Swanson, Norman R.; Chao, John C.

Working Paper
Asymptotic Normality of Single-Equation Estimators for the Case with a Large Number of Weak Instruments

Working Paper, No. 2003-12

Provided in Cooperation with:
Department of Economics, Rutgers University

Suggested Citation: Swanson, Norman R.; Chao, John C. (2003) : Asymptotic Normality of Single-Equation Estimators for the Case with a Large Number of Weak Instruments, Working Paper, No. 2003-12, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:
http://hdl.handle.net/10419/23169

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.
If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Asymptotic Normality of Single-Equation Estimators for the Case with a Large Number of Weak Instruments

John C. Chao
University of Maryland

Norman R. Swanson
Rutgers University

June 2003

Abstract

This paper analyzes conditions under which various single-equation estimators are asymptotically normal in a simultaneous equations framework with many weak instruments. In particular, our paper adds to the many instruments asymptotic normality literature, including papers by Morimune (1983), Bekker (1994), Angrist and Krueger (1995), Donald and Newey (2001), Hahn, Hausman, and Kuersteiner (2001), and Stock and Yogo (2003). We consider the case where instrument weakness is such that \( r_n \), the rate of growth of the concentration parameter, is slower than \( K_n \), the growth rate of the number of instruments, but such that \( \frac{\sqrt{K_n}}{r_n} \to 0 \) as \( n \to \infty \). In this case, the rate of convergence is shown to be \( \frac{r_n}{\sqrt{K_n}} \). We also show that formulae for the asymptotic variances of various single-equation estimators are different from those obtained under assumptions of stronger instruments, i.e., cases where \( r_n \) is assumed to grow at the same rate or at a faster rate than \( K_n \). An interesting finding of this paper is that, for the case we study here, both the LIML and the Fuller estimators can be shown to be asymptotically more efficient than the \( B2SLS \) estimator not just for the case where the error distributions are assumed to be Gaussian but for all error distributions that lie within the elliptical family.

*JEL classification:* C13, C31.
*Keywords:* CLT for bilinear forms, instrumental variables, \( k \)-class estimator, local-to-zero framework, pathwise asymptotics, weak instruments.

*John C. Chao: Department of Economics, University of Maryland, College Park, MD, USA 20742, chao@econ.umd.edu. Norman R. Swanson: Department of Economics, Rutgers University, New Brunswick, NJ, USA 08901, nswanson@econ.rutgers.edu.*
1 Introduction

Amongst Peter C. B. Phillips’s many contributions to econometrics are two papers which explore the theoretical properties of conventional econometric procedures in models which suffer from a lack of identification. These papers, Phillips (1989) and Choi and Phillips (1992), were the first to derive both finite sample and asymptotic distributions of the instrumental variables (IV) estimator in a simultaneous equations system with identification failure. One of the key findings of Phillips (1989) and Choi and Phillips (1992) is that, when the model is underidentified, the IV estimator is inconsistent and converges to a random variable, reflecting the fact that even in the limit the estimation uncertainty does not go away due to the lack of identification.

Since the work of Phillips and Choi and Phillips, research on econometric models with identification problems has picked up steam and the area is currently one of the most active ones in econometrics. In particular, econometricians have become interested in the case where the model is weakly identified (or nearly unidentified), which, in the context of an IV regression, translates to the case where the instruments are only weakly correlated with the endogenous explanatory variables. Indeed, in recent years, it has become popular to model weak instruments using the local-to-zero asymptotic framework of Staiger and Stock (1994), which takes the coefficients of the instruments in the first-stage regression to be in a $n^{-1/2}$ shrinking neighborhood of the origin, where $n$ denotes the sample size\(^1\). An essential feature of the Staiger-Stock local-to-zero device is that it keeps the so-called concentration parameter from diverging as the sample size approaches infinity, so that, under their framework, conventional k-class estimators, such as the two-stage least squares (2SLS) and the limited information maximum likelihood (LIML) estimator, exhibit asymptotic behaviour similar to that which occurs in the underidentified case, at least when the number of instruments is held fixed as the sample size is allowed to approach infinity. More specifically, under the Staiger-Stock local-to-zero framework, conventional k-estimators can be shown to be inconsistent and, in fact, converges weakly to nonstandard distributions.

More recently, Chao and Swanson (2002b) argue that there may be benefits to using a large number of instruments when the available instruments are of poor quality. In particular, they show that by allowing the number of instruments to increase to infinity with the sample size, the growth of the concentration parameter may be accelerated sufficiently, so that consistent estimation may become achievable even when all available instruments are weak in the local-to-zero sense. In this case, the choice of estimator becomes important, as not all estimators are equally susceptible to instrument weakness. Along these lines, Chao and Swanson show that single-equation estimators satisfying certain conditions, such as the LIML estimator and the Jackknife Instrumental Variables Estimators (JIVE), are consistent even when instrument weakness is such that the rate at which the concentration parameter grows, say $r_n$, is slower than the rate of expansion of the number of instruments, say $K_n$, so long as $\frac{\sqrt{K_n}}{r_n} \to 0$ as $n \to \infty^2$. On the other hand, the 2SLS estimator

\(^1\)Other interesting papers which make use of the local-to-zero setup include Wang and Zivot (1998) and Kleibergen (2002).

\(^2\)One version of the JIVE estimator was introduced by Phillips (1977). Other versions of JIVE have since been...
is only consistent if \( r_n \) approaches infinity faster than \( K_n \). In addition, asymptotic distributions for various \( k \)-class estimators in the case where \( r_n \) approaches infinity at the same rate as \( K_n \) (i.e., \( \frac{K_n}{r_n} \to \alpha \) for \( 0 < \alpha < \infty \)) have now been derived by Stock and Yogo (2003). Overall, the work of Chao and Swanson (2002b) and Stock and Yogo (2003) can be viewed as adding to the many instrument asymptotic results of Morimune (1983), Bekker (1994), Angrist and Krueger (1995), Donald and Newey (2001), and Hahn, Hausman, and Kuersteiner (2001) to a weakly identified IV regression model with a local-to-zero structure.

The purpose of the present paper is to further extend the results presented in Chao and Swanson (2002b) and in Stock and Yogo (2003). More precisely, we extend the asymptotic normality results obtained by Stock and Yogo (2003) for LIML, Fuller’s modified LIML (FLIML, henceforth), and the Bias-adjusted Two Stage Least Squares (B2SLS) estimators to the case where instrument weakness is such that the rate of growth of the concentration parameter \( r_n \) is slower than the rate of growth of the number of instruments \( K_n \) but such that \( \sqrt{K_n} r_n \to 0 \) as \( n \to \infty \). Thus, we obtain asymptotic normality results in situations with weaker instruments than has been assumed by other papers using the many instruments setup. The rate of convergence in our case is shown to be \( \frac{r_n}{\sqrt{K_n}} \), which is slower than the rate of convergence to normality obtained by other authors, and which reflects our assumption of weaker instruments. Formulae for the asymptotic variances of the estimators are also shown to be different from those obtained under assumptions of stronger instruments, i.e., cases where \( r_n \) is assumed to grow at the same rate or at a faster rate than \( K_n \). An additional finding of this paper is that, for the case studied in this paper, both the LIML and the FLIML estimators can be shown to be asymptotically more efficient than the B2SLS estimator not just for the case where the error distributions are assumed to be Gaussian but for all error distributions that lie within the elliptical family.

The rest of the paper proceeds as follows. Section 2 sets up the model and discusses our assumptions. Section 3 presents the main results of the paper and briefly comments on the implications of these results. Concluding remarks are given in Section 4, and all proofs are gathered in two appendices. The following notation is used in the remainder of the paper: \( \text{Tr}(\cdot) \) denotes the trace of a matrix, \( \sim > 0 \) denotes positive definiteness when applied to matrices, \( \lim_{n \to \infty} a_n \) denotes the limit inferior of the sequence \( \{a_n\} \), and \( \overline{\lim}_{n \to \infty} a_n \) denotes the limit superior of the sequence \( \{a_n\} \). In addition, \( P_X = X(X'X)^{-1}X' \) denotes the matrix which projects orthogonally onto the range space of \( X \) and \( M_X = I - P_X \).

### 2 Model and Assumptions

Consider the following two-equation simultaneous equations model (SEM)

\[
\begin{align*}
y_{1n} &= y_{2n} \beta + X_n \gamma + u_n, \\
y_{2n} &= Z_n \pi + X_n \varphi + v_n,
\end{align*}
\]

introduced and studied independently by Angrist, Imbens, and Krueger (1999) and Blomquist and Dahlberg (1999).
where \(y_{1n}\) and \(y_{2n}\) are \(n \times 1\) vectors of observations on the 2 endogenous variables of the system, \(X_n\) is an \(n \times J\) matrix of observations on the \(J\) exogenous variables included in the structural equation (1), \(Z_n\) is an \(n \times K_n\) matrix of observations on the \(K_n\) instrumental variables, or exogenous variables excluded from the structural equation (1), and \(u_n\) and \(v_n\) are \(n \times 1\) vectors of random disturbances. Further, let \(\eta_i = (u_i, v_i)'\) where \(u_i\) and \(v_i\) are, respectively, the \(i\)th component of the random vectors \(u_n\) and \(v_n\), respectively. The following assumptions are used in the sequel.

**Assumption 1:** \(\pi = \pi_n = \frac{c_n}{b_n}\) for some sequence of positive real numbers \(\{b_n\}\), nondecreasing in \(n\), and for some sequence of nonrandom, \(K_n \times 1\) parameter matrices \(\{c_n\}\).

**Assumption 2:** Let \(\{Z_{i,n} : i = 1, \ldots, n; n \geq 1\}\) be a triangular array of \(R^{K_n+J}\)-valued random variables, where \(Z_{i,n} = (Z_{i,n}', X_{i,n}')'\) with \(Z_{i,n}'\) and \(X_{i,n}'\) denoting the \(i\)th row of the matrices \(Z_n\) and \(X_n\), respectively. Moreover, suppose that:

(a) \(K_n \to \infty\) as \(n \to \infty\) such that \(\frac{K_n}{n} \to \alpha\) for some constant \(\alpha\) satisfying \(0 \leq \alpha < 1\).

(b) Let \(m_{1n} \not\to \infty\) as \(n \to \infty\), and suppose that there exist constants \(D_\lambda\) and \(\bar{D}_\lambda\), with \(0 < D_\lambda \leq \bar{D}_\lambda < \infty\), such that

\[
D_\lambda \leq \lim_{n \to \infty} \lambda_{\min} \left( \frac{Z_n'Z_n}{m_{1n}} \right) a.s. \tag{3}
\]

and

\[
\overline{\lim}_{n \to \infty} \lambda_{\max} \left( \frac{Z_n'Z_n}{m_{1n}} \right) \leq \bar{D}_\lambda \ a.s., \tag{4}
\]

where \(Z_n = (Z_n' X_n)\).

(c) There exist a sequence of positive real numbers \(\{m_{2n}\}\), nondecreasing in \(n\), and constants \(D_c\) and \(\bar{D}_c\), with \(0 < D_c \leq \bar{D}_c < \infty\), such that

\[
D_c \leq \lim_{n \to \infty} \left( \frac{c_n'c_n}{m_{2n}} \right) \tag{5}
\]

and

\[
\overline{\lim}_{n \to \infty} \left( \frac{c_n'c_n}{m_{2n}} \right) \leq \bar{D}_c. \tag{6}
\]

**Assumption 3:** \(\bar{Z}_n\) and \(\eta_i\) are independent for all \(i\) and \(n\).
Assumption 4:

(a) $\eta_i \equiv i.i.d.(0, \Sigma)$, where $\Sigma > 0$, and partition $\Sigma$ conformably with $(u_i, v_i)'$ as $\Sigma = \begin{pmatrix} \sigma_{uu} & \sigma_{uv} \\ \sigma_{vu} & \sigma_{vv} \end{pmatrix}$.

(b) There exists some constant $D_\eta$, with $0 < D_\eta < \infty$, such that $\max\{E(u_i^8), E(v_i^8)\} \leq D_\eta$.

(c) $E(u_i^3) = E(v_i^3) = E(u_i^2v_i) = E(u_i^2) = 0$.

Assumption 5: Define the ratio $\frac{r_n}{b_n^2} = \frac{m_1n^{m_2n}}{b_n^2}$. Suppose that, as $n \to \infty$, $r_n \to \infty$ such that $D_\pi \frac{r_n}{b_n^2} \to 0$ but $\sqrt{\frac{r_n}{b_n^2}} \to 0$.

Remark 2.1: (i) Assumptions 1 and 2 are the same as corresponding assumptions that were made in Chao and Swanson (2002a). As explained in that paper, these assumptions imply that there exists a positive integer $N$ such that, for all $n \geq N$, $0 < D_\lambda D_c \leq \frac{\pi'_n Z'_n M_{X_n} Z_n \pi_n}{r_n} \leq D_\lambda D_c < \infty$ with probability one, so that the concentration parameter $\pi'_n Z'_n M_{X_n} Z_n \pi_n$ grows at the rate $r_n = \frac{m_1n^{m_2n}}{b_n^2}$.

(ii) Assumption 4(c) impose a certain symmetry on the distribution of the disturbances of the simultaneous equations model given by equations (1) and (2). Similar conditions have also been assumed in the paper by Koenker and Machado (1999), which examines the asymptotic properties of a GMM estimator as the number of moment conditions goes to infinity with the sample size. Note also that our Assumption 4 is satisfied by all distributions within the elliptical family which have finite eighth moments.

(iii) Assumption 6 focuses attention on the case where the concentration parameter grows at a slower rate than the number of instruments $K_n$ but at a faster rate than $\sqrt{K_n}$. To the best of our knowledge, this is a case for which the asymptotic normality of various IV estimators, such as LIML, FLIML, and B2SLS, has not been established previously. In particular, earlier papers by Morimune (1983) and Bekker (1994) studied the case where $r_n \sim n$, i.e., the case where concentration parameter diverges at the same rate as the sample size, so that those papers consider situations where the concentration parameter either grows at the same rate as $K_n$ (if $\frac{K_n}{n} \to \alpha$ for some constant $\alpha$ such that $0 < \alpha < 1$) or at a faster rate than $K_n$ (if $\frac{K_n}{n} \to 0$). In addition, as part of a larger paper on choosing the number of instruments using (asymptotic) mean-square error formulae of various IV estimators, Donald and Newey (2001) also present a proof of the asymptotic normality of LIML in a many-instruments setup when $r_n \sim n$. Finally, a recent paper by Stock and Yogo (2003), which derives the limiting distributions of LIML, FLIML, and B2SLS within a many weak instruments framework, also considers a case different from ours, as these authors assume that $r_n$ and $K_n$ grow at the same rate. Since the concentration parameter is a natural measure of instrument weakness, as pointed out by Phillips (1983), Rothenberg (1983), Stock and Yogo (2001), and others, our analysis here can be viewed as considering cases where the instruments are weaker than that investigated by other authors using a many-instruments asymptotic framework. As we will show in the next section of the paper, the case we study here is also interesting because the weaker instruments lead to rate of convergence and asymptotic variances that are different vis-à-vis that obtained by assuming faster growth of the concentration parameter relative to $K_n$. 
(iv) Note that our assumptions involve a tradeoff of conditions relative to Donald and Newey (2001) and Stock and Yogo (2003). In particular, we do not make i.i.d. assumptions on the triangular array of exogenous variables $Z_{i,n}$. Thus, our assumptions on the exogenous variables are weaker than those made in Donald and Newey (2001) and Stock and Yogo (2003). On the other hand, we make more stringent assumptions on the moments of the error distributions. In addition to the symmetry condition discussed in Remark 2.1(ii) above, our Assumption 4(b) require the error distributions to possess finite eighth moments, whereas Donald and Newey (2001) and Stock and Yogo (2003) only assume finite fourth moments. Finally, our Assumption 2(a) impose a less stringent condition on the rate of increase of the number of instruments relative to Donald and Newey (2001) and Stock and Yogo (2003). While Donald and Newey (2001) require that $K_n/n \to 0$ as $n \to \infty$ in deriving their asymptotic normality result for $LIML$ and while Stock and Yogo (2003) require that $K_n^2/n \to 0$, we require only that $K_n/n \to \alpha$, with $0 \leq \alpha < 1$, so that the results of this paper will hold with $K_n$ growing either at the same rate as $n$ or at a slower rate relative to $n$.

