行政院國家科學委員會專題研究計畫成果報告

廣義 LISREL 模式（IV）
Generalized LISREL Models (IV)

計畫類別：☑ 個別型計畫　　□ 整合型計畫

計畫編號：NSC 89－2118－M－002－001

執行期間：88 年 8 月 1 日至 89 年 7 月 31 日

個別型計畫：計畫主持人：胡賦強 助理教授
　　　　　　共同主持人：

整合型計畫：總計畫主持人：
　　　　　　子計畫主持人：

註：整合型計畫總報告與子計畫成果報告請分開編印各成一冊，彙整一起繳送國科會。

處理方式：□ 可立即對外提供參考
（請打✓）☑ 一年後可對外提供參考
　　　　　□ 兩年後可對外提供參考
　　　　　（必要時，本會得展延發表時限）

執行單位：國立台灣大學 公共衛生學院 流行病學研究所
　　　　　生物醫學統計組

中華民國　90　年　3　月　15　日

(Submitted to the *Journal of the American Statistical Association*)

Generalized Simultaneous Equations Model and Generalized Path Analysis for Recursive Systems (I): The Indirect Least Squares and Two-Stage Least Squares Estimators

Tien-Lung Tsai, Wen-Yi Shau, and Fu-Chang Hu[1]

Division of Biostatistics
Graduate Institute of Epidemiology
College of Public Health
National Taiwan University
Taipei, Taiwan 100
R.O.C.

December 14, 2000

[1]Correspondence: Fu-Chang Hu, Division of Biostatistics, Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, 1 Jen-Ai Road, Section 1, Room 1541, Taipei, Taiwan 100, R.O.C., Phone: +886 2 / 2394-2050 or +886 2 / 2312-3456 × 8349, Fax: +886 2 / 2351-1955, E-mail: fchu@ha.mc.ntu.edu.tw.

This paper is based on the first part of Tsai's doctoral dissertation under Hu's advice. We are very grateful to the Doctoral Oral Examination Committee members for their helpful discussions and comments. And, Hu wishes to thank Professor Karl Jöreskog for his interesting introduction of the statistical software *LISREL* (Version 8.0) in an invited workshop held in Taipei, Taiwan, R.O.C., August 29-31, 1995. This research project was financially supported by the National Science Council of the Republic of China (NSC 86-2115-M-002-016, NSC 87-2118-M-002-003, NSC 88-2118-M-002-008, and NSC 89-2118-M-002-001).

Contents

1 INTRODUCTION

In the past 50 years or so, econometricians have developed *simultaneous equations models* (SiEM) for exploring and/or examining the plausible structural relationships among several endogenous variables given a set of exogenous variables. Independently, *path analysis* (PA) has been used extensively in various areas of social sciences for exploring and/or examining the plausible causal relationships among several response variables given a set of independent variables. These two kinds of statistical models are essentially the same except that their estimation methods are different.

Specifically, there are two major classes of SiEM/PA models:

1. Recursive SiEM/PA Models:
   (a) *Fully* recursive SiEM/PA models.
   (b) *Partially* recursive SiEM/PA models.
2. Non-Recursive SiEM/PA Models.

Given a set of independent variables (or exogenous variables), a system without any reciprocal effects between the response variables (or endogenous variables) of the structural equations is called the *recursive* model, which is particularly useful in analyzing longitudinal data due to the temporal order of the responses (Finkel 1995). Otherwise, it is called the *non-recursive* model. When all the error terms of the structural equations in a recursive model are mutually uncorrelated, it is called the *fully* recursive model, for which consistent estimates of the structural coefficients can be obtained equation-by-equation separately. A recursive model with correlated error terms between some of the structural equations is called the *partially* recursive model, for which the equation-by-equation approach is usually not valid and the estimation of the structural coefficients should be based on the whole system of equations. The terminology — *recursive* versus *non-recursive* — may not be so intuitively understandable. Greene (2000, p. 659) provides the following explanation: "The joint determination of the variables in this (recursive) model is *recursive*. The first is completely determined by the exogenous factors. Then, given the first, the second is likewise determined, and so on." In this study, we shall focus on the recursive models and reserve the more complicated non-recursive models in a following research project. And, note that we shall make no distinction between independent variables and exogenous variables and between response variables and endogenous variables in this paper.

1.1 Motivation

Both SiEM and PA require that all the response variables be continuous random variables and their joint distribution be multivariate normal (or, at least, symmetric). However, this distributional assumption may not be appropriate especially in many biological, medical, social, and public health studies. As an example, the following hypothetical causal model might be of interest to the investigator of a community-based observational study on the health of the elderly in Taiwan (Wu 1995):

$$
\begin{array}{ccc}
\mathbf{X_1} & \longrightarrow & \text{Utilization of Health Care Facilities} \\
 & \nearrow & \text{(Yes/No)} \\
\mathbf{X_2} & & \downarrow \uparrow \\
 & \searrow & \\
\mathbf{X_3} & \longrightarrow & \text{Health Status}
\end{array}
$$

where $\mathbf{X_1}$, $\mathbf{X_2}$, and $\mathbf{X_3}$ are three sets of independent variables, "Utilization of Health Care Facilities" (Y_1) is a *binary* response variable, and "Health Status" (Y_2) is a *continuous* response variable. We assume that the observed y_1 and y_2 are the equilibrium values satisfying the above non-recursive model (see, e.g., Amemiya (1985, pp. 228-229)). In order to make reasonable suggestions for the remedies of the health care policy, the investigator is interested not only in the effects of the independent variables $\mathbf{X_1}$ and $\mathbf{X_2}$ on the binary response Y_1, but also in the effects of the intermediate variable Y_1 with the independent variables $\mathbf{X_2}$ and $\mathbf{X_3}$ on the continuous response Y_2.

In fact, the importance of discrete response data is evident by the popularity of the logistic and Poisson regressions in the biostatistical applications. Thus, as inspired by the generalization of the linear models to the *generalized linear models* (GLMs) (Nelder and Wedderburn 1972), we are interested in generalizing the linear SiEM and PA to deal with continuous, binary, counts, or mixed responses in partially recursive models and call them the "generalized simultaneous equations models" (GSiEM) or equivalently the "generalized path analysis" (GPA).

1.2 Problem

Yet, as said by Davidson and MacKinnon (1993, p. 662),

> "The problem with models that are nonlinear in the endogenous variables is that for such models there is nothing equivalent to the unrestricted reduced form for a linear simultaneous equations model. It is generally difficult or impossible to solve for the endogenous variables as functions of the exogenous variables and the error terms. Even when it is possible, $\mathbf{Y}_t$ will almost always depend nonlinearly on both the exogenous variables and the error terms."

Hence, developing GSiEM/GPA models is a challenging task due to

1. the difficulty in obtaining the reduced form of the model and
2. the potential nonlinearity in variables, parameters, and/or errors in some of the reduced-form equations.

In this set of two papers, we will develop the estimation methods for the *partially* recursive GSiEM/GPA models and discuss the statistical properties of our estimators. Specifically, in this paper, we combine the *indirect least squares* (ILS) and *two-stage least squares* (2SLS) estimation methods of SiEM with the *iterative reweighted least squares* (IRLS) algorithm of GLMs to estimate the structural coefficients of a partially recursive model with the response variables of mixed types. In particular, with the aid of the IRLS algorithm, we derive the *reduced form* of such a nonlinear recursive model, which is crucial for the ILS and 2SLS estimators. The performances of various estimators are compared in the simulations. The applications including a real example from a medical study are presented in another separate paper due to the restriction of the paper length.

2 REVIEW

The linear SiEM and PA models for continuous responses have been developed independently in the fields of economy and social sciences. In fact, they have the same model specification, but differ in the estimation methods. The estimation of the structural coefficients in a SiEM is based on the *first moments* of the response variables, whereas the estimation of the structural coefficients in a PA model is based on the *second moments* of the response variables. See, for example, Greene (2000, Chap. 16, pp. 652-711) and Bollen (1989, Chap. 4, pp. 80-150) for details. In the following two subsections, we shall give a brief review of nonlinear and discrete SiEM/PA models respectively. Finally, we end this section with an introduction to GLMs.

2.1 Simultaneous Equations Models (SiEM)

2.1.1 Nonlinear SiEM

A *nonlinear* SiEM for the *i*th endogenous variable Y_i $(i = 1, 2, \cdots, M)$ is of the form

$$Y_i = f_i(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta}_i) + \epsilon_i$$

where $\mathbf{Y}_i$ is a vector of endogenous variables, $\mathbf{X}_i$ is a vector of exogenous variables, $\boldsymbol{\beta}_i$ is a vector of unknown parameters, and ϵ_i is a scalar i.i.d. random variable with mean 0 and variance σ^2. See, for example, Goldfeld and Quandt (1968), Zellner, Huang, and Chau (1965), Kelejian (1971), Greene (2000, Subsec. 16.5.2.f, pp. 689-690, esp., Note #36), Davidson and MacKinnon (1993, pp. 661-667), Amemiya (1985, Chap. 8, pp. 245-266), and Bowden and Turkington (1984, Chap. 5, pp. 156-201) for details.

2.1.2 Discrete SiEM

We define the *discrete* SiEM as the SiEM of which some of the responses are discrete such as "Yes/No" and "counts." First, Schmidt and Strauss (1975) proposed a *simultaneous logit model*, which was labeled by Nerlove and Press (1973) as a *multivariate logit model*. As specified by Maddala (1983, p. 108), one such model with two binary response variables is

$$\begin{aligned} \text{logit}\big[Pr(Y_1 = 1) \mid Y_2, \mathbf{X_1}\big] &= \mathbf{X_1}\boldsymbol{\beta}_1 + \gamma Y_2 \\ \text{logit}\big[Pr(Y_2 = 1) \mid Y_1, \mathbf{X_2}\big] &= \mathbf{X_2}\boldsymbol{\beta}_2 + \gamma Y_1 \end{aligned}$$

which requires that all endogenous variables be binary and the coefficients of Y_1 and Y_2 be the same.

Maddala (1983, pp. 117-125) discussed a series of six SiEM with *mixed* responses of continuous and qualitative variables, but the restriction is that an underlying continuous variable from a normal distribution, although unobserved, was assumed for each of the observed qualitative variables. For example, the observed binary response Y_2 is assumed to be generated from its underlying continuous variable Y_2^*, which has a normal distribution.

- **Model 1:**

$$\begin{aligned} Y_1 &= \mathbf{X_1}\boldsymbol{\beta}_1 + \gamma_1 Y_2 + \epsilon_1 \\ Y_2^* &= \mathbf{X_2}\boldsymbol{\beta}_2 + \gamma_2 Y_1 + \epsilon_2 \end{aligned}$$

where

$$Y_2 = \begin{cases} 1 & \text{if } Y_2^* > 0, \\ 0 & \text{otherwise.} \end{cases}$$

- **Model 2:**

$$\begin{aligned} Y_1^* &= \mathbf{X_1}\boldsymbol{\beta}_1 + \gamma_1 Y_2^* + \epsilon_1 \\ Y_2^* &= \mathbf{X_2}\boldsymbol{\beta}_2 + \gamma_2 Y_1 + \epsilon_2 \end{aligned}$$

where

$$Y_1 = \begin{cases} 1 & \text{if } Y_1^* > 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$Y_2 = \begin{cases} 1 & \text{if } Y_2^* > 0, \\ 0 & \text{otherwise.} \end{cases}$$

- **Model 3:**

$$\begin{aligned} Y_1^* &= \mathbf{X_1}\boldsymbol{\beta}_1 + \gamma_1 Y_2 + \epsilon_1 \\ Y_2^* &= \mathbf{X_2}\boldsymbol{\beta}_2 + \gamma_2 Y_1 + \epsilon_2 \end{aligned}$$

where the definitions of Y_1 and Y_2 are the same as those in Model 2.

- **Model 4:**

$$\begin{aligned} Y_1 &= \mathbf{X_1}\boldsymbol{\beta}_1 + \epsilon_1 \\ Y_2^* &= \mathbf{X_2}\boldsymbol{\beta}_2 + \gamma_2 Y_1 + \epsilon_2 \end{aligned}$$

where the definition of Y_2 is the same as that in Model 1. This model is a special case of Model 1 with $\gamma_1 = 0$.

- **Model 5:**

$$\begin{aligned} Y_1 &= \mathbf{X_1}\boldsymbol{\beta}_1 + \gamma_1 Y_2 + \epsilon_1 \\ Y_2^* &= \mathbf{X_2}\boldsymbol{\beta}_2 + \epsilon_2 \end{aligned}$$

where $V(\epsilon_2) = 1$ and the definition of Y_2 is the same as that in Model 1. This model is a special case of Model 1 with $\gamma_2 = 0$.

- **Model 6:**

$$\begin{aligned} Y_1^* &= \mathbf{X_1}\boldsymbol{\beta}_1 + \epsilon_1 \\ Y_2^* &= \mathbf{X_2}\boldsymbol{\beta}_2 + \gamma_2 Y_1 + \epsilon_2 \end{aligned}$$

where $V(\epsilon_1) = V(\epsilon_2) = 1$ and the definitions of Y_1 and Y_2 are the same as those in Model 2. This model is a special case of Models 2 and 3 with $\gamma_1 = 0$.

Moreover, the SiEM with categorical observed variables were discussed in Manski and McFadden (1981, pp. 345-472). In particular, Lee (1981) considered the SiEM with different endogenous variables, which include observable continuous variables, truncated continuous variables, underlying continuous variables, and censored dependent variables. For example, suppose that there are M endogenous variables in the model and $0 \leq M_1 \leq M_2 \leq M_3 \leq M$.

- $Y_1, \cdots, Y_{M_1}$ are the *observable* continuous variables.
- $Y_{M_1+1}, \cdots, Y_{M_2}$ are the *truncated* continuous variables, which can be observed only when $Y_i > 0$ for $M_1 < i \leq M_2$.
- $Y_{M_2+1}, \cdots, Y_{M_3}$ are the unobserved *underlying* continuous variables, but the corresponding binary indicators $I_{M_2+1}, \cdots, I_{M_3}$ are observable, where

$$I_i = \begin{cases} 1 & \text{if } Y_i > 0, \\ 0 & \text{otherwise.} \end{cases}$$

  for $M_2 < i \leq M_3$.
- $Y_{M_3+1}, \cdots, Y_M$ are the *censored* dependent variables.

Again, the estimation of the structural coefficients in the equations for the limited dependent variables such as $Y_{M_2+1}, \cdots, Y_{M_3}$ mainly makes use of the *probit* link function, which relies on the underlying normality assumption.

On the other hand, Greene (2000, pp. 135-137) considered a SiEM with the expected value of Y_1 in the second equation:

$$\begin{aligned}\log(\mu_1) &= \mathbf{X_1}\boldsymbol{\beta}_1 \\ \text{logit}\big[Pr(Y_2 = 1)\big] &= \mathbf{X_2}\boldsymbol{\beta}_2 + \gamma\,\mu_1\end{aligned}$$

where $Y_1 \sim$ Poisson (μ_1). Specifying the expected value of a discrete Y_1 in the second equation for Y_2 avoids the estimation problem, but it limits its applications in the situations where the observed value of Y_1, instead of its expected value, actually affects the probability of Y_2 being 1 as in many biomedical studies.

2.2 Path Analysis (PA)

2.2.1 Nonlinear PA Models

Kenny and Judd (1984) considered a structural equation model with *interaction effects* of two latent variables:

$$Y = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \zeta$$

and

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{bmatrix} + \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & 0 \\ 0 & \lambda_3 \\ 0 & \lambda_4 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}$$

where ξ_1 and ξ_2 are latent variables having a joint bivariate normal distribution with zero means, $\zeta \sim N(0, \psi)$, and $\delta_i \sim N(0, \theta_i)$, for $i = 1, 2, 3,$ and 4. They proposed an estimation method using the products of the corresponding observed variables. The intercepts are added into the model since Jöreskog and Yang (1996) found that the intercepts must be nonzero. In fact, Jöreskog and Yang (1996) had developed a general structural equation model with *polynomial* relationships between latent variables. For example, they considered the following second-order polynomial model

$$Y = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_{11}\xi_1^2 + \gamma_{22}\xi_2^2 + \gamma_{12}\xi_1\xi_2 + \zeta$$

of which the Kenny-Judd model (with intercepts) is a special case. This type of nonlinear structural equation models have a different structure from that of our GSiEM/GPA models.

2.2.2 Discrete PA Models

Again, we define the *discrete* PA model as the PA model of which some of the responses are discrete such as "Yes/No" and "counts." As shown in Muthén (1984), Arminger and Küsters (1985), and their later works, a typical approach to dealing with the discrete response variables in PA models is to use the *probit* link function by assuming the existences of the underlying continuous variables. On the other hand, Hellevik (1988) and Hagenaars (1993) developed the loglinear models with latent variables respectively with the restriction that all the responses and covariates be *categorical* variables or *discretized* continuous variables. And, we note that Bentler and Newcomb (1991) introduced linear structural equation models with *nonnormal* continuous response variables for the study of human health-related issues.

2.3 Generalized Linear Models (GLMs)

Let i index observations, $i = 1, 2, \cdots, n$. For the ith observation, a GLM is

$$g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}$$

where μ_i is the mean of the response variable Y_i from the *exponential family of distributions* (of the same form for all i), $\mathbf{x}_i$ is a vector of covariates, $\boldsymbol{\beta}$ is a vector of regression coefficients, and $g(\cdot)$ is a monotonic and differentiable function called the *link function* (to link μ_i with the linear combination of $\boldsymbol{\beta}$ and $\mathbf{x}_i$).

Linear regression (*identity* link), logistic regression (*logit* link), probit regression (*probit* link), and

Poisson regression (*log* link) are all special cases of GLMs. The maximum likelihood estimates (MLEs) of the unknown regression coefficients $\boldsymbol{\beta}$ in a GLM can be obtained by applying the *unified* IRLS algorithm (see, e.g., Dobson (1990, Sec. 4.4, pp. 39-42) and McCullagh and Nelder (1989, Sec. 2.5, pp. 40-43)), of which an outline is provided in **Appendix**. The statistical inference on the regression coefficients $\boldsymbol{\beta}$ is based on the asymptotic distribution of the score function $U(\boldsymbol{\beta})$ (by a central limit theorem) and the first-order Taylor expansion of $U(\boldsymbol{\beta})$ for $\widehat{\boldsymbol{\beta}}$ (see, e.g., Dobson (1990, Chap. 5, pp. 49-67)). The reader may consult Dobson (1990), McCullagh and Nelder (1989), and Fahrmeir and Tutz (1994) for more details.

3 MODEL SPECIFICATION, ASSUMPTIONS, AND INTERPRETATION

The estimation of the structural coefficients in a *fully* recursive GSiEM/GPA model is trivial since they can be estimated equation-by-equation separately. In fact, since all response variables are measured on the same group of subjects, the error terms of the equations are likely correlated as in the longitudinal data. Thus, to begin with, we consider the following *partially* recursive two-equation GSiEM/GPA model:

$$g_1(\mu_1) = \beta_{10} + \beta_{1x_1}X_1 + \beta_{1x_2}X_2 \quad (3.1)$$

$$\mu_2 = \beta_{20} + \beta_{2x_1}X_1 + \beta_{2x_3}X_3 + \gamma_{2y_1}Y_1 \quad (3.2)$$

where X_1, X_2, and X_3 are the independent variables, μ_1 and μ_2 are the means of the response variables Y_1 and Y_2 respectively,

$$Y_1 \sim \text{The exponential family of distributions, e.g., Binomial } (m, \mu_1),$$

$$Y_2 \sim \text{Normal } (\mu_2, \sigma_2^2),$$

and $g_1(\cdot)$ is the *link* function for μ_1. The subscript i, which indexes the observations ($i = 1, 2, \cdots, n$), is dropped for simplicity. It is assumed that the error term ϵ_2 of Eq. (3.2) has mean 0 and it is independent of the independent variables X_1, X_2, and X_3 respectively, i.e., $\epsilon_2 \perp X_1$, $\epsilon_2 \perp X_2$, and $\epsilon_2 \perp X_3$. A simplified path diagram for Eqs. (3.1) and (3.2) is listed below.

$$
\begin{array}{ccc}
\mathbf{X_2} & \longrightarrow & \mathbf{Y_1} \\
 & \nearrow & \\
\mathbf{X_1} & & \downarrow \\
 & \searrow & \\
\mathbf{X_3} & \longrightarrow & \mathbf{Y_2}
\end{array}
$$

As in the usual SiEM/PA models, the effects of the covariates on the corresponding response variables can be classified into three types in the GSiEM/GPA models:

1. **Direct Effect:**

   The structural coefficients in an equation represent the *direct* effects of the covariates in that equation on the mean of the response variable of the equation. For example, β_{2x_1}, β_{2x_3}, and γ_{2y_1} in Eq. (3.2) are the direct effects of X_1, X_3, and Y_1 on μ_2 respectively.

2. **Indirect Effect:**

   An *indirect* effect is the effect of the covariate in an equation on the mean of the response variable of a different equation mediated by some other covariates in the causal pathway. For example, X_1 has a direct effect on $g_1(\mu_1)$ in Eq. (3.1) and Y_1 has a direct effect on μ_2; and thus, X_1 has an indirect effect on μ_2 through Y_1.

3. **Total Effect:**

   The *total* effect of a covariate on the mean of a response variable is the "sum" of its direct and indirect effects.

Remarks. Although we consider a two-equation case, the statistical methods to be introduced in the sequel can be applied to multi-equation GSiEM/GPA Models — at least, recursively to each equation. At this time, we require that the link function for μ_2 be *identity*. The other link functions for μ_2 (e.g., *log*) and different types of Y_2's (e.g., binary responses) will be considered later. Moreover, notice that if the roles of Y_1 and Y_2 in this two-equation partially recursive GSiEM/GPA model are switched over, then the new GSiEM/GPA model

$$\begin{aligned} \mu_1 &= \beta_{10} + \beta_{1x_1} X_1 + \beta_{1x_2} X_2 \\ g_2(\mu_2) &= \beta_{20} + \beta_{2x_1} X_1 + \beta_{2x_3} X_3 + \gamma_{2y_1} Y_1 \end{aligned}$$

would be an easier case since the reduced form of the second equation is readily obtained by plugging the first equation (with its error term ϵ_1) into the second equation at the place of Y_1. Finally, as regard to the correlation between the error terms of the two structural equations Eqs. (3.1) and (3.2), we shall discuss this issue in Subsection 6.1, when we explain how to generate such data in simulations.

4 ESTIMATION

We find the estimation methods of PA based on the *second moments* of the observed variables more difficult than those of SiEM based on the *first moments* of the observed variables to be adopted for the GSiEM/GPA models. The heterogeneous variance of the response variable in a GLM due to the dependence of the *variance function* on the mean of the response variable makes the estimation based on the second moments of the observed variables even harder.

However, we may make use of the following two important tools to develop suitable estimation methods for the GSiEM/GPA models.

1. **Tool A:** The IRLS algorithm can be used to linearize GLMs.

   (a) At each iteration indexed by (m), the IRLS algorithm takes a special transformation on the original response variable Y_i to obtain a *pseudo*-response variable Z_i, no matter what type of Y_i is, to modify the property of the original response variable such that the original GLM

$$g(\mu_i) = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}$$

becomes a linear regression model

$$E\left[Z_i^{(m)}\right] = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}^{(m)}$$

until the convergence of $\widehat{\boldsymbol{\beta}}$ (see **Appendix**).

(b) Most importantly, the derived linear regression model for the pseudo-response variable Z_i at the *last* iteration of the IRLS algorithm provides an "equivalent" linear regression model for the original GLM in the sense that they have the same values of the regression coefficients.

(c) And, instead of writing specific computing programs for a variety of GLMs according to their likelihood functions to obtain MLEs, one can just apply one single *unified* IRLS algorithm for all GLMs. See, for example, the `PROC GENMOD` in the *SAS for Windows* and the `glm` function in the *S-PLUS for Windows*.

2. **Tool B:** The estimation methods previously developed for linear SiEM/PA models may be applied to the derived linear regression model for the pseudo-response variable Z_i at each iteration, including the last one, of the IRLS algorithm.

   (a) Treat the pseudo-response variable Z_i constructed at the last iteration of the IRLS algorithm for a GLM as if it is a "continuous" response variable of a linear regression model.

   (b) Then, apply the usual estimation methods for a linear SiEM/PA model to obtain estimates of the structural coefficients.

Specifically, we have the following thought in mind.

