regressor associated with the structural error E_i. The regression coefficients (α, β) are structural model parameters capturing the causal or independent effects of regressors.

The association between the endogenous regressor P_i and the structure error E_i can lead to biased estimates of model parameters (i.e., endogeneity bias) when using estimation methods ignoring 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 positing a Gaussian copula model for the joint distribution of (E_i, P_i)

$$H(E_i, P_i) \quad = \quad C(H_E(E_i), H_P(P_i)) = \Psi_\rho(\Phi^{-1}(H_E(E_i), \Phi^{-1}(H_P(P_i)), \tag{2}$$

where $H(\cdot, \cdot)$, $H_E(\cdot)$ and $H_P(\cdot)$ are the cumulative distribution functions (CDFs) for (E, P), E and P, respectively; $\Phi(\cdot)$ is the CDF of the standard normal distribution; $\Psi_\rho(\cdot, \cdot)$ is the CDF of the bivariate standard normal distribution with a Pearson correlation coefficient of ρ; $C(\cdot, \cdot)$ is a copula function that maps two uniform marginal distributions on $[0, 1]$ to a CDF. In the above, the first equation means that the joint cumulative distribution of (E, P) can be written as a copula function $C(\cdot, \cdot)$ of marginal CDFs of E and P (Sklar, 1959), and the second equation assumes a Gaussian copula function for $C(\cdot, \cdot)$.

Park and Gupta (2012) proposed two equivalent estimation methods based on the copula model. By maximizing the likelihood function derived from the joint CDF of (E, P) from Eqn (2), one obtains consistent structural model parameter estimates that correct for endogeneity bias. Alternatively one can use a generated regressor approach. The Gaussian copula approach assumes the transformed data $E_i^* = \Phi^{-1}(H_E(E_i))$ and $P_i^* = \Phi^{-1}(H_P(P_i))$ jointly follow a bivariate normal distribution with variances of one and the correlation coefficient ρ, and, using the Cholesky transformation converting correlated variables to uncorrelated ones, they can be re-written as

$$\begin{pmatrix} P_i^* \\ E_i^* \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} w_{1i} \\ w_{2i} \end{pmatrix}, \qquad \begin{pmatrix} w_{1i} \\ w_{2i} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right), \tag{3}$$

where w_{1i} and w_{2i} are independent standard normal variables. Under the assumption of normal structural error: $E_i \sim N(0, \sigma^2)$, we have $E_i = \sigma E_i^* = \sigma\rho w_{1i} + \sigma\sqrt{1-\rho^2} w_{2i} = \sigma\rho P_i^* + \sigma\sqrt{1-\rho^2} w_{2i}$ and thus

$$Y_i \quad = \quad \alpha_0 + P_i\alpha + X_i^T\beta + \sigma\rho P_i^* + \sigma\sqrt{1-\rho^2} w_{2i}. \tag{4}$$

where $P_i^* = w_{1i}$ and w_{2i} are two independent standard normal random variables under the Gaussian copula model. As a result w_{2i} is also independent of P_i which is a function of P_i^*. Based on Eqn (4), Park and Gupta (2012) suggest including the transformed variable P_i^* as an additional control variable in the structural model: given P_i^* in the model, the OLS regression yields consistent model estimates as the new error term w_{2i} is uncorrelated with other terms on the right side of the above equation. The computation of the generated regressor $P_i^* = \Phi^{-1}(H_P(P_i))$ requires an estimate of $H_P(\cdot)$, the unknown marginal CDF of the endogenous regressor P_i. One simple approach is to estimate $H_P(\cdot)$ with the empirical CDF, $\hat{H}_P(\cdot)$, that assigns probability mass to the uniquely observed values of P_i in the sample according to their sample frequencies. The above derivation also makes clear that the distribution of the endogenous regressor P must be different from that of the generated regressor P^* which is normal. If the distribution of P approaches the normal distribution, P becomes approximately a linear transformation of P^*. The resulting colinearity between the two regressors P and P^* can lead to large standard errors of estimates and render the precise evaluation of the independent effect of P impossible with a finite sample size.