3 Asymptotic Normality of Single-Equation Estimators

We focus our analysis on the following three estimators:

1. Limited Information Maximum Likelihood ($LIML$) Estimator

$$
\hat{\beta}_{LIML,n} = \left( y_{2n}' M_{X,n} y_{2n} - \hat{\lambda}_{LIML,n} y_{2n}' M_{Z,n} y_{2n} \right)^{-1} \left( y_{2n}' M_{X,n} y_{1n} - \hat{\lambda}_{LIML,n} y_{2n}' M_{Z,n} y_{1n} \right),
$$

where $\hat{\lambda}_{LIML,n}$ is the smallest root of the determinantal equation:

$$
\det \left\{ \begin{pmatrix} y_{1n}' M_{X,n} y_{1n} & y_{1n}' M_{X,n} y_{2n} \\ y_{2n}' M_{X,n} y_{1n} & y_{2n}' M_{X,n} y_{2n} \end{pmatrix} - \lambda_n \begin{pmatrix} y_{1n}' M_{Z,n} y_{1n} & y_{1n}' M_{Z,n} y_{2n} \\ y_{2n}' M_{Z,n} y_{1n} & y_{2n}' M_{Z,n} y_{2n} \end{pmatrix} \right\} = 0
$$

2. Fuller’s Modified LIML ($FLIML$) Estimator:

$$
\hat{\beta}_{FLIML,n} = \left( y_{2n}' M_{X,n} y_{2n} - \hat{k}_{FLIML,n} y_{2n}' M_{Z,n} y_{2n} \right)^{-1} \left( y_{2n}' M_{X,n} y_{1n} - \hat{k}_{FLIML,n} y_{2n}' M_{Z,n} y_{1n} \right),
$$

where $\hat{k}_{FLIML,n} = \hat{\lambda}_{LIML,n} - \frac{a}{n-K_n-J}$ for some positive constant $a$.

3. Bias-Corrected Two-Stage Least Squares ($B2SLS$) Estimator:

$$
\hat{\beta}_{FLIML,n} = \left( y_{2n}' M_{X,n} y_{2n} - \left( \frac{n}{n-K_n+2} \right) y_{2n}' M_{Z,n} y_{2n} \right)^{-1} \times \left( y_{2n}' M_{X,n} y_{1n} - \left( \frac{n}{n-K_n+2} \right) y_{2n}' M_{Z,n} y_{1n} \right).
$$
All three of these estimators are, of course, special cases of the k-class estimator defined by

$\hat{\beta}_{k,n} = (y'_{2n}M_{X_n}y_{2n} - ky'_{2n}M_{Z_n}y_{2n})^{-1}(y'_{2n}M_{X_n}y_{1n} - ky'_{2n}M_{Z_n}y_{1n})$.

These three estimators are three of the most well-known k-class estimators, and the asymptotic properties of one or more of these estimators have been studied previously in the many instruments context by Morimune (1983), Bekker (1994), Donald and Newey (2001) and Stock and Yogo (2003). However, as discussed above, the purpose of this paper is to derive the asymptotic distributions of these estimators in the case where the instruments are weaker than that assumed in these earlier papers.

The following theorems present the main asymptotic results of this paper

**Theorem 3.1: (LIML)**

Let $\hat{\beta}_{LIML,n}$ be as defined in equation (7) above. Then, under assumptions 1-5,

$$\left(\frac{\Psi_n}{\sigma_{L,n}}\right) \left(\hat{\beta}_{LIML,n} - \beta_0\right) \xrightarrow{d} N(0, 1) \quad \text{as} \quad n \to \infty,$$

where $\Psi_n = b_n^{-2}c_n'Z_n'M_{X_n}Z_n c_n$, where

$$\sigma_{L,n}^2 = \left[ E(u_j^2v_j^2) - \sigma_{uv}^2 \right] \sum_{j=1}^n E(g_{jj,n}^2) + \frac{\sigma_{uv}^2}{\sigma_{uu}^2} \left[ E(u_j^4) - \sigma_{uu}^2 \right] \sum_{j=1}^n E(g_{jj,n}^2)$$

$$+ 2\frac{\sigma_{uv}^2}{\sigma_{uu}^2} \left[ E(u_j^2v_j) - \sigma_{uv}\sigma_{uw} \right] \sum_{j=1}^n E(g_{jj,n}^2) - 2\frac{\sigma_{uv}^2}{\sigma_{uu}^2} \left[ E(u_j^2v_j) - \sigma_{uv}\sigma_{uw} \right] \sum_{j=1}^n E(g_{jj,n}^2),$$

and where $g_{jj,n}$ and $g_{ij,n}$ denote, respectively, the $j^{th}$ diagonal element and the $(i, j)^{th}$ element of the matrix $G_n = P_{Z_n} - P_{X_n} - \left(\frac{K_n}{n-K_n-J}\right)M_{Z_n}$.

**Theorem 3.2: (FLIML)**

Let $\hat{\beta}_{FLIML,n}$ be as defined in equation (9) above. Then, under assumptions 1-5,

$$\left(\frac{\Psi_n}{\sigma_{L,n}}\right) \left(\hat{\beta}_{FLIML,n} - \beta_0\right) \xrightarrow{d} N(0, 1) \quad \text{as} \quad n \to \infty,$$

where $\Psi_n$ and $\sigma_{L,n}$ are as defined in Theorem 3.1 above.

**Theorem 3.3: (B2SLS)**

Let $\hat{\beta}_{B2SLS,n}$ be as defined in equation (10) above. Then, under assumptions 1-5,

$$\left(\frac{\Psi_n}{\sigma_{B,n}}\right) \left(\hat{\beta}_{B2SLS,n} - \beta_0\right) \xrightarrow{d} N(0, 1) \quad \text{as} \quad n \to \infty,$$
employing a many-instruments setup. A comparison illuminates the differences between our results and those obtained by other authors.

where \( \Psi_n \) is as defined in Theorem 3.1 and where

\[
\sigma_{B,n}^2 = \left[ E(u_j^2 v_j^2) - \sigma_{uu}^2 \right] \sum_{j=1}^n E(g_{jj,n}^2) + 2 \left( \sigma_{uu} \sigma_{vv} + \sigma_{uv}^2 \right) \sum_{1 \leq i < j \leq n} E(g_{ij,n}^2),
\]

with \( g_{jj,n} \) and \( g_{ij,n} \) is as defined in Theorem 3.1.

**Remark 3.2:**

(i) Note that Lemma shows that \( \sigma_{L,n}^2 \) and \( \sigma_{B,n}^2 \) grow at the same rate as \( K_n \) as \( n \to \infty \). If we make the additional assumptions that, as \( n \to \infty \), \( \Psi_n = r_n^{-1} \Psi \), \( \sigma_{L,n}^2 \to \sigma_L^2 \), and \( \sigma_{B,n}^2 \to \sigma_B^2 \) for positive constant \( \Psi \), \( \sigma_L^2 \), and \( \sigma_B^2 \); then the asymptotic normality results given in Theorems 3.1-3.3 can be restated as

\[
\frac{r_n}{\sqrt{K_n}} \left( \beta_{\text{LIML},n} - \beta_0 \right) \overset{d}{\to} N \left( 0, \sigma_B^2 \Psi^{-2} \right),
\]

\[
\frac{r_n}{\sqrt{K_n}} \left( \beta_{\text{FLIML},n} - \beta_0 \right) \overset{d}{\to} N \left( 0, \sigma_L^2 \Psi^{-2} \right),
\]

\[
\frac{r_n}{\sqrt{K_n}} \left( \beta_{\text{B2SLS},n} - \beta_0 \right) \overset{d}{\to} N \left( 0, \sigma_B^2 \Psi^{-2} \right).
\]

Interestingly, under Assumption 6, \( \beta_{\text{LIML},n} \), \( \beta_{\text{FLIML},n} \), and \( \beta_{\text{B2SLS},n} \) are all consistent, but the rate of convergence is \( \frac{r_n}{\sqrt{K_n}} \), which depends both on the rate of growth of the concentration parameter \( r_n \) and on the rate of increase of the number of instruments. Note further that under Assumptions 2(a) and 6, \( \frac{r_n}{\sqrt{K_n}} = o \left( \sqrt{n} \right) \), so this rate of convergence is slower than the usual \( \sqrt{n} \) rate of convergence. This slower rate of convergence, in turn, reflects the fact that here we are studying the case where the instruments are weaker than that under the conventional strong identification case, where the concentration parameter grows at the rate \( n \).

(ii) It is of interest to briefly compare the results we obtained here under Assumption 5 with results which occur in cases where \( r_n \) is assumed to grow at the same rate or at a faster rate than \( K_n \). Such a comparison illuminates the differences between our results and those obtained by other authors employing a many-instruments setup.

To begin, note that, in general, it can be shown that the three estimators studied here have the generic (asymptotic) representation

\[
\frac{\Psi_n}{\sigma_{.,n}} \left( \beta - \beta_0 \right) = \frac{f_n' u_n + d_1 v_n' G_n u_n + d_2 v_n' G_n u_n}{\sigma_{.,n}} + o_p(1),
\]

where

\[
d_2 = \left\{ \begin{array}{ll}
-\frac{\sigma_{uv}}{\sigma_{uu}} & \text{for } \beta_{\text{LIML},n}, \beta_{\text{FLIML},n} \\
0 & \text{for } \beta_{\text{B2SLS},n}
\end{array} \right.
\]

where

\[
\sigma_{.,n}^2 = \left\{ \begin{array}{ll}
\sigma_{L,n}^2 + \sigma_{uu} E(f_n' f_n) & \text{for } \beta_{\text{LIML},n}, \beta_{\text{FLIML},n} \\
\sigma_{B,n}^2 + \sigma_{uu} E(f_n' f_n) & \text{for } \beta_{\text{B2SLS},n}
\end{array} \right.
\]

with \( \sigma_{L,n}^2 \) and \( \sigma_{B,n}^2 \) as defined in expressions (12) and (13) above, and where \( \Psi_n = \frac{c_n Z_n' M_{X_n} Z_n c_n}{b_n^2 r_n} \), \( f_n = b_n^{-1} M_{X_n} Z_n c_n \), and \( d_1 = 1 \) for all three estimators. Under Assumption 5, \( \frac{f_n' u_n}{\sigma_{.,n}} = o_p(1) \), so that
the asymptotic distributions of the estimators depend only on the bilinear part of (14), i.e.,

\[
\frac{\Psi_n}{\sigma_{.,n}} \left( \hat{\beta} - \beta_0 \right) = \frac{d_1 v'_n G_n u_n + d_2 u'_n G_n u_n}{\sigma_{.,n}} + o_p(1).
\]

It is of interest to first compare our case with the case studied recently by Stock and Yogo (2003), which assumes that \( r_n \) grows at the same rate as \( K_n \). In the Stock-Yogo case, the asymptotic distributions of LIML, FLIML, and B2SLS depend on both the linear part, \( \frac{f'_n u_n}{\sigma_{.,n}} \), and the bilinear part \( \frac{d_1 v'_n G_n u_n + d_2 u'_n G_n u_n}{\sigma_{.,n}} \). Thus, the general form of the asymptotic variance for these estimators in the Stock-Yogo case is different from that which we obtained in Theorems 3.1-3.3 and in Remark 3.2(i) above, as the asymptotic variance in their case also depends on contribution from the linear component. In addition, Stock and Yogo (2003) find the rate of convergence in their case to be \( \sqrt{K_n} \). This is the same as our rate of convergence of \( \frac{r_n}{\sqrt{K_n}} \) in the case where \( r_n \sim K_n \). However, for \( r_n = o(K_n) \), our rate of convergence is slower than theirs, reflecting the fact that we treat a case with weaker instruments.

It should be noted that earlier papers by Morimune (1983) and Bekker (1994) have also examined the case where the concentration parameter grows at the same rate as the number of instruments, but those papers differ from Stock and Yogo (2003) and also from this paper in that they assume \( r_n \) and \( K_n \) to grow at the same rate as the sample size \( n \). Hence, the situation studied in those papers might be better characterized as one with strong, as opposed to weak, instruments.

Finally, in the case where \( r_n \) grows faster than \( K_n \),

\[
\frac{d_1 v'_n G_n u_n + d_2 u'_n G_n u_n}{\sigma_{.,n}} = o_p(1),
\]

and the asymptotic distributions depend only on the linear part, \( \frac{f'_n u_n}{\sigma_{.,n}} \), and not on the bilinear component at all. Thus, the general form of the asymptotic variance of LIML, FLIML, and B2SLS in this case is also qualitatively different from what we derived under Assumption 5. The case where \( r_n \) grows faster than \( K_n \) is one which has been well studied in the literature. In particular, and as mentioned above, Donald and Newey (2001) derive asymptotic normality results for LIML under the assumptions that \( r_n \sim n \) and \( K_n \rightarrow 0 \), as \( n \rightarrow \infty \). Note also that the case where \( r_n \) grows faster than \( K_n \) includes the conventional case with full identification and \( \sqrt{n} \) convergence of estimators to asymptotic normal distributions, since the conventional setup can be obtained by assuming \( r_n \sim n \) and taking \( K_n \) to be fixed for all \( n \).

(iii) Note further that Theorem 3.1-3.3 show that LIML and FLIML are asymptotically equivalent. However, the B2SLS estimator is not asymptotically equivalent to LIML or FLIML. Indeed, the following result shows that if the distribution of the disturbances of the simultaneous equations system (1)-(2) are taken to belong to the family of elliptically distributions with finite eighth moments, then LIML and FLIML can be shown to be asymptotically more efficient than B2SLS.

**Theorem 3.4:** Suppose that Assumptions 1-5 hold. Suppose, in addition, that \( \eta_i \sim E_2(0,\Xi) \), where \( \Xi = \tau \Sigma \) for some positive constant \( \tau \) and where \( E_2(0,\Xi) \) is as defined in Definition A1 of
Appendix A with \(m = 2\). Then, there exists a positive integer \(N\) such that for all \(n \geq N\),

\[
\sigma_{B,n}^2 > \sigma_{L,n}^2.
\]  

(15)

Note that when the error distribution is Gaussian, \(LIML\) and \(FLIML\) have interpretations as maximum likelihood (ML) estimators, so one would expect \(LIML\) and \(FLIML\) to be more efficient than \(B2SLS\) within a many-weak-instruments asymptotic framework. However, our result shows that even when the errors are non-Gaussian but lie within the elliptical family, in which case \(LIML\) and \(FLIML\) do not have strict interpretations as ML estimators, these estimators are still asymptotically more efficient than \(B2SLS\) within the local-to-zero, many instruments framework studied in this paper. This result is consistent with the asymptotic mean square error results obtained by Donald and Newey (2001) for these estimators under the assumption of \(i.i.d\). instruments. With regard to the relative efficiency of \(LIML\) vis-à-vis the \(B2SLS\) estimator, our results might be viewed as extending the work of Donald and Newey (2001) both to the case with weaker instruments and to the case where the instruments are possibly not \(i.i.d\).

(iv) Another well-known \(k\)-class estimator is the (unadjusted) Two-Stage Least Squares (2SLS) estimator. However, we did not derive the asymptotic distribution of this estimator here because, as shown in Chao and Swanson (2002b), the 2SLS estimator is inconsistent under Assumption 5. More specifically, part (a) of Theorem 3.4 of Chao and Swanson (2002b) shows that, when \(\frac{K_n}{n} \to 0\) as \(n \to \infty\),

\[
\hat{\beta}_{2SLS,n} \xrightarrow{p} \beta_0 + \frac{\sigma_{uv}}{\sigma_{vv}}
\]

Note further that, as shown in Chao and Swanson (2002a), \(\beta_0 + \frac{\sigma_{uv}}{\sigma_{vv}}\) is also the probability limit of the Ordinary Least Squares (OLS) estimator in a local-to-zero framework, so that the 2SLS and the OLS estimators have the same asymptotic bias in the case where the concentration parameter grows at a slower rate than the number of instruments. Hence, under Assumption 5, both 2SLS and OLS are asymptotically deficient relative to the three estimators studied in this paper.

4 Concluding Remarks

This paper derives the limiting distributions of the \(LIML\), \(FLIML\), and \(B2SLS\) estimators in a many weak instruments setup where the concentration parameter is assumed to grow at a slower rate than the number of instruments \(K_n\) but at a faster rate than \(\sqrt{K_n}\). Thus, we have obtained asymptotic normality results for these estimators in situations with weaker instruments than in previous papers that use the many instruments asymptotic framework. In our context, both the rate of convergence and the form of the variance of the limiting distributions are different than for cases where the instruments are stronger, i.e., cases where the instruments grow at the same rate or at a faster rate than \(K_n\). In addition, in contrast to the conventional full-identification case where all three estimators are asymptotically equivalent, we find that the \(B2SLS\) estimator is not asymptotically equivalent to \(LIML\) and \(FLIML\) under the weak instruments scenario studied in
this paper. In particular, we show that LIML and FLIML are asymptotically more efficient than B2SLS if the distribution of the disturbances of the underlying instrumental variables regression model is assumed to belong to the elliptical family.

5 Appendix

Appendix A

In this appendix, we collect some definitions and preliminary lemmas, which we will use to prove our main results.

Definition A1: The \( m \times 1 \) random vector \( X \) is said to have an elliptical distribution with parameters \( \mu \ (m \times 1) \) and \( \Xi \ (m \times m) \) if its density function is of the form

\[
k_m (\det \Xi)^{-\frac{1}{2}} h \left( (x - \mu)' \Xi^{-1} (x - \mu) \right)
\]

for some normalizing constant \( k_m \) and some function \( h (\cdot) \), where \( \Xi \) is positive definite. (Note: A similar definition appears in Muirhead, 1982, page 34.)

