Instrumental variables estimation and inference in the presence of many exogenous regressors

Stanislav Anatolyev[*]

New Economic School

March 2010
This version: March 2012

Abstract

We consider a standard instrumental variables model contaminated by the presence of a large number of exogenous regressors. In an asymptotic framework where this number is proportional to the sample size, we study the impact of their ratio on the validity of existing estimators and tests. When the instruments are few, the inference using the conventional 2SLS estimator and associated t and J statistics, as well as the Anderson–Rubin and Kleibergen tests, is still valid. When the instruments are many, the LIML estimator remains consistent, but the presence of many exogenous regressors changes its asymptotic variance. Moreover, the conventional bias correction of the 2SLS estimator is no longer appropriate. We provide asymptotically correct versions of bias correction for the 2SLS estimator, derive its asymptotically correct variance estimator, extend the Hansen–Hausman–Newey LIML variance estimator to the case of many exogenous regressors, and propose asymptotically valid modifications of the J overidentification tests based on the LIML and bias corrected 2SLS estimators.

Keywords: instrumental variables regression, many instruments, many exogenous regressors, 2SLS estimator, LIML estimator, bias correction, t test, J test, Anderson–Rubin test, Kleibergen test.

JEL codes: C12, C21

---

[*]Address: New Economic School, Nakhimovsky Prospekt, 47, room 1721, Moscow, 117418, Russia; sanatoly@nes.ru; http://www.nes.ru/~sanatoly/

1 Introduction

Often an instrumental variables regression contains many exogenous regressors. These variables are not of interest, but are included primarily as a guard against model misspecification and omitted variables bias, in order to approximate as closely as possible an important but unobservable factor by throwing in many observable characteristics, or in order to capture possible non-linearities in a semi-nonparametric manner. As a result, if the sample is not too large, the proportion of such regressors to the number of observations may be perceptible. This may adversely affect the inference because a part of information in the sample has to be spent on estimation of a large number of nuisance parameters. Hahn and Hausman (2002, Section 6) in their illustrative application to a demand function compute a number of estimators and run a number of tests where they include 134 predetermined variables, the sample size being 1459. The coefficients of these variables are of no interest, and the authors even do not mention what they are, while of interest is the price elasticity only. The concern that the ratio of the number of exogenous variables to the sample size may be non-negligible is relevant for estimation of a demand function for various products (e.g., for water, see Billings, 1987, or for electricity, see Kamerschen and Porter, 2004), estimation of the hedonic price model (e.g., see Witte, Sumka and Erekson, 1979), or estimation of the household production function (e.g., see Rosenzweig and Schultz, 1983), to name a few.

In this paper we analyze the impact of a large number of exogenous regressors on existing estimators and tests for a linear model estimated by instrumental variables methods. For a standard instrumental variables regression with few or many instruments we consider an asymptotic framework where the number of exogenous regressors and possibly instruments is proportional to the sample size. The literature on estimation and inference in the presence of many (possibly weak) instruments is rapidly growing (see, among others, Chao and Swanson, 2005, 2006; Hansen, Hausman and Newey, 2008; van Hasselt, 2010; Anderson, Kunimoto and Matsushita, 2010; Hausman, Newey, Woutersen, Chao, and Swanson, 2010; Anatolyev and Gospodinov, 2011; Lee and Okui, 2012; Chao, Swanson, Hausman, Newey, and Woutersen, 2012), and the present paper aims to contribute to this literature. Among the estimators we consider are conventional 2SLS, bias corrected 2SLS, LIML and K-class estimators. Among the tests we consider are conventional t and J tests, as well as those recently proposed in the context of weak or many instruments, such as Anderson–Rubin (Anderson and Rubin, 1949) and Kleibergen (Kleibergen, 2002) tests for parameter restrictions, and the Wald test with variance estimates of Hansen, Hausman and Newey (2008). In those cases when the presence of many exogenous regressors invalidates the estimators or tests under consideration, we propose their modified versions. We do not assume error normality; all modifications are constructed in the general case

of possibly non-normal model errors. Some of our results are new in the literature even for the special case when the exogenous regressors are not many.

A brief preview of results follows. When the instruments are few, the inference using the conventional 2SLS estimator and associated t and J statistics, as well as the Anderson–Rubin and Kleibergen tests, is still valid. When the instruments are many, the LIML estimator remains consistent, but the presence of many exogenous regressors changes its asymptotic variance. Moreover, the conventional bias correction of the 2SLS estimator is no longer appropriate. We provide asymptotically correct versions of bias correction for the 2SLS estimator, derive its asymptotically correct variance estimator, extend the Hansen–Hausman–Newey LIML variance estimator to the case of many exogenous regressors, and propose asymptotically valid modifications of the J overidentification tests based on the LIML and bias corrected 2SLS estimators. A small Monte-Carlo experiment shows good performance of proposed modifications in moderately sized samples in a model with non-normal errors.

The paper is structured as follows. Section 2 describes the model and states the assumptions. Section 3 analyzes behavior of conventional estimators and tests when exogenous regressors are many but instruments are few. Sections 4 and 5 discuss the bias corrected 2SLS and LIML estimators as well as their variance estimators and associated tests when instruments and exogenous regressors are numerous. Section 6 presents simulation results in finite samples. Some useful computational remarks are contained in Section 7. Finally, Section 8 concludes. All proofs are relegated to the Appendix.

2 The setup

2.1 Model

We are interested in estimating and testing the structural equation with p endogenous and m included exogenous regressors:

$$Y = X\beta_0 + W\delta_0 + e,$$

where $Y = (y_1, ..., y_n)'$ is $n \times 1$, $X = (x_1, ..., x_n)'$ is $n \times p$, $W = (w_1, ..., w_n)'$ is $n \times m$, and $e = (e_1, ..., e_n)'$ is $n \times 1$. The object of primary interest is the vector of structural parameters β_0, while the vector δ_0 contains nuisance parameters. There is additionally an $n \times \ell$ matrix of instruments ('excluded exogenous variables') $Z = (z_1, ..., z_n)'$, $\ell \geq p$. For convenience, the data on instruments Z and exogenous regressors W will be treated as nonrandom. Because the column dimensions of Z and W will grow with sample size n, their elements implicitly depend on n.

Let the reduced form be

$$X = \Upsilon_{XZ} + \Upsilon_{XW} + U,$$

where $\Upsilon_{XZ} + \Upsilon_{XW} = Z\Pi_1 + W\Pi_2$ is the linear projection on the space of (Z, W), and $U = (u_1, ..., u_n)'$ is $n \times p$. We assume throughout that (Z, W) has full column rank: $\mathrm{rk}\,(Z, W) = \ell + m$. This excludes redundant columns from data on instruments and exogenous regressors. It also means that $\ell + m$ does not exceed n.

2.2   Notation and assumptions

Let us denote by P_W and M_W the orthogonal projection matrices associated with W

$$P_W = W\,(W'W)^{-1}\,W', \quad M_W = I_n - P_W,$$

by P_{ZW} and M_{ZW} the projection matrices associated with (Z, W), and by $P_{Z^\perp}$ and $M_{Z^\perp}$ the projection matrices associated with $Z^\perp = M_W Z$. Note that $P_{Z^\perp} = P_{ZW} - P_W$ (see Lemma A1 in the Appendix). Similarly, denote $X^\perp = M_W X$ and $Y^\perp = M_W Y$. Additionally, let us introduce the following notation to be used throughout:

$$P_\zeta = P_{Z^\perp} - \zeta M_W$$

for any real number ζ.

Lower index i will point at the row number of a matrix, e.g. $(M_W \Upsilon_{XZ})_i$ is the transposed i^{th} row of $M_W \Upsilon_{XZ}$. Upper indexes will denote corresponding elements of a matrix, e.g. P_W^{ij} is the $(i, j)^{th}$ element of P_W. Next, a bar will denote taking an average over the index present, e.g. $\overline{P_W^{ii}}$ is n^{-1} times the trace of P_W and $\overline{\sum_j P_W^{ij}}$ is n^{-1} times the sum of all elements of P_W. Also, let n-vector d_A contain diagonal elements of an $n \times n$ matrix A.

We adapt the following asymptotic framework.

Assumption 1 *Asymptotically, as $n \to \infty$, $m/n = \mu + o\,(1/\sqrt{n})$ with $0 < \mu < 1$, and either ℓ is fixed, or $\ell/n = \lambda + o\,(1/\sqrt{n})$ with $0 \le \lambda < 1 - \mu$.*

Assumption 1 is reminiscent of the classical many instruments asymptotic framework of Bekker (1994). It is critical that the number of exogenous regressors and possibly instruments grows proportionately with the sample size rather than slower than that. We associate the word 'many' with such proportional growth, the qualifier 'moderately many' with a slower growth, and the word 'few' with a fixed number. We exclude the case of few or moderately many exogenous regressors $\mu = 0$ here, but some classical textbook results can be considered as limiting cases of ours when μ is put to zero. Let us also introduce a fundamental quantity

$$\alpha = \frac{\lambda}{1 - \mu},$$

which is (asymptotically) a number of instruments per degrees of freedom. Note that $0 \leq \alpha < 1$ given the restrictions on λ and μ.

Next we make assumptions about data generation.

Assumption 2 *The errors (e_i, u_i) are zero mean IID across i having finite eighth moments, with $E\left[e_i^2\right] = \sigma_e^2$, $E\left[u_i u_i'\right] = \Sigma_u$ and $E\left[u_i e_i\right] = \Sigma_{ue}$.*

Henceforth, by 'lim' we understand taking a limit under Assumption 1.

Assumption 3 *When ℓ is fixed, $\lim n^{-1}(Z^\perp)'Z^\perp = Q_{Z^\perp}$, where $Q_{Z^\perp}$ is finite and positive definite, Π_1 is of full column rank p, and $\sum_{i=1}^n \left\|Z_i^\perp\right\|^4 \leq o\left(n^2\right)$. When $\ell \to \infty$, $\lim n^{-1}\Upsilon_{XZ}' M_W \Upsilon_{XZ} = Q_{XZ^\perp}$, where $Q_{XZ^\perp}$ is finite and positive definite, $\sum_{i=1}^n \left\|(M_W \Upsilon_{XZ})_i\right\|^4 \leq o\left(n^2\right)$, and the limits $\rho_\alpha = \lim \overline{\left(P_\alpha^{ii}\right)^2}$ and $\pi_\alpha = \lim \overline{P_\alpha^{ii} \left(M_W \Upsilon_{XZ}\right)_i}$ exist and are finite.*

Assumption 3 means in particular that the excluded instruments are, as a group, strong after controlling for the explanatory power provided by the exogenous regressors. Other requirements in Assumption 3 are technical, they are useful for various large sample results to go through, and also are helpful in constructing estimators of various moments.

Note that if there are no exogenous regressors ($m = 0$ and $\alpha = \lambda$), then $P_\alpha = P_Z - \lambda I_n$, $\overline{\left(P_\alpha^{ii}\right)^2} = \overline{\left(P_Z^{ii}\right)^2} - \lambda^2 + o(1)$ and $\overline{P_\alpha^{ii}\left(M_W \Upsilon_{XZ}\right)_i} = \overline{\left(P_Z^{ii} - \lambda\right)\left(\Upsilon_{XZ}\right)_i}$, hence $\rho_\alpha = \lim \overline{\left(P_Z^{ii}\right)^2} - \lambda^2$ and $\pi_\alpha = \lim \overline{\left(P_Z^{ii} - \lambda\right)\left(\Upsilon_{XZ}\right)_i}$.

Let also
$$\hat{\lambda} = \frac{\ell}{n}, \quad \hat{\mu} = \frac{m}{n}, \quad \hat{\alpha} = \frac{\ell}{n - m}$$
be finite sample analogs of λ, μ and α. Also, put
$$\hat{\rho}_\alpha = \overline{\left(P_{\hat{\alpha}}^{ii}\right)^2}, \quad \hat{\pi}_\alpha = \overline{P_{\hat{\alpha}}^{ii}\left(P_{Z^\perp} X\right)_i}$$
to be estimates of ρ_α and π_α.

Denote also for future use
$$\Gamma \equiv \frac{\Sigma_{ue}'}{\sigma_e^2},$$
the coefficients in a linear projection of reduced form errors on structural errors, and let
$$\tilde{u}_i = u_i - \Gamma' e_i$$
be corresponding population least squares residuals, or in a matrix form,
$$\tilde{U} = U - e\Gamma.$$

3 Estimators and tests under few instruments

3.1 2SLS estimation

In this subsection we consider the standard 2SLS estimator and associated J test. The 2SLS estimator for the parameter of interest is

$$\hat{\beta}_{2SLS} = (I_p, 0_{p \times m}) \left((X, W)' P_{ZW} (X, W) \right)^{-1} (X, W)' P_{ZW} Y.$$

By the partitioned matrix inverse,

$$\hat{\beta}_{2SLS} = (X' P_{Z^{\perp}} X)^{-1} X' P_{Z^{\perp}} Y.$$

Note that the nuisance parameters can be computed as

$$\hat{\delta}_{2SLS} = (W'W)^{-1} W'(Y - X \hat{\beta}_{2SLS}),$$

so the 2SLS residuals are then may be computed directly from $\hat{\beta}_{2SLS}$:

$$\hat{e}_{2SLS} = Y^{\perp} - X^{\perp} \hat{\beta}_{2SLS}.$$

Define the residual variance

$$\hat{\sigma}^2_{2SLS} = \frac{\hat{e}'_{2SLS} \hat{e}_{2SLS}}{n - m - p}.$$

The t statistic for the jth component of β_0 based on 2SLS estimation is

$$
\begin{aligned}
t^{(j)}_{2SLS} &= \frac{(\hat{\beta}_{2SLS})_j - (\beta_0)_j}{\sqrt{\hat{\sigma}^2_{2SLS} \left[\left((X, W)' P_{ZW} (X, W) \right)^{-1} \right]^{jj}}} \\
&= \frac{(\hat{\beta}_{2SLS})_j - (\beta_0)_j}{\sqrt{\hat{\sigma}^2_{2SLS} \left[(X' P_{Z^{\perp}} X)^{-1} \right]^{jj}}}.
\end{aligned}
$$

Further, the standard J test for overidentifying restrictions is given by

$$J_{2SLS} = \frac{\hat{e}'_{2SLS} P_{ZW} \hat{e}_{2SLS}}{\hat{\sigma}^2_{2SLS}}.$$

A classical textbook result is that when there are few exogenous regressors (so that $\mu = 0$) and few instruments (so that $\lambda = 0$), the 2SLS estimator is consistent and asymptotically normal, and under the null of correct moment restrictions J_{2SLS} is asymptotically distributed as $\chi^2 (\ell - p)$. When the instruments are many (so that $\lambda > 0$), the 2SLS estimator is inconsistent (Bekker, 1994; Newey, 2004). The following Proposition establishes the asymptotic properties of 2SLS in the presence of many exogenous regressors (so that $\mu > 0$) when the instruments are few or many.

Proposition 1: Suppose assumptions 1–3 hold. Then

$$\hat{\beta}_{2SLS} \xrightarrow{p} \beta_0 + (Q_{XZ^\perp} + \lambda \Sigma_u)^{-1} \lambda \Sigma_{ue}$$

and

$$\frac{J_{2SLS}}{n} \xrightarrow{p} \alpha J^\infty,$$

where the formula for the constant J^∞ can be found in the Appendix.

Thus, the 2SLS estimator is consistent only if $\lambda = 0$, i.e. when there are few instruments. The J statistic has a random limit only when $\lambda = 0$, i.e. when there are few instruments and the 2SLS estimator is consistent. In the rest of this Section we focus on the case of few instruments.

Theorem 1: Suppose assumptions 1–3 hold, and ℓ is fixed. Then

$$\sqrt{n} \left(\hat{\beta}_{2SLS} - \beta_0 \right) \xrightarrow{d} N \left(0, \sigma_e^2 \left(\Pi_1' Q_{Z^\perp} \Pi_1 \right)^{-1} \right),$$

$$t_{2SLS}^{(j)} \xrightarrow{d} N(0, 1), \quad j = 1, ..., p,$$

and

$$J_{2SLS} \xrightarrow{d} \chi^2 (\ell - p).$$

Thus, the presence of many exogenous regressors does not affect the form of the asymptotic variance of the 2SLS estimator. In the case when the exogenous regressors are all orthogonal to the excluded instruments, the asymptotic variance does not depend on their number (assuming that the unexplained variance of the dependent variable does not either). However, if the excluded instruments get better and better explained by the exogenous regressors as the number of those grows, the effect of numerosity of regressors is to increase the asymptotic variance. The conventional standard errors, however, take account of these variations in the asymptotic variance. The conventional J test is also asymptotically valid.

Note that in the special case of a classical linear regression when $Z = X$ the asymptotic validity of t or Wald tests is consistent with results of Anatolyev (2012) who establishes, although under more restrictive assumptions, that the classical tests are valid in the many regressors framework when the number of restrictions is asymptotically fixed.

3.2 Anderson–Rubin and Kleibergen tests

In this subsection we consider the celebrated Anderson–Rubin (AR) and Kleibergen (K) tests for testing parameter restrictions. In the usual circumstances these tests are robust to the quality of instruments, and are correctly sized when the instruments are strong, weak or irrelevant. We study the influence of the numerosity of exogenous regressors on the asymptotics of the corresponding test statistics. We still focus on the case of fixed ℓ.

Consider the null hypothesis $H_0 : \beta = \beta_0$. The Anderson–Rubin AR statistic (Anderson and Rubin, 1949) is

$$AR = \frac{n - \ell - m}{\ell} \frac{\left(Y^\perp - X^\perp \beta_0\right)' P_{Z^\perp} \left(Y^\perp - X^\perp \beta_0\right)}{\left(Y^\perp - X^\perp \beta_0\right)' M_{Z^\perp} \left(Y^\perp - X^\perp \beta_0\right)},$$

and is asymptotically distributed as $\chi^2 (\ell) / \ell$ under the null. The Kleibergen K statistic (Kleibergen, 2002) is

$$K = (n - \ell - m) \frac{\left(Y^\perp - X^\perp \beta_0\right)' P_{\tilde{Z}} \left(Y^\perp - X^\perp \beta_0\right)}{\left(Y^\perp - X^\perp \beta_0\right)' M_{Z^\perp} \left(Y^\perp - X^\perp \beta_0\right)},$$

where

$$\tilde{Z} = P_{Z^\perp} \left(X^\perp - \left(Y^\perp - X^\perp \beta_0\right) \frac{\left(Y^\perp - X^\perp \beta_0\right)' M_{Z^\perp} X^\perp}{\left(Y^\perp - X^\perp \beta_0\right)' M_{Z^\perp} \left(Y^\perp - X^\perp \beta_0\right)} \right),$$

and is asymptotically distributed as $\chi^2 (p)$ under the null regardless of the strength of the instruments. The difference with the AR statistic is that instead of projecting $Y^\perp - X^\perp \beta_0$ onto the ℓ columns of matrix Z, the K statistic projects onto the vector $\tilde{Z}$ whose column dimension is given by the number of endogenous variables. This reduces the number of degrees of freedom and enhances the power properties (Kleibergen, 2002).

It turns out that both tests are robust to the presence of many exogenous regressors.

Theorem 2: Suppose assumptions 1–3 hold, and ℓ is fixed. Then

$$AR \xrightarrow{d} \frac{\chi^2 (\ell)}{\ell}$$

and

$$K \xrightarrow{d} \chi^2 (p).$$

The conclusion also holds if the assumption of fixed Π_1 is replaced by $\Pi_1 = \mathring{\Pi}/\sqrt{n}$, where $\mathring{\Pi}$ is fixed.

4 Bias-corrected 2SLS estimation

4.1 Construction

Consider the existing bias corrected versions of the 2SLS estimator:

$$\breve{\beta}_{B2SLS} = (I_p, 0_{p \times m}) \left((X, W)' (P_{ZW} - \hat{\xi} I_n) (X, W) \right)^{-1} (X, W)' (P_{ZW} - \hat{\xi} I_n) Y,$$

where either

$$\hat{\xi} = \frac{\ell + m}{n}$$

where the total number of instruments is counted (e.g., Newey, 2004; van Hasselt, 2010), or

$$\hat{\xi} = \frac{\ell - 2}{n}$$

where the number of additional instruments only is counted (e.g., Donald and Newey, 2001; Hahn and Hausman, 2002). By the partitioned matrix inverse,

$$\check{\beta}_{B2SLS} = (X'P_{\hat{\xi}}X)^{-1}X'P_{\hat{\xi}}Y,$$

These estimators are inconsistent because they do not account for many exogenous regressors as the following proposition shows.

Proposition 2: Suppose assumptions 1–3 hold, and $\lambda > 0$. Then the asymptotic biases of the bias corrected 2SLS estimators $\check{\beta}_{B2SLS}$ are

$$-\mu \left(Q_{XZ^{\perp}} - \mu\Sigma_u\right)^{-1}\Sigma_{ue}$$

and

$$\lambda\mu \left((1-\lambda)Q_{XZ^{\perp}} + \lambda\mu\Sigma_u\right)^{-1}\Sigma_{ue}$$

for the two choices of $\hat{\xi}$.

One can easily see that the conventional bias correction works only if either there is no endogeneity ($\Sigma_{ue} = 0$) and there is no need to correct for bias, or the exogenous regressors are not many ($\mu = 0$).

Let us instead make the following correction to the 2SLS estimator:

$$\begin{aligned}
\hat{\beta}_{B2SLS} &= (I_p, 0_{p\times m})\left((X,W)'\left(P_{ZW} - \hat{\alpha}I_n\right)(X,W)\right)^{-1}(X,W)'\left(P_{ZW} - \hat{\alpha}I_n\right)Y \\
&= (X'P_{\hat{\alpha}}X)^{-1}X'P_{\hat{\alpha}}Y.
\end{aligned}$$

Note that this is similar to the standard bias correction for 2SLS above, but the factor $\hat{\alpha}$ is different from either conventional factor $\hat{\xi}$ which asymptotically is equivalent to either $\hat{\lambda} + \hat{\mu}$ or $\hat{\lambda}$. For the bias-corrected 2SLS to be consistent, that factor should be adjusted for the numerosity of exogenous regressors in a proper way. If there are no exogenous regressors ($m = 0$ and $\alpha = \lambda$), then $\hat{\mu} = 0$, $\hat{\alpha} = \hat{\lambda}$, $P_{\hat{\alpha}} = P_Z - \hat{\lambda}I$, and the estimator is equivalent to either version of the conventional bias corrected 2SLS.

