NBER WORKING PAPER SERIES

HANDLING ENDOGENOUS MARKETING MIX REGRESSORS IN CORRELATED HETEROGENEOUS PANELS WITH COPULA AUGMENTED MEAN GROUP ESTIMATION

Liying Yang
Yi Qian
Hui Xie

Working Paper 33265
http://www.nber.org/papers/w33265

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

We acknowledge the support by Social Sciences and Humanities Research Council of Canada [grants 435-2018-0519 and 435-2023-0306], Natural Sciences and Engineering Research Council of Canada [grant RGPIN-2018-04313 and RGPIN-2024-06629] and US National Institute of Health [grant R01CA178061]. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Handling Endogenous Marketing Mix Regressors in Correlated Heterogeneous Panels with Copula Augmented Mean Group Estimation
Liying Yang, Yi Qian, and Hui Xie
NBER Working Paper No. 33265
December 2024
JEL No. C10, C5, L16, L2

ABSTRACT

Endogeneity is a primary concern when evaluating causal effects using observational panel data. While unit-specific intercepts control for unobserved time-invariant confounders, dependence between (i) regressors (e.g., marketing mix strategy of interests) and the current error term (regressor endogeneity) and/or between (ii) regressors and heterogeneous slope coefficients (slope endogeneity) can introduce significant estimation bias, resulting in misleading inference. This paper proposes a two-stage copula endogeneity correction mean group (2sCOPE-MG) estimator for panel data models, simultaneously addressing both endogeneity concerns. We generalize the IV-free copula control function, employing a general location Gaussian copula that effectively captures the panel structure. The heterogeneous coefficients are treated as unit-specific fixed parameters without distributional assumptions. Consequently, the 2sCOPE-MG estimator allows for arbitrary dependence structure between heterogeneous coefficients and regressors. Unlike Haschka (2022), 2sCOPE-MG requires neither a normal error distribution nor a Gaussian copula regressor-error dependence structure and is more robust, easier to implement, and capable of addressing slope endogeneity. The 2sCOPE-MG estimator is extended to dynamic panels, where intertemporal dependence in the outcome process can be suitably captured. We study its asymptotic properties and provide an analytical variance formula for inference without the need to bootstrap. For short dynamic panels, a Jackknife bias-corrected 2sCOPE-MG estimator is provided to ensure unbiased inference. The usage of the 2sCOPE-MG estimator is demonstrated by Monte Carlo simulations and a marketing mix response application across 21 categories to account for regressor and slope endogeneities in store-panel sales data.

Liying Yang
Sauder School of Business
University of British Columbia
liying.yang@sauder.ubc.ca

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

Hui Xie
Department of Biostatistics
Faculty of Health Sciences
Simon Fraser University
huixie@uic.edu

1 Introduction

With technology advancement, high-quality panel data are now widely available in marketing and other related fields. For example, population-level administrative data are routinely captured by firms, stores, web platforms, hospitals, governments, and organizations as part of their daily operations. Particularly, scanner panel data provide detailed and high-fidelity real-life purchase data, capturing individual purchase behaviors linked with concomitant marketing mix variables in real time. These high-quality panel datasets offer tremendous opportunities to infer causal relationships among relevant variables, facilitating optimal decision-making (e.g., setting prices or advertisement budgets to maximize profits). However, except in the case of experimentally generated data (e.g., field or conjoint experiments), these data sets are often observational that complicates causal inference.

Researchers and practitioners often encounter two prominent challenges that plague standard panel regression analysis based on observational data. The first challenge is the potential regressor-error dependence. Frequently, the regressors of primary interests such as marketing mix variables, are endogenous and set by managers based on relevant demand shocks (e.g., certain product characteristics) unobserved to data analysts. Such unmeasured confounders generate dependence between these key regressors and the error term in the panel regression model, which can cause severely biased effect estimates of these marketing mix variables if ignored. In a meta-analysis, Bijmolt et al. (2005) found substantial differences in price elasticity estimates, depending on whether price endogeneity is accounted for. By including unit-specific intercepts in the model, the fixed-effects (FE) approach can eliminate confounding effects of time-constant unobservables. However, dependence between the regressors and the current error term can persist and lead to inconsistent estimation because of time-varying unmeasured confounders. Examples of such confounders are consumer tastes or unobserved brand attributes evolving over time when modeling consumer choices (Chintagunta et al., 2005), and purchase intent when studying effects of Internet advertising (Hoban and Bucklin, 2015; Blake et al., 2015). Such endogeneity also arises with correlated measurement errors in variables, simultaneity, or reverse causality. We name it "contemporaneous regressor endogeneity."

The second challenge is the potential regressor-coefficient dependence. The values of marketing mix regressors observed in historical data can be set by managers who possess private information about marketing response coefficients. For example, store managers may charge higher prices in markets with lower price elasticity. Broadly speaking, the correlations between heterogeneous responses coefficients and marketing mix regressors are naturally induced by targeted promotions and advertisements (Manchanda et al. 2004, Luan and Sudhir 2010, Goldfarb and Tucker 2011, Blake et al. 2015, Hoban and Bucklin 2015, Esteves and Resende 2016) and behavior-based pricing (Li, 2018).[1] Neglecting such regressor-coefficient correlations can yield biased estimates for the average effects (i.e., the mean of response coefficients), as shown in Wooldridge (2005) and Pesaran and Yang (2024). Hence, we term it as "slope endogeneity" (Luan and Sudhir, 2010).

This paper introduces a copula control function approach to estimating the average effects of potentially endogenous regressors in heterogeneous panel data models with correlated random coefficients. A two-stage copula endogeneity corrected mean group (2sCOPE-MG) estimator is proposed, simultaneously addressing both concerns of contemporaneous regressor endogeneity and slope endogeneity without using instrumental variables (IVs). Extending the two-stage copula endogeneity correction approach in Yang et al. (2024) to the panel data setting, we capture the dependence between the current error term and endogenous regressors using a general location Gaussian copula that employs nonparametric marginals and

[1] Game-theoretical models are employed to examine the impacts of targeted advertisements (Iyer et al., 2005) and behavior-based pricing (Li, 2018) on equilibrium sales, profits, etc. In these models, consumers are targeted based on their purchase history or attributes, where the impact of consumer heterogeneity on equilibrium outcomes is highlighted. Similarly, it is crucial to incorporate correlated heterogeneous responses in empirical studies of causal effects of marketing mix variables. For example, Goldfarb and Tucker (2011), Blake et al. (2015), and Hoban and Bucklin (2015) estimate the causal effect of Internet advertising using data from field experiments. With the concern of regressor endogeneity being relieved, they capture heterogeneous treatment effects by interacting the treatment dummy with observed characteristics of ads and consumers. For observational studies, this strategy may not fully account for heterogeneity in consumer responses, which can be correlated with regressors in unknown functional forms.

accounts for the panel data structure. Assuming either the error term or the endogenous part of the error term is locally Guassian distributed within each panel unit, we can decompose the error term as a linear combination of first-stage residuals computed using copula transformed regressors, known as the "copula control function (CCF)", plus a new independent error term. Then we augment the panel data model with CCF in the second stage. As CCF captures the regressor-error dependence, the new error term in the augmented panel data model is orthogonal to all the regressors, addressing the regressor endogeneity problem. Given this augmented panel data model with heterogeneous coefficients, we adopt the mean group (MG) estimator (Pesaran and Smith, 1995; Pesaran and Yang, 2024) for the average effects.

Compared with the likelihood-based copula FE estimator for panel data by Haschka (2022), 2sCOPE-MG requires neither a normal error distribution nor a Gaussian copula regressor-error dependence structure and is more robust, easier to implement, and more general by allowing for slope endogeneity. Treating heterogeneous coefficients as fixed parameters, the estimator we propose is agnostic to (i) the underlying distributions of heterogeneous coefficients and (ii) the dependence structure between heterogeneous coefficients and regressors (such as marketing mix variables, customer characteristics, and product attributes). Moreover, the 2sCOPE-MG estimator is extended to dynamic panels with lagged outcome variables as regressors. Dynamic panel models allow researchers to separately analyze short-term and long-term effects, making them especially valuable for studying the impact of marketing mix activities on the growth and market potential of a new brand. As the incidental parameter problem is inherent in estimating dynamic panels with fixed effects, we correct the small time-period (T) bias by the Jackknife (JK) method.

By Monte Carlo (MC) simulations, we examine finite sample properties of the 2sCOPE-MG estimator, compared with the FE and two-stage copula augmented fixed effects (2sCOPE-FE) estimators. In static panel models, it is shown that the 2sCOPE-MG estimator remains unbiased in the presence of regressor endogeneity, slope endogeneity, or both and is robust under various data generating processes of the error term and regressors. We also caution practitioners about the substantial bias exhibited in the FE estimates in these scenarios, as well as bias in the 2sCOPE-FE estimator that addresses regressor endogeneity but neglects slope correlated heterogeneity. In dynamic panels, while FE and 2sCOPE-FE estimators exhibit more pronounced bias, the 2sCOPE-MGJK estimator continues to provide unbiased inference. Moreover, the MC results also demonstrate the effectiveness of including additional lagged variables as regressors to resolve the regressor endogeneity induced by serially correlated errors. Accompanied by a detailed guideline in subsection 3.7, these simulations aim to assist researchers in applying the 2sCOPE-MG estimator in practice.

We apply the 2sCOPE-MG method to estimate price elasticity and promotion effects (bonus and price reduction) in a dynamic sales response model, analyzing store-week panel data from Dominick's scanner data (1991–1994) for each of the 21 categories separately. We find that failing to account for price endogeneity leads to significant attenuation bias in the price elasticity for 19 categories. The price elasticity estimates after endogeneity correction can be twice the size of the uncorrected ones in certain categories. Comparisons with alternative methods show that ignoring slope endogeneity alone can lead to either overestimation or underestimation in the price elasticity estimates, depending on the correlation between store-specific price elasticities and within-store price variations. Consistent with reference price theory, the result shows that consumer response parameters vary not only with the levels but also with the variability of prices and promotions over time. Overall, averaging over 21 categories, the estimated price elasticity, which does not account for either endogeneity issue, is 0.258 smaller in size than the 2sCOPE-MG estimate. The difference amounts to 18.4% of the category-average price elasticity estimate of -1.407 based on the 2sCOPE-MG method. Moreover, the FE estimated price reduction effect for all categories has a large upward bias, that is, on average, 40.6% of the corresponding 2sCOPE-MG estimate of 0.233.[2] These results indicate that ignoring endogeneity in observational data can distort

[2]For the average bonus effect and persistence of sales (measured by the mean autoregressive coefficients), the bias direction varies across categories.

our understanding of marketing mix responses and misguide the design of future marketing strategies.

The rest of the paper is organized as follows. Section 2 reviews the literature on estimation methods handling regressor endogeneity and/or slope endogeneity. Section 3 sets out the model and derives the 2sCOPE-MG estimator. Section 4 presents MC evidence. Section 5 shows the application of estimating the dynamic sales response model using Dominick's scanner data. Section 6 concludes.

2 Literature review

Existing estimation methods in the literature are often designed to solve a single type of endogeneity. While considerable progress has been made, these methods often require the availability of auxiliary data (e.g., IVs) to identify causal effects or accurate knowledge of the dependence between consumer response parameters and marketing mix variables to address slope endogeneity. As an alternative, the 2sCOPE-MG method requires neither condition, enabling straightforward and potentially broader applications using observational panel data.

There is a rich set of estimation methods explicitly modeling heterogeneity in consumer responses by the *random* coefficient approach, with a focus on optimizing the marketing mix strategy. The Hierarchical Bayesian (HB) method reviewed in Rossi and Allenby (2003) is widely used in the marketing literature, particularly with a Gaussian prior of random coefficients. The HB models have been used in modeling discrete choices (Rossi et al., 1996; Allenby and Rossi, 1999; Andrews et al., 2002), discrete choices jointly with consideration sets (Van Nierop et al., 2010) or with selectively missing values in marketing mix variables (Qian and Xie, 2011), demand with marketing mix variables (Manchanda et al., 2004; Fok et al., 2006), customer channel migration (Ansari et al., 2008), and optimal price targeting (Smith et al., 2023). Alternatives to parametric HB include semiparametric finite mixture models assuming discrete random coefficients (Allenby and Rossi, 1999; Andrews et al., 2002) and more flexible prior distributions for heterogeneous coefficients (Fiebig et al., 2010; Ebbes et al., 2015). Besides the Bayesian approach, in conjoint analysis, Evgeniou et al. (2007) introduce a new approach to model heterogeneity using convex optimization and ridge regressions with unit-specific coefficients, and Chen et al. (2017) further develop a sparse learning approach for a multimodal continuous heterogeneity distribution. The above approaches have improved model fits and predictions by explicitly modeling consumer heterogeneity.

However, these methods are designed for experimental data or otherwise assume no endogeneity issues (i.e., independence between marketing mix regressors and response coefficients as well as between these regressors and the error term conditional on observables) and may yield significant estimation bias in the presence of slope or regressor endogeneity, with only a few exceptions described below. To address slope endogeneity, Manchanda et al. (2004), Fok et al. (2006), and Luan and Sudhir (2010) model the relationships between the latent heterogeneous coefficients and marketing mix variables, which may additionally require the availability of IVs (Luan and Sudhir, 2010). In contrast, the proposed 2sCOPE-MG estimator can handle arbitrary types of slope endogeneity, while requiring neither IVs nor knowledge about the nature of slope endogeneity. Given this advantage, Dubois et al. (2020) also use the MG approach to estimate individual preferences in a logit demand model with consumer-level purchase panel data. The flexibility in heterogeneity distributions enables the evaluation of soda taxes' effectiveness in reducing sugar consumption among *targeted* groups. Moreover, only Luan and Sudhir (2010) and our 2sCOPE-MG method simultaneously consider the problem of regressor endogeneity (regressor-error dependence), which we discuss next.

To tackle regressor endogeneity, the conventional approach is to use IVs, where available data on exogenous variations play a key role. However, good IVs that affect the outcome only through the focal regressor (i.e., the exclusion restriction) can be hard to find and validate in practice (Rossi, 2014). Even with theoretical guidance, ruling out alternative causal pathways requires significant efforts, especially when an IV has been used across different studies (Mellon, 2024). Given these challenges, the recent copula-modeling approaches (Park and Gupta, 2024; Qian et al., 2024) have the comparative advantage when facing a data problem. The IV-free copula approach to correcting regressor endogeneity was first introduced in Park and Gupta (2012), who proposed the maximum likelihood (MLE) and least squares estimators with copula-generated regressors. Later, Haschka (2022) generalizes the approach to linear panel models with fixed effects but homogeneous slopes and derives an alternative MLE that permits correlations between endogenous and exogenous regressors. The generalized least squares transformation employed by Haschka (2022) can only eliminate impacts of unit-specific intercepts but not those of correlated heterogeneous slopes. Hence, the MLE estimator will be susceptible to bias due to slope endogeneity. Yang et al. (2024) propose a two-stage copula control function estimator that can handle endogenous regressors with insufficient nonnormality and/or correlated with exogenous regressors. Its greater usage over various regressor distributions and robustness lay a good foundation for our further development in heterogeneous panel data models.[3] Nevertheless, none of the existing copula approaches or standard IV procedures handles the salient slope endogeneity problem in heterogeneous panel data models.[4]

We also adapt the copula procedure for dynamic panel models with heterogeneous slopes. The advantages of leveraging historical purchase data to enhance prediction accuracy are emphasized in Rossi et al. (1996), where they showed that the optimal customization procedure predicted using individual observations over all periods yields the largest net revenue. An alternative approach involves constructing aggregate/average measures and using them as regressors (Bucklin and Sismeiro, 2003; Smith et al., 2023). However, state dependence is a key feature in many outcome variables, like consumption and income. It is less clear how

[3]For reviews of other IV-free endogeneity correction methods, see Ebbes et al. (2009), Park and Gupta (2012) (p. 568) and Yang et al. (2024) (p. 10). Lewbel et al. (2024) propose a higher moments approach, where they assume the endogenous regressor can be decomposed linearly into an endogenous component and an exogenous component that does not affect the outcome.

[4]While Park and Gupta (2012) and Yang et al. (2024) use copula models to address slope endogeneity in cross-sectional data, we employ the model-free MG approach to handle slope endogeneity in heterogeneous panel models.

serial correlations in the outcome and regressor processes can be dealt with in the above papers. We capture dynamics using lagged variables as regressors (Montgomery et al., 2004; Fok et al., 2006) and discuss conditions and appropriate copula procedures to handle serial correlations in the error term, ensuring they do not pose difficulties in applying our method.

3 Models and estimation procedures

3.1 Static panels with regressor and slope endogeneity

Consider the following static panel data model for an outcome variable:

$$y_{it} = \alpha_i + \boldsymbol{\beta}_i' \boldsymbol{p}_{it} + \boldsymbol{\gamma}_{i,1}' \boldsymbol{w}_{it,1} + \boldsymbol{\gamma}_2' \boldsymbol{w}_{it,2} + \xi_{it}, \text{ for } i = 1, 2, ..., n, \text{ and } t = 1, 2, ..., T, \tag{1}$$

where $\boldsymbol{p}_{it}$ is a $K_p \times 1$ vector of continuous *endogenous* regressors (e.g., price), and $\boldsymbol{w}_{it} = (\boldsymbol{w}_{it,1}', \boldsymbol{w}_{it,2}')'$ is a $K_w \times 1$ vector of *strictly exogenous* control variables, with possibly heterogeneous coefficients. Let $\boldsymbol{x}_{it} = (1, \boldsymbol{p}_{it}', \boldsymbol{w}_{it,1}')'$. Stacking (1) over time, we have

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\theta}_i + \boldsymbol{W}_{i,2} \gamma_2 + \boldsymbol{\xi}_i, \tag{2}$$

where $\boldsymbol{y}_i = (y_{i1}, y_{i2}, ..., y_{iT})'$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, ..., \boldsymbol{x}_{iT})'$, $\boldsymbol{W}_{i,2} = (\boldsymbol{w}_{i1,2}, \boldsymbol{w}_{i2,2}, ..., \boldsymbol{w}_{iT,2})'$, $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, ..., \xi_{iT})'$, and $\boldsymbol{\theta}_i = (\alpha_i, \boldsymbol{\beta}_i', \boldsymbol{\gamma}_{i,1}')'$. One is often interested in estimating the mean coefficients (i.e., the average partial effects): $\boldsymbol{\theta}_0 = E\left[E(\boldsymbol{\theta}_i | \boldsymbol{X}_i, \boldsymbol{W}_{i,2})\right]$. For example, in category demand models, these mean coefficients may represent average category price or advertising elasticity across a population of stores or markets. Such population-averaged category elasticity estimates are often a key piece of information for policymakers to design policy interventions (e.g., soda tax) or for marketers to set optimal product pricing strategy and advertising budget.

When using the historical data to estimate the above panel data model, the popular FE panel data estimator assumes that the fixed effects α_i and the time-varying control variables in $(\boldsymbol{w}_{it,1}, \boldsymbol{w}_{it,2})$ capture all time-constant and time-varying confounders, respectively. When this assumption is violated and regressors and errors (or unit-specific coefficients) are actually correlated (e.g., because of potential omission of relevant time-varying confounders (Germann et al., 2015)), the FE panel data estimator can be severely biased. Furthermore, when slope coefficients are heterogeneous and correlated with regressors, the FE estimator that neglects such correlation between regressors and slopes, also resulting in severe bias and misleading inference.

Remedies require availability of additional data (measuring and controlling for all time-varying confounders or obtaining valid IVs) that is often impossible or difficult to obtain in practice. In these cases, we address endogenous regressors in panel data using a general and feasible IV-free copula endogeneity correction approach as described next. Then we demonstrate how to use the MG estimator to address slope endogeneity.

3.2 A generalized framework for IV-free copula correction

The rationale for the proposed approach is to correct the endogeneity bias via directly accounting for the dependence between the regressors and the structural error using copulas. A primary reason for such regressor-error dependence in a regression model is due to omitted variables. For example, in the sales response model, the structural error term may contain unmeasured managerial knowledge in decision-making (e.g., demand shocks, unmeasured product characteristics, or the cost of production) affecting both consumer purchases and retailer price decisions, leading to the regressor-error dependence.

In these cases, it appears reasonable to decompose the structural error term as $\xi_{it} = \sigma_i \xi^*_{it} + v_{it}$, where ξ^*_{it} is the error's (rescaled) endogenous part that captures the combined effects of all omitted variables mentioned above, and v_{it} is a disturbance term independent of the regressors and omitted variables such that $E(v_{it}|\boldsymbol{p}_{is}, \boldsymbol{w}_{is}, \xi^*_{is}) = 0$ for all i, t, and s. As ξ^*_{it} represents the combined effects of many omitted variables, it is reasonable to assume it approximately follows a normal distribution with $\xi^*_{it} \overset{iid}{\sim} N(0,1)$. The distribution of the disturbance term is left unspecified, so the error term does not need to follow a normal distribution.

To account for contemporaneous regressor endogeneity, we propose to capture the dependence between the regressors and the endogenous part of the error term (ξ^*_{it}) using the following *general location* Gaussian copula (GC) model that takes into consideration the panel data structure:

$$\boldsymbol{p}_{it} = \boldsymbol{\alpha}_{ip} + \boldsymbol{\phi}_i^P \boldsymbol{z}_{it} + \boldsymbol{e}_{it,p}, \quad \boldsymbol{w}_{it} = \boldsymbol{\alpha}_{iw} + \boldsymbol{\phi}_i^W \boldsymbol{z}_{it} + \boldsymbol{e}_{it,w}, \tag{3}$$

and

$$\begin{pmatrix} \boldsymbol{e}^*_{it,w} \\ \boldsymbol{e}^*_{it,p} \\ \xi^*_{it} \end{pmatrix} \sim IIDN\left(\mathbf{0}, \boldsymbol{V}_{i,\rho}\right) \text{ with } \boldsymbol{V}_{i,\rho} = \begin{pmatrix} \boldsymbol{V}_{i,w} & \boldsymbol{V}'_{i,pw} & \mathbf{0} \\ \boldsymbol{V}_{i,pw} & \boldsymbol{V}_{i,p} & \boldsymbol{\rho}_i \\ \mathbf{0}' & \boldsymbol{\rho}'_i & 1 \end{pmatrix}, \tag{4}$$

where $\boldsymbol{e}^*_{it,p} = \Phi^{-1}\left(F_p(\boldsymbol{e}_{it,p})\right)$ and $\boldsymbol{e}^*_{it,w} = \Phi^{-1}\left(F_w(\boldsymbol{e}_{it,w})\right)$, $\boldsymbol{V}_{i,\rho}$, $\boldsymbol{V}_{i,w}$, and $\boldsymbol{V}_{i,p}$ are $(K_p + K_w + 1) \times (K_p + K_w + 1)$, $K_w \times K_w$, and $K_p \times K_p$ positive definite and bounded matrices with diagonal elements being one and possibly non-zero elements off the diagonal, $\boldsymbol{V}_{i,pw}$ is a $K_p \times K_w$ matrix, and $\boldsymbol{\rho}_i$ is a $K_p \times 1$ vector.

Notations. *For a random variable x with a continuous distribution function (CDF) $F(\cdot)$, denote $x^* = \Phi^{-1}(F(x))$, where $\Phi(\cdot)$ is the standard normal CDF.*

We propose using the above general location GC model for a number of reasons. Broadly speaking, a GC model has a number of merits that makes it widely applicable and flexible to adequately capture multivariate dependence (Danaher and Smith, 2011; Park and Gupta, 2012; Christopoulos et al., 2021; Eckert and Hohberger, 2023; Qian and Xie, 2024). The GC model links marginal distributions of the variables in Equation 4 to form their joint distribution, even when these variables follow arbitrarily disparate marginal distributions, such as bounded supports, multi-modals, or skewed distributions. By using the nonparametric empirical CDF estimates of $F(\cdot)$, the GC model does not require these variables to take particular distributional forms, is capable to faithfully maintain the important marginal distributional features of regressors for model identification while simultaneously capturing the dependence of focal variables in Equation 4 separately from their marginal distributions.

The above general location GC model also explicitly accounts for panel data structure and possible heterogeneous endogeneity across panel units. In Equation 3, the regressors

$\{\boldsymbol{p}_{it}\}$ and $\{\boldsymbol{w}_{it}\}$ are allowed to depend on unit-specific mean levels $(\boldsymbol{\alpha}'_{ip}, \boldsymbol{\alpha}'_{iw})'$ and *observed exogenous covariates* in $\boldsymbol{z}_{it}$ (such as time trends) with the respective coefficients $\boldsymbol{\phi}_i^P$ and $\boldsymbol{\phi}_i^W$, and $\boldsymbol{e}_{it,p} \perp \boldsymbol{z}_{it}$, and $\boldsymbol{e}_{it,w} \perp \boldsymbol{z}_{it}$. We do not require $\boldsymbol{z}_{it}$ to contain IVs that meet the exclusion restriction condition. In fact, $\boldsymbol{z}_{it}$ can be null. The error terms in (2) and (3) then follow a Gaussian copula model described in Equation 4, capturing the regressor endogeneity of $\boldsymbol{p}_{it}$ and the dependence among endogenous and exogenous regressors. Thus, the general location GC model captures both linear and nonlinear effects of exogenous regressors on the endogenous ones while taking into consideration panel data structure and strengthening identification for endogenous regressors with Gaussian errors as will be shown later. Furthermore, in Equation 4, the correlations between $\boldsymbol{e}_{it,w}$, $\boldsymbol{e}_{it,p}$, and ξ^*_{it} is characterized by a possibly heterogeneous Gaussian copula model that permits the GC dependence structure to vary by panel units. Thus, the above general location model includes prior copula correction models (Park and Gupta, 2012; Haschka, 2022; Yang et al., 2024; Breitung et al., 2024) as special cases.[5]

Finally, the above general location GC model has the desirable property of *double robustness*. It is important to note that Equation 4 does not involve the error's exogenous part, v_{it}. Therefore, the dependence between the structural error ξ_{it} and regressors' errors needs not to follow a GC model and is left unspecified. Alternatively, one can assume $\xi_{it} = \sigma_i \xi^*_{it}$ where ξ^*_{it} is simply the standardized error term. Then Equation 4 does impose the assumption of GC regressor-error dependence while not assuming the error term can be decomposed to exogenous and exogenous parts. The proposed copula correction approach can work under either set of assumption and thus possesses the property of double robustness. Furthermore, even when both assumptions are wrong, the copula correction demonstrates robustness to a range of departures to the violation of both assumptions.

[5] Breitung et al. (2024) consider a degenerated GC model between the structural error's endogenous part and the error of endogenous regressor in which the correlation coefficient in the GC model is fixed at ± 1. Such a one-to-one deterministic linear relationship appears to be a too strong assumption to hold in practice.