Lemma A2:

Let

\[
G_n = P_{Z_n} - P_{X_n} - \left( \frac{K_n}{n - K_n - J} \right) M_{Z_n}
\]

and let \( g_{jj,n} \) and \( g_{ij,n} \) denote, respectively, the \( j^{th} \) diagonal element and the \((i, j)^{th}\) off-diagonal element of the matrix \( G_n \). Then, under Assumptions 2(a) and 2(b), the following statements hold as \( n \to \infty \)

(a) \( Tr(G^1_n) = O_{a.s.}(K_n) \)
(b) \( \sum_{i=1}^{n} \sum_{j=1}^{n} g^4_{ij,n} = O_{a.s.}(K_n) \),
(c) \( \sum_{1 \leq i \leq n} \left[ \sum_{1 \leq j < k \leq n} g^2_{ij,n} g^2_{ik,n} \right] = O_{a.s.}(K_n) \),
(d) \( \sum_{1 \leq i < j \leq n} g^2_{ii,n} g^2_{ij,n} = O_{a.s.}(K_n) \),
(e) \( \sum_{1 \leq i < j \leq n} g^2_{jj,n} g^2_{ij,n} = O_{a.s.}(K_n) \),
(f) \( \sum_{1 \leq i < j < k \leq n} g^2_{ij,n} g^2_{ik,n} = O_{a.s.}(K_n) \),

10
(g) \[ \sum_{1 \leq i < j < k \leq n} g_{ij,n}^2 g_{jk,n}^2 = O_{a.s.}(K_n), \]
(h) \[ \sum_{1 \leq i < j < k \leq n} g_{ik,n}^2 g_{jk,n}^2 = O_{a.s.}(K_n), \]
(i) \[ \text{Tr}(G_n^2) = O_{a.s.}(K_n), \]
(j) \[ \sum_{j=1}^{n} g_{jj,n}^2 = O_{a.s.}(K_n), \]
(k) \[ \sum_{1 \leq i < j \leq n} g_{ij,n}^2 = O_{a.s.}(K_n). \]

Proof of Lemma A2:

To show part (a), note that, by direct calculation,
\[ G_n^4 = P_{Z_n} - P_{X_n} + \left( \frac{K_n}{n - K_n - J} \right)^4 M_{Z_n}, \]
where \( P_{Z_n} \) and \( P_{X_n} \), and thus \( G_n^4 \), are well-defined with probability one for \( n \) sufficiently large given Assumption 2(b). It follows that, with probability one for \( n \) sufficiently large,
\[
\frac{1}{K_n} \text{Tr} \left( G_n^4 \right) = \frac{1}{K_n} \left[ \text{Tr} \left( P_{Z_n} - P_{X_n} \right) + \left( \frac{K_n}{n - K_n - J} \right)^4 \text{Tr} \left( M_{Z_n} \right) \right] = 1 + \frac{K_n^3}{(n - K_n - J)^3},
\]
so that \( \text{Tr} \left( G_n^4 \right) = O_{a.s.}(K_n) \) as required.

To show (b), note that, for \( n \) sufficiently large with probability one, we have
\[
\text{Tr}(G_n^4) = \sum_{1 \leq i \leq n} \left( \sum_{1 \leq j \leq n} g_{ij,n}^2 \right)^2 + \sum_{1 \leq i < j \leq n} \left( \sum_{1 \leq k \leq n} g_{ki,n} g_{kj,n} \right)^2 \\
+ \sum_{1 \leq j < i \leq n} \left( \sum_{1 \leq k \leq n} g_{ki,n} g_{kj,n} \right)^2 \\
\geq \sum_{1 \leq i \leq n} \left( \sum_{1 \leq j \leq n} g_{ij,n}^4 \right)^2 \\
\geq \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij,n}^4,
\]
where \( g_{ij,n} \) denotes the \((i, j)^{th}\) element of \( G_n \). It follows from the result given in part (a) that
\[ O_{a.s.}(K_n) = \text{Tr}(G_n^4) \geq \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij,n}^4. \]
Similarly, for part (c), we have, for \( n \) sufficiently large with probability one, that

\[
\text{Tr}(G_n^4) \geq \sum_{1 \leq i \leq n} \left( \sum_{1 \leq j \leq n} g_{ij,n}^2 \right)^2 \geq 2 \sum_{1 \leq i \leq n} \left[ \sum_{1 \leq j < k \leq n} g_{ij,n}^2 g_{ik,n}^2 \right];
\]

so again the result given in part (a) implies that

\[
O_{a.s.}(K_n) = \frac{1}{2} \text{Tr}(G_n^4) \geq \sum_{1 \leq i \leq n} \left[ \sum_{1 \leq j < k \leq n} g_{ij,n}^2 g_{ik,n}^2 \right].
\]

To show parts (d)-(h), we note that part (c) of this lemma implies that

\[
O_{a.s.}(K_n) = \frac{1}{2} \text{Tr}(G_n^4) \geq \sum_{1 \leq i \leq n} \left[ \sum_{1 \leq j < k \leq n} g_{ij,n}^2 g_{ik,n}^2 \right].
\]

The results stated in parts (d)-(h) then follow directly from the expression on the right-hand side of the last equality in (18) above since each term of the sum which comprises that expression is non-negative.

The proofs for parts (i)-(k) are very similar to the proofs for parts (a)-(h) by noting that

\[
G_n^2 = P_{Z_n} - P_{X_n} + \left( \frac{K_n}{n - K_n - J} \right)^2 M_{Z_n}.
\]

Hence, to avoid redundancy, we omit these proofs. □

**Lemma A3:**

Let \( G_n \) and \( g_{ij,n} \) and \( g_{ij,n} \) be as defined in Lemma A2. Then, under Assumptions 2(a) and 2(b) as \( n \to \infty \),

\[
\sum_{1 \leq i < j \leq n} \left( E\left( g_{ij,n}^2 \right) \right)^2 = O(K_n) \quad \text{and} \quad \sum_{1 \leq i < j < k \leq n} E\left( g_{ij,n}^2 \right) E\left( g_{ik,n}^2 \right) = O(K_n).
\]
Proof of Lemma A3: To proceed, note that part (a) of Lemma A2 implies that
\[
O(K_n) = E[Tr(G_n^4)] \\
\geq E \left[ \sum_{1 \leq i \leq n} \left( \sum_{1 \leq j \leq n} g_{ij,n}^2 \right)^2 \right] \\
\geq \sum_{i=1}^{n} \left( \sum_{j=1}^{n} E(g_{ij,n}^2) \right)^2 \\
\geq \sum_{1 \leq i < j \leq n} \left( E(g_{ij,n}^2) \right)^2 + \sum_{1 \leq i < k \leq n} E(g_{ij,n}^2) E(g_{ik,n}^2),
\tag{19}
\]
where the second inequality above follows from application of the Jensen’s inequality. The desired result follows immediately from (19) by noting that both \( \sum_{1 \leq i < j \leq n} \left( E(g_{ij,n}^2) \right)^2 \) and
\[
\sum_{1 \leq i < j < k \leq n} E(g_{ij,n}^2) E(g_{ik,n}^2)
\]
are non-negative, so they cannot be of an order greater than \( K_n \).

Lemma A4: Define the bilinear form
\[
W_n = d_1 v_n G_n u_n + d_2 u_n G_n v_n,
\tag{20}
\]
where \( d_1 \) and \( d_2 \) are constants and \( G_n \) is as defined in (17) above. Let \( \sigma_{W_n}^2 \) denote the variance of \( W_n \). Suppose Assumptions 2-4 hold, then
\[
\begin{align*}
(a) \quad & \sigma_{W_n}^2 = \sum_{j=1}^{n} d_1^2 E \left( g_{jj,n}^2 \right) \left[ E \left( u_j^2 v_j^2 \right) - \sigma_{uv}^2 \right] + \sum_{j=1}^{n} d_2^2 E \left( g_{jj,n}^2 \right) \left[ E \left( u_j^4 \right) - \sigma_{uu}^2 \right] \\
& \quad + 2 \left\{ \sum_{j=1}^{n} d_1 d_2 E \left( g_{jj,n}^2 \right) \left[ E \left( u_j^2 v_j \right) - \sigma_{uv} \sigma_{uv} \right] \right\} + 2 \left\{ \sum_{1 \leq i < j \leq n} d_1^2 E \left( g_{ij,n}^2 \right) (\sigma_{uu} \sigma_{uv} + \sigma_{uv}^2) \right\} \\
& \quad + 2 \left( \sum_{1 \leq i < j \leq n} d_2^2 E \left( g_{ij,n}^2 \right) \sigma_{uu}^2 \right) + 4 \sum_{1 \leq i < j \leq n} d_1 d_2 E \left( g_{ij,n}^2 \right) \sigma_{uu} \sigma_{uv} \\
(b) \quad & \sigma_{W_n}^2 \asymp K_n,
\end{align*}
\]
where \( g_{jj,n} \) and \( g_{ij,n} \) denote, respective, the \( j^{th} \) diagonal element and the \( (i,j)^{th} \) off-diagonal element of the matrix \( G_n \) and where, for two sequences \( x_n \) and \( y_n \), the notation “\( x_n \asymp y_n \)” means that \( x_n \) is of the same order as \( y_n \), i.e., \( x_n \asymp y_n \) if and only if \( x_n = O(y_n) \) and \( y_n = O(x_n) \).

Proof of Lemma A4:
To show part (a), note that we can write \( W_n = \sum_{j=1}^{n} W_{jn} \), where
\[
W_{jn} = d_1 g_{jj,n} (u_j v_j - \sigma_{uv}) + \sum_{1 \leq i < j} d_1 g_{ij,n} (v_i u_j + v_j u_i) \\
+ d_2 g_{jj,n} (u_j^2 - \sigma_{uu}) + 2 \sum_{1 \leq i < j} d_2 g_{ij,n} u_i u_j
\tag{21}
\]
and where expression (21) has made use of the fact that $G_n$ is a symmetric matrix. Moreover, given that $\eta_i = (u_i, v_i)$ is an independent sequence by Assumption 4(a), it is easy to see that

$$\sigma_W^2 = \sum_{j=1}^n E \left( W_{jn}^2 \right).$$

It follows by straightforward calculation that

$$\sigma_W^2 = \sum_{j=1}^n d_1^2 E \left( g_{jj,n}^2 \right) \left[ E \left( u_j^2 v_j^2 \right) - \sigma_{uv}^2 \right] + \sum_{j=1}^n d_2^2 E \left( g_{jj,n}^2 \right) \left[ E \left( u_j^4 \right) - \sigma_{uu}^2 \right]$$

$$+ 2 \left\{ \sum_{j=1}^n d_1 d_2 E \left( g_{jj,n}^2 \right) \left[ E \left( u_j^2 v_j^2 \right) - \sigma_{uu} \sigma_{uv} \right] \right\}$$

$$+ 2 \left\{ \sum_{1 \leq i < j \leq n} d_1^2 E \left( g_{ij,n}^2 \right) \left( \sigma_{uu} \sigma_{uv} + \sigma_{uv}^2 \right) + 2 \sum_{1 \leq i < j \leq n} d_2^2 E \left( g_{ij,n}^2 \right) \sigma_{uu}^2 \right\}$$

$$+ 4 \sum_{1 \leq i < j \leq n} d_1 d_2 E \left( g_{ij,n}^2 \right) \sigma_{uu} \sigma_{uv} \right\} \right.$$ (22)

as required.

To show part (b), we first show that $\sigma_W^2$ is at most of order $K_n$. To show this, note that

$$\sigma_W^2 = \left\{ \sum_{j=1}^n d_1^2 E \left( g_{jj,n}^2 \right) \left[ E \left( u_j^2 v_j^2 \right) - \sigma_{uv}^2 \right] + \sum_{j=1}^n d_2^2 E \left( g_{jj,n}^2 \right) \left[ E \left( u_j^4 \right) - \sigma_{uu}^2 \right] \right\}$$

$$2 \sum_{j=1}^n d_1 d_2 E \left( g_{jj,n}^2 \right) \left[ E \left( u_j^2 v_j^2 \right) - \sigma_{uu} \sigma_{uv} \right]$$

$$+ 2 \left\{ \sum_{1 \leq i < j \leq n} d_1^2 E \left( g_{ij,n}^2 \right) \left( \sigma_{uu} \sigma_{uv} + \sigma_{uv}^2 \right) + 2 \sum_{1 \leq i < j \leq n} d_2^2 E \left( g_{ij,n}^2 \right) \sigma_{uu}^2 \right\}$$

$$+ 4 \sum_{1 \leq i < j \leq n} d_1 d_2 E \left( g_{ij,n}^2 \right) \sigma_{uu} \sigma_{uv} \right\}$$

$$\leq 2 \left( d_1^2 + d_2^2 + 2d_1 d_2 \right) D_2 \sum_{j=1}^n E \left( g_{jj,n}^2 \right)$$

$$+ 4 \left( d_1^2 + d_2^2 + 4d_1 d_2 \right) D_2 \sum_{1 \leq i < j \leq n} E \left( g_{ij,n}^2 \right)$$

$$= O \left( K_n \right), \quad (23)$$

where the last equality is implied by parts (j) and (k) of Lemma A2.
Next, we show that $\sigma^2_{W_n}$ is not of an order lower than $K_n$. To proceed, note that

$$
\sigma^2_{W_n} = \varpi_1^2 \left( \sum_{j=1}^{n} E \left[ g_{jj,n}^2 \right] \right) + 2 \varpi_2^2 \left( \sum_{1 \leq i < j \leq n} E \left[ g_{ij,n}^2 \right] \right)
$$

$$
\geq \varpi_*^2 \left( \sum_{i=1}^{n} \sum_{j=1}^{n} E \left[ g_{ij,n}^2 \right] \right)
$$

$$
= \varpi_*^2 E \left[ \text{Tr} \left( G_n^2 \right) \right]
$$

$$
= \varpi_*^2 \left( K_n + \frac{K_n^2}{n - K_n - j} \right),
$$

(24)

where $\varpi_1^2 = E \left( d_2 \left[ u_j v_j - \sigma_{uv} \right] \right)$, $\varpi_2^2 = E \left( d_2 \left[ u_j v_j + u_i v_j \right] + d_3 u_i u_j \right)^2$, and $\varpi_*^2 = \min \{ \varpi_1^2, \varpi_2^2 \}$ and where the last equality follows from direct calculation. The desired result follows immediately from expressions (23) and (24) given Assumption 2(a). $\square$

**Lemma A5:** Let $G_n$ be as defined in (17) above and let $g_{jj,n}$ and $g_{ij,n}$ denote, respective, the $j^{th}$ diagonal element and the $(i, j)^{th}$ off-diagonal element of the matrix $G_n$. Then, under Assumption 2-4 as $n \to \infty$,

$$
\frac{1}{K_n^2} \sum_{1 \leq i < j < k < l \leq n} E \left( g_{ik,n} g_{jk,n} g_{il,n} g_{jl,n} \right) = o(1)
$$

(25)

**Proof of Lemma A5:**

We will prove this lemma in two steps. First, we will show that

$$
\frac{1}{K_n^2} \sum_{1 \leq i < j < k < l \leq n} \left[ g_{ik,n} g_{jk,n} g_{il,n} g_{jl,n} + g_{ij,n} g_{jk,n} g_{il,n} g_{kl,n} + g_{ij,n} g_{ik,n} g_{jl,n} g_{kl,n} \right] = o_a.s. (1).
$$

(26)

We will then use (26) to show the desired result (25). To proceed, first define

$$
\overline{G}_n = G_n - dg \left( G_n \right),
$$

where $dg \left( G_n \right) = \text{diag} \left( g_{11,n}, \ldots, g_{nn,n} \right)$, i.e., $dg \left( G_n \right)$ is an $n \times n$ diagonal matrix whose diagonal elements are the same as that of $G_n$. Now, note that, by direct calculation, we obtain

$$
\text{Tr} \left( G_n^4 \right) = \sum_{1 \leq i \leq n} \left( \sum_{j \neq i} g_{ij,n}^2 \right)^2 + 2 \sum_{1 \leq i < j \leq n} \left( \sum_{k \neq i, k \neq j} g_{ki,n} g_{kj,n} \right)^2
$$

$$
= 2 \sum_{1 \leq i < j \leq n} g_{ij,n}^4 + 4 \sum_{1 \leq i < j < k \leq n} \left[ g_{ik,n}^2 g_{jk,n}^2 + g_{ij,n}^2 g_{kl,n}^2 + g_{ij,n}^2 g_{jk,n}^2 \right] + 8 \sum_{1 \leq i < j < k < l \leq n} \left[ g_{ik,n} g_{jk,n} g_{il,n} g_{jl,n} + g_{ij,n} g_{jk,n} g_{il,n} g_{kl,n} + g_{ij,n} g_{ik,n} g_{jl,n} g_{kl,n} \right],
$$

(27)
where $P_{Z_n}$ and $P_{X_n}$ and, thus, $\overline{C}_n$ and $\overline{C}_n^4$ are each well-defined with probability one for $n$ sufficiently large in light of Assumption 2(b). Now, let $\lambda_{1,n} \leq \lambda_{2,n} \leq \cdots \leq \lambda_{n,n}$ be the eigenvalues of the matrix $\overline{C}_n$, and note that

$$\text{Tr}(\overline{C}_n^4) = \sum_{i=1}^{n} \lambda_{i,n}^4.$$  \hfill (28)