1. In solving a linear SiEM/PA model, the first step is usually to obtain its *reduced form* by substituting the endogenous variables on the right-hand side of each equation with the functions of the exogenous variables in their own equations and regrouping the corresponding coefficients. Since in the reduced form of a SiEM/PA model, all the terms on the right-hand side of each equation are exogenous variables, which have no correlations with the error term of the equation by assumption, so that the regrouped coefficients

can be estimated by the least squares method. Then, the second step is to recover the estimates of the original coefficients from the estimates of the regrouped coefficients in the reduced form. We can follow the same steps to solve the GSiEM/GPA models.

2. However, in order to substitute the various kinds of endogenous variables, e.g., a "Yes/No" response, on the right-hand side of each equation with their own set of exogenous variables, we use the *IRLS algorithm* as a tool to transform each subject's original response variable Y_i into a pseudo-response variable Z_i by the following formula:

$$Z_i = \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + g'(\widehat{\mu}_i)\,(Y_i - \widehat{\mu}_i) = \widehat{\mu}_i^* + \epsilon_i^*$$

so that we have

$$Y_i = \widehat{\mu}_i + \frac{Z_i - \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}}{g'(\widehat{\mu}_i)} = g^{-1}\left(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\right) + \epsilon_i$$

for substitution (see **Appendix**). By doing so, plugging the estimated mean of Y_1, $\widehat{\mu}_i$ (or, equivalently, denoted as $\widehat{Y}_i$), into the equations where Y_1 stays on the right-hand side as a covariate for the other endogenous variables, e.g., Y_2, will probably result in consistent estimates of the structural coefficients for Y_1 in those equations.

4.1 The Indirect Least Squares (ILS) Estimator

Now, we introduce the *indirect least squares* (ILS) method for estimating the structural coefficients in the GLM-LM GSiEM/GPA model (3.1) and (3.2).

- **Step 1:**

  Fit the GLM for the response Y_1 on the independent variables X_1 and X_2 using the IRLS algorithm to obtain the estimates $\widehat{\beta}_{10}$, $\widehat{\beta}_{1x_1}$, and $\widehat{\beta}_{1x_2}$ of the corresponding coefficients in the first equation, Eq. (3.1).

- **Step 2:**

In fact, in **Step 1**, the estimates $\widehat{\beta}_{10}$, $\widehat{\beta}_{1x_1}$, and $\widehat{\beta}_{1x_2}$ are obtained by fitting a multiple linear regression of the *pseudo*-response variable Z_1 (defined below) on the original independent variables X_1 and X_2 on the convergence:

$$\begin{aligned} Z_1 &= \widehat{\beta}_{10} + \widehat{\beta}_{1x_1} X_1 + \widehat{\beta}_{1x_2} X_2 + g_1'(\widehat{\mu}_1)(Y_1 - \widehat{\mu}_1) \\ &= \widehat{\beta}_{10} + \widehat{\beta}_{1x_1} X_1 + \widehat{\beta}_{1x_2} X_2 + \epsilon_{1,1}^*. \end{aligned}$$

Thus, we can rewrite the original response variable Y_1 as a function of the pseudo-response variable Z_1 in the following way:

$$\begin{aligned} Z_1 &= \left(\widehat{\beta}_{10} + \widehat{\beta}_{1x_1} X_1 + \widehat{\beta}_{1x_2} X_2\right) + g_1'(\widehat{\mu}_1)(Y_1 - \widehat{\mu}_1) \\ &= g_1(\widehat{\mu}_1) + g_1'(\widehat{\mu}_1)(Y_1 - \widehat{\mu}_1), \\ Z_1 - g_1(\widehat{\mu}_1) &= g_1'(\widehat{\mu}_1)(Y_1 - \widehat{\mu}_1), \\ Y_1 - \widehat{\mu}_1 &= \frac{Z_1 - g_1(\widehat{\mu}_1)}{g_1'(\widehat{\mu}_1)}, \\ Y_1 &= \widehat{\mu}_1 + \frac{Z_1 - g_1(\widehat{\mu}_1)}{g_1'(\widehat{\mu}_1)} \end{aligned}$$

where

$$\widehat{\mu}_1 = \widehat{Y}_1 = g_1^{-1}\left(\widehat{\beta}_{10} + \widehat{\beta}_{1x_1} X_1 + \widehat{\beta}_{1x_2} X_2\right).$$

In this step, we just compute $\widehat{\mu}_1$ from the first equation, Eq. (3.1).

- **Step 3:**

Notice that by plugging

$$Y_1 = \widehat{Y}_1 + \frac{Z_1 - g_1(\widehat{\mu}_1)}{g_1'(\widehat{\mu}_1)}$$

from **Step 2** into the second equation, Eq. (3.2), we have

$$\begin{aligned}
Y_2 &= \beta_{20} + \beta_{2x_1}X_1 + \beta_{2x_3}X_3 + \gamma_{2y_1}Y_1 + \epsilon_2 \\
&= \beta_{20} + \beta_{2x_1}X_1 + \beta_{2x_3}X_3 + \gamma_{2y_1}\left[\widehat{Y}_1 + \frac{Z_1 - g_1(\widehat{\mu}_1)}{g_1'(\widehat{\mu}_1)}\right] + \epsilon_2 \\
&= \beta_{20} + \beta_{2x_1}X_1 + \beta_{2x_3}X_3 + \gamma_{2y_1}\widehat{Y}_1 + \left[\gamma_{2y_1}\left(\frac{Z_1 - g_1(\widehat{\mu}_1)}{g_1'(\widehat{\mu}_1)}\right) + \epsilon_2\right] \\
&= \beta_{20} + \beta_{2x_1}X_1 + \beta_{2x_3}X_3 + \gamma_{2y_1}\widehat{Y}_1 + \epsilon^*_{2,1} \qquad (4.1)
\end{aligned}$$

where ϵ_2 is the error term of the second equation, Eq. (3.2), and as the "proxy" of Y_1,

$$\widehat{Y}_1 = g_1^{-1}\left(\widehat{\beta}_{10} + \widehat{\beta}_{1x_1}X_1 + \widehat{\beta}_{1x_2}X_2\right)$$

is a function of the independent variables (X_1, X_2) only. Then, we may choose one of the following three methods to obtain the ILS estimate for the unknown parameters in the second equation, Eq. (3.2):

1. For simplicity, we can just use the *ordinary least squares* (OLS) method to obtain

$$\widehat{\beta}_{2,ILS} = \left(\mathbf{A}_2^{+\mathrm{T}}\mathbf{A}_2^{+}\right)^{-1}\mathbf{A}_2^{+\mathrm{T}}\mathbf{Y}_2$$

where the design matrix is $\mathbf{A}_2^{+} = \left[\mathbf{1}, \mathbf{X}_1, \mathbf{X}_3, \widehat{\mathbf{Y}}_1\right]$.

2. Moreover, in contrast to the OLS estimator, we may consider the WLS estimator due to the heteroscedasticity of $\epsilon^*_{2,1}$. From Eq. (4.1), we know that

$$\begin{aligned}
\epsilon^*_{2,1} &= \gamma_{2y_1}\left(\frac{Z_1 - g_1(\widehat{\mu}_1)}{g_1'(\widehat{\mu}_1)}\right) + \epsilon_2 \\
&= \gamma_{2y_1}(Y_1 - \widehat{\mu}_1) + \epsilon_2,
\end{aligned}$$

and thus

$$Var(\epsilon^*_{2,1}) = \gamma^2_{2y_1}Var(Y_1 - \widehat{\mu}_1) + Var(\epsilon_2) + 2Cov\left(\gamma_{2y_1}(Y_1 - \widehat{\mu}_1), \epsilon_2\right).$$

Clearly, $Var(Y_1 - \widehat{\mu}_1)$ is *not* constant over all observations when Y_1 belongs to the binomial or Poisson distributions. Therefore, in regressing $\mathbf{Y}_2$ on $\mathbf{A}_2^{+} = \left[\mathbf{1}, \mathbf{X}_1, \mathbf{X}_3, \widehat{\mathbf{Y}}_1\right]$, we may use the *iterated feasible weighted least squares* (IFWLS) method, instead of the OLS method, to obtain the estimate $\widetilde{\beta}_{2,ILS}$.

3. Alternatively, to gain even more efficiency, we may also regress $[\mathbf{Z}_1, \mathbf{Y}_2]^{\mathrm{T}}$ on $\mathbf{A}_1$ and $\mathbf{A}_2^+$ jointly as in a *regression system* (RS) or *seemingly unrelated regressions* (SUR) model, i.e.,

$$\begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2^+ \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \epsilon_{1,1}^* \\ \epsilon_{2,1}^* \end{bmatrix}$$

using the *iterated feasible generalized least squares* (IFGLS) method for this "stacked" model to estimate $\boldsymbol{\beta}$, where $\mathbf{A}_1 = [\mathbf{1}, \mathbf{X}_1, \mathbf{X}_2]$, $\mathbf{A}_2^+ = \left[\mathbf{1}, \mathbf{X}_1, \mathbf{X}_3, \widehat{\mathbf{Y}}_1\right]$, and $\boldsymbol{\beta}^{\mathrm{T}} = \left(\boldsymbol{\beta}_1^{\mathrm{T}}, \boldsymbol{\beta}_2^{\mathrm{T}}\right) = (\beta_{10}, \beta_{1x_1}, \beta_{1x_2}, \beta_{20}, \beta_{2x_1}, \beta_{2x_3}, \gamma_{2y_1})$.

In Section 5, we shall prove the crucial condition that as the "proxy" of Y_1 in the equation for Y_2, $\widehat{Y}_1$ is indeed uncorrelated with the "combined" error term $\epsilon_{2,1}^*$.

4.2 The Two-Stage Least Squares (2SLS) Estimator

Next, we introduce the *two-stage least squares* (2SLS) method for estimating the structural coefficients in the GLM-LM GSiEM/GPA model (3.1) and (3.2).

- **Step 1:**

  Fit the GLM for the response Y_1 on *all* the independent variables including X_1, X_2, and X_3 using the IRLS algorithm to obtain the estimates $\widehat{\beta}_{10}^*$, $\widehat{\beta}_{1x_1}^*$, $\widehat{\beta}_{1x_2}^*$, and $\widehat{\beta}_{1x_3}^*$.

- **Step 2:**

  In fact, in **Step 1**, the estimates $\widehat{\beta}_{10}^*$, $\widehat{\beta}_{1x_1}^*$, $\widehat{\beta}_{1x_2}^*$, and $\widehat{\beta}_{1x_3}^*$ are obtained by fitting a multiple linear regression of the *pseudo*-response variable Z_1^* (defined below) on the original independent variables X_1,

X_2, and X_3 on the convergence:

$$\begin{aligned} Z_1^* &= \widehat{\beta}_{10}^* + \widehat{\beta}_{1x_1}^* X_1 + \widehat{\beta}_{1x_2}^* X_2 + \widehat{\beta}_{1x_3}^* X_3 + g_1'(\widehat{\mu}_1^*)(Y_1 - \widehat{\mu}_1^*) \\ &= \widehat{\beta}_{10}^* + \widehat{\beta}_{1x_1}^* X_1 + \widehat{\beta}_{1x_2}^* X_2 + \widehat{\beta}_{1x_3}^* X_3 + \epsilon_{1,2}^*. \end{aligned}$$

Thus, we can rewrite the original response variable Y_1 as a function of the pseudo-response variable Z_1^* in the following way:

$$\begin{aligned} Z_1^* &= \left(\widehat{\beta}_{10}^* + \widehat{\beta}_{1x_1}^* X_1 + \widehat{\beta}_{1x_2}^* X_2 + \widehat{\beta}_{1x_3}^* X_3\right) + g_1'(\widehat{\mu}_1^*)(Y_1 - \widehat{\mu}_1^*) \\ &= g_1(\widehat{\mu}_1^*) + g_1'(\widehat{\mu}_1^*)(Y_1 - \widehat{\mu}_1^*), \\ Z_1^* - g_1(\widehat{\mu}_1^*) &= g_1'(\widehat{\mu}_1^*)(Y_1 - \widehat{\mu}_1^*), \\ Y_1 - \widehat{\mu}_1^* &= \frac{Z_1^* - g_1(\widehat{\mu}_1^*)}{g_1'(\widehat{\mu}_1^*)}, \\ Y_1 &= \widehat{\mu}_1^* + \frac{Z_1^* - g_1(\widehat{\mu}_1^*)}{g_1'(\widehat{\mu}_1^*)} \end{aligned}$$

where

$$\widehat{\mu}_1^* = \widehat{Y}_1^* = g_1^{-1}\left(\widehat{\beta}_{10}^* + \widehat{\beta}_{1x_1}^* X_1 + \widehat{\beta}_{1x_2}^* X_2 + \widehat{\beta}_{1x_3}^* X_3\right).$$

In this step, we just compute $\widehat{\mu}_1^*$ by overfitting Y_1. These two steps are similar to **Steps 1** and **2** of the ILS estimator.

- **Step 3:**

  Notice that by plugging

$$Y_1 = \widehat{Y}_1^* + \frac{Z_1^* - g_1(\widehat{\mu}_1^*)}{g_1'(\widehat{\mu}_1^*)}$$

from **Step 2** into the second equation, Eq. (3.2), we have

$$\begin{aligned} Y_2 &= \beta_{20} + \beta_{2x_1} X_1 + \beta_{2x_3} X_3 + \gamma_{2y_1} Y_1 + \epsilon_2 \\ &= \beta_{20} + \beta_{2x_1} X_1 + \beta_{2x_3} X_3 + \gamma_{2y_1} \widehat{Y}_1^* + \left[\gamma_{2y_1}\left(\frac{Z_1^* - g_1(\widehat{\mu}_1^*)}{g_1'(\widehat{\mu}_1^*)}\right) + \epsilon_2\right] \\ &= \beta_{20} + \beta_{2x_1} X_1 + \beta_{2x_3} X_3 + \gamma_{2y_1} \widehat{Y}_1^* + \epsilon_{2,1}^{**} \end{aligned} \qquad (4.2)$$

where ϵ_2 is the error term of the second equation, Eq. (3.2), and as the "proxy" of Y_1,

$$\widehat{Y}_1^* = g_1^{-1}\left(\widehat{\beta}_{10}^* + \widehat{\beta}_{1x_1}^* X_1 + \widehat{\beta}_{1x_2}^* X_2 + \widehat{\beta}_{1x_3}^* X_3\right)$$

is a function of the independent variables (X_1, X_2, X_3) only. Then, for simplicity, we can just use the OLS method to obtain the 2SLS estimate for the unknown parameters in the second equation, Eq. (3.2):

$$\widehat{\boldsymbol{\beta}}_{2,2SLS} = \left(\mathbf{A}_2^{*\mathbf{T}}\mathbf{A}_2^*\right)^{-1}\mathbf{A}_2^{*\mathbf{T}}\mathbf{Y}_2$$

where the design matrix is $\mathbf{A}_2^* = \left[\mathbf{1}, \mathbf{X}_1, \mathbf{X}_3, \widehat{\mathbf{Y}}_1^*\right]$. Similarly, to gain some efficiency, we may also use the IFWLS method considered in the previous subsection to obtain the estimate $\widetilde{\boldsymbol{\beta}}_{2,2SLS}$.

Again, in Section 5, we shall prove the crucial condition that as the "proxy" of Y_1 in the equation for Y_2, $\widehat{Y}_1^*$ is indeed uncorrelated with the "combined" error term $\epsilon_{2,1}^{**}$.

5 JUSTIFICATION

In this section, we will show that as a "proxy" of Y_1, $\widehat{Y}_1$ is *uncorrelated* with the "combined" error term $\epsilon_{2,1}^*$ to justify our ILS estimator. Then, by the same token, we can justify the 2SLS estimator.

5.1 Proof

Specifically, we prove that $Cov\left(\widehat{Y}_1, \epsilon_{2,1}^*\right) = 0$ asymptotically in two steps.

Since

$$\begin{aligned}
\epsilon^*_{2,1} &= \gamma_{2y_1}\left(\frac{Z_1 - g_1(\widehat{\mu}_1)}{g_1'(\widehat{\mu}_1)}\right) + \epsilon_2 \qquad \text{from (4.1)} \\
&= \gamma_{2y_1}\left(\frac{\epsilon^*_{1,1}}{g_1'(\widehat{\mu}_1)}\right) + \epsilon_2 \\
&= \gamma_{2y_1}\left[\frac{g_1'(\widehat{\mu}_1)(Y_1 - \widehat{\mu}_1)}{g_1'(\widehat{\mu}_1)}\right] + \epsilon_2 \\
&= \gamma_{2y_1}(Y_1 - \widehat{\mu}_1) + \epsilon_2,
\end{aligned}$$

we have

$$Cov\left(\widehat{Y}_1, \epsilon^*_{2,1}\right) = \gamma_{2y_1} Cov\left(\widehat{Y}_1, Y_1 - \widehat{Y}_1\right) + Cov\left(\widehat{Y}_1, \epsilon_2\right).$$

Part 1: Proof of $Cov\left(\widehat{Y}_1, \epsilon_2\right) = 0$.

Notice that $\widehat{Y}_1$ (or, equivalently, $\widehat{\mu}_1$) is a function of (X_1, X_2). By the model assumption, $X_1 \perp \epsilon_2$ and $X_2 \perp \epsilon_2$, where the symbol "$\perp$" denotes "being independent of." Since any functions of independent variables are still independent to each other (see, e.g., Casella and Berger (1990, Theorem 4.3.2, p. 150)), it is clear that $\widehat{Y}_1 \perp \epsilon_2$, which implies $Cov\left(\widehat{Y}_1, \epsilon_2\right) = 0$.

Part 2: Proof of $Cov\left(\widehat{Y}_1, Y_1 - \widehat{Y}_1\right) = 0$ asymptotically.

According to the properties of MLEs, $\widehat{Y}_1$ (or, equivalently, $\widehat{\mu}_1$) is asymptotically unbiased for μ_1. Thus, when sample size n is large,

$$E\left(Y_1 - \widehat{Y}_1\right) = E\left[(Y_1 - \mu_1) - \left(\widehat{Y}_1 - \mu_1\right)\right] \cong 0.$$

Therefore, as $n \to \infty$,

$$\begin{aligned} Cov\left(\widehat{Y}_1, Y_1 - \widehat{Y}_1\right) &= E\left[\widehat{Y}_1\left(Y_1 - \widehat{Y}_1\right)\right] - E\left(\widehat{Y}_1\right)E\left(Y_1 - \widehat{Y}_1\right) \\ &= E\left[\widehat{Y}_1\left(Y_1 - \widehat{Y}_1\right)\right] \\ &= E\left\{E\left[\widehat{Y}_1\left(Y_1 - \widehat{Y}_1\right) \mid X_1, X_2\right]\right\} \\ &= E\left\{\widehat{Y}_1 E\left[\left(Y_1 - \widehat{Y}_1\right) \mid X_1, X_2\right]\right\} \\ &= E\left\{\widehat{Y}_1 \times 0\right\} \\ &= 0 \end{aligned}$$

by the *double expectation theorem*, which ends the proof.

To summarize, we wish to verify the crucial condition required in Subsection 4.1 that as a "proxy" of Y_1, $\widehat{Y}_1$ is *uncorrelated* with the "combined" error term $\epsilon^*_{2,1}$, where $\widehat{Y}_1$ is a function of (X_1, X_2) and $\epsilon^*_{2,1}$ is a function of $\epsilon^*_{1,1}$ and ϵ_2 respectively. By the model assumption, X_1 and X_2 are independent of ϵ_2. Hence, we only need to verify that $\widehat{Y}_1$ is uncorrelated with the error term $\epsilon^*_{1,1} = Z_1 - g_1(\widehat{\mu}_1) = g'_1(\widehat{\mu}_1)(Y_1 - \widehat{\mu}_1)$ by checking if $Cov\left(\widehat{Y}_1, Y_1 - \widehat{Y}_1\right) = 0$, which is true if the sample size is large.

5.2 Numerical Results

Next, we show that $Cor\left(\widehat{Y}_1, Y_1 - \widehat{Y}_1\right) = 0$ numerically in the following simulations, where $\widehat{Y}_1$ is defined in Subsection 4.1 for the ILS estimator.

1. Binomial Distribution:

The data are generated from the following logistic regression model

$$\text{logit}(\mu_1) = 3X_1 + X_2$$

where the distributions of the covariates X_1 and X_2 are independent Normal $(0,1)$ respectively. Given the sample size $n = 500$, averaging over 1000 repetitions yields $Cor\left(\widehat{Y}_1, Y_1 - \widehat{Y}_1\right) = 2.074 \times 10^{-5}$.

2. Poisson Distribution:

The data are generated from the following Poisson regression model

$$\log(\mu_1) = 1X_1 + X_2$$

where the distributions of the covariates X_1 and X_2 are independent Normal $(0,1)$ respectively. Given the sample size $n = 500$, averaging over 1000 repetitions yields $Cor\left(\widehat{Y}_1, Y_1 - \widehat{Y}_1\right) = -5.020 \times 10^{-7}$.

6 SIMULATIONS

In the following two simulation studies, the estimation of the structural coefficients in a partially recursive Binomial-Normal GSiEM/GPA model (with logit-identity links) and a partially recursive Poisson-Normal GSiEM/GPA model (with log-identity links) are examined with the comparisons among various estimators including two *instrumental variable* (IV) estimators, of which the technical details will be discussed in the accompany paper.

6.1 A Strategy for Data Generation

Following Hsiao (1986, Sec. 5.4, pp. 112-125), we can think that the data of the *partially* recursive two-equation GSiEM/GPA model (3.1) and (3.2) are generated from the following two equations:

$$\begin{aligned} g_1(\mu_1) &= \beta_{10} + \beta_{1x_1}X_1 + \beta_{1x_2}X_2 + h \\ \mu_2 &= \beta_{20} + \beta_{2x_1}X_1 + \beta_{2x_3}X_3 + \gamma_{2y_1}Y_1 + \alpha h \end{aligned}$$

where the *latent* variable h is generated independently from a common distribution such as

$$h \sim \text{Normal}\,(0,1)$$

for each subject i so that it is independent of all the independent variables X_1, X_2, and X_3. After the data are generated, the latent variable h is unknown to the data analyst. The chosen value of the *latent* coefficient α on the latent variable h in the second equation controls the degree of correlatedness between the random errors of Eqs. (3.1) and (3.2). Yet, the other terms in the above two equations remain the same as in the specification of the original partially recursive GSiEM/GPA model (3.1) and (3.2). By adding an extra "unobserved" latent variable h to the equations of a partially recursive GSiEM/GPA model, we find a feasible way to generating the data for simulations.

6.2 A Partially Recursive *Binomial*-Normal GSiEM/GPA Model

For simplicity, we specify the following partially recursive two-equation Binomial-Normal GSiEM/GPA model

$$\begin{aligned} \text{logit}(\mu_1) &= \beta_{10} + \beta_{11}X_1 \\ \mu_2 &= \beta_{20} + \beta_{22}X_2 + \gamma_{21}Y_1 \end{aligned}$$

where the two response variables are

$$\begin{aligned} Y_1 &\sim \text{Binomial}\,(1, \mu_1), \\ Y_2 &\sim \text{Normal}\,(\mu_2, 1). \end{aligned}$$

And, the data are actually generated from the following two equations

$$\text{logit}(\mu_1) = 3X_1 + h, \tag{6.1}$$

$$\mu_2 = 2X_2 - 2Y_1 + \alpha h \tag{6.2}$$

where the independent variables X_1 and X_2, the "unobserved" latent variable h, and the error term e_2 of the second equation for Y_2 are generated independently from Normal $(0, 1)$. The true values of the coefficients are all explicitly listed in the above two equations. Then, the equation-by-equation OLS, ILS, 2SLS, IV-1, and IV-2 estimators for estimating the coefficients in the second equation are computed in the following simulations for comparison. See **Table 1** for the list of their formulas. Note that the IV-1 estimator uses the predicted value of Y_1 based on X_1 only as the instrument for Y_1, but the IV-2 estimator uses the predicted value of Y_1 based on X_1 and X_2 together as the instrument for Y_1.

