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.
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 Devereaux (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

\[
\begin{align*}
\mathbf{y} & \sim \mathbf{X} \ \mathbf{\delta}_0 + \mathbf{\varepsilon}, \\
\mathbf{X} & \sim \mathbf{Y} + \mathbf{U},
\end{align*}
\]

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
In the Monte Carlo results given below, we try different values of $C$ with the corresponding columns of $X$ being 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 $\pi$. The model also permits $Z$ to approximate the reduced form. For example, let $X_i', Y_i'$, and $Z_i'$ denote the $i$th 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 $Z_i = (p_{1K}(w_i), \ldots, p_{KK}(w_i))'$ 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 $ij$th element of $P$, and let $\bar{X} = [y, X]$. Let $\bar{X}_i'$ be a row vector consisting of the $i$th row of $\bar{X}$ and let

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

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

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

In the Monte Carlo results given below, we try different values of $C$ and recommend $C = 1$.\footnote{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.} HFUL is given by

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

(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 $\hat{\delta}_i = y_i - X_i'\hat{\delta}$, $\hat{\gamma} = X'\hat{\delta}/\hat{\alpha}'\hat{\delta}$, $\hat{X} = X - \hat{\delta}\hat{\gamma}'$, $\hat{X} = P\hat{X}$, and $\hat{Z} = Z(Z'Z)^{-1}$. Also let

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

\footnote{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 $\det((\bar{X}'\bar{X})^{-1/2}(\bar{X}'P\bar{X} - \sum_{i=1}^n P_{ii}\bar{X}_i\bar{X}_i')(\bar{X}'\bar{X})^{-1/2} - \lambda I) = 0$. Moreover, this is equivalent to solving $\det((\bar{X}'\bar{X})^{-1/2}\det((\bar{X}'\bar{X})^{-1/2}(\bar{X}'P\bar{X} - \sum_{i=1}^n P_{ii}\bar{X}_i\bar{X}_i')(\bar{X}'\bar{X})^{-1/2} - \lambda I) = 0$. 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, $\lambda$ is real almost surely.}
\[ \hat{\Sigma} = \sum_{i=1}^{n} (\hat{X}_i \hat{X}_i' - \hat{X}_i \hat{P}_{ii} \hat{X}_i') \hat{\epsilon}_i^2 + K \sum_{k=1}^{K} \sum_{\ell=1}^{K} \left( \sum_{i=1}^{n} \hat{Z}_{ik} \hat{Z}_{i\ell} \hat{X}_i \hat{\epsilon}_i \right) \left( \sum_{j=1}^{n} \hat{Z}_{jk} \hat{Z}_{j\ell} \hat{X}_j \hat{\epsilon}_j \right)'. \]

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 \( \delta_0 \) and variance \( \hat{V} \). Asymptotic \( t \)-ratios \( \hat{\delta}_j / \sqrt{V_{jj}} \) will be asymptotically normal. Also, defining \( q_\alpha \) as the \( 1 - \alpha/2 \) quantile of a \( N(0, 1) \) distribution, an asymptotic \( 1 - \alpha \) confidence interval for \( \delta_0 \) is given by \( \hat{\delta}_k \pm q_\alpha \sqrt{V_{kk}} \). 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

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

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

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

\[ A_n = Y' P Y / n - \sum_{i=1}^{n} P_{ii} Y_i Y_i' / n. \]
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 $\delta_0$. 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 $\delta_0$ 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 $\delta_0$. This leads to an objective function of the form

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

The expected value of this objective function is

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

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

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

This is the JIVE2 estimator of Angrist, Imbens, and Krueger (1999). Since the objective function for $\hat{\delta}_{JIV}$ has expectation minimized at $\delta_0$, we expect that $\hat{\delta}_{JIV}$ 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

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

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

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

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

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

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 \( \hat{\alpha} = \hat{Q}_{HLIM}(\tilde{\delta}) \) is the smallest eigenvalue of

\[ (\bar{X}'\bar{X})^{-1}(\bar{X}'\bar{P}\bar{X} - \sum_{i=1}^{n} P_{ii}\bar{X}_i\bar{X}_i'). \]

Solving the first order conditions gives

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

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 \( \hat{d} = (1, -\tilde{\delta}') \) solves

\[ \min_{d, d_1 = 1} \frac{d'(\bar{X}'\bar{X} - \sum_{i=1}^{n} P_{ii}\bar{X}_i\bar{X}_i')d}{d'X'Xd}. \]

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 \( \hat{\alpha} \) in the HLIM estimator with \( \hat{\alpha} = \hat{\alpha} - (1 - \hat{\alpha})C/n \), 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

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

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 \( \hat{\alpha} = \hat{\alpha} \) and JIV does with \( \hat{\alpha} = 0 \).
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

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

The first term following the equality will be minimized at $\delta_0$. The second term may not have a critical value at $\delta_0$, so the objective function will not be minimized at $\delta_0$. To see this, let $\sigma_i^2 = E[\epsilon_i^2]$, $\gamma_i = E[X_i\epsilon_i]/\sigma_i^2$, and $\tilde{\gamma} = \sum_{i=1}^{n} E[X_i\epsilon_i]/\sum_{i=1}^{n} \sigma_i^2 = \sum_i \gamma_i\sigma_i^2/\sum_i \sigma_i^2$. Then

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

$$= \frac{-2}{\sum_{i=1}^{n} \sigma_i^2} \left[ \sum_{i=1}^{n} P_{ii}(\gamma_i - \tilde{\gamma})\sigma_i^2 \right] = \frac{-2}{\sum_{i=1}^{n} \sigma_i^2} \text{Cov}_{\sigma_i^2}(P_{ii}, \gamma_i),$$

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

$$\lim_{n \to \infty} \text{Cov}_{\sigma_i^2}(P_{ii}, \gamma_i) \neq 0,$$

the LIML objective function will not have a zero derivative at $\delta_0$ asymptotically, so it is not minimized at $\delta_0$. 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 $\lim_{n \to \infty} \text{Cov}_{\sigma_i^2}(P_{ii}, \gamma_i) = 0$ for consistency of LIML.

Note that $\text{Cov}_{\sigma_i^2}(P_{ii}, \gamma_i) = 0$ when either $\gamma_i$ or $P_{ii}$ does not depend on $i$. Thus, it is variation in $\gamma_i = E[X_i\epsilon_i]/\sigma_i^2$, the coefficients from the projection of $X_i$ on $\epsilon_i$, 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,

$$\hat{Q}_{\text{LIML}}(\delta) = \hat{Q}_{\text{HLIM}}(\delta) + \sum_{i} P_{ii}(y_{i} - X'_{i}\delta)^{2} / (y - X\delta)^{2}(y - X\delta)' = \hat{Q}_{\text{HLIM}}(\delta) + P_{11},$$

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 $Z'_{i}$, $e_{i}$, $U'_{i}$, and $Y'_{i}$ denote the $i$th row of $Z$, $e$, $U$, 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, rank($Z$) = $K$, and there is a constant $C$ such that $P_{ii} \leq C < 1$ ($i = 1, \ldots, n$), $K \to \infty$.

The restriction that 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 $p_{K}(w) = w^{K-1}$, it can be shown that $Z'Z$ is nonsingular with probability 1 for $K < n$. The condition $P_{ii} \leq C < 1$ implies that $K/n \leq C$, because $K/n = \sum_{i=1}^{n} P_{ii}/n \leq C$.