The estimator $\hat{\beta}_{B2SLS}$ allows for a variety of interpretations. For example, it is an instrumental variables estimator in the regression of Y on X only, using the instrument $\hat{Z}_1 = P_{\hat{\alpha}}X$. Or, it is an instrumental variables estimator in the regression of $Y^{\perp}$ on $X^{\perp}$ only, using the instrument $\hat{Z}_2 = P_{\hat{\alpha}}X^{\perp}$. Note also that, similarly to the case of 2SLS, $\hat{\delta}_{B2SLS}$ may be concentrated out:

$$\hat{\delta}_{B2SLS} = (W'W)^{-1}W'(Y - X\hat{\beta}_{B2SLS}).$$

As a result, the B2SLS residuals then may be computed directly from $\hat{\beta}_{B2SLS}$:

$$\hat{e}_{B2SLS} = Y^\perp - X^\perp \hat{\beta}_{B2SLS}.$$

Define the residual variance

$$\hat{\sigma}^2_{B2SLS} = \frac{\hat{e}'_{B2SLS}\hat{e}_{B2SLS}}{n - m - p}.$$

Note the degrees of freedom adjustment.

4.2   Asymptotic properties

The following theorem establishes the asymptotic properties of the bias-corrected 2SLS in the presence of many exogenous regressors.

Theorem 3: Suppose assumptions 1–3 hold, and $\lambda > 0$. Then the bias-corrected 2SLS estimator $\hat{\beta}_{B2SLS}$ is consistent and zero mean asymptotically normal with the asymptotic variance

$$V_{B2SLS} = \Phi_2^{B2SLS} + \Phi_3^{B2SLS} + \Phi_4^{B2SLS},$$

where

$$\Phi_2^{B2SLS} = \sigma_e^2 Q_{XZ^\perp}^{-1} + \frac{\lambda}{1-\alpha} Q_{XZ^\perp}^{-1} \left(\sigma_e^2 \Sigma_u + \Sigma_{ue}\Sigma'_{ue} \right) Q_{XZ^\perp}^{-1},$$

$$\Phi_3^{B2SLS} = \frac{1}{1-\alpha} Q_{XZ^\perp}^{-1} \left(\pi_\alpha E\left[e_i^2 u_i' \right] + E\left[e_i^2 u_i \right] \pi'_\alpha \right) Q_{XZ^\perp}^{-1},$$

and

$$\Phi_4^{B2SLS} = \frac{\rho_\alpha}{(1-\alpha)^2} Q_{XZ^\perp}^{-1} \left(E\left[e_i^2 u_i u_i' \right] - \sigma_e^2 \Sigma_u - 2\Sigma_{ue}\Sigma'_{ue} \right) Q_{XZ^\perp}^{-1}.$$

The composition of the asymptotic variance is the same as that in van Hasselt (2010): one term Φ_2^{B2SLS} is present even under error normality, and two terms Φ_3^{B2SLS} and Φ_4^{B2SLS} are responsible for possible deviations of third and fourth moments from their values under normality. Moreover, the presence of many exogenous regressors is reflected only in changes in the scalar factors, otherwise the forms of the components are the same.

Under error normality when $\Phi_3^{B2SLS} = \Phi_4^{B2SLS} = 0$, the presence of numerous exogenous regressors is reflected, apart from possible reductions in $Q_{XZ^\perp}$, in the factor $\lambda/(1-\alpha)$ instead of $\lambda/(1-\lambda)$ scaling up the unconventional term in Φ_2^{B2SLS} contributed by the numerosity of instruments. In the case of error non-normality, the terms Φ_3^{B2SLS} and Φ_4^{B2SLS} are inflated because of numerous exogenous regressors. Of course, these effects may be partially offset or further exacerbated by changes in π_α and ρ_α.

4.3 Variance estimation

To estimate V_{B2SLS}, we need to construct consistent estimates of its three components, Φ_2^{B2SLS}, Φ_3^{B2SLS} and Φ_4^{B2SLS}; under error normality, consistent estimation of only Φ_2^{B2SLS} suffices. Its estimate is straightforward to construct:

$$\hat{\Phi}_2^{B2SLS} = (1 - \hat{\alpha}) \, n \hat{\sigma}_{B2SLS}^2 \left(X' P_{\hat{\alpha}} X\right)^{-1} + \frac{\hat{\alpha}}{1 - \hat{\mu} - \hat{\lambda}} \left(X' P_{\hat{\alpha}} X\right)^{-1} \hat{D}_2 \left(X' P_{\hat{\alpha}} X\right)^{-1},$$

where

$$\hat{D}_2 = \left(\hat{e}'_{B2SLS} M_{ZW} \hat{e}_{B2SLS}\right) \left(X' M_{ZW} X\right) + \left(X' M_{ZW} \hat{e}_{B2SLS}\right) \left(\hat{e}'_{B2SLS} M_{ZW} X\right).$$

For estimation of Φ_3^{B2SLS} and Φ_4^{B2SLS} it is necessary to construct various third and fourth cross-moments of e_i and u_i using the B2SLS residuals and regressors. This turns out to be not a straightforward task. Introduce

$$\tilde{X} = X - \hat{e}_{B2SLS} \hat{\Gamma}_{B2SLS},$$

where

$$\hat{\Gamma}_{B2SLS} = \frac{\hat{e}'_{B2SLS} X}{\hat{e}'_{B2SLS} \hat{e}_{B2SLS}}$$

estimates Γ. Then form an estimate of Φ_3^{B2SLS} as

$$\hat{\Phi}_3^{B2SLS} = n^2 \left(1 - \hat{\alpha}\right) \left(X' P_{\hat{\alpha}} X\right)^{-1} \left(\hat{D}_3 \hat{\pi}'_\alpha + \hat{\pi}_\alpha \hat{D}'_3\right) \left(X' P_{\hat{\alpha}} X\right)^{-1},$$

where

$$\hat{D}_3 = \frac{\overline{(\hat{e}_{B2SLS})_i^3}}{\sum_j \left(M_W^{ij}\right)^3} \hat{\Gamma}'_{B2SLS} + \frac{\overline{(\hat{e}_{B2SLS})_i^2 (M_{ZW} \tilde{X})_i}}{\sum_j \left(M_W^{ij}\right)^2 M_{ZW}^{ij}}.$$

Finally, form an estimate of Φ_4^{B2SLS} as

$$\hat{\Phi}_4^{B2SLS} = n^2 \hat{\rho}_\alpha \left(X' P_{\hat{\alpha}} X\right)^{-1} \hat{D}_4 \left(X' P_{\hat{\alpha}} X\right)^{-1},$$

where

$$\begin{aligned}
\hat{D}_4 &= \frac{\overline{(\hat{e}_{B2SLS})_i^4} - 3 \overline{(M_W^{ii})^2} \hat{\sigma}_{B2SLS}^4}{\overline{\sum_j \left(M_W^{ij}\right)^4}} \hat{\Gamma}'_{B2SLS} \hat{\Gamma}_{B2SLS} \\
&+ \hat{\Gamma}'_{B2SLS} \frac{\overline{(\hat{e}_{B2SLS})_i^2 (M_{ZW} \tilde{X})'_i}}{\sum_j \left(M_W^{ij}\right)^2 M_{ZW}^{ij}} + \frac{\overline{(\hat{e}_{B2SLS})_i^2 (M_{ZW} \tilde{X})_i}}{\sum_j \left(M_W^{ij}\right)^2 M_{ZW}^{ij}} \hat{\Gamma}_{B2SLS} \\
&+ \frac{\overline{\left((\hat{e}_{B2SLS})_i^2 - \hat{\phi} \hat{\sigma}_{B2SLS}^2\right) (M_{ZW} \tilde{X})_i (M_{ZW} \tilde{X})'_i}}{\overline{\sum_j \left(M_W^{ij} M_{ZW}^{ij}\right)^2}}
\end{aligned}$$

and

$$\hat{\phi} = \frac{\overline{M_W^{ii} M_{ZW}^{ii}}}{1 - \hat{\mu} - \hat{\lambda}}.$$

The variance estimator is computed as

$$\hat{V}_{B2SLS} = \hat{\Phi}_2^{B2SLS} + \hat{\Phi}_3^{B2SLS} + \hat{\Phi}_4^{B2SLS}.$$

The following theorem shows the asymptotic validity of our construct in the case of many exogenous regressors.

Theorem 4: Suppose assumptions 1–3 hold, and $\lambda > 0$. Then

$$\hat{V}_{B2SLS} \xrightarrow{p} V_{B2SLS}.$$

Thus, the standard errors obtained with the use of $\hat{V}_{B2SLS}$ will be asymptotically valid, and hypothesis testing on its basis will be asymptotically correct. Of course, the variance estimator $\hat{V}_{B2SLS}$ is robust to the numerosity of exogenous regressors and may be used when their number is large, moderately large, small, or zero.

Consider the limiting case when there are no exogenous regressors ($m = 0$ and $\alpha = \lambda$). Then $\hat{\mu} = 0$, $\hat{\alpha} = \hat{\lambda}$, $P_{\hat{\alpha}} = P_Z - \hat{\lambda} I_n$, $\hat{\rho}_\alpha = \overline{(P_Z^{ii})^2} - \hat{\lambda}^2$, $\overline{\sum_j (M_W^{ij})^2 M_{ZW}^{ij}} = \overline{M_Z^{ii}} = 1 - \hat{\lambda}$, $\overline{\sum_j (M_W^{ij})^3} = 1$, $\overline{\sum_j (M_W^{ij})^4} = 1$, $\overline{(M_W^{ii})^2} = 1$, $\overline{\sum_j (M_W^{ij} M_{ZW}^{ij})^2} = \overline{(M_Z^{ii})^2} = 1 - 2\hat{\lambda} + \overline{(P_Z^{ii})^2}$, $\overline{M_W^{ii} M_{ZW}^{ii}} = \overline{M_Z^{ii}} = 1 - \hat{\lambda}$ so that $\hat{\phi} = 1$, and the three components of the variance estimator can be simplified to

$$
\begin{aligned}
\hat{\Phi}_2^{B2SLS} &= (1 - \hat{\lambda}) (\hat{e}'_{B2SLS} \hat{e}_{B2SLS}) \left(X'(P_Z - \hat{\lambda} I_n) X \right)^{-1} \\
&\quad + \frac{\hat{\lambda}}{1 - \hat{\lambda}} \left(X'(P_Z - \hat{\lambda} I_n) X \right)^{-1} \hat{D}_2 \left(X'(P_Z - \hat{\lambda} I_n) X \right)^{-1},
\end{aligned}
$$

$$\hat{\Phi}_3^{B2SLS} = n^2 (1 - \hat{\lambda}) \left(X'(P_Z - \hat{\lambda} I_n) X \right)^{-1} \left(\hat{D}_3 \hat{\pi}'_\alpha + \hat{\pi}_\alpha \hat{D}'_3 \right) \left(X'(P_Z - \hat{\lambda} I_n) X \right)^{-1},$$

$$\hat{\Phi}_4^{B2SLS} = n^2 \left(\overline{(P_Z^{ii})^2} - \hat{\lambda}^2 \right) \left(X'(P_Z - \hat{\lambda} I_n) X \right)^{-1} \hat{D}_4 \left(X'(P_Z - \hat{\lambda} I_n) X \right)^{-1},$$

where

$$\hat{D}_2 = (\hat{e}'_{B2SLS} M_Z \hat{e}_{B2SLS}) (X' M_Z X) + (X' M_Z \hat{e}_{B2SLS}) (\hat{e}'_{B2SLS} M_Z X),$$

$$\hat{D}_3 = \overline{(\hat{e}_{B2SLS})_i^3} \hat{\Gamma}'_{B2SLS} + \frac{\overline{(\hat{e}_{B2SLS})_i^2 (M_Z \tilde{X})_i}}{1 - \hat{\lambda}},$$

and

$$
\begin{aligned}
\hat{D}_4 &= \left(\overline{(\hat{e}_{B2SLS})_i^4} - 3\hat{\sigma}^4_{B2SLS} \right) \hat{\Gamma}'_{B2SLS} \hat{\Gamma}_{B2SLS} \\
&\quad + \frac{\hat{\Gamma}'_{B2SLS} \overline{(\hat{e}_{B2SLS})_i^2 (M_Z \tilde{X})'_i} + \overline{(\hat{e}_{B2SLS})_i^2 (M_Z \tilde{X})_i} \hat{\Gamma}_{B2SLS}}{1 - \hat{\lambda}} \\
&\quad + \frac{\overline{\left((\hat{e}_{B2SLS})_i^2 - \hat{\sigma}^2_{B2SLS} \right) (M_Z \tilde{X})_i (M_Z \tilde{X})'_i}}{1 - 2\hat{\lambda} + \overline{(P_Z^{ii})^2}}.
\end{aligned}
$$

This asymptotic variance estimator in the special case of no exogenous regressors is new to the literature on many instruments.

4.4 Specification testing

Now consider a modified J statistic, a suitable quadratic form in the B2SLS residuals:

$$J_{B2SLS} = \frac{\hat{e}'_{B2SLS}\left(P_{ZW} - \hat{\alpha}I_n\right)\hat{e}_{B2SLS}}{\hat{\sigma}^2_{B2SLS}} = \frac{\hat{e}'_{B2SLS}P_{\hat{\alpha}}\hat{e}_{B2SLS}}{\hat{\sigma}^2_{B2SLS}}.$$

This is an adaptation of the Sargan type statistic of Lee and Okui (2012) to the case of many exogenous regressors.[1]

Theorem 5: Suppose assumptions 1–3 hold, and $\lambda > 0$. Then

$$\frac{J_{B2SLS}}{\sqrt{n}} \xrightarrow{d} N\left(0, V^J_{B2SLS}\right),$$

where the asymptotic variance is

$$V^J_{B2SLS} = 2\lambda\left(1 - \alpha\right) + \rho_\alpha\left(\frac{E\left[e_i^4\right]}{\sigma_e^4} - 3\right).$$

The estimator of the asymptotic variance in the general case can be constructed as

$$\hat{V}^J_{B2SLS} = 2\hat{\lambda}\left(1 - \hat{\alpha}\right) + \frac{\hat{\rho}_\alpha}{\sum_j\left(M_W^{ij}\right)^4}\left(\frac{\overline{(\hat{e}_{B2SLS})_i^4}}{\hat{\sigma}^4_{B2SLS}} - 3\overline{(M_W^{ii})^2}\right),$$

while under error normality it is sufficient to use only the first term and set the second term to zero.

Corollary to Theorem 5: Suppose assumptions 1–3 hold, and $\lambda > 0$. Then

$$\hat{V}^J_{B2SLS} \xrightarrow{p} V^J_{B2SLS}.$$

The J type test is one-sided (see Lee and Okui, 2012): we reject the null if the value of

$$\frac{J_{B2SLS}}{\sqrt{n\hat{V}^J_{B2SLS}}}$$

exceeds $q_\eta^{N(0,1)}$, the $(1 - \eta)$-quantile of the standard normal.

Note that when there are no exogenous regressors ($m = 0$ and $\alpha = \lambda$),

$$V^J_{B2SLS} = 2\lambda\left(1 - \lambda\right) + \left(\lim\overline{(P_Z^{ii})^2} - \lambda^2\right)\left(\frac{E\left[e_i^4\right]}{\sigma_e^4} - 3\right),$$

---

[1]Among other things, Lee and Okui (2012) prove that a test based on such statistic is in fact asymptotically equivalent to the Hahn and Hausman (2002) test for instrument validity.

and the test coincides with that in Lee and Okui (2012). In this limiting case one can simply put

$$\hat{V}^J_{B2SLS} = 2\hat{\lambda}(1-\hat{\lambda}) + \left(\overline{(P^{ii}_Z)^2} - \hat{\lambda}^2\right)\left(\frac{\overline{(\hat{e}_{B2SLS})^4_i}}{\hat{\sigma}^4_{B2SLS}} - 3\right).$$

If one uses only the first term in this estimate, the test is similar to the Anatolyev and Gospodinov (2011) J test. That is, the Anatolyev and Gospodinov (2011) J test is not robust to the numerosity of exogenous regressors even under error normality, in contrast to the general test presented here.

5 LIML estimation

5.1 Construction

Consider now the LIML estimator (Anderson and Rubin, 1949, Bekker 1994, Donald and Newey 2001, Hansen, Hausman and Newey, 2008, van Hasselt 2010):

$$\hat{\beta}_{LIML} = \arg\min_\beta \frac{(Y - X\beta)' M_W (Y - X\beta)}{(Y - X\beta)' M_{ZW} (Y - X\beta)},$$

or, equivalently,

$$\hat{\beta}_{LIML} = \arg\min_\beta F(\beta),$$

where

$$F(\beta) = \frac{(Y - X\beta)' P_{Z^\perp} (Y - X\beta)}{(Y - X\beta)' M_W (Y - X\beta)}.$$

The second formulation is more convenient because the probability limit of the minimized value of the objective function is exactly α (see the proof of Theorem 6). Note also that the estimator $\hat{\beta}_{LIML}$ may be interpreted as a LIML estimator in a transformed equation where the left side variable is $Y^\perp$ and the right side variables are $X^\perp$ using the instruments (Z, W).

Let us write out the first order conditions for $\hat{\beta}_{LIML}$ as

$$X'P_{Z^\perp}\left(Y - X\hat{\beta}_{LIML}\right) - F(\hat{\beta}_{LIML}) \cdot X'M_W\left(Y - X\hat{\beta}_{LIML}\right) = 0.$$

This implies

$$\hat{\beta}_{LIML} = \frac{X'P_{\tilde{\alpha}}Y}{X'P_{\tilde{\alpha}}X},$$

where

$$\tilde{\alpha} = F(\hat{\beta}_{LIML}).$$

It is easy to show (see the proof of Theorem 6 below) that $\tilde{\alpha} \xrightarrow{p} \alpha$. The value of $\tilde{\alpha}$, which provides an alternative way of calculating the LIML estimator, may be computed as the smallest eigenvalue of the matrix $(\mathring{X}'\mathring{X})^{-1}\mathring{X}'P_{ZW}\mathring{X}$, where $\mathring{X} = \left(Y^\perp, X^\perp\right)$.

Let us define the residual variance

$$\hat{\sigma}^2_{LIML} = \frac{\hat{e}'_{LIML}\hat{e}_{LIML}}{n-m-p}.$$

Note the degrees of freedom adjustment.

5.2 Asymptotic properties

The following theorem establishes the asymptotic properties of the LIML estimator in the asymptotic framework under consideration.

Theorem 6: Suppose assumptions 1–3 hold, and $\lambda > 0$. Then the corrected LIML estimator $\hat{\beta}_{LIML}$ is consistent and zero mean asymptotically normal with the asymptotic variance

$$V_{LIML} = \Phi_2^{LIML} + \Phi_3^{LIML} + \Phi_4^{LIML},$$

where

$$\Phi_2^{LIML} = \sigma_e^2 Q_{XZ^\perp}^{-1} + \frac{\lambda}{1-\alpha} Q_{XZ^\perp}^{-1} \left(\sigma_e^2 \Sigma_u - \Sigma_{ue}\Sigma'_{ue}\right) Q_{XZ^\perp}^{-1},$$

$$\Phi_3^{LIML} = \frac{1}{1-\alpha} Q_{XZ^\perp}^{-1} \left(\pi_\alpha E\left[e_i^2 \tilde{u}_i'\right] + E\left[e_i^2 \tilde{u}_i\right] \pi'_\alpha\right) Q_{XZ^\perp}^{-1},$$

and

$$\Phi_4^{LIML} = \frac{\rho_\alpha}{(1-\alpha)^2} Q_{XZ^\perp}^{-1} E\left[\left(e_i^2 - \sigma_e^2\right)\tilde{u}_i\tilde{u}_i'\right] Q_{XZ^\perp}^{-1}.$$

The effects of the numerous exogenous regressors on the asymptotic variance are similar to those for the appropriately bias corrected 2SLS.

The difference between V_{B2SLS} and V_{LIML} is composed of three differences, $\Phi_2^{B2SLS} - \Phi_2^{LIML}$, $\Phi_3^{B2SLS} - \Phi_3^{LIML}$ and $\Phi_4^{B2SLS} - \Phi_4^{LIML}$, where

$$\Phi_2^{B2SLS} - \Phi_2^{LIML} = \frac{2\lambda}{1-\alpha} Q_{XZ^\perp}^{-1} \Sigma_{ue}\Sigma'_{ue} Q_{XZ^\perp}^{-1}$$

$$\Phi_3^{B2SLS} - \Phi_3^{LIML} = \frac{1}{1-\alpha}\frac{E\left[e_i^3\right]}{\sigma_e^2} Q_{XZ^\perp}^{-1} \left(\pi_\alpha \Sigma'_{ue} + \Sigma_{ue}\pi'_\alpha\right) Q_{XZ^\perp}^{-1}$$

$$\Phi_4^{B2SLS} - \Phi_4^{LIML} = \frac{\rho_\alpha}{(1-\alpha)^2} Q_{XZ^\perp}^{-1} \left(\frac{E\left[e_i^3 u_i\right]}{\sigma_e^2}\Sigma'_{ue} + \Sigma_{ue}\frac{E\left[e_i^3 u_i'\right]}{\sigma_e^2}\right.$$
$$\left. -\Sigma_{ue}\Sigma'_{ue}\left(\frac{E\left[e_i^4\right]}{\sigma_e^4} + 3\right)\right) Q_{XZ^\perp}^{-1}.$$

The difference $\Phi_2^{B2SLS} - \Phi_2^{LIML}$ is positive semidefinite, hence the LIML estimator is asymptotically more efficient than the bias-corrected 2SLS estimator under error normality. This result has a well known analog when the number of exogenous regressors is small (e.g., Chao and Swanson 2006; van Hasselt, 2010). Under error non-normality, the differences $\Phi_3^{B2SLS} - \Phi_3^{LIML}$ and $\Phi_4^{B2SLS} - \Phi_4^{LIML}$ can potentially take any sign. Even in the few exogenous regressors framework, van Hasselt (2010) finds it difficult (if not

to say impossible) and unproductive to analyze the additional terms in the asymptotic variance, while Hansen, Hausman and Newey (2008) mention that the adjustments for non-normality Φ_3^{LIML} and Φ_4^{LIML} tend to be small, at least when λ is small, in contrast to the second term in Φ_2^{LIML} responsible for numerosity of instruments. Anderson, Kunimoto and Matsushita (2010), Anatolyev and Gospodinov (2011), and Lee and Okui (2012) find via simulations that the effects of deviation from normality are barely noticeable for non-extreme error distributions. Thus, it is quite unlikely that the sum of the second and third differences of asymptotic variance components, if negative, will overweigh the first difference.

