Quantitative Economics 3 (2012), 211–255 1759-7331/20120211

Instrumental variable estimation with heteroskedasticity and many instruments

Jerry A. Hausman
Department of Economics, M.I.T.

Whitney K. Newey
Department of Economics, M.I.T.

Tiemen Woutersen
Department of Economics, University of Arizona

John C. Chao
Department of Economics, University of Maryland

Norman R. Swanson
Department of Economics, Rutgers University

This paper gives a relatively simple, well behaved solution to the problem of many instruments in heteroskedastic data. Such settings are common in microeconometric applications where many instruments are used to improve efficiency and allowance for heteroskedasticity is generally important. The solution is a Fuller (1977) like estimator and standard errors that are robust to heteroskedasticity and many instruments. We show that the estimator has finite moments and high asymptotic efficiency in a range of cases. The standard errors are easy to compute, being like White's (1982), with additional terms that account for many instruments. They are consistent under standard, many instrument, and many weak instrument asymptotics. We find that the estimator is asymptotically as efficient as the limited-information maximum likelihood (LIML) estimator under many weak instruments. In Monte Carlo experiments, we find that the estimator performs as well as LIML or Fuller (1977) under homoskedasticity, and has much lower bias and dispersion under heteroskedasticity, in nearly all cases considered.

Jerry A. Hausman: jhausman@mit.edu
Whitney K. Newey: wnewey@mit.edu
Tiemen Woutersen: woutersen@email.arizona.edu
John C. Chao: chao@econ.umd.edu
Norman R. Swanson: nswanson@econ.rutgers.edu
The NSF provided financial support for this paper under Grant 0136869. Helpful comments were provided by A. Chesher and participants in seminars at CalTech, CEMMAP, Harvard, MIT, Pittsburgh, UC Berkeley, UCL, and USC. Capable research assistance was provided by H. Arriizumi, S. Chang, A. Kowalski, R. Lewis, N. Lott, and K. Menzel. K. Menzel derived the vectorized form of the variance estimator.

DOI: 10.3982/QE89

Keywords. Instrumental variables, heteroskedasticity, many instruments, jackknife.

JEL classification. C12, C13, C23.

1. Introduction

This paper gives a relatively simple, well behaved solution to the problem of many instruments in heteroskedastic data. Such settings are common in microeconometric applications where many instruments are used to improve efficiency and allowance for heteroskedasticity is generally important. The solution is a Fuller (1977) like estimator and standard errors that are robust to heteroskedasticity and many instruments. We show that the estimator has finite moments and high asymptotic efficiency in a range of cases. The standard errors are easy to compute, being like White's (1982), with additional terms that account for many instruments. They are consistent under standard, many instrument, and many weak instrument asymptotics. They extend Bekker's (1994) standard errors to the heteroskedastic case.

The estimator that we refer to as HFUL is based on a jackknife version of the limited-information maximum likelihood (LIML) estimator, referred to as HLIM. The name HFUL is an abbreviation for the heteroskedasticity robust version of the Fuller (1977) estimator, while HLIM stands for the heteroskedasticity robust version of the LIML estimator. We show that HFUL has moments and, in Monte Carlo experiments, has much lower dispersion than HLIM with weak identification, an advantage analogous to that of the Fuller (1977) estimator over LIML with homoskedasticity. Hahn, Hausman, and Kuersteiner (2004) pointed out this problem for LIML and we follow them in referring to it as the "moments problem," because large dispersion corresponds to nonexistence of moments there.

HFUL is robust to heteroskedasticity and many instruments because of its jackknife form. Previously proposed jackknife instrumental variable (JIV) estimators are also known to be robust to heteroskedasticity and many instruments; see Phillips and Hale (1977), Blomquist and Dahlberg (1999), Angrist, Imbens, and Krueger (1999), Ackerberg and Deveraux (2003), and Chao and Swanson (2004). HFUL is better than these estimators because it is as efficient as LIML under many weak instruments and homoskedasticity, and so overcomes the efficiency problems for JIV noted in Davidson and MacKinnon (2006). Thus, HFUL provides a relatively efficient estimator for many instruments with heteroskedasticity that does not suffer from the moments problem.

Bekker and van der Ploeg (2005) proposed an interesting consistent estimators with many dummy instrumental variables and group heteroskedasticity, but these results are restrictive. For high efficiency, it is often important to use instruments that are not dummy variables. For example, linear instrumental variables can be good first approximations to optimal nonlinear instruments. HFUL allows for general instrumental variables and unrestricted heteroskedasticity, as does the asymptotics given here.

Newey and Windmeijer (2009) showed that the continuously updated generalized method of moments estimator and other generalized empirical likelihood estimators are robust to heteroskedasticity and many weak instruments, and asymptotically efficient under that asymptotics relative to JIV. However, this efficiency depends on using a heteroskedasticity consistent weighting matrix that can degrade the finite sample performance of continuously updated estimators (CUE) with many instruments, as shown in Monte Carlo experiments here. HFUL continues to have good properties under many instrument asymptotics, rather than just many weak instruments. The properties of CUE are likely to be poor under many instruments asymptotics due to the heteroskedasticity consistent weighting matrix. Also CUE is quite difficult to compute and tends to have large dispersion under weak identification, which HFUL does not. Thus, relative to CUE, HFUL provides a computationally simpler solution with better finite sample properties.

The need for HFUL is motivated by the inconsistency of LIML and the Fuller (1977) estimator under heteroskedasticity and many instruments. The inconsistency of LIML was pointed out by Bekker and van der Ploeg (2005) and Chao and Swanson (2004) in special cases. We give a characterization of the inconsistency here, showing the precise restriction on the heteroskedasticity that would be needed for LIML to be consistent.

The asymptotic theory we consider allows for many instruments as in Kunitomo (1980) and Bekker (1994) or many weak instruments as in Chao and Swanson (2004, 2005), Stock and Yogo (2005), Han and Phillips (2006), and Andrews and Stock (2007). The asymptotic variance estimator will be consistent for any of standard, many instrument, or many weak instrument asymptotics. Asymptotic normality is obtained via a central limit theorem that imposes weak conditions on instrumental variables, given by Chao, Swanson, Hausman, Newey, and Woutersen (2012b). Although the inference methods will not be valid under the weak instrument asymptotics of Staiger and Stock (1997), we do not consider this to be very important. Hansen, Hausman, and Newey's (2008) survey of the applied literature suggests that the weak instrument approximation is not needed very often in microeconomic data, where we focus our attention.

In Section 2, the model is outlined and a practitioner's guide to the estimator is given. We give there simple formulae for HFUL and its variance estimator. Section 3 motivates HLIM and HFUL as jackknife forms of LIML and Fuller (1977) estimators, and discusses some of their properties. Section 5 presents our Monte Carlo findings, while our conclusion is given in Section 6. The Appendix gives a theorem on the existence of moments of HFUL and also presents proofs of our asymptotic results. The proof of the existence of moments theorem can be found in a supplementary file on the journal website, http://qeconomics.org/supp/89/supplement.pdf.

2. The model and HFUL

The model we consider is given by

$$\underset{n\times 1}{y} = \underset{n\times G}{X}\,\underset{G\times 1}{\delta_0} + \underset{n\times 1}{\varepsilon},$$
$$X = Y + U,$$

where n is the number of observations, G is the number of right-hand side variables, Y is a matrix of observations on the reduced form, and U is the matrix of reduced form disturbances. For our asymptotic approximations, the elements of Y will be implicitly allowed to depend on n, although we suppress dependence of Y on n for notational convenience. Estimation of δ_0 will be based on an $n \times K$ matrix, Z, of instrumental variable observations with $\text{rank}(Z) = K$. We will assume that Z is nonrandom and that observations (ε_i, U_i) are independent across i and have mean zero. Alternatively, we could allow Z to be random, but condition on it, as in Chao et al. (2012b).

In this model, some columns of X may be exogenous, with the corresponding columns of U being zero. Also, this model allows for Y to be a linear combination of Z, that is, $Y = Z\pi$ for some $K \times G$ matrix π. The model also permits Z to approximate the reduced form. For example, let X_i', Y_i', and Z_i' denote the ith row (observation) of X, Y, and Z, respectively. We could let $Y_i = f_0(w_i)$ be a vector of unknown functions of a vector w_i of underlying instruments and let $Z_i = (p_{1K}(w_i), \dots, p_{KK}(w_i))'$ be approximating functions $p_{kK}(w)$, such as power series or splines. In this case, linear combinations of Z_i may approximate the unknown reduced form (e.g., as in Newey (1990)).

To describe HFUL, let

$$P = Z(Z'Z)^{-1}Z',$$

let P_{ij} denote the ijth element of P, and let $\bar{X} = [y, X]$. Let $\bar{X}_i'$ be a row vector consisting of the ith row of $\bar{X}$ and let

$$\tilde{\alpha} \text{ be the smallest eigenvalue of } (\bar{X}'\bar{X})^{-1}\Big(\bar{X}'P\bar{X} - \sum_{i=1}^n P_{ii}\bar{X}_i\bar{X}_i'\Big).$$

Although this matrix is not symmetric, it has real eigenvalues.[1] For a constant C, let

$$\hat{\alpha} = [\tilde{\alpha} - (1-\tilde{\alpha})C/n]/[1-(1-\tilde{\alpha})C/n].$$

In the Monte Carlo results given below, we try different values of C and recommend $C = 1$.[2] HFUL is given by

$$\hat{\delta} = \left(X'PX - \sum_{i=1}^n P_{ii}X_iX_i' - \hat{\alpha}X'X\right)^{-1}\left(X'Py - \sum_{i=1}^n P_{ii}X_iy_i - \hat{\alpha}X'y\right). \tag{1}$$

Thus, HFUL can be computed by finding the smallest eigenvalue of a matrix and then using this explicit formula.

To describe the asymptotic variance estimator, let $\hat{\varepsilon}_i = y_i - X_i'\hat{\delta}$, $\hat{\gamma} = X'\hat{\varepsilon}/\hat{\varepsilon}'\hat{\varepsilon}$, $\hat{X} = X - \hat{\varepsilon}\hat{\gamma}'$, $\dot{X} = P\hat{X}$, and $\tilde{Z} = Z(Z'Z)^{-1}$. Also let

$$\hat{H} = X'PX - \sum_{i=1}^n P_{ii}X_iX_i' - \hat{\alpha}X'X,$$

[1] Note that solving $\det\{(\bar{X}'\bar{X})^{-1}(\bar{X}'P\bar{X} - \sum_{i=1}^n P_{ii}\bar{X}_i\bar{X}_i') - \lambda I\} = 0$ is equivalent to solving $\det\{(\bar{X}'P\bar{X} - \sum_{i=1}^n P_{ii}\bar{X}_i\bar{X}_i') - \lambda(\bar{X}'\bar{X})\} = 0$. Moreover, this is equivalent to solving $\det\{(\bar{X}'\bar{X})^{1/2}\}\det\{(\bar{X}'\bar{X})^{-1/2}(\bar{X}'P\bar{X} - \sum_{i=1}^n P_{ii}\bar{X}_i\bar{X}_i')(\bar{X}'\bar{X})^{-1/2} - \lambda I\}\det\{(\bar{X}'\bar{X})^{1/2}\} = 0$. As is well know, a sufficient condition for the eigenvalue to be real is that the matrix is real and symmetric, and this condition is satisfied here almost surely. Thus, λ is real almost surely.

[2] Fuller (1977) made a degrees of freedom correction in the choice of C; the existence of finite sample moments and the large sample properties of the estimator are not affected by this correction.

$$\hat{\Sigma}=\sum_{i=1}^{n}(\dot{X}_i\dot{X}_i'-\hat{X}_iP_{ii}\dot{X}_i'-\dot{X}_iP_{ii}\hat{X}_i')\hat{\varepsilon}_i^2$$
$$+\sum_{k=1}^{K}\sum_{\ell=1}^{K}\left(\sum_{i=1}^{n}\tilde{Z}_{ik}\tilde{Z}_{i\ell}\hat{X}_i\hat{\varepsilon}_i\right)\left(\sum_{j=1}^{n}Z_{jk}Z_{j\ell}\hat{X}_j\hat{\varepsilon}_j\right)',$$

The formula for $\hat{\Sigma}$ is vectorized in such a way that it can easily be computed even when the sample size, n, is very large. The asymptotic variance estimator is

$$\hat{V}=\hat{H}^{-1}\hat{\Sigma}\hat{H}^{-1}.$$

This asymptotic variance estimator will be consistent under standard, many instrument, and many weak instrument asymptotics.

This asymptotic variance estimator can be used to do large sample inference in the usual way under the conditions of Section 4. This is done by treating $\hat{\delta}$ as if it were normally distributed with mean δ_0 and variance $\hat{V}$. Asymptotic t-ratios $\hat{\delta}_j/\sqrt{\hat{V}_{jj}}$ will be asymptotically normal. Also, defining q_α as the $1-\alpha/2$ quantile of a $N(0,1)$ distribution, an asymptotic $1-\alpha$ confidence interval for δ_{0k} is given by $\hat{\delta}_k\pm q_\alpha\sqrt{\hat{V}_{kk}}$. More generally, a confidence interval for a linear combination $c'\delta$ can be formed as $c'\hat{\delta}\pm q_\alpha\sqrt{c'\hat{V}c}$. We find in the Monte Carlo results that these asymptotic confidence intervals are very accurate in a range of finite sample settings.

3. Consistency with many instruments and heteroskedasticity

In this section, we explain the HFUL estimator, why it has moments, why it is robust to heteroskedasticity and many instruments, and why it has high efficiency under homoskedasticity. We also compare it with other estimators and briefly discuss some of their properties. To do so, it is helpful to consider each estimator as a minimizer of an objective function. As usual, the limit of the minimizer will be the minimizer of the limit under appropriate regularity conditions, so estimator consistency can be analyzed using the limit of the objective function. This amounts to a modern version of method of moments interpretations of consistency that has now become common in econometrics (Amemiya (1973, 1984), Newey and McFadden (1994)).

To motivate HFUL, it is helpful to begin with two-stage least squares (2SLS). The 2SLS estimator minimizes

$$\hat{Q}_{\mathrm{2SLS}}(\delta)=(y-X\delta)'P(y-X\delta)/n.$$

The limit of this function will equal the limit of its expectation under general conditions. With independent observations,

$$E[\hat{Q}_{\mathrm{2SLS}}(\delta)]=(\delta-\delta_0)'A_n(\delta-\delta_0)+\sum_{i=1}^{n}P_{ii}E[(y_i-X_i'\delta)^2]/n,$$

$$A_n=Y'PY/n-\sum_{i=1}^{n}P_{ii}Y_iY_i'/n.$$

The matrix A_n will be positive definite under conditions given below, so that the first term $(\delta - \delta_0)' A_n (\delta - \delta_0)$ will be minimized at δ_0. The second term $\sum_{i=1}^n P_{ii} E[(y_i - X_i'\delta)^2]/n$ is an expected squared residual that will not be minimized at δ_0 due to endogeneity. With many (weak) instruments, P_{ii} does not shrink to zero (relative to the first term), so that the second term does not vanish asymptotically (relative to the first). Hence, with many (weak) instruments, 2SLS is not consistent, even under homoskedasticity, as pointed out by Bekker (1994). This objective function calculation for 2SLS is also given in Han and Phillips (2006), though the following analysis is not.

A way to modify the objective function so that it gives a consistent estimator is to remove the term whose expectation is not minimized at δ_0. This leads to an objective function of the form

$$\hat{Q}_{\text{JIV}}(\delta) = \sum_{i \neq j} (y_i - X_i'\delta)' P_{ij} (y_j - X_j'\delta)/n.$$

The expected value of this objective function is

$$E[\hat{Q}_{\text{JIV}}(\delta)] = (\delta - \delta_0)' A_n (\delta - \delta_0),$$

which is minimized at $\delta = \delta_0$. Thus, the estimator minimizing $\hat{Q}_{\text{JIV}}(\delta)$ should be consistent. Solving the first order conditions gives

$$\hat{\delta}_{\text{JIV}} = \left(\sum_{i \neq j} X_i' P_{ij} X_j \right)^{-1} \sum_{i \neq j} X_i' P_{ij} y_j.$$

This is the JIVE2 estimator of Angrist, Imbens, and Krueger (1999). Since the objective function for $\hat{\delta}_{\text{JIV}}$ has expectation minimized at δ_0, we expect that $\hat{\delta}_{\text{JIV}}$ is consistent, as has already been shown by Ackerberg and Deveraux (2003) and Chao and Swanson (2004). Other JIV estimators have also been shown to be consistent in these papers.

So far, we have only used the objective function framework to describe previously known consistency results. We now use it to motivate the form of HFUL (and HLIM).

A problem with JIV estimators, pointed out by Davidson and MacKinnon (2006), is that they can have low efficiency relative to LIML under homoskedasticity. This problem can be avoided by using a jackknife version of LIML. The LIML objective function is

$$\hat{Q}_{\text{LIML}}(\delta) = \frac{(y - X\delta)' P (y - X\delta)}{(y - X\delta)'(y - X\delta)}.$$

The numerator of $\hat{Q}_{\text{LIML}}(\delta)$ is $n\hat{Q}_{\text{2SLS}}(\delta)$. If we replace this numerator with $n\hat{Q}_{\text{JIV}}(\delta)$, we obtain

$$\hat{Q}_{\text{HLIM}}(\delta) = \frac{\sum_{i \neq j} (y_i - X_i'\delta)' P_{ij} (y_j - X_j'\delta)}{(y - X\delta)'(y - X\delta)}.$$

The minimizer of this objective function is the HLIM estimator that we denote by $\tilde{\delta}$. This estimator is consistent with many instruments and heteroskedasticity. It is also as efficient asymptotically and performs as well in our Monte Carlo results as LIML under homoskedasticity, thus overcoming the Davidson and MacKinnon (2006) objection to JIV.

The use of the JIV objective function in the numerator makes this estimator consistent with heteroskedasticity and many instruments. In large samples, the HLIM objective function will be close to

$$\frac{E[n\hat{Q}_{\mathrm{JIV}}(\delta)]}{E[(y-X\delta)'(y-X\delta)]}=\frac{(\delta-\delta_0)'A_n(\delta-\delta_0)}{E[(y-X\delta)'(y-X\delta)]}.$$

This function is minimized at $\delta=\delta_0$, even with heteroskedasticity and many instruments, leading to consistency of HLIM.

Computation of HLIM is straightforward. For $\bar{X}=[y,X]$, the minimized objective function $\tilde{\alpha}=\hat{Q}_{\mathrm{HLIM}}(\tilde{\delta})$ is the smallest eigenvalue of $(\bar{X}'\bar{X})^{-1}(\bar{X}'P\bar{X}-\sum_{i=1}^{n}P_{ii}\bar{X}_i\bar{X}_i')$. Solving the first order conditions gives

$$\tilde{\delta}=\left(X'PX-\sum_{i=1}^{n}P_{ii}X_iX_i'-\tilde{\alpha}X'X\right)^{-1}\left(X'Py-\sum_{i=1}^{n}P_{ii}X_iy_i-\tilde{\alpha}X'y\right).$$

The formula for HLIM is analogous to that of LIML where the own observation terms have been removed from the double sums involving P. Also, HLIM is invariant to normalization, similarly to LIML, although HFUL is not. The vector $\tilde{d}=(1,-\tilde{\delta}')'$ solves

$$\min_{d:d_1=1}\frac{d'\left(\bar{X}'P\bar{X}-\sum_{i=1}^{n}P_{ii}\bar{X}_i\bar{X}_i'\right)d}{d'\bar{X}'\bar{X}d}.$$

Because of the ratio form of the objective function, another normalization, such as imposing that another d is equal to 1, would produce the same estimator, up to the normalization.

Like LIML, the HLIM estimator suffers from the moments problem, having large dispersion with weak instruments, as shown in the Monte Carlo results below. Hahn, Hausman, and Kuersteiner (2004) suggested the Fuller (1977) estimator as a solution to this problem for LIML. We suggest the HFUL as a solution to this potential problem with HLIM. HFUL is obtained analogously to Fuller (1977) by replacing the eigenvalue $\tilde{\alpha}$ in the HLIM estimator with $\hat{\alpha}=[\tilde{\alpha}-(1-\tilde{\alpha})C/n]/[1-(1-\tilde{\alpha})C/n]$, giving the HFUL estimator of equation (1). We show that this estimator does have moments and low dispersion with weak instruments, thus providing a solution to the moments problem.