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 $S_{n} = \hat{S} \text{ diag}(\mu_{1n}, \ldots, \mu_{Gn})$ and $\hat{S}$ is nonsingular. Also, for each $j$, either $\mu_{jn} = \sqrt{n}$ or $\mu_{jn}/\sqrt{n} \to 0$, $\mu_{n} = \min_{1 \leq j \leq G} \mu_{jn} \to \infty$, and $\sqrt{K}/\mu_{n}^{2} \to 0$. Also, there is $C > 0$ such that $\|\sum_{i=1}^{n} z_{i}z_{i}'/n\| \leq C$ and $\lambda_{\min}(\sum_{i=1}^{n} z_{i}z_{i}'/n) \geq 1/C$ for $n$ sufficiently large.

---

3The observations $w_{1}, \ldots, w_{n}$ are distinct with probability 1 and, therefore, by $K < n$, cannot all be roots of a $K$th degree polynomial. It follows that for any nonzero $a$, there must be some $i$ with $a'Z_{i} = a'p^{K}(w_{i}) \neq 0$, implying that $a'Z'Z_{a} > 0$. 
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 \( \delta \), 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 \( \mu_n \) determines how strongly \( \delta_G \) is identified and absorb into \( z_{iG} \) any other terms, such as unknown coefficients. For Assumption 2, we let \( z_i = (Z_{i1}', z_i) \) 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 \( \delta_G \) be determined by \( \mu_n \). For example, if \( \mu_n = \sqrt{n} \), then the coefficient on \( z_{iG} \) does not shrink, corresponding to strong identification of \( \delta_G \). If \( \mu_n \) grows slower than \( \sqrt{n} \), then \( \delta_G \) will be more weakly identified. Indeed, \( 1/\mu_n \) will be the convergence rate for estimators of \( \delta_G \). We require \( \mu_n \to \infty \) to avoid the weak instrument setting of Staiger and Stock (1997), where \( \delta_G \) is not asymptotically identified.

For this model, the reduced form is

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

This reduced form is as specified in Assumption 2 with

\[ S_n = \begin{bmatrix} I \\ \pi_1 \end{bmatrix}, \quad \mu_{jn} = \sqrt{n}, \quad 1 \leq j \leq G_1, \quad \mu_{Gn} = \mu_n. \]

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

$$z_{iG} = \pi'_2 Z_{i2}/\sqrt{K - G_1}, \quad \mu_n = \sqrt{K - G_1}.$$  

Assumption 2 will then require that

$$\sum_i z_{iG}^2/n = (K - G_1)^{-1} \frac{\sum_i (\pi'_2 Z_{i2})^2}{n}$$

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 $\delta_{G_1}$ 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 $p^K(w_i) \equiv (p_{1K}(w_i), \ldots, p_{K-G_1,K}(w_i))'$ 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 $\mu_n/\sqrt{n} \rightarrow 0$, it is a modified version where $\delta_G$ 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 $\mu_n^2$ can be thought of as a version of the concentration parameter, determining the convergence rate of estimators of $\delta_{G_1}$, 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 $\mu_n^2$ 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 $(\varepsilon_1, U_1), \ldots, (\varepsilon_n, U_n)$ are independent, with $E[\varepsilon_i] = 0$, $E[U_i] = 0$, $E[\varepsilon_i^2] < C$, $E[||U_i||^2] \leq C$, $\text{Var}((\varepsilon_i, U_i)') = \text{diag}(\Omega_i^*, 0)$, 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 $\pi_{Kn}$ such that $\sum_{i=1}^n ||z_i - \pi_{Kn} Z_i||^2/n \rightarrow 0$. 

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

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

$$= \sum_{i=1}^{n} (1 - P_{ii})Y_i'Y_i/n + o(1) \geq (1 - C) \sum_{i=1}^{n} Y_i'Y_i/n,$$

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 $X_i$. 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 $\mu_n^{-1}S_n^i(\hat{\delta} - \delta_0) \xrightarrow{P} 0$ 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 $\pi'_1\hat{\delta}_1 + \hat{\delta}^2 = o_p(\mu_n/\sqrt{n})$.

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, $\sum_{i=1}^{n} \|z_i\|^4 / n^2 \rightarrow 0$, $E[z_i^n] \leq C$, and $E[\|U_i\|^4] \leq C$.

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 $\mu_n^2$. Also, the asymptotic variance of the estimator will depend on the growth rate of $K$ relative to $\mu_n^2$. Let $\sigma_i^2 = E[\varepsilon_i^2]$ and $\gamma_n = \sum_{i=1}^{n} E[U_i\varepsilon_i]/\sum_{i=1}^{n} \sigma_i^2$, $\bar{U} = U - \varepsilon_\gamma'$, having ith row $\bar{U}_i$; and let $\bar{\Omega}_i = E[\bar{U}_i\bar{U}_i']$.

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

Case I. $K/\mu_n^2 \rightarrow \alpha$ for finite $\alpha$.

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

Also $H_P = \lim_{n \rightarrow \infty} \sum_{i=1}^{n} (1 - P_{ii})z_i^2/\gamma_n$, $\Sigma_P = \lim_{n \rightarrow \infty} \sum_{i=1}^{n} (1 - P_{ii})^2z_i^2/\gamma_n$, and $\Psi = \lim_{n \rightarrow \infty} \sum_{i \neq j} P_{ij}^2(\sigma_i^2 E[U_j\bar{U}_j] + E[U_j\varepsilon_i]E[\varepsilon_j\bar{U}_j])/K$ 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

$$S_n^i(\hat{\delta} - \delta_0) \xrightarrow{d} N(0, A_1), \quad (3)$$

where

$$A_1 = H_P^{-1}\Sigma_P H_P^{-1} + \alpha H_P^{-1}S_0\Psi S_0' H_P^{-1}. $$
In Case II, we have that

\[(\mu_n/\sqrt{K})S_0'(\hat{\delta} - \delta_0) \xrightarrow{d} N(0, A_{II}),\] (4)

where

\[A_{II} = H_p^{-1}S_0\Psi S_0'H_p^{-1}.\]

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, \(A_{I, corr}\), 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, \(A_{II}\), only contains the adjustment for many instruments. This is because \(K\) is growing faster than \(\mu_n^2\). Also, \(A_{II}\) 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 \(\hat{\alpha} = \bar{\alpha} + O_p(1/n)\) 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 \((\epsilon_i, U_i')\) will not depend on \(i\) (e.g., so that \(\sigma_i^2 = \sigma^2\)). Then \(\gamma_n = E[X_i'\epsilon_i]/\sigma^2 = \gamma\) and \(E[U_i'\epsilon_i] = E[U_i'\epsilon_i] - \gamma\sigma^2 = 0\), so that

\[\Sigma_p = \sigma^2\tilde{H}_p, \quad \tilde{H}_p = \lim_{n \to \infty} \sum_{i=1}^n (1 - P_{ii})^2 z_i z_i'/n,
\]

