

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

STANISLAV ANATOLYEV\*  
New Economic School

March 2010  
This version: March 2011

## 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, and the associated Hahn–Hausman test is not valid. 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 Hahn–Hausman and  $J$  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, Hahn–Hausman test.

JEL CODES: C12, C21

---

\*Address: New Economic School, Nakhimovskiy Prospekt, 47, room 1721, Moscow, 117418, Russia; sanatoly@nes.ru; <http://www.nes.ru/~sanatoly/>

# 1 Introduction

Often a structural 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, 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 (the authors even do not mention what they are), and of interest is the price elasticity only. Possibly, in this particular example the ratio of the number of exogenous variables to the sample size is not that large for the possible distortions to be perceptible, but in actual empirical studies this ratio may turn out well bigger in practice. This concern may be 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; Lee and Okui, 2009; van Hasselt, 2010; Anderson, Kunimoto and Matsushita, 2010; Chao, Swanson, Hausman, Newey, and Woutersen, 2010; Hausman, Newey, Woutersen, Chao, and Swanson, 2010; Anatolyev and Gospodinov, 2011), and the present paper is aimed 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, the test for strong instruments of Hahn and Hausman (2002), 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, and the associated Hahn–Hausman test is not valid. 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 Hahn–Hausman and  $J$  tests based on the LIML and bias corrected 2SLS estimators.

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. Some useful computational remarks are contained in Section 6. Finally, Section 7 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  $Y = (y_1, \dots, y_n)'$  is  $n \times 1$ ,  $X = (x_1, \dots, x_n)'$  is  $n \times p$ ,  $W = (w_1, \dots, w_n)'$  is  $n \times m$ , and  $e = (e_1, \dots, e_n)'$  is  $n \times 1$ . The object of primary interest is the vector of structural parameters  $\beta_0$ , while the vector  $\delta_0$  contains nuisance parameters. There is additionally an  $n \times \ell$  matrix of instruments (‘excluded exogenous variables’)  $Z = (z_1, \dots, z_n)'$ ,  $\ell \geq p$ . 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  $U = (u_1, \dots, u_n)'$  is  $n \times p$ .

## 2.2 Notation

Denote

$$\Gamma \equiv \frac{\sum' u_e e}{\sigma_e^2},$$

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,

$$\tilde{U} = U - e\Gamma.$$

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 A0 in the Appendix). Additionally, let us introduce the following notation to be used throughout:

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

for any real number  $\zeta$ .

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  $A$ .

Let us also introduce a fundamental quantity

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

which is (asymptotically) a number of instruments per degrees of freedom.

## 2.3 Assumptions

We adapt the following asymptotic framework.

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

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. Even though we exclude the case of few or moderately many exogenous regressors  $\mu = 0$ , it will still be possible to consider it as a limiting case. Henceforth, by ‘lim’ we understand taking a limit under Assumption 1.

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  $E[e_i^2] = \sigma_e^2$ ,  $E[u_i u_i'] = \Sigma_u$  and  $E[u_i e_i] = \Sigma_{ue}$ .*

**Assumption 3** *(i)  $\text{rk}(Z, W) = \ell + m$ , (ii) when  $\ell$  is fixed,  $\lim n^{-1} Z' M_W Z = Q_{Z^\perp}$  and  $\Pi_1$  is of full column rank  $p$ , and when  $\ell \rightarrow \infty$ ,  $\lim n^{-1} \Upsilon'_{XZ} M_W \Upsilon_{XZ} = Q_{XZ^\perp}$ , where  $Q_{Z^\perp}$  and  $Q_{XZ^\perp}$  are finite and positive definite, (iii)  $\sup_{n \geq 1} \sup_{1 \leq i \leq n} \sum_{j=1}^n (|P_W^{i,j}| + |P_{ZW}^{i,j}|) < C < \infty$  and  $\sup_{i \geq 1} \max_{1 \leq j \leq p} (|\Upsilon_{XZ}^{ij}| + |\Upsilon_{XW}^{ij}|) < C < \infty$ , (iv) the limits  $\rho_\alpha = \overline{(P_\alpha^{ii})^2}$  and  $\pi_\alpha = \overline{P_\alpha^{ii} (M_W \Upsilon_{XZ})_i}$  exist and are finite.*

Assumption 3(i) excludes redundant columns from data on instruments and exogenous regressors. Assumption 3(ii) means that the excluded instruments are, as a group, strong after controlling for the explanatory power provided by the exogenous regressors. Assumptions 3(iii,iv) 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  $P_\alpha = P_Z - \lambda I_n$ ,  $\overline{(P_\alpha^{ii})^2} = \overline{(P_Z^{ii})^2} - \lambda^2$  and  $\overline{P_\alpha^{ii} (M_W \Upsilon_{XZ})_i} = \overline{(P_Z^{ii} - \lambda) (\Upsilon_{XZ})_i}$ , hence  $\rho_\alpha = \lim \overline{(P_Z^{ii})^2} - \lambda^2$  and  $\pi_\alpha = \lim \overline{(P_Z^{ii} - \lambda) (\Upsilon_{XZ})_i}$ .

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  $\lambda$ ,  $\mu$  and  $\alpha$ . Also, put

$$\hat{\rho}_\alpha = \overline{(P_{\hat{\alpha}}^{ii})^2}, \quad \hat{\pi}_\alpha = \overline{P_{\hat{\alpha}}^{ii} (P_{Z^\perp} X)_i}$$

to be estimates of  $\rho_\alpha$  and  $\pi_\alpha$ .

### 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

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

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

$$\hat{e}_{2SLS} = M_W \left( Y - X \hat{\beta}_{2SLS} \right).$$

Define the residual variance

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

The t statistic for the  $j^{\text{th}}$  component of  $\beta_0$  based on 2SLS estimation is

$$\begin{aligned} t_{2SLS}^{(j)} &= \frac{(\hat{\beta}_{2SLS})_j - (\beta_0)_j}{\sqrt{\hat{\sigma}_{2SLS}^2 \left[ ((X, W)' P_{ZW} (X, W))^{-1} \right]^{jj}}} \\ &= \frac{(\hat{\beta}_{2SLS})_j - (\beta_0)_j}{\sqrt{\hat{\sigma}_{2SLS}^2 \left[ (X' P_{Z^\perp} X)^{-1} \right]^{jj}}}. \end{aligned}$$

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

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

When there are few exogenous regressors ( $\mu = 0$ ) and few instruments ( $\lambda = 0$ ), 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 ( $\lambda > 0$ ), 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 ( $\mu > 0$ ).

**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  $J^\infty$  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 this case.

**Theorem 1:** Suppose assumptions 1–3 hold, and  $\ell$  is fixed. Then

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

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

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 (2011) 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  $\ell$ .

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

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

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

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

where

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

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  $\ell$  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  $\ell$  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  $\Pi_1$  is replaced by  $\Pi_1 = \hat{\Pi}/\sqrt{n}$ , where  $\hat{\Pi}$  is fixed.

