SSE-DP-2023-4

An Asymptotically Optimal Two-Sample IV Estimation with Many Instruments

Naoto Kunitomo

(The Institute of Statistical Mathematics)

and

Ryota Yuasa

(The Institute of Statistical Mathematics)

June 2023

An Asymptotically Optimal Two-Sample IV Estimation with Many Instruments *

Naoto Kunitomo †
and
Ryota Yuasa ‡

June 17, 2023

Abstract

We consider the statistical estimation of the coefficients of a linear structural equation in a simultaneous equation system when we use two-sample data and there are many instrumental variables. We derive some asymptotic properties of the Two-Sample Least Variance Ratio (2SLVR) estimator, which is an extension of the limited information maximum likelihood (LIML) estimator in one-sample, when we have two-sample data with many insrumental variables. It has been known that there is a non-negligible bias in the one-sample two stage least squares (TSLS) estimator and the generalized moment method (GMM), which are widely used in practice. They often lose even consistency when we have many instruments. We have found that the variance-covariance matrix of the limiting distribution of the 2SLVR estimator and its modifications often attain the asymptotic lower bound when the number of instruments is large and the disturbance terms are not necessarily normally distributed. The results would be useful for applications in econometrics and biometrics including Mendelian Randomization (RM) using DNA data analysis.

Key Words

Structural Equation, Instrumental Variables (IV), Many Instruments, Two-Sample Least Variance Ratio (2SLVR) Estimation, Asymptotic Optimality, Econometrics, Mendelian Randomization (MR), DNA Data Analysis.

---

*KY2023-6-17. This is a preliminary memorandum. The research has been supported by a project of Consortium for training experts in statistical sciences at ISM and it has been also supported by JSPS-Grant 22K01428.

†Institute of Statistical Mathematics

‡Institute of Statistical Mathematics

1. Introduction

There has been a growing interest and research on the instrumental variables (IV) estimation method of a single structural equation when the number of instruments (the number of instrumental variables excluded from the structural equation), say K_2 (or K_{2n}), is large relative to the sample size n. The relevance of such models is due to collection of large data sets and the development of computational equipment capable of analysis of such data sets in econometrics and biometrics including Mendelian Randomization (MR). Traditionally, many econometricians have been interested in the instrumental variables methods, but recently there have been many applications reported in the DNA data analysis of MR particularly. (Burgess and Thomspon (2020), Zhao, Q., J. Wang, G. Hemani, J. Bowden, and D. Small (2020), Ye, T., J. Shao and H. Kang (2021), for instance.)

In recent statistical applications, we sometimes cannot obtain the full data sets of endogenous variables and instrumental variables as one-sample, but we can obtain two-sample data set to investigate statistical causal relationships among endogenous variables. Then several estimation methods have been proposed. Among them, the two-sample two-stage least squares (2STSLS) method has been sometimes used since Inoue and Solon (2010) in econometrics. However, it has been known that in one-sample standard case the bias of the TSLS estimation is not negligible when there are many instruments (see Anderson, Kunitomo and Sawa (1982), and Anderson, Kunitomo and Matsushita (AKM, 2010), for instance). In the one sample case, AKM (2010) have investigated the asymptotic properties of some instrumental variables methods and shown the asymptotic efficiency of the LIML estimation when there are many instruments.

The main purpose of the present paper is to show that one method, the two-sample least variance ratio (2SLVR) estimator, has an optimum property when both the sample size n and K_2, which is the number of excluded instrumental variables. The 2SLVR estimator is a modified version of the one-sample LIML estimator to the two-sample situation. As a background of the statistical problem, we state and derive some asymptotic distributions of the 2LVR estimator and the 2TSLS estimator as $K_{2n} \to \infty$ and $n \to \infty$ while $K_{2n} < n$. Some of these results are improvements on AKM (2010) and earier studies in econometrics and they are presented in a unified notation.

In addition to the 2SLVR and 2STSLS estimators there are other instrumental variables (IV) methods. See Kunitomo (1980), Morimune (1983), Anderson, Kunitomo, and Morimune (1986) on the earlier studies of the finite sample properties of alternative (one-sample) estimation methods, for instance. Several semiparametric estimation methods have been developed including the generalized method of moments (GMM) estimation and the empirical likelihood (EL) method. (See Hayashi (2000) or Angrist and Pishke (2009), for instance.) However, it has been recently recognized that the classical methods have some advantages in some situations with many instruments.

In this paper, we shall give the results on the asymptotic properties of the 2SLVR estimator when the number of instruments is large and we develop *the large-K_2 asymptotic theory* or *the many instruments asymptotics*. The 2STSLS estimator is badly biased and it loses even consistency in some situation. Our results on the asymptotic properties and optimality of the 2SLVR estimator and its variants give a new interpretation on the numerical information of the finite sample properties, A guidance on the use of alternative estimation methods is provided for the use of structural equation model with *many instruments* in micro-econometric models and RM in biometrics. There is a number of related studies both in econometrics and biometrics. Some examples in econometrics were Donald and Newey (2001), Stock and Yogo (2005), Chao and Swanson (2005, 2006), and Hansen et al. (2008).

In Section 2, we state the formulation of a linear structural model and the two-sample alternative estimation methods of unknown structural parameters with possibly many instruments. In Section 3, we develop the large-K_2 asymptotics (or *many instruments* asymptotics) and give some results on the asymptotic normality of the 2SLVR estimator when n and K_2 are large. Then we shall present some results on the asymptotic optimality of the 2SLVR estimation in the sense that it attains the lower bound of the asymptotic variance in a class of consistent estimators with *many instruments* under reasonable assumptions. Also we discuss some issues such as a finite sample correction, the case when there are included instruments, the case when we have different sample sizes, and the problem of weak instruments in Section 4. In Section 5, we give some numerical results and show that the asymptotic results in Section 3 agree with the finite sample properties of estimators. Concluding remarks will be given in Section 6. The derivations of theorems are given in the Appendix.

2. Alternative Two-Sample IV Estimation Methods in Structural Equation Models with Many Instruments

We investigate the estimation problem of a structural equation in the classical linear simultaneous equation framework when we have two-sample data. Let a single linear structural equation be

$$y_{1i} = \boldsymbol{\beta}_2' \boldsymbol{y}_{2i} + \boldsymbol{\gamma}_1' \boldsymbol{z}_{1i} + u_i \quad (i = 1, \cdots, n), \tag{2.1}$$

where y_{1i} and $\boldsymbol{y}_{2i}$ are a scalar and a vector of G_2 endogenous variables, $\boldsymbol{z}_{1i}$ is a vector of K_1 (included) exogenous variables in (2.1), $\boldsymbol{\gamma}_1$ and $\boldsymbol{\beta}_2$ are $K_1 \times 1$ and $G_2 \times 1$ vectors of unknown parameters, and $u_1, \cdots, u_n$ are independent disturbance terms with $\mathbf{E}[u_i | \mathbf{z}_{1i}, \mathbf{z}_{2n,i}] = 0$ and $\mathbf{E}[u_i^2 | \mathbf{z}_{1i}, \mathbf{z}_{2n,i}] = \sigma_*^2$ $(i = 1, \cdots, n)$, where $\mathbf{z}_{2n,i}$ ($K_{2n} \times 1$ vector) is excluded from (2.1), but it is in the set of instrumental variables (IV) to explain the endogenous variables y_{1i} and $\mathbf{y}_{2i}$ in the simultaneous equation system.

The structural equation of (2.1) is not a regression because the endogenous variables y_{1i} and $\mathbf{y}_{2i}$ are correlated with the noise term and they are (probably linear) jointly explained by a set of included exogenous variables $\mathbf{z}_{1i}$, and the instrumental variables $\mathbf{z}_{2n,i}$, which are excluded from (2.1). An interesting feature of the present problem is the situation when there are many instrumental variables called *many instruments* in econometrics available as data and thus we use the notation $\mathbf{z}_{2n,i}$ $(i = 1, \cdots, n)$ depending on n.

The main interest of this paper is the estimation of structural coefficient vector (say, $\boldsymbol{\beta}_2$), but the data set of endogenous variables $\boldsymbol{y}_i^{'} = (y_{1i}, \boldsymbol{y}_{2i}^{'})^{'}$ for individual i is not available and we have y_{1i} $(i = 1, \cdots, n_1)$ and $\boldsymbol{y}_{2j}$ $(j = 1, \cdots, n_2)$ in different samples. In addition to endogenous variables, however, we have a large number of data sets of the instrumental variables in two-sample data and these instrumental variables are two samples on the same set of explanatory variables. This makes the present problem different from the traditional structural equation estimation and the instrumental variables (IV) methods.
In particular. we consider the situation that in the first sample we have the observations of $y_{1i}^{(1)}$, $\mathbf{z}_{1i}^{(1)}$ ($K_1 \times 1$ vector) and a $K_{2n} \times 1$ vector $\mathbf{z}_{2n,i}^{(1)}$ $(i = 1, \cdots, n_1)$, and in the second sample we have the observations of a $G_2 \times 1$ vector $\mathbf{y}_{2j}^{(2)}$, $\mathbf{z}_{1j}^{(2)}$ ($K_1 \times 1$ vector) and a $K_{2n} \times 1$ vector $\mathbf{z}_{2n,j}^{(2)}$ $(j = 1, \cdots, n_2)$. We investigate the estimation problem of the vector $\boldsymbol{\beta}_2$, when the reduced form of the endogenous variables $\boldsymbol{y}_i^{'} = (y_{1i}^{(1)}, \boldsymbol{y}_{2i}^{(2)'})^{'}$ in the two-sample data is given by

$$
(2.2) \qquad \mathbf{y}_1^{(1)} = (\mathbf{Z}_1^{(1)}, \mathbf{Z}_{2n}^{(1)})\boldsymbol{\Pi}_{1n} + \mathbf{v}_1 \ , \ \mathbf{Y}_2^{(2)} = (\mathbf{Z}_1^{(2)}, \mathbf{Z}_{2n}^{(2)})\boldsymbol{\Pi}_{2n} + \mathbf{V}_2 \ ,
$$

where $\mathbf{y}_1^{(1)}$ and $\mathbf{Y}_2^{(2)}$ are $n_1 \times 1$ vector and $n_2 \times G_2$ matrix of endogenous variables, $(\mathbf{Z}_1^{(k)}, \mathbf{Z}_{2n}^{(k)})$ $(k = 1, 2)$ are the $n_k \times (K_1 + K_{2n})$ matrices of vectors of instrumental variables, $\mathbf{v}_1$ $(= (v_{1i}))$ and $\mathbf{V}_2$ $(= (\mathbf{v}_{2i}^{'}))$ are the $n_1 \times 1$ vector and $n_2 \times G_2$ matrix of disturbances,

$$
\boldsymbol{\Pi}_{1n} = \begin{pmatrix} \boldsymbol{\pi}_{11} \\ \boldsymbol{\pi}_{21}^{(n)} \end{pmatrix} , \boldsymbol{\Pi}_{2n} = \begin{pmatrix} \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{22}^{(n)} \end{pmatrix}
$$

is the $(K_1 + K_{2n}) \times (1 + G_2)$ matrix of coefficients, and we assume that the coefficients $(\boldsymbol{\Pi}_{1n}, \boldsymbol{\Pi}_{2n})$ are the same as for $(y_{1i}, \mathbf{y}_{2i}^{'})^{'}$ in (2.1) and $(y_{1i}^{(k)}, \mathbf{y}_{2i}^{(k)})$ $(k = 1, 2)$ in (2.2). The disturbance term v_{1i} has $\mathbf{E}[v_{1i}] = 0$ and $\mathbf{E}[v_{1i}^2] = \omega_{11}$ $(i = 1, \cdots, n_1)$, and the disturbance vector $\mathbf{v}_{2j}$ has $\mathbf{E}[\mathbf{v}_{2j}] = \mathbf{0}$ and $\mathbf{E}[\mathbf{v}_{2j}\mathbf{v}_{2j}^{'}] = \boldsymbol{\Omega}_{22}$ $(j = 1, \cdots, n_2)$. Because the data sources of two-sample are ususally different, we assume that v_{1i} and $\mathbf{v}_{2j}$ are mutually independent. The vector of K_n $(= K_1 + K_{2n})$ instrumental variables $\mathbf{z}_{1i}^{(k)}, \mathbf{z}_{2n,i}^{(k)}$ $(k = 1, 2)$ satisfies the orthogonality condition $\mathbf{E}[v_{1i} | \boldsymbol{z}_{1i}^{(1)}, \mathbf{z}_{2n,i}^{(1)}] = 0$ $(i = 1, \cdots, n_1)$ and $\mathbf{E}[\mathbf{v}_{2j} | \boldsymbol{z}_{1j}^{(2)}, \mathbf{z}_{2n,j}^{(2)}] = \mathbf{0}$ $(j = 1, \cdots, n_2)$.
In Sections 2 and 3, we assume that the data size of two sample is the same, that is, $n_1 = n_2 = n$ and $n > K_1 + K_{2n}$ $(n > 2)$. We shall discuss the case when $n_1 \neq n_2$

in Section 4.
To remove the effects of included instrumental variables $\mathbf{Z}_1^{(k)}$ $(k=1,2)$, we define $\mathbf{Z}_{2n}^{(k)*}=\mathbf{Z}_{2n}^{(k)}-\mathbf{P}_{Z_1^{(k)}}\mathbf{Z}_{2n}^{(k)}=\bar{\mathbf{P}}_{Z_1^{(k)}}\mathbf{Z}_{2n}^{(k)}$ and the projection matrices $\mathbf{P}_{Z_1^{(k)}}=\mathbf{Z}_1^{(k)}(\mathbf{Z}_1^{(k)'}\mathbf{Z}_1^{(k)})^{-1}\mathbf{Z}_1^{(k)'}$, $\bar{\mathbf{P}}_{Z_1^{(k)}}=\mathbf{I}_n-\mathbf{P}_{Z_1^{(k)}}$ $(k=1,2)$. The estimation of the reduced form coefficients in (2.2) can be done by using
(2.3)

$$[\hat{\boldsymbol{\pi}}_{21},\hat{\boldsymbol{\Pi}}_{22}]=[(\mathbf{Z}_{2n}^{(1)'}\bar{\mathbf{P}}_{Z_1^{(1)}}\mathbf{Z}_{2n}^{(1)})^{-1}\mathbf{Z}_{2n}^{(1)'}\bar{\mathbf{P}}_{Z_1^{(1)}}\mathbf{y}_1^{(1)},(\mathbf{Z}_{2n}^{(2)'}\bar{\mathbf{P}}_{Z_1^{(2)}}\mathbf{Z}_{2n}^{(2)})^{-1}\mathbf{Z}_{2n}^{(2)'}\bar{\mathbf{P}}_{Z_1^{(2)}}\mathbf{Y}_2^{(2)}]\ .$$

We impose the condition that that there exists a $G_2\times 1$ (unknown) structural parameter vector $\boldsymbol{\beta}_2$ such as

$$\begin{pmatrix}\boldsymbol{\pi}_{11} & \boldsymbol{\Pi}_{12}\\ \boldsymbol{\pi}_{21}^{(n)} & \boldsymbol{\Pi}_{22}^{(n)}\end{pmatrix}\begin{pmatrix}1\\ -\boldsymbol{\beta}_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{\gamma}_1\\ \mathbf{0}\end{pmatrix}\ , \tag{2.4}$$

where $\boldsymbol{\gamma}_1$ $(K_1\times 1)$ is defined by the upper-part of (2.4).
It may be natural to use the lower part of the relation in (2.4) and (2.3) to estimate the structural coefficients $\boldsymbol{\beta}^{'}=(1,-\boldsymbol{\beta}_2^{'})$. (We have taken 1 as the first component of $\boldsymbol{\beta}$ as normalization.)

A Remark on Two-Sample Model : For the simplicity, we take the case when $K_1=0$ and $n_1=n_2=n$. In the first sample, we have the regression relation with the estimated coefficient vector $\hat{\boldsymbol{\Pi}}_{1n}$ of instruments while in the second sample we have the regression relation with the estimated coefficient matrix $\hat{\boldsymbol{\Pi}}_{2n}$ of instruments. The expected values of $\hat{\boldsymbol{\pi}}_{21}^{(n)}$ and $\hat{\boldsymbol{\Pi}}_{22}^{(n)}$ are $\boldsymbol{\pi}_{21}^{(n)}$ and $\boldsymbol{\Pi}_{22}^{(n)}$, respectively. If there exists a linear statistical relation between unknown regression coefficients $\boldsymbol{\pi}_{21}^{(n)}$ and $\boldsymbol{\Pi}_{22}^{(n)}$, we have $\boldsymbol{\beta}_2$ ($G_2\times 1$ unknown non-zero vector) such that $\boldsymbol{\pi}_{21}^{(n)}=\boldsymbol{\Pi}_{22}^{(n)}\boldsymbol{\beta}_2$. Since $\hat{\boldsymbol{\pi}}_{21}^{(n)}$ $(K_{2n}\times 1)$ and $\hat{\boldsymbol{\Pi}}_{22}^{(n)}$ $(K_{2n}\times G_2)$ are estimated from two-sample data and random, we represent as

$$\hat{\boldsymbol{\pi}}_{21}^{(n)}=\boldsymbol{\pi}_{21}^{(n)}+\mathbf{x}_1^{(n)}\ ,\ \hat{\boldsymbol{\Pi}}_{22}^{(n)}=\boldsymbol{\Pi}_{22}^{(n)}+\mathbf{X}_2^{(n)}\ , \tag{2.5}$$

where $(\mathbf{x}_1^{(n)},\mathbf{X}_2^{(n)})$ is a $K_{2n}\times(1+G_2)$ random matrix with zero means and heteroskedasticity. We need to consider the sampling variation both in $\mathbf{x}_1^{(n)}$ and $\mathbf{X}_2^{(n)}$ to estimate the true structural relationship of $\boldsymbol{\pi}_{21}^{(n)}=\boldsymbol{\Pi}_{22}^{(n)}\boldsymbol{\beta}_2$.
In one-sample case, we take $\mathbf{y}_1=\mathbf{y}_1^{(1)}$, $\mathbf{Y}_2=\mathbf{Y}_2^{(2)}$, and $(\mathbf{Z}_1^{(k)},\mathbf{Z}_{2n}^{(k)})=(\mathbf{Z}_1,\mathbf{Z}_{2n})$ $(k=1,2)$, and then we recover (2.1) with $u_i=v_{1i}-\boldsymbol{\beta}_2^{'}\mathbf{v}_{2i}$ by multiplying $(1,-\boldsymbol{\beta}_2^{'})'$ from the right to $(\mathbf{y}_1,\mathbf{Y}_2)$. In two-sample case, however, we do not have such structural representation for the $i-$th individual $(i=1,\cdots,n)$ observation. Nonetheless, the situation can be interpreted as a classical (linear) errors-in-variables model and then, we may expect that the estimated regression coefficients of $\hat{\boldsymbol{\pi}}_{21}^{(n)}$ on $\hat{\boldsymbol{\Pi}}_{22}^{(n}$ (in the second stage) are badly biased when K_{2n} is large, that is, when the number of instruments

is large. (See Anderson (1984) on the classical errors-in-variables models.)