Next, observe that part (b) and parts (f)-(h) of Lemma A2 imply that

$$\frac{1}{K^2} \sum_{1 \leq j < k \leq n} g_{jk,n}^4 = O_{a.s.} \left( K_n^{-1} \right),$$  \hfill (29)

$$\frac{1}{K^2} \sum_{1 \leq i < j < k \leq n} \left( g_{ik,n}^2 g_{jk,n}^2 + g_{ij,n}^2 g_{jk,n}^2 + g_{ij,n}^2 g_{ik,n}^2 \right) = O_{a.s.} \left( K_n^{-1} \right)$$  \hfill (30)

It follows from equations (27)-(30) that showing that

$$\frac{1}{K^2} \sum_{1 \leq i < j < k < l \leq n} \left[ g_{ik,n} g_{jk,n} g_{il,n} g_{jl,n} + g_{ij,n} g_{jk,n} g_{il,n} g_{kl,n} + g_{ij,n} g_{ik,n} g_{jl,n} g_{kl,n} \right] = o_{a.s.} (1), \quad \text{as } n \to \infty,$$

is equivalent to showing that

$$\frac{1}{K^2} \sum_{i=1}^{n} \lambda_{i,n}^4 = o_{a.s.} (1) \quad \text{as } n \to \infty.$$  \hfill (31)

To show (31), we first note that, for each $n$,

$$\lambda_{n,n}^2 \leq \sum_{i=1}^{n} \lambda_{i,n}^2 \leq \lambda_{n,n}^2 \left( \sum_{i=1}^{n} \lambda_{i,n}^2 \right)$$  \hfill (32)
and

\[
\frac{1}{K_n} \sum_{i=1}^{n} \lambda_{i,n}^2 = \frac{1}{K_n} \text{Tr} \left( G_n^2 \right)
\]

\[
\leq \frac{1}{K_n} \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij,n}^2
\]

\[
= \frac{1}{K_n} \text{Tr} \left( G_n^2 - dg(G_n) G_n - G_n dg(G_n) + [dg(G_n)]^2 \right)
\]

\[
= \frac{1}{K_n} \text{Tr} \left( G_n^2 - [dg(G_n)]^2 \right)
\]

\[
= \frac{1}{K_n} \left\{ \text{Tr} \left[ P_{Z,n} - P_{X,n} + \left( \frac{K_n}{n-K_n-J} \right)^2 M_{ZX} \right] \right\}
\]

\[
+ \sum_{j=1}^{n} \left[ \left( \frac{n-J}{n-K_n-J} \right) p_{j,j,n}^Z - p_{j,j,n}^X - \left( \frac{K_n}{n-K_n-J} \right)^2 \right] \right\}
\]

\[
\leq \frac{1}{K_n} \left\{ \text{Tr} \left[ P_{Z,n} - P_{X,n} + \left( \frac{K_n}{n-K_n-J} \right)^2 M_{ZX} \right] \right\}
\]

\[
+ \left( \frac{n-J}{n-K_n-J} \right)^2 \sum_{j=1}^{n} \left( p_{j,j,n}^Z \right)^2 \right\}
\]

\[
\leq \frac{1}{K_n} \left\{ K_n + \frac{K_n^2}{n-K_n-J} + \left( \frac{n-J}{n-K_n-J} \right)^2 K_n \right\}
\]

\[
= 1 + \left( \frac{n-J}{n-K_n-J} \right)^2 + \frac{K_n}{n-K_n-J}
\]

where \( p_{j,j,n}^Z \) and \( p_{j,j,n}^X \) are the \( j^{th} \) diagonal elements of the projection matrices \( P_{Z,n} \) and \( P_{X,n} \). It follows from Assumption 2(a) that \( \frac{1}{K_n} \sum_{i=1}^{n} \lambda_{i,n}^2 = O_a.s. (1) \). Hence, to show (31), we need to show that

\[
\frac{1}{K_n} \lambda_{n,n}^2 = o_a.s. (1) \quad \text{as} \quad n \to \infty.
\] (34)

To show (34), we proceed as follows: let \( x_n \) be any \( n \times 1 \) vector such that \( \|x_n\| = 1 \) and let \( x_{j,n} \) denote the \( j^{th} \) element of \( x_n \). Now, consider the quadratic form

\[
x_n^T G_n^2 x_n = x_n^T G_n^2 x_n - x_n^T [dg(G_n)] G_n x_n
\]

\[
- x_n^T G_n [dg(G_n)] x_n + x_n^T [dg(G_n)]^2 x_n
\]

\[
\leq x_n^T G_n^2 x_n + |x_n^T [dg(G_n)] G_n x_n|
\]

\[
+ |x_n^T [G_n dg(G_n)] x_n| + x_n^T [dg(G_n)]^2 x_n
\]

Note that, for \( n \) sufficiently large so that \( P_{Z,n} \) and \( P_{X,n} \) are well-defined with probability one, we
have that

$$x_n' [dg(G_n)]^2 x_n = \sum_{j=1}^{n} \left[ \left( \frac{n-J}{n-K_n-J} \right) p_{jj,n}^Z - p_{jj,n}^X - \frac{K_n}{n-K_n-J} \right]^2 x_{j,n}^2$$

$$\leq \left( \frac{n-J}{n-K_n-J} \right)^2 \sum_{i=1}^{n} x_{i,n}^2$$

$$= \left( \frac{n-J}{n-K_n-J} \right)^2 x_n' x_n$$

$$= \left( \frac{n-J}{n-K_n-J} \right)^2,$$  \hspace{1cm} (36)

where inequality above follows from the fact that $0 \leq p_{jj,n}^Z \leq 1$, $0 \leq p_{jj,n}^X \leq 1$, and $\left( \frac{K_n}{n-K_n-J} \right) > 0$, and note that

$$x_n' G_n^2 x_n \leq x_n' P_{Z_n} x_n + x_n' P_{X_n} x_n + \left( \frac{K_n}{n-K_n-J} \right)^2 x_n' M_{Z_n} x_n$$

$$\leq 2 + \left( \frac{K_n}{n-K_n-J} \right)^2,$$ \hspace{1cm} (37)

where the inequality follows from the Rayleigh quotient by making use of the fact that $\lambda_{\max}(P_{Z_n}) = \lambda_{\max}(P_{X_n}) = \lambda_{\max}(M_{Z_n}) = 1$ since $P_{Z_n}$, $P_{X_n}$, and $M_{Z_n}$ are idempotent matrices. (See pages 203-204 of Magnus and Neudecker, 1988, for a statement of the Rayleigh quotient.) It then follows from the Cauchy-Schwarz inequality that

$$|x_n' [dg(G_n) G_n] x_n| \leq \sqrt{x_n' [dg(G_n)]^2 x_n} \sqrt{x_n' G_n^2 x_n}$$

$$= \left( \frac{n-J}{n-K_n-J} \right) \sqrt{2 + \left( \frac{K_n}{n-K_n-J} \right)^2},$$  \hspace{1cm} (38)

Define

$$\Delta_n = \left( \frac{n-J}{n-K_n-J} \right)^2 + 2 + \left( \frac{K_n}{n-K_n-J} \right)^2$$

$$+ 2 \left( \frac{n-J}{n-K_n-J} \right) \sqrt{2 + \left( \frac{K_n}{n-K_n-J} \right)^2}$$

and note that, for $n$ sufficiently large so that $\overline{G_n}$ is well-defined with probability one, expressions (36), (37), and (38) imply that $x_n' \overline{G_n} x_n \leq \Delta_n$ for any $n \times 1$ vector $x_n$ such that $\|x_n\| = 1$. Moreover, since Assumption 2(a) implies that

$$\lim_{n \to \infty} \Delta_n = 2 + \left( \frac{1}{1-\alpha} \right)^2 + \left( \frac{\alpha}{1-\alpha} \right)^2 + 2 \left( \frac{1}{1-\alpha} \right) \sqrt{2 + \left( \frac{\alpha}{1-\alpha} \right)^2}$$

$$< \infty,$$
so that there exist a positive constant $\Delta_\alpha$ and a positive integer $N$ such that for all $n \geq N$

$$\Delta_n \leq \Delta_\alpha < \infty.$$ 

It then follows that, for all $n \geq N$,

$$\lambda^2_{n,n} = \max_{x_n: \|x_n\|=1} x_n^T \tilde{G}_n x_n \leq \Delta_n \leq \Delta_\alpha < \infty$$

with probability one, from which (34) and thus (26) follow immediately as $K_n \to \infty$.

Next, we show that (26) implies the desired result (25). To proceed, first define

$$\zeta_{1n} = \sum_{1 \leq i < j < k \leq n} g_{ik,n}g_{jk,n} (v_i v_j \sigma_{uu} + u_i v_j \sigma_{uv} + v_i u_j \sigma_{uv} + u_i u_j \sigma_{vv}),$$

and note that

$$E (\zeta^2_{1n} = (2\sigma^2_{uu} \sigma^2_{vv} + 12 \sigma_{uu} \sigma_{uv} \sigma^2_{uv} + 2 \sigma^4_{uv}) \left[ \sum_{1 \leq i < j < k \leq n} E (g^2_{ik,n} g^2_{jk,n}) + 2 \sum_{1 \leq i < j < k < l \leq n} E (g_{ik,n} g_{jk,n} g_{il,n} g_{jl,n}) \right].$$

Since part (h) of Lemma A2 implies that

$$\frac{1}{K^2_n} \sum_{1 \leq i < j < k \leq n} E (g^2_{ik,n} g^2_{jk,n}) = o(1),$$

and Assumption 4, that

$$\frac{1}{K^2_n} \sum_{1 \leq i < j < k < l \leq n} E (g_{ik,n} g_{jk,n} g_{il,n} g_{jl,n}) = o(1) \quad (39)$$

if and only if

$$\frac{1}{K^2_n} E (\zeta^2_{1n}) = o(1). \quad (40)$$

To show equation (40), further define

$$\zeta_{2n} = \sum_{1 \leq i < j < k \leq n} \left[ g_{ik,n}g_{jk,n} (v_i v_j \sigma_{uu} + u_i v_j \sigma_{uv} + v_i u_j \sigma_{uv} + u_i u_j \sigma_{vv}) + g_{ij,n}g_{jk,n} (v_i v_k \sigma_{uu} + u_i v_k \sigma_{uv} + v_i u_k \sigma_{uv} + u_i u_k \sigma_{vv}) + g_{ij,n}g_{ik,n} (v_j v_k \sigma_{uu} + v_j u_k \sigma_{uv} + v_j u_k \sigma_{uv} + v_j u_k \sigma_{vv}) \right],$$

$$\zeta_{3n} = \sum_{1 \leq i < j < k \leq n} \left[ g_{ij,n}g_{jk,n} (v_i v_k \sigma_{uu} + u_i v_k \sigma_{uv} + v_i u_k \sigma_{uv} + u_i u_k \sigma_{vv}) + g_{ij,n}g_{ik,n} (v_j v_k \sigma_{uu} + u_j v_k \sigma_{uv} + v_j u_k \sigma_{uv} + u_j u_k \sigma_{vv}) \right],$$

and note that $\zeta_{1n} = \zeta_{2n} - \zeta_{3n}$, so that

$$E (\zeta^2_{1n}) = E (\zeta^2_{2n}) + E (\zeta^2_{3n}) - 2E (\zeta_{2n}\zeta_{3n}). \quad (41)$$
By direct calculation, we obtain

\[ E(\zeta^2_n) = T_1 + 4T_2, \]

where

\[
T_1 = \left( 2\sigma^2_{uu}\sigma^2_{vv} + 12\sigma_{uu}\sigma_{vv}\sigma^2_{uv} + 2\sigma^4_{uv} \right) \times \sum_{1 \leq i < j < k \leq n} E\left[ g_{ik,n}g_{jk,n}^2 + g_{ij,n}g_{jk,n}^2 + g_{ij,n}g_{ik,n}^2 \right],
\]

\[
T_2 = \left( 2\sigma^2_{uu}\sigma^2_{vv} + 12\sigma_{uu}\sigma_{vv}\sigma^2_{uv} + 2\sigma^4_{uv} \right) \left[ \sum_{1 \leq i < j < k < l \leq n} E\left( g_{ik,n}g_{il,n}g_{jk,n}g_{jl,n} \right) + \sum_{1 \leq i < j < k < l \leq n} E\left( g_{ij,n}g_{ik,n}g_{jl,n}g_{kl,n} \right) \right].
\]

and

\[ E(\zeta^3_n) = T_3 + 2T_2, \]

where

\[
T_3 = \left( 2\sigma^2_{uu}\sigma^2_{vv} + 12\sigma_{uu}\sigma_{vv}\sigma^2_{uv} + 2\sigma^4_{uv} \right) \sum_{1 \leq i < j < k \leq n} E\left[ g_{ij,n}g_{jk,n}^2 + g_{ij,n}g_{ik,n}^2 \right],
\]

Next, observe that Assumption 4 and Lemma A2 parts (f)-(h) imply that \( K^{-2}T_1 = o(1) \) and \( K^{-2}T_3 = o(1) \). In addition, (26) implies that

\[
\frac{1}{K_n^2} \sum_{1 \leq i < j < k < l \leq n} \left[ E\left( g_{ik,n}g_{jk,n}g_{il,n}g_{jl,n} \right) + E\left( g_{ij,n}g_{jk,n}g_{il,n}g_{kl,n} \right) + E\left( g_{ij,n}g_{ik,n}g_{jl,n}g_{kl,n} \right) \right] = o(1),
\]

so that \( K^{-2}T_2 = o(1) \) given Assumption 4. It follows that

\[
K_n^{-2}E(\zeta^2_n) \to 0, \quad (42)
\]

\[
K_n^{-2}E(\zeta^3_n) \to 0. \quad (43)
\]

The Jensen and Cauchy-Schwarz inequalities then imply that, as \( n \to \infty \),

\[
K_n^{-2}|E(\zeta_{2n}\zeta_{3n})| \leq K_n^{-2}E|\zeta_{2n}\zeta_{3n}| \leq \sqrt{K_n^{-2}E(\zeta^2_{2n})} \sqrt{K_n^{-2}E(\zeta^2_{3n})} \to 0. \quad (44)
\]

\( K_n^{-2}E(\zeta^2_{1n}) \to 0 \) then follows as a direct consequence of (42), (43), and (44) in view of equation (41). □

**Lemma A6:** Under Assumptions 1-5, \( b_n^{-1}K_n^{-\frac{1}{2}}\zeta_n\Gamma_n'M_nu_n \xrightarrow{P} 0 \) as \( n \to \infty \).

20
Proof of Lemma A6:

We will show the mean square convergence of \( b_n^{-1} K_n^{-\frac{1}{2}} c_n' Z_n' M X_n u_n \) to zero. To proceed, note that Assumptions and the law of iterated expectations imply that

\[
E \left[ \frac{c_n' Z_n' M X_n u_n}{b_n \sqrt{K_n}} \right] = E_{Z_n} \left[ \frac{c_n' Z_n' M X_n E(u_n | Z_n)}{b_n \sqrt{K_n}} \right] = E_{Z_n} \left[ \frac{c_n' Z_n' M X_n E(u_n)}{b_n \sqrt{K_n}} \right] = 0 \tag{45}
\]

\[
E \left[ \left( \frac{c_n' Z_n' M X_n u_n}{b_n \sqrt{K_n}} \right)^2 \right] = E_{Z_n} \left[ \frac{c_n' Z_n' M X_n E(u_n u'_n | Z_n) M X_n Z_n c_n}{b_n^2 K_n} \right] = \sigma_{uu} \left( \frac{r_n}{K_n} \right) E_{Z_n} \left[ \frac{c_n' Z_n' M X_n Z_n c_n}{b_n^2 r_n} \right] = O \left( \frac{r_n}{K_n} \right) = o(1) \tag{46}
\]

given that \( \frac{r_n}{K_n} \to 0 \) as \( n \to \infty \), where the expectation \( E_{Z_n} \left[ \frac{c_n' Z_n' M X_n Z_n c_n}{b_n^2 r_n} \right] \) exists for \( n \) sufficiently large in light of Assumptions 2. The desired result follows immediately from (45) and (46). □

Lemma A7: (Gänsler and Stute, 1977)

Let \( \{ X_{i,n}, \mathcal{F}_{i,n}, 1 \leq i \leq l_n, n \geq 1 \} \) be a square integrable martingale difference array. Also, let \( l_n \not\to \infty \) as \( n \to \infty \), and suppose that for all \( \varepsilon > 0 \)

\[
\sum_{i=1}^{l_n} E \left[ X_{i,n}^2 I(|X_{i,n}| > \varepsilon) | \mathcal{F}_{i-1,n} \right] \overset{P}{\to} 0 \tag{C1}
\]

and

\[
\sum_{i=1}^{l_n} E \left[ X_{i,n}^2 | \mathcal{F}_{i-1,n} \right] \overset{P}{\to} 1. \tag{C2}
\]

Then, \( \sum_{i=1}^{l_n} X_{i,n} \overset{d}{\to} N(0, 1) \).