## 4 Bias-corrected 2SLS estimation

### 4.1 Construction

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

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

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 inconsistencies of the bias corrected 2SLS estimators  $\hat{\beta}_{B2SLS}$  are

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

and

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

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:

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

Note that this is similar to the standard bias correction for 2SLS above, but the factor  $\hat{\alpha}$  is different from either conventional factor  $\hat{\xi}$  that asymptotically is equivalent to  $\hat{\lambda}$ . For the bias-corrected 2SLS to be consistent, that factor should be adjusted for the numerosity of exogenous regressors. 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  $\tilde{Y}$  on  $\tilde{X}$  only, using the instrument  $\hat{Z}_2 = P_{\hat{\alpha}}\tilde{X}$ , where  $\tilde{X} = M_W X$  and  $\tilde{Y} = M_W Y$ . 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  $\hat{\beta}_{B2SLS}$ :

$$\hat{e}_{B2SLS} = M_W(Y - X\hat{\beta}_{B2SLS}).$$

Define the residual variance

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

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

$$\begin{aligned}\Phi_2^{B2SLS} &= \sigma_e^2 Q_{XZ^\perp}^{-1} + \frac{\lambda}{1 - \alpha} Q_{XZ^\perp}^{-1} (\sigma_e^2 \Sigma_u + \Sigma_{ue} \Sigma'_{ue}) Q_{XZ^\perp}^{-1}, \\ \Phi_3^{B2SLS} &= \frac{1}{1 - \alpha} Q_{XZ^\perp}^{-1} (\pi_\alpha E[e_i^2 u_i'] + E[e_i^2 u_i] \pi'_\alpha) Q_{XZ^\perp}^{-1},\end{aligned}$$

and

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

The composition of the asymptotic variance is the same as that in van Hasselt (2010): one term  $\Phi_2^{B2SLS}$  is present even under error normality, and two terms  $\Phi_3^{B2SLS}$  and  $\Phi_4^{B2SLS}$  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  $\Phi_2^{B2SLS}$  contributed by the numerosity of instruments. In the case of error non-normality, the terms  $\Phi_3^{B2SLS}$  and  $\Phi_4^{B2SLS}$  are inflated because of numerous exogenous regressors. Of course, these effects may be partially offset or further exacerbated by changes in  $\pi_\alpha$  and  $\rho_\alpha$ .

### 4.3 Variance estimation

To estimate  $V_{B2SLS}$ , we need to construct consistent estimates of its three components,  $\Phi_2^{B2SLS}$ ,  $\Phi_3^{B2SLS}$  and  $\Phi_4^{B2SLS}$ ; under error normality, consistent estimation of only  $\Phi_2^{B2SLS}$  suffices. Its estimate is straightforward to construct:

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

where

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

For estimation of  $\Phi_3^{B2SLS}$  and  $\Phi_4^{B2SLS}$  is it 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  $\Gamma$ . Then form an estimate of  $\Phi_3^{B2SLS}$  as

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

where

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

Finally, form an estimate of  $\Phi_4^{B2SLS}$  as

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

where

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

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 general 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 special 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

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

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:

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

This is an adaptation of the Sargan type statistic of Lee and Okui (2009) to the case of many exogenous regressors.

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

$$\frac{J_{B2SLS}}{\sqrt{n}} \xrightarrow{d} N(0, V_{B2SLS}^J),$$

where the asymptotic variance is

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

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

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

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

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

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

The J type test is one-sided (see Lee and Okui, 2009): we reject the null if the value of

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

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  $\alpha = \lambda$ ),

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

and the test coincides with that in Lee and Okui (2009). In this special case one can simply put

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

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.

## 4.5 Hahn–Hausman test

Consider the case  $p = 1$  and  $\beta_0 \neq 0$ . The celebrated Hahn–Hausman (Hahn and Hausman, 2002) test for the null of strong instruments (HH henceforth) is based on the difference between the usual (forward) and reverse bias corrected 2SLS estimators, i.e. between  $\check{\beta}_{B2SLS}$  with  $\hat{\xi} = (\ell - 2)/n$  and its reverse version

$$\check{\beta}_{B2SLS}^R = \frac{Y' P_{\hat{\xi}} Y}{X' P_{\hat{\xi}} Y}.$$

Because the bias correction in the presence of many exogenous regressors is asymptotically invalid, the HH test will exhibit size distortions.

To appropriately correct for bias when there are many exogenous regressors, we define the modified reverse B2SLS estimator as

$$\hat{\beta}_{B2SLS}^R = \frac{Y' P_{\hat{\alpha}} Y}{X' P_{\hat{\alpha}} Y}$$

**Theorem 6:** Suppose assumptions 1–3 hold, and  $\lambda > 0$ . Then the corrected Hahn–Hausman test statistic

$$HH = \sqrt{\frac{n}{\hat{V}^{HH}}} \left( \hat{\beta}_{B2SLS} - \hat{\beta}_{B2SLS}^R \right),$$

where

$$\hat{V}^{HH} = \frac{n^2 \hat{\sigma}_{B2SLS}^4}{(X' P_{\hat{\alpha}} Y)^2} \hat{V}_{B2SLS}^J,$$

is asymptotically standard normal.

Lee and Okui (2009) show that a one-sided test is more appropriate rather than a two sided test used in Hahn and Hausman (2002). Thus, for a given sample we reject the null if  $HH > q_{\eta}^{N(0,1)}$ , the  $(1 - \eta)$ -quantile of the standard normal. Lee and Okui (2009) also point out that it is more appropriate to interpret the HH test as a test for overidentifying restrictions rather than a test for strong instruments, which explains its low power found in Hausman, Stock and Yogo (2005).

## 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  $\alpha$  (see the proof of Theorem 7). Note also that the estimator  $\hat{\beta}_{LIML}$  may be interpreted as a LIML estimator in a transformed equation where the left side variable are  $M_W Y$  and the right side variables are  $M_W X$  using the instruments  $(Z, W)$ .

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

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

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 7 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  $(\hat{X}' \hat{X})^{-1} \hat{X}' P_{ZW} \hat{X}$ , where  $\hat{X} = (M_W Y, M_W X)$ .

Let us define the residual variance

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

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 7:** 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

$$\begin{aligned}\Phi_2^{LIML} &= \sigma_e^2 Q_{XZ^\perp}^{-1} + \frac{\lambda}{1-\alpha} Q_{XZ^\perp}^{-1} (\sigma_e^2 \Sigma_u - \Sigma_{ue} \Sigma'_{ue}) Q_{XZ^\perp}^{-1}, \\ \Phi_3^{LIML} &= \frac{1}{1-\alpha} Q_{XZ^\perp}^{-1} (\pi_\alpha E[e_i^2 \tilde{u}'_i] + E[e_i^2 \tilde{u}_i] \pi'_\alpha) Q_{XZ^\perp}^{-1},\end{aligned}$$

and

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

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

$$\begin{aligned}\Phi_2^{B2SLS} - \Phi_2^{LIML} &= \frac{2\lambda}{1-\alpha} Q_{XZ^\perp}^{-1} \Sigma_{ue} \Sigma'_{ue} Q_{XZ^\perp}^{-1} \\ \Phi_3^{B2SLS} - \Phi_3^{LIML} &= \frac{1}{1-\alpha} \frac{E[e_i^3]}{\sigma_e^2} Q_{XZ^\perp}^{-1} (\pi_\alpha \Sigma'_{ue} + \Sigma_{ue} \pi'_\alpha) Q_{XZ^\perp}^{-1} \\ \Phi_4^{B2SLS} - \Phi_4^{LIML} &= \frac{\rho_\alpha}{(1-\alpha)^2} Q_{XZ^\perp}^{-1} \left( \frac{E[e_i^3 u_i]}{\sigma_e^2} \Sigma'_{ue} + \Sigma_{ue} \frac{E[e_i^3 u'_i]}{\sigma_e^2} \right. \\ &\quad \left. - \Sigma_{ue} \Sigma'_{ue} \left( \frac{E[e_i^4]}{\sigma_e^4} + 3 \right) \right) Q_{XZ^\perp}^{-1}.\end{aligned}$$

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  $\Phi_3^{LIML}$  and  $\Phi_3^{LIML}$  tend to be small, at least when  $\lambda$  is small, in contrast to the second term in  $\Phi_2^{LIML}$  responsible for numerosity of instruments. Anderson, Kunitomo and Matsushita (2010), Lee and Okui (2009) and Anatolyev and Gospodinov (2011)



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 outweigh 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:

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

where

$$\begin{aligned} \bar{H} &= (X, W)' (P_{ZW} - \bar{\alpha} I_n) (X, W), \\ \bar{\alpha} &= \frac{\hat{e}'_{LIML} P_{ZW} \hat{e}_{LIML}}{\hat{e}'_{LIML} \hat{e}_{LIML}}, \end{aligned}$$

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

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

and

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

The following proposition formalizes the asymptotic validity of our modification.