2.2 *Handling higher-order terms of endogenous regressors*

An issue not considered before is how to handle higher-order terms when using the above copula endogeneity-correction approach. To illustrate, consider the following model:

$$Y_i \quad = \quad \alpha_0 + P_i\alpha_1 + P_i^2\alpha_2 + X_i^T\beta + E_i, \tag{5}$$

As both P and P^2 are endogenous regressors, it is tempting to control the endogeneity of both endogenous regressors by adding two transformed variables $C_P = P_i^* = \Phi^{-1}(H_P(P_i))$ and $C_{P^2} = P_i^{2*} = \Phi^{-1}(H_{P^2}(P^2))$. However, the point of not needing the copula correction term C_{P^2} for the higher-order term P^2 is clear once we note that the joint distribution of (E, P, P^2) reduces to the joint distribution of (E, P), given that P^2 is determined by P. This point can also be seen as follows: under the same bivariate normal model for transformed data P_i^* and E_i^* as in Eqn (3), we obtain

$$Y_i \quad = \quad \alpha_0 + P_i\alpha_1 + P_i^2\alpha_2 + X_i^T\beta + \sigma\rho P_i^* + \sigma\sqrt{1-\rho^2}w_{2i}, \tag{6}$$

where P_i^* and w_{2i} are independent standard normal random variables under the Gaussian copula model, and thus w_{2i} is also independent of P_i and P_i^2, both of which are functions of P_i^*. Once the copula correction term $C_P = P_i^*$ is included as a control variable into the structural model of Eqn (5), the new (standardized) error term w_{2i} is already independent of and thus uncorrelated with P_i^2, and the correction term C_{P^2} is not needed. This simplicity of the copula correction to handling higher-order endogenous terms resembles that of the control function approach using IVs (Papies *et al.*, 2017).

More importantly, we show next the harmful effects of adding unnecessary copula transformations for high-order terms. The question now is: what if we include additional copula generated regressors for the higher-order terms? Will doing this lead to better or worse performance of the copula correction? Including copula correction terms for both P and P^2 is equivalent to the following copula joint model for multiple endogenous regressors (Park and Gupta, 2012):

$$\begin{aligned}\begin{pmatrix} C_P \\ C_{P^2} \\ E^* \end{pmatrix} &= N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{P,P^2} & \rho_{E,P} \\ \rho_{P,P^2} & 1 & \rho_{E,P^2} \\ \rho_{E,P} & \rho_{E,P^2} & 1 \end{bmatrix}\right) \\ &= \begin{bmatrix} 1 & 0 & 0 \\ \rho_{P,P^2} & \sqrt{1-\rho^2_{P,P^2}} & 0 \\ \rho_{E,P} & \frac{\rho_{E,P^2}-\rho_{E,P}\rho_{P,P^2}}{\sqrt{1-\rho_{P,P^2}}} & \sqrt{1-\rho^2_{E,P}-\frac{(\rho_{E,P^2}-\rho_{E,P}\rho_{P,P^2})^2}{1-\rho_{P,P^2}}} \end{bmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix},\end{aligned}$$

$$\text{where} \quad \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right).$$

We can then re-express the model in Eqn (5) as

$$\begin{aligned} Y_i &= \alpha_0 + P_i\alpha_1 + P_i^2\alpha_2 + X_i^T\beta + \sigma\frac{\rho_{E,P}-\rho_{E,P^2}\rho_{P,P^2}}{1-\rho_{P,P^2}}C_P \\ &\quad +\sigma\frac{\rho_{E,P^2}-\rho_{E,P}\rho_{P,P^2}}{1-\rho_{P,P^2}}C_{P^2} + \sigma\sqrt{1-\rho^2_{E,P}-\frac{(\rho_{E,P^2}-\rho_{E,P}\rho_{P,P^2})^2}{1-\rho_{P,P^2}}}w_{3i}, \end{aligned} \tag{7}$$

which yields the alternative generated regressor approach that includes both C_P and C_{P^2} as additional control variables with w_{3i} as the new error term. Both this approach and the simpler approach including only $C_P = P_i^*$ as the control variable can yield consistent model estimates.