HFUL, HLIM, and JIV are members of a class of estimators of the form

$$\bar{\delta}=\left(X'PX-\sum_{i=1}^{n}P_{ii}X_iX_i'-\bar{\alpha}X'X\right)^{-1}\left(X'Py-\sum_{i=1}^{n}P_{ii}X_iy_i-\bar{\alpha}X'y\right).$$

This might be thought of as a type of k-class estimator that is robust to heteroskedasticity and many instruments. HFUL takes this form as in equation (1); HLIM does with $\bar{\alpha}=\tilde{\alpha}$ and JIV does with $\bar{\alpha}=0$.

HLIM can also be interpreted as a jackknife version of the continuously updated GMM estimator, and as an optimal linear combination of forward and reverse JIV estimators, analogously to Hahn and Hausman's (2002) interpretation of LIML as an optimal linear combination of forward and reverse bias corrected estimators. For brevity, we do not give these interpretations here.

HFUL is motivated by the inconsistency of LIML and Fuller (1977) with many instruments and heteroskedasticity. To give precise conditions for LIML inconsistency, note that in large samples, the LIML objective function will be close to

$$\frac{E[\hat{Q}_{2SLS}(\delta)]}{E[(y-X\delta)'(y-X\delta)]} = \frac{(\delta-\delta_0)'A_n(\delta-\delta_0)}{E[(y-X\delta)'(y-X\delta)]} + \frac{\sum_{i=1}^{n} P_{ii}E[(y_i-X_i'\delta)^2]}{E[(y-X\delta)'(y-X\delta)]}.$$

The first term following the equality will be minimized at δ_0. The second term may not have a critical value at δ_0, so the objective function will not be minimized at δ_0. To see this, let $\sigma_i^2 = E[\varepsilon_i^2]$, $\gamma_i = E[X_i\varepsilon_i]/\sigma_i^2$, and $\bar{\gamma} = \sum_{i=1}^n E[X_i\varepsilon_i]/\sum_{i=1}^n \sigma_i^2 = \sum_i \gamma_i\sigma_i^2/\sum_i \sigma_i^2$. Then

$$\left.\frac{\partial}{\partial\delta}\frac{\sum_{i=1}^{n} P_{ii}E[(y_i-X_i\delta)^2]}{\sum_{i=1}^{n} E[(y_i-X_i\delta)^2]}\right|_{\delta=\delta_0} = \frac{-2}{\sum_{i=1}^{n}\sigma_i^2}\left[\sum_{i=1}^{n} P_{ii}E[X_i\varepsilon_i] - \sum_{i=1}^{n} P_{ii}\sigma_i^2\bar{\gamma}\right]$$

$$= \frac{-2\sum_{i=1}^{n} P_{ii}(\gamma_i-\bar{\gamma})\sigma_i^2}{\sum_{i=1}^{n}\sigma_i^2} = -2\widehat{\mathrm{Cov}_{\sigma^2}(P_{ii}},\gamma_i),$$

where $\widehat{\mathrm{Cov}_{\sigma^2}(P_{ii}},\gamma_i)$ is the covariance between P_{ii} and γ_i, for the distribution with probability weight $\sigma_i^2/\sum_{i=1}^n \sigma_i^2$ for the ith observation. When

$$\lim_{n\longrightarrow\infty} \widehat{\mathrm{Cov}_{\sigma^2}(P_{ii}},\gamma_i) \neq 0,$$

the LIML objective function will not have a zero derivative at δ_0 asymptotically, so it is not minimized at δ_0. Bekker and van der Ploeg (2005) and Chao and Swanson (2004) pointed out that LIML can be inconsistent with heteroskedasticity; the contribution here is to give the exact condition $\lim_{n\longrightarrow\infty} \widehat{\mathrm{Cov}_{\sigma^2}(P_{ii}},\gamma_i) = 0$ for consistency of LIML.

Note that $\widehat{\mathrm{Cov}_{\sigma^2}(P_{ii}},\gamma_i) = 0$ when either γ_i or P_{ii} does not depend on i. Thus, it is variation in $\gamma_i = E[X_i\varepsilon_i]/\sigma_i^2$, the coefficients from the projection of X_i on ε_i, that leads to inconsistency of LIML, and not just any heteroskedasticity. Also, the case where P_{ii} is constant occurs with dummy instruments and equal group sizes. It was pointed out by

Bekker and van der Ploeg (2005) that LIML is consistent in this case, under heteroskedasticity. Indeed, when P_{ii} is constant,

$$\hat{Q}_{\text{LIML}}(\delta) = \hat{Q}_{\text{HLIM}}(\delta) + \frac{\sum_i P_{ii}(y_i - X_i'\delta)^2}{(y - X\delta)'(y - X\delta)} = \hat{Q}_{\text{HLIM}}(\delta) + P_{11},$$

so that the LIML objective function equals the HLIM objective function plus a constant, and hence HLIM equals LIML.

When the instrumental variables are dummy variables, HLIM can be related to the method of moments (MM) estimator of Bekker and van der Ploeg (2005). Both are minimizers of ratios of quadratic forms. In notes that are available from the authors, we show that the numerator of the quadratic form for the MM estimator can be interpreted as an objective function that is minimized by the JIVE1 estimator of Angrist, Imbens, and Krueger (1999). In this sense, the MM estimator can be thought of as being related to the JIVE1 estimator, while we use the numerator from the JIVE2 estimator. The denominator of the MM estimator is different than $(y - X\delta)'(y - X\delta)$, but has a similar effect of making the MM estimator have properties similar to LIML under homoskedasticity and Gaussian disturbances.

4. Asymptotic theory

Theoretical justification for the estimators is provided by asymptotic theory where the number of instruments grows with the sample size. Some regularity conditions are important for this theory. Let Z_i', ε_i, U_i', and Y_i' denote the ith row of Z, ε, U, and Y, respectively. Here, we will consider the case where Z is constant, which can be viewed as conditioning on Z (see, e.g., Chao et al. (2012b)).

Assumption 1. *Z includes among its columns a vector of ones*, $\text{rank}(Z) = K$, *and there is a constant C such that $P_{ii} \le C < 1$ $(i = 1, \dots, n)$, $K \longrightarrow \infty$.*

The restriction that $\text{rank}(Z) = K$ is a normalization that requires excluding redundant columns from Z. It can be verified in particular cases. For instance, when w_i is a continuously distributed scalar, $Z_i = p^K(w_i)$, and $p_{kK}(w) = w^{k-1}$, it can be shown that $Z'Z$ is nonsingular with probability 1 for $K < n$.[3] The condition $P_{ii} \le C < 1$ implies that $K/n \le C$, because $K/n = \sum_{i=1}^n P_{ii}/n \le C$.

The next condition specifies that the reduced form Y_i is a linear combination of a set of variables z_i having certain properties.

Assumption 2. *$Y_i = S_n z_i/\sqrt{n}$, where $S_n = \tilde{S}\,\text{diag}(\mu_{1n}, \dots, \mu_{Gn})$ and $\tilde{S}$ is nonsingular. Also, for each j, either $\mu_{jn} = \sqrt{n}$ or $\mu_{jn}/\sqrt{n} \longrightarrow 0$, $\mu_n = \min_{1\le j\le G} \mu_{jn} \longrightarrow \infty$, and $\sqrt{K}/\mu_n^2 \longrightarrow 0$. Also, there is $C > 0$ such that $\|\sum_{i=1}^n z_i z_i'/n\| \le C$ and $\lambda_{\min}(\sum_{i=1}^n z_i z_i'/n) \ge 1/C$ for n sufficiently large.*

[3]The observations $w_1, \dots, w_n$ are distinct with probability 1 and, therefore, by $K < n$, cannot all be roots of a Kth degree polynomial. It follows that for any nonzero a, there must be some i with $a'Z_i = a'p^K(w_i) \ne 0$, implying that $a'Z'Za > 0$.

This condition is similar to Assumption 2 of Hansen, Hausman, and Newey (2008). It accommodates linear models where included instruments (e.g., a constant) have fixed reduced form coefficients and excluded instruments have coefficients that can shrink as the sample size grows. A leading example of such a model is a linear structural equation with one endogenous variable,

$$y_i = Z_{i1}'\delta_{01} + \delta_{0G}X_{iG} + \varepsilon_i, \tag{2}$$

where Z_{i1} is a $G_1 \times 1$ vector of included instruments (e.g., including a constant) and X_{iG} is an endogenous variable. Here the number of right-hand side variables is $G_1 + 1 = G$. Let the reduced form be partitioned conformably with δ, as $Y_i = (Z_{i1}', Y_{iG})'$ and $U_i = (0, U_{iG})'$. Here the disturbances for the reduced form for Z_{i1} are zero because Z_{i1} is taken to be exogenous. Suppose that the reduced form for X_{iG} depends linearly on the included instrumental variables Z_{i1} and on an excluded instrument z_{iG} as in

$$X_{iG} = Y_{iG} + U_{iG}, \quad Y_{iG} = \pi_1 Z_{i1} + (\mu_n/\sqrt{n})z_{iG}.$$

Here we normalize z_{iG} so that μ_n determines how strongly δ_G is identified and absorb into z_{iG} any other terms, such as unknown coefficients. For Assumption 2, we let $z_i = (Z_{i1}', z_{iG})'$ and require that the second moment matrix of z_i is bounded and bounded away from zero. This is the normalization that makes the strength of identification of δ_G be determined by μ_n. For example, if $\mu_n = \sqrt{n}$, then the coefficient on z_{iG} does not shrink, corresponding to strong identification of δ_G. If μ_n grows slower than $\sqrt{n}$, then δ_G will be more weakly identified. Indeed, $1/\mu_n$ will be the convergence rate for estimators of δ_G. We require $\mu_n \longrightarrow \infty$ to avoid the weak instrument setting of Staiger and Stock (1997), where δ_G is not asymptotically identified.

For this model, the reduced form is

$$Y_i = \begin{bmatrix} Z_{i1} \\ \pi_1 Z_{i1} + (\mu_n/\sqrt{n})z_{iG} \end{bmatrix} = \begin{bmatrix} I & 0 \\ \pi_1 & 1 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \mu_n/\sqrt{n} \end{bmatrix} \begin{pmatrix} Z_{i1} \\ z_{iG} \end{pmatrix}.$$

This reduced form is as specified in Assumption 2 with

$$\tilde{S}_n = \begin{bmatrix} I & 0 \\ \pi_1 & 1 \end{bmatrix}, \qquad \mu_{jn} = \sqrt{n}, \quad 1 \le j \le G_1, \qquad \mu_{Gn} = \mu_n.$$

Note how this somewhat complicated specification is needed to accommodate fixed reduced form coefficients for included instrumental variables and excluded instruments with identifying power that depends on n. We have been unable to simplify Assumption 2 while maintaining the generality needed for such important cases.

We will not require that z_{iG} be known, only that it be approximated by a linear combination of the instrumental variables $Z_i = (Z_{i1}', Z_{i2}')'$. Implicitly, it is also allowed to depend on n, as is Z_{i1}. One important case is where the excluded instrument z_{iG} is an unknown linear combination of the instrumental variables $Z_i = (Z_{i1}', Z_{i2}')'$. For example, one of the cases examined in the many weak instrument setting of Chao and Swanson (2005) is where the reduced form is given by

$$Y_{iG} = \pi_1 Z_{i1} + (\pi_2/\sqrt{n})' Z_{i2}$$

for a $K - G_1$ dimensional vector Z_{i2} of excluded instrumental variables. This particular case can be folded into our framework by specifying that

$$z_{iG} = \pi_2' Z_{i2}/\sqrt{K - G_1}, \qquad \mu_n = \sqrt{K - G_1}.$$

Assumption 2 will then require that

$$\sum_i z_{iG}^2/n = (K - G_1)^{-1} \sum_i (\pi_2' Z_{i2})^2/n$$

is bounded and bounded away from zero. Thus, the second moment $\sum_i (\pi_2' Z_{i2})^2/n$ of the term in the reduced form that identifies δ_{0G} must grow linearly in K, leading to a convergence rate of $1/\sqrt{K - G_1} = 1/\mu_n$.

In another important case, the excluded instrument z_{iG} could be an unknown function that can be approximated by a linear combination of Z_i. For instance, suppose that $z_{iG} = f_0(w_i)$ for an unknown function $f_0(w_i)$ of variables w_i. In this case, the instrumental variables could include a vector $p^K(w_i) \stackrel{\text{def}}{=} (p_{1K}(w_i), \dots, p_{K-G_1,K}(w_i))'$ of approximating functions, such as polynomials or splines. Here the vector of instrumental variables would be $Z_i = (Z_{i1}', p^K(w_i)')'$. For $\mu_n = \sqrt{n}$, this example is like Newey (1990), where Z_i includes approximating functions for the reduced form but the number of instruments can grow as fast as the sample size. Alternatively, if $\mu_n/\sqrt{n} \longrightarrow 0$, it is a modified version where δ_G is more weakly identified.

Assumption 2 also allows for multiple endogenous variables with a different strength of identification for each one, leading to different convergence rates. In the above example, we maintained the scalar endogenous variable for simplicity.

The μ_n^2 can be thought of as a version of the concentration parameter, determining the convergence rate of estimators of δ_{0G}, just as the concentration parameter does in other settings. For $\mu_n^2 = n$, the convergence rate will be $\sqrt{n}$, where Assumptions 1 and 2 permit K to grow as fast as the sample size, corresponding to a many instrument asymptotic approximation like Kunitomo (1980), Morimune (1983), and Bekker (1994). For μ_n^2 growing slower than n, the convergence rate will be slower than $1/\sqrt{n}$, leading to an asymptotic approximation like that of Chao and Swanson (2005).

Assumption 3. *There is a constant $C > 0$ such that $(\varepsilon_1, U_1), \dots, (\varepsilon_n, U_n)$ are independent, with $E[\varepsilon_i] = 0$, $E[U_i] = 0$, $E[\varepsilon_i^2] < C$, $E[\|U_i\|^2] \leq C$, $\operatorname{Var}((\varepsilon_i, U_i')') = \operatorname{diag}(\Omega_i^*, 0)$, and $\lambda_{\min}(\sum_{i=1}^n \Omega_i^*/n) \geq 1/C$.*

This assumption requires the second conditional moments of disturbances to be bounded. It also imposes uniform nonsingularity of the variance of the reduced form disturbances, which is useful in the consistency proof, to help the denominator of the objective function stay away from zero.

Assumption 4. *There is a π_{K_n} such that $\sum_{i=1}^n \|z_i - \pi_{K_n} Z_i\|^2/n \longrightarrow 0$.*

This condition and $P_{ii} \le C < 1$ will imply that for a large enough sample,

$$A_n = Y'PY/n - \sum_{i=1}^{n} P_{ii} Y_i Y_i'/n = \sum_{i=1}^{n}(1 - P_{ii}) Y_i Y_i'/n - Y'(I - P)Y/n$$
$$= \sum_{i=1}^{n}(1 - P_{ii}) Y_i Y_i'/n + o(1) \ge (1 - C)\sum_{i=1}^{n} Y_i Y_i'/n,$$

so that A_n is positive definite in large enough samples. Also, Assumption 4 is not very restrictive, because flexibility is allowed in the specification of Y_i. If we simply make Y_i the expectation of X_i given the instrumental variables, then Assumption 4 holds automatically.

These conditions imply estimator consistency:

Theorem 1. *If Assumptions 1–4 are satisfied and either* (i) $\hat{\delta}$ *is HLIM,* (ii) $\hat{\delta}$ *is HFUL, or* (iii) $\hat{\alpha} = o_p(\mu_n^2/n)$*, then* $\mu_n^{-1} S_n'(\hat{\delta} - \delta_0) \xrightarrow{p} 0$ *and* $\hat{\delta} \xrightarrow{p} \delta_0$.

This result gives convergence rates for linear combinations of $\hat{\delta}$. For instance, in the above example, it implies that $\hat{\delta}_1$ is consistent and that $\pi_{11}'\hat{\delta}^1 + \hat{\delta}^2 = o_p(\mu_n/\sqrt{n})$.

For asymptotic normality, it is helpful to strengthen the conditions on moments.

Assumption 5. *There is a constant,* $C > 0$*, such that with probability* 1, $\sum_{i=1}^{n} \|z_i\|^4/n^2 \longrightarrow 0$, $E[\varepsilon_i^4] \le C$*, and* $E[\|U_i\|^4] \le C$.

To state a limiting distribution result, it is helpful to also assume that certain objects converge and to allow for two cases of growth rates of K relative to μ_n^2. Also, the asymptotic variance of the estimator will depend on the growth rate of K relative to μ_n^2. Let $\sigma_i^2 = E[\varepsilon_i^2]$ and $\gamma_n = \sum_{i=1}^{n} E[U_i\varepsilon_i]/\sum_{i=1}^{n}\sigma_i^2$, $\tilde{U} = U - \varepsilon\gamma_n'$, having ith row $\tilde{U}_i'$; and let $\tilde{\Omega}_i = E[\tilde{U}_i\tilde{U}_i']$.

Assumption 6. $\mu_n S_n^{-1} \longrightarrow S_0$ *and either of the following cases holds*:

Case I. $K/\mu_n^2 \longrightarrow \alpha$ *for finite* α.

Case II. $K/\mu_n^2 \longrightarrow \infty$.

Also $H_P = \lim_{n\longrightarrow\infty}\sum_{i=1}^{n}(1 - P_{ii}) z_i z_i'/n$, $\Sigma_P = \lim_{n\longrightarrow\infty}\sum_{i=1}^{n}(1 - P_{ii})^2 z_i z_i'\sigma_i^2/n$*, and* $\Psi = \lim_{n\longrightarrow\infty}\sum_{i\ne j} P_{ij}^2(\sigma_i^2 E[\tilde{U}_j\tilde{U}_j'] + E[\tilde{U}_i\varepsilon_i]E[\varepsilon_j\tilde{U}_j'])/K$ *exist.*

This convergence condition can be replaced by an assumption that certain matrices are uniformly positive definite without affecting the limiting distribution result for t-ratios given in Theorem 3 below (see Chao et al. (2012b)).

We can now state the asymptotic normality results. In Case I, we have that

$$S_n'(\hat{\delta} - \delta_0) \xrightarrow{d} N(0, \Lambda_{\mathrm{I}}), \tag{3}$$

where

$$\Lambda_{\mathrm{I}} = H_P^{-1}\Sigma_P H_P^{-1} + \alpha H_P^{-1} S_0 \Psi S_0' H_P^{-1}.$$

In Case II, we have that

$$(\mu_n/\sqrt{K})S_n'(\hat{\delta}-\delta_0)\xrightarrow{d} N(0,\Lambda_{\mathrm{II}}), \tag{4}$$

where

$$\Lambda_{\mathrm{II}} = H_P^{-1}S_0\Psi S_0' H_P^{-1}.$$

The asymptotic variance expressions allow for the many instrument sequence of Kunitomo (1980) and Bekker (1994), and the many weak instrument sequence of Chao and Swanson (2004, 2005). In Case I, the first term in the asymptotic variance, Λ_{I}, corresponds to the usual asymptotic variance, and the second term is an adjustment for the presence of many instruments. In Case II, the asymptotic variance, Λ_{II}, only contains the adjustment for many instruments. This is because K is growing faster than μ_n^2. Also, Λ_{II} will be singular when included exogenous variables are present.

We can now state an asymptotic normality result.

Theorem 2. *If Assumptions 1–6 are satisfied and $\hat{\alpha}=\tilde{\alpha}+O_p(1/n)$ or $\hat{\delta}$ is HLIM or HFUL, then in Case* I, *equation* (3) *is satisfied, and in Case* II, *equation* (4) *is satisfied.*

It is interesting to compare the asymptotic variance of the HFUL estimator with that of LIML when the disturbances are homoskedastic. First, note that the disturbances are not restricted to be Gaussian and that the asymptotic variance does not depend on third or fourth moments of the disturbances. In contrast, the asymptotic variance of LIML does depend on third and fourth moment terms for non-Gaussian disturbances; see Bekker and van der Ploeg (2005), Hansen, Hausman, and Newey (2008), and van Hasselt (2010). This makes estimation of the asymptotic variance simpler for HFUL than for LIML. It appears that the jackknife form of the numerator has this effect on HFUL. Deleting the own observation terms in effect removes moment conditions that are based on squared residuals. Bekker and van der Ploeg (2005) also found that the limiting distribution of their MM estimator for dummy instruments and group heteroskedasticity did not depend on third and fourth moments.