6.2.1 Comparison 1: Performances of the Equation-by-Equation OLS and ILS Estimators

The value of the latent coefficient α is set to 2.0 and 5.0 respectively for varying the degree of the association between the random components of the two equations in the partially recursive GSiEM/GPA model. The sample sizes n are 100, 500, and 1000 in three separate simulations. And, 1000 repetitions are performed in each setting. For each data set, the coefficient β_{11} in Eq. (6.1) is estimated directly by the IRLS algorithm of GLMs. Since the latent variable h is assumed to be unknown to us, it can not be included in the fitted models. Thus, even though h is independent of X_1, the "correct" value of the coefficient β_{11} in the first equation does not equal 3.0 due to the underfitting of the logistic regression model, which is the so-called *omitted-variable bias* in GLMs as discussed by Breslow and Day (1980), Lee (1982), Gail, Wieand, and Piantadosi (1984), and Neuhaus and Jewell (1993).

The comparison between the equation-by-equation OLS estimate and our ILS estimate of the structural coefficient γ_{21} on the covariate Y_1 in Eq. (6.2), which has the true value -2.0, is our main interest. The means and standard deviations (SDs) of these two estimates from the 1000 repetitions are listed in **Table 2**. As we expected, the equation-by-equation OLS estimate of γ_{21} is seriously *biased* even though the sample size n is large. And, the asymptotic bias gets bigger as the association between the random components of these two equations gets higher (e.g., α = 5.0). In contrast, our ILS estimates of γ_{21} are all close to the true parameter value -2.0 without being interfered by the association between the random components of these two equations. According to the histograms and quantile-normal plots (not shown here), the sampling distribution of our ILS estimate of γ_{21} in Eq. (6.2) from the 1000 repetitions is approximately normal.

6.2.2 Comparison 2: Performances of the Equation-by-Equation OLS, ILS, 2SLS, IV-1, and IV-2 Estimators

The value of α is set to 2.0 and the sample size n is 500 in this simulation. Again, 1000 repetitions are performed in this setting. For each data set, the coefficients β_{10} and β_{11} in Eq. (6.1) are estimated directly by the IRLS algorithm of GLMs. Omitting the latent variable h leads to the unbiased MLE of β_{10} and the biased MLE of β_{11}, even though h is independent of X_1 (see, esp., Gail, Wieand, and Piantadosi (1984)). Adding the additional covariate X_2 into the first equation causes very little change on the MLEs of β_{10} and β_{11} due to the independence between Y_1 and X_2.

The comparisons among the equation-by-equation OLS, ILS, 2SLS, IV-1, and IV-2 estimates of the structural coefficient γ_{21} on the covariate Y_1 in Eq. (6.2), which has the true value -2.0, is our main interest. The means and standard deviations (SDs) of these estimates from the 1000 repetitions are listed in **Table 3**. In addition, the mean of the standard error (SE) of $\widehat{\gamma}_{21}$ from the 1000 repetitions, which is estimated naively by the usual OLS formula (ignoring the fact that the proxy $\widehat{Y}_1$ is actually estimated from the data) for the ILS and 2SLS estimators and using the correct formula derived in the accompany paper for the IV-1 and IV-2 estimators, is compared with the standard deviation (SD) of the sampling distribution of $\widehat{\gamma}_{21}$. Again, we can see that the equation-by-equation OLS estimate of γ_{21} is seriously *biased*. However, the ILS, 2SLS, IV-1, and IV-2 estimates of γ_{21} are very close to each other and they all are *asymptotically unbiased*. It has been known in econometrics that when the equation of Y_1 is a linear regression model, the ILS, 2SLS, and IV-2 estimates are *exactly* the same. The difference between the mean of the estimated standard error (SE) of $\widehat{\gamma}_{21}$ from the 1000 repetitions and the corresponding standard deviation (SD) of the sampling distribution of $\widehat{\gamma}_{21}$ is relatively small in all these five estimates. Finally, according to the histograms and quantile-normal plots (see **Figure 1**), the sampling distributions of the five different estimates of γ_{21} from the 1000 repetitions are all approximately normal, but the center of the sampling distribution for the equation-by-equation OLS estimate of γ_{21} is clearly away from the true value -2.0. The similar results have also been seen when X_1 is a binary covariate.

6.3 A Partially Recursive *Poisson*-Normal GSiEM/GPA Model

For simplicity, we specify the following partially recursive two-equation Poisson-Normal GSiEM/GPA model

$$\begin{aligned}\log(\mu_1) &= \beta_{10} + \beta_{11} X_1 \\ \mu_2 &= \beta_{20} + \beta_{22} X_2 + \gamma_{21} Y_1\end{aligned}$$

where the two response variables are

$$\begin{aligned}Y_1 &\sim \text{Poisson}\,(\mu_1), \\ Y_2 &\sim \text{Normal}\,(\mu_2, 1).\end{aligned}$$

And, the data are actually generated from the following two equations

$$\log(\mu_1) = 1X_1 + h, \tag{6.3}$$

$$\mu_2 = 10X_2 + 0.5Y_1 + \alpha h \tag{6.4}$$

where the independent variables X_1 and X_2, the "unobserved" latent variable h, and the error term e_2 of the second equation for Y_2 are generated independently from Normal $(0, 1)$. The true values of the coefficients are all explicitly listed in the above two equations. Then, the equation-by-equation OLS, ILS, 2SLS, IV-1, and IV-2 estimators for estimating the coefficients in the second equation are computed in the following simulations for comparison. See **Table 1** for the list of their formulas. Again, note that the IV-1 estimator uses the predicted value of Y_1 based on X_1 only as the instrument for Y_1, but the IV-2 estimator uses the predicted value of Y_1 based on X_1 and X_2 together as the instrument for Y_1.

6.3.1 Comparison 1: Performances of the Equation-by-Equation OLS and ILS Estimators

The value of the latent coefficient α is set to 0.5 and 2.0 respectively for varying the degree of the association between the random components of the two equations in the partially recursive GSiEM/GPA model. The sample sizes n are 100, 500, and 1000 in three separate simulations. And, 1000 repetitions are performed in each setting. For each data set, the coefficient β_{11} in Eq. (6.3) is estimated directly by the IRLS algorithm of GLMs. Since the latent variable h is assumed to be unknown to us, it can not be included in the fitted models. However, although h is omitted from the fitted first equation, the "correct" value of the coefficient β_{11} still equals 1.0 without the *omitted-variable bias* due to the *log* link function of the Poisson regression model as discussed by Gail, Wieand, and Piantadosi (1984) and Neuhaus and Jewell (1993).

The comparison between the equation-by-equation OLS estimate and our ILS estimate of the structural coefficient γ_{21} on the covariate Y_1 in Eq. (6.4), which has the true value 0.5, is our main interest. The means and standard deviations (SDs) of these two estimates from the 1000 repetitions are listed in **Table 4**. Again, as we expected, the equation-by-equation OLS estimate of γ_{21} is seriously *biased* even though the sample size n is large. And, the asymptotic bias gets bigger as the association between the random components of these two equations gets higher (e.g., α = 2.0). In contrast, our ILS estimates of γ_{21} are all relatively close to the true parameter value 0.5 without being interfered by the association between the random components of these two equations. Yet, according to the histograms and quantile-normal plots (not shown here), the sampling distribution of our ILS estimate of γ_{21} in Eq. (6.4) from the 1000 repetitions is a little skewed to the right, which deserves a further investigation. Finally, we note that unlike the results listed in **Table 2**, the standard deviations (SDs) of these two estimates from the 1000 repetitions do not necessarily decrease as the sample size n increases. This odd phenomenon is also seen in the comprehensive simulation study presented in the accompany paper (see **Table 4** there).

6.3.2 Comparison 2: Performances of the Equation-by-Equation OLS, ILS, 2SLS, IV-1, and IV-2 Estimators

The value of α is set to 2.0 and the sample size n is 500 in this simulation. Again, 1000 repetitions are performed in this setting. For each data set, the coefficients β_{10} and β_{11} in Eq. (6.3) are estimated directly by the IRLS algorithm of GLMs. Omitting the latent variable h leads to the biased MLE of β_{10} and the unbiased MLE of β_{11}, since Y_1 has a Poisson distribution and the link function is *log* (see, esp., Gail, Wieand, and Piantadosi (1984)). Adding the additional covariate X_2 into the first equation causes very little change on the MLEs of β_{10} and β_{11} due to the independence between Y_1 and X_2.

The comparisons among the equation-by-equation OLS, ILS, 2SLS, IV-1, and IV-2 estimates of the structural coefficient γ_{21} on the covariate Y_1 in Eq. (6.4), which has the true value 0.5, is our main interest. The means and standard deviations (SDs) of these estimates from the 1000 repetitions are listed in **Table 5**. In addition, the mean of the standard error (SE) of $\widehat{\gamma}_{21}$ from the 1000 repetitions, which is estimated naively by the usual OLS formula (ignoring the fact that the proxy $\widehat{Y}_1$ is actually estimated from the data) for the ILS and 2SLS estimators and using the correct formula derived in the accompany paper for the IV-1 and IV-2 estimators, is compared with the standard deviation (SD) of the sampling distribution of $\widehat{\gamma}_{21}$. Again, we can see that the equation-by-equation OLS estimate of γ_{21} is seriously *biased*. However, the ILS, 2SLS, IV-1, and IV-2 estimates of γ_{21} are relatively close to each other and they all, especially the latter two, seem to be *asymptotically unbiased*. As mentioned before, it has been known in econometrics that when the equation of Y_1 is a linear regression model, the ILS, 2SLS, and IV-2 estimates are *exactly* the same. On the other hand, notice that the ILS and 2SLS estimates are much less efficient than the IV-1 and IV-2 estimates in the estimation of γ_{21} by examining the standard deviations (SDs) of their sampling distributions. And, the difference between the mean of the estimated standard error (SE) of $\widehat{\gamma}_{21}$ from the 1000 repetitions and the corresponding standard deviation (SD) of the sampling distribution of $\widehat{\gamma}_{21}$ is small only in the IV-1 and IV-2 estimates, which indicates that the formula for the estimates of their standard errors is correct. Finally, according to the histograms and quantile-normal plots (see **Figure 2**), the sampling distribution of the equation-by-equation OLS estimate of γ_{21} from the 1000 repetitions is approximately normal, but its center is clearly away from the true value 0.5. In contrast, the sampling distributions of the ILS and 2SLS estimates of γ_{21} are a little skewed to the right, but the sampling distributions of the IV-1 and IV-2 estimates of γ_{21} are a little skewed to the left, which also deserve a further investigation.

7 DISCUSSION

7.1 Summary

In this study, we try to combine the estimation methods of SiEM and the IRLS algorithm of GLMs to develop suitable estimation methods for estimating the structural coefficients in a partially recursive GSiEM/GPA model especially with responses of a mixed type. Specifically, we have developed the ILS and 2SLS estimators in this paper for a partially recursive two-equation GSiEM/GPA model, in which the first equation, Eq. (3.1), is a GLM and the second equation, Eq. (3.2), is a linear regression model. However, it is a straightforward task to apply the ILS and 2SLS estimators to (1) a partially recursive two-equation GSiEM/GPA model in which both equations are GLMs by replacing Y_1 as a covariate in the second equation with its proxy and (2) a partially recursive multi-equation GSiEM/GPA model by solving its equations either recursively or jointly. Moreover, with the aid of the IRLS algorithm of GLMs, we have obtained from the development of the ILS method a direct way to deriving the *reduced form* of a GSiEM/GPA model as a by-product, although it is usually difficult to obtain the reduced form of a nonlinear SiEM/PA model (Davidson and MacKinnon 1993, p. 662). Finally, we remark that up to now, we have not seen the need of adding any new or extra constraints to the current available rules for the model identification. See, for example, Greene (2000, Sec. 16.3, pp. 663-676) for details.

7.2 Future Work

Nonetheless, the ILS and 2SLS estimators for the partial recursive GSiEM/GPA models suffer two drawbacks. First, it is difficult to estimate their asymptotic variances. Second, the above simulation studies give us two different results: (1) When the distribution of the first response Y_1 is *binomial*, the performances of the three estimators are almost the same. (2) Yet, when the distribution of the first response Y_1 is *Poisson*, these two estimators are asymptotically less efficient than the IV estimator. We shall develop the IV estimator for the partially recursive GSiEM/GPA models in the accompany paper.

8 APPENDIX: THE ITERATIVELY REWEIGHTED LEAST SQUARES (IRLS) ALGORITHM FOR GLMS

Let i index observations, $i = 1, 2, \cdots, n$ and j index parameters, $j = 1, 2, \cdots, p$. As defined in the review section, a GLM for the *i*th observation is

$$g(\mu_i) = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}$$

where μ_i is the mean of the response variable Y_i from the *exponential family of distributions* (of the same form for all i), $\mathbf{x}_i$ is a vector of covariates, $\boldsymbol{\beta}$ is a vector of regression coefficients, and $g(\cdot)$ is a monotonic and differentiable function called the *link function* (to link μ_i with the linear combination of $\boldsymbol{\beta}$ and $\mathbf{x}_i$). The MLEs for the unknown regression coefficients $\boldsymbol{\beta}$ in the above GLM can be obtained by applying the following unified algorithm (see, e.g., Dobson (1990, Sec. 4.4, pp. 39-42) and McCullagh and Nelder (1989, Sec. 2.5, pp. 40-43)).

Recall that for a random variable Y having a probability density function (or probability mass function) $f(y; \theta)$ with a single parameter θ, the *log-likelihood function* is defined as

$$l(\theta; y) \equiv \log f(y; \theta).$$

Then, the *score function* for the parameter θ is

$$U(\theta) \equiv \frac{\partial l}{\partial \theta}.$$

It can be shown that the mean of U is

$$E(U) = 0$$

and the variance of U, called the *information*, is

$$Var(U) \equiv E(U^2) = -\,E(U')$$

where

$$U' \equiv \frac{\partial U}{\partial \theta}$$

(see, e.g., Dobson (1990, Appendix A, pp. 142-144)). The same results can be generalized to a set of independent random variables $Y_1, Y_2, \cdots, Y_n$ from the distributions of the same form with parameters $\theta_1, \theta_2, \cdots, \theta_p$, where $p \leq n$.

Now, given a random sample of size n, let the *joint* log-likelihood function for the independent response variables $Y_1, Y_2, \cdots, Y_n$ from the (one-parameter) linear exponential family of distributions of the same form with a single parameter θ_i (which may have a different value for each observation) be written, in the canonical form, as

$$l\,(\theta_1, \theta_2, \cdots, \theta_n; y_1, y_2, \cdots, y_n) = \sum_{i=1}^{n} l_i = \sum_{i=1}^{n} y_i\, b(\theta_i) + \sum_{i=1}^{n} c(\theta_i) + \sum_{i=1}^{n} d(y_i).$$

Based on the above results for $E(U)$ and $Var(U)$, it can be shown that

$$\mu_i \equiv E(Y_i) = -\frac{c'(\theta_i)}{b'(\theta_i)}$$

and

$$Var(Y_i) = \frac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{[b'(\theta_i)]^3}$$

where the superscript "$\prime$" denotes the derivative with respect to θ (see, e.g., Dobson (1990, pp. 28-30)). If $b(\theta_i) = \theta_i$, then the formulas for $E(Y_i)$ and $Var(Y_i)$ reduce to $-c'(\theta_i)$ and $-c''(\theta_i)$ respectively. And, we define the *linear predictor* η_i as the linear combination of $\mathbf{x}_i$ and $\boldsymbol{\beta}$, i.e.,

$$g(\mu_i) = \eta_i \equiv \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}.$$

Then, by the chain rule, it can be shown that the jth *score equation* for the regression coefficient β_j is just

$$U_j \equiv \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \theta_i} \left(\frac{\partial \mu_i}{\partial \theta_i} \right)^{-1} \left(\frac{\partial \eta_i}{\partial \mu_i} \right)^{-1} \frac{\partial \eta_i}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{Var(Y_i)} \left[g'(\mu_i) \right]^{-1} x_{ij} = 0$$

and the elements of the *information matrix* $\boldsymbol{I}$ are

$$I_{jk} \equiv E[U_j U_k] \equiv E\left[\frac{\partial l}{\partial \beta_j} \frac{\partial l}{\partial \beta_k} \right] = \sum_{i=1}^{n} \frac{x_{ij} x_{ik}}{Var(Y_i)} \left\{ \left[g'(\mu_i) \right]^{-1} \right\}^2$$

where

$$g'(\mu_i) \equiv \frac{\partial g(\mu_i)}{\partial \mu_i} = \frac{\partial \eta_i}{\partial \mu_i}$$

(see, e.g., Dobson (1990, Appendix B, pp. 145-146)). Thus, putting things together yields the total *score equation*

$$\boldsymbol{U} \equiv \frac{\partial l}{\partial \boldsymbol{\beta}} = \boldsymbol{D}^{\mathrm{T}} \boldsymbol{V}(\boldsymbol{\beta})^{-1} (\boldsymbol{Y} - \boldsymbol{\mu}(\boldsymbol{\beta})) = 0$$

and the *information matrix*

$$\boldsymbol{I} \equiv Var(\boldsymbol{U}) = \boldsymbol{D}^{\mathrm{T}} \boldsymbol{V}(\boldsymbol{\beta})^{-1} \boldsymbol{D}$$

where $\boldsymbol{V}(\boldsymbol{\beta})$ is the variance-covariance matrix of $Y_1, Y_2, \cdots, Y_n$ and $\boldsymbol{D}$, called the *derivative matrix*, is

$$\boldsymbol{D} \equiv \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} = \begin{bmatrix} \ddots & & 0 \\ & \left[g'(\mu_i) \right]^{-1} & \\ 0 & & \ddots \end{bmatrix} \boldsymbol{X}.$$

Notice that both the means $\boldsymbol{\mu}(\boldsymbol{\beta})$ and the variance-covariance matrix $\boldsymbol{V}(\boldsymbol{\beta})$ of $Y_1, Y_2, \cdots, Y_n$ in general depend on $\boldsymbol{\beta}$.

If the *Newton-Raphson method* is used to obtain the MLEs $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, then at the mth iteration,

$$\widehat{\boldsymbol{\beta}}^{(m)} = \widehat{\boldsymbol{\beta}}^{(m-1)} - \left[\frac{\partial^2 l}{\partial\beta_j \partial\beta_k}\right]^{-1}_{\widehat{\boldsymbol{\beta}}^{(m-1)}} \boldsymbol{U}^{(m-1)}. \tag{8.1}$$

Alternatively, the *Fisher's scoring method*, which replaces the matrix of the second derivatives (called the *Hessian matrix*) in the above equation (Eq. (8.1)) by its expected value, can be used to obtain the MLEs $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$. Then, at the mth iteration,

$$\widehat{\boldsymbol{\beta}}^{(m)} = \widehat{\boldsymbol{\beta}}^{(m-1)} + \left[\boldsymbol{I}^{(m-1)}\right]^{-1} \boldsymbol{U}^{(m-1)} \tag{8.2}$$

due to the *information equality*

$$E\left[\frac{\partial l}{\partial\beta_j}\frac{\partial l}{\partial\beta_k}\right] = -\,E\left[\frac{\partial^2 l}{\partial\beta_j \partial\beta_k}\right]$$

(see, e.g., McCullagh and Nelder (1989, p. 42)). Premultiplying both sides of Eq. (8.2) by $\boldsymbol{I}^{(m-1)}$ yields

$$\boldsymbol{I}^{(m-1)}\widehat{\boldsymbol{\beta}}^{(m)} = \boldsymbol{I}^{(m-1)}\widehat{\boldsymbol{\beta}}^{(m-1)} + \boldsymbol{U}^{(m-1)} \tag{8.3}$$

which can be converted into a *normal equation* of *weighted least squares* (WLS).

First, we can rewrite the information matrix $\boldsymbol{I}$ as a *weighted* sums-of-squares-and-products matrix of the covariates

$$\boldsymbol{I} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}$$

where the weights $\boldsymbol{W}$ is an $n \times n$ diagonal matrix with elements

$$W_{ii} \equiv \sum_{i=1}^{n} \frac{1}{Var(Y_i)} \left\{\left[g'(\mu_i)\right]^{-1}\right\}^2.$$

Then, the left-hand side of Eq. (8.3) can be rewritten as

$$\boldsymbol{I}^{(m-1)}\widehat{\boldsymbol{\beta}}^{(m)} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}^{(m-1)}\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{(m)}.$$

Next, we define a *pseudo*-response variable Z_i by making the following transformation on the original response variable Y_i (for each i)

$$Z_i \equiv \sum_j x_{ij}\widehat{\beta}_j^{(m-1)} + g'(\mu_i)\,(Y_i - \mu_i)$$

where μ_i and $g'(\mu_i)$ are evaluated at $\widehat{\boldsymbol{\beta}}^{(m-1)}$. As noted by McCullagh and Nelder (1989, p. 40), Z_i is just a *linearized* form of the link function applied to Y_i to the first order

$$g(Y_i) \cong g(\mu_i) + g'(\mu_i)\,(Y_i - \mu_i)\,.$$

Then, the right-hand side of Eq. (8.3), which is a $p \times 1$ vector with elements

$$\sum_k \sum_i \frac{x_{ij}x_{ik}}{Var(Y_i)}\left\{\left[g'(\mu_i)\right]^{-1}\right\}^2 \widehat{\beta}_k^{(m-1)} + \sum_i \frac{(y_i - \mu_i)}{Var(Y_i)}\left[g'(\mu_i)\right]^{-1} x_{ij},$$

can be rewritten as

$$\boldsymbol{I}^{(m-1)}\widehat{\boldsymbol{\beta}}^{(m-1)} + \boldsymbol{U}^{(m-1)} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}^{(m-1)}\boldsymbol{Z}^{(m-1)}.$$

Finally, assuming that $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ are known and $\boldsymbol{X}$ is fixed, it is straightforward to show

$$E(\boldsymbol{Z}) = \boldsymbol{X}\boldsymbol{\beta}$$

$$Var(\boldsymbol{Z}) = \boldsymbol{W}^{-1}$$

by direct calculations. Hence, the iterative equation (Eq. (8.3)) for the method of scoring can be rewritten as a normal equation

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}^{(m-1)}\boldsymbol{X}\widehat{\boldsymbol{\beta}}^{(m)} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}^{(m-1)}\boldsymbol{Z}^{(m-1)} \tag{8.4}$$

for obtaining the WLS estimates of the regression coefficients $\boldsymbol{\beta}$ in the derived linear regression for the pseudo-response variable $\boldsymbol{Z}^{(m-1)}$ with weights $\boldsymbol{W}^{(m-1)}$. Yet, since $\boldsymbol{Z}^{(m-1)}$ and $\boldsymbol{W}^{(m-1)}$ depend on $\widehat{\boldsymbol{\beta}}^{(m-1)}$, Eq. (8.4) must also be solved *iteratively* until $\widehat{\boldsymbol{\beta}}^{(m)}$ converges. Thus, this is called the *iterative reweighted least squares* (IRLS) algorithm for obtaining the MLEs of the regression coefficients $\boldsymbol{\beta}$ in a GLM.

9 REFERENCES

1. Amemiya, T. (1985), *Advanced Econometrics*, Cambridge, MA: Harvard University Press.

2. Arminger, G. and Küsters, U. (1985), "Simultaneous Equation Systems with Categorical Observed Variables," In: R. Gilchrist, B. Francis, and J. Whittaker (eds.) *Generalized Linear Models: Proceedings of the GLIM 85 Conference, Held in Lancaster, UK, September 16-19, 1985*, Berlin: Springer-Verlag, pp. 15-26.

3. Bentler, P. M. and Newcomb, M. D. (1991), "Chapter 6: Linear Structure Equation Modelling with Non-Normal Continuous Variables," In: J. H. Dwyer, M. Feinleib, P. Lippert, and H. Hoffmeister (eds.) *Statistical Models for Longitudinal Studies of Health*, New York: Oxford University Press, pp. 132-160.

4. Bollen, K. A. (1989), *Structural Equations with Latent Variables*, New York: John Wiley & Sons.

5. Bowden, R. J. and Turkington, D. A. (1984), *Instrumental Variables*, Cambridge: Cambridge University Press.

6. Breslow, N. E. and Day, N. E. (1980), *Statistical Methods in Cancer Research, Volume 1 – The Analysis of Case-Control Studies*, Lyon: International Agency for Research on Cancer.

7. Casella, G. and Berger, R. L. (1990), *Statistical Inference*, Belmont, CA: Duxbury Press.

8. Davidson, R. and Mackinnon, J. G. (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press.

9. Dobson, A. J. (1990), *An introduction to Generalized Linear Models*, London: Chapman and Hall.

10. Fahrmeir, L. and Tutz, G. (1994), *Multivariate Statistical Modelling Based on Generalized Linear Models*, New York: Springer-Verlag.