5.3   Variance estimation

Hansen, Hausman and Newey (2008) propose a variance estimator (HHN henceforth) in the case of many instruments that takes into account error non-normality. Denote

$$\bar{X} = (X, W) - \hat{e}_{LIML}\bar{\Gamma}_{LIML},$$

where

$$\bar{\Gamma}_{LIML} = \frac{\hat{e}'_{LIML}(X, W)}{\hat{e}'_{LIML}\hat{e}_{LIML}}$$

estimates $(\Gamma, 0'_{m \times 1})$. The HHN estimator has the following structure:

$$\hat{V}_{LIML} = n \left[\bar{H}^{-1}\bar{\Sigma}\bar{H}^{-1} \right]_{1:p,1:p},$$

where

$$\bar{H} = (X, W)' \left(P_{ZW} - \bar{\alpha}I_n \right) (X, W),$$

$$\bar{\alpha} = \frac{\hat{e}'_{LIML}P_{ZW}\hat{e}_{LIML}}{\hat{e}'_{LIML}\hat{e}_{LIML}},$$

and

$$\bar{\Sigma} = \hat{\Sigma}_0 + \hat{A} + \hat{A}' + \hat{B},$$

where the terms $\hat{A}$ and $\hat{B}$ estimate the terms related to the third and fourth moments, respectively, that are present under error non-normality in addition to the term $\hat{\Sigma}_0$ that is present under error normality. It turns out that under error normality the term

$$\hat{\Sigma}_0 = \hat{\sigma}_{LIML}^2 \left((1 - \bar{\alpha})^2 \bar{X}'P_{ZW}\bar{X} + \bar{\alpha}^2 \bar{X}'M_{ZW}\bar{X} \right)$$

given in Hansen, Hausman and Newey (2008) still correctly estimates the relevant portion of the asymptotic variance when there are many exogenous regressors. However, the terms $\hat{A}$ and $\hat{B}$ given in Hansen, Hausman and Newey (2008, p.399) do not correctly estimate their corresponding portions when exogenous regressors are numerous. We therefore

change those to

$$\hat{A} = n(1 - \hat{\alpha})\hat{\pi}_\alpha \overline{\frac{(\hat{e}_{LIML})_i^2 \left(M_{ZW}\bar{X}\right)_i'}{\sum_j \left(M_W^{ij}\right)^2 M_{ZW}^{ij}}}$$

and

$$\hat{B} = n\hat{\rho}_\alpha \overline{\frac{\left(\left(\hat{e}_{LIML}\right)_i^2 - \hat{\phi}\hat{\sigma}_{LIML}^2\right)\left(M_{ZW}\bar{X}\right)_i \left(M_{ZW}\bar{X}\right)_i'}{\sum_j \left(M_W^{ij} M_{ZW}^{ij}\right)^2}}.$$

The following proposition formalizes the asymptotic validity of our modification.

Theorem 7: Suppose assumptions 1–3 hold, and $\lambda > 0$. Then the modified HHN estimator $\hat{V}_{LIML}$ is consistent for V_{LIML}.

Suppose there are no exogenous regressors ($m = 0$ and $\alpha = \lambda$), then $\hat{\mu} = 0$, $\hat{\alpha} = \hat{\lambda}$, $P_{\hat{\alpha}} = P_Z - \hat{\lambda}I_n$, $\hat{\rho}_\alpha = \overline{(P_Z^{ii})^2} - \hat{\lambda}^2$, $\overline{\sum_j \left(M_W^{ij}\right)^2 M_{ZW}^{ij}} = \overline{M_Z^{ii}} = 1 - \hat{\lambda}$, $\overline{(M_Z^{ii})^2} = 1 - 2\hat{\lambda} + \overline{(P_Z^{ii})^2}$, $\overline{M_W^{ii} M_{ZW}^{ii}} = \overline{M_Z^{ii}} = 1 - \hat{\lambda}$, and it is easily seen that $\hat{A}$ and $\hat{B}$ become exactly as those in Hansen, Hausman and Newey (2008, p.399):

$$\hat{A} = \sum_{i=1}^n \left(P_Z^{ii} - \hat{\lambda}\right)(P_Z X)_i \cdot \frac{1}{n}\sum_{i=1}^n (\hat{e}_{LIML})_i^2 \left(M_Z \bar{X}\right)_i',$$

$$\hat{B} = \frac{\overline{(P_Z^{ii})^2} - \hat{\lambda}^2}{1 - 2\hat{\lambda} + \overline{(P_Z^{ii})^2}} \sum_{i=1}^n \left((\hat{e}_{LIML})_i^2 - \hat{\sigma}_{LIML}^2\right)\left(M_Z \bar{X}\right)_i \left(M_Z \bar{X}\right)_i'.$$

5.4 Specification testing

Now consider a modified J statistic, a suitable quadratic form in the LIML residuals:

$$J_{LIML} = \frac{\hat{e}_{LIML}' \left(P_{ZW} - \hat{\alpha}I_n\right)\hat{e}_{LIML}}{\hat{\sigma}_{LIML}^2} = \frac{\hat{e}_{LIML}' P_{\hat{\alpha}} \hat{e}_{LIML}}{\hat{\sigma}_{LIML}^2}.$$

This an adaptation of the Sargan type statistic of Lee and Okui (2012) to the case of many exogenous regressors and LIML estimation.

Theorem 8: Suppose assumptions 1–3 hold, and $\lambda > 0$. Then

$$\frac{J_{LIML}}{\sqrt{n}} \xrightarrow{d} N\left(0, V_{LIML}^J\right),$$

where the asymptotic variance is

$$V_{LIML}^J = 2\lambda(1 - \alpha) + \rho_\alpha \left(\frac{E[e_i^4]}{\sigma_e^4} - 3\right).$$

Note the equivalence to the B2SLS-based modified J test. The estimator of the asymptotic variance in the general case can be constructed analogously as

$$\hat{V}_{LIML}^J = 2\hat{\lambda}(1 - \hat{\alpha}) + \frac{\hat{\rho}_\alpha}{\sum_j \left(M_W^{ij}\right)^4}\left(\frac{\overline{(\hat{e}_{LIML})_i^4}}{\hat{\sigma}_{LIML}^4} - 3\overline{(M_W^{ii})^2}\right).$$

while under error normality it is sufficient to use only the first term and set the second term to zero.

Corollary to Theorem 8: Suppose assumptions 1–3 hold, and $\lambda > 0$. Then

$$\hat{V}^J_{LIML} \xrightarrow{p} V^J_{LIML}.$$

Note that if there are no exogenous regressors ($m = 0$ and $\alpha = \lambda$), then

$$V^J_{LIML} = 2\lambda\,(1 - \lambda) + \left(\lim \overline{(P_Z^{ii})^2} - \lambda^2\right)\left(\frac{E\,[e_i^4]}{\sigma_e^4} - 3\right).$$

In this limiting case one can simply put

$$\hat{V}^J_{LIML} = 2\hat{\lambda}\left(1 - \hat{\lambda}\right) + \left(\overline{(P_Z^{ii})^2} - \hat{\lambda}^2\right)\left(\frac{\overline{(\hat{e}_{LIML})_i^4}}{\hat{\sigma}^4_{LIML}} - 3\right).$$

If one uses only the first term in this estimate, the test becomes similar to the Anatolyev and Gospodinov (2011) J test. That is, the Anatolyev and Gospodinov (2011) J test is not robust to the numerosity of exogenous regressors, in contrast to the general test presented here.

The J type test is one-sided (see Lee and Okui, 2012): we reject the null if the value of

$$\frac{J_{LIML}}{\sqrt{n\hat{V}^J_{LIML}}}$$

exceeds $q_\eta^{N(0,1)}$, the $(1 - \eta)$-quantile of the standard normal.

Another possible statistic to test for the model specification is based on the difference between the minimized objective function $\tilde{\alpha} = F(\hat{\beta}_{LIML})$ and a direct estimate $\hat{\alpha}$ of α, but this yields an asymptotically equivalent test. Moreover, inspection of the proofs of Theorems 5, 6 and 8 reveals that the statistics $\sqrt{n}J_{B2SLS}$, $\sqrt{n}J_{LIML}$ and $\sqrt{n}\,(\tilde{\alpha} - \hat{\alpha})$ all are based on the asymptotic behavior of the same quantity,

$$\frac{e'P_\alpha e}{\sqrt{n}},$$

and hence the corresponding tests are asymptotically equivalent.

5.5 K-class estimation

One can also extend the notion of a K-class estimator (Nagar, 1959; Hansen, Hausman and Newey, 2008) to the present situation with numerous exogenous regressors:

$$\hat{\beta}_\kappa = \frac{X'P_\kappa Y}{X'P_\kappa X},$$

18

where κ is a constant or data-dependent quantity. From inspection of proofs it follows that $\hat{\beta}_\kappa$ is consistent only if $\kappa \xrightarrow{p} \alpha$. Further, as

$$\frac{X'P_\kappa e}{\sqrt{n}} = \frac{X'P_{\tilde{\alpha}} e}{\sqrt{n}} - \sqrt{n}\,(\kappa - \tilde{\alpha})\,\frac{(X^\perp)' e}{\sqrt{n}},$$

$\hat{\beta}_\kappa$ is asymptotically equivalent to $\hat{\beta}_{LIML}$ if $\sqrt{n}\,(\kappa - \tilde{\alpha}) \xrightarrow{p} 0$. One popular choice

$$\kappa = \frac{\tilde{\alpha} - (1 - \tilde{\alpha})\,c/n}{1 - (1 - \tilde{\alpha})\,c/n}$$

leads to the Fuller (1977) type estimator, which is asymptotically equivalent to $\hat{\beta}_{LIML}$ under many exogenous regressors.

6   Simulation experiment

In this section we carry out a small Monte-Carlo experiment. Of primary interest to us is the amount of distortions arising from the presence of many exogenous regressors when they play a role, and to verify for how big their proportion to the sample size may make the distortions sizable. It is also of interest to see how well the asymptotic theory works for a typical sample size.

The data for the Monte Carlo experiment are generated from

$$y = \beta x + \sum_{j=1}^{m} \delta_j w_j + e,$$

$$x = \sum_{j=1}^{\ell} \pi_{1j} z_j + \sum_{j=1}^{m} \pi_{2j} w_j + u,$$

Here, the parameter values are $\beta = 1$, $\delta_j = \pi_{2j} = 1/\sqrt{m}$ for $j = 1, ..., m$, and $\pi_{1j} = (1, 0'_{\ell-1})'$. The local-to-zero δ_j's and π_{1j}'s keep the information contained in included exogenous variables fixed as m changes. The exogenous regressors w_j are independent and distributed according to the standard normal except for one exogenous regressor which is set to unity. When the instruments z_j's are few (small ℓ), they are independent standard normal; when they are many (large ℓ), they are generated as in Hausman, Newey, Woutersen, Chao, and Swanson (2010): $z = (\varsigma, \varsigma^2, \varsigma^3, \varsigma^4, \varsigma D_1, ..., \varsigma D_{\ell-4})$, where $D_k \in \{0, 1\}$, $\Pr\{D_k = 1\} = \frac{1}{2}$ for $k = 1, ..., \ell - 4$, and ς is standard normal. Next, each of errors e and u is drawn from the skew Student's distribution of Azzalini and Capitanio (2003) transformed to have zero means and unit variances, with the parameters implying the skewness coefficient 1.38 and the kurtosis coefficient 6.54.[2] The covariance between e and u is 0.9. The sample size is $n = 400$. The results are based on 5,000 simulations.

---

[2] We used the procedure ST from the Gauss library SKEW by Roncalli and Lagache (2004).

First we report the actual rejection rates at the 5% nominal level of the 2SLS based t test and AR and K tests when there are $\ell = 2$ instruments. The last two columns correspond to the AR and K tests when we additionally divide π_{1j} by $\sqrt{n}$ making the set of instruments weak.

μ	t-2SLS	J-2SLS	AR	K	AR(W)	K(W)
0.1	4.9%	5.3%	5.0%	4.8%	5.0%	5.3%
0.2	5.3%	4.8%	5.1%	4.7%	5.1%	4.7%
0.3	5.1%	4.9%	4.8%	4.6%	4.8%	4.9%
0.4	5.2%	4.9%	4.8%	4.5%	4.8%	4.6%
0.6	5.2%	4.8%	5.0%	4.9%	5.0%	5.0%
0.8	6.1%	5.3%	6.1%	5.8%	6.1%	5.4%

Clearly, all considered tests are robust to the numerosity of exogenous regressors exhibiting rejection rates very close to the nominal one. When μ gets very large though, the tests acquire a slight tendency to overreject.

Now we switch to the many instrument case and look at the medians of simulated bias corrected 2SLS estimators and the LIML estimator for some values of λ and μ.

$\lambda = \mu$	med-B2SLS$_1$	med-B2SLS$_2$	med-B2SLS	med-LIML
0.1	-0.114	0.014	-0.002	0.001
0.2	-0.303	0.057	-0.001	0.000
0.3	-0.665	0.140	-0.003	0.000
0.4	-1.226	0.276	0.000	-0.002

The median bias is obviously present in the "inappropriately" corrected 2SLS estimators, and is practically absent in the properly corrected one, as well as in the LIML estimator. The first version of "inappropriate" correction leads to a much higher bias than the second version, and these biases are of opposite signs.

Next, we present the actual rejection rates at the 5% nominal level of t and J tests. The third and fourth columns of the following table show rejection rates based on the appropriately corrected 2SLS estimator using the proposed standard errors, and the LIML estimator using the modified HHN standard errors. The fifth and sixth columns show, respectively, the null rejection rates for the Anatolyev and Gospodinov (2011, AG henceforth) and Lee and Okui (2012, LO henceforth) J type tests that account for the numerosity of instruments but do not for the numerosity of exogenous regressors. The decision rule

of the AG test has the following form: $J > q^{\chi^2(\ell-1)}_{\Phi\left(\sqrt{1-\lambda}\Phi^{-1}(\alpha)\right)}$, where J is the conventional J statistic based on LIML residuals. Finally, the seventh and eighth columns represent the proposed J type tests based on the bias corrected 2SLS and LIML estimators, respectively.

λ	μ	t-B2SLS	t-LIML	J-AG	J-LO	J-B2SLS	J-LIML
0.1	0.1	5.0%	5.2%	6.5%	5.8%	5.7%	5.3%
	0.2	4.5%	5.1%	7.1%	6.2%	6.2%	5.3%
	0.3	5.2%	4.9%	5.6%	6.0%	5.8%	5.0%
	0.4	4.9%	5.3%	5.2%	6.0%	5.8%	5.1%
0.2	0.1	5.0%	5.1%	5.8%	5.8%	5.5%	5.0%
	0.2	5.0%	5.2%	6.5%	6.7%	6.1%	5.5%
	0.3	4.4%	4.8%	5.2%	6.3%	5.7%	4.9%
	0.4	4.8%	4.5%	4.5%	6.6%	6.5%	5.2%
0.3	0.1	4.8%	4.5%	7.1%	6.1%	6.1%	4.9%
	0.2	4.3%	5.1%	6.1%	6.7%	6.3%	5.1%
	0.3	5.0%	4.4%	4.2%	6.2%	5.8%	4.5%
	0.4	5.6%	5.0%	3.4%	6.4%	6.4%	4.8%
0.4	0.1	5.1%	4.7%	6.6%	5.9%	5.7%	4.5%
	0.2	5.5%	5.1%	5.2%	6.6%	6.6%	4.6%
	0.3	6.2%	5.4%	2.9%	6.0%	6.0%	4.3%
	0.4	6.5%	4.4%	1.2%	5.2%	6.1%	3.9%

Both t tests exhibit just slight distortions even for a moderate sample size, with an exception of the B2SLS-based one when both μ and λ are big. The LIML based standard errors seem to be a bit more reliable than those based on B2SLS. The (asymptotically incorrect) AG test behaves unstably, in some cases exhibiting perceptible overrejection, in some cases perceptible underrejection. The (also asymptotically incorrect) LO test behaves more stably showing consistent overrejection, which turns out non-severe for a moderate degree of non-normality. Our J type B2SLS based test also tends to slightly overreject, while actual rejection rates for the J type LIML based test are very close to the nominal level, again except when both μ and λ are big.

7 Computational note

The tests proposed in this paper are easy and straightforward to implement. A computational difficulty may be to program various objects which are functions of elements of

matrices $P_{\hat{\alpha}}$, M_W, M_{ZW} and others. In GAUSS, they can be coded in the following way. Let the matrices P_W and P_{ZW} be coded as `Pw` and `Pzw`, then the matrices M_W, M_{ZW} and $P_{\hat{\alpha}}$ are coded as `Mw=eye(n)-Pw`, `Mzw=eye(n)-Pzw` and `Pa=Pzw-Pw-ahat*(eye(n)-Pw)`, where scalar `n` contains n, and `ahat` contains $\hat{\alpha}$. The following table represents statements for various objects, functions of elements of these matrices.

$\overline{\left(P_{\hat{\alpha}}^{ii}\right)^2}$	`meanc(diag(Pa)^2)`
$\overline{\left(M_W^{ii}\right)^2}$	`meanc(diag(Mw)^2)`
$\overline{M_W^{ii} M_{ZW}^{ii}}$	`meanc(diag(Mw.*Mzw))`
$\overline{\sum_j \left(M_W^{ij}\right)^3}$	`meanc(sumc(Mw^3))`
$\overline{\sum_j \left(M_W^{ij}\right)^2 M_{ZW}^{ij}}$	`meanc(sumc((Mw^2).*Mzw))`
$\overline{\sum_j \left(M_W^{ij}\right)^4}$	`meanc(sumc(Mw^4))`
$\overline{\sum_j \left(M_W^{ij} M_{ZW}^{ij}\right)^2}$	`meanc(sumc((Mw^2).*(Mzw^2)))`

Other objects of interest involving these matrices may be formed similarly. For example, let vector `ehat` contain the elements of $\hat{e}$, and `X` contain elements of X. Then the coding table is

$\overline{P_{\hat{\alpha}}^{ii} \left(P_{Z^\perp} X\right)_i}$	`meanc(diag(Pa).*((Pzw-Pw)*X))`
$\hat{e}_i^2 \left(M_{ZW} X\right)_i'$	`(ehat^2)'(Mzw*X)/n`
$\hat{e}_i^2 \left(M_{ZW} X\right)_i \left(M_{ZW} X\right)_i'$	`X'Mzw*diagrv(eye(N),ehat^2)*Mzw*X/n`

8 Conclusion

We have considered a standard linear instrumental variables regression with few or many instruments where the number of exogenous regressors may be large and comparable to the sample size. Within an asymptotic framework where the number of exogenous regressors and possibly instruments is proportional to the sample size, some existing estimators and tests are robust to their numerosity, which is however not true for others. In those cases when the presence of many exogenous regressors invalidates an estimator or test under consideration, we have proposed its modified asymptotically valid version. Future research may extend to instrumental variables models with heteroskedasticity, with the literature for the many instrument case still being developed (e.g., Chao, Swanson, Hausman, Newey, and Woutersen, 2012; Hausman, Newey, Woutersen, Chao, and Swanson, 2010), and accommodate serial correlation in model errors.

9 Acknowledgements

A Appendix: proofs

Lemma A1: Suppose (Z, W) has full column rank. The following is true: $P_{Z^\perp} = P_{ZW} - P_W$, $P_{Z^\perp} M_W = M_W P_{Z^\perp} M_W = P_{Z^\perp}$, $M_W M_{Z^\perp} M_W = M_{Z^\perp} M_W = M_W - P_{Z^\perp}$.

Proof of Lemma A1.

By the partitioned matrix inverse $P_{ZW} = (Z, W) \left((Z, W)' (Z, W) \right)^{-1} (Z, W)'$ is

$$(Z, W) \begin{bmatrix} \Psi_n^{-1} & -\Psi_n^{-1} Z_W' \\ -Z_W \Psi_n^{-1} & (W'W)^{-1} + Z_W \Psi_n^{-1} Z_W' \end{bmatrix} (Z, W)',$$

where $Z_W = (W'W)^{-1} W'Z$ and $\Psi_n = Z' M_W Z$. Hence,

$$P_{ZW} - P_W = M_W Z (Z' M_W Z)^{-1} Z' M_W = P_{Z^\perp}.$$

Next, $P_{Z^\perp} M_W = (P_{ZW} - P_W) M_W = P_{ZW} - P_W = P_{Z^\perp}$, $M_W P_{Z^\perp} M_W = M_W P_{Z^\perp} = (P_{Z^\perp} M_W)' = P_{Z^\perp}' = P_{Z^\perp}$. Finally, $M_W M_{Z^\perp} M_W = M_W M_W - M_W P_{Z^\perp} M_W = M_W - P_{Z^\perp}$.

Lemma A2: Suppose assumptions 1–3 hold. Then

$$\frac{X' P_{Z^\perp} X}{n} \xrightarrow{p} Q_{XZ^\perp} + \lambda \Sigma_u, \qquad \frac{X' P_{Z^\perp} e}{n} \xrightarrow{p} \lambda \Sigma_{ue}, \qquad \frac{e' P_{Z^\perp} e}{n} \xrightarrow{p} \lambda \sigma_e^2,$$

$$\frac{X' M_W X}{n} \xrightarrow{p} Q_{XZ^\perp} + (1 - \mu) \Sigma_u, \qquad \frac{X' M_W e}{n} \xrightarrow{p} (1 - \mu) \Sigma_{ue}, \qquad \frac{e' M_W e}{n} \xrightarrow{p} (1 - \mu) \sigma_e^2,$$

$$\frac{X' M_{ZW} X}{n} \xrightarrow{p} (1 - \mu - \lambda) \Sigma_u, \qquad \frac{X' M_{ZW} e}{n} \xrightarrow{p} (1 - \mu - \lambda) \Sigma_{ue}, \qquad \frac{e' M_{ZW} e}{n} \xrightarrow{p} (1 - \mu - \lambda) \sigma_e^2.$$

Proof of Lemma A2. First, by Lemma A1 of Hansen, Hausman and Newey (2008) using Assumptions 2 and 3,

$$\frac{U' P_{ZW} U}{n} \xrightarrow{p} (\lambda + \mu) \Sigma_u, \qquad \frac{U' P_W U}{n} \xrightarrow{p} \mu \Sigma_u, \qquad \frac{U' P_{ZW} e}{n} \xrightarrow{p} (\lambda + \mu) \Sigma_{ue},$$

$$\frac{U' P_W e}{n} \xrightarrow{p} \mu \Sigma_{ue}, \qquad \frac{e' P_{ZW} e}{n} \xrightarrow{p} (\lambda + \mu) \sigma_e^2, \qquad \frac{e' P_W e}{n} \xrightarrow{p} \mu \sigma_e^2.$$

It follows immediately that $n^{-1} e' P_{Z^\perp} e \xrightarrow{p} \lambda \sigma_e^2$ and $n^{-1} e' M_W e \xrightarrow{p} (1 - \mu) \sigma_e^2$.