Under homoskedasticity, the variance of (ε_i, U_i') will not depend on i (e.g., so that $\sigma_i^2=\sigma^2$). Then $\gamma_n = E[X_i\varepsilon_i]/\sigma^2=\gamma$ and $E[\tilde{U}_i\varepsilon_i]=E[U_i\varepsilon_i]-\gamma\sigma^2=0$, so that

$$\Sigma_P=\sigma^2\tilde{H}_P,\quad \tilde{H}_P=\lim_{n\longrightarrow\infty}\sum_{i=1}^n(1-P_{ii})^2 z_i z_i'/n,$$

$$\Psi=\sigma^2 E[\tilde{U}_j\tilde{U}_j']\left(1-\lim_{n\longrightarrow\infty}\sum_{i=1}^n P_{ii}^2/K\right).$$

Focusing on Case I, letting $\Gamma=\alpha\sigma^2 S_0 E[\tilde{U}_i\tilde{U}_i']S_0'$, the asymptotic variance of HLIM is

$$V=\sigma^2 H_P^{-1}\tilde{H}_P H_P^{-1}+\lim_{n\longrightarrow\infty}\left(1-\sum_{i=1}^n P_{ii}^2/K\right)H_P^{-1}\Gamma H_P^{-1}.$$

For the variance of LIML, assume that third and fourth moments obey the same restrictions that they do under normality. Then from Hansen, Hausman, and Newey (2008), for $H=\lim_{n\longrightarrow\infty}\sum_{i=1}^{n} z_i z_i'/n$ and $\tau=\lim_{n\longrightarrow\infty} K/n$, the asymptotic variance of LIML is

$$V^{*}=\sigma^{2}H^{-1}+(1-\tau)^{-1}H^{-1}\Gamma H^{-1}.$$

With many weak instruments, where $\tau=0$ and $\max_{i\le n} P_{ii}\longrightarrow 0$, we will have $H_P=\tilde{H}_P=H$ and $\lim_{n\longrightarrow\infty}\sum_i P_{ii}^2/K\longrightarrow 0$, so that the asymptotic variances of HLIM and LIML are the same and are equal to $\sigma^2H^{-1}+H^{-1}\Gamma H^{-1}$. This case is most important in practical applications, where K is usually very small relative to n. In such cases, we would expect from the asymptotic approximation to find that the variances of LIML and HLIM are very similar.

In the many instruments case, where K and μ_n^2 grow as fast as n, it turns out that we cannot rank the asymptotic variances of LIML and HLIM. To show this, consider an example where $p=1$, z_i alternates between $-\bar{z}$ and $\bar{z}$ for $\bar{z}\neq 0$, $S_n=\sqrt{n}$ (so that $Y_i=z_i$), and z_i is included among the elements of Z_i. Then, for $\tilde{\Omega}=E[\tilde{U}_i^2]$ and $\kappa=\lim_{n\longrightarrow\infty}\sum_{i=1}^{n} P_{ii}^2/K$, we find that

$$V-V^{*}=\frac{\sigma^{2}}{\bar{z}^{2}(1-\tau)^{2}}(\tau\kappa-\tau^{2})\left(1-\frac{\tilde{\Omega}}{\bar{z}^{2}}\right).$$

Since $\tau\kappa-\tau^2$ is the limit of the sample variance of P_{ii}, which we assume to be positive, $V\ge V^{*}$ if and only if $\bar{z}^2\ge\tilde{\Omega}$. Here, $\bar{z}^2$ is the limit of the sample variance of z_i. Thus, the asymptotic variance ranking can go either way, depending on whether the sample variance of z_i is greater than the variance of $\tilde{U}_i$. In applications where the sample size is large relative to the number of instruments, these efficiency differences will tend to be quite small, because P_{ii} is small.

With many instruments and homoskedasticity, HLIM is asymptotically efficient relative to JIV. As shown in Chao et al. (2012b), the asymptotic variance of JIV is

$$\begin{aligned}V_{\mathrm{JIV}}&=\sigma^{2}H_P^{-1}\tilde{H}_P H_P^{-1}\\&\quad+\lim_{n\longrightarrow\infty}\left(1-\sum_{i=1}^{n}P_{ii}^{2}/K\right)H_p^{-1}(\Gamma+2\alpha S_0E[U_i\varepsilon_i]E[\varepsilon_iU_i']S_0')H_P^{-1},\end{aligned}$$

which is greater than the asymptotic variance of HLIM because $E[U_i\varepsilon_i]E[\varepsilon_iU_i']$ is positive semidefinite.

It remains to establish the consistency of the asymptotic variance estimator and to show that confidence intervals can be formed for linear combinations of the coefficients in the usual way. The following theorem accomplishes this under additional conditions on z_i.

Theorem 3. *If Assumptions 1–6 are satisfied, and $\hat{\alpha}=\tilde{\alpha}+O_p(1/n)$ or $\hat{\delta}$ is HLIM or HFUL, there exists a C with $\|z_i\|\le C$ for all i, and there exists a π_n such that* $\max_{i\le n}\|z_i-$

$\pi_n Z_i\| \longrightarrow 0$, *then in Case* I, $S_n'\hat{V}S_n \xrightarrow{p} \Lambda_{\mathrm{I}}$ *and in Case* II, $\mu_n^2 S_n'\hat{V}S_n/K \xrightarrow{p} \Lambda_{\mathrm{II}}$. *Also, if* $c'S_0'\Lambda_{\mathrm{I}}S_0c \neq 0$ *in Case* I *or* $c'S_0'\Lambda_{\mathrm{II}}S_0c \neq 0$ *in Case* II, *then*

$$\frac{c'(\hat{\delta}-\delta_0)}{\sqrt{c'\hat{V}c}} \xrightarrow{d} N(0,1).$$

This result allows us to form confidence intervals and test statistics for a single linear combination of parameters in the usual way. To show how the conditions of this result can be checked, we return to the previous example with one right-hand side endogenous variable. The following result gives primitive conditions in that example for the last conclusion of Theorem 3, that is, for asymptotic normality of a t-ratio.

Corollary 1. *If equation* (2) *holds, Assumptions* 1–6 *are satisfied for* $z_i = (Z_{i1}', z_{i2})$, $\hat{\alpha} = \tilde{\alpha} + O_p(1/n)$ *or* $\hat{\delta}$ *is HLIM or HFUL, there exists a* C *with* $\|z_i\| \leq C$ *for all* i, *and there exists a* π_n *such that* $\max_{i\leq n}\|z_i - \pi_n Z_i\| \longrightarrow 0$, $\inf_i \sigma_i^2 \geq C$ *and either* (a) $c \neq 0$ *and* $\mu_n^2 = n$, (b) K/μ_n^2 *is bounded and* $(-\pi_1, 1)c \neq 0$, *or* (c) $K/\mu_n^2 \longrightarrow \infty$, $(-\pi_1, 1)c \neq 0$, $\inf_i E[\tilde{U}_{iG}^2] > 0$, *and the sign of* $E[\varepsilon_i\tilde{U}_{iG}]$ *is constant, then*

$$\frac{c'(\hat{\delta}-\delta_0)}{\sqrt{c'\hat{V}c}} \xrightarrow{d} N(0,1).$$

The conditions of Corollary 1 are quite primitive. We have previously described how Assumption 2 is satisfied in the model of equation (2). Assumptions 1 and 3–6 are also quite primitive.

This result can be applied to show that t-ratios are asymptotically correct when the many instrument robust variance estimator is used. For the coefficient δ_2 of the endogenous variable, note that $c = e_G$ (the $G \times 1$ unit vector with 1 in the last position) so that $(-\pi_1, 1)c = 1 \neq 0$. Therefore, if $E[U_{iG}^2]$ is bounded away from zero and the sign of $E[\varepsilon_i U_{iG}]$ is constant, it follows from Corollary 1 that

$$\frac{\hat{\delta}_G - \delta_{0G}}{\sqrt{\hat{V}_{GG}}} \xrightarrow{d} N(0,1).$$

Thus the t-ratio for the coefficient of the endogenous variable is asymptotically correct across a wide range of different growth rates for μ_n and K. The analogous result holds for each coefficient δ_j, $j \leq G_1$, of an included instrument as long as $\pi_{1j} \neq 0$ is not zero. If $\pi_{1j} = 0$, then the asymptotics is more complicated. For brevity we will not discuss this unusual case here.

5. Monte Carlo results

In this Monte Carlo simulation, we provide evidence concerning the finite sample behavior of HLIM and HFUL. The model that we consider is

$$y_i = \delta_{10} + \delta_{20}x_{2i} + \varepsilon_i, \quad x_{2i} = \pi z_{1i} + U_{2i},$$

where $z_{i1} \sim N(0, 1)$ and $U_{2i} \sim N(0, 1)$. The ith instrument observation is

$$Z_i' = (1, z_{1i}, z_{1i}^2, z_{1i}^3, z_{1i}^4, z_{1i}D_{i1}, \dots, z_{1i}D_{i,K-5}),$$

where $D_{ik} \in \{0, 1\}$, $\Pr(D_{ik} = 1) = 1/2$, and $z_{i1} \sim N(0, 1)$. Thus, the instruments consist of powers of a standard normal up to the fourth power plus interactions with dummy variables. Only z_1 affects the reduced form, so that adding the other instruments does not improve asymptotic efficiency of HFUL, though the powers of z_{i1} do help with asymptotic efficiency of the CUE.

The structural disturbance, ε, is allowed to be heteroskedastic, being given by

$$\varepsilon = \rho U_2 + \sqrt{\frac{1-\rho^2}{\phi^2 + (0.86)^4}}(\phi v_1 + 0.86 v_2), \quad v_1 \sim N(0, z_1^2), v_2 \sim N(0, (0.86)^2),$$

where v_1 and v_2 are independent of U_2. This is a design that will lead to LIML being inconsistent with many instruments. Here, $E[X_i\varepsilon_i]$ is constant and σ_i^2 is quadratic in z_{i1}, so that $\gamma_i = (C_1 + C_2 z_{i1} + C_3 z_{i1}^2)^{-1} A$ for a constant vector A and constants C_1, C_2, C_3. In this case, P_{ii} will be correlated with $\gamma_i = E[X_i\varepsilon_i]/\sigma_i^2$ so that LIML is not consistent.

We report properties of estimators and t-ratios for δ_2. We set $n = 800$ and $\rho = 0.3$ throughout and let the number of instrumental variables be $K = 2, 30$. For $K = 2$, the instruments are $(1, z_i)$. We choose π so that the concentration parameter is $n\pi^2 = \mu^2 = 8, 32$. We also ran experiments with $K = 10$ and $\mu^2 = 16$. We also choose ϕ so that the R-squared for the regression of ε^2 on the instruments is 0, 0.1, or 0.2.

In Tables 1–6, we report results on median bias, the range between the 0.05 and 0.95 quantiles, and the nominal 0.05 rejection frequencies for a Wald test on δ_2 for LIML, HLIM, Fuller (1977), HFUL ($C = 1$), JIVE, and CUE. Interquartile range results were similar. We find that under homoskedasticity, HFUL is much less dispersed than LIML but slightly more biased. Under heteroskedasticity, HFUL is much less biased and also much less dispersed than LIML. Thus, we find that heteroskedasticity can bias LIML. We also find that the dispersion of LIML is substantially larger than HFUL. Thus we find a lower bias for HFUL under heteroskedasticity and many instruments, as predicted by the theory, as well as substantially lower dispersion, which, though not predicted by the theory, may be important in practice.

TABLE 1. Median bias: $\mathcal{R}^2_{\varepsilon^2|z_1^2} = 0$.[a]

μ^2	K	LIML	HLIM	FULL1	HFUL	JIVE	CUE
8	2	0.005	0.005	0.042	0.043	−0.034	0.005
8	10	0.024	0.023	0.057	0.057	0.053	0.025
8	30	0.065	0.065	0.086	0.091	0.164	0.071
32	2	0.002	0.002	0.011	0.011	−0.018	0.002
32	10	0.002	0.001	0.011	0.011	−0.019	0.002
32	30	0.003	0.002	0.013	0.013	−0.014	0.006

[a]Results based on 20,000 simulations.

TABLE 2. Nine decile range: 0.05 to 0.95 $\mathcal{R}^2_{\varepsilon^2|z_1^2} = 0$.[a]

μ^2	K	LIML	HLIM	FULL1	HFUL	JIVE	CUE
8	2	1.470	1.466	1.072	1.073	3.114	1.470
8	10	2.852	2.934	1.657	1.644	5.098	3.101
8	30	5.036	5.179	2.421	2.364	6.787	6.336
32	2	0.616	0.616	0.590	0.589	0.679	0.616
32	10	0.715	0.716	0.679	0.680	0.816	0.770
32	30	0.961	0.985	0.901	0.913	1.200	1.156

[a]Results based on 20,000 simulations.

TABLE 3. 0.05 rejection frequencies: $\mathcal{R}^2_{\varepsilon^2|z_1^2} = 0$.[a]

μ^2	K	LIML	HLIM	FULL1	HFUL	JIVE	CUE
8	2	0.025	0.026	0.021	0.034	0.051	0.012
8	10	0.035	0.037	0.029	0.044	0.063	0.027
8	30	0.045	0.049	0.040	0.054	0.068	0.051
32	2	0.041	0.042	0.037	0.044	0.038	0.030
32	10	0.041	0.042	0.038	0.044	0.046	0.041
32	30	0.042	0.047	0.039	0.050	0.057	0.062

[a]Results based on 20,000 simulations.

TABLE 4. Median bias: $\mathcal{R}^2_{\varepsilon^2|z_1^2} = 0.2$.[a]

μ^2	K	LIML	HLIM	FULL1	HFUL	JIVE	CUE
8	2	−0.001	0.050	0.041	0.078	−0.031	−0.001
8	10	−0.623	0.094	−0.349	0.113	0.039	0.003
8	30	−1.871	0.134	−0.937	0.146	0.148	−0.034
32	2	−0.001	0.011	0.008	0.020	−0.021	−0.001
32	10	−0.220	0.015	−0.192	0.024	−0.021	0.000
32	30	−1.038	0.016	−0.846	0.027	−0.016	−0.017

[a]Results based on 20,000 simulations.

TABLE 5. Nine decile range: 0.05 to 0.95 $\mathcal{R}^2_{\varepsilon^2|z_1^2} = 0.2$.[a]

μ^2	K	LIML	HLIM	FULL1	HFUL	JIVE	CUE
8	2	2.219	1.868	1.675	1.494	4.381	2.219
8	10	26.169	5.611	4.776	2.664	7.781	16.218
8	30	60.512	8.191	7.145	3.332	9.975	1.5E+012
32	2	0.941	0.901	0.903	0.868	1.029	0.941
32	10	3.365	1.226	2.429	1.134	1.206	1.011
32	30	18.357	1.815	5.424	1.571	1.678	3.563

[a]Results based on 20,000 simulations.

TABLE 6. 0.05 rejection frequencies: $\mathcal{R}^2_{\varepsilon^2|z_1^2} = 0.2$.[a]

μ^2	K	LIML	HLIM	FULL1	HFUL	JIVE	CUE
8	2	0.097	0.019	0.075	0.023	0.026	0.008
8	10	0.065	0.037	0.080	0.041	0.036	0.043
8	30	0.059	0.051	0.118	0.055	0.046	0.094
32	2	0.177	0.040	0.162	0.040	0.039	0.024
32	10	0.146	0.042	0.120	0.044	0.033	0.030
32	30	0.128	0.049	0.107	0.051	0.039	0.073

[a]Results based on 20,000 simulations.

In addition, in Tables 3 and 6 we find that the rejection frequencies for HFUL are quite close to their nominal values, being closer than all the rest throughout much of the tables. Thus, the standard errors we have given work very well in accounting for many instruments and heteroskedasticity.

6. Conclusion

We have considered the situation of many instruments with heteroskedastic data. In this situation, both 2SLS and LIML are inconsistent. We have proposed two new estimators, HLIM and HFUL, that are consistent in this situation. We derive the asymptotic normal distributions for both estimators with many instruments and many weak instrument sequences. We find that the variances of the asymptotic distributions take a convenient form, which are straightforward to estimate consistently. A problem with the HLIM (and LIML) estimator is the wide dispersion caused by the "moments problem." We demonstrate that HFUL has finite sample moments so that the moments problem does not exist.

In Monte Carlo experiments, we find these properties hold. With heteroskedasticity and many instruments we find that both LIML and Fuller have significant median bias (Table 4). We find that HLIM, HFUL, JIVE, and CUE do not have this median bias. However, HLIM, JIVE, and CUE all suffer from very large dispersions arising from the moments problem (Table 5). Indeed, the nine decile range for CUE exceeds 10^{12}! The dispersion of the HFUL estimate is much less than these alternative consistent estimators. Thus, we recommend that HFUL be used in the many instruments situation when heteroskedasticity is present, which is the common situation in microeconometrics.

Appendix A

This appendix is divided into two parts. In the first part, we discuss the existence of moments of HFUL. In the second part, we give proofs for the theorems stated in the body of the paper as well as state and prove some preliminary lemmas.

A.1 *Existence of moments of HFUL*

Here, we give a formal result that shows the existence of moments of the HFUL estimator. Some additional assumptions must be specified for this result, and we introduce these conditions below.

Assumption 7. (a) $K = O(n^a)$ *for some real constant* a *such that* $0 \le a \le 1$; *if* $a = 1$, *then* $n - K \to \infty$ *as* $n \to \infty$; *and for all* n *sufficiently large, there exists a positive constant* C_P *such that* $P_{ii} \le C_P(K/n) < 1$ $(i = 1, \dots, n)$. (b) $\mu_n^2 \sim n^b$ *for some real constant* b *such that* $a/2 < b \le 1$. (c) *If* K *is fixed, then* $z_i = \pi Z_i$. (d) $\delta_0 \in \mathcal{D} \subset \mathbb{R}^G$, *where* $\mathcal{D}$ *is bounded.* (e) $\lambda_{\max}(\widetilde{S}'\widetilde{S})$ *is bounded.*

Next, let $x \vee y = \max(x, y)$ and $x \wedge y = \min(x, y)$, and define

$$\varphi(a, b) = \frac{a \vee (1 - 4\psi_1(a, b))/2}{\psi_1(a, b)} \mathbb{I}\left\{\frac{a}{2} < b \le \frac{1}{2}\right\} + \frac{a}{\psi_2(a, b)} \mathbb{I}\left\{\frac{1}{2} < b \le 1\right\}, \tag{5}$$

where $\psi_1(a, b) = 2b - a \wedge b$, $\psi_2(a, b) = 2b - a \wedge \frac{1}{2}$, and a and b are as specified in Assumption 7. Also, we take the (restricted) reduced form of the instrumental variable (IV) regression model to be

$$\overline{X} = Y\Delta + \overline{V},$$

where $\overline{X} = [y \ \ X]$, $\Delta = [\delta_0 \ \ I_G]$, and $\overline{V} = [v \ \ U]$.

Assumption 8. (i) *Let* p *be a positive integer, let* η *be a positive constant, and define*

$$q = (1 + \eta)[2(G + 1) + \varphi(a, b)].$$

There is $\widetilde{C} > 0$ *such that* $E[\|\overline{V}_i\|^{2pq}] \le \widetilde{C}$ *and* $\sum_{i=1}^n \|z_i\|^{2pq}/n \le \widetilde{C}$, *where* $\overline{V}_i'$ *denotes the* ith *row of* $\overline{V}$. (ii) $\lambda_{\min}(\frac{1}{n}\sum_{i=1}^n E[\overline{V}_i\overline{V}_i'])$ *is bounded away from zero for* n *sufficiently large.*

Assumption 8 specifies the moment condition on the error process $\{\overline{V}_i\}$ as dependent on the number of endogenous regressors G, instrument weakness as parameterized by b, and an upper bound on the rate at which the number of instruments grows, as parameterized by a. Although the function $\varphi(a, b)$ that enters into the moment condition seems complicated, it actually depends on a and b in an intuitive way, so that everything else being equal, more stringent moment conditions are needed in cases with weaker instruments and/or faster growing K. More stringent moment conditions are also needed in situations with a larger number of endogenous regressors.

To get more intuition about Assumption 8, consider the following two special cases. First, consider the conventional case where the instruments are strong and the number of instruments is fixed, so that $a = 0$ and $b = 1$. In this case, it is easy to see that $\varphi(a, b) = \varphi(0, 1) = 0$, and Assumption 8 requires finite moments up to the order

$$2pq = 4p(G + 1)(1 + \eta).$$

If we further consider the case with one endogenous regressor ($G = 1$) and where η can be taken to be small, then Assumption 8 requires a bit more than an 8th order moment condition (on the errors) for the existence of the first moment of HFUL and a bit more than a 16th order moment condition for the existence of the second moment.