However, a well-known issue in models containing higher-order regression terms is the significant colinearity between higher-order terms and their co-varying constituents, making it difficult to detect the main and higher-order effects (Echambadi and Hess, 2007). Including the unnecessary generated regressor C_{P^2} into the model introduces additional and potentially strong colinearity between C_{P^2} and those regressors that are functions of C_{P^2}: P, P^2 and C_P. The exacerbated colinearity problem will make it more difficult to distinguish the independent effects of the first- and higher-order terms involving the endogenous regressors, negatively affect the precision of the regression coefficient estimates, and make copula correction method perform worse than otherwise, as shown in a set of simulation studies below.

3. SIMULATION STUDIES

In this section we conduct a set of simulation studies to demonstrate empirically (1) that there is no need to add correction terms for higher-order terms of endogenous regressors to control for their endogeneity, and (2) more importantly, harmful effects occur if correction terms for higher-order terms are added to control for their endogeneity. We examine the cases of quadratic terms of endogenous regressors (Case I), interaction of two endogenous regressors (Case II) and interaction between endogenous and exogenous regressors (Case III). Finally we evaluate if the results change when we mean-center regressors.

3.1 *Case I: Quadratic Function of endogenous regressors*

Data were simulated from the following model

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

where $H_P(\cdot)$ is the CDF for the marginal distribution of P, $\alpha_0 = 0, \alpha_1 = -1, \alpha_2 = 1$ and $\rho = 0.7$ with a sample size of 500. We consider two marginal distributions of P: (1) $H_P(\cdot)$ is the CDF of the truncated standard normal distribution on $[-0.5, 1]$; (2) $H_P(\cdot)$ is the CDF of the uniform distribution on $[-1, 1]$. For each simulated data set, we then apply the following three estimation procedures by performing OLS regression of Y on the following sets of regressors:

$$\begin{aligned} \text{OLS:} \qquad & P, P^2 \\ \text{Copula-Main:} \qquad & P, P^2, C_P \\ \text{Copula-All:} \qquad & P, P^2, C_P, C_{P^2} \end{aligned}$$

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 observed sample, respectively; and Copula-Main and Copula-All denote procedures including copula correction terms for the main effect only and for all terms involving endogenous regressor P, respectively.

Results over 5000 simulated samples are summarized in Table 2. As expected, the OLS regression without any correction term yields large bias for the estimates of regression parameters and the error standard deviation σ in the structural regression model. Both Copula-Main and Copula-All can correct for the endogenous bias, demonstrating that there is no need to additionally include C_{P^2}. Furthermore, Copula-All yields less precise estimates (i.e. larger standard errors for all estimates) than Copula-Main (Table 2). To summarize efficiency of Copula-All and Copula-Main, we examine the D-error measure (Arora and Huber, 2001). 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 value of D-error 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 2, the D-error inflation for Copula-All is about 2-folds and 3-folds when P follows the truncated normal distribution and uniform distribution, respectively, meaning that in this case Copula-All requires 2–3 times of sample sizes in order to achieve approximately the same accuracy for estimating α_1 and α_2 jointly as Copula-Main. The variance inflation for the Copula-All estimate of α_2, the coefficient for the squared term, is much larger and equals $(\frac{0.383}{0.226})^2 \approx 3$ and $(\frac{0.354}{0.124})^2 \approx 8$ when P follows the truncated normal and the uniform distributions, respectively (Table 2). This means up to 8 times of sample size is required for Copula-All to achieve the same estimation accuracy of the squared term as Copula-Main.