**Theorem 8:** 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  $\hat{\mu} = 0$ ,  $\hat{\alpha} = \hat{\lambda}$ ,  $P_{\hat{\alpha}} = P_Z - \hat{\lambda}I_n$ ,  $\hat{\rho}_{\alpha} = \overline{(P_Z^{ii})^2} - \hat{\lambda}^2$ ,  $\sum_j (M_W^{ij})^2 M_{ZW}^{ij} = \overline{M_Z^{ii}} = 1 - \hat{\lambda}$ ,  $\overline{(M_Z^{ii})^2} = 1 - 2\hat{\lambda} + \overline{(P_Z^{ii})^2}$ ,  $\overline{M_W^{ii}M_{ZW}^{ii}} = \overline{M_Z^{ii}} = 1 - \hat{\lambda}$ , and it is easily seen that  $\hat{A}$  and  $\hat{B}$  become exactly as those in Hansen, Hausman and Newey (2008, p.399):

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

## 5.4 Specification testing

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

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

This an adaptation of the Sargan type statistic of Lee and Okui (2009) to the case of many exogenous regressors and LIML estimation.

**Theorem 9:** Suppose assumptions 1–3 hold, and  $\lambda > 0$ . Then

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

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

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

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

**Corollary to Theorem 9:** Suppose assumptions 1–3 hold, and  $\lambda > 0$ . Then

$$\hat{V}_{LIML}^J \xrightarrow{p} V_{LIML}^J.$$

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

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

In this special case one can simply put

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

If in addition 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, in contrast to the general test presented here.

The J type test is one-sided (see Lee and Okui, 2009): we reject the null if the value of

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

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  $\alpha$ , but this yields an asymptotically equivalent test. Moreover, inspection of the proofs of Theorems 5, 6, 7 and 9 reveals that the statistics  $\sqrt{n}J_{B2SLS}$ ,  $\sqrt{n}J_{LIML}$ ,  $\sqrt{n}(\tilde{\alpha} - \hat{\alpha})$  and  $HH$  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 (cf. Lee and Okui, 2009).

## 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:

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

where  $\kappa$  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

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

$\hat{\beta}_\kappa$  is asymptotically equivalent to  $\hat{\beta}_{LIML}$  if  $\sqrt{n}(\kappa - \tilde{\alpha}) \xrightarrow{p} 0$ . In the latter case the modified J test statistic

$$J_\kappa = \frac{\hat{e}'_\kappa P_\kappa \hat{e}_\kappa}{\hat{e}'_\kappa \hat{e}_\kappa}$$

is also asymptotically equivalent to  $J_{B2SLS}$  and  $J_{LIML}$ . 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 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  $P_{\tilde{\alpha}}$ ,  $M_W$ ,  $M_{ZW}$  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  $P_{\tilde{\alpha}}$  are coded as `Mw=eye(n)-Pw`, `Mzw=eye(n)-Pzw` and `Pa=Pzw-Pw-ahat*(eye(n)-Pw)`, where scalar `n` contains  $n$ . The following table represents statements for various objects, functions of elements of these matrices.

$\overline{(P_{\tilde{\alpha}}^{ii})^2}$	<code>meanc(diag(Pa)^2)</code>
$\overline{(M_W^{ii})^2}$	<code>meanc(diag(Mw)^2)</code>
$\overline{M_W^{ii} M_{ZW}^{ii}}$	<code>meanc(diag(Mw.*Mzw))</code>
$\overline{\sum_j (M_W^{ij})^3}$	<code>meanc(sumc(Mw^3))</code>
$\overline{\sum_j (M_W^{ij})^2 M_{ZW}^{ij}}$	<code>meanc(sumc(Mw^2.*Mzw))</code>
$\overline{\sum_j (M_W^{ij})^4}$	<code>meanc(sumc(Mw^4))</code>
$\overline{\sum_j (M_W^{ij} M_{ZW}^{ij})^2}$	<code>meanc(sumc(Mw^2.*Mzw^2))</code>

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

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

## 7 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, 2010; Hausman, Newey, Woutersen, Chao, and Swanson, 2010), and accommodate serial correlation in model errors.

## 8 Acknowledgements

My thanks go to the Editor and two anonymous referees for numerous useful suggestions that significantly improved the paper. Also, I am grateful to Victor Chernozhukov, Nikolay Gospodinov and Grigory Kosenok for helpful discussions.

## A Appendix: proofs

**Lemma A0:** 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 A0.**

By the partitioned matrix inverse  $P_{ZW} = (Z, W) ((Z, W)'(Z, W))^{-1} (Z, W)'$  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 A1:** Suppose assumptions 1–3 hold. Then

$$\begin{aligned} \frac{X'P_{Z^\perp}X}{n} &\xrightarrow{p} Q_{XZ^\perp} + \lambda\Sigma_u, & \frac{X'P_{Z^\perp}e}{n} &\xrightarrow{p} \lambda\Sigma_{ue}, & \frac{e'P_{Z^\perp}e}{n} &\xrightarrow{p} \lambda\sigma_e^2, \\ \frac{X'M_W X}{n} &\xrightarrow{p} Q_{XZ^\perp} + (1-\mu)\Sigma_u, & \frac{X'M_W e}{n} &\xrightarrow{p} (1-\mu)\Sigma_{ue}, & \frac{e'M_W e}{n} &\xrightarrow{p} (1-\mu)\sigma_e^2, \\ \frac{X'M_{ZW}X}{n} &\xrightarrow{p} (1-\mu-\lambda)\Sigma_u, & \frac{X'M_{ZW}e}{n} &\xrightarrow{p} (1-\mu-\lambda)\Sigma_{ue}, & \frac{e'M_{ZW}e}{n} &\xrightarrow{p} (1-\mu-\lambda)\sigma_e^2. \end{aligned}$$

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

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

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

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

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

$$\begin{aligned} \frac{X'P_{Z^\perp}X}{n} &= \frac{(P_{Z^\perp}X)'P_{Z^\perp}X}{n} = \frac{\Upsilon'_{XZ}\Upsilon_{XZ}}{n} - \frac{\Upsilon'_{XZ}P_W\Upsilon_{XZ}}{n} + \frac{\Upsilon'_{XZ}P_{Z^\perp}U}{n} + \frac{U'P_{Z^\perp}\Upsilon_{XZ}}{n} \\ &\quad + \frac{U'P_{Z^\perp}U}{n} \xrightarrow{p} Q_{XZ^\perp} + \lambda\Sigma_u \end{aligned}$$

