



Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Handling Endogenous Regressors by Joint Estimation Using Copulas

Sungho Park, Sachin Gupta,

To cite this article:

Sungho Park, Sachin Gupta, (2012) Handling Endogenous Regressors by Joint Estimation Using Copulas. Marketing Science 31(4):567-586. <https://doi.org/10.1287/mksc.1120.0718>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright ©2012, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Handling Endogenous Regressors by Joint Estimation Using Copulas

Sungho Park

W. P. Carey School of Business, Arizona State University, Tempe, Arizona 85287, spark104@asu.edu

Sachin Gupta

Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, New York 14853, sg248@cornell.edu

We propose a new statistical instrument-free method to tackle the endogeneity problem. The proposed method models the joint distribution of the endogenous regressor and the error term in the structural equation of interest (the structural error) using a copula method, and it makes inferences on the model parameters by maximizing the likelihood derived from the joint distribution. Similar to the “exclusion restriction” in instrumental variable methods, extant instrument-free methods require the assumption that the unobserved instruments are exogenous, a requirement that is difficult to meet. The proposed method does not require such an assumption. Other benefits of the proposed method are that it allows the modeling of discrete endogenous regressors and offers a new solution to the slope endogeneity problem. In addition to linear models, the method is applicable to the popular random coefficient logit model with either aggregate-level or individual-level data. We demonstrate the performance of the proposed method via a series of simulation studies and an empirical example.

Key words: endogeneity; copula method; instrumental variables; two-stage least squares; linear regression model; logit model; random coefficient

History: Received: December 30, 2010; accepted: March 18, 2012; Preyas Desai served as the editor-in-chief and Michel Wedel served as associate editor for this article. Published online in *Articles in Advance* May 22, 2012.

1. Introduction

A prominent challenge faced by empirical researchers in marketing, economics, and many other social sciences is endogenous regressors. Fundamentally, the problem arises because researchers are interested in questions of cause and effect, yet data in these disciplines are frequently obtained in nonexperimental settings. The most common form of this problem is correlation between regressors and random shocks. The canonical example arises in simultaneous-equations models of supply and demand. Researchers are interested in estimating the slope of the demand curve. However, ordinary least squares (OLS) estimates are biased because both the supply and the demand schedules are influenced by common shocks. Similarly, if there are unmeasured product characteristics that influence consumer decisions but are not observed by the researcher, then while the marketer observes these and incorporates them into decision making, an endogeneity problem could arise (Berry 1994, Villas-Boas and Winer 1999). Endogeneity problems also occur when there are omitted variables in a model of a causal relationship, such as the effect of schooling on wages (in this relationship, omitted variables may be motivation, ability, and socioeconomic status, which affect both schooling and wages), the

impact of medical treatments on health, or the effect of social insurance programs on labor supply.

Three broad classes of solutions to the problem of endogenous regressors have been discussed in the literature. The textbook solution uses observed instrumental variables to correct for endogeneity (Angrist and Krueger 2001 provide a survey of the history). The ideal instrumental variable meets two requirements: it is correlated with the endogenous regressor through a relationship that the researcher can explain and validate, and it is uncorrelated with the error term. The theory of instrumental variables is well developed, but applied researchers face the critical practical problem that good instruments are very hard to find. The common prescription that good instruments come from detailed knowledge of the economic mechanism and institutions that determine the (endogenous) regressor of interest does not often help in practice. Even if an instrumental variable is available, it is often subject to potential pitfalls because it fails to meet the two requirements adequately. The most serious problem is the violation of the “exclusion restriction”: the instrument should be uncorrelated with the error term in the structural equation of interest. Researchers routinely devote a great deal of effort toward convincing the reader that

their assumed exclusion restriction is justifiable, but this claim is often debated (Conley et al. 2012). Note that this claim is impossible to test directly because of unobservability of the error.

Another important problem arises as a result of “weak instruments”—instruments that are insufficiently correlated with the endogenous regressors, which may result in biased estimates. In a landmark study of the effect of schooling on wages, Angrist and Krueger (1991) used census data with more than 329,000 observations and more than 180 instruments to find significant results, but they had small F -statistics from first-stage regressions of endogenous regressors on instruments. Bound et al. (1995) reexamined Angrist and Krueger’s (1991) data and showed a finite sample bias even with the huge sample sizes. Further, in a rather dramatic demonstration of the problem of weak instruments, they replicated their results using random numbers as instruments. In fact, researchers sometimes have to worry if the “cure is worse than the disease”—is the bias caused by the use of weak instrumental variables worse than the bias caused by endogenous regressors?

The second class of solutions accomplishes correction of the endogeneity problem by supplementing the analysis with a supply-side model that explains observed realizations of the endogenous variables. Because these variables are often marketing variables, the model typically reflects firms’ goal-directed behavior as theorized by the researcher. A number of examples of this approach are found in the marketing literature; a very partial list includes Yang et al. (2003), Dotson and Allenby (2010), and Otter et al. (2011). A specific case in this setting is that of slope endogeneity—namely, random coefficients being correlated with the explanatory variable. Manchanda et al. (2004) provides an example of dealing with slope endogeneity via a supply-side model.

Yang et al. (2003) refer to this as a “full-information approach” because in all these cases the researcher needs to have prior knowledge or a theory of the process that generates the endogeneity. For instance, Yang et al. assume a particular game-theoretic model of price setting. The key concern with this approach is that incorrect specification of the supply-side model can lead to biased estimates of demand-side parameters (Chintagunta et al. 2006). Hartmann et al. (2011, p. 1081) note that “this approach is feasible only if full information is available to the analyst about how the firm allocates its marketing efforts. In the absence of such information, the analysis is sensitive to misspecification bias.”

The third class of solutions to the endogeneity problem includes statistical approaches that require neither observed instruments nor specification of

a supply-side model.¹ These are sometimes called “instrument-free” approaches. Ebbes et al. (2009) provide an excellent review and compares three extant instrument-free approaches: the *higher moments* (HM) approach (Erickson and Whited 2002, Lewbel 1997), the *identification through heteroscedasticity* (IH) estimator (Rigobon 2003, Hogan and Rigobon 2003), and the *latent instrumental variables* (LIV) method (Ebbes et al. 2005). In the HM approach, instruments are generated using the available data by exploiting higher-order moments. Similar to other method-of-moments approaches, this method does not make strong assumptions about the distributions of variables. However, this method works under certain higher-order moment conditions that may or may not hold (e.g., the requirement that the second moments of the mean-centered dependent variable and a mean-centered endogenous regressor are uncorrelated). The IH estimator also exploits the higher-order moments for model identification and is based on the method-of-moments approach. It requires researchers to have some information on heteroscedasticity: an observable grouping variable has to be available that describes the heteroscedastic error structure. However, such a variable may not always be available. The LIV method assumes that the endogenous regressor can be separated into an exogenous part and an endogenous part. The exogenous part is approximated using a latent discrete variable with a finite number of levels. Unlike the HM and IH methods, the LIV method is likelihood based. It does not require information on higher-order moment conditions or heteroscedasticity and thus is more generally applicable (see recent applications in Rutz and Trusov 2011, Rutz et al. 2012, and Zhang et al. 2009).

In all three extant instrument-free approaches, the endogenous regressor P can be represented as $P = f(Z) + \nu$, where $f(Z)$ is treated as a random variable and unique structures are imposed on $f(Z)$ for model identification. For instance, $f(Z)$ is assumed to be discrete in the LIV method (leading to a latent class model), skewed in the HM method, and heteroscedastic in the IH method. These assumptions ensure the distribution of P to be different from that of the structural error and thus provide model identification. A key assumption of all three extant instrument-free approaches is that random variable $f(Z)$ is exogenous. This assumption is similar to the

¹ There are in addition a few published approaches that do not neatly fit into our classification. For instance, Hartmann et al. (2011) applied a regression discontinuity approach to handle the endogeneity problem. Bronnenberg and Mahajan (2001) and van Dijk et al. (2004) used spatial econometrics approaches to tackle endogeneity. However, these methods are context-specific and often also require detailed knowledge of the process that generates the endogeneity.

“exclusion restriction” for observed instrumental variable methods.

There are two reasons for possible concern with the exogeneity assumption. First, in applications of instrument-free methods, researchers face a significant hurdle in arguing that the exclusion restriction is reasonable because they cannot observe $f(Z)$ and do not know what it represents. Second, in some situations one could offer conceptual challenges to the exogeneity assumption based on context-specific knowledge. For instance, reasons for possible endogeneity of retail prices that have been discussed in the literature are manufacturer coupons, TV advertising, and retailer’s strategic behavior (e.g., shelf-space allocation or special display), which are known to the retailer and consumer but unobserved by the researcher (Besanko et al. 1998); hence they are omitted variables in the demand model. Manufacturer coupons affect prices, and the latent discrete instrument variable in the LIV approach is then essentially capturing the presence versus absence of the coupon. The structural error in the demand model captures the omitted variables discussed previously. If manufacturers strategically plan both TV advertising and manufacturer coupons, or retailers allocate good shelf space for items with manufacturer coupons, then the latent discrete instrument variable might be correlated with the structural error and thus endogenous.

Our proposed method, which we describe next, falls into the third class of approaches for correcting for endogeneity; it is a statistical, instrument-free method. We model the joint distribution of the endogenous regressor and the error term in the structural equation of interest (we refer to this as structural error) and make inferences on the model parameters by maximizing the likelihood from the joint distribution. We use a nonparametric density estimation method to build the marginal distribution function of the endogenous regressor. The suggested procedure allows the observed data to determine the marginal distribution flexibly and does not impose any restrictive assumptions on the stochastic behavior of the data. To build the joint distribution function of nonparametrically determined endogenous regressors and error terms in the structural equation, we use a copula method. Copulas are functions that join or “couple” multivariate distributions to their one-dimensional marginal distribution functions (Balakrishnan and Lai 2009). In particular, using a Gaussian copula model, we construct a multivariate distribution that effectively captures the correlation between the regressor and the structural error. Once this correlation is properly handled, the proposed model does not suffer from the endogeneity problem, and we can obtain consistent estimates for model parameters.

Next we compare and contrast our proposed approach with extant instrument-free approaches. An important advantage of our proposed method is that it does not require the assumption of exogeneity of $f(Z)$ because it directly addresses the correlation between P and the structural error using a copula. It only requires the distinctiveness between the distributions of P and the structural error, which is a requirement in other instrument-free methods as well. In §3.9, we illustrate this advantage of the proposed method in a simulation study.

We believe that among extant instrument-free methods, the LIV method is most closely related to our proposed method. Both the LIV and the proposed method assume that the structural error follows a normal distribution. If the endogenous regressor is also normally distributed, both models encounter difficulty in separating variation as a result of the endogenous regressor from variation due to the structural error; hence an identification problem arises.² An important strength of the copula model is that it can capture the correlation between discrete variables and continuous variables. As a result, the proposed model can handle *discrete* endogenous regressors, a noteworthy advantage over the LIV and other instrument-free methods. Another advantage of the proposed method is that it can be used to handle the “slope endogeneity” problem that occurs when a regressor is correlated with its random coefficient. Researchers have typically used a “supply-side” model or instruments to tackle this problem (Manchanda et al. 2004, Kuksov and Villas-Boas 2008, Luan and Sudhir 2010). As noted previously, supply-side approaches are subject to possible misspecification if the researcher does not have adequate knowledge of the process generating the endogeneity. To the best of our knowledge, ours is the first statistical, instrument-free method to handle the slope endogeneity problem.

The method we propose is applicable both in linear models and in models that are nonlinear in variables and parameters. As an example of the latter, we apply the model to the random coefficient logit (RCL) model that has now become the workhorse of demand analysis in marketing and economics. We show that our proposed approach can be applied to correct for endogeneity when the RCL model is estimated using either individual-level or aggregate-level data. In simulation studies, we find that the proposed method provides unbiased and consistent estimates of model parameters, not only when the dependence structure between the endogenous regressor and the structural error is what the Gaussian copula assumes but also when this is not the case. That is, the proposed method

² We provide more details on model identification in Online Appendices 1 and 2 (at <http://dx.doi.org/10.1287/mksc.1120.0718>).

captures the dependence structure well enough to overcome the endogeneity problem. We also show that the proposed method is robust to misspecification of the distribution of the structural error. In an empirical application to paper towel sales data, we find that the proposed instrument-free method and the classical method using instrumental variables lead to essentially the same results on the endogeneity of retail price. This provides convergent validity to the proposed method. All in all, our results suggest that the proposed method handles the endogeneity problem well.