For the vector of included instrumental variables in the structural equation $\boldsymbol{\gamma}_1$ is based on the upper part of the relation in (2.4) as

$$\hat{\boldsymbol{\gamma}}_1 = \hat{\boldsymbol{\pi}}_{11} - \hat{\boldsymbol{\Pi}}_{12}\hat{\boldsymbol{\beta}}_2 \ , \tag{2.6}$$

where $\hat{\boldsymbol{\beta}}_2$ is the estimator of β_2,
(2.7)

$$[\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\Pi}}_{12}] = [(\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{Z}_1^{(1)})^{-1}\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{y}_1^{(1)}, (\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{Z}_1^{(2)})^{-1}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{Y}_2^{(2)}] \ ,$$

and the projection matrices $\mathbf{P}_{Z_{2n}^{(k)}} = \mathbf{Z}_{2n}^{(k)}(\mathbf{Z}_{2n}^{(k)'}\mathbf{Z}_{2n}^{(k)})^{-1}\mathbf{Z}_{2n}^{(k)'}$, $\bar{\mathbf{P}}_{Z_{2n}^{(k)}} = \mathbf{I}_n - \mathbf{P}_{Z_{2n}^{(k)}}$ $(k = 1, 2)$.

Let $\mathbf{A}_{22.1}^{(k)} = \mathbf{Z}_{2n}^{(k)'}\bar{\mathbf{P}}_{Z_1^{(k)}}\mathbf{Z}_{2n}^{(k)}$ $(k = 1, 2)$ be sequences of $K_{2n} \times K_{2n}$ matrices, which remove the effects of included instrumental variables. We define the $(1+G_2)\times(1+G_2)$ matrix by

$$\mathbf{G} = \begin{bmatrix} g_{11} & \mathbf{g}_{12} \\ \mathbf{g}_{21} & \mathbf{G}_{22} \end{bmatrix} \tag{2.8}$$

where $g_{11} = \hat{\boldsymbol{\pi}}_{21}^{(n)'}\mathbf{A}_{22.1}^{(1)}\hat{\boldsymbol{\pi}}_{21}^{(n)}$, $\mathbf{G}_{22} = \hat{\boldsymbol{\Pi}}_{22}^{(n)'}\mathbf{A}_{22.1}^{(2)}\hat{\boldsymbol{\Pi}}_{22}^{(n)}$, $\mathbf{g}_{12} = \hat{\boldsymbol{\pi}}_{21}^{(n)'}\mathbf{A}_{22.1}^{(1)}\mathbf{A}_{22.1}^{-1}\mathbf{A}_{22.1}^{(2)}\hat{\boldsymbol{\Pi}}_{22}^{(n)}$, $\mathbf{g}_{12} = \mathbf{g}_{12}^{'}$, and $\mathbf{A}_{22.1} = w_1\mathbf{A}_{22.1}^{(1)} + w_2\mathbf{A}_{22.1}^{(2)}$ is a sequence of $K_{2n} \times K_{2n}$ positive definite matrices with $w_1 + w_2 = 1, w_1 \geq 0, w_2 \geq 2$ in general. But we take the case of $w_1 = 1$ in the following analysis because of the resulting simplicity. (We shall impose a condition that $\mathbf{A}_{22.1}^{(k)}$ $(k = 1, 2)$ are essentially the same in Section 3.)
For the estimation of the variance-covariance matrix $\boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Omega}_{22} \end{pmatrix}$ ($(1 + G_2) \times (1 + G_2)$ matrix), we use the $(1 + G_2) \times (1 + G_2)$ matrix by

$$\mathbf{H} = \begin{bmatrix} h_{11} & \mathbf{0}' \\ \mathbf{0} & \mathbf{H}_{22} \end{bmatrix} \ , \tag{2.9}$$

where $h_{11} = \mathbf{y}_1^{(1)'}[\mathbf{I}_n - \mathbf{P}_{(Z_1^{(1)}, Z_{2n}^{(1)})}]\mathbf{y}_1^{(1)}$, ($G_2 \times G_2$ matrix) $\mathbf{H}_{22} = \mathbf{Y}_2^{(2)'}[\mathbf{I}_n - \mathbf{P}_{(Z_1^{(2)}, Z_{2n}^{(2)})}]\mathbf{Y}_2^{(2)}$, and $\mathbf{I}_n - \mathbf{P}_{(Z_1^{(k)}, Z_{2n}^{(k)})}$ $(k = 1, 2)$ are the $n \times n$ projection matrices.

We investigate the class of estimators depending on $\mathbf{G}$ and $\mathbf{H}$ ($(1+G_2)\times(1+G_2)$ stochastic matrices) in two sample situation. In one sample case when K_{2n} does not depend on n, they constitute the sufficient statistics known in statistical multivariate analysis (see Anderson (2003)). They differ from the corresponding ones in the one-sample case because $(\mathbf{Z}_1^{(1)}, \mathbf{Z}_{2n}^{(1)}, \mathbf{y}_1^{(1)})$ is different from $(\mathbf{Z}_1^{(2)}, \mathbf{Z}_{2n}^{(2)}, \mathbf{Y}_2^{(2)})$. Because

we have two samples and they are different samples from the same population, we assume that there is no correlation among v_{1i} and $\mathbf{v}_{2i}$ $(i = 1, \cdots, n)$.
We define the 2SLVR (Two-Sample Least Variance Ratio) estimator $\hat{\boldsymbol{\beta}}_{LVR}$ $(= (1, -\hat{\boldsymbol{\beta}}^{'}_{2.LVR})^{'})$ of $\boldsymbol{\beta}$ $(= (1, -\boldsymbol{\beta}^{'}_2)^{'})$ as the solution of

$$(2.10) \qquad (\frac{1}{n}\mathbf{G} - \lambda_n \frac{1}{q_n}\mathbf{H})\hat{\boldsymbol{\beta}}_{LVR} = \mathbf{0} \ ,$$

where $q_n = n - K_n$ $(q_n > 2)$ and λ_n is the smallest root of

$$(2.11) \qquad |\frac{1}{n}\mathbf{G} - l\frac{1}{q_n}\mathbf{H}| = 0 \ .$$

The solution minimizes the variance ratio

$$(2.12) \qquad \mathrm{VR}_n = \frac{(1, -\boldsymbol{\beta}^{'}_2)\mathbf{G}\begin{pmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{pmatrix}}{(1, -\boldsymbol{\beta}^{'}_2)\mathbf{H}\begin{pmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{pmatrix}}$$

with respect to the structural parameter vector $\boldsymbol{\beta}$. The 2STSLS estimator $\hat{\boldsymbol{\beta}}_{TS}$ $(= (1, -\hat{\boldsymbol{\beta}}^{'}_{2.TS})^{'})$ of $\boldsymbol{\beta} = (1, -\boldsymbol{\beta}^{'}_2)^{'}$ is given by

$$(2.13) \qquad [\mathbf{g}_{21}, \mathbf{G}_{22}]\begin{pmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.TS} \end{pmatrix} = \mathbf{0} \ ,$$

where $\mathbf{Y}_2 = (\mathbf{y}^{'}_{2i})$ is an $n \times G_2$ matrix. The 2STSLS estimator minimizes the numerator of the variance ratio.

For the one sample data, the 2SLVR estimator corresponds to the LIML (limited information maximum likelihood) estimator while the 2STSLS estimator corresponds to the TSLS (two-stage least squares) estimator. Their properties in the general case were originally developed by Anderson and Rubin (1949, 1950). See also Anderson (2005).

3 Asymptotic Optimality of 2SLVR Estimator

3.1 Asymptotic Normality of the 2SLVR Estimator

We state the limiting distribution of the 2SLVR estimator under a set of alternative assumptions when K_{2n} and $\mathbf{\Pi}_{2n}$ depend on n and $n \to \infty$. We first consider the case when

$$(\mathbf{I}) \qquad \frac{K_{2n}}{n} \longrightarrow c \;\; (0 \le c < 1),$$

$$(\mathbf{II}) \qquad \frac{1}{n}\mathbf{\Pi}_{22}^{(n)'}\mathbf{A}_{22.1}^{(k)}\mathbf{\Pi}_{22}^{(n)} \stackrel{p}{\longrightarrow} \mathbf{\Phi}_{22.1} \quad (k=1,2),$$

$$(\mathbf{III}) \qquad \eta = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\left[p_{ii}^{(n)}(\mathbf{P}_{Z_{2n}^{*(k)}}) - c\right]^2 \quad (k=1,2)\ ,$$

and $\eta_c = \left[\frac{1}{1-c}\right]^2 \eta$ (η is a non-negative constant),
where $p_{ii}^{(n)}(\mathbf{P}_{Z_{2n}^*}^{(k)})$ is the $(i,i)-$diagonal element of $\mathbf{P}_{Z_{2n}^*}^{(k)}$ $(= \mathbf{Z}_{2n}^{*(k)}(\mathbf{A}_{22.1}^{(k)})^{-1}\mathbf{Z}_{2n}^{*(k)'})$ and $\mathbf{\Phi}_{22.1}$ is a $G_2 \times G_2$ nonsingular matrix.

Condition (I) implies that the number of coefficient parameters is proportional to the number of observations. Because we want to estimate the covariance matrix of $(v_{1i}, \mathbf{v}_{2i}^{'})'$ $(i = 1, \cdots, n)$ consistently, we need $0 \le c < 1$. Then, Condition (I) implies $q_n \longrightarrow \infty$ as $n \longrightarrow \infty$. Condition (II) controls the non-centrality (or concentration) parameter to be proportional to the sample size. Since K_{2n} grows, it is often called the case of *many instruments*. These conditions define the rates of growth of the number of incidental parameters in the statistical model. Condition (I) can be weakened to Conditions with $c = 0$, where K_{2n} increases with n but at a smaller rate (see Theorem 2). Condition (II) implies that two sets of instruments share the same consistent structure on average, which may be reasonable if the effects of the instruments are similar in two samples. Further, we may impose conditions that for any $g, h = 1, 2$

$$(\mathbf{II}^*) \quad \frac{1}{n}\mathbf{\Pi}_{22}^{(n)'}\mathbf{A}_{22.1}^{(g)}\mathbf{A}_{22.1}^{(h)-1}\mathbf{A}_{22.1}^{(g)}\mathbf{\Pi}_{22}^{(n)} \stackrel{p}{\longrightarrow} \mathbf{\Phi}_{22.1}\ ,$$

and

$$(\mathbf{II}^{**}) \quad \frac{1}{n}\mathbf{\Pi}_{22}^{(n)'}\mathbf{A}_{22.1}^{(g)}\mathbf{\Pi}_{22}^{(n)} = \mathbf{\Phi}_{22.1} + o_p(\frac{1}{\sqrt{n}})\ .$$

These are useful for the simplification of the resulting formulae of the asymptotic variance-covariances of the 2SLVLR estimator. As an example, we take $\mathbf{A}_{22.1}^{(k)}/n \sim c_a\mathbf{I}_{K_{2n}}$ for some c_a (> 0) and $k = 1, 2$. Then, a sequence of $G_2 \times G_2$ (or $G_2 \times K_{2n} \times K_{2n} \times G_2$) matrices $\mathbf{\Pi}_{22}^{(n)'}\mathbf{\Pi}_{22}^{(n)} < \infty$, which means that the additional coefficients in the reduced form (2.2) become negligible as n and K_{2n} increase. However, these conditions can be weakened with some cost of notational complications without changing the results essentially. Condition (III) is needed for analyzing the effects of non-normal disturbances on estimation.

We summarize the basic results on the asymptotic distributions of the 2SLVR and 2STSLS estimators for $\boldsymbol{\beta}_2$ when n and K_{2n} are large. To state our results in a compact way, we transform $\mathbf{v}_i = (v_{1i}, \mathbf{v}_{2i}^{'})^{'}$ $(= (v_{ji}), i = 1, \cdots, n; j = 1, \cdots, G_2 + 1)$ to

$$\text{(3.1)} \qquad \mathbf{w}_{2i} = \mathbf{A}\ \mathbf{v}_i\ ,\ \mathbf{A} = (\mathbf{0}, \mathbf{I}_{G_2})[\mathbf{I}_{G_2+1} - \frac{\mathbf{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta}}]\ ,$$

and $\sigma^2 = \mathbf{E}[(= \boldsymbol{\beta}^{'}\mathbf{v}_i)^2]$.
Then, the vector $\mathbf{w}_{2i}$ is uncorrelated with $u_i = \boldsymbol{\beta}^{'}\mathbf{v}_i$ $(i = 1, \cdots, n)$, which is the disturbance term of the structural equation.
To measure the effects of the non-normality of disturbance terms, we use the fourth order moments and we define a $(1 + G_2) \times (1 + G_2)$ matrix

$$\text{(3.2)} \qquad \mathbf{\Gamma}(\mathbf{v}) = \begin{bmatrix} \mathbf{\Gamma}(v_1) & \mathbf{0}^{'} \\ \mathbf{0} & \mathbf{\Gamma}(\mathbf{v}_2) \end{bmatrix}\ ,$$

where $\mathbf{\Gamma}(v_1) = \mathbf{E}[v_{1i}^4] - 3\omega_{11}^2$ and $\mathbf{\Gamma}(\mathbf{v}_2) = \mathbf{E}[\mathbf{v}_{2i}\mathbf{v}_{2i}^{'}(\boldsymbol{\beta}_2^{'}\mathbf{v}_{2i})^2] - [2\mathbf{\Omega}_{22}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'}\mathbf{\Omega}_{22} + \boldsymbol{\beta}_2^{'}\mathbf{\Omega}_{22}\boldsymbol{\beta}_2\mathbf{\Omega}_{22}]$. (When $G_2 = 1$, $\mathbf{\Gamma}(v_2) = \mathbf{E}[(v_{2i}^4 - 3\Omega_{22}^2)\beta_2^2]$.)
When $\mathbf{v}_i$ are normally distributed, $\mathbf{\Gamma}(v_1) = 0$, $\mathbf{\Gamma}(\mathbf{v}_2) = \mathbf{O}$ and then $\mathbf{\Gamma}(\mathbf{v}_2) = \mathbf{O}$. Because we assume that the noise terms in two-sample data v_{1i} and $\mathbf{v}_{2i}$ are mutually independent sequences, $\mathbf{\Gamma}(\mathbf{v})$ is a block-diagonal matrix.

Theorem 1 : Let $\mathbf{z}_{2n,i}^{*(k)'}$ $(i = 1, 2, \cdots, n; k = 1, 2)$ be the $i-$th row vector (a $K_{2n} \times 1$) of $\mathbf{Z}_{2n}^{*(k)}$ ($n \times K_{2n}$ matrix of normalized instrumental variables). Let $\mathbf{v}_i = (v_{1i}, \mathbf{v}_{2i}^{'})^{'}, i = 1, 2, \cdots, n$, be a set of $(1 + G_2) \times 1$ independent random vectors, which are orthogonal to $\mathbf{z}_{2n,1}^{*(k)}, \cdots, \mathbf{z}_{2n,n}^{*(k)}$ such that $\mathbf{E}(\mathbf{v}_i|\mathbf{z}_{2n,i}^{*(k)}) = \mathbf{0}$ and $\mathbf{E}(\mathbf{v}_i\mathbf{v}_i^{'}) = \mathbf{\Omega}$ $(a.s.)$.
Suppose that Conditions (I), (II)*, (II)**, (III) hold. In addition, we assume

$$\textbf{(IV)} \qquad \frac{1}{n}\max_{1 \le i \le n} \|\mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{2n,i}^{*(k)}\|^2 \xrightarrow{p} 0\ (k = 1, 2)\ .$$

Let $\hat{\boldsymbol{\beta}}_{2.LVR}$ be the 2SLVR estimator of $\boldsymbol{\beta}_2$.
(i) For $c = 0$, then

$$\text{(3.3)} \qquad \sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LVR} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Psi}^*)\ ,$$

where $\mathbf{\Psi}^* = \sigma^2\mathbf{\Phi}_{22.1}^{-1}$ and $\sigma^2 = \boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta}$.
(ii) For $0 < c < 1$, we further assume that $\mathbf{E}[v_{ji}v_{ki}v_{li}] = 0$ for any j, k, l and $i = 1, \cdots, n$ [1], where v_{ji} is the j-th element of $\mathbf{v}_i$, and assume that $\mathbf{E}[\|\mathbf{v}_i\|^{4+\epsilon}] < \infty$ for some $\epsilon > 0$ (and $\mathbf{E}[\|\mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{2n,i}^{*(k)}\|^{2+\delta}] < \infty$ for some $\delta > 0$ when $\mathbf{z}_i^{(n)}$ are stochastic). Then

$$\text{(3.4)} \qquad \sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LVR} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Psi}^{**})\ ,$$

[1] We make this assumption for the simplicity. It is possible to relax this condition with some complication of notation. See AKM (2010) for the one-sample case.

where

$$\begin{aligned}\boldsymbol{\Psi}^{**} &= \sigma^2\boldsymbol{\Phi}_{22.1}^{-1} + \boldsymbol{\Phi}_{22.1}^{-1}\Big\{c\Big[\boldsymbol{\Omega}_{22}\sigma^2 - \boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'}\boldsymbol{\Omega}_{22}\Big] \\ &+ cc_*\Big[(\boldsymbol{\beta}_2^{'}\boldsymbol{\Omega}_{22}\boldsymbol{\beta})\boldsymbol{\Omega}_{22} + \boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'}\boldsymbol{\Omega}_{22}[4(\frac{\omega_{11}}{\sigma^2})^2 - 1]\Big] + \eta_c\mathbf{A}\boldsymbol{\Gamma}(\mathbf{v})\mathbf{A}^{'}\Big\}\boldsymbol{\Phi}_{22.1}^{-1}\end{aligned}$$

and $c_* = c/(1-c)$.