11. Finkel, S. E. (1995), *Causal Analysis with Panel Data*, (Sage University Paper series on Quantitative Applications in the Social Sciences, 07-105), Thousand Oaks, CA: Sage.

12. Gail, M. H., Wieand, S., and Piantadosi, S. (1984), "Biased Estimates of Treatment Effect in Randomized Experiments with Nonlinear Regressions and Omitted Covariates," *Biometrika*, 71, 431-444.

13. Goldfeld, S. M. and Quandt, R. E. (1968), "Nonlinear Simultaneous Equations Estimation and Prediction," *International Economic Review*, 9, 113-136.

14. Greene, W. H. (2000), *Econometric Analysis* (4th ed.), Upper Saddle River, NJ: Prentice-Hall.

15. Hagenaars, J. A. (1993), *Loglinear Models with Latent Variables*, (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-094), Newbury Park, CA: Sage.

16. Hellevik, O. (1988), *Introduction to Causal Analysis: Exploring Survey Data by Crosstabulation*, Oslo: Norwegian University press.

17. Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge: Cambridge University Press.

18. Jöreskog, K. G. and Yang, F. (1996), "Chapter 3: Nonlinear Structural Equation Models: The Kenny-Judd Model with Interaction Effects," In: G. A. Marcoulides and R. E. Schumacker (eds.) *Advanced Structural Equation Modeling: Issues and Techniques*, Mahwah, NJ: Lawrence Erlbaum Associates, pp. 57-88.

19. Kelejian, H. H. (1971), "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables," *Journal of American Statistical Association*, 66, 373-374.

20. Kenny, D. A. and Judd, C. M. (1984), "Estimating the Nonlinear and Interactive Effects of Latent Variables," *Psychological Bulletin*, 96, 201-210.

21. Lee, L. -F. (1981), "Chapter 9: Simultaneous Equations Models with Discrete and Censored Dependent Variables," In: C. F. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA: The MIT Press, pp. 346-364.

22. Lee, L. -F. (1982), "Specification Error in Multinomial Logit Models: Analysis of the Omitted Variable Bias," *Journal of Econometrics*, 20, 197-209.

23. Maddala, G. S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press.

24. Manski, C. F. and McFadden, D. (1981), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA: The MIT Press.

25. McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman and Hall.

26. Muthén, B. (1984), "A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators," *Psychometrika*, 49, 115-132.

27. Nelder, J. A. and Wedderburn, R. W. M. (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society*, Ser. A, 135, 370-384.

28. Nerlove, M. and Press, S. J. (1973), *Univariate and Multivariate Log-Linear and Logistic Models*, (Report R-1306), Monica, CA: RAND Corporation.

29. Neuhaus, J. M. and Jewell, N. P. (1993), "A Geometric Approach to Assess Bias Due to Omitted Covariates in Generalized Linear Models," *Biometrika*, 80, 807-815.

30. Schmidt, P. and Strauss, R. P. (1975), "Estimation of Models with Jointly Dependent Qualitative Variables: A Simultaneous Logit Approach," *Econometrica*, 43, 745-755.

31. Wu, S. -C. (1995), Personal communication.

32. Zellner, A., Huang, D. S., and Chau, L. C. (1965), "Further Analysis of the Short-Run Consumption Function with Emphasis on the Role of Liquid Assets," *Econometrica*, 33, 571-581.

10 TABLES

Table 1: The OLS, ILS, 2SLS, IV-1, and IV-2 Estimators Used in the Simulation Study.

	Estimator	Proxy of $\mathbf{Y}_1$: $\widehat{\mathbf{Y}}_1$ or $\widehat{\mathbf{Y}}_1^*$?	Formula for $\widehat{\beta}_2$
1	**OLS**	$\mathbf{Y}_1$: Use itself.	$(\mathbf{A}_2^T\ \mathbf{A}_2)^{-1}\ \mathbf{A}_2^T\ \mathbf{Y}_2$
2	**ILS**	$\widehat{\mathbf{Y}}_1$: Use the X's in Eqs. (6.1) or (6.3) only.	$(\mathbf{A}_2^{+T}\mathbf{A}_2^{+})^{-1}\ \mathbf{A}_2^{+T}\mathbf{Y}_2$
3	**2SLS**	$\widehat{\mathbf{Y}}_1^*$: Use all X's.	$(\mathbf{A}_2^{*T}\ \mathbf{A}_2^*)^{-1}\ \mathbf{A}_2^{*T}\ \mathbf{Y}_2$
4	**IV-1**	$\widehat{\mathbf{Y}}_1$: Use the X's in Eqs. (6.1) or (6.3) only.	$(\mathbf{A}_2^{+T}\mathbf{A}_2)^{-1}\ \mathbf{A}_2^{+T}\mathbf{Y}_2$
5	**IV-2**	$\widehat{\mathbf{Y}}_1^*$: Use all X's.	$(\mathbf{A}_2^{*T}\ \mathbf{A}_2)^{-1}\ \mathbf{A}_2^{*T}\ \mathbf{Y}_2$

Notation

(1) $\mathbf{A}_2 = \left[\mathbf{1}, \mathbf{X}_2, \mathbf{Y}_1\right]$.

(2) $\mathbf{A}_2^+ = \left[\mathbf{1}, \mathbf{X}_2, \widehat{\mathbf{Y}}_1\right]$.

(3) $\mathbf{A}_2^* = \left[\mathbf{1}, \mathbf{X}_2, \widehat{\mathbf{Y}}_1^*\right]$.

Table 2: Means (Row 1) and SDs (Row 2) of the OLS and ILS Estimates from 1000 Repetitions for the Partially Recursive Binomial-Normal GSiEM/GPA Model.

n	MLE	OLS		ILS		Asymptotic Bias of OLS	Asymptotic Bias of ILS
	β_{11} (3.0)	β_{22} (2.0)	γ_{21} (-2.0)	β_{22} (2.0)	γ_{21} (-2.0)		
$\alpha = 2.0$							
100							
	2.670	2.000	**-2.440**	2.000	**-1.990**	-0.440	0.010
	0.576	0.221	0.302	0.249	0.387	0.302	0.387
500							
	2.590	2.000	**-2.440**	2.000	**-2.000**	-0.440	0.000
	0.237	0.098	0.134	0.112	0.178	0.134	0.178
1000							
	2.590	2.000	**-2.440**	2.000	**-2.000**	-0.440	0.000
	0.165	0.069	0.101	0.079	0.127	0.101	0.127
$\alpha = 5.0$							
100							
	2.720	2.020	**-3.130**	2.020	**-2.050**	-1.130	-0.050
	0.555	0.505	0.738	0.538	0.885	0.738	0.885
500							
	2.610	2.000	**-3.110**	2.000	**-2.010**	-1.110	-0.010
	0.232	0.228	0.313	0.242	0.391	0.313	0.391
1000							
	2.590	2.000	**-3.110**	1.990	**-2.010**	-1.110	-0.010
	0.169	0.157	0.231	0.169	0.279	0.231	0.279

Table 3: Means (Row 1) and SDs (Row 2) of the OLS, ILS, 2SLS, IV-1, and IV-2 Estimates from 1000 Repetitions for the Partially Recursive Binomial-Normal GSiEM/GPA Model at α = 2.0 with n = 500.

	MLE (1)		MLE (2)			OLS			
	β_{10} (0)	β_{11} (3.0)	β_{10} (0)	β_{11} (3.0)	β_{12} (0)	β_{20} (0)	γ_{21} (-2.0)	β_{22} (2.0)	$\widehat{\gamma}_{21}$ SE
Mean	-0.0067	2.6092	-0.0067	2.6172	-0.0015	-0.4451	**-1.1130**	2.0017	**0.1966**
SD	0.1304	0.2453	0.1312	0.2471	0.1286	0.1373	**0.1946**	0.0979	0.0063

	ILS				2SLS			
	β_{20} (0)	γ_{21} (-2.0)	β_{22} (2.0)	$\widehat{\gamma}_{21}$ SE	β_{20} (0)	γ_{21} (-2.0)	β_{22} (2.0)	$\widehat{\gamma}_{21}$ SE
Mean	-0.0057	**-1.9935**	2.0017	**0.2768**	-0.0066	**-1.9917**	2.0013	**0.2766**
SD	0.1705	**0.2809**	0.0957	0.0135	0.1705	**0.2808**	0.1004	0.0135

	IV-1				IV-2			
	β_{20} (0)	γ_{21} (-2.0)	β_{22} (2.0)	$\widehat{\gamma}_{21}$ SE	β_{20} (0)	γ_{21} (-2.0)	β_{22} (2.0)	$\widehat{\gamma}_{21}$ SE
Mean	-0.0037	**-1.9976**	2.0013	**0.2886**	-0.0045	**-1.9959**	2.0013	**0.2884**
SD	0.1707	**0.2819**	0.1004	0.0160	0.1706	**0.2815**	0.1004	0.0160

1. **SD**: The *standard deviation* of the estimates from the 1000 repetitions.
2. **SE**: The *standard error* of $\widehat{\gamma}_{21}$ estimated naively by the usual OLS formula (ignoring the fact that the proxy $\widehat{Y}_1$ is actually estimated from the data) for the ILS and 2SLS estimators and using the correct formula derived in the accompany paper for the IV-1 and IV-2 estimators at each repetition.

Table 4: Means (Row 1) and SDs (Row 2) of the OLS and ILS Estimates from 1000 Repetitions for the Partially Recursive Poisson-Normal GSiEM/GPA Model.

n	MLE β_{11} (1.0)	OLS β_{22} (10.0)	OLS γ_{21} (0.5)	ILS β_{22} (10.0)	ILS γ_{21} (0.5)	Asymptotic Bias of OLS	Asymptotic Bias of ILS
			$\alpha = 0.5$				
100							
	0.976	9.996	**0.549**	9.997	**0.482**	0.048	-0.018
	0.217	0.114	0.032	0.311	0.089	0.030	0.089
500							
	0.990	9.998	**0.537**	9.995	**0.490**	0.037	-0.010
	0.118	0.045	0.000	0.145	0.063	0.014	0.062
1000							
	0.997	10.00	**0.534**	10.00	**0.495**	0.034	-0.005
	0.089	0.032	0.000	0.105	0.063	0.011	0.061
			$\alpha = 2.0$				
100							
	0.976	9.997	**0.695**	10.00	**0.461**	0.195	-0.039
	0.219	0.205	0.089	0.436	0.134	0.091	0.136
500							
	0.992	10.00	**0.646**	10.01	**0.488**	0.146	-0.012
	0.114	0.095	0.045	0.187	0.077	0.049	0.076
1000							
	0.993	10.00	**0.637**	10.00	**0.490**	0.137	-0.010
	0.084	0.063	0.045	0.134	0.063	0.040	0.064

Table 5: Means (Row 1) and SDs (Row 2) of the OLS, ILS, 2SLS, IV-1, and IV-2 Estimates from 1000 Repetitions for the Partially Recursive Poisson-Normal GSiEM/GPA Model at α = 2.0 with n = 500.

	MLE (1)		MLE (2)			OLS			
	β_{10} (0)	β_{11} (1.0)	β_{10} (0)	β_{11} (1.0)	β_{12} (0)	β_{20} (0)	γ_{21} (0.5)	β_{22} (10.0)	$\widehat{\gamma}_{21}$ SE
Mean	0.4975	0.9931	0.4932	0.9937	-0.0006	-0.3882	**0.6459**	10.0003	**0.0153**
SD	0.0946	0.1142	0.0956	0.1148	0.0917	0.1517	**0.0491**	0.0929	0.0038

	ILS				2SLS			
	β_{20} (0)	γ_{21} (0.5)	β_{22} (10.0)	$\widehat{\gamma}_{21}$ SE	β_{20} (0)	γ_{21} (0.5)	β_{22} (10.0)	$\widehat{\gamma}_{21}$ SE
Mean	0.0270	**0.4867**	10.0000	**0.0563**	0.0126	**0.4917**	9.9971	**0.0560**
SD	0.2406	**0.0787**	0.1913	0.0110	0.2440	**0.0794**	0.1090	0.0111

	IV-1				IV-2			
	β_{20} (0)	γ_{21} (0.5)	β_{22} (10.0)	$\widehat{\gamma}_{21}$ SE	β_{20} (0)	γ_{21} (0.5)	β_{22} (10.0)	$\widehat{\gamma}_{21}$ SE
Mean	0.0088	**0.4947**	9.9998	**0.0323**	0.0053	**0.4961**	9.9998	**0.0320**
SD	0.1320	**0.0325**	0.1008	0.0104	0.1313	**0.0322**	0.1006	0.0104

1. **SD**: The *standard deviation* of the estimates from the 1000 repetitions.
2. **SE**: The *standard error* of $\widehat{\gamma}_{21}$ estimated naively by the usual OLS formula (ignoring the fact that the proxy $\widehat{Y}_1$ is actually estimated from the data) for the ILS and 2SLS estimators and using the correct formula derived in the accompany paper for the IV-1 and IV-2 estimators at each repetition.

11 FIGURES

Figure 1: Histograms and Quantile-Normal Plots for $\hat{\gamma}_{21}$ in a Partially Recursive *Binomial*-Normal GSiEM/GPA Model from 1000 Repetitions

From the top to the bottom: OLS, ILS, 2SLS, IV-1, and IV-2.

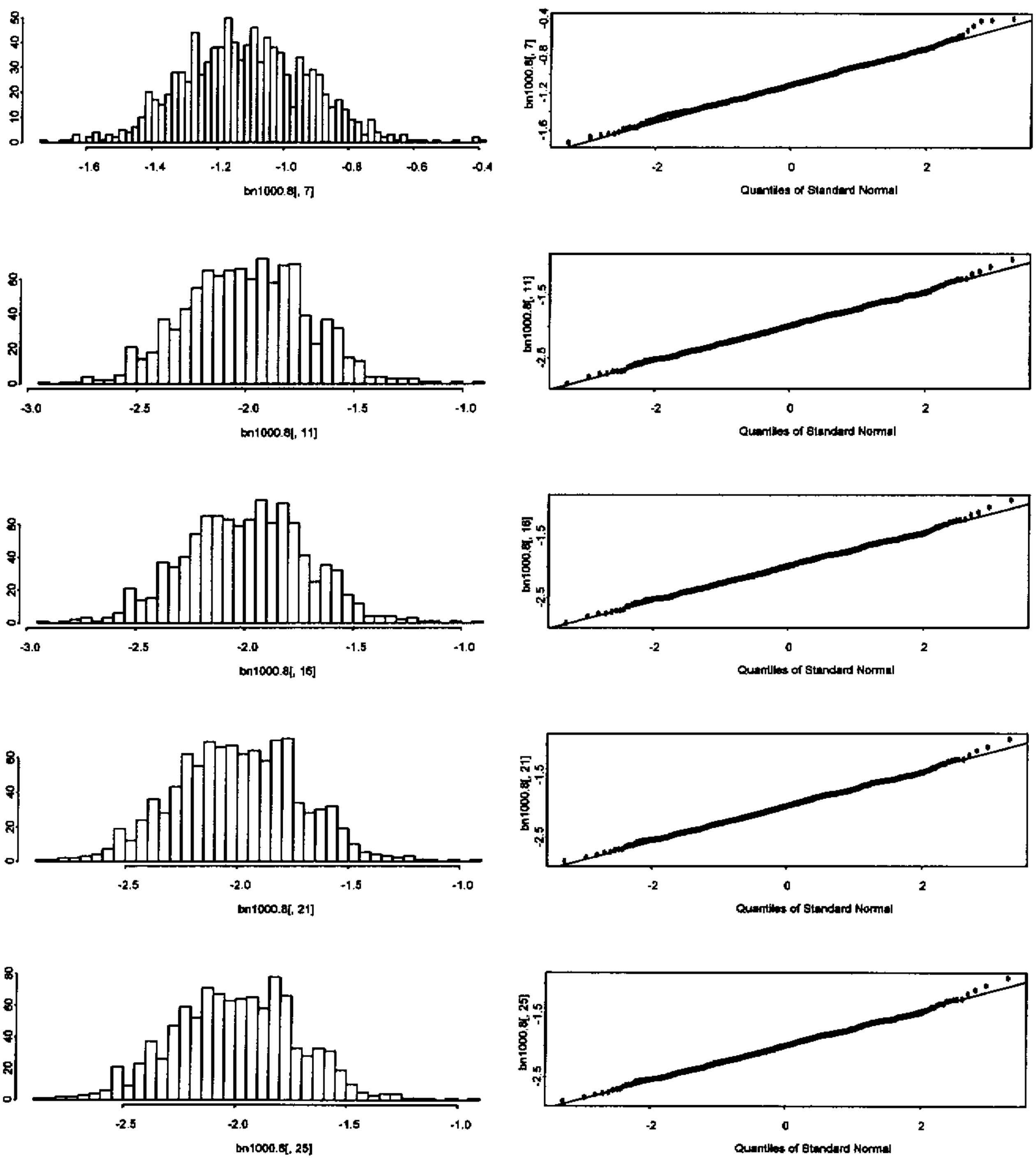
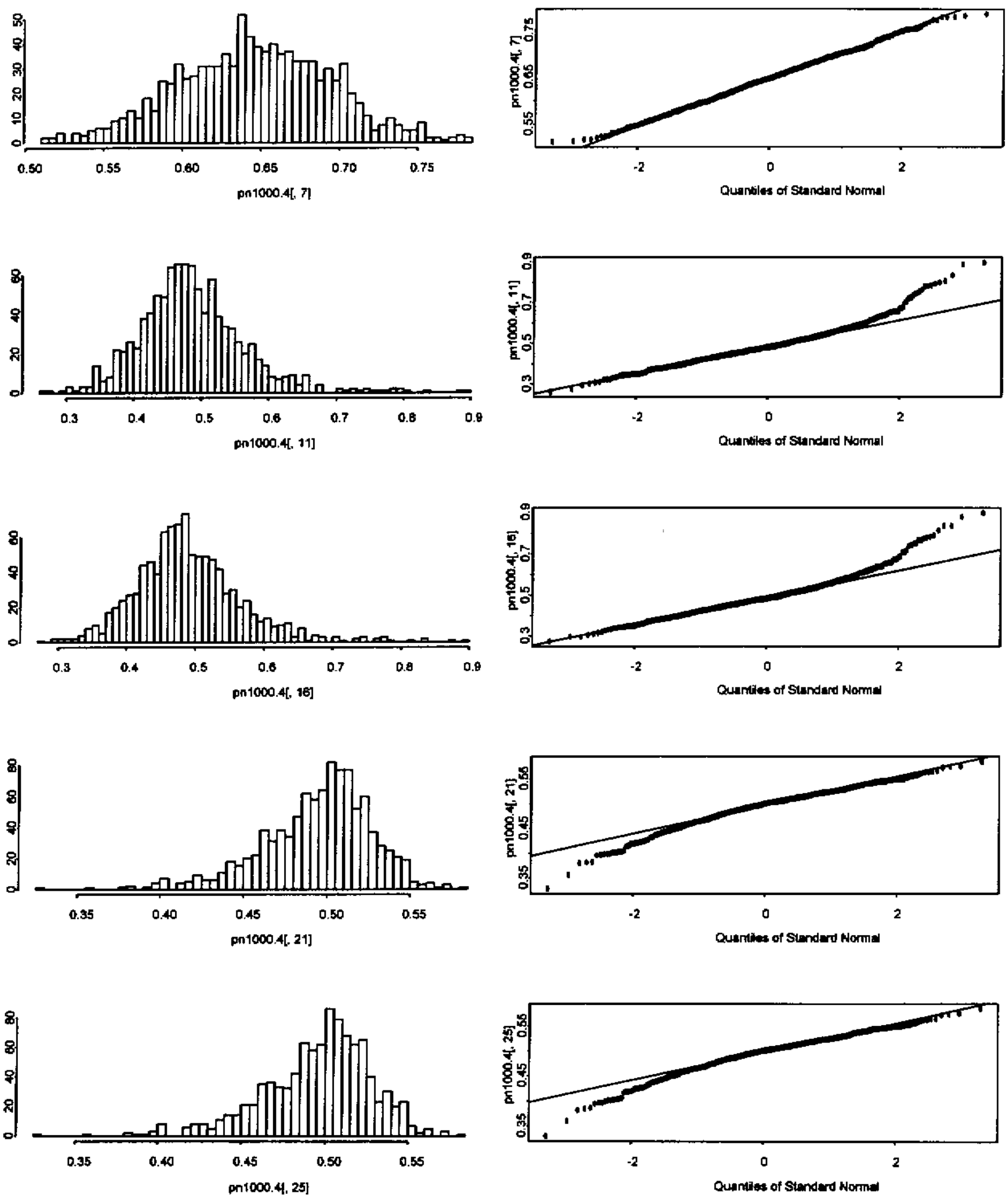


Figure 2: Histograms and Quantile-Normal Plots for $\hat{\gamma}_{21}$ in a Partially Recursive *Poisson*-Normal GSiEM/GPA Model from 1000 Repetitions

From the top to the bottom: OLS, ILS, 2SLS, IV-1, and IV-2.



(Submitted to the *Journal of the American Statistical Association*)

Generalized Simultaneous Equations Model and Generalized Path Analysis for Recursive Systems (II): The Instrumental Variable Estimator

Tien-Lung Tsai, Wen-Yi Shau, and Fu-Chang Hu[1]

Division of Biostatistics
Graduate Institute of Epidemiology
College of Public Health
National Taiwan University
Taipei, Taiwan 100
R.O.C.

December 14, 2000

[1]Correspondence: Fu-Chang Hu, Division of Biostatistics, Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, 1 Jen-Ai Road, Section 1, Room 1541, Taipei, Taiwan 100, R.O.C., Phone: +886 2 / 2394-2050 or +886 2 / 2312-3456 × 8349, Fax: +886 2 / 2351-1955, E-mail: fchu@ha.mc.ntu.edu.tw. This paper is based on the second part of Tsai's doctoral dissertation under Hu's advice. We are very grateful to the Doctoral Oral Examination Committee members for their helpful discussions and comments. And, Hu wishes to thank Professor Karl Jöreskog for his interesting introduction of the statistical software *LISREL* (Version 8.0) in an invited workshop held in Taipei, Taiwan, R.O.C., August 29-31, 1995. This research project was financially supported by the National Science Council of the Republic of China (NSC 86-2115-M-002-016, NSC 87-2118-M-002-003, NSC 88-2118-M-002-008, and NSC 89-2118-M-002-001).

ABSTRACT

The statistical models for a system of linear regression-like structural equations with observed variables, which include *simultaneous equations models* (SiEM) and *path analysis* (PA), have been used extensively for exploring and/or examining the plausible causal relationships among several continuous response variables (or endogenous variables) by economists and social scientists. These two kinds of statistical models are essentially the same except that their estimation methods differ. Given a set of independent variables (or exogenous variables), a system without any reciprocal effects between the response variables of the structural equations is called the *recursive* model, which is particularly useful in analyzing longitudinal data due to the temporal order of the responses. In this study, as inspired by the *generalized linear models* (GLMs), we generalize the linear SiEM and PA to deal with the situations in which the responses of a partially recursive system are a mixture of discrete and continuous variables. In a previous work, we have developed the *indirect least squares* (ILS) and *two-stage least squares* (2SLS) estimators for the recursive *generalized* simultaneous equations models (GSiEM) and *generalized* path analysis (GPA) models. In this paper, we combine the *instrumental variable* (IV) estimation method of SiEM with the *iterative reweighted least squares* (IRLS) algorithm of GLMs to estimate the structural coefficients of a partially recursive model with the response variables of mixed types. The IV estimation method has been developed in econometrics specifically for fixing the problem that one or more covariates are correlated with the error term of the equation. And, we prove in the paper that our IV estimator is consistent and asymptotically normally distributed. Unlike the linear cases, the ILS and 2SLS estimators are no longer equivalent to the corresponding IV estimators. The simulations, in which the performances of various estimators are further compared, show more promising results.

KEY WORDS:

Causal analysis; Structural equations model; Recursive model; Generalized linear models; Discrete response; Mixed responses; Indirect least squares; Two-stage least squares.