Now, consider $n^{-1} \Upsilon_{XZ}' P_{Z^\perp} e$ whose mean is zero and whose variance is

$$E \left[\frac{\Upsilon_{XZ}' P_{Z^\perp} e e' P_{Z^\perp} \Upsilon_{XZ}}{n^2} \right] = \frac{\sigma_e^2}{n} \frac{\Upsilon_{XZ}' P_{Z^\perp} \Upsilon_{XZ}}{n} \to 0,$$

hence $n^{-1} \Upsilon_{XZ}' P_{Z^\perp} e \xrightarrow{p} 0$. Similarly, one can show that $n^{-1} \Upsilon_{XZ}' M_W e \xrightarrow{p} 0$, $n^{-1} \Upsilon_{XZ}' P_{Z^\perp} U \xrightarrow{p} 0$ and $n^{-1} \Upsilon_{XZ}' M_W U \xrightarrow{p} 0$.

Now, $P_{Z^\perp} X = P_{Z^\perp} \Upsilon_{XZ} + P_{Z^\perp} U$, hence

$$\frac{X' P_{Z^\perp} X}{n} = \frac{(P_{Z^\perp} X)' P_{Z^\perp} X}{n} = \frac{\Upsilon_{XZ}' \Upsilon_{XZ}}{n} - \frac{\Upsilon_{XZ}' P_W \Upsilon_{XZ}}{n} + \frac{\Upsilon_{XZ}' P_{Z^\perp} U}{n} + \frac{U' P_{Z^\perp} \Upsilon_{XZ}}{n}$$

$$+ \frac{U' P_{Z^\perp} U}{n} \xrightarrow{p} Q_{XZ^\perp} + \lambda \Sigma_u$$

and

$$\frac{X'P_{Z^\perp}e}{n} = \frac{\Upsilon'_{XZ}P_{Z^\perp}e}{n} + \frac{U'P_{Z^\perp}e}{n} \xrightarrow{p} \lambda\Sigma_{ue}.$$

Similarly, $M_W X = M_W \Upsilon_{XZ} + M_W U$, hence

$$\frac{X'M_W X}{n} = \frac{(M_W X)' M_W X}{n} = \frac{\Upsilon'_{XZ}M_W \Upsilon_{XZ}}{n} + \frac{\Upsilon'_{XZ}M_W U}{n} + \frac{U'M_W \Upsilon_{XZ}}{n} + \frac{U'M_W U}{n}$$
$$\xrightarrow{p} Q_{XZ^\perp} + (1-\mu)\Sigma_u$$

and

$$\frac{X'M_W e}{n} = \frac{\Upsilon'_{XZ}M_W e}{n} + \frac{U'M_W e}{n} \xrightarrow{p} (1-\mu)\Sigma_{ue}.$$

The last three results directly follow from the previous ones.

Proof of Proposition 1. Straightforward computations using Lemma A2 yield

$$\hat{\beta}_{2SLS} = \beta_0 + \left(n^{-1}X'P_{Z^\perp}X\right)^{-1} n^{-1}X'P_{Z^\perp}e \xrightarrow{p} \beta_0 + (Q_{XZ^\perp} + \lambda\Sigma_u)^{-1}\lambda\Sigma_{ue}.$$

Next, recall that $\hat{e}_{2SLS} = M_W(Y - X\hat{\beta}_{2SLS})$, or

$$\hat{e}_{2SLS} = M_W e - M_W X \left(X'P_{Z^\perp}X\right)^{-1} X'P_{Z^\perp}e.$$

Then, using Lemma A2,

$$\frac{\hat{e}'_{2SLS}P_{ZW}\hat{e}_{2SLS}}{n} = \frac{e'P_{Z^\perp}e}{n} - \frac{e'P_{Z^\perp}X}{n}\left(\frac{X'P_{Z^\perp}X}{n}\right)^{-1}\frac{X'P_{Z^\perp}e}{n}$$
$$\xrightarrow{p} \lambda\sigma_e^2 - \lambda^2\Sigma'_{ue}\tilde{\Sigma}\Sigma_{ue},$$

where $\tilde{\Sigma} = (Q_{XZ^\perp} + \lambda\Sigma_u)^{-1}$. Next,

$$\frac{\hat{e}'_{2SLS}\hat{e}_{2SLS}}{n} = \frac{e'M_W e}{n} - 2\frac{e'M_W X}{n}\left(\frac{X'P_{Z^\perp}X}{n}\right)^{-1}\frac{X'P_{Z^\perp}e}{n}$$
$$+\frac{e'P_{Z^\perp}X}{n}\left(\frac{X'P_{Z^\perp}X}{n}\right)^{-1}\frac{X'M_W X}{n}\left(\frac{X'P_{Z^\perp}X}{n}\right)^{-1}\frac{X'P_{Z^\perp}e}{n}$$
$$\xrightarrow{p} (1-\mu)\sigma_e^2 - 2\lambda(1-\mu)\Sigma'_{ue}\tilde{\Sigma}\Sigma_{ue}$$
$$+\lambda^2(1-\mu)\Sigma'_{ue}\tilde{\Sigma}(Q_{XZ^\perp} + (1-\mu)\Sigma_u)\tilde{\Sigma}\Sigma_{ue}$$
$$= (1-\mu)\left(\sigma_e^2 + \lambda\Sigma'_{ue}\tilde{\Sigma}((\lambda-2)Q_{XZ^\perp} - \lambda(1+\mu)\Sigma_u)\tilde{\Sigma}\Sigma_{ue}\right).$$

Summarizing, the J statistic has the following limit:

$$\frac{J_{2SLS}}{n} \xrightarrow{p} \alpha\frac{\sigma_e^2 - \lambda\Sigma'_{ue}\tilde{\Sigma}\Sigma_{ue}}{\sigma_e^2 + \lambda\Sigma'_{ue}\tilde{\Sigma}((\lambda-2)Q_{XZ^\perp} - \lambda(1+\mu)\Sigma_u)\tilde{\Sigma}\Sigma_{ue}} \equiv \alpha J^\infty.$$

Proof of Theorem 1. The probability limit of the residual variance is σ_e^2. The estimator $\hat{\beta}_{2SLS}$ satisfies

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta_0) = \left(\frac{X'P_{Z^\perp}X}{n}\right)^{-1} \frac{X'P_{Z^\perp}e}{\sqrt{n}}.$$

As in Lemma A2, $P_{Z^\perp}X = M_W Z \Pi_1 + P_{Z^\perp}U$ and hence $n^{-1}X'P_{Z^\perp}X \xrightarrow{p} \Pi_1' Q_{Z^\perp} \Pi_1$ and

$$\frac{X'P_{Z^\perp}e}{\sqrt{n}} = \Pi_1' \frac{(Z^\perp)'e}{\sqrt{n}} + \frac{U'P_{Z^\perp}e}{\sqrt{n}}.$$

Because $U'P_{Z^\perp}e = n^{-1/2}U'Z^\perp \left(n^{-1}(Z^\perp)'Z^\perp\right)^{-1} n^{-1/2}(Z^\perp)'e = O_P(1) Q_{Z^\perp} O_P(1)$, the second term above is $o_P(1)$. Now let

$$W_i = \frac{Z_i^\perp e_i}{\sqrt{n}}.$$

Note that by Assumptions 2 and 3, we have $\sum_{i=1}^n E[\|Z_i^\perp e_i\|^4] \leq E[e_i^4] \sum_{i=1}^n \|Z_i^\perp\|^4 \leq o(n^2)$. Hence $\sum_{i=1}^n E[\|W_i\|^4] \leq n^{-2}o(n^2) \to 0$. Then we have that $\sum_{i=1}^n W_i W_i'$ converges to $\sigma_e^2 Q_{Z^\perp}$. By Lemma A2 of Hansen, Hausman and Newey (2008), $(Z^\perp)'e/\sqrt{n}$ is asymptotically normal with the asymptotic variance $\sigma_e^2 Q_{Z^\perp}$. Together,

$$\frac{X'P_{Z^\perp}e}{\sqrt{n}} \xrightarrow{d} N\left(0, \sigma_e^2 \Pi_1' Q_{Z^\perp} \Pi_1\right)$$

and

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta_0) \xrightarrow{d} N\left(0, \sigma_e^2 \left(\Pi_1' Q_{Z^\perp} \Pi_1\right)^{-1}\right).$$

The t statistic for the jth component based on 2SLS is

$$t_{2SLS}^{(j)} = \frac{\sqrt{n}(\hat{\beta}_{2SLS} - \beta_0)_j}{\sqrt{\hat{\sigma}_{2SLS}^2 \left[(n^{-1}X'P_{Z^\perp}X)^{-1}\right]^{jj}}} \xrightarrow{d} \frac{N\left(0, \sigma_e^2 \left[(\Pi_1' Q_{Z^\perp} \Pi_1)^{-1}\right]^{j,j}\right)}{\sqrt{\sigma_e^2 \left[(\Pi_1' Q_{Z^\perp} \Pi_1)^{-1}\right]^{jj}}} \stackrel{d}{=} N(0,1).$$

Next, the numerator of the J statistic is

$$
\begin{aligned}
\hat{e}_{2SLS}' P_{ZW} \hat{e}_{2SLS} &= e'P_{Z^\perp}e - \frac{e'P_{Z^\perp}X}{\sqrt{n}} \left(\frac{X'P_{Z^\perp}X}{n}\right)^{-1} \frac{X'P_{Z^\perp}e}{\sqrt{n}} \\
&= \frac{e'Z^\perp}{\sqrt{n}} Q_{Z^\perp}^{-1} \frac{(Z^\perp)'e}{\sqrt{n}} - \frac{e'Z^\perp}{\sqrt{n}} \Pi_1 \left(\Pi_1' Q_{Z^\perp} \Pi_1\right)^{-1} \Pi_1' \frac{(Z^\perp)'e}{\sqrt{n}} + o_P(1) \\
&= \psi'\psi + o_P(1),
\end{aligned}
$$

where

$$\psi = \left(I_\ell - Q_{Z^\perp}^{1/2} \Pi_1 \left(\Pi_1' Q_{Z^\perp} \Pi_1\right)^{-1} \Pi_1' Q_{Z^\perp}^{1/2}\right) Q_{Z^\perp}^{-1/2} \frac{(Z^\perp)'e}{\sqrt{n}}.$$

Taking account of the fact that the first factor in ψ is an idempotent matrix of rank $\ell - p$, we have by the standard argumentation that $\psi'\psi \xrightarrow{d} \sigma_e^2 \chi^2 (\ell - p)$. As the denominator of the J statistic is consistent for σ_e^2, the two pieces together yield the conclusion.

Proof of Theorem 2. Note that $Y^\perp - X^\perp \beta_0 = M_W (Y - X\beta_0) = M_W (e + W\delta_0) = M_W e$. Now, irrespective of the strength of instruments,

$$\frac{\left(Y^\perp - X^\perp \beta_0\right)' M_{Z^\perp} \left(Y^\perp - X^\perp \beta_0\right)}{n - \ell - m} = \frac{e' M_W e}{n - \ell - m} - \frac{e' P_{Z^\perp} e}{n - \ell - m} \xrightarrow{p} \sigma_e^2$$

and

$$\frac{\left(Y^\perp - X^\perp \beta_0\right)' M_{Z^\perp} X^\perp}{n - \ell - m} = \frac{e' M_W X}{n - \ell - m} - \frac{e' P_{Z^\perp} X}{n - \ell - m} \xrightarrow{p} \Sigma'_{ue}$$

using Lemmas A1 and A2. By Lemma A1, $\left(Y^\perp - X^\perp \beta_0\right)' P_{Z^\perp} \left(Y^\perp - X^\perp \beta_0\right) = e' P_{Z^\perp} e$, and from the proof of Theorem 1 $e' P_{Z^\perp} e \xrightarrow{d} \sigma_e^2 \chi^2 (\ell)$. Hence,

$$AR = \frac{1}{\ell} \frac{e' P_{Z^\perp} e}{\left(Y^\perp - X^\perp \beta_0\right)' M_{Z^\perp} \left(Y^\perp - X^\perp \beta_0\right) / (n - \ell - m)} \xrightarrow{d} \frac{1}{\ell} \frac{\sigma_e^2 \chi^2 (\ell)}{\sigma_e^2} \sim \frac{\chi^2 (\ell)}{\ell}.$$

Next, again using Lemma A1,

$$\tilde{Z} = P_{Z^\perp} \left(M_W X - M_W (Y - X\beta_0) \frac{\Sigma'_{ue} + o_P (1)}{\sigma_e^2 + o_P (1)} \right) = P_{Z^\perp} \left(X - e \left(\Gamma + o_P (1) \right) \right).$$

When the instruments are strong,

$$\frac{\tilde{Z}'\tilde{Z}}{n} = \frac{X' P_{Z^\perp} X}{n} + \frac{e' P_{Z^\perp} e}{n} \left(\Gamma + o_P (1) \right)^2 - 2\frac{X' P_{Z^\perp} e}{n} \left(\Gamma + o_P (1) \right)$$

$$\xrightarrow{p} \Pi'_1 Q_{Z^\perp} \Pi_1,$$

using Lemma A2. Hence

$$\left(Y^\perp - X^\perp \beta_0\right)' P_{\tilde{Z}} \left(Y^\perp - X^\perp \beta_0\right) = e' M_W P_{\tilde{Z}} M_W e$$
$$= \psi'_1 \left(\Pi'_1 Q_{Z^\perp} \Pi_1\right)^{-1} \psi_1 + o_P (1),$$

where

$$\psi_1 \equiv \frac{1}{\sqrt{n}} (X - e\Gamma)' P_{Z^\perp} e = \frac{X' P_{Z^\perp} e}{\sqrt{n}} - \Gamma' \frac{e' P_{Z^\perp} e}{\sqrt{n}}$$
$$\xrightarrow{d} N \left(0, \sigma_e^2 \Pi'_1 Q_{Z^\perp} \Pi_1\right)$$

using the proof of Theorem 1. To summarize,

$$K = \frac{\psi'_1 \left(\Pi'_1 Q_{Z^\perp} \Pi_1\right)^{-1} \psi_1 + o_P (1)}{\sigma_e^2 + o_P (1)} \xrightarrow{d} \chi^2 (p).$$

Now let fixed Π_1 be replaced by $\Pi_1 = \mathring{\Pi}/\sqrt{n}$ so that the instruments are weak or irrelevant. Recall that $\tilde{U} = U - e\Gamma$, and note that $\tilde{U}$ is uncorrelated with e. Then jointly

$$
\Psi \equiv \begin{pmatrix} \Psi_{\tilde{U}} \\ \Psi_e \end{pmatrix} \equiv \mathrm{vec}\left(\left(\frac{(Z^\perp)'Z^\perp}{n} \right)^{-1/2} \frac{(Z^\perp)'(\tilde{U}, e)}{\sqrt{n}} \right)
$$

$$
\xrightarrow{d} \begin{pmatrix} \xi_{\tilde{U}} \\ \xi_e \end{pmatrix} \sim N\left(0, \mathrm{diag}\left\{ \Sigma_u - \sigma_e^2 \Gamma' \Gamma, \sigma_e^2 \right\} \otimes I_\ell \right).
$$

Using Lemma A1,

$$
\tilde{Z} = P_{Z^\perp}\left(Z \frac{\mathring{\Pi}}{\sqrt{n}} + \tilde{U} \right) + P_{Z^\perp} e \cdot o_P(1),
$$

therefore

$$
\tilde{Z}'\tilde{Z} = \left(Z \frac{\mathring{\Pi}}{\sqrt{n}} + \tilde{U} \right)' P_{Z^\perp}\left(Z \frac{\mathring{\Pi}}{\sqrt{n}} + \tilde{U} \right) + o_P(1) = \Psi_2' \Psi_2 + o_P(1),
$$

where

$$
\begin{aligned}
\Psi_2 &\equiv \left(\frac{(Z^\perp)'Z^\perp}{n} \right)^{-1/2} \left(\frac{(Z^\perp)'Z^\perp}{n} \mathring{\Pi} + \frac{(Z^\perp)'\tilde{U}}{\sqrt{n}} \right) \\
&= \left(\frac{(Z^\perp)'Z^\perp}{n} \right)^{1/2} \mathring{\Pi} + \Psi_{\tilde{U}} \xrightarrow{d} Q_{Z^\perp}^{1/2}\mathring{\Pi} + \xi_{\tilde{U}} \equiv \xi_2.
\end{aligned}
$$

Hence,

$$
e' M_W P_{\tilde{Z}} M_W e = \Psi_e' \Psi_2 \left(\Psi_2' \Psi_2 + o_P(1) \right)^{-1} \Psi_2' \Psi_e \xrightarrow{d} \xi_e' P_{\xi_2} \xi_e.
$$

Then, due to independence of ξ_2 and ξ_e and idempotence of P_{ξ_2}, we have $P_{\xi_2}\xi_e|\xi_2 \sim N\left(0, \sigma_e^2 P_{\xi_2} \right)$ and $\xi_e' P_{\xi_2}\xi_e|\xi_2 \sim \sigma_e^2 \chi^2\left(\mathrm{rk}\left(P_{\xi_2} \right) \right) = \sigma_e^2 \chi^2(p)$, and hence $\xi_e' P_{\xi_2}\xi_e \sim \sigma_e^2 \chi^2(p)$ and $K \xrightarrow{d} \chi^2(p)$.

   Proof of Proposition 2. For the first version,

$$
\check{\beta}_{B2SLS} - \beta_0 = (X' P_{\lambda+\mu} X)^{-1} X' P_{\lambda+\mu} e + o_P(1).
$$

From Lemma A2(a),

$$
\begin{aligned}
\check{\beta}_{B2SLS} - \beta_0 &= \left(\frac{X' P_{Z^\perp} X}{n} - (\lambda + \mu) \frac{X' M_W X}{n} \right)^{-1}\left(\frac{X' P_{Z^\perp} e}{n} - (\lambda + \mu) \frac{X' M_W e}{n} \right) + o_P(1) \\
&\xrightarrow{p} \left(Q_{XZ^\perp} + \lambda \Sigma_u - (\lambda + \mu)\left(Q_{XZ^\perp} + (1 - \mu)\Sigma_u \right) \right)^{-1} \\
&\quad \times \left(\lambda \Sigma_{ue} - (\lambda + \mu)(1 - \mu)\Sigma_{ue} \right) = -\mu\left(Q_{XZ^\perp} - \mu\Sigma_u \right)^{-1}\Sigma_{ue}.
\end{aligned}
$$

Similarly, for the second version,

$$\check{\beta}_{B2SLS} - \beta_0 = \left(\frac{X'P_\lambda X}{n}\right)^{-1}\frac{X'P_\lambda e}{n} + o_P(1)$$

$$\xrightarrow{p} (Q_{XZ^\perp} + \lambda\Sigma_u - \lambda(Q_{XZ^\perp} + (1-\mu)\Sigma_u))^{-1}$$

$$\times (\lambda\Sigma_{ue} - \lambda(1-\mu)\Sigma_{ue})$$

$$= \lambda\mu((1-\lambda)Q_{XZ^\perp} + \lambda\mu\Sigma_u)^{-1}\Sigma_{ue}.$$

Lemma A3: Suppose assumptions 1–3 hold, and $\lambda > 0$. Then

$$\frac{X'P_\alpha X}{n} \xrightarrow{p} (1-\alpha)Q_{XZ^\perp}, \quad \frac{X'P_\alpha e}{n} \xrightarrow{p} 0, \quad \frac{e'P_\alpha e}{n} \xrightarrow{p} 0.$$

Proof of Lemma A3. Using Lemma A2,

$$\frac{X'P_\alpha X}{n} = \frac{X'P_{Z^\perp}X}{n} - \alpha\frac{X'M_W X}{n} \xrightarrow{p} Q_{XZ^\perp} + \lambda\Sigma_u - \alpha(Q_{XZ^\perp} + (1-\mu)\Sigma_u)$$

$$= (1-\alpha)Q_{XZ^\perp}$$

and similarly

$$\frac{X'P_\alpha e}{n} \xrightarrow{p} \lambda\Sigma_{ue} - \alpha(1-\mu)\Sigma_{ue} = 0.$$

Last,

$$\frac{e'P_\alpha e}{n} = \frac{e'P_{ZW}e}{n} - (1-\alpha)\frac{e'P_W e}{n} - \alpha\frac{e'e}{n} \xrightarrow{p} (\lambda+\mu)\sigma_e^2 - (1-\alpha)\mu\sigma_e^2 - \alpha\sigma_e^2 = 0.$$

Lemma A4. Under Assumption 1, the following quantities are bounded from above by $O(n)$: $\sum_{i=1}^n\sum_{j=1}^n (P_\alpha^{ij})^4$, $\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n (P_\alpha^{ij})^2(P_\alpha^{ik})^2$, $\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n |(P_\alpha^{ij})^2 P_\alpha^{ik}P_\alpha^{jk}|$ and $\text{tr}((P_\alpha - D_\alpha)^4)$, where $D_\alpha = \text{diag}(P_\alpha^{11}, ..., P_\alpha^{nn})$.