If $G_2 = 1$ in Theorem 1, $[\boldsymbol{\Omega}\sigma^2 - \boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}]_{22} = \omega_{11}\omega_{22} = |\boldsymbol{\Omega}|$ because $\mathbf{E}[v_{1i}v_{2i}] = \omega_{12} = 0$.
This theorem is an extension of Theorem 1 of AKM (2010) for the one sample case to the two-sample case. Since we have some different model structure in two-sample case, the resulting formula is slightly different from Theorem 1 of AKM (2010). The variance-covariance matrix of the limiting distribution of the 2SLVR estimator depends on the third and fourth order moments of the disturbance terms in the general case. Instead of making an assumption on the distribution of disturbance terms except the existence of their moments, alternatively we may assume

$$(\mathbf{III}^{'}) \qquad \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Big[p_{ii}^{(n)}(\mathbf{P}_{Z_{2n}^{*(k)}}) - c\Big]^2 = 0 \ (k=1,2).$$

Then the asymptotic distribution does not depend on the fourth-order moments. A simple example for Condition $(\mathbf{III}^{'})$ is the case when we have $p_{ii}((\mathbf{P}_{Z_{2n}^{*(k)}}) = \mathbf{z}_{2n,i}^{*(k)'}(\mathbf{Z}_{2n}^{*'}\mathbf{Z}_{2n}^{*})^{-1}\mathbf{z}_{2n,i}^{*(k)} \sim K_{2n}/n$. Condition $(\mathbf{III}^{'})$ is the same as $\eta = 0$ in Condition (III) and the direct consequence of Condition $(\mathbf{III}^{'})$ implies the following result :

Corollary 1 : For $0 \le c < 1$ assume Conditions (I), (II*), (II**) and (III$^{'}$). Furthermore, assume that $\mathbf{E}[\|\mathbf{v}_i\|^{2+\epsilon}] < \infty$ for some $\epsilon > 0$ (and $\mathbf{E}[\|\boldsymbol{\Pi}_{22}^{(n)'}\mathbf{z}_{2n,i}^{*(k)}\|^{2+\delta}] < \infty$ for some $\delta > 0$ when $\mathbf{z}_i^{(n)}$ are stochastic). Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LVR} - \boldsymbol{\beta}_2) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Psi}^{**}) \ , \tag{3.5}$$

where

$$\begin{aligned}\boldsymbol{\Psi}^{**} &= \sigma^2\boldsymbol{\Phi}_{22.1}^{-1} + \boldsymbol{\Phi}_{22.1}^{-1}\Big\{c\Big[\boldsymbol{\Omega}_{22}\sigma^2 - \boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'}\boldsymbol{\Omega}_{22}\Big] \\ &+ cc_*\Big[(\boldsymbol{\beta}_2^{'}\boldsymbol{\Omega}_{22}\boldsymbol{\beta})\boldsymbol{\Omega}_{22} + \boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'}\boldsymbol{\Omega}_{22}[4(\frac{\omega_{11}}{\sigma^2})^2 - 1]\Big]\Big\}\boldsymbol{\Phi}_{22.1}^{-1}\end{aligned}$$

and $c_* = c/(1-c)$.

The asymptotic properties of the 2SLVR estimator also hold when K_{2n} increases as $n \to \infty$ and $K_{2n}/n \to 0$. In this case the limiting distribution of the 2SLVR estimator can be still different from that of the 2TSLS estimator, depending on the relative magnitude of n and K_{2n}.

Theorem 2 : Let $\mathbf{z}_{2n,i}^{*(k)'}$ $(i = 1, 2, \cdots, n; k = 1, 2)$ be the $i-$th row vector (a $K_{2n} \times 1$) of $\mathbf{Z}_{2n}^{*(k)}$ ($n \times K_{2n}$ matrix of normalized instrumental variables). Let $\mathbf{v}_i = (v_{1i}, \mathbf{v}_{2i}^{'})^{'}, i = 1, 2, \cdots, n$, be a set of $(1+G_2) \times 1$ independent random vectors, which are orthogonal to $\mathbf{z}_{2n,1}^{*(k)}, \cdots, \mathbf{z}_{2n,n}^{*(k)}$ such that $\mathbf{E}(\mathbf{v}_i|\mathbf{z}_{2n,i}^{*(k)}) = \mathbf{0}$ and $\mathbf{E}(\mathbf{v}_i\mathbf{v}_i^{'}|\mathbf{z}_{2n,i}^{*(k)}) = \mathbf{\Omega}_i^{(n)}$ $(a.s.)$ is a function of $\mathbf{z}_{2n,i}^{*(k)}$ $(k = 1, 2)$. The further assumptions on $(\mathbf{v}_i, \mathbf{z}_{2n,i}^{*(k)})$ $(\mathbf{v}_i = (v_{ji}))$ are that $\mathbf{E}(v_{ji}^4|\mathbf{z}_{2n,i}^{*(k)})$ are bounded, there exists a constant matrix $\mathbf{\Omega}$ such that $\sqrt{n}\|\mathbf{\Omega}_i^{(n)} - \mathbf{\Omega}\|$ is bounded and $\sigma^2 = \boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta} > 0$. Assume Conditions (II*), (II**), and

$$(\mathbf{I}^{'}) \qquad \frac{K_{2n}}{n^{\nu}} \longrightarrow c_{\nu} \ (0 \leq \nu < 1,\ 0 < c_{\nu} < \infty),$$

$$(\mathbf{IV}) \qquad \frac{1}{n} \max_{1 \leq i \leq n} \|\mathbf{\Pi}_{22}^{(n)'} \mathbf{z}_{2n,i}^{*(k)}\|^2 \stackrel{p}{\longrightarrow} 0 \ \ (k = 1, 2),$$

where $\mathbf{\Phi}_{22.1}$ is a nonsingular constant matrix.
Let $\hat{\boldsymbol{\beta}}_{2.LVR}$ and $\hat{\boldsymbol{\beta}}_{2.TS}$ be the 2SLVR estimator and the 2STSLS estimator of $\boldsymbol{\beta}_2$, respectively.
(i) Then for the 2SLVR estimator when $0 \leq \nu < 1$,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LVR} - \boldsymbol{\beta}_2) \stackrel{d}{\longrightarrow} N(\mathbf{0}, \sigma^2 \mathbf{\Phi}_{22.1}^{-1}) \ , \tag{3.6}$$

where $\sigma^2 = \boldsymbol{\beta}^{'}\mathbf{\Omega}\boldsymbol{\beta}$.
(ii) For the 2STSLS estimator when $1/2 < \nu < 1$,

$$n^{1-\nu}(\hat{\boldsymbol{\beta}}_{2.TS} - \boldsymbol{\beta}_2) \stackrel{p}{\longrightarrow} \mathbf{\Phi}_{22.1}^{-1}\ (-c_{\nu})\mathbf{\Omega}_{22})\boldsymbol{\beta}_2 \ , \tag{3.7}$$

when $\nu = 1/2$,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.TS} - \boldsymbol{\beta}_2) \stackrel{d}{\longrightarrow} N\left[(-c_{\nu})\mathbf{\Phi}_{22.1}^{-1}\mathbf{\Omega}_{22}\boldsymbol{\beta}_2, \sigma^2 \mathbf{\Phi}_{22.1}^{-1}\right] \ , \tag{3.8}$$

where $\mathbf{\Omega}_{22}$ is the $G_2 \times G_2$ lower-left submatrix of $\mathbf{\Omega}$. When $0 \leq \nu < 1/2$,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.TS} - \boldsymbol{\beta}_2) \stackrel{d}{\longrightarrow} N(\mathbf{0}, \sigma^2 \mathbf{\Phi}_{22.1}^{-1}) \ . \tag{3.9}$$

It is possible to interpret the standard large sample theory as a special case of *Theorem 2*. When K_{2n} is a fixed number, it has been known that the LIML and TSLS estimators in the one-sample are asymptotically normal and equivalent. When K_{2n} is large, they are substantially different and there were several studies on their asymptotic as well as finite sample properties in one-sample case. Since the TSLS estimator is a special case of the GMM estimator and both share the similar asymptotic properties when K_{2n} is large. For the one-sample case of structural equation estimation, there have been earlier studies on the finite sample properties such as as Anderson, Kunitomo and Sawa (1982). See Angrist and Pischke (2009) as a textbook on applied econometrics. The asymptotic properties of one-sample IV methods have been discussed in econometric literature including AKM (2011), Donald and Newey (2001), Stock and Yogo (2005), Chao and Swanson (2005), Hansen et al. (2008), among many others.

3.2 Asymptotic Optimality with Many Instruments

For the estimation of the vector of structural parameters $\boldsymbol{\beta}$, it seems natural to consider procedures based on the two $(1+G_2)\times(1+G_2)$ matrices $\mathbf{G}$ and $\mathbf{H}$, which are sufficient statistics in the classical one-sample standard situation. (See Anderson (2003).) We shall consider a class of estimators which are functions of these matrices for the two sample structural equation estimation. The typical examples of this class are the Two-Sample OLS estimator, the 2STSLS estimator, and the 2SLVR estimator. Then we have a basic result on the asymptotic optimality of the 2LVR estimator and its (asymptotically equivalent) modifications, which attains the lower bound of the asymptotic covariance under alternative assumptions in most cases. The proof is essentially the same as the one-sample case given as the proof of Theorem 4 in AKM (2010), but we give its skech in the Appendix for the sake of completeness.

Theorem 3 : Define a class of consistent estimators for $\boldsymbol{\beta}_2$ in the form of

$$\hat{\boldsymbol{\beta}}_2=\phi(\frac{1}{n}\mathbf{G},\ \frac{1}{q_n}\mathbf{H})\ , \tag{3.10}$$

where ϕ is continuously differentiable and its derivatives are bounded at the probability limits of $(1/n)\mathbf{G}$ and $(1/q_n)\mathbf{H}$ as $K_{2n}\to\infty$ and $n\to\infty$ and $0\le c<1$. Then under the assumptions of *Theorem 1*, *Corollary 1*, or *Theorem 2*.

$$\mathcal{AE}\left[n(\hat{\boldsymbol{\beta}}_2-\boldsymbol{\beta}_2)(\hat{\boldsymbol{\beta}}_2-\boldsymbol{\beta}_2)^{'}\right]\ge\boldsymbol{\Psi}^{*}\ (\text{or }\boldsymbol{\Psi}^{**})\ , \tag{3.11}$$

and $\boldsymbol{\Psi}^{*}$ (or $\boldsymbol{\Psi}^{**}$) is given in Theorem 1, Corollary 1, and Theorem 2, where the right-hand side of (3.11) is the covariance matrix of the limiting distribution of the normalized estimator $\sqrt{n}(\hat{\boldsymbol{\beta}}_2-\boldsymbol{\beta}_2)$ for the class of (3.10).

The inequality (3.11) holds in the sense of non-negative definiteness of matrices. As a consequence, we have $\mathcal{AE}[n\|\hat{\boldsymbol{\beta}}_2-\boldsymbol{\beta}_2\|^2]\ge\mathrm{Tr}(\boldsymbol{\Psi}^{**})$. It is a fundamental result on the asymptotic optimality of the Two-Sample LVR estimator when there are many instruments K_{2n} along with the sample size. Although in some applications such as econometrics as well as Menderian Randomization (MR), there have been many studies on the asymptotic properties of alternative IV estimation methods, we could not find any asymptotic optimality results as far as we know even in the special case when $G_2 = 1$.

We have an important remark on Theorem 3 that the asymptotic optimality of the 2SLVR estimation holds as far as we have the asymptotic normality of the 2SLVR estimation. Although the asymptotic variance-covariances become more complicated, Conditions (II*) (II**), and assumption on the third order moments of $\mathbf{v}_i$ could be relaxed considerably. This problem is currently under investigation.

4 Discussion and Generalizations

4.1 A Finite Sample Correction

Although we have shown that the 2SLVR estimator has an asymptotic optimality when K_{2n} and n are large in Theorem 3, we need to have some care to evaluate its finite sample property. We may propose a finite sample correction of the 2SLVR estimator $\hat{\boldsymbol{\beta}}_{CLVR}^{'} = (1, -\hat{\boldsymbol{\beta}}_{2.CLVR})$ of $\boldsymbol{\beta}$, which is the solution of

$$(4.1) \qquad [\mathbf{0}, \mathbf{I}_{G_2}][\frac{1}{n}\mathbf{G} - \lambda_n^{(f)}\frac{1}{q_n}\mathbf{H}]\hat{\boldsymbol{\beta}}_{CLVR} = \mathbf{0} \ ,$$

where $\lambda_n^{(f)} = \lambda_n - f/n$ for a positive number f and λ_n is the smallest root of the determinantal equation $|\frac{1}{n}\mathbf{G} - l\frac{1}{q_n}\mathbf{H}| = 0$. The choice of f can be either 1, 2 or 4 in practice.

This modified 2SLVR estimator is asymptotically equivalent to the 2SLVR estimator in the standard asymptotic framework, however, we may expect that it may improve the finite sample property in some situation. It has been known that in the one sample case, Fuller (1977) proposed to take $f = 1$ to improve the LIML estimation. Later, Anderson, Kunitomo and Morimune (1986) have investigated the effect of choosing f on the finite sample distribution of estimation methods when the disturbance terms are normally distributed and K_{2n} is fixed.

4.2 On Estimation of $\boldsymbol{\gamma}_1$

By using the upper-part of the relation (2.4) and the regression coefficients (2.7) in the reduced form equations, the 2SLVR estimator of $\boldsymbol{\beta}_2$ is given as (2.10).

Then, by using (2.2), we find that

$$
\begin{aligned}
\hat{\gamma}_1 &= (\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{Z}_{2n}^{(1)})^{-1}\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{y}_1^{(1)} - (\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{Z}_1^{(2)})^{-1}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{Y}_2^{(2)}]\hat{\boldsymbol{\beta}}_2 \\
&= [\boldsymbol{\Pi}_{11} + (\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{Z}_{2n}^{(1)})^{-1}\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{v}_1] \\
&\quad -[\boldsymbol{\Pi}_{12} + (\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{Z}_1^{(2)})^{-1}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{V}_2][\boldsymbol{\beta}_2 + (\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)] \\
&= \boldsymbol{\gamma}_1 + [(\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{Z}_{2n}^{(1)})^{-1}\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{v}_1 - (\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{Z}_1^{(1)})^{-1}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{V}_2\boldsymbol{\beta}_2] \\
&\quad -[\boldsymbol{\Pi}_{12} + (\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{Z}_1^{(2)})^{-1}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{V}_2](\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)\ .
\end{aligned}
$$

and

$$
\begin{aligned}
\sqrt{n}[\hat{\gamma}_1 - \boldsymbol{\gamma}_1] &= \frac{1}{n}\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{Z}_{2n}^{(1)})^{-1}\frac{1}{\sqrt{n}}\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{v}_1 \\
&\quad -(\frac{1}{n}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{Z}_1^{(2)})^{-1}\frac{1}{\sqrt{n}}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{V}_2\boldsymbol{\beta}_2 \\
&\quad -[\boldsymbol{\Pi}_{12} + (\frac{1}{n}(\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{Z}_1^{(2)})^{-1}\frac{1}{n}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{V}_2][\sqrt{n}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)\ .
\end{aligned}
$$

Then, from the proof of Theorem 1 in Section 3, we obtain the asymptotic normality under a set of conditions when both n and K_{2n} are large.
For the asymptotic covariance-variance, we assume Condition II in Section 3. Then, by using the derivations outlined in the Appendix, the asymptotic variance-covariance matrix is given by

$$
\text{(4.2)} \qquad \mathbf{V}[\sqrt{n}(\hat{\gamma}_1 - \boldsymbol{\gamma}_1)] = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}[\mathbf{M}_{11.2}^{-1} + \boldsymbol{\Pi}_{12}\Phi_{22.1}^{-1}(\frac{1}{\sigma^2}\boldsymbol{\Psi}^{**})\boldsymbol{\Phi}_{22.1}^{-1}\boldsymbol{\Pi}_{12}' + \mathbf{F} + \mathbf{F}']\ ,
$$

where $\mathbf{F} = \mathbf{M}_{11.2}^{-1}\mathbf{M}_{12.2}\sigma^2\Phi_{22.1}^{-1}\boldsymbol{\Pi}_{12}'$, provided that there exist $\mathbf{M}_{11.2}$ (a non-singular $K_1 \times K_1$ matrix) and a $K_1 \times G_2$ matrix $\mathbf{M}_{12.2}$ such that

$$
\text{(4.3)} \qquad \frac{1}{n}\mathbf{Z}_1^{(k)'}\bar{\mathbf{P}}_{Z_{2n}^{(k)}}\mathbf{Z}_1^{(k)} \xrightarrow{p} \mathbf{M}_{11.2}\ ,\ \frac{1}{n}\mathbf{Z}_1^{(k)'}\bar{\mathbf{P}}_{Z_{2n}^{(k)}}\bar{\mathbf{P}}_{Z_1^{(k)}}\mathbf{Z}_{2n}^{(k)}\boldsymbol{\Pi}_{22}^{(n)} \xrightarrow{p} \mathbf{M}_{11.2}\ (k = 1, 2).
$$

Furthermore, if we have the orthogonal condition that $\mathbf{Z}_1^{(k)'}\mathbf{Z}_{2n}^{(k)} = \mathbf{O}$ and $\mathbf{F} = \mathbf{O}$, (4.2) is simplified as

$$
\text{(4.4)} \qquad \mathbf{V}[\sqrt{n}(\hat{\gamma}_1 - \boldsymbol{\gamma}_1)] = \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}[\mathbf{M}_{11}^{-1} + \boldsymbol{\Pi}_{12}\Phi_{22.1}\boldsymbol{\Pi}_{12}']\ ,
$$

where $\frac{1}{n}\mathbf{Z}_1^{(k)'}\mathbf{Z}_1^{(k)} \xrightarrow{p} \mathbf{M}_{11}$ and $\mathbf{M}_{11}$ is assumed to be positive definite.

4.3 Weak Instruments

When there are many instruments, some economists were interested in the case when the explanatory power of additional instruments is not zero, but is small. This situation has been called as the problem of *weak instruments* in the econometric literature. One way to formulate this problem may be to cinsider the conditions that there exist ν ($0 < \nu \leq 1$) and δ ($0 < \delta \leq 1$) such that

$$(\mathbf{I}^{'}) \quad \frac{K_{2n}}{n^{\nu}} \longrightarrow c_{\nu} \ (0 \leq c_{\nu} < 1) \ , \ (\mathbf{II}^{'}) \quad \frac{1}{n^{\delta}} \mathbf{\Pi}_{22}^{(n)'} \mathbf{A}_{22.1}^{(k)} \mathbf{\Pi}_{22}^{(n)} \xrightarrow{p} \mathbf{\Phi}_{22.1} \ ,$$

where $\mathbf{\Phi}_{22.1}$ is positive definite.
Then, we need to take the normalization as $n^{\delta/2}[\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2]$ instead of $\sqrt{n}[\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2]$, and it is possible to obtain the results by modifying the derivations given in the Appendix. The results are similar to Theorem 2 and Theorem 3 in Section 3 with slightly complicated notations and we omitted the details.

4.4 On Estimation When $n_1 \neq n_2$

In applications the sample size from two samples are often different and we need to deal with this problem. There can be two ways to handle the problem.

First, we have obtained the asymptotic results when $n_1 = n_2$, and there is a natural way to take two-sample data with the same sample size by deleting extra observations. If we have the situation that $n_1 > n_2$, then we take two samples with the sample size $n = \min\{n_1, n_2\}$. Then, one obvious problem is a loss of information in two-sample data.

Because we have different samples, but there is a relationship between two endogenous variables in the structural equation. There should be a way to recover the information in two-sample data. Let $n \leq \min\{n_1, n_2\}$. Then, we can construct a set of n samples of $(y_{1i}, \mathbf{y}_{2i} \ ; i = 1, \cdots, n)$ randomly and apply the 2SLVR estimation method. We repeat the same procedure and apply the 2SLVR estimation method. It may be possible to combine many sets of estimates in subsampling to make a statistical inference in a efficient way by using this type of statistical resampling procedure.

5 Numerical Evaluation of Finite Distributions

It is important to investigate the finite sample properties of estimators partly because they are not necessarily the same as their asymptotic properties. One simple example would be the fact that the exact moments of the LIML estimator in one-sample case do not necessarily exist. In that case it is meaningless to compare the exact MSEs of alternative estimators and their Monte Carlo analogues. Although we discuss the asymptotic properties of the 2SLVR estimator, we need to investigate their relevance for practical applications.