3.3 An overview of how it works

The copula model given by (4) can be rewritten as

$$\begin{pmatrix} \boldsymbol{e}^*_{it,w} \\ \boldsymbol{e}^*_{it,p} \\ \xi^*_{it} \end{pmatrix} = \begin{pmatrix} \boldsymbol{V}^{1/2}_{i,w} & \mathbf{0}_{K_w \times K_p} & \mathbf{0}_{K_w \times 1} \\ \boldsymbol{V}_{i,pw}\boldsymbol{V}^{-1/2}_{i,w} & \left(\boldsymbol{V}_{i,p} - \boldsymbol{V}_{i,pw}\boldsymbol{V}^{-1}_{i,w}\boldsymbol{V}'_{i,pw}\right)^{1/2} & \mathbf{0}_{K_p \times 1} \\ \mathbf{0}'_{K_w \times 1} & \boldsymbol{\rho}'_{i,1} & \rho_{i,2} \end{pmatrix} \begin{pmatrix} \boldsymbol{\omega}^*_{it,w} \\ \boldsymbol{\omega}^*_{it,p} \\ \omega^*_{it} \end{pmatrix}, \tag{5}$$

with $\boldsymbol{\rho}_{i,1} = \left(\boldsymbol{V}_{i,p} - \boldsymbol{V}_{i,pw}\boldsymbol{V}^{-1}_{i,w}\boldsymbol{V}'_{i,pw}\right)^{-\frac{1}{2}}\boldsymbol{\rho}_i$, $\rho_{i,2} = \left[1 - \boldsymbol{\rho}'_i\left(\boldsymbol{V}_{i,p} - \boldsymbol{V}_{i,pw}\boldsymbol{V}^{-1}_{i,w}\boldsymbol{V}'_{i,pw}\right)^{-1}\boldsymbol{\rho}_i\right]^{\frac{1}{2}}$, and $\left(\boldsymbol{\omega}^{*\prime}_{it,p}, \boldsymbol{\omega}^{*\prime}_{it,w}, \omega^*_{it}\right)' \sim IIDN\left(\mathbf{0}, \boldsymbol{I}_k\right)$. Given (5), the transformed error terms of the endogenous regressors can be decomposed into two orthogonal components:

$$\boldsymbol{e}^*_{it,p} = \boldsymbol{V}_{i,pw}\boldsymbol{V}^{-1}_{i,w}\boldsymbol{e}^*_{it,w} + \boldsymbol{\epsilon}_{it,p}, \tag{6}$$

where $\boldsymbol{\epsilon}_{it,p}$ is independent of $\boldsymbol{e}^*_{it,w}$ but possibly correlated with ξ^*_{it}, given by

$$\boldsymbol{\epsilon}_{it,p} = \left(\boldsymbol{V}_{i,p} - \boldsymbol{V}_{i,pw}\boldsymbol{V}^{-1}_{i,w}\boldsymbol{V}'_{i,pw}\right)^{1/2}\boldsymbol{\omega}^*_{it,p} = \boldsymbol{e}^*_{it,p} - \boldsymbol{\Pi}_{i,pw}\boldsymbol{e}^*_{it,w}, \tag{7}$$

with $\boldsymbol{\Pi}_{i,pw} = \boldsymbol{V}_{i,pw}\boldsymbol{V}^{-1}_{i,w}$. Finally, we obtain a decomposition of the structural error term:

$$\xi^*_{it} = \boldsymbol{\rho}'_{i,1}\boldsymbol{\omega}^*_{it,p} + \rho_{i,2}\omega^*_{it} = \tilde{\boldsymbol{\rho}}'_i\boldsymbol{\epsilon}_{it,p} + \rho_{i,2}\omega^*_{it}, \tag{8}$$

with $\tilde{\boldsymbol{\rho}}'_i = \boldsymbol{\rho}'_{i,1}\left(\boldsymbol{V}_{i,p} - \boldsymbol{V}_{i,pw}\boldsymbol{V}^{-1}_{i,w}\boldsymbol{V}'_{i,pw}\right)^{-1/2}$.

Stacking (8) over time, we have $\boldsymbol{\xi}^*_i = \boldsymbol{\epsilon}_{ip}\tilde{\boldsymbol{\rho}}_i + \rho_{i,2}\boldsymbol{\omega}^*_i$, with $\boldsymbol{\epsilon}_{ip} = (\boldsymbol{\epsilon}_{i1,p}, \boldsymbol{\epsilon}_{i2,p}, .., \boldsymbol{\epsilon}_{iT,p})'$ and $\boldsymbol{\omega}^*_i = (\omega^*_{i1}, \omega^*_{i2}, ..., \omega^*_{iT})'$, and then plugging it into (2), we obtain the following panel model augmented by the copula generated regressors, $\boldsymbol{\epsilon}_{ip}$, given by

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\theta}_i + \boldsymbol{W}_{i,2}\boldsymbol{\gamma}_2 + \boldsymbol{\epsilon}_{ip}\boldsymbol{\delta}_i + \boldsymbol{u}_i, \tag{9}$$

with $\boldsymbol{\delta}_i = \sigma_i\tilde{\boldsymbol{\rho}}_i$ and $\boldsymbol{u}_i = \sigma_i\rho_{i,2}\boldsymbol{\omega}^*_i + \boldsymbol{v}_i$. Given (5) and recognizing v_{it} as the exogenous component of the structural error, the new error u_{it} is *exogenous* to all regressors in (9):

$$E_i(u_{it}|\boldsymbol{X}_i, \boldsymbol{W}_{i,2}, \boldsymbol{\epsilon}_{ip}) = 0, \text{ for } t = 1, 2, ..., T. \tag{10}$$

Thus, the augmented panel regression model in (9) is free from the regressor endogeneity problem. Based on the Gaussian copula model, we only need to estimate $\boldsymbol{\Pi}_{i,pw}$ in (7) to obtain the generated regressors and control for regressor endogeneity.

3.4 Identification and estimation

The identification of the augmented panel regression model in (9) requires the full rank condition of the predictor matrix. Specifically, the copula generated regressor $\boldsymbol{\epsilon}_{ip}$ cannot be perfectly collinear with the existing regressors in (2). Thus, we need to further impose certain distributional assumptions and rank condition on $\boldsymbol{e}_{it,p}$ and $\boldsymbol{e}_{it,w}$, summarized below.

Theorem 1 (Identification)**.** *Given Equations (2), (3) and (4) and Assumptions A.1 (random sampling), A.2(either the structural error or its endogenous part is normally distributed), A.4(The regressors' errors have bounded marginal density functions) and A.5(Either the structure error or its endogenous part follows a GC model jointly with regressors' errors.) in the online appendix hold. Then* $(\boldsymbol{\theta}', \boldsymbol{\gamma}_2', \boldsymbol{\delta}')'$ *are identified if and only if for each endogenous regressor, either (a) its error term has a marginal non-Gaussian distribution, or (b) its error term has a marginal Gaussian distribution but is correlated with non-Gaussian error terms of at least one distinct exogenous regressors errors. (See Assumption A.6 in the online appendix.)*

As noted previously, Assumptions A.2 and A.5 means that the proposed approach does not require structural error be normally distributed or have a joint GC dependence with regressors' errors. Also, when the GC dependence structure are allowed to vary by panel units, ξ_{it}^* is normally distributed conditional on unit-specific effects and over $t = 1, 2, ..., T$. Effectively, we do not impose normality assumption on the unconditional distribution of ξ_{it}^* across units. Furthermore, Assumptions A.2 and A.5 are working assumptions used in the derivation of the control functions. 2sCOPE-MG demonstrates robustness to a range of violations of both assumptions. With the identification conditions in place, we are ready to formulate a new estimator based on regressions that addresses slope endogeneity. As shown above, we adopt a general framework that allows for multiple heterogeneous effects in the outcome model (2), the regressor decomposition (3), and the Gaussian copula (4), consistent estimation of which requires a moderately large number of periods, T, of a panel dataset.

Otherwise, small-sample bias may lead to poor estimates.

Table 1 presents our estimation algorithm. The copula correction assuming homogenous GC dependence has shown robustness to heterogeneous endogeneity (Haschka, 2022). For sufficiently long panel data, one can explicitly permit heterogeneous GC dependence across panel units and obtain unit-specific endogeneity estimates. Specifically, we consider a group-specific Gaussian copula model. When T is sufficiently large, each unit can be viewed as a group. When T is moderately short, we cluster cross-sectional units into a much smaller number of groups, where the group structure can be identified based on some categorical variables (e.g., store-specific price tiers or consumer demographic characteristics) or informed by prior beliefs. In the first stage, the copula generated regressors are computed as residuals given group-specific estimates of the first-stage panel regression model. In the second stage, we augment the original panel model with these generated regressors and estimate the mean coefficients by averaging over the unit-specific OLS estimates.

3.5 Standard errors and inference

The asymptotic properties of the 2sCOPE-MG estimator for the mean coefficients are summarized in Theorem 2. A proof is provided in the online appendix. Unlike Yang et al. (2024), the proof needs to take into account the fact that copula transformations are performed on the unobserved errors in the general location GC model for the regressors instead of on the observed regressors themselves. As $n, T \to \infty$ and $\frac{n}{T^2} \to 0$, $\hat{\boldsymbol{\theta}}$ in (14) converges to the same asymptotic distribution, regardless of whether a pooling or a known grouping strategy is used in the estimation process.

Unlike all the previous copula endogeneity approaches, our inference does not require bootstrapped standard errors for inference. The variance estimator given by (15) already incorporates the estimation errors associated with $\hat{\boldsymbol{\theta}}_i$ as well as the generated regressors. Also, it is a consistent estimator for the asymptotic variance of $\hat{\boldsymbol{\theta}}$. As confirmed by our MC simulations, the estimated standard errors based on (15) are unbiased for the true variation of the estimator (e.g., see Table 3).

Table 1: Algorithm of the 2sCOPE-MG estimator

Data: $\{y_{it}, \boldsymbol{p}_{it}, \boldsymbol{w}_{it}\}$ for $i = 1, 2, \ldots, n$, and $t = 1, 2, \ldots, T$

Stage 1: Estimation of copula generated regressors.

1. Unit-specific demeaning: $\hat{\boldsymbol{e}}_{it,p} = \boldsymbol{p}_{it} - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{p}_{it}$ and $\hat{\boldsymbol{e}}_{it,w} = \boldsymbol{w}_{it} - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{w}_{it}$.
2. Apply group-specific Gaussian copula transformation based on empirical CDFs: $\hat{\boldsymbol{e}}^*_{it,p} = \Phi^{-1}\left(\hat{F}_{g(i),p}(\hat{\boldsymbol{e}}_{it,p})\right)$ and $\hat{\boldsymbol{e}}^*_{it,w} = \Phi^{-1}\left(\hat{F}_{g(i),w}(\hat{\boldsymbol{e}}_{it,w})\right)$.
3. Obtain residuals from group-specific regressions:

$$\hat{\boldsymbol{\epsilon}}_{ip} = \hat{\boldsymbol{P}}^*_i - \hat{\boldsymbol{W}}^*_i \hat{\boldsymbol{\Pi}}'_{g(i),pw}, \tag{11}$$

with $\hat{\boldsymbol{P}}^*_i = (\hat{\boldsymbol{e}}^*_{i1,p}, \hat{\boldsymbol{e}}^*_{i2,p}, ..., \hat{\boldsymbol{e}}^*_{iT,p})'$, $\hat{\boldsymbol{W}}^*_i = (\hat{\boldsymbol{e}}^*_{i1,w}, \hat{\boldsymbol{e}}^*_{i2,w}, ..., \hat{\boldsymbol{e}}^*_{iT,w})'$, and $\hat{\boldsymbol{\Pi}}'_{g,pw} = \left(\frac{1}{n_g}\sum_{i\in[n_g]}^{n_g} \hat{\boldsymbol{W}}^{*\prime}_i \hat{\boldsymbol{W}}^*_i\right)^{-1} \left(\frac{1}{n_g}\sum_{i\in n_g}^{n_g} \hat{\boldsymbol{W}}^{*\prime}_i \hat{\boldsymbol{P}}^*_i\right)$.

Stage 2: Estimation and statistical inference of the average partial effects in Model (9), including $\hat{\boldsymbol{\epsilon}}_{ip}$ as the control function. The estimation procedure involves four steps.

1. Estimate the homogeneous effects:

$$\hat{\boldsymbol{\gamma}}_2 = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{W}'_{i,2}\boldsymbol{M}_{ix,2}\boldsymbol{W}_{i,2}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{W}'_{i,2}\boldsymbol{M}_{ix,2}\boldsymbol{y}_i\right), \tag{12}$$

with $\boldsymbol{M}_{ix,2} = \boldsymbol{I}_T - \boldsymbol{X}_{i,2}(\boldsymbol{X}'_{i,2}\boldsymbol{X}_{i,2})^{-1}\boldsymbol{X}'_{i,2}$ and $\boldsymbol{X}_{i,2} = (\boldsymbol{X}_i, \hat{\boldsymbol{\epsilon}}_{ip})$.
2. Estimate group-specific coefficients of the copula generated regressors:

$$\hat{\boldsymbol{\delta}}_g = \left(\frac{1}{n_g}\sum_{i\in[n_g]}^{n_g} \hat{\boldsymbol{\epsilon}}'_{ip}\boldsymbol{M}_{ix}\hat{\boldsymbol{\epsilon}}_{ip}\right)^{-1} \left[\frac{1}{n_g}\sum_{i\in[n_g]}^{n_g} \hat{\boldsymbol{\epsilon}}'_{ip}\boldsymbol{M}_{ix}(\boldsymbol{y}_i - \boldsymbol{W}_{i,2}\hat{\boldsymbol{\gamma}}_2)\right], \tag{13}$$

with $\boldsymbol{M}_{ix} = \boldsymbol{I}_T - \boldsymbol{X}_i(\boldsymbol{X}'_i\boldsymbol{X}_i)^{-1}\boldsymbol{X}'_i$.
3. Estimate the mean coefficients:

$$\hat{\boldsymbol{\theta}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{\theta}}_i, \tag{14}$$

where $\hat{\boldsymbol{\theta}}_i = (\boldsymbol{X}'_i\boldsymbol{X}_i)^{-1}\boldsymbol{X}'_i\left(\boldsymbol{y}_i - \boldsymbol{W}_{i,2}\hat{\boldsymbol{\gamma}}_2 - \hat{\boldsymbol{\epsilon}}_{ip}\hat{\boldsymbol{\delta}}_{g(i)}\right)$.
4. The inference for $\hat{\boldsymbol{\theta}}$ is based on a consistent estimator of its asymptotic variance:

$$\hat{\boldsymbol{\Omega}}_\theta = \frac{1}{n}\sum_{i=1}^{n}(\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}})'. \tag{15}$$

Notes: (i) $g(i)$ and $i \in [n_g]$ denote that unit i belongs to Group g with n_g number of units. (ii) For group-specific coefficients, when T is sufficiently large, each cross-section unit can be a group. If there is no heterogeneity in the Gaussian copula dependence structure, $\hat{\boldsymbol{\epsilon}}_{ip}$ can be obtained given a homogeneous estimator using all observations. (iii) In the second stage, the homogeneous effects $\boldsymbol{\gamma}_2$ are estimated using pooled OLS after projecting out the influence of regressors with heterogeneous coefficients.

Theorem 2 (Asymptotic distributions and consistent variance estimator). *Suppose Equations (2), (3) and (4) and Assumptions A.1–A.7 in the online appendix hold. For the estimator of mean coefficients, $\hat{\boldsymbol{\theta}}$ given by (14), as $n, T \to \infty$ and $n/T^2 \to 0$, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to_d N(\mathbf{0}, \boldsymbol{\Omega}_\theta)$, where $\boldsymbol{\Omega}_\theta = Var(\boldsymbol{\theta}_i) \succ \mathbf{0}$, which can be consistently estimated by $\hat{\boldsymbol{\Omega}}_\theta$ in (15).*

3.6 Extensions to dynamic panel data models

3.6.1 Dynamic heterogeneous panels with serially *uncorrelated* errors

It is natural to extend the above analysis to heterogeneous dynamic panels with weakly exogenous regressors. Unlike strictly exogenous regressors, there is feedback from past outcomes to these regressors in the current period, resulting in a non-zero correlation between them and past errors. Some examples are lagged outcome and independent variables, which capture state dependence, expectation, and dynamic effects. To illustrate, consider a first-order autoregressive panel data model with covariates (namely, an ARX(1) panel model):

$$y_{it} = \alpha_i + \phi_i y_{i,t-1} + \boldsymbol{\beta}_i' \boldsymbol{p}_{it} + \boldsymbol{\gamma}_{i,1}' \boldsymbol{w}_{it,1} + \boldsymbol{\gamma}_2' \boldsymbol{w}_{it,2} + \xi_{it}, \tag{16}$$

where $y_{i,t-1}$ is the first lagged outcome variable. Consistent with the earlier discussion, $\boldsymbol{p}_{it}$ and $\boldsymbol{w}_{it}$ are contemporaneously endogenous and strictly exogenous, respectively. Our focus is on stationary dynamic models with possible time effects and trends, where $|\phi_i| < 1$ for all i. See Assumption A.8 in the online appendix.[6]

For dynamic panels, the errors are usually assumed to be serially uncorrelated, particularly given a sufficient number of lagged variables used to model the outcome process. Moreover, as the lagged variables are realized in the past, it is unlikely they are affected by the current shock. Thus, we assume $E(y_{i,t-1}\xi_{it}) = 0$ with ξ_{it} independently distributed over time for all i and t, under which there is no need to modify the Gaussian copula in (4).

However, we need to highlight the differences between imposing homogeneity on the au-

[6] As extensively studied in the time series and panel model literature, the unit root process ($\phi_i = 1$) has distinctive properties, requiring different estimation approaches. Researchers often apply the stationary dynamic panel model given by (16) after first or higher order differencing the outcome variable, which then becomes stationary.

toregressive coefficients, ϕ_i, and not. As well documented in the literature (Pesaran and Smith, 1995), the FE estimator can be severely biased with presence of heterogeneity, even for randomly distributed ϕ_i. In this case, the MG estimator is to be applied to estimate $\phi_0 = E(\phi_i)$. Moreover, there exists a small-T bias in estimating dynamic panels whether ϕ_i is heterogeneous or not. As inference based on the asymptotic distribution will be distorted when the time dimension is relatively smaller than the cross-sectional dimension, bias correction should be implemented in this case.

In summary, the copula-generated regressors are computed as previously mentioned. The mean coefficients of weakly exogenous regressors are estimated in Stage 2, along with other regressors with heterogeneous coefficients, namely, $\boldsymbol{x}_{it} = (1, y_{i,t-1}, \boldsymbol{p}'_{it}, \boldsymbol{w}'_{it,1})'$ with $\boldsymbol{\theta}_i = (\alpha_i, \phi_i, \boldsymbol{\beta}'_i, \boldsymbol{\gamma}'_{i,1})'$. For panel datasets with T relatively small to n, it is recommended to apply the Jackknife bias-correction given by

$$\hat{\boldsymbol{\theta}}_{MG-JK} = \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{\theta}}_{i,JK}, \tag{17}$$

where $\hat{\boldsymbol{\theta}}_{i,JK} = 2\hat{\boldsymbol{\theta}}_i - \frac{1}{2}(\hat{\boldsymbol{\theta}}_{iH_{2,1}} + \hat{\boldsymbol{\theta}}_{iH_{2,2}})$, and $\hat{\boldsymbol{\theta}}_i$, $\hat{\boldsymbol{\theta}}_{iH_{2,1}}$, and $\hat{\boldsymbol{\theta}}_{iH_{2,2}}$ are the unit-specific OLS estimates based on individual i's all $(t = 1, 2, ..., T)$, the first half $(t = 1, 2, ..., h)$, and the second half $(t = h + 1, h + 2, ..., T)$ of time-series observations.

3.6.2 Heterogeneous autoregressive panels with serially correlated errors

Though we abstract from serially correlated errors in the previous subsection, the Gaussian copula model can also be used to address endogeneity due to dynamic misspecification under stationarity. We first provide a new estimation strategy for a heterogeneous AR(1) panel with no other time-varying covariates, then extend the estimation method to the ARX(1) panel model. Consider a heterogeneous panel AR(1) model given by

$$y_{it} = \alpha_i + \phi_i y_{i,t-1} + \xi_{it}, \tag{18}$$

with $\xi_{it} = \rho_e \xi_{i,t-1} + \xi_{it}^*$ and $\xi_{it}^* \sim IIDN(0,1)$ for all t. Then (18) can be written as

$$y_{it} = \alpha_i + \phi_i y_{i,t-1} + \rho_e \left(y_{i,t-1} - \alpha_i - \phi_i y_{i,t-2}\right) + \xi_{it}^* = \tilde{\alpha}_i + \tilde{\phi}_i y_{i,t-1} + \delta_i \Delta y_{i,t-1} + \xi_{it}^*, \quad (19)$$

with $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$, $\tilde{\alpha}_i = \alpha_i(1-\rho_e)$, $\tilde{\phi}_i = \phi_i + \rho_e - \rho_e\phi_i$, and $\delta_i = \rho_e\phi_i$, where $y_{i,t-1}$ and $\Delta y_{i,t-1}$ are exogenous to ξ_{it}^*, and the process $\{y_{it}\}$ is stationary for all i. Note that the autoregressive structure of the error term, ξ_{it}, allows the serially uncorrelated shocks ξ_{it}^* to be the new error term by including a *finite* number of lagged outcome variables as regressors. Since ξ_{it}^* is uncorrelated with lagged outcome variables, the mean coefficient $E(\tilde{\phi}_i)$ can be estimated consistently by adding $\Delta y_{i,t-1}$ into the regression. For AR(1) panels with a general AR(p) error term, the contemporaneous endogeneity can be eliminated by including $(\Delta y_{i,t-1}, \Delta y_{i,t-2}, ..., \Delta y_{i,t-p})'$ as additional regressors.

Alternatively, note that ξ_{it}^* is Gaussian distributed for all t. It can be shown that under Assumption A.8 of stationarity, Δy_{it} is also Gaussian distributed for all t. Then the contemporaneous endogeneity between ξ_{it} and $y_{i,t-1}$ can be represented by a Gaussian copula model, which only involves a *finite* number of further lags, $y_{i,t-2}$ in the case of the AR(1) error process. Hence, the transformed first difference, $(\Delta y_{i,t-1})^*$, can be used as a generated regressor, and we can estimate the mean coefficients in the following augmented model consistently by the MG or MG-JK estimators,

$$y_{it} = \tilde{\alpha}_i + \tilde{\phi}_i y_{i,t-1} + \tilde{\delta}_i \left(\Delta y_{i,t-1}\right)^* + \xi_{it}^*, \quad (20)$$

with $\tilde{\delta}_i = \delta_i \tilde{\sigma}$ and $\tilde{\sigma}^2 = E\left[(\Delta y_{it})^2\right]$.

3.6.3 Heterogeneous dynamic panels with covariates and serially correlated errors

The above analysis shows why further lags of the outcome variable need to be added to AR(1) panels with serially correlated errors. Now consider the ARX(1) panel model in (16) with contemporaneous regressor endogeneity given by (4) and serially correlated errors $\xi_{it} = \rho_e \xi_{i,t-1} + \xi_{it}^*$. With the presence of other covariates in the model, we directly substitute

$\xi_{i,t-1}$ by $(y_{i,t-1} - \boldsymbol{\theta}_i' \boldsymbol{x}_{i,t-1} - \boldsymbol{\gamma}_2' \boldsymbol{w}_{i,t-1,2})$ into (16), then we have

$$y_{it} = (1-\rho_e)\alpha_i + (\phi_i+\rho_e)y_{i,t-1} + \boldsymbol{\beta}_i'\boldsymbol{p}_{it} + \boldsymbol{\gamma}_{i,1}'\boldsymbol{w}_{it,1} + \boldsymbol{\gamma}_2'\boldsymbol{w}_{it,2} - \rho_e(\boldsymbol{\theta}_i'\boldsymbol{x}_{i,t-1} + \boldsymbol{\gamma}_2'\boldsymbol{w}_{i,t-1,2}) + \xi_{it}^*. \quad (21)$$

As $y_{i,t-1}$ is weakly exogenous in (21), to address contemporaneous endogeneity between $\boldsymbol{p}_{it}$ and ξ_{it}^*, we can directly apply the 2sCOPE-MG or 2sCOPE-MGJK estimators for the (mean) coefficients. In the first stage, the copula-generated regressors are constructed in the same way by regressing $\hat{\boldsymbol{e}}_{it,p}^*$ on $\hat{\boldsymbol{e}}_{it,w}^*$. Then the mean coefficients of all regressors with coefficient heterogeneity in (21) including $\boldsymbol{x}_{i,t-1}$ and $\boldsymbol{w}_{i,t-1}$, are estimated in the second stage by the MG (or MG-JK) estimator. Whether the Jackknife bias correction is applied or not, the analytical variance estimator in (15) provides asymptotically unbiased inference.

3.7 Guidelines for using the 2sCOPE-MG estimation approach

For both dynamic and static panels, incorporating heterogeneous effects eliminates the concern of slope endogeneity bias.[7] The choice between dynamic and static model specifications hinges on whether individual decisions are influenced by past outcomes or if information sets of decision-makers include historical outcomes. It is context-specific and relies on institutional knowledge. For dynamic panels, the presence of serially correlated errors leads to the regressor endogeneity problem, to which a common solution is by including higher-order lags of regressors. We have illustrated the rationality of this procedure in sections 3.6.2 and 3.6.3. Its effectiveness in restoring consistency is shown by MC simulations in section 4.3.

To preprocess the data, we remove unit-specific means from the regressors based on the general location Gaussian copula before applying the Gaussian copula transformation. Then people can employ the procedures suggested by Yang et al. (2024) to empirically validate the non-Gaussionality (and relevance) conditions on the residuals. That is, we first check the non-Gaussianality of $\hat{\boldsymbol{e}}_{it,p}$ by the Kolmogorov-Smirnov (KS) test. If the p-value is smaller than 0.05, we are on the safe side to proceed with the 2sCOPE-MG estimation. Otherwise, we further examine the non-Gaussianality of $\hat{\boldsymbol{e}}_{it,w}$ by the KS test and the strength

[7] Formal tests for slope heterogeneity can be found in Pesaran and Yamagata (2008) for dynamic panels and Pesaran and Yang (2024) for static panels, where the exogeneity condition is assumed to hold.

of correlation between $\hat{\boldsymbol{e}}_{it,p}$ and $\hat{\boldsymbol{e}}_{it,w}$ by the F test. The sufficient identification condition for the 2sCOPE-MG estimation requires that there is at least one variable in $\hat{\boldsymbol{e}}_{it,w}$ with sufficient non-Gaussianality, p-value < 0.001, and explanatory power of $\hat{\boldsymbol{e}}_{it,p}$, F-statistic > 10. When such conditions also fail, we suggest people collect more data on other control variables satisfying the non-Gaussianality and relevance conditions. Different from IVs, these control variables can be some of the omitted variables researchers are concerned about – with non-zero impacts on the outcome. Validating the rank condition using the aforementioned rules of thumb can enhance our confidence in proceeding with the 2sCOPE-MG estimation.