Proof. Note that $|P_\alpha^{ij}| \le |P_{Z^\perp}^{ij}| + \alpha|M_W^{ij}| \le 1+\alpha$. Using the inequality $(a-b)^2 \le 2(a^2+b^2)$ which follows from the inequality $(a+b)^2 \ge 0$, we get $\sum_{j=1}^n (P_\alpha^{ij})^2 = 2\sum_{j=1}^n (P_{Z^\perp}^{ij})^2 + 2\alpha^2\sum_{j=1}^n (M_W^{ij})^2 \le 2P_{Z^\perp}^{ii} + 2\alpha^2 M_W^{ii}$. Then $\sum_{i=1}^n\sum_{j=1}^n (P_\alpha^{ij})^4 \le (1+\alpha)^2\sum_{i=1}^n\sum_{j=1}^n (P_\alpha^{ij})^2 \le 2(1+\alpha)^2\left(\sum_{i=1}^n P_{Z^\perp}^{ii} + \alpha^2\sum_{i=1}^n M_W^{ii}\right) = O(n)$ and $\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n (P_\alpha^{ij})^2(P_\alpha^{ik})^2 = \sum_{i=1}^n\left(\sum_{j=1}^n (P_\alpha^{ij})^2\right)^2 \le 4\sum_{i=1}^n\left(P_{Z^\perp}^{ii} + \alpha^2 M_W^{ii}\right)^2 \le 4(1+\alpha^2)\left(\sum_{i=1}^n P_{Z^\perp}^{ii} + \alpha^2\sum_{i=1}^n M_W^{ii}\right) = O(n)$. Further, $\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n |(P_\alpha^{ij})^2 P_\alpha^{ik}P_\alpha^{jk}| \le \sum_{i=1}^n\sum_{j=1}^n (P_\alpha^{ij})^2\left(\sum_{k=1}^n (P_\alpha^{ik})^2\sum_{k=1}^n (P_\alpha^{jk})^2\right)^{1/2} \le 4\sum_{i=1}^n\sum_{j=1}^n (P_\alpha^{ij})^2\left((P_{Z^\perp}^{ii} + \alpha^2 M_W^{ii})(P_{Z^\perp}^{jj} + \alpha^2 M_W^{jj})\right)^{1/2} \le 4(1+\alpha^2)\sum_{i=1}^n\sum_{j=1}^n (P_\alpha^{ij})^2 \le 8(1+\alpha^2)\sum_{i=1}^n\left(P_{Z^\perp}^{ii} + \alpha^2 M_W^{ii}\right) = O(n)$.

Next, consider the matrix trace inequality $\text{tr}((P_\alpha - D_\alpha)^4) \le 2\text{tr}(P_\alpha^4) + 2\text{tr}(D_\alpha^4) + 8\text{tr}(P_\alpha^2 D_\alpha^2) + 4\text{tr}((P_\alpha D_\alpha)^2)$ which follows from the inequality $\text{tr}((P_\alpha + D_\alpha)^4) \ge 0$ after expansion of the fourth power and collecting terms of equal trace. From Baumgartner (2011,

theorem 2) it follows that $\mathrm{tr}(P_\alpha^2 D_\alpha^2) \leq (\mathrm{tr}\,(P_\alpha^4)\,\mathrm{tr}\,(D_\alpha^4))^{1/2}$. From Baumgartner (2011, theorem 3) and by positive semidefiniteness of $P_\alpha |D_\alpha| P_\alpha$ it follows that $|\mathrm{tr}((P_\alpha D_\alpha)^2)| \leq \max_{1\leq i\leq n}|D_\alpha^{ii}|\,\mathrm{tr}\,(P_\alpha \mathrm{diag}(|P_\alpha^{11}|,...,|P_\alpha^{nn}|)P_\alpha) \leq (\max_{1\leq i\leq n}|P_\alpha^{ii}|)^2 \mathrm{tr}\,(P_\alpha^2) \leq (1+\alpha)^2 \mathrm{tr}\,(P_\alpha^2)$. Now, $\mathrm{tr}\,(D_\alpha^4) = \sum_{i=1}^n (P_\alpha^{ii})^4 \leq (1+\alpha)^4 n = O(n)$. Then, $\mathrm{tr}\,(P_\alpha^2) = \sum_{i=1}^n \sum_{j=1}^n (P_\alpha^{ij})^2 \leq 2\sum_{i=1}^n \left(P_{Z^\perp}^{ii} + \alpha^2 M_W^{ii}\right) = O(n)$. Finally, $\mathrm{tr}\,(P_\alpha^4) = \mathrm{tr}\,((P_{Z^\perp} - \alpha M_W)^4) \leq 2\mathrm{tr}\,(P_{Z^\perp}^4) + 2\alpha^4 \mathrm{tr}\,(M_W^4) + 8\alpha^2 \mathrm{tr}(P_{Z^\perp}^2 M_W^2) + 4\alpha^2 \mathrm{tr}\,((P_{Z^\perp} M_W)^2) = (2+12\alpha^2)\,\mathrm{tr}\,(P_{Z^\perp}) + 2\alpha^4 \mathrm{tr}\,(M_W) = O(n)$. Collecting the pieces, $\mathrm{tr}\,((P_\alpha - D_\alpha)^4) \leq 2O(n) + 2O(n) + 8\,(O(n)O(n))^{1/2} + 4(1+\alpha)^2 O(n) = O(n)$.

Lemma A5: Suppose assumptions 1–3 hold, and $\lambda > 0$. Then for a constant matrix $B = \begin{pmatrix} B_1 \\ b_2' \end{pmatrix}$ with row dimension $p + 1$,

$$\frac{B'\,(X,e)'\,P_\alpha e}{\sqrt{n}} \xrightarrow{d} N\left(0, (1-\alpha)\,\Xi_2 + (1-\alpha)\,\Xi_3 + \rho_\alpha \Xi_4\right),$$

where

$$
\begin{aligned}
\Xi_2 &= B_1'\left((1-\alpha)\,\sigma_e^2 Q_{XZ^\perp} + \lambda \sigma_e^2 \Sigma_u + \lambda \Sigma_{ue} \Sigma_{ue}'\right) B_1 \\
&\quad + 2\lambda \sigma_e^2 \left(B_1'\Sigma_{ue} b_2' + b_2 \Sigma_{ue}' B_1 + \sigma_e^2 b_2 b_2'\right), \\
\Xi_3 &= B_1'\pi_\alpha E\left[e_i^2 (u_i', e_i)\right] B + B'E\left[e_i^2 (u_i', e_i)'\right]\pi_\alpha' B_1, \\
\Xi_4 &= B'\left(E\left[e_i^2 (u_i', e_i)'(u_i', e_i)\right]\right) B - B_1'\left(\sigma_e^2 \Sigma_u + 2\Sigma_{ue} \Sigma_{ue}'\right) B_1 \\
&\quad - 3\sigma_e^2 \left(B_1'\Sigma_{ue} b_2' + b_2 \Sigma_{ue}' B_1 + \sigma_e^2 b_2 b_2'\right).
\end{aligned}
$$

Proof. Let us represent the quantity of interest as

$$
\begin{aligned}
\frac{B'\,(X,e)'\,P_\alpha e}{\sqrt{n}} &= \frac{B'\,(P_\alpha X, P_\alpha e)'\,e}{\sqrt{n}} = \frac{((1-\alpha)\,M_W \Upsilon_{XZ} B_1 + P_\alpha U B_1, P_\alpha e b_2')'\,e}{\sqrt{n}} \\
&= \frac{((1-\alpha)\,M_W \Upsilon_{XZ} B_1 + P_\alpha U B_1 + P_\alpha e b_2')'\,e}{\sqrt{n}} + \frac{(U B_1 + e b_2')'\,P_\alpha e}{\sqrt{n}} \\
&= \sum_{i=1}^n \frac{P_\alpha^{ii}}{\sqrt{n}}\left(B_1'\Sigma_{ue} + b_2 \sigma_e^2\right) + (1-\alpha)\sum_{i=1}^n \frac{(B_1'\Upsilon_{XZ}' M_W)_i\, e_i}{\sqrt{n}} \\
&\quad + \sum_{i=1}^n \frac{B_1' P_\alpha^{ii}\,(u_i e_i - \Sigma_{ue})}{\sqrt{n}} + \sum_{i=1}^n \frac{b_2 P_\alpha^{ii}\,(e_i^2 - \sigma_e^2)}{\sqrt{n}} \\
&\quad + \sum_{i\neq j} \frac{(B_1' u_i + b_2 e_i)\,P_\alpha^{ij} e_j}{\sqrt{n}} \\
&= S_1 + S_2 + S_3 + S_4 + S_5
\end{aligned}
$$

First,

$$
\begin{aligned}
\sum_{i=1}^n \frac{P_\alpha^{ii}}{\sqrt{n}} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(P_{Z^\perp}^{ii} - \alpha M_W^{ii}\right) = \frac{1}{\sqrt{n}}\,(\ell - \alpha\,(n-m)) \\
&= \sqrt{n}\left(\lambda + o\,(1/\sqrt{n}) - \frac{\lambda}{1-\mu}\,(1 - \mu + o\,(1/\sqrt{n}))\right) = o\,(1),
\end{aligned}
$$

hence $S_1 = o(1)$.

Next, the inspection of Lemma A2 of Chao, Swanson, Hausman, Newey, and Woutersen (2012) (cf. its fixed-instrument-design analog, Lemma A2 of Hansen, Hausman and Newey, 2008) reveals that this central limit theorem also holds when P in it is replaced by $P_\alpha = P_{Z^\perp} - \alpha M_W$ thanks to the inequalities for P_α from Lemma A4, with an appropriate adjustment in the expression for the asymptotic variance. In particular, the multiplier $\overline{\sum_{j=1}^n (P^{ij})^2} - \overline{(P^{ii})^2} = n^{-1}\mathrm{rk}(P) - \overline{(P^{ii})^2}$ in the variance expression becomes $\overline{\sum_{j=1}^n (P_\alpha^{ij})^2} - \overline{(P_\alpha^{ii})^2} = (1 - 2\alpha) n^{-1}\mathrm{rk}(P_{Z^\perp}^{ii}) + \alpha^2 n^{-1}\mathrm{rk}(M_W^{ii}) - \overline{(P_\alpha^{ii})^2}$ whose limit value is $(1 - \alpha)\lambda - \rho_\alpha$.

Now let

$$W_i = \frac{1}{\sqrt{n}} \begin{pmatrix} (B_1' \Upsilon_{XZ}' M_W)_i \, e_i \\ B_1' P_\alpha^{ii} (u_i e_i - \Sigma_{ue}) \\ b_2 P_\alpha^{ii} (e_i^2 - \sigma_e^2) \end{pmatrix}.$$

Note that using Assumptions 2 and 3 and that $\sum_{i=1}^n (P_\alpha^{ii})^4 \le O(n)$ (see the proof of Lemma A4), we have that $\sum_{i=1}^n E[\|(B_1' \Upsilon_{XZ}' M_W)_i \, e_i\|^4] \le E[e_i^4] \|B_1\|^4 \sum_{i=1}^n \|(\Upsilon_{XZ}' M_W)_i\|^4 \le o(n^2)$. Next, $\sum_{i=1}^n E[\|B_1' P_\alpha^{ii} (u_i e_i - \Sigma_{ue})\|^4] \le \sum_{i=1}^n (P_\alpha^{ii})^4 \sum_d \|(B_1)_d\|^4 E[(u_i e_i - \Sigma_{ue})_d^4] \le O(n)$, and $\sum_{i=1}^n E[\|b_2 P_\alpha^{ii} (e_i^2 - \sigma_e^2)\|^4] \le \|b_2\|^4 \sum_{i=1}^n (P_\alpha^{ii})^4 E[(e_i^2 - \sigma_e^2)^4] \le O(n)$. Hence $\sum_{i=1}^n E[\|W_i\|^4] \le n^{-2} o(n^2) \to 0$. Then we have that $\sum_{i=1}^n W_i W_i'$ converges to

$$\begin{pmatrix} \sigma_e^2 B_1' Q_{XZ^\perp} B_1 & B_1' \pi_\alpha E[e_i^2 u_i'] B_1 & B_1' \pi_\alpha b_2' E[e_i^3] \\ B_1' E[u_i e_i^2] \pi_\alpha' B_1 & \rho_\alpha B_1' E[(u_i e_i - \Sigma_{ue})(u_i e_i - \Sigma_{ue})'] B_1 & \rho_\alpha B_1' E[u_i' e_i (e_i^2 - \sigma_e^2)] b_2' \\ b_2 \pi_\alpha' B_1 E[e_i^3] & \rho_\alpha b_2 E[u_i e_i (e_i^2 - \sigma_e^2)] B_1 & \rho_\alpha b_2 b_2' E[(e_i^2 - \sigma_e^2)^2] \end{pmatrix} \equiv \bar{\Psi}.$$

Next, for any comfortable vector c,

$$E\left[\left(\sum_{i \ne j} \frac{c'(B_1' u_i + b_2 e_i) P_\alpha^{ij} e_j}{\sqrt{n}}\right)^2\right]$$

$$= \frac{1}{n} \sum_{i \ne j} \sum_{l \ne k} P_\alpha^{ij} P_\alpha^{lk} c' E\left[(B_1' u_i + b_2 e_i)(B_1' u_l + b_2 e_l)' e_j e_k\right] c$$

$$\to ((1 - \alpha)\lambda - \rho_\alpha) c' \begin{pmatrix} \sigma_e^2 E[(B_1' u_i + b_2 e_i)(B_1' u_i + b_2 e_i)'] + \\ E[(B_1' u_i + b_2 e_i) e_i] E[(B_1' u_i + b_2 e_i)' e_i] \end{pmatrix} c$$

$$\equiv c' \bar{\Omega} c.$$

Hence we have

$$c'(S_2 + S_3 + S_4 + S_5) \xrightarrow{d} N\left(0, c' \mathrm{diag}\{(1 - \alpha) I, I, I\} \bar{\Psi} \mathrm{diag}\{(1 - \alpha) I, I, I\} c + c' \bar{\Omega} c\right),$$

and it follows that

$$\frac{B'(X, e)' P_\alpha e}{\sqrt{n}} \xrightarrow{d} N(0, \bar{V})$$

31

$$\begin{aligned}
\bar{V} &= \text{diag}\left\{(1-\alpha)I, I, I\right\} \bar{\Psi} \text{diag}\left\{(1-\alpha)I, I, I\right\} + \bar{\Omega} \\
&= (1-\alpha)^2 \sigma_e^2 B_1' Q_{XZ^\perp} B_1 + (1-\alpha)\left(B_1' E\left[u_i e_i^2\right]\pi_\alpha' B_1 + B_1'\pi_\alpha E\left[e_i^2 u_i'\right]B_1\right) \\
&\quad + (1-\alpha)\left(b_2\pi_\alpha' B_1 E\left[e_i^3\right] + B_1'\pi_\alpha b_2' E\left[e_i^3\right]\right) \\
&\quad + \rho_\alpha b_2 E\left[u_i e_i\left(e_i^2 - \sigma_e^2\right)\right]B_1 + \rho_\alpha B_1' E\left[u_i'e_i\left(e_i^2 - \sigma_e^2\right)\right]b_2' \\
&\quad + \rho_\alpha B_1' E\left[(u_i e_i - \Sigma_{ue})(u_i e_i - \Sigma_{ue})'\right]B_1 + \rho_\alpha b_2 b_2' E\left[\left(e_i^2 - \sigma_e^2\right)^2\right] \\
&\quad + ((1-\alpha)\lambda - \rho_\alpha)\sigma_e^2 E\left[(B_1'u_i + b_2 e_i)(B_1'u_i + b_2 e_i)'\right] \\
&\quad + ((1-\alpha)\lambda - \rho_\alpha) E\left[(B_1'u_i + b_2 e_i)e_i\right] E\left[(B_1'u_i + b_2 e_i)'e_i\right],
\end{aligned}$$

which coincides with the variance expression in the statement of the Lemma.

Lemma A6: Suppose assumptions 1–3 hold, and $\lambda > 0$. Then $\hat{\pi}_\alpha \to \pi_\alpha$ and $\hat{\rho}_\alpha \to \rho_\alpha$.

Proof of Lemma A6. Recall that $P_{Z^\perp}X = M_W \Upsilon_{XZ} + P_{Z^\perp}U$. Now, $E\left[\overline{P_\alpha^{ii}(P_{Z^\perp}U)_i}\right] = 0$ and

$$\begin{aligned}
\text{var}\left(\overline{P_\alpha^{ii}(P_{Z^\perp}U)_i}\right) &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n P_\alpha^{ii}P_\alpha^{jj}\sum_{l=1}^n\sum_{q=1}^n P_{Z^\perp}^{il}P_{Z^\perp}^{jq}E\left[u_l u_q'\right] \\
&= \frac{\Sigma_u}{n^2}\sum_{i=1}^n\sum_{j=1}^n P_\alpha^{ii}P_\alpha^{jj}\sum_{l=1}^n P_{Z^\perp}^{il}P_{Z^\perp}^{jl} = \frac{\Sigma_u}{n^2}\sum_{i=1}^n\sum_{j=1}^n P_\alpha^{ii}P_\alpha^{jj}P_{Z^\perp}^{ij},
\end{aligned}$$

which in absolute value is no greater than

$$\frac{\Sigma_u}{n^2}\left(\sup_{n\geq 1}\sup_{1\leq i\leq n}\left|P_\alpha^{ii}\right|\right)^2\sum_{i=1}^n\sum_{j=1}^n\left|P_{Z^\perp}^{ij}\right| \leq \frac{\Sigma_u(1+\alpha)^2}{n^2}n\sqrt{n} \to 0$$

as $\sum_{j=1}^n\left|P_{Z^\perp}^{ij}\right| \leq \|P_{Z^\perp}\|_1 \leq \sqrt{n}\|P_{Z^\perp}\|_2 = \sqrt{n}\lambda_{\max}^{1/2}\left(P_{Z^\perp}'P_{Z^\perp}\right) = \sqrt{n}$. Hence,

$$\overline{P_{\hat{\alpha}}^{ii}(P_{Z^\perp}X)_i} = \overline{P_\alpha^{ii}(M_W\Upsilon_{XZ})_i} + \overline{P_\alpha^{ii}(P_{Z^\perp}U)_i} + o_p(1) \to \pi_\alpha$$

according to Assumption 3. Also, by Assumption 3, $\hat{\rho}_\alpha \to \rho_\alpha$.

Proof of Theorem 3. From Lemma A3,

$$\hat{\beta}_{B2SLS} - \beta_0 = \left(\frac{X'P_\alpha X}{n}\right)^{-1}\frac{X'P_\alpha e}{n} + o_P(1) \xrightarrow{p} 0.$$

Next,

$$\sqrt{n}\left(\hat{\beta}_{B2SLS} - \beta_0\right) = \left(\frac{X'P_\alpha X}{n}\right)^{-1}\frac{X'P_\alpha e}{\sqrt{n}} + o_P(1).$$

Application of Lemma A3, Lemma A5 with $B = (I_p, 0_p)'$ and the Slutsky Theorem yields

$$\sqrt{n}\left(\hat{\beta}_{B2SLS} - \beta_0\right) \xrightarrow{d} ((1-\alpha)Q_{XZ^\perp})^{-1} N\left(0, \begin{array}{l} (1-\alpha)^2\sigma_e^2 Q_{XZ^\perp} + ((1-\alpha)\lambda - \rho_\alpha)\sigma_e^2\Sigma_u \\ + ((1-\alpha)\lambda - 2\rho_\alpha)\Sigma_{ue}\Sigma_{ue}' + \rho_\alpha E\left[e_i^2 u_i u_i'\right] \\ + (1-\alpha)\left(\pi_\alpha E\left[e_i^2 u_i'\right] + E\left[e_i^2 u_i\right]\pi_\alpha'\right) \end{array}\right)$$

$$\xoverset{d}{=} N\left(0, \Phi_2^{B2SLS} + \Phi_3^{B2SLS} + \Phi_4^{B2SLS}\right),$$

where Φ_2^{B2SLS}, Φ_3^{B2SLS} and Φ_4^{B2SLS} are as in the statement of the Theorem.

Lemma A7: Suppose assumptions 1–3 hold, and $\lambda > 0$. The following expressions are $o_P(1)$:

I. $\overline{(M_W e)_i^3} - \overline{\sum_j (M_W^{ij})^3 E[e_i^3]}$,

II. $\overline{(M_W e)_i^2 (M_{ZW} \tilde{U})_i} - \overline{\sum_j (M_W^{ij})^2 M_{ZW}^{ij} E[e_i^2 \tilde{u}_i]}$,

III. $\overline{(M_W e)_i^4} - \overline{\sum_j (M_W^{ij})^4 E[e_i^4 - 3\sigma_e^4]} - 3\overline{(M_W^{ii})^2} \sigma_e^4$,

IV. $\overline{(M_W e)_i^2 (M_{ZW} \tilde{U})_i (M_{ZW} \tilde{U})_i'} - \overline{\sum_j (M_W^{ij} M_{ZW}^{ij})^2 E[(e_i^2 - \sigma_e^2) \tilde{u}_i \tilde{u}_i']} - \overline{M_W^{ii} M_{ZW}^{ii}} \sigma_e^2 E[\tilde{u}_i \tilde{u}_i']$.

Proof of Lemma A7. We heavily use the properties of M_W such as $|M_W^{ij}| < 1$, $\sum_j (M_W^{ij})^2 = M_W^{ii} \leq 1$, $\sum_k |M_W^{i_1 k} M_W^{i_2 k}| \leq \left(\sum_k (M_W^{i_1 k})^2 \sum_k (M_W^{i_2 k})^2 \right)^{1/2} \leq (M_W^{i_1 i_1} M_W^{i_2 i_2})^{1/2} \leq 1$ and similarly $\sum_k |M_W^{i_1 k} M_{ZW}^{i_2 k}| \leq 1$, $\sum_j |M_W^{ij}| \leq \|M_W\|_1 \leq \sqrt{n} \|M_W\|_2 = \sqrt{n} \lambda_{\max}^{1/2}(M_W' M_W) = \sqrt{n}$, Lemma B1 in Chao, Swanson, Hausman, Newey, and Woutersen (2012), and the following and similar inequalities:

$$\left| \sum_{l \notin \{j_1, j_2, k\}} M_W^{i_1 l} M_W^{i_2 l} \right| \leq \left| \sum_l M_W^{i_1 l} M_W^{i_2 l} - M_W^{i_1 j_1} M_W^{i_2 j_1} - M_W^{i_1 j_2} M_W^{i_2 j_2} - M_W^{i_1 k} M_W^{i_2 k} \right|$$
$$= \left| M_W^{i_1 i_2} \right| + \left| M_W^{i_1 j_1} M_W^{i_2 j_1} \right| + \left| M_W^{i_1 j_2} M_W^{i_2 j_2} \right| + \left| M_W^{i_1 k} M_W^{i_2 k} \right|.$$

We will show in detail how to obtain parts II and III; the other two are handled analogously. The critical statement for part I is $E\left[\overline{(M_W e)_i^3} \right] = \overline{\sum_j (M_W^{ij})^3 E[e_i^3]}$, while the critical statements for part VI are $E\left[\overline{(M_W e)_i^2 (M_{ZW} \tilde{U})_i (M_{ZW} \tilde{U})_i'} \right] = \overline{\sum_j (M_W^{ij} M_{ZW}^{ij})^2 E[e_i^2 \tilde{u}_i \tilde{u}_i']} + \overline{\sum_j \sum_{k \neq j} (M_W^{ij} M_{ZW}^{ik})^2 \sigma_e^2 E[\tilde{u}_i \tilde{u}_i']}$ and $\sum_j \sum_{k \neq j} (M_W^{ij} M_{ZW}^{ik})^2 = M_W^{ii} M_{ZW}^{ii} - \sum_j (M_W^{ij} M_{ZW}^{ij})^2$.