Remark: Note that, as discussed in Kelejian and Prucha (1999), a sufficient condition for condition (C1) is the following:
Condition C1' :
\[ \sum_{j=1}^{k_n} E \left\{ E \left[ |X_{j,n}|^{2+\delta} \mid \mathcal{F}_{j-1,n} \right] \right\} \rightarrow 0 \]
for some \( \delta > 0 \).

Since condition C1' is easier to verify in our case, in the proofs which follow, we will be verifying condition C1' instead of condition C1 for the case \( \delta = 2 \).

**Lemma A8:** Let \( W_n \) be as defined in (20) above and let \( \sigma_{W_n}^2 \) be the variance of \( W_n \) with explicit formula given in expression (22). Define
\[ B_n = \sigma_{W_n}^{-1} W_n. \]  
(47)

Then, under Assumptions 2-4,
\[ B_n \overset{d}{\rightarrow} N(0, 1) \quad \text{as} \quad n \rightarrow \infty. \]

**Proof of Lemma A8:**

The proof of this lemma involves verifying conditions C1' and C2 which jointly imply the central limit theorem given in Lemma A7. As discussed in the Remark above, we shall verify conditions C1' in lieu of condition C1. The proof is, thus, divided into two parts: in part I, we check condition C1' and, in part II, we check condition C2.

**I. Checking Condition C1':**

As in the proof of Lemma A3, we can write \( W_n = \sum_{j=1}^{n} W_{jn} \), where \( W_{jn} \) is as defined in (21) above. To verify condition C1' for \( \delta = 2 \), we need to show that \( \sum_{j=1}^{n} E \left\{ (\sigma_{W_n}^{-1} W_{jn})^4 \right\} \rightarrow 0 \) as \( n \rightarrow \infty \). In light of Lemma A4 part (b), this is equivalent to showing that \( K_n^{-2} \sum_{j=1}^{n} E \left( W_{jn}^4 \right) \rightarrow 0 \) as \( n \rightarrow \infty \). To proceed, note that direct calculation yields the following expression for the fourth moment of \( W_{jn} \)
\[ E \left( W_{jn}^4 \right) = \sum_{i=1}^{8} \mathcal{E}_{ij,n}, \]
where
\[ \mathcal{E}_{1j,n} = d_1^4 E \left( g_{jj,n}^4 \right) E \left( u_j v_j - \sigma_{uv} \right)^4 + 6d_1^2 d_2^2 E \left( g_{jj,n}^4 \right) E \left[ (u_j^2 - \sigma_{uu})^2 (u_j v_j - \sigma_{uv})^2 \right] \]
\[ + d_2^4 E \left( g_{jj,n}^4 \right) E \left( u_j^2 - \sigma_{uu} \right)^4 + 4d_1^3 d_2 E \left( g_{jj,n}^4 \right) E \left[ (u_j v_j - \sigma_{uv})^3 (u_j^2 - \sigma_{uu}) \right] \]
\[ + 4d_1 d_2^3 E \left( g_{jj,n}^4 \right) E \left[ (u_j v_j - \sigma_{uv}) (u_j^2 - \sigma_{uu})^3 \right], \]
\[ \varepsilon_{2j,n} = 4 \left\{ \sum_{1 \leq i < j} d_1^4 E \left(g_{jj,n}^2 g_{ij,n}^2\right) E \left[ (u_j v_j - \sigma_{uv})^2 (v_i u_j + v_j u_i)^2 \right] \right. \\
+ 4 \sum_{1 \leq i < j} d_1^3 d_2 E \left(g_{jj,n}^2 g_{ij,n}^2\right) E \left[ u_i^2 u_j^2 (u_j^2 - \sigma_{uu})^2 \right] \\
+ 4 \sum_{1 \leq i < j} d_1^2 d_2^2 E \left(g_{jj,n}^2 g_{ij,n}^2\right) E \left[ u_i^2 u_j^2 (u_j v_j - \sigma_{uv})^2 \right] \\
+ \sum_{1 \leq i < j} d_1^2 d_2^2 E \left(g_{jj,n}^2 g_{ij,n}^2\right) E \left[ (u_j^2 - \sigma_{uu})^2 (v_i u_j + v_j u_i)^2 \right] \right\} 
\]

\[ \varepsilon_{3j,n} = 4 \left\{ \sum_{1 \leq i < j} d_1^2 d_2^2 E \left(g_{jj,n}^2 g_{ij,n}^2\right) E \left[ u_i u_j (u_j^2 - \sigma_{uu}) (u_j v_j - \sigma_{uv}) (v_i u_j + v_j u_i) \right] \right. \\
+ 2 \sum_{1 \leq i < j} d_1^3 d_2 E \left(g_{jj,n}^2 g_{ij,n}^2\right) E \left[ u_i u_j (u_j v_j - \sigma_{uv})^2 (v_i u_j + v_j u_i) \right] \\
+ \sum_{1 \leq i < j} d_1^3 d_2 E \left(g_{jj,n}^2 g_{ij,n}^2\right) E \left[ (u_j^2 - \sigma_{uu}) (u_j v_j - \sigma_{uv}) (v_i u_j + v_j u_i)^2 \right] \\
+ 4 \sum_{1 \leq i < j} d_1 d_2 E \left(g_{jj,n}^2 g_{ij,n}^2\right) E \left[ u_i^2 u_j^2 (u_j^2 - \sigma_{uu}) (u_j v_j - \sigma_{uv}) \right] \\
+ 2 \sum_{1 \leq i < j} d_1 d_2 E \left(g_{jj,n}^2 g_{ij,n}^2\right) E \left[ u_i u_j (u_j^2 - \sigma_{uu})^2 (v_i u_j + v_j u_i) \right] \\
+ \sum_{1 \leq i < j} d_1 d_2 E \left(g_{jj,n}^2 g_{ij,n}^2\right) E \left[ u_i u_j (u_j v_j - \sigma_{uv}) (u_j^2 - \sigma_{uu}) (v_i u_j + v_j u_i) \right] \right\} 
\]

\[ \varepsilon_{4j,n} = \sum_{1 \leq i < j} d_1^4 E \left(g_{ij,n}^4\right) E \left[(v_i u_j + v_j u_i)^4\right] + 96 \sum_{1 \leq h < i < j} d_1^4 E \left(g_{hj,n}^2 g_{ij,n}^2\right) \sigma_{uu}^2 E \left(u_j^4\right) \\
+ 6 \sum_{1 \leq h < i < j} d_1^4 E \left(g_{hj,n}^2 g_{ij,n}^2\right) E \left[(v_h u_j + v_j u_h)^2 (v_i u_j + v_j u_i)^2\right] \\
+ 16 \sum_{1 \leq i < j} d_1^2 d_2 E \left(g_{ij,n}^4\right) E \left(u_i^4\right) E \left(u_j^4\right) + 16 \sum_{1 \leq i < j} d_1^2 d_2 E \left(g_{ij,n}^4\right) E \left[u_i^2 u_j^2 (v_i u_j + v_j u_i)^2\right] \\
+ 32 \sum_{1 \leq h < i < j} d_1^2 d_2 E \left(g_{hj,n}^2 g_{ij,n}^2\right) E \left[u_h u_i u_j^2 (v_h u_j + v_j u_h) (v_i u_j + v_j u_i)\right] \\
+ 16 \sum_{1 \leq h < i < j} d_1^2 d_2 E \left(g_{hj,n}^2 g_{ij,n}^2\right) E \left[u_i^2 u_j^2 (v_h u_j + v_j u_h)^2\right] \\
+ 16 \sum_{1 \leq h < i < j} d_1^2 d_2 E \left(g_{hj,n}^2 g_{ij,n}^2\right) E \left[u_h^2 u_j^2 (v_i u_j + v_j u_i)^2\right] \]
\[ \mathcal{E}_{ij,n} = 8 \sum_{1 \leq i < j} d_1^2 d_2^2 E \left( g_{ij,n}^4 \right) E \left[ u_i^2 u_j^2 (v_i u_j + v_j u_i)^2 \right] \\
+ 8 \sum_{1 \leq h < i < j} d_1^2 d_2^2 E \left( g_{h,n}^2 g_{ij,n}^2 \right) E \left[ u_h^2 u_i^2 (v_i u_j + v_j u_i)^2 \right] \\
+ 8 \sum_{1 \leq h < i < j} d_1^2 d_2^2 E \left( g_{h,n}^2 g_{ij,n}^2 \right) E \left[ u_i^2 u_j^2 (v_h u_j + v_j u_h)^2 \right] \\
+ 8 \sum_{1 \leq i < j} d_1^3 d_2 E \left( g_{ij,n}^4 \right) E \left[ u_i u_j (v_i u_j + v_j u_i)^3 \right] \\
+ 32 \sum_{1 \leq i < j} d_1^3 d_2^3 E \left( g_{ij,n}^4 \right) E \left[ u_i^3 u_j^3 (v_i u_j + v_j u_i) \right] \\
+ 8 \sum_{1 \leq h < i < j} d_1^2 d_2 E \left( g_{h,n}^2 g_{ij,n}^2 \right) E \left[ u_i u_j (v_i u_j + v_j u_i) (v_h u_j + v_j u_h)^2 \right] \\
+ 8 \sum_{1 \leq h < i < j} d_1^2 d_2 E \left( g_{h,n}^2 g_{ij,n}^2 \right) E \left[ u_h u_j (v_i u_j + v_j u_i)^2 (v_h u_j + v_j u_h) \right] \\
+ 96 \sum_{1 \leq h < i < j} d_1 d_2^3 E \left( g_{h,n}^2 g_{ij,n}^2 \right) E \left[ u_i^2 u_j^2 (v_h u_j + v_j u_h) \right] \\
+ 96 \sum_{1 \leq h < i < j} d_1 d_2^3 E \left( g_{h,n}^2 g_{ij,n}^2 \right) E \left[ u_i^2 u_j (v_i u_j + v_j u_i) \right] \\
+ 32 \sum_{1 \leq h < i < j} d_1^3 d_2^2 E \left( g_{h,n}^2 g_{ij,n}^2 \right) E \left[ u_i^2 u_j (v_i u_j + v_j u_i) u_h (v_h u_j + v_j u_h) \right] \\
+ 16 \sum_{1 \leq h < i < j} d_1^3 d_2 E \left( g_{h,n}^2 g_{ij,n}^2 \right) E \left[ u_i u_j (v_i u_j + v_j u_i) (v_h u_j + v_j u_h)^2 \right] \\
+ 16 \sum_{1 \leq h < i < j} d_1^3 d_2 E \left( g_{h,n}^2 g_{ij,n}^2 \right) E \left[ u_h u_j (v_i u_j + v_j u_i)^2 (v_h u_j + v_j u_h) \right] \\
+ 32 \sum_{1 \leq h < i < j} d_1^3 d_2^2 E \left( g_{h,n}^2 g_{ij,n}^2 \right) E \left[ u_h u_i u_j^2 (v_i u_j + v_j u_i) (v_h u_j + v_j u_h) \right] \\
+ 2 \sum_{1 \leq i < j} d_1^3 d_2^2 E \left( g_{ij,n}^2 g_{ij,n}^2 \right) E \left[ (u_j v_j - \sigma_{uv})^2 (v_i u_j + v_j u_i)^2 \right] \\
+ \sum_{1 \leq i < j} d_1^3 d_2^2 E \left( g_{ij,n}^2 g_{ij,n}^2 \right) E \left[ (u_j^2 - \sigma_{uu})^2 (v_i u_j + v_j u_i)^2 \right] \\
+ 2 \sum_{1 \leq i < j} d_1^3 d_2 E \left( g_{ij,n}^2 g_{ij,n}^2 \right) E \left[ (u_j v_j - \sigma_{uv}) (u_j^2 - \sigma_{uu}) (v_i u_j + v_j u_i)^2 \right] \right\}
\[ \mathcal{E}_{ij,n} = 8 \left\{ \sum_{1 \leq i < j} d_1^i d_2 E (g_{j,j,n}^2 g_{i,j,n}^2) \sigma_{uu} E \left[ u_j^2 (u_j^2 - \sigma_{uu})^2 \right] \\
+ \sum_{1 \leq i < j} d_1^i d_2 E (g_{j,j,n}^2 g_{i,j,n}^2) \sigma_{uu} E \left[ u_j^2 (u_j^2 - \sigma_{uu})^2 \right] \\
+ 2 \sum_{1 \leq i < j} d_1^i d_2 E (g_{j,j,n}^2 g_{i,j,n}^2) \sigma_{uu} E \left[ u_j^2 (u_j^2 - \sigma_{uu}) (u_j^2 - \sigma_{uu}) \right] \right\} \]

\[ \mathcal{E}_{8j,n} = 8 \left\{ \sum_{1 \leq i < j} d_1^i d_2 E (g_{j,j,n}^2 g_{i,j,n}^2) E \left[ u_i u_j (u_j v_j - \sigma_{uu})^2 (v_i u_j + v_j u_i) \right] \\
+ \sum_{1 \leq i < j} d_1^i d_2 E (g_{j,j,n}^2 g_{i,j,n}^2) E \left[ u_i u_j (u_j^2 - \sigma_{uu})^2 (v_i u_j + v_j u_i) \right] \\
+ 2 \sum_{1 \leq i < j} d_1^i d_2 E (g_{j,j,n}^2 g_{i,j,n}^2) E \left[ u_i u_j (u_j v_j - \sigma_{uu}) (u_j^2 - \sigma_{uu}) (v_i u_j + v_j u_i) \right] \right\} \]

Now, making use of Lemmas and Assumption, we see that

\[ \frac{1}{K_n^2} \sum_{j=1}^{n} |\mathcal{E}_{ij,n}| \leq 16 D_\eta \frac{1}{K_n^2} \sum_{j=1}^{n} E \left( g_{j,j,n}^4 \right) \]

\[ = o(1), \quad (48) \]

\[ \frac{1}{K_n^2} \sum_{j=1}^{n} |\mathcal{E}_{8j,n}| \leq 16 \sqrt{14} D_\eta \left( d_1^2 + d_2^2 + 2 d_1^2 d_2^2 \right) \frac{1}{K_n^2} \sum_{1 \leq i < j \leq n} E \left( g_{j,j,n}^2 g_{i,j,n}^2 \right) \]

\[ = o(1), \quad (49) \]

\[ \frac{1}{K_n^2} \sum_{j=1}^{n} |\mathcal{E}_{3j,n}| \leq 16 \sqrt{14} D_\eta \left( 3 d_1^2 d_2^2 + 2 d_1^3 d_2 + d_1 d_2^3 \right) \frac{1}{K_n^2} \sum_{1 \leq i < j \leq n} E \left( g_{j,j,n}^2 g_{i,j,n}^2 \right) \]

\[ + 16 \sqrt{2} (14)^{1/4} D_\eta^{7/2} d_1 d_2 \frac{1}{K_n^2} \sum_{1 \leq i < j \leq n} E \left( g_{j,j,n}^2 g_{i,j,n}^2 \right) \]

\[ = o(1), \quad (50) \]

\[ \frac{1}{K_n^2} \sum_{j=1}^{n} |\mathcal{E}_{4j,n}| \leq 16 D_\eta \left( d_1^4 + d_2^4 + 4 d_1^2 d_2^2 \right) \frac{1}{K_n^2} \sum_{1 \leq i < j \leq n} E \left( g_{i,j,n}^4 \right) \]

\[ + 32 D_\eta \left( 3 d_1^4 + 3 d_2^4 + 8 d_1^2 d_2^2 \right) \frac{1}{K_n^2} \sum_{1 \leq h < i < j \leq n} E \left( g_{j,j,n}^2 g_{i,j,n}^2 \right) \]

\[ = o(1), \quad (51) \]
\[
\frac{1}{K^2_n} \sum_{j=1}^{n} |E_{5j,n}| \leq 32 D_\eta \left( d_1^2 d_2^2 + 2d_1^3 d_2 + 2d_1 d_2^3 \right) \frac{1}{K^2_n} \sum_{1 \leq i < j \leq n} E \left( g_{ij,n}^4 \right) + 64 D_\eta \left( 5d_2^2 d_3^2 + 6d_1^2 d_2 + 3d_1 d_2^3 \right) \frac{1}{K^2_n} \sum_{1 \leq h < i < j \leq n} E \left( g_{hj,n}^2 g_{ij,n}^2 \right) = o(1), \tag{52}
\]

\[
\frac{1}{K^2_n} \sum_{j=1}^{n} |E_{6j,n}| \leq 8 \sqrt{14} D_\eta \left( d_1^4 + d_1^2 d_2^2 + 2d_1^3 d_2 \right) \frac{1}{K^2_n} \sum_{1 \leq i < j \leq n} E \left( g_{ij,n}^2 g_{ij,n}^2 \right) = o(1), \tag{53}
\]

\[
\frac{1}{K^2_n} \sum_{j=1}^{n} |E_{7j,n}| \leq 8 \sqrt{14} D_\eta \left( d_1^2 d_2^2 + d_1^3 d_2 + 4d_1 d_2^3 \right) \frac{1}{K^2_n} \sum_{1 \leq i < j \leq n} E \left( g_{ij,n}^2 g_{ij,n}^2 \right) = o(1), \tag{54}
\]

\[
\frac{1}{K^2_n} \sum_{j=1}^{n} |E_{8j,n}| \leq 16 \sqrt{14} D_\eta \left( d_1^3 d_2 + d_1^2 d_2^2 + 2d_1 d_2^3 \right) \frac{1}{K^2_n} \sum_{1 \leq i < j \leq n} E \left( g_{ij,n}^2 g_{ij,n}^2 \right) = o(1), \tag{55}
\]

where the inequalities in expressions (48)-(55) are obtained by repeated applications of the Cauchy-Schwarz and the triangle inequalities. From expressions (48)-(55), it follows immediately that

\[
\frac{1}{K^2_n} \sum_{j=1}^{n} E \left( W_{jn}^4 \right) \leq \sum_{i=1}^{8} \left( \frac{1}{K^2_n} \sum_{j=1}^{n} |E_{ij,n}| \right) = o(1) \text{ as } n \to \infty. \tag{56}
\]

**II. Checking Condition C2:**

First define

\[ B_{jn} = \sigma_{W_{jn}}^{-1} W_{jn}, \]

where \( W_{jn} \) and \( \sigma_{W_{jn}}^2 \) are as defined in expression (21) and (22), respectively. Now, consider the \( \sigma \)-fields \( \mathcal{F}_{j,n} = \sigma \left( \eta_1, ..., \eta_j, Z_n \right), i = 1, ..., n \), and take \( \mathcal{F}_{0,n} \) to be the trivial \( \sigma \)-field. It follows that by construction that \( \mathcal{F}_{j-1,n} \subseteq \mathcal{F}_{j,n} \). Moreover, note that \( W_{jn} \) is \( \mathcal{F}_{j,n} \)-measurable, and straightforward calculation shows that \( E \left( W_{jn} \mid \mathcal{F}_{j-1,n} \right) = 0 \), so that \( \{W_{jn}, \mathcal{F}_{j,n}, 1 \leq j \leq n, n \geq 1 \} \) forms a martingale difference array.