Last but not least, for unbiased inference in dynamic panel data models, we suggest applying the 2sCOPE-MGJK estimator given by (17) to the copula augmented model, particularly with short T (relative to n) panel datasets. A sufficient condition is provided for the existence of its finite second-order moments in heterogeneous dynamic panel models in Yang (2023). Researchers are recommended to check this condition before estimation.

4 Monte Carlo simulations

We design MC simulations to inspect the effects of (i) regressor endogeneity, (ii) slope endogeneity induced by correlated heterogeneous coefficients, and (iii) both regressor and slope endogeneity on the estimation of the mean coefficients, separately. For the data generating process (DGP) of the outcome variable, y_{it}, we consider both static and dynamic panel data models, without and with serial correlations in the error process. In the following, we first focus on static panel data models. A number of DGPs are considered to examine whether our proposed estimator is robust to heteroskedastic errors, serially correlated errors, and different distributions of individual fixed effects in the regressors processes, which are likely to present in various practical data sets. Then we investigate the finite-sample properties of our estimation approach in dynamic panels with lagged dependent variables as explanatory variables. Finally, for the case of dynamic panels with autoregressive errors, subsections 3.6.2 and 3.6.3 introduce a different estimation strategy, whose finite-sample properties are shown by MC simulations below. Table 2 summarizes the MC designs.

Table 2: Summary of MC designs

1. Static panel data models	
1.1.	One endogenous regressor and one exogenous regressor
1.2.	Two exogenous regressors with correlated heterogeneous coefficients
1.3.	One endogenous regressor and one exogenous regressor with correlated heterogeneous coefficients
1.4.	Robustness to heteroskedastic errors
1.5.	Robustness to serially correlated errors
1.6.	Robustness to uniform-distributed fixed effects in regressor processes
2. Dynamic panel data models with serially uncorrelated errors	
2.1.	A lagged dependent variable, one endogenous covariate, and one exogenous covariate as regressors
2.2.	A lagged dependent variable and two exogenous covariates as regressors with correlated heterogeneous coefficients
2.3.	A lagged dependent variable, one endogenous covariate, and one exogenous covariate as regressors with correlated heterogeneous coefficients
3. Dynamic panel data models with serially correlated errors	
3.1.	A lagged dependent variable as a regressor with homogeneous autoregressive coefficients
3.2.	A lagged dependent variable as a regressor with heterogeneous autoregressive coefficients
3.3.	A lagged dependent variable, one endogenous covariate, and one exogenous covariate as regressors with correlated heterogeneous coefficients

Two alternative estimation approaches are considered, including the FE estimator and the two-stage Copula augmented fixed effects (2sCOPE-FE) estimator computed by applying the FE estimator to the copula augmented panel model. As the FE estimator for heterogeneous intercepts is one of the most popular estimators for panel data, its performance in the MC simulations provides relevant measures of the estimation bias and size distortions when ignoring regressors and/or slope endogeneity. For the 2sCOPE-FE estimator, while the second-stage regression does not suffer from regressor endogeneity, it does not account for correlated heterogeneity in slope coefficients. For dynamic panel data models, following our recommendation to apply the Jackknife method for correcting small-T bias, we present simulation results of the 2sCOPE-MGJK estimator given by (17) instead of the 2sCOPE-MG estimator. We report the mean bias, standard deviation of the estimates (SD), mean estimated standard error ($\hat{se}$), root mean squared errors (RMSE), mean size of testing the null hypothesis $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$, and the ratio of the absolute mean bias to the mean estimated standard error (t_{bias}) over 1,000 replications for each simulation.[8] We consider a combination of sample sizes: $n = 100$ and $T \in \{10, 50, 100\}$.

4.1 Static panel data models

We generate $\{y_{it}\}$ by a static panel data model as

$$y_{it} = \alpha_i + \beta_i p_{it} + \gamma_i w_{it} + \xi_{it}, \text{ for } i = 1, 2, ..., n, \text{ and } t = 1, 2, ..., T, \tag{22}$$

where p_{it} and w_{it} are endogenous and exogenous regressors, respectively, with possibly serially correlated errors, $\xi_{it} = \phi_e \xi_{i,t-1} + \sigma_i \xi_{it}^*$, and individual fixed effects. The slope coefficients for regressors, $(\beta_i, \gamma_i)'$, are individual-specific, which is an important feature of panel data but hasn't been considered in Haschka (2022). The regressors are generated from the following general location Gaussian copula model: $p_{it} = \alpha_{i,p} + e_{it,p}$, and $w_{it} = \alpha_{i,w} + e_{it,w}$, with $\alpha_{i,p} \sim IIDN(1, 1)$ and $\alpha_{i,w} \sim IIDN(1, 1)$. Following Assumption A.6, we generate $e_{it,p}$ from a mixture normal distribution, $e_{it,p} = v_{it,1} + v_{it,2}$ with $v_{it,1} \sim IIDN(\mu_{p1}, \sigma_{p1})$ and $v_{it,2} \sim IIDN(\mu_{p2}, \sigma_{p2})$, and $e_{it,w}$ from an exponential distribution, $e_{it,w} \sim IIDExp(\mu_w)$, with $|\rho_{pw}| > 0$. We set $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0, \gamma_0)' = (1, 1, -1)'$.

4.1.1 Regressor endogeneity

To induce regressor endogeneity, the shocks in the processes of dependent variables and regressors are jointly generated based on a Gaussian copula model for Case 1.1,

$$\begin{pmatrix} e_{it,p}^* \\ e_{it,w}^* \\ \xi_{it}^* \end{pmatrix} \sim IIDN \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{pw} & \rho_{pe} \\ \rho_{pw} & 1 & 0 \\ \rho_{pe} & 0 & 1 \end{bmatrix} \right), \tag{23}$$

[8] The standard errors of FE and 2sCOPE-FE estimators are computed based on the "clustered" covariance estimator on p. 654 of Pesaran (2015), shown to be consistent under the respective modeling assumptions. When these underlying assumptions for derivation do not hold in our DGP, the estimated standard errors may not be consistent. The standard errors of the 2sCOPE-MG estimator are computed by (15).

with $\rho_{pw} = \rho_{pe} = 0.5$. Then we generate $e_{it,p} = H^{-1}\left(\Phi(e^*_{it,p})\right)$ and $e_{it,w} = G^{-1}\left(\Phi(e^*_{it,w})\right)$. $\{y_{it}\}$ is generated by (22) with individual fixed effects that are correlated with $\{p_{it}\}$ as $\alpha_i = 0.25\left(\frac{1}{T}\sum_{t=1}^{T} p_{it}\right) + \eta_{i\alpha}$, and $\eta_{i\alpha} \sim IIDN(0, 0.5)$, homogeneous slopes $\beta_i = \beta_0 = 1$ and $\gamma_i = \gamma_0 = -1$ for all i, and homoskedastic serially uncorrelated errors $\xi_{it} = \xi^*_{it}$ for all i and t. For all MC simulations, α_i are always generated to be correlated with regressors. With homogeneous coefficients in (23), we pool all panel observations to compute $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}_g$ in (13).

Table 3 shows simulation results of FE, 2sCOPE-FE, and 2sCOPE-MG estimators under regressor endogeneity. The FE estimator is severely biased. Particularly for slope coefficients, the inferences using the default estimator of standard errors incorrectly reject the true value with a 100% chance over 1,000 replications. As the generated correlation between error term and endogenous regressor is constant for each period, the bias of the FE estimator does not vary by T. For 2sCOPE-FE and 2sCOPE-MG estimators, the contemporaneous regressor endogeneity is eliminated in the augmented panel regression. Hence, for the slope coefficients under homogeneity, they are both unbiased, with size around the 5% nominal level. It is worth noting that the 2sCOPE-MG estimator exhibits better small-sample (T) performance compared with 2sCOPE-FE, displaying a smaller finite-sample bias when $T = 10$.

4.1.2 Slope endogeneity

To investigate the impacts of slope endogeneity without regressor endogeneity, we set $\rho_{pw} = \rho_{pe} = 0$ in (23) and generate $\boldsymbol{\theta}_i$ to be correlated with regressors for Case 1.2:

$$\boldsymbol{\theta}_i = (\alpha_i, \beta_i, \gamma_i)' = \boldsymbol{\theta}_0 + \boldsymbol{\psi}_\mu \bar{e}_{ip} + \boldsymbol{\psi}_\lambda \lambda_i + \boldsymbol{\eta}_i, \tag{24}$$

where $\bar{e}_{ip} = T^{-1}\sum_{t=1}^{T} e_{it,p}$, $\lambda_i = \frac{\boldsymbol{e}'_{i,p}\boldsymbol{e}_{i,p} - n^{-1}\sum_{j=1}^{n}\boldsymbol{e}'_{j,p}\boldsymbol{e}_{j,p}}{\sqrt{(\boldsymbol{e}'_{i,p}\boldsymbol{e}_{i,p} - n^{-1}\sum_{j=1}^{n}\boldsymbol{e}'_{j,p}\boldsymbol{e}_{j,p})^2}}$ with $\boldsymbol{e}_{i,p} = (e_{i1,p}, e_{i2,p}, ..., e_{iT,P})'$, $\boldsymbol{\eta}_i = (\eta_{i\alpha}, \eta_{i\beta}, \eta_{i\gamma})' \sim IIDN(\mathbf{0}, 0.5\boldsymbol{I}_k)$, $\boldsymbol{\psi}_\mu = (\psi_{\mu\alpha}, \psi_{\mu\beta}, \psi_{\mu\gamma})'$, and $\boldsymbol{\psi}_\lambda = (\psi_{\lambda\alpha}, \psi_{\lambda\beta}, \psi_{\lambda\gamma})'$.[9] The three unit-specific terms in the right-hand side of (24) correspond to (i) a component correlated with the levels of regressors (i.e., $\bar{e}_{ip}$), (ii) a component correlated with the within-unit variations of regressors (i.e., λ_i), and (iii) an idiosyncratic component, respectively. For

[9] λ_i is constructed to have zero mean and unit variance s.t. $E(\boldsymbol{\theta}_i) = \boldsymbol{\theta}_0$ and $Var(\boldsymbol{\theta}_i)$ is constant over T.

Table 3: MC results of FE, 2sCOPE-FE and 2sCOPE-MG estimators in Case 1.1: static panels under regressor endogeneity

		Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}
		$T = 10, n = 100$						$T = 50, n = 100$						$T = 100, n = 100$					
FE	$\hat{\alpha}$	0.128	0.090	0.057	0.156	0.56	2.23	0.130	0.074	0.025	0.149	0.87	5.26	0.130	0.073	0.017	0.149	0.92	7.52
	$\hat{\beta}$	0.443	0.022	0.022	0.444	1.00	20.34	0.443	0.009	0.009	0.443	1.00	47.40	0.443	0.007	0.007	0.443	1.00	67.00
	$\hat{\gamma}$	-0.283	0.033	0.032	0.285	1.00	8.73	-0.283	0.014	0.014	0.283	1.00	20.23	-0.282	0.010	0.010	0.282	1.00	28.71
2sCOPE-FE	$\hat{\alpha}$	-0.062	0.115	0.091	0.131	0.18	0.68	-0.021	0.077	0.029	0.079	0.46	0.72	-0.012	0.074	0.020	0.075	0.61	0.61
	$\hat{\beta}$	-0.015	0.150	0.143	0.151	0.07	0.11	0.010	0.044	0.043	0.045	0.06	0.24	0.011	0.030	0.030	0.032	0.07	0.38
	$\hat{\gamma}$	0.036	0.106	0.110	0.112	0.06	0.33	0.004	0.031	0.032	0.031	0.05	0.13	0.001	0.020	0.022	0.020	0.05	0.03
2sCOPE-MG	$\hat{\alpha}$	-0.004	0.148	0.148	0.148	0.05	0.03	-0.005	0.078	0.084	0.078	0.03	0.06	-0.003	0.075	0.079	0.075	0.04	0.04
	$\hat{\beta}$	-0.011	0.150	0.144	0.151	0.07	0.08	0.012	0.044	0.043	0.045	0.07	0.28	0.012	0.030	0.030	0.032	0.08	0.41
	$\hat{\gamma}$	-0.008	0.118	0.116	0.119	0.06	0.07	-0.008	0.032	0.032	0.033	0.06	0.25	-0.006	0.021	0.022	0.022	0.05	0.28

Notes: The table shows the mean bias, standard deviation (SD) of the respective coefficients, mean estimated standard error ($\hat{se}$), root mean squared error (RMSE), mean size of the test, and the ratio of the absolute mean bias divided by the mean estimated standard error (t_{bias}). The number of replications is 1000.

Table 4: MC results of FE, 2sCOPE-FE and 2sCOPE-MG estimators in Cass 1.2: static panels under slope endogeneity

		Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}
		$T = 10, n = 100$						$T = 50, n = 100$						$T = 100, n = 100$					
FE	$\hat{\alpha}$	0.184	0.205	0.250	0.276	0.07	0.74	0.204	0.143	0.197	0.249	0.10	1.03	0.213	0.138	0.190	0.253	0.12	1.12
	$\hat{\beta}$	0.101	0.086	0.091	0.133	0.18	1.11	0.043	0.072	0.081	0.084	0.06	0.54	0.030	0.069	0.080	0.075	0.04	0.38
	$\hat{\gamma}$	-0.001	0.111	0.114	0.111	0.05	0.01	-0.001	0.082	0.087	0.082	0.03	0.01	0.000	0.074	0.083	0.074	0.04	0.00
2sCOPE-FE	$\hat{\alpha}$	0.196	0.522	0.556	0.558	0.06	0.35	0.204	0.211	0.254	0.294	0.09	0.80	0.207	0.174	0.219	0.270	0.11	0.95
	$\hat{\beta}$	0.090	0.542	0.534	0.549	0.06	0.17	0.043	0.180	0.187	0.185	0.05	0.23	0.036	0.130	0.139	0.135	0.05	0.26
	$\hat{\gamma}$	-0.002	0.114	0.118	0.114	0.05	0.02	-0.001	0.082	0.087	0.082	0.04	0.01	0.000	0.074	0.083	0.074	0.04	0.00
2sCOPE-MG	$\hat{\alpha}$	0.009	0.513	0.516	0.513	0.03	0.02	-0.001	0.177	0.183	0.177	0.04	0.00	-0.006	0.132	0.137	0.132	0.04	0.04
	$\hat{\beta}$	-0.010	0.538	0.533	0.538	0.05	0.02	-0.001	0.180	0.186	0.180	0.05	0.00	0.006	0.130	0.139	0.130	0.03	0.04
	$\hat{\gamma}$	-0.005	0.091	0.099	0.091	0.03	0.05	-0.002	0.074	0.080	0.074	0.04	0.02	0.000	0.070	0.080	0.070	0.03	0.00

Notes: See notes to Table 3.

example, consumers' responses to ads may vary with income levels, price variability, and personal emotions at the moment.

To examine the impacts of slope correlated heterogeneity with respect to the unit-specific means and variations of regressors, we set $\boldsymbol{\psi}_\mu = \boldsymbol{\psi}_\lambda = (0.25, 0.25, 0.25)'$. Table 4 reports the simulation results with $\xi_{it} = \xi_{it}^*$ for all i and t. The magnitudes of bias and bias ratio for both FE and 2sCOPE-FE estimators of β are large, showing that both approaches cannot address the issue of slope endogeneity. In contrast, the 2sCOPE-MG estimator is shown to be consistent, exhibiting negligible small-sample bias and a near-zero bias ratio, and its size is at the 5% nominal level uniformly across all sample sizes.

4.1.3 Regressor and slope endogeneity

In Case 1.3, to induce both regressor and slope endogeneity, we generate the errors by (23) with $\rho_{pw} = \rho_{pe} = 0.5$ where $\xi_{it} = \xi_{it}^*$ for all i and t, and heterogeneous coefficients by (24) with $\boldsymbol{\psi}_\mu = \boldsymbol{\psi}_\lambda = (0.25, 0.25, 0.25)'$. Under both regressor and slope endogeneity, Table 5 shows that the bias of the FE estimator closely matches the sum of its biases from Tables 3 and 4. Since both regressor and slope endogeneity induce positive bias, the overall bias increases when both are present. The 2sCOPE-FE estimator also displays substantial bias, which is greater than its bias in Case 1.2. However, the 2sCOPE-MG estimator continues to be unbiased and has comparable performances in Cases 1.1 and 1.3. As discussed earlier, the standard errors and RMSE of the 2sCOPE-MG estimator increase in copula-augmented panel regressions when there is no regressor endogeneity, as in Case 1.2. As it is designed for heterogeneous panels, its estimation errors do not rise much. The increase in RMSEs for the slope estimates can be attributed to the positive variances of the heterogeneous coefficients. In summary, the 2sCOPE-MG estimator demonstrates unbiasedness, exhibits the smallest RMSE and bias ratios, and consistently offers valid inference across various sample sizes, under either regressor endogeneity, slope endogeneity, or both.

4.1.4 Robustness to different processes of errors and regressors

The Gaussian copula we impose is on the primitive shocks, and it does not forbid either serial correlation or heteroskedasticity in the error term of the outcome process, nor individual fixed effects or time effects in the regressors processes. For practitioners, we experiment beyond the cases where the error term in (22) and regressors are independently and identically distributed, to illustrate the robustness of the 2sCOPE-MG estimator. The DGPs for Cases 1.4–1.6 are the same as that in Case 1.3 except for the following variations:

- Case 1.4. Heteroskedastic errors: $\xi_{it} = \sigma_i \xi^*_{it}$ and $\sigma_i^2 = 0.5 + 0.5\nu_i^2$ with $\nu_i \sim IIDN(0, 1)$
- Case 1.5. Serial correlated errors: $\xi_{it} = \rho_e \xi_{i,t-1} + \xi^*_{it}$ with $\rho = 0.4$
- Case 1.6. Uniformly distributed individual fixed effects in regressors:
  $\alpha_{i,p} \sim IIDUniform\left(1 - \sqrt{3}, 1 + \sqrt{3}\right)$ and $\alpha_{i,w} \sim IIDUniform\left(1 - \sqrt{3}, 1 + \sqrt{3}\right)$

The simulation results of Cases 1.4–1.6 are reported in Table A.1 of the online appendix. The performances of the estimators are similar to those in Case 1.3, where the 2sCOPE-MG estimator continues to be the only unbiased estimator of the mean coefficients under both regressor and slope endogeneity. Particularly, the simulation results of Case 1.4 show that the 2sCOPE-MG estimator can accommodate cross-sectional random heteroskedasticity in the error term. Results of Case 1.5 show that the 2sCOPE-MG estimator can handle contemporaneous endogeneity and provides a valid inference even with serially correlated errors. Case 1.6 shows that as we de-mean the covariates separately for each unit, the 2sCOPE-MG estimator can allow for arbitrary distributions of individual-specific means in the regressor processes with bounded moments.

4.2 Dynamic panel data models with serially uncorrelated errors

It has been shown that the 2sCOPE-MG estimator effectively restores exogeneity for unbiased estimation in static panels. For dynamic panel data models containing lagged dependent variables as regressors, the autoregressive structure automatically introduces non-zero correlations between the lagged dependent variables and past error terms. Thus, consistent

Table 5: MC results of FE, 2sCOPE-FE and 2sCOPE-MG estimators in Case 1.3: static panels under regressor and slope endogeneity

		Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}
		$T = 10, n = 100$						$T = 50, n = 100$						$T = 100, n = 100$					
FE	$\hat{\alpha}$	0.209	0.214	0.261	0.299	0.10	0.80	0.295	0.144	0.200	0.328	0.28	1.48	0.317	0.139	0.191	0.346	0.33	1.66
	$\hat{\beta}$	0.563	0.088	0.094	0.570	1.00	5.99	0.496	0.073	0.081	0.501	1.00	6.11	0.480	0.069	0.080	0.485	1.00	6.00
	$\hat{\gamma}$	-0.227	0.128	0.127	0.261	0.47	1.79	-0.264	0.085	0.090	0.278	0.84	2.94	-0.270	0.076	0.084	0.281	0.92	3.21
2sCOPE-FE	$\hat{\alpha}$	0.020	0.222	0.269	0.223	0.02	0.07	0.145	0.145	0.200	0.205	0.05	0.73	0.175	0.140	0.191	0.223	0.08	0.91
	$\hat{\beta}$	0.105	0.175	0.170	0.204	0.12	0.62	0.063	0.083	0.092	0.105	0.09	0.69	0.048	0.075	0.085	0.090	0.06	0.57
	$\hat{\gamma}$	0.092	0.162	0.165	0.186	0.09	0.56	0.023	0.090	0.094	0.093	0.05	0.24	0.012	0.079	0.086	0.080	0.03	0.14
2sCOPE-MG	$\hat{\alpha}$	-0.004	0.148	0.149	0.148	0.05	0.03	-0.005	0.078	0.088	0.078	0.03	0.06	-0.003	0.075	0.083	0.075	0.04	0.04
	$\hat{\beta}$	-0.012	0.169	0.164	0.170	0.06	0.07	0.012	0.082	0.090	0.082	0.03	0.13	0.011	0.074	0.084	0.075	0.03	0.14
	$\hat{\gamma}$	-0.010	0.141	0.141	0.141	0.05	0.07	-0.009	0.079	0.085	0.079	0.04	0.10	-0.006	0.073	0.082	0.073	0.03	0.07

Notes: See notes to Table 3.

Table 6: MC results of FE, 2sCOPE-FE, and 2sCOPE-MGJK estimators in Case 2.3: dynamic panels under regressor and slope endogeneity with serial uncorrelated errors

		Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}
		$T = 10, n = 100$						$T = 50, n = 100$						$T = 100, n = 100$					
FE	$\hat{\alpha}$	0.321	0.427	0.262	0.534	0.37	1.23	0.338	0.312	0.204	0.460	0.44	1.65	0.355	0.289	0.196	0.458	0.49	1.81
	$\hat{\phi}$	-0.048	0.042	0.040	0.064	0.28	1.21	0.065	0.035	0.033	0.073	0.50	1.97	0.082	0.034	0.032	0.089	0.72	2.58
	$\hat{\beta}$	0.522	0.088	0.094	0.530	1.00	5.54	0.491	0.073	0.081	0.496	1.00	6.04	0.478	0.069	0.080	0.483	1.00	5.97
	$\hat{\gamma}$	-0.197	0.128	0.127	0.235	0.37	1.55	-0.260	0.086	0.090	0.274	0.82	2.88	-0.269	0.076	0.085	0.279	0.92	3.18
2sCOPE-FE	$\hat{\alpha}$	0.131	0.435	0.273	0.455	0.25	0.48	0.187	0.314	0.205	0.365	0.29	0.92	0.213	0.289	0.196	0.359	0.30	1.09
	$\hat{\phi}$	-0.048	0.042	0.040	0.064	0.27	1.20	0.065	0.035	0.033	0.073	0.49	1.97	0.082	0.034	0.032	0.089	0.72	2.58
	$\hat{\beta}$	0.063	0.185	0.179	0.195	0.07	0.35	0.059	0.084	0.092	0.103	0.08	0.64	0.046	0.075	0.085	0.088	0.05	0.54
	$\hat{\gamma}$	0.123	0.165	0.170	0.206	0.10	0.72	0.027	0.090	0.095	0.094	0.05	0.28	0.014	0.079	0.087	0.080	0.03	0.16
2sCOPE-MGJK	$\hat{\alpha}$	-0.036	0.846	0.694	0.847	0.04	0.05	-0.003	0.093	0.100	0.093	0.04	0.03	-0.002	0.079	0.086	0.079	0.02	0.02
	$\hat{\phi}$	-0.015	0.081	0.065	0.082	0.06	0.23	-0.001	0.024	0.024	0.024	0.05	0.04	-0.001	0.023	0.023	0.023	0.06	0.04
	$\hat{\beta}$	-0.024	0.203	0.197	0.205	0.06	0.12	0.012	0.082	0.090	0.082	0.03	0.13	0.012	0.074	0.085	0.075	0.03	0.14
	$\hat{\gamma}$	0.006	0.273	0.243	0.273	0.04	0.02	-0.008	0.079	0.086	0.080	0.04	0.10	-0.005	0.073	0.082	0.073	0.03	0.07

Notes: See notes to Table 3.

(asymptotically unbiased) estimation of the mean autoregressive coefficients relies on the weak exogeneity assumption, where the current error term is not correlated with the current and lagged dependent variables. It also implies that the current error term would be uncorrelated with past errors, a scenario for which we conduct MC simulations to examine performances of the 2sCOPE-MGJK estimator introduced in subsection 3.6.1 as follows.

We consider a first-order autoregressive panel model with covariates (namely, an ARX(1) panel model) as a leading example, where the first lag of the outcome variable, $y_{i,t-1}$, is one of the regressors in (25), with weak exogeneity, $E(\xi_{it}|y_{i,t-1}, y_{i,t-2}, ...) = 0$. The weakly exogenous condition essentially implies that $y_{i,t-1}$ is sufficient to represent the information set of a decision maker i at time t. As the past errors, $\xi_{i,t-h}$ for $h = 1, 2, ...$, naturally correlate with $y_{i,t-1}$, weak exogeneity further dictates the absence of serial correlation in the error processes.

$$y_{it} = \alpha_i + \phi_i y_{i,t-1} + \beta_i p_{it} + \gamma_i w_{it} + \xi_{it} \tag{25}$$

For the DGP of Cases 2.1–2.3, the contemporaneously endogenous regressor p_{it} and strictly exogenous regressor w_{it} as well as $(\alpha_i, \beta_i, \gamma_i)'$ are generated according to Cases 1.1–1.3 in subsection 4.1, respectively. For Case 2.1, we consider homogeneous AR(1) coefficients, $\phi_i = \phi_0 = 0.5$ for all i. For Cases 2.2 and 2.3, we also generate random AR(1) coefficients as $\phi_i \sim IIDUniform(0.1, 0.9)$. Tables A.2 and A.3 in the online appendix report the results for Cases 2.1 and 2.2, respectively, and the results of Case 2.3 under both regressor and slope endogeneity are summarized in Table 6.