Next, consider the many weak instrument case where $a = 1/2$ and $a/2 = 1/4 < b \leq 1/2$. In this case, note that since $2b - a \leq b$, we have

$$\psi_1(a, b) = \psi_1(1/2, b) = 2b - 1/2$$

and

$$\varphi(a, b) = \varphi(1/2, b) = \frac{1}{4b - 1} \quad \text{for } b \in (1/4, 1/2],$$

so that $1 \leq \varphi(1/2, b) < \infty$ and $\varphi(1/2, b)$ is a decreasing function of b. In particular, note that the required strength of the moment condition grows without bound as b approaches 1/4.

Proving the existence of moments of HFUL requires showing the existence of certain inverse moments of $\det[S_{*,n}]$, where $S_{*,n} = X_*' M X_*/(n - K)$ and $X_* = [\varepsilon \;\; X]$. We shall explicitly assume the existence of such inverse moments below.

Assumption 9. *There exists a positive constant* $\overline{C}$ *and a positive integer* N *such that*

$$E\left[(\det[S_{*,n}])^{-2p(1+\eta)/\eta}\right] \leq \overline{C} < \infty \tag{6}$$

for all $n \geq N$, *where* $\eta > 0$ *is as specified in Assumption* 8.

In Chao, Hausman, Newey, Swanson, and Woutersen (2012a), we gave an example of a probability density function for which inverse moments of the form (6) do not exist and, hence, some condition such as Assumption 9 is needed to rule out pathological cases. On the other hand, Assumption 9 is also not vacuous. In particular, it can be verified, as we do in Theorem 4 below, that this assumption holds for an IV regression model with heteroskedastic, Gaussian error distributions. However, it should be noted that normality is not necessary for Assumption 9 to hold, as has been discussed in Chao et al. (2012a).

Theorem 4. *Suppose that Assumptions 1, 2, and 4 hold. In addition, suppose that the IV regression model has heteroskedastic, Gaussian errors, that is,* $\{\overline{U}_i\} \equiv i.n.i.d.N(0, \Omega_i^*)$, *where* $\overline{U}_i$ *is the* i*th row of* $\overline{U} = [\varepsilon \;\; U]$, *and suppose that there exists a constant* $\underline{C} > 0$ *such that* $\lambda_* = \min_i \lambda_{\min}(\Omega_i^*) \geq \underline{C}$. *Then Assumption 9 holds.*

A proof of this theorem is given in Appendix A.2.

We now state our existence of moments result.

Theorem 5. *Suppose that Assumptions 1–4 and 7 hold. In addition, suppose that Assumptions 8 and 9 are satisfied for some positive* p. *Then there exists a positive constant* $\overline{C}$ *such that*

$$E[\|\hat{\delta}_{\text{HFUL}}\|^p] \leq \overline{C} < \infty$$

for n sufficiently large.

A proof of this theorem can be found in a supplementary file on the journal website.

A.2 *Proofs of consistency and asymptotic distributional results*

Throughout, let C denote a generic positive constant that may be different in different uses and let "Markov inequality" denote the conditional Markov inequality. The first lemma is Lemma A0 from Hansen, Hausman, and Newey (2008).

Lemma A0. *If Assumption 2 is satisfied and* $\|S_n'(\hat{\delta}-\delta_0)/\mu_n\|^2/(1+\|\hat{\delta}\|^2)\xrightarrow{p}0$, *then* $\|S_n'(\hat{\delta}-\delta_0)/\mu_n\|\xrightarrow{p}0$.

We next give a result from Chao et al. (2012b) that is used in the proof of consistency.

Lemma A1 (Special Case of Lemma A1 of Chao et al. (2012b)). *If* (W_i, Y_i) $(i=1,\dots,n)$ *are independent,* W_i *and* Y_i *are scalars, and* P *is symmetric, idempotent of* rank K, *then for* $\bar{w}=E[(W_1,\dots,W_n)']$, $\bar{y}=E[(Y_1,\dots,Y_n)']$, $\bar{\sigma}_{Wn}=\max_{i\le n}\operatorname{Var}(W_i)^{1/2}$, $\bar{\sigma}_{Yn}=\max_{i\le n}\operatorname{Var}(Y_i)^{1/2}$,

$$E\biggl[\biggl(\sum_{i\neq j}P_{ij}W_iY_j-\sum_{i\neq j}P_{ij}\bar{w}_i\bar{y}_j\biggr)^2\biggr]\le C(K\bar{\sigma}_{W_n}^2\bar{\sigma}_{Y_n}^2+\bar{\sigma}_{W_n}^2\bar{y}'\bar{y}+\bar{\sigma}_{Y_n}^2\bar{w}'\bar{w}).$$

For the next result, let $\bar{S}_n=\operatorname{diag}(\mu_n, S_n)$, $\tilde{X}=[\varepsilon, X]\bar{S}_n^{-1\prime}$, and $H_n=\sum_{i=1}^n(1-P_{ii})z_iz_i'/n$.

Lemma A2. *If Assumptions 1–4 are satisfied and* $\sqrt{K}/\mu_n^2\longrightarrow 0$, *then*

$$\sum_{i\neq j}\tilde{X}_iP_{ij}\tilde{X}_j'=\operatorname{diag}(0,H_n)+o_p(1).$$

Proof. Note that

$$\tilde{X}_i=\begin{pmatrix}\mu_n^{-1}\varepsilon_i\\ S_n^{-1}X_i\end{pmatrix}=\begin{pmatrix}0\\ z_i/\sqrt{n}\end{pmatrix}+\begin{pmatrix}\mu_n^{-1}\varepsilon_i\\ S_n^{-1}U_i\end{pmatrix}.$$

Since $\|S_n^{-1}\|\le C\mu_n^{-1}$, we have $\operatorname{Var}(\tilde{X}_{ik})\le C\mu_n^{-2}$ for any element $\tilde{X}_{ik}$ of $\tilde{X}_i$. Then applying Lemma A1 to each element of $\sum_{i\neq j}\tilde{X}_iP_{ij}\tilde{X}_j'$ gives

$$\begin{aligned}\sum_{i\neq j}\tilde{X}_iP_{ij}\tilde{X}_j'&=\operatorname{diag}\biggl(0,\sum_{i\neq j}z_iP_{ij}z_j'/n\biggr)+O_p\biggl(K^{1/2}/\mu_n^2+\mu_n^{-1}\biggl(\sum_i\|z_i\|^2/n\biggr)^{1/2}\biggr)\\&=\operatorname{diag}\biggl(0,\sum_{i\neq j}z_iP_{ij}z_j'/n\biggr)+o_p(1).\end{aligned}$$

Also note that

$$
\begin{aligned}
H_n - \sum_{i\neq j} z_i P_{ij} z_j'/n &= \sum_i z_i z_i'/n - \sum_i P_{ii} z_i z_i'/n - \sum_{i\neq j} z_i P_{ij} z_j'/n = z'(I-P)z/n \\
&= (z - Z\pi_{Kn}')'(I-P)(z - Z\pi_{Kn}')/n \\
&\le (z - Z\pi_{Kn}')'(z - Z\pi_{Kn}')/n \\
&\le I_G \sum_i \|z_i - \pi_{Kn} Z_i\|^2/n \longrightarrow 0,
\end{aligned}
$$

where the third equality follows by $PZ = Z$, the first inequality follows by $I - P$ idempotent, and the last inequality follows by $A \le \operatorname{tr}(A)I$ for any positive semidefinite (p.s.d.) matrix A. Since this equation shows that $H_n - \sum_{i\neq j} z_i P_{ij} z_j'/n$ is p.s.d. and is less than or equal to another p.s.d. matrix that converges to zero, it follows that $\sum_{i\neq j} z_i P_{ij} z_j'/n = H_n + o_p(1)$. The conclusion follows by the triangle inequality. □

In what follows, it is useful to prove directly that the HLIM estimator $\tilde{\delta}$ satisfies $S_n'(\tilde{\delta} - \delta_0)/\mu_n \xrightarrow{p} 0$.

Lemma A3. *If Assumptions 1–4 are satisfied, then* $S_n'(\tilde{\delta} - \delta_0)/\mu_n \xrightarrow{p} 0$.

Proof. Let $\bar{Y} = [0, Y]$, $\bar{U} = [\varepsilon, U]$, and $\bar{X} = [y, X]$, so that $\bar{X} = (\bar{Y} + \bar{U})D$ for

$$
D = \begin{bmatrix} 1 & 0 \\ \delta_0 & I \end{bmatrix}.
$$

Let $\hat{B} = \bar{X}'\bar{X}/n$. Note that $\|S_n/\sqrt{n}\| \le C$ and, by standard calculations, $z'U/n \xrightarrow{p} 0$. Then

$$
\|\bar{Y}'\bar{U}/n\| = \|(S_n/\sqrt{n})z'U/n\| \le C\|z'U/n\| \xrightarrow{p} 0.
$$

Let $\bar{\Omega}_n = \sum_{i=1}^n E[\bar{U}_i\bar{U}_i']/n = \operatorname{diag}(\sum_{i=1}^n \Omega_i^*/n, 0) \ge C \operatorname{diag}(I_{G_2+1}, 0)$ by Assumption 3, where $G_2 + 1$ is the dimension of the number of included endogenous variables. By the Markov inequality, we have $\bar{U}'\bar{U}/n - \bar{\Omega}_n \xrightarrow{p} 0$, so it follows that with probability approaching 1 (w.p.a.1),

$$
\hat{B} = (\bar{U}'\bar{U} + \bar{Y}'\bar{U} + \bar{U}'\bar{Y} + \bar{Y}'\bar{Y})/n = \bar{\Omega}_n + \bar{Y}'\bar{Y}/n + o_p(1) \ge C \operatorname{diag}\big(I_{G-G_2+1}, 0\big).
$$

Since $\bar{\Omega}_n + \bar{Y}'\bar{Y}/n$ is bounded, it follows that w.p.a.1,

$$
C \le (1, -\delta')\hat{B}(1, -\delta')' = (y - X\delta)'(y - X\delta)/n \le C\|(1, -\delta')\|^2 = C(1 + \|\delta\|^2).
$$

Next, as defined preceding Lemma A2, let $\bar{S}_n = \operatorname{diag}(\mu_n, S_n)$ and $\tilde{X} = [\varepsilon, X]\bar{S}_n^{-1\prime}$. Note that by $P_{ii} \le C < 1$ and uniform nonsingularity of $\sum_{i=1}^n z_i z_i'/n$, we have $H_n \ge (1 - C)\sum_{i=1}^n z_i z_i'/n \ge CI_G$. Then by Lemma A2, w.p.a.1,

$$
\hat{A} \stackrel{\text{def}}{=} \sum_{i\neq j} P_{ij}\tilde{X}_i\tilde{X}_j' \ge C \operatorname{diag}(0, I_G).
$$

Note that $\bar{S}_n' D(1,-\delta')' = (\mu_n, (\delta_0-\delta)'S_n)'$ and $\bar{X}_i = D'\bar{S}_n\tilde{X}_i$. Then w.p.a.1 for all δ,

$$\begin{aligned}
&\mu_n^{-2}\sum_{i\neq j} P_{ij}(y_i - X_i'\delta)(y_j - X_j'\delta) \\
&\quad = \mu_n^{-2}(1,-\delta')\Bigl(\sum_{i\neq j} P_{ij}\bar{X}_i\bar{X}_j'\Bigr)(1,-\delta')' \\
&\quad = \mu_n^{-2}(1,-\delta')D'\bar{S}_n\hat{A}\bar{S}_n'D(1,-\delta')' \geq C\|S_n'(\delta-\delta_0)/\mu_n\|^2.
\end{aligned}$$

Let $\hat{Q}(\delta) = (n/\mu_n^2)\sum_{i\neq j}(y_i - X_i'\delta)P_{ij}(y_j - X_j'\delta)/(y-X\delta)'(y-X\delta)$. Then by the upper-left element of the conclusion of Lemma A2, $\mu_n^{-2}\sum_{i\neq j}\varepsilon_i P_{ij}\varepsilon_j \xrightarrow{p} 0$. Then w.p.a.1,

$$|\hat{Q}(\delta_0)| = \left|\mu_n^{-2}\sum_{i\neq j}\varepsilon_i P_{ij}\varepsilon_j \Big/ \sum_{i=1}^n \varepsilon_i^2/n\right| \xrightarrow{p} 0.$$

Since $\hat{\delta} = \arg\min_\delta \hat{Q}(\delta)$, we have $\hat{Q}(\hat{\delta}) \leq \hat{Q}(\delta_0)$. Therefore, w.p.a.1, by $(y - X\delta)'(y - X\delta)/n \leq C(1+\|\delta\|^2)$, it follows that

$$0 \leq \frac{\|S_n'(\hat{\delta}-\delta_0)/\mu_n\|^2}{1+\|\hat{\delta}\|^2} \leq C\hat{Q}(\hat{\delta}) \leq C\hat{Q}(\delta_0) \xrightarrow{p} 0,$$

implying $\|S_n'(\hat{\delta}-\delta_0)/\mu_n\|^2/(1+\|\hat{\delta}\|^2) \xrightarrow{p} 0$. Lemma A0 gives the conclusion. □

LEMMA A4. *If Assumptions 1–4 are satisfied,* $\hat{\alpha} = o_p(\mu_n^2/n)$, *and* $S_n'(\hat{\delta}-\delta_0)/\mu_n \xrightarrow{p} 0$, *then for* $H_n = \sum_{i=1}^n (1-P_{ii})z_i z_i'/n$,

$$S_n^{-1}\Bigl(\sum_{i\neq j} X_i P_{ij} X_j' - \hat{\alpha}X'X\Bigr)S_n^{-1\prime} = H_n + o_p(1),$$

$$S_n^{-1}\Bigl(\sum_{i\neq j} X_i P_{ij}\hat{\varepsilon}_j - \hat{\alpha}X'\hat{\varepsilon}\Bigr)/\mu_n \xrightarrow{p} 0.$$

PROOF. By the Markov inequality and standard arguments, $X'X = O_p(n)$ and $X'\hat{\varepsilon} = O_p(n)$. Therefore, by $\|S_n^{-1}\| = O(\mu_n^{-1})$,

$$\hat{\alpha}S_n^{-1}X'XS_n^{-1\prime} = o_p(\mu_n^2/n)O_p(n/\mu_n^2) \xrightarrow{p} 0,$$

$$\hat{\alpha}S_n^{-1}X'\hat{\varepsilon}/\mu_n = o_p(\mu_n^2/n)O_p(n/\mu_n^2) \xrightarrow{p} 0.$$

Lemma A2 (lower right-hand block) and the triangle inequality then give the first conclusion. By Lemma A2 (off diagonal), we have $S_n^{-1}\sum_{i\neq j} X_i P_{ij}\varepsilon_j/\mu_n \xrightarrow{p} 0$, so that

$$S_n^{-1}\sum_{i\neq j} X_i P_{ij}\hat{\varepsilon}_j/\mu_n = o_p(1) - \Bigl(S_n^{-1}\sum_{i\neq j} X_i P_{ij} X_j' S_n^{-1\prime}\Bigr)S_n'(\hat{\delta}-\delta_0)/\mu_n \xrightarrow{p} 0.$$

□

LEMMA A5. *If Assumptions 1–4 are satisfied and* $S_n'(\hat{\delta}-\delta_0)/\mu_n \xrightarrow{p} 0$, *then* $\sum_{i\neq j}\hat{\varepsilon}_i P_{ij}\hat{\varepsilon}_j/\hat{\varepsilon}'\hat{\varepsilon} = o_p(\mu_n^2/n)$.

PROOF. Let $\hat{\beta} = S_n'(\hat{\delta}-\delta_0)/\mu_n$ and $\breve{\alpha} = \sum_{i\neq j}\varepsilon_i P_{ij}\varepsilon_j/\varepsilon'\varepsilon = o_p(\mu_n^2/n)$. Note that $\hat{\sigma}_\varepsilon^2 = \hat{\varepsilon}'\hat{\varepsilon}/n$ satisfies $1/\hat{\sigma}_\varepsilon^2 = O_p(1)$ by the Markov inequality. By Lemma A4 with $\hat{\alpha} = \breve{\alpha}$, we have $\tilde{H}_n = S_n^{-1}(\sum_{i\neq j}X_i P_{ij} X_j' - \breve{\alpha}X'X)S_n^{-1\prime} = O_p(1)$ and $W_n = S_n^{-1}(\sum_{i\neq j}X_i P_{ij}\varepsilon_j - \breve{\alpha}X'\varepsilon)/\mu_n \xrightarrow{p} 0$, so

$$\frac{\sum_{i\neq j}\hat{\varepsilon}_i P_{ij}\hat{\varepsilon}_j}{\hat{\varepsilon}'\hat{\varepsilon}} - \breve{\alpha} = \frac{1}{\hat{\varepsilon}'\hat{\varepsilon}}\left(\sum_{i\neq j}\hat{\varepsilon}_i P_{ij}\hat{\varepsilon}_j - \sum_{i\neq j}\varepsilon_i P_{ij}\varepsilon_j - \breve{\alpha}(\hat{\varepsilon}'\hat{\varepsilon} - \varepsilon'\varepsilon)\right)$$
$$= \frac{\mu_n^2}{n}\frac{1}{\hat{\sigma}_\varepsilon^2}(\hat{\beta}'\tilde{H}_n\hat{\beta} - 2\hat{\beta}'W_n) = o_p(\mu_n^2/n),$$

so the conclusion follows by the triangle inequality. □

PROOF OF THEOREM 1. First note that if $S_n'(\hat{\delta}-\delta_0)/\mu_n \xrightarrow{p} 0$, then by $\lambda_{\min}(S_n S_n'/\mu_n^2) \geq \lambda_{\min}(\tilde{S}\tilde{S}') > 0$, we have

$$\|S_n'(\hat{\delta}-\delta_0)/\mu_n\| \geq \lambda_{\min}(S_n S_n'/\mu_n^2)^{1/2}\|\hat{\delta}-\delta_0\| \geq C\|\hat{\delta}-\delta_0\|,$$

implying $\hat{\delta} \xrightarrow{p} \delta_0$. Therefore, it suffices to show that $S_n'(\hat{\delta}-\delta_0)/\mu_n \xrightarrow{p} 0$. For HLIM, this follows from Lemma A3. For HFUL, note that $\tilde{\alpha} = \hat{Q}(\tilde{\delta}) = \sum_{i\neq j}\tilde{\varepsilon}_i P_{ij}\tilde{\varepsilon}_j/\tilde{\varepsilon}'\tilde{\varepsilon} = o_p(\mu_n^2/n)$ by Lemma A5, so by the formula for HFUL, $\hat{\alpha} = \tilde{\alpha} + O_p(1/n) = o_p(\mu_n^2/n)$. Thus, the result for HFUL will follow from the most general result for any $\hat{\alpha}$ with $\hat{\alpha} = o_p(\mu_n^2/n)$. For any such $\hat{\alpha}$, by Lemma A4 we have

$$\begin{aligned}
&S_n'(\hat{\delta}-\delta_0)/\mu_n \\
&\quad = S_n'\left(\sum_{i\neq j}X_i P_{ij}X_j' - \hat{\alpha}X'X\right)^{-1}\sum_{i\neq j}(X_i P_{ij}\varepsilon_j - \hat{\alpha}X'\varepsilon)/\mu_n \\
&\quad = \left[S_n^{-1}\left(\sum_{i\neq j}X_i P_{ij}X_j' - \hat{\alpha}X'X\right)S_n^{-1\prime}\right]^{-1} S_n^{-1}\sum_{i\neq j}(X_i P_{ij}\varepsilon_j - \hat{\alpha}X'\varepsilon)/\mu_n \\
&\quad = (H_n + o_p(1))^{-1}o_p(1) \xrightarrow{p} 0.
\end{aligned}$$ □

Now we move on to asymptotic normality results. The next result is a central limit theorem that was proven in Chao et al. (2012b).