Hence, to verify condition 2, we need to show

\[
\sum_{j=1}^{n} E \left[ B_{jn}^2 \mid \mathcal{F}_{j-1,n} \right] \overset{P}{\to} 1, \text{ as } n \to \infty, \tag{57}
\]
or, alternatively,

\[ \sum_{j=1}^{n} \left( \frac{E \left[ W_{j,n}^2 \mid \mathcal{F}_{j-1,n} \right] - E \left[ W_{j,n}^2 \right]}{\sigma_{W_n}^2} \right)^p \to 0, \quad \text{as } n \to \infty. \]  

(58)

(58), in turn, is implied by

\[ \frac{1}{K_n^2} E \left( \sum_{j=1}^{n} \left\{ E \left[ W_{j,n}^2 \mid \mathcal{F}_{j-1,n} \right] - E \left[ W_{j,n}^2 \right] \right\}^2 \right) \to 0, \quad \text{as } n \to \infty, \]  

(59)
in light of the result we obtained in part (b) of Lemma A4. To show (59), we proceed by noting that

\[
\sum_{j=1}^{n} \left( E \left[ W_{j,n}^2 \mid \mathcal{F}_{j-1,n} \right] - E \left[ W_{j,n}^2 \right] \right) \\
= \sum_{1 \leq i < j \leq n} d_1^2 \left[ \sigma_{uu} \left( g_{ij,n}^2 v_i^2 - E \left( g_{ij,n}^2 \right) \sigma_{uv} \right) + 2 \sigma_{uv} \left( g_{ij,n}^2 u_i v_i - E \left( g_{ij,n}^2 \right) \sigma_{uv} \right) \right] \\
+ \sum_{1 \leq i < j \leq n} d_2^2 \sigma_{uu} \left( g_{ij,n}^2 u_i^2 - E \left( g_{ij,n}^2 \right) \sigma_{uu} \right) \\
+ 2 \sum_{1 \leq h < i < j \leq n} d_1 g_{hj,n} g_{ij,n} \left[ \sigma_{uu} v_h v_i + \sigma_{uv} u_h v_i + \sigma_{uu} u_i v_h + \sigma_{uv} u_i v_h \right] \\
+ 8 \sum_{1 \leq h < i < j \leq n} d_2 g_{hj,n} g_{ij,n} \sigma_{uu} u_h u_i \\
+ 4 \sum_{1 \leq i < j \leq n} d_1 d_2 \left[ \sigma_{uu} \left( g_{ij,n}^2 u_i v_i - E \left( g_{ij,n}^2 \right) \sigma_{uv} \right) + \sigma_{uv} \left( g_{ij,n}^2 u_i^2 - E \left( g_{ij,n}^2 \right) \sigma_{uu} \right) \right] \\
+ 4 \sum_{1 \leq h < i < j \leq n} d_1 d_2 g_{hj,n} g_{ij,n} \left[ \sigma_{uu} v_h u_i + \sigma_{uv} u_h u_i \right] \\
+ 4 \sum_{1 \leq h < i < j \leq n} d_1 d_2 g_{hj,n} g_{ij,n} \left[ \sigma_{uu} u_h v_i + \sigma_{uv} u_h u_i \right]
\]

Further calculations yield

\[
\frac{1}{K_n^2} E \left( \sum_{j=1}^{n} \left\{ E \left[ W_{j,n}^2 \mid \mathcal{F}_{j-1,n} \right] - E \left[ W_{j,n}^2 \right] \right\} ^2 \right) = \frac{1}{K_n^2} \sum_{i=1}^{4} A_{i,n},
\]

where

\[ A_{i,n} \]
\[ A_{1,n} = \sum_{1 \leq i < j \leq n} d_1^4 E (g_{ij,n}^4) \left[ \sigma_{uu}^2 E (v_i^4) + 4 \sigma_{uu}^2 E (u_i^2 v_i^2) + \sigma_{vv}^2 E (u_i^4) \right] \\
+ 4 \sigma_{uu} \sigma_{uu} E (u_i v_i^3) + 2 \sigma_{uu} \sigma_{vv} E (u_i^2 v_i^2) + 4 \sigma_{vv} \sigma_{uu} E (u_i^3 v_i) \right] \\
- 4 \sum_{1 \leq i < j \leq n} d_1^4 E (g_{ij,n}^2) \left[ \sigma_{uu}^2 \sigma_{vv}^2 + \sigma_{uv}^4 + 2 \sigma_{uu} \sigma_{vv} \sigma_{uv}^2 \right] \\
+ 16 \sum_{1 \leq i < j \leq n} d_1^4 E (g_{ij,n}^4) \sigma_{uu}^2 E (u_i^4) - (E (g_{ij,n}^2))^2 \sigma_{uu}^4 \right] \\
+ 8 \sum_{1 \leq h < i < j \leq n} d_1^3 E (u_i^4) \left[ \sigma_{uu}^2 \sigma_{vv}^2 + 6 \sigma_{uu}^2 \sigma_{uu} \sigma_{vv} + \sigma_{uv}^4 \right] \\
+ 64 \sum_{1 \leq h < i < j \leq n} d_1^2 d_2^2 E (g_{ij,n}^4) \left[ \sigma_{uu}^2 E (u_i^2 v_i^2) + 16 \sum_{1 \leq i < j \leq n} d_1^2 d_2^2 E (g_{ij,n}^4) \left[ \sigma_{uu}^2 E (u_i^2 v_i^2) \\
+ \sigma_{uu}^2 E (u_i^4) + 2 \sigma_{uu} \sigma_{uv} E (u_i^2 v_i) \right] - 64 \sum_{1 \leq i < j \leq n} d_1^2 d_2^2 E (g_{ij,n}^2) \right] \sigma_{uu}^2 \sigma_{uv}^2 \\
+ 32 \sum_{1 \leq h < i < j \leq n} d_1^2 d_2^2 E (g_{ij,n}^2 g_{hk,n}^2) \left[ \sigma_{uu}^3 \sigma_{vv} + 3 \sigma_{uu}^2 \sigma_{uv}^2 \right] \\
+ 2 \left\{ \sum_{1 \leq i < j < k \leq n} d_1^4 E (g_{ij,n}^4) \left[ \sigma_{uu}^2 E (v_i^4) + 4 \sigma_{uu}^2 E (u_i^2 v_i^2) + \sigma_{vv}^2 E (u_i^4) \right] \\
+ 4 \sigma_{uu} \sigma_{uu} E (u_i v_i^3) + 2 \sigma_{uu} \sigma_{vv} E (u_i^2 v_i^2) + 4 \sigma_{vv} \sigma_{uu} E (u_i^3 v_i) \right] \\
- 4 \sum_{1 \leq i < j < k \leq n} d_1^4 E (g_{ij,n}^2) \left[ \sigma_{uu}^2 \sigma_{vv}^2 + \sigma_{uv}^4 + 2 \sigma_{uu} \sigma_{vv} \sigma_{uv}^2 \right] \\
+ 16 \sum_{1 \leq i < j < k \leq n} d_1^4 E (g_{ij,n}^2) \left[ \sigma_{uu}^2 E (u_i^4) - (E (g_{ij,n}^2))^2 \sigma_{uu}^4 \right] \\
+ 8 \sum_{1 \leq h < i < j < k \leq n} d_1^3 E (g_{ij,n}^2 g_{jk,n}^2 g_{hk,n}^2) \left[ \sigma_{uu}^2 \sigma_{vv}^2 + 6 \sigma_{uu}^2 \sigma_{uu} \sigma_{vv} + \sigma_{uv}^4 \right] \\
+ 64 \sum_{1 \leq h < i < j < k \leq n} d_1^2 d_2^2 E (g_{ij,n}^4) \left[ \sigma_{uu}^2 E (u_i^2 v_i^2) + 16 \sum_{1 \leq i < j < k \leq n} d_1^2 d_2^2 E (g_{ij,n}^4) \right] \sigma_{uu}^2 \sigma_{uv}^2 \\
+ 32 \sum_{1 \leq h < i < j < k \leq n} d_1^2 d_2^2 E (g_{ij,n}^2 g_{jk,n}^2 g_{hk,n}^2) \left[ \sigma_{uu}^3 \sigma_{vv} + 3 \sigma_{uu}^2 \sigma_{uv}^2 \right] \right\} \]
$$A_{3,n} = 2 \left\{ \sum_{1 \leq i < j \leq n} d_1^2 d_2^2 E (g_{ij,n}) \left[ \sigma_{uu}^2 E (u_i^2 v_i^2) + 2 \sigma_{uu} \sigma_{uv} E (u_i^3 v_i) + \sigma_{uu} \sigma_{vv} E (u_i^4) \right] \\
-2 \sum_{1 \leq i < j \leq n} d_1^2 d_2^2 (E (g_{ij,n}))^2 \left[ \sigma_{uu}^3 \sigma_{vv} + \sigma_{uu}^2 \sigma_{uv}^2 \right] \\
+4 \sum_{1 \leq i < j \leq n} d_1^3 d_2 E (g_{ij,n}) \left[ \sigma_{uu}^2 E (u_i v_i^3) + 3 \sigma_{uu} \sigma_{uv} E (u_i^2 v_i^2) + \sigma_{uv}^3 \sigma_{vv} E (u_i^4) \right] \\
-16 \sum_{1 \leq i < j \leq n} d_1^3 d_2 (E (g_{ij,n}))^2 \left[ \sigma_{uu}^2 \sigma_{vv} \sigma_{uv} + \sigma_{uu} \sigma_{uv}^3 \right] \\
+16 \sum_{1 \leq i < j \leq n} d_1 d_2^2 E (g_{ij,n}) \left[ \sigma_{uu}^2 E (u_i^3 v_i) + \sigma_{uu} \sigma_{uv} E (u_i^4) \right] \\
-32 \sum_{1 \leq i < j \leq n} d_1 d_2^2 (E (g_{ij,n}))^2 \sigma_{uu}^3 \sigma_{uv} \\
+16 \sum_{1 \leq h < i < j \leq n} d_1^2 d_2^2 E (g_{hj,n} g_{ij,n}) \left[ 3 \sigma_{uu}^2 \sigma_{uv}^2 + \sigma_{uu} \sigma_{uv}^3 \right] \\
+64 \sum_{1 \leq h < i < j \leq n} d_1^3 d_2 E (g_{hj,n} g_{ij,n}) \left[ \sigma_{uu}^2 \sigma_{vv} \sigma_{uv} + \sigma_{uu} \sigma_{uv}^3 \right] \\
+128 \sum_{1 \leq h < i < j \leq n} d_1 d_2^3 E (g_{hj,n} g_{ij,n}) \sigma_{uu}^3 \sigma_{uv} + 64 \sum_{1 \leq h < i < j \leq n} d_1 d_2^3 E (g_{hj,n} g_{ij,n}) \sigma_{uu}^2 \sigma_{uv} \right\}$$
\[
A_{1,n} = 4 \left\{ \sum_{1 \leq i < j < k \leq n} d_i^2 d_j^2 E \left( g_{ij,n}^2 g_{ik,n}^2 \right) \left[ \sigma_{uu}^2 E \left( u_i^2 v_i^2 \right) + 2\sigma_{uu} \sigma_{uv} E \left( u_i^3 v_i \right) + \sigma_{uu} \sigma_{vv} E \left( u_i^4 \right) \right] \\
-2 \sum_{1 \leq i < j < k \leq n} d_i^2 d_j^2 E \left( g_{ij,n}^2 \right) E \left( g_{ik,n}^2 \right) \left[ \sigma_{uu}^3 \sigma_{uv} + 2\sigma_{uu} \sigma_{vv}^2 \right] \\
+4 \sum_{1 \leq i < j < k \leq n} d_i^2 d_2 E \left( g_{ij,n}^2 g_{ik,n}^2 \right) \left[ \sigma_{uu}^2 E \left( u_i v_i^3 \right) + 3\sigma_{uu} \sigma_{uv} E \left( u_i^2 v_i^2 \right) \right] \\
+ 2\sigma_{uu}^2 E \left( u_i^3 v_i \right) + \sigma_{uu} \sigma_{vv} E \left( u_i^3 v_i \right) + \sigma_{vv} \sigma_{uv} E \left( u_i^4 \right) \right] \\
-16 \sum_{1 \leq i < j < k \leq n} d_i^2 d_2 E \left( g_{ij,n}^2 \right) E \left( g_{ik,n}^2 \right) \left[ \sigma_{uu}^2 \sigma_{uv} \sigma_{uv} + \sigma_{uu} \sigma_{vv}^3 \right] \\
+16 \sum_{1 \leq i < j < k \leq n} d_i^2 d_2 E \left( g_{ij,n}^2 g_{ik,n}^2 \right) \left[ \sigma_{uu}^2 E \left( u_i^3 v_i \right) + \sigma_{uu} \sigma_{uu} E \left( u_i^4 \right) \right] \\
-32 \sum_{1 \leq i < j < k \leq n} d_i^2 d_2 E \left( g_{ij,n}^2 \right) E \left( g_{ik,n}^2 \right) \sigma_{uu}^2 \sigma_{uv} \right). \\
\]

Again, making use of Lemma A2 and Assumption 3 and 4, we see that
\[
\frac{1}{K_n^2} |A_{1,n}| \leq 16 \left( d_1^4 + d_2^4 + 4d_1^2 d_2^2 \right) D_n \frac{1}{K_n^2} \sum_{1 \leq i < j \leq n} E \left( g_{ij,n}^4 \right) \\
+16 \left( d_1^4 + d_2^4 + 4d_1^2 d_2^2 \right) D_n \frac{1}{K_n^2} \sum_{1 \leq i < j \leq n} \left( E \left( g_{ij,n}^2 \right) \right)^2 \\
+64 \left( d_1^4 + d_2^4 + 2d_1^2 d_2^2 \right) D_n \frac{1}{K_n^2} \sum_{1 \leq h < i < j \leq n} E \left( g_{ij,n}^2 g_{ij,n}^2 \right) \\
= O \left( K_n^{-1} \right) = o(1), \quad (60)
\]

\[
\frac{1}{K_n^2} |A_{2,n}| \leq 32 \left( d_1^4 + d_2^4 + 4d_1^2 d_2^2 \right) D_n \frac{1}{K_n^2} \sum_{1 \leq i < j < k \leq n} E \left( g_{ij,n}^2 g_{ik,n}^2 \right) \\
+32 \left( d_1^4 + d_2^4 + 4d_1^2 d_2^2 \right) D_n \frac{1}{K_n^2} \sum_{1 \leq i < j < k \leq n} E \left( g_{ij,n}^2 \right) E \left( g_{ik,n}^2 \right) \\
+128 \left( d_1^4 + d_2^4 + 2d_1^2 d_2^2 \right) D_n \frac{1}{K_n^2} \sum_{1 \leq h < i < j < k \leq n} E \left( g_{ij,n}^2 g_{ij,n}^2 \right) g_{hk,n}^2 g_{ik,n}^2 \right) \\
= o(1), \quad (61)
\]
\[
\frac{1}{K_n^2} |A_{4,n}| \leq 8 \left( d_1^2 d_2 + 8d_1^3 d_2 + 8d_1 d_2^3 \right) D_{\eta} \frac{1}{K_n^2} \sum_{1 \leq i < j \leq n} E (g_{ij,n}^2) + 8 \left( d_1^2 d_2 + 8d_1^3 d_2 + 8d_1 d_2^3 \right) D_{\eta} \frac{1}{K_n^2} \sum_{1 \leq i < j \leq n} \left( E (g_{ij,n}^2) \right)^2 + 128 \left( d_1^2 d_2 + 2d_1^3 d_2 + 3d_1 d_2^3 \right) D_{\eta} \frac{1}{K_n^2} \sum_{1 \leq h < i < j \leq n} E (g_{h,i,n}^2 g_{j,n}^2) = O \left( K_n^{-1} \right) = o(1),
\]

where the inequalities in expressions (60)-(63) have been obtained by repeated applications of the Cauchy-Schwarz and the triangle inequalities. The (59) follows directly from expressions ((60)-(63)).

\[\square\]

**Lemma A9:** Under assumptions, let \( \tilde{\lambda}_{LIML,n} \) be the smallest root of the determinantal equation given by (8). Then, under Assumptions 1-5,

\[
\tilde{\lambda}_{LIML,n} = \frac{n - J}{n - K_n - J} + \left( \frac{\sqrt{K_n}}{n - K_n - J} \right) \frac{s_{uu}^G}{\sigma_{uu}} + o_p \left( \frac{\sqrt{K_n}}{n - K_n - J} \right),
\]

where

\[ s_{uu}^G = \frac{u_n' G_n u_n}{\sqrt{K_n}} \]

and where \( G_n \) is defined in (17) above.