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

Focusing on Case I, letting \(\Gamma = \alpha\sigma^2 S_0 E[U_i'U_i']S_0'\), the asymptotic variance of HLIM is

\[V = \sigma^2 H_p^{-1}\tilde{H}_p H_p^{-1} + \lim_{n \to \infty} \left(1 - \sum_{i=1}^n P_{ii}^2/K\right)H_p^{-1} \Gamma H_p^{-1}.
\]
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 $H = \lim_{n \to \infty} \sum_{i=1}^{n} z_i z_i' / n$ and $\tau = \lim_{n \to \infty} K / n$, 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 $\max_{i \leq n} P_{ii} \to 0$, we will have $H_p = \hat{H}_p = H$ and $\lim_{n \to \infty} \sum_i P_{ii}^2 / K \to 0$, so that the asymptotic variances of HLIM and LIML are the same and are equal to $\sigma^2 H^{-1} + H^{-1} \Gamma H^{-1}$. 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 $\mu^2_n$ 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 $\Omega = E[U_i^2]$ and $\kappa = \lim_{n \to \infty} \sum_{i=1}^{n} P_{ii}^2 / K$, we find that

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

Since $\tau \kappa - \tau^2$ is the limit of the sample variance of $P_{ii}$, which we assume to be positive, $V \geq V^*$ if and only if $\bar{z}^2 \geq \hat{\Omega}$. 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 $\hat{U}_i$. 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

$$ V_{JIV} = \sigma^2 H_p^{-1} \hat{H}_p H_p^{-1} + \lim_{n \to \infty} \left(1 - \sum_{i=1}^{n} P_{ii}^2 / K\right) H_p^{-1} (\Gamma + 2a_s S_0 E[U_i \varepsilon_i] E[\varepsilon_i U_i'] S_0') H_p^{-1}, $$

which is greater than the asymptotic variance of HLIM because $E[U_i \varepsilon_i] E[\varepsilon_i U_i']$ 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 $\hat{\alpha} = \hat{\alpha} + O_p(1/n)$ or $\hat{\delta}$ is HLIM or HFUL, there exists a $C$ with $\|z_i\| \leq C$ for all $i$, and there exists a $\pi_n$ such that $\max_{i \leq n} \|z_i -
\[ \pi_n Z_i \| \to 0, \text{ then in Case I, } S_n' \hat{V} S_n \overset{p}{\to} \Lambda_1 \text{ and in Case II, } \mu_n^2 S_n' \hat{V} S_n / K \overset{p}{\to} \Lambda_{II}. \]  Also, if \( c' S_0' \Lambda_1 S_0 c \neq 0 \) in Case I or \( c' S_0' \Lambda_{II} S_0 c \neq 0 \) in Case II, then
\[ \frac{c' (\hat{\delta} - \delta_0)}{\sqrt{c' \hat{V} c}} \overset{d}{\to} N(0, 1). \]

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 \( z_i = (Z_{i1}', z_{i2}) \), \( \hat{\alpha} = \alpha + O_p(1/n) \) or \( \hat{\delta} \) is HLIM or HFUL, there exists a \( C \) with \( \| z_i \| \leq C \) for all \( i \), and there exists a \( \pi_n \) such that \( \max_{i \leq n} \| z_i - \pi_n Z_i \| \to 0 \), \( \inf_i \sigma_i^2 \geq C \) and either (a) \( c \neq 0 \) and \( \mu_n^2 = n \), (b) \( K / \mu_n^2 \) is bounded and \( (-\pi_1, 1) c \neq 0 \), or (c) \( K / \mu_n^2 \to \infty \), \( (-\pi_1, 1) c \neq 0 \), \( \inf_i E[e_i U_{iG}] > 0 \), and the sign of \( E[e_i U_{iG}] \) is constant, then
\[ \frac{c' (\hat{\delta} - \delta_0)}{\sqrt{c' \hat{V} c}} \overset{d}{\to} N(0, 1). \]

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 \( \delta_2 \) 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 \( E[e_i U_{iG}] \) is constant, it follows from Corollary 1 that
\[ \frac{\hat{\delta}_G - \delta_{0G}}{\sqrt{\hat{V}_{GG}}} \overset{d}{\to} N(0, 1). \]

Thus the \( t \)-ratio for the coefficient of the endogenous variable is asymptotically correct across a wide range of different growth rates for \( \mu_n \) and \( K \). The analogous result holds for each coefficient \( \delta_j, j \leq G_1 \), 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 \( i \)th instrument observation is

\[
Z_i' = (1, z_{i1}, z_{i1}^2, z_{i1}^3, z_{i1}^4, z_{i1}D_{i1}, \ldots, z_{i1}D_{iK-5}),
\]

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, \( \varepsilon \), is allowed to be heteroskedastic, being given by

\[
\varepsilon = \rho U_2 + \sqrt{\frac{1 - \rho^2}{\phi^2 + (0.86)^4}} (\phi v_1 + 0.86 v_2), \quad v_1 \sim N(0, z_{i1}^2), \quad v_2 \sim N(0, (0.86)^2),
\]

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 \( \sigma^2_{\varepsilon} \) 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 \( \gamma_i = E[X_i\varepsilon_i]/\sigma^2_{\varepsilon} \) so that LIML is not consistent.

We report properties of estimators and \( t \)-ratios for \( \delta_2 \). 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 \( \pi \) 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 \( \phi \) so that the \( R \)-squared for the regression of \( \varepsilon^2 \) 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 \( \delta_2 \) 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>
<thead>
<tr>
<th>( \mu^2 )</th>
<th>( K )</th>
<th>LIML</th>
<th>HLIM</th>
<th>FULL1</th>
<th>HFUL</th>
<th>JIVE</th>
<th>CUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2</td>
<td>0.005</td>
<td>0.005</td>
<td>0.042</td>
<td>0.043</td>
<td>-0.034</td>
<td>0.005</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>0.024</td>
<td>0.023</td>
<td>0.057</td>
<td>0.057</td>
<td>0.053</td>
<td>0.025</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>0.065</td>
<td>0.065</td>
<td>0.086</td>
<td>0.091</td>
<td>0.164</td>
<td>0.071</td>
</tr>
<tr>
<td>32</td>
<td>2</td>
<td>0.002</td>
<td>0.002</td>
<td>0.011</td>
<td>0.011</td>
<td>-0.018</td>
<td>0.002</td>
</tr>
<tr>
<td>32</td>
<td>10</td>
<td>0.002</td>
<td>0.001</td>
<td>0.011</td>
<td>0.011</td>
<td>-0.019</td>
<td>0.002</td>
</tr>
<tr>
<td>32</td>
<td>30</td>
<td>0.003</td>
<td>0.002</td>
<td>0.013</td>
<td>0.013</td>
<td>-0.014</td>
<td>0.006</td>
</tr>
</tbody>
</table>

*Results based on 20,000 simulations.*
Table 2. Nine decile range: 0.05 to 0.95 $R^2$ $\varepsilon^2 | z_1 | = 0.$