Since in the DGP, $y_{i,t-1}$ is not correlated with either the current error term, p_{it}, or w_{it}, having $y_{i,t-1}$ as a regressor or not does not affect performances of FE and 2sCOPE-FE estimators for β and γ, comparing the results in Tables A.2, A.3, and 6 of Cases 2.1–2.3 with those in Tables 3–5 of Cases 1.1–1.3, respectively. $y_{i,t-1}$ can be viewed as an exogenous regressor which is not correlated with the endogenous regressor. Thus, the estimated AR(1) coefficient of $y_{i,t-1}$ is not affected by whether the regression is augmented by the copula generated regressor.[10] As can be seen in Tables A.2, A.3, and 6, the simulation results of FE and 2sCOPE-FE estimators for ϕ are almost identical. With homogeneous AR(1) coefficients in Case 2.1, there is a negative small-T bias for $T = 10$, which shrinks over T. But with heterogeneous AR(1) coefficients in Cases 2.2 and 2.3, there is an upward bias for relatively large $T = 50, 100$ (resembling the large T asymptotics).[11]

Now focusing on the 2sCOPE-MGJK estimator which addresses both contemporaneous endogeneity and slope endogeneity, when the Jackknife method is exploited to correct for the small-T bias particularly of $\hat{\phi}$, it introduces greater sampling errors such that the estimated standard errors and RMSEs of $\hat{\beta}$ and $\hat{\gamma}$ increases slightly. We aim to obtain asymptotically unbiased inference, despite a cost in estimation precision measured by RMSE. As the 2sCOPE-MGJK estimator always delivers a size around the 5% nominal level with a much smaller bias than the 2sCOPE-FE estimator even with homogeneous slopes, it is recommended to apply the 2sCOPE-MGJK estimator to dynamic panels.

4.3 Dynamic panel data models with serially correlated errors

For the DGP of regressor endogeneity, the MC simulations above focus on the scenario where the endogenous regressor is correlated with the current error term but not with past errors. For dynamic panels, it rules out non-zero correlations between lagged dependent variables as regressors and the current error term. Subsection 3.6.1 has introduced an estimation approach to dealing with autoregressive error processes in dynamic panels, where the 2sCOPE-MGJK estimator given by (17) is used after including further lags of regressors in the model. We assess the effectiveness of this approach by MC simulations as follows.

We first consider the AR(1) panel model given by (18), $y_{it} = \alpha_i + \phi_i y_{i,t-1} + \xi_{it}$, with $\xi_{it} = \rho_e \xi_{i,t-1} + \xi^*_{it}$, and we set $\rho_e = 0.4$. For the AR(1) coefficients, we consider both homogeneous and heterogeneous DGPs: $\phi_i = 0.5$ for all i and $\phi_i \sim IIDUniform(0.1, 0.9)$ for all i, respectively. Table 7 summarizes the MC results of both cases.

[10]This finding is consistent with that in Yang et al. (2024).

[11]This result is in line with the theoretical results on p. 725 of Pesaran (2015).

Table 7: MC results of FE, 2sCOPE-FE, and 2sCOPE-MGJK estimators in Case 3: AR(1) panels with serially correlated errors

		Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}
		$T=10, n=100$						$T=50, n=100$						$T=100, n=100$					
Case 3.1: Homogeneous AR(1) coefficients																			
FE	$\hat{\alpha}$	0.448	0.136	0.121	0.468	0.94	3.69	-0.126	0.046	0.042	0.134	0.84	3.04	-0.189	0.033	0.028	0.192	1.00	6.74
	$\hat{\tilde{\phi}}$	-0.090	0.025	0.024	0.093	0.95	3.69	0.025	0.008	0.008	0.027	0.86	3.04	0.038	0.006	0.006	0.038	1.00	6.74
	$\hat{\tilde{\delta}}$	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
2sCOPE-FE	$\hat{\alpha}$	1.086	0.179	0.158	1.100	1.00	6.86	0.172	0.059	0.053	0.182	0.88	3.23	0.083	0.041	0.036	0.093	0.60	2.31
	$\hat{\tilde{\phi}}$	-0.217	0.033	0.032	0.220	1.00	6.87	-0.034	0.011	0.011	0.036	0.90	3.23	-0.017	0.007	0.007	0.018	0.64	2.31
	$\hat{\tilde{\delta}}$	0.072	0.035	0.034	0.080	0.56	2.12	0.015	0.014	0.014	0.020	0.20	1.11	0.007	0.010	0.010	0.013	0.15	0.77
2sCOPE-MG-JK	$\hat{\alpha}$	0.181	1.494	0.755	1.505	0.06	0.24	-0.024	0.076	0.077	0.080	0.07	0.31	-0.008	0.047	0.048	0.048	0.06	0.17
	$\hat{\tilde{\phi}}$	-0.035	0.264	0.143	0.266	0.06	0.24	0.005	0.014	0.014	0.015	0.06	0.33	0.002	0.008	0.009	0.009	0.05	0.19
	$\hat{\tilde{\delta}}$	0.069	0.228	0.119	0.238	0.13	0.58	0.000	0.015	0.016	0.015	0.05	0.00	-0.001	0.011	0.010	0.011	0.05	0.06
Case 3.2: Heterogeneous AR(1) coefficients																			
FE	$\hat{\alpha}$	0.963	0.221	0.212	0.988	0.99	4.54	-0.038	0.087	0.119	0.095	0.01	0.32	-0.162	0.060	0.108	0.173	0.21	1.51
	$\hat{\tilde{\phi}}$	-0.060	0.033	0.031	0.068	0.49	1.94	0.087	0.018	0.017	0.089	1.00	4.99	0.105	0.016	0.016	0.106	1.00	6.68
	$\hat{\tilde{\delta}}$	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
2sCOPE-FE	$\hat{\alpha}$	1.691	0.283	0.263	1.715	1.00	6.42	0.211	0.107	0.134	0.237	0.35	1.57	0.048	0.074	0.119	0.088	0.01	0.40
	$\hat{\tilde{\phi}}$	-0.167	0.040	0.038	0.171	0.98	4.33	0.050	0.021	0.020	0.055	0.73	2.58	0.074	0.018	0.017	0.076	0.98	4.28
	$\hat{\tilde{\delta}}$	0.027	0.039	0.037	0.047	0.14	0.74	-0.033	0.019	0.018	0.038	0.43	1.80	-0.044	0.016	0.015	0.047	0.82	2.85
2sCOPE-MG-JK	$\hat{\alpha}$	0.230	1.310	1.076	1.330	0.06	0.21	-0.035	0.102	0.104	0.108	0.07	0.34	-0.010	0.059	0.059	0.060	0.07	0.17
	$\hat{\tilde{\phi}}$	-0.027	0.161	0.137	0.164	0.06	0.20	0.005	0.021	0.020	0.021	0.07	0.25	0.001	0.016	0.016	0.016	0.06	0.08
	$\hat{\tilde{\delta}}$	0.059	0.125	0.110	0.139	0.10	0.54	0.003	0.018	0.018	0.018	0.05	0.15	0.001	0.014	0.014	0.014	0.05	0.04

Notes: The outcome variable is generated by an AR(1) panel in (18) with $\xi_{it} = \rho_e \xi_{i,t-1} + \xi_{it}^*$ and $\rho_e = 0.4$. The AR(1) coefficients are generated as homogeneous in Case 3.1: $\phi_i = 0.5$ for all i and heterogeneous in Case 3,2 $\phi_i \sim IIDUniform(0.1, 0.9)$, respectively. The model being estimated is given by (20), with the true parameters $(\tilde{\phi}_0, \tilde{\delta}_0)' = (0.7, 0.2)'$. See also notes to Table 3.

With homogeneous AR(1) coefficients, there are substantial bias and size distortions in the FE estimator. With different values of T, its bias direction varies, due to its vulnerability to regressor endogeneity caused by serially correlated errors and the Nickell bias in estimating dynamic panels. As we augment the homogeneous AR(1) model with the copula generated regressor, we may expect both 2sCOPE-FE and 2sCOPE-MG estimator to be consistent. Unfortunately, the 2sCOPE-FE estimator still exhibits substantial bias, particularly when T is short ($T = 10$) and severe size distortions. In contrast, the 2sCOPE-MGJK estimator has negligible bias and delivers size around the 5% nominal level across different sample sizes. Note that the Jackknfe bias correction approach faces a bias-variance trade-off such that the RMSE of the 2sCOPE-MGJK estimator can be quite large with small $T = 10$. Thus, we recommend applying the Jackknife bias correction when T is relatively short compared with n. In the case of heterogeneous AR(1) coefficients, the performances of FE and 2sCOPE-FE estimators remain undesirable, with even greater bias and RMSEs for $T = 50, 100$. However, the performance of the 2sCOPE-MGJK estimator is shown to be robust, irrespective of whether the AR(1) coefficients are homogeneous or heterogeneous.

We further extend our investigation to an ARX(1) panel model given by

$$y_{it} = \alpha_i + \phi_i y_{i,t-1} + \beta_i p_{it} + \gamma_i w_{it} + \xi_{it}, \tag{26}$$

with $\xi_{it} = \rho_e \xi_{i,t-1} + \xi^*_{it}$. The estimation approach in subsection 3.6.1 requires us to estimate the following model given by

$$y_{it} = \tilde{\alpha}_i + \phi_{i1} y_{i,t-1} + \beta_i p_{it} + \gamma_i w_{it} + \phi_{i2} y_{i,t-2} + \beta_{i1} p_{i,t-1} + \gamma_{i1} w_{i,t-1} + (\delta_p \epsilon_{it,p} + u_{it}), \tag{27}$$

where $\phi_{i1} = \phi_i + \rho_e$, $\phi_{i2} = -\rho_e \phi_i$, $\beta_{i1} = -\rho_e \beta_i$, $\gamma_{i1} = -\rho_e \gamma_i$, and $\epsilon_{it,p}$ is the copula-generated regressor. Accordingly, we set $\rho_e = 0.4$ and generate the heterogeneous coefficients $(\alpha_i, \phi_i, \beta_i, \gamma_i)'$ and the Gaussian copula among $(p_{it}, w_{it}, \xi_{it})'$ as in Case 2.3. Note that the number of regressors in (27) is much more than the ones in the other cases. Hence, we perform MC simulations with sample sizes of $T = 20$, 50, and 100, taking into account the implementation of the Jackknife bias correction on our estimator. The simulation results

Table 8: MC results of FE, 2sCOPE-FE, and 2sCOPE-MGJK estimators in Case 3.3: ARX(1) panels with heterogeneous coefficients and serially correlated errors

		Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}
		$T=20, n=100$						$T=50, n=100$						$T=100, n=100$					
FE	$\hat{\alpha}$	0.665	0.301	0.230	0.730	0.78	2.89	0.677	0.249	0.204	0.721	0.86	3.32	0.694	0.228	0.195	0.731	0.92	3.56
	$\hat{\phi}_1$	-0.282	0.034	0.031	0.284	1.00	8.96	-0.240	0.029	0.028	0.242	1.00	8.57	-0.224	0.028	0.027	0.226	1.00	8.36
	$\hat{\beta}$	0.515	0.079	0.087	0.521	1.00	5.89	0.493	0.074	0.082	0.498	1.00	6.04	0.479	0.070	0.080	0.484	1.00	5.97
	$\hat{\gamma}$	-0.242	0.105	0.107	0.264	0.62	2.27	-0.263	0.086	0.091	0.277	0.83	2.89	-0.270	0.076	0.085	0.281	0.91	3.19
2sCOPE-FE	$\hat{\alpha}$	0.418	0.298	0.240	0.514	0.43	1.75	0.377	0.220	0.191	0.437	0.52	1.97	0.377	0.193	0.177	0.423	0.57	2.13
	$\hat{\phi}_1$	-0.281	0.049	0.045	0.285	1.00	6.25	-0.227	0.040	0.038	0.230	1.00	6.00	-0.209	0.037	0.035	0.212	1.00	5.98
	$\hat{\beta}$	0.072	0.121	0.126	0.141	0.08	0.57	0.061	0.085	0.093	0.105	0.09	0.66	0.047	0.076	0.086	0.089	0.06	0.55
	$\hat{\gamma}$	0.062	0.123	0.125	0.138	0.07	0.49	0.025	0.091	0.096	0.094	0.05	0.26	0.012	0.080	0.087	0.081	0.03	0.14
	$\hat{\phi}_2$	0.202	0.028	0.027	0.204	1.00	7.61	0.208	0.021	0.020	0.209	1.00	10.61	0.210	0.018	0.017	0.211	1.00	12.65
	$\hat{\beta}_1$	0.382	0.063	0.062	0.387	1.00	6.20	0.314	0.050	0.048	0.318	1.00	6.49	0.292	0.045	0.043	0.296	1.00	6.77
	$\hat{\gamma}_1$	-0.354	0.077	0.071	0.362	1.00	4.95	-0.290	0.056	0.054	0.295	1.00	5.32	-0.267	0.050	0.047	0.271	1.00	5.64
2sCOPE-MGJK	$\hat{\alpha}$	-0.039	0.430	0.401	0.432	0.05	0.10	-0.009	0.093	0.097	0.093	0.05	0.09	-0.002	0.065	0.067	0.065	0.06	0.03
	$\hat{\phi}_1$	0.034	0.064	0.064	0.073	0.07	0.53	0.005	0.030	0.029	0.030	0.07	0.16	0.000	0.025	0.025	0.025	0.06	0.01
	$\hat{\beta}$	0.012	0.123	0.129	0.123	0.04	0.09	0.013	0.082	0.091	0.083	0.03	0.15	0.012	0.074	0.085	0.075	0.03	0.14
	$\hat{\gamma}$	-0.012	0.128	0.130	0.128	0.04	0.09	-0.010	0.080	0.086	0.081	0.04	0.11	-0.006	0.073	0.082	0.074	0.03	0.07
	$\hat{\phi}_2$	-0.018	0.046	0.045	0.050	0.06	0.39	-0.003	0.016	0.015	0.016	0.05	0.18	0.000	0.012	0.012	0.012	0.05	0.01
	$\hat{\beta}_1$	-0.035	0.115	0.113	0.120	0.06	0.31	-0.008	0.043	0.044	0.043	0.05	0.18	-0.001	0.033	0.037	0.033	0.03	0.02
	$\hat{\gamma}_1$	0.047	0.126	0.125	0.135	0.06	0.38	0.007	0.044	0.045	0.044	0.05	0.16	0.001	0.033	0.037	0.033	0.03	0.02

Notes: The outcome variable is generated by an ARX(1) panel in (26) with $\xi_{it} = \rho_e \xi_{i,t-1} + \xi_{it}^*$ and $\rho_e = 0.4$. The model being estimated is given by (27), with the true parameters $(\phi_{1,0}, \beta_0, \gamma_0, \phi_{2,0}, \beta_{1,0}, \gamma_{1,0})' = (0.9, 1, -1, -0.2, -0.4, 0.4)'$. See also notes to Table 3.

are summarized in Table 8. Remarkably, even with the presence of slope, contemporaneous, and dynamic endogeneity, the 2sCOPE-MGJK estimator continues to deliver valid inference around the 5% level for $n = 100$, while the FE and 2sCOPE-FE estimators continue to be severely biased. With larger n, the 2sCOPE-MGJK is expected to perform better as the estimation precision is enhanced.

5 An application of price elasticity on Dominick's scanner data

Price elasticity is a key factor for managers to optimize the marketing mix strategy and for policymakers to evaluate taxation policies and regulations. While correlation analysis provides some information, its insights are limited. Moreover, these estimates fail to yield credible causal estimates, when prices are confounded by unobservables in the errors, or when correlated heterogeneity in consumer response parameters is disregarded.

Using Dominick's scanner data from 1990–1994, we apply the 2sCOPE-MG method to estimate average price elasticities and promotion effects. In subsection 5.1, we first illustrate the use of the 2sCOPE-MG estimator in the cereal category following the guidelines in subsection 3.7, and by comparison, we show the bias magnitudes in the FE estimates. Also, we discuss factors influencing store-specific price elasticities and promotion effects. Then in subsection 5.2, we examine the presence of regressor and/or slope endogeneity across 21 categories systematically.

5.1 2sCOPE-MG v.s. alternative estimators: the cereal category

For the cereal category, we construct a balanced panel data set of 80 stores observed over 170 weeks. Given the size information of different products with the universal product code (UPC) as an identifier, we first standardize prices and sales quantities based on each UPC's size information, then compute the aggregate sales and market-share weighted prices and promotions (including bonus promotion and direct price reduction) at the store level. The variables related to promotions are represented as dummy variables at the store-UPC level. When weighted by market shares, they indicate the proportions of sales associated with each promotion at the store level. The summary statistics of the outcome and independent variables are reported in Table 9, computed before demeaning and detrending the variables.

Table 9: Summary statistics of outcome and independent variables of 80 stores over 170 weeks in the cereal category

	Mean	SD	Min	25% Quantile	Median	75% Quantile	Max	No. obs.
log(Sales)	8.219	0.363	5.374	7.967	8.229	8.468	9.536	13,600
log(Price)	0.789	0.087	-0.067	0.753	0.794	0.843	1.004	13,600
Bonus	0.119	0.117	0.000	0.046	0.087	0.154	1.000	13,600
Price reduction	0.061	0.095	0.000	0.000	0.017	0.087	0.700	13,600

We consider a dynamic sales response model, which accommodates potential effects of past outcomes on the current outcome. Such a dynamic structure is suitable when the persistence of outcome processes, like sales, is a parameter of interest itself. It also enables researchers to disentangle short-run and long-run effects. But for consistent estimation of dynamic panels, relatively large T panels are required such as the one used in our empirical application. The logarithm of sales of store i in week t, $\log(Sales_{it})$, varies according to

$$\log(Sales_{it}) = \alpha_i+\tau_t+\phi_i \log(Sales_{i,t-1})+\beta_{i1} \log(Price_{it})+\beta_{i2}Bonus_{it}+\beta_{i3}PriceRedu_{it}+\xi_{it}, \tag{28}$$

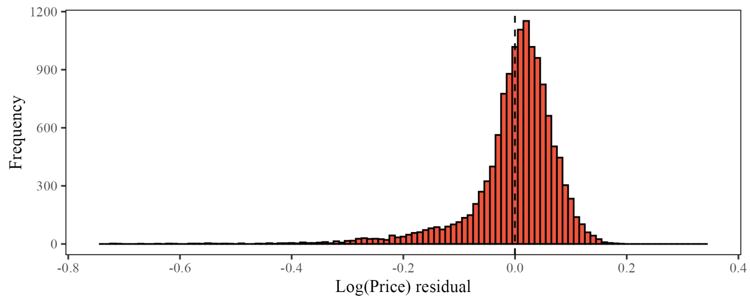
where $\log(Price_{it})$, $Bonus_{it}$, and $PriceRedu_{it}$ are the logarithm of market-share weighted price, and market-share weighted bonus and price reduction of store i in week t, respectively, and α_i and τ_t denote store- and time-fixed effects, respectively. The structural error term, ξ_{it}, is possibly correlated with the price due to unobserved demand shocks among other confounders, as shown in Park and Gupta (2012), Haschka (2022), and Yang et al. (2024). On the other hand, following Sriram et al. (2007), we consider bonus and price reduction promotions as exogenous due to the typical quarterly decision-making process and the lead time required for implementation. We allow for heterogeneous coefficients for all regressors.

The 2sCOPE-MG approach does not rely on any IVs. However, as a commonly used alternative to handle endogeneity, we also explore estimation using an IV for price. The IV is constructed based on the average prices of UPCs (with positive sales throughout the sample periods) over different stores for each week. Then for each store, we aggregate the prices of

UPCs using market shares averaged over the first 26 weeks (excluded from estimation) as weights. The weights are predetermined, not reacting to demand shocks in the later period.[12] Given possible time trends in sales and prices, we use detrended log sales and log price (and its instrument) in estimation. The correlation coefficient between the detrended IV and price is 0.362, thereby meeting the relevance condition.

We first validate the non-normality assumption using the data on demeaned price (after detrending), bonus, and price reduction. For all three regressors, the resulting p-values are close to zero up to seven decimal points, indicating significant non-normality. The empirical distribution of log price residuals exhibits left-skewness with a heavy tail on the left-hand side (Figure 1). Now we can proceed with 2sCOPE-MG estimation.

Figure 1: Histogram of the distribution of log price residuals



Notes: The histogram shows the empirical distribution of $Log(Price_{it})$ residuals after removing store-specific means and trends.

Table 10 presents estimated average price elasticity and promotion effects in Model (28). Columns (1) and (2) display results of applying FE and MG estimators directly. Columns (3)–(4) and (5)–(6) show FE and MG estimation results using copula-generated regressors

[12]This IV is constructed as the Bartik instrument in Goldsmith-Pinkham et al. (2020) (on p. 2592). For estimation consistency, a sufficient assumption requires that the predetermined weights (or initial shares) are strictly exogenous to *changes* in the outcome variable over time. In our case, these weights are linked to UPCs, with a large number of UPCs used in the calculation. It is unlikely that these weights directly affect sales changes in later periods, particularly when the regressors, market-share weighted prices and promotions, already account for the effects of current-period market shares. An example of the Bartik IV can also be found in Li et al. (2014) (p. 318). Moreover, Chevalier et al. (2003) found that prices, on average, tend to be countercyclical and largely varied with retail margins, indicating that variations in average prices are largely driven by wholesale costs. The UPCs are typically more established, and according to Nevo (2001), they are less susceptible to systematic demand shocks. These, combined with the predetermined weights, lend greater credence to the satisfaction of the exclusion restriction.

Table 10: Estimated average price elasticity and promotion effects on cereal sales in a dynamic sales response model

			Copula generated regressors		IV	
	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	FE	MG	2sCOPE-FE	2sCOPE-MG	FE-IV	MG-IV
$\log(Price_{it})$	-1.200	-1.089	-1.700	-1.550	-1.674	-1.478
	(0.056)	(0.022)	(0.089)	(0.100)	(0.131)	(0.049)
$Bonus_{it}$	0.205	0.199	0.160	0.160	0.193	0.215
	(0.024)	(0.020)	(0.028)	(0.032)	(0.023)	(0.022)
$PriceRedu_{it}$	0.153	0.150	0.091	0.103	0.114	0.125
	(0.030)	(0.014)	(0.033)	(0.033)	(0.027)	(0.023)
$\log(Sales_{i,t-1})$	0.089	0.117	0.089	0.116	0.067	0.112
	(0.021)	(0.015)	(0.021)	(0.016)	(0.017)	(0.016)
Store fixed effects	Y	Y	Y	Y	Y	Y
Week fixed effects	Y	Y	Y	Y	Y	Y
Test of price endogeneity						
Pearson cor.	-	-	0.269	0.222	-	-
p-value	-	-	0.000	0.000	-	-
Slope endogeneity	-	Y	-	Y	-	Y
Regressor endogeneity	-	-	Y	Y	Y	Y
No. observations	13,520	13,520	13,520	13,520	11,440	11,440

Notes: The estimates are computed based on a balanced panel of 80 stores over 169 weeks of the cereals category from Dominick's database. The dynamic sales response model is given by (28). The coefficients of the Gaussian copula model are assumed to be homogeneous across stores. (i) Price, bonus, and price reduction are computed as market share weighted averages over UPCs sold in each store. (ii) $\log(Sales_{it})$ and $\log(Price_{it})$ are detrended prior to estimation with linear and quadratic trends, respectively. (iii) To construct an IV for the price, we consider the weekly prices of UPCs average over different stores (for those UPCs whose prices are observed over all periods) and further aggregate the prices over UPCs for each store with "predetermined" weights. The weights are computed as market shares of these UPCs in each store average over the first 26 weeks, which are excluded from the sample used in IV estimation ($T = 143$). The first stage regression includes time fixed effects and exogenous regressors and assumes homogeneous and heterogeneous slopes of the IV for FE-IV and MG-IV estimators, respectively.

and the IV for price, respectively. Standard errors are calculated using consistent estimators of asymptotic variance, given the respective underlying assumptions.

We first examine the effects of price endogeneity on estimation outcomes. Given the potential bias from slope endogeneity in the FE approach, we focus on comparing MG, 2sCOPE-MG, and MG-IV estimates. Similar patterns are likewise observed in the FE estimation outcomes. The MG estimated price elasticity, assuming price exogeneity, is lower compared to the 2sCOPE-MG and MG-IV estimates, which account for price endogeneity.

Specifically, the average price elasticity shifts from -1.089 (0.022) to -1.478 (0.049) based on the MG-IV estimator. We denote standard errors in brackets throughout the paper. Remarkably, even without external exogenous variations provided by an IV, the 2sCOPE-MG estimator recovers the causal average price effect at -1.550 (0.1), which closely aligns with the MG-IV estimate. The average bonus effect and sales persistence estimates remain comparable whether adjusting for price endogeneity or not. Conversely, the average effect of price production decreases slightly after accounting for price endogeneity at 0.103 (0.033) of the 2sCOPE-MG estimator, showing contamination bias.

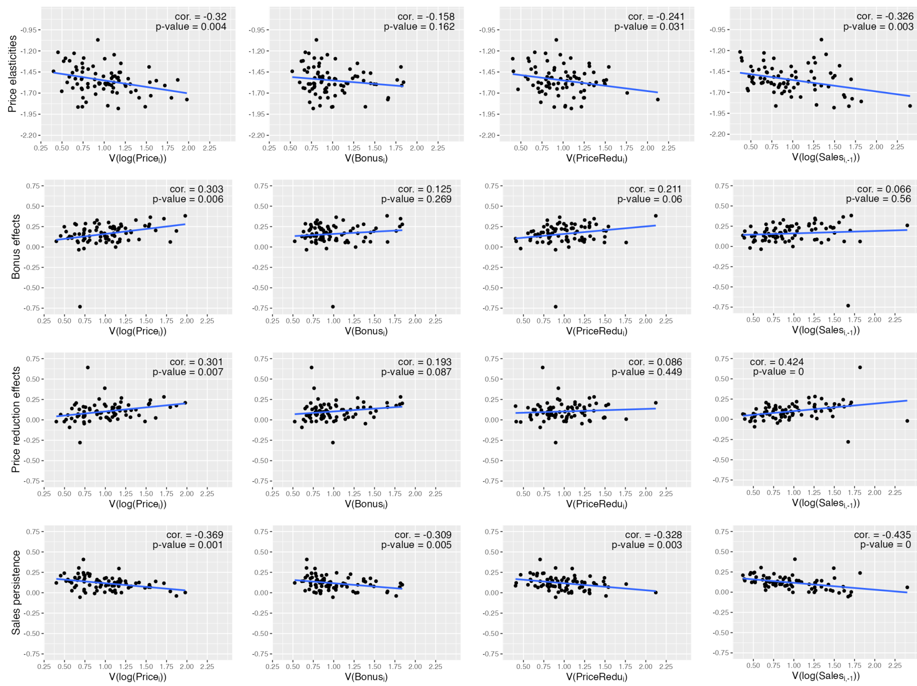
Moreover, the 2sCOPE-MG method can be utilized to measure correlations between endogenous variables and the structural error. The Pearson correlation coefficient between price and estimated structural error terms is 0.222, with a near-zero p-value, strongly rejecting the hypothesis of no price endogeneity. The positive correlation suggests that as prices rise, store managers often intensify efforts to enhance consumers' willingness to purchase, for example, by offering higher-quality products, to counteract potential negative price effects. As researchers may not always have access to observations of the kind of variables, the 2sCOPE-MG approach is equipped to handle their absence when estimating causal effects.