Contents

1 INTRODUCTION

The statistical models for a system of linear regression-like structural equations with observed variables, which include *simultaneous equations models* (SiEM) and *path analysis* (PA), have been used extensively for exploring and/or examining the plausible causal relationships among several continuous response variables (or endogenous variables) by economists and social scientists. These two kinds of statistical models are essentially the same except that their estimation methods differ. Given a set of independent variables (or exogenous variables), a system without any reciprocal effects between the response variables of the structural equations is called the *recursive* model, which is particularly useful in analyzing longitudinal data due to the temporal order of the responses. When all the error terms of the structural equations in a recursive SiEM/PA model are mutually uncorrelated, it is called the *fully* recursive model, for which consistent estimates of the structural coefficients can be obtained equation-by-equation separately. A recursive model with correlated error terms between some of the structural equations is called the *partially* recursive model, for which the equation-by-equation approach is usually not valid and the estimation of the structural coefficients should be based on the whole system of equations. Note that we shall make no distinction between independent variables and exogenous variables and between response variables and endogenous variables in this paper.

In this study, as inspired by *generalized linear models* (GLMs), we generalize the linear SiEM and PA to deal with the situations in which the responses of a partially recursive system are a mixture of discrete and continuous variables, but we focus on the recursive models and reserve the more complicated non-recursive models in a following research project. Since discrete responses such as "Yes/No" and "counts" are very popular in biological, medical, social, and public health studies, the development of *generalized* simultaneous equations models (GSiEM) and *generalized* path analysis (GPA) is important. As presented in the previous paper, we have combined the *indirect least squares* (ILS) and *two-stage least squares* (2SLS) estimation methods of SiEM with the *iterative reweighted least squares* (IRLS) algorithm of GLMs to estimate the structural coefficients of a partially recursive model with the response variables of mixed types. In particular, with the aid of the

IRLS algorithm, we successfully derive the *reduced form* of such a nonlinear recursive model, which is not only crucial for the ILS and 2SLS estimators but also important in its own right.

However, in this paper, we shall develop the *instrumental variable* (IV) estimator for the structural coefficients of a partially recursive model with the response variables of mixed types (see Sections 3 and 4) and show that the IV estimator is preferred due to the following four reasons. First, as reported in the previous paper, the ILS and 2SLS estimators for the partially recursive GSiEM/GPA models suffer two drawbacks: (1) It is difficult to estimate the asymptotic variances of the ILS and 2SLS estimators for a partially recursive GSiEM/GPA model. (2) When the distribution of the first response Y_1 in a partially recursive GSiEM/GPA model is *Poisson*, the performances of the ILS and 2SLS estimators are asymptotically worse than that of the IV estimator. Next, as would be shown in the simulation of this paper (Section 6), when the distribution of the first response Y_1 is *Poisson*, the ILS and 2SLS estimates of the structural coefficient γ_{21} on Y_1 in the equation for the second response Y_2 of a partially recursive GSiEM/GPA model are consistent only in certain limited situations, which depend on the values of several parameters. Moreover, although the ILS and 2SLS estimators for a linear SiEM are special cases of the IV estimator, the ILS and 2SLS estimators for a GSiEM/GPA model are *not* necessarily equivalent to any IV estimators (see Section 7). On the other hand, the IV estimation method for a *linear* and *nonlinear* SiEM, which actually minimizes an appropriate quadratic form, is more general and easier to be extended for more complex GSiEM/GPA models (see Section 2). Also, we shall prove in Section 5 that our IV estimator is consistent and asymptotically normally distributed. The simulations presented in Section 6, in which the performances of various estimators are further compared, show more promising results.

2 REVIEW

A brief review of nonlinear and discrete SiEM and PA models as well as GLMs has been given in the previous paper. In this section, we shall focus on the IV estimators. For easy computation, let the "convergence in probability" (i.e., $\xrightarrow{p}$) be denoted by "plim" hereinafter.

2.1 The Instrumental Variable (IV) Estimators for a Linear Equation

A *linear* equation is specified as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\mathbf{Y}$ is an $n \times 1$ vector of response variables, $\mathbf{X}$ is an $n \times p$ matrix of covariates, $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of errors. Suppose that

$$\text{plim}\left(\frac{1}{n}\mathbf{X}^{\mathrm{T}}\boldsymbol{\epsilon}\right) \neq \mathbf{0}$$

so that the *ordinary least squares* (OLS) or *generalized least squares* (GLS) estimators of $\boldsymbol{\beta}$ are not consistent. However, if we can find a set of q ($q \geq p$) variables to form a matrix $\mathbf{Z}$ such that

$$\text{plim}\left(\frac{1}{n}\mathbf{Z}^{\mathrm{T}}\boldsymbol{\epsilon}\right) = \mathbf{0}$$

then we may use $\mathbf{Z}$ as an instrument for $\mathbf{X}$ and minimize the following quadratic form with respect to $\boldsymbol{\beta}$

$$\begin{aligned} \mathbf{S}_1(\boldsymbol{\beta} \mid \mathbf{W}) &= \left(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\epsilon}\right)^{\mathrm{T}} \mathbf{W} \left(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\epsilon}\right) \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{Z}\mathbf{W}\mathbf{Z}^{\mathrm{T}}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

to obtain a consistent IV estimator

$$\widehat{\boldsymbol{\beta}}_{IV} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{P}_w\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{P}_w\mathbf{Y}$$

where $\mathbf{W}$ is a weight matrix and $\mathbf{P}_w = \mathbf{Z}\mathbf{W}\mathbf{Z}^{\mathrm{T}}$.

The single-equation IV estimators for a linear equation with (1) homogenous and independent ϵ_i and (2) heteroscedastic or dependent ϵ_i (without transformations) are listed in **Table 1**. And, the single-equation IV estimators for a linear equation with heteroscedastic or dependent ϵ_i (with transformations) are listed in **Table 2**. From these two tables, we can see that the IV estimation method provides a very rich class of estimators for linear equations. Most of the derivations are straightforward, which are available upon request from the authors, and the technical details can also be found in the references given in the tables.

2.2 The Instrumental Variable (IV) Estimators for a Nonlinear Equation

Next, a *nonlinear* equation can be specified as

$$\mathbf{Y} = \mathbf{f}(\mathbf{X}; \boldsymbol{\beta}) + \boldsymbol{\epsilon}.$$

One way to generalize the IV estimators from linear equations to nonlinear ones is to replace $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ by $\big(\mathbf{Y} - \mathbf{f}(\mathbf{X}; \boldsymbol{\beta})\big)$ in the minimization of the corresponding quadratic form. Thus, given an instrument $\mathbf{Z}$ for $\mathbf{X}$, a consistent IV estimator $\widetilde{\boldsymbol{\beta}}_{IV}$ is the value of $\boldsymbol{\beta}$ that minimizes the following quadratic form

$$\begin{aligned} \mathbf{S}_2(\boldsymbol{\beta} \mid \mathbf{W}) &= \Big[\mathbf{Z}^{\mathrm{T}}(\mathbf{Y} - \mathbf{f}(\mathbf{X}; \boldsymbol{\beta}))\Big]^{\mathrm{T}} \mathbf{W} \Big[\mathbf{Z}^{\mathrm{T}}(\mathbf{Y} - \mathbf{f}(\mathbf{X}; \boldsymbol{\beta}))\Big] \\ &= (\mathbf{Y} - \mathbf{f}(\mathbf{X}; \boldsymbol{\beta}))^{\mathrm{T}} \mathbf{Z}\mathbf{W}\mathbf{Z}^{\mathrm{T}} (\mathbf{Y} - \mathbf{f}(\mathbf{X}; \boldsymbol{\beta})) \end{aligned}$$

which may have to be solved iteratively. See Bowden and Turkington (1984, Chap. 5, pp. 156-201), Amemiya (1985, Chap. 8, pp. 245-266), and Davidson and MacKinnon (1993, pp. 661-667) for details.

3 MODEL SPECIFICATION, ASSUMPTIONS, AND INTERPRETATION

As mentioned in the previous paper, the estimation of the structural coefficients in a *fully* recursive GSiEM/GPA model is trivial since they can be estimated equation-by-equation separately. In fact, since all response variables are measured on the same group of subjects, the error terms of the equations are likely correlated as in the longitudinal data. Thus, to begin with, we consider the following *partially* recursive two-equation GSiEM/GPA model:

$$g_1(\mu_1) = \beta_{10} + \beta_{1x_1}X_1 + \beta_{1x_2}X_2 \tag{3.1}$$

$$\mu_2 = \beta_{20} + \beta_{2x_1}X_1 + \beta_{2x_3}X_3 + \gamma_{2y_1}Y_1 \tag{3.2}$$

where X_1, X_2, and X_3 are the independent variables, μ_1 and μ_2 are the means of the response variables Y_1 and Y_2 respectively,

$$Y_1 \sim \text{The exponential family of distributions, e.g., Binomial } (m, \mu_1),$$

$$Y_2 \sim \text{Normal } (\mu_2, \sigma_2^2),$$

and $g_1(\cdot)$ is the *link* function for μ_1. The subscript i, which indexes the observations ($i = 1, 2, \cdots, n$), is dropped for simplicity. It is assumed that the error term ϵ_2 of Eq. (3.2) has mean 0 and it is *independent* of the independent variables X_1, X_2, and X_3 respectively, i.e., $\epsilon_2 \perp X_1$, $\epsilon_2 \perp X_2$, and $\epsilon_2 \perp X_3$. As in the usual SiEM/PA models, the effects of the covariates on the corresponding response variables can be classified into three types — the *direct*, *indirect*, and *total* effects — in the GSiEM/GPA models, which have been defined in the previous paper.

Remarks. Although we consider a two-equation case, the statistical methods to be introduced in the sequel can be applied to multi-equation GSiEM/GPA Models — at least, recursively to each equation. At this time, we require that the link function for μ_2 be *identity*. The other link functions for μ_2 (e.g., *log*) and different types of Y_2's (e.g., binary responses) will be considered later. Moreover, notice that if the roles of Y_1 and Y_2 in this two-equation partially recursive GSiEM/GPA model are switched over, then the new GSiEM/GPA model

$$\begin{aligned} \mu_1 &= \beta_{10} + \beta_{1x_1} X_1 + \beta_{1x_2} X_2 \\ g_2(\mu_2) &= \beta_{20} + \beta_{2x_1} X_1 + \beta_{2x_3} X_3 + \gamma_{2y_1} Y_1 \end{aligned}$$

would be an easier case since the reduced form of the second equation is readily obtained by plugging the first equation (with its error term ϵ_1) into the second equation at the place of Y_1. Finally, as regard to the correlation between the error terms of the two structural equations Eqs. (3.1) and (3.2), we shall discuss this issue in Subsection 6.1, when we explain how to generate such data in simulations.

4 ESTIMATION: THE INSTRUMENTAL VARIABLE (IV) ESTIMATOR

In the previous paper, we have discussed: (1) two important tools for the estimation of structural coefficients in a GSiEM/GPA model, (2) the developments of the ILS and 2SLS estimators for a partially recursive GSiEM/GPA model, and (3) the derivation of the *reduced form* of a recursive GSiEM/GPA model with the aid of the IRLS algorithm of GLMs. Now, we introduce the IV method for estimating the structural coefficients in the partially recursive GLM-LM GSiEM/GPA model (3.1) and (3.2).

4.1 Steps

- **Step 1:**

  Fit the GLM for the response Y_1 on *all* the independent variables including X_1, X_2, and X_3 using the IRLS algorithm to obtain the estimates $\widehat{\beta}^*_{10}$, $\widehat{\beta}^*_{1x_1}$, $\widehat{\beta}^*_{1x_2}$, and $\widehat{\beta}^*_{1x_3}$.

- **Step 2:**

  In fact, in **Step 1**, the estimates $\widehat{\beta}^*_{10}$, $\widehat{\beta}^*_{1x_1}$, $\widehat{\beta}^*_{1x_2}$, and $\widehat{\beta}^*_{1x_3}$ are obtained by fitting a multiple linear regression of the *pseudo*-response variable Z^*_1 (defined below) on the original independent variables X_1, X_2, and X_3 on the convergence:

$$\begin{aligned} Z^*_1 &= \widehat{\beta}^*_{10} + \widehat{\beta}^*_{1x_1}X_1 + \widehat{\beta}^*_{1x_2}X_2 + \widehat{\beta}^*_{1x_3}X_3 + g'_1\left(\widehat{\mu}^*_1\right)\left(Y_1 - \widehat{\mu}^*_1\right) \\ &= \widehat{\beta}^*_{10} + \widehat{\beta}^*_{1x_1}X_1 + \widehat{\beta}^*_{1x_2}X_2 + \widehat{\beta}^*_{1x_3}X_3 + \epsilon^*_{1,2}. \end{aligned}$$

  Thus, we can rewrite the original response variable Y_1 as a function of the pseudo-response variable Z^*_1, i.e.,

$$\begin{aligned}
Z_1^* &= \left(\widehat{\beta}_{10}^* + \widehat{\beta}_{1x_1}^* X_1 + \widehat{\beta}_{1x_2}^* X_2 + \widehat{\beta}_{1x_3}^* X_3\right) + g_1'\left(\widehat{\mu}_1^*\right)\left(Y_1 - \widehat{\mu}_1^*\right) \\
&= g_1\left(\widehat{\mu}_1^*\right) + g_1'\left(\widehat{\mu}_1^*\right)\left(Y_1 - \widehat{\mu}_1^*\right), \\
Z_1^* - g_1\left(\widehat{\mu}_1^*\right) &= g_1'\left(\widehat{\mu}_1^*\right)\left(Y_1 - \widehat{\mu}_1^*\right), \\
Y_1 - \widehat{\mu}_1^* &= \frac{Z_1^* - g_1\left(\widehat{\mu}_1^*\right)}{g_1'\left(\widehat{\mu}_1^*\right)}, \\
Y_1 &= \widehat{\mu}_1^* + \frac{Z_1^* - g_1\left(\widehat{\mu}_1^*\right)}{g_1'\left(\widehat{\mu}_1^*\right)}
\end{aligned}$$

where

$$\widehat{\mu}_1^* = \widehat{Y}_1^* = g_1^{-1}\left(\widehat{\beta}_{10}^* + \widehat{\beta}_{1x_1}^* X_1 + \widehat{\beta}_{1x_2}^* X_2 + \widehat{\beta}_{1x_3}^* X_3\right).$$

In this step, we just compute $\widehat{\mu}_1^*$ by overfitting Y_1.

- **Step 3:**

From Eq. (3.2), we have

$$Y_2 = \beta_{20} + \beta_{2x_1} X_1 + \beta_{2x_3} X_3 + \gamma_{2y_1} Y_1 + \epsilon_2$$

or in matrix notation

$$\mathbf{Y}_2 = \mathbf{A}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2.$$

Then, the IV estimate for the unknown parameters in the second equation, Eq. (3.2), is

$$\widehat{\boldsymbol{\beta}}_{2,IV} = \left(\mathbf{A}_2^{*\mathrm{T}} \mathbf{A}_2\right)^{-1} \mathbf{A}_2^{*\mathrm{T}} \mathbf{Y}_2 \tag{4.1}$$

where $\mathbf{A}_2 = \left[\mathbf{1}, \mathbf{X}_1, \mathbf{X}_3, \mathbf{Y}_1\right]$, $\mathbf{A}_2^* = \left[\mathbf{1}, \mathbf{X}_1, \mathbf{X}_3, \widehat{\mathbf{Y}}_1^*\right]$, and $\mathbf{A}_2^*$ satisfies the following requirements for being an IV:

$$\begin{aligned}
\text{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}} \epsilon_2\right) &= \mathbf{0}, \\
\text{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}} \mathbf{A}_2\right) &= \mathbf{Q}_{A_2^* A_2} \quad \text{(a finite nonsingular matrix)}, \\
\text{plim}\left(\frac{1}{n}\mathbf{A}_2^{*T} \mathbf{A}_2^*\right) &= \mathbf{Q}_{A_2^* A_2^*} \quad \text{(a positive definite matrix)}.
\end{aligned}$$

The justification for $\widehat{\beta}_{2,IV}$ being a valid estimator will be given in the following subsection.

4.2 Justification

By taking $\mathbf{A}_2^*$ as the IV for estimating β_2, we can see intuitively that from Eq. (4.1),

$$\begin{aligned}
\widehat{\beta}_{2,IV} &= \left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2 \\
&= \left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}(\mathbf{A}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2) \\
&= \left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\boldsymbol{\beta}_2 + \left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\boldsymbol{\epsilon}_2) \\
&= \boldsymbol{\beta}_2 + \left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\boldsymbol{\epsilon}_2
\end{aligned}$$

is a reasonable estimator of $\boldsymbol{\beta}_2$ as long as $\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\boldsymbol{\epsilon}_2$ is close to zero.

To show in our case that the IV $\mathbf{A}_2^*$ is indeed *uncorrelated* with the error term ϵ_2, we argue that $\widehat{Y}_1^*$ (or, equivalently, $\widehat{\mu}_1^*$) is a function of (X_1, X_2, X_3), which are independent of ϵ_2 by the model assumption, so that $\mathbf{A}_2^* = \left[X_1, X_3, \widehat{Y}_1^*\right]$ is independent of ϵ_2.

Specifically, the key requirement for $\widehat{Y}_1$ being an IV is

$$\text{plim}\left(\frac{1}{n}\widehat{\mathbf{Y}}_1^{\mathrm{T}}\boldsymbol{\epsilon}_2\right) = 0$$

but we shall prove that $Cov\left(\widehat{Y}_1, \epsilon_2\right) = 0$, which implies plim $\left(\frac{1}{n}\sum_i \widehat{Y}_{i1}\epsilon_{i2}\right) = 0$ since $E(\epsilon_{i2}) = 0$.

Proof: $Cov\left(\widehat{Y}_1, \epsilon_2\right) = 0$.

Note that the symbol "$\perp$" stands for "being independent of." By the model assumption, $X_1 \perp \epsilon_2$ and $X_2 \perp \epsilon_2$. And, $\widehat{Y}_1$ (or, equivalently, $\widehat{\mu}_1$) is a function of (X_1, X_2). Since any functions of independent variables are still independent to each other (see, e.g., Casella and Berger 1990, Theorem 4.3.2, p. 150), we have $\widehat{Y}_1 \perp \epsilon_2$, which implies $Cov\left(\widehat{Y}_1, \epsilon_2\right) = 0$. Thus, the key requirement for $\widehat{Y}_1$ being an IV is fulfilled. By the same token, the key requirement for $\widehat{Y}_1^*$ being an IV is also fulfilled.

5 STATISTICAL INFERENCE

The statistical properties of our IV estimator for the structural coefficients in a partially recursive GLM-LM GSiEM/GPA model will be investigated analytically as well as numerically. The IV estimator introduced in Subsection 4.1, $\widehat{\boldsymbol{\beta}}_{2,IV} = \left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2$, is *consistent*, *asymptotically unbiased*, and *asymptotically normally distributed*. The proofs are given in the following subsections.

5.1 Consistency

First, we prove that $\widehat{\boldsymbol{\beta}}_{2,IV}$ is a *consistent* estimator, i.e., $\operatorname{plim}\widehat{\boldsymbol{\beta}}_{2,IV} = \boldsymbol{\beta}_2$.

Proof:

We have

$$\begin{aligned} \operatorname{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2\right) &= \left[\operatorname{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)\right]\boldsymbol{\beta}_2 + \operatorname{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\boldsymbol{\epsilon}_2\right) \\ &= \mathbf{Q}_{A_2^*A_2}\boldsymbol{\beta}_2 \end{aligned}$$

since $\operatorname{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\boldsymbol{\epsilon}_2\right) = 0$. And,

$$\begin{aligned} \operatorname{plim}\widehat{\boldsymbol{\beta}}_{2,IV} &= \operatorname{plim}\left[\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2\right] \\ &= \operatorname{plim}\left[\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\left(\mathbf{A}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2\right)\right] \\ &= \operatorname{plim}\left[\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)\boldsymbol{\beta}_2 + \left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\boldsymbol{\epsilon}_2\right)\right] \\ &= \boldsymbol{\beta}_2. \end{aligned}$$

Hence, $\widehat{\boldsymbol{\beta}}_{2,IV}$ is a consistent estimator.

5.2 Asymptotic Unbiasedness

There are several definitions for asymptotic unbiasedness of an estimator. Firstly, Amemiya (1985, pp. 93-95) and Davidson and MacKinnon (1993, pp. 124-125) defined the *asymptotic unbiasedness* for $\widehat{\beta}$ as

$$AE\left(\widehat{\beta}\right) = \int_{-\infty}^{\infty} \widehat{\beta}\, dF\left(\widehat{\beta}\right) = \beta,$$

where AE stands for the "asymptotic expectation" (or the "asymptotic mean") and $F(\widehat{\beta})$ is the cumulative distribution function (cdf) of the limiting distribution for $\widehat{\beta}$. Then, under this definition, a consistent estimator is asymptotically unbiased, but not vice versa (Amemiya 1985, p. 95).

Secondly, Bickel and Doksum (1977, pp. 133-135) gave the following definition of *asymptotic unbiasedness* for $\widehat{\beta}$ is

$$\lim_{n\to\infty} \left[\frac{E\left(\widehat{\beta}\right) - \beta}{\sqrt{Var\left(\widehat{\beta}\right)}} \right] = 0.$$

In this definition, asymptotic unbiasedness is a stronger property — a consistent estimator may be asymptotically biased. They (1977, p. 150) provided an example in their exercise problem #8.

Similarly, Serfling (1980, p. 48) provided the following definition of *asymptotic unbiasedness* for $\widehat{\beta}$ is

$$\lim_{n\to\infty} E\left(\widehat{\beta}\right) = \beta.$$

Yet, as argued by Davidson and MacKinnon (1993, pp. 124-125), the definition of asymptotic unbiasedness from Amemiya (1985) and Davidson and MacKinnon (1993) is preferred; and thus, we adopt their definition. Since our IV estimator $\widehat{\beta}_{2,IV}$ is *consistent* from Subsection 5.1, it is also *asymptotically unbiased.* Specifically, as shown in our simulations to be discussed in Section 6, the IV estimator $\widehat{\beta}_{2,IV}$ is indeed asymptotically unbiased.

Finally, to see if our IV estimator is *unbiased,* we partition $\widehat{\beta}_{2,IV}$ into two parts

$$\begin{aligned}\widehat{\beta}_{2,IV} &= (\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2)^{-1}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2 \\ &= (\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2)^{-1}\mathbf{A}_2^{*\mathrm{T}}(\mathbf{A}_2\beta_2+\epsilon_2) \\ &= \beta_2+(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2)^{-1}\mathbf{A}_2^{*\mathrm{T}}\epsilon_2.\end{aligned}$$

By the model specification, $X_1 \perp \epsilon_2$, $X_2 \perp \epsilon_2$, $X_3 \perp \epsilon_2$, and $E(\epsilon_2)=0$. Since $\widehat{Y}_1^*$ is a function of X_1, X_1, and X_3, we have

$$E\left[\mathbf{A}_2^{*\mathrm{T}}\epsilon_2\right]=\mathbf{0}.$$

Then, it seems that

$$E\left[(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2)^{-1}\mathbf{A}_2^{*\mathrm{T}}\epsilon_2\right]=\mathbf{0}$$

but, in fact, $\widehat{Y}_1^*$ is random and $Cov(Y_1,\epsilon_2)\neq 0$ may disprove it. The issue about whether a consistent IV estimator is also unbiased has been discussed by White (1984, pp. 9-10) and Davidson and MacKinnon (1993, p. 217) respectively.