**Proof of Lemma A9:** To proceed, note first that, by definition, \( \tilde{\lambda}_{LIML,n} \) is the smallest root of the determinantal equation

\[
\det \left\{ \begin{pmatrix} y_{1n} M_{X_n} y_{1n} & y_{1n} M_{X_n} y_{2n} \\ y_{2n} M_{X_n} y_{1n} & y_{2n} M_{X_n} y_{2n} \end{pmatrix} - \lambda_n \begin{pmatrix} y_{1n} M_{Z_n} y_{1n} & y_{1n} M_{Z_n} y_{2n} \\ y_{2n} M_{Z_n} y_{1n} & y_{2n} M_{Z_n} y_{2n} \end{pmatrix} \right\} = 0
\]

or, in more succinct notation,

\[
\det \left\{ Y_n' M_{X_n} Y_n - \lambda_n Y_n' M_{Z_n} Y_n \right\} = 0,
\]

where \( Y_n = [y_{1n}, y_{2n}] \) and where the elements of the determinantal equation given above are all well-defined with probability one for \( n \) sufficiently large, as a consequence of Assumption 2. Now,
define \( \Upsilon = \begin{pmatrix} 1 & 0 \\ -\beta_0 & 1 \end{pmatrix} \) and note that the smallest root of equation (64) is the same as the smallest root of the equation

\[
\det \left\{ \Upsilon' Y_n' M_{X_n} Y_n - \lambda_n \Upsilon' \tilde{Y}_n' M_{\tilde{Z}_n} \tilde{Y}_n \Upsilon \right\} = 0,
\]

(66)

where

\[
\Upsilon' Y_n' M_{X_n} Y_n \Upsilon = \begin{pmatrix} 1 & -\beta_0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{1n} M_{X_n} y_{1n} & y_{1n} M_{X_n} y_{2n} \\ y_{2n} M_{X_n} y_{1n} & y_{2n} M_{X_n} y_{2n} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\beta_0 & 1 \end{pmatrix}
\]

(67)

and

\[
\Upsilon' \tilde{Y}_n' M_{\tilde{Z}_n} \tilde{Y}_n \Upsilon = \begin{pmatrix} u_n M_{\tilde{Z}_n} u_n & u_n M_{\tilde{Z}_n} v_n \\ v_n M_{\tilde{Z}_n} u_n & v_n M_{\tilde{Z}_n} v_n \end{pmatrix}.
\]

(68)

Now, let \( \lambda_n = \frac{n - J}{n - K_n - J} + \frac{\tau_n r_n J}{n - K_n - J} \) and rewrite (66) as

\[
\det \left\{ \begin{pmatrix} u_n' M_{X_n} y_{1n} & u_n' M_{X_n} y_{2n} \\ y_{2n} M_{X_n} y_{1n} & y_{2n} M_{X_n} y_{2n} \end{pmatrix} - (\frac{n - J}{n - K_n - J}) \begin{pmatrix} u_n' M_{\tilde{Z}_n} u_n & u_n' M_{\tilde{Z}_n} v_n \\ v_n' M_{\tilde{Z}_n} u_n & v_n' M_{\tilde{Z}_n} v_n \end{pmatrix} \\
- \tau_n \begin{pmatrix} r_n u_n M_{\tilde{Z}_n} V_n \cr r_n M_{\tilde{Z}_n} V_n \cr r_n M_{\tilde{Z}_n} V_n \cr r_n M_{\tilde{Z}_n} V_n \end{pmatrix} \right\} = 0,
\]

(69)

which, in turn, can be shown, by straightforward manipulation, to be equivalent to the determinantal equation

\[
\det \left\{ \begin{pmatrix} u_n' M_{X_n} y_{1n} & u_n' M_{X_n} y_{2n} \\ y_{2n} M_{X_n} y_{1n} & y_{2n} M_{X_n} y_{2n} \end{pmatrix} - (\frac{n - J}{n - K_n - J}) \begin{pmatrix} u_n' M_{\tilde{Z}_n} u_n & u_n' M_{\tilde{Z}_n} v_n \\ v_n' M_{\tilde{Z}_n} u_n & v_n' M_{\tilde{Z}_n} v_n \end{pmatrix} \\
- \tau_n \begin{pmatrix} r_n u_n M_{\tilde{Z}_n} V_n \cr r_n M_{\tilde{Z}_n} V_n \cr r_n M_{\tilde{Z}_n} V_n \cr r_n M_{\tilde{Z}_n} V_n \end{pmatrix} \right\} = 0.
\]

(70)

Moreover, it is apparent that \( \hat{\lambda}_{LIML,n} \), the smallest root of equation (64), is related to \( \hat{\tau}_{LIML,n} \), the smallest root of (70), by the equation

\[
\hat{\lambda}_{LIML,n} = \frac{n - J}{n - K_n - J} + \frac{\hat{\tau}_{LIML,n} r_n}{n - K_n - J}.
\]

(71)

Furthermore, note that \( \hat{\tau}_{LIML,n} \) is also the smallest root of the determinantal equation

\[
\det \left\{ \begin{pmatrix} u_n' G_n u_n & u_n' M_{X_n} Z_n c_n \\ c_n' M_{X_n} Z_n u_n & c_n' M_{X_n} Z_n c_n + c_n' G_n v_n \\ u_n' M_{X_n} Z_n c_n & u_n' G_n v_n \\ c_n' G_n u_n & c_n' G_n v_n + c_n' Z_n M_{X_n} Z_n c_n \end{pmatrix} - \tau_n \begin{pmatrix} r_n u_n M_{\tilde{Z}_n} V_n \cr r_n M_{\tilde{Z}_n} V_n \cr r_n M_{\tilde{Z}_n} V_n \cr r_n M_{\tilde{Z}_n} V_n \end{pmatrix} \right\} = 0.
\]

(72)
Now, rewrite (72) as follows

\[
\det \left\{ \begin{pmatrix} \frac{\sqrt{K_n}}{r_n} s_{uu} & \frac{\sqrt{K_n}}{r_n} s_{uv} \\ \frac{1}{\sqrt{n^*}}x_{cu} + \frac{\sqrt{K_n}}{r_n} s_{uv} & \frac{1}{\sqrt{n^*}}x_{cu} + \frac{\sqrt{K_n}}{r_n} s_{uv} \end{pmatrix} + \frac{\sqrt{K_n}}{r_n} s_{uv} \right\} - \tau_n \left( \begin{pmatrix} \sigma_{uu} + \frac{1}{\sqrt{n^*}} s_{uu}^M & \sigma_{uv} + \frac{1}{\sqrt{n^*}} s_{uv}^M \\ \sigma_{uv} + \frac{1}{\sqrt{n^*}} s_{uv}^M & \sigma_{vv} + \frac{1}{\sqrt{n^*}} s_{vv}^M \end{pmatrix} \right) = 0,
\]

(73)

where \( n^* = n - K_n - J, s_{uu}^G = \frac{u'_n G_n u_n}{\sqrt{K_n}}, s_{uv}^G = \frac{u'_n G_n v_n}{\sqrt{K_n}}, s_{vv}^G = \frac{v'_n G_n v_n}{\sqrt{K_n}}, x_{cu} = \frac{c'_n Z_{X_n} u_n}{b_n \sqrt{r_n}}, \) and the quadratic relationships

\[
\begin{align*}
\left( \frac{\sqrt{K_n}}{r_n} s_{uu} - \tau_n \left[ \sigma_{uu} + O_p \left( \frac{1}{\sqrt{n^*}} \right) \right] \right) & \left( \Psi_n + \frac{\sqrt{K_n}}{r_n} s_{uv} + O_p \left( \frac{\sqrt{K_n}}{r_n} \right) \right) - \tau_n \left[ \sigma_{uv} + O_p \left( \frac{1}{\sqrt{n^*}} \right) \right] \left( \frac{\sqrt{K_n}}{r_n} s_{uv} + O_p \left( \frac{1}{\sqrt{n^*}} \right) \right) \\
- \left( \frac{\sqrt{K_n}}{r_n} s_{uv} + O_p \left( \frac{\sqrt{K_n}}{r_n} \right) \right) & - \tau_n \left[ \sigma_{uv} + O_p \left( \frac{1}{\sqrt{n^*}} \right) \right] \right)^2 = 0,
\end{align*}
\]

so that by rearranging terms, we obtain, up to terms of order \( O_p \left( K_n^2 r_n^{-1} \right) \), the quadratic relationship

\[
\left( \sigma_{uu} \sigma_{vv} - \sigma_{uv}^2 + O_p \left( \frac{1}{\sqrt{n^*}} \right) \right) \tau_n^2 \\
- \left( \sigma_{uu} \Psi_n + \sqrt{K_n} r_n s_{uv} \sigma_{uu} + \sqrt{K_n} r_n s_{uv} \sigma_{vv} - 2 \sqrt{K_n} r_n s_{uv} \sigma_{uv} + O_p \left( \frac{\sqrt{K_n}}{r_n} \right) \right) \tau_n \\
+ \sqrt{K_n} r_n s_{uu} \Psi_n + O_p \left( \frac{K_n}{r_n} \right) = 0.
\]
It follows from the quadratic formula that

\[ \hat{\tau}_{LIML,n} = \left[ 2 \left( \sigma_{uu} \sigma_{vv} - \sigma_{uv}^2 + O_p \left( \frac{1}{\sqrt{n}} \right) \right) \right]^{-1} \times \left\{ \sigma_{uu} \Psi_n + \sqrt{\frac{K_n}{r_n}} G \sigma_{uu} + \sqrt{\frac{K_n}{r_n}} G \sigma_{vv} - 2 \sqrt{\frac{K_n}{r_n}} G \sigma_{uv} + o_p \left( \frac{\sqrt{K_n}}{r_n} \right) \right\} \]

\[ = \left[ 2 \left( \sigma_{uu} \sigma_{vv} - \sigma_{uv}^2 + O_p \left( \frac{1}{\sqrt{n}} \right) \right) \right]^{-1} \times \left\{ \sigma_{uu} \Psi_n + \frac{\sqrt{K_n}}{r_n} \left[ G \sigma_{uu} + G \sigma_{vv} - 2 G \sigma_{uv} \right] + o_p \left( \frac{\sqrt{K_n}}{r_n} \right) \right\} \]

Now, focusing on the square root function

\[ R_{1n} = \sigma_{uu} \Psi_n \left[ \sigma_{uu} \Psi_n + 2 \frac{\sqrt{K_n}}{r_n} \left[ G \sigma_{uu} - G \sigma_{vv} - 2 G \sigma_{uv} \right] + o_p \left( \frac{\sqrt{K_n}}{r_n} \right) \right] \]

we note that we can expand \( R_{1n} \) as a power series as follows:

\[ R_{1n} = \sigma_{uu} \Psi_n \left( 1 + \frac{\sqrt{K_n}}{r_n} \sigma_{uu}^{-1} \Psi_n^{-1} \left[ G \sigma_{uu} - G \sigma_{vv} - 2 G \sigma_{uv} \right] + o_p \left( \frac{\sqrt{K_n}}{r_n} \right) \right) \]  \hspace{1cm} (76)

Inserting (76) into (75), we obtain, after minor manipulations,

\[ \hat{\tau}_{LIML,n} = \left[ 2 \left( \sigma_{uu} \sigma_{vv} - \sigma_{uv}^2 \right) \right]^{-1} \left( 1 + O_p \left( \frac{1}{\sqrt{n}} \right) \right) \times \left\{ \frac{\sqrt{K_n}}{r_n} \sigma_{uu} \left[ G \sigma_{uu}^2 - G \sigma_{uv}^2 \right] + o_p \left( \frac{\sqrt{K_n}}{r_n} \right) \right\} \]

\[ = \frac{\sqrt{K_n} G \sigma_{uu}}{r_n} \left( 1 + o_p \left( \sqrt{n} \right) \right) \]  \hspace{1cm} (77)

The desired result follows immediately by substituting (77) into (71). □
Appendix B

This part of the appendix contains proofs of the main theorems of this paper.

Proof of Theorem 3.1:

By the usual regression algebra, we can write

\[ \hat{\beta}_{LIML,n} - \beta_0 = \left( y_{2n}' [ P_{Z_n} - P_{X_n} - \tilde{\lambda}_{LIML,n} M_{Z_n} ] y_{2n} \right)^{-1} \times \left( y_{2n}' [ P_{Z_n} - P_{X_n} - \tilde{\lambda}_{LIML,n} M_{Z_n} ] u_n \right), \]

so that

\[ \frac{\Psi_n}{\sigma_{L,n}} ( \hat{\beta}_{LIML,n} - \beta_0 ) = \left( \frac{y_{2n}' [ P_{Z_n} - P_{X_n} - \tilde{\lambda}_{LIML,n} M_{Z_n} ] y_{2n}}{r_n} \right)^{-1} \times \left( \frac{y_{2n}' [ P_{Z_n} - P_{X_n} - \tilde{\lambda}_{LIML,n} M_{Z_n} ] u_n}{\sigma_{L,n}} \right), \] (78)

where the inverse in (78) exists in probability as \( n \to \infty \) in the sense of White (1984) given our assumptions, as will be shown in expression (81) below. (See page 24 of White, 1984, for a definition of “existence in probability”) To derive the limiting distribution of (78), first write

\[ \frac{y_{2n}' [ P_{Z_n} - P_{X_n} - \tilde{\lambda}_{LIML,n} M_{Z_n} ] u_n}{\sigma_{L,n}} = \frac{y_{2n}' [ P_{Z_n} - P_{X_n} - (\tilde{\lambda}_{LIML,n} - 1) M_{Z_n} ] u_n}{\sigma_{L,n}} \]

where the first equality above follows from the definition of \( \tilde{\lambda}_{LIML,n} \). It then follows from Lemmas A6 and A9 that

\[ \frac{y_{2n}' [ P_{Z_n} - P_{X_n} - \tilde{\lambda}_{LIML,n} M_{Z_n} ] u_n}{\sigma_{L,n}} = \left( \frac{\sqrt{K_n}}{\sigma_{L,n}} \right) \left( \frac{\sqrt{\eta_n}}{K_n} \right) \frac{c_n' Z_n' M_{X_n} u_n}{b_n \sqrt{K_n}} + o_p(1) \]

\[ = \left( \frac{\sqrt{K_n}}{\sigma_{L,n}} \right) \left( \frac{\sqrt{\eta_n}}{K_n} \right) \frac{c_n' Z_n' M_{X_n} u_n}{b_n \sqrt{K_n}} + o_p(1) \]

\[ = \frac{v_n' G_n u_n}{\sigma_{L,n}} + o_p(1) \]

\[ = \frac{v_n' G_n u_n}{\sigma_{L,n}} + o_p(1) \] (79)
where $G_n = PZ_n - P_{X_n} - \left(\frac{K_n}{n-K_n-1}\right) M_{Z_n}$, where the second equality above from part (d) of Lemma A2 of Chao and Swanson (2002b), which show that $\frac{v_n\cdot M_{Z_n} u_n}{\sigma_{uv}} \rightarrow \sigma_{uv}$, and where the last equality above follows from arguments similar to that given in part (e) of Lemma A1 of Chao and Swanson (2002b), which can be used to show that $\frac{c_n Z_n^t M_{X_n} u_n}{b_n \sqrt{r_n}} = O_p(1)$ and from part (d) of Lemma A2 of Chao and Swanson (2002b). Note also that by setting $d_1 = 1$ and $d_2 = -\frac{\sigma_{uv}}{\sigma_{uu}}$ in the general formula (22), we deduce that

$$
\sigma_{L,n}^2 = \left[ E(u_j^2 v_j^2) - \sigma_{uv}^2 \right] \sum_{j=1}^n E(g_{jj,n}^2) + \frac{\sigma_{uu}^2}{\sigma_{uv}^2} \left[ E(u_j^4) - \sigma_{uu}^2 \right] \sum_{j=1}^n E(g_{jj,n}^2) 
$$

$$
-2 \left\{ \frac{\sigma_{uv}}{\sigma_{uu}} \left[ E(u_j^3 v_j) - \sigma_{uu} \sigma_{uv} \right] \sum_{j=1}^n E(g_{jj,n}^2) \right\}
$$

$$
+2 \left( \sigma_{uu} \sigma_{vv} - \sigma_{uv}^2 \right) \sum_{1 \leq i < j \leq n} E(g_{ij,n}^2)
$$

is the variance of the bilinear form $v_n^t G_n u_n - \frac{\sigma_{uv}}{\sigma_{uu}} u_n^t G_n u_n$. It follows from Lemma A8 above that, as $n \rightarrow \infty$,

$$
y_{2n}^t \left[ PZ_n - P_{X_n} - \bar{\lambda}_{LIML,n} M_{Z_n} \right] u_n \frac{\sigma_{L,n}}{\sigma_{L,n}} = v_n^t G_n u_n - \frac{\sigma_{uv}}{\sigma_{uu}} u_n^t G_n u_n + o_p(1) \xrightarrow{d} N(0, 1).
$$