<table>
<thead>
<tr>
<th>$\mu^2$</th>
<th>$K$</th>
<th>LIML</th>
<th>HLIM</th>
<th>FULL1</th>
<th>HFUL</th>
<th>JIVE</th>
<th>CUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2</td>
<td>1.470</td>
<td>1.466</td>
<td>1.072</td>
<td>1.073</td>
<td>3.114</td>
<td>1.470</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>2.852</td>
<td>2.934</td>
<td>1.657</td>
<td>1.644</td>
<td>5.098</td>
<td>3.101</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>5.036</td>
<td>5.179</td>
<td>2.421</td>
<td>2.364</td>
<td>6.787</td>
<td>6.336</td>
</tr>
<tr>
<td>32</td>
<td>2</td>
<td>0.616</td>
<td>0.616</td>
<td>0.590</td>
<td>0.589</td>
<td>0.679</td>
<td>0.616</td>
</tr>
<tr>
<td>32</td>
<td>10</td>
<td>0.715</td>
<td>0.716</td>
<td>0.679</td>
<td>0.680</td>
<td>0.816</td>
<td>0.770</td>
</tr>
<tr>
<td>32</td>
<td>30</td>
<td>0.961</td>
<td>0.985</td>
<td>0.901</td>
<td>0.913</td>
<td>1.200</td>
<td>1.156</td>
</tr>
</tbody>
</table>

*Results based on 20,000 simulations.*

Table 3. 0.05 rejection frequencies: $R^2$ $\varepsilon^2 | z_1 | = 0.$

<table>
<thead>
<tr>
<th>$\mu^2$</th>
<th>$K$</th>
<th>LIML</th>
<th>HLIM</th>
<th>FULL1</th>
<th>HFUL</th>
<th>JIVE</th>
<th>CUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2</td>
<td>0.025</td>
<td>0.026</td>
<td>0.021</td>
<td>0.034</td>
<td>0.051</td>
<td>0.012</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>0.035</td>
<td>0.037</td>
<td>0.029</td>
<td>0.044</td>
<td>0.063</td>
<td>0.027</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>0.045</td>
<td>0.049</td>
<td>0.040</td>
<td>0.054</td>
<td>0.068</td>
<td>0.051</td>
</tr>
<tr>
<td>32</td>
<td>2</td>
<td>0.041</td>
<td>0.042</td>
<td>0.037</td>
<td>0.044</td>
<td>0.038</td>
<td>0.030</td>
</tr>
<tr>
<td>32</td>
<td>10</td>
<td>0.041</td>
<td>0.042</td>
<td>0.038</td>
<td>0.044</td>
<td>0.046</td>
<td>0.041</td>
</tr>
<tr>
<td>32</td>
<td>30</td>
<td>0.042</td>
<td>0.047</td>
<td>0.039</td>
<td>0.050</td>
<td>0.057</td>
<td>0.062</td>
</tr>
</tbody>
</table>

*Results based on 20,000 simulations.*

Table 4. Median bias: $R^2$ $\varepsilon^2 | z_1 | = 0.$

<table>
<thead>
<tr>
<th>$\mu^2$</th>
<th>$K$</th>
<th>LIML</th>
<th>HLIM</th>
<th>FULL1</th>
<th>HFUL</th>
<th>JIVE</th>
<th>CUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2</td>
<td>−0.001</td>
<td>0.050</td>
<td>0.041</td>
<td>0.078</td>
<td>−0.031</td>
<td>−0.001</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>−0.623</td>
<td>0.094</td>
<td>−0.349</td>
<td>0.113</td>
<td>0.039</td>
<td>0.003</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>−1.871</td>
<td>0.134</td>
<td>−0.937</td>
<td>0.146</td>
<td>0.148</td>
<td>−0.034</td>
</tr>
<tr>
<td>32</td>
<td>2</td>
<td>−0.001</td>
<td>0.011</td>
<td>0.008</td>
<td>0.020</td>
<td>−0.021</td>
<td>−0.001</td>
</tr>
<tr>
<td>32</td>
<td>10</td>
<td>−0.220</td>
<td>0.015</td>
<td>−0.192</td>
<td>0.024</td>
<td>−0.021</td>
<td>0.000</td>
</tr>
<tr>
<td>32</td>
<td>30</td>
<td>−1.038</td>
<td>0.016</td>
<td>−0.846</td>
<td>0.027</td>
<td>−0.016</td>
<td>−0.017</td>
</tr>
</tbody>
</table>

*Results based on 20,000 simulations.*

Table 5. Nine decile range: 0.05 to 0.95 $R^2$ $\varepsilon^2 | z_1 | = 0.$

<table>
<thead>
<tr>
<th>$\mu^2$</th>
<th>$K$</th>
<th>LIML</th>
<th>HLIM</th>
<th>FULL1</th>
<th>HFUL</th>
<th>JIVE</th>
<th>CUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2</td>
<td>2.219</td>
<td>1.868</td>
<td>1.675</td>
<td>1.494</td>
<td>4.381</td>
<td>2.219</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>26.169</td>
<td>5.611</td>
<td>4.776</td>
<td>2.664</td>
<td>7.781</td>
<td>16.218</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>60.512</td>
<td>8.191</td>
<td>7.145</td>
<td>3.332</td>
<td>9.975</td>
<td>15.6412</td>
</tr>
<tr>
<td>32</td>
<td>2</td>
<td>0.941</td>
<td>0.901</td>
<td>0.903</td>
<td>0.868</td>
<td>1.029</td>
<td>0.941</td>
</tr>
<tr>
<td>32</td>
<td>10</td>
<td>3.365</td>
<td>1.226</td>
<td>2.429</td>
<td>1.134</td>
<td>1.206</td>
<td>1.011</td>
</tr>
<tr>
<td>32</td>
<td>30</td>
<td>18.357</td>
<td>1.815</td>
<td>5.424</td>
<td>1.571</td>
<td>1.678</td>
<td>3.563</td>
</tr>
</tbody>
</table>

*Results based on 20,000 simulations.*
Table 6. 0.05 rejection frequencies: $R^2_{x|x_2} = 0.2$.a

<table>
<thead>
<tr>
<th>$\mu^2$</th>
<th>$K$</th>
<th>LIML</th>
<th>HLIM</th>
<th>FULL1</th>
<th>HFUL</th>
<th>JIVE</th>
<th>CUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2</td>
<td>0.097</td>
<td>0.019</td>
<td>0.075</td>
<td>0.023</td>
<td>0.026</td>
<td>0.008</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>0.065</td>
<td>0.037</td>
<td>0.080</td>
<td>0.041</td>
<td>0.036</td>
<td>0.043</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>0.059</td>
<td>0.051</td>
<td>0.118</td>
<td>0.055</td>
<td>0.046</td>
<td>0.094</td>
</tr>
<tr>
<td>32</td>
<td>2</td>
<td>0.177</td>
<td>0.040</td>
<td>0.162</td>
<td>0.040</td>
<td>0.039</td>
<td>0.024</td>
</tr>
<tr>
<td>32</td>
<td>10</td>
<td>0.146</td>
<td>0.042</td>
<td>0.120</td>
<td>0.044</td>
<td>0.033</td>
<td>0.030</td>
</tr>
<tr>
<td>32</td>
<td>30</td>
<td>0.128</td>
<td>0.049</td>
<td>0.107</td>
<td>0.051</td>
<td>0.039</td>
<td>0.073</td>
</tr>
</tbody>
</table>

aResults 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 \( 0 \leq a \leq 1 \); 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 \( P_{ii} \leq C_P (K/n) < 1 \) \( (i = 1, \ldots, n) \). (b) \( \mu_n^2 \sim n^b \) for some real constant \( b \) such that \( a/2 < b \leq 1 \). (c) If \( K \) is fixed, then \( z_i = \pi Z_i \). (d) \( \delta_0 \in D \subset \mathbb{R}^{G} \), where \( D \) is bounded. (e) \( \lambda_{\max}(\tilde{S}^2 \tilde{S}) \) is bounded.