LEMMA A6 (Lemma A2 of Chao et al. (2012b)). *If* (i) P *is a symmetric, idempotent matrix with* $\operatorname{rank}(P) = K$, $P_{ii} \leq C < 1$, (ii) $(W_{1n}, U_1, \varepsilon_1), \ldots, (W_{nn}, U_n, \varepsilon_n)$ *are independent and* $D_n = \sum_{i=1}^n E[W_{in}W_{in}']$ *is bounded,* (iii) $E[W_{in}'] = 0$, $E[U_i] = 0$, $E[\varepsilon_i] = 0$, *and there exists a constant* C *such that* $E[\|U_i\|^4] \leq C$, $E[\varepsilon_i^4] \leq C$, (iv) $\sum_{i=1}^n E[\|W_{in}\|^4] \longrightarrow 0$, *and*

(v) $K \longrightarrow \infty$, *then for* $\bar{\Sigma}_n \stackrel{\text{def}}{=} \sum_{i\neq j} P_{ij}^2(E[U_iU_i']E[\varepsilon_j^2] + E[U_i\varepsilon_i]E[\varepsilon_jU_j'])/K$ *and for any sequence of bounded nonzero vectors* c_{1n} *and* c_{2n} *such that* $\Xi_n = c_{1n}'D_nc_{1n} + c_{2n}'\bar{\Sigma}_nc_{2n} > C$, *it follows that*

$$Y_n = \Xi_n^{-1/2}\left(\sum_{i=1}^n c_{1n}'W_{in} + c_{2n}'\sum_{i\neq j} U_iP_{ij}\varepsilon_j/\sqrt{K}\right) \xrightarrow{d} N(0,1).$$

Let $\tilde{\alpha}(\delta) = \sum_{i\neq j}\varepsilon_i(\delta)P_{ij}\varepsilon_j(\delta)/\varepsilon(\delta)'\varepsilon(\delta)$ and

$$\begin{aligned}\hat{D}(\delta) &= -\left[\frac{\varepsilon(\delta)'\varepsilon(\delta)}{2}\right]\frac{\partial}{\partial\delta}\left[\frac{\sum_{i\neq j}\varepsilon_i(\delta)P_{ij}\varepsilon_j(\delta)}{\varepsilon(\delta)'\varepsilon(\delta)}\right] \\ &= \sum_{i\neq j}X_iP_{ij}\varepsilon_j(\delta) - \tilde{\alpha}(\delta)X'\varepsilon(\delta).\end{aligned}$$

A couple of other intermediate results are also useful.

Lemma A7. *If Assumptions 1–4 are satisfied and* $S_n'(\bar{\delta}-\delta_0)/\mu_n \xrightarrow{p} 0$, *then*

$$-S_n^{-1}[\partial\hat{D}(\bar{\delta})/\partial\delta]S_n^{-1\prime} = H_n + o_p(1).$$

Proof. Let $\bar{\varepsilon} = \varepsilon(\bar{\delta}) = y - X\bar{\delta}$, $\bar{\gamma} = X'\bar{\varepsilon}/\bar{\varepsilon}'\bar{\varepsilon}$, and $\bar{\alpha} = \tilde{\alpha}(\bar{\delta})$. Then differentiating gives

$$\begin{aligned}-\frac{\partial\hat{D}}{\partial\delta}(\bar{\delta}) &= \sum_{i\neq j}X_iP_{ij}X_j' - \bar{\alpha}X'X - \bar{\gamma}\sum_{i\neq j}\bar{\varepsilon}_iP_{ij}X_j' - \sum_{i\neq j}X_iP_{ij}\bar{\varepsilon}_j\bar{\gamma}' + 2(\bar{\varepsilon}'\bar{\varepsilon})\bar{\alpha}\bar{\gamma}\bar{\gamma}' \\ &= \sum_{i\neq j}X_iP_{ij}X_j' - \bar{\alpha}X'X + \bar{\gamma}\hat{D}(\bar{\delta})' + \hat{D}(\bar{\delta})\bar{\gamma}',\end{aligned}$$

where the second equality follows by $\hat{D}(\bar{\delta}) = \sum_{i\neq j}X_iP_{ij}\bar{\varepsilon}_j - (\bar{\varepsilon}'\bar{\varepsilon})\bar{\alpha}\bar{\gamma}$. By Lemma A5, we have $\bar{\alpha} = o_p(\mu_n^2/n)$. By standard arguments, $\bar{\gamma} = O_p(1)$ so that $S_n^{-1}\bar{\gamma} = O_p(1/\mu_n)$. Then by Lemma A4 and $\hat{D}(\bar{\delta}) = \sum_{i\neq j}X_iP_{ij}\bar{\varepsilon}_j - \bar{\alpha}X'\bar{\varepsilon}$,

$$S_n^{-1}\left(\sum_{i\neq j}X_iP_{ij}X_j' - \bar{\alpha}X'X\right)S_n^{-1\prime} = H_n + o_p(1), S_n^{-1}\hat{D}(\bar{\delta})\bar{\gamma}'S_n^{-1\prime} \xrightarrow{p} 0.$$

The conclusion then follows by the triangle inequality. □

Lemma A8. *If Assumptions 1–4 are satisfied, then for* $\gamma_n = \sum_i E[U_i\varepsilon_i]/\sum_i E[\varepsilon_i^2]$ *and* $\tilde{U}_i = U_i - \gamma_n\varepsilon_i$,

$$S_n^{-1}\hat{D}(\delta_0) = \sum_{i=1}^n(1-P_{ii})z_i\varepsilon_i/\sqrt{n} + S_n^{-1}\sum_{i\neq j}\tilde{U}_iP_{ij}\varepsilon_j + o_p(1).$$

Proof. Note that for $W = z'(P-I)\varepsilon/\sqrt{n}$, by $I-P$ idempotent and $E[\varepsilon\varepsilon'] \le CI_n$, we have

$$E[WW'] \le Cz'(I-P)z/n = C(z - Z\pi'_{Kn})'(I-P)(z - Z\pi'_{Kn})/n$$
$$\le CI_G \sum_{i=1}^{n} \|z_i - \pi_{Kn}Z_i\|^2/n \longrightarrow 0,$$

so $z'(P-I)\varepsilon/\sqrt{n} = o_p(1)$. Also, by the Markov inequality,

$$X'\varepsilon/n = \sum_{i=1}^{n} E[X_i\varepsilon_i]/n + O_p(1/\sqrt{n}), \qquad \varepsilon'\varepsilon/n = \sum_{i=1}^{n} \sigma_i^2/n + O_p(1/\sqrt{n}).$$

Also, by Assumption 3, $\sum_{i=1}^{n} \sigma_i^2/n \ge C > 0$. The delta method then gives $\tilde{\gamma} = X'\varepsilon/\varepsilon'\varepsilon = \gamma_n + O_p(1/\sqrt{n})$. Therefore, it follows by Lemma A1 and $\hat{D}(\delta_0) = \sum_{i\neq j} X_i P_{ij}\varepsilon_j - \varepsilon'\varepsilon\tilde{\alpha}(\delta_0)\tilde{\gamma}$ that

$$S_n^{-1}\hat{D}(\delta_0) = \sum_{i\neq j} z_i P_{ij}\varepsilon_j/\sqrt{n} + S_n^{-1}\sum_{i\neq j}\tilde{U}_i P_{ij}\varepsilon_i - S_n^{-1}(\tilde{\gamma} - \gamma_n)\varepsilon'\varepsilon\tilde{\alpha}(\delta_0)$$
$$= z'P\varepsilon/\sqrt{n} - \sum_i P_{ii} z_i\varepsilon_i/\sqrt{n} + S_n^{-1}\sum_{i\neq j}\tilde{U}_i P_{ij}\varepsilon_j$$
$$+ O_p(1/(\sqrt{n}\mu_n))o_p(\mu_n^2/n)$$
$$= \sum_{i=1}^{n}(1-P_{ii})z_i\varepsilon_i/\sqrt{n} + S_n^{-1}\sum_{i\neq j}\tilde{U}_i P_{ij}\varepsilon_j + o_p(1).$$

□

Proof of Theorem 2. Consider first the case where $\hat{\delta}$ is HLIM. Then by Theorem 1, $\hat{\delta} \xrightarrow{p} \delta_0$. First order conditions for LIML are $\hat{D}(\hat{\delta}) = 0$. Expanding gives

$$0 = \hat{D}(\delta_0) + \frac{\partial \hat{D}}{\partial \delta}(\bar{\delta})(\hat{\delta} - \delta_0),$$

where $\bar{\delta}$ lies on the line joining $\hat{\delta}$ and δ_0, and hence $\bar{\beta} = \mu_n^{-1}S_n'(\bar{\delta} - \delta_0) \xrightarrow{p} 0$. Then by Lemma A7, $\bar{H}_n = S_n^{-1}[\partial\hat{D}(\bar{\delta})/\partial\delta]S_n^{-1\prime} = H_P + o_p(1)$. Then $\partial\hat{D}(\bar{\delta})/\partial\delta$ is nonsingular w.p.a.1 and solving gives

$$S_n'(\hat{\delta} - \delta) = -S_n'[\partial\hat{D}(\bar{\delta})/\partial\delta]^{-1}\hat{D}(\delta_0) = -\bar{H}_n^{-1}S_n^{-1}\hat{D}(\delta_0).$$

Next, apply Lemma A6 with $U_i = \tilde{U}_i$ and

$$W_{in} = (1-P_{ii})z_i\varepsilon_i/\sqrt{n}.$$

By ε_i having bounded fourth moment and for $P_{ii} \le 1$,

$$\sum_{i=1}^{n} E[\|W_{in}\|^4] \le C\sum_{i=1}^{n}\|z_i\|^4/n^2 \longrightarrow 0.$$

By Assumption 6, we have $\sum_{i=1}^n E[W_{in}W_{in}'] \longrightarrow \Sigma_P$. Let $\Gamma = \operatorname{diag}(\Sigma_P, \Psi)$ and

$$A_n = \begin{pmatrix} \sum_{i=1}^n W_{in} \\ \sum_{i\neq j} \tilde{U}_i P_{ij}\varepsilon_j/\sqrt{K} \end{pmatrix}.$$

Consider c such that $c'\Gamma c > 0$. Then by the conclusion of Lemma A6, we have $c'A_n \xrightarrow{d} N(0, c'\Gamma c)$. Also, if $c'\Gamma c = 0$, then it is straightforward to show that $c'A_n \xrightarrow{p} 0$. Then it follows by the Cramer–Wold device that

$$A_n = \begin{pmatrix} \sum_{i=1}^n W_{in} \\ \sum_{i\neq j} \tilde{U}_i P_{ij}\varepsilon_j/\sqrt{K} \end{pmatrix} \xrightarrow{d} N(0, \Gamma), \qquad \Gamma = \operatorname{diag}(\Sigma_P, \Psi).$$

Next, we consider the two cases. Case I has K/μ_n^2 bounded. In this case, $\sqrt{K}S_n^{-1} \longrightarrow S_0$, so that

$$F_n \stackrel{\text{def}}{=} [I, \sqrt{K}S_n^{-1}] \longrightarrow F_0 = [I, \sqrt{\alpha}S_0], \qquad F_0\Gamma F_0' = \Sigma_P + \alpha S_0 \Psi S_0'.$$

Then by Lemma A8,

$$S_n^{-1}\hat{D}(\delta_0) = F_n A_n + o_p(1) \xrightarrow{d} N(0, \Sigma_P + \alpha S_0\Psi S_0'),$$
$$S_n'(\hat{\delta} - \delta_0) = -\bar{H}_n^{-1} S_n^{-1}\hat{D}(\delta_0) \xrightarrow{d} N(0, \Lambda_{\mathrm{I}}).$$

In Case II, we have $K/\mu_n^2 \longrightarrow \infty$. Here

$$(\mu_n/\sqrt{K})F_n \longrightarrow \bar{F}_0 = [0, S_0], \qquad \bar{F}_0\Gamma\bar{F}_0' = S_0\Psi S_0'$$

and $(\mu_n/\sqrt{K})o_p(1) = o_p(1)$. Then by Lemma A8,

$$(\mu_n/\sqrt{K})S_n^{-1}\hat{D}(\delta_0) = (\mu_n/\sqrt{K})F_n A_n + o_p(1) \xrightarrow{d} N(0, S_0\Psi S_0'),$$
$$(\mu_n/\sqrt{K})S_n'(\hat{\delta} - \delta_0) = -\bar{H}_n^{-1}(\mu_n/\sqrt{K})S_n^{-1}\hat{D}(\delta_0) \xrightarrow{d} N(0, \Lambda_{\mathrm{II}}). \qquad \square$$

The next two results are useful for the proof of consistency of the variance estimator, and these results are taken from Chao et al. (2012b). Let $\bar{\mu}_{Wn} = \max_{i\leq n}|E[W_i]|$ and $\bar{\mu}_{Yn} = \max_{i\leq n}|E[Y_i]|$.

Lemma A9 (Lemma A3 of Chao et al. (2012b)). *If (W_i, Y_i) $(i = 1, \dots, n)$ are independent, and W_i and Y_i are scalars, then*

$$\sum_{i\neq j} P_{ij}^2 W_i Y_j = E\Bigl[\sum_{i\neq j} P_{ij}^2 W_i Y_j\Bigr] + O_p(\sqrt{K}(\bar{\sigma}_{Wn}\bar{\sigma}_{Yn} + \bar{\sigma}_{Wn}\bar{\mu}_{Yn} + \bar{\mu}_{Wn}\bar{\sigma}_{Yn})).$$

Lemma A10 (Lemma A4 of Chao et al. (2012b)). *If W_i, Y_i, and η_i, are independent across i with $E[W_i] = a_i/\sqrt{n}$, $E[Y_i] = b_i/\sqrt{n}$, $|a_i| \leq C$, $|b_i| \leq C$, $E[\eta_i^2] \leq C$,* $\mathrm{Var}(W_i) \leq C\mu_n^{-2}$, *and* $\mathrm{Var}(Y_i) \leq C\mu_n^{-2}$, *there exists π_n such that* $\max_{i \leq n} |a_i - Z_i'\pi_n| \longrightarrow 0$, *and $\sqrt{K}/\mu_n^2 \longrightarrow 0$, then*

$$A_n = E\Bigg[\sum_{i \neq j \neq k} W_i P_{ik} \eta_k P_{kj} Y_j\Bigg] = O(1), \qquad \sum_{i \neq j \neq k} W_i P_{ik} \eta_k P_{kj} Y_j - A_n \xrightarrow{p} 0.$$

Next, recall that $\hat{\varepsilon}_i = Y_i - X_i'\hat{\delta}$, $\hat{\gamma} = X'\hat{\varepsilon}/\hat{\varepsilon}'\hat{\varepsilon}$, and $\gamma_n = \sum_i E[X_i\varepsilon_i]/\sum_i \sigma_i^2$, and let

$$\begin{aligned}
\breve{X}_i &= S_n^{-1}(X_i - \hat{\gamma}\hat{\varepsilon}_i) = S_n^{-1}\hat{X}_i, \qquad \dot{X}_i = S_n^{-1}(X_i - \gamma_n\varepsilon_i), \\
\breve{\Sigma}_1 &= \sum_{i \neq j \neq k} \breve{X}_i P_{ik} \hat{\varepsilon}_k^2 P_{kj} \breve{X}_j', \qquad \breve{\Sigma}_2 = \sum_{i \neq j} P_{ij}^2 (\breve{X}_i \breve{X}_i' \hat{\varepsilon}_j^2 + \breve{X}_i \hat{\varepsilon}_i \hat{\varepsilon}_j \breve{X}_j'), \\
\dot{\Sigma}_1 &= \sum_{i \neq j \neq k} \dot{X}_i P_{ik} \varepsilon_k^2 P_{kj} \dot{X}_j', \qquad \dot{\Sigma}_2 = \sum_{i \neq j} P_{ij}^2 (\dot{X}_i \dot{X}_i' \varepsilon_j^2 + \dot{X}_i \varepsilon_i \varepsilon_j \dot{X}_j').
\end{aligned}$$

Note that for $\hat{\Delta} = S_n'(\hat{\delta} - \delta_0)$, we have

$$\begin{aligned}
&\hat{\varepsilon}_i - \varepsilon_i = -X_i'(\hat{\delta} - \delta_0) = -X_i'S_n^{-1\prime}\hat{\Delta}, \\
&\hat{\varepsilon}_i^2 - \varepsilon_i^2 = -2\varepsilon_i X_i'(\hat{\delta} - \delta_0) + [X_i'(\hat{\delta} - \delta_0)]^2, \\
&\breve{X}_i - \dot{X}_i = -S_n^{-1}\hat{\gamma}(\hat{\varepsilon}_i - \varepsilon_i) - S_n^{-1}(\hat{\gamma} - \gamma_n)\varepsilon_i \\
&\qquad = S_n^{-1}\hat{\gamma}X_i'S_n^{-1\prime}\hat{\Delta} - S_n^{-1}\mu_n(\hat{\gamma} - \gamma_n)(\varepsilon_i/\mu_n), \\
&\breve{X}_i\hat{\varepsilon}_i - \dot{X}_i\varepsilon_i = X_i\hat{\varepsilon}_i - \hat{\gamma}\hat{\varepsilon}_i^2 - X_i\varepsilon_i + \gamma_n\varepsilon_i^2 \\
&\qquad = -X_iX_i'(\hat{\delta} - \delta_0) - \hat{\gamma}\big\{-2\varepsilon_i X_i'(\hat{\delta} - \delta_0) + [X_i'(\hat{\delta} - \delta_0)^2]\big\} \\
&\qquad\quad - (\hat{\gamma} - \gamma_n)\varepsilon_i^2, \\
&\|\breve{X}_i\breve{X}_i' - \dot{X}_i\dot{X}_i'\| \leq \|\breve{X}_i - \dot{X}_i\|^2 + 2\|\dot{X}_i\|\|\breve{X}_i - \dot{X}_i\|.
\end{aligned}$$

Lemma A11. *If the hypotheses of Theorem 3 are satisfied, then* $\breve{\Sigma}_2 - \dot{\Sigma}_2 = o_p(K/\mu_n^2)$.