There is a notable difference between the results in Theorems in Section 3, that is, the asymptotic variance depends on the 3rd and 4th order moments of the disturbance terms in the former. The finite sample properties of the LIML estimator for one-sample have been investigated by Anderson, Kunitomo and Matsushita (2011) in a systematic way. As typical examples, we present only 9 figures (Figures 1-9) among many other simulation cases we have done. We have used the numerical estimation of the cumulative distribution function (cdf) of the 2SLVR estimator based on the simulation and we have enough numerical accuracy in most cases.

To make interpretation easier, we set $n_1 = n_2 = n = 100$, $\beta_2 = 1$ ($G_2 = 1$), $\omega_{11} = 1, \omega_{12} = 0$ and $K_1 = 0$. Then, the key parameters in figures are K_2 (or K_{2n}), $n - K$ (or q_n and $\delta^2 = \mathbf{\Pi}_{22}^{(n)\prime}(\frac{1}{2}[\mathbf{A}_{22.1}^{(1)} + \mathbf{A}_{22.1}^{(2)})\mathbf{\Pi}_{22}^{(n)})/\sigma^2$. The empirical distributions of the normalized 2SLVR and 2STSLS estimators are traced by the red-curve and blue-curve, respectively. In addition to the 2SLVR estimator, we have added the distribution function of the 2STSLS estimator and the normal distribution for comparisons. The figures (Figures 1-9) show the estimated cdf of two estimators in the standard form, that is,

$$\frac{1}{\sigma}[\mathbf{\Pi}_{22}^{(n)\prime}\frac{1}{2}(\mathbf{A}_{22.1}^{(1)} + \mathbf{A}_{22.1}^{(2)})\mathbf{\Pi}_{22}^{(n)}]^{1/2}\left(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right) . \tag{5.1}$$

In our simulations, we first generate the normal random numbers for $\mathbf{\Pi}_{22}^{(n)}$, $\mathbf{Z}_{2n}^{(k)}$ ($k = 1, 2$), which are fixed. Then, we draw the empirical distribution function of estimators with 10,000 replications. (It is 100,000 when $K_{2n} = 2$.) Our numerical computation method is similar to the one used in Anderson et al. (2011) except that all computations were done by a new R-program.

The limiting distributions of the 2SLVR and 2STSLS estimators are $N(0, 1)$ in the standard large sample asymptotics, that is, when K_1 is a fixed number while n does to ∞. The corresponding limiting distributions of the 2SLVR estimators in the large K_2 asymptotics are $N(0, a)$ ($a = \mathbf{\Psi}^{*-1}\mathbf{\Psi}^{**}, a \geq 1$), which are traced by the black-curve in Figures 1-9.

From these figures, we have found that the effects of many instruments on the cdfs of the estimators are significant and the approximations based on the standard large sample asymptotics are often inferior while the large-K_2 asympptotics give reasonable approximations when K_2 is large. The empirical distribution function of the normalized 2SLVR estimator is close to the standard normal distribution and it has slightly more variation around the true value. It corresponds the result of Theorem 1 and $a \geq 1$. On the other hand, the empirical distribution of the normalized 2STSLS estimator has often bias when K_2 is not small. When Ω_{22} is large and $c_\nu > 0$, the negative bias becomes large and it agrees with the results of Theorem 2. At the same time, we also have found that the effects of non-normality of disturbance terms on the cdf of the 2SLVR estimator are often very small. (The dashed curves and x are almost identical.) The distribution of the 2STSLS estimator has significant bias with many instruments.

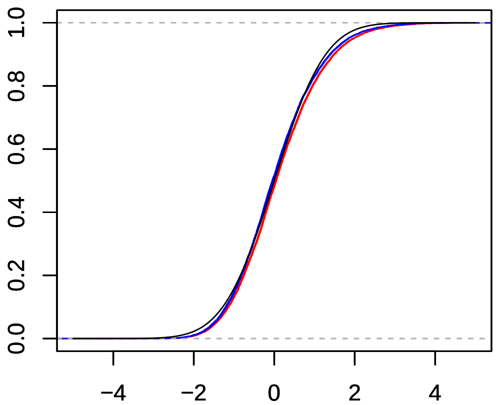

Figure 1: Ω_{22}=1, K_2=2
(Distributions of 2SLVR and 2STSLS, n=100.)

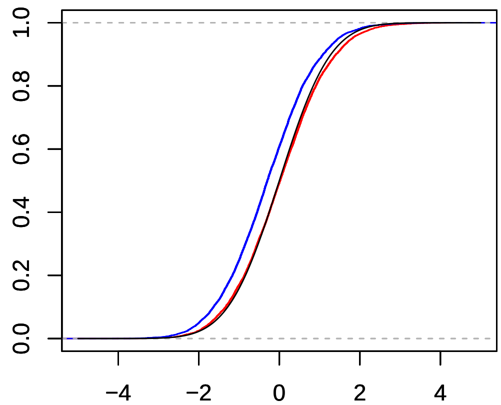

Figure 2: Ω_{22}=1, K_2=20
(Distributions of 2SLVR and 2STSLS, n=100.)

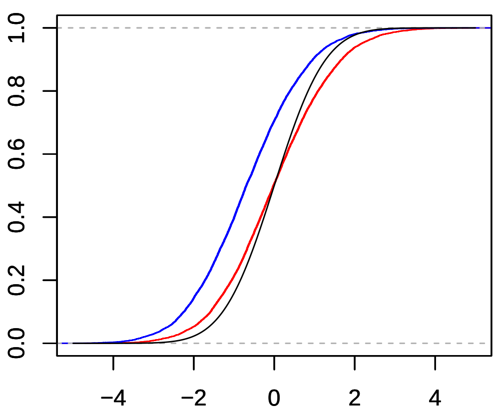

Figure 3: Ω_{22}=1, K_2=50
(Distributions of 2SLVR and 2STSLS, n=100.)

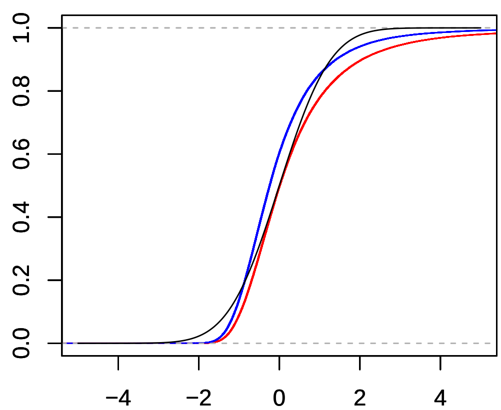

Figure 4: Ω_{22}=10, K_2=2
(Distributions of 2SLVR and 2STSLS, n=100.)

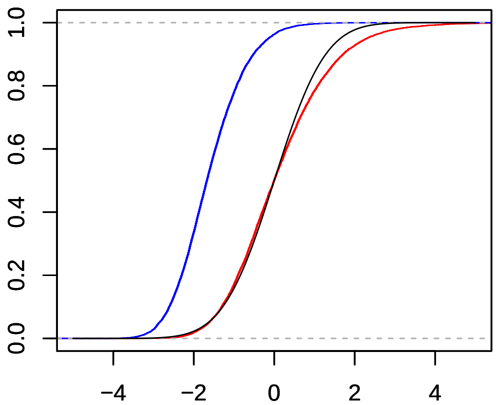

Figure 5: Ω_{22}=10, K_2=2
(Distributions of 2SLVR and 2STSLS, n=100.)

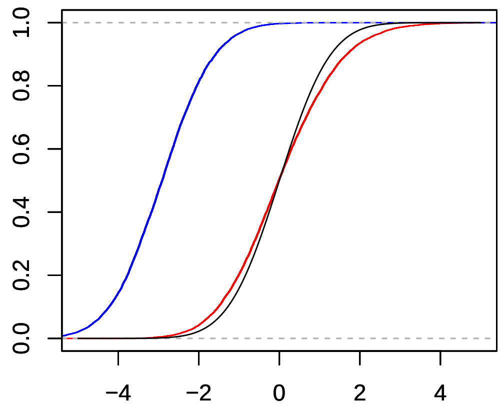

Figure 6: Ω_{22}=10, K_2=50
(Distributions of 2SLVR and 2STSLS, n=100.)

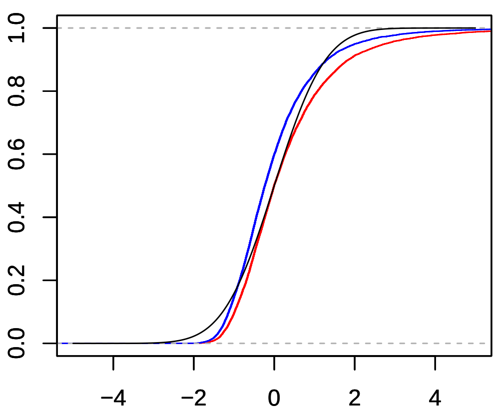

Figure 7: Ω_{22}=10, K_2=2,correlated instruments
(Distributions of 2SLVR and 2STSLS, n=100.)

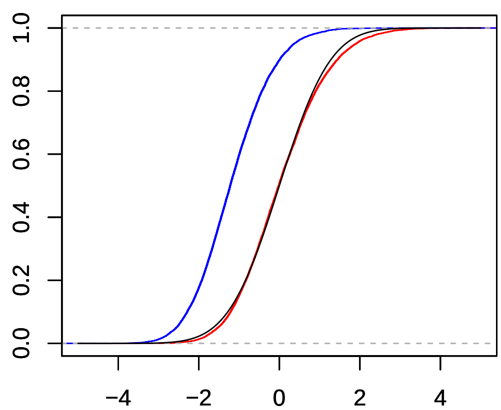

Figure 8: Ω_{22}=10, K_2=50,correlated instruments
(Distributions of 2SLVR and 2STSLS, n=100.)

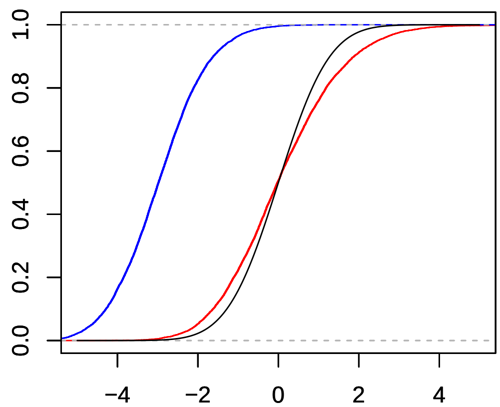

Figure 9: Ω_{22}=10, K_2=50, correlated instruments
(Distributions of 2SLVR and 2STSLS, n=100.)

6 Concluding Remarks

In this paper we have investigated on the asymptotic properties of the 2SLVR estimator when both the sample size and the number of instruments are large, that is, the case of *many instruments*. While two-sample version of the 2STSLS is badly biased, the 2SLVR estimator has the asymptotic optimality in the case of many instrumental variables. Our numerical results in Section 5 agree with the asymptotic results when the sample size is finite.

Hence, our results of this paper has suggested a guidance on the use of alternative estimation methods in structural equation estimation in econometrics and the Mendelian Randomization (MR) in biometrics where we may have *many instruments*.

There are a number of problems to be investigated. First, there are conditions imposed may be restrictive, we will need to investigate our results by relaxing them. Second, there are some literature in econometrics to extend the one-sample LIML, GMM and TSLS estimation methods such as Kunitomo (2011) for possible heteroscedasticity of noise terms, for instance.. It may be interesting to extend this line of study to the two-sample structural equation estimation. Third, the validity of instruments when we have many possible candidates may be important. This problem is related to the model selection when we have many instruments. Fourth, the linear structural equation estimation can be generalized to the non-linear models, but there may be some problem when there are many instruments. Finally, since there are a growing number of applications both in econometrics and bio-metrics. We are currently investigating these issues.

References

[1] Anderson, T.W. (1984), "Estimating linear statistical relationships." *Annals of Statistics*, 12, 1-45.

[2] Anderson, T.W. (2005), "Origins of the Limited Information Maximum Likelihood and Two-Stage Least Squares Estimators," *Journal of Econometrics*, Vol. 127, 1-16.

[3] Anderson, T.W. (2003), *An Introduction to Multivariate Statistical Analysis*, 3rd Ediition, Wiley.

[4] Anderson, T.W. and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, Vol. 20, 46-63.

[5] Anderson, T.W. and H. Rubin (1950), "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equation," *Annals of Mathematical Statistics*, Vol. 21, 570-582.

[6] Anderson, T.W., N. Kunitomo and T. Sawa (1982), "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Economerica*, Vol. 50, 1009-1026.

[7] Anderson, T.W., N. Kunitomo and K. Morimune (1986), "Comparing Single-Equation Estimators in a Simultaneous Equation System," *Economeric Theory*, Vol. 2, 1-32.

[8] Anderson, T.W., Kunitomo, N. and Matsushita, Y. (2010), "On the asymptotic optimality of the LIML estimator with possibly many instruments", Journal of Econometrics, Vol. 157, 191-204.

[9] Anderson, T.W., Kunitomo, N. and Matsushita, Y. (2011), "On finite sample properties of alternative estimators of coefficients in a structural equation with many instruments," Journal of Econometrics, Vol. 165-1, 58-69.

[10] Angrist, J. D. and J.S. Pischke (2009), Mostly Harmkless Econometrics : An Empiricist's Companion, Princeton University Press.

[11] Burgess, S. and Thompson, S. (2020), *Mendelian Randomization : Methods for Causal Inference Using Genetic Variants*, 2nd Edition, CRC Press.

[12] Chao, J. and N. Swanson (2005), "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, Vol. 73, 1673-1692.

[13] Donald, S. and W. Newey (2001), "Choosing The Number of Instruments," *Econometrica*, Vol. 69-5, 1161-1191.

[14] Fuller, W. (1977), "Some properties of a modification of the limited information estimator," *Econometrica*, 45 (1977): 939-953.

[15] Hansen, C. J. Hausman, and Newey, W. K. (2008), "Estimation with Many Instrumental Variables," *Journal of Business and Economic Statistics*, 26-4, 398-422.

[16] Hayashi, F. (2000) *Econometrics*, Princeton University Press.

[17] Hall, P. and Heyde, C. (1980), *Martingale Limit Theory and its Applications*, Academic Press.

[18] Inoue, A. and G. Solon (2010), "Two-Sample instrumental variables Estimators," *The Review of Economics and Statistics*, 92-3, 557-561.

[19] Kunitomo, N. (1980), "Asymptotic Expansions of the Distributions of Estimators in a Linear Functional Relationship and a Simultaneous Equation," *Journal of the American Statistical Association*, Vo.75-371, 693-700.

[20] Kunitomo, N. (2011), "An optimal modification of the LIML estimation for many instruments and persistent heteroscedasticity," *Annals of Institute of Statistical Mathematics*, DOI 10.1007/s10463-011-0336-7.

[21] Morimune, K. (1983), "Approximate Distributions of k-Class Estimators when the degrees of Overidentification is LargeCompared with the Sample Size," *Econometrica*, 51-3, 821-841.

[22] Stock, J. and Yogo, M. (2005), "Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments," In *Identification and Inference for Econometric Models*, Cambridge University Press (edited by D. Andrews and J. Stock).

[23] Ye, T., J. Shao and H. Kang (2021), "Debiased Inverse-Variance Weighted Estimator in Two-Sample Summary-Data Mendelian Randomization," *Annals of Statistics*, 49-4,2079-2100.

[24] Zhao, Q., J. Wang, G. Hemani, J. Bowden, and D. Small (2020), "Statistical Inference in Two-Sample Summary-Data Mendelian Randomization Using Robust Adjusted Profile Score," *Annals of Statistics*, 48-3, 1742-1769.

APPENDIX A : Mathematical Derivations

In this Appendix we give the proofs of *Theorems* and the mathematical derivation in Section 3. Many arguments are similar to those in AKM (2010) for the one-sample case and the proof of Theorem 3 is almost the same as the one in AKM (2010), which is added for the sake of expository completeness.

(i) **Proof of Theorem 1** :

We take $\mathbf{A}_{22.1} = \mathbf{A}_{22.1}^{(1)}$ and use the $n \times K_{2n}$ matrix $\mathbf{Z}_{2n}^{*(k)}$, which is orthogonal to the $n \times K_1$ matrix $\mathbf{Z}_1^{(k)}$ $(k = 1, 2)$, extensively. We intestigate the asymptotic behaviors of $(1/n)\mathbf{G}$ and $(1/q_n)\mathbf{H}$ when both n and K_{2n} are large. From (2.2) and (2.3),

$$\begin{aligned} g_{11} &= [\mathbf{\Pi}_{1n}'(\mathbf{Z}_1^{(1)'}, \mathbf{Z}_{2n}^{(1)})' + \mathbf{v}_1']\bar{\mathbf{P}}_{Z_1^{(1)}}\mathbf{Z}_{2n}^{(1)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)})^{-1}\mathbf{Z}_{2n}^{(1)}\bar{\mathbf{P}}_{Z_1^{(1)}}[(\mathbf{Z}_1^{(1)}, \mathbf{Z}_{2n}^{(1)})\mathbf{\Pi}_{1n} + \mathbf{v}_1] \\ &= \boldsymbol{\pi}_{21}^{(n)'}\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)}\boldsymbol{\pi}_{21}^{(n)} + \boldsymbol{\pi}_{21}^{(n)'}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 + \mathbf{v}_1'\mathbf{Z}_{2n}^{*(1)}\boldsymbol{\pi}_{21} + \mathbf{v}_1'\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 \ . \end{aligned}$$

Then

$$\text{(A.1)} \qquad \begin{aligned} & g_{11} - [\boldsymbol{\pi}_{21}^{(n)'}\mathbf{A}_{22.1}^{(1)}\boldsymbol{\pi}_{21}^{(n)} + K_{2n}\boldsymbol{\omega}_{11}] \\ &= \boldsymbol{\pi}_{21}^{(n)'}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 + \mathbf{v}_1'\mathbf{Z}_{2n}^{*(1)}\boldsymbol{\pi}_{21}^{(n)} + [\mathbf{v}_1'\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 - K_{2n}\,\omega_{11}] \ . \end{aligned}$$

Conditions (I) and (II) imply that as $n \longrightarrow \infty$

$$\text{(A.2)} \qquad \frac{1}{n}\boldsymbol{\pi}_{21}^{(n)'}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 \xrightarrow{p} \mathbf{0} \ ,$$

and

$$\text{(A.3)} \qquad \frac{1}{n}\left[\mathbf{v}_1'\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 - K_{2n}\boldsymbol{\omega}_{11}\right] \xrightarrow{p} 0 \ .$$

It is because (A.2) is the result of direct evaluation of

$$\mathbf{Var}[\frac{1}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{\pi}_{21}^{(n)'}\mathbf{z}_{2n.i}^{*(1)}v_{1i}] = \frac{1}{n}\omega_{11}\boldsymbol{\pi}_{21}^{(n)'}\sum_{i=1}^n \mathbf{z}_{2n.i}^*\mathbf{z}_{2n.i}^*\boldsymbol{\pi}_{21} = \omega_{11}\boldsymbol{\pi}_{21}^{(n)'}\frac{1}{n}\mathbf{A}_{22.1}^{(1)}\boldsymbol{\pi}_{21}^{(n)} \ .$$

For (A.3), we use the notation $p_{ij}(1) = (\mathbf{P}_{Z_{2n}^{*(1)}})_{ij}$ $(i, j = 1, \cdots, K_{2n})$ and $\sum_{i,j=1}^n [v_{1i}v_{1j}p_{ij}(1) - \omega_{11}p_{ij}(1)\delta(i,j)] = \sum_{i=1}^n [(v_{1i}^2 - \omega_{11})p_{ii}(1)] + 2\sum_{i>j=1}^{n-1} v_{1i}v_{1j}p_{ij}(1)$. Then the variance becomes $\sum_{i=1}^n \mathbf{E}[(v_{1i}^2 - \omega_{11})p_{ii}(1)]^2 + 4\sum_{i>j=1}^{n-1}\omega_{11}^2 p_{ij}(1)^2 = \mathbf{E}[v_{1i}^4 - 3\omega_{11}^2]\sum_{i=1}^n p_{ii}(1)]^2 + 2\sum_{i,j=1}^n \omega_{11}^2 p_{ij}(1)^2$. By using the relations $0 \leq p_{ii} < 1$ and $\sum_{i,j=1}^n p_{ij}^2 = K_{2n}$, we have the result.