We now address the issue of slope endogeneity, the presence of which leads to disparities between FE and MG estimates. As demonstrated in Section A.2.3 of the online appendix, under the assumption of regressor exogeneity, the FE estimator can be written as a weighted average of unit-specific OLS estimators. These weights are proportional to the *variance-covariance* matrix of each unit's regressor processes, capturing within-unit variations like price and promotion activities over time. For the current application, stores with larger temporal variations in their regressor processes receive higher weights on their store-specific estimates. However, non-zero correlations between these weights and the heterogeneous slope coefficients introduce bias in the FE estimation, while the different weights may not make a difference with homogeneous coefficients.

Under the assumption of heterogeneous slope coefficients, the 2sCOPE-MG estimator shows a smaller-sized price elasticity compared with the 2sCOPE-FE estimator. This observation remains consistent when comparing MG vs FE and MG-IV vs FE-IV, regardless of whether price endogeneity is considered. The presence of an upward neglected heterogeneity bias implies a negative correlation between within-store regressor variations and store-specific estimates. In other words, stores with more frequent or intense price variations or promotional activities exhibit higher store-specific price elasticities. Given that consumers frequently visit retail stores, they are more likely to make purchases at lower prices or during promotions when they expect such activities to occur more frequently or encounter greater price variations. Conversely, when consumers visit stores with fewer price variations or promotions, their purchase decisions may exhibit less variability across different prices. On the other hand, the 2sCOPE-MG and 2sCOPE-FE estimates of bonus and price reduction promotion effects are comparable, with a mild difference in sales persistence, indicating homogeneous effects across different stores. The bonus effect is at 0.16, slightly higher than the price reduction effect, (remembering that bonus and price reduction variables are measured in terms of the proportions of brands under promotion in a store). Tables A.4 and A.5 in the online appendix present estimation results of other models. The biases due to regressor and slope endogeneity remain consistent across three model specifications.

In light of the direction of slope endogeneity bias, Figure 2 illustrates the relationship between the 2sCOPE-MG estimated store-specific coefficients and scaled within-store variance of regressors. The correlation coefficient and the associated p-value between each pair of coefficients and regressor variations are presented in the upper right corner of each subplot. In the first row, the estimated price elasticities exhibit substantial heterogeneity across stores, ranging between -1.882 and -1.07. The correlation coefficients between price elasticities and within-store variations in price and price reduction promotion are significant, at -0.32 and -0.241, respectively. The results provide further evidence that the intensity of price variation and promotion frequency can affect consumers' response parameters as shown in Kalwani et al. (1990) and Fok et al. (2006). The fact that consumers' price consciousness rises with the degree of variations in price and discount can be rationalized by a learning model of price formation or a reference price adaptation model as in the literature.

Figure 2: Scatter plots of 2sCOPE-MG store-specific estimates on within-store regressor variations



Notes: The x-axis is computed as $\frac{1/T\sum_{t=1}^{T}\left(x_{it,(j)}-\bar{x}_{i,(j)}\right)^2}{1/(nT)\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it,(j)}-\bar{x}_{i,(j)}\right)^2}$ with $\bar{x}_{i,(j)} = \frac{1}{T}\sum_{t=1}^{T} x_{it,(j)}$ for each regressor j, denoted as $V(\log(Price_i))$, $V(Bonus_i)$, $V(PriceRedu_i)$, and $V(\log(Sales_{i,-1}))$, respectively. In each subplot, "cor." and "p-value" denote the Pearson correlation coefficient between the 2sCOPE-MG store-specific estimates and within-store regressor variations, and the associated p-value. See notes under Table 10 for the model and estimation procedure.

For store-specific bonus and price reduction effects in the second and third rows, the variation across different stores is relatively modest, leading to close estimates between the 2sCOPE-MG and 2sCOPE-FE estimators. Since promotion activities are typically planned and implemented uniformly across different stores, it is expected that they have rather homogeneous impacts. However, we still observe significant correlations between price variations and responses to promotions. Specifically, as price variations increase, the influences of promotions on consumer purchasing decisions are strengthened. It is also worth noting that the effect of each promotion does not vary much with its own degree of variation, but with the degree of variations in the other types of promotion activity. In the last row, the estimated autoregressive coefficients of sales are negative for certain stores. This suggests two counteracting effects: one leads to persistence in sales, and another one drives sales back to the long-term level. When prices and promotional activities exhibit greater variability, it is more likely to observe mean reversions in sales. It is important to note that these two kinds of sales dynamics have significantly distinct implications for forecasting future sales. It can be seen in the last column that not only the mean but the variance of sales varies with price and promotions. As expected, greater price elasticity, greater responses to price reduction, and less persistence are tied to increased sales volatility.

As sales response parameters are often modelled as linear in store-specific regressor mean levels in the HB approach, we also examine the correlations between store-specific coefficients and regressor levels, shown in Figure A.1 in the online appendix. The results reveal that price sensitivity and bonus effects increase with the proportion of brands sold under price reduction promotions, but remain relatively constant with price and bonus levels. The impact of price reduction diminishes with lower price levels, which may explain why price reduction is not as effective as marketing managers desire. Furthermore, sales persistence tends to decrease with a larger fraction of brands under bonus or price reduction promotions. Based on the 2sCOPE-MG store-specific estimates, store managers can evaluate whether the current sales response function and marketing mix strategy are desirable to maximize revenues/profits. In cases where an adjustment is needed, the above analysis suggests directions to proceed.

5.2 Regressor and slope endogeneity across different categories

Table 11 summarizes the 2sCOPE-MG estimation results across 21 categories in Dominick's scanner data. The corresponding results based on the FE method (not accounting for either endogeneity issues) are shown in Table A.6 in the online appendix. First, we find prevalent price-error dependence, specifically a significant positive correlation between price and the structural error term across 19 categories. Failing to account for this regressor endogeneity properly results in *attenuation* bias in the estimates of price elasticity. Second, 18 categories show non-negligible correlated heterogeneity in store-specific price elasticities.

Table 11: 2sCOPE-MG estimated average price elasticity, promotion effects on sales, and sales persistence in a dynamic sales response model across 21 categories using a homogeneous Gaussian copula

	Sample size		Price elasticity				Bonus		Price reduction		Sales persistence	
Category	No. stores	No. weeks	Estimate	$\hat{se}$	Regressor endo.	Slope endo.	Estimate	$\hat{se}$	Estimate	$\hat{se}$	Estimate	$\hat{se}$
Food categories												
Beer	67	101	-1.056	(0.155)	+	+	0.163	(0.046)	0.438	(0.151)	0.202	(0.026)
Bottled juices	81	165	-1.565	(0.099)	+	-	0.088	(0.021)	0.149	(0.035)	0.162	(0.018)
Cereals	80	169	-1.550	(0.100)	+	-	0.160	(0.032)	0.103	(0.033)	0.116	(0.016)
Cheeses	81	170	-2.388	(0.290)	+	-	0.035	(0.026)	0.013	(0.024)	0.164	(0.022)
Cookies	81	159	-1.256	(0.101)	+	-	0.107	(0.026)	0.215	(0.037)	0.176	(0.019)
Crackers	81	155	-0.845	(0.102)	+	-	0.249	(0.039)	0.313	(0.051)	0.161	(0.018)
Canned soup	81	162	-1.145	(0.203)	+	-	0.167	(0.047)	0.197	(0.051)	0.169	(0.030)
Frozen dinners	80	173	-1.973	(0.060)	+	-	0.041	(0.024)	0.235	(0.030)	0.150	(0.014)
Frozen entrees	81	170	-1.052	(0.131)	+	-	0.261	(0.042)	0.536	(0.061)	0.134	(0.023)
Oatmeal	79	133	-1.417	(0.098)	+	0	-0.115	(0.030)	0.025	(0.045)	0.196	(0.029)
Soft drinks	80	171	-1.274	(0.227)	+	-	0.074	(0.038)	0.299	(0.045)	0.106	(0.017)
Canned tuna	81	165	-1.309	(0.102)	+	-	-0.039	(0.027)	0.081	(0.035)	0.121	(0.016)
Non-food categories												
Analgesics	81	173	-1.096	(0.194)		+	0.277	(0.074)	0.179	(0.068)	0.159	(0.019)
Dish detergent	81	166	-1.548	(0.204)	+	-	0.112	(0.027)	0.203	(0.042)	0.103	(0.017)
Grooming products	80	133	-0.897	(0.087)		-	0.112	(0.029)	0.461	(0.052)	0.202	(0.029)
Laundry detergents	81	170	-1.984	(0.133)	+	-	0.109	(0.028)	0.197	(0.045)	0.083	(0.016)
Paper towels	80	177	-1.746	(0.122)	+	-	0.048	(0.021)	0.160	(0.025)	0.102	(0.014)
Shampoos	80	121	-0.965	(0.118)	+	0	0.115	(0.044)	0.363	(0.059)	0.206	(0.021)
Soaps	80	175	-1.205	(0.326)	+	0	0.067	(0.026)	0.155	(0.068)	0.264	(0.021)
Toothpastes	81	170	-1.933	(0.138)	+	+	0.242	(0.041)	0.244	(0.056)	0.123	(0.014)
Bathroom tissues	81	166	-1.744	(0.106)	+	+	0.119	(0.025)	0.118	(0.032)	0.076	(0.017)
Average bias			0.258	(0.235)	0.344	-0.135	0.028	(0.035)	0.091	(0.053)	0.009	(0.045)

Notes: This table summarizes the estimation results of the dynamic sales response model in (28), using a balanced store-week panel dataset for each of 21 categories. The results are based on the 2sCOPE-MG estimator, assuming homogeneous coefficients in the Gaussian copula model across stores. In the regressor endogeneity column, "+" denotes a significant positive correlation between price and estimated structural errors, leading to underestimation bias. In the slope endogeneity column, "+" or "-" indicates whether the 2sCOPE-MG estimate is higher or lower than the 2sCOPE-FE estimate, respectively, while "0" denotes an absolute difference smaller than 0.1. In the last row, bias is calculated as the FE estimate minus the 2sCOPE-MG estimate, averaged over 21 categories, with standard deviations of biases shown in brackets. For price elasticity, regressor endogeneity bias is computed as the MG estimate minus the 2sCOPE-MG estimate, and slope endogeneity bias is computed as the 2sCOPE-FE estimate minus the respective 2sCOPE-MG estimate.

The slope endogeneity bias, calculated as the difference between 2sCOPE-FE and 2sCOPE-MG estimates, is negative in 14 categories. As discussed earlier, this indicates that consumer price elasticity increases with price variability in these categories. The 2sCOPE-MG estimates of average price elasticity range from -2.388 (the cheese category) to -0.845 (the cracker category) across categories. Hence, not only do we find substantial cross-store heterogeneity, there presents cross-category heterogeneity. Overall, the bias due to both endogeneity (calculated as the difference between FE and 2sCOPE-MG estimates) ranges from 0.007 to 0.904 across categories with a median of 0.187, highlighting potentially large deviations from the intended targets when setting marketing mix variables based on the biased estimates.[13] Additionally, our estimation results show that consumers, in general, preferred price reductions over bonuses, given the higher estimated effects of price reduction compared to bonus effects. The last two columns provide evidence of sales persistence. The estimates range from 0.076–0.264, which is reasonable given the relatively high-frequency weekly data versus monthly or annual data.

The results above assume homogeneous coefficients in the Gaussian copula across stores. Table A.8 in the online appendix presents 2sCOPE-MG estimates with store-specific Gaussian copula coefficients. The estimates for each category are very similar to those in Table 11, supporting the homogeneity assumption. It also shows the robustness of the 2sCOPE-MG estimator. Averaged across the 21 categories, the total, regressor, and slope endogeneity bias are approximately 0.25, 0.34, and -0.14, respectively.

6 Conclusions

The undesirable bias caused by regressor endogeneity and slope endogeneity has been widely recognized in the two strands of the literature as reviewed. To our knowledge, the proposed estimator in this study is the first remedy for both types of endogeneity biases in the estimation of the average effects for panel regressions without IVs. Given the general location

[13] The overall bias is somewhat reduced as positive bias from regressor endogeneity is offset by negative bias from slope endogeneity in most categories. Note that the biases can have potentially opposite signs.

Gaussian copula model with possibly heterogeneous dependence structure, the 2sCOPE-MG estimator can cope with contemporaneous regressor endogeneity, which challenges the validity of hypothesis testing using observational data. The homogeneity assumption on slope coefficients is also relaxed, which is less realistic as individual responses may be governed by different parameters. Treated as unit-specific fixed parameters, the heterogeneous slope coefficients can be functions of individual characteristics and depend on the relative magnitudes of changes and even the entire path of dependent and independent variables.

By a comprehensive set of MC simulations, we illustrate the use of the 2sCOPE-MG and 2sCOPE-MGJK estimators in static and dynamic panels with contemporaneous regressor endogeneity, correlated slope heterogeneity, and even dynamic misspecification, separately and jointly. While the FE and 2sCOPE-FE estimators exhibit severe biases and size distortions, the 2sCOPE-MG and 2sCOPE-MGJK estimators are shown to provide unbiased inferences.

Using Dominick's scanner data, we apply the 2sCOPE-MG approach to consistently estimate the average causal effects of price and promotions in dynamic sales response panel models, addressing both regressor and slope endogeneity without IVs. We highlight the ubiquitous presence of regressor and slope endogeneities in the conventional estimation method across different categories, and the resulting bias can be substantial. Our findings complement the existing studies on consumer heterogeneity in the literature, suggesting that micro-marketing pricing strategies at the store level need to account for both the level and variability of past prices and promotional activities. Specifically, sales and marketing managers could integrate the indirect effects of past pricing and advertising strategies on sales response parameters into their future pricing and marketing strategies.

References

Allenby, G. M. and Rossi, P. E. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics*, 89(1-2):57–78.

Andrews, R. L., Ainslie, A., and Currim, I. S. (2002). An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity. *Journal of Marketing Research*, 39(4):479–487.

Ansari, A., Mela, C. F., and Neslin, S. A. (2008). Customer channel migration. *Journal of Marketing Research*, 45(1):60–76.

Bijmolt, T. H., Van Heerde, H. J., and Pieters, R. G. (2005). New empirical generalizations on the determinants of price elasticity. *Journal of Marketing Research*, 42(2):141–156.

Blake, T., Nosko, C., and Tadelis, S. (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica*, 83(1):155–174.

Breitung, J., Mayer, A., and Wied, D. (2024). Asymptotic properties of endogeneity corrections using nonlinear transformations. *The Econometrics Journal*, 27:362–383.

Bucklin, R. E. and Sismeiro, C. (2003). A model of web site browsing behavior estimated on clickstream data. *Journal of Marketing Research*, 40(3):249–267.

Chen, Y., Iyengar, R., and Iyengar, G. (2017). Modeling multimodal continuous heterogeneity in conjoint analysis—a sparse learning approach. *Marketing Science*, 36(1):140–156.

Chevalier, J. A., Kashyap, A. K., and Rossi, P. E. (2003). Why don't prices rise during periods of peak demand? Evidence from scanner data. *American Economic Review*, 93(1):15–37.

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

Christopoulos, D., McAdam, P., and Tzavalis, E. (2021). Dealing with endogeneity in threshold models using copulas. *Journal of Business & Economic Statistics*, 39(1):166–178.

Chudik, A. and Pesaran, M. H. (2019). Mean group estimation in presence of weakly cross-correlated estimators. *Economics Letters*, 175:101–105.

Danaher, P. J. and Smith, M. S. (2011). Modeling multivariate distributions using copulas: Applications in marketing. *Marketing Science*, 30(1):4–21.

Dubois, P., Griffith, R., and O'Connell, M. (2020). How well targeted are soda taxes? *American Economic Review*, 110(11):3661–3704.

Ebbes, P., Liechty, J. C., and Grewal, R. (2015). Attribute-level heterogeneity. *Management Science*, 61(4):885–897.

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

Eckert, C. and Hohberger, J. (2023). Addressing endogeneity without instrumental variables: An evaluation of the gaussian copula approach for management research. *Journal of Management*, 49(4):1460–1495.

Esteves, R.-B. and Resende, J. (2016). Competitive targeted advertising with price discrimination. *Marketing Science*, 35(4):576–587.

Evgeniou, T., Pontil, M., and Toubia, O. (2007). A convex optimization approach to modeling consumer heterogeneity in conjoint estimation. *Marketing Science*, 26(6):805–818.

Fiebig, D. G., Keane, M. P., Louviere, J., and Wasi, N. (2010). The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science*, 29(3):393–421.

Fok, D., Horváth, C., Paap, R., and Franses, P. H. (2006). A hierarchical Bayes error correction model to explain dynamic effects of price changes. *Journal of Marketing Research*, 43(3):443–461.

Germann, F., Ebbes, P., and Grewal, R. (2015). The chief marketing officer matters! *Journal of Marketing*, 79(3):1–22.

Goldfarb, A. and Tucker, C. (2011). Online display advertising: Targeting and obtrusiveness. *Marketing Science*, 30(3):389–404.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624.

Haschka, R. E. (2022). Handling endogenous regressors using copulas: A generalization to linear panel models with fixed effects and correlated regressors. *Journal of Marketing Research*, 59(4):860–881.

Hoban, P. R. and Bucklin, R. E. (2015). Effects of internet display advertising in the purchase funnel: Model-based insights from a randomized field experiment. *Journal of Marketing Research*, 52(3):375–393.

Iyer, G., Soberman, D., and Villas-Boas, J. M. (2005). The targeting of advertising. *Marketing Science*, 24(3):461–476.

Kalwani, M. U., Yim, C. K., Rinne, H. J., and Sugita, Y. (1990). A price expectations model of customer brand choice. *Journal of Marketing Research*, 27(3):251–262.

Klaassen, C. A. and Wellner, J. A. (1997). Efficient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli*, 3(1):55–77.

Lewbel, A., Schennach, S. M., and Zhang, L. (2024). Identification of a triangular two equation system without instruments. *Journal of Business & Economic Statistics*, 42(1):14–25.

Li, K. J. (2018). Behavior-based pricing in marketing channels. *Marketing Science*, 37(2):310–326.

Li, S., Linn, J., and Muehlegger, E. (2014). Gasoline taxes and consumer behavior. *American Economic Journal: Economic Policy*, 6(4):302–342.

Luan, Y. J. and Sudhir, K. (2010). Forecasting marketing-mix responsiveness for new products.

Journal of Marketing Research, 47(3):444–457.

Manchanda, P., Rossi, P. E., and Chintagunta, P. K. (2004). Response modeling with nonrandom marketing-mix variables. *Journal of Marketing Research*, 41(4):467–478.

Mellon, J. (2024). Rain, rain, go away: 194 potential exclusion-restriction violations for studies using weather as an instrumental variable. *American Journal of Political Science*.

Montgomery, A. L., Li, S., Srinivasan, K., and Liechty, J. C. (2004). Modeling online browsing and path analysis using clickstream data. *Marketing Science*, 23(4):579–595.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342.

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

Park, S. and Gupta, S. (2024). A review of copula correction methods to address regressor–error correlation. *Impact at JMR*.

Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics*. Oxford University Press.

Pesaran, M. H. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68:79–113.

Pesaran, M. H. and Yamagata, T. (2008). Testing slope homogeneity in large panels. *Journal of Econometrics*, 142(1):50–93.

Pesaran, M. H. and Yang, L. (2024). Trimmed mean group estimation of average effects in ultra short T panels under correlated heterogeneity. arXiv preprint arXiv:2310.11680.

Qian, Y., Koschmann, A., and Xie, H. (2024). A practical guide to endogeneity correction using copulas. Technical report, NBER Working Paper 32231, http://www.nber.org/papers/w32231.

Qian, Y. and Xie, H. (2011). No customer left behind: A distribution-free Bayesian approach to accounting for missing Xs in marketing models. *Marketing Science*, 30(4):717–736.

Qian, Y. and Xie, H. (2024). Correcting regressor-endogeneity bias via instrument-free joint estimation using semiparametric odds ratio models. *Journal of Marketing Research*, 61(5):914–936.

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

Rossi, P. E. and Allenby, G. M. (2003). Bayesian statistics and marketing. *Marketing Science*, 22(3):304–328.

Rossi, P. E., McCulloch, R. E., and Allenby, G. M. (1996). The value of purchase history data in target marketing. *Marketing Science*, 15(4):321–340.

Smith, A. N., Seiler, S., and Aggarwal, I. (2023). Optimal price targeting. *Marketing Science*, 42(3):476–499.

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

Van Nierop, E., Bronnenberg, B., Paap, R., Wedel, M., and Franses, P. H. (2010). Retrieving unobserved consideration sets from household panel data. *Journal of Marketing Research*, 47(1):63–74.

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Review of Economics and Statistics*, 87(2):385–390.

Yang, F., Qian, Y., and Xie, H. (2024). Express: Addressing endogeneity using a two-stage copula generated regressor approach. *Journal of Marketing Research*, 0(ja):https://doi.org/10.1177/00222437241296453.

Yang, L. (2023). Jackknife bias-corrected mean group estimator for heterogeneous dynamic panels with short T: An application to effects of minimum wages on employment. Working paper.

Zhao, Y., Gijbels, I., and van Keilegom, I. (2020). Inference for semiparametric Gaussian copula model adjusted for linear regression using residual ranks. *Bernoulli*, 26(4):2815–2846.

Online Appendix to

"Handling Endogenous Marketing Mix Regressors in Correlated Heterogeneous Panels with Copula Augmented Mean Group Estimation"

Liying Yang

Postdoctoral research fellow, Sauder School of Business, University of British Columbia

Yi Qian

Sauder School of Business, University of British Columbia

Hui Xie

Faculty of Health Sciences, Simon Fraser University

December 15, 2024

A.1 Introduction

The online appendix is organized as follows. Section A.2 details the assumptions for the panel data models we consider and the two-stage copula augmented mean group (2sCOPE-MG) estimator, provides proof for the theorems in the main paper, and illustrates biases in the pooled OLS and standard fixed effect estimators with correlated heterogeneous coefficients. Sections A.3 and A.4 present supplementary Monte Carlo (MC) evidence and empirical results.

A.2 Mathematical appendix

Notations. *Generic positive finite constants are denoted by C when large and c when small. They can take different values at different instances. $[K]$ denotes a set of K indices. For a symmetric matrix $\boldsymbol{A}$, $\boldsymbol{A} \succ \boldsymbol{0}$ denotes that $\boldsymbol{A}$ is positive definite. $\odot$ denotes the element-wise matrix product. Suppose $\{f_n\}_{n=1}^{\infty}$ is any real sequence and $\{g_n\}_{n=1}^{\infty}$ is a sequences of positive real numbers, then $f_n = O(g_n)$ if there exists a constant C such that $|f_n|/g_n \leq C$ for all n; and $f_n = o(g_n)$ if $f_n/g_n \to 0$ as $n \to \infty$. Similarly, $f_n = O_p(g_n)$ if f_n/g_n is stochastically bounded, and $f_n = o_p(g_n)$ if $f_n/g_n \to_p 0$. $\to_p$ denotes convergence in probability, and $\to_d$ denotes convergence in distribution.*

A.2.1 Assumptions

Assumption A.1 (Sampling)**.** *$\{\boldsymbol{y}_i, \boldsymbol{P}_i, \boldsymbol{W}_i, \boldsymbol{\theta}_i\}_{i=1}^{n}$ is an i.i.d random sample from the population distributions of interest.*

Assumption A.2 (Errors)**.** *(i) For $i = 1, 2, ..., n$ and $t = 1, 2, ..., T$, there exists a decomposition of the structural error term: $\xi_{it} = \sigma_i \xi_{it}^* + v_{it}$, where $\xi_{it}^* \sim IIDN(0, 1)$, $E(\xi_{it}^*|\boldsymbol{W}_i) = 0$, and $E(v_{it}|\boldsymbol{P}_i, \boldsymbol{W}_i) = 0$, with $0 < E(v_{it}^2|\boldsymbol{p}_{is}, \boldsymbol{w}_{is}) < C$ and $E(v_{it}^4) < C$. (ii) $E(\xi_{it}\xi_{is}) = 0$ for $t \neq s$. (iii) σ_i^2 is distributed independently of ξ_{it}^*, with $0 < \inf_i \sigma_i^2 < \sup_i \sigma_i^2 < C$.*

Assumption A.3 (Correlated random coefficients)**.** *For $i = 1, 2, ..., n$, there exists a decomposition of the correlated random coefficients: $\boldsymbol{\theta}_i = E(\boldsymbol{\theta}_i|\boldsymbol{P}_i, \boldsymbol{W}_i, \alpha_i) + \boldsymbol{\eta}_i$, where $E(\boldsymbol{\eta}_i) = \boldsymbol{0}$,*

$E(\boldsymbol{\theta}_i) = \boldsymbol{\theta}_0$ *with* $\|\boldsymbol{\theta}_0\| < C$, $Var(\boldsymbol{\theta}_i) = \boldsymbol{\Omega}_\theta \succ \mathbf{0}$ *is bounded, and* $E\|\boldsymbol{\theta}_i\|^4 < C$.

Assumption A.4 (Regressors)**.** *(i) For each* $k \in [K_p + K_w]$*, the* k*-th regressor error term,* $e_{it,k}$*, is identically and independently drawn from an absolutely continuous marginal distribution function (CDF)* F_k *for all* i *and* t*, with a Lipschitz continuous and bounded marginal density function* f_k*,* $E(e_{it,k}) = 0$*,* $Var(e_{it,k}) = \sigma_k^2 > 0$*, and* $E(e_{it,k}^4) < C$*. (ii)* $e_{it,k}$ *is uncorrelated with* $\boldsymbol{z}_{it}$ *in (3). (iii) For* $i = 1, 2, ..., n$*, there exists a* T_0 *such that for all* $T > T_0$*,* $\frac{1}{T}(\boldsymbol{P}_i, \boldsymbol{W}_i)'(\boldsymbol{P}_i, \boldsymbol{W}_i) \succ \mathbf{0}$*. As* $T \to \infty$*,* $\frac{1}{T}(\boldsymbol{P}_i, \boldsymbol{W}_i)'(\boldsymbol{P}_i, \boldsymbol{W}_i) \to_p \boldsymbol{\Sigma}_i \succ \mathbf{0}$ *with* $\sup_i \|\boldsymbol{\Sigma}_i\| < C$.