3.2 *Case II: Interaction of two endogenous regressors*

We simulated data from the following structural regression model with interaction between two endogenous regressors P_1 and P_2

$$\begin{aligned}
Y &= \alpha_0 + \alpha_1 P_1 + \alpha_2 P_2 + \alpha_3 P_1 * P2 + E \\
\begin{pmatrix} E^* \\ P_1^* \\ P_2^* \end{pmatrix} &= N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho_{E1} & \rho_{E2} \\ \rho_{E1} & 1 & \rho_{12} \\ \rho_{E2} & \rho_{12} & 1 \end{bmatrix}\right) \\
E = H_E^{-1}(\Phi(E^*)) &= \Phi^{-1}(\Phi(E^*)), \qquad P_1 = H_{P_1}^{-1}(\Phi(P_1^*)), \quad P_2 = H_{P_2}^{-1}(\Phi(P_2^*)).
\end{aligned} \tag{8}$$

In the simulation we set $H_{P_1}(\cdot)$ to be the CDF of the uniform distribution on $[4, 6]$, $H_{P_2}(\cdot)$ to be the CDF of the truncated standard normal with a lower bound of 0, $\alpha_0 = 0, \alpha_1 = 1, \alpha_2 = -1, \alpha_3 = 1, \rho_{E1} = \rho_{E2} = 0.5, \rho_{12} = -0.5$ with a sample size of 200, 500, 5000 and 50000. For each sample size, we generated 5000 repeated samples. For each generated sample, we then apply three estimation procedures: OLS, Copula-Main and Copula-All. The OLS 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

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

Besides C_{P_1} and C_{P_2}, Copula-All adds the following copula correction term $C_{P_1 * P_2}$, where

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

$\hat{H}_{P_1}$, $\hat{H}_{P_2}$ 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.

Results over 5000 simulated samples are summarized in Table 3. As expected, the OLS regression without any correction terms yields large bias for the estimates of regression parameters and the error standard deviation σ in the structural regression model. Copula-Main corrects for the endogenous bias, demonstrating that there is no need to additionally include the copula correction term, $C_{P_1 * P_2}$. Furthermore, Copula-Main performs substantially better than Copula-All that does include $C_{P_1*P_2}$. In fact, Copula-All yields significantly biased parameter estimates (bold numbers in Table 3) with bias becoming smaller as sample size increases but remaining for a sample size as large as 50000. Copula-All also yields less precise estimates (larger standard errors) than Copula-Main: the standard errors for the estimates can be increased by more than 5-folds (underlined numbers in Table 3) for Copula-All compared with those from Copula-Main. The ratio of the overall D-error measure for Copula-All to that of Copula-Main over all regression coefficient estimates ranges from $\sim$1.3-folds for a sample size of 200 to $\sim$5-fold for a sample size of 50000.

3.3 *Case III: Interaction between endogenous and exogenous regressors*

We also considered the case of a structural model with an interaction term between an endogenous regressor and an exogenous regressor using the data generating process described in the Online Appendix A. The simulation results are summarized in Online Appendix Table 4. The overall conclusions remain the same in that Copula-Main outperforms Copula-All for correcting the endogeneity bias of OLS estimates. Compared with Copula-Main, Copula-All yields substantially less precise estimates and significant finite-sample estimation bias that becomes smaller as sample sizes increase but remains even if sample size is as large as 50000.

3.4 *Mean-centering does not affect the comparison conclusion*

A common practice for researchers in marketing 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 between the linear and high-order terms (e.g., quadratic terms or interaction terms) and the resulting colinearity problem is reduced. However, Echambadi and Hess (2007) showed that mean-centering regressors does not alleviate colinearity problems in moderated regression models in that none of the parameter estimates and sampling accuracy of main effects, simple effects and interactions as well as R^2 is changed by mean-centering.