Proof. Note first that $S_n/\sqrt{n}$ is bounded, so by the Cauchy–Schwarz inequality, $\|Y_i\| = \|S_n z_i/\sqrt{n}\| \leq C$. Let $d_i = C + |\varepsilon_i| + \|U_i\|$. Note that $\hat{\gamma} - \gamma_n \xrightarrow{p} 0$ by standard arguments. Then for $\hat{A} = (1 + \|\hat{\gamma}\|)(1 + \|\hat{\delta}\|) = O_p(1)$ and $\hat{B} = \|\hat{\gamma} - \gamma_n\| + \|\hat{\delta} - \delta_0\| \xrightarrow{p} 0$, we have

$$\begin{aligned}
&\|X_i\| \leq C + \|U_i\| \leq d_i, \qquad |\hat{\varepsilon}_i| \leq |X_i'(\delta_0 - \hat{\delta}) + \varepsilon_i| \leq Cd_i\hat{A}, \\
&\|\dot{X}_i\| = \|S_n^{-1}(X_i - \gamma_n\varepsilon_i)\| \leq C\mu_n^{-1}d_i, \\
&\|\breve{X}_i\| = \|S_n^{-1}(X_i - \hat{\gamma}\hat{\varepsilon}_i)\| \leq C\mu_n^{-1}d_i\hat{A}, \\
&\|\breve{X}_i\breve{X}_i' - \dot{X}_i\dot{X}_i'\| \leq (\|\breve{X}_i\| + \|\dot{X}_i\|)\|\breve{X}_i - \dot{X}_i\| \\
&\qquad \leq C\mu_n^{-2}d_i\hat{A}\|\hat{\gamma}\|\|\hat{\varepsilon}_i - \varepsilon_i\| + \|\hat{\gamma} - \gamma_n\||\varepsilon_i| \\
&\qquad \leq C\mu_n^{-2}d_i^2\hat{A}^2\hat{B},
\end{aligned}$$

$$
\begin{aligned}
&|\hat{\varepsilon}_i^2-\varepsilon_i^2| \le (|\varepsilon_i|+|\hat{\varepsilon}_i|)|\hat{\varepsilon}_i-\varepsilon_i| \le C d_i^2 \hat{A}\hat{B},\\
&\|\breve{X}_i\hat{\varepsilon}_i-\dot{X}_i\varepsilon_i\| = \|S_n^{-1}(X_i\hat{\varepsilon}_i-\hat{\gamma}\hat{\varepsilon}_i^2-X_i\varepsilon_i+\gamma_n\varepsilon_i^2)\|\\
&\qquad \le C\mu_n^{-1}(\|X_i\||\hat{\varepsilon}_i-\varepsilon_i|+\|\hat{\gamma}\||\hat{\varepsilon}_i^2-\varepsilon_i^2|+|\varepsilon_i^2|\|\hat{\gamma}-\gamma_n\|)\\
&\qquad \le C\mu_n^{-1}d_i^2(\hat{B}+\hat{A}^2\hat{B}+\hat{B}) \le C d_i^2\hat{A}^2\hat{B},\\
&\|\breve{X}_i\hat{\varepsilon}_i\| \le C\mu_n^{-1}d_i^2\hat{A}^2, \qquad \|\dot{X}_i\varepsilon_i\| \le C\mu_n^{-1}d_i^2.
\end{aligned}
$$

Also note that

$$
E\left[\sum_{i\neq j} P_{ij}^2 d_i^2 d_j^2 \mu_n^{-2}\right] \le C\mu_n^{-2}\sum_{i,j}P_{ij}^2 = C\mu_n^{-2}\sum_i P_{ii} = C\mu_n^{-2}K,
$$

so that $\sum_{i\neq j}P_{ij}^2 d_i^2 d_j^2\mu_n^{-2} = O_p(K/\mu_n^2)$ by the Markov inequality. Then it follows that

$$
\begin{aligned}
\left\|\sum_{i\neq j}P_{ij}^2(\breve{X}_i\breve{X}_i'\hat{\varepsilon}_j^2-\dot{X}_i\dot{X}_i'\varepsilon_j^2)\right\| &\le \sum_{i\neq j}P_{ij}^2(|\hat{\varepsilon}_j^2|\|\breve{X}_i\breve{X}_i'-\dot{X}_i\dot{X}_i'\|+\|\dot{X}_i\|^2|\hat{\varepsilon}_j^2-\varepsilon_j^2|)\\
&\le C\mu_n^{-2}\sum_{i\neq j}P_{ij}^2 d_i^2 d_j^2(\hat{A}^4\hat{B}+\hat{A}\hat{B}) = o_p(K/\mu_n^2).
\end{aligned}
$$

We also have

$$
\begin{aligned}
&\left\|\sum_{i\neq j}P_{ij}^2(\breve{X}_i\hat{\varepsilon}_i\hat{\varepsilon}_j\breve{X}_j'-\dot{X}_i\varepsilon_i\varepsilon_j\dot{X}_j)\right\|\\
&\quad \le \sum_{i\neq j}P_{ij}^2(\|\breve{X}_i\hat{\varepsilon}_i\|\|\breve{X}_j\hat{\varepsilon}_j-\dot{X}_j\varepsilon_j\|+\|\dot{X}_j\varepsilon_j\|\|\breve{X}_i\hat{\varepsilon}_i-\dot{X}_i\varepsilon_i\|)\\
&\quad \le C\mu_n^{-2}\sum_{i\neq j}P_{ij}^2 d_i^2 d_j^2(1+\hat{A}^2)\hat{A}^2\hat{B} = o_p\left(\frac{K}{\mu_n^2}\right).
\end{aligned}
$$

The conclusion then follows by the triangle inequality. □

Lemma A12. *If the hypotheses of Theorem 3 are satisfied, then* $\breve{\Sigma}_1-\dot{\Sigma}_1 = o_p(K/\mu_n^2)$.

Proof. Note first that

$$
\hat{\varepsilon}_i-\varepsilon_i = -X_i'(\hat{\delta}-\delta_0) = -X_i'S_n^{-1\prime}S_n'(\hat{\delta}-\delta_0) = -(z_i/\sqrt{n}+S_n^{-1}U_i)'\hat{\Delta} = -D_i'\hat{\Delta},
$$

where $D_i = z_i/\sqrt{n}+S_n^{-1}U_i$ and $\hat{\Delta} = S_n'(\hat{\delta}-\delta_0)$. Also

$$
\begin{aligned}
&\hat{\varepsilon}_i^2-\varepsilon_i^2 = -2\varepsilon_i X_i'(\hat{\delta}-\delta_0)+[X_i'(\hat{\delta}-\delta_0)]^2,\\
&\breve{X}_i-\dot{X}_i = -\hat{\gamma}\hat{\varepsilon}_i+\gamma_n\varepsilon_i = S_n^{-1}\hat{\gamma}D_i'\hat{\Delta}-S_n^{-1}\mu_n(\hat{\gamma}-\gamma_n)\varepsilon_i/\mu_n.
\end{aligned}
$$

We now have $\breve{\Sigma}_1-\dot{\Sigma}_1 = \sum_{r=1}^7 T_r$, where

$$
T_1 = \sum_{i\neq j\neq k}(\breve{X}_i-\dot{X}_i)P_{ik}(\hat{\varepsilon}_k^2-\varepsilon_k^2)P_{kj}(\breve{X}_j-\dot{X}_j)',
$$

$$T_2=\sum_{i\neq j\neq k}\dot{X}_iP_{ik}(\hat{\varepsilon}_k^2-\varepsilon_k^2)P_{kj}(\check{X}_j-\dot{X}_j)',$$

$$T_3=\sum_{i\neq j\neq k}(\check{X}_i-\dot{X}_i)P_{ik}\varepsilon_k^2P_{kj}(\check{X}_j-\dot{X}_j)',\qquad T_4=T_2',$$

$$T_5=\sum_{i\neq j\neq k}(\check{X}_i-\dot{X}_i)P_{ik}\varepsilon_k^2P_{kj}\dot{X}_j',$$

$$T_6=\sum_{i\neq j\neq k}\dot{X}_iP_{ik}(\hat{\varepsilon}_k^2-\varepsilon_k^2)P_{kj}\dot{X}_j',\qquad T_7=T_5'.$$

From the above expression for $\hat{\varepsilon}_i^2-\varepsilon_i^2$, we see that T_6 is a sum of terms of the form $\hat{B}\sum_{i\neq j\neq k}\dot{X}_iP_{ik}\eta_iP_{kj}\dot{X}_j'$, where $\hat{B}\xrightarrow{p}0$ and η_i is a component either of $-2\varepsilon_iX_i$ or of X_iX_i'. By Lemma A10, we have $\sum_{i\neq j\neq k}\dot{X}_iP_{ik}\eta_iP_{kj}\dot{X}_j'=O_p(1)$, so by the triangle inequality, $T_6\xrightarrow{p}0$. Also, note that

$$T_5=S_n^{-1}\hat{\gamma}\hat{\Delta}'\sum_{i\neq j\neq k}D_iP_{ik}\varepsilon_k^2P_{kj}\dot{X}_j'+S_n^{-1}\mu_n(\hat{\gamma}-\gamma_n)\sum_{i\neq j\neq k}(\varepsilon_i/\mu_n)P_{ik}\varepsilon_k^2P_{kj}\dot{X}_j'.$$

Note that $S_n^{-1}\hat{\gamma}\hat{\Delta}'\xrightarrow{p}0$, $E[D_i]=z_i/\sqrt{n}$, $\mathrm{Var}(D_i)=O(\mu_n^{-2})$, $E[\dot{X}_i]=z_i/\sqrt{n}$, and $\mathrm{Var}(\dot{X})=O(\mu_n^{-2})$. Then by Lemma A10, it follows that $\sum_{i\neq j\neq k}D_iP_{ik}\varepsilon_k^2P_{kj}\dot{X}_j'=O_p(1)$, so that $S_n^{-1}\hat{\gamma}\hat{\Delta}'\sum_{i\neq j\neq k}D_iP_{ik}\varepsilon_k^2P_{kj}\dot{X}_j'\xrightarrow{p}0$. A similar argument applied to the second term and the triangle inequality then give $T_5\xrightarrow{p}0$. Also $T_7=T_5'\xrightarrow{p}0$.

Next, analogous arguments apply to T_2 and T_3, except that there are four terms in each of them rather than two, and also apply to T_1, except there are eight terms in T_1. For brevity we omit the details. □

Lemma A13. *If the hypotheses of Theorem 3 are satisfied, then*

$$\dot{\Sigma}_2=\sum_{i\neq j}P_{ij}^2z_iz_i'\sigma_j^2/n+S_n^{-1}\sum_{i\neq j}P_{ij}^2(E[\tilde{U}_i\tilde{U}_i']\sigma_j^2+E[\tilde{U}_i\varepsilon_i]E[\varepsilon_j\tilde{U}_j'])S_n^{-1\prime}$$
$$+o_p(K/\mu_n^2).$$

Proof. Note that $\mathrm{Var}(\varepsilon_i^2)\le C$ and $\mu_n^2\le Cn$, so that for $u_{ki}=e_k'S_n^{-1}U_i$,

$$E[(\dot{X}_{ik}\dot{X}_{i\ell})^2]\le CE[\dot{X}_{ik}^4+\dot{X}_{i\ell}^4]\le C\{z_{ik}^4/n^2+E[u_k^4]+z_{i\ell}^4/n^2+E[u_\ell^4]\}\le C\mu_n^{-4},$$
$$E[(\dot{X}_{ik}\varepsilon_i)^2]\le CE[(z_{ik}^2\varepsilon_i^2/n+u_{ki}^2\varepsilon_i^2)]\le Cn^{-1}+C\mu_n^{-2}\le C\mu_n^{-2}.$$

Also, we have, for $\tilde{\Omega}_i=E[\tilde{U}_i\tilde{U}_i']$,

$$E[\dot{X}_i\dot{X}_i']=z_iz_i'/n+S_n^{-1}\tilde{\Omega}_iS_n^{-1\prime},\qquad E[\dot{X}_i\varepsilon_i]=S_n^{-1}E[\tilde{U}_i\varepsilon_i].$$

Next let W_i be $e_j'\dot{X}_i\dot{X}_i'e_k$ for some j and k, so that

$$E[W_i]=e_j'S_n^{-1}E[\tilde{U}_i\tilde{U}_i']S_n^{-1\prime}e_k+z_{ij}z_{ik}/n,\qquad |E[W_i]|\le C\mu_n^{-2},$$

$$\begin{aligned}\operatorname{Var}(W_i) &= \operatorname{Var}\{(e_j' S_n^{-1} U_i + z_{ij}/\sqrt{n})(e_k' S_n^{-1} U_i + z_{ik}/\sqrt{n})\} \\ &\le C/\mu_n^4 + C/n\mu_n^2 \le C/\mu_n^4.\end{aligned}$$

Also let $Y_i = \varepsilon_i^2$. Then $\sqrt{K}(\bar{\sigma}_{Wn}\bar{\sigma}_{Yn} + \bar{\sigma}_{Wn}\bar{\mu}_{Yn} + \bar{\mu}_{Wn}\bar{\sigma}_{Yn}) \le CK^{1/2}/\mu_n^2$, so applying Lemma A9 for this W_i and Y_i gives

$$\sum_{i\neq j} P_{ij}^2 \dot{X}_i \dot{X}_i' \varepsilon_j^2 = \sum_{i\neq j} P_{ij}^2 (z_i z_i'/n + S_n^{-1}\tilde{\Omega}_i S_n^{-1\prime})\sigma_j^2 + O_p(\sqrt{K}/\mu_n^2).$$

It follows similarly from Lemma A9 with W_i and Y_i equal to elements of $\dot{X}_i\varepsilon_i$ that

$$\sum_{i\neq j} P_{ij}^2 \dot{X}_i \varepsilon_i \varepsilon_j \dot{X}_j' = S_n^{-1} \sum_{i\neq j} P_{ij}^2 E[\tilde{U}_i \varepsilon_i] E[\varepsilon_j \tilde{U}_j'] S_n^{-1\prime} + O_p(\sqrt{K}/\mu_n^2).$$

Also, by $K \longrightarrow \infty$, we have $O_p(\sqrt{K}/\mu_n^2) = o_p(K/\mu_n^2)$. The conclusion then follows by the triangle inequality. □

Lemma A14. *If the hypotheses of Theorem 3 are satisfied, then*

$$\dot{\Sigma}_1 = \sum_{i\neq j\neq k} z_i P_{ik} \sigma_k^2 P_{kj} z_j'/n + o_p(1).$$

Proof. Apply Lemma A10 with W_i equal to an element of $\dot{X}_i$, Y_j equal to an element of $\dot{X}_j$, and η_k equal to ε_k^2. □

Proof of Theorem 3. Note that $\bar{X}_i = \sum_{j=1}^n P_{ij}\hat{X}_j$,

$$\begin{aligned}&\sum_{i=1}^n (\bar{X}_i \bar{X}_i' - \hat{X}_i P_{ii} \bar{X}_i' - \bar{X}_i P_{ii} \hat{X}_i')\hat{\varepsilon}_i^2 \\ &\quad= \sum_{i,j,k=1}^n \hat{X}_i P_{ik} \hat{\varepsilon}_k^2 P_{kj} \hat{X}_j' - \sum_{i,j=1}^n \hat{X}_i P_{ii} \hat{\varepsilon}_i^2 P_{ij} \hat{X}_j' - \sum_{i,j=1}^n \hat{X}_i P_{ij} \hat{\varepsilon}_j^2 P_{jj} \hat{X}_j' \\ &\quad= \sum_{i,j,k=1}^n \hat{X}_i P_{ik} \hat{\varepsilon}_k^2 P_{kj} \hat{X}_j' - \sum_{i\neq j} \hat{X}_i P_{ii} \hat{\varepsilon}_i^2 P_{ij} \hat{X}_j' \\ &\qquad - \sum_{i\neq j} \hat{X}_i P_{ij} \hat{\varepsilon}_j^2 P_{jj} \hat{X}_j' - 2\sum_{i=1}^n \hat{X}_i P_{ii}^2 \hat{\varepsilon}_i^2 \hat{X}_i' \\ &\quad= \sum_{i,j,k\notin\{i,j\}}^n \hat{X}_i P_{ik} \hat{\varepsilon}_k^2 P_{kj} \hat{X}_j' - \sum_{i=1}^n \hat{X}_i P_{ii}^2 \hat{\varepsilon}_i^2 \hat{X}_i' \\ &\quad= \sum_{i\neq j\neq k}^n \hat{X}_i P_{ik} \hat{\varepsilon}_k^2 P_{kj} \hat{X}_j' + \sum_{i\neq j}^n P_{ij}^2 \hat{X}_i \hat{X}_i \hat{\varepsilon}_j^2 - \sum_{i=1}^n \hat{X}_i P_{ii}^2 \hat{\varepsilon}_i^2 \hat{X}_i'.\end{aligned}$$

Also, for Z_i' and $\tilde{Z}_i'$ equal to the ith row of Z and $\tilde{Z} = Z(Z'Z)^{-1}$, we have

$$
\begin{aligned}
&\sum_{k=1}^{K}\sum_{\ell=1}^{K}\left(\sum_{i=1}^{n}\tilde{Z}_{ik}\tilde{Z}_{i\ell}\hat{X}_i\hat{\varepsilon}_i\right)\left(\sum_{j=1}^{n}Z_{jk}Z_{j\ell}\hat{X}_j\hat{\varepsilon}_j\right)' \\
&\quad = \sum_{i,j=1}^{n}\left(\sum_{k=1}^{K}\sum_{\ell=1}^{K}\tilde{Z}_{ik}Z_{jk}\tilde{Z}_{i\ell}Z_{j\ell}\right)\hat{X}_i\hat{\varepsilon}_i\hat{\varepsilon}_j\hat{X}_j' = \sum_{i,j=1}^{n}\left(\sum_{k=1}^{K}\tilde{Z}_{ik}Z_{jk}\right)^2\hat{X}_i\hat{\varepsilon}_i\hat{\varepsilon}_j\hat{X}_j' \\
&\quad = \sum_{i,j=1}^{n}(\tilde{Z}_i'Z_j)^2\hat{X}_i\hat{\varepsilon}_i\hat{\varepsilon}_j\hat{X}_j' = \sum_{i,j=1}^{n}P_{ij}^2\hat{X}_i\hat{\varepsilon}_i\hat{\varepsilon}_j\hat{X}_j'.
\end{aligned}
$$

Adding this equation to the previous one then gives

$$
\begin{aligned}
\hat{\Sigma} &= \sum_{i\neq j\neq k}\hat{X}_iP_{ik}\hat{\varepsilon}_k^2P_{kj}\hat{X}_j' + \sum_{i\neq j}P_{ij}^2\hat{X}_i\hat{X}_i'\hat{\varepsilon}_j^2 - \sum_{i=1}^{n}\hat{X}_iP_{ii}^2\hat{\varepsilon}_i^2\hat{X}_i' + \sum_{i,j=1}^{n}P_{ij}^2\hat{X}_i\hat{\varepsilon}_i\hat{\varepsilon}_j\hat{X}_j' \\
&= \sum_{i\neq j\neq k}\hat{X}_iP_{ik}\hat{\varepsilon}_k^2P_{kj}\hat{X}_j' + \sum_{i\neq j}P_{ij}^2(\hat{X}_i\hat{X}_i'\hat{\varepsilon}_j^2 + \hat{X}_i\hat{\varepsilon}_i\hat{\varepsilon}_j\hat{X}_j').
\end{aligned}
$$

It then follows that $S_n^{-1}\hat{\Sigma}S_n^{-1\prime} = \breve{\Sigma}_1 + \breve{\Sigma}_2$, so that

$$
\begin{aligned}
S_n'\hat{V}S_n &= (S_n^{-1}\hat{H}S_n^{-1\prime})^{-1}S_n^{-1}\hat{\Sigma}S_n^{-1\prime}(S_n^{-1}\hat{H}S_n^{-1\prime})^{-1} \\
&= (S_n^{-1}\hat{H}S_n^{-1\prime})^{-1}(\breve{\Sigma}_1 + \breve{\Sigma}_2)(S_n^{-1}\hat{H}S_n^{-1\prime})^{-1}.
\end{aligned}
$$

By Lemma A4 we have $S_n^{-1}\hat{H}S_n^{-1\prime} \xrightarrow{p} H_P$. Also, note that for $\bar{z}_i = \sum_j P_{ij}z_i = e_i'Pz$,

$$
\begin{aligned}
&\sum_{i\neq j\neq k} z_iP_{ik}\sigma_k^2P_{kj}z_j'/n \\
&\quad = \sum_i\sum_{j\neq i}\sum_{k\notin\{i,j\}} z_iP_{ik}\sigma_k^2P_{kj}z_j'/n \\
&\quad = \sum_i\sum_{j\neq i}\left(\sum_k z_iP_{ik}\sigma_k^2P_{kj}z_j' - z_iP_{ii}\sigma_i^2P_{ij}z_j' - z_iP_{ij}\sigma_j^2P_{jj}z_j'\right)/n \\
&\quad = \left(\sum_k \bar{z}_k\sigma_k^2\bar{z}_k' - \sum_{i,k}P_{ik}^2z_iz_i'\sigma_k^2 - \sum_i z_iP_{ii}\sigma_i^2\bar{z}_i' + \sum_i z_iP_{ii}\sigma_i^2P_{ii}z_i'\right. \\
&\qquad \left. - \sum_j \bar{z}_j\sigma_j^2P_{jj}z_j' + \sum_i z_jP_{jj}\sigma_j^2P_{jj}z_j'\right)/n \\
&\quad = \sum_i \sigma_i^2(\bar{z}_i\bar{z}_i' - P_{ii}z_i\bar{z}_i' - P_{ii}\bar{z}_iz_i' + P_{ii}^2z_iz_i')/n - \sum_{i\neq j}P_{ij}^2z_iz_i'\sigma_j^2/n.
\end{aligned}
$$

Also, it follows similarly to the proof of Lemma A8 that $\sum_i \|z_i - \bar{z}_i\|^2/n \leq z'(I-P)z/n \longrightarrow 0$. Then by σ_i^2 and P_{ii} bounded, we have

$$
\begin{aligned}
&\left\| \sum_i \sigma_i^2 (\bar{z}_i \bar{z}_i' - z_i z_i')/n \right\| \\
&\quad \leq \sum_i \sigma_i^2 (2\|z_i\| \|z_i - \bar{z}_i\| + \|z_i - \bar{z}_i\|^2)/n \\
&\quad \leq C\left(\sum_i \|z_i\|^2/n\right)^{1/2} \left(\sum_i \|z_i - \bar{z}_i\|^2/n\right)^{1/2} + C \sum_i \|z_i - \bar{z}_i\|^2/n \\
&\quad \longrightarrow 0, \\
&\left\| \sum_i \sigma_i^2 P_{ii} (z_i \bar{z}_i' - z_i z_i')/n \right\| \leq \left(\sum_i \sigma_i^4 P_{ii}^2 \|z_i\|^2/n\right)^{1/2} \left(\sum_i \|z_i - \bar{z}_i\|^2/n\right)^{1/2} \longrightarrow 0.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\sum_{i \neq j \neq k} z_i P_{ik} \sigma_k^2 P_{kj} z_j'/n &= \sum_i \sigma_i^2 (1 - P_{ii})^2 z_i z_i'/n + o(1) - \sum_{i \neq j} P_{ij}^2 z_i z_i' \sigma_j^2/n \\
&= \Sigma_P - \sum_{i \neq j} P_{ij}^2 z_i z_i' \sigma_j^2/n + o(1).
\end{aligned}
$$