Similarly, we expand $\mathbf{g}_{21}$ and $\mathbf{G}_{22}$ as

$$\begin{aligned} g_{21} &= [\mathbf{\Pi}_{1n}'(\mathbf{Z}_1^{(1)'}, \mathbf{Z}_{2n}^{(1)})' + \mathbf{v}_1']\bar{\mathbf{P}}_{Z_1^{(1)}}\mathbf{Z}_{2n}^{(1)}\mathbf{A}_{22.1}^{(2)}\mathbf{Z}_{2n}^{*(1)'}\bar{\mathbf{P}}_{Z_2^{(2}}[(\mathbf{Z}_1^{(2)}, \mathbf{Z}_{2n}^{(2)})\mathbf{\Pi}_{2n} + \mathbf{V}_2] \\ &= \boldsymbol{\pi}_{21}^{(n)'}\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)}\mathbf{\Pi}_{22}^{(n)} + \boldsymbol{\pi}_{21}^{'(n)}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2 + \mathbf{v}_1'\mathbf{Z}_{2n}^{*(1)}\mathbf{A}_{22.1}^{(1)-1}\mathbf{A}_{22.1}^{(2)}\mathbf{\Pi}_{22}^{(n)} \\ &\qquad + \mathbf{v}_1'\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2 \ , \end{aligned}$$

and

$$\begin{aligned}\mathbf{G}_{22} &= [\mathbf{\Pi}_{2n}^{'}(\mathbf{Z}_1^{(2)'},\mathbf{Z}_{2n}^{(2)})'+\mathbf{V}_2^{'}]\bar{\mathbf{P}}_{Z_1^{(2)}}\mathbf{Z}_{2n}^{(2)}\mathbf{A}_{22.1}^{(2)}\mathbf{Z}_{2n}^{*(2)'}\bar{\mathbf{P}}_{Z_2^{(2}}[(\mathbf{Z}_1^{(2)},\mathbf{Z}_{2n}^{(2)})\mathbf{\Pi}_{2n}+\mathbf{V}_2] \\ &= \mathbf{\Pi}_{22}^{(n)'}\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)}\mathbf{\Pi}_{22}^{(n)}+\mathbf{\Pi}_{22}^{(n)'}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2+\mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}\mathbf{\Pi}_{22}^{(n)} \\ &\qquad +\mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\ .\end{aligned}$$

For h_{11} and $\mathbf{H}_{22}$, we use the expressions :

$$h_{11}=[\mathbf{\Pi}_{1n}^{'}(\mathbf{Z}_1^{(1)'},\mathbf{Z}_{2n}^{(1)})'+\mathbf{v}_1^{'}]\bar{\mathbf{P}}_{(Z_1^{(1)},Z_{2n}^{(1)})}[(\mathbf{Z}_1^{(1)},\mathbf{Z}_{2n}^{(1)})\mathbf{\Pi}_{1n}+\mathbf{v}_1]=\mathbf{v}_1^{'}\bar{\mathbf{P}}_{(Z_1^{(1)},Z_{2n}^{(1)})}\mathbf{v}_1\ ,$$

and

$$\mathbf{H}_{22}=[\mathbf{\Pi}_{2n}^{'}(\mathbf{Z}_1^{(2)'},\mathbf{Z}_{2n}^{(2)})'+\mathbf{V}_2^{'}]\bar{\mathbf{P}}_{(Z_1^{(2)},Z_{2n}^{(2)})}[(\mathbf{Z}_1^{(2)},\mathbf{Z}_{2n}^{(2)})\mathbf{\Pi}_{2n}+\mathbf{V}_2]=\mathbf{V}_2^{'}\bar{\mathbf{P}}_{(Z_1^{(2)},Z_{2n}^{(2)})}\mathbf{V}_2\ .$$

Then, we evaluate each terms of $\mathbf{G}/n$ and $\mathbf{H}/q_n$ as n and K_{2n} increase. The first and last terms of each components of $\mathbf{G}/n$ are

$$\text{(A.4)}\qquad \frac{1}{n}\begin{bmatrix}\boldsymbol{\pi}_{21}^{(n)'}\mathbf{A}_{22.1}^{(1)}\boldsymbol{\pi}_{21}^{(n)} & \boldsymbol{\pi}_{21}^{(n)'}\mathbf{A}_{22.1}^{(2)}\mathbf{\Pi}_{22}^{(n)}\\ \mathbf{\Pi}_{22}^{(n)'}\mathbf{A}_{22.1}^{(2)}\boldsymbol{\pi}_{21}^{(n)} & \mathbf{\Pi}_{22}^{(n)'}\mathbf{A}_{22.1}^{(2)}\mathbf{\Pi}_{22}^{(n)}\end{bmatrix}\ ,$$

and

$$\text{(A.5)}\qquad \frac{1}{n}\begin{bmatrix}\mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\\ \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\end{bmatrix}\ ,$$

respectively.
The block-diagonal elements of the second matrix are $(K_{2n}/n)(1/K_{2n})\mathbf{v}_1^{'}\mathbf{P}_{Z_{2n}^{*(1)}}\mathbf{v}_1$ and $(K_{2n}/n)(1/K_{2n})\mathbf{V}_2^{'}\mathbf{P}_{Z_{2n}^{*(2}}\mathbf{V}_2$, and then, by using $\mathrm{tr}(\mathbf{P}_{Z_{2n}^{*(k)}})=K_{2n}$ $(k=1,2)$, they converge to $c\ \omega_{11}$ and $c\ \mathbf{\Omega}_{22}$, respectively. Since $\mathbf{v}_1$ and $\mathbf{V}_2$ are mutually independent, the off-block-diagonal parts converge to zeros in probability. By using the relation (2.4) and Condition (II), as $n\longrightarrow\infty$,

$$\text{(A.6)}\qquad \frac{1}{n}\mathbf{G}\stackrel{p}{\longrightarrow}\mathbf{G}_0=\begin{bmatrix}\boldsymbol{\beta}_2^{'}\\ \mathbf{I}_{G_2}\end{bmatrix}\mathbf{\Phi}_{22.1}(\boldsymbol{\beta}_2,\mathbf{I}_{G_2})+c\ \mathbf{\Omega}\ .$$

Since h_{11} and $\mathbf{H}_{22}$ are quadratic forms of $\mathbf{v}_1$ and $\mathbf{V}_2$, and $\mathrm{tr}(\bar{\mathbf{P}}_{(Z_1^{(k)},Z_{2n}^{(k)})})=q_n$ $(k=1,2)$, by using a similar argument, we have

$$\text{(A.7)}\qquad \frac{1}{q_n}\mathbf{H}\stackrel{p}{\longrightarrow}\mathbf{\Omega}=\begin{pmatrix}\omega_{11} & \mathbf{0}'\\ \mathbf{0} & \mathbf{\Omega}_{22}\end{pmatrix}\ .$$

Then, by using that λ_n is the minimum root of (2.11), $\lambda_n\stackrel{p}{\to}c$ as $n\to\infty$ and we find that $\hat{\boldsymbol{\beta}}_{2.LVR}\stackrel{p}{\to}\boldsymbol{\beta}$ as $n\to\infty$.

Define $\mathbf{G}_1$, $\mathbf{H}_1$, λ_{1n} and $\mathbf{b}_1$ by $\mathbf{G}_1 = \sqrt{n}(\frac{1}{n}\mathbf{G} - \mathbf{G}_0)$, $\mathbf{H}_1 = \sqrt{q_n}(\frac{1}{q_n}\mathbf{H} - \boldsymbol{\Omega})$, $\lambda_{1n} = \sqrt{n}(\lambda_n - c)$, $\mathbf{b}_1 = \sqrt{n}(\hat{\boldsymbol{\beta}}_{LVR} - \boldsymbol{\beta})$. From (2.10),

$$(A.8)\qquad [\mathbf{G}_0 - c\,\boldsymbol{\Omega}]\boldsymbol{\beta} + \frac{1}{\sqrt{n}}[\mathbf{G}_1 - \lambda_{1n}\boldsymbol{\Omega}]\boldsymbol{\beta} + \frac{1}{\sqrt{n}}[\mathbf{G}_0 - c\,\boldsymbol{\Omega}]\mathbf{b}_1 + \frac{1}{\sqrt{q_n}}[-c\mathbf{H}_1]\boldsymbol{\beta} = o_p(\frac{1}{\sqrt{n}}) .$$

Since $(\mathbf{G}_0 - c\,\boldsymbol{\Omega})\boldsymbol{\beta} = \mathbf{0}$ and $\hat{\boldsymbol{\beta}}_{LVR}^{'} = (1, -\hat{\boldsymbol{\beta}}_{2.LVR}^{'})$, (2.10) gives

$$(A.9)\qquad \begin{bmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{bmatrix} \boldsymbol{\Phi}_{22.1}\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LVR} - \boldsymbol{\beta}_2) = (\mathbf{G}_1 - \lambda_{1n}\boldsymbol{\Omega} - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\beta} + o_p(1) .$$

Multiplication of (A.8) from the left by $\boldsymbol{\beta}^{'} = (1, -\boldsymbol{\beta}_2^{'})$ yields

$$(A.10)\qquad \lambda_{1n} = \frac{\boldsymbol{\beta}^{'}(\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\beta}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}} + o_p(1) .$$

Also multiplication of (A.8) from the left by $(\mathbf{0}, \mathbf{I}_{G_2})$ and substitution for λ_{1n} from (A.10) yields

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LVR} - \boldsymbol{\beta}_2) &= \boldsymbol{\Phi}_{22.1}^{-1}(\mathbf{0}, \mathbf{I}_{G_2})(\mathbf{G}_1 - \lambda_{1n}\boldsymbol{\Omega} - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\beta} + o_p(1) \\ &= \boldsymbol{\Phi}_{22.1}^{-1}(\mathbf{0}, \mathbf{I}_{G_2})[\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}}](\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\beta} + o_p(1) . \end{aligned}$$

Define the vector $\mathbf{u}$ such that $\mathbf{V}\boldsymbol{\beta} = \mathbf{u}$ and $\mathbf{V} = (\mathbf{v}_1, \mathbf{V}_2)$. From the above expression, we need to evaluate each terms of $(\mathbf{G}_1 - \sqrt{cc_*}\mathbf{H}_1)\boldsymbol{\beta}$ and there are five terms, but we have evaluated each elements of the first term by (A.6) and (A.7). The second term

$$(A.11)\qquad \frac{1}{\sqrt{n}} \begin{bmatrix} \mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}\boldsymbol{\pi}_{21}^{(n)} & \mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)'})^{-1}\mathbf{A}_{22.1}^{(2)}\boldsymbol{\Pi}_{22}^{(n)} \\ \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}\boldsymbol{\pi}_{21}^{(n)} & \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}\boldsymbol{\Pi}_{22}^{(n)} \end{bmatrix} \begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{bmatrix}$$

is stochastically negligible due to (2.4) under Condition (II) because
$(1/\sqrt{n})\mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(1)'}[\mathbf{I}_{G_2} - \mathbf{Z}_{2n}^{*(1)'})^{-1}\mathbf{A}_{22.1}^{(2)}]\boldsymbol{\Pi}_{22}^{(n)} \xrightarrow{p} 0$.
Then, we need to evaluate three terms for the asymptotic distribution of the normalized 2SLVR estimator. (Two terms due to $\mathbf{G}_1$ and one term due to $\mathbf{H}_1$.) The first term has the form

$$\mathbf{gh}(1) = \frac{1}{\sqrt{n}} \begin{bmatrix} \boldsymbol{\pi}_{21}^{(n)'}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \boldsymbol{\pi}_{21}^{(n)'}\mathbf{Z}_{2n}^{*(1)'}\mathbf{V}_2 \\ \boldsymbol{\Pi}_{22}^{'(n)}\mathbf{A}_{22.1}^{(2)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)'})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \boldsymbol{\Pi}_{22}^{(n)'}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2 \end{bmatrix} \begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{bmatrix} ,$$

which is asymptotically equivalent to

$$(A.12)\qquad \mathbf{gh}^*(1) = \begin{bmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{bmatrix} \frac{1}{\sqrt{n}}\boldsymbol{\Pi}_{22}^{(n)'}\left[\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 - \mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\boldsymbol{\beta}_2\right] .$$

It is because the variance-covariance matrix of the difference $(1/\sqrt{n})\mathbf{\Pi}_{22}^{'(n)}[\mathbf{A}_{22.1}^{(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)'})^{-1}\mathbf{Z}_{2n}^{*(1)'}-\mathbf{I}_n]\mathbf{v}_1$ is given by

$$
\begin{aligned}
& \omega_{11}\frac{1}{n}\mathbf{\Pi}_{22}^{(n)'}[\mathbf{A}_{22.1}^{(2)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)'})^{-1}\mathbf{Z}_{2n}^{*(1)'}-\mathbf{I}_n]\mathbf{\Pi}_{22}^{(n)'}[(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)'})^{-1}\mathbf{A}_{22.1}^{(2)}-\mathbf{I}_n]\mathbf{Z}_{2n}^{*(1)'}\mathbf{\Pi}_{22}^{(n)} \\
= \; & \omega_{11}\frac{1}{n}\mathbf{\Pi}_{22}^{(n)'}[\mathbf{A}_{22.1}^{(2)}(\mathbf{A}_{22.1}^{*(1)'})^{-1}\mathbf{A}_{22.1}^{(2)}-\mathbf{A}_{22.1}^{(1)}]\mathbf{\Pi}_{22}^{(n)} ,
\end{aligned}
$$

which is stochastically $o_p(1)$ under Condition (II*).
The asymptotic covariance matrix of the first term ($\mathbf{A}\times\mathbf{gh}^*(1)$) is given by

$$
\frac{1}{n}\mathbf{A}[\begin{array}{c}\boldsymbol{\pi}_{21}^{(}n)'\\ \mathbf{\Pi}_{22}^{(n)'}\end{array}][\omega_{11}\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)}+(\boldsymbol{\beta}_2^{'}\mathbf{\Omega}_{22}\boldsymbol{\beta}_2)\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)}][\boldsymbol{\pi}_{21}^{(n)},\mathbf{\Pi}_{22}^{(n)}]\mathbf{A}'\xrightarrow{p}\sigma^2\mathbf{\Phi}_{22.1}
$$

by using Condition (II).
We use the standard CLT to $\mathbf{gh}^*(1)$ instead of the first tem $\mathbf{gh}(1)$. By applying the CLT to

$$
(\frac{1}{\sqrt{n}})\mathbf{a}'\mathbf{\Pi}_{22}^{(n)'}[\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1-\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\boldsymbol{\beta}_2]=(\frac{1}{\sqrt{n}})\sum_{i=1}^{n}[t_{11,i}^{(n)}v_{1i}-t_{12,i}^{(n)}\boldsymbol{\beta}_2^{'}\mathbf{v}_{2i}] \tag{A.13}
$$

for any vector $\mathbf{a}$, $t_{11,i}^{(n)}=\mathbf{a}'\Phi_{22.1}^{-1}\mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{2n,i}^{*(1)}$ and $t_{12,i}^{(n)}=\mathbf{a}'\Phi_{22.1}^{-1}\mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{2n,i}^{*(2)}$ with Condition (IV) to have the asymptotic normality. (Condition (IV) is needed to have the Lindberg-type condition for Part (i).

The second term has the form

$$
\mathbf{gh}(2)=\frac{1}{\sqrt{K_{2n}}}\left[\begin{array}{cc} gh_{11}(2) & \mathbf{gh}_{12}(2)\\ \mathbf{gh}_{21}(2) & \mathbf{gh}_{22}(2)\end{array}\right]\left[\begin{array}{c}1\\ -\boldsymbol{\beta}_2\end{array}\right] , \tag{A.14}
$$

where $gh_{11}(2)=\mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)'})^{-1}\mathbf{Z}_{2n}^{*(1)}\mathbf{v}_1-K_{2n}\omega_{11}$, $\mathbf{gh}_{12}(2)=\mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{A}_{22.1}^{(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2$, $\mathbf{gh}_{21}(2)=\mathbf{gh}_{12}(2)'$, and $\mathbf{gh}_{22}(2)=\mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2-K_{2n}\mathbf{\Omega}_{22}$.
The third term has the form

$$
\mathbf{gh}(3)=\frac{1}{\sqrt{q_n}}\left[\begin{array}{c} gh_1(3)\\ \mathbf{gh}_2(3)\end{array}\right]+o_p(1) , \tag{A.15}
$$

where $gh_1(3)=\mathbf{v}_1^{'}[\mathbf{I}_n-\mathbf{P}_{(Z_1^{(1)},Z_{2n}^{(1)})}]\mathbf{v}_1-q_n\omega_{11}$, and $\mathbf{gh}_2(3)=-(\mathbf{V}_2^{'}[\mathbf{I}_n-\mathbf{P}_{(Z_2^{(2)},Z_{2n}^{(2)})}]\mathbf{V}_2\boldsymbol{\beta}_2-q_n\mathbf{\Omega}_{22}\boldsymbol{\beta}_2)$.

Because each of the second and third terms are quadratic forms of $\mathbf{v}_1$ and $\mathbf{V}_2$, their covariances depend on the fourth-order moments of disturbances in the general case.
We utilize the representation of

$$
\sqrt{n}\mathbf{\Phi}_{22.1}(\hat{\boldsymbol{\beta}}_{2.LVR}-\boldsymbol{\beta}_2)=\mathbf{A}[\mathbf{gh}(1)+\sqrt{c}\mathbf{gh}(2)-\sqrt{cc_*}\mathbf{gh}(3)]+o_p(1) , \tag{A.16}
$$

where $\mathbf{A} = (\mathbf{0}, \mathbf{I}_{G_2})[\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}}]$.

To evaluate the covariance matrices of the second and third terms ($\mathbf{A}\sqrt{c}\mathbf{gh}(2)$ and $\mathbf{A}(\sqrt{cc_*})\mathbf{gh}(3)$), and their asymptotic distributions, we prepare two lemmas.