The remainder of this paper unfolds as follows. In §2, we explain the proposed method in detail. In §3, we evaluate the performance of the proposed method in several simulation studies. In §4, we apply the proposed method to real marketing data. We conclude in §5.

2. Model and Estimation Procedure

We first present the proposed method in the context of the familiar linear regression model. We show how to handle multiple endogenous regressors and discrete endogenous regressors in §§2.2 and 2.3, respectively. We then explain how the method can be applied in RCL models, to individual-level data in §2.4, and to aggregate-level data in §2.5. Finally, we explain how the slope endogeneity problem can be handled using the proposed method in §2.6.

2.1. Linear Regression Model

Consider the following linear regression model with an endogenous regressor:

$$Y_t = X_t' \beta + P_t \alpha + \xi_t, \quad (1)$$

where $t = 1, \dots, T$ indexes either time or cross-sectional units, Y_t is a 1×1 dependent variable, X_t is a $(k \times 1)$ vector of exogenous regressors, P_t is a (1×1) continuous endogenous regressor, ξ_t is a structural error term, and α and β are model parameters. P_t is correlated with ξ_t , and this correlation generates the endogeneity problem. To handle this problem, one can use instrumental variables that are correlated with P_t but uncorrelated with ξ_t . The instrument equation is

$$P_t = Z_t' \gamma + \nu_t, \quad (2)$$

where Z_t is a $l \times 1$ vector of instruments, γ is a $l \times 1$ vector of parameters, and ν_t is a random error term correlated with ξ_t . Two classical instrumental variables estimators are the two-stage least squares (TSLS) estimator and the generalized method-of-moments (GMM) estimator. To obtain the TSLS estimator, in the first stage the endogenous regressor P_t is regressed on Z_t , and in the second stage predicted values of P_t are plugged into (1) instead of P_t .

The GMM estimator is based on the orthogonality condition between Z_t and ξ_t . Neither of these methods imposes specific assumptions regarding the distribution of the error terms ξ_t and ν_t . Ebbes et al. (2005; i.e., the LIV method) and some other researchers make the assumption that ξ_t and ν_t are jointly bivariate normal (e.g., Chao and Phillips 1998, Kleibergen and Zivot 2003, Rossi et al. 2005).

Given “good” instrumental variables, one can obtain consistent estimates for the model parameters using TSLS, GMM, or other methods. However, even without instruments, one can obtain consistent estimates if the joint distribution of the endogenous regressor P_t and the structural error term ξ_t is known. To illustrate this, let us assume that P_t and ξ_t are generated from a certain bivariate distribution $f(p, \xi)$ and researchers know this. Then, one can obtain consistent estimates of the model parameters by maximizing the log-likelihood function derived from $f(p, \xi)$. As this simple example illustrates, researchers can resort to the information in the joint distribution of the endogenous regressor and the structural error term instead of instrumental variables. The proposed method is based on this essential idea.

To implement the proposed method, researchers need to know the joint distribution of the endogenous regressors P_t and the structural error term ξ_t , which is often unknown. To address this problem, we propose a unique copula model. In traditional multivariate distributions, once the parametric joint distribution function has been selected, the marginals are derived by integration. In other words, there is no flexibility for the marginals, and they are determined entirely by their parent joint multivariate distributions (Danaher and Smith 2011). Thus, traditional multivariate distributions limit the choice of marginals for the endogenous regressor and the structural error term and can impose restrictive assumptions on the stochastic behavior of the observed data, resulting in biases. Copula models overcome these shortcomings of traditional multivariate models by starting from the marginals. In the proposed approach, using information contained in the observed data, we first select marginal distributions for the endogenous regressor and the structural error term, respectively. Then, the copula model enables us to construct a flexible multivariate joint distribution from the marginals, allowing a wide range of correlations between the two marginals.

We assume that the marginal distribution of the structural error term is normal. The normality assumption for the structural error seems reasonable and has been widely used in the literature (e.g., Villas-Boas and Winer 1999, Yang et al. 2003, Chao and Phillips 1998, Kleibergen and Zivot 2003, Rossi et al. 2005, Ebbes et al. 2005). Of course,

other assumptions are possible. For example, if a researcher believes that the actual error distribution may have thicker tails than the normal, Student's t -distribution can be used. In §3, we assess robustness of the proposed model to possible misspecification in this respect. We find that model estimates are fairly robust to violations of the distributional assumption of ξ_t .

We use nonparametric density estimation to identify the marginal distribution of the endogenous regressor P_t . Unlike the structural error term, which is unobserved, we see sample data from the true distribution of P_t . Nonparametric density estimation allows the data to speak for themselves in determining the density function, thus avoiding problems of distributional misspecification. Suppose we are given independent and identically distributed observations $\{P_1, \dots, P_T\}$ with the density function $h(p)$. The simplest nonparametric density estimator is a naïve estimator (Silverman 1986):

$$\hat{h}(p) = \frac{1}{T \cdot b} \sum_{t=1}^T W\left(\frac{p - P_t}{b}\right),$$

where $W(x) = 0.5 \cdot I(|x| \leq 1)$, $I(\cdot)$ is an indicator function, and b is the bandwidth parameter that determines the smoothness of the kernel estimator. The larger the bandwidth, the smoother the density estimate. This estimator can be considered as a sum of “boxes” centered at the observations. Note that this estimator has some undesirable properties: $\hat{h}(p)$ is a step function and weights each observation inside the box equally. Because of these shortcomings, the kernel density estimator is preferred by many researchers. The kernel density estimator replaces $W(x)$, or “boxes,” by smooth “bumps,” or a kernel function $K(x)$, that is continuous and symmetric and satisfies $\int K(x) = 1$. One of the most commonly used kernel functions is the Epanechnikov kernel, and we use this kernel function in the proposed method. So the marginal density estimator for the endogenous regressor can be written as

$$\hat{h}(p) = \frac{1}{T \cdot b} \sum_{t=1}^T K\left(\frac{p - P_t}{b}\right), \quad (3)$$

where $K(x) = 0.75 \cdot (1 - x^2) \cdot I(|x| \leq 1)$.

Bandwidth selection is of crucial importance in density estimation, and various methods have been proposed in the literature. Silverman (1986) has suggested a frequently used data-driven bandwidth $b = 0.9 \cdot T^{-1/5} \cdot \min(s, IQR/1.34)$, where s is the sample standard deviation of the data and IQR is the interquartile range (the difference between the 75th and 25th percentile). This rule is regarded as a good starting point (DiNardo and Tobias 2001), and we use

this bandwidth in the proposed method. The kernel estimator $\hat{h}(p)$ can be shown to be consistent for $h(p)$ if b goes to 0 at a suitable rate as the sample size T goes to infinity.

Now we are ready to construct the joint distribution function from the marginal distributions of the endogenous regressor and the structural error term, denoted by $H(p)$ and $G(\xi)$, respectively. Also, let $F(p, \xi)$ be a joint distribution function with marginals $H(p)$ and $G(\xi)$. Sklar's theorem (Sklar 1959) states that there exists a copula function C such that for all p and ξ ,

$$F(p, \xi) = C(H(p), G(\xi)) = C(U_p, U_\xi), \quad (4)$$

where $U_p = H(p)$ and $U_\xi = G(\xi)$. This theorem is central to the theory of copulas and clarifies the role that copulas play in the relationship between multivariate distribution functions and their univariate marginals. Since $H(\cdot)$ and $G(\cdot)$ are marginal distribution functions, $U_p = H(p)$ and $U_\xi = G(\xi)$ are probability integral transformations, and thus U_p and U_ξ are uniform(0, 1) random variables and the copula can be viewed as a function $[0, 1]^2 \rightarrow [0, 1]$. Now it becomes clear that the copula provides a way to study and model the scale-free dependence between variables after filtering out the influence of marginal distributions. Let $f(p, \xi)$ denote the joint density function. Then, from (4), we have

$$f(p, \xi) = c(U_p, U_\xi) h(p) g(\xi), \quad (5)$$

where $c(U_p, U_\xi) = \partial^2 C / (\partial p \partial \xi)$ and $g(\xi)$ is the marginal density function of the structural error.

In the statistics literature, various copula functions have been studied (see Nelsen 2006 or Balakrishnan and Lai 2009). In the proposed method, we use the Gaussian copula:

$$\begin{aligned} C(U_p, U_\xi) &= \Psi_\rho(\Phi^{-1}(U_p), \Phi^{-1}(U_\xi)) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \\ &\cdot \int_{-\infty}^{\Phi^{-1}(U_p)} \int_{-\infty}^{\Phi^{-1}(U_\xi)} \exp\left[\frac{-(s^2 - 2\rho \cdot s \cdot t + t^2)}{2(1-\rho^2)}\right] ds dt, \end{aligned} \quad (6)$$

where Φ denotes the univariate standard normal distribution function and Ψ_ρ denotes the bivariate standard normal distribution function with the correlation coefficient ρ . If we let $P^* = \Phi^{-1}(U_p)$ and $\xi^* = \Phi^{-1}(U_\xi)$, the Gaussian copula models $[P^* \xi^*]'$ as the standard bivariate normal distribution with the correlation coefficient ρ . The Gaussian copula is a general and robust copula for most applications (Song 2000) and has many desirable properties (see Danaher and Smith 2011). There are numerous applications of Gaussian copulas, particularly in hydrology and finance. Recently, Danaher and Smith applied the

Gaussian copula to multiple marketing data sets and showed that it performs better than existing probability models. In §3, we assess robustness of the Gaussian copula in capturing the dependence between the endogenous regressor and the structural error. From (5) and (6), we can write the joint density function of P_t and ξ_t :

$$f(P_t, \xi_t) = \frac{1}{(1-\rho^2)^{1/2}} \exp \left[-\frac{\rho^2(\Phi^{-1}(U_{p,t})^2 + \Phi^{-1}(U_{\xi,t})^2)}{2(1-\rho^2)} + \frac{\rho\Phi^{-1}(U_{p,t})\Phi^{-1}(U_{\xi,t})}{(1-\rho^2)} \right] h(P_t)g(\xi_t). \quad (7)$$

We can obtain consistent estimates by maximizing the following log-likelihood function, derived from (7), with respect to the model parameters $\Theta = \{\alpha, \beta, \sigma_\xi, \rho\}$:

$$\ln l(\{P_t, \xi_t\} | \Theta) = -\frac{T}{2} \ln(1-\rho^2) - \sum_{t=1}^T \left[\frac{\rho^2(\Phi^{-1}(U_{p,t})^2 + \Phi^{-1}(U_{\xi,t})^2)}{2(1-\rho^2)} - \frac{\rho\Phi^{-1}(U_{p,t})\Phi^{-1}(U_{\xi,t})}{(1-\rho^2)} \right] + \ln \phi_{(0, \sigma_\xi^2)}(\xi_t), \quad (8)$$

where $\phi_{(0, \sigma_\xi^2)}(\cdot)$ is the normal density with mean 0 and variance σ_ξ^2 . Note that the nonparametric density $h(p)$ does not include any parameters and thus disappears from the log-likelihood function. However, it influences the log-likelihood function via $U_{p,t}$. By definition, $U_{p,t} = H(P_t) = \int_{-\infty}^{P_t} h(x) dx$. To obtain $U_{p,t}$, we integrate the nonparametric density estimate $\hat{h}(p)$ using a simple numerical method.³ $U_{\xi,t}$ can be easily calculated from the normal distribution function. For the implementation of the proposed method in the simulation study and empirical application, we use the GAUSS package. Specifically, we use the quasi-Newton numerical optimization routines (the Broyden–Fletcher–Goldfarb–Shanno method) for the maximization of the likelihood function.