5.3 Asymptotic Normality

Also, we show that $\widehat{\beta}_{2,IV}$ is *asymptotically normally distributed,* i.e.,

$$\sqrt{n}\left(\widehat{\beta}_{2,IV}-\beta_2\right) \xrightarrow{d} \mathbf{N}_p\left(\mathbf{0},\Sigma_{\widehat{\beta}_{2,IV}}\right).$$

Proof:

First, we have

$$\begin{aligned}\sqrt{n}\left(\widehat{\beta}_{2,IV}-\beta_2\right) &= \sqrt{n}\left[\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2-\beta_2\right] \\ &= \sqrt{n}\left[\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\left(\mathbf{Y}_2-\mathbf{A}_2\beta_2\right)\right] \\ &= \sqrt{n}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\epsilon_2\right).\end{aligned}$$

Next, we shall apply the multivariate version of the *Lindeberg-Feller central limit theorem* (CLT) for independent, but not identically distributed, random vectors (see, e.g., Serfling (1980, Subsec. 1.9.2, Theorem B, pp. 30-31) and Rao (1973, Example 4.7, p. 147)) to obtain the asymptotic normality of $\sqrt{n}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\epsilon_2\right)$. In our case, that theorem states the following. Let $\{\mathbf{A}_{i2}^{*}\epsilon_{i2}\}$ be independent random vectors with means $\{0\}$, covariance matrices $\{\boldsymbol{\Sigma}_i\}$, and the distribution functions $\{F_i\}$. Suppose that as $n \to \infty$,

$$\frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 + \cdots + \boldsymbol{\Sigma}_n}{n} \longrightarrow \boldsymbol{\Sigma} \neq \mathbf{0}$$

and the *Lindeberg condition*

$$\frac{1}{n}\sum_{i=1}^{n}\int_{||\mathbf{A}_{i2}^{*}\epsilon_{i2}|| > \varepsilon\sqrt{n}} ||\mathbf{A}_{i2}^{*}\epsilon_{i2}||^2 \, dF_i \longrightarrow 0 \quad \text{for every } \varepsilon > 0$$

is satisfied. Then,

$$\sqrt{n}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\epsilon_2\right) = \sqrt{n}\left(\frac{\sum_{i=1}^{n}\mathbf{A}_{i2}^{*}\epsilon_{i2}}{n}\right) \xrightarrow{d} \mathbf{N}_p\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$$

where $\mathbf{N}_p$ is a p-variate Normal distribution, p is the column dimension of the design matrix $\mathbf{A}_2$, $\boldsymbol{\Sigma} = \sigma_2^2\mathbf{Q}_{A_2^*A_2^*}$, and $\mathbf{Q}_{A_2^*A_2^*} = \text{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^{*}\right)$. The proof is left in **Appendix 1**.

Finally, we obtain

$$\sqrt{n}\left(\widehat{\beta}_{2,IV} - \beta_2\right) \xrightarrow{d} \mathbf{N}_p\left(\mathbf{0}, \sigma_2^2\mathbf{Q}_{A_2^*A_2}^{-1}\mathbf{Q}_{A_2^*A_2^*}\mathbf{Q}_{A_2A_2^*}^{-1}\right)$$

where $\mathbf{Q}_{A_2^*A_2}^{-1}$ and $\mathbf{Q}_{A_2A_2^*}^{-1}$ are $\left[\text{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)\right]^{-1}$ and $\left[\text{plim}\left(\frac{1}{n}\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\right)\right]^{-1}$ respectively. Note that $\mathbf{Q}_{A_2^*A_2} = \text{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)$ is a finite nonsingular matrix as required in **Step 3** of Subsection 4.1. As given in **Appendix 2**, it is straightforward to show that $\text{plim}\left[\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\right] = \left[\text{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)\right]^{-1}$ by applying the *Application D* of the *Corollary* in Section 1.7 of Serfling (1980, p. 26), which ends the proof.

5.4 The Estimated Asymptotic Variance

The asymptotic variance of the IV estimator $\widehat{\beta}_{2,IV}$ can be consistently estimated by

$$\widehat{Var}\left(\widehat{\beta}_{2,IV}\right) = \widehat{\sigma}_{2,IV}^2\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^{*}\right)\left(\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\right)^{-1}$$

where

$$\widehat{\sigma}_{2,IV}^2 = \frac{\left(\mathbf{Y}_2 - \mathbf{A}_2\widehat{\boldsymbol{\beta}}_{2,IV}\right)^{\mathrm{T}}\left(\mathbf{Y}_2 - \mathbf{A}_2\widehat{\boldsymbol{\beta}}_{2,IV}\right)}{n-p}$$

with $p =$ the column dimension of the design matrix $\mathbf{A}_2$. By comparing the mean of the estimated standard error (SE) of $\widehat{\boldsymbol{\beta}}_{2,IV}$ from 1000 repetitions with the corresponding standard deviation (SD) of the sampling distribution of $\widehat{\boldsymbol{\beta}}_{2,IV}$, the correctness of the above formula for estimating the asymptotic variance of $\widehat{\boldsymbol{\beta}}_{2,IV}$ has already been examined in our simulations presented in the previous paper. In contrast, since the ILS and 2SLS estimators $\widehat{\boldsymbol{\beta}}_{2,ILS}$ and $\widehat{\boldsymbol{\beta}}_{2,2SLS}$ introduced in the previous paper take a "two-step" estimation procedure, the formulas for estimating their asymptotic variances are much more complicated, although a formula for estimating the asymptotic variance of the second-step parameter estimate has been worked out by Murphy and Topel (1985) as discussed in Greene (2000, pp. 133-137 and pp. 433-438).

5.5 The Determinants of $\widehat{Var}(\widehat{\gamma}_{21,IV})$

In this subsection, we shall find out the major determinants of the estimated asymptotic variance for the estimate of the structural coefficient $\gamma_{21,IV}$ on the covariate Y_1 in the second equation. Without loss of generality, let all variables be centered for simplicity. Recall that the estimated asymptotic variance of $\widehat{\boldsymbol{\beta}}_{2,IV}$ is

$$\widehat{Var}\left(\widehat{\boldsymbol{\beta}}_{2,IV}\right) = \widehat{\sigma}_{2,IV}^2\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^*\right)\left(\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^*\right)^{-1}$$

where, after centering, $\mathbf{A}_2 = \left[\mathbf{X}_2, \mathbf{Y}_1\right]$, $\mathbf{A}_2^* = \left[\mathbf{X}_2, \widehat{\mathbf{Y}}_1^*\right]$, and

$$\widehat{\sigma}_{2,IV}^2 = \frac{\left(\mathbf{Y}_2 - \mathbf{A}_2\widehat{\boldsymbol{\beta}}_{2,IV}\right)^{T}\left(\mathbf{Y}_2 - \mathbf{A}_2\widehat{\boldsymbol{\beta}}_{2,IV}\right)}{n-p}$$

with $p =$ the column dimension of the design matrix $\mathbf{A}_2$. By rewriting

$$\mathbf{Y}_1 = \widehat{\mathbf{Y}}_1^* + \left[\frac{\boldsymbol{\epsilon}_{1,2}^*}{g_1'\left(\widehat{\boldsymbol{\mu}}_1\right)}\right] = \widehat{\mathbf{Y}}_1^* + \boldsymbol{\epsilon}_{Y_1}^*,$$

we can partition $\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^*$ into two parts:

$$\begin{aligned}
\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^* &= \mathbf{A}_2^{*\mathrm{T}}\left(\mathbf{A}_2 + \left[\,\mathbf{0},\ -\boldsymbol{\epsilon}_{Y_1}^*\right]\right) \\
&= \mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2 + \left[\,\mathbf{0},\ -\mathbf{A}_2^{*\mathrm{T}}\boldsymbol{\epsilon}_{Y_1}^*\right].
\end{aligned}$$

Then, since

$$\operatorname{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\boldsymbol{\epsilon}_{Y_1}^*\right) = \mathbf{0},$$

we have

$$\operatorname{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^*\right) = \operatorname{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)$$

and thus

$$\operatorname{plim}\left[\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^*\right)\left(\frac{1}{n}\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^*\right)^{-1}\right] = \operatorname{plim}\left[\left(\frac{1}{n}\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^*\right)^{-1}\right].$$

Next, we make the following decomposition

$$\begin{aligned}
(\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^*)^{-1} &= \left(\begin{bmatrix}\mathbf{X}_2^{\mathrm{T}} \\ \mathbf{Y}_1^{\mathrm{T}}\end{bmatrix}\begin{bmatrix}\mathbf{X}_2 & \widehat{\mathbf{Y}}_1^*\end{bmatrix}\right)^{-1} \\
&= \begin{bmatrix}\sum X_{i2}^2 & \sum X_{i2}\widehat{Y}_{i1}^* \\ \sum X_{i2}Y_{i1} & \sum Y_{i1}\widehat{Y}_{i1}^*\end{bmatrix}^{-1} \\
&= \frac{1}{|\,A\,|}\begin{bmatrix}\sum Y_{i1}\widehat{Y}_{i1}^* & -\sum X_{i2}\widehat{Y}_{i1}^* \\ -\sum X_{i2}Y_{i1} & \sum X_{i2}^2\end{bmatrix}
\end{aligned}$$

where

$$|\,A\,| = \sum X_{i2}^2 \sum Y_{i1}\widehat{Y}_{i1}^* - \sum X_{i2}Y_{i1}\sum X_{i2}\widehat{Y}_{i1}^*.$$

Thus, the estimated asymptotic variance of $\widehat{\gamma}_{21,IV}$ equals

$$\widehat{Var}\left(\widehat{\gamma}_{21,IV}\right) = \widehat{\sigma}_{2,IV}^2\left(\frac{\sum X_{i2}^2}{\sum X_{i2}^2\sum Y_{i1}\widehat{Y}_{i1}^* - \sum X_{i2}Y_{i1}\sum X_{i2}\widehat{Y}_{i1}^*}\right)$$

which reduces to

$$\frac{\widehat{\sigma}^2_{2,IV}}{n\,\widehat{Cov}\left(Y_1, \widehat{Y}_1^*\right)}$$

if, in addition, $X_2 \perp Y_1$. Hence, as one would expect, $\widehat{Cov}\left(Y_1, \widehat{Y}_1^*\right)$ is actually one of the most important determinants for the estimated asymptotic variance of $\widehat{\gamma}_{21,IV}$. The better $\widehat{Y}_1^*$ fits Y_1, the smaller $\widehat{Var}\left(\widehat{\gamma}_{21,IV}\right)$ would be, which will be demonstrated in the simulations of Section 6.

5.6 The Optimal IV Estimator

As discussed by White (1984, Sec. 4.3, pp. 78-106, esp., pp. 99-100), there are three ways to improving the *efficiency* of an IV estimator, which are: (1) use weighted IV's and all extra available IV's, (2) take a transformation for a nonspherical model, and (3) use linear or nonlinear constraints. This provides a guide for the following discussions (see, also, **Tables 1** and **2**).

5.6.1 The Choice of the Weight Matrix

Recall that the key requirement for $\widehat{Y}_1^*$ being an IV for Y_1 is

$$\text{plim}\left[\frac{\sum_{i=1}^{n} \widehat{Y}_{i1}^* \left(Y_{i2} - \mathbf{A}_{i2}\boldsymbol{\beta}_2\right)}{n}\right] = 0.$$

Under such condition, we wish to choose a $\widehat{\boldsymbol{\beta}}_2$ to make $\frac{1}{n}\sum_{i=1}^{n} \widehat{Y}_{i1}^* \left(Y_{i2} - \mathbf{A}_{i2}\widehat{\boldsymbol{\beta}}_2\right)$ as small as possible. As in an OLS estimation, we minimize the following quadratic form

$$\begin{aligned}
\mathbf{S}_1(\boldsymbol{\beta}_2) &= \left[\mathbf{A}_2^{*\mathrm{T}}\left(\mathbf{Y}_2 - \mathbf{A}_2\boldsymbol{\beta}_2\right)\right]^{\mathrm{T}}\left[\mathbf{A}_2^{*\mathrm{T}}\left(\mathbf{Y}_2 - \mathbf{A}_2\boldsymbol{\beta}_2\right)\right] \\
&= \left(\mathbf{Y}_2 - \mathbf{A}_2\boldsymbol{\beta}_2\right)^{\mathrm{T}}\mathbf{A}_2^*\mathbf{A}_2^{*\mathrm{T}}\left(\mathbf{Y}_2 - \mathbf{A}_2\boldsymbol{\beta}_2\right)
\end{aligned}$$

with respect to $\boldsymbol{\beta}_2$. Differentiating $\mathbf{S}_1(\boldsymbol{\beta}_2)$ by $\boldsymbol{\beta}_2$ and setting it to zero

$$\frac{\partial \mathbf{S}_1(\boldsymbol{\beta}_2)}{\partial \boldsymbol{\beta}_2} = -2\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^*\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2 + 2\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^*\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\boldsymbol{\beta}_2 = \mathbf{0}$$

yields

$$\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\boldsymbol{\beta}_2 = \mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2$$

so that

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{2,IV} &= \left(\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2 \\ &= \left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\left(\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\right)^{-1}\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2 \\ &= (\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2)^{-1}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2\end{aligned}$$

where $\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2$ and $\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}$ are square and nonsingular. In general, the instrument $\mathbf{A}_2^{*}$ can have a larger rank than $\mathbf{A}_2$. If $Rank(\mathbf{A}_2^{*}) > Rank(\mathbf{A}_2)$, then

$$\widehat{\boldsymbol{\beta}}_{2,IV}^{*} = \left(\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2.$$

Moreover, since $Var\left[\mathbf{A}_2^{*\mathrm{T}}(\mathbf{Y}_2 - \mathbf{A}_2\boldsymbol{\beta}_2)\right]$ is not homogenous, we can use the optimal weight matrix $\left\{Var\left[\mathbf{A}_2^{*\mathrm{T}}(\mathbf{Y}_2 - \mathbf{A}_2\boldsymbol{\beta}_2)\right]\right\}^{-1}$ to improve the efficiency as in a GLS estimation. Thus, we minimize the following quadratic form

$$\begin{aligned}\mathbf{S}_2(\boldsymbol{\beta}_2 \mid \mathbf{W}_2) &= \left[\mathbf{A}_2^{*\mathrm{T}}(\mathbf{Y}_2 - \mathbf{A}_2\boldsymbol{\beta}_2)\right]^{\mathrm{T}}\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^{*}\right)^{-1}\left[\mathbf{A}_2^{*\mathrm{T}}(\mathbf{Y}_2 - \mathbf{A}_2\boldsymbol{\beta}_2)\right] \\ &= (\mathbf{Y}_2 - \mathbf{A}_2\boldsymbol{\beta}_2)^{\mathrm{T}}\mathbf{A}_2^{*}\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^{*}\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}(\mathbf{Y}_2 - \mathbf{A}_2\boldsymbol{\beta}_2)\end{aligned}$$

with respect to $\boldsymbol{\beta}_2$. Differentiating $\mathbf{S}_2(\boldsymbol{\beta}_2 \mid \mathbf{W}_2)$ by $\boldsymbol{\beta}_2$ and setting it to zero

$$\frac{\partial \mathbf{S}_2(\boldsymbol{\beta}_2 \mid \mathbf{W}_2)}{\partial \boldsymbol{\beta}_2} = -2\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^{*}\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2 + 2\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^{*}\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\boldsymbol{\beta}_2 = \mathbf{0}$$

yields

$$\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^{*}\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\boldsymbol{\beta}_2 = \mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^{*}\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2$$

so that

$$\begin{aligned}\widetilde{\boldsymbol{\beta}}_{2,IV} &= \left(\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^{*}\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2^{*}\left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^{*}\right)^{-1}\mathbf{A}_2^{*\mathrm{T}}\mathbf{Y}_2 \\ &= \left(\mathbf{A}_2^{\mathrm{T}}\mathbf{P}_{A^*}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{\mathrm{T}}\mathbf{P}_{A^*}\mathbf{Y}_2\end{aligned}$$

where $\mathbf{P}_{A^*} = \mathbf{A}_2^* \left(\mathbf{A}_2^{*T}\mathbf{A}_2^*\right)^{-1} \mathbf{A}_2^{*T}$. See Bowden and Turkington (1984, pp. 13-14) and White (1984, Sec. 4.3, pp. 78-106) for more details.

However, if $Rank(\mathbf{A}_2^*) = Rank(\mathbf{A}_2)$, then

$$\begin{aligned}
\widetilde{\beta}_{2,IV} &= \left(\mathbf{A}_2^T\mathbf{A}_2^* \left(\mathbf{A}_2^{*T}\mathbf{A}_2^*\right)^{-1} \mathbf{A}_2^{*T}\mathbf{A}_2\right)^{-1} \mathbf{A}_2^T\mathbf{A}_2^* \left(\mathbf{A}_2^{*T}\mathbf{A}_2^*\right)^{-1} \mathbf{A}_2^{*T}\mathbf{Y}_2 \\
&= \left(\mathbf{A}_2^{*T}\mathbf{A}_2\right)^{-1} \mathbf{A}_2^{*T}\mathbf{A}_2^* \left(\mathbf{A}_2^T\mathbf{A}_2^*\right)^{-1} \mathbf{A}_2^T\mathbf{A}_2^* \left(\mathbf{A}_2^{*T}\mathbf{A}_2^*\right)^{-1} \mathbf{A}_2^{*T}\mathbf{Y}_2 \\
&= \left(\mathbf{A}_2^{*T}\mathbf{A}_2\right)^{-1} \mathbf{A}_2^{*T}\mathbf{Y}_2
\end{aligned}$$

which reduces to $\widehat{\beta}_{2,IV}$. Thus, when $Rank(\mathbf{A}_2^*) = Rank(\mathbf{A}_2)$, the optimal weighted IV estimator and the unweighted IV estimator are exactly the same.

5.6.2 The Choice of Overfitting and/or Larger Rank

Although the asymptotic variance of the IV estimator will usually get smaller by adding extra IV's, the gain in efficiency may be very little after including an additional IV. As pointed out by White (1984, p. 82), if the extra IV is *uncorrelated* with the residual $\left[\mathbf{I} - \mathbf{A}_2^* \left(\mathbf{A}_2^{*T}\mathbf{A}_2^*\right)^{-1} \mathbf{A}_2^{*T}\right] \mathbf{A}_2$ in our notation, then it would not be useful.

5.6.3 The Choice of the Best IV Estimator for a Nonlinear Equation

Amemiya (1985, p. 248) has provided the *best nonlinear two-stage least squares (BNL2S) estimator* for a nonlinear SiEM, which is actually an IV estimator (Davidson and MacKinnon 1993, p. 663). The optimal choice of an IV for a nonlinear equation

$$Y = f(\mathbf{X}; \beta) + \epsilon$$

is

$$\bar{\mathbf{G}} \equiv E(\mathbf{G}_0) \equiv E\left[\frac{\partial f}{\partial \beta}\right]$$

evaluated at the true value of β. Then, the best IV estimator for this nonlinear equation can be obtained by minimizing the corresponding quadratic form as mentioned in the review section.

In our case, since f of the second equation is just the *identity* function, we have

$$\frac{\partial f}{\partial \beta_2} = \mathbf{A}_2.$$

Thus, the optimal choice of an IV for our second equation is

$$\tilde{\mathbf{G}} \equiv E(\mathbf{G}_0) = E(\mathbf{A}_2)$$

of which the best proxy is $\mathbf{A}_2^*$. And, the best IV estimator for β_2 can be obtained by minimizing the following quadratic form

$$\begin{aligned} \mathbf{S}_3(\beta_2 \mid \mathbf{W}_3) &= (\mathbf{Y}_2 - f(\mathbf{A}_2\beta_2))^{\mathrm{T}} \mathbf{A}_2^* \left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^*\right)^{-1} \mathbf{A}_2^{*\mathrm{T}} (\mathbf{Y}_2 - f(\mathbf{A}_2\beta_2)) \\ &= (\mathbf{Y}_2 - \mathbf{A}_2\beta_2)^{\mathrm{T}} \mathbf{A}_2^* \left(\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^*\right)^{-1} \mathbf{A}_2^{*\mathrm{T}} (\mathbf{Y}_2 - \mathbf{A}_2\beta_2) \end{aligned}$$

which reduces to $\mathbf{S}_2(\beta_2 \mid \mathbf{W}_2)$.

To summarize, we remark that for the second equation in our partially recursive two-equation GSiEM/GPA model (3.1) and (3.2), the IV $\mathbf{A}_2^*$ is the best choice for $\mathbf{A}_2$, in which the IV $\widehat{Y}_1^*$ is the best choice for Y_1, and our $\widehat{\beta}_{2,IV}$ is the optimal IV estimator in this setting.

6 SIMULATIONS

In the following two simulation studies, the estimation of the structural coefficients in a partially recursive Binomial-Normal GSiEM/GPA model (with logit-identity links) and a partially recursive Poisson-Normal GSiEM/GPA model (with log-identity links) are examined with the comparisons among various estimators including the ILS and 2SLS estimators developed in the previous paper.

6.1 A Strategy for Data Generation

Following Hsiao (1986, Sec. 5.4, pp. 112-125), we can think that the data of the *partially* recursive two-equation GSiEM/GPA model (3.1) and (3.2) are generated from the following two equations:

$$
\begin{aligned}
g_1(\mu_1) &= \beta_{10} + \beta_{1x_1}X_1 + \beta_{1x_2}X_2 + h \\
\mu_2 &= \beta_{20} + \beta_{2x_1}X_1 + \beta_{2x_3}X_3 + \gamma_{2y_1}Y_1 + \alpha h
\end{aligned}
$$

where the *latent* variable h is generated independently from a common distribution such as

$$h \sim \text{Normal}\,(0, 1)$$

for each subject i so that it is independent of all the independent variables X_1, X_2, and X_3. After the data are generated, the latent variable h is unknown to the data analyst. The chosen value of the *latent* coefficient α on the latent variable h in the second equation controls the degree of correlatedness between the random errors of Eqs. (3.1) and (3.2). Yet, the other terms in the above two equations remain the same as in the specification of the original partially recursive GSiEM/GPA model (3.1) and (3.2). By adding an extra "unobserved" latent variable h to the equations of a partially recursive GSiEM/GPA model, we find a feasible way to generating the data for simulations.

6.2 A Partially Recursive *Binomial*-Normal GSiEM/GPA Model

For simplicity, we specify the following partially recursive two-equation Binomial-Normal GSiEM/GPA model

$$
\begin{aligned}
\text{logit}(\mu_1) &= \beta_{10} + \beta_{11}X_1 \\
\mu_2 &= \beta_{20} + \beta_{22}X_2 + \gamma_{21}Y_1
\end{aligned}
$$

where the two response variables are

$$
\begin{aligned}
Y_1 &\sim \text{Binomial}\,(1, \mu_1), \\
Y_2 &\sim \text{Normal}\,(\mu_2, 1).
\end{aligned}
$$

And, the data are actually generated from the following two equations

$$\begin{aligned}\text{logit}(\mu_1) &= \beta_{11}X_1 + h,\\ \mu_2 &= 2X_2 - 2Y_1 + \alpha h\end{aligned}$$

where the independent variables X_1 and X_2, the "unobserved" latent variable h, and the error term e_2 of the second equation for Y_2 are generated independently from Normal $(0, 1)$. The true values of the coefficients are all explicitly listed in the above two equations except β_{11}. The value of the latent coefficient α is set to 2.0 for fixing the degree of the association between the random components of the two equations in the partially recursive GSiEM/GPA model. The sample sizes n are 50, 100, 250, 500, and 1000 in five separate simulations. And, 1000 repetitions are performed in each setting.

Specifically, we choose seven different values of β_{11}, which are 0.2, 0.4, 0.6, 0.8, 1.0, 1.5, and 2.0 respectively, to see their impacts on our ILS, 2SLS, IV-1, and IV-2 estimates of γ_{21} by looking at the mean and standard deviation of the sampling distribution of $\widehat{\gamma}_{21}$ from the 1000 repetitions. The major part of the simulation results are listed in **Table 3**. For an easy grasp of the vast outputs from the simulations, we subjectively consider the estimates which satisfy the following two criteria as being *good*:

- **Criterion A** (biasedness): $\Delta = |$ Mean of $\widehat{\gamma}_{21}$'s from the 1000 repetitions $- \gamma_{21} | < 0.1$.
- **Criterion B** (efficiency): Standard deviation of $\widehat{\gamma}_{21}$'s from the 1000 repetitions < 1.0.

In other words, an estimate is *not* good if its bias is too big or its variance is too large. Then, a summary of our findings is given below: All the ILS, 2SLS, IV-1, and IV-2 estimates are *good* when (1) $\beta_{11} = 0.4$ and sample size $n = 1000$, (2) $\beta_{11} = 0.6$ and sample size $n = 500$ or 1000, (3) $\beta_{11} = 0.8$ or 1.0 and sample size $n = 250$, 500, or 1000, and (4) $\beta_{11} = 1.5$ or 2.0 and sample size $n = 100$, 250, 500, or 1000. Therefore, we can see that not only the ILS, 2SLS, IV-1, and IV-2 estimators are all asymptotically unbiased and equally efficient, but, more importantly, their performances heavily depend on the value of β_{11} in the first equation since it determines the correlation between Y_1 and its proxy $\widehat{Y}_1$ to be used in the second equation.