Next, let \( x \lor y = \max(x, y) \) and \( x \land y = \min(x, y) \), and define

\[
\varphi(a, b) = \frac{a \lor (1 - 4\psi_1(a, b))}{\psi_1(a, b)} \mathbb{1}\left\{ \frac{a}{2} < b \leq \frac{1}{2} \right\} + \frac{a}{\psi_2(a, b)} \mathbb{1}\left\{ \frac{1}{2} < b \leq 1 \right\},
\]

where \( \psi_1(a, b) = 2b - a \land b \), \( \psi_2(a, b) = 2b - a \land \frac{1}{2} \), 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

\[
\tilde{X} = \tilde{Y} \Delta + \tilde{V},
\]

where \( \tilde{X} = [y \ X], \Delta = [\delta_0 \ I_G], \) and \( \tilde{V} = [v \ U] \).

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

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

There is \( \tilde{C} > 0 \) such that \( E[\|\tilde{V}_i\|_{p q}^2] \leq \tilde{C} \) and \( \sum_{i=1}^n \|z_i\|_{2pq}^2 / n \leq \tilde{C} \), where \( \tilde{V}_i \) denotes the \( i \)th row of \( \tilde{V} \). (ii) \( \lambda_{\min}(\frac{1}{n} \sum_{i=1}^n E[\tilde{V}_i \tilde{V}_i^\top]) \) is bounded away from zero for \( n \) sufficiently large.

Assumption 8 specifies the moment condition on the error process \( \{\tilde{V}_i\} \) 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 \(\eta\) 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 \(\det\left[ S_{*,n} \right]\), where \(S_{*,n} = X^*_n M X_n/(n - K)\) and \(X_n = [\epsilon \ X]\). 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

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

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, \(\{U_i\} \equiv i.i.d. N(0, \Omega_i)\), where \(U_i\) is the \(i\)th row of \(U = [\epsilon \ U]\), and suppose that there exists a constant \(C > 0\) such that \(\lambda_* = \min_i \lambda_{\min}(\Omega_i) \geq C\). 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 \(C\) 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 \( (W_i, Y_i) (i = 1, \ldots, n) \) are independent, \( W_i \) and \( Y_i \) are scalars, and \( P \) is symmetric, idempotent of rank \( K \), then for \( \bar{w} = E[(W_1, \ldots, W_n)'], \bar{y} = E[(Y_1, \ldots, Y_n)'] \), \( \bar{\sigma}_W = \max_{i \leq n} \text{Var}(W_i)^{1/2}, \bar{\sigma}_Y = \max_{i \leq n} \text{Var}(Y_i)^{1/2} \),

\[
E \left[ \left( \sum_{i \neq j} P_{ij} W_i Y_j - \sum_{i \neq j} P_{ij} \bar{w}_i \bar{y}_j \right)^2 \right] \leq C(K \bar{\sigma}_W^2 \bar{\sigma}_Y^2 + \bar{\sigma}_W^2 \bar{\sigma}_Y^2 + \bar{\sigma}_W^2 \bar{\sigma}_Y^2 \bar{w} \bar{w}').
\]

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

**LEMMA A2.** If Assumptions 1–4 are satisfied and \( \sqrt{K} / \mu_n^2 \longrightarrow 0 \), then \( \sum_{i \neq j} \tilde{X}_i P_{ij} \tilde{X}_j' = \text{diag}(0, H_n) + o_p(1) \).

**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 \( \| S_n^{-1} \| \leq C \mu_n^{-1} \), we have \( \text{Var}(X_{ik}) \leq C \mu_n^{-2} \) for any element \( X_{ik} \) of \( \tilde{X}_i \). Then applying Lemma A1 to each element of the sum of \( X_{ij} P_{ij} X_j' \) gives

\[
\sum_{i \neq j} \tilde{X}_i P_{ij} \tilde{X}_j' = \text{diag} \left( 0, \sum_{i \neq j} z_i P_{ij} z_j' / n \right) + O_p \left( K^{1/2} / \mu_n^2 + \mu_n^{-1} \left( \sum_i \| z_i \|^2 / n \right)^{1/2} \right)
\]

\[
= \text{diag} \left( 0, \sum_{i \neq j} z_i P_{ij} z_j' / n \right) + o_p(1).
\]
Also note that

\[
H_n - \sum_{i \neq j} z_i P_{ij} z'_j / n = \sum_i z_i z'_i / n - \sum_i P_{ii} z_i z'_i / n - \sum_{i \neq j} z_i P_{ij} z'_j / n = z'(I - P)z / n
\]

\[
= (z - Z \pi_{K_n}')(I - P)(z - Z \pi_{K_n}) / n
\]

\[
\leq (z - Z \pi_{K_n}')(z - Z \pi_{K_n}) / n
\]

\[
\leq I_G \sum_i \|z_i - \pi_{K_n} z_i\|^2 / n \rightarrow 0,
\]

where the third equality follows by \(PZ = Z\), the first inequality follows by \(I - P\) idempotent, and the last inequality follows by \(A \leq \text{tr}(A)I\) 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 \(\tilde{Y} = [0, Y]\), \(\tilde{U} = [\varepsilon, U]\), and \(\tilde{X} = [y, X]\), so that \(\tilde{X} = (\tilde{Y} + \tilde{U})D\) for

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

Let \(\tilde{B} = \tilde{X}' \tilde{X} / n\). Note that \(\|S_n/\sqrt{n}\| \leq C\) and, by standard calculations, \(z'U / n \xrightarrow{p} 0\). Then

\[
\|\tilde{Y}' \tilde{U} / n\| = \|(S_n/\sqrt{n})z'U / n\| \leq C\|z'U / n\| \xrightarrow{p} 0.
\]

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

\[
\tilde{B} = (\tilde{U}' \tilde{U} + \tilde{Y}' \tilde{Y} + \tilde{U}' \tilde{Y} + \tilde{Y}' \tilde{Y}) / n = \tilde{\Omega}_n + \tilde{Y}' \tilde{Y} / n + o_p(1) \geq C \text{diag}(I_{G - G_2 + 1}, 0).
\]

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

\[
C \leq (1, -\delta')\tilde{B}(1, -\delta)' = (y - X \delta)'(y - X \delta) / n \leq C\|(1, -\delta')\|^2 = C(1 + \|\delta\|^2).
\]

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

\[
A \overset{\text{def}}{=} \sum_{i \neq j} P_{ij} \tilde{X}_i \tilde{X}_j' \geq C \text{diag}(0, I_G).
\]
Note that $\tilde{S}_n D(1, -\delta)' = (\mu_n, (\delta_0 - \delta)' S_n)'$ and $\tilde{X}_i = D^T \tilde{S}_n \tilde{X}_i$. Then w.p.a.1 for all $\delta$,

$$
\mu_n^{-2} \sum_{i \neq j} P_{ij} (y_i - X_i' \delta)(y_j - X_j' \delta)
= \mu_n^{-2} (1, -\delta') \left( \sum_{i \neq j} P_{ij} \tilde{X}_i \tilde{X}_j' \right) (1, -\delta')'
= \mu_n^{-2} (1, -\delta') D^T \tilde{S}_n \tilde{A} \tilde{S}_n^T D(1, -\delta')' \geq C \| \bar{S}_n' (\delta - \delta_0)/\mu_n \|^2.
$$