Part II. Observe that

$$\overline{(M_W e)_i^2 (M_{ZW} \tilde{U})_i} = \overline{(\sum_j M_W^{ij} e_j)^2 \sum_j M_{ZW}^{ij} \tilde{u}_j} = \overline{\sum_j (M_W^{ij})^2 M_{ZW}^{ij} E[e_i^2 \tilde{u}_i]} + T_1 + T_2 + T_3 + T_4,$$

where the formulas for T_1, T_2 and T_3 appear below. We will show that $T_1 + T_2 + T_3 + T_4 = O_P(1/n)$. First, take

$$T_1 = \sum_j \sum_{k \neq j} \overline{(M_W^{ij})^2 M_{ZW}^{ik} e_j^2 \tilde{u}_k},$$

whose variance

$$\sum_{j_1} \sum_{k_1 \neq j_1} \sum_{j_2} \sum_{k_2 \neq j_2} \overline{(M_W^{ij_1})^2 M_{ZW}^{ik_1} (M_W^{ij_2})^2 M_{ZW}^{ik_2} E\left[e_{j_1}^2 \tilde{u}_{k_1} e_{j_2}^2 \tilde{u}_{k_2} \right]}$$

contains terms that are not zero, in the following cases:
(i) $j_1 = j_2$, $k_1 = k_2 \neq j_1$ leading to scaled $E[e_j^4] E[\tilde{u}_k \tilde{u}_k']$, then the variance component is proportional to

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j (M_W^{i_1 j})^2 (M_W^{i_2 j})^2 \sum_{k \neq j} M_{ZW}^{i_1 k} M_{ZW}^{i_2 k},$$

33

which in absolute value is no greater than

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left(M_W^{i_1 j}\right)^2 \left(M_W^{i_2 j}\right)^2 \sum_{k \neq j} \left|M_{ZW}^{i_1 k} M_{ZW}^{i_2 k}\right| \leq \frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left(M_W^{i_1 j}\right)^2 \left(M_W^{i_2 j}\right)^2$$

$$= O\left(\frac{n-m}{n^2}\right) = o(1);$$

(ii) $j_1 = k_2$, $j_2 = k_1 \neq j_1$ leading to scaled $E\left[e_j^2 \tilde{u}_j\right] E\left[e_j^2 \tilde{u}_j'\right]$, then the variance component is proportional to

$$\frac{1}{n^2} \sum_j \sum_{k \neq j} \sum_{i_1} \sum_{i_2} \left(M_W^{i_1 j}\right)^2 M_{ZW}^{i_1 k} \left(M_W^{i_2 k}\right)^2 M_{ZW}^{i_2 j},$$

which in absolute value is no greater than

$$\frac{1}{n^2} \sum_j \sum_{i_1} \left(M_W^{i_1 j}\right)^2 \sum_{i_2} \left|M_{ZW}^{i_2 j}\right| \sum_k \left(M_W^{i_2 k}\right)^2 \leq \frac{1}{n^2} \sum_j \sqrt{n} = o(1).$$

Second, take

$$\frac{T_2}{2} = \sum_j \sum_{k \neq j} \overline{M_W^{ij} M_W^{ik} M_{ZW}^{ik}} e_j e_k \tilde{u}_k,$$

whose variance

$$\sum_{j_1} \sum_{k_1 \neq j_1} \sum_{j_2} \sum_{k_2 \neq j_2} \overline{M_W^{ij_1} M_W^{ik_1} M_{ZW}^{ik_1} M_W^{ij_2} M_W^{ik_2} M_{ZW}^{ik_2}} E\left[e_{j_1} e_{k_1} e_{j_2} e_{k_2} \tilde{u}_{k_1} \tilde{u}_{k_2}'\right]$$

contains terms that are not zero, in the following cases:

(i) $j_1 = j_2$, $k_1 = k_2 \neq j_1$ leading to scaled $E\left[e_k^2 \tilde{u}_k \tilde{u}_k'\right] \sigma_e^2$, then the variance component is proportional to

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_k M_W^{i_1 k} M_{ZW}^{i_1 k} M_W^{i_2 k} M_{ZW}^{i_2 k} \sum_{j \neq k} M_W^{i_1 j} M_W^{i_2 j},$$

which in absolute value is no greater than

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_k \left|M_W^{i_1 k} M_{ZW}^{i_1 k} M_W^{i_2 k} M_{ZW}^{i_2 k}\right| \sum_{j \neq k} \left|M_W^{i_1 j} M_W^{i_2 j}\right|$$

$$\leq \frac{1}{n^2} \sum_k \sum_{i_1} \left|M_W^{i_1 k} M_{ZW}^{i_1 k}\right| \sum_{i_2} \left|M_W^{i_2 k} M_{ZW}^{i_2 k}\right| = O\left(\frac{1}{n}\right) = o(1);$$

(ii) $j_1 = k_2$, $j_2 = k_1 \neq j_1$ leading to scaled $E\left[e_j^2 \tilde{u}_j\right] E\left[e_j^2 \tilde{u}_j'\right]$, then the variance component is proportional to

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j M_W^{i_1 j} M_W^{i_2 j} M_{ZW}^{i_2 j} \sum_{k \neq j} M_W^{i_1 k} M_{ZW}^{i_1 k} M_W^{i_2 k},$$

which in absolute value is no greater than

$$\frac{1}{n^2} \sum_j \sum_{i_1} \left|M_W^{i_1 j}\right| \sum_{i_2} \left|M_W^{i_2 j} M_{ZW}^{i_2 j}\right| \sum_k \left|M_W^{i_1 k} M_W^{i_2 k}\right| \leq \frac{1}{n^2} \sum_j \sqrt{n} = o(1).$$

Third, take

$$T_3 = \sum_j \sum_{k \neq j} \sum_{l \notin \{j,k\}} \overline{M_W^{ij} M_W^{ik} M_{ZW}^{il}} e_j e_k \tilde{u}_l,$$

34

whose variance

$$\sum_{j_1} \sum_{k_1 \neq j_1} \sum_{l_1 \notin \{j_1, k_1\}} \sum_{j_2} \sum_{k_2 \neq j_2} \sum_{l_2 \notin \{j_2, k_2\}} \overline{M_W^{ij_1} M_W^{ik_1} M_{ZW}^{il_1}}$$
$$\overline{M_W^{ij_2} M_W^{ik_2} M_{ZW}^{il_2}} E \left[e_{j_1} e_{k_1} \tilde{u}_{l_1} e_{j_2} e_{k_2} \tilde{u}_{l_2} \right]$$

contains terms that are not zero only in the case $j_1 = j_2$, $k_1 = k_2 \neq j_1$, $l_1 = l_2 \notin \{j_1, k_1\}$ or similar leading to scaled $\sigma_e^4 E\left[\tilde{u}_l \tilde{u}_l'\right]$, then the variance component is proportional to

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j M_W^{i_1 j} M_W^{i_2 j} \sum_{k \neq j} M_W^{i_1 k} M_W^{i_2 k} \sum_{l \notin \{j,k\}} M_{ZW}^{i_1 l} M_{ZW}^{i_2 l},$$

which in absolute value is no greater than

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left| M_W^{i_1 j} \right| \left| M_W^{ji_2} \right| \sum_l \left| M_{ZW}^{i_1 l} M_{ZW}^{i_2 l} \right| \left(\left| M_W^{i_1 i_2} \right| + \left| M_W^{i_1 j} M_W^{i_2 j} \right| \right)$$
$$\leq \quad \frac{1}{n^2} \sum_{i_1} \sum_{i_2} \left| M_W^{i_1 i_2} \right| \sum_j \left| M_W^{i_1 j} M_W^{j i_2} \right| + \frac{1}{n^2} \sum_j \sum_{i_1} \left(M_W^{i_1 j} \right)^2 \sum_{i_2} \left(M_W^{i_2 j} \right)^2.$$

Hence the variance component is no greater than proportional to

$$\frac{1}{n^2} \sum_{i_1} \sqrt{n} + \frac{1}{n^2} \sum_j 1 = o\left(1\right).$$

Fourth, take

$$T_4 = \overline{\sum_j \left(M_W^{ij} \right)^2 M_{ZW}^{ij}} \left(e_j^2 \tilde{u}_j - E[e_j^2 \tilde{u}_j] \right),$$

whose variance is

$$\sum_{j_1} \sum_{j_2} \overline{\left(M_W^{ij_1} \right)^2 M_{ZW}^{ij_1} \left(M_W^{ij_2} \right)^2 M_{ZW}^{ij_2}} E\left[\left(e_{j_1}^2 \tilde{u}_{j_1} - E[e_j^2 \tilde{u}_j] \right) \left(e_{j_2}^2 \tilde{u}_{j_2} - E[e_j^2 \tilde{u}_j] \right) \right]$$

This is non-zero only when $j_1 = j_2$ leading to scaled $E\left[(e_j^2 \tilde{u}_j - E[e_j^2 \tilde{u}_j])(e_j^2 \tilde{u}_j - E[e_j^2 \tilde{u}_j])' \right]$, and proportional to

$$\frac{1}{n^2} \sum_j \sum_{i_1} \left(M_W^{i_1 j} \right)^2 M_{ZW}^{i_1 j} \sum_{i_2} \left(M_W^{i_2 j} \right)^2 M_{ZW}^{i_2 j},$$

which in absolute value is no greater than

$$\frac{1}{n^2} \sum_j \sum_{i_1} \left(M_W^{i_1 j} \right)^2 \sum_{i_2} \left(M_W^{i_2 j} \right)^2 \leq O\left(\frac{1}{n} \right) = o(1).$$

Part III. Observe that

$$\overline{(M_W e)_i^4} = \overline{\left(\sum_j M_W^{ij} e_j \right)^4} = \overline{\sum_j \left(M_W^{ij} \right)^4} E\left[e_i^4 - 3\sigma_e^4 \right] + 3\overline{\left(M_W^{ii} \right)^2} \sigma_e^4$$
$$+ T_1 + T_2 + T_3 + T_4 + T_5,$$

where the formulas for T_1, T_2, T_3, T_4 and T_5 appear below. We will show that $T_1 + T_2 + T_3 + T_4 + T_5 = O_P\left(1/n\right)$. First, take

$$\frac{T_1}{4} = \sum_j \sum_{k \neq j} \overline{\left(M_W^{ij} \right)^3 M_W^{ik}} e_j^3 e_k,$$

35

whose variance

$$\sum_{j_1} \sum_{k_1 \neq j_1} \sum_{j_2} \sum_{k_2 \neq j_2} \overline{\left(M_W^{ij_1}\right)^3 M_W^{ik_1} \left(M_W^{ij_2}\right)^3 M_W^{ik_2}} E\left[e_{j_1}^3 e_{k_1} e_{j_2}^3 e_{k_2}\right]$$

contains terms that are not zero, in the following cases:

(i) $j_1 = j_2$, $k_1 = k_2 \neq j_1$ leading to scaled $E[e_j^6]\sigma_e^2$, then the variance component is proportional to

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left(M_W^{i_1 j}\right)^3 \left(M_W^{i_2 j}\right)^3 \sum_{k \neq j} M_W^{i_1 k} M_W^{i_2 k},$$

which in absolute value is no greater than

$$
\begin{aligned}
\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left|M_W^{i_1 j}\right|^3 \left|M_W^{i_2 j}\right|^3 \sum_{k \neq j} \left|M_W^{i_1 k} M_W^{i_2 k}\right| &\leq \frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left(M_W^{i_1 j}\right)^2 \left(M_W^{i_2 j}\right)^2 \\
&= O\left(\frac{n-m}{n^2}\right) = o(1);
\end{aligned}
$$

(ii) $j_1 = k_2$, $j_2 = k_1 \neq j_1$ leading to scaled $E[e_j^4]^2$, then the variance component is proportional to

$$\frac{1}{n^2} \sum_j \sum_{k \neq j} \sum_{i_1} \sum_{i_2} \left(M_W^{i_1 j}\right)^3 M_W^{i_1 k} \left(M_W^{i_2 k}\right)^3 M_W^{i_2 j},$$

which in absolute value is no greater than

$$
\begin{aligned}
\frac{1}{n^2} \sum_j \sum_{i_1} \sum_{i_2} \left|M_W^{i_1 j}\right|^3 \left|M_W^{i_2 j}\right| \sum_k \left(M_W^{i_2 k}\right)^2 &\leq \frac{1}{n^2} \sum_j \sum_{i_2} \left|M_W^{i_2 j}\right| \sum_{i_1} \left(M_W^{i_1 j}\right)^2 \\
&\leq \frac{1}{n^2} \sum_j \sqrt{n} = o(1).
\end{aligned}
$$

(iii) $j_1 \neq j_2$, $k_1 = k_2 \notin \{j_1, j_2\}$ leading to scaled $E[e_j^3]^2 \sigma_e^2$, then the variance component is proportional to

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_{j_1} \sum_{j_2 \neq j_1} \sum_{k \notin \{j_1, j_2\}} \left(M_W^{i_1 j_1}\right)^3 M_W^{i_1 k} \left(M_W^{i_2 j_2}\right)^3 M_W^{i_2 k},$$

which in absolute value is no greater than

$$
\begin{aligned}
&\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_{j_1} \left|M_W^{i_1 j_1}\right|^3 \sum_{j_2 \neq j_1} \left|M_W^{i_2 j_2}\right|^3 \left(\left|M_W^{i_1 i_2}\right| + \left|M_W^{i_1 j_1} M_W^{i_2 j_1}\right| + \left|M_W^{i_1 j_2} M_W^{i_2 j_2}\right|\right) \\
&\leq \frac{1}{n^2} \sum_{i_1} \sum_{i_2} \left|M_W^{i_1 i_2}\right| \sum_{j_1} \left(M_W^{i_1 j_1}\right)^2 \sum_{j_2} \left(M_W^{i_2 j_2}\right)^2 \\
&\quad + \frac{1}{n^2} \sum_{i_1} \sum_{j_1} \left(M_W^{i_1 j_1}\right)^4 \sum_{i_2} \left|M_W^{i_2 j_1}\right| \sum_{j_2} \left(M_W^{i_2 j_2}\right)^2 \\
&\quad + \frac{1}{n^2} \sum_{i_2} \sum_{j_2} \left(M_W^{i_2 j_2}\right)^4 \sum_{i_1} \left|M_W^{i_1 j_2}\right| \sum_{j_1} \left(M_W^{i_1 j_1}\right)^2 \\
&\leq \frac{1}{n^2} \sum_{i_1} \sqrt{n} + \frac{n-m}{n^2} \sqrt{n} + \frac{n-m}{n^2} \sqrt{n} = o(1).
\end{aligned}
$$

Second, take

$$\frac{T_2}{6} = \sum_j \sum_{k \neq j} \sum_{l \notin \{j,k\}} \overline{\left(M_W^{ij}\right)^2 M_W^{ik} M_W^{il}} e_j^2 e_k e_l,$$

whose variance

$$\sum_{j_1} \sum_{k_1 \neq j_1} \sum_{l_1 \notin \{j_1,k_1\}} \sum_{j_2} \sum_{k_2 \neq j_2} \sum_{l_2 \notin \{j_2,k_2\}} \overline{\left(M_W^{ij_1}\right)^2 M_W^{ik_1} M_W^{il_1}} \\ \left(M_W^{ij_2}\right)^2 M_W^{ik_2} M_W^{il_2} E\left[e_{j_1}^2 e_{k_1} e_{l_1} e_{j_2}^2 e_{k_2} e_{l_2}\right]$$

contains terms that are not zero, in the following cases:

(i) $j_1 = j_2$, $k_1 = k_2 \neq j_1$, $l_1 = l_2 \notin \{j_1, k_1\}$ or similar leading to scaled $E[e_j^4]\sigma_e^4$, then the variance component is proportional to

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left(M_W^{i_1j}\right)^2 \left(M_W^{i_2j}\right)^2 \sum_{k \neq j} M_W^{i_1k} M_W^{i_2k} \sum_{l \notin \{j,k\}} M_W^{i_1l} M_W^{i_2l},$$

which in absolute value is no greater than

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left(M_W^{i_1j}\right)^2 \left(M_W^{ji_2}\right)^2 \sum_{k \neq j} \left| M_W^{i_1k} M_W^{i_2k} \right|.$$

Hence the variance component is proportional to

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left(M_W^{i_1j}\right)^2 \left(M_W^{ji_2}\right)^2 = O\left(\frac{n-m}{n^2}\right) = o(1).$$

(ii) $j_1 \neq j_2$, $k_1 = k_2 \notin \{j_1, j_2\}$, $l_1 = l_2 \notin \{j_1, j_2, k_1\}$ or similar leading to scaled σ_e^8, then the variance component is proportional to

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_{j_1} \left(M_W^{i_1j_1}\right)^2 \sum_{j_2} \left(M_W^{i_2j_2}\right)^2 \sum_{k \neq j_1,j_2} M_W^{i_1k} M_W^{i_2k} \sum_{l \notin \{j_1,j_2,k\}} M_W^{i_1l} M_W^{i_2l},$$

which is no greater in absolute value than

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_{j_1} \left(M_W^{i_1j_1}\right)^2 \sum_{j_2} \left(M_W^{i_2j_2}\right)^2 \\ \sum_{k \neq j_1,j_2} \left| M_W^{i_1k} M_W^{i_2k} \right| \left(\left| M_W^{i_1i_2} \right| + 2 \left| M_W^{i_1j_1} M_W^{i_2j_1} \right| + \left| M_W^{i_1k} M_W^{i_2k} \right| \right),$$

because of symmetry between j_1 and j_2. Consider the three terms in turn. The first is

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \left| M_W^{i_1i_2} \right| \sum_{j_1} \left(M_W^{i_1j_1}\right)^2 \sum_{j_2} \left(M_W^{i_2j_2}\right)^2 \sum_{k \neq j_1,j_2} \left| M_W^{i_1k} M_W^{i_2k} \right| \leq \frac{1}{n^2} \sum_{i_1} \sqrt{n} \leq o(1),$$

the second does not exceed

$$\frac{2}{n^2} \sum_{i_1} \sum_{i_2} \sum_{j_1} \left| M_W^{i_1j_1} \right|^3 \sum_{j_2} \left(M_W^{i_2j_2}\right)^2 \left| M_W^{i_2j_1} \right| \sum_{k \neq j_1,j_2} \left| M_W^{i_1k} M_W^{i_2k} \right|$$
$$\leq \frac{2}{n^2} \sum_{i_2} \sum_{j_1} \sum_{j_2} \left(M_W^{i_2j_2}\right)^2 \left| M_W^{i_2j_1} \right| \sum_{i_1} \left(M_W^{i_1j_1}\right)^2$$
$$\leq \frac{2}{n^2} \sum_{i_2} \sum_{j_2} \left(M_W^{i_2j_2}\right)^2 \sum_{j_1} \left| M_W^{i_2j_1} \right| \leq \frac{2\sqrt{n}}{n^2} \sum_{i_2} \sum_{j_2} \left(M_W^{i_2j_2}\right)^2 = o(1),$$

and the third is

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_{j_1} \left(M_W^{i_1j_1}\right)^2 \sum_{j_2} \left(M_W^{i_2j_2}\right)^2 \sum_{k \notin \{j_1,j_2\}} \left(M_W^{i_1k} M_W^{i_2k}\right)^2$$
$$\leq \frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_k \left(M_W^{i_1k}\right)^2 \left(M_W^{i_2k}\right)^2 = O\left(\frac{n-m}{n^2}\right) = o(1).$$

(iii) $j_1 = k_2 \neq j_2$, $j_2 = k_1 \neq j_1$, $l_1 = l_2 \notin \{j_1, j_2\}$ or similar leading to scaled $E[e_j^3]^2 \sigma_e^2$, then the variance component is proportional to

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left(M_W^{i_1j}\right)^2 M_W^{i_2j} \sum_{k \neq j} M_W^{i_1k} \left(M_W^{i_2k}\right)^2 \sum_{l \notin \{j,k\}} M_W^{i_1l} M_W^{i_2l},$$

which is no greater in absolute value than

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left(M_W^{i_1j}\right)^2 \left|M_W^{i_2j}\right| \sum_{k \neq j} \left|M_W^{i_1k}\right| \left(M_W^{i_2k}\right)^2 \left(\left|M_W^{i_1i_2}\right| + \left|M_W^{i_1j} M_W^{i_2j}\right| + \left|M_W^{i_1k} M_W^{i_2k}\right|\right)$$

$$\leq \frac{1}{n^2} \sum_{i_1} \sum_{i_2} \left|M_W^{i_1i_2}\right| \sum_j \left(M_W^{i_1j}\right)^2 \sum_k \left(M_W^{i_2k}\right)^2$$

$$+ \frac{1}{n^2} \sum_{i_1} \sum_j \left|M_W^{i_1j}\right|^3 \sum_{i_2} \left(M_W^{i_2j}\right)^2 \sum_k \left|M_W^{i_1k}\right| + \frac{1}{n^2} \sum_{i_2} \sum_k \left|M_W^{i_2k}\right|^3 \sum_j \left|M_W^{i_2j}\right| \sum_{i_1} \left(M_W^{i_1k}\right)^2$$

$$\leq \frac{1}{n^2} \sum_{i_1} \sqrt{n} + \frac{n}{n^2} \sqrt{n} + \frac{n}{n^2} \sqrt{n} = o(1).$$