Nonetheless, in this section we apply regressor mean-centering in our evaluation. This differs from the case of OLS regression of moderated regression models considered in Echambadi and Hess (2007) in that we consider the more general case of endogeneity-bias correction of structural regression models with endogenous higher-order regressors. Specifically, 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.

The results are summarized in Online Appendix Tables 5 to 7. 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 centered data will not be the same as those in Tables 2 to 4 using uncentered data. Nonetheless, we have the same conclusion of inflated variability of estimates for Copula-All in that the ratios of the D-error measure for Copula-All to that for Copula-Main over all regression coefficient estimates are all well above 1 and on average $\sim$3-folds (Online Appendix Tables 5 to 7). Furthermore, mean-centering seems to shift the variance inflation from the estimates of regression coefficients of first-order terms to those of the high-order terms (Online Appendix Tables 6 and 7) and may hurt the estimation of the high-order terms in some cases (e.g. the significantly inflated standard error of the estimate of α_2^c for Copula-All when P follows a truncated normal distribution in Online Appendix Table 5). It is, however, important to note that this does not imply that mean-centering helps with the estimation of the *same* first-order effects. As explained in Online Appendix B, 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: they represent the main effect when mean-centering regressors and the simple effect when using uncentered data. They are thus 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.

4. CONCLUSION

Estimation bias due to the presence of endogenous regressors is very prevalent and important to address in marketing and many other fields. Copula correction approach has been increasingly used to address endogeneity-bias due to its feasibility to use. There is, however, a lack of clear guidelines regarding the good practice of using the method to handle endogenous higher-order terms.

We investigate this issue here and show that copula correction terms for higher-order regression terms should not be included once the copula correction terms for the first-order endogenous regressors are included as generated regressors to control endogeneity. More importantly, adding the unnecessary copula correction terms for higher-order endogenous regressors will exacerbate the colinearity problem. This results in sub-optimal performance of copula correction method manifested as substantially increased standard errors (by up to 5 times as shown in our simulation studies) and potentially significant finite-sample bias of parameter estimates. 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 corrections terms should not be included.

References

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: An inverted u relationship. *Quarterly Journal of Economics*, **120**, 701–728.

Arora, N. and Huber, J. (2001). Improving parameter estimates and model prediction by aggregate customization in choice experiments. *Journal of Consumer Research*, **28**, 273–83.

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

Blauw, S. and Hans Franses, P. (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.

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

Ebbes, P., Wedel, M., Böckenholt, U., and Steerneman, A. (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. *Quant. Marketing Econom.*, **3(4)**, 365–392.

Ebbes, P., Wedel, M., and Böckenholt, U. (2009). Frugal iv alternatives to identify the parameter for an endogenous regressors. *J. Appl. Econometrics*, **24(3)**, 446–468.

Echambadi, R. and Hess, J. D. (2007). Mean-centering does not alleviate collinearity problems in moderated multiple regression models. *Marketing Science*, **26(3)**, 438–445.

Gielens, K., Geyskens, I., Deleersnyder, B., and Nohe, M. (2018). The new regulator in town: The effect of walmarts sustainability mandate on supplier shareholder value. *Journal of Marketing*, **82**, 124–141.

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

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

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

Lenz, I., Wetzel, H. A., and Hammerschmidt, M. (2017). Can doing good lead to doing poorly? firm value implications of csr in the face of csi. *J. of the Acad. Mark. Sci.*, **45**, 677–697.

Lewbel, A. (1997). Constructing instruments for regressions with measurement error when no additional data are available. *Econometrica*, **65(5)**, 1201–1213.

Luan, Y. and Sudhir, J. (2010). Forecasting marketing-mix responsiveness for new products. *J. Marketing Res.*, **47(3)**, 444–457.

Manchanda, P., Rossi, P., and Chintagunta, P. (2004). Response modeling with nonrandom marketing-mix variables. 41(november):4. *J. Marketing Res.*, **42**, 467–478.