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, $\mu_n^{-2} \sum_{i \neq j} e_i P_{ij} e_j \rightarrow 0$. Then w.p.a.1,

$$
|\hat{Q}(\delta_0)| = \left| \mu_n^{-2} \sum_{i \neq j} e_i P_{ij} e_j \right| \left( \sum_{i=1}^n e_i^2 \right) / n \rightarrow 0.
$$

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

$$
0 \leq \| \bar{S}_n' (\hat{\delta} - \delta_0)/\mu_n \|^2 / (1 + \| \hat{\delta} \|^2) \leq C \hat{Q}(\hat{\delta}) \leq C \hat{Q}(\delta_0) \rightarrow 0,
$$

implying $\| \bar{S}_n' (\hat{\delta} - \delta_0)/\mu_n \|^2 / (1 + \| \hat{\delta} \|^2) \rightarrow 0$. Lemma A0 gives the conclusion. \(\square\)

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

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

$$
S_n^{-1} \left( \sum_{i \neq j} X_i P_{ij} \hat{e}_j - \hat{\alpha} X' \hat{e} \right) / \mu_n \rightarrow 0.
$$

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

$$
\hat{\alpha} S_n^{-1} X' X S_n^{-1} V = o_p(\mu_n^2/n) O_p(n/\mu_n^2) \rightarrow 0,
$$

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

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

$$
S_n^{-1} \sum_{i \neq j} X_i P_{ij} \hat{e}_j / \mu_n = o_p(1) - \left( S_n^{-1} \sum_{i \neq j} X_i P_{ij} X_j' S_n^{-1} V \right) S_n' (\hat{\delta} - \delta_0)/\mu_n \rightarrow 0.
$$
L E M M A A 5. If Assumptions 1–4 are satisfied and $S'_n(\hat{\delta} - \delta_0)/\mu_n \xrightarrow{p} 0$, then $\sum_{i \neq j} \hat{\epsilon}_i P_{ij} \hat{\epsilon}_j / \hat{\epsilon}' \hat{\epsilon} = o_p(\mu_n^2/n)$.

P R O O F. Let $\hat{\beta}' = S'_n(\hat{\delta} - \delta_0)/\mu_n$ and $\hat{\alpha} = \sum_{i \neq j} \epsilon_i P_{ij} \epsilon_j / \epsilon' \epsilon = o_p(\mu_n^2/n)$. Note that $\hat{\sigma}_e^2 = \hat{\epsilon}' \hat{\epsilon}/n$ satisfies $1/\hat{\alpha}^2 = O_p(1)$ by the Markov inequality. By Lemma A4 with $\hat{\alpha} = \alpha$, we have $\tilde{H}_n = S_n^{-1}(\sum_{i \neq j} X_i P_{ij} X'_j - \tilde{\alpha} X' X S_n^{-1}) = O_p(1)$ and $W_n = S_n^{-1}(\sum_{i \neq j} X_i P_{ij} \epsilon_j - \tilde{\alpha} X' \epsilon)/\mu_n \xrightarrow{p} 0$, so the conclusion follows by the triangle inequality.

P R O O F O F T H E O R E M 1. First note that if $S'_n(\hat{\delta} - \delta_0)/\mu_n \xrightarrow{p} 0$, then by $\lambda_{\min}(S_n S'_n/\mu_n^2) \geq \lambda_{\min}(SS') > 0$, 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 $\tilde{\alpha} = \tilde{Q}(\tilde{\delta}) = \sum_{i \neq j} \tilde{\epsilon}_i P_{ij} \tilde{\epsilon}_j / \tilde{\epsilon}' \tilde{\epsilon} = o_p(\mu_n^2/n)$ 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 $\tilde{\alpha} = o_p(\mu_n^2/n)$. For any such $\tilde{\alpha}$, by Lemma A4 we have

$$S'_n(\hat{\delta} - \delta_0)/\mu_n$$

$$= S_n\left(\sum_{i \neq j} X_i P_{ij} X'_j - \tilde{\alpha} X' X\right)^{-1} \sum_{i \neq j} (X_i P_{ij} \epsilon_j - \tilde{\alpha} X' \epsilon)/\mu_n$$

$$= \left[S_n^{-1}\left(\sum_{i \neq j} X_i P_{ij} X'_j - \tilde{\alpha} X' X\right)S_n^{-1}\right]^{-1} \sum_{i \neq j} (X_i P_{ij} \epsilon_j - \tilde{\alpha} X' \epsilon)/\mu_n$$

$$= (H_n + o_p(1))^{-1} o_p(1) \xrightarrow{p} 0. \quad \square$$

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

L E M M A A 6 (Lemma A2 of Chao et al. (2012b)). If (i) $P$ is a symmetric, idempotent matrix with rank($P$) = $K$, $P_{ii} \leq C < 1$, (ii) $(W_{1n}, U_1, \epsilon_1), \ldots, (W_{nn}, U_n, \epsilon_n)$ are independent and $D_n = \sum_{i=1}^n E[W_{in} W_{in}^\prime]$ is bounded, (iii) $E[W_{in}^\prime] = 0$, $E[U_i] = 0$, $E[\epsilon_i] = 0$, and there exists a constant $C$ such that $E[\|U_i\|^4] \leq C, E[\epsilon_i^4] \leq C$, (iv) $\sum_{i=1}^n E[\|W_{in}\|^4] \rightarrow 0$, and
(v) $K \to \infty$, then for \( \tilde{S}_n \) defined as $\tilde{S}_n = \sum_{i \neq j} P_{ij}^2 (E[U_i U_j] E[\varepsilon_i^2] + E[U_i \varepsilon_i] E[\varepsilon_j U_j]) / K$ and for any sequence of bounded nonzero vectors $c_{1n}$ and $c_{2n}$ such that $\Xi_n = c_{1n} D_n c_{1n} + c_{2n} \tilde{S}_n c_{2n} > C$, it follows that

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

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

$$\hat{D}(\delta) = \left[ \begin{array}{c} \varepsilon(\delta)^{\prime} \varepsilon(\delta) \\ 2 \end{array} \right] \left[ \begin{array}{c} \sum_{i \neq j} \varepsilon_i \varepsilon_j (\delta) \\ \varepsilon(\delta)^{\prime} \varepsilon(\delta) \end{array} \right] = \sum_{i \neq j} X_i P_{ij} \varepsilon_j (\delta) - \tilde{\alpha}(\delta) X^{\prime} \varepsilon(\delta).$$

A couple of other intermediate results are also useful.