Third, take

$$T_3 = \sum_j \sum_{k \neq j} \sum_{l \notin \{k,j\}} \sum_{q \notin \{j,k,l\}} \overline{M_W^{ij} M_W^{ik} M_W^{il} M_W^{iq}} e_j e_k e_l e_q,$$

whose variance

$$\sum_{j_1} \sum_{k_1 \neq j_1} \sum_{l_1 \notin \{k_1,j_1\}} \sum_{q_1 \notin \{j_1,k_1,l_1\}} \sum_{j_2} \sum_{k_2 \neq j_2} \sum_{l_2 \notin \{k_2,j_2\}} \sum_{q_1 \notin \{j_2,k_2,l_2\}}$$
$$\overline{M_W^{ij_1} M_W^{ik_1} M_W^{il_1} M_W^{iq_1} M_W^{ij_2} M_W^{ik_2} M_W^{il_2} M_W^{iq_2}} E\left[e_{j_1} e_{k_1} e_{l_1} e_{q_1} e_{j_2} e_{k_2} e_{l_2} e_{q_2}\right]$$

contains terms that are not zero only in the case $j_1 = j_2$, $k_1 = k_2$, $l_1 = l_2$, $q_1 = q_2$ and similar leading to scaled σ_e^8, so the variance component is proportional to

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j M_W^{i_1j} M_W^{i_2j} \sum_{k \neq j} M_W^{i_1k} M_W^{i_2k} \sum_{l \notin \{k,j\}} M_W^{i_1l} M_W^{i_2l} \sum_{q \notin \{j,k,l\}} M_W^{i_1q} M_W^{i_2q},$$

which in absolute value is no greater than

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left|M_W^{i_1j} M_W^{i_2j}\right| \left(\left|M_W^{i_1i_2}\right| + 3\left|M_W^{i_1j} M_W^{i_2j}\right|\right)$$

$$\leq \frac{1}{n^2} \sum_{i_1} \sum_{i_2} \left|M_W^{i_1i_2}\right| + \frac{3}{n^2} \sum_j \sum_{i_1} \sum_{i_2} \left(M_W^{i_1j}\right)^2 \left(M_W^{i_2j}\right)^2 \leq O\left(\frac{1}{n}\right) = o(1),$$

because of and symmetry among j, k and l. Fourth, take

$$T_4 = \overline{\sum_j \left(M_W^{ij}\right)^4} \left(e_j^4 - E[e_j^4]\right),$$

whose variance is

$$\sum_{j_1} \sum_{j_2} \overline{\left(M_W^{ij_1}\right)^4 \left(M_W^{ij_2}\right)^4} E\left[\left(e_{j_1}^4 - E[e_j^4]\right)\left(e_{j_2}^4 - E[e_j^4]\right)\right]$$

This is non-zero only when $j_1 = j_2$ leading to scaled $E\left[(e_j^4 - E[e_j^4])^2\right]$, and proportional to

$$\frac{1}{n^2} \sum_j \sum_{i_1} \left(M_W^{i_1j}\right)^4 \sum_{i_2} \left(M_W^{i_2j}\right)^4,$$

which is no greater than

$$\frac{1}{n^2} \sum_j \sum_{i_1} \left(M_W^{i_1 j}\right)^2 \sum_{i_2} \left(M_W^{i_2 j}\right)^2 \le O\left(\frac{1}{n}\right) = o(1).$$

Fifth, take

$$\frac{T_5}{3} = \sum_j \sum_{k \neq j} \overline{\left(M_W^{ij}\right)^2 \left(M_W^{ik}\right)^2} \left(e_j^2 e_k^2 - \sigma_e^4\right),$$

whose variance is

$$\sum_{j_1} \sum_{k_1 \neq j_1} \sum_{j_2} \sum_{k_2 \neq j_2} \overline{\left(M_W^{ij_1}\right)^2 \left(M_W^{ik_1}\right)^2} \overline{\left(M_W^{ij_2}\right)^2 \left(M_W^{ik_2}\right)^2} E\left[\left(e_{j_1}^2 e_{k_1}^2 - \sigma_e^4\right)\left(e_{j_2}^2 e_{k_2}^2 - \sigma_e^4\right)\right].$$

It contains terms that are not zero only in the following cases:

(i) $j_1 = j_2$, $k_1 = k_2 \neq j_1$ or similar leading to scaled $E[e_j^4]^2 - \sigma_e^8$, then the variance component is proportional to

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left(M_W^{i_1 j}\right)^2 \left(M_W^{i_2 j}\right)^2 \sum_{k \neq j} \left(M_W^{i_1 k}\right)^2 \left(M_W^{i_2 k}\right)^2$$

$$\le \frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left(M_W^{i_1 j}\right)^2 \left(M_W^{i_2 j}\right)^2 \sum_k \left|M_W^{i_1 k} M_W^{i_2 k}\right| \le O\left(\frac{n-m}{n^2}\right) = o(1),$$

(ii) $j_1 = j_2$, $k_2 \notin \{k_1, j_1\}$ or similar leading to scaled $\left(E[e_j^4] - \sigma_e^4\right)\sigma_e^4$, then the variance component is proportional to

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j \left(M_W^{i_1 j}\right)^2 \left(M_W^{i_2 j}\right)^2 \sum_{k_1 \neq j} \left(M_W^{i_1 k_1}\right)^2 \sum_{k_2 \notin \{k_1, j\}} \left(M_W^{i_2 k_2}\right)^2 \le O\left(\frac{n-m}{n^2}\right) = o(1).$$

Lemma A8: The following is true:

$$
\begin{aligned}
E\left[\tilde{u}_i \tilde{u}_i'\right] &= \Sigma_u - \sigma_e^2 \Gamma' \Gamma, \\
E\left[e_i^2 \tilde{u}_i\right] &= E\left[e_i^2 u_i\right] - E\left[e_i^3\right] \Gamma, \\
E\left[e_i^2 \tilde{u}_i \tilde{u}_i'\right] &= E\left[e_i^2 u_i u_i'\right] - \Gamma' E\left[e_i^2 u_i'\right] - E\left[e_i^2 u_i\right] \Gamma + E\left[e_i^4\right] \Gamma' \Gamma.
\end{aligned}
$$

Proof of Lemma A8. By straightforward computation.

Proof of Theorem 4. Because $M_{ZW} \hat{e}_{B2SLS} = M_{ZW} e - M_{ZW} X \left(\hat{\beta}_{B2SLS} - \beta_0\right)$, we have $n^{-1} \hat{e}_{B2SLS}' M_{ZW} \hat{e}_{B2SLS} \xrightarrow{p} (1 - \mu - \lambda) \sigma_e^2$ and $n^{-1} X' M_{ZW} \hat{e}_{B2SLS} \xrightarrow{p} (1 - \mu - \lambda) \Sigma_u$ using Lemma A2 and consistency of $\hat{\beta}_{B2SLS}$. Thus, again using Lemma A2,

$$
\begin{aligned}
\frac{\hat{D}_2}{n^2} &= \frac{\hat{e}_{B2SLS}' M_{ZW} \hat{e}_{B2SLS}}{n} \frac{X' M_{ZW} X}{n} + \frac{X' M_{ZW} \hat{e}_{B2SLS}}{n} \frac{\hat{e}_{B2SLS}' M_{ZW} X}{n} \\
&\xrightarrow{p} (1 - \mu - \lambda)^2 \left(\sigma_e^2 \Sigma_u + \Sigma_{ue} \Sigma_{ue}'\right).
\end{aligned}
$$

Hence, also using Lemma A3,

$$\hat{\Phi}_2^{B2SLS} \xrightarrow{p} (1-\alpha)\,\sigma_e^2\left((1-\alpha)\,Q_{XZ^\perp}\right)^{-1} + \frac{1}{1-\mu}\frac{\alpha}{1-\alpha}\left((1-\alpha)\,Q_{XZ^\perp}\right)^{-1}$$
$$\times\left((1-\mu-\lambda)^2\left(\sigma_e^2\Sigma_u + \Sigma_{ue}\Sigma_{ue}'\right)\right)\left((1-\alpha)\,Q_{XZ^\perp}\right)^{-1})$$
$$= \sigma_e^2 Q_{XZ^\perp}^{-1} + \frac{\lambda}{1-\alpha}Q_{XZ^\perp}^{-1}\left(\sigma_e^2\Sigma_u + \Sigma_{ue}\Sigma_{ue}'\right)Q_{XZ^\perp}^{-1} = \Phi_2^{B2SLS}.$$

Next, using Lemma A3

$$\hat{\Gamma}_{B2SLS} = \frac{\hat{e}_{B2SLS}'X}{\hat{e}_{B2SLS}'\hat{e}_{B2SLS}} = \frac{e_{B2SLS}'M_W X}{e_{B2SLS}'M_W e_{B2SLS}} = \frac{\Sigma_{ue}'}{\sigma_e^2} + o_P(1),$$

so $\tilde{X} = X - \hat{e}_{B2SLS}\Gamma + o_P(1)$. Note that

$$\overline{(M_{ZW}\tilde{X})_i(M_{ZW}\tilde{X})_i'} = \frac{\tilde{X}'M_{ZW}\tilde{X}}{n} \xrightarrow{p} (1-\mu-\lambda)\,E\left[\tilde{u}_i\tilde{u}_i'\right].$$

By Lemma A7(I,II),

$$\overline{(\hat{e}_{B2SLS})_i^3} = \overline{(M_W e)_i^3} + o_P(1) = \overline{\sum_j\left(M_W^{ij}\right)^3 E\left[e_i^3\right]} + o_P(1)$$

and, using that $M_{ZW}\tilde{X} = M_{ZW}\tilde{U} + o_P(1)$,

$$\overline{(\hat{e}_{B2SLS})_i^2\,(M_{ZW}\tilde{X})_i} = \overline{(M_W e)_i^2\,(M_{ZW}\tilde{U})_i} + o_P(1) = \overline{\sum_j\left(M_W^{ij}\right)^2 M_{ZW}^{ij}E\left[e_i^2\tilde{u}_i\right]} + o_P(1),$$

we have using Lemma A8

$$\hat{D}_3 = \frac{\overline{\sum_j\left(M_W^{ij}\right)^3 E\left[e_i^3\right]}}{\overline{\sum_j\left(M_W^{ij}\right)^3}}\Gamma' + \frac{\overline{\sum_j\left(M_W^{ij}\right)^2 M_{ZW}^{ij}E\left[e_i^2\tilde{u}_i\right]}}{\overline{\sum_j\left(M_W^{ij}\right)^2 M_{ZW}^{ij}}} + o_P(1) \xrightarrow{p} E\left[e_i^2 u_i\right].$$

Then

$$\hat{\Phi}_3^{B2SLS} \xrightarrow{p} (1-\alpha)\left((1-\alpha)\,Q_{XZ^\perp}\right)^{-1}\left(E\left[e_i^2 u_i\right]\pi_\alpha' + \pi_\alpha E\left[e_i^2 u_i'\right]\right)\left((1-\alpha)\,Q_{XZ^\perp}\right)^{-1} = \Phi_3^{B2SLS}.$$

Next, by Lemma A7(III,IV),

$$\overline{(\hat{e}_{B2SLS})_i^4} = \overline{(M_W e)_i^4} + o_P(1) = \overline{\sum_j\left(M_W^{ij}\right)^4 E\left[e_i^4 - 3\sigma_e^4\right]} + 3\overline{(M_W^{ii})^2}\sigma_e^4 + o_P(1)$$

and

$$\overline{(\hat{e}_{B2SLS})_i^2\,(M_{ZW}\tilde{X})_i(M_{ZW}\tilde{X})_i'} = \overline{(M_W e)_i^2\,(M_{ZW}\tilde{U})_i(M_{ZW}\tilde{U})_i'} + o_P(1)$$
$$= \overline{\sum_j\left(M_W^{ij}M_{ZW}^{ij}\right)^2 E\left[\left(e_i^2 - \sigma_e^2\right)\tilde{u}_i\tilde{u}_i'\right]}$$
$$+ \overline{M_W^{ii}M_{ZW}^{ii}}\sigma_e^2 E\left[\tilde{u}_i\tilde{u}_i'\right] + o_P(1).$$

40

Then, using also Lemma A8,

$$
\begin{aligned}
\hat{D}_4 &= \left(E\left[e_i^4 - 3\sigma_e^4 \right] + \frac{3\overline{(M_W^{ii})^2}\sigma_e^4}{\sum_j \left(M_W^{ij}\right)^4} - 3\frac{\overline{(M_W^{ii})^2}}{\sum_j \left(M_W^{ij}\right)^4}\sigma_e^4 \right) \Gamma'\Gamma \\
&\quad + \frac{1}{\overline{\sum_j \left(M_W^{ij}\right)^2 M_{ZW}^{ij}}} \left(\Gamma'\overline{\sum_j \left(M_W^{ij}\right)^2 M_{ZW}^{ij}} E\left[e_i^2 \tilde{u}_i' \right] + \overline{\sum_j \left(M_W^{ij}\right)^2 M_{ZW}^{ij}} E\left[e_i^2 \tilde{u}_i \right] \Gamma \right) \\
&\quad + \left(E\left[\left(e_i^2 - \sigma_e^2 \right) \tilde{u}_i \tilde{u}_i' \right] + \frac{\overline{M_W^{ii} M_{ZW}^{ii}}}{\sum_j \left(M_W^{ij} M_{ZW}^{ij}\right)^2}\sigma_e^2 E\left[\tilde{u}_i \tilde{u}_i' \right] \right) \\
&\quad - \frac{1}{\overline{\sum_j \left(M_W^{ij} M_{ZW}^{ij}\right)^2}}\frac{\overline{M_W^{ii} M_{ZW}^{ii}}}{1 - \mu - \lambda}\sigma_e^2 \left(1 - \mu - \lambda \right) E\left[\tilde{u}_i \tilde{u}_i' \right] + o_P(1) \\
&= E\left[e_i^4 \right] \Gamma'\Gamma - 3\sigma_e^4 \Gamma'\Gamma + \left(\Gamma' E\left[e_i^2 \tilde{u}_i' \right] + E\left[e_i^2 \tilde{u}_i \right] \Gamma \right) + \left(E\left[\left(e_i^2 - \sigma_e^2 \right) \tilde{u}_i \tilde{u}_i' \right] \right) + o_P(1) \\
&= E\left[e_i^2 u_i u_i' \right] - \sigma_e^2 \left(\Sigma_u - \sigma_e^2 \Gamma'\Gamma \right) - 3\sigma_e^4 \Gamma'\Gamma = E\left[e_i^2 u_i u_i' \right] - \sigma_e^2 \Sigma_u - 2\Sigma_{ue}\Sigma_{ue}',
\end{aligned}
$$

and

$$
\hat{\Phi}_4^{B2SLS} \overset{p}{\to} \rho_\alpha \left((1-\alpha) Q_{XZ^\perp} \right)^{-1} \left(E\left[e_i^2 u_i u_i' \right] - \sigma_e^2 \Sigma_u - 2\Sigma_{ue}\Sigma_{ue}' \right) \left((1-\alpha) Q_{XZ^\perp} \right)^{-1} = \Phi_4^{B2SLS}.
$$

Proof of Theorem 5. Because $M_W \left(P_{ZW} - \hat{\alpha} I_n \right) M_W = M_W P_{\hat{\alpha}} = P_{\hat{\alpha}}$ and using Lemma A3, we have

$$
\begin{aligned}
\frac{\hat{e}_{B2SLS}' \left(P_{ZW} - \hat{\alpha} I_n \right) \hat{e}_{B2SLS}}{\sqrt{n}} &= \frac{e' P_\alpha e}{\sqrt{n}} - \frac{1}{\sqrt{n}} \left(\frac{X' P_\alpha e}{\sqrt{n}} \right)' \left(\frac{X' P_\alpha X}{n} \right)^{-1} \frac{X' P_\alpha e}{\sqrt{n}} + o_P(1) \\
&= \frac{e' P_\alpha e}{\sqrt{n}} + o_P(1).
\end{aligned}
$$

Using asymptotic normality of $e' P_\alpha e / \sqrt{n}$ from Lemma A5 with $B = (0_p', 1)'$, we get that the above quantity is asymptotically normal with the asymptotic variance $2\lambda \left(1 - \alpha \right) \sigma_e^4 + \rho_\alpha \left(E\left[e_i^4 \right] - 3\sigma_e^4 \right)$.

As $\hat{\sigma}_{B2SLS}^2 \overset{p}{\to} \sigma_e^2$, we summarize that

$$
\begin{aligned}
\frac{J_{B2SLS}}{\sqrt{n}} &= \frac{\hat{e}_{B2SLS}' \left(P_{ZW} - \hat{\alpha} I_n \right) \hat{e}_{B2SLS}}{\sqrt{n}\hat{\sigma}_{B2SLS}^2} \overset{d}{\to} \frac{N\left(0, 2\lambda \left(1 - \alpha \right) \sigma_e^4 + \rho_\alpha \left(E\left[e_i^4 \right] - 3\sigma_e^4 \right) \right)}{\sigma_e^2} \\
&\overset{d}{=} N\left(0, 2\lambda \left(1 - \alpha \right) + \rho_\alpha \left(\frac{E\left[e_i^4 \right]}{\sigma_e^4} - 3 \right) \right).
\end{aligned}
$$

Proof of Corollary to Theorem 5. The first term clearly converges to $2\lambda \left(1 - \alpha \right)$. In the second term, $\hat{\rho}_\alpha$ is consistent for ρ_α by Lemma A6, and $\hat{\sigma}_{B2SLS}^4$ is consistent for σ_e^4.

Next, by Lemma A7(III),

$$\frac{1}{\overline{\sum_j \left(M_W^{ij}\right)^4}} \left(\frac{\overline{(\hat{e}_{B2SLS})_i^4}}{\hat{\sigma}_{B2SLS}^4} - 3\overline{(M_W^{ii})^2} \right) \xrightarrow{p} \frac{E\left[e_i^4\right]}{\sigma_e^4} - 3.$$

Summarizing, $\hat{V}_{B2SLS}^J \xrightarrow{p} V_{B2SLS}^J$.

Proof of Theorem 6. Consider the normalized (by n^{-1}) numerator and denominator of $F(\beta)$. Using Lemma A2, $n^{-1}\left(e - X(\beta - \beta_0)\right)' P_{Z^\perp}\left(e - X(\beta - \beta_0)\right)$ equals

$$\frac{e'P_{Z^\perp}e}{n} - (\beta - \beta_0)'\frac{X'P_{Z^\perp}e}{n} - \frac{e'P_{Z^\perp}X}{n}(\beta - \beta_0) + (\beta - \beta_0)'\frac{X'P_{Z^\perp}X}{n}(\beta - \beta_0)$$
$$\xrightarrow{p} \lambda\sigma_e^2 - \lambda(\beta - \beta_0)'\Sigma_{ue} - \lambda\Sigma_{ue}'(\beta - \beta_0) + (\beta - \beta_0)'\left(Q_{XZ^\perp} + \lambda\Sigma_u\right)(\beta - \beta_0),$$

while $n^{-1}\left(e - X(\beta - \beta_0)\right)' M_W\left(e - X(\beta - \beta_0)\right)$ equals

$$\frac{e'M_W e}{n} - (\beta - \beta_0)'\frac{X'M_W e}{n} - \frac{e'M_W X}{n}(\beta - \beta_0) + (\beta - \beta_0)'\frac{X'M_W X}{n}(\beta - \beta_0)$$
$$\xrightarrow{p} (1 - \mu)\sigma_e^2 - (1 - \mu)(\beta - \beta_0)'\Sigma_{ue} - (1 - \mu)\Sigma_{ue}'(\beta - \beta_0)$$
$$+(\beta - \beta_0)'\left(Q_{XZ^\perp} + (1 - \mu)\Sigma_u\right)(\beta - \beta_0).$$

To summarize, the probability limit of the objective function is

$$\frac{1}{1 - \mu}\left(\lambda + \frac{(1 - \alpha)(\beta - \beta_0)'Q_{XZ^\perp}(\beta - \beta_0)}{\sigma_e^2 - (\beta - \beta_0)'\Sigma_{ue} - \Sigma_{ue}'(\beta - \beta_0) + (\beta - \beta_0)'\left((1 - \mu)^{-1}Q_{XZ^\perp} + \Sigma_u\right)(\beta - \beta_0)}\right).$$

Indeed, it is minimized when $\beta = \beta_0$. The formal proof of consistency is standard and follows, for example, Hansen, Hausman and Newey (2008).