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

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

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 A1 yield

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

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

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

Then, using Lemma A1,

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

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

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

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  $\sigma_e^2$ . 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 A1,  $P_{Z^\perp}X = M_W Z \Pi_1 + P_{Z^\perp}U$  and hence  $n^{-1}X'P_{Z^\perp}X \xrightarrow{p} \Pi_1' Q_{Z^\perp} \Pi_1$  and

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

Because  $U'P_{Z^\perp}e = n^{-1/2}U'M_W Z (n^{-1}Z'M_W Z)^{-1} n^{-1/2}Z'M_W e = O_P(1)Q_{Z^\perp}O_P(1)$ , the second term above is  $o_P(1)$ . Let  $a = (0'_p, 1)'$  and  $b = (I_p, 0_p)'$ , then we have

$$\frac{Z'M_W e}{\sqrt{n}} = b' \frac{(Z, e)' M_W (Z, e) a}{\sqrt{n}}.$$

Note first that  $E[Z'M_W e] = 0$ . Our Assumptions 2 and 3 make sure that Assumption 1 of Theorem 1 in van Hasselt (2010) is satisfied. We apply the central limit theorem in Theorem 1 of van Hasselt (2010) with  $C = M_W$ ,  $\tau_C = \tau_{C^2} = 1 - \mu$ ,  $M = (Z, 0)$ ,  $V = (0, e)$ ,  $Q_{CM} = \text{diag}\{Q_{Z^\perp}, 0\}$  and  $\Omega \equiv \text{var}((0, e_i)') = \text{diag}\{0, \sigma_e^2\}$ . Note that  $a'Q_{CM} = 0$  which makes the first, third and fourth terms in the asymptotic variance equal zero. Next, because  $b'\Omega = 0$ , the next term is also zero. In addition,  $b'v_t = 0$  and  $b'\Omega = 0$  together make all the other terms also zero. Now, because  $a'\Omega a = \sigma_e^2$  and  $b'Q_{CM} = Q_{Z^\perp}$ , we conclude that  $Z'M_W e/\sqrt{n}$  is asymptotically normal with the asymptotic variance  $\sigma_e^2 Q_{Z^\perp}$ . Together,

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

and

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

The t statistic for the  $j^{\text{th}}$  component based on 2SLS is

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

Next, the numerator of the J statistic is

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



where

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

Taking account of the fact that the first factor in  $\psi$  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  $\sigma_e^2$ , 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,

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

and

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

using Lemmas A0 and A1. By Lemma A0,  $(Y^\perp - X^\perp \beta_0)' P_{Z^\perp} (Y^\perp - X^\perp \beta_0) = e' P_{Z^\perp} e$ , and from the proof of Theorem 1  $e' P_{Z^\perp} e \xrightarrow{d} \sigma_e^2 \chi^2(\ell)$ . Hence,

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

Next, again using Lemma A0,

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

When the instruments are strong,

$$\begin{aligned} \frac{\tilde{Z}' \tilde{Z}}{n} &= \frac{X' P_{Z^\perp} X}{n} + \frac{e' P_{Z^\perp} e}{n} (\Gamma + o_P(1))^2 - 2 \frac{X' P_{Z^\perp} e}{n} (\Gamma + o_P(1)) \\ &\xrightarrow{p} \Pi_1' Q_{Z^\perp} \Pi_1, \end{aligned}$$

using Lemma A1. Hence

$$\begin{aligned} (Y^\perp - X^\perp \beta_0)' P_{\tilde{Z}} (Y^\perp - X^\perp \beta_0) &= e' M_W P_{\tilde{Z}} M_W e \\ &= \psi_1' (\Pi_1' Q_{Z^\perp} \Pi_1)^{-1} \psi_1 + o_P(1), \end{aligned}$$

where

$$\begin{aligned} \psi_1 &\equiv \frac{1}{\sqrt{n}} (X - e\Gamma)' P_{Z^\perp} e = \frac{X' P_{Z^\perp} e}{\sqrt{n}} - \Gamma' \frac{e' P_{Z^\perp} e}{\sqrt{n}} \\ &\xrightarrow{d} N(0, \sigma_e^2 \Pi_1' Q_{Z^\perp} \Pi_1) \end{aligned}$$

using the proof of Theorem 1. To summarize,

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

Now let fixed  $\Pi_1$  be replaced by  $\Pi_1 = \dot{\Pi}/\sqrt{n}$  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

$$\begin{aligned}\Psi &\equiv \begin{pmatrix} \Psi_{\tilde{U}} \\ \Psi_e \end{pmatrix} \equiv \text{vec} \left( \left( \frac{Z'M_W Z}{n} \right)^{-1/2} \frac{Z'M_W(\tilde{U}, e)}{\sqrt{n}} \right) \\ &\xrightarrow{d} \begin{pmatrix} \xi_{\tilde{U}} \\ \xi_e \end{pmatrix} \sim N(0, \text{diag} \{ \Sigma_u - \sigma_e^2 \Gamma' \Gamma, \sigma_e^2 \} \otimes I_\ell).\end{aligned}$$

Using Lemma A0,

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

therefore

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

where

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

Hence,

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

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

**Lemma A2:** 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 A2.** Using Lemma A1,

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

and similarly

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

Last,

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

**Lemma A3:** 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  $p + 1$ ,

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

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 of Lemma A3.** Note first that, using the manipulations in the proof of Lemma A1,

$$E \left[ (X, e)' P_\alpha e \right] = E \left[ \begin{pmatrix} \Upsilon_{XZ}' P_\alpha e + U' P_\alpha e \\ e' P_\alpha e \end{pmatrix} \right] = 0_{p+1}.$$

Let  $a = (0_p', 1)'$  then we have

$$B' \frac{(X, e)' P_\alpha e}{\sqrt{n}} = B' \frac{(X, e)' P_\alpha (X, e) a}{\sqrt{n}}.$$

We apply Theorem 1 of van Hasselt (2010) with  $C = P_\alpha$ ,  $\text{tr}(C) = \ell + m - m - \alpha(n - m) = 0$ ,  $\tau_C = 0$ ,  $\text{tr}(C^2) = \text{tr}((1 - 2\alpha) P_{Z^\perp} + \alpha^2 M_W) = (1 - 2\alpha)(\ell + m - m) + \alpha^2(n - m) = (1 - \alpha)\ell$  so that  $\text{tr}(C^2)/n \rightarrow (1 - \alpha)\lambda \equiv \tau_{C^2}$ . Next, the fixed part is  $M = (\Upsilon_{XZ} + \Upsilon_{XW}, 0)$  while the random part is  $(U, e)$ , so

$$CM = P_\alpha (\Upsilon_{XZ} + \Upsilon_{XW}, 0) = (1 - \alpha) (M_W \Upsilon_{XZ}, 0)$$

and

$$\frac{(CM)' CM}{n} \xrightarrow{p} (1 - \alpha)^2 \begin{bmatrix} Q_{XZ^\perp} & 0_p \\ 0_p' & 0 \end{bmatrix} \equiv Q_{CM}.$$

According to Assumption 3(iv),

$$\frac{1}{n} d_{P_\alpha}' d_{P_\alpha} \xrightarrow{p} \rho_\alpha \equiv \delta_C$$

and

$$\frac{1}{n} (CM)' d_{P_\alpha} = (1 - \alpha) \frac{1}{n} \begin{pmatrix} \Upsilon'_{XZ} M_W d_{P_\alpha} \\ 0 \end{pmatrix} \xrightarrow{p} (1 - \alpha) \begin{pmatrix} \pi_\alpha \\ 0 \end{pmatrix} \equiv \mu_{CM}$$

according to Assumption 3(iv).