It then follows by Lemmas A10–A14 and the triangle inequality that

$$
\begin{aligned}
\breve{\Sigma}_1 + \breve{\Sigma}_2 &= \sum_{i \neq j \neq k} z_i P_{ik} \sigma_k^2 P_{kj} z_j'/n + \sum_{i \neq j} P_{ij}^2 z_i z_i' \sigma_j^2/n \\
&\quad + S_n^{-1} \sum_{i \neq j} P_{ij}^2 (E[\tilde{U}_i \tilde{U}_i'] \sigma_j^2 + E[\tilde{U}_i \varepsilon_i] E[\varepsilon_j \tilde{U}_j']) S_n^{-1\prime} + o_p(1) + o_p(K/\mu_n^2) \\
&= \Sigma_P + K S_n^{-1} (\Psi + o(1)) S_n^{-1\prime} + o_p(1) + o_p(K/\mu_n^2) \\
&= \Sigma_P + K S_n^{-1} \Psi S_n^{-1\prime} + o_p(1) + o_p(K/\mu_n^2).
\end{aligned}
$$

Then in Case I, we have $o_p(K/\mu_n^2) = o_p(1)$, so that

$$
S_n' \hat{V} S_n = H^{-1} (\Sigma_P + K S_n^{-1} \Psi S_n^{-1\prime}) H^{-1} + o_p(1) = \Lambda_{\mathrm{I}} + o_p(1).
$$

In Case II, we have $(\mu_n^2/K) o_p(1) \xrightarrow{p} 0$, so that

$$
(\mu_n^2/K) S_n' \hat{V} S_n = H^{-1} ((\mu_n^2/K) \Sigma_P + \mu_n^2 S_n^{-1} \Psi S_n^{-1\prime}) H^{-1} + o_p(1) = \Lambda_{\mathrm{II}} + o_p(1).
$$

Next, consider Case I and note that $S_n'(\hat{\delta} - \delta_0) \xrightarrow{d} Y \sim N(0, \Lambda_{\mathrm{I}})$, $S_n' \hat{V} S_n \xrightarrow{p} \Lambda_{\mathrm{I}}$, $c' \mu_n S_n^{-1\prime} \to c' S_0'$, and $c' S_0' \Lambda_{\mathrm{I}} S_0 c \neq 0$. Then by the continuous mapping and Slutzky theorems,

$$
\frac{c'(\hat{\delta} - \delta_0)}{\sqrt{c' \hat{V} c}} = \frac{c' \mu_n S_n^{-1\prime} S_n' (\hat{\delta} - \delta_0)}{\sqrt{c' \mu_n S_n^{-1\prime} S_n' \hat{V} S_n S_n^{-1} \mu_n c}} \xrightarrow{d} \frac{c' S_0' Y}{\sqrt{c' S_0' \Lambda_{\mathrm{I}} S_0 c}} \sim N(0, 1).
$$

For Case II, $(\mu_n/\sqrt{K})S_n'(\hat{\delta}-\delta_0)\xrightarrow{d}\bar{Y}\sim N(0,\Lambda_{\mathrm{II}})$ and $(\mu_n^2/K)S_n'\hat{V}S_n\xrightarrow{p}\Lambda_{\mathrm{II}}$. Then by $c'S_0'\Lambda_{\mathrm{II}}S_0c\neq 0$,

$$\begin{aligned}\frac{c'(\hat{\delta}-\delta_0)}{\sqrt{c'\hat{V}c}}&=\frac{c'S_n^{-1\prime}(\mu_n/\sqrt{K})S_n'(\hat{\delta}-\delta_0)}{\sqrt{c'S_n^{-1\prime}(\mu_n^2/K)S_n'\hat{V}S_nS_n^{-1}c}}\\&=\frac{c'\mu_nS_n^{-1\prime}(\mu_n/\sqrt{K})S_n'(\hat{\delta}-\delta_0)}{\sqrt{c'\mu_nS_n^{-1\prime}(\mu_n^2/K)S_n'\hat{V}S_nS_n^{-1}\mu_nc}}\xrightarrow{d}\frac{c'S_0'\bar{Y}}{\sqrt{c'S_0'\Lambda_{\mathrm{II}}S_0c}}\sim N(0,1).\end{aligned}$$ □

Proof of Corollary 1. Note first that in this example, $\mu_n=\mu_{Gn}$, so that

$$\mu_nS_n^{-1}=\operatorname{diag}(\mu_n/\sqrt{n},\ldots,\mu_n/\sqrt{n},1)\tilde{S}^{-1},\qquad \tilde{S}^{-1}=\begin{bmatrix}I&0\\-\pi_1&1\end{bmatrix}.$$

By Assumption 2, $\gamma=\lim_{n\longrightarrow\infty}(\mu_n^2/n)$ exists and is either 0 or 1. By $\sigma_i^2\geq C>0$, $H_P^{-1}\Sigma_PH_P^{-1}$ is positive definite (p.d.). Also $\Psi=\lim_{n\longrightarrow\infty}E[(\sum_{i\neq j}P_{ij}\tilde{U}_i\varepsilon_j)(\sum_{i\neq j}P_{ij}\tilde{U}_i\times\varepsilon_j)']/K$ is positive semidefinite, implying Λ_{I} is p.d.

If (a) $\mu_n=n$, then we have Case I and $S_0=\tilde{S}^{-1}$, which is nonsingular so that $S_0c\neq 0$. Hence

$$c'S_0'\Lambda_{\mathrm{I}}S_0c\neq 0$$

and the conclusion follows from Theorem 3.

Next let $b=(-\pi_1,1)'$, so $b'c\neq 0$ under (b) and (c). Note that if $\mu_n\neq n$, then by Assumption 2, it follows that $S_0=\operatorname{diag}(0,\ldots,0,1)\tilde{S}^{-1}=e_Gb'$ for the Gth unit vector e_g. Then under (b), where K/μ_n^2 is bounded,

$$S_0c=e_Gb'c\neq 0.$$

Therefore, $c'S_0'\Lambda_{\mathrm{I}}S_0c\neq 0$ and the conclusion follows from Theorem 3.

Finally, if (c) holds and $\mu_n^2/n\longrightarrow 0$, note that because $E[\varepsilon_i\tilde{U}_{iG}]$ has the same sign for all i, that $E[\varepsilon_i\tilde{U}_{iG}]E[\varepsilon_j\tilde{U}_{jG}]\geq 0$ for all i and j, and hence

$$\begin{aligned}&\sum_{i\neq j}P_{ij}^2(\sigma_i^2E[\tilde{U}_{jG}^2]+E[\varepsilon_i\tilde{U}_{iG}]E[\varepsilon_j\tilde{U}_{jG}])/K\\&\quad\geq\sum_{i\neq j}P_{ij}^2\sigma_i^2E[\tilde{U}_{jG}^2]/K\geq C\sum_{i\neq j}P_{ij}^2/K\geq C\Bigl(1-\sum_iP_{ii}^2/K\Bigr)\geq C.\end{aligned}$$

Then $\Psi\geq Ce_Ge_G'$ in the positive semidefinite sense, and hence by $b'e_G=1$,

$$\Lambda_{\mathrm{II}}\geq CH_P^{-1}S_0e_Ge_G'S_0'H_P^{-1}=CH_P^{-1}e_Gb'e_Ge_G'be_G'H_P^{-1}=CH_P^{-1}e_Ge_G'H_P^{-1}.$$

It then follows that

$$c'S_0'\Lambda_{\mathrm{II}}S_0c=c'be_G'\Lambda_{\mathrm{II}}e_Gb'c\geq C(c'b)^2(e_G'H_P^{-1}e_G)^2\neq 0,$$

so the conclusion follows from Theorem 3. □

Proof of Theorem 4. It suffices for us to show that for all n sufficiently large and for any fixed positive real number ρ, there exists a constant $\overline{C}$ such that

$$E\left[\det\left(\frac{X_*' M X_*}{n-K}\right)\right]^{-\rho} \leq \overline{C} < \infty.$$

To proceed, note that it is convenient here to change variables in the following way: define $H_Z = Z(Z'Z)^{-1/2} \in V_{K,n}$ (where $V_{K,n}$ denotes the Stiefel manifold, i.e., the set (or space) of $n \times K$ matrices such that $X'X = I_K$) and partition

$$H_Z = \begin{pmatrix} Z_{1\cdot}(Z'Z)^{-1/2} \\ Z_{2\cdot}(Z'Z)^{-1/2} \end{pmatrix} = \begin{pmatrix} \underset{K\times K}{H_{Z,1}} \\ \underset{(n-K)\times K}{H_{Z,2}} \end{pmatrix} \quad \text{(say)}.$$

Now define

$$\begin{aligned} H_Z^{\perp} &= \begin{bmatrix} -(H_{Z,1}')^{-1} H_{Z,2}' \\ I_{n-K} \end{bmatrix} [I_{n-K} + H_{Z,2}(H_{Z,1}' H_{Z,1})^{-1} H_{Z,2}']^{-1/2} \\ &= \begin{bmatrix} -(Z_{1\cdot}')^{-1} Z_{2\cdot}' \\ I_{n-K} \end{bmatrix} [I_{n-K} + Z_{2\cdot}(Z_{1\cdot}' Z_{1\cdot})^{-1} Z_{2\cdot}']^{-1/2} \in V_{n-K,n}. \end{aligned}$$

Note that the implicit assumption that $H_{Z,1}$ is nonsingular is really without loss of generality since $\text{rank}(Z) = K$ by Assumption 1; hence, the invertibility of Z_1 (and, thus, $H_{Z,1}$) can always be achieved, if necessary, by a repermutation of the rows of Z. Note also that, by construction,

$$P = H_Z H_Z', \qquad M = H_Z^{\perp} H_Z^{\perp\prime}, \quad \text{and} \quad (H_Z \;\; H_Z^{\perp}) \in \mathcal{O}(n),$$

where $\mathcal{O}(n)$ denotes the orthogonal group of $n \times n$ orthogonal matrices. Next, define

$$W_n = \frac{1}{\sqrt{n-K}} H_Z^{\perp\prime}[\varepsilon \;\; X] = \frac{1}{\sqrt{n-K}} H_Z^{\perp\prime} X_*,$$

so that

$$W_n' W_n = \frac{X_*' H_Z^{\perp} H_Z^{\perp\prime} X_*}{n-K} = \frac{X_*' M X_*}{n-K}.$$

Under the Gaussian heteroskedastic error assumption, the probability density function of W_n has the representation

$$\begin{aligned} f_n(W_n) &= (2\pi)^{-(n-K)L/2} (\det[\Xi_W])^{-1/2} (n-K)^{(n-K)L/2} \\ &\quad \times \exp\left\{-\frac{(n-K)}{2} (\text{vec}(W_n) - \varphi_n)' \Xi_W^{-1} (\text{vec}(W_n) - \varphi_n)\right\}, \end{aligned}$$

where

$$\Xi_W = (I_L \otimes H_Z^{\perp\prime}) K_{nL}' D_{\Omega^*} K_{nL} (I_L \otimes H_Z^{\perp}),$$

$$D_{\Omega^*} = \begin{pmatrix} \Omega_1^* & 0 & \cdots & 0 \\ 0 & \Omega_2^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Omega_n^* \end{pmatrix}, \quad \Omega_i^* = E[\overline{U}_i \overline{U}_i'] \quad \text{for } i = 1, \ldots, n,$$

$$\varphi_n = \mathrm{vec}(\Phi_n), \quad \text{and} \quad \Phi_n = \begin{pmatrix} 0_{(n-K)\times 1} & \underset{(n-K)\times G}{\frac{1}{\sqrt{n(n-K)}} H_Z^{\perp\prime} z D_\mu \widetilde{S}_n'} \end{pmatrix},$$

with K_{nL} being the $nL \times nL$ commutation matrix and $L = G + 1$. Note that

$$\begin{aligned} \lambda_{\min}(\Xi_W) &= (I_L \otimes H_Z^{\perp\prime}) K_{nL}' K_{nL} (I_L \otimes H_Z^{\perp}) \lambda_{\min}(D_{\Omega^*}) \\ &= \lambda_{\min}(D_{\Omega^*}) \quad (\text{since } H_Z^{\perp\prime} H_Z^{\perp} = I_{n-K} \text{ and } K_{nL}' K_{nL} = I_{nL}) \\ &= \min_i \lambda_{\min}(\Omega_i^*) \\ &= \lambda_*. \end{aligned}$$

Hence, we can bound $f_n(W_n)$ from above as

$$\begin{aligned} & f_n(W_n) \\ &\quad \le \lambda_*^{-(n-K)L/2} (2\pi)^{-(n-K)L/2} (n-K)^{(n-K)L/2} \\ &\qquad \times \exp\left\{ -\frac{(n-K)}{2\lambda_*} (\mathrm{vec}(W_n) - \varphi_n)' (\mathrm{vec}(W_n) - \varphi_n) \right\} \\ &\quad \le \lambda_*^{-(n-K)L/2} (2\pi)^{-(n-K)L/2} (n-K)^{(n-K)L/2} \\ &\qquad \times \exp\left\{ -\frac{(n-K)}{2\lambda_*} \mathrm{tr}[(W_n - \Phi_n)'(W_n - \Phi_n)] \right\} \\ &\quad = f_n^*(W) \quad (\text{say}). \end{aligned}$$

It follows that to complete the proof, we need to show that there exists constant $\overline{C}$ such that for all n sufficiently large and for fixed $\rho > 0$,

$$\int_{\mathbb{R}^{(n-K)L}} [\det(W_n' W_n)]^{-\rho} f_n^*(W_n)(dW_n) \le \overline{C} < \infty.$$

Write

$$\begin{aligned} f_n^*(W_n) &= \lambda_*^{-(n-K)L/2} (2\pi)^{-(n-K)L/2} (n-K)^{(n-K)L/2} \\ &\quad \times \exp\left\{ -\frac{(n-K)}{2\lambda_*} \mathrm{tr}[(W_n - \Phi_n)'(W_n - \Phi_n)] \right\} \\ &= \lambda_*^{-(n-K)L/2} (2\pi)^{-(n-K)L/2} (n-K)^{(n-K)L/2} \\ &\quad \times \exp\left\{ -\frac{(n-K)}{2\lambda_*} \mathrm{tr}[W_n' W_n] \right\} \exp\left\{ -\frac{(n-K)}{2\lambda_*} \mathrm{tr}[\Phi_n' \Phi_n] \right\} \\ &\quad \times \exp\left\{ \frac{(n-K)}{\lambda_*} \mathrm{tr}[\Phi_n W_n] \right\}. \end{aligned}$$

Next, we consider the change of variables

$$W_n = Q_1 T_n,$$

where $Q_1 \in V_{L,n-K}$ so that $Q_1' Q_1 = I_L$ and T_n is a $L \times L$ upper-triangular matrix with positive diagonal elements. From Theorem 2.1.13 of Muirhead (1982), we obtain the Jacobian of the transformation $W_n \to (Q_1, T_n)$ as

$$(dW_n) = \prod_{g=1}^{L} t_{gg}^{(n-K-g)} (dT_n)(dQ_1),$$

where t_{gg} is the gth diagonal element of T_n, so that

$$\begin{aligned}
&\int_{\mathbb{R}^{(n-K)L}} [\det(W_n' W_n)]^{-\rho} f_n^*(W_n)(dW_n) \\
&\quad = \lambda_*^{-(n-K)L/2} (2\pi)^{-(n-K)L/2} (n-K)^{(n-K)L/2} \\
&\qquad \times \exp\left\{-\frac{(n-K)}{2\lambda_*} \operatorname{tr}[\Phi_n' \Phi_n]\right\} \\
&\qquad \times \int \underset{t_{ij}}{\cdots} \int \left[\int_{V_{L,n-K}} \exp\left\{\frac{(n-K)}{\lambda_*} \operatorname{tr}[\Phi_n Q_1 T_n]\right\}(dQ_1)\right] \\
&\qquad \times \exp\left\{-\frac{(n-K)}{2\lambda_*} \operatorname{tr}(T_n' T_n)\right\} [\det(T_n' T_n)]^{-\rho} \prod_{g=1}^{L} t_{gg}^{(n-K-g)} (dT_n) \\
&\quad = \lambda_*^{-(n-K)L/2} (2\pi)^{-(n-K)L/2} (n-K)^{(n-K)L/2} \exp\left\{-\frac{(n-K)}{2\lambda_*} \operatorname{tr}[\Phi_n' \Phi_n]\right\} \\
&\qquad \times \int \underset{t_{ij}}{\cdots} \int \left[\int_{V_{L,n-K}} \exp\left\{\frac{(n-K)}{\lambda_*} \operatorname{tr}[\Phi_n Q_1 T_n]\right\}(dQ_1)\right] \\
&\qquad \times \exp\left\{-\frac{(n-K)}{2\lambda_*} \operatorname{tr}(T_n' T_n)\right\} [\det(T_n)]^{-2\rho} \prod_{g=1}^{L} t_{gg}^{(n-K-g)} (dT_n) \\
&\quad = \lambda_*^{-(n-K)L/2} (2\pi)^{-(n-K)L/2} (n-K)^{(n-K)L/2} \\
&\qquad \times \exp\left\{-\frac{(n-K)}{2\lambda_*} \operatorname{tr}[\Phi_n' \Phi_n]\right\} \\
&\qquad \times \int \underset{t_{ij}}{\cdots} \int \left[\int_{V_{L,n-K}} \exp\left\{\frac{(n-K)}{\lambda_*} \operatorname{tr}[\Phi_n Q_1 T_n]\right\}(dQ_1)\right] \\
&\qquad \times \exp\left\{-\frac{(n-K)}{2\lambda_*} \operatorname{tr}(T_n' T_n)\right\} \prod_{g=1}^{L} t_{gg}^{(n-K-2\rho-g)} (dT_n).
\end{aligned} \tag{7}$$

Consider first the integral with respect to the invariant measure defined by the differential (dQ_1). In this case, applying Lemma 9.5.3 of Muirhead (1982), we have

$$\begin{aligned}
&\int_{Q_1 \in V_{L,n-K}} \exp\left\{\frac{(n-K)}{\lambda_*} \operatorname{tr}[\Phi_n Q_1 T_n]\right\}(dQ_1) \\
&= \frac{\Gamma_{n-K-L}\left[\frac{1}{2}(n-K-L)\right]}{2^{(n-K-L)}\pi^{(n-K-L)^2/2}} \\
&\quad \times \int_{Q_1 \in V_{L,n-K}} \int_{H \in O(n-K-L)} \exp\left\{\frac{(n-K)}{\lambda_*} \operatorname{tr}[\Phi_n Q_1 T_n]\right\}(dH)(dQ_1) \\
&= \frac{2^L \pi^{(n-K)L/2}}{\Gamma_L\left[\frac{1}{2}(n-K)\right]} \int_{Q \in O(n-K)} \exp\left\{\frac{(n-K)}{\lambda_*} \operatorname{tr}[\Phi_n Q_1 T_n]\right\}[dQ] \\
&= \frac{2^L \pi^{(n-K)L/2}}{\Gamma_L\left[\frac{1}{2}(n-K)\right]} {}_0F_1\left(\frac{1}{2}(n-K); \frac{1}{4}\left(\frac{n-K}{\lambda_*}\right)^2 T_n \Phi_n' \Phi_n T_n'\right) \\
&= \frac{2^L \pi^{(n-K)L/2}}{\Gamma_L\left[\frac{1}{2}(n-K)\right]} {}_0F_1\left(\frac{1}{2}(n-K); \frac{1}{4}\left(\frac{n-K}{\lambda_*}\right)^2 \Phi_n' \Phi_n R_n\right),
\end{aligned} \tag{8}$$

where $R_n = T_n' T_n = W_n' W_n$ and where ${}_0F_1(\frac{1}{2}(n-K); \frac{1}{4}(\frac{n-K}{\lambda_*})^2 \Phi_n' \Phi_n R_n)$ denotes a hypergeometric function with matrix argument. (See Muirhead (1982, Chapter 7.3) for definition and discussion on the hypergeometric function with matrix argument.) Now, from Theorem 2.1.9 of Muirhead (1982), we have that

$$(dR_n) = 2^L \prod_{g=1}^{L} t_{gg}^{(L+1-g)}(dT_n),$$

so that the Jacobian of the transformation $T_n \to R_n$ is given by

$$(dT_n) = 2^{-L} \prod_{g=1}^{L} t_{gg}^{(-L-1+g)}(dR_n). \tag{9}$$

Making the change of variables $T_n \to R_n$ and substituting (9) and (8) into (7), we get

$$\begin{aligned}
&\int_{\mathbb{R}^{(n-K)L}} [\det(W_n' W_n)]^{-\rho} f_n^*(W_n)(dW_n) \\
&= 2^{-L} \lambda_*^{-(n-K)L/2} (2\pi)^{-(n-K)L/2} (n-K)^{(n-K)L/2} \\
&\quad \times \exp\left\{-\frac{(n-K)}{2\lambda_*} \operatorname{tr}[\Phi_n' \Phi_n]\right\} \frac{2^L \pi^{(n-K)L/2}}{\Gamma_L\left[\frac{1}{2}(n-K)\right]}
\end{aligned} \tag{10}$$

$$\begin{aligned}
&\times \int_{R_n>0} {}_0F_1\left(\frac{1}{2}(n-K); \frac{1}{4}\left(\frac{n-K}{\lambda_*}\right)^2 \Phi_n'\Phi_n R_n\right) \\
&\times (\det R_n)^{(n-K-2\rho-L-1)/2} \exp\left\{-\frac{(n-K)}{2\lambda_*}\operatorname{tr}(R_n)\right\}(dR_n) \\
&\quad \left(\text{using the fact that } \prod_{g=1}^{L} t_{gg} = (\det T_n'T_n)^{1/2} = (\det R_n)^{1/2}\right) \\
&= \lambda_*^{-(n-K)L/2}(2\pi)^{-(n-K)L/2}(n-K)^{(n-K)L/2} \\
&\times \exp\left\{-\frac{(n-K)}{2\lambda_*}\operatorname{tr}[\Phi_n'\Phi_n]\right\} \frac{\pi^{(n-K)L/2}}{\Gamma_L\left[\frac{1}{2}(n-K)\right]} \\
&\times \int_{R_n>0} \exp\left\{-\frac{(n-K)}{2\lambda_*}\operatorname{tr}(R_n)\right\}(\det R_n)^{(n-K-2\rho-L-1)/2} \\
&\times {}_0F_1\left(\frac{1}{2}(n-K); \frac{1}{4}\left(\frac{n-K}{\lambda_*}\right)^2 \Phi_n'\Phi_n R_n\right)(dR_n),
\end{aligned}$$

where the integral converges for all n such that $n-K > 2\rho + L - 1$.