Lemma 1 : Let $\mathbf{w}_i$ are p-dimensional i.i.d. random vectors with $\mathbf{E}[\mathbf{w}_i] = \mathbf{0}$ and $\mathbf{E}[\mathbf{w}_i\mathbf{w}_i^{'}] = \boldsymbol{\Sigma}$ and the fourth-order moments exist. Let $\mathbf{W} = (\mathbf{w}_i^{'})$ be $n \times p$ random matrix and $\mathbf{w}_i$ are mutually uncorrelated. For any symmetric matrices $\mathbf{A}$,$\mathbf{B}$ and any vectors $\mathbf{a}_k, \mathbf{b}_k$ $(k = 1, 2)$,

$$\mathbf{Cov}[\mathbf{a}_1^{'}\mathbf{W}^{'}\mathbf{A}\mathbf{W}\mathbf{a}_2, \mathbf{b}_1^{'}\mathbf{W}^{'}\mathbf{B}\mathbf{W}\mathbf{b}_2] = \sum_{i=1}^{n} a_{ii}b_{ii}\mathbf{F} + \mathrm{tr}(\mathbf{A}\mathbf{B})[\mathbf{a}_1^{'}\boldsymbol{\Sigma}\mathbf{b}_1\mathbf{a}_2^{'}\boldsymbol{\Sigma}\mathbf{b}_2 + \mathbf{a}_1^{'}\boldsymbol{\Sigma}\mathbf{b}_2\mathbf{a}_2^{'}\boldsymbol{\Sigma}\mathbf{b}_1]$$

and

$$\mathbf{F} = \mathbf{a}_1^{'}\left[\mathbf{E}[\mathbf{w}\mathbf{w}^{'}\mathbf{a}_2\mathbf{b}_2^{'}\mathbf{w}\mathbf{w}^{'}] - [\boldsymbol{\Sigma}(\mathbf{a}_2\mathbf{b}_2^{'} + \mathbf{b}_2\mathbf{a}_2^{'})\boldsymbol{\Sigma} + \boldsymbol{\Sigma}(\mathbf{a}_2^{'}\boldsymbol{\Sigma}\mathbf{b}_2]\right]\mathbf{b}_1 \ .$$

Proof of Lemma 1 : We re-write the summation

$$\mathbf{a}_1^{'}\mathbf{W}^{'}\mathbf{A}\mathbf{W}\mathbf{a}_2\mathbf{b}_1^{'}\mathbf{W}^{'}\mathbf{B}\mathbf{W}\mathbf{b}_2 = \sum_{i,j,k,l=1}^{n} \mathbf{a}_1^{'}\mathbf{w}_i a_{ij}\mathbf{a}_2^{'}\mathbf{w}_j\mathbf{b}_1^{'}\mathbf{w}_k b_{kl}\mathbf{b}_2^{'}\mathbf{w}_l \tag{A.17}$$

Then, we take the expectations of each terms of summations with the cases of (i) $i = j = k = l$, (ii) $i = j \neq k = l$, (iii) $i = k \neq j = l$ and (iv) $i = l \neq j = k$. By direct calculations, the first term is $\sum_{i=1}^{n} a_{ii}b_{ii}\mathbf{E}[(\mathbf{a}_1^{'}\mathbf{w}_i)(\mathbf{a}_2^{'}\mathbf{w}_i)(\mathbf{b}_1^{'}\mathbf{w}_i\mathbf{b}_2^{'}\mathbf{w}_i]$, the second term is $\sum_{i=j\neq k=l}^{n} a_{ii}b_{kk}\mathbf{E}[\mathbf{a}_1^{'}\boldsymbol{\Sigma}\mathbf{a}_2\mathbf{b}_1^{'}\boldsymbol{\Sigma}(\mathbf{b}_2^{'}]$, the third term is $\sum_{i=k\neq j=l}^{n} a_{ij}b_{ij}\mathbf{E}[(\mathbf{a}_1^{'}\boldsymbol{\Sigma}\mathbf{b}_1\mathbf{a}_2^{'}\boldsymbol{\Sigma}\mathbf{b}_2]$, and the fourth term is $\sum_{i=l\neq j=k}^{n} a_{ij}b_{ji}\mathbf{E}[\mathbf{a}_1^{'}\boldsymbol{\Sigma}\mathbf{b}_2\mathbf{a}_2^{'}\boldsymbol{\Sigma}\mathbf{b}_1]$.
By using $\mathbf{E}[\mathbf{a}_1^{'}\mathbf{W}^{'}\mathbf{A}\mathbf{W}\mathbf{a}_2] = \mathbf{E}[\sum_{i,j=1}^{n}\mathbf{a}_1^{'}\mathbf{w}_i a_{ij}\mathbf{a}_2^{'}\mathbf{w}_j] = \sum_{i=,j=1}^{n} a_{ii}\mathbf{a}_1^{'}\boldsymbol{\Sigma}\mathbf{a}_2 = \mathrm{tr}(\mathrm{A})\mathbf{a}_1^{'}\boldsymbol{\Sigma}\mathbf{a}_2$, $\mathbf{E}[\mathbf{b}_1^{'}\mathbf{W}^{'}\mathbf{B}\mathbf{W}\mathbf{b}_2] = \mathrm{tr}(\mathrm{B})\mathbf{b}_1^{'}\boldsymbol{\Sigma}\mathbf{b}_2$, and by summarizing each terms, we have the result.
(**Q.E.D. of Lemma 1**)

Lemma 2 : Let $t_{11}^{(n)} = (1/\sqrt{n})\sum_{i=1}^{n} t_{11,i}v_{1i}$ and $t_{12}^{(n)} = (1/\sqrt{n})\sum_{i=1}^{n} t_{12,i}\boldsymbol{\beta}_2^{'}\mathbf{v}_{2i}$ in (A.13).
For (A.14), let

$$t_{21}^{(n)} = \frac{1}{\sqrt{K_{2n}}}\sum_{i,j=1}^{n} p_{ij}(1)[v_{1i}v_{1j} - \omega_{11}\delta(i,j)] - \frac{1}{\sqrt{K_{2n}}}\sum_{i,j=1}^{n} p_{ij}^{*}(1,2)[v_{1i}(\boldsymbol{\beta}_2^{'}\mathbf{v}_{2j})] \ ,$$

and

$$t_{22}^{(n)} = \frac{1}{\sqrt{K_{2n}}}\sum_{i,j=1}^{n} p_{ij}^{*}(2,1)\mathbf{v}_{2i}v_{1j} - \frac{1}{\sqrt{K_{2n}}}\sum_{i,j=1}^{n} p_{ij}(2)[\mathbf{v}_{2i}\mathbf{v}_{2j}^{'} - \boldsymbol{\Omega}_{22}]\boldsymbol{\beta}_2 \ ,$$

where we define

$$p^*_{ij}(1,2)\ (=p^*_{ji}(2,1)) = (\mathbf{Z}^{*(1)}_{2n}(\mathbf{A}^{(2)}_{22.1})^{-1}\mathbf{Z}^{*(2)'}_{2n})_{ij}\ .$$

For (A.15), let

$$t^{(n)}_{31} = \frac{1}{\sqrt{q_n}}\sum_{i,j=1}^{n}[\delta(i,j)-q_{ij}(1)][v_{1i}v_{1j}-\omega_{11}\delta(i,j)]\ ,$$

and

$$t^{(n)}_{32} = \frac{1}{\sqrt{q_n}}\sum_{i,j=1}^{n}[\delta(i,j)-q_{ij}(2)][\mathbf{v}_{2i}\mathbf{v}'_{2j}-\delta(i,j)\mathbf{\Omega}_{22}]\boldsymbol{\beta}_2\ ,$$

where $q_{ij}(k) = (\mathbf{P}_{(Z^{(k)}_1, Z^{(k)}_{2n})})_{ij}$ $(k=1,2)$.
Then, under the conditions in Theorem 1, $t^{(n)}_{11}$, $t^{(n)}_{12}$, $t^{(n)}_{21}$, $t^{(n)}_{22}$, $t^{(n)}_{31}$ and $t^{(n)}_{32}$ have the asymptotic normality.

A Scketch of Proof of Lemma 2 :
The proof of Lemma 2 is based on straightforward, but the results of lengthy calculations, which are quite similar to Lemma 3 of AKM (2011). We have shown the asymptotic normality of $t^{(n)}_{11}$ and $t^{(n)}_{12}$ already. The proof of other terms is the results of applying the martingale CLT (Theorem 3.5 of Hall and Heyde (1980) for instance). We illustrate an outline of the proof for the first term of $t^{(n)}_{21}$ as a typical case. (The derivations of the asymptotic normality for other terms are similar and we have omitted the details.)
Let $\mathcal{F}_{n.i}$ be the σ-field generated by the random variables $\mathbf{v}_{1j}$ $(j \le i \le n)$ and $\mathcal{F}_{n,0}$ be the initial $\sigma-$ field. We define a sequence of martingale differences as $X_{ni} = (1/\sqrt{K_{2n}})[(v^2_{1i}-\omega_{11})p_{ii}(1) + 2\sum_{i>j=1} v_{1i}v_{1j}p_{ij}(1)]$ for $i=1,\cdots,n$. Then $\mathbf{E}[X_{ni}|\mathcal{F}_{n,i-1}] = 0$ and $\mathbf{E}[X^2_{ni}|\mathcal{F}_{n,i-1}] = \frac{1}{K_{2n}}\mathbf{E}[(v_{1i}-\omega_{11})^2]p_{ii}(1)^2 + 4\omega^2_{11}\frac{1}{K_{2n}}\sum_{i>j=1}p_{ij}(1)^2$.
Then, we need to show that

$$\sum_{i=1}^{n}\mathbf{E}[X^2_{ni}|\mathcal{F}_{n,i-1}] - \sum_{i=1}^{n}\mathbf{E}[X^2_{ni}] \xrightarrow{p} 0\ . \tag{A.18}$$

This can be done by decomposing

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}[\sum_{j=1}^{i-1}v_{1j}p_{ij}(1)]^2 - \frac{1}{n}[\omega_{11}\sum_{j=1}^{i-1}p_{ij}(1)^2] \\
= \ & \frac{1}{n}[\sum_{j=j'=1}^{i-1}(v^2_{1j}-\omega_{11})p_{ij}(1)^2 + 2\sum_{j>j'=1}^{i-1}v_{1j}v_{11j'}p_{ij}(1)p_{ij'}(1)]
\end{aligned}$$

and evaluating the variances of each terms.
The variance of the first term is less than $(1/n)^2\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{E}[(v^2_{1i}-\omega_{11})^2]p_{ij}(1)^2$,

which converges to zero in probability because the matrix (p_{ij}) is a projection. The variance of second term is less than $\omega_{11}^2(1/n)^2\sum_{i=1}^n\sum_{j,j'=1}^n p_{ij}(1)^2p_{ij'}(1)^2$ which converges to zero in probability because the matrix $(p_{ij}(1))$ is a projection. Finally, we utilize that for any $\xi>0$ and $\nu>0$, $\sum_{i=1}^n \mathbf{E}[X_{ni}^2 \mathrm{I}(|X_{ni}\geq \xi)] \leq (1/\xi)^{\nu}\sum_{i=1}^n \mathbf{E}[X_{ni}^{2+\nu}]$ and the moment condition $\mathbf{E}[\|\mathbf{v}_i\|^{4+\epsilon}]<\infty$ $(\epsilon>0)$ when $c\neq 0$ in Section 3 to show the Lindeberg-type condition $(\xi\to\infty)$ in the present situation. (**End of A Scketch of Proof of Lemma 2**)

By using Lemma 1, we calculate the asymptotic variance-covariances of $\mathbf{gh}(2)$ $((1/\sqrt{K_{2n}})[gh_{11}(2)-\mathbf{gh}_{12}(2)\boldsymbol{\beta}_2]$ and $(1/\sqrt{K_{2n}})[\mathbf{gh}_{21}(2)-\mathbf{gh}_{22}(2)\boldsymbol{\beta}_2)$, which is given by

$$\begin{aligned}\mathbf{D}_2 &= \begin{bmatrix} \boldsymbol{\Gamma}(v_1)\frac{1}{K_{2n}}\sum_{i=1}^n p_{ii}^2(1) & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Gamma}(\mathbf{v}_2)\frac{1}{K_{2n}}\sum_{i=1}^n p_{ii}^2(2)\end{bmatrix} \\ &+ \begin{bmatrix} 2\omega_{11}^2+\omega_{11}\boldsymbol{\beta}_2'\boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2 & -\boldsymbol{\beta}_2'\boldsymbol{\Omega}_{22} \\ -\boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2 & \omega_{11}\boldsymbol{\Omega}_{22}+(\boldsymbol{\beta}_2'\boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2)\boldsymbol{\Omega}_{22}+\boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{\Omega}_{22}\end{bmatrix},\end{aligned}$$

where we use the notation $p_{ii}(k)=p_{ii}^{(n)}(\mathbf{P}_{Z_{2n}^{*(k)}})$ $(k=1,2;i=1,\cdots,n)$.
By using a simple calculation, the second term of $\mathbf{D}_2$ is equivalent to $\mathbf{D}_2^{(2)}=\sigma^2\boldsymbol{\Omega}+\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\Omega}$. Then, since $\sigma^2=\omega_{11}+\boldsymbol{\beta}_2'\boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2$, we have the relation

$$\mathbf{A}\mathbf{D}_2^{(2)}\mathbf{A}'=\sigma^2[\boldsymbol{\Omega}_{22}-\frac{1}{\sigma^2}\boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2\boldsymbol{\beta}_2\boldsymbol{\Omega}_{22}]\ .$$

By using Lemma 1, we calculate the asymptotic variance-covariances of $\mathbf{gh}(3)$ $(gh_1(3)=(1/\sqrt{q_n})\mathbf{v}_1'[\mathbf{I}_n-\mathbf{P}_{(Z_1^{(1)},Z_{2n}^{(1)})}]\mathbf{v}_1-q_n\omega_{11}$ and
$\mathbf{gh}_2(3)=(1/\sqrt{q_n})[-(\mathbf{V}_2'[\mathbf{I}_n-\mathbf{P}_{(Z_2^{(2)},Z_{2n}^{(2)})}]\mathbf{V}_2\boldsymbol{\beta}_2-q_n\boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2)]$, which is given by

$$\begin{aligned}\mathbf{D}_3 &= \begin{bmatrix} \boldsymbol{\Gamma}(\mathbf{v}_1)\frac{1}{q_n}\sum_{i=1}^n(1-p_{ii}(1))^2 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Gamma}(\mathbf{v}_2)\frac{1}{q_n}\sum_{i=1}^n(1-p_{ii}(2))^2\end{bmatrix} \\ &+ \begin{bmatrix} 2\omega_{11}^2 & \mathbf{0}' \\ \mathbf{0} & (\boldsymbol{\beta}_2'\boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2)\boldsymbol{\Omega}_{22}+\boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{\Omega}_{22}\end{bmatrix}.\end{aligned}$$

We denote the second term of $\mathbf{D}_3$ as $\mathbf{D}_3^{(2)}$ and use the fact that

$$\mathbf{D}_3^{(2)}=2(\omega_{11})^2\mathbf{e}_1\mathbf{e}_2'+\begin{pmatrix}0 & \mathbf{0}'\\ \mathbf{0} & \boldsymbol{\Omega}_{22}\end{pmatrix}\boldsymbol{\beta}\boldsymbol{\beta}'\begin{pmatrix}0 & \mathbf{0}'\\ \mathbf{0} & \boldsymbol{\Omega}_{22}\end{pmatrix}+\boldsymbol{\beta}_2'\boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2\begin{pmatrix}0 & \mathbf{0}'\\ \mathbf{0} & \boldsymbol{\Omega}_{22}\end{pmatrix}\ .$$

Then, by using a simple calculation, we have

$$\mathbf{A}\mathbf{D}_3^{(2)}\mathbf{A}'=\boldsymbol{\beta}_2'\boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2\boldsymbol{\Omega}_{22}+\boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{\Omega}_{22}[4(\frac{\omega_{11}}{\sigma^2})^2-1]\ .$$

For the asymptotic distribution of estimator under non-normal case, we need to evaluate the cross products of the second and third terms, that is, $-\mathbf{A}\mathbf{E}[\mathbf{gh}(2)\mathbf{gh}^{'}(3)+\mathbf{gh}(3)\mathbf{gh}^{'}(2)]\mathbf{A}^{'} = -\mathbf{A}[\mathbf{D}_{23}+\mathbf{D}_{32}]\mathbf{A}^{'}$, where

$$\mathbf{D}_{23} = \begin{bmatrix} \mathbf{\Gamma}(v_1)\frac{1}{\sqrt{K_{2n}q_n}}\sum_{i=1}^{n} p_{ii}(1)(1-p_{ii}(1)) & \mathbf{0}^{'} \\ \mathbf{0} & \mathbf{\Gamma}(\mathbf{v}_2)\frac{1}{\sqrt{K_{2n}q_n}}\sum_{i=1}^{n} p_{ii}(2)(1-p_{ii}(2)) \end{bmatrix}$$

$$+o(1)\ .$$

We consider the effects of fourth-order moments of disturbance terms. Define $c_n = K_{2n}/n$ and then the coefficient of $\mathbf{\Gamma}(v_1)$ becomes

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}[p_{ii}(1)]^2+(\frac{c_n}{1-c_n})^2\frac{1}{n}\sum_{i=1}^{n}[1-p_{ii}(1)]^2-2(\frac{c_n}{1-c_n})\frac{1}{n}\sum_{i=1}^{n}p_{ii}(1)][1-p_{ii}(1)] \\ = \ & (\frac{1}{1-c_n})^2\frac{1}{n}\sum_{i=1}^{n}[p_{ii}(1)-c_n]^2 \end{aligned}$$

by using $\sum_{i=1}^{n} p_{ii}(X_1) = K_{2n}$ and $(1/n)\sum_{i=1}^{n}(p_{ii}(1)-c)^2 = (1/n)\sum_{i=1}^{n}(p_{ii}(1))^2-c_n^2$. Similarly, we find that the coefficient of $\mathbf{\Gamma}(\mathbf{v}_2)$ becomes $(\frac{1}{1-c_n})^2\frac{1}{n}\sum_{i=1}^{n}[p_{ii}(2)-c_n]^2$.