6.3 A Partially Recursive *Poisson*-Normal GSiEM/GPA Model

For simplicity, we specify the following partially recursive two-equation Poisson-Normal GSiEM/GPA model

$$\begin{aligned} \log(\mu_1) &= \beta_{10} + \beta_{11} X_1 \\ \mu_2 &= \beta_{20} + \beta_{22} X_2 + \gamma_{21} Y_1 \end{aligned}$$

where the two response variables are

$$\begin{aligned} Y_1 &\sim \text{Poisson}\,(\mu_1), \\ Y_2 &\sim \text{Normal}\,(\mu_2, 1). \end{aligned}$$

And, the data are actually generated from the following two equations

$$\begin{aligned} \log(\mu_1) &= \beta_{11} X_1 + h, \\ \mu_2 &= 2X_2 - 2Y_1 + \alpha h \end{aligned}$$

where the independent variables X_1 and X_2, the "unobserved" latent variable h, and the error term e_2 of the second equation for Y_2 are generated independently from Normal $(0, 1)$. The true values of the coefficients are all explicitly listed in the above two equations except β_{11}. The value of the latent coefficient α is set to 2.0 for fixing the degree of the association between the random components of the two equations in the partially recursive GSiEM/GPA model. The sample sizes n are 50, 100, 250, 500, and 1000 in five separate simulations. And, 1000 repetitions are performed in each setting.

Specifically, we choose seven different values of β_{11}, which are 0.2, 0.4, 0.6, 0.8, 1.0, 1.5, and 2.0 respectively, to see their impacts on our ILS, 2SLS, IV-1, and IV-2 estimates of γ_{21} by looking at the mean and standard deviation of the sampling distribution of $\widehat{\gamma}_{21}$ from the 1000 repetitions. The major part of the simulation results are listed in **Table 4**. Again, for an easy grasp of the vast outputs from the simulations, we subjectively consider the estimates which satisfy the above-mentioned two criteria as being *good*. In other words, an estimate is *not* good if its bias is too big or its variance is too large. Then, a summary of our findings is given below: All the ILS, 2SLS, IV-1, and IV-2 estimates are *good* when (1) $\beta_{11} = 0.2$ and sample size $n =$ 1000, (2) $\beta_{11} = 0.4$ and sample size $n = 250$, 500, or 1000, (3) $\beta_{11} = 0.6$ and sample size $n = 100$, 250, 500, or 1000, and (4) $\beta_{11} = 0.8$, 1.0, 1.5, or 2.0 and sample size $n = 50$, 100, 250, 500, or 1000. Therefore, we can see that the performances of the ILS, 2SLS, IV-1, and IV-2 estimators again heavily depend on the value of β_{11} in the first equation since it determines the correlation between Y_1 and its proxy $\widehat{Y}_1$ to be used in the second equation. However, only the IV-1 and IV-2 estimators are asymptotically unbiased, whereas the ILS and 2SLS estimators are not in this situation. In fact, the biases of the ILS and 2SLS estimators, which may still be within the range of $\pm$ 0.1 (see **Criteria A**), depend on the value of β_{11} — they have shown a "U" shape with the bottom at the value of β_{11} around 0.8, which has puzzled us and deserves a further investigation. Moreover, the IV-1 and IV-2 estimators are surprisingly much more efficient than the ILS and 2SLS estimators across almost all good settings by comparing the standard deviations of their sampling distributions. Finally, we note that unlike the results listed in **Table 3**, the standard deviations of the ILS and 2SLS estimates from the 1000 repetitions do not necessarily decrease not only as the sample size n increases but also as the value of β_{11} increases. The first odd phenomenon has already been seen in the simulation study presented in the previous paper (see **Table 4** there). In contrast, the standard deviations of the IV-1 and IV-2 estimates from the 1000 repetitions show very reasonable results, which decrease as the sample size n increases and as the value of β_{11} increases.

7 COMPARISON: THE RELATIONSHIP BETWEEN THE 2SLS AND IV ESTIMATORS

It has been known in econometrics that the 2SLS estimator is a special case of the IV estimator in a linear SiEM/PA model. Yet, as seen in the above simulations, they can be quite different in our partially recursive two-equation GSiEM/GPA model. We shall explore the relationship between our 2SLS and IV estimators in this section. See **Table 1** of the previous paper for their formulas.

7.1 Proof

To prove analytically that $\widehat{\beta}_{2,2SLS} \neq \widehat{\beta}_{2,IV}$, we just need to show that

$$\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^* \neq \mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2.$$

First, we compute

$$\begin{aligned}
\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^* &= \begin{bmatrix} \mathbf{X}_1^{\mathrm{T}} \\ \mathbf{X}_3^{\mathrm{T}} \\ \widehat{\mathbf{Y}}_1^{*\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_3 & \widehat{\mathbf{Y}}_1^* \end{bmatrix} \\
&= \left[\begin{array}{cc|c} \mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1 & \mathbf{X}_1^{\mathrm{T}}\mathbf{X}_3 & \mathbf{X}_1^{\mathrm{T}}\widehat{\mathbf{Y}}_1^* \\ \mathbf{X}_3^{\mathrm{T}}\mathbf{X}_1 & \mathbf{X}_3^{\mathrm{T}}\mathbf{X}_3 & \mathbf{X}_3^{\mathrm{T}}\widehat{\mathbf{Y}}_1^* \\ \hline \widehat{\mathbf{Y}}_1^{*\mathrm{T}}\mathbf{X}_1 & \widehat{\mathbf{Y}}_1^{*\mathrm{T}}\mathbf{X}_3 & \widehat{\mathbf{Y}}_1^{*\mathrm{T}}\widehat{\mathbf{Y}}_1^* \end{array}\right] \\
&= \left[\begin{array}{c|c} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \hline \mathbf{B}_{21} & \mathbf{B}_{22} \end{array}\right].
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2 &= \begin{bmatrix} \mathbf{X}_1^{\mathrm{T}} \\ \mathbf{X}_3^{\mathrm{T}} \\ \widehat{\mathbf{Y}}_1^{*\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_3 & \mathbf{Y}_1 \end{bmatrix} \\
&= \left[\begin{array}{cc|c} \mathbf{X}_1^{\mathrm{T}}\mathbf{X}_1 & \mathbf{X}_1^{\mathrm{T}}\mathbf{X}_3 & \mathbf{X}_1^{\mathrm{T}}\mathbf{Y}_1 \\ \mathbf{X}_3^{\mathrm{T}}\mathbf{X}_1 & \mathbf{X}_3^{\mathrm{T}}\mathbf{X}_3 & \mathbf{X}_3^{\mathrm{T}}\mathbf{Y}_1 \\ \hline \widehat{\mathbf{Y}}_1^{*\mathrm{T}}\mathbf{X}_1 & \widehat{\mathbf{Y}}_1^{*\mathrm{T}}\mathbf{X}_3 & \widehat{\mathbf{Y}}_1^{*\mathrm{T}}\mathbf{Y}_1 \end{array}\right] \\
&= \left[\begin{array}{c|c} \mathbf{B}_{11} & \mathbf{B}_{12}^* \\ \hline \mathbf{B}_{21} & \mathbf{B}_{22}^* \end{array}\right].
\end{aligned}$$

Then, we must check if $\mathbf{B}_{12} = \mathbf{B}_{12}^*$ or, equivalently,

$$\begin{cases} \mathbf{X}_1^{\mathrm{T}}\left(\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^*\right) = 0 \\ \mathbf{X}_3^{\mathrm{T}}\left(\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^*\right) = 0 \end{cases}$$

and if $\mathbf{B}_{22} = \mathbf{B}_{22}^*$ or, equivalently,

$$\widehat{\mathbf{Y}}_1^{*\mathrm{T}}\left(\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^*\right) = 0.$$

Nonetheless, it is not clear how this can be done analytically without any further information. We simply note that the score equation for a GLM is of the form

$$\mathbf{D}^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\beta})) = \mathbf{0}$$

so that in our case,

$$\mathbf{D}_1^{\mathrm{T}}\mathbf{V}_1^{-1}\left(\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^*\right) \cong \mathbf{0}$$

when $\widehat{\boldsymbol{\beta}}_1$ converges. Then, we resort to the simulations with an attempt to falsify these equalities (by finding counterexamples) in the next subsection.

7.2 Numerical Result

To check on the above three equalities, we specify the following two GLMs for simulations.

- **Binomial Distribution:**

$$\text{logit}(\mu_1) = 3X_1 + h.$$

- **Poisson Distribution:**

$$\log(\mu_1) = 1X_1 + h.$$

The covariate X_1 and the latent variable h are generated independently from $\text{Normal}\,(0,1)$. The sample size is 500. And, 1000 repetitions are performed.

Then, we consider the following three situations.

1. When the distribution of Y_1 is *Normal* and the link function is *identity*, it is clear that

$$\mathbf{D}^{\mathrm{T}}\mathbf{V}^{-1} = \mathbf{A}_1^{\mathrm{T}} = \begin{bmatrix}\mathbf{1} & \mathbf{X}_1\end{bmatrix}^{\mathrm{T}}$$

so that

$$\begin{cases} \mathbf{1}^{\mathrm{T}}\left(\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^*\right) = 0, \\ \mathbf{X}_1^{\mathrm{T}}\left(\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^*\right) = 0. \end{cases}$$

And, we have

$$\begin{aligned} \widehat{\mathbf{Y}}_1^{*\mathrm{T}}\left(\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^*\right) &= \widehat{\mathbf{Y}}_1^{*\mathrm{T}}\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^{*\mathrm{T}}\widehat{\mathbf{Y}}_1^* \\ &= \widehat{\mathbf{Y}}_1^{*\mathrm{T}}\widehat{\mathbf{Y}}_1^* - \widehat{\mathbf{Y}}_1^{*\mathrm{T}}\widehat{\mathbf{Y}}_1^* \\ &= 0 \end{aligned}$$

due to the *idempotency* of the projection matrix. Therefore, the 2SLS and IV estimators are identical in this situation, which has been known for years.

2. When the distribution of Y_1 is *Binomial* and the link function is *logit*, it can be shown that

$$\mathbf{D}^{\mathrm{T}}\mathbf{V}^{-1} = \mathbf{A}_1^{\mathrm{T}} = \begin{bmatrix}\mathbf{1} & \mathbf{X}_1\end{bmatrix}^{\mathrm{T}}$$

so that

$$\begin{cases} \mathbf{1}^{\mathrm{T}}\left(\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^*\right) = 1.585 \times 10^{-3} \cong 0 \\ \mathbf{X}_1^{\mathrm{T}}\left(\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^*\right) = 4.465 \times 10^{-5} \cong 0 \end{cases}$$

on the convergence of the IRLS algorithm in our simulation. And, we find that

$$\widehat{\mathbf{Y}}_1^{*\mathrm{T}}\left(\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^*\right) = 3.206 \times 10^{-4} \cong 0$$

from our simulation. Therefore, the values of the 2SLS and IV estimators are very close to each other in this situation.

3. When the distribution of Y_1 is *Poisson* and the link function is *log*, it can also be shown that

$$\mathbf{D}^{\mathrm{T}}\mathbf{V}^{-1} = \mathbf{A}_1^{\mathrm{T}} = \begin{bmatrix} \mathbf{1} & \mathbf{X}_1 \end{bmatrix}^{\mathrm{T}}$$

so that

$$\begin{cases} \mathbf{1}^{\mathrm{T}} \left(\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^* \right) = -2.633 \times 10^{-2} \cong 0 \\ \mathbf{X}_1^{\mathrm{T}} \left(\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^* \right) = 3.810 \times 10^{-4} \cong 0 \end{cases}$$

on the convergence of the IRLS algorithm in our simulation. However, we find that

$$\widehat{\mathbf{Y}}_1^{*\mathrm{T}} \left(\mathbf{Y}_1 - \widehat{\mathbf{Y}}_1^* \right) = 279.196 \ncong 0$$

from our simulation. Therefore, the values of the 2SLS and IV estimators are different in this situation. This discrepancy ends the proof of this section.

8 DISCUSSION

8.1 Summary

In this study, we try to combine the estimation methods of SiEM and the IRLS algorithm of GLMs to develop suitable estimation methods for estimating the structural coefficients in a partially recursive GSiEM/GPA model especially with responses of a mixed type. Specifically, we have developed (1) the ILS and 2SLS estimators in the previous paper and (2) the IV estimator in this paper for a partially recursive two-equation GSiEM/GPA model, in which the first equation, Eq. (3.1), is a GLM and the second equation, Eq. (3.2), is a linear regression model. And, as mentioned in the beginning of this paper, the IV estimator is preferred in terms of asymptotic performance and generality. However, unlike the ILS and 2SLS estimators, it is *not* a straightforward task to apply the IV estimator to a partially recursive two-equation GSiEM/GPA model in which both equations are GLMs. When the second equation in a partially recursive GSiEM/GPA model is a GLM, some additional difficulties arise for the IV method because (1) the link function of a GLM makes it nonlinear and (2) the variance function of the response variable in a GLM usually depends on its mean. Thus, the IV estimator for a GLM derived by minimizing a quadratic form like the one defined in the review section for a nonlinear equation would be *inconsistent*. By applying the *theory of the estimating function*, we have developed a general IV for GLMs and will present the result in a forthcoming paper due to its general applicability in the error-in-variable problem and the simultaneous equations of GLMs. Then, with the general IV for GLMs, the multi-equation partially recursive GSiEM/GPA models are straightforward extensions of the two-equation cases, which can be solved either recursively or jointly. Again, up to now, we have not seen the need of adding any new or extra constraints to the current available rules for the model identification. See, for example, Greene (2000, Sec. 16.3, pp. 663-676) for details. Finally, we note that a real data set from a clinical study has been analyzed using our methods and reported in a separate paper.

8.2 Future Work

The ILS, 2SLS, and IV estimators that we have developed are all the single-equation methods. It will be a further challenge to work on the following problems: (1) How can the single-equation IV estimator be applied to a *non-recursive* GSiEM/GPA model? (2) And, can the system methods be developed for estimating the structural coefficients in a *partially recursive* or *non-recursive* GSiEM/GPA model to gain efficiency?

9 APPENDICES

9.1 Appendix 1: $\sqrt{n}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\epsilon_2\right) = \sqrt{n}\left(\frac{\sum_{i=1}^n \mathbf{A}_{i2}^*\epsilon_{i2}}{n}\right) \xrightarrow{d} \mathbf{N}_p\left(\mathbf{0}, \sigma_2^2\mathbf{Q}_{A_2^*A_2^*}\right)$.

We shall prove that

$$\sqrt{n}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\epsilon_2\right) = \sqrt{n}\left(\frac{\sum_{i=1}^n \mathbf{A}_{i2}^*\epsilon_{i2}}{n}\right) \xrightarrow{d} \mathbf{N}_p\left(0, \sigma_2^2\mathbf{Q}_{A_2^*A_2^*}\right)$$

where $\mathbf{N_p}$ is a p-variate Normal distribution, p is the column dimension of the design matrix $\mathbf{A}_2$, and $\mathbf{Q}_{A_2^* A_2^*} =$ plim $\left(\frac{1}{n}\mathbf{A}_2^{*T}\mathbf{A}_2^*\right)$.

Proof:

We know that $X_{i1} \perp \epsilon_{i2}$, $X_{i2} \perp \epsilon_{i2}$, $\widehat{Y}_{i1}^* = f(X_{i1}, X_{i2})$, and $\epsilon_{i2} \overset{iid}{\sim} N\left(0, \sigma^2\right)$, where i.i.d. stands for "independent and identically distributed." And, we assume

$$E\left(X_{i1}^{2v}\right) < \infty \qquad \text{and} \qquad E\left(X_{i2}^{2v}\right) < \infty$$

for $v = 1, 2, \cdots$. Now, we write

$$\sqrt{n}\left(\frac{\sum_{i=1}^n \mathbf{A}_{i2}^* \epsilon_{i2}}{n}\right) = \sqrt{n}\begin{bmatrix} \frac{\sum_{i=1}^n \epsilon_{i2}}{n} \\ \frac{\sum_{i=1}^n X_{i2}\epsilon_{i2}}{n} \\ \frac{\sum_{i=1}^n \widehat{Y}_{i1}^*\epsilon_{i2}}{n} \end{bmatrix} = \sqrt{n}\begin{bmatrix} (1) \\ (2) \\ (3) \end{bmatrix}.$$

Term (1):

$$\begin{aligned} \sum_{i=1}^n E\left|\epsilon_{i2} - E(\epsilon_{i2})\right|^4 &= \sum_{i=1}^n E\left(\epsilon_{i2}^4\right) \qquad (\text{because } E(\epsilon_{i2}) = 0) \\ &= \sum_{i=1}^n 3\sigma^4 \qquad \left(\text{because } \epsilon_{i2} \overset{iid}{\sim} N\left(0, \sigma^2\right)\right) \\ &= 3n\sigma^4 \end{aligned}$$

and

$$\begin{aligned} B_{1n}^2 &= \sum_{i=1}^n Var(\epsilon_{i2}) \\ &= \sum_{i=1}^n \sigma^2 \\ &= n\sigma^2 \end{aligned}$$

so that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n E\left|\epsilon_{i2} - E(\epsilon_{i2})\right|^4}{B_{1n}^4} = \lim_{n\to\infty} \frac{3n\sigma^4}{n^2\sigma^4} = 0.$$

Then, by Serfling (1980, Subsec. 1.9.2, Theorem A and Corollary, pp. 29-30), we obtain

$$\sqrt{n}\left(\frac{\sum_{i=1}^{n}\epsilon_{i2}}{n}\right) \xrightarrow{d} N\left(0,\sigma^2\right).$$

In fact, we can directly apply the *Lindeberg-Lévy CLT* for i.i.d. random variables (see, e.g., Serfling (1980, Subsec. 1.9.1, Theorem A, p. 28)) to get this result, but the above proof is illustrative for the proofs of **Terms (2)** and **(3)**.

Term (2):

$$\begin{aligned}\sum_{i=1}^{n} E\left|X_{i2}\epsilon_{i2}-E(X_{i2}\epsilon_{i2})\right|^4 &= \sum_{i=1}^{n} E\left(X_{i2}^4\epsilon_{i2}^4\right) \quad \text{(because } E(X_{i2}\epsilon_{i2})=0\text{)}\\ &= \sum_{i=1}^{n} E\left(X_{i2}^4\right)E\left(\epsilon_{i2}^4\right) \quad \text{(because } X_{i2}\perp\epsilon_{i2}\text{)}\\ &= 3\sigma^4\sum_{i=1}^{n} E\left(X_{i2}^4\right)\end{aligned}$$

and

$$\begin{aligned}B_{2n}^2 &= \sum_{i=1}^{n} Var(X_{i2}\epsilon_{i2})\\ &= \sum_{i=1}^{n} E\left(X_{i2}^2\epsilon_{i2}^2\right)\\ &= \sum_{i=1}^{n} E\left(X_{i2}^2\right)E\left(\epsilon_{i2}^2\right)\\ &= \sigma^2\sum_{i=1}^{n} E\left(X_{i2}^2\right)\end{aligned}$$

so that

$$\lim_{n\to\infty}\frac{\sum_{i=1}^{n} E\left|X_{i2}\epsilon_{i2}-E(X_{i2}\epsilon_{i2})\right|^4}{B_{2n}^4} = \lim_{n\to\infty}\frac{3\sigma^4\sum_{i=1}^{n} E\left(X_{i2}^4\right)}{\sigma^4\left[\sum_{i=1}^{n} E\left(X_{i2}^2\right)\right]^2} = 0$$

since the numerator and the denominator are $O(n)$ and $O(n^2)$ respectively. Then, by the same token, we obtain

$$\sqrt{n}\left(\frac{\sum_{i=1}^{n}X_{i2}\epsilon_{i2}}{n}\right) \xrightarrow{d} N\left(0, \lim_{n\to\infty}\left[\frac{\sum_{i=1}^{n} E\left(X_{i2}^2\right)}{n}\right]\sigma^2\right).$$

Term (3):

$$\begin{aligned}
\sum_{i=1}^{n} E\left|\widehat{Y}_{i1}^{*}\epsilon_{i2} - E\left(\widehat{Y}_{i1}^{*}\epsilon_{i2}\right)\right|^{4} &= \sum_{i=1}^{n} E\left(\widehat{Y}_{i1}^{*4}\epsilon_{i2}^{4}\right) \quad \left(\text{because } E\left(\widehat{Y}_{i1}^{*}\epsilon_{i2}\right)=0\right) \\
&= \sum_{i=1}^{n} E\left(\widehat{Y}_{i1}^{*4}\right) E\left(\epsilon_{i2}^{4}\right) \quad \left(\text{because } X_{i1} \perp \epsilon_{i2}, X_{i2} \perp \epsilon_{i2}, \text{ and } \widehat{Y}_{i1}^{*} = f(X_{i1}, X_{i2})\right) \\
&= 3\sigma^{4} \sum_{i=1}^{n} E\left(\widehat{Y}_{i1}^{*4}\right)
\end{aligned}$$

and

$$\begin{aligned}
B_{3n}^{2} &= \sum_{i=1}^{n} Var\left(\widehat{Y}_{i1}^{*}\epsilon_{i2}\right) \\
&= \sum_{i=1}^{n} E\left(\widehat{Y}_{i1}^{*2}\epsilon_{i2}^{2}\right) \\
&= \sum_{i=1}^{n} E\left(\widehat{Y}_{i1}^{*2}\right) E\left(\epsilon_{i2}^{2}\right) \\
&= \sigma^{2} \sum_{i=1}^{n} E\left(\widehat{Y}_{i1}^{*2}\right)
\end{aligned}$$

so that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} E\left|\widehat{Y}_{i1}^{*}\epsilon_{i2} - E\left(\widehat{Y}_{i1}^{*}\epsilon_{i2}\right)\right|^{4}}{B_{3n}^{4}} = \lim_{n\to\infty} \frac{3\sigma^{4}\sum_{i=1}^{n} E\left(\widehat{Y}_{i1}^{*4}\right)}{\sigma^{4}\left[\sum_{i=1}^{n} E\left(\widehat{Y}_{i1}^{*2}\right)\right]^{2}} = 0$$

since the numerator and the denominator are $O(n)$ and $O(n^2)$ respectively. Then, by the same token, we obtain

$$\sqrt{n}\left(\frac{\sum_{i=1}^{n} \widehat{Y}_{i1}^{*}\epsilon_{i2}}{n}\right) \xrightarrow{d} N\left(0, \lim_{n\to\infty}\left[\frac{\sum_{i=1}^{n} E\left(\widehat{Y}_{i1}^{*2}\right)}{n}\right]\sigma^{2}\right).$$

Moreover, for any real λ_1, λ_2, and λ_3 which are not all equal to zero,

$$\begin{aligned}
&\sum_{i=1}^{n} E\left|\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^{*}\right)\epsilon_{i2} - E\left[\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^{*}\right)\epsilon_{i2}\right]\right|^{4} \\
= &\sum_{i=1}^{n} E\left[\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^{*}\right)^{4}\epsilon_{i2}^{4}\right] \quad \left(\text{because } E\left[\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^{*}\right)\epsilon_{i2}\right]=0\right) \\
= &\sum_{i=1}^{n} E\left[\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^{*}\right)^{4}\right] E\left(\epsilon_{i2}^{4}\right) \quad \left(\text{because } X_{i1} \perp \epsilon_{i2}, X_{i2} \perp \epsilon_{i2}, \text{ and } \widehat{Y}_{i1}^{*} = f(X_{i1}, X_{i2})\right) \\
= &\ 3\sigma^{4}\sum_{i=1}^{n} E\left[\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^{*}\right)^{4}\right]
\end{aligned}$$

and

$$
\begin{aligned}
B_{4n}^2 &= \sum_{i=1}^{n} Var\left[\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^*\right)\epsilon_{i2}\right] \\
&= \sum_{i=1}^{n} E\left[\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^*\right)^2 \epsilon_{i2}^2\right] \\
&= \sum_{i=1}^{n} E\left[\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^*\right)^2\right] E\left(\epsilon_{i2}^2\right) \\
&= \sigma^2 \sum_{i=1}^{n} E\left[\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^*\right)^2\right]
\end{aligned}
$$

so that

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{\sum_{i=1}^{n} E\left|\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^*\right)\epsilon_{i2} - E\left[\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^*\right)\epsilon_{i2}\right]\right|^4}{B_{4n}^4} \\
&= \lim_{n\to\infty} \frac{3\sigma^4 \sum_{i=1}^{n} E\left[\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^*\right)^4\right]}{\sigma^4 \left\{\sum_{i=1}^{n} E\left[\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^*\right)^2\right]\right\}^2} \\
&= 0
\end{aligned}
$$

since the numerator and the denominator are $O(n)$ and $O(n^2)$ respectively. Thus, by the same token, we obtain

$$
\sqrt{n}\left(\frac{\sum_{i=1}^{n}\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^*\right)\epsilon_{i2}}{n}\right) \xrightarrow{d} N\left(0, \lim_{n\to\infty}\left\{\frac{\sum_{i=1}^{n} E\left[\left(\lambda_1 + \lambda_2 X_{i2} + \lambda_3 \widehat{Y}_{i1}^*\right)^2\right]}{n}\right\}\sigma^2\right).
$$

Hence, by the *Cramer-Wold device* (see, e.g., Serfling (1980, Subsec. 1.5.2, Theorem, p. 18)),

$$
\sqrt{n}\left(\frac{\sum_{i=1}^{n} \mathbf{A}_{i2}^* \epsilon_{i2}}{n}\right) \xrightarrow{d} \mathbf{N}_p\left(0, \sigma_2^2 \mathbf{Q}_{A_2^* A_2^*}\right)
$$

where $\mathbf{Q}_{A_2^* A_2^*} = \text{plim }\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2^*\right)$.