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

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

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

$$\frac{\partial \hat{D}}{\partial \delta}(\delta) = \sum_{i \neq j} X_i P_{ij} X_j^{\prime} - \tilde{\alpha} X^\prime X - \tilde{\gamma} \sum_{i \neq j} \tilde{\varepsilon}_i P_{ij} X_j^{\prime} - \sum_{i \neq j} X_i P_{ij} \tilde{\varepsilon}_j \tilde{\gamma}^{\prime} + 2 (\tilde{\varepsilon}^{\prime} \tilde{\varepsilon}) \tilde{\alpha} \tilde{\gamma} \tilde{\gamma}^{\prime},$$

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

$$S_n^{-1} \left( \sum_{i \neq j} X_i P_{ij} X_j^{\prime} - \tilde{\alpha} X^\prime X \right) S_n^{-1} = H_n + o_p(1), S_n^{-1} \hat{D}(\delta) \tilde{\gamma} S_n^{-1} \tilde{\gamma} S_n^{-1} \xrightarrow{p} 0.$$
Proof. Note that for \( W = z'(P - I) \varepsilon / \sqrt{n} \), by \( I - P \) idempotent and \( E[\varepsilon \varepsilon'] \leq C I_n \), we have
\[
E[WW'] \leq C z'(I - P) z / n = C(z - Z \pi'_{K_n})(I - P)(z - Z \pi'_{K_n}) / n
\]
\[
\leq C I n \sum_{i=1}^{n} \| z_i - \pi_{K_n} z_i \|^2 / n \rightarrow 0,
\]
so \( z'(P - I) \varepsilon / \sqrt{n} = o_p(1) \). Also, by the Markov inequality,
\[
X'\varepsilon / n = \sum_{i=1}^{n} E[X_i \varepsilon_i] / n + O_p(1 / \sqrt{n}), \quad \varepsilon' / n = \sum_{i=1}^{n} \sigma_i^2 / n + O_p(1 / \sqrt{n}).
\]
Also, by Assumption 3, \( \sum_{i=1}^{n} \sigma_i^2 / n \geq C > 0 \). The delta method then gives \( \tilde{\gamma} = X'\varepsilon / \varepsilon' = \gamma_n + O_p(1 / \sqrt{n}) \). Therefore, it follows by Lemma A1 and \( \hat{D}(\delta_0) = \sum_{i \neq j} X_i P_{ij} e_j - \varepsilon' \varepsilon \tilde{\alpha}(\delta_0) \tilde{\gamma} \)
that
\[
S^{-1}_n \hat{D}(\delta_0) = \sum_{i \neq j} z_i P_{ij} e_j / \sqrt{n} + S^{-1}_n \sum_{i \neq j} \hat{U}_i P_{ij} e_i - S^{-1}_n (\tilde{\gamma} - \gamma_n) \varepsilon' \varepsilon \tilde{\alpha}(\delta_0)
\]
\[
= z' P \varepsilon / \sqrt{n} - \sum_{i} P_{ii} z_i \varepsilon_i / \sqrt{n} + S^{-1}_n \sum_{i \neq j} \hat{U}_i P_{ij} e_j
\]
\[
+ O_p(1 / (\sqrt{n} \mu_n)) o_p(\mu_n^2 / n)
\]
\[
= \sum_{i=1}^{n} (1 - P_{ii}) z_i \varepsilon_i / \sqrt{n} + S^{-1}_n \sum_{i \neq j} \check{U}_i P_{ij} e_j + o_p(1).
\]
\( \square \)

Proof of Theorem 2. Consider first the case where \( \hat{\delta} \) is HLIM. Then by Theorem 1, \( \hat{\delta} \overset{p}{\rightarrow} \delta_0 \). First order conditions for LIIML are \( \hat{D}(\delta) = 0 \). Expanding gives
\[
0 = \hat{D}(\delta_0) + \frac{\partial \hat{D}(\delta_0)}{\partial \delta}(\hat{\delta} - \delta_0),
\]
where \( \hat{\delta} \) lies on the line joining \( \hat{\delta} \) and \( \delta_0 \), and hence \( \tilde{\beta} = \mu_n^{-1} S_n^{-1}(\hat{\delta} - \delta_0) \overset{p}{\rightarrow} 0 \). Then by Lemma A7, \( \check{H}_n = S_n^{-1}[\partial \hat{D}(\hat{\delta}) / \partial \delta] S_n^{-1} = H_P + o_p(1) \). Then \( \partial \hat{D}(\hat{\delta}) / \partial \delta \) is nonsingular w.p.a.1 and solving gives
\[
S_n^{-1}(\hat{\delta} - \delta) = -S_n^{-1}[\partial \hat{D}(\hat{\delta}) / \partial \delta]^{-1} \hat{D}(\delta_0) = -\check{H}_n S_n^{-1} \hat{D}(\delta_0).
\]
Next, apply Lemma A6 with \( U_i = \check{U}_i \) and
\[
W_{in} = (1 - P_{ii}) z_i \varepsilon_i / \sqrt{n}.
\]
By \( \varepsilon_i \) having bounded fourth moment and for \( P_{ii} \leq 1 \),
\[
\sum_{i=1}^{n} E[||W_{in}||^4] \leq C \sum_{i=1}^{n} ||z_i||^4 / n^2 \rightarrow 0.
\]
By Assumption 6, we have $\sum_{i=1}^{n} E[W_{in}W_{in}'] \rightarrow \Sigma_P$. Let $\Gamma = \text{diag}(\Sigma_P, \Psi)$ and

$$A_n = \left( \begin{array}{c} \sum_{i=1}^{n} W_{in} \\ \sum_{i \neq j} \hat{U}_i P_{ij} e_j / \sqrt{K} \end{array} \right).$$

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

$$A_n = \left( \begin{array}{c} \sum_{i=1}^{n} W_{in} \\ \sum_{i \neq j} \hat{U}_i P_{ij} e_j / \sqrt{K} \end{array} \right) \xrightarrow{d} N(0, \Gamma), \quad \Gamma = \text{diag}(\Sigma_P, \Psi).$$

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

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

Then by Lemma A8,

$$S_n^{-1} \hat{D}(\delta_0) = F_n A_n + o_p(1) \xrightarrow{d} N(0, \Sigma_P + \alpha S_0 \Psi S_0'),$$

and $S_n'(\hat{\delta} - \delta_0) = - \hat{H}_n^{-1} S_n^{-1} \hat{D}(\delta_0) \xrightarrow{d} N(0, \Lambda_1)$.

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

$$(\mu_n/\sqrt{K}) F_n \rightarrow \tilde{F}_0 = [0, S_0], \quad \tilde{F}_0 \Gamma \tilde{F}_0' = S_0 \Psi S_0'$$

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

$$(\mu_n/\sqrt{K}) S_n^{-1} \hat{D}(\delta_0) = (\mu_n/\sqrt{K}) F_n A_n + o_p(1) \xrightarrow{d} N(0, S_0 \Psi S_0'),$$

and $$(\mu_n/\sqrt{K}) S_n'(\hat{\delta} - \delta_0) = - H_n^{-1}(\mu_n/\sqrt{K}) S_n^{-1} \hat{D}(\delta_0) \xrightarrow{d} N(0, \Lambda_2).$$

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 $\mu_{W_n} = \max_{i \leq n} |E[W_i]|$ and $\mu_{Y_n} = \max_{i \leq n} |E[Y_i]|$.