The standard first order asymptotic stochastic expansion leads to

$$\sqrt{n}\left(\hat{\beta}_{LIML} - \beta_0\right) = -\left(\frac{\partial^2 F(\beta_0)}{\partial\beta\partial\beta'}\right)^{-1}\sqrt{n}\frac{\partial F(\beta_0)}{\partial\beta} + o_P(1).$$

The first derivative of $F(\beta)$ evaluated at β_0 is

$$\frac{\partial F(\beta_0)}{\partial\beta} = \frac{\partial}{\partial\beta}\left.\frac{(Y - X\beta)'P_{Z^\perp}(Y - X\beta)}{(Y - X\beta)M_W(Y - X\beta)}\right|_{\beta=\beta_0} = -2\frac{X'P_{Z^\perp}e}{e'M_W e} + 2\frac{e'P_{Z^\perp}e}{(e'M_W e)^2}X'M_W e.$$

Note that

$$\frac{X'P_{Z^\perp}e}{e'M_W e} - \frac{e'P_{Z^\perp}e}{(e'M_W e)^2}X'M_W e$$
$$= \frac{1}{(e'M_W e/n)^2}\left(\frac{X'P_{Z^\perp}e \cdot e'M_W e}{n^2} - \frac{e'P_{Z^\perp}e \cdot X'M_W e}{n^2}\right)$$
$$= \frac{1}{(e'M_W e/n)^2}\left(\frac{X'P_\alpha e}{n}\frac{e'M_W e}{n} - \frac{e'P_\alpha e}{n}\frac{X'M_W e}{n}\right),$$

so, using Lemma A3,

$$
\begin{aligned}
\sqrt{n}\frac{\partial F\left(\beta_0\right)}{\partial \beta} &= -\frac{2}{\left(\left(1-\mu\right)\sigma_e^2\right)^2}\left(\frac{e'M_W e}{n}\frac{X'P_\alpha e}{\sqrt{n}} - \frac{X'M_W e}{n}\frac{e'P_\alpha e}{\sqrt{n}}\right) + o_P\left(1\right) \\
&= -\frac{2}{\left(1-\mu\right)\sigma_e^2}\frac{\left(X-e\Gamma\right)'P_\alpha e}{\sqrt{n}} + o_P\left(1\right),
\end{aligned}
$$

Next we apply Lemma A5 with $B = \begin{pmatrix} I_p \\ -\Gamma \end{pmatrix}$ to get that

$$
\frac{\left(X-e\Gamma\right)'P_\alpha e}{\sqrt{n}} \xrightarrow{d} N\left(0, \left(1-\alpha\right)\Xi_2 + \left(1-\alpha\right)\Xi_3 + \rho_\alpha\Xi_4\right),
$$

where, exploiting Lemma A8,

$$
\begin{aligned}
\Xi_2 &= \left(1-\alpha\right)\sigma_e^2 Q_{XZ^\perp} + \lambda\left(\sigma_e^2\Sigma_u - \Sigma_{ue}\Sigma_{ue}'\right), \\
\Xi_3 &= \pi_\alpha\left(E\left[e_i^2 u_i'\right] - E\left[e_i^3\right]\Gamma\right) + \left(E\left[e_i^2 u_i\right] - E\left[e_i^3\right]\Gamma'\right)\pi_\alpha' \\
&= \pi_\alpha E\left[e_i^2 \tilde{u}_i'\right] + E\left[e_i^2 \tilde{u}_i\right]\pi_\alpha', \\
\Xi_4 &= E\left[e_i^2 u_i u_i'\right] - E\left[e_i^3 u_i\right]\Gamma - \Gamma'E\left[e_i^3 u_i'\right] + \Gamma'\Gamma E\left[e_i^4\right] - \sigma_e^2\Sigma_u + \Sigma_{ue}\Sigma_{ue}' \\
&= E\left[\left(e_i^2 - \sigma_e^2\right)\tilde{u}_i\tilde{u}_i'\right].
\end{aligned}
$$

The second derivative of $F\left(\beta\right)$ is

$$
\begin{aligned}
\frac{\partial^2 F\left(\beta\right)}{\partial\beta\partial\beta'} &= \frac{\partial}{\partial\beta'}\left(-2\frac{X'P_{Z^\perp}\left(Y-X\beta\right)}{\left(Y-X\beta\right)'M_W\left(Y-X\beta\right)}\right) \\
&\quad + \frac{\partial}{\partial\beta'}\left(2\frac{\left(Y-X\beta\right)'P_{Z^\perp}\left(Y-X\beta\right)}{\left(\left(Y-X\beta\right)'M_W\left(Y-X\beta\right)\right)^2}X'M_W\left(Y-X\beta\right)\right) \\
&= 2\frac{X'P_{Z^\perp}X}{\left(Y-X\beta\right)'M_W\left(Y-X\beta\right)} \\
&\quad -8\frac{X'P_{Z^\perp}\left(Y-X\beta\right)}{\left(\left(Y-X\beta\right)'M_W\left(Y-X\beta\right)\right)^2}\left(Y-X\beta\right)'M_W X \\
&\quad +8\frac{\left(Y-X\beta\right)'P_{Z^\perp}\left(Y-X\beta\right)}{\left(\left(Y-X\beta\right)'M_W\left(Y-X\beta\right)\right)^3}X'M_W\left(Y-X\beta\right)\left(Y-X\beta\right)'M_W X \\
&\quad -2\frac{\left(Y-X\beta\right)'P_{Z^\perp}\left(Y-X\beta\right)}{\left(\left(Y-X\beta\right)'M_W\left(Y-X\beta\right)\right)^2}X'M_W X.
\end{aligned}
$$

When it is evaluated at β_0, the probability limit is, using Lemma A2,

$$
\begin{aligned}
&2\frac{Q_{XZ^\perp} + \lambda\Sigma_u}{\left(1-\mu\right)\sigma_e^2} - 8\frac{\lambda\Sigma_{ue}}{\left(\left(1-\mu\right)\sigma_e^2\right)^2}\left(1-\mu\right)\Sigma_{ue}' \\
&+8\frac{\lambda\sigma_e^2}{\left(\left(1-\mu\right)\sigma_e^2\right)^3}\left(1-\mu\right)\Sigma_{ue}\left(1-\mu\right)\Sigma_{ue}' - 2\frac{\lambda\sigma_e^2}{\left(\left(1-\mu\right)\sigma_e^2\right)^2}\left(Q_{XZ^\perp} + \left(1-\mu\right)\Sigma_u\right) \\
&= \frac{2\left(1-\alpha\right)}{1-\mu}\frac{Q_{XZ^\perp}}{\sigma_e^2}.
\end{aligned}
$$

Summarizing, the asymptotics for the LIML estimator is

$$\sqrt{n}\left(\hat{\beta}_{LIML} - \beta_0\right) \xrightarrow{d} \left(\frac{2(1-\alpha)}{1-\mu}\frac{Q_{XZ^\perp}}{\sigma_e^2}\right)^{-1}\frac{2}{(1-\mu)\sigma_e^2}N\left(0, (1-\alpha)\Xi_2 + \alpha(1-\alpha)\Xi_3 + \rho_\alpha\Xi_4\right)$$

$$\stackrel{d}{=} N\left(0, \Phi_2^{LIML} + \Phi_3^{LIML} + \Phi_4^{LIML}\right),$$

where Φ_2^{LIML}, Φ_3^{LIML} and Φ_4^{LIML} are as in the statement of the theorem.

Proof of Theorem 7. First, using that $\hat{e}_{LIML} = M_W e - M_W X\left(\hat{\beta}_{LIML} - \beta_0\right)$, we get using Lemma A2 that $\hat{\sigma}_{LIML}^2 \xrightarrow{p} \sigma_e^2$ and

$$\bar{\alpha} = \frac{\hat{e}'_{LIML}P_{ZW}\hat{e}_{LIML}}{\hat{e}'_{LIML}\hat{e}_{LIML}} = \frac{n^{-1}e'P_{Z^\perp}e + o_P(1)}{n^{-1}e'M_W e + o_P(1)} \xrightarrow{p} \frac{\lambda\sigma_e^2}{(1-\mu)\sigma_e^2} = \alpha.$$

Second, by the partitioned matrix inverse, using that $(P_{ZW} - \bar{\alpha}I_n)W = (1-\bar{\alpha})W$ and letting $X_W = (W'W)^{-1}W'X$,

$$\bar{H}^{-1} = \begin{bmatrix} (X'P_{\bar{\alpha}}X)^{-1} & -(X'P_{\bar{\alpha}}X)^{-1}X'_W \\ -X_W(X'P_{\bar{\alpha}}X)^{-1} & (1-\bar{\alpha})^{-1}(W'W)^{-1} + X_W(X'P_{\bar{\alpha}}X)^{-1}X'_W \end{bmatrix}.$$

Using Lemma A3,

$$\left(\frac{\bar{H}}{n}\right)^{-1} = (1-\alpha)^{-1}\begin{bmatrix} Q_{XZ^\perp}^{-1} & -Q_{XZ^\perp}^{-1}X'_W \\ -X_W Q_{XZ^\perp}^{-1} & n(W'W)^{-1} + X_W Q_{XZ^\perp}^{-1}X'_W \end{bmatrix} + o_P(1).$$

Next,

$$\begin{aligned}
\frac{\bar{X}'P_{ZW}\bar{X}}{n} &= \frac{(X,W)'P_{ZW}(X,W)}{n} + \frac{n^{-1}\hat{e}'_{LIML}P_{ZW}\hat{e}_{LIML}}{(n^{-1}\hat{e}'_{LIML}\hat{e}_{LIML})^2}\frac{(X,W)'\hat{e}_{LIML}}{n}\frac{\hat{e}'_{LIML}(X,W)}{n} \\
&\quad - \frac{(X,W)'P_{ZW}\hat{e}_{LIML}}{n}\frac{n^{-1}\hat{e}'_{LIML}(X,W)}{n^{-1}\hat{e}'_{LIML}\hat{e}_{LIML}} - \frac{n^{-1}(X,W)'\hat{e}_{LIML}}{n^{-1}\hat{e}'_{LIML}\hat{e}_{LIML}}\frac{\hat{e}'_{LIML}P_{ZW}(X,W)}{n} \\
&= \frac{(X,W)'(P_{ZW}X,W)}{n} + \frac{n^{-1}e'P_{Z^\perp}e}{(n^{-1}e'M_W e)^2}\frac{(e'M_W X,0)'}{n}\frac{(e'M_W X,0)}{n} \\
&\quad - \frac{(X'P_{Z^\perp}e,0)'}{n}\frac{n^{-1}(e'M_W X,0)}{n^{-1}e'M_W e} - \frac{n^{-1}(e'M_W X,0)'}{n^{-1}e'M_W e}\frac{(X'P_{Z^\perp}e,0)}{n} + o_P(1) \\
&= \frac{(X,W)'(P_{ZW}X,W)}{n} + \frac{\lambda\sigma_e^2}{((1-\mu)\sigma_e^2)^2}((1-\mu)\Sigma'_{ue},0)'((1-\mu)\Sigma'_{ue},0)(\lambda\Sigma'_{ue},0) \\
&\quad - (\lambda\Sigma'_{ue},0)'\frac{((1-\mu)\Sigma'_{ue},0)}{(1-\mu)\sigma_e^2} - \frac{((1-\mu)\Sigma'_{ue},0)'}{(1-\mu)\sigma_e^2} + o_P(1) \\
&= \frac{(X,W)'(P_{ZW}X,W)}{n} - \lambda\frac{\text{diag}\{\Sigma_{ue}\Sigma'_{ue}, 0_{m\times m}\}}{\sigma_e^2}
\end{aligned}$$

and similarly

$$\frac{\bar{X}'M_{ZW}\bar{X}}{n} = \frac{(X,0)'(M_{ZW}X,0)}{n} - (1-\mu-\lambda)\frac{\text{diag}\{\Sigma_{ue}\Sigma'_{ue}, 0_{m\times m}\}}{\sigma_e^2} + o_P(1).$$

Further, using Lemma A2,

$$\frac{(1-\bar{\alpha})^2 \, \bar{X}' P_{ZW} \bar{X} + \bar{\alpha}^2 \bar{X}' M_{ZW} \bar{X}}{n}$$

$$= (1-\bar{\alpha})^2 \frac{(X,W)' (P_{ZW} X, W)}{n} + \bar{\alpha}^2 \frac{(X,0)' (M_{ZW} X, 0)}{n}$$

$$- \left((1-\bar{\alpha})^2 \lambda + \bar{\alpha}^2 (1-\mu-\lambda) \right) \frac{\operatorname{diag}\{\Sigma_{ue} \Sigma'_{ue}, 0_{m\times m}\}}{\sigma_e^2} + o_P(1).$$

After multiplication of the sandwich, the northwest block becomes

$$\hat{V}_{LIML} = \left[\left(\frac{\bar{H}}{n} \right)^{-1} \left(\hat{\sigma}_{LIML}^2 \frac{(1-\bar{\alpha})^2 \, \bar{X}' P_{ZW} \bar{X} + \bar{\alpha}^2 \bar{X}' M_{ZW} \bar{X}}{n} + \frac{\hat{A}}{n} + \frac{\hat{A}'}{n} + \frac{\hat{B}}{n} \right) \left(\frac{\bar{H}}{n} \right)^{-1} \right]_{1:p,1:p}$$

$$= Q_{XZ^\perp}^{-1} \left(\frac{\sigma_e^2 (1-\alpha)^2 (Q_{XZ^\perp} + \lambda \Sigma_u) + \sigma_e^2 \alpha^2 (1-\mu-\lambda) \Sigma_u}{(1-\alpha)^2} \right.$$

$$\left. - \frac{\Sigma_{ue} \Sigma'_{ue} \left((1-\alpha)^2 \lambda + \alpha^2 (1-\mu-\lambda) \right)}{(1-\alpha)^2} + \frac{\hat{A}_{1:p,1:p} + \hat{A}'_{1:p,1:p} + \hat{B}_{1:p,1:p}}{n (1-\alpha)^2} \right) Q_{XZ^\perp}^{-1} + o_P(1)$$

$$= \bar{V}_2 + \bar{V}_3 + \bar{V}_4 + o_P(1),$$

where

$$\bar{V}_2 = \sigma_e^2 Q_{XZ^\perp}^{-1} + \frac{\lambda}{1-\alpha} Q_{XZ^\perp}^{-1} \left(\sigma_e^2 \Sigma_u - \Sigma_{ue} \Sigma'_{ue} \right) Q_{XZ^\perp}^{-1} = \Phi_2^{LIML},$$

$$\bar{V}_3 = Q_{XZ^\perp}^{-1} \frac{\hat{A}_{1:p,1:p} + \hat{A}'_{1:p,1:p}}{n (1-\alpha)^2} Q_{XZ^\perp}^{-1}, \quad \bar{V}_4 = Q_{XZ^\perp}^{-1} \frac{\hat{B}_{1:p,1:p}}{n (1-\alpha)^2} Q_{XZ^\perp}^{-1}.$$

We need to find the limits of $\bar{V}_3$ and $\bar{V}_4$. Taking into account that $M_{ZW} X = M_{ZW} U$,

$$M_{ZW} \bar{X} = (M_{ZW} X, M_{ZW} W) - M_{ZW} \hat{e}_{LIML} \bar{\Gamma}_{LIML}$$

$$= (M_{ZW} X, 0) - M_{ZW} e \frac{e' (M_W X, 0)}{\hat{e}'_{LIML} \hat{e}_{LIML}}$$

$$+ M_{ZW} e \left(X' M_W X, 0 \right) \cdot o_P(1) + (M_{ZW} X) \hat{e}'_{LIML} (X, W) \cdot o_P(1)$$

$$= \left(M_{ZW} \tilde{U}, 0 \right) + o_P(1),$$

so $\left(M_{ZW} \bar{X} \right)_i = \left((M_{ZW} \tilde{U})'_i, 0'_{m\times 1} \right)' + o_P(1)$ and

$$\left(M_{ZW} \bar{X} \right)_i \left(M_{ZW} \bar{X} \right)'_i = \operatorname{diag} \left\{ (M_{ZW} \tilde{U})_i (M_{ZW} \tilde{U})'_i, 0_{m\times m} \right\} + o_P(1).$$

Using Lemmas A6 and A7(II),

$$\overline{(\hat{e}_{LIML})_i^2 \left(M_{ZW} \bar{X} \right)_i} = \left(\overline{(M_W e)_i^2 (M_{ZW} \tilde{U})_i}, 0'_{m\times 1} \right)' + o_P(1)$$

$$= \sum_j \left(M_W^{ij} \right)^2 M_{ZW}^{ij} \left(E\left[e_i^2 \tilde{u}'_i \right], 0'_{m\times 1} \right)' + o_P(1),$$

45

then

$$\frac{A_{1:p,1:p}}{n} = \frac{(1-\hat{\alpha})}{\sum_j \left(M_W^{ij}\right)^2 M_{ZW}^{ij}} \left[\hat{\pi}_\alpha \overline{(\hat{e}_{LIML})_i^2 \left(M_{ZW}\bar{X}\right)_i'}\right]_{1:p,1:p} = (1-\alpha)\,\pi_\alpha E\left[e_i^2 \tilde{u}_i'\right] + o_P(1),$$

and so

$$\bar{V}_3 = Q_{XZ\perp}^{-1} \frac{\hat{A}_{1:p,1:p} + \hat{A}_{1:p,1:p}'}{n(1-\alpha)^2} Q_{XZ\perp}^{-1} \xrightarrow{p} Q_{XZ\perp}^{-1} \frac{\pi_\alpha E\left[e_i^2 \tilde{u}_i'\right] + E\left[e_i^2 \tilde{u}_i'\right]\pi_\alpha'}{1-\alpha} Q_{XZ\perp}^{-1} = \Phi_3^{LIML}.$$

Next,

$$\begin{aligned}
\overline{\left(M_{ZW}\bar{X}\right)_i \left(M_{ZW}\bar{X}\right)_i'} &= \operatorname{diag}\left\{\overline{\frac{\tilde{U}'M_{ZW}\tilde{U}}{n}}, 0_{m\times m}\right\} \\
&= \operatorname{diag}\left\{(1-\mu-\lambda)\,E\left[\tilde{u}_i\tilde{u}_i'\right], 0_{m\times m}\right\} + o_P(1).
\end{aligned}$$

Using Lemmas A6 and A7(II),

$$\begin{aligned}
\overline{(\hat{e}_{LIML})_i^2 \left(M_{ZW}\bar{X}\right)_i \left(M_{ZW}\bar{X}\right)_i'} &= \overline{\sum_j \left(M_W^{ij} M_{ZW}^{ij}\right)^2}\operatorname{diag}\left\{E\left[\left(e_i^2 - \sigma_e^2\right)\tilde{u}_i\tilde{u}_i'\right], 0_{m\times m}\right\} \\
&\quad + \overline{M_W^{ii} M_{ZW}^{ii}}\operatorname{diag}\left\{\sigma_e^2 E\left[\tilde{u}_i\tilde{u}_i'\right], 0_{m\times m}\right\} + o_P(1),
\end{aligned}$$

and thus

$$\begin{aligned}
\frac{\hat{B}_{1:p,1:p}}{n} &= \frac{\hat{\rho}_\alpha}{\sum_j \left(M_W^{ij} M_{ZW}^{ij}\right)^2}\left[\overline{(\hat{e}_{LIML})_i^2 \left(M_{ZW}\bar{X}\right)_i \left(M_{ZW}\bar{X}\right)_i'} - \hat{\phi}\overline{\left(M_{ZW}\bar{X}\right)_i \left(M_{ZW}\bar{X}\right)_i'}\right]_{1:p,1:p} \\
&\xrightarrow{p} \rho_\alpha E\left[\left(e_i^2 - \sigma_e^2\right)\tilde{u}_i\tilde{u}_i'\right],
\end{aligned}$$

hence

$$\bar{V}_4 = Q_{XZ\perp}^{-1} \frac{\hat{B}_{1:p,1:p}}{n(1-\alpha)^2} Q_{XZ\perp}^{-1} \xrightarrow{p} \rho_\alpha Q_{XZ\perp}^{-1} \frac{E\left[\left(e_i^2 - \sigma_e^2\right)\tilde{u}_i\tilde{u}_i'\right]}{(1-\alpha)^2} Q_{XZ\perp}^{-1} = \Phi_4^{LIML}.$$

Proof of Theorem 8. The properly normalized numerator of J_{LIML} equals

$$\frac{\hat{e}_{LIML}'P_{\hat{\alpha}}\hat{e}_{LIML}}{\sqrt{n}} = \frac{e'P_\alpha e}{\sqrt{n}} + o_P(1).$$

Following the proof of Theorem 5, one gets the conclusion.

Proof of Corollary to Theorem 8. Follow the proof of Corollary to Theorem 5.

References

ANATOLYEV, S. (2012): "Inference in regression models with many regressors," *Journal of Econometrics*, forthcoming.

ANATOLYEV, S. AND N. GOSPODINOV (2011): "Specification testing in models with many instruments," *Econometric Theory*, 27, 427–441.

ANDERSON, T.W. AND H. RUBIN (1949): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Mathematical Statistics*, 20, 46–63.

ANDERSON, T.W., N. KUNIMOTO AND Y. MATSUSHITA (2010): "On the asymptotic optimality of the LIML estimator with possibly many instruments," *Journal of Econometrics*, 157, 191–204.

AZZALINI, A. AND A. CAPITANIO (2003): "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution," *Journal of the Royal Statistical Society Series B*, 65, 367–389.

BAUMGARTNER, B. (2011): "An inequality for the trace of matrix products, using absolute values," manuscript, Universität Wien. Available at `arXiv:1106.6189v2`.

BEKKER, P. A. (1994): "Alternative approximations to the distributions of instrumental variable estimators," *Econometrica*, 62, 657–681.

BILLINGS, R.B. (1987): "Alternative demand model estimators for block rate pricing," *Water Resources Bulletin*, 23, 341–345.

CHAO, J.C. AND N.R. SWANSON (2005): "Consistent estimation with a large number of weak instruments," *Econometrica*, 73, 1673–1692.

CHAO, J.C. AND N.R. SWANSON (2006): "Asymptotic normality of single-equation estimators for the case with a large number of weak instruments," in: D. CORBAE, S.N. DURLAUF, AND B.E.HANSEN (eds.) *Econometric Theory and Practice: Frontier of Analysis and Applied Research*. New York: Cambridge University Press.

CHAO, J.C., N.R. SWANSON, J.A. HAUSMAN, W.K. NEWEY, AND T. WOUTERSEN (2012): "Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments," *Econometric Theory*, 28, 42–86.

DONALD, S. G. AND W.K. NEWEY (2001): "Choosing the number of instruments," *Econometrica*, 69, 1161–1191.

FULLER, W.A. (1977): "Some properties of a modification of the limited information estimator," *Econometrica*, 45, 939–954.

HAHN, J. AND J. HAUSMAN (2002): "A new specification test for the validity of instrumental variables," *Econometrica*, 70, 163–189.

HANSEN, C., J. HAUSMAN AND W.K. NEWEY (2008): "Estimation with many instrumental variables," *Journal of Business & Economics Statistics*, 26, 398–422.

HAUSMAN, J.A., W.K. NEWEY, T. WOUTERSEN, J.C. CHAO, AND N.R. SWANSON (2010): "Instrumental variable estimation with heteroskedasticity and many instruments," *Quantitative Economics*, forthcoming.

KAMERSCHEN, D.R. AND D.V. PORTER (2004): "The demand for residential, industrial and total electricity, 1973–1998," *Energy Economics*, 26, 87–100.

KELEJIAN, H.H. AND I.R. PRUCHA (2001): "On the asymptotic distribution of the Moran I test statistic with applications," *Journal of Econometrics*, 104, 219–257.

LEE, Y. AND R. OKUI (2012): "Hahn–Hausman test as a specification test," *Journal of Econometrics*, 167, 133–139.

NAGAR, A. L. (1959): "The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations," *Econometrica*, 27, 573–595.

NEWEY, W. K. (2004): "Many instrument asymptotics," manuscript, MIT.

RONCALLI, T. AND T. LAGACHE (2004) "A Gauss implementation of skew normal/Student distributions: The Skew library," manuscript, University of Evry.

ROSENZWEIG, M.R. AND T.P. SCHULTZ (1983) "Estimating a household production function: heterogeneity, the demand for health inputs, and their effects on birth weight," *Journal of Political Economy*, 91, 723–746.

VAN HASSELT, M. (2010): "Many instruments asymptotic approximations under nonnormal error distributions," *Econometric Theory*, 26, 633–645.

WITTE, A.D., H.J. SUMKA, AND H. EREKSON (1979): "An estimate of a structural hedonic price model of the housing market: an application of Rosen's theory of implicit markets," *Econometrica*, 47, 1151–1173.