Assumption A.5 (Semiparametric Gaussian copula)**.** *(i)* $(\boldsymbol{e}_{it,p}^{*\prime}, \boldsymbol{e}_{it,w}^{*\prime}, \xi_{it}^*)'$ *follows a Gaussian copula given by (4) with a positive definite covariance matrix,* $\boldsymbol{V}_{i,\rho} \succ \mathbf{0}$ *for all* i*. (ii)* $E(\boldsymbol{V}_{i,\rho}) = \boldsymbol{V}_\rho \succ \mathbf{0}$ *with* $\|\boldsymbol{V}_\rho\| < C$*,* $Var(\boldsymbol{V}_{i,\rho}) \succeq \mathbf{0}$*, and* $E\|\boldsymbol{V}_{i,\rho}\|^4 < C$.

Assumption A.6 (Identification)**.** *The following holds for* $i = 1, 2, ..., n$*. The* K_p *endogenous regressors are divided into two sets:* $[K_{p,NG}]$ *and* $[K_{p,G}]$*. (i) For each* $k \in [K_{p,NG}]$*, the marginal distribution of the* k*-th endogenous regressor error term,* $e_{it,p(k)}$ *in (3), is not Gaussian. (ii) For each* $k \in [K_{p,G}]$*, the marginal distribution of the* k*-th endogenous regressor error term is Gaussian. A similar partition applies to the* K_w *exogenous regressors. Define* $\boldsymbol{\Pi}_{i,[K_{p,G}][K_{w,NG}]}$ *as the submatrix of* $\boldsymbol{\Pi}_{i,pw} = \boldsymbol{V}_{i,pw}\boldsymbol{V}_{i,w}^{-1}$ *(see Equation (6)) containing entries with row indices in* $[K_{p,G}]$ *and column indices in* $[K_{w,NG}]$*.* $\text{Rank}\left(\boldsymbol{\Pi}_{i,[K_{p,G}][K_{w,NG}]}\right) = K_{p,G}$.

Assumption A.7. *There exists a constant* γ *satisfying* $1/2 < \gamma < \min\{2\tau, 1\}$ $(\tau > 1/4)$ *such that for each* $k \in [K_p + K_w]$*, the following holds for the density function of the* k*-th regressor error term: as* $a \to 0$*,*

$$\sup_{u \in (a, 1-a)} \frac{f_k(F_k^{-1}(u))}{\min\{u, 1-u\}} = o\left(a^{-\frac{1}{3\gamma}}\right).$$

Assumption A.8 (Stationarity for dynamic panel models)**.** *For* $i = 1, 2, ..., n$*, (i)* $\sup_i |\phi_i| < 1$*; (ii) and the initial observation,* y_{i0}*, is drawn from the stationary distribution of* $\{y_{it}\}_{t=1}^T$.

Remark A.1. *Assumption A.2 (ii) assumes serially uncorrelated* ξ_{it}*. For dynamic panel models, it implies that there is no misspecification in the order of lagged variables. When* *estimating dynamic panels, a standard procedure to address endogeneity due to serially correlated errors is to include higher order lags of the regressors. The Akaike and Bayesian information criteria can be used to select the appropriate lag orders.*

Remark A.2. *Assumption A.6 requires that any Gaussian distributed $e_{it,k}$ of an endogneous regressor ($k \in [K_p]$) must be correlated with at least one distinct non-Gaussian distributed $e_{it,k'}$ of an exogenous regressor ($k' \in [K_w]$). This is analogous to the IV identification condition that the number of IVs must be at least equal to the number of endogenous regressors.*

Remark A.3. *Assumption A.7 imposes a restriction on the marginal density of regressor error terms to facilitate the derivation of asymptotic properties of the 2sCOPE-MG estimator. It is used to bound the difference between the residual ranks $\hat{F}_{g(i),k}(\hat{e}_{it,k})$, based on unit-specific demeaned residuals, and the oracle ranks $\hat{F}_{g(i),k}(e_{it,k})$, based on unobserved true $e_{it,k}$.*

A.2.2 Proof

A.2.2.1 Proof of Theorem 1

Proof. To begin with, note that under Assumption A.5, for each $k \in [K_p + K_w]$, if F_k is Gaussian, then $e^*_{it,k} = \Phi^{-1}(F_k(e_{it,k})) = e_{it,k}/\sigma_k$, where σ_k is the standard deviation of $e_{it,k}$ with $0 < \sigma_k < C$. Suppose $(\boldsymbol{\theta}', \boldsymbol{\gamma}_2', \boldsymbol{\delta}')'$ in (9) are not identified under Assumptions A.1, A.2, A.4, and A.5, then there exists a non-zero $(2K_p + K_w) \times 1$ vector $\boldsymbol{\kappa} = \left(\boldsymbol{\kappa}_\epsilon', \boldsymbol{\kappa}_p', \boldsymbol{\kappa}_w'\right)'$ such that

$$\boldsymbol{\kappa}_\epsilon' \boldsymbol{\epsilon}_{it,p} + \boldsymbol{\kappa}_p' \boldsymbol{e}_{it,p} + \boldsymbol{\kappa}_w' \boldsymbol{e}_{it,w} = \boldsymbol{\kappa}_\epsilon' \left(\boldsymbol{e}^*_{it,p} - \boldsymbol{\Pi}_{pw} \boldsymbol{e}^*_{it,w}\right) + \boldsymbol{\kappa}_p' \boldsymbol{e}_{it,p} + \boldsymbol{\kappa}_w' \boldsymbol{e}_{it,w} = 0. \tag{A.1}$$

Under Assumptions A.5 ($\boldsymbol{V}_\rho \succ \mathbf{0}$) and A.4, if $\boldsymbol{\kappa}_\epsilon = \mathbf{0}$, then (A.1) holds only when $\left(\boldsymbol{\kappa}_p', \boldsymbol{\kappa}_w'\right)' = \mathbf{0}$, and vice versa. Thus, for (A.1) to hold with a non-zero $\boldsymbol{\kappa}$, there must be at least two non-zero entries in the two vectors, $\boldsymbol{\kappa}_\epsilon$ and $\left(\boldsymbol{\kappa}_p', \boldsymbol{\kappa}_w'\right)'$, separately.

First, suppose for all $k \in [K_p + K_w]$, F_k is Gaussian, then (A.1) is solved with $\boldsymbol{\sigma}_p \odot \boldsymbol{\kappa}_\epsilon = -\boldsymbol{\kappa}_p$ and $\boldsymbol{\sigma}_w \odot \boldsymbol{\Pi}_{pw}' \boldsymbol{\kappa}_\epsilon = -\boldsymbol{\kappa}_w$, where $\boldsymbol{\sigma}_p = (\sigma_{p,1}, ..., \sigma_{p,K_p})'$, $\boldsymbol{\sigma}_w = (\sigma_{w,1}, ..., \sigma_{w,K_w})'$, and $\odot$ denotes the element-wise product.

Next, we focus on the case where not all $e_{it,k}$ for $k \in [K_p + K_w]$ have marginal Gaussian distributions. The K_p endogenous regressors are divided into two sets: $[K_{p,NG}]$ and $[K_{p,G}]$, where for each $k \in [K_{p,NG}]$, the marginal density of $e_{it,k}$ is not Gaussian, and for each $k' \in [K_{p,G}]$, the marginal distribution of $e_{it,k'}$ is Gaussian; similarly for the K_w exogenous regressors. For any non-Gaussian distributed regressor $k \in [K_{p,NG}]$ (or $k \in [K_{w,NG}]$), since $\sigma_k > 0$, it requires that $\kappa_{p,k} = 0$ (or $\kappa_{w,k} = 0$) for (A.1) to hold, i.e., $\boldsymbol{\kappa}_{p,[K_{p,NG}]} = \mathbf{0}$ and $\boldsymbol{\kappa}_{w,[K_{w,NG}]} = \mathbf{0}$. Furthermore, as $\boldsymbol{\kappa}_{p,[K_{p,NG}]} = \mathbf{0}$ and $\boldsymbol{V}_\rho \succ \mathbf{0}$, it also requires $\boldsymbol{\kappa}_{\epsilon,[K_{p,NG}]} = \mathbf{0}$.

Then for $k \in [K_{p,G}]$, given a fixed non-zero $\boldsymbol{\kappa}_{\epsilon,[K_{p,G}]}$, we can choose $\boldsymbol{\kappa}_{p,[K_{p,G}]} = -\boldsymbol{\sigma}_{p,[K_{p,G}]} \odot \boldsymbol{\kappa}_{\epsilon,[K_{p,G}]}$ such that $\boldsymbol{\kappa}'_\epsilon \boldsymbol{\epsilon}_{it,p} + \boldsymbol{\kappa}'_p \boldsymbol{e}_{it,p} = 0$. Let $\boldsymbol{\Pi}_{[K_p],[K_{w,G}]}$ denotes the submatrix of $\boldsymbol{\Pi}_{pw} = \boldsymbol{V}_{pw}\boldsymbol{V}_w^{-1}$ (see Equation (6)) containing entries with row indices in $[K_p]$ and column indices in $[K_{w,G}]$. For $k \in [K_{w,G}]$, we can choose $\boldsymbol{\sigma}_{w,[K_{w,G}]} \odot \boldsymbol{\Pi}'_{[K_p][K_{w,G}]}\boldsymbol{\kappa}_\epsilon = \boldsymbol{\kappa}_{w,[K_{w,G}]}$ such that $-\boldsymbol{\kappa}'_\epsilon \boldsymbol{\Pi}_{[K_p],[K_{w,G}]} \boldsymbol{e}^*_{it,w[K_{w,G}]} + \boldsymbol{\kappa}'_{w,[K_{w,G}]} \boldsymbol{e}_{it,w[K_{w,G}]} = 0$. Now the only remaining term is given by $-\boldsymbol{\kappa}'_{\epsilon,[K_{p,G}]} \boldsymbol{\Pi}_{[K_{p,G}],[K_{w,NG}]} \boldsymbol{e}^*_{it,w[K_{w,NG}]}$, if $[K_{w,NG}]$ is not an empty set. Since $\boldsymbol{\kappa}_{\epsilon,[K_{p,G}]}$ is a non-zero vector, $\boldsymbol{\kappa}'_{\epsilon,[K_{p,G}]} \boldsymbol{\Pi}_{[K_{p,G}],[K_{w,NG}]} = \mathbf{0}$ holds if and only if $\text{Rank}(\boldsymbol{\Pi}_{[K_{p,G}],[K_{w,NG}]}) < K_{p,G}$, where $K_{p,G}$ is the number of $e_{it,k}$ ($k \in [K_p]$) with a marginal Gaussian distribution. Therefore, Assumption A.6 provides the identification condition for $(\boldsymbol{\theta}', \boldsymbol{\gamma}'_2, \boldsymbol{\delta}')'$. □

A.2.2.2 Lemmas for the proof of Theorem 2

Lemma A.1. *Suppose Equation (3) in the main paper and Assumptions A.1, A.2, A.4, A.5, and A.7 hold. Let n_g denote the number of units in Group g, for $g = 1, 2, ..., G$. For the estimator of group-specific coefficients in the first stage regression, $\hat{\boldsymbol{\Pi}}_{g,pw}$ in (11) of the main paper, as $n, T \to \infty$ and $T/n \to \infty$, $\sqrt{n_g T}(\hat{\boldsymbol{\Pi}}_{g,pw} - \boldsymbol{\Pi}_{g,pw}) \to_d N(\mathbf{0}, \boldsymbol{\Sigma}_{\Pi,g})$ where $\boldsymbol{\Sigma}_{\Pi,g} \succ \mathbf{0}$.*

Proof. The asymptotic property of $\hat{\boldsymbol{\Pi}}_{g,pw}$ can be established using Theorem 3.4 of Zhao et al. (2020). To begin, we derive the order of the estimation errors for $\hat{\boldsymbol{e}}_{it,p}$ and $\hat{\boldsymbol{e}}_{it,w}$. Given Equation (3), $\hat{\boldsymbol{e}}_{it,p} = \boldsymbol{p}_{it} - \hat{\boldsymbol{\alpha}}_{ip}$ with $\hat{\boldsymbol{\alpha}}_{ip} = \frac{1}{T}\sum_{t=1}^T \boldsymbol{p}_{it}$. As $T \to \infty$, $\hat{\boldsymbol{\alpha}}_{ip} - \boldsymbol{\alpha}_{ip} = O_p\left(T^{-1/2}\right)$. Similarly, $\hat{\boldsymbol{\alpha}}_{iw} - \boldsymbol{\alpha}_{iw} = O_p\left(T^{-1/2}\right)$. Next, we investigate the conditions under which Assumption 3.4 in Zhao et al. (2020) holds under two practical grouping schemes.

When T is sufficiently large such that each unit can be viewed as a group, i.e., $n_g = 1$. Then it follows that $O(n_g T) = O(T)$, and Assumption 3.4 in Zhao et al. (2020) holds trivially. For moderately short T and a finite number of groups, $O(n_g) = O(n)$. In this case, Assumption 3.4 in Zhao et al. (2020) is satisfied if $T/n \to \infty$.

By Theorem 3.4 of Zhao et al. (2020), the asymptotic distribution of $\sqrt{n_g T}(\hat{\boldsymbol{\Pi}}_{g,pw} - \boldsymbol{\Pi}_{g,pw})$ coincides with the asymptotic distribution of $\sqrt{n_g T}(\tilde{\boldsymbol{\Pi}}_{g,pw} - \boldsymbol{\Pi}_{g,pw})$, where each element in $\tilde{\boldsymbol{\Pi}}_{g,pw}$ is the normal scores rank correlation coefficient estimator based on the empirical CDFs of the unobserved $\boldsymbol{e}_{it,p}$ and $\boldsymbol{e}_{it,w}$. As established in Theorem 3.1 of Klaassen and Wellner (1997), it is asymptotically normally distributed with mean zero and a positive variance. $\square$

Lemma A.2. *Suppose Equations (2) and (3) in the main paper and Assumptions A.1–A.7 hold. For the estimator of the homogeneous coefficients in the augmented panel data model, $\hat{\boldsymbol{\gamma}}_2$ in (12), as $n, T \to \infty$, $\sqrt{nT}(\hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_2) \to_d N(\mathbf{0}, \boldsymbol{\Sigma}_{\gamma,2})$ where $\boldsymbol{\Sigma}_{\gamma,2} \succ \mathbf{0}$.*

Proof. Given $\hat{\boldsymbol{\gamma}}_2$ in (12), we have

$$\sqrt{nT}(\hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_2) = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{W}'_{i,2}\boldsymbol{M}_{ix,2}\boldsymbol{W}_{i,2}\right)^{-1} \left[\frac{1}{\sqrt{nT}}\sum_{i=1}^{n} \boldsymbol{W}'_{i,2}\boldsymbol{M}_{ix,2}\left(\boldsymbol{u}_i + \boldsymbol{\epsilon}_{ip}\boldsymbol{\delta}_{g(i)}\right)\right].$$

Given (10) in the main paper, $E(\boldsymbol{W}'_{i,2}\boldsymbol{M}_{ix,2}\boldsymbol{u}_i) = E\left[\boldsymbol{W}'_{i,2}\boldsymbol{M}_{ix,2}E(\boldsymbol{u}_i|\boldsymbol{X}_i, \boldsymbol{W}_{i,2}, \boldsymbol{\epsilon}_{ip})\right] = \mathbf{0}$. Under Assumption A.5, $\boldsymbol{\epsilon}_{it,p}$ is distributed independently of $\boldsymbol{e}_{is,w}$ with a zero mean such that conditional on $\boldsymbol{M}_{ix,2}$ and $\boldsymbol{\delta}_{g(i)}$, $E\left(\boldsymbol{w}_{it,2}\boldsymbol{\epsilon}'_{is,p}\right) = \mathbf{0}$ for all t, s, and thus, $E_i(\boldsymbol{W}'_{i,2}\boldsymbol{M}_{ix,2}\boldsymbol{\epsilon}_{ip})\boldsymbol{\delta}_{g(i)} = \mathbf{0}$. Combined these two terms, as $n, T \to \infty$, we have $E(\hat{\boldsymbol{\gamma}}_2) = \boldsymbol{\gamma}_2 + o(1)$. Hence, under Assumptions A.1–A.7, by the Central Limit Theorem, $\sqrt{nT}(\hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_2) \to_d N(\mathbf{0}, \boldsymbol{\Sigma}_{\gamma,2})$, where

$$\boldsymbol{\Sigma}_{\gamma,2} = \boldsymbol{\Psi}_{w,2}^{-1}\left[\lim_{n,T\to\infty}\frac{1}{nT}\sum_{i=1}^{n} \boldsymbol{W}'_{i,2}\boldsymbol{M}_{ix,2}\left(\boldsymbol{u}_i + \boldsymbol{\epsilon}_{ip}\boldsymbol{\delta}_{g(i)}\right)\left(\boldsymbol{u}_i + \boldsymbol{\epsilon}_{ip}\boldsymbol{\delta}_{g(i)}\right)'\boldsymbol{M}_{ix,2}\boldsymbol{W}_{i,2}\right]\boldsymbol{\Psi}_{w,2}^{-1},$$

with $\boldsymbol{\Psi}_{w,2} = \lim_{n,T\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\boldsymbol{W}'_{i,2}\boldsymbol{M}_{ix,2}\boldsymbol{W}_{i,2} \succ \mathbf{0}$. $\square$

Lemma A.3. *Suppose Equations (2) and (3) in the main paper and Assumptions A.1–A.7 hold. For the estimator of group-specific coefficients of the copula generated regressors in the augmented panel data model, $\hat{\boldsymbol{\delta}}_g$ in (13), as $n, T \to \infty$ and $T/n \to \infty$, $\sqrt{n_g T}(\hat{\boldsymbol{\delta}}_g - \boldsymbol{\delta}_g) \to_d N(\mathbf{0}, \boldsymbol{\Sigma}_{\delta,g})$ where $\boldsymbol{\Sigma}_{\delta,g} \succ \mathbf{0}$, and n_g denotes the number of units in Group g, for $g = 1, 2, ..., G$.*

Proof. Without loss of the generality, we consider the case with a scalar endogenous regressor, p_{it}, and a scalar exogenous regressor, w_{it}, to simplify the mathematical exposition. $\hat{\boldsymbol{\delta}}_g$ in (13) can be rewritten as $\hat{\delta}_g = \Psi^{-1}_{\epsilon_p,N_g}\left(\frac{1}{N_g}\sum_{i\in[n_g]}^{n_g}\hat{\boldsymbol{\epsilon}}'_{ip}\boldsymbol{M}_{ix}\boldsymbol{y}_i\right)$ with $\Psi_{\epsilon_p,N_g} = \frac{1}{N_g}\sum_{i\in[n_g]}^{n_g}\hat{\boldsymbol{\epsilon}}'_{ip}\boldsymbol{M}_{ix}\hat{\boldsymbol{\epsilon}}_{ip}$ and $N_g = n_gT$. Then we have

$$\Psi_{\epsilon_p,N_g}\sqrt{N_g}\left(\hat{\delta}_g - \delta_g\right) = \frac{1}{\sqrt{N_g}}\sum_{i\in[n_g]}^{n_g}\hat{\boldsymbol{\epsilon}}'_{ip}\boldsymbol{M}_{ix}\left[\boldsymbol{u}_i + (\boldsymbol{\epsilon}_{ip} - \hat{\boldsymbol{\epsilon}}_{ip})\,\delta_g\right] = A_{N_g} + B_{N_g},$$

where $A_{N_g} = \frac{1}{\sqrt{N_g}}\sum_{i\in[n_g]}^{n_g}\hat{\boldsymbol{\epsilon}}'_{ip}\boldsymbol{M}_{ix}\boldsymbol{u}_i$, and $B_{N_g} = \frac{1}{\sqrt{N_g}}\sum_{i\in[n_g]}^{n_g}\hat{\boldsymbol{\epsilon}}'_{ip}\boldsymbol{M}_{ix}\left(\boldsymbol{\epsilon}_{ip} - \hat{\boldsymbol{\epsilon}}_{ip}\right)\delta_g$.

We first analyze $A_{N_g} = A_{N_g,1} + A_{N_g,2} + A_{N_g,3}$, where $A_{N_g,1} = \frac{1}{\sqrt{N_g}}\sum_{i\in[n_g]}^{n_g}\boldsymbol{\epsilon}'_{ip}\boldsymbol{M}_{ix}\boldsymbol{u}_i$, $A_{N_g,2} = \frac{1}{\sqrt{N_g}}\sum_{i\in[n_g]}^{n_g}(\tilde{\boldsymbol{\epsilon}}_{ip} - \boldsymbol{\epsilon}_{ip})'\boldsymbol{M}_{ix}\boldsymbol{u}_i$, and $A_{N_g,3} = \frac{1}{\sqrt{N_g}}\sum_{i\in[n_g]}^{n_g}(\hat{\boldsymbol{\epsilon}}_{ip} - \tilde{\boldsymbol{\epsilon}}_{ip})'\boldsymbol{M}_{ix}\boldsymbol{u}_i$, with

$$\tilde{\boldsymbol{\epsilon}}_{ip} = \tilde{\boldsymbol{e}}^*_{ip} - \Pi_{g(i),pw}\tilde{\boldsymbol{e}}^*_{iw}, \tag{A.2}$$

$\tilde{\boldsymbol{e}}^*_{ip} = (\tilde{e}^*_{i1,p}, ..., \tilde{e}^*_{iT,p})'$, $\tilde{\boldsymbol{e}}^*_{iw} = (\tilde{e}^*_{i1,w}, ..., \tilde{e}^*_{iT,w})'$, $\tilde{e}^*_{it,p} = \hat{F}_{g(i),p}(e_{it,p})$, and $\tilde{e}^*_{it,w} = \hat{F}_{g(i),w}(e_{it,w})$. Given (10), $A_{N_g,1} \to_d N(0, \Sigma_{A,1})$ with $\Sigma_{A,1} = \lim_{N_g\to\infty}\frac{1}{N_g}\sum_{i\in[n_g]}^{n_g}\boldsymbol{\epsilon}'_{ip}\boldsymbol{M}_{ix}\boldsymbol{u}_i\boldsymbol{u}'_i\boldsymbol{M}_{ix}\boldsymbol{\epsilon}_{ip}$.

To analyze $A_{N_g,2}$, let $A_{N_g,2(1)} = \frac{1}{\sqrt{N_g}}\sum_{i\in[N_g]}^{N_g}(\tilde{\boldsymbol{\epsilon}}_{ip} - \boldsymbol{\epsilon}_{ip})'\boldsymbol{\tau}_T$, where $\boldsymbol{\tau}_T$ is a $T\times 1$ vector of ones. There exist two orderings of the indices $\{1, 2, ..., N_g\}$ such that for Group g,

$$\begin{aligned}
&A_{N_g,2(1)}\\
=&\frac{1}{\sqrt{N_g}}\sum_{s\in[N_g]}^{N_g}[\Phi^{-1}(\hat{F}^{-1}_{g,p}(s/N_g)) - \Phi^{-1}(s/\tilde{N}_g)] - \frac{\Pi_{pw}}{\sqrt{N_g}}\sum_{j\in[N_g]}^{N_g}[\Phi^{-1}(\hat{F}^{-1}_{g,w}(j/N_g)) - \Phi^{-1}(j/\tilde{N}_g)]\\
=&\frac{1}{\sqrt{N_g}}\sum_{s\in[N_g]}^{N_g}\frac{\hat{F}^{-1}_{g,p}(s/\tilde{N}_g) - s/\tilde{N}_g}{\phi(\Phi^{-1}(s/\tilde{N}_g))} - \frac{\Pi_{pw}}{\sqrt{N_g}}\sum_{j\in[N_g]}^{N_g}\frac{\hat{F}^{-1}_{g,w}(j/\tilde{N}_g) - j/\tilde{N}_g}{\phi(\Phi^{-1}(j/\tilde{N}_g))} + o_p(1),
\end{aligned}$$

with $\tilde{N}_g = N_g + 1$.[A1] By the weak convergence of the uniform quantile process,

$$\frac{1}{\sqrt{N_g}}\sum_{s\in[N_g]}^{N_g}\frac{\hat{F}^{-1}_{g,p}(s/\tilde{N}_g) - s/\tilde{N}_g}{\phi(\Phi^{-1}(s/\tilde{N}_g))} \to_d H_1 = \int_0^1\frac{B(u)}{\phi(\Phi^{-1}(u))}du,$$

and

$$\frac{\Pi_{pw}}{\sqrt{N_g}}\sum_{j\in[N_g]}^{N_g}\frac{\hat{F}^{-1}_{g,w}(j/\tilde{N}_g) - j/\tilde{N}_g}{\phi(\Phi^{-1}(j/\tilde{N}_g))} \to_d H_2 = \Pi_{pw}\int_0^1\frac{B(v)}{\phi(\Phi^{-1}(v))}dv,$$

where $B(\cdot)$ is a standard Brownian bridge. Thus, $A_{N_g,2(1)} \to_d H = H_1 - H_2$, where $E(H) =$

[A1] Here we use s and j to replace the double indices it, since we assume $e_{it,p}$ and $e_{it,w}$ are i.i.d draws.