To evaluate the integral with respect to R_n, we make use of the well known fact that the hypergeometric function ${}_0F_1(\cdot)$ has an infinite series representation in terms of zonal polynomials, namely

$$\begin{aligned}
&{}_0F_1\left(\frac{1}{2}(n-K); (n-K)^2 A_n R_n\right) \\
&= \sum_{k=0}^{\infty}\sum_{\kappa} \frac{1}{((n-K)/2)_\kappa} \frac{C_\kappa((n-K)^2 A_n R_n)}{k!},
\end{aligned} \tag{11}$$

where, in the definition above, we let

$$A_n = \left(\frac{1}{2\lambda_*}\right)^2 \Phi_n'\Phi_n.$$

Here, $\sum_\kappa$ denotes summation over all partitions $\kappa = (k_1, \ldots, k_L)$ of k such that $k_1 \geq \cdots \geq k_L \geq 0$ and $\sum_{g=1}^{L} k_g = k$, $C_\kappa((n-K)^2 A_n R_n)$ is the zonal polynomial of $(n-K)^2 A_n R_n$ corresponding to κ, and the generalized hypergeometric coefficient $((n-K)/2)_\kappa$ is defined by

$$\left(\frac{n-K}{2}\right)_\kappa = \prod_{g=1}^{L}\left(\frac{n-K-g+1}{2}\right)_{k_g}$$

with $(\frac{n-K-i+1}{2})_{k_g}$ being Pocchammer's symbol or forward factorial, so that

$$\begin{aligned}
\left(\frac{n-K-g+1}{2}\right)_{k_g} &= \left(\frac{n-K-g+1}{2}\right)\left(\frac{n-K-g+2}{2}\right) \times \cdots \\
&\quad \times \left(\frac{n-K-g+2k_g-1}{2}\right)
\end{aligned}$$

and $(\frac{n-K-g+1}{2})_0 = 1$. Making use of the infinite series representation (11), we apply Theorem 7.2.7 of Muirhead (1982) to integrate term-by-term to obtain, for all n sufficiently large such that $n-K > 2\rho + L - 1$,

$$\begin{aligned}
&\int_{R_n>0} {}_0F_1\left(\frac{1}{2}(n-K); (n-K)^2 A_n R_n\right)(\det R_n)^{(n-K-2\rho-L-1)/2} \\
&\qquad \times \exp\left\{-\frac{(n-K)}{2\lambda_*}\operatorname{tr}(R_n)\right\}(dR_n) \\
&\quad = \sum_{k=0}^{\infty}\sum_{\kappa}\frac{1}{((n-K)/2)_\kappa} \\
&\qquad \times \int_{R_n>0}\frac{C_\kappa((n-K)^2 A_n R_n)}{k!}(\det R_n)^{(n-K-2\rho)/2-(L+1)/2} \\
&\qquad \times \exp\left\{-\frac{(n-K)}{2\lambda_*}\operatorname{tr}(R_n)\right\}(dR_n) \\
&\quad = \Gamma_L\left(\frac{n-K-2\rho}{2}\right)\left(\det\left\{\frac{(n-K)}{2\lambda_*}\cdot I_L\right\}\right)^{-(n-K-2\rho)/2} \\
&\qquad \times \sum_{k=0}^{\infty}\sum_{\kappa}\frac{([n-K-2\rho]/2)_\kappa}{([n-K]/2)_\kappa}\frac{C_\kappa(2\lambda_*(n-K)A_n)}{k!} \\
&\quad = \Gamma_L\left(\frac{n-K-2\rho}{2}\right)\left(\frac{n-K}{2\lambda_*}\right)^{-(n-K-2\rho)L/2} \\
&\qquad \times {}_1F_1\left(\frac{n-K-2\rho}{2}; \frac{n-K}{2}; \left(\frac{n-K}{2\lambda_*}\right)\Phi_n'\Phi_n\right).
\end{aligned}$$

Next, we analyze the asymptotic behavior of the hypergeometric function ${}_1F_1(\frac{n-K-2\rho}{2}; \frac{n-K}{2}; (\frac{n-K}{2\lambda_*})\Phi_n'\Phi_n)$. Using (generalized) Kummer's relation, we can write

$$\begin{aligned}
&{}_1F_1\left(\frac{n-K-2\rho}{2}; \frac{n-K}{2}; \left(\frac{n-K}{2\lambda_*}\right)\Phi_n'\Phi_n\right) \\
&\quad = \exp\left(\operatorname{tr}\left[\left(\frac{n-K}{2\lambda_*}\right)\Phi_n'\Phi_n\right]\right)\, {}_1F_1\left(\rho; \frac{n-K}{2}; -\left(\frac{n-K}{2\lambda_*}\right)\Phi_n'\Phi_n\right).
\end{aligned}$$

Now, using (A.6.17) of Chikuse (2003), we obtain

$$\begin{aligned}
&{}_1F_1\left(\rho; \frac{n-K}{2}; -\left(\frac{n-K}{2\lambda_*}\right)\Phi_n'\Phi_n\right) \\
&\quad = \left[\det\left(I_L + \frac{\Phi_n'\Phi_n}{\lambda_*}\right)\right]^{-\rho}(1+O(n^{-1})) \\
&\quad = O(1).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\int_{R_n>0} {}_0F_1\left(\frac{1}{2}(n-K);(n-K)^2A_nR_n\right)(\det R_n)^{(n-K-2\rho-L-1)/2} \\
&\qquad\times \exp\left\{-\frac{(n-K)}{2\lambda_*}\operatorname{tr}(R_n)\right\}(dR_n) \\
&\quad=\Gamma_L\left(\frac{n-K-2\rho}{2}\right)\left(\frac{n-K}{2\lambda_*}\right)^{-(n-K-2\rho)L/2}\exp\left(\operatorname{tr}\left[\left(\frac{n-K}{2\lambda_*}\right)\Phi_n'\Phi_n\right]\right) \\
&\qquad\times\left[\det\left(I_L+\frac{\Phi_n'\Phi_n}{\lambda_*}\right)\right]^{-\rho}(1+O(n^{-1})).
\end{aligned} \tag{12}$$

Substituting (12) into (10), we have

$$\begin{aligned}
&\int_{\mathbb{R}^{(n-K)L}}[\det(W_n'W_n)]^{-\rho}f_n^*(W_n)(dW_n) \\
&\quad=\lambda_*^{-(n-K)L/2}(2\pi)^{-(n-K)L/2}(n-K)^{(n-K)L/2} \\
&\qquad\times\exp\left\{-\frac{(n-K)}{2\lambda_*}\operatorname{tr}[\Phi_n'\Phi_n]\right\}\frac{\pi^{(n-K)L/2}}{\Gamma_L\left[\frac{1}{2}(n-K)\right]} \\
&\qquad\times\Gamma_L\left(\frac{n-K-2\rho}{2}\right)\left(\frac{n-K}{2\lambda_*}\right)^{-(n-K-2\rho)L/2}\exp\left(\operatorname{tr}\left[\left(\frac{n-K}{2\lambda_*}\right)\Phi_n'\Phi_n\right]\right) \\
&\qquad\times\left[\det\left(I_L+\frac{\Phi_n'\Phi_n}{\lambda_*}\right)\right]^{-\rho}(1+O(n^{-1})) \\
&\quad=\frac{\Gamma_L\left[\frac{1}{2}(n-K-2\rho)\right]}{\Gamma_L\left[\frac{1}{2}(n-K)\right]}(n-K)^{\rho L}(2\lambda_*)^{-\rho L}\left[\det\left(I_L+\frac{\Phi_n'\Phi_n}{\lambda_*}\right)\right]^{-\rho}(1+O(n^{-1})).
\end{aligned} \tag{13}$$

Now, multivariate gamma function can be written as a product of ordinary gamma functions as

$$\Gamma_L\left[\frac{1}{2}(n-K-2\rho)\right]=\pi^{L(L-1)/4}\prod_{g=1}^{L}\Gamma\left[\frac{n-K-2\rho-g+1}{2}\right].$$

(See, for example, Muirhead (1982, Theorem 2.1.12).) Using the Stirling approximation, we have, for $g=1,\dots,L$,

$$\begin{aligned}
&\Gamma\left[\frac{n-K-2\rho-g+1}{2}\right] \\
&\quad=\left(\frac{4\pi}{n-K-2\rho-g+1}\right)^{1/2}\left(\frac{n-K-2\rho-g+1}{2e}\right)^{(n-K-2\rho-g+1)/2}(1+O(n^{-1})) \\
&\quad=(4\pi)^{1/2}(2e)^{-(n-K-2\rho-g+1)/2}(n-K-2\rho-g+1)^{(n-K-2\rho-g)/2}(1+O(n^{-1}))
\end{aligned}$$

$$
\begin{aligned}
&= (4\pi)^{1/2}(2e)^{-(n-K-2\rho-g+1)/2}(n-K)^{(n-K-2\rho-g)/2} \\
&\quad\times \left(1-\frac{2\rho+g-1}{n-K}\right)^{(n-K)/2}(1+O(n^{-1})) \\
&= (4\pi)^{1/2}(2e)^{-(n-K-2\rho-g+1)/2}(n-K)^{(n-K-2\rho-g)/2}e^{-(2\rho+g-1)/2}(1+O(n^{-1})) \\
&= \sqrt{\pi}2^{g/2}(n-K)^{-g/2}2^{-(n-K-2\rho-1)/2}e^{-(n-K)/2}(n-K)^{(n-K-2\rho)/2}(1+O(n^{-1})),
\end{aligned}
$$

so that

$$
\begin{aligned}
&\Gamma_L\left[\frac{1}{2}(n-K-2\rho)\right] \\
&\quad= \pi^{L(L-1)/4}\prod_{g=1}^{L}\Gamma\left[\frac{n-K-2\rho-g+1}{2}\right] \\
&\quad= \pi^{L(L-1)/4}\prod_{g=1}^{L}\sqrt{\pi}2^{g/2}(n-K)^{-g/2}2^{-(n-K-2\rho-1)/2} \\
&\quad\quad\times e^{-(n-K)/2}(n-K)^{(n-K-2\rho)/2}(1+O(n^{-1})) \\
&\quad= (2\pi)^{L(L+1)/4}(n-K)^{-L(L+1)/4}2^{-(n-K-2\rho-1)L/2} \\
&\quad\quad\times e^{-(n-K)L/2}(n-K)^{(n-K-2\rho)L/2}(1+O(n^{-1})) \\
&\quad= (2\pi)^{L(L+1)/4}2^{-(n-K-2\rho-1)L/2} \\
&\quad\quad\times e^{-(n-K)L/2}(n-K)^{L[n-K-2\rho-(L+1)/2]/2}(1+O(n^{-1})).
\end{aligned}
\tag{14}
$$

Similarly, we have

$$
\begin{aligned}
\Gamma_L\left[\frac{1}{2}(n-K)\right] &= \pi^{L(L-1)/4}\prod_{g=1}^{L}\Gamma\left[\frac{n-K-g+1}{2}\right] \\
&= (2\pi)^{L(L+1)/4}2^{-(n-K-1)L/2}e^{-(n-K)L/2} \\
&\quad\times (n-K)^{L[n-K-(L+1)/2]/2}(1+O(n^{-1})).
\end{aligned}
\tag{15}
$$

Applying (14) and (15) to (13), we obtain

$$
\begin{aligned}
&\int_{\mathbb{R}^{(n-K)L}}[\det(W_n'W_n)]^{-\rho}f_n^*(W_n)(dW_n) \\
&\quad= \frac{\Gamma_L\left[\frac{1}{2}(n-K-2\rho)\right]}{\Gamma_L\left[\frac{1}{2}(n-K)\right]}(n-K)^{\rho L}(2\lambda_*)^{-\rho L}\left[\det\left(I_L+\frac{\Phi_n'\Phi_n}{2\lambda_*}\right)\right]^{-\rho}(1+O(n^{-1})) \\
&\quad= (2\pi)^{L(L+1)/4}2^{-(n-K-2\rho-1)L/2}e^{-(n-K)L/2} \\
&\quad\quad\times (n-K)^{L[n-K-2\rho-(L+1)/2]/2}(2\pi)^{-L(L+1)/4}2^{(n-K-1)L/2}
\end{aligned}
$$

$$
\begin{aligned}
&\times e^{(n-K)L/2}(n-K)^{-L[n-K-(L+1)/2]/2}(n-K)^{\rho L} \\
&\times (2\lambda_*)^{-\rho L}\left[\det\left(I_L + \frac{\Phi_n' \Phi_n}{2\lambda_*}\right)\right]^{-\rho}(1+O(n^{-1})) \\
&= \lambda_*^{-\rho L}\left[\det\left(I_L + \frac{\Phi_n' \Phi_n}{2\lambda_*}\right)\right]^{-\rho}(1+O(n^{-1})) \\
&\leq \lambda_*^{-\rho L}(1+O(n^{-1})) \\
&= O(1). \qquad \square
\end{aligned}
$$

References

Ackerberg, D. A. and P. Devereux (2003), "Improved JIVE estimators for overidentified models with and without heteroskedasticity." Working paper, UCLA. [212, 216]

Amemiya, T. (1973), "Regression analysis when the dependent variable is truncated normal." *Econometrica*, 41, 997–1016. [215]

Amemiya, T. (1984), "Tobit models: A survey." *Journal of Econometrics*, 24, 3–61. [215]

Andrews, D. W. K. and J. H. Stock (2007), "Testing with many weak instruments." *Journal of Econometrics*, 138, 24–46. [213]

Angrist, J. D., G. W. Imbens, and A. Krueger (1999), "Jackknife instrumental variables estimation." *Journal of Applied Econometrics*, 14, 57–67. [212, 216, 219]

Bekker, P. A. (1994), "Alternative approximations to the distributions of instrumental variables estimators." *Econometrica*, 62, 657–681. [212, 213, 216, 221, 223]

Bekker, P. A. and J. van der Ploeg (2005), "Instrumental variable estimation based on grouped data." *Statistica Neerlandica*, 59, 239–267. [212, 213, 218, 219, 223]

Blomquist, S. and M. Dahlberg (1999), "Small sample properties of LIML and jackknife IV estimators: Experiments with weak instruments." *Journal of Applied Econometrics*, 14, 69–88. [212]

Chao, J. C., J. A. Hausman, W. K. Newey, N. R. Swanson, and T. Woutersen (2012a), "An expository note on the existence of moments of Fuller and HFUL estimators." Working paper, University of Arizona. [230]

Chao, J. C. and N. R. Swanson (2004), "Estimation and testing using jackknife IV in heteroskedastic regressions with many weak instruments." Working paper, University of Maryland and Rutgers University. [212, 213, 216, 218, 223]

Chao, J. C. and N. R. Swanson (2005), "Consistent estimation with a large number of weak instruments." *Econometrica*, 73, 1673–1692. [213, 220, 221, 223]

Chao, J. C., N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen (2012b), "Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments." *Econometric Theory*, 28, 42–86. [213, 214, 219, 222, 224, 231, 234, 237, 238]

Chikuse, Y. (2003), *Statistics on Special Manifolds*. Lecture Notes in Statistics, Vol. 174. Springer-Verlag, New York. [250]

Davidson, R. and J. G. MacKinnon (2006), "The case against JIVE" (with discussion and reply). *Journal of Applied Econometrics*, 21, 827–833. [212, 216, 217]

Fuller, W. A. (1977), "Some properties of a modification of the limited information estimator." *Econometrica*, 45, 939–954. [211, 212, 213, 214, 217, 218, 226]

Hahn, J. and J. A. Hausman (2002), "A new specification test for the validity of instrumental variables." *Econometrica*, 70, 163–189. [218]

Hahn, J., J. A. Hausman, and G. M. Kuersteiner (2004), "Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations." *Econometrics Journal*, 7, 272–306. [212, 217]

Han, C. and P. C. B. Phillips (2006), "GMM with many moment conditions." *Econometrica*, 74, 147–192. [213, 216]

Hansen, C., J. A. Hausman, and W. K. Newey (2008), "Estimation with many instrumental variables." *Journal of Business & Economic Statistics*, 26, 398–422. [213, 220, 223, 224, 231]

Kunitomo, N. (1980), "Asymptotic expansions of distributions of estimators in a linear functional relationship and simultaneous equations." *Journal of the American Statistical Association*, 75, 693–700. [213, 221, 223]

Morimune, K. (1983), "Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size." *Econometrica*, 51, 821–841. [221]

Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory*. Wiley, New York. [247, 248, 250, 251]

Newey, W. K. (1990), "Efficient instrumental variables estimation of nonlinear models." *Econometrica*, 58, 809–837. [214, 221]

Newey, W. K. and D. McFadden (1994), "Large sample estimation and hypothesis testing." In *Handbook of Econometrics*, Vol. 4 (R. F. Engle and D. L. McFadden, eds.), Chapter 36, Elsevier, Amsterdam, The Netherlands. [215]

Newey, W. K. and F. Windmejer (2009), "Generalized method of moments with many weak moment conditions." *Econometrica*, 77, 687–719. [212]

Phillips, G. D. A. and C. Hale (1977), "The bias of instrumental variable estimators of simultaneous equation systems." *International Economic Review*, 18, 219–228. [212]

Staiger, D. and J. Stock (1997), "Instrumental variables regression with weak instruments." *Econometrica*, 65, 557–586. [213, 220]

Stock, J. and M. Yogo (2005), "Asymptotic distributions of instrumental variables statistics with many instruments." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (D. W. K. Andrews and J. H. Stock, eds.), Chapter 6, Cambridge University Press, Cambridge. [213]

van Hasselt, M. (2010), "Many instruments asymptotic approximations under nonnormal error distributions." *Econometric Theory*, 26, 633–645. [223]

White, H. (1982), "Instrumental variables regression with independent observations." *Econometrica*, 50, 483–500. [211, 212]

Submitted June, 2010. Final version accepted February, 2012.