Finally, by using the central limit theorem (CLT) in *Lemma 2* for every constant vector $\mathbf{a}^{'}$ from the left, we have the asymptotic normality with the asymptoticvariance $\mathbf{a}^{'}\Psi^{**}\mathbf{a}$ and it proves (ii) of *Theorem 1.*
Q.E.D

(ii) **Proof of Theorem 2** :

(**Step 1**, The case of $0 \leq \nu < 1/2$, 2SLVR and 2STSLS.)
For the 2SLVR estimation, we use the similar arguments as the proof of Theorem 1, but we need additional derivations because of the possible heteroscedasticity of disturbances. The limiting distribution of the 2STSLS estimator is the same as the2SLVR estimator when $0 \leq \nu < 1/2$.
For the 2STSLS estimation, from (2.13) and (A.6), we use the representation

$$\text{(A.19)} \qquad \left[(\mathbf{\Phi}_{22.1}\boldsymbol{\beta}_2, \mathbf{\Phi}_{22.1}) + \frac{1}{\sqrt{n}}\mathbf{G}_1^{**}\right]\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.TS} \end{bmatrix} = \mathbf{0}\ ,$$

where $\mathbf{G}_1^{**} = \sqrt{n}\left[(\frac{1}{n}\mathbf{g}_{21}, \frac{1}{n}\mathbf{G}_{22}) - (\mathbf{\Phi}_{22.1}\boldsymbol{\beta}_2, \mathbf{\Phi}_{22.1})\right]$.
We make use of the fact that $\mathbf{P}_{Z_{2n}^{*(k)}}$ $(k=1,2)$ is idempotent of rank K_{2n} and the boundedness of $\mathbf{E}[v_{ji}^4|\mathbf{z}_{2n,i}^{*(k)}]$ (k=1,2) implies a Lindeberg condition $\sup_{1\leq i\leq n}\mathbf{E}\left[\mathbf{v}_i^{'}\mathbf{v}_i\mathbf{I}(\mathbf{v}_i^{'}\mathbf{v}_i > a)|\mathbf{z}_{2n,1}^{*(k)}, \cdots, \mathbf{z}_{2n,n}^{*(k)}\right] \xrightarrow{p} 0 \ \ (a \to \infty)$. By taking the expectation of $\mathbf{v}_1^{'}\mathbf{P}_{\mathbf{Z}_{2n}^{*(1)}}\mathbf{v}_1$, and $\mathbf{E}[\mathbf{v}_i^{(n)}\mathbf{v}_i^{(n)'}|z_{2n,1}^{*(1)}, \cdots, z_{2n,n}^{*(1)}]$ is bounded. Then, there is a

(constant) $\bar{\omega}_{11}$ such that

$$\text{(A.20)} \qquad \mathbf{E}[\frac{1}{\sqrt{n}}\mathbf{v}_1^{'}\mathbf{P}_{\mathbf{Z}_{2n}^{*(1)}}\mathbf{v}_1] = \mathbf{E}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_{11,i}^{(n)}p_{ii}^{(n)}(1)] \leq \frac{K_{2n}}{\sqrt{n}}\bar{\omega}_{11} \longrightarrow 0$$

when $0 < \nu < 1/2$.
Hence, by using the same argument, we find that $(1/\sqrt{n})\mathbf{v}_1^{'}\mathbf{P}_{\mathbf{Z}_{2n}^{*(1)}}\mathbf{v}_1 \xrightarrow{p} 0$ and $(1/\sqrt{n})\mathbf{V}_2^{'}\mathbf{P}_{\mathbf{Z}_{2n}^{*(2)}}\mathbf{V}_2 \xrightarrow{p} \mathbf{O}$ as $n \to \infty$.
For the 2SLVR estimation, (2.13) implies

$$\text{(A.21)} \qquad \left[\begin{pmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{pmatrix} \boldsymbol{\Phi}_{22.1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}) + (\frac{1}{\sqrt{n}}\mathbf{G}_1^* - \lambda_n\frac{1}{q_n}\mathbf{H})\right] \begin{pmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.LVR} \end{pmatrix} = \mathbf{0}\ ,$$

where $\mathbf{G}_1^* = \sqrt{n}[(1/n)\mathbf{G} - \mathbf{G}_0]$.
By using the facts that $(1/\sqrt{n})\mathbf{G}_1^* \xrightarrow{p} \mathbf{O}$, $\lambda_n \xrightarrow{p} 0$ (Lemma 3 below) and $[1/q_n]\mathbf{H} \xrightarrow{p} \mathbf{\Omega}$, we have

$$\boldsymbol{\Phi}_{22.1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2})\text{plim}_{n\to\infty}\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.TS} \end{bmatrix} = \mathbf{0}\ ,\ \boldsymbol{\Phi}_{22.1}(\boldsymbol{\beta}_2, \mathbf{I}_{G_2})\text{plim}_{n\to\infty}\begin{bmatrix} 1 \\ -\hat{\boldsymbol{\beta}}_{2.LVR} \end{bmatrix} = \mathbf{0}\ ,$$

which imply $\text{plim}_{n\to\infty}\hat{\boldsymbol{\beta}}_{2.TS} = \boldsymbol{\beta}_2$ and $\text{plim}_{n\to\infty}\hat{\boldsymbol{\beta}}_{2.LVR} = \boldsymbol{\beta}_2$ because $\boldsymbol{\Phi}_{22.1}$ is positive definite under Condition (II).
Then, by using (2.13), we use the representation

$$\sqrt{n}\left[\begin{pmatrix} \boldsymbol{\beta}_2^{'} \\ \mathbf{I}_{G_2} \end{pmatrix} \boldsymbol{\Phi}_{22.1}\left(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}\right) + (\frac{1}{\sqrt{n}}\mathbf{G}_1^* - \lambda_n\frac{1}{q_n}\mathbf{H})\right]\left[\boldsymbol{\beta} + (\hat{\boldsymbol{\beta}}_{LVR} - \boldsymbol{\beta})\right] = \mathbf{0}.$$

We prepare the following lemma, which is the same as Lemma 4 in AKM (2010) and the proof is omitted)

Lemma 3 : Let λ_n $(n > 2)$ be the smallest root of (2.11). (i) For $0 < \xi < 1 - \nu$ and $0 \leq \nu < 1$,

$$\text{(A.22)} \qquad n^{\xi}\lambda_n \xrightarrow{p} 0$$

as $n \to \infty$. (ii) For $0 \leq \nu < 1$,

$$\text{(A.23)} \qquad \sqrt{n}\left[\lambda_n - \frac{K_{2n}}{n}\right] \xrightarrow{p} 0$$

as $n \longrightarrow \infty$.

Since $\sqrt{n}\,\lambda_n \xrightarrow{p} 0$ when $0 \le \nu < 1/2$ due to Lemma 3, the asymptotic distributions of the 2LVR and 2TSLS estimators are equivalent, and we have

$$\text{(A.24)} \qquad \mathbf{\Phi}_{22.1}\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LVR} - \boldsymbol{\beta}_2) - (\mathbf{0}, \mathbf{I}_{G_2})\mathbf{G}_1^{*}\boldsymbol{\beta} \xrightarrow{p} \mathbf{0} \; .$$

We notice that for $k = 1, 2$

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n} \mathbf{\Omega}_i^{(n)} \otimes \mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{2n,i}^{*(k)}\mathbf{z}_{2n,i}^{*(k)'}\mathbf{\Pi}_{22}^{(n)} - \mathbf{\Omega} \otimes \mathbf{\Phi}_{22.1} \\ = \; & \frac{1}{n}\sum_{i=1}^{n}(\mathbf{\Omega}_i^{(n)} - \mathbf{\Omega}) \otimes \mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{2n,i}^{*(k)}\mathbf{z}_{2n,i}^{*(k)'}\mathbf{\Pi}_{22}^{(n)} \\ & + \frac{1}{n}\sum_{i=1}^{n} \mathbf{\Omega} \otimes \left[\mathbf{\Pi}_{22}^{(n)'}\mathbf{z}_{2n,i}^{*(k)}\mathbf{z}_{2n,i}^{*(k)'}\mathbf{\Pi}_{22}^{(n)} - \mathbf{\Phi}_{22.1}\right] \xrightarrow{p} \mathbf{O} \end{aligned}$$

because Condition (II) and the conditions imposed on $\mathbf{\Omega}_i^{(n)}$ $(i = 1, \cdots, n)$.
With the notation of the proof of Theorem 1, $\sqrt{n}[\hat{\boldsymbol{\beta}}_{2.LVR} - \boldsymbol{\beta}_2]$ is asymptotically equivalent to $\mathbf{A} \times \mathbf{gh}(1)$, which in turn is asymptotically equivalent to $\mathbf{A} \times \mathbf{gh}^{*}(1)$. By applying the CLT to $(1/\sqrt{n})\mathbf{\Pi}_{22}^{(n)'}[\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 - \mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\boldsymbol{\beta}_2]$, we obtain the limiting normal distribution $N(\mathbf{0}, \sigma^2\mathbf{\Phi}_{22.1})$. (Condition (IV) is needed to have the Lindberg-type condition.)
This proves the first part of (i) in *Theorem 2* for the 2SLVR estimator and the last part (ii) for the 2STSLS estimator when $0 \le \nu < 1/2$.

(**Step 2**, The case of $1/2 < \nu < 1$, 2SLVR.)
We consider the asymptotic distribution of the 2SLVR estimator when $1/2 \le \nu < 1$. In this case, there is a complication because there are four terms in each elements of $\mathbf{G}$ (g_{11}, $\mathbf{g}_{21}$, $\mathbf{G}_{22}$) and their stochastic orders are $O_p(n)$, $O_p(\sqrt{n})$, $O_p(\sqrt{n})$ and $O_p(n^{\nu})$ with possible heteroscedasticity of disturbances.
By using Part (ii) of Lemma 3 and the facts that $\lambda_n \xrightarrow{p} 0$ and $[n/K_{2n}]\times$(A.8) converges to $\mathbf{\Omega}$ in probability, we have $\hat{\boldsymbol{\beta}}_{2.LVR} - \boldsymbol{\beta}_2 \xrightarrow{p} \mathbf{0}$. By multiplying $\boldsymbol{\beta}^{'}$ from the left to (A.8), we have

$$\begin{aligned} & \boldsymbol{\beta}^{'}\left\{\sqrt{n}[\frac{K_{2n}}{n} - \lambda_n]\mathbf{\Omega} + \frac{1}{\sqrt{n}}\begin{bmatrix} \mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}\boldsymbol{\pi}_{21}^{(n)} & \mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)'})^{-1}\mathbf{A}_{22.1}^{(2)}\mathbf{\Pi}_{22}^{(n)} \\ \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}\boldsymbol{\pi}_{21}^{(n)} & \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}\mathbf{\Pi}_{22}^{(n)} \end{bmatrix}\right. \\ & \left. + \frac{1}{\sqrt{n}}\begin{bmatrix} \boldsymbol{\pi}_{21}^{(n)'}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \boldsymbol{\pi}_{21}^{(n)'}\mathbf{Z}_{2n}^{*(1)'}\mathbf{V}_2 \\ \mathbf{\Pi}_{22}^{'(n)}\mathbf{A}_{22.1}^{(2)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)'})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \mathbf{\Pi}_{22}^{(n)'}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2 \end{bmatrix} - \lambda_n\sqrt{\frac{n}{q_n}}\mathbf{H}_1\right\} \\ & \times\left[\boldsymbol{\beta} + (\hat{\boldsymbol{\beta}}_{2.LVR} - \boldsymbol{\beta})\right] \sim o_p(1) \; . \end{aligned}$$

Multiply (A.8) with c_ν instead of c on the left by $(\mathbf{0}, \mathbf{I}_{G_2})$ to obtain approximately

$$(\mathbf{0},\mathbf{I}_{G_2})\sqrt{n}\left\{\left[\left(\begin{array}{c}\boldsymbol{\beta}_2^{'}\\ \mathbf{I}_{G_2}\end{array}\right)\boldsymbol{\Phi}_{22.1}\left(\boldsymbol{\beta}_2^{'},\mathbf{I}_{G_2}\right)+\frac{K_{2n}}{n}\boldsymbol{\Omega}\right]\right.$$
$$+\frac{1}{\sqrt{n}}\left[\left(\begin{array}{c}\boldsymbol{\beta}_2^{'}\\ \mathbf{I}_{G_2}\end{array}\right)\boldsymbol{\Pi}_{22}^{(n)'}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1,\mathbf{Z}_{2n}^{*(1)'}\mathbf{V}_2)+\frac{1}{\sqrt{n}}\mathbf{V}^{'}(\mathbf{Z}_{2n}^{*(1)},\mathbf{Z}_{2n}^{*(2)})(\boldsymbol{\pi}_{21}^{(n)},\boldsymbol{\Pi}_{22}^{(n)})\right]$$
$$+\frac{1}{\sqrt{n}}\left[\begin{array}{cc}\boldsymbol{\pi}_{21}^{(n)'}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \boldsymbol{\pi}_{21}^{(n)'}\mathbf{Z}_{2n}^{*(1)'}\mathbf{V}_2\\ \boldsymbol{\Pi}_{22}^{(n)'}\mathbf{A}_{22.1}^{(2)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)'})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \boldsymbol{\Pi}_{22}^{(n)'}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\end{array}\right]\left.-\lambda_n\frac{1}{q_n}\mathbf{H}\right\}$$
$$\times\left[\boldsymbol{\beta}+(\hat{\boldsymbol{\beta}}_{2.LVR}-\boldsymbol{\beta})\right]\sim o_p(1)\ .$$

From the above expression, we need to evaluate the effects of heteroscedasticity of disturbance terms $(v_{1i}, \mathbf{v}_{2i})$ $(i = 1, \cdots, n)$ are negligible for the results of asymptotic distributions. For any constant vectors $\mathbf{a}$ and $\mathbf{b}$, there exists a positive constant A_1 such that for $k = 1, 2$ we have

$$\frac{1}{n}\mathbf{E}\left[\sum_{i,j=1}^{n}[p_{ij}(1)+p_{ij}(2)]\times\mathbf{a}^{'}(\mathbf{v}_i^{(n)}\mathbf{v}_j^{(n)'}-\delta_i^j\boldsymbol{\Omega}_i^{(n)})\mathbf{b}\right]^2$$
$$\leq\ 2\frac{1}{n}\mathbf{E}\left[\sum_{i=1}^{n}[p_{ii}(1)^2+p_{ii}(2)^2][\mathbf{a}^{'}(\mathbf{v}_i\mathbf{v}_i^{'}-\boldsymbol{\Omega}_i^{(n)})\mathbf{b}]^2+\sum_{i\neq j}[p_{ij}(1)^2+p_{ij}(1)^2][\mathbf{a}^{'}\mathbf{v}_i\mathbf{v}_j\mathbf{b}]^2\right.$$
$$\left.+\sum_{i\neq j}[p_{ij}(1)^2+p_{ij}(1)^2][\mathbf{a}^{'}\mathbf{v}_i\mathbf{v}_j^{'}\mathbf{b}\mathbf{a}^{'}\mathbf{v}_j\mathbf{v}_i^{'}\mathbf{b}]\right]$$
$$\leq\ A_1\frac{K_{2n}}{n}\longrightarrow 0\ .$$

It is because the conditional moments of the disturbance terms v_{ji}^4 are bounded, $\sum_{i=1}^n p_{ii}^{(n)} = K_{2n}$ and we have the relation $\sum_{i=1}^n p_{ii}^{(n)2} \leq K_{2n}$ due to the projection matrix.

Then, we find

$$\frac{1}{\sqrt{n}}\left[\left(\begin{array}{cc}\mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\\ \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\end{array}\right)-K_{2n}\boldsymbol{\Omega}\right]$$
$$\stackrel{p}{\longrightarrow}\mathbf{O}$$

when $0 \leq \nu < 1$. We use (2.10) and the relation that

$$\left[\left(\begin{array}{c}\boldsymbol{\beta}_2^{'}\\ \mathbf{I}_{G_2}\end{array}\right)\boldsymbol{\Phi}_{22.1}\left(\boldsymbol{\beta}_2^{'},\mathbf{I}_{G_2}\right)+\frac{K_{2n}}{n}\boldsymbol{\Omega}-\lambda_n\frac{1}{q_n}\mathbf{H}\right]\boldsymbol{\beta}=o_p(\frac{1}{\sqrt{n}})\ .$$

By multiplying the preceding equation out to separate the terms with factor $\boldsymbol{\beta}$ and with the factor $\sqrt{n}\,(\hat{\boldsymbol{\beta}}_{2.LVR} - \boldsymbol{\beta})$, we have

$$(\mathbf{0},\mathbf{I}_{G_2})\left[\left(\begin{array}{c}\boldsymbol{\beta}_2^{'}\\ \mathbf{I}_{G_2}\end{array}\right)\boldsymbol{\Phi}_{22.1}\left(\boldsymbol{\beta}_2^{'},\mathbf{I}_{G_2}\right)\sqrt{n}(\hat{\boldsymbol{\beta}}_{LVR}-\boldsymbol{\beta})+\frac{1}{\sqrt{n}}\boldsymbol{\Pi}_{2*}^{'}(\mathbf{Z}_{2n}^{*(1)},\mathbf{Z}_{2n}^{*(1)})^{'}\mathbf{V}\boldsymbol{\beta}\right]\stackrel{p}{\rightarrow}\mathbf{0}\ ,$$

which is asymptotically equivalent to

$$\mathbf{\Phi}_{22.1}\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.LI}-\boldsymbol{\beta}_2)-\frac{1}{\sqrt{n}}\mathbf{\Pi}_{22}^{(n)'}[\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1-\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\boldsymbol{\beta}_2]\xrightarrow{p}\mathbf{0}\ .$$

By applying the CLT to the second term, we complete the proof of (i) of *Theorem 2* for the 2SLVR estimator of $\boldsymbol{\beta}$ when $1/2\leq\nu<1$.