$E(H_1 - H_2) = 0$ as $E(H_1) = 0$ and $E(H_2) = 0$. For the variance,

$$V(H) = V(H_1) + V(H_2) - 2E(H_1 H_2),$$

where $V(H_1) = 1$, $V(H_2) = \Pi_{pw}^2$,

$$E(H_1 H_2) = -\Pi_{pw} \int_0^1 \int_0^1 B(u)B(v) f\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) du dv,$$

and $f(x, y)$ is the density function of a standard bivariate normal distribution with a correlation coefficient Π_{pw}. By some algebra, it can be shown that $E(H_1 H_2) = E\left[\min(\Phi(x), \Phi(y))\right] - E\left[\Phi(x)\Phi(y)\right]$, where $\Phi(x)$ and $\Phi(y)$ are the standard normal CDFs of x and y. Since $E\left[\min(\Phi(x), \Phi(y))\right] = \frac{1}{2} - \frac{1}{2\pi}\sin^{-1}(\Pi_{pw})$ and $E\left[\Phi(x)\Phi(y)\right] = \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(\Pi_{pw})$, $E(H_1 H_2) = \frac{1}{4} - \frac{1}{\pi}\sin^{-1}(\Pi_{pw})$ where $|E(H_1 H_2)| \leq 1/4$ for all $\Pi_{pw} \in (0, 1)$. Thus, $V(H) = 1 + \Pi_{pw}^2 - 2\left(\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}(\Pi_{pw})\right) \geq 1 + \Pi_{pw}^2 - 1/2 > 0$, and $A_{N_g,2(1)}$ converges to a non-degenerate random variable with mean zero. Furthermore, since u_{it} is mean zero and distributed independently of $e_{it,p}$ and $e_{it,w}$, conditional on the weight matrix $\boldsymbol{M}_{ix} = \{m_{j,x}\}_{j=1,\ldots,T^2}$,

$$\begin{aligned} A_{N_g,2} &= \frac{1}{\sqrt{N_g}} \sum_{s\in[N_g]}^{N_g} u_s m_{s,x} \left(\frac{\hat{F}_{g,p}^{-1}(s/\tilde{N}_g) - s/\tilde{N}_g}{\phi(\Phi^{-1}(s/\tilde{N}_g))}\right) - \frac{\Pi_{pw}}{\sqrt{N_g}} \sum_{j\in[N_g]}^{N_g} u_j m_{j,x} \left(\frac{\hat{F}_{g,p}^{-1}(j/\tilde{N}_g) - j/\tilde{N}_g}{\phi(\Phi^{-1}(j/\tilde{N}_g))}\right) \\ &= O_p(N_g^{-1/2}) = o_p(1). \end{aligned}$$

For the term $A_{N_g,3}$, let $A_{N_g,3(1)} = \frac{1}{\sqrt{N_g}} \sum_{i\in[N_g]}^{N_g} (\hat{\boldsymbol{\epsilon}}_{ip} - \tilde{\boldsymbol{\epsilon}}_{ip})' \boldsymbol{\tau}_T$. As $\hat{\Pi}_{g,pw} - \Pi_{g,pw} = O_p(N_g^{-1/2})$ by Lemma A.1, we have $A_{N_g,3(1)} = \frac{1}{\sqrt{N_g}} \sum_{i\in[N_g]}^{N_g} \left[(\hat{\boldsymbol{e}}_{ip}^* - \tilde{\boldsymbol{e}}_{ip}^*) - \Pi_{g,pw}(\hat{\boldsymbol{e}}_{iw}^* - \tilde{\boldsymbol{e}}_{iw}^*)\right]' \boldsymbol{\tau}_T + o_p(1)$. Using a first-order Taylor expansion and the arguments from the proof of Theorem 3.4 in Zhao et al. (2020), it follows that $A_{N_g,3(1)} = O_p\left[\log^{1/2}(N_g) N_g^{1/2} T^{-1}\right] = O_p\left[\log^{1/2}(nT)(n/T)^{1/2}\right]$. As $T/n \to \infty$, this implies $A_{N_g,3(1)} = o_p(1)$. Analogously to the derivation of the $A_{N_g,2}$ term, conditional on the weight matrix $\boldsymbol{M}_{ix}$, $A_{N_g,3} = o_p(1)$. Thus, combining the three terms, we have $A_{N_g} = A_{N_g,1} + o_p(1)$.

We now analyze $B_{N_g} = B_{N_g,1} + B_{N_g,2} + B_{N_g,3} + B_{N_g,4}$, where

$$B_{N_g,1} = \frac{1}{\sqrt{N_g}} \sum_{i\in[n_g]}^{n_g} \boldsymbol{\epsilon}_{ip}' \boldsymbol{M}_{ix} \left(\boldsymbol{\epsilon}_{ip} - \tilde{\boldsymbol{\epsilon}}_{ip}\right) \delta_g,$$

$B_{N_g,2} = \frac{1}{\sqrt{N_g}} \sum_{i\in[n_g]}^{n_g} (\hat{\boldsymbol{\epsilon}}_{ip}' - \tilde{\boldsymbol{\epsilon}}_{ip}') \boldsymbol{M}_{ix} \left(\boldsymbol{\epsilon}_{ip} - \tilde{\boldsymbol{\epsilon}}_{ip}\right) \delta_g$, $B_{N_g,3} = \frac{1}{\sqrt{N_g}} \sum_{i\in[n_g]}^{n_g} (\hat{\boldsymbol{\epsilon}}_{ip}' - \tilde{\boldsymbol{\epsilon}}_{ip}') \boldsymbol{M}_{ix} \left(\hat{\boldsymbol{\epsilon}}_{ip} - \tilde{\boldsymbol{\epsilon}}_{ip}\right) \delta_g =$

$o_p(1)$, and $B_{N_g,4} = \frac{1}{\sqrt{N_g}} \sum_{i\in[n_g]}^{n_g} \hat{\boldsymbol{\epsilon}}'_{ip} \boldsymbol{M}_{ix} \left(\tilde{\boldsymbol{\epsilon}}_{ip} - \hat{\boldsymbol{\epsilon}}_{ip} \right) \delta_g$. Using the Cauchy–Schwarz's inequality repeatedly and $\frac{1}{\sqrt{N_g}} \sum_{i\in[n_g]}^{n_g} (\tilde{\boldsymbol{\epsilon}}'_{ip} - \boldsymbol{\epsilon}'_{ip}) \boldsymbol{M}_{ix} \left(\boldsymbol{\epsilon}_{ip} - \tilde{\boldsymbol{\epsilon}}_{ip} \right) \delta_g = o_p(1)$, $B_{N_g,2} = o_p(1)$. Given $A_{N_g,2(1)} = O_p(N_g^{1/2})$, $B_{N_g,3} = o_p(1)$. Similarly, $B_{N_g,4} = o_p(1)$. Thus, $B_{N_g} = B_{N_g,1} + o_p(1)$.

For the $B_{N_g,1}$ term, there exist two orderings of the indices $\{1, 2, ..., N_g\}$ such that

$$
\begin{aligned}
B_{N_g,1} =& \frac{1}{\sqrt{N_g}} \sum_{s\in[N_g]}^{N_g} \Phi^{-1}(\hat{F}_{g,p}^{-1}(s/\tilde{N}_g)) \frac{\hat{F}_{g,p}^{-1}(s/\tilde{N}_g) - s/\tilde{N}_g}{\phi(\Phi^{-1}(s/\tilde{N}_g))} \\
&- \frac{\Pi_{pw}}{\sqrt{N_g}} \sum_{j\in[N_g]}^{N_g} \Phi^{-1}(\hat{F}_{g,w}^{-1}(j/\tilde{N}_g)) \frac{\hat{F}_{g,w}^{-1}(j/\tilde{N}_g) - j/\tilde{N}_g}{\phi(\Phi^{-1}(j/\tilde{N}_g))} + o_p(1),
\end{aligned}
$$

with $\tilde{N}_g = N_g + 1$. By the weak convergence of the uniform quantile process,

$$
\frac{1}{\sqrt{N_g}} \sum_{s\in[N_g]}^{N_g} \Phi^{-1}(\hat{F}_{g,p}^{-1}(s/\tilde{N}_g)) \frac{\hat{F}_{g,p}^{-1}(s/\tilde{N}_g) - s/\tilde{N}_g}{\phi(\Phi^{-1}(s/\tilde{N}_g))} \to_d H_3 = \int_0^1 h(u)B(u)du,
$$

and

$$
\frac{\Pi_{pw}}{\sqrt{N_g}} \sum_{j\in[N_g]}^{N_g} \Phi^{-1}(\hat{F}_{g,w}^{-1}(j/\tilde{N}_g)) \frac{\hat{F}_{g,w}^{-1}(j/\tilde{N}_g) - j/\tilde{N}_g}{\phi(\Phi^{-1}(j/\tilde{N}_g))} \to_d H_4 = \Pi_{pw} \int_0^1 h(v)B(v)dv,
$$

where $h(x) = \frac{\Phi^{-1}(x)}{\phi(\Phi^{-1}(x))}$, and $B(\cdot)$ is a standard Brownian bridge. Thus, as $N_g \to \infty$, $B_{N_g,1} \to_d (H_3 - H_4)$, where $E(H_3 - H_4) = 0$ and $Var(H_3 - H_4) = \Sigma_{B,1} > 0$.

The above results also imply that $\Psi_{\epsilon_p, N_g} \to_p \Psi_{\epsilon_p} = \lim_{N_g \to \infty} \frac{1}{N_g} \sum_{i\in[n_g]}^{n_g} \boldsymbol{\epsilon}'_{ip} \boldsymbol{M}_{ix} \boldsymbol{\epsilon}_{ip}$. Thus, as $n, T \to \infty$ and $\frac{T}{n} \to \infty$, $\sqrt{n_g T}(\hat{\delta}_g - \delta_g) \to_d N(0, \Sigma_{\delta,g})$, with $\Sigma_{\delta,g} = \Psi_{\epsilon_p}^{-2}(\Sigma_{A,1} + \Sigma_{B,1}) > 0$. $\square$

A.2.2.3 Proof of Theorem 2

Proof. From Lemmas A.1–A.3, it follow that $\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i = (\boldsymbol{X}'_i \boldsymbol{X}_i)^{-1} \boldsymbol{X}'_i \boldsymbol{u}_i + O_p\left[(n_g T)^{-1/2}\right] = (\boldsymbol{X}'_i \boldsymbol{X}_i)^{-1} \boldsymbol{X}'_i \boldsymbol{u}_i + o_p(1)$. Also, $E(\hat{\boldsymbol{\theta}}_i) = \boldsymbol{\theta}_i + O\left(T^{-1}\right)$, where the incidental parameter bias arises from the estimation of $\boldsymbol{\alpha}_{ip}$ and $\boldsymbol{\alpha}_{iw}$. Then under Assumptions A.1–A.7, the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ in (14) and a consistent estimator of its variance are derived using Theorems 3.1 and 3.2 from Chudik and Pesaran (2019), respectively. $\square$

A.2.3 Comparisons of Pooled OLS, FE, and MG estimators in heterogeneous panel data models

Consider a heterogeneous panel data model given by

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta}_i + \xi_{it},$$

for $i = 1, 2, ..., n$ and $t = 1, 2, ..., T$. Assuming the regressors are exogenous, the pooled ordinary least squares (OLS) and fixed effects (FE) estimators can be rewritten as weighted averages over individual-specific OLS estimators, with non-uniform weights shown below.

For the slope coefficients, the pooled OLS estimator is given by

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{POLS} &= \left[(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})'\right]^{-1}\left[(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})(y_{it}-\bar{y})\right] \\ &= \sum_{i=1}^{n}\underbrace{\frac{1}{n}\left[\frac{\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)'}{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})'}\right]\left[\frac{\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})(y_{it}-\bar{y})}{\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(y_{it}-\bar{y}_i)}\right]}_{\text{the weight on the individual-specific OLS estimate}}\hat{\boldsymbol{\beta}}_i,\end{aligned}$$

where $\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\bar{\boldsymbol{x}}_i$ with $\bar{\boldsymbol{x}}_i = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}$, $\bar{y} = (nT)^{-1}\sum_{i=1}^{n}\bar{y}_i$ with $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T}y_{it}$, and

$$\hat{\boldsymbol{\beta}}_i = \left[\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(y_{it}-\bar{y}_i)\right].$$

The FE estimator for panel data models is given by

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{FE} &= \left[(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)'\right]^{-1}\left[(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(y_{it}-\bar{y}_i)\right] \\ &= \sum_{i=1}^{n}\underbrace{\frac{1}{n}\left[\frac{\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)'}{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it}-\bar{\boldsymbol{x}})'}\right]}_{\text{the weight on the individual-specific OLS estimate}}\hat{\boldsymbol{\beta}}_i.\end{aligned}$$

Also, the mean group (MG) estimator is given by

$$\hat{\boldsymbol{\beta}}_{MG} = \sum_{i=1}^{n}\frac{1}{n}\hat{\boldsymbol{\beta}}_i,$$

with a uniform weight of $1/n$ on all individual-specific estimates.

When the weights in pooled OLS and FE estimators are correlated with individual-specific slope coefficients, they may not serve as unbiased or consistent estimators for the population mean of heterogeneous coefficients, given by $\boldsymbol{\beta}_0 = plim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\beta}_i$. In contrast, the MG estimator is consistent and can be viewed as a sample analog of $\boldsymbol{\beta}_0$.

A.3 Supplementary MC evidence

Table A.1 shows simulation results of FE, 2sCOPE-FE, and 2sCOPE-MG estimators in Cases 1.4–1.6: static panel data models under both regressor and slope endogeneity with different error and regressor processes. For detailed data generating processes and discussions, see subsection 4.1.4 in the main paper. Tables A.2 and A.3 present simulation results in Cases 2.1 and 2.2 of dynamic panel data models under regressor endogeneity and slope endogeneity, respectively. See subsection 4.2 in the main paper for details.

Table A.1: MC results of FE, 2sCOPE-FE, and 2sCOPE-MG estimators in Cases 1.4–1.6: static panels under regressor and slope endogeneity with different error and regressor processes

		Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}
		$T = 10, n = 100$						$T = 50, n = 100$						$T = 100, n = 100$					
Case 1.4: correlated heterogeneous slopes, endogenous regressors, and heteroskedastic errors																			
FE	$\hat{\alpha}$	0.204	0.214	0.261	0.295	0.09	0.78	0.291	0.144	0.200	0.324	0.27	1.46	0.312	0.139	0.191	0.341	0.32	1.63
	$\hat{\beta}$	0.547	0.088	0.095	0.554	1.00	5.76	0.479	0.072	0.082	0.484	1.00	5.83	0.463	0.069	0.081	0.468	1.00	5.72
	$\hat{\gamma}$	-0.217	0.129	0.127	0.252	0.43	1.70	-0.254	0.085	0.090	0.267	0.81	2.81	-0.259	0.076	0.084	0.270	0.90	3.07
2sCOPE-FE	$\hat{\alpha}$	0.022	0.222	0.269	0.223	0.02	0.08	0.146	0.145	0.200	0.206	0.05	0.73	0.175	0.139	0.191	0.224	0.07	0.92
	$\hat{\beta}$	0.106	0.175	0.170	0.205	0.11	0.62	0.063	0.083	0.093	0.104	0.08	0.68	0.048	0.076	0.087	0.090	0.06	0.56
	$\hat{\gamma}$	0.090	0.164	0.165	0.187	0.08	0.55	0.022	0.089	0.095	0.092	0.05	0.24	0.012	0.079	0.087	0.080	0.03	0.14
2sCOPE-MG	$\hat{\alpha}$	-0.003	0.146	0.150	0.146	0.05	0.02	-0.004	0.078	0.089	0.078	0.03	0.04	-0.002	0.075	0.084	0.075	0.03	0.02
	$\hat{\beta}$	-0.011	0.170	0.165	0.170	0.05	0.07	0.011	0.081	0.091	0.082	0.02	0.12	0.011	0.074	0.086	0.075	0.02	0.13
	$\hat{\gamma}$	-0.010	0.142	0.142	0.142	0.04	0.07	-0.009	0.078	0.086	0.079	0.04	0.10	-0.006	0.073	0.082	0.074	0.03	0.07
Case 1.5: correlated heterogeneous slopes, endogenous regressors, and serially correlated errors																			
FE	$\hat{\alpha}$	0.202	0.217	0.261	0.296	0.09	0.77	0.293	0.145	0.200	0.327	0.29	1.47	0.316	0.139	0.191	0.345	0.33	1.65
	$\hat{\beta}$	0.536	0.088	0.094	0.543	1.00	5.69	0.490	0.073	0.081	0.495	1.00	6.02	0.477	0.070	0.080	0.482	1.00	5.96
	$\hat{\gamma}$	-0.210	0.128	0.127	0.246	0.41	1.65	-0.260	0.086	0.090	0.274	0.83	2.89	-0.269	0.076	0.084	0.279	0.91	3.18
2sCOPE-FE	$\hat{\alpha}$	0.025	0.226	0.271	0.227	0.02	0.09	0.145	0.147	0.200	0.206	0.05	0.72	0.175	0.140	0.191	0.224	0.08	0.91
	$\hat{\beta}$	0.107	0.182	0.177	0.212	0.11	0.61	0.063	0.086	0.094	0.107	0.09	0.67	0.048	0.077	0.086	0.091	0.07	0.56
	$\hat{\gamma}$	0.088	0.166	0.170	0.188	0.07	0.52	0.023	0.091	0.095	0.094	0.05	0.25	0.012	0.079	0.087	0.080	0.03	0.14
2sCOPE-MG	$\hat{\alpha}$	-0.002	0.158	0.160	0.158	0.05	0.01	-0.005	0.082	0.091	0.082	0.03	0.05	-0.003	0.077	0.084	0.077	0.03	0.04
	$\hat{\beta}$	-0.009	0.176	0.172	0.176	0.05	0.05	0.011	0.085	0.092	0.085	0.04	0.12	0.011	0.076	0.086	0.077	0.03	0.13
	$\hat{\gamma}$	-0.012	0.147	0.147	0.147	0.05	0.08	-0.008	0.080	0.087	0.081	0.04	0.09	-0.006	0.073	0.083	0.074	0.03	0.07
Case 1.6: correlated heterogeneous slopes, endogenous regressors, and uniformly distributed fixed effects in regressors processes																			
FE	$\hat{\alpha}$	0.194	0.217	0.268	0.291	0.09	0.73	0.287	0.147	0.207	0.322	0.23	1.39	0.320	0.139	0.198	0.349	0.32	1.62
	$\hat{\beta}$	0.562	0.088	0.093	0.569	1.00	6.01	0.496	0.072	0.081	0.501	1.00	6.12	0.480	0.069	0.080	0.485	1.00	6.00
	$\hat{\gamma}$	-0.227	0.125	0.127	0.259	0.48	1.79	-0.265	0.086	0.090	0.279	0.83	2.94	-0.270	0.076	0.084	0.281	0.91	3.21
2sCOPE-FE	$\hat{\alpha}$	0.005	0.228	0.276	0.228	0.02	0.02	0.138	0.149	0.207	0.203	0.04	0.67	0.180	0.139	0.198	0.227	0.06	0.91
	$\hat{\beta}$	0.103	0.175	0.170	0.203	0.12	0.61	0.063	0.084	0.091	0.105	0.09	0.69	0.048	0.075	0.085	0.089	0.06	0.57
	$\hat{\gamma}$	0.093	0.160	0.165	0.185	0.08	0.56	0.022	0.091	0.094	0.094	0.05	0.24	0.012	0.079	0.086	0.080	0.03	0.14
2sCOPE-MG	$\hat{\alpha}$	-0.003	0.149	0.151	0.149	0.05	0.02	-0.004	0.081	0.088	0.081	0.03	0.04	-0.003	0.075	0.083	0.075	0.03	0.03
	$\hat{\beta}$	-0.012	0.169	0.164	0.170	0.06	0.07	0.012	0.082	0.090	0.082	0.03	0.13	0.011	0.074	0.084	0.075	0.03	0.14
	$\hat{\gamma}$	-0.010	0.141	0.141	0.141	0.05	0.07	-0.009	0.079	0.085	0.079	0.04	0.10	-0.006	0.073	0.082	0.073	0.03	0.07

Notes: The table shows the mean bias, standard deviation (SD) of the respective coefficients, mean estimated standard error ($\hat{se}$), root mean squared error (RMSE), mean size of the test, and the ratio of the absolute mean bias divided by the mean estimated standard error (t_{bias}). The number of replications is 1000.

Table A.2: MC results of FE, 2sCOPE-FE, and 2sCOPE-MGJK estimators in Case 2.1: dynamic panels under regressor endogeneity

		Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}
		$T = 10, n = 100$						$T = 50, n = 100$						$T = 100, n = 100$					
FE	$\hat{\alpha}$	0.112	0.207	0.066	0.235	0.59	1.69	0.128	0.168	0.031	0.211	0.79	4.07	0.133	0.159	0.023	0.207	0.84	5.76
	$\hat{\phi}$	0.000	0.030	0.031	0.030	0.05	0.01	0.074	0.028	0.027	0.079	0.76	2.71	0.087	0.027	0.027	0.091	0.88	3.23
	$\hat{\beta}$	0.420	0.026	0.026	0.421	1.00	16.46	0.439	0.011	0.011	0.439	1.00	38.37	0.441	0.008	0.008	0.441	1.00	54.19
	$\hat{\gamma}$	-0.264	0.039	0.037	0.267	1.00	7.08	-0.280	0.017	0.017	0.280	1.00	16.66	-0.280	0.012	0.012	0.281	1.00	23.56
2sCOPE-FE	$\hat{\alpha}$	-0.076	0.223	0.102	0.236	0.40	0.75	-0.023	0.170	0.035	0.171	0.69	0.65	-0.009	0.159	0.025	0.160	0.78	0.35
	$\hat{\phi}$	0.001	0.030	0.031	0.030	0.05	0.03	0.074	0.028	0.027	0.079	0.76	2.71	0.087	0.027	0.027	0.091	0.88	3.23
	$\hat{\beta}$	-0.037	0.163	0.154	0.167	0.07	0.24	0.007	0.045	0.044	0.045	0.07	0.15	0.010	0.030	0.030	0.032	0.06	0.32
	$\hat{\gamma}$	0.055	0.116	0.119	0.128	0.06	0.46	0.007	0.033	0.033	0.034	0.05	0.22	0.002	0.021	0.023	0.021	0.04	0.10
2sCOPE-MGJK	$\hat{\alpha}$	-0.042	0.786	0.584	0.788	0.04	0.07	-0.005	0.085	0.095	0.085	0.04	0.05	-0.003	0.077	0.085	0.077	0.03	0.03
	$\hat{\phi}$	-0.014	0.083	0.068	0.084	0.05	0.20	-0.001	0.024	0.024	0.024	0.05	0.03	-0.001	0.023	0.023	0.023	0.06	0.05
	$\hat{\beta}$	-0.018	0.195	0.182	0.196	0.06	0.10	0.013	0.044	0.044	0.046	0.07	0.29	0.012	0.030	0.030	0.032	0.07	0.41
	$\hat{\gamma}$	0.007	0.310	0.231	0.310	0.05	0.03	-0.008	0.033	0.033	0.034	0.06	0.24	-0.006	0.021	0.022	0.022	0.05	0.27

Notes: The data generating process is (25). See also notes to Table A.1.

Table A.3: MC results of FE, 2sCOPE-FE, and 2sCOPE-MGJK estimators in Case 2.2: dynamic panels under slope endogeneity

		Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}	Bias	SD	$\hat{se}$	RMSE	Size	t_{bias}
		$T = 10, n = 100$						$T = 50, n = 100$						$T = 100, n = 100$					
FE	$\hat{\alpha}$	0.300	0.433	0.253	0.527	0.36	1.19	0.248	0.315	0.202	0.401	0.35	1.23	0.254	0.290	0.195	0.385	0.35	1.30
	$\hat{\phi}$	-0.056	0.042	0.040	0.070	0.34	1.40	0.064	0.035	0.033	0.072	0.49	1.94	0.081	0.034	0.032	0.088	0.71	2.54
	$\hat{\beta}$	0.070	0.085	0.091	0.111	0.11	0.77	0.040	0.072	0.081	0.083	0.06	0.50	0.028	0.069	0.080	0.075	0.04	0.36
	$\hat{\gamma}$	0.025	0.112	0.114	0.114	0.05	0.22	0.003	0.082	0.087	0.082	0.04	0.03	0.001	0.074	0.083	0.074	0.04	0.02
2sCOPE-FE	$\hat{\alpha}$	0.313	0.704	0.590	0.770	0.15	0.53	0.248	0.351	0.259	0.430	0.25	0.96	0.248	0.307	0.223	0.395	0.29	1.11
	$\hat{\phi}$	-0.056	0.042	0.040	0.070	0.34	1.40	0.064	0.035	0.033	0.072	0.49	1.94	0.081	0.034	0.032	0.088	0.71	2.54
	$\hat{\beta}$	0.058	0.597	0.573	0.599	0.06	0.10	0.040	0.183	0.188	0.187	0.05	0.21	0.034	0.132	0.140	0.136	0.05	0.25
	$\hat{\gamma}$	0.024	0.115	0.118	0.118	0.05	0.21	0.003	0.082	0.087	0.082	0.04	0.03	0.001	0.074	0.083	0.074	0.04	0.02
2sCOPE-MGJK	$\hat{\alpha}$	0.006	1.097	0.983	1.097	0.04	0.01	0.000	0.191	0.194	0.191	0.05	0.00	-0.004	0.137	0.141	0.137	0.04	0.03
	$\hat{\phi}$	-0.013	0.104	0.075	0.105	0.05	0.18	-0.001	0.025	0.024	0.025	0.05	0.04	-0.001	0.023	0.024	0.023	0.06	0.04
	$\hat{\beta}$	-0.023	0.608	0.584	0.608	0.06	0.04	0.000	0.182	0.188	0.182	0.05	0.00	0.006	0.131	0.140	0.131	0.04	0.04
	$\hat{\gamma}$	0.010	0.305	0.224	0.306	0.03	0.04	-0.002	0.075	0.081	0.075	0.04	0.03	0.000	0.070	0.080	0.070	0.03	0.00

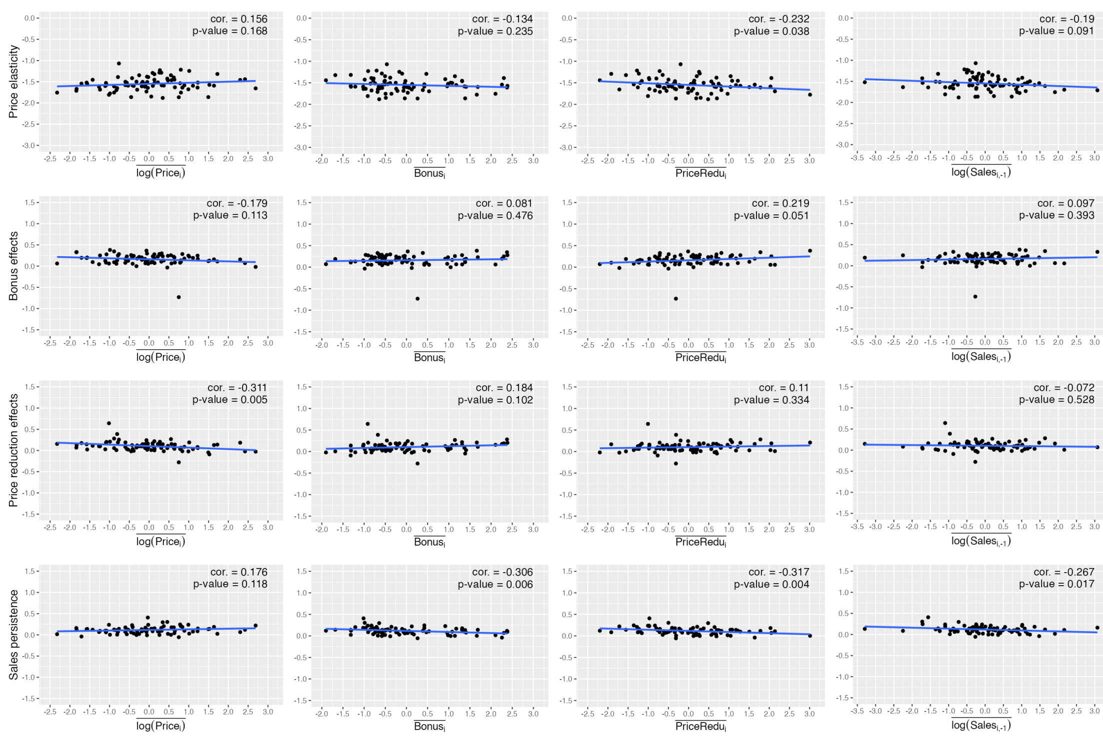
Notes: The data generating process is (25). See also notes to Table A.1.