The proposed copula model implies that $[P^* \xi^*]'$ follows the standard bivariate normal distribution with correlation coefficient ρ . This can be rewritten as follows:

$$\begin{pmatrix} P_t^* \\ \xi_t^* \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} \varpi_{1,t} \\ \varpi_{2,t} \end{pmatrix},$$

³ Specifically, we use the trapezoidal rule for numerical integration. This method works by constructing a grid on the integration variable and thus is easy to implement. And instead of the integration of density estimate, one can use the empirical distribution function to obtain $U_{p,t}$. We found that this approach provides comparable results to the proposed method.

$$\begin{pmatrix} \varpi_{1,t} \\ \varpi_{2,t} \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right),$$

or $\xi_t^* = \rho \cdot \varpi_{1,t} + \sqrt{1-\rho^2} \cdot \varpi_{2,t} = \rho \cdot P_t^* + \sqrt{1-\rho^2} \cdot \varpi_{2,t}$. Also, the normality assumption of the structural error states that $U_{\xi,t} = G(\xi_t) = \Phi_{(0, \sigma_\xi^2)}(\xi_t)$, where $\Phi_{(0, \sigma_\xi^2)}(\cdot)$ is the distribution function of the normal with mean 0 and variance σ_ξ^2 . Then,

$$\begin{aligned} \xi_t &= \Phi_{(0, \sigma_\xi^2)}^{-1}(U_{\xi,t}) = \Phi_{(0, \sigma_\xi^2)}^{-1}(\Phi(\xi_t^*)) = \sigma_\xi \cdot \xi_t^* \\ &= \sigma_\xi \cdot \rho \cdot P_t^* + \sigma_\xi \cdot \sqrt{1-\rho^2} \cdot \varpi_{2,t}. \end{aligned} \quad (9)$$

Finally, we can rewrite (1) as follows:

$$Y_t = X_t' \beta + P_t \alpha + \sigma_\xi \cdot \rho \cdot P_t^* + \sigma_\xi \cdot \sqrt{1-\rho^2} \cdot \varpi_{2,t}. \quad (10)$$

This representation provides another way to estimate the model. Given P_t^* as an additional regressor, $\varpi_{2,t}$ is not correlated with any other terms on the right-hand side of (10), and thus we can consistently estimate the model using the least squares estimator. P_t^* can be obtained by transforming $U_{p,t}$ using numerical integration of the nonparametric density estimator $\hat{h}(p)$. Note that P_t^* corrects for the endogeneity bias and in this respect is similar to the “control function” (Heckman 1978, Hausman 1978).

As we explained in §1, the nonnormality of an endogenous regressor is required for the model identification. We can easily understand the role of nonnormality in model identification using (10). If P_t follows a normal distribution, P_t^* is a linear transformation of P_t since $P_t^* = \Phi^{-1}(H(P_t))$. Hence, we cannot separately identify α and $\sigma_\xi \cdot \rho$ in (10). As the true distribution of P_t approaches a normal distribution, the correlation between P_t and P_t^* increases, causing a multicollinearity problem in (10). As is well known, multicollinearity does not create bias, only inefficiency. It appears that there is no simple remedy to this problem. If the standard errors of estimates are too large to draw meaningful conclusions, researchers might be able to mitigate this problem by increasing the sample size.

The inference procedure of the proposed method occurs in two stages. In the first stage, the density of the endogenous regressor is estimated using a nonparametric method, and then $U_{p,t}$ or P_t^* are computed from the estimated density. In the second stage, computed $U_{p,t}$ or P_t^* are used to construct the likelihood function, and the parameters are estimated. Here, computed $U_{p,t}$ and P_t^* are the “generated regressors” (Pagan 1984). However, the usual standard errors of parameters based on the information matrix regard the generated regressors as the given observations, and thus the standard errors are incorrect. We propose a simple bootstrap method to compute the correct

standard errors.⁴ The bootstrap method is a sample reuse technique designed to estimate standard errors and confidence intervals (Efron 1979). To illustrate the method, from the original data with T observations, we draw T observations with replacement. Using these resampled data, estimate the model parameters and save them. This process is repeated many times. The bootstrap standard error is the sample standard deviation of the saved estimates.⁵

Shih and Louis (1995) and Genest et al. (1995) examine general properties of two-step semiparametric copula estimators such as our method, and they show that the estimators are consistent and asymptotically normal under common mild regularity conditions if the first-step nonparametric estimation is consistent. In our first-step nonparametric estimation, we use the Epanechnikov kernel with Silverman's bandwidth selection rule, which has been shown to be consistent (Silverman 1986). In the simulations discussed in §3, we provide further empirical evidence of consistency.

2.2. Multiple Endogenous Regressors

Given (10), we can easily extend the proposed method to handle multiple endogenous regressors. Let us consider the following linear model with two endogenous regressors ($P_{1,t}$ and $P_{2,t}$):

$$Y_t = X_t' \beta + P_{1,t} \alpha_1 + P_{2,t} \alpha_2 + \xi_t. \quad (11)$$

The proposed model assumes the following relationship among variables:

$$\begin{pmatrix} P_{1,t}^* \\ P_{2,t}^* \\ \xi_t^* \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{12} & \rho_{\xi 1} \\ \rho_{12} & 1 & \rho_{\xi 2} \\ \rho_{\xi 1} & \rho_{\xi 2} & 1 \end{bmatrix} \right),$$

where $P_{1,t}^* = \Phi^{-1}(H_1(P_{1,t}))$, $P_{2,t}^* = \Phi^{-1}(H_2(P_{2,t}))$, and $H_1(\cdot)$ and $H_2(\cdot)$ are distribution functions of $P_{1,t}$ and $P_{2,t}$, respectively. This can be rewritten as follows:

$$\begin{pmatrix} P_{1,t}^* \\ P_{2,t}^* \\ \xi_t^* \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \rho_{12} & \sqrt{1-\rho_{12}^2} & 0 \\ \rho_{\xi 1} & \frac{\rho_{\xi 2}-\rho_{12}\rho_{\xi 1}}{\sqrt{1-\rho_{12}^2}} & \sqrt{1-\rho_{\xi 1}^2-\frac{(\rho_{\xi 2}-\rho_{12}\rho_{\xi 1})^2}{1-\rho_{12}^2}} \end{pmatrix} \cdot \begin{pmatrix} \varpi_{1,t} \\ \varpi_{2,t} \\ \varpi_{3,t} \end{pmatrix},$$

⁴ The proposed method can be classified as a semiparametric method because it combines a nonparametric part (the distribution of endogenous regressor) and a parametric part (the distribution of the structural shock). In semiparametric models, analytical derivation of asymptotic standard error can be extremely difficult, and one can use bootstrap standard error instead.

⁵ In the simulation study and empirical application, we use 100 replications to estimate the standard error. One exception is Simulation Case 5 ($T=1,000$), where we use 50 replications because of computation time. It is reported that a total of 50–200 replications is generally adequate for estimates of standard error (Mooney and Duval 1993).

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

Then,

$$\xi_t = \sigma_\xi \cdot \xi_t^* = \sigma_\xi \cdot \frac{\rho_{\xi 1} - \rho_{12}\rho_{\xi 2}}{1 - \rho_{12}^2} \cdot P_{1,t}^* + \sigma_\xi \cdot \frac{\rho_{\xi 2} - \rho_{12}\rho_{\xi 1}}{1 - \rho_{12}^2} \cdot P_{2,t}^* + \sigma_\xi \cdot \sqrt{1 - \rho_{\xi 1}^2 - \frac{(\rho_{\xi 2} - \rho_{12}\rho_{\xi 1})^2}{1 - \rho_{12}^2}} \cdot \varpi_{3,t}.$$

Also, we can rewrite (11) as follows:

$$Y_t = X_t' \beta + P_{1,t} \alpha_1 + P_{2,t} \alpha_2 + \sigma_\xi \cdot \frac{\rho_{\xi 1} - \rho_{12}\rho_{\xi 2}}{1 - \rho_{12}^2} \cdot P_{1,t}^* + \sigma_\xi \cdot \frac{\rho_{\xi 2} - \rho_{12}\rho_{\xi 1}}{1 - \rho_{12}^2} \cdot P_{2,t}^* + \sigma_\xi \cdot \sqrt{1 - \rho_{\xi 1}^2 - \frac{(\rho_{\xi 2} - \rho_{12}\rho_{\xi 1})^2}{1 - \rho_{12}^2}} \cdot \varpi_{3,t}. \quad (12)$$

Given $P_{1,t}^*$ and $P_{2,t}^*$ as additional regressors, $\varpi_{3,t}$ is not correlated with any other terms on the right-hand side of (12), and thus we can consistently estimate the model using the least squares estimator. $P_{1,t}^*$ and $P_{2,t}^*$ can be obtained from the nonparametric densities of $P_{1,t}$ and $P_{2,t}$. Note that we are not estimating two-dimensional nonparametric distribution in the case of two endogenous regressors. Copulas combine one-dimensional marginals to construct a joint distribution. Thus, the proposed copula model does not suffer from the “curse of dimensionality problem,” which is frequently encountered in nonparametric estimation of multivariate distribution functions. Further extension to more endogenous regressors is possible using a similar construction as in (12).

2.3. Discrete Endogenous Regressor

Now assume that endogenous regressor P_t is a discrete variable. The distribution function $H(\cdot)$ is a step function, and we can define $U_{p,t}$ to be related to the discrete variable P_t through the following inequality:

$$H(P_t - 1) < U_{p,t} < H(P_t).$$

$U_{p,t}$ is uniformly distributed on $[0, 1]$ and can be considered as a latent variable that determines the discrete outcome of P_t . The above inequality implies the following relationship between $P_t^* = \Phi^{-1}(U_{p,t})$ and $H(P_t)$:

$$\Phi^{-1}(H(P_t - 1)) < P_t^* < \Phi^{-1}(H(P_t)).$$

$H(P_t)$ can be easily estimated from the frequencies of observed data, and thus we do not need to perform the nonparametric density estimation nor the numerical integration to obtain P_t^* as in the continuous

regressor cases. By introducing P_t^* as an additional regressor, we can consistently estimate the model parameters as in (10).

The proposed method decomposes the structural error into one that is correlated with the endogenous regressor and the other that is unrelated to the endogenous regressor by relying on the information contained in the distribution of the endogenous regressor. Therefore, the precision of the proposed method depends on the quality of information contained in the distribution of the endogenous regressor. Some commonly used discrete distributions in the marketing literature are the Poisson, the negative binomial, and the beta-binomial distribution. These have been employed to model, for instance, the number of exposures to a TV advertisement (Rust 1986). Although these distributions contain rich information, some discrete variables do not contain enough information to decompose the structural error into the correlated and uncorrelated parts. For example, a Bernoulli random variable has only two possible outcomes (0 and 1) and does not contain rich enough information. But as the support of the discrete variable increases, the performance of the proposed method improves quickly. We demonstrate with simulation results in §3.

2.4. Random Coefficient Logit Model Using Individual-Level Data

The RCL model has been used in numerous empirical studies in marketing and economics and has been applied to individual-level data (e.g., Chintagunta et al. 2005). We briefly describe this model next. Let us specify an individual's latent utilities for choice alternatives as follows:

$$\begin{aligned} u_{h0t} &= \varepsilon_{h0t}, \quad \text{if no purchase,} \\ u_{hjt} &= X'_{jt}\beta_h + P_{jt}\alpha_h + \xi_{jt} + \varepsilon_{hjt} \\ &= X'_{jt}(\bar{\beta} + b_h) + P_{jt}(\bar{\alpha} + a_h) + \xi_{jt} + \varepsilon_{hjt}, \\ &\quad j = 1, \dots, J, \end{aligned} \quad (13)$$

where $t = 1, \dots, T$ denotes time, j denotes choice alternatives, and $h = 1, \dots, H$ denotes individuals. ε_{h0t} and ε_{hjt} are Type I extreme value errors, and (13) becomes a logit model. The structural error ξ_{jt} is correlated with P_{jt} , and this correlation causes the endogeneity problem as in the linear regression case. ξ_{jt} is also referred to as “common shock” or “unmeasured product characteristic” in the literature. Consumers' heterogeneous preference is considered via a_h and b_h in the model.

To estimate the model, we follow a two-step estimation method. The basic approach to handle the endogeneity problem remains the same as in the linear

regression case. The second equation in (13) can be rewritten as follows:

$$\begin{aligned} u_{hjt} &= (X'_{jt}\bar{\beta} + P_{jt}\bar{\alpha} + \xi_{jt}) + X'_{jt}b_h + P_{jt}a_h + \varepsilon_{hjt} \\ &= \delta_{jt} + X'_{jt}b_h + P_{jt}a_h + \varepsilon_{hjt}, \end{aligned}$$

where $\delta_{jt} = X'_{jt}\bar{\beta} + P_{jt}\bar{\alpha} + \xi_{jt}$. δ_{jt} is common to all individuals. We treat $\{\delta_{jt}\}_{j=1, \dots, J, t=1, \dots, T}$ as parameters in the first step and estimate these along with the other parameters. Because we have individual-level data, the first step estimation is straightforward. $\{\delta_{jt}\}$ are regarded as time- and alternative-specific fixed effects, and a simulation method is applied to handle random deviations a_h and b_h . Also, note that the endogeneity problem does not arise in the first stage because the endogenous regressor and the structural error are both encapsulated in δ_{jt} . Let us denote the first-step estimates of δ_{jt} as $\hat{\delta}_{jt} = \delta_{jt} + s_{jt}$, where s_{jt} is the estimation error, which is asymptotically normal. Now we have

$$\begin{aligned} \hat{\delta}_{jt} &= \delta_{jt} + s_{jt} = X'_{jt}\bar{\beta} + P_{jt}\bar{\alpha} + \xi_{jt} + s_{jt} \\ &= X'_{jt}\bar{\beta} + P_{jt}\bar{\alpha} + \bar{\xi}_{jt}, \end{aligned} \quad (14)$$

where $\bar{\xi}_{jt} = \xi_{jt} + s_{jt}$. Equation (14) is the same as the linear regression model (1) except that an additional dimension for choice alternatives j is added in the data. At the second step, we can consistently estimate the model parameters of interest $\bar{\alpha}$ and $\bar{\beta}$ using the copula method described in the previous section.