9.2 Appendix 2: $\text{plim }\left[\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\right] = \left[\text{plim }\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)\right]^{-1}$.

We shall show that

$$
\text{plim }\left[\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\right] = \left[\text{plim }\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)\right]^{-1}
$$

where $\mathbf{A}_2 = \begin{bmatrix} \mathbf{1} & \mathbf{X}_2 & \mathbf{Y}_1 \end{bmatrix}$ and $\mathbf{A}_2^* = \begin{bmatrix} \mathbf{1} & \mathbf{X}_2 & \widehat{\mathbf{Y}}_1^* \end{bmatrix}$.

Proof:

First, we have

$$\begin{aligned} \frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2 &= \frac{1}{n}\begin{bmatrix} \mathbf{1}^{\mathrm{T}} \\ \mathbf{X}_2^{\mathrm{T}} \\ \widehat{\mathbf{Y}}_1^{*\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{X}_2 & \mathbf{Y}_1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & \frac{\sum X_{i2}}{n} & \frac{\sum Y_{i1}}{n} \\ \frac{\sum X_{i2}}{n} & \frac{\sum X_{i2}^2}{n} & \frac{\sum X_{i2}Y_{i1}}{n} \\ \frac{\sum \widehat{Y}_{i1}^*}{n} & \frac{\sum X_{i2}\widehat{Y}_{i1}^*}{n} & \frac{\sum Y_{i1}\widehat{Y}_{i1}^*}{n} \end{bmatrix}. \end{aligned}$$

Then, we set

$$\operatorname{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right) = \operatorname{plim}\begin{bmatrix} 1 & \frac{\sum X_{i2}}{n} & \frac{\sum Y_{i1}}{n} \\ \frac{\sum X_{i2}}{n} & \frac{\sum X_{i2}^2}{n} & \frac{\sum X_{i2}Y_{i1}}{n} \\ \frac{\sum \widehat{Y}_{i1}^*}{n} & \frac{\sum X_{i2}\widehat{Y}_{i1}^*}{n} & \frac{\sum Y_{i1}\widehat{Y}_{i1}^*}{n} \end{bmatrix} = \begin{bmatrix} 1 & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$$

so that

$$\left[\operatorname{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)\right]^{-1} = \frac{1}{C_{A_2^*A_2}}\begin{bmatrix} c_{22}c_{33} - c_{32}c_{23} & c_{13}c_{32} - c_{12}c_{33} & c_{12}c_{23} - c_{13}c_{22} \\ c_{31}c_{23} - c_{21}c_{33} & c_{33} - c_{13}c_{31} & c_{21}c_{13} - c_{23} \\ c_{21}c_{32} - c_{31}c_{22} & c_{31}c_{12} - c_{32} & c_{22} - c_{21}c_{12} \end{bmatrix}$$

where

$$C_{A_2^*A_2} = \left|\operatorname{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)\right| \neq 0.$$

On the other hand, $\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}$ equals

$$\frac{1}{C^{\triangle}_{A_2^*A_2}}\begin{bmatrix} \frac{\sum X_{i2}^2}{n}\frac{\sum Y_{i1}\widehat{Y}_{i1}^*}{n} - \frac{\sum X_{i2}\widehat{Y}_{i1}^*}{n}\frac{\sum X_{i2}Y_{i1}}{n} & \frac{\sum X_{i2}\widehat{Y}_{i1}^*}{n}\frac{\sum Y_{i1}}{n} - \frac{\sum X_{i2}}{n}\frac{\sum Y_{i1}\widehat{Y}_{i1}^*}{n} & \frac{\sum X_{i2}}{n}\frac{\sum X_{i2}Y_{i1}}{n} - \frac{\sum X_{i2}^2}{n}\frac{\sum Y_{i1}}{n} \\ \frac{\sum \widehat{Y}_{i1}^*}{n}\frac{\sum X_{i2}Y_{i1}}{n} - \frac{\sum X_{i2}}{n}\frac{\sum Y_{i1}\widehat{Y}_{i1}^*}{n} & \frac{\sum Y_{i1}\widehat{Y}_{i1}^*}{n} - \frac{\sum Y_{i1}}{n}\frac{\sum \widehat{Y}_{i1}^*}{n} & \frac{\sum X_{i2}}{n}\frac{\sum Y_{i1}}{n} - \frac{\sum X_{i2}Y_{i1}}{n} \\ \frac{\sum X_{i2}}{n}\frac{\sum X_{i2}\widehat{Y}_{i1}^*}{n} - \frac{\sum X_{i2}^2}{n}\frac{\widehat{Y}_{i1}^*}{n} & \frac{\sum X_{i2}}{n}\frac{\sum \widehat{Y}_{i1}^*}{n} - \frac{\sum X_{i2}\widehat{Y}_{i1}^*}{n} & \frac{\sum X_{i2}^2}{n} - \frac{\sum X_{i2}}{n}\frac{\sum X_{i2}}{n} \end{bmatrix}$$

where

$$\begin{aligned} C^{\triangle}_{A_2^* A_2} &= \left|\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right| \\ &= \frac{\sum X_{i2}^2}{n}\frac{\sum Y_{i1}\widehat{Y}_{i1}^*}{n} + \frac{\sum X_{i2}}{n}\frac{\sum X_{i2}\widehat{Y}_{i1}^*}{n}\frac{\sum Y_{i1}}{n} + \frac{\sum X_{i2}}{n}\frac{\sum X_{i2}Y_{i1}}{n}\frac{\sum \widehat{Y}_{i1}^*}{n} - \\ &\quad \frac{\sum Y_{i1}}{n}\frac{\sum X_{i2}^2}{n}\frac{\sum \widehat{Y}_{i1}^*}{n} - \frac{\sum X_{i2}Y_{i1}}{n}\frac{\sum X_{i2}\widehat{Y}_{i1}^*}{n} - \frac{\sum Y_{i1}\widehat{Y}_{i1}^*}{n}\left(\frac{\sum X_{i2}}{n}\right)^2. \end{aligned}$$

Then, by the *Application D* of the *Corollary* in Section 1.7 of Serfling (1980, p. 26) regarding sums and products of random variables converging in probability, we have

$$\text{plim}\, C^{\triangle}_{A_2^* A_2} = \text{plim}\left|\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right| = C_{A_2^* A_2}$$

and

$$\text{plim}\left[\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\right] = \frac{1}{C_{A_2^* A_2}}\begin{bmatrix} c_{22}c_{33} - c_{32}c_{23} & c_{13}c_{32} - c_{12}c_{33} & c_{12}c_{23} - c_{13}c_{22} \\ c_{31}c_{23} - c_{21}c_{33} & c_{33} - c_{13}c_{31} & c_{21}c_{13} - c_{23} \\ c_{21}c_{32} - c_{31}c_{22} & c_{31}c_{12} - c_{32} & c_{22} - c_{21}c_{12} \end{bmatrix}.$$

Hence,

$$\text{plim}\left[\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)^{-1}\right] = \left[\text{plim}\left(\frac{1}{n}\mathbf{A}_2^{*\mathrm{T}}\mathbf{A}_2\right)\right]^{-1}.$$

10 REFERENCES

1. Amemiya, T. (1985), *Advanced Econometrics*, Cambridge, MA: Harvard University Press.

2. Bickel, P. J. and Doksum, K. A. (1977), *Mathematical Statistics: Basic Ideas and Selected Topics*, San Francisco, CA: Holden-Day.

3. Bowden, R. J. and Turkington, D. A. (1984), *Instrumental Variables*, Cambridge: Cambridge University Press.

4. Casella, G. and Berger, R. L. (1990), *Statistical Inference*, Belmont, CA: Duxbury Press.

5. Davidson, R. and Mackinnon, J. G. (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press.

6. Greene, W. H. (2000), *Econometric Analysis* (4th ed.), Upper Saddle River, NJ: Prentice-Hall.

7. Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge: Cambridge University Press.

8. Murphy, K. and Topel, R. (1985), "Estimation and Inference in Two Step Econometric Models," *Journal of Business and Economic Statistics*, 3, 370-379.

9. Rao, C. R. (1976), *Linear Statistical Inference and Its Applications* (2nd ed.), New York: Wiley.

10. Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

11. White, H. (1984), *Asymptotic Theory for Econometricians*, San Diego, CA: Academic Press.

11 TABLES

Table 1: The Single-Equation Instrumental Variable (IV) Estimators for a Linear Equation with (1) Homogenous and Independent ϵ_i and (2) Heteroscedastic or Dependent ϵ_i (without Transformations).

Var(ϵ_i)	Unweighted IV-OLS Analogs	Weighted IV-OLS Analogs
	[1] $Rank(\mathbf{Z}) \geq Rank(\mathbf{X})$	
I	$\widehat{\beta}_{IV} = (\mathbf{X}^T\mathbf{P}_1\mathbf{X})^{-1}\mathbf{X}^T\mathbf{P}_1\mathbf{Y}$, where $\mathbf{P}_1 = \mathbf{Z}\mathbf{Z}^T$. Bowden and Turkington (1984, p. 13).	$\widehat{\beta}_{IV} = (\mathbf{X}^T\mathbf{P}_2\mathbf{X})^{-1}\mathbf{X}^T\mathbf{P}_2\mathbf{Y}$, where $\mathbf{P}_2 = \mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T$. Bowden and Turkington (1984, p. 14) and White (1984, p. 9).
Σ	$\widehat{\beta}_{IV} = (\mathbf{X}^T\mathbf{P}_1\mathbf{X})^{-1}\mathbf{X}^T\mathbf{P}_1\mathbf{Y}$, where $\mathbf{P}_1 = \mathbf{Z}\mathbf{Z}^T$. Bowden and Turkington (1984, p. 13).	$\widehat{\beta}_{IV} = (\mathbf{X}^T\mathbf{P}_3\mathbf{X})^{-1}\mathbf{X}^T\mathbf{P}_3\mathbf{Y}$, where $\mathbf{P}_3 = \mathbf{Z}\left(\mathbf{Z}^T\mathbf{\Sigma}\mathbf{Z}\right)^{-1}\mathbf{Z}^T$. Bowden and Turkington (1984, p. 15) and Davidson and MacKinnon (1993, p. 663).
	[2] $Rank(\mathbf{Z}) = Rank(\mathbf{X})$	
I	$\widehat{\beta}_{IV} = (\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{Y}$. Bowden and Turkington (1984, p. 13).	$\widehat{\beta}_{IV} = (\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{Y}$. Bowden and Turkington (1984, p. 14).
Σ	$\widehat{\beta}_{IV} = (\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{Y}$. Bowden and Turkington (1984, p. 13).	$\widehat{\beta}_{IV} = (\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{Y}$. Bowden and Turkington (1984, p. 15).
	[3] $\mathbf{Z} = \mathbf{X}$	
I and **Σ**	$\widehat{\beta}_{IV} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$. Bowden and Turkington (1984, p. 70) and White (1984, p. 9).	$\widehat{\beta}_{IV} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$. Bowden and Turkington (1984, p. 70) and White (1984, p. 9).

1. The linear equation is specified as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$.
2. $\mathbf{Z}$ is the instrumental variable for $\mathbf{X}$.

Table 2: The Single-Equation Instrumental Variable (IV) Estimators for a Linear Equation with Heteroscedastic or Dependent ϵ_i (with Transformations).

$\mathbf{Var}(\epsilon_i)$	Unweighted IV-GLS Analogs	Weighted IV-GLS Analogs
	[1] $Rank(\mathbf{Z}) \geq Rank(\mathbf{X})$	
$\mathbf{\Sigma}$	$\widehat{\beta}_{IV1} = (\mathbf{X}^T\mathbf{P}_4\mathbf{X})^{-1}\mathbf{X}^T\mathbf{P}_4\mathbf{Y}$, where $\mathbf{P}_4 = \mathbf{\Sigma}^{-1}\mathbf{Z}\mathbf{Z}^T\mathbf{\Sigma}^{-1}$. $\widehat{\beta}_{IV2} = (\mathbf{X}^T\mathbf{P}_5\mathbf{X})^{-1}\mathbf{X}^T\mathbf{P}_5\mathbf{Y}$, where $\mathbf{P}_5 = \mathbf{C}^T\mathbf{Z}\mathbf{Z}^T\mathbf{C}$, (For example, $\mathbf{C} = \mathbf{\Sigma}^{-1/2}$.)	$\widehat{\beta}_{IV1} = (\mathbf{X}^T\mathbf{P}_6\mathbf{X})^{-1}\mathbf{X}^T\mathbf{P}_6\mathbf{Y}$, where $\mathbf{P}_6 = \mathbf{\Sigma}^{-1}\mathbf{Z}(\mathbf{Z}^T\mathbf{\Sigma}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{\Sigma}^{-1}$. $\widehat{\beta}_{IV2} = (\mathbf{X}^T\mathbf{P}_7\mathbf{X})^{-1}\mathbf{X}^T\mathbf{P}_7\mathbf{Y}$, where $\mathbf{P}_7 = \mathbf{C}^T\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{C}$. Bowden and Turkington (1984, p. 70) and White (1984, p. 98).
	[2] $Rank(\mathbf{Z}) = Rank(\mathbf{X})$	
$\mathbf{\Sigma}$	$\widehat{\beta}_{IV1} = (\mathbf{Z}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{\Sigma}^{-1}\mathbf{Y}$. $\widehat{\beta}_{IV2} = (\mathbf{Z}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{C}\mathbf{Y}$.	$\widehat{\beta}_{IV1} = (\mathbf{Z}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{\Sigma}^{-1}\mathbf{Y}$. $\widehat{\beta}_{IV2} = (\mathbf{Z}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{C}\mathbf{Y}$. Bowden and Turkington (1984, p. 71).
	[3] $\mathbf{Z} = \mathbf{X}$	
$\mathbf{\Sigma}$	$\widehat{\beta}_{IV1} = (\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{Y}$. $\widehat{\beta}_{IV2} = (\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{C}\mathbf{Y}$.	$\widehat{\beta}_{IV1} = (\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{Y}$. $\widehat{\beta}_{IV2} = (\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{C}\mathbf{Y}$. Bowden and Turkington (1984, p. 71).

1. The linear equation is specified as $\mathbf{Y} = \mathbf{X}\beta + \epsilon$.
2. $\mathbf{Z}$ is the instrumental variable for $\mathbf{X}$.
3. In this table, "IV1" is $\mathbf{Z}^* = \mathbf{C}\mathbf{Z}$ and "IV2" is $\mathbf{Z}$.

Table 3: The Means and Standard Deviations of the Four Estimates of the Coefficient γ_{21} (= -2.0) from 1000 Repetitions for the *Binomial*-Normal GSiEM/GPA Model at α = 2.0.

n	β_{11}	Mean				Standard Deviation			
		ILS	2SLS	IV-1	IV-2	ILS	2SLS	IV-1	IV-2
50	0.2	0.6220	-0.3604	2.1040	-2.3562	122.0623	44.6926	59.8189	59.7691
	0.4	-2.7037	1.2227	2.5394	4.8437	61.2024	73.8298	171.6641	173.0208
	0.6	-0.5021	-3.1253	-4.9218	-2.5843	72.0780	39.4788	102.8768	16.9492
	0.8	-2.3544	-2.2051	-1.7302	-2.2360	5.9763	5.3581	16.8234	6.1378
	1.0	-2.3858	-2.2326	-2.2561	-2.2875	4.7537	3.5737	3.5158	3.8961
	1.5	-2.1370	-2.1456	-2.1666	-2.1567	1.7247	1.9873	2.0318	2.0468
	2.0	-2.0000	-1.9845	-2.0133	-1.9951	1.1893	1.1883	1.1992	1.1926
100	0.2	-0.7737	-1.7080	-7.4731	4.1729	65.4209	40.8639	156.1715	131.8802
	0.4	-2.0060	-1.2719	-1.1144	-1.7986	19.3989	20.3911	41.7691	15.1956
	0.6	-2.2241	-2.4977	-2.4614	-2.8410	5.7528	5.1915	5.7290	15.4473
	0.8	-2.6559	-2.2763	-2.2720	-2.2263	16.0896	4.5728	4.5519	3.8566
	1.0	-2.2723	-2.1388	-2.1428	-2.1130	6.9785	2.9252	2.8641	2.1827
	1.5	-2.0348	-2.0297	-2.0451	-2.0379	0.9784	0.9814	0.9891	0.9883
	2.0	-1.9958	-1.9920	-2.0040	-1.9997	0.8185	0.8163	0.8260	0.8233
250	0.2	-0.1197	-0.4207	-1.5144	-4.5752	84.3684	47.9877	40.4075	52.6492
	0.4	-2.3474	-2.5767	-2.5737	-2.7678	14.1987	12.7347	12.9134	9.7287
	0.6	-2.1255	-2.1195	-2.1217	-2.1201	1.4654	1.4454	1.4463	1.4451
	0.8	-1.9951	-1.9942	-1.9959	-1.9965	0.9815	0.9792	0.9810	0.9807
	1.0	-1.9973	-1.9964	-2.0012	-1.9989	0.8027	0.8047	0.8062	0.8065
	1.5	-1.9991	-1.9985	-2.0041	-2.0029	0.5824	0.5823	0.5859	0.5849
	2.0	-2.0175	-2.0155	-2.0227	-2.0201	0.5135	0.5132	0.5149	0.5144
500	0.2	-1.8226	-2.1636	-2.1788	-1.9409	23.0826	14.3063	14.1890	16.9720
	0.4	-2.1447	-2.1438	-2.1447	-2.1440	1.4818	1.4798	1.4797	1.4793
	0.6	-2.0799	-2.0804	-2.0813	-2.0810	0.8782	0.8800	0.8809	0.8807
	0.8	-2.0346	-2.0341	-2.0366	-2.0359	0.6833	0.6830	0.6838	0.6835
	1.0	-2.0340	-2.0327	-2.0348	-2.0335	0.5684	0.5670	0.5674	0.5665
	1.5	-2.0153	-2.0150	-2.0193	-2.0185	0.4193	0.4192	0.4199	0.4197
	2.0	-2.0169	-2.0154	-2.0215	-2.0202	0.3467	0.3464	0.3478	0.3477
1000	0.2	-1.5910	-2.5351	-2.5399	-2.7862	24.8714	6.9869	7.0859	13.6898
	0.4	-2.0821	-2.0830	-2.0831	-2.0830	0.9418	0.9441	0.9439	0.9439
	0.6	-1.9947	-1.9940	-1.9941	-1.9940	0.6259	0.6249	0.6251	0.6253
	0.8	-1.9963	-1.9960	-1.9969	-1.9966	0.4771	0.4772	0.4770	0.4770
	1.0	-2.0010	-2.0007	-2.0023	-2.0019	0.3953	0.3954	0.3954	0.3953
	1.5	-2.0063	-2.0057	-2.0086	-2.0082	0.2814	0.2812	0.2817	0.2816
	2.0	-2.0041	-2.0036	-2.0080	-2.0074	0.2401	0.2400	0.2404	0.2401

Table 4: The Means and Standard Deviations of the Four Estimates of the Coefficient γ_{21} (= -2.0) from 1000 Repetitions for the *Poisson*-Normal GSiEM/GPA Model at α = 2.0.

n	β_{11}	Mean				Standard Deviation			
		ILS	2SLS	IV-1	IV-2	ILS	2SLS	IV-1	IV-2
50	0.2	-1.6211	-1.7040	-4.7418	-2.0372	49.9497	7.3601	78.4194	7.3625
	0.4	-2.6557	-2.1114	-2.8568	-1.9522	7.7178	2.4277	16.7946	9.0721
	0.6	-1.9258	-1.9596	-1.9936	-1.9235	4.9979	3.8789	3.9265	7.5392
	0.8	-2.0126	-2.0270	-2.1007	-2.0584	0.4143	0.4014	0.9186	0.3559
	1.0	-1.9641	-1.9941	-2.0541	-2.0415	0.3142	0.3224	0.2027	0.1969
	1.5	-1.9472	-1.9837	-2.0230	-2.0159	0.2604	0.2594	0.0888	0.0844
	2.0	-1.9192	-1.9666	-2.0119	-2.0084	0.2541	0.2359	0.0388	0.0351
100	0.2	-1.8489	-1.8147	-1.9319	-1.8583	5.1869	3.4710	4.6789	8.5031
	0.4	-2.1094	-2.1055	-2.1093	-2.1060	1.4952	0.6457	0.8686	0.7029
	0.6	-2.0235	-2.0421	-2.0667	-2.0579	0.3218	0.3403	0.2794	0.2815
	0.8	-1.9840	-2.0049	-2.0377	-2.0303	0.2717	0.2754	0.1632	0.1589
	1.0	-1.9790	-2.0062	-2.0281	-2.0214	0.2597	0.2681	0.1191	0.1145
	1.5	-1.9429	-1.9801	-2.0112	-2.0086	0.2777	0.2732	0.0453	0.0440
	2.0	-1.9051	-1.9530	-2.0040	-2.0026	0.2722	0.2600	0.0168	0.0161
250	0.2	-2.9145	-2.1028	-2.3145	-2.0623	26.1916	2.6276	5.7663	2.6912
	0.4	-2.0344	-2.0470	-2.0432	-2.0365	0.2895	0.3029	0.2573	0.2539
	0.6	-2.0069	-2.0175	-2.0224	-2.0187	0.2228	0.2314	0.1332	0.1326
	0.8	-1.9833	-2.0016	-2.0097	-2.0066	0.2250	0.2366	0.0805	0.0791
	1.0	-1.9516	-1.9780	-2.0099	-2.0072	0.2564	0.2676	0.0516	0.0508
	1.5	-1.9109	-1.9462	-2.0050	-2.0037	0.2782	0.2736	0.0194	0.0190
	2.0	-1.8908	-1.9318	-2.0020	-2.0015	0.3005	0.2853	0.0071	0.0068
500	0.2	-2.1511	-2.1489	-2.1596	-2.3568	1.2985	1.2771	1.3289	7.0380
	0.4	-2.0150	-2.0207	-2.0226	-2.0205	0.1853	0.1928	0.1499	0.1492
	0.6	-1.9996	-2.0053	-2.0133	-2.0117	0.1674	0.1739	0.0868	0.0865
	0.8	-1.9917	-2.0050	-2.0089	-2.0070	0.1938	0.2017	0.0538	0.0533
	1.0	-1.9612	-1.9763	-2.0057	-2.0042	0.2402	0.2432	0.0346	0.0343
	1.5	-1.9427	-1.9627	-2.0020	-2.0014	0.2970	0.2914	0.0111	0.0110
	2.0	-1.8933	-1.9321	-2.0011	-2.0008	0.3044	0.2923	0.0040	0.0039
1000	0.2	-2.0599	-2.0629	-2.0586	-2.0567	0.3208	0.3216	0.3155	0.3171
	0.4	-2.0079	-2.0109	-2.0088	-2.0075	0.1336	0.1374	0.0990	0.0989
	0.6	-1.9992	-2.0041	-2.0015	-2.0004	0.1454	0.1495	0.0566	0.0562
	0.8	-1.9954	-2.0044	-2.0039	-2.0031	0.1798	0.1911	0.0359	0.0359
	1.0	-1.9771	-1.9880	-2.0021	-2.0014	0.2174	0.2214	0.0222	0.0220
	1.5	-1.9478	-1.9688	-2.0013	-2.0010	0.3149	0.3171	0.0070	0.0068
	2.0	-1.9060	-1.9399	-2.0004	-2.0003	0.3188	0.3179	0.0020	0.0019