A.4 Supplementary empirical results using Dominick's scanner data

For the cereal category, Figure A.1 shows scatter plots of 2sCOPE-MG store-specific estimates of the dynamic sales response model in (28) of the main paper against store-specific regressor means. The estimation results using a static sales response model are reported in Table A.4. Table A.5 presents results using a dynamic sales response model with lagged sales and lagged price, and the corresponding scatter plots are shown in Figures A.2 and A.3. See subsection 5.1 in the main paper for details.

Tables A.6 and A.7 provide FE, MG, and 2sCOPE-FE estimation results across 21 categories in Dominick's scanner database, respectively, for the dynamic sales response model in (28) of the main paper. Table A.8 shows the 2sCOPE-MG estimation results assuming store-specific coefficients in the Gaussian copula model, where the estimates are comparable to those in Table 11 in the main paper with homogeneous Gaussian copula coefficients. See subsection 5.2 in the main paper for our discussions.

Figure A.1: Scatter plots of 2sCOPE-MG store-specific estimates on store-specific mean of regressors in a dynamic panel data model with lagged sales as regressors



Notes: The x-axis is computed as $\frac{\bar{x}_{i,(j)} - n^{-1}\sum_{i=1}^{n}\bar{x}_{i,(j)}}{\sqrt{n^{-1}\sum_{i=1}^{n}(\bar{x}_{i,(j)} - n^{-1}\sum_{i=1}^{n}\bar{x}_{i,(j)})^2}}$ with $\bar{x}_{i,(j)} = T^{-1}\sum_{t=1}^{T} x_{it,(j)}$ for each regressor j, denoted as $\overline{\log(Price_i)}$, $\overline{Bonus_i}$, $\overline{PriceRedu_i}$, and $\overline{\log(Sales_{i,-1})}$, respectively. In each subplot, "cor." and "p-value" denote the Pearson correlation coefficient between the 2sCOPE-MG store-specific estimates and store-specific mean of regressors, and the associated p-value. See notes under Table 10 in the main paper for the model and estimation procedure.

Table A.4: Estimates of average price elasticity and promotion effects on cereal sales in a static panel data model using a homogeneous Gaussian copula

			Copula generated regressors		IV	
	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	FE	MG	2sCOPE-FE	2sCOPE-MG	FE-IV	MG-IV
$\log(Price_{it})$	-1.199	-1.106	-1.702	-1.587	-1.679	-1.554
	(0.055)	(0.028)	(0.090)	(0.103)	(0.130)	(0.050)
$Bonus_{it}$	0.203	0.199	0.158	0.159	0.190	0.212
	(0.025)	(0.017)	(0.029)	(0.032)	(0.023)	(0.020)
$PriceRedu_{it}$	0.152	0.143	0.089	0.094	0.113	0.113
	(0.030)	(0.017)	(0.033)	(0.034)	(0.027)	(0.023)
Store fixed effects	Y	Y	Y	Y	Y	Y
Week fixed effects	Y	Y	Y	Y	Y	Y
Test of price endogeneity						
Pearson cor.	-	-	0.269	0.232	-	-
p-value	-	-	0.000	0.000	-	-
Slope endogeneity	-	Y	-	Y	-	Y
Regressor endogeneity	-	-	Y	Y	Y	Y
No. observations	13,600	13,600	13,600	13,600	11,520	11,520

Notes: The estimates are based on a balanced panel of 80 stores over 170 weeks (1990–1994) of the cereals category from Dominick's database. The static sales response model is given by $\log(Sales_{it}) = \alpha_i + \tau_t + \beta_{i1}\log(Price_{it}) + \beta_{i2}Bonus_{it} + \beta_{i3}PriceRedu_{it} + \xi_{it}$, where α_i and τ_t denotes store and week fixed effects, respectively. The coefficients of the Gaussian copula model are assumed to be homogeneous across stores. (i) Price, bonus, and price reduction are computed as market share weighted averages over UPCs sold in each store. (ii) $\log(Sales_{it})$ and $\log(Price_{it})$ are detrended prior to estimation, using linear and quadratic trends, respectively. (iii) To construct an instrument for price, we consider the weekly prices of UPCs average over different stores (for those UPCs whose prices are observed over all periods), and we further aggregate the prices over UPCs for each store with "predetermined" weights. The weights are computed as market shares of these UPCs in each store average over the first 26 weeks, which are excluded from the sample used in IV estimation ($T = 144$). The first stage regression includes time fixed effects and the exogenous variables, and assumes homogeneous and heterogeneous slopes of the IV for FE and MG estimators, respectively.

Table A.5: Estimates of average price elasticity and promotion effects on cereal sales on cereal sales in a dynamic panel data model with lagged sales and lagged price as regressors using a homogeneous Gaussian copula

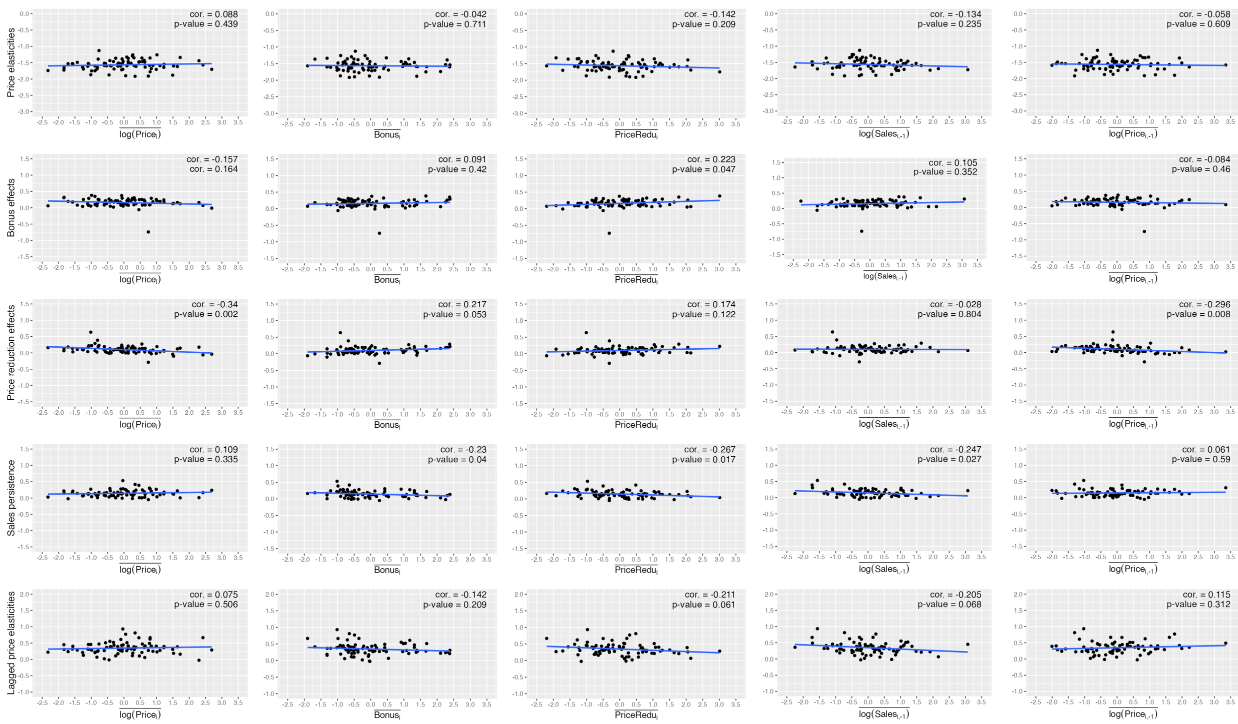
			Copula generated regressors		IV	
	(1)	(2)	(3)	(4)	(5)	(6)
Estimator	FE	MG	2sCOPE-FE	2sCOPE-MG	FE-IV	MG-IV
$\log(Sales_{i,t-1})$	0.115	0.148	0.112	0.145	0.108	0.114
	(0.024)	(0.015)	(0.024)	(0.017)	(0.020)	(0.020)
$\log(Price_{it})$	-1.289	-1.214	-1.675	-1.566	-1.868	-1.670
	(0.058)	(0.021)	(0.087)	(0.097)	(0.144)	(0.047)
$\log(Price_{i,t-1})$	0.360	0.375	0.321	0.346	0.095	0.465
	(0.049)	(0.023)	(0.051)	(0.055)	(0.027)	(0.027)
$Bonus_{it}$	0.195	0.189	0.161	0.159	0.608	0.203
	(0.024)	(0.019)	(0.028)	(0.031)	(0.062)	(0.027)
$PriceRedu_{it}$	0.143	0.136	0.094	0.100	0.180	0.103
	(0.029)	(0.016)	(0.032)	(0.032)	(0.023)	(0.022)
Store fixed effects	Y	Y	Y	Y	Y	Y
Week fixed effects	Y	Y	Y	Y	Y	Y
Test of price endogeneity						
Pearson cor.	-	-	0.218	0.179	-	-
p-value	-	-	0.000	0.000	-	-
Slope endogeneity	-	Y	-	Y	-	Y
Regressor endogeneity	-	-	Y	Y	Y	Y
No. observations	13,520	13,520	13,520	13,520	11,440	11,440

Notes: The estimates are based on a balanced panel of 80 stores over 169 weeks (1990–1994) of the cereals category from Dominick's database. The dynamic sales repsonse model is given by $\log(Sales_{it}) = \alpha_i + \tau_t + \phi_i \log(Sales_{i,t-1}) + \beta_{i1} \log(Price_{it}) + \phi_{i2} \log(Price_{i,t-1}) + \beta_{i2} Bonus_{it} + \beta_{i3} PriceRedu_{it} + \xi_{it}$, where α_i and τ_t denotes store and week fixed effects, respectively. The coefficients of the Gaussian copula model are assumed to be homogeneous across stores. (i) Price, bonus, and price reduction are computed as market share weighted averages over UPCs sold in each store. (ii) $\log(Sales_{it})$ and $\log(Price_{it})$ are detrended prior to estimation, using linear and quadratic trends, respectively. (iii) To construct an instrument for price, we consider the weekly prices of UPCs average over different stores (for those UPCs whose prices are observed over all periods), and we further aggregate the prices over UPCs for each store with "predetermined" weights. The weights are computed as market shares of these UPCs in each store average over the first 26 weeks, which are excluded from the sample used in IV estimation ($T = 143$). The first stage regression includes time fixed effects and the exogenous variables, and assumes homogeneous and heterogeneous slopes of the IV for FE and MG estimators, respectively.

Figure A.2: Scatter plots of 2sCOPE-MG store-specific estimates on within-store regressor variations in a dynamic panel data model with lagged sales and lagged price as regressors

Notes: The x-axis is computed as $\frac{T^{-1}\sum_{t=1}^{T}\left(x_{it,(j)}-\bar{x}_{i,(j)}\right)^2}{(nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{it,(j)}-\bar{x}_{i,(j)}\right)^2}$ with $\bar{x}_{i,(j)} = T^{-1}\sum_{t=1}^{T} x_{it,(j)}$ for each regressor j, denoted as $V(\log(Price_i))$, $V(Bonus_i)$, $V(PriceRedu_i)$, $V(\log(Sales_{i,-1}))$, and $V(\log(Price_{i,-1}))$, respectively. In each sub-plot, "cor." and "p-value" denote the Pearson correlation coefficient between the 2sCOPE-MG store-specific estimates and within-store regressor variations, and the associated p-value, respectively. See notes under Table A.5 for the model and estimation procedure.

Figure A.3: Scatter plots of 2sCOPE-MG store-specific estimates on store-specific mean of regressors in a dynamic panel data model with lagged sales and lagged price as regressors



Notes: The x-axis is computed as $\frac{\bar{x}_{i,(j)} - n^{-1}\sum_{i=1}^{n}\bar{x}_{i,(j)}}{\sqrt{n^{-1}\sum_{i=1}^{n}(\bar{x}_{i,(j)} - n^{-1}\sum_{i=1}^{n}\bar{x}_{i,(j)})^2}}$ with $\bar{x}_{i,(j)} = T^{-1}\sum_{t=1}^{T} x_{it,(j)}$ for each regressor j, denoted as $\overline{\log(Price_i)}$, $\overline{Bonus_i}$, $\overline{PriceRedu_i}$, $\overline{\log(Sales_{i,-1})}$, and $\overline{\log(Price_{i,-1})}$, respectively. In each sub-plot, "cor." and "p-value" denote the Pearson correlation coefficient between the 2sCOPE-MG store-specific estimates and store-specific mean of regressors, and the associated p-value. See notes under Table A.5 for the model and estimation procedure.

Table A.6: FE estimated average price elasticity, promotion effects on sales, and sales persistence in a dynamic sales response model across 21 categories using a homogeneous Gaussian copula

	Price elasticity			Bonus effect			Promotion effect			Sales persistence		
Category	FE estimate	$\hat{se}$	Bias	FE estimate	$\hat{se}$	Bias	FE estimate	$\hat{se}$	Bias	FE estimate	$\hat{se}$	Bias
Food categories												
Beer	-0.962	(0.052)	0.094	0.203	(0.032)	0.040	0.580	(0.134)	0.142	0.215	(0.041)	0.013
Bottled juices	-1.349	(0.063)	0.216	0.089	(0.022)	0.001	0.194	(0.033)	0.045	0.149	(0.026)	-0.013
Cereals	-1.200	(0.056)	0.350	0.205	(0.024)	0.045	0.153	(0.030)	0.050	0.089	(0.021)	-0.027
Cheeses	-1.484	(0.070)	0.904	0.072	(0.022)	0.037	0.053	(0.024)	0.040	0.131	(0.028)	-0.033
Cookies	-1.151	(0.063)	0.105	0.121	(0.025)	0.014	0.252	(0.032)	0.037	0.162	(0.025)	-0.014
Crackers	-0.724	(0.055)	0.121	0.318	(0.043)	0.069	0.409	(0.044)	0.096	0.158	(0.022)	-0.003
Canned soup	-0.951	(0.170)	0.194	0.166	(0.036)	-0.001	0.270	(0.041)	0.073	0.184	(0.034)	0.015
Frozen dinners	-1.634	(0.045)	0.339	0.137	(0.025)	0.096	0.379	(0.031)	0.144	0.142	(0.017)	-0.008
Frozen entrees	-0.968	(0.063)	0.084	0.288	(0.032)	0.027	0.600	(0.060)	0.064	0.128	(0.030)	-0.006
Oatmeal	-1.021	(0.051)	0.396	-0.115	(0.026)	0.000	0.179	(0.055)	0.154	0.235	(0.031)	0.039
Soft drinks	-1.267	(0.085)	0.007	0.064	(0.029)	-0.010	0.303	(0.040)	0.004	0.089	(0.023)	-0.017
Canned tuna	-1.210	(0.076)	0.099	0.008	(0.024)	0.047	0.157	(0.045)	0.076	0.141	(0.018)	0.020
Non-food categories												
Analgesics	-1.081	(0.048)	0.015	0.353	(0.084)	0.076	0.282	(0.078)	0.103	0.345	(0.132)	0.186
Dish detergent	-1.281	(0.106)	0.267	0.167	(0.028)	0.055	0.424	(0.049)	0.221	0.096	(0.020)	-0.007
Grooming products	-0.888	(0.026)	0.009	0.122	(0.030)	0.010	0.569	(0.059)	0.108	0.222	(0.038)	0.020
Laundry detergents	-1.831	(0.083)	0.153	0.065	(0.026)	-0.044	0.239	(0.045)	0.042	0.075	(0.016)	-0.008
Paper towels	-1.574	(0.065)	0.172	0.035	(0.021)	-0.013	0.179	(0.027)	0.019	0.100	(0.020)	-0.002
Shampoos	-0.778	(0.057)	0.187	0.187	(0.043)	0.072	0.503	(0.051)	0.140	0.241	(0.040)	0.035
Soaps	-0.552	(0.069)	0.653	0.105	(0.023)	0.038	0.264	(0.060)	0.109	0.287	(0.043)	0.023
Toothpastes	-1.090	(0.052)	0.443	0.282	(0.041)	0.040	0.377	(0.044)	0.133	0.115	(0.019)	-0.008
Bathroom tissues	-1.129	(0.045)	0.615	0.107	(0.026)	-0.012	0.219	(0.027)	0.101	0.062	(0.018)	-0.014

Notes: The table summarizes estimation results of the dynamic sales response model given by (28) in the main paper, based on the FE method using a balanced store-week panel dataset for each category from Dominick's scanner database. The bias is calculated by subtracting the respective 2sCOPE-MG estimate from the FE estimate.

Table A.7: MG and 2sCOPE-FE estimated average price elasticity, promotion effects on sales, and sales persistence in a dynamic sales response model across 21 categories using a homogeneous Gaussian copula

	Price elasticity				Bonus effect				Price reduction effect				Sales persistence			
	MG		2sCOPE-FE		MG		2sCOPE-FE		MG		2sCOPE-FE		MG		2sCOPE-FE	
Category	estimate	$\hat{se}$	estimate	$\hat{se}$	estimate	$\hat{se}$	estimate	$\hat{se}$	estimate	$\hat{se}$	estimate	$\hat{se}$	estimate	$\hat{se}$	estimate	$\hat{se}$
Food categories																
Beer	-1.013	(0.033)	-0.911	(0.123)	0.172	(0.024)	0.214	(0.043)	0.442	(0.043)	0.585	(0.139)	0.202	(0.017)	0.215	(0.041)
Bottled juices	-1.253	(0.025)	-1.700	(0.095)	0.097	(0.017)	0.078	(0.022)	0.203	(0.017)	0.128	(0.036)	0.161	(0.013)	0.150	(0.026)
Cereals	-1.089	(0.022)	-1.700	(0.089)	0.199	(0.020)	0.160	(0.028)	0.150	(0.014)	0.091	(0.033)	0.117	(0.015)	0.089	(0.021)
Cheeses	-1.245	(0.033)	-2.608	(0.240)	0.082	(0.021)	0.027	(0.021)	0.050	(0.016)	0.022	(0.024)	0.163	(0.013)	0.132	(0.028)
Cookies	-1.072	(0.027)	-1.386	(0.098)	0.133	(0.017)	0.086	(0.025)	0.251	(0.014)	0.198	(0.032)	0.175	(0.014)	0.163	(0.025)
Crackers	-0.650	(0.038)	-1.017	(0.087)	0.269	(0.022)	0.290	(0.042)	0.357	(0.020)	0.321	(0.051)	0.161	(0.012)	0.158	(0.022)
Canned soup	-0.953	(0.045)	-1.001	(0.393)	0.198	(0.031)	0.156	(0.054)	0.234	(0.028)	0.257	(0.087)	0.168	(0.025)	0.184	(0.034)
Frozen dinners	-1.531	(0.026)	-2.109	(0.060)	0.127	(0.014)	0.044	(0.025)	0.362	(0.014)	0.222	(0.031)	0.149	(0.009)	0.140	(0.017)
Frozen entrees	-0.905	(0.034)	-1.160	(0.108)	0.297	(0.018)	0.242	(0.035)	0.583	(0.016)	0.527	(0.064)	0.134	(0.010)	0.127	(0.030)
Oatmeal	-0.972	(0.029)	-1.490	(0.096)	-0.061	(0.027)	-0.173	(0.029)	0.090	(0.019)	0.095	(0.055)	0.198	(0.016)	0.235	(0.031)
Soft drinks	-1.189	(0.022)	-1.464	(0.225)	0.086	(0.019)	0.037	(0.038)	0.311	(0.015)	0.273	(0.050)	0.106	(0.016)	0.089	(0.023)
Canned tuna	-1.044	(0.036)	-1.667	(0.129)	0.001	(0.014)	-0.059	(0.027)	0.139	(0.016)	0.034	(0.050)	0.121	(0.010)	0.141	(0.018)
Non-food categories																
Analgesics	-1.093	(0.029)	-0.986	(0.334)	0.277	(0.028)	0.360	(0.960)	0.180	(0.032)	0.301	(0.109)	0.159	(0.021)	0.345	(0.132)
Dish detergent	-0.920	(0.054)	-2.313	(0.225)	0.169	(0.018)	0.064	(0.029)	0.287	(0.019)	0.225	(0.054)	0.103	(0.010)	0.095	(0.020)
Grooming products	-0.866	(0.017)	-1.021	(0.089)	0.117	(0.012)	0.097	(0.035)	0.463	(0.021)	0.558	(0.061)	0.202	(0.021)	0.222	(0.038)
Laundry detergents	-1.630	(0.021)	-2.448	(0.089)	0.114	(0.017)	0.064	(0.026)	0.240	(0.018)	0.157	(0.043)	0.083	(0.011)	0.075	(0.016)
Paper towels	-1.457	(0.022)	-1.973	(0.117)	0.057	(0.018)	0.023	(0.022)	0.178	(0.014)	0.150	(0.026)	0.100	(0.013)	0.101	(0.020)
Shampoos	-0.730	(0.025)	-1.029	(0.147)	0.179	(0.021)	0.116	(0.053)	0.451	(0.021)	0.393	(0.080)	0.206	(0.022)	0.242	(0.040)
Soaps	-0.560	(0.040)	-1.142	(0.340)	0.105	(0.013)	0.071	(0.027)	0.220	(0.024)	0.187	(0.079)	0.263	(0.017)	0.287	(0.043)
Toothpastes	-0.972	(0.029)	-1.669	(0.143)	0.292	(0.012)	0.232	(0.043)	0.378	(0.016)	0.219	(0.057)	0.123	(0.015)	0.115	(0.019)

Notes: The table summarizes estimation results of the dynamic sales response model given by (28) in the main paper, based on the MG and 2sCOPE-FE methods using a balanced store-week panel dataset for each category from Dominick's scanner database. The coefficients of the Gaussian copula model are assumed to be homogeneous across stores.

Table A.8: 2sCOPE-MG estimated average price elasticity, promotion effects on sales, and sales persistence in a dynamic sales response model across 21 categories using a heterogeneous Gaussian copula

	Price elasticity				Bonus effect		Price reduction effect		Sales persistence	
	2sCOPE-MG		Regressor	Slope	2sCOPE-MG		2sCOPE-MG		2sCOPE-MG	
Category	estimate	$\hat{se}$	endogenity	endogeneity	estimate	$\hat{se}$	estimate	$\hat{se}$	estimate	$\hat{se}$
Food categories										
Beer	-1.014	(0.148)	0	+	0.171	(0.048)	0.433	(0.152)	0.200	(0.026)
Bottled juices	-1.550	(0.097)	+	-	0.095	(0.022)	0.154	(0.036)	0.162	(0.017)
Cereals	-1.638	(0.113)	+	0	0.155	(0.032)	0.089	(0.035)	0.115	(0.016)
Cheeses	-2.477	(0.305)	+	-	0.036	(0.026)	0.011	(0.024)	0.162	(0.023)
Cookies	-1.301	(0.106)	+	0	0.107	(0.026)	0.218	(0.039)	0.175	(0.019)
Crackers	-0.811	(0.103)	+	-	0.254	(0.038)	0.327	(0.052)	0.161	(0.018)
Canned soup	-1.328	(0.160)	+	+	0.147	(0.048)	0.169	(0.042)	0.166	(0.031)
Frozen dinners	-2.009	(0.063)	+	-	0.038	(0.024)	0.226	(0.030)	0.148	(0.014)
Frozen entrees	-1.073	(0.132)	+	0	0.262	(0.043)	0.534	(0.061)	0.132	(0.023)
Oatmeal	-1.480	(0.109)	+	0	-0.125	(0.029)	0.007	(0.046)	0.184	(0.029)
Soft drinks	-1.318	(0.210)	+	-	0.067	(0.035)	0.293	(0.038)	0.108	(0.017)
Canned tuna	-1.292	(0.103)	+	-	-0.037	(0.028)	0.079	(0.036)	0.121	(0.016)
Non-food categories										
Analgesics	-1.027	(0.217)	-	0	0.275	(0.073)	0.181	(0.062)	0.158	(0.019)
Dish detergent	-1.055	(0.229)	0	-	0.158	(0.028)	0.290	(0.042)	0.097	(0.017)
Grooming products	-0.846	(0.095)	-	-	0.120	(0.029)	0.463	(0.052)	0.202	(0.029)
Laundry detergents	-2.002	(0.145)	+	-	0.113	(0.028)	0.205	(0.048)	0.083	(0.016)
Paper towels	-1.749	(0.124)	+	-	0.050	(0.021)	0.163	(0.026)	0.102	(0.014)
Shampoos	-1.184	(0.123)	+	+	0.054	(0.054)	0.283	(0.064)	0.203	(0.021)
Soaps	-1.160	(0.197)	+	0	0.079	(0.025)	0.146	(0.062)	0.262	(0.021)
Toothpastes	-1.501	(0.147)	+	-	0.241	(0.039)	0.236	(0.055)	0.123	(0.014)
Bathroom tissues	-1.583	(0.124)	+	0	0.124	(0.024)	0.134	(0.024)	0.079	(0.016)
Average bias	0.251	(0.271)	0.337	-0.142	0.028	(0.041)	0.093	(0.055)	0.011	(0.046)

Notes: This table summarizes estimation results of the dynamic sales response model in (28) in the main paper, using a balanced store-week panel dataset for each of 21 categories. The results are based on the 2sCOPE-MG estimator, assuming store-specific Gaussian copula coefficients. In the regressor endogeneity column, "+"/"-" denotes a significant positive/negative correlation between price and estimated structural errors. In the slope endogeneity column, "+"/"-" indicates whether the 2sCOPE-MG estimate is higher/lower than the 2sCOPE-FE estimate, respectively, while "0" denotes an absolute difference smaller than 0.1. In the last row, bias is calculated as the FE estimate minus the 2sCOPE-MG estimate, averaged over 21 categories, with standard deviations of biases shown in brackets. For price elasticity, regressor endogeneity bias is computed as the MG estimate minus the 2sCOPE-MG estimate, and slope endogeneity bias is computed as the 2sCOPE-FE estimate minus the respective 2sCOPE-MG estimate.