Note that our two-step estimation strategy is similar to the method proposed by Chintagunta et al. (2005). The first-step estimation is the same as their approach. In the second step, we apply the proposed copula method to $\hat{\delta}_{jt}$ to handle the endogeneity problem, whereas Chintagunta et al. (2005) applies the instrumental variable method to $\hat{\delta}_{jt}$.

2.5. Random Coefficient Logit Model Using Aggregate-Level Data

The RCL model has been applied to aggregate-level data by numerous researchers (e.g., Chintagunta 2001, Sudhir 2001) beginning with Berry et al. (1995). The utility specification for the RCL model with aggregate-level data is the same as (13). However, what we observe in the data is different. If individual-level data are available, a researcher knows the actual choice of individual h at time t . Aggregate-level data add up individual-level outcomes to $(J \times T)$ market shares $S_{jt} = n_{jt}/H$, where n_{jt} is the count of purchases for alternative j in time period t . Note that the two-step estimation of the previous section is no longer feasible because the number of first-step parameters is greater than the number of observations $(J \times T)$. Let us assume that $\xi_{jt} \sim N(0, \sigma_{\xi_j}^2)$ and $[P_{jt}^* \xi_{jt}^*]'$ follows

the standard bivariate normal distribution with correlation coefficient ρ_j for $j = 1, \dots, J$. Then, as in (9), ξ_{jt} can be written as $\sigma_{\xi_j} \cdot \rho_j \cdot P_{jt}^* + \sigma_{\xi_j} \cdot \sqrt{1 - \rho_j^2} \cdot \varpi_{2,jt}$. Consequently, we can rewrite the second equation of (13) as follows:

$$u_{hjt} = X'_{jt}\beta_h + P_{jt}\alpha_h + \sigma_{\xi_j} \cdot \rho_j \cdot P_{jt}^* + \sigma_{\xi_j} \cdot \sqrt{1 - \rho_j^2} \cdot \varpi_{2,jt} + \varepsilon_{hjt}. \quad (15)$$

$\varpi_{2,jt}$ is an independent and identically distributed random shock with a standard normal distribution and is uncorrelated with any other term in (15). Therefore, we can treat $\varpi_{2,jt}$ as an additional error component and integrate it out of the demand function. P_{jt}^* can be obtained from $U_{p,jt}$ using the numerical integration of the nonparametric density estimator $\hat{h}_j(p_j)$, as we have explained in §2.1. From (15) we can derive the logit probability that individual h chooses alternative j :

$$\begin{aligned} \Pr_{h0t} &= (1) \cdot \left(1 + \sum_{i=1}^J \exp \left(X'_{it}(\bar{\beta} + b_h) + P_{it}(\bar{\alpha} + a_h) \right. \right. \\ &\quad \left. \left. + \sigma_{\xi_i} \cdot \rho_i \cdot P_{it}^* + \sigma_{\xi_i} \cdot \sqrt{1 - \rho_i^2} \cdot \varpi_{2,it} \right) \right)^{-1}, \\ \Pr_{hjt} &= \left(\exp \left(X'_{jt}(\bar{\beta} + b_h) + P_{jt}(\bar{\alpha} + a_h) \right. \right. \\ &\quad \left. \left. + \sigma_{\xi_j} \cdot \rho_j \cdot P_{jt}^* + \sigma_{\xi_j} \cdot \sqrt{1 - \rho_j^2} \cdot \varpi_{2,jt} \right) \right) \\ &\quad \cdot \left(1 + \sum_{i=1}^J \exp \left(X'_{it}(\bar{\beta} + b_h) + P_{it}(\bar{\alpha} + a_h) \right. \right. \\ &\quad \left. \left. + \sigma_{\xi_i} \cdot \rho_i \cdot P_{it}^* + \sigma_{\xi_i} \cdot \sqrt{1 - \rho_i^2} \cdot \varpi_{2,it} \right) \right)^{-1}, \\ &\quad j=1, \dots, J. \end{aligned}$$

Conditional on $\varpi_{2,t} = \{\varpi_{2,1t}, \dots, \varpi_{2,Jt}\}$, we can write the likelihood of the observed aggregate-level data in time period t :

$$L_t(\varpi_{2,t}) = \frac{H!}{n_{0t}! \dots n_{Jt}!} \prod_{j=0}^J \left(\int \Pr_{hjt}(\varpi_{2,t}, a_h, b_h) \cdot \varphi(a_h, b_h) d(a_h, b_h) \right)^{n_{jt}}, \quad (16)$$

where $\varphi(a_h, b_h)$ is the density function of the heterogeneity distribution. Since $\varpi_{2,t}$ are unknown, we integrate them out:

$$L_t = \int L_t(\varpi_{2,t}) \phi(\varpi_{2,t}) d\varpi_{2,t},$$

where $\phi(\cdot)$ is the standard normal density. The likelihood function for the sample of T weeks becomes $\prod_{t=1}^T L_t$. Monte Carlo simulation methods can be used

to calculate the likelihood (see Keane 1993), and the resulting estimator becomes a simulated maximum likelihood (SML) estimator. To be more specific, the integrals in the likelihood function can be calculated by taking averages of computed integrands using random normal draws for a_h , b_h , and $\varpi_{2,t}$. The main idea of this estimation scheme is in line with Park and Gupta's (2009) SML estimator, in which the authors decompose ξ_{jt} into an endogenous part and an exogenous part and account for the endogenous part using instrumental variables via control function and then integrate out the exogenous part. A critical differentiating advantage of our proposed method is that it does not require instruments, whereas Park and Gupta's SML does. One computational issue with the proposed SML is that the evaluation of the likelihood function may encounter numerical problems because $(\int \Pr_{hjt}(\varpi_{2,t}, a_h, b_h) \varphi(a_h, b_h) d(a_h, b_h))^{n_{jt}}$ in (16) reaches machine 0 fairly quickly. If such a problem occurs, researchers can use a subsampling method.⁶

2.6. Slope Endogeneity

Consider the following random coefficient model:

$$Y_t = X'_t\beta_t + P_t\tilde{\xi}_t + \eta_t = X'_t\beta_t + P_t(\tilde{\xi} + \xi_t) + \eta_t, \quad (17)$$

where $t = 1, \dots, T$ indexes cross-sectional units (i.e., markets or households), and β_t is a vector of random coefficients. Unlike the previous specifications, η_t is an exogenous normal shock, and thus endogeneity as a result of correlation between the structural error and regressors is not an issue here.⁷ We assume that $\xi_t \sim N(0, \sigma_\xi^2)$, and by doing so we capture the heterogeneity of cross-sectional units with respect to P_t . More importantly, we assume that ξ_t is correlated with P_t . In the presence of correlation between coefficient and regressor, both the OLS estimator and the standard IV estimator are inconsistent (Heckman 1976), and this problem in estimation is referred to as "slope endogeneity." This problem occurs commonly in price discrimination situations in which the level of P_t is strategically set based on the value of ξ_t . For example, assume that P_t is price, $\tilde{\xi}_t$ represents price sensitivity, and t indexes geographic markets. The marketing manager of market t might use knowledge of price sensitivity of the market area to set market-specific prices. This strategic decision results in correlation between P_t and ξ_t , and the slope endogeneity problem arises.

⁶ See Park and Gupta (2009) for a detailed discussion of the subsampling method. Musalem et al. (2009) reported a similar numerical problem when the number of individuals underlying the aggregate share is larger than about 500. To circumvent this numerical problem, they suggested a subsampling method that is conceptually similar to Park and Gupta's (2009) approach.

⁷ In the literature, this common type of endogeneity is referred to as "intercept endogeneity" in contrast to "slope endogeneity."

The proposed Gaussian copula implies that $[P_t^* \xi_t^*]'$ follows the standard bivariate normal distribution with correlation coefficient ρ , and from this we can rewrite (17) as the following:

$$\begin{aligned} Y_t &= X_t' \beta_t + P_t (\bar{\xi} + \sigma_{\xi} \cdot \rho \cdot P_t^* + \sigma_{\xi} \sqrt{1 - \rho^2} \cdot \varpi_{2,t}) + \eta_t \\ &= X_t' \beta_t + \bar{\xi} \cdot P_t + \sigma_{\xi} \cdot \rho \cdot P_t \times P_t^* \\ &\quad + \sigma_{\xi} \sqrt{1 - \rho^2} \cdot P_t \times \varpi_{2,t} + \eta_t. \end{aligned} \quad (18)$$

The interaction term $P_t \times P_t^*$ handles the correlation between P_t and ξ_t . $\varpi_{2,t}$ is uncorrelated with P_t , and thus we can regard $\varpi_{2,t}$ as the usual random effect term and use standard methods for random effect models in estimation.

3. Simulation Study

In this section, we describe Monte Carlo simulation experiments that are designed with the following goals: (a) to assess the performance of the proposed method, (b) to examine robustness of the proposed estimator to key distributional assumptions about the structural error, (c) to assess generalizability and robustness of the suggested Gaussian copula in capturing various dependence structures between the endogenous regressor and the structural error, and (d) to compare the performance of the proposed method with existing methods. To achieve these goals, we consider nine cases that are summarized in Table 1.

For all models, we measure bias as absolute deviation of the mean of the sampling distribution from the true parameter value expressed in terms of number of standard errors of the parameter estimate, and we denote this quantity as t_{bias} . To infer statistical significance, we use the t -test. If the t_{bias} is smaller

than two standard errors (corresponding to the critical t -statistic for a test at 95% confidence), we infer that this deviation could be due to chance; hence the estimate is consistent.

3.1. Case 1: Linear Regression Model

The specific data-generating process (DGP) and parameter values we assign are summarized below:

$$\begin{aligned} \begin{pmatrix} \xi_t^* \\ P_t^* \\ Z_t^* \end{pmatrix} &\sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_2 \\ 0 & \rho_2 & 1 \end{bmatrix} \right) \\ &= N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0.8 \\ 0 & 0.8 & 1 \end{bmatrix} \right), \end{aligned} \quad (19)$$

$$\xi_t = G^{-1}(U_{\xi,t}) = G^{-1}(\Phi(\xi_t^*)) = \Phi^{-1}(\Phi(\xi_t^*)), \quad (20)$$

$$P_t = H^{-1}(U_{P,t}) = H^{-1}(\Phi(P_t^*)) = \Phi(P_t^*), \quad (21)$$

$$Y_t = \alpha \cdot P_t + \xi_t = -1 \cdot P_t + \xi_t. \quad (22)$$

We consider two levels of sample size, $T = 200$ and $T = 400$. ξ_t^* and P_t^* are correlated, and thus ξ_t and P_t are correlated, generating the endogeneity problem. By letting $Z_t = \Phi(Z_t^*)$, we generate an instrumental variable Z_t . The instrumental variable is ideal since it is uncorrelated with ξ_t but highly correlated with P_t . Note that ξ_t follows the standard normal distribution, and P_t and Z_t follow the uniform distribution. We consider four different estimation methods: (i) OLS, (ii) TSLS, (iii) the proposed method in the form of (8) (we refer to this as Proposed I), and (iv) the proposed method in the form of (10) (we refer to this as Proposed II). One thousand data sets are generated as replicates, and models are estimated on each data set

Table 1 Description of Simulation Studies

Case	Data-generating process	Goals
1.	Linear regression model, structural error from the normal distribution, Gaussian copula	Examine the performance of the proposed method in linear regression models
2.	Linear regression model, structural error from the uniform distribution, Gaussian copula	Assess robustness of the proposed method to misspecification of distribution of structural error
3.	Linear regression model, structural error from the normal distribution, non-Gaussian copulas	Assess robustness of the proposed method to misspecification of the copula
4.	RCL model, individual-level data, structural error from the normal distribution, Gaussian copula	Examine performance of the proposed method in RCL models using individual-level data
5.	RCL model, aggregate-level data, structural error from the normal distribution, Gaussian copula	Examine performance of the proposed method in RCL models using aggregate-level data
6.	Linear regression model, two endogenous regressors	Examine performance of the proposed method when the model has multiple endogenous regressors
7.	Linear regression model, discrete endogenous regressor	Examine performance of the proposed method when an endogenous regressor is discrete
8.	Linear random coefficient regression model, slope endogeneity	Examine performance of the proposed method in handling slope endogeneity problem
9.	Linear regression model, endogeneity based on LIV model	Compare performance of the proposed method with the LIV method

Table 2 Results of the Simulation Study Case 1

<i>T</i>	Parameters	True values	OLS			TSLS			Proposed I			Average bootstrap SE
			Mean	SE	<i>t</i> _{bias}	Mean	SE	<i>t</i> _{bias}	Mean	SE	<i>t</i> _{bias}	
200	α	−1	−0.589	0.112	3.656	−1.010	0.119	0.085	−1.018	0.139	0.130	0.139
	σ_ξ	1	—	—	—	—	—	—	1.012	0.059	0.205	0.058
	ρ_1	0.5	—	—	—	—	—	—	0.514	0.063	0.217	0.062
400	α	−1	−0.571	0.080	5.363	−0.997	0.093	0.033	−1.000	0.098	0.002	0.097
	σ_ξ	1	—	—	—	—	—	—	1.007	0.039	0.175	0.038
	ρ_1	0.5	—	—	—	—	—	—	0.508	0.049	0.156	0.050

Note. Mean and SE are the average and standard deviation of the estimates, respectively.

to obtain the empirical sampling distribution of the parameter estimates.