From the above and our Assumptions 2 and 3 it follows that Assumption 1 of Theorem 1 in van Hasselt (2010) is satisfied. Next note that  $a'Q_{CM} = 0$  which makes the first, third and fourth terms in the asymptotic variance of van Hasselt (2010, Theorem 1) equal zero. Next,  $\Omega \equiv \text{var}((u'_i, e_i)')$  obeys  $a'\Omega a = \text{var}((0'_p, 1)(u'_i, e_i)') = \sigma_e^2$ ,  $\Omega a = (\Sigma'_{ue}, \sigma_e^2)'$  and  $a'\mu_{CM} = 0$ . Hence the component of the asymptotic variance that comes from the second and fifth terms equals

$$\begin{aligned} & \sigma_e^2 B' Q_{CM} B + \tau_{C^2} B' (\sigma_e^2 \Omega + (\Sigma'_{ue}, \sigma_e^2)' (\Sigma'_{ue}, \sigma_e^2)) B \\ = & (1 - \alpha)^2 \sigma_e^2 B'_1 Q_{XZ^\perp} B_1 \\ & + (1 - \alpha) \lambda (\sigma_e^2 B' \Omega B + (B'_1 \Sigma_{ue} + \sigma_e^2 b_2) (\Sigma'_{ue} B_1 + \sigma_e^2 b'_2)). \end{aligned}$$

Next, the component associated with third moments equals

$$\begin{aligned} & B' \mu_{CM} E [(a' (u'_i, e_i)')^2 (u'_i, e_i)] B + B' E [(a' (u'_i, e_i)')^2 (u'_i, e_i)'] \mu'_{CM} B \\ = & (1 - \alpha) (B'_1 \pi_\alpha E [e_i^2 (u'_i, e_i)] B + B' E [e_i^2 (u'_i, e_i)'] \pi'_\alpha B_1). \end{aligned}$$

Finally, the component associated with fourth moments equals

$$\begin{aligned} & \delta_C B' (E [(a' (u'_i, e_i)')^2 (u'_i, e_i)' (u'_i, e_i)] - \sigma_e^2 \Omega - 2(\Sigma'_{ue}, \sigma_e^2)' (\Sigma'_{ue}, \sigma_e^2)) B \\ = & \rho_\alpha (B' (E [e_i^2 (u'_i, e_i)' (u'_i, e_i)]) B - \sigma_e^2 B' \Omega B - 2(B'_1 \Sigma_{ue} + \sigma_e^2 b_2) (\Sigma'_{ue} B_1 + \sigma_e^2 b'_2)). \end{aligned}$$

Rearranging terms yields the statement.

**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 A1(a),

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

Similarly, for the second version,

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

**Lemma A4:** Suppose assumptions 1–3 hold, and  $\lambda > 0$ . Then  $\hat{\pi}_\alpha \rightarrow \pi_\alpha$  and  $\hat{\rho}_\alpha \rightarrow \rho_\alpha$ .

**Proof of Lemma A4.** 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

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

which in absolute value is no greater than

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

for  $C$  from Assumption 3(iii). Hence,

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

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

**Proof of Theorem 3.** From Lemma A2,

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

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 A2, Lemma A3 with  $B = (I_p, 0_p)'$  and the Slutsky Theorem yields

$$\begin{aligned} \sqrt{n} \left( \hat{\beta}_{B2SLS} - \beta_0 \right) &\xrightarrow{d} ((1 - \alpha) Q_{XZ^\perp})^{-1} N \left( 0, \begin{pmatrix} (1 - \alpha)^2 \sigma_e^2 Q_{XZ^\perp} + ((1 - \alpha) \lambda - \rho_\alpha) \sigma_e^2 \Sigma_u \\ + ((1 - \alpha) \lambda - 2\rho_\alpha) \Sigma_{ue} \Sigma'_{ue} + \rho_\alpha E [e_i^2 u_i u'_i] \\ + (1 - \alpha) (\pi_\alpha E [e_i^2 u'_i] + E [e_i^2 u_i] \pi'_\alpha) \end{pmatrix} \right) \\ &\stackrel{d}{=} N \left( 0, \Phi_2^{B2SLS} + \Phi_3^{B2SLS} + \Phi_4^{B2SLS} \right), \end{aligned}$$

where  $\Phi_2^{B2SLS}$ ,  $\Phi_3^{B2SLS}$  and  $\Phi_4^{B2SLS}$  are as in the statement of the theorem.

**Lemma A5:** Suppose assumptions 1–3 hold, and  $\lambda > 0$ . The following expressions are at most  $O_P(1/n)$ :

- I.  $\overline{(M_W e)_i^3} - \sum_j (M_W^{ij})^3 E [e_i^3]$ ,
- II.  $\overline{(M_W e)_i^2 (M_{ZW} \tilde{U})_i} - \sum_j (M_W^{ij})^2 M_{ZW}^{ij} E [e_i^2 \tilde{u}_i]$ ,

$$\begin{aligned} \text{III. } & \overline{(M_W e)_i^4} - \overline{\sum_j (M_W^{ij})^4} E[e_i^4 - 3\sigma_e^4] - 3\overline{(M_W^{ii})^2} \sigma_e^4, \\ \text{IV. } & \overline{(M_W e)_i^2 (M_{ZW} \tilde{U})_i (M_{ZW} \tilde{U})'_i} - \overline{\sum_j (M_W^{ij} M_{ZW}^{ij})^2} E[(e_i^2 - \sigma_e^2) \tilde{u}_i \tilde{u}'_i] - \overline{M_W^{ii} M_{ZW}^{ii}} \sigma_e^2 E[\tilde{u}_i \tilde{u}'_i]. \end{aligned}$$

**Proof of Lemma A5.** We heavily use the properties of  $M_W$  such as  $|M_W^{ij}| < 1$ ,  $\sum_j (M_W^{ij})^2 = M_W^{ii} \leq 1$ ,  $\sum_j |M_W^{ij}| < C$  with  $C$  from Assumption 3,  $\sum_k |M_W^{i_1 k} M_W^{i_2 k}| \leq \left(\sum_k (M_W^{i_1 k})^2 \sum_k (M_W^{i_2 k})^2\right)^{1/2} \leq (M_W^{i_1 i_1} M_W^{i_2 i_2})^{1/2} \leq 1$  and similarly  $\sum_k |M_W^{i_1 k} M_{ZW}^{i_2 k}| \leq 1$ , Lemma B1 in Chao et al (2009), and the following and similar inequalities:

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

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

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

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

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

which in absolute value is no greater than

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

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

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

which in absolute value is no greater than

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

Second, take

$$\frac{T_2}{2} = \sum_j \sum_{k \neq j} \overline{M_W^{i_1 j} M_W^{i_1 k} M_{ZW}^{i_2 k}} e_j e_k \tilde{u}_k,$$

whose variance

$$\sum_{j_1} \sum_{k_1 \neq j_1} \sum_{j_2} \sum_{k_2 \neq j_2} \overline{M_W^{i_1 j_1} M_W^{i_1 k_1} M_{ZW}^{i_2 k_1} M_W^{i_2 j_2} M_W^{i_2 k_2} M_{ZW}^{i_2 k_2}} E [e_{j_1} e_{k_1} e_{j_2} e_{k_2} \tilde{u}_{k_1} \tilde{u}'_{k_2}]$$