Papies, D., Ebbes, P., and Van Heerde, H. J. (2017). Addressing endogeneity in marketing models. In T. B. P. Leeflang, J. Wieringa and K. Pauwels, editors, *Advanced methods for modeling markets*. Springer.

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

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

Qian, Y. (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.

Rigobon, R. (2003). Identification through heteroskedasticity. *Rev. Econom. Statist.*, **85(4)**, 777–

792.

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

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, **8**, 229–231.

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

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

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

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

Yang, S., Chen, Y., and Allenby, G. (2003). Bayesian analysis of simultaneous demand and supply. *Quant. Marketing Econom.*, **1(3)**, 251–275.

Yoon, H., Kim, N., Buisson, B., and Phillips, F. (2018). A cross-national study of knowledge, government intervention, and innovative nascent entrepreneurship. *Journal of Business Research*, **84**, 243–252.

Table 1

Examples of Applications Involving Higher-order Terms of Endogenous Regressors.

Applications	Examples of higher-order terms of endogenous regressors	Copula correction added for endogenous higher-order terms
Burmester *et al.* (2015)	AdStock * PubStock	Yes
Blauw and Hans Franses (2016)	Mobile Phone Ownership^2	Yes
Lenz *et al.* (2017)	Corporate Social Responsibility * Corporate Social Irresponsibility	No
Lamey *et al.* (2018)	New Product Price Premium (or Promotion Intensity) * Store context variables	No
Gielens *et al.* (2018)	R& D * Retailer Power	No
Yoon *et al.* (2018)	(Scientific or Technological) Knowledge * Areas of Government Activity	Yes
Atefi *et al.* (2018)	Trained Percentage^2	Yes
	Trained Percentage *Performance Diversity	
Guitart *et al.* (2018)	Advertising * Price	No
Wetzel *et al.* (2018)	Recruitment Spend * Brand Age	No
Heitmann *et al.* (2020)	Complexity *Segment Typicality	No

Table 2

Results from Case I: Endogenous Squared Terms.

Distribution	Method	$\alpha_0(=0)$	$\alpha_1(=-1)$	$\alpha_2(=1)$	$\sigma(=1)$	D-error
Trunc. Norm.	OLS	-0.327	0.705	0.884	0.726	
		(0.044)	(0.125)	(0.212)	(0.023)	—
	Copula-Main	-0.005	-0.974	0.997	1.005	
		(0.102)	(0.476)	(0.226)	(0.131)	0.098
	Copula-All	-0.010	-0.968	0.993	1.005	
		(0.143)	(0.503)	<u>(0.383)</u>	(0.153)	0.198
Uniform	OLS	0.001	0.183	0.997	0.728	
		(0.049)	(0.058)	(0.117)	(0.023)	—
	Copula-Main	0.000	-0.997	0.995	1.003	
		(0.061)	(0.296)	(0.124)	(0.119)	0.036
	Copula-All	-0.003	-0.992	0.995	1.006	
		<u>(0.126)</u>	(0.296)	<u>(0.354)</u>	(0.119)	0.105

Table presents the averages of the estimates and standard errors in the parenthesis over the repeated samples. 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.

Table 3

Results from Case II: 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	
		(1.36)	(0.270)	(0.468)	(0.107)	(0.134)	0.0368
500	OLS	-7.624	2.281	-1.546	1.432	0.297	
		(0.29)	(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)	(0.08)	(0.015)	(0.024)	0.0011
	Copula-All	**0.202**	0.968	**-0.713**	**0.929**	1.044	
		(0.318)	(0.061)	(0.24)	(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 note under Table 2. In addition, bold numbers highlight the estimates from Copula-Main or Copula-All with bias of at least 0.05.