Estimation results are summarized in Table 2. We first discuss the results in the upper panel based on $T = 200$. As expected, OLS estimates of α are biased ($t_{\text{bias}} = 3.656$ standard errors) toward 0 as a result of the endogeneity problem. Also as expected, TSLS corrects for the bias by using the instrument Z_t and provides a consistent estimate of α . Estimates from methods Proposed I and II are almost identical (confirming their equivalence); hence, we only show Proposed I. Both proposed methods provide consistent estimates without the use of instruments. The standard error of α is 0.139 in the proposed methods, which is 17% larger than the standard error of the TSLS estimate. This reduced efficiency is because TSLS uses the “ideal” instrumental variable, whereas the proposed methods do not use any instruments. The estimates of ρ_1 are tightly distributed near the true value 0.5. Also, the estimates are significantly different from zero, implying significant dependence between ξ_t^* and P_t^* , or significant dependence between ξ_t and P_t subsequently. The last column of Table 2 reports the average of bootstrap standard errors, which for all parameters are very close to the standard errors from the proposed method, confirming that the bootstrap method provides precise standard errors.⁸

The lower panel shows results for the larger sample size $T = 400$. Although essential conclusions are maintained, it is noteworthy that the bias (defined as the absolute deviation of the true parameter value from the estimate; not shown in table) and standard error of both TSLS and Proposed I estimates move toward 0, supporting our claim of consistency.

3.2. Case 2: Misspecification of ξ_t

The proposed methods assume that the structural error ξ_t is normally distributed. Whereas the normality of ξ_t is a reasonable statistical assumption, the true

distribution of ξ_t is unknown to the researcher, thus leaving room for the possibility of misspecification. In this simulation, we examine the robustness of the proposed methods to violations of the normality of ξ_t . To this end, we generate 1,000 data sets using the following DGP:

$$\begin{pmatrix} \xi_t^* \\ P_t^* \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \\ = N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right), \quad (23)$$

$$\xi_t = G^{-1}(U_{\xi,t}) = U_{\xi,t} - 0.5 = \Phi(\xi_t^*) - 0.5, \quad (24)$$

$$P_t = H^{-1}(U_{P,t}) = \Phi_{(1,1)}^{-1}(U_{P,t}) = \Phi_{(1,1)}^{-1}(\Phi(P_t^*)), \quad (25)$$

$$Y_t = \alpha \cdot P_t + \xi_t = -1 \cdot P_t + \xi_t, \quad (26)$$

where $T = 200$ and $\Phi_{(1,1)}^{-1}(\cdot)$ is the inverse distribution function of the normal with mean 1 and variance 1. Note that $\xi_t \sim U(-0.5, 0.5)$. For estimation, we use the OLS estimator along with the proposed methods (Proposed I and Proposed II), all of which assume normality of ξ_t .

Table 3 reports estimation results. As in Case 1, OLS estimates are biased ($t_{\text{bias}} = 5.385$). By contrast, both proposed methods successfully recover the true parameter values despite the misspecification of ξ_t (again, we only show results for Proposed I because these are identical to Proposed II results).⁹ We believe this is because the normal distribution approximates other distributions flexibly.

3.3. Case 3: Misspecification of Copula

In the proposed methods, we use the Gaussian copula to capture the dependence structure between U_p and U_ξ . In real data, the dependence between U_p and U_ξ is a consequence of an economic mechanism (such as marketing strategy decisions); hence, the

⁸ In all simulation cases, we found the same result—namely, that bootstrap standard errors are very close to the empirical standard errors of the proposed method. We provide the bootstrap standard errors for the other cases in Online Appendix 3. Note that the bootstrap standard errors are the same for Proposed I and Proposed II.

⁹ We also specified the triangular distribution and the Student's *t*-distribution for the marginal distribution of the structural error. Simulation results using these distributions echoed what we found with the uniform marginal: the proposed methods provided consistent estimates of the model parameters. Results are available from the authors upon request.

Table 3 Results of the Simulation Study Case 2

Parameters	True values	OLS			Proposed I		
		Mean	SE	t_{bias}	Mean	SE	t_{bias}
α	-1	-0.93	0.013	5.385	-0.999	0.021	0.048
σ_ξ	0.289	—	—	—	0.290	0.014	0.071
ρ	0.5	—	—	—	0.486	0.077	0.182

dependence structure might be different from what the Gaussian copula generates. In this section, we examine the robustness of the Gaussian copula in simulated data. We generate the dependence between U_p and U_ξ using several copula models that are different from the Gaussian. Specifically, we consider the following copula models:

(i) Ali-Mikhail-Haq distribution (with $\theta = 1$):

$$C(U_{p,t}, U_{\xi,t}) = \frac{U_{p,t} U_{\xi,t}}{1 - \theta(1 - U_{p,t})(1 - U_{\xi,t})}.$$

(ii) Plackett distribution (with $\theta = 20$):

$$\begin{aligned} C(U_{p,t}, U_{\xi,t}) &= \left(1 + (\theta - 1)(U_{p,t} + U_{\xi,t})\right. \\ &\quad \left. - \sqrt{[1 + (\theta - 1)(U_{p,t} + U_{\xi,t})]^2 - 4U_{p,t} U_{\xi,t} \theta(\theta - 1)}\right) \\ &\quad \cdot (2(\theta - 1))^{-1}. \end{aligned}$$

(iii) Farlie-Gumbel-Morgenstern distribution (with $\theta = 1$):

$$C(U_{p,t}, U_{\xi,t}) = U_{p,t} U_{\xi,t} (1 + \theta(1 - U_{p,t})(1 - U_{\xi,t})).$$

(iv) Clayton copula (with $\theta = 2$):

$$C(U_{p,t}, U_{\xi,t}) = (U_{p,t}^\theta + U_{\xi,t}^\theta - 1)^{-1/\theta}.$$

(v) Frank copula (with $\theta = 2$):

$$\begin{aligned} C(U_{p,t}, U_{\xi,t}) &= -\frac{1}{\theta} \ln \left(1 + \frac{(\exp(-\theta U_{p,t}) - 1)(\exp(-\theta U_{\xi,t}) - 1)}{\exp(-\theta) - 1} \right). \end{aligned}$$

Figure 1 shows the scatterplots of randomly generated $(U_{p,t}, U_{\xi,t})$ pairs from the above-mentioned copulas, as well as the Gaussian copula with the correlation coefficient 0.5. We can see that each copula generates a unique dependence structure between U_p and U_ξ . After randomly generating $\{U_{p,t}\}$ and $\{U_{\xi,t}\}$ from the above-mentioned copulas, we use the following process to generate $\{P_t\}$ and $\{Y_t\}$:

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

$$P_t = H^{-1}(U_{p,t}) = \Phi_{(1,1)}^{-1}(U_{p,t}), \quad (28)$$

$$Y_t = \alpha \cdot P_t + \xi_t = -1 \cdot P_t + \xi_t. \quad (29)$$

We set $T = 200$. For each copula, we generate 100 data sets and estimate the parameters using the OLS

estimator and Proposed I (Proposed II yields the same results).

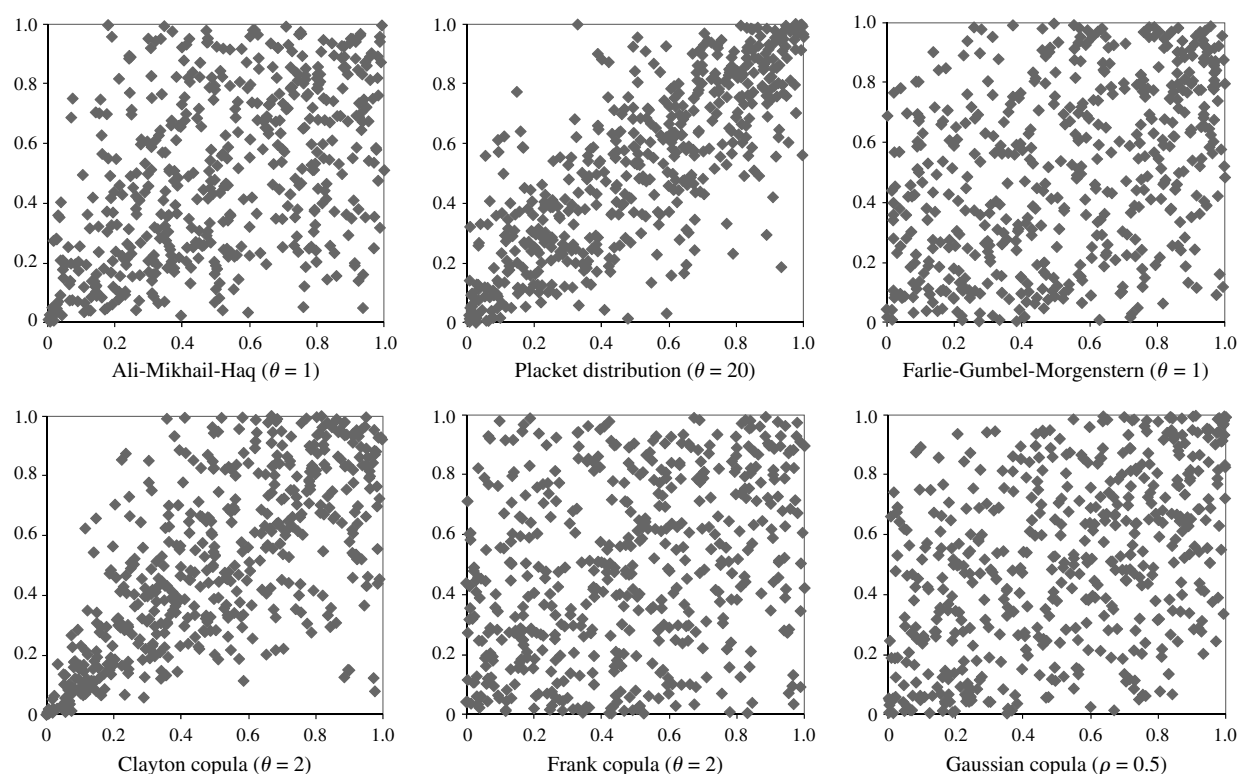
Table 4 reports estimation results. OLS estimates of α are biased in all cases. By contrast, estimates from the proposed model are tightly distributed around the true values, and the bias is not significantly different from 0 in all cases. In summary, we find that the proposed method based on the Gaussian copula is robust to misspecification of the dependence structure between the endogenous regressor and the structural error. Our results echo findings about robustness of the Gaussian copula model that have been documented in the literature (see Song 2000, Danaher and Smith 2011, Balakrishnan and Lai 2009).