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  $E [e_k^2 \tilde{u}_k \tilde{u}'_k] \sigma_e^2$ , 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

$$\begin{aligned} & \frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_k |M_W^{i_1 k} M_{ZW}^{i_1 k} M_W^{i_2 k} M_{ZW}^{i_2 k}| \sum_{j \neq k} |M_W^{i_1 j} M_W^{i_2 j}| \\ & \leq \frac{1}{n^2} \sum_k \sum_{i_1} |M_W^{i_1 k} M_{ZW}^{i_1 k}| \sum_{i_2} |M_W^{i_2 k} M_{ZW}^{i_2 k}| = O\left(\frac{1}{n}\right); \end{aligned}$$

(ii)  $j_1 = k_2, j_2 = k_1 \neq j_1$  leading to scaled  $E [e_j^2 \tilde{u}_j] E [e_j^2 \tilde{u}'_j]$ , 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

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

Third, take

$$T_3 = \sum_j \sum_{k \neq j} \sum_{l \notin \{j, k\}} \overline{M_W^{i_1 j} M_W^{i_1 k} M_{ZW}^{i_2 l}} e_j e_k \tilde{u}_l,$$

whose variance

$$\sum_{j_1} \sum_{k_1 \neq j_1} \sum_{l_1 \notin \{j_1, k_1\}} \sum_{j_2} \sum_{k_2 \neq j_2} \sum_{l_2 \notin \{j_2, k_2\}} \overline{M_W^{i_1 j_1} M_W^{i_1 k_1} M_{ZW}^{i_2 l_1}} \overline{M_W^{i_2 j_2} M_W^{i_2 k_2} M_{ZW}^{i_2 l_2}} E [e_{j_1} e_{k_1} \tilde{u}_{l_1} e_{j_2} e_{k_2} \tilde{u}_{l_2}]$$

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  $\sigma_e^4 E [\tilde{u}_l \tilde{u}'_l]$ , then the variance component is proportional to

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

which in absolute value is no greater than

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j |M_W^{i_1 j}| |M_W^{j i_2}| \sum_{k \neq j} |M_W^{i_1 k} M_W^{i_2 k}|.$$

Hence the variance component is proportional to

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

Fourth, take

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

whose variance is

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

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

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

which in absolute value is no greater than

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

Part III. Observe that

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

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

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

whose variance

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

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

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



which in absolute value is no greater than

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

(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

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

which in absolute value is no greater than

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

(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

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

which in absolute value is no greater than

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

Second, take

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

whose variance

$$\begin{aligned} &\sum_{j_1} \sum_{k_1 \neq j_1} \sum_{l_1 \notin \{j_1, k_1\}} \sum_{j_2} \sum_{k_2 \neq j_2} \sum_{l_2 \notin \{j_2, k_2\}} \overline{(M_W^{i j_1})^2 M_W^{i k_1} M_W^{i l_1}} \\ &\quad \overline{(M_W^{i j_2})^2 M_W^{i k_2} M_W^{i l_2}} E[e_{j_1}^2 e_{k_1} e_{l_1} e_{j_2}^2 e_{k_2} e_{l_2}] \end{aligned}$$

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

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

which in absolute value is no greater than

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

Hence the variance component is proportional to

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

(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  $\sigma_e^8$ , then the variance component is proportional to

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

which is no greater in absolute value than

$$\begin{aligned} & \frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_{j_1} (M_W^{i_1 j_1})^2 \sum_{j_2} (M_W^{i_2 j_2})^2 \\ & \sum_{k \neq j_1, j_2} |M_W^{i_1 k} M_W^{i_2 k}| (|M_W^{i_1 i_2}| + 2|M_W^{i_1 j_1} M_W^{i_2 j_1}| + |M_W^{i_1 k} M_W^{i_2 k}|), \end{aligned}$$

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

$$\frac{1}{n^2} \sum_{i_1} \sum_{i_2} |M_W^{i_1 i_2}| \sum_{j_1} (M_W^{i_1 j_1})^2 \sum_{j_2} (M_W^{i_2 j_2})^2 \sum_{k \neq j_1, j_2} |M_W^{i_1 k} M_W^{i_2 k}| \leq \frac{1}{n^2} \sum_{i_1} C \leq O\left(\frac{1}{n}\right),$$

the second does not exceed

$$\begin{aligned} & \frac{2}{n^2} \sum_{i_1} \sum_{i_2} \sum_{j_1} |M_W^{i_1 j_1}|^3 \sum_{j_2} (M_W^{i_2 j_2})^2 |M_W^{i_2 j_1}| \sum_{k \neq j_1, j_2} |M_W^{i_1 k} M_W^{i_2 k}| \\ & \leq \frac{2}{n^2} \sum_{i_2} \sum_{j_1} \sum_{j_2} (M_W^{i_2 j_2})^2 |M_W^{i_2 j_1}| \sum_{i_1} (M_W^{i_1 j_1})^2 \\ & \leq \frac{2}{n^2} \sum_{i_2} \sum_{j_2} (M_W^{i_2 j_2})^2 \sum_{j_1} |M_W^{i_2 j_1}| \leq \frac{2C}{n^2} \sum_{i_2} \sum_{j_2} (M_W^{i_2 j_2})^2 = O\left(\frac{n-m}{n^2}\right), \end{aligned}$$

and the third is

$$\begin{aligned} & \frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_{j_1} (M_W^{i_1 j_1})^2 \sum_{j_2} (M_W^{i_2 j_2})^2 \sum_{k \notin \{j_1, j_2\}} (M_W^{i_1 k} M_W^{i_2 k})^2 \\ & \leq \frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_k (M_W^{i_1 k})^2 (M_W^{i_2 k})^2 = O\left(\frac{n-m}{n^2}\right). \end{aligned}$$

(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

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

which is no greater in absolute value than

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

Third, take

$$T_3 = \sum_j \sum_{k \neq j} \sum_{l \notin \{k, j\}} \sum_{q \notin \{j, k, l\}} \overline{M_W^{i_1 j} M_W^{i_2 k} M_W^{i_3 l} M_W^{i_4 q}} e_j e_k e_l e_q,$$

whose variance

$$\begin{aligned}
& \sum_{j_1} \sum_{k_1 \neq j_1} \sum_{l_1 \notin \{k_1, j_1\}} \sum_{q_1 \notin \{j_1, k_1, l_1\}} \sum_{j_2} \sum_{k_2 \neq j_2} \sum_{l_2 \notin \{k_2, j_2\}} \sum_{q_2 \notin \{j_2, k_2, l_2\}} \\
& \quad \overline{M_W^{i_1 j_1} M_W^{i_2 k_1} M_W^{i_3 l_1} M_W^{i_4 q_1} M_W^{i_1 j_2} M_W^{i_2 k_2} M_W^{i_3 l_2} M_W^{i_4 q_2}} E[e_{j_1} e_{k_1} e_{l_1} e_{q_1} e_{j_2} e_{k_2} e_{l_2} e_{q_2}]
\end{aligned}$$

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  $\sigma_e^8$ , so the variance component is proportional to

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

which in absolute value is no greater than

$$\begin{aligned}
& \frac{1}{n^2} \sum_{i_1} \sum_{i_2} \sum_j |M_W^{i_1 j} M_W^{i_2 j}| (|M_W^{i_1 i_2}| + 3 |M_W^{i_1 j} M_W^{i_2 j}|) \\
& \leq \frac{1}{n^2} \sum_{i_1} \sum_{i_2} |M_W^{i_1 i_2}| + \frac{3}{n^2} \sum_j \sum_{i_1} \sum_{i_2} (M_W^{i_1 j})^2 (M_W^{i_2 j})^2 \leq O\left(\frac{1}{n}\right),
\end{aligned}$$