(80)

Note further that

$$
\frac{y_{2n}^t \left[ PZ_n - P_{X_n} - \bar{\lambda}_{LIML,n} M_{Z_n} \right] y_{2n}}{r_n} = \frac{c_n^t Z_n^t M_{X_n} Z_n c_n}{b_n^2 r_n} + 2 \frac{c_n^t Z_n^t M_{X_n} v_n}{b_n r_n} + \frac{v_n^t \left[ PZ_n - P_{X_n} - \bar{\lambda}_{LIML,n} M_{Z_n} \right] v_n}{r_n}
$$

$$
= \frac{c_n^t Z_n^t M_{X_n} Z_n c_n}{b_n^2 r_n} + 2 \frac{c_n^t Z_n^t M_{X_n} v_n}{b_n r_n} + \frac{v_n^t \left[ PZ_n - P_{X_n} - \left(\frac{K_n}{n-K_n-1}\right) M_{Z_n} \right] v_n}{r_n} + o_p(1)
$$

$$
= \Psi_n + o_p(1),
$$

(81)

where $\Psi_n = r_n^{-1} \Psi_n$ is nonsingular with probability one for $n$ sufficiently large given Assumption 2, where the second equality above follows from Theorem 3.3 of Chao and Swanson (2002b), and
where the third equality follows from parts (c) and (f) of Lemma A1. (80) and (81) imply that

\[
(\Psi_n \frac{\beta_{LIML,n} - \beta_0}{\sigma_{L,n}}) = \Psi_n \left( y_{2n} \left[ P_{Z_n} - P_{X_n} - \lambda_{LIML,n} M_{Z_n} \right] y_{2n} \right)^{-1} \times \left( y_{2n} \left[ P_{Z_n} - P_{X_n} - \lambda_{LIML,n} M_{Z_n} \right] u_n \right) \sigma_{L,n} \]

\[
= \frac{v_n' G_n u_n}{\sigma_{L,n}} - \frac{u_n' G_n u_n}{\sigma_{L,n}} + o_p(1),
\]

so that

\[
\left( \Psi_n \frac{\beta_{LIML,n} - \beta_0}{\sigma_{L,n}} \right) \overset{d}{\rightarrow} N(0, 1) \text{ as } n \rightarrow \infty,
\]
as required. □

**Proof of Theorem 3.2:**

By the usual regression algebra, we can write

\[
\hat{\beta}_{FLIML,n} - \beta_0 = \left( y_{2n} \left[ M_{X_n} - \hat{k}_{FLIML,n} M_{Z_n} \right] y_{2n} \right)^{-1} \times \left( y_{2n} \left[ M_{X_n} - \hat{k}_{FLIML,n} M_{Z_n} \right] u_n \right) \sigma_{L,n}
\]

\[
= \left( y_{2n} \left[ P_{Z_n} - P_{X_n} - \left( \lambda_{LIML,n} - 1 - \frac{a}{n - K_n - J} \right) M_{Z_n} \right] y_{2n} \right)^{-1} \times \left( y_{2n} \left[ P_{Z_n} - P_{X_n} - \left( \lambda_{LIML,n} - 1 - \frac{a}{n - K_n - J} \right) M_{Z_n} \right] u_n \right), \quad (82)
\]

where again the inverse in (82) exists in probability as \( n \rightarrow \infty \) in the sense of White (1984) given our assumptions, as will be shown in expression (84) below. Note that the second equality above follows from the fact that \( \hat{k}_{FLIML,n} = \lambda_{LIML,n} - \frac{a}{n - K_n - J} \) by definition. It follows from calculations similar to that used to derive expressions (79) and (81) above that

\[
y_{2n} \left[ P_{Z_n} - P_{X_n} - \left( \lambda_{LIML,n} - 1 - \frac{a}{n - K_n - J} \right) M_{Z_n} \right] u_n
\]

\[
= \left( \sqrt{K_n} \right) \left( \frac{c_n Z_n' M_{X_n} u_n}{b_n \sqrt{K_n}} + \frac{v_n' G_n u_n}{\sigma_{L,n}} \right) + \frac{u_n' G_n u_n}{\sigma_{L,n}} \left( \frac{v_n' M_{Z_n} u_n}{n - K_n - J} \right) + \frac{a}{\sigma_{L,n}} \left( \frac{v_n' M_{Z_n} u_n}{n - K_n - J} \right) + o_p(1)
\]

\[
= \frac{v_n' G_n u_n}{\sigma_{L,n}} - \frac{u_n' G_n u_n}{\sigma_{uu} \sigma_{L,n}} + o_p(1), \quad (83)
\]
and
\[
\frac{y'_{2n} \left[ P_{Z_n} - P_{X_n} - \left( \hat{\lambda}_{LIML,n} - 1 - \frac{a}{n-K_n+J} \right) M_{Z_n} \right] y_{2n}}{r_n} = \frac{c'_{n} Z'_{n} M_{X_{n}} Z_{n} c_{n}}{b_{n} r_{n}} + 2 \frac{c'_{n} Z'_{n} M_{X_{n}} v_{n}}{b_{n} r_{n}} + \frac{v'_{n} \left[ P_{Z_n} - P_{X_n} - \left( \hat{\lambda}_{LIML,n} - 1 \right) M_{Z_n} \right] v_{n}}{r_n} + \frac{a}{r_{n}} \left( \frac{v'_{n} M_{Z_n} v_{n}}{n-K_n-J} \right) \Psi_{n} = \Psi_{n} + o_{p}(1),
\]

where \( \Psi_{n} \) is nonsingular with probability one for \( n \) sufficiently large given Assumption 2. It follows immediately from (83) and (84) that
\[
\left( \frac{\Psi_{n}}{\sigma_{L,n}} \right) \left( \tilde{\beta}_{FLIML,n} - \beta_{0} \right) \overset{d}{\to} N(0,1) \text{ as } n \to \infty,
\]
as required. \( \square \)

Proof of Theorem 3.3:

To proceed, note first that, using the usual regression algebra, we can write
\[
\tilde{\beta}_{B2SLS,n} - \beta_{0} = \left( \frac{y'_{2n} \left[ P_{Z_n} - P_{X_n} - \left( \frac{K_n-2}{n-K_n+2} \right) M_{Z_n} \right] y_{2n}}{r_n} \right)^{-1} \times \left( \frac{y'_{2n} \left[ P_{Z_n} - P_{X_n} - \left( \frac{K_n-2}{n-K_n+2} \right) M_{Z_n} \right] u_{n}}{r_n} \right),
\]
so that
\[
\left( \frac{\Psi_{n}}{\sigma_{B,n}} \right) \left( \tilde{\beta}_{B2SLS,n} - \beta_{0} \right) = \left( \frac{\Psi_{n}}{r_{n}} \right) \left( \frac{y'_{2n} \left[ P_{Z_n} - P_{X_n} - \left( \frac{K_n-2}{n-K_n+2} \right) M_{Z_n} \right] y_{2n}}{r_n} \right)^{-1} \times \left( \frac{y'_{2n} \left[ P_{Z_n} - P_{X_n} - \left( \frac{K_n-2}{n-K_n+2} \right) M_{Z_n} \right] u_{n}}{\sigma_{B,n}} \right),
\]
where the inverse in (85) exists in probability as \( n \to \infty \) in the sense of White (1984) given our assumptions, as will be shown in expression (89) below. Next, note that
\[
\frac{K_n-2}{n-K_n+2} = \left( \frac{K_n}{n-K_n-j} \right) \left( \frac{n-K_n-j}{n-K_n+2} \right) \left( \frac{K_n-2}{K_n} \right) = \left( \frac{K_n}{n-K_n-j} \right) (1 + O(K_n^{-1})).
\]
To derive the limiting distribution of (85), we write

\[
\frac{y'_{2n} \left[ P_{Z_n} - P_{X_n} - \left( \frac{K_n - 2}{n-K_n+2} \right) M_{Z_n} \right] u_n}{\sigma_{B,n}} = \frac{\left( \sqrt{K_n} \right)}{\sigma_{B,n}} \frac{y'_{2n} \left[ P_{Z_n} - P_{X_n} - \left( \frac{K_n - 2}{n-K_n+2} \right) M_{Z_n} \right] u_n}{\sqrt{K_n}} \\
= \frac{v'_{2n} \left[ P_{Z_n} - P_{X_n} - \left( \frac{K_n - 2}{n-K_n+2} \right) M_{Z_n} \right] u_n}{\sigma_{B,n}} + O_{p}(1) \quad (87)
\]

where the fourth equality above follows from arguments similar to that given in part (e) of Lemma A1 of Chao and Swanson (2002), which can be used to show that 

\[
\left( \frac{1}{n-K_n-J} \right) \frac{v'_{n} M_{Z_n} u_n}{\sigma_{B,n}} = \left( \frac{\sqrt{K_n}}{\sigma_{B,n}} \right) \left( \frac{1}{\sqrt{K_n}} \right) \frac{v'_{n} M_{Z_n} u_n}{n - K_n - J} = O_{p} \left( K_{n}^{-\frac{1}{2}} \right).
\]

given Lemma A4 part (b). Note also that by setting \(d_1 = 1\) and \(d_2 = 0\) in the general formula (22), we deduce that

\[
\sigma_{B,n}^2 = \left[ E \left( u_j^2 v_j^2 \right) - \sigma_{uv}^2 \right] \sum_{j=1}^{n} E \left( g_{jj,n}^2 \right) + 2 \left( \sigma_{uu} \sigma_{vv} + \sigma_{uv}^2 \right) \sum_{1 \leq i < j \leq n} E \left( g_{ij,n}^2 \right).
\]

is the variance of the bilinear form \(v'_{n} G_{n} u_{n}\). It follows from Lemma A8 above that, as \(n \to \infty\),

\[
\frac{y'_{2n} \left[ P_{Z_n} - P_{X_n} - \left( \frac{K_n - 2}{n-K_n+2} \right) M_{Z_n} \right] u_n}{\sigma_{B,n}} = \frac{v'_{n} G_{n} u_{n}}{\sigma_{B,n}} + o_{p}(1) \xrightarrow{d} N(0, 1) \quad (88)
\]
where $G_n = P_{Z_n} - P_{X_n} - \left(\frac{K_n}{n-K_n-J}\right)M_{Z_n}$. Moreover, note that

$$y_n' \left[ P_{Z_n} - P_{X_n} - \left(\frac{K_n-2}{n-K_n+2}\right)M_{Z_n} \right] y_n = c_n' Z_n' M_{X_n} Z_n c_n + 2 c_n' Z_n' M_{X_n} v_n + v_n' \left[ P_{Z_n} - P_{X_n} - \left(\frac{K_n-2}{n-K_n+2}\right)M_{Z_n} \right] v_n$$

$$= c_n' Z_n' M_{X_n} Z_n c_n + 2 c_n' Z_n' M_{X_n} v_n + v_n' \left[ P_{Z_n} - P_{X_n} - \left(\frac{K_n}{n-K_n-J}\right)M_{Z_n} \right] v_n$$

$$= -2 \left(\frac{1}{n-K_n-J}\right) v_n' M_{Z_n} u_n + O_p\left(\frac{\sqrt{K_n}}{n}\right)$$

$$= \Psi_n + o_p(1) \quad (89)$$

where $\Psi_n = r^{-1}_n \Psi_n$ is nonsingular with probability one for $n$ sufficiently large given Assumption 2 and where the third equality follows from parts (a), (c), and (f) of Lemma A1 of Chao and Swanson (2002b) and from the fact that

$$\left(\frac{1}{n-K_n-J}\right) v_n' M_{Z_n} u_n = \left(\frac{\sqrt{K_n}}{\sigma_{B,n}}\right) \left(\frac{1}{\sqrt{K_n}}\right) v_n' M_{Z_n} v_n$$

$$= O_p\left(K_n^{-\frac{3}{2}}\right).$$

using arguments similar to that given to prove part (e) of Lemma A2 of Chao and Swanson (2002b). (88) and (89) imply that

$$\left(\frac{\Psi_n}{\sigma_{B,n}} \left(\hat{\beta}_{B2SLS,n} - \beta_0\right)\right) = \Psi_n \left(\frac{y_n' \left[ P_{Z_n} - P_{X_n} - \left(\frac{K_n-2}{n-K_n+2}\right)M_{Z_n} \right] y_n}{r_n}\right)^{-1}$$

$$\times \left(\frac{y_n' \left[ P_{Z_n} - P_{X_n} - \left(\frac{K_n-2}{n-K_n+2}\right)M_{Z_n} \right] u_n}{\sigma_{B,n}}\right)$$

$$= v_n' G_n u_n + o_p(1),$$

so that

$$\left(\frac{\Psi_n}{\sigma_{B,n}} \left(\hat{\beta}_{B2SLS,n} - \beta_0\right)\right) \overset{d}{\rightarrow} N(0,1) \quad \text{as } n \rightarrow \infty,$$

as required. \Box
Proof of Theorem 3.4:

Making use of expressions (12) and (13), we see that

\[
\sigma_{B,n}^2 - \sigma_{L,n}^2 = 2 \frac{\sigma_{uv}}{\sigma_{uu}} \left[ E \left( u_j^3 v_j \right) - \sigma_{uu} \sigma_{uv} \right] \sum_{j=1}^{n} E \left( g_{jj,n}^2 \right) + 4 \sigma_{uv}^2 \sum_{1 \leq i < j \leq n} E \left( g_{ij,n}^2 \right) \\
- \frac{\sigma_{uv}^2}{\sigma_{uu}^2} \left[ E \left( u_j^4 \right) - \sigma_{uu}^2 \right] \sum_{j=1}^{n} E \left( g_{jj,n}^2 \right). \\
\tag{90}
\]

Since by assumption \( \eta_i \) is \( E \left( 0, \Xi \right) \), we have, as a result of special properties of elliptical distributions, that

\[
E \left( u_j^4 \right) = 3 \left( \kappa + 1 \right) \sigma_{uu}^2, \\
E \left( u_j^3 v_j \right) = 3 \left( \kappa + 1 \right) \sigma_{uu} \sigma_{uv},
\]

where \( \kappa \) here denotes the kurtosis parameter of an elliptical distribution as defined in Muirhead (1982) page 41. It follows that we can rewrite (90) as

\[
\sigma_{B,n}^2 - \sigma_{L,n}^2 = (3\kappa + 2) \sigma_{uv}^2 \sum_{j=1}^{n} E \left( g_{jj,n}^2 \right) + 4 \sigma_{uv}^2 \sum_{1 \leq i < j \leq n} E \left( g_{ij,n}^2 \right).
\]

Moreover, Bentler and Berkane (1986) show that the kurtosis parameter \( \kappa \) for a \( m \)-variate continuous elliptical distribution with real positive definite covariance matrix \( \Sigma = \tau \Xi \) must be greater than \(-2 / (m + 2)\). Setting \( m = 2 \), we have that

\[
\sigma_{B,n}^2 - \sigma_{L,n}^2 = (3\kappa + 2) \sigma_{uv}^2 \sum_{j=1}^{n} E \left( g_{jj,n}^2 \right) + 4 \sigma_{uv}^2 \sum_{1 \leq i < j \leq n} E \left( g_{ij,n}^2 \right) \\
> \frac{1}{2} \sigma_{uv}^2 \sum_{j=1}^{n} E \left( g_{jj,n}^2 \right) + 4 \sigma_{uv}^2 \sum_{1 \leq i < j \leq n} E \left( g_{ij,n}^2 \right) \\
> 0,
\]

as required. □
6 References