3.4. Case 4: RCL Model Using Individual-Level Data

We examine the performance of the proposed method in RCL models using individual-level data. The specific DGP and the parameter values we assign are as follows:

$$\begin{aligned} u_{h0t} &= \varepsilon_{h0t}, \\ u_{h1t} &= \bar{\beta}_1 + \bar{\beta}_3 \cdot X_{1t} + (\bar{\alpha} + a_h) \cdot P_{1t} + \xi_{1t} + \varepsilon_{h1t} \\ &= 0.7 + 1 \cdot X_{1t} + (-1 + a_h) \cdot P_{1t} + \xi_{1t} + \varepsilon_{h1t}, \\ u_{h2t} &= \bar{\beta}_2 + \bar{\beta}_3 \cdot X_{2t} + (\bar{\alpha} + a_h) \cdot P_{2t} + \xi_{2t} + \varepsilon_{h2t} \\ &= 0.7 + 1 \cdot X_{2t} + (-1 + a_h) \cdot P_{2t} + \xi_{2t} + \varepsilon_{h2t}, \\ X_{jt} &= I(Q_{jt} > 0.7), \quad Q_{jt} \sim U(0, 1) \quad \text{for } j = \{1, 2\}, \\ a_h &\sim N(0, \sigma_a^2) = N(0, 0.5^2), \\ \begin{pmatrix} \xi_{jt}^* \\ P_{jt}^* \end{pmatrix} &\sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \\ &= N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right) \quad \text{for } j = \{1, 2\}, \\ \xi_{jt} &= G^{-1}(\Phi(\xi_{jt}^*)) = \Phi_{(0,0.2)}^{-1}(\Phi(\xi_{jt}^*)) \quad \text{for } j = \{1, 2\}, \\ P_{jt} &= H^{-1}(\Phi(P_{jt}^*)) = 1.5 + 0.2 \cdot \ln(-\ln(1 - \Phi(P_{jt}^*))) \\ &\quad \text{for } j = \{1, 2\}, \end{aligned}$$

where $t = 1, \dots, 200$ and $h = 1, \dots, 100$. Note that $H^{-1}(\cdot)$ in this case is the inverse distribution function of the extreme value distribution with mean 1.5 and scale parameter 0.2. So the marginal distribution of P_{jt} is the extreme value distribution. X_{jt} is a regressor that is uncorrelated with ξ_{jt} . Consumers have heterogeneous preferences with respect to P_{jt} only. We generate individual-level choices as described above and use them as data to estimate the model parameters. In the estimation, in addition to the proposed method, we also use a benchmark method where we ignore the endogeneity by applying the OLS estimator to the first-step estimates $\hat{\delta}_{jt}$.

Figure 1 Scatterplots of Randomly Generated $(U_{p,t}, U_{\xi,t})$ Pairs for Considered Copulas

Estimation results are in Table 5. Again, the OLS estimates of three of the five parameters are biased. Estimates from the proposed method are tightly distributed around the true values, and the bias is not significantly different from 0 for all six parameters.

3.5. Case 5: RCL Model Using Aggregate-Level Data

We investigate the performance of the proposed method in RCL models with aggregate-level data. The

DGP is the same as that of Case 4 except that $G^{-1}(\cdot) = \Phi_{(0, 0.52)}^{-1}(\cdot)$, $H = 1,000$, and $T = 500$ and 1,000. To generate the aggregate-level data to be used in estimation, we first generate individual-level choices from the latent utilities and then add them up. We generate 100 data sets and estimate the model using the proposed method. Table 6 summarizes the estimation results. As in the other cases, estimates from the proposed model are closely distributed around the true parameter values and show no significant bias. As T

Table 4 Results of the Simulation Study Case 3

DGP	Parameters	True values	OLS			Proposed I		
			Mean	SE	t_{bias}	Mean	SE	t_{bias}
Ali-Mikhail-Haq	α	-1	-0.750	0.046	5.435	-1.001	0.067	0.015
	α_{ξ}	1	—	—	—	1.006	0.060	0.100
	ρ	0.478	—	—	—	0.501	0.085	0.271
Plackett distribution	α	-1	-0.620	0.043	8.837	-1.004	0.069	0.058
	σ_{ξ}	1	—	—	—	1.013	0.075	0.173
	ρ	0.774	—	—	—	0.760	0.046	0.304
Farlie-Gumbel-Morgenstern	α	-1	-0.839	0.046	3.500	-0.997	0.064	0.047
	σ_{ξ}	1	—	—	—	0.999	0.052	0.019
	ρ	0.337	—	—	—	0.322	0.080	0.188
Clayton copula	α	-1	-0.658	0.041	8.341	-0.999	0.072	0.014
	σ_{ξ}	1	—	—	—	1.004	0.069	0.058
	ρ	0.679	—	—	—	0.684	0.066	0.076
Frank copula	α	-1	-0.849	0.047	3.213	-0.998	0.070	0.029
	σ_{ξ}	1	—	—	—	0.999	0.051	0.020
	ρ	0.317	—	—	—	0.303	0.081	0.173

Table 5 Results of the Simulation Study Case 4

Parameters	True values	OLS			Proposed II		
		Mean	SE	t_{bias}	Mean	SE	t_{bias}
$\bar{\beta}_1$	0.7	0.178	0.084	6.214	0.727	0.323	0.084
$\bar{\beta}_2$	0.7	0.181	0.083	6.253	0.728	0.324	0.086
$\bar{\beta}_3$	1	1.004	0.030	0.133	1.004	0.030	0.133
$\bar{\alpha}$	−1	−0.627	0.057	6.544	−1.044	0.210	0.210
σ_a	0.5	0.527	0.088	0.307	0.527	0.088	0.307
$\sigma_\xi \cdot \rho$	0.1	—	—	—	0.112	0.050	0.240

Notes. σ_a is estimated in the first step, and thus OLS and Proposed II have the same estimates. We report results of Proposed II. We obtained the same results from Proposed I.

increases from 500 to 1,000, the bias (defined as the absolute deviation of true parameter value from the estimate) of all estimates moves toward 0 (one exception is the bias of $\bar{\alpha}$, which increases slightly from 0.008 to 0.010). This observation supports the consistency of the proposed method.

3.6. Case 6: Multiple Endogenous Regressors

We examine the performance of the proposed method when the model has multiple endogenous regressors. To this end, we generate the data using the following DGP:

$$Y_t = \alpha_1 \cdot P_{1,t} + \alpha_2 \cdot P_{2,t} + \xi_t = -1 \cdot P_{1,t} + 1 \cdot P_{2,t} + \xi_t,$$

$$\begin{pmatrix} \xi_t^* \\ P_{1,t}^* \\ P_{2,t}^* \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{\xi 1} & \rho_{\xi 2} \\ \rho_{\xi 1} & 1 & \rho_{12} \\ \rho_{\xi 2} & \rho_{12} & 1 \end{bmatrix} \right) \\ = N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.1 \\ 0.5 & 0.1 & 1 \end{bmatrix} \right),$$

$$\xi_t = G^{-1}(\Phi(\xi_t^*)) = \Phi_{(0, 0.5^2)}^{-1}(\Phi(\xi_t^*)),$$

$$P_{1,t} = H_1^{-1}(\Phi(P_{1,t}^*)) = 1.5 + 0.2 \cdot \ln(-\ln(1 - \Phi(P_{1,t}^*))),$$

$$P_{2,t} = H_2^{-1}(\Phi(P_{2,t}^*)) = 1.0 + 0.3 \cdot \ln(-\ln(1 - \Phi(P_{2,t}^*))),$$

where $T = 500$. H_1^{-1} and H_2^{-1} are the inverse distribution functions of the extreme value distributions.

Table 6 Results of the Simulation Study Case 5

Parameters	True values	Proposed method					
		$T = 500$			$T = 1,000$		
		Mean	SE	t_{bias}	Mean	SE	t_{bias}
$\bar{\beta}_1$	0.7	0.698	0.328	0.006	0.702	0.219	0.009
$\bar{\beta}_2$	0.7	0.689	0.328	0.034	0.702	0.217	0.009
$\bar{\beta}_3$	1	0.977	0.054	0.426	0.990	0.042	0.238
$\bar{\alpha}$	−1	−0.992	0.235	0.034	−1.010	0.157	0.064
σ_a	0.5	0.489	0.108	0.102	0.492	0.088	0.091
$\sigma_\xi \cdot \rho$	0.25	0.261	0.077	0.143	0.259	0.044	0.205
$\sigma_\xi \cdot (1 - \rho^2)^{1/2}$	0.433	0.443	0.021	0.476	0.440	0.016	0.438

So the marginal distributions of $P_{1,t}$ and $P_{2,t}$ are extreme value distributions. Note that both $P_{1,t}$ and $P_{2,t}$ are correlated with ξ_t . We generate 1,000 data sets and estimate the model using the proposed method. Table 7 reports estimation results. The OLS estimates of both parameters are biased, whereas estimates of all five parameters from the proposed method are unbiased, suggesting that the proposed method provides consistent estimates when the model has two endogenous regressors.

3.7. Case 7: Discrete Endogenous Regressor

We investigate the performance of the proposed method when the endogenous regressor is a discrete variable. To this end, we consider four different discrete distributions. We generated the data using the following DGP:

$$Y_t = \alpha \cdot P_t + \xi_t = -1 \cdot P_t + \xi_t$$

$$\begin{pmatrix} \xi_t^* \\ P_t^* \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) = N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \right),$$

$$\xi_t = G^{-1}(\Phi(\xi_t^*)) = \Phi_{(0, \sigma_\xi^2)}^{-1}(\Phi(\xi_t^*)),$$

$$\Phi^{-1}(H(P_t - 1)) < P_t^* < \Phi^{-1}(H(P_t)).$$

(i) Discrete 1 (Bernoulli distribution):

$$H(P_t = 0) = 0.5, \quad H(P_t = 1) = 1, \quad \sigma_\xi^2 = 0.25.$$

(ii) Discrete 2:

$$H(P_t = 0) = 0.2, \quad H(P_t = 1) = 0.4,$$

$$H(P_t = 2) = 0.6, \quad H(P_t = 3) = 0.8,$$

$$H(P_t = 4) = 1; \quad \sigma_\xi^2 = 2.$$

(iii) Discrete 3 (Poisson distribution):

$$H(P_t) = \exp(-4) \sum_{i=0}^{P_t} \frac{4^i}{i!}, \quad \sigma_\xi^2 = 4.$$

(iv) Discrete 4 (Negative binomial distribution):

$$H(P_t) = \sum_{i=0}^{P_t} \binom{i+4-1}{i} \times 0.5^4 \times 0.5^i, \quad \sigma_\xi^2 = 8.$$

We set $T = 200$. Note that σ_ξ^2 is set to be equal to the variance of P_t in all cases. For each discrete distribution, we generate 1,000 data sets and estimate the parameters using the OLS estimator and the proposed method. Table 8 summarizes the results.

When the endogenous regressor is generated from the Bernoulli distribution (Discrete 1), the average OLS estimate of α is -0.801 (SE, 0.047), and the average of the proposed method is -0.895 (SE, 0.056). Both

Table 7 Results of the Simulation Study Case 6

Parameters	True values	OLS			Proposed method		
		Mean	SE	t_{bias}	Mean	SE	t_{bias}
α_1	−1	−1.174	0.031	5.613	−0.985	0.117	0.128
α_2	1	1.354	0.05	7.080	0.978	0.194	0.113
$\sigma_\xi \cdot \frac{\rho_{\xi 1} - \rho_{12}\rho_{\xi 2}}{1 - \rho_{12}^2}$	0.227	—	—	—	0.225	0.036	0.056
$\sigma_\xi \cdot \frac{\rho_{\xi 2} - \rho_{12}\rho_{\xi 1}}{1 - \rho_{12}^2}$	0.227	—	—	—	0.236	0.077	0.117
$\sigma_\xi \cdot \left(1 - \rho_{\xi 1}^2 - \frac{(\rho_{\xi 2} - \rho_{12}\rho_{\xi 1})^2}{1 - \rho_{12}^2}\right)^{0.5}$	0.369	—	—	—	0.366	0.013	0.231

estimators are biased, although the estimates from the proposed method are distributed nearer to the true value. As noted previously, this is because the support of the Bernoulli distribution is only $\{0, 1\}$, and the information contained in this variable is not rich enough for the proposed method to handle the endogeneity problem properly. In Discrete 2, the support of the endogenous regressor expands to $\{0, 1, 2, 3, 4\}$, providing richer information. We observe that the OLS estimator is still biased, but the proposed method provides unbiased estimates for all three parameters. We find similar results in Discrete 3 (Poisson) and Discrete 4 (Negative binomial). To summarize, the proposed method shows good performance with discrete endogenous regressors if the support of the regressor is reasonably large.