(**Step 3**, The case of $1/2\leq\nu<1$, 2STSLS.)
Next, we shall investigate the asymptotic property of the 2STSLS estimator. We set $\hat{\boldsymbol{\beta}}_{TS}^{'}=(1,-\hat{\boldsymbol{\beta}}_{2.TS}^{'})$, which is the solution of (2.13). By evaluating each term of

$$(\mathbf{0},\mathbf{I}_{G_2})\sqrt{n}\left[\left(\begin{array}{c}\boldsymbol{\beta}_2^{'}\\ \mathbf{I}_{G_2}\end{array}\right)\mathbf{\Phi}_{22.1}\left(\boldsymbol{\beta}_2^{'},\mathbf{I}_{G_2}\right)+\frac{1}{\sqrt{n}}\mathbf{G}_1^{*}\right]\left[\boldsymbol{\beta}+(\hat{\boldsymbol{\beta}}_{TS}-\boldsymbol{\beta})\right]=\mathbf{0}\ ,$$

we have

$$\text{(A.25)}\qquad [\mathbf{\Phi}_{22.1}]\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.TS}-\boldsymbol{\beta}_2)-\mathbf{G}_1^{**}\boldsymbol{\beta}=o_p(1)\ .$$

Then the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_{2.TS}-\boldsymbol{\beta}_2)$ is the same as that of $\mathbf{\Phi}_{22.1}^{-1}\mathbf{G}_1^{**}\boldsymbol{\beta}$. By using

$$\frac{1}{K_{2n}}\left[\begin{array}{cc}\mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\\ \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\end{array}\right]\xrightarrow{p}\mathbf{\Omega}\ ,$$

and then the asymptotic bias becomes $[K_{2n}/\sqrt{n}]\mathbf{\Omega}\boldsymbol{\beta}\rightarrow(-c_\nu)\mathbf{\Omega}_{22}\boldsymbol{\beta}_2$.
We apply the CLT as (i), we have the result for the 2STSLS estimator of $\boldsymbol{\beta}$ when $\nu=1/2$.
When $1/2<\nu<1$, we notice

$$\begin{aligned}&n^{1-\nu}\left[\frac{1}{n}\mathbf{G}-\left(\begin{array}{c}\boldsymbol{\beta}_2^{'}\\ \mathbf{I}_{G_2}\end{array}\right)\mathbf{\Phi}_{22.1}\left(\boldsymbol{\beta}_2^{'},\mathbf{I}_{G_2}\right)\right]\boldsymbol{\beta}\\ =\ &\frac{K_{2n}}{n^\nu}\mathbf{\Omega}\boldsymbol{\beta}+\frac{1}{n^\nu}\mathbf{\Pi}_{22}^{(n)'}[\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1-\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\boldsymbol{\beta}_2]\\ &+\frac{1}{n^\nu}\left[(\begin{array}{cc}\mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\\ \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\end{array})-K_{2n}\mathbf{\Omega}\right]\boldsymbol{\beta}.\end{aligned}$$

Because the last two terms of the right-hand side except the first term are of the order $o_p(n^{-\nu})$, we have

$$\begin{aligned}&n^{1-\nu}\left[\frac{1}{n}\mathbf{G}-\frac{1}{n}\left(\begin{array}{cc}\mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \mathbf{v}_1^{'}\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\\ \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \mathbf{V}_2^{'}\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\end{array}\right)\right]\\ &\xrightarrow{p}c_\nu\mathbf{\Omega}\boldsymbol{\beta}\end{aligned}$$

as $n \longrightarrow \infty$. Hence by using the similar arguments as (i),

$$(\mathbf{0}, \mathbf{I}_{G_2})\frac{1}{n}\begin{pmatrix} \mathbf{v}_1'\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(1)'}\mathbf{Z}_{2n}^{*(1)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \mathbf{v}_1'\mathbf{Z}_{2n}^{*(1)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2 \\ \mathbf{V}_2'\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 & \mathbf{V}_2'\mathbf{Z}_{2n}^{*(2)}(\mathbf{Z}_{2n}^{*(2)'}\mathbf{Z}_{2n}^{*(2)})^{-1}\mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2 \end{pmatrix}$$

$$\times \quad n^{1-\nu}(\hat{\boldsymbol{\beta}}_{2.TS} - \boldsymbol{\beta}_2) - (\mathbf{0}, \mathbf{I}_{G_2})c\boldsymbol{\Omega}\boldsymbol{\beta} \xrightarrow{p} \mathbf{0}$$

and we complete the proof of (ii) of *Theorem 2* for the 2STSLS estimator when $1/2 \leq \nu < 1$. **Q.E.D.**

(iii) **On the Proof of Theorem 3** :
We set the vector of true parameters $\boldsymbol{\beta}' = (1, -\boldsymbol{\beta}_2') = (1, -\beta_2, \cdots, -\beta_{1+G_2})$. For the estimator of $\boldsymbol{\beta}_2$ to be consistent we need the conditions

$$\beta_k = \phi_k\left[\begin{pmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{pmatrix}\boldsymbol{\Phi}_{22.1}\left(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}\right) + c\boldsymbol{\Omega}, \boldsymbol{\Omega}\right] \quad (k = 2, \cdots, 1 + G_2) \tag{A.26}$$

as identities in $\boldsymbol{\beta}_2$, $\boldsymbol{\Phi}_{22.1}$, and $\boldsymbol{\Omega}$. Let a $(1 + G_2) \times (1 + G_2)$ matrix

$$\mathbf{T}^{(k)} = \left(\frac{\partial \phi_k}{\partial g_{ij}}\right) = (\tau_{ij}^{(k)}) \quad (k = 2, \cdots, 1 + G_2; i, j = 1, \cdots, 1 + G_2) \tag{A.27}$$

evaluated at the probability limits of (A.26). We denote a $(1+G_2)\times(1+G_2)$ matrix $\boldsymbol{\Theta}$ $(= (\theta_{ij}))$

$$\boldsymbol{\Theta} = \begin{pmatrix} \boldsymbol{\beta}_2' \\ \mathbf{I}_{G_2} \end{pmatrix}\boldsymbol{\Phi}_{22.1}\left(\boldsymbol{\beta}_2, \mathbf{I}_{G_2}\right) = \begin{bmatrix} \boldsymbol{\beta}_2'\boldsymbol{\Phi}_{22.1}\boldsymbol{\beta}_2 & \boldsymbol{\beta}_2'\boldsymbol{\Phi}_{22.1} \\ \boldsymbol{\Phi}_{22.1}\boldsymbol{\beta}_2 & \boldsymbol{\Phi}_{22.1} \end{bmatrix},$$

where $\boldsymbol{\Phi}_{22.1} = (\rho_{m,l})$ $(m, l = 2, \cdots, 1+G_2)$, $(\boldsymbol{\Phi}_{22.1}\boldsymbol{\beta}_2)_l = \sum_{j=2}^{1+G_2} \beta_j\rho_{lj}$ $(l = 2, \cdots, 1+G_2)$, $(\boldsymbol{\beta}_2'\boldsymbol{\Phi}_{22.1})_m = \sum_{i=2}^{1+G_2} \beta_i\rho_{im}$ $(m = 2, \cdots, 1+G_2)$, and $\boldsymbol{\beta}_2'\boldsymbol{\Phi}_{22.1}\boldsymbol{\beta}_2 = \sum_{i,j=2}^{1+G_2} \rho_{ij}\beta_i\beta_j$. By differentiating each components of $\boldsymbol{\Theta}$ with respect to β_j $(j = 2, \cdots, G_2)$, we have

$$\frac{\partial \boldsymbol{\Theta}}{\partial \beta_j} = (\frac{\partial \theta_{lm}}{\partial \beta_j}) , \tag{A.28}$$

where $\frac{\partial \theta_{11}}{\partial \beta_j} = 2\sum_{i=2}^{1+G_2} \rho_{ji}\beta_i$ $(j = 2, \cdots, 1 + G_2)$, $\frac{\partial \theta_{1m}}{\partial \beta_j} = \rho_{jm}$ $(m = 2, \cdots, 1 + G_2)$, $\frac{\partial \theta_{l1}}{\partial \beta_j} = \rho_{lj}$ $(l = 2, \cdots, 1 + G_2)$, and $\frac{\partial \theta_{lm}}{\partial \beta_j} = 0$ $(l, m = 2, \cdots, 1 + G_2)$.
Hence

$$\mathbf{tr}\left(\mathbf{T}^{(k)}\frac{\partial \boldsymbol{\Theta}}{\partial \beta_j}\right) = 2\tau_{11}^{(k)}\sum_{i=2}^{1+G_2} \rho_{ji}\beta_i + 2\sum_{i=2}^{1+G_2} \rho_{ji}\tau_{ji}^{(k)} = \delta_j^k , \tag{A.29}$$

where we define $\delta_k^k = 1$ and $\delta_j^k = 0$ $(k \neq j)$. Define a $(1 + G_2) \times (1 + G_2)$ partitioned matrix

$$\mathbf{T}^{(k)} = \begin{bmatrix} \tau_{11}^{(k)} & \boldsymbol{\tau}_2^{(k)'} \\ \boldsymbol{\tau}_2^{(k)} & \mathbf{T}_{22}^{(k)} \end{bmatrix} .$$

Then, (A.29) is represented as $2\tau_{11}^{(k)}\mathbf{\Phi}_{22.1}\boldsymbol{\beta}+2\mathbf{\Phi}_{22.1}\boldsymbol{\tau}_2^{(k)}=\boldsymbol{\epsilon}_k$, where $\boldsymbol{\epsilon}_k^{'}=(0,\cdots,0,1,0,\cdots,0)$ with 1 in the k-th place and zeros in other elements.
Since $\mathbf{\Phi}_{22.1}$ is positive definite, we have

$$\boldsymbol{\tau}_2^{(k)}=\frac{1}{2}\mathbf{\Phi}_{22.1}^{-1}\boldsymbol{\epsilon}_k-\tau_{11}^{(k)}\boldsymbol{\beta}_2\ . \tag{A.30}$$

Further, by differentiating $\mathbf{\Theta}$ with respect to ρ_{ij}, we have

$$\frac{\partial\mathbf{\Theta}}{\partial\rho_{ii}}=(\frac{\partial\theta_{lm}}{\partial\rho_{ii}})\ , \tag{A.31}$$

where $\frac{\partial\theta_{11}}{\partial\rho_{ii}}=\beta_i^2$, $\frac{\partial\theta_{1m}}{\partial\rho_{ii}}=\beta_i\ (m=i),0\ (m\neq i)$, $\frac{\partial\theta_{l1}}{\partial\rho_{ii}}=\beta_i\ (l=i),0\ (l\neq i)$ and $\frac{\partial\theta_{lm}}{\partial\rho_{ii}}=1\ (l=m=i),0$ (otherwise). For $i\neq j$

$$\frac{\partial\mathbf{\Theta}}{\partial\rho_{ij}}=(\frac{\partial\theta_{lm}}{\partial\rho_{ij}})\ , \tag{A.32}$$

where $\frac{\partial\theta_{11}}{\partial\rho_{ij}}=2\beta_i\beta_j$, $\frac{\partial\theta_{1m}}{\partial\rho_{ij}}=\beta_j\ (m=i),\beta_i\ (m=j),0\ (m\neq i,j)$, $\frac{\partial\theta_{l1}}{\partial\rho_{ij}}=\beta_j\ (l=i),\beta_i\ (l=j),0\ (l\neq i,j)$, and $\frac{\partial\theta_{lm}}{\partial\rho_{ij}}=1\ (l=i,m=j\ \mathit{or}\ l=j,m=i),0$ (otherwise) for $(2\leq l,m\leq 1+G_2)$.
Then we have the representation

$$\mathrm{tr}\left(\mathbf{T}^{(k)}\frac{\partial\mathbf{\Theta}}{\partial\rho_{ij}}\right)=\begin{cases}\beta_i^2\tau_{11}^{(k)}+2\tau_{1i}^{(k)}\beta_i+\tau_{ii}^{(k)} & (i=j)\\ \\ 2\beta_i\beta_j\tau_{11}^{(k)}+2\tau_{1j}^{(k)}\beta_i+2\tau_{1i}^{(k)}\beta_j+2\tau_{ij}^{(k)} & (i\neq j)\end{cases}.$$

In the matrix form we have as $\tau_{11}^{(k)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'}+\boldsymbol{\tau}_2^{(k)}\boldsymbol{\beta}_2^{'}+\boldsymbol{\beta}_2\boldsymbol{\tau}_2^{(k)'}+\mathbf{T}_{22}^{(k)}=\mathbf{O}$, and then, we have the representation

$$\begin{aligned}\mathbf{T}_{22}^{(k)}&=-\tau_{11}^{(k)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'}-\boldsymbol{\tau}_2^{(k)}\boldsymbol{\beta}_2^{'}-\boldsymbol{\beta}_2\boldsymbol{\tau}_2^{(k)'}\\ &=\tau_{11}^{(k)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'}-\frac{1}{2}\left[\mathbf{\Phi}_{22.1}^{-1}\boldsymbol{\epsilon}_k\boldsymbol{\beta}_2^{'}+\boldsymbol{\beta}_2\boldsymbol{\epsilon}_k^{'}\mathbf{\Phi}_{22.1}^{-1}\right]\ .\end{aligned}$$

Next, we consider the role of the second matrix in (A.26). By differentiating (A.26) with respect to ω_{ij} $(i,j=1,\cdots,1+G_2)$, we have the condition

$$c\frac{\partial\phi_k}{\partial g_{ij}}=-\frac{\partial\phi_k}{\partial h_{ij}}\ (k=2,\cdots,1+G_2;i,j=1,\cdots,1+G_2)$$

evaluated at the probability limits. Let

$$\mathbf{S}=\mathbf{G}_1-\sqrt{cc_*}\mathbf{H}_1=\begin{bmatrix}s_{11}&\mathbf{s}_2^{'}\\ \mathbf{s}_2&\mathbf{S}_{22}\end{bmatrix}\ . \tag{A.33}$$

Since $\phi(\,\cdot\,)$ is differentiable and its first derivatives are bounded at the true parameters by assumption, the linearized estimator of β_k in the class of our concern can be represented as

$$\begin{aligned}
\sum_{g,h=1}^{1+G_2} \tau_{gh}^{(k)} s_{gh} &= \tau_{11}^{(k)} s_{11} + 2\boldsymbol{\tau}_2^{(k)'} \mathbf{s}_2 + \mathbf{tr}\left[\mathbf{T}_{22}^{(k)}\mathbf{S}_{22}\right] \\
&= \tau_{11}^{(k)} s_{11} + \left(\boldsymbol{\epsilon}_k^{'}\boldsymbol{\Phi}_{22.1}^{-1} - 2\tau_{11}^{(k)}\boldsymbol{\beta}_2^{'}\right)\mathbf{s}_2 + \mathbf{tr}\left[\left(\tau_{11}^{(k)}\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{'} - \boldsymbol{\Phi}_{22.1}^{-1}\boldsymbol{\epsilon}_k\boldsymbol{\beta}_2^{'}\right)\mathbf{S}_{22}\right] \\
&= \tau_{11}^{(k)}\left[s_{11} - 2\boldsymbol{\beta}_2^{'}\mathbf{s}_2 + \boldsymbol{\beta}_2^{'}\mathbf{S}_{22}\boldsymbol{\beta}_2\right] + \boldsymbol{\epsilon}_k^{'}\boldsymbol{\Phi}_{22.1}^{-1}(\mathbf{s}_2 - \mathbf{S}_{22}\boldsymbol{\beta}_2) \\
&= \tau_{11}^{(k)}\boldsymbol{\beta}^{'}\mathbf{S}\boldsymbol{\beta} + \boldsymbol{\epsilon}_k^{'}\boldsymbol{\Phi}_{22.1}^{-1}(\mathbf{s}_2, \mathbf{S}_{22})\boldsymbol{\beta}\ .
\end{aligned}$$

Let

$$\text{(A.34)}\qquad \boldsymbol{\tau}_{11} = \begin{bmatrix} \tau_{11}^{(2)} \\ \vdots \\ \tau_{11}^{(1+G_2)} \end{bmatrix}$$

and we consider the asymptotic behavior of the normalized estimator $\sqrt{n}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$ as

$$\text{(A.35)}\qquad \hat{\mathbf{e}} = \left[\boldsymbol{\tau}_{11}\boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\right]\mathbf{S}\boldsymbol{\beta}\ .$$

Since the asymptotic variance-covariance matrix of $\mathbf{S}\boldsymbol{\beta}$ has been obtained by the proof of *Theorem 1*, *Corollary 1*, and *Theorem 2*, we have

$$\begin{aligned}
&\mathcal{E}\left[\hat{\mathbf{e}}\,\hat{\mathbf{e}}^{'}\right] \\
=\ &\left[(\boldsymbol{\tau}_{11} + \frac{1}{\sigma^2}(\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\boldsymbol{\Omega}\boldsymbol{\beta})\boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})(\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}})\right] \\
&\times\mathcal{E}[\mathbf{S}\boldsymbol{\beta}\boldsymbol{\beta}^{'}\mathbf{S}] \times \left[(\boldsymbol{\tau}_{11} + \frac{1}{\sigma^2}(\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\boldsymbol{\Omega}\boldsymbol{\beta})\boldsymbol{\beta}^{'} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})(\mathbf{I}_{G_2+1} - \frac{\boldsymbol{\Omega}\boldsymbol{\beta}\boldsymbol{\beta}^{'}}{\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\boldsymbol{\beta}})\right]^{'} \\
=\ &\boldsymbol{\Psi}^{**} + \mathcal{E}\left[(\boldsymbol{\beta}^{'}\mathbf{S}\boldsymbol{\beta})^2\right]\left[\boldsymbol{\tau}_{11} + (\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\frac{1}{\sigma^2}\boldsymbol{\Omega}\boldsymbol{\beta}\right]\left[\boldsymbol{\tau}_{11}^{'} + \frac{1}{\sigma^2}\boldsymbol{\beta}^{'}\boldsymbol{\Omega}\begin{pmatrix} \mathbf{0}^{'} \\ \boldsymbol{\Phi}_{22.1}^{-1} \end{pmatrix}\right] + o(1)\ ,
\end{aligned}$$

where $\boldsymbol{\Psi}^{**}$ has been given by *Theorem 1* or *Theorem 2*. This covariance matrix is the sum of a positive semi-definite matrix of rank 1 and a positive definite matrix. It has a minimum if

$$\text{(A.36)}\qquad \boldsymbol{\tau}_{11} = -\frac{1}{\sigma^2}(\mathbf{0}, \boldsymbol{\Phi}_{22.1}^{-1})\boldsymbol{\Omega}\boldsymbol{\beta}\ .$$

This completes the proof of *Theorem 3*.
Q.E.D.

(iv) **On the Derivation of (4.2)** :
In the general case, there is a complication with $\mathbf{F}$. It is because the term of $\sqrt{n}[\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_1]$ as $\frac{1}{n}\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{Z}_{2n}^{(1)})^{-1}\frac{1}{\sqrt{n}}\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{v}_1 - (\frac{1}{n}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{Z}_1^{(1)})^{-1}\frac{1}{\sqrt{n}}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{V}_2\boldsymbol{\beta}_2$ and the first term of $\sqrt{n}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)$ as $\mathbf{gh}_1$ may have some correlation asymptotically. We only evaluate the asymptotic covariance of

$$\begin{aligned} &\frac{1}{n}\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{Z}_{2n}^{(1)})^{-1}\frac{1}{\sqrt{n}}\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{v}_1 - (\frac{1}{n}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{Z}_1^{(1)})^{-1}\frac{1}{\sqrt{n}}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{V}_2\boldsymbol{\beta}_2 \\ \sim\ &\mathbf{M}_{11.2}^{-1}[\frac{1}{\sqrt{n}}\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{v}_1 - \frac{1}{\sqrt{n}}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{V}_2\boldsymbol{\beta}_2] \end{aligned}$$

and $-\boldsymbol{\Pi}_{12}\boldsymbol{\Phi}_{22.1}^{-1}\mathbf{A}\mathbf{gh}(1)$, because the effects of $\boldsymbol{\Phi}_{22.1}^{-1}\mathbf{A}[\mathbf{gh}(2) + \mathbf{gh}(3)]$ are asymptotically negligible due to the assumption of third moments of disturbances. By using (A.12), we evaluate the covariance of $(\frac{1}{\sqrt{n}})(\mathbf{Z}_{2n}^{*(1)'}\mathbf{v}_1 - \mathbf{Z}_{2n}^{*(2)'}\mathbf{V}_2\boldsymbol{\beta}_2)$ and $\frac{1}{n}\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{Z}_{2n}^{(1)})^{-1}\frac{1}{\sqrt{n}}\mathbf{Z}_1^{(1)'}\bar{\mathbf{P}}_{Z_{2n}^{(1)}}\mathbf{v}_1 - (\frac{1}{n}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{Z}_1^{(2)})^{-1}\frac{1}{\sqrt{n}}\mathbf{Z}_1^{(2)'}\bar{\mathbf{P}}_{Z_{2n}^{(2)}}\mathbf{V}_2\boldsymbol{\beta}_2$. Since $\sigma^2 = \omega_{11} + \boldsymbol{\beta}_2'\boldsymbol{\Omega}_{22}\boldsymbol{\beta}_2$, we use the conditions in (4.3) and we arrange several terms to obtain (4.2).