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

$$T_4 = \overline{\sum_j (M_W^{i_1 j})^4} (e_j^4 - E[e_j^4]),$$

whose variance is

$$\sum_{j_1} \sum_{j_2} \overline{(M_W^{i_1 j_1})^4 (M_W^{i_1 j_2})^4} E[(e_{j_1}^4 - E[e_{j_1}^4]) (e_{j_2}^4 - E[e_{j_2}^4])]$$

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

$$\frac{1}{n^2} \sum_j \sum_{i_1} (M_W^{i_1 j})^4 \sum_{i_2} (M_W^{i_2 j})^4,$$

which is no greater than

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

Fifth, take

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

whose variance is

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

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

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

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

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

**Lemma A6:** The following is true:

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

**Proof of Lemma A6.** By straightforward computation.

**Proof of Theorem 4.** Because  $M_{ZW} \hat{e}_{B2SLS} = M_{ZW} e - M_{ZW} X (\hat{\beta}_{B2SLS} - \beta_0)$ , 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 A1 and consistency of  $\hat{\beta}_{B2SLS}$ . Thus, again using Lemma A1,

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

Hence, also using Lemma A2,

$$\begin{aligned} \hat{\Phi}_2^{B2SLS} &\xrightarrow{p} (1 - \alpha) \sigma_e^2 ((1 - \alpha) Q_{XZ^\perp})^{-1} + \frac{1}{1 - \mu} \frac{\alpha}{1 - \alpha} ((1 - \alpha) Q_{XZ^\perp})^{-1} \\ &\quad \times ((1 - \mu - \lambda)^2 (\sigma_e^2 \Sigma_u + \Sigma_{ue} \Sigma'_{ue})) ((1 - \alpha) Q_{XZ^\perp})^{-1} \\ &= \sigma_e^2 Q_{XZ^\perp}^{-1} + \frac{\lambda}{1 - \alpha} Q_{XZ^\perp}^{-1} (\sigma_e^2 \Sigma_u + \Sigma_{ue} \Sigma'_{ue}) Q_{XZ^\perp}^{-1} = \Phi_2^{B2SLS}. \end{aligned}$$

Next, using Lemma A2

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

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

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

By Lemma A5(I,II),

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

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

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

we have using Lemma A6

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

Then

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

Next, by Lemma A5(III,IV),

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

and

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

Then, using also Lemma A6,

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

and

$$\hat{\Phi}_4^{B2SLS} \xrightarrow{p} \rho_\alpha ((1 - \alpha) Q_{XZ^\perp})^{-1} (E [e_i^2 u_i u'_i] - \sigma_e^2 \Sigma_u - 2\Sigma_{ue} \Sigma'_{ue}) ((1 - \alpha) Q_{XZ^\perp})^{-1} = \Phi_4^{B2SLS}.$$

**Proof of Theorem 5.** Because  $M_W (P_{ZW} - \hat{\alpha} I_n) M_W = M_W P_{\hat{\alpha}} = P_{\hat{\alpha}}$  and using Lemma A2, we have

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

Using asymptotic normality of  $e' P_\alpha e / \sqrt{n}$  from Lemma A3 with  $B = (0'_p, 1)'$ , we get that the above quantity is asymptotically normal with the asymptotic variance  $2\lambda(1 - \alpha) \sigma_e^4 + \rho_\alpha (E [e_i^4] - 3\sigma_e^4)$ .

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

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

**Proof of Corollary to Theorem 5.** The first term clearly converges to  $2\lambda(1 - \alpha)$ . In the second term,  $\hat{\rho}_\alpha$  is consistent for  $\rho_\alpha$  by Lemma A4, and  $\hat{\sigma}_{B2SLS}^4$  is consistent for  $\sigma_e^4$ .

Next, by Lemma A5(III),

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

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

**Proof of Theorem 6.** The limit of the reverse B2SLS estimator is, using Assumption 1 and that  $P_\alpha W = 0$ ,

$$\begin{aligned} \hat{\beta}_{B2SLS}^R &= \frac{(X\beta_0 + W\delta_0 + e)' P_\alpha (X\beta_0 + W\delta_0 + e)}{X' P_\alpha (X\beta_0 + W\delta_0 + e)} + o_P \left( \frac{1}{\sqrt{n}} \right) \\ &= \beta_0 + \frac{e' P_\alpha X \beta_0 + e' P_\alpha e}{X' P_\alpha X \beta_0 + X' P_\alpha e} + o_P \left( \frac{1}{\sqrt{n}} \right). \end{aligned}$$

From Lemma A2,

$$\begin{aligned} \hat{\beta}_{B2SLS}^R &= \beta_0 + \left( \frac{X' P_\alpha X}{n} \beta_0 + \frac{X' P_\alpha e}{n} \right)^{-1} \left( \frac{e' P_\alpha X}{n} \beta_0 + \frac{e' P_\alpha e}{n} \right) + o_P \left( \frac{1}{\sqrt{n}} \right) \\ &= \beta_0 + (((1 - \alpha) Q_{XZ^\perp} + o_P(1)) \beta_0 + o_P(1))^{-1} (o_P(1) \cdot \beta_0 + o_P(1)) \\ &\xrightarrow{p} \beta_0. \end{aligned}$$

The difference between the estimators, using that  $X' P_\alpha Y = X' P_\alpha M_W Y = X' P_\alpha X \beta_0 + X' P_\alpha e$  and  $Y' P_\alpha Y = Y' M_W P_\alpha M_W Y = X' P_\alpha X \beta_0^2 + 2e' P_\alpha X \beta_0 + e' P_\alpha e$ , is, apart from the  $o_P(1/\sqrt{n})$  term,

$$\begin{aligned} \hat{\beta}_{B2SLS} - \hat{\beta}_{B2SLS}^R &= \frac{(X' P_\alpha Y)^2 - X' P_\alpha X \cdot Y' P_\alpha Y}{X' P_\alpha X \cdot X' P_\alpha Y} \\ &= \frac{(X' P_\alpha X \beta_0 + X' P_\alpha e)^2 - X' P_\alpha X (X' P_\alpha X \beta_0^2 + 2e' P_\alpha X \beta_0 + e' P_\alpha e)}{X' P_\alpha X \cdot X' P_\alpha Y} \\ &= \frac{(X' P_\alpha e)^2 - X' P_\alpha X \cdot e' P_\alpha e}{X' P_\alpha X \cdot X' P_\alpha Y} = \frac{(X' P_\alpha X)^{-1} (X' P_\alpha e)^2 - e' P_\alpha e}{X' P_\alpha Y}, \end{aligned}$$

so, using repeatedly Lemma A2, we get

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

Following the proof of Theorem 5, we get

$$\sqrt{n} \left( \hat{\beta}_{B2SLS} - \hat{\beta}_{B2SLS}^R \right) \left| \frac{X' P_\alpha Y}{n} \right| \xrightarrow{d} N \left( 0, 2\lambda(1 - \alpha) \sigma_e^4 + \rho_\alpha (E[e_i^4] - 3\sigma_e^4) \right).$$

As  $\hat{\sigma}_{B2SLS}^4 \hat{V}_{B2SLS}^J$  is consistent for this asymptotic variance (see proof of Corollary to Theorem 5),  $\hat{V}^{HH} \xrightarrow{p} V^{HH}$ , and standardization by  $\sqrt{\hat{V}^{HH}}$  leads to the standard normal distribution.