3.8. Case 8: Slope Endogeneity

We examine the performance of the proposed method in handling the slope endogeneity problem. To this end, we generated the data using the following DGP:

$$Y_t = P_t(\bar{\xi} + \xi_t) + \eta_t = P_t(1 + \xi_t) + \eta_t, \\ \eta_t \sim N(0, \sigma_\eta^2) = N(0, 1),$$

$$\begin{pmatrix} \xi_t^* \\ P_t^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right),$$

$$\xi_t = G^{-1}(\Phi(\xi_t^*)) = \Phi_{(0,1)}^{-1}(\Phi(\xi_t^*)),$$

$$P_t = H^{-1}(\Phi(P_t^*)) = \Phi_{(1,1)}^{-1}(\Phi(P_t^*)).$$

We generate 1,000 data sets with $T = 300$ and estimate the parameters ignoring the slope endogeneity and using the proposed method. To handle the random coefficient, we use numerical integration using 500 random draws. Table 9 summarizes the results. When we ignore the slope endogeneity, the estimate of $\bar{\xi}$ is biased (overstated). All estimates from the proposed method are unbiased.

3.9. Case 9: Comparison with LIV

The goal of this simulation is to compare the proposed copula method with the most similar extant instrument-free method, the LIV method. To briefly describe the LIV method, we reproduce Equations (1) and (2) from §2.1:

$$Y_t = X_t'\beta + P_t\alpha + \xi_t, \\ P_t = Z_t'\gamma + \nu_t.$$

Table 8 Results of the Simulation Study Case 7

DGP	Parameters	True values	OLS			Proposed method		
			Mean	SE	t_{bias}	Mean	SE	t_{bias}
Discrete 1 (Bernoulli)	α	−1	−0.801	0.047	4.234	−0.895	0.056	1.875
	$\sigma_\xi \cdot \rho$	0.25	—	—	—	0.117	0.04	3.325
	$\sigma_\xi \cdot \sqrt{1 - \rho^2}$	0.433	—	—	—	0.468	0.024	1.458
Discrete 2	α	−1	−0.84	0.036	4.361	−0.98	0.046	0.478
	$\sigma_\xi \cdot \rho$	0.707	—	—	—	0.606	0.116	0.871
	$\sigma_\xi \cdot \sqrt{1 - \rho^2}$	1.225	—	—	—	1.259	0.061	0.557
Discrete 3 (Poisson)	α	−1	−0.9	0.031	3.097	−1	0.036	0.056
	$\sigma_\xi \cdot \rho$	1	—	—	—	0.958	0.156	0.269
	$\sigma_\xi \cdot \sqrt{1 - \rho^2}$	1.732	—	—	—	1.738	0.085	0.071
Discrete 4 (Negative binomial)	α	−1	−0.84	0.036	4.444	−0.99	0.047	0.149
	$\sigma_\xi \cdot \rho$	1.414	—	—	—	1.352	0.231	0.268
	$\sigma_\xi \cdot \sqrt{1 - \rho^2}$	2.449	—	—	—	2.45	0.131	0.008

Table 9 Results of the Simulation Study Case 8

Parameters	True values	Slope endogeneity ignored			Proposed method		
		Mean	SE	t_{bias}	Mean	SE	t_{bias}
$\bar{\xi}$	1	1.254	0.097	2.619	1.003	0.101	0.030
σ_{ξ}	1	0.976	0.102	0.235	—	—	—
$\sigma_{\xi} \cdot \rho$	0.5	—	—	—	0.493	0.085	0.082
$\sigma_{\xi} \sqrt{1 - \rho^2}$	0.866	—	—	—	0.846	0.095	0.211
σ_{η}	1	1.027	0.084	0.321	1.019	0.078	0.244

The LIV method models $Z_t'\gamma$ as a discrete, exogenous random variable with an unknown number of possible outcomes; ν_t is an endogenous error and is usually assumed to be Gaussian. The endogeneity issue is handled by modeling the correlation between ξ_t and ν_t . The finite mixture approach is used to estimate the model.

Note that the LIV method is based on two key assumptions: $Z_t'\gamma$ is a discrete variable and $Z_t'\gamma$ is exogenous. Based on simulations, Ebbes et al. (2009) reported that the model successfully approximated some continuous $Z_t'\gamma$; hence, the first assumption may not be overly restrictive. The second assumption or the exogeneity of $Z_t'\gamma$ is similar to the exclusion restriction in usual IV models. In the LIV method, we do not observe Z_t , and $Z_t'\gamma$ is estimated using data. Therefore it is difficult to justify or test the exogeneity of $Z_t'\gamma$.

We generate data from an LIV model and compare the estimation performance of the copula and LIV methods. The specific DGP is as follows:

$$\begin{pmatrix} \xi_t \\ \nu_t \\ M_t \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & \rho \\ 0.5 & 1 & 0 \\ \rho & 0 & 1 \end{bmatrix} \right),$$

$$Z_t = \begin{cases} 0 & \text{if } M_t < 0, \\ 4 & \text{if } M_t \geq 0, \end{cases}$$

$$P_t = Z_t + \nu_t,$$

$$Y_t = P_t \alpha + \xi_t = P_t \cdot 1 + \xi_t.$$

We consider two different values for ρ : $\rho = 0$ and $\rho = 0.5$. When $\rho = 0$, the latent instrument Z_t is exogenous. In this case, the LIV method is correctly specified but the proposed method is not because the correlation between P_t and ξ_t is different from what the Gaussian copula assumes. When $\rho = 0.5$, the LIV method is misspecified because Z_t is correlated with ξ_t . The proposed method continues to be misspecified because the correlation between P_t and ξ_t is different from what the Gaussian copula assumes. In both cases, we generate 1,000 data sets with $T=300$. We estimate the LIV model assuming that the latent instrument has two categories. Table 10 shows the

Table 10 Results of the Simulation Study Case 9

DGP	Method	True value	Estimate	SE	t_{bias}
$\rho = 0$	OLS	1	1.058	0.022	2.668
	LIV	1	1.002	0.022	0.094
	Proposed	1	0.973	0.033	-0.819
$\rho = 0.5$	OLS	1	1.144	0.015	9.596
	LIV	1	1.089	0.017	5.396
	Proposed	1	0.992	0.030	-0.280

Note. The estimation results for other model parameters are available from the authors on request.

performance of OLS, LIV, and the proposed copula methods in estimating α . When $\rho = 0$, as expected, the OLS estimate is biased, and the LIV estimate is unbiased. Importantly, the proposed method yields unbiased estimates despite misspecification. When $\rho = 0.5$, the OLS estimate continues to be biased. Importantly, the LIV estimate is also biased, whereas the estimate from the copula model is unbiased, supporting the flexibility of the proposed approach.

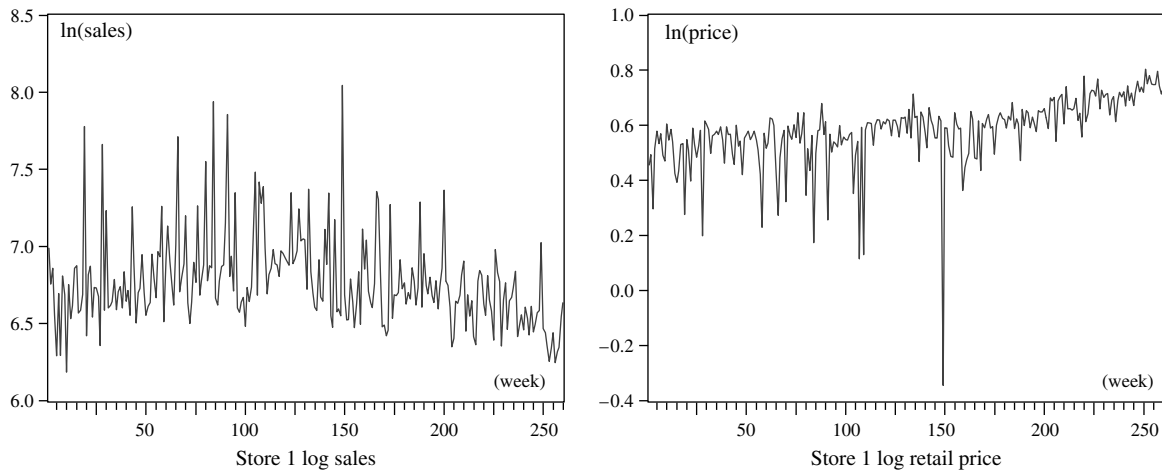
4. Empirical Application

We illustrate empirical application of the proposed method to handle price endogeneity in demand estimation using data on store-level sales of paper towels (category sales) in Eau Claire, Wisconsin over 260 weeks from 2001 to 2005.¹⁰ The argument for endogeneity of retail price is that there are unmeasured product characteristics, or demand shocks, that influence not only consumer decisions but also retailer pricing decisions. Because they are unobserved by researchers, these variables are incorporated into the structural error, leading to the endogeneity problem. We focus on category sales in the two largest independent stores in the market (we refer to them as Stores 1 and 2). Retail price is defined on a per-roll basis. Retail price, in-store display, and feature advertising at the category level are computed as market share-weighted averages of UPC-level variables.

We use the TSLS estimator as a benchmark to assess the validity of the proposed method. To implement the TSLS model, we use retail price at the other store as an instrument (that is, Store 2 price is used as instrument for Store 1 price, and vice versa). Retail prices across stores in a market are likely to be highly correlated because competing retailers are typically offered the same wholesale prices for a brand. The Pearson correlation between the retail prices (category average) of paper towels at Stores 1 and 2 is 0.83. However, unmeasured product characteristics, especially those determined at retail (e.g., shelf-space allocation, shelf

¹⁰ We obtained the data from the IRI marketing data set (Bronnenberg et al. 2008).

Figure 2 Log Sales and Log Retail Prices of Paper Towel in Store 1



location), are not likely to be systematically related to wholesale prices. If this is true, the instrumental variable is exogenous and thus is valid for controlling for the endogeneity of retail price.

We estimate the following model:

$$\log(\text{Sales}_t) = \log(\text{Retail Price}_t) \cdot \alpha + X_t' \beta + \xi_t,$$

where the vector X_t includes a constant, three dummy variables to represent quarters (Q2, Q3, and Q4), feature, and display. Figure 2 shows log sales and log retail prices at store 1 (Store 2 is very similar). The log retail price shows an increasing trend over the five-year sample period. Therefore, we use detrended log retail prices for estimation and also as instrumental variables. We estimate the model using the OLS estimator and also apply the two methods that correct for price endogeneity—the TSLS estimator and the proposed method. As we explained in §2, for the proposed approach, we first estimate the marginal distributions of prices using a nonparametric method and then use these distributions to construct the likelihood functions. Figure 3 shows the nonparametric densities of detrended log retail prices

at the two stores. We use the Epanechnikov kernel for the nonparametric density estimation. Following Silverman's (1986) data-driven bandwidth selection rule, we set the bandwidth parameters to be equal to 0.036 and 0.039 for Stores 1 and 2, respectively. We can observe that both distributions are left-skewed and quite different from the normal distribution. Therefore, we expect that this difference will provide sufficient information for model identification.