Online Appendix A: 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

$$\begin{aligned} Y &= \alpha_0 + \beta_1 X + \alpha_1 P + \alpha_2 X * P + E \\ \begin{pmatrix} P^* \\ E^* \end{pmatrix} &= N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \\ E = H_E^{-1}(\Phi(E^*)) &= \Phi^{-1}(\Phi(E^*)), \qquad P = H_P^{-1}(\Phi(P^*)), \end{aligned} \tag{9}$$

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 a sample size of 200, 500, 5000 and 50000. For each sample size, we generated 5000 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 $X * P$. Besides 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 5000 simulated samples are summarized in Table 4.

Table 4

Results from Case III: 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)	0.00155
	Copula-All	**0.075**	0.983	**-0.836**	0.969	1.003	
		(0.361)	(0.077)	(0.654)	(0.121)	(0.028)	0.00640
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)	0.00015
	Copula-All	0.028	0.994	**-0.942**	0.989	1.000	
		(0.169)	(0.037)	(0.329)	(0.061)	(0.01)	0.00086

Table presents the averages of the estimates and standard errors in the parenthesis over the repeated samples. Bold numbers highlight the estimates from Copula-Main or Copula-All 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.

Online Appendix B: Mean-centering Regressors

Consider the following structural regression model with a 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

$$Y = \alpha_0^c + \alpha_1^c(P_1 - \bar{P}_1) + \alpha_2^c(P_2 - \bar{P}_2) + \alpha_3^c(P_1 - \bar{P}_1) * (P_2 - \bar{P}_2) + E, \tag{10}$$

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

$$\begin{aligned} \alpha_0^c &= \alpha_0 + \alpha_1 \bar{P}_1 + \alpha_2 \bar{P}_2 + \alpha_3 \bar{P}_1 \bar{P}_2 \\ \alpha_1^c &= \alpha_1 + \alpha_3 \bar{P}_2 \\ \alpha_2^c &= \alpha_2 + \alpha_3 \bar{P}_1 \\ \alpha_3^c &= \alpha_3. \end{aligned} \tag{11}$$

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 zeros (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 thus due to entirely that the two coefficients have 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.

For the models giving results in Tables 2, 3 and 4, 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 5, 6 and 7. The true values for the parameters in the models after mean-centering are listed in Tables 5 to 7. The means 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 $[a, b]$.

[h!]

Table 5

Results from Case I: Endogenous Squared Terms With Mean-Centering

Distribution	Method	$\alpha_0^c(=-0.164)$	$\alpha_1^c(=-0.587)$	$\alpha_2^c(=1)$	$\sigma(=1)$	D-error
Trunc. Norm.	OLS	-0.143	1.07	0.884	0.727	
		(0.049)	(0.086)	(0.209)	(0.023)	—
	Copula-Main	-0.163	-0.584	0.977	1.006	
		(0.053)	(0.466)	(0.224)	(0.133)	0.103
	Copula-All	-0.160	-0.587	0.960	1.012	
		<u>(0.113)</u>	(0.469)	<u>(0.621)</u>	(0.134)	0.288

Distribution	Method	$\alpha_0^c(=0)$	$\alpha_1^c(=-1)$	$\alpha_2^c(=1)$	$\sigma(=1)$	D-error
Uniform	OLS	0.001	0.186	1.002	0.729	
		(0.049)	(0.077)	(0.115)	(0.023)	—
	Copula-Main	0.001	-0.995	0.997	1.002	
		(0.051)	(0.304)	(0.122)	(0.120)	0.037
	Copula-All	-0.001	-0.991	0.995	1.007	
		<u>(0.120)</u>	(0.306)	<u>(0.351)</u>	(0.120)	0.107

See the note under Table 2 for the meaning of underlined numbers.

Table 6

Results from Case II: 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)	0.00011
	Copula-All	8.192	1.799	3.999	1.002	1.000	
		(0.014)	(0.018)	(0.017)	(0.017)	(0.008)	0.00030

See the note under Table 3 for the meanings of bold and underlined numbers.

Table 7

Results from Case III: 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 note under Table 3 for the meanings of bold and underlined numbers.