**Proof of Theorem 7.** Consider the normalized (by  $n^{-1}$ ) numerator and denominator of  $F(\beta)$ . Using Lemma A1,  $n^{-1}(e - X(\beta - \beta_0))' P_{Z^\perp} (e - X(\beta - \beta_0))$  equals

$$\begin{aligned} & \frac{e' P_{Z^\perp} e}{n} - (\beta - \beta_0)' \frac{X' P_{Z^\perp} e}{n} - \frac{e' P_{Z^\perp} X}{n} (\beta - \beta_0) + (\beta - \beta_0)' \frac{X' P_{Z^\perp} X}{n} (\beta - \beta_0) \\ & \xrightarrow{p} \lambda \sigma_e^2 - \lambda (\beta - \beta_0)' \Sigma_{ue} - \lambda \Sigma'_{ue} (\beta - \beta_0) + (\beta - \beta_0)' (Q_{XZ^\perp} + \lambda \Sigma_u) (\beta - \beta_0), \end{aligned}$$

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

$$\begin{aligned} & \frac{e' M_W e}{n} - (\beta - \beta_0)' \frac{X' M_W e}{n} - \frac{e' M_W X}{n} (\beta - \beta_0) + (\beta - \beta_0)' \frac{X' M_W X}{n} (\beta - \beta_0) \\ & \xrightarrow{p} (1 - \mu) \sigma_e^2 - (1 - \mu) (\beta - \beta_0)' \Sigma_{ue} - (1 - \mu) \Sigma'_{ue} (\beta - \beta_0) \\ & + (\beta - \beta_0)' (Q_{XZ^\perp} + (1 - \mu) \Sigma_u) (\beta - \beta_0). \end{aligned}$$

To summarize, the probability limit of the objective function is

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

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

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

The first derivative of  $F(\beta)$  evaluated at  $\beta_0$  is

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

Note that

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

so, using Lemma A2,

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



Next we apply Lemma A3 with  $B = \begin{pmatrix} I_p \\ -\Gamma \end{pmatrix}$  to get that

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

where, exploiting Lemma A6,

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

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

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

When it is evaluated at  $\beta_0$ , the probability limit is, using Lemma A1,

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

Summarizing, the asymptotics for the LIML estimator is

$$\begin{aligned} \sqrt{n} (\hat{\beta}_{LIML} - \beta_0) &\xrightarrow{d} \left( \frac{2(1 - \alpha) Q_{XZ^\perp}}{1 - \mu \sigma_e^2} \right)^{-1} \frac{2}{(1 - \mu) \sigma_e^2} N(0, (1 - \alpha) \Xi_2 + \alpha(1 - \alpha) \Xi_3 + \rho_\alpha \Xi_4) \\ &\stackrel{d}{=} N(0, \Phi_2^{LIML} + \Phi_3^{LIML} + \Phi_4^{LIML}), \end{aligned}$$

where  $\Phi_2^{LIML}$ ,  $\Phi_3^{LIML}$  and  $\Phi_4^{LIML}$  are as in the statement of the theorem.

**Proof of Theorem 8.** First, using that  $\hat{e}_{LIML} = M_W e - M_W X \left( \hat{\beta}_{LIML} - \beta_0 \right)$ , we get using Lemma A1 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 A2,

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

Next,

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

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 A1,

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

After multiplication of the sandwich, the northwest block becomes

$$\begin{aligned}
\hat{V}_{LIML} &= \left[ \left( \frac{\bar{H}}{n} \right)^{-1} \left( \hat{\sigma}_{LIML}^2 \frac{(1-\bar{\alpha})^2 \bar{X}' P_{ZW} \bar{X} + \bar{\alpha}^2 \bar{X}' M_{ZW} \bar{X}}{n} + \frac{\hat{A}}{n} + \frac{\hat{A}'}{n} + \frac{\hat{B}}{n} \right) \left( \frac{\bar{H}}{n} \right)^{-1} \right]_{1:p,1:p} \\
&= Q_{XZ^\perp}^{-1} \left( \frac{\sigma_e^2 (1-\alpha)^2 (Q_{XZ^\perp} + \lambda \Sigma_u) + \sigma_e^2 \alpha^2 (1-\mu-\lambda) \Sigma_u}{(1-\alpha)^2} \right. \\
&\quad \left. - \frac{\Sigma_{ue} \Sigma'_{ue} ((1-\alpha)^2 \lambda + \alpha^2 (1-\mu-\lambda))}{(1-\alpha)^2} + \frac{\hat{A}_{1:p,1:p} + \hat{A}'_{1:p,1:p} + \hat{B}_{1:p,1:p}}{n(1-\alpha)^2} \right) Q_{XZ^\perp}^{-1} + o_P(1) \\
&= \bar{V}_2 + \bar{V}_3 + \bar{V}_4 + o_P(1),
\end{aligned}$$

where

$$\begin{aligned}
\bar{V}_2 &= \sigma_e^2 Q_{XZ^\perp}^{-1} + \frac{\lambda}{1-\alpha} Q_{XZ^\perp}^{-1} (\sigma_e^2 \Sigma_u - \Sigma_{ue} \Sigma'_{ue}) Q_{XZ^\perp}^{-1} = \Phi_2^{LIML}, \\
\bar{V}_3 &= Q_{XZ^\perp}^{-1} \frac{\hat{A}_{1:p,1:p} + \hat{A}'_{1:p,1:p}}{n(1-\alpha)^2} Q_{XZ^\perp}^{-1}, \quad \bar{V}_4 = Q_{XZ^\perp}^{-1} \frac{\hat{B}_{1:p,1:p}}{n(1-\alpha)^2} Q_{XZ^\perp}^{-1}.
\end{aligned}$$

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

$$\begin{aligned}
M_{ZW} \bar{X} &= (M_{ZW}X, M_{ZW}W) - M_{ZW} \hat{e}_{LIML} \bar{\Gamma}_{LIML} \\
&= (M_{ZW}X, 0) - M_{ZW} e \frac{e' (M_W X, 0)}{\hat{e}'_{LIML} \hat{e}_{LIML}} \\
&\quad + M_{ZW} e (X' M_W X, 0) \cdot o_P(1) + (M_{ZW}X) e'_{LIML} (X, W) \cdot o_P(1) \\
&= (M_{ZW} \tilde{U}, 0) + o_P(1),
\end{aligned}$$

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

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

Using Lemmas A4 and A5(II),

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

then

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

and so

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

Next,

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

Using Lemmas A4 and A5(II),

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

and thus

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

hence

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

**Proof of Theorem 9.** 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 3, one gets the conclusion.

**Proof of Corollary to Theorem 9.** Follow the proof of Corollary to Theorem 5.

## References

- ANATOLYEV, S. (2011): “Inference in regression models with many regressors,” *Journal of Econometrics*, accepted for publication.
- 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.
- 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 (2010): “Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments,” University of Maryland and MIT Working Paper.
- 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,” University of Maryland and MIT Working Paper.
- HAUSMAN, J.A., J.H. STOCK, AND M. YOGO (2005): “Asymptotic properties of the Hahn–Hausman test for weak instruments,” *Economics Letters*, 89, 333–342.

- 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 (2009): “A specification test for instrumental variables regression with many instruments,” Manuscript, University of Michigan.
- 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.
- NEWBY, W. K. (2004): “Many instrument asymptotics,” manuscript, MIT.
- 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.