Table 11 reports estimation results for the three methods. Beginning with the results from Store 1, by comparing OLS estimates with TSLS estimates, we can verify whether prices are endogenous. The price coefficients are substantially different, supporting the case for retail price endogeneity. Also, the Hausman test for endogeneity (Wooldridge 2001, p. 118) confirms that the retail price is endogenous. Instrumenting for retail price changes its coefficient estimate from -0.676 to -0.928 , implying that there is a positive correlation between unmeasured product characteristics and the price. This direction of correlation is consistent with previous empirical findings (e.g., Villas-Boas and Winer 1999, Chintagunta et al. 2005). Estimates from

Figure 3 Estimated Nonparametric Densities of Detrended Log Retail Prices

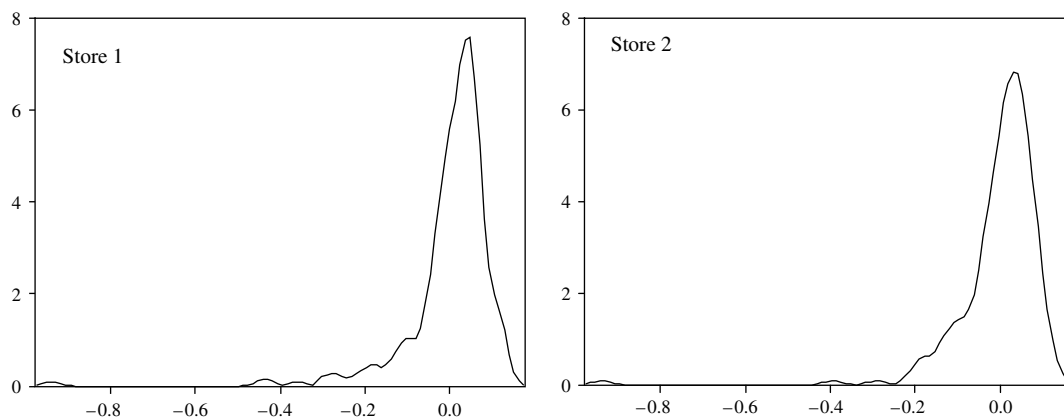


Table 11 Estimation Results for Three Methods: Paper Towel Sales

Parameters	Store 1			Store 2		
	Estimate	SE	t-value	Estimate	SE	t-value
OLS						
Constant	6.607	0.031	213.9	6.549	0.024	270.7
Q2	0.094	0.033	2.8	0.089	0.027	3.3
Q3	0.055	0.033	1.7	0.116	0.027	4.3
Q4	−0.067	0.033	−2.0	0.060	0.027	2.2
Price	−0.676	0.149	−4.5	−0.780	0.125	−6.3
Promotion	0.407	0.041	9.8	0.433	0.030	14.4
Display	0.173	0.080	2.2	0.158	0.061	2.6
TSLS						
Constant	6.613	0.030	217.0	6.553	0.025	266.1
Q2	0.078	0.033	2.3	0.081	0.028	2.9
Q3	0.042	0.033	1.3	0.110	0.027	4.0
Q4	−0.071	0.032	−2.2	0.057	0.028	2.1
Price	−0.928	0.170	−5.4	−0.653	0.153	−4.3
Promotion	0.428	0.041	10.3	0.429	0.030	14.1
Display	0.159	0.079	2.0	0.164	0.062	2.6
Proposed						
Constant	6.600	0.035	212.4	6.549	0.021	212.4
Q2	0.099	0.036	3.0	0.090	0.027	3.0
Q3	0.056	0.033	1.7	0.115	0.024	1.7
Q4	−0.066	0.032	−2.0	0.059	0.030	−2.0
Price	−0.934	0.240	−4.1	−0.883	0.236	−4.1
Promotion	0.395	0.043	9.4	0.430	0.038	9.4
Display	0.209	0.080	2.5	0.163	0.070	2.5
ρ	0.197	0.110	1.8	0.080	0.134	0.6

the proposed method are very close to those from TSLS. The estimate of ρ is significant and positive, consistent with the TSLS finding of positive correlation between unmeasured product characteristics and the retail price. As we have observed in the simulation studies, the standard errors of the proposed method are slightly larger than those of TSLS but small enough for use to draw meaningful conclusion on model parameters. This implies that there is no problem in model identification in this case.

Unlike Store 1, the results from Store 2 provide multiple convergent indications that the retail price is exogenous. The OLS estimates are not significantly different from the TSLS estimates, the Hausman test of endogeneity confirms the conclusion, the estimates from the proposed method are not significantly different from the OLS estimates, and finally, the estimated value of ρ is not significantly different from 0.

We interpret our results to mean that retailer pricing decisions are sensitive to changes in unmeasured product characteristics in Store 1 but less sensitive in Store 2. The convergence of results between TSLS and the proposed method in both stores lends validity to the proposed method to the extent that our assumption regarding exogeneity of the instruments is true.

5. Conclusion

We propose a new instrument-free method to handle the problem of endogenous regressors. Our method uses a Gaussian copula to model the joint distribution

of the endogenous regressor and the structural error term, and it makes inferences on model parameters by maximizing the resulting likelihood function. We use flexible nonparametric density estimation to construct the marginal distribution of the endogenous regressor. In simulation studies we find that the proposed method provides consistent estimates not only when the actual correlation structure between the regressor and structural error follows what the Gaussian copula assumes but also when the correlation structure is different from what the Gaussian copula assumes. This finding supports robustness of the proposed method. We show how the proposed method can be applied to three popular models in marketing and economics—the linear regression model, the RCL model using individual-level data, and the RCL model using aggregate-level data. Also, we show how to extend the proposed method to (1) multiple endogenous regressors, (2) discrete endogenous regressors, and (3) slope endogeneity problems. In an empirical application to paper towel sales, we again find that the proposed method successfully overcomes the endogeneity bias.

The proposed method is not without limitations. The method exploits the information contained in the endogenous regressor for identification. For the proposed method to work properly, the distribution of the endogenous regressor must contain adequate information *and* must be different from that of the structural error (assumed to be normal). The former condition is violated when the endogenous regressor follows the Bernoulli distribution, and the latter is violated when the endogenous regressor follows the normal distribution. However, despite these limitations, we expect that the proposed method will provide a useful alternative in many empirical problems in which instruments are not available.

Electronic Companion

An electronic companion to this paper is available as part of the online version at <http://dx.doi.org/10.1287/mksc.1120.0718>.

Acknowledgments

The authors acknowledge helpful comments and suggestions from Peter Ebbes, Duk Bin Jun, and Chang-Jin Kim; the seminar participants at Korea University, Korea Advanced Institute of Science and Technology, and the 2011 Marketing Science Conference; and the editor, area editor, and two reviewers. The authors thank Symphony IRI for making the data available. All estimates or analyses in this paper based on Symphony IRI Group, Inc., data are by the authors and not Symphony IRI Group, Inc.

References

- Angrist JD, Krueger AB (1991) Does compulsory school attendance affect schooling and earnings? *Quart. J. Econom.* 106(4):979–1014.

- Angrist JD, Krueger AB (2001) Instrumental variables and the search for identification: From supply and demand to natural experiments. *J. Econom. Perspect.* 15(4):69–85.
- Balakrishnan N, Lai C-D (2009) *Continuous Bivariate Distributions*, 2nd ed. (Springer, New York).
- Berry ST (1994) Estimating discrete-choice models of product differentiation. *RAND J. Econom.* 25(2):242–262.
- Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica* 63(4):841–890.
- Besanko D, Gupta S, Jain D (1998) Logit demand estimation under competitive pricing behavior: An equilibrium framework. *Management Sci.* 44(11, Part 1):1533–1547.
- Bound J, Jaeger D, Baker R (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J. Amer. Statist. Assoc.* 90(430):443–450.
- Bronnenberg BJ, Mahajan V (2001) Unobserved retailer behavior in multimarket data: Joint spatial dependence in market shares and promotion variables. *Marketing Sci.* 20(3):284–299.
- Bronnenberg BJ, Kruger MW, Mela CF (2008) The IRI marketing data set. *Marketing Sci.* 27(4):745–748.
- Chao JC, Phillips PCB (1998) Posterior distributions in limited information analysis of the simultaneous equations model using the Jeffreys prior. *J. Econometrics* 87(1):49–86.
- Chintagunta PK (2001) Endogeneity and heterogeneity in a probit demand model: Estimation using aggregate data. *Marketing Sci.* 20(4):442–456.
- Chintagunta PK, Dubé J-P, Goh KY (2005) Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice model. *Management Sci.* 51(5):832–849.
- Chintagunta PK, Erdem T, Rossi PE, Wedel M (2006) Structural modeling in marketing: Review and assessment. *Marketing Sci.* 25(6):604–616.
- Conley T, Hansen C, Rossi PE (2012) Plausibly exogenous. *Rev. Econom. Statist.* 94(1):260–272.
- Danaher PJ, Smith MS (2011) Modeling multivariate distributions using copulas: Applications in marketing. *Marketing Sci.* 30(1):4–21.
- DiNardo J, Tobias JL (2001) Nonparametric density and regression estimation. *J. Econom. Perspect.* 15(4):11–28.
- Dotson JP, Allenby GM (2010) Investigating the strategic influence of customer and employee satisfaction on firm financial performance. *Marketing Sci.* 29(5):895–908.
- Ebbes P, Wedel M, Böckenholt U (2009) Frugal IV alternatives to identify the parameter for an endogenous regressors. *J. Appl. Econometrics* 24(3):446–468.
- Ebbes P, Wedel M, Böckenholt U, Steerneman AGM (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.
- Efron B (1979) Bootstrap methods: Another look at the jackknife. *Ann. Statist.* 7(1):1–26.
- Erickson T, Whited TM (2002) Two-step GMM estimation of the errors-in-variables model using high-order moments. *Econometric Theory* 18(3):776–799.
- Genest C, Ghoudi K, Rivest LP (1995) A semiparametric estimation procedure for dependence parameters in multivariate families of distributions. *Biometrika* 82(3):543–552.
- Hartmann W, Nair HS, Narayanan S (2011) Identifying causal marketing mix effects using a regression discontinuity design. *Marketing Sci.* 30(6):1079–1097.
- Hausman JA (1978) Specification tests in econometrics. *Econometrica* 46(6):1251–1272.
- Heckman JJ (1976) Common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Ann. Econom. Soc. Measurement* 5(4):475–492.
- Heckman JJ (1978) Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46(4):931–959.
- Hogan V, Rigobon R (2003) Using unobserved supply shocks to estimate the returns to education. Technical report, University College Dublin, Dublin, Ireland.
- Keane MP (1993) Simulation estimation for panel data models with limited dependent variables. Rao CR, Maddala GS, Vinod D, eds. *The Handbook of Statistics and Econometrics* (Elsevier Science, Amsterdam), 545–572.
- Kleibergen F, Zivot E (2003) Bayesian and classical approaches to instrumental variable regression. *J. Econometrics* 114(1):29–72.
- Kuksov D, Villas-Boas JM (2008) Endogeneity and individual consumer choice. *J. Marketing Res.* 45(6):702–714.
- Lewbel A (1997) Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica* 65(5):1201–1213.
- Luan Y, Sudhir JK (2010) Forecasting marketing-mix responsiveness for new products. *J. Marketing Res.* 47(3):444–457.
- Manchanda P, Rossi PE, Chintagunta PK (2004) Response modeling with nonrandom marketing-mix variables. *J. Marketing Res.* 41(November):467–478.
- Mooney CZ, Duval RD (1993) *Bootstrapping: A Nonparametric Approach to Statistical Inference* (Sage, Newbury Park, CA).
- Musalem A, Bradlow ET, Raju JS (2009) Bayesian estimation of random-coefficients choice models using aggregate data. *J. Appl. Econometrics* 24(3):490–516.
- Nelsen RB (2006) *An Introduction to Copulas*, 2nd ed. (Springer, New York).
- Otter T, Gilbride T, Allenby GM (2011) Testing models of strategic behavior characterized by conditional likelihoods. *Marketing Sci.* 30(4):686–701.
- Pagan A (1984) Econometric issues in the analysis of regressions with generated regressors. *Internat. Econom. Rev.* 25(1):221–247.
- Park S, Gupta S (2009) Simulated maximum likelihood estimator for the random coefficient logit model using aggregate data. *J. Marketing Res.* 46(August):531–542.
- Rigobon R (2003) Identification through heteroskedasticity. *Rev. Econom. Statist.* 85(4):777–792.
- Rossi PE, Allenby GM, McCulloch R (2005) *Bayesian Statistics and Marketing* (John Wiley & Sons, New York).
- Rust RT (1986) *Advertising Media Models: A Practical Guide* (Lexington Books, Lexington, MA).
- Rutz OJ, Trusov M (2011) Zooming in on paid search ads—An individual-level model calibrated on aggregated data. *Marketing Sci.* 30(5):789–800.
- Rutz OJ, Bucklin RE, Sonnier GP (2012) A latent instrumental variables approach to modeling keyword conversion in paid search advertising. *J. Marketing Res.* 49(3):306–319.
- Shih JH, Loius TA (1995) Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* 51(4):1384–1399.
- Silverman BW (1986) *Density Estimation for Statistics and Data Analysis* (Chapman & Hall, London).
- Sklar A (1959) Fonctions de répartition n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* 8:229–231.
- Song PX-K (2000) Multivariate dispersion models generated from gaussian copula. *Scand. J. Statist.* 27(2):305–320.

- Sudhir K (2001) Competitive pricing behavior in the US auto market: A structural analysis. *Marketing Sci.* 20(1):42–60.
- van Dijk A, van Heerde HJ, Leeflang PSH, Wittink DR (2004) Similarity-based spatial methods to estimate shelf space elasticities. *Quant. Marketing Econom.* 2(3):257–277.
- Villas-Boas JM, Winer R (1999) Endogeneity in brand choice models. *Management Sci.* 45(10):1324–1338.
- Wooldridge JM (2001) *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Cambridge, MA).
- Yang S, Chen Y, Allenby GM (2003) Bayesian analysis of simultaneous demand and supply. *Quant. Marketing Econom.* 1(3):251–275.
- Zhang J, Wedel M, Pieters R (2009) Sales effects of visual attention to feature ads: A Bayesian mediation analysis. *J. Marketing Res.* 46(3):669–681.