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

$$\sum_{i \neq j} P_{ij}^2 W_i Y_j = E \left[ \sum_{i \neq j} P_{ij}^2 W_i Y_j \right] + O_p(\sqrt{K}(\tilde{\sigma}_{W_n} \tilde{\sigma}_{Y_n} + \tilde{\sigma}_{W_n} \tilde{\mu}_{Y_n} + \tilde{\mu}_{W_n} \tilde{\sigma}_{Y_n})).$$
LEMMA A10 (Lemma A4 of Chao et al. (2012b)). If $W_i$, $Y_i$, and $\eta_i$, are independent across $i$ with $E[W_i] = a_i/\sqrt n$, $E[Y_i] = b_i/\sqrt n$, $|a_i| \leq C$, $|b_i| \leq C$, $E[\eta_i^2] \leq C$, $\text{Var}(W_i) \leq C\mu_n^{-2}$, and $\text{Var}(Y_i) \leq C\mu_n^{-2}$, there exists $\pi_n$ such that $\max_{i \leq n} |a_i - Z'_i\pi_n| \rightarrow 0$, and $\sqrt K/\mu_n^2 \rightarrow 0$, then

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

Next, recall that $\hat{\epsilon}_i = Y_i - X'_i\hat{\delta}, \tilde{\epsilon}_i = X'_i\hat{\delta}/\hat{\epsilon}', \epsilon_i = X'_i\epsilon/\epsilon'$, and $\gamma_n = \sum_i E[X_i\epsilon_i]/\sum_i \epsilon_i^2$, and let

$$\hat{X}_i = S_n^{-1}(X_i - \gamma_n \epsilon_i), \quad \hat{X}_i = S_n^{-1}(X_i - \gamma_n \epsilon_i),$$

$$\tilde{\epsilon}_i = \sum_{i \neq j \neq k} \hat{X}_i P_{ik} \epsilon^2_j P_{kj} \hat{X}'_j, \quad \tilde{\epsilon}_i = \sum_{i \neq j} P^2_{ij}(\hat{X}_i \hat{X}'_j \epsilon^2_j + \hat{X}_i \hat{\epsilon}_i \hat{\epsilon}'_j \hat{X}'_j),$$

$$\hat{S}_1 = \sum_{i \neq j \neq k} \hat{X}_i P_{ik} \epsilon^2_j P_{kj} \hat{X}'_j, \quad \hat{S}_2 = \sum_{i \neq j} P^2_{ij}(\hat{X}_i \hat{X}'_j \epsilon^2_j + \hat{X}_i \epsilon_i \epsilon'_j \hat{X}'_j).$$

Note that for $\hat{A} = S_n^{-1}(\hat{\delta} - \delta_0)$, we have

$$\hat{\epsilon}_i - \epsilon_i = -X'_i(\hat{\delta} - \delta_0) = -X'_i S_n^{-1} \hat{\delta},$$

$$\hat{\epsilon}'_i - \epsilon'_i = \epsilon_i X'_i(\hat{\delta} - \delta_0) + [X'_i(\hat{\delta} - \delta_0)]^2,$$

$$\hat{X}_i = X_i - S_n^{-1}(\hat{\epsilon}_i - \epsilon_i) = S_n^{-1}(\hat{x}_i - \gamma_n \epsilon_i),$$

$$\hat{X}_i = X_i - S_n^{-1}(\hat{\epsilon}_i - \epsilon_i) = S_n^{-1}(\hat{x}_i - \gamma_n \epsilon_i) = S_n^{-1}(\gamma(\hat{\epsilon}_i - \epsilon_i) - S_n^{-1}(\gamma - \gamma_n) \epsilon_i,$$

$$\hat{X}_i \hat{\epsilon}_i - X_i \epsilon_i = X_i \hat{\epsilon}_i - \gamma \epsilon_i^2 - X_i \epsilon_i + \gamma \epsilon_i^2,$$

$$\|\hat{X}_i \hat{x}'_i - \hat{X}_i \hat{x}'_i\| \leq \|\hat{X}_i - \hat{X}_i\| + \|\hat{X}_i\| \|\hat{X}_i - \hat{x}_i\|.$$

LEMMA A11. If the hypotheses of Theorem 3 are satisfied, then $\hat{S}_2 - \hat{S}_2 = O_p(K/\mu_n^2)$.

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 $d_i = C + |\epsilon_i| + |U_i|$. Note that $\hat{\gamma} - \gamma_n \xrightarrow{p} 0$ by standard arguments. Then for $\hat{A} = (1 + \|\hat{x}\|)(1 + \|\hat{\delta}\|) = O_p(1)$ and $\hat{B} = \|\hat{\gamma} - \gamma_n\| + \|\hat{\delta} - \delta_0\| \xrightarrow{p} 0$, we have

$$\|X_i\| \leq C + |U_i| \leq d_i, \quad |\hat{\epsilon}_i| \leq |X'_i(\delta_0 - \hat{\delta}) + \epsilon_i| \leq C d_i \hat{A},$$

$$\|\hat{X}_i\| = \|S_n^{-1}(X_i - \gamma_n \epsilon_i)\| \leq C \mu_n^{-1} d_i,$$

$$\|\hat{X}_i\| = \|S_n^{-1}(X_i - \gamma_n \epsilon_i)\| \leq C \mu_n^{-1} d_i,$$

$$\|\hat{X}_i \hat{x}'_i - \hat{X}_i \hat{x}'_i\| \leq (\|\hat{X}_i\| + \|\hat{X}_i\|) \|\hat{X}_i - \hat{X}_i\|$$

$$\leq C \mu_n^{-2} d_i \hat{A} \|\hat{x}'_i - \epsilon_i\| + \|\hat{\gamma} - \gamma_n\| \|\epsilon_i\|$$

$$\leq C \mu_n^{-2} d_i \hat{A} \epsilon_i^2,$$
Also note that

\[
|\hat{\epsilon}_i^2 - \epsilon_i^2| \leq (|\epsilon_i| + |\hat{\epsilon}_i|)|\hat{\epsilon}_i - \epsilon_i| \leq C d_i^2 \hat{A}\hat{B},
\]

\[
\|\hat{X}_i \hat{\epsilon}_i - \hat{X}_i \epsilon_i\| = \|S_n^{-1}(X_i \hat{\epsilon}_i - \hat{\gamma}\hat{\epsilon}_i^2 - X_i \epsilon_i + \gamma_n\epsilon_i^2)\|
\leq C \mu_n^{-1}(\|X_i\|\|\hat{\epsilon}_i - \epsilon_i\| + \|\hat{\gamma}\|\|\hat{\epsilon}_i^2 - \epsilon_i^2\| + |\epsilon_i|^2\|\hat{\gamma} - \gamma_n\|)
\leq C \mu_n^{-1} d_i^2(\hat{B} + \hat{A}^2 \hat{B} + \hat{B}) \leq C d_i^2 \hat{A}^2 \hat{B},
\]

\[
\|\hat{X}_i \hat{\epsilon}_i\| \leq C \mu_n^{-1} d_i^2 \hat{A}^2, \quad \|\hat{X}_i \epsilon_i\| \leq C \mu_n^{-1} d_i^2.
\]

Also note that

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

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

\[
\left\|\sum_{i,j} P_{ij}^2(\hat{X}_i \hat{\epsilon}_i \hat{X}_j - \hat{X}_i \epsilon_i \hat{X}_j)\right\|
\leq \sum_{i,j} P_{ij}^2(\|\hat{X}_i \hat{\epsilon}_i\| \|\hat{X}_j \hat{\epsilon}_j\| + \|\hat{X}_j \hat{\epsilon}_j\| \|\hat{X}_i \hat{\epsilon}_i\|) \leq C \mu_n^{-2} \sum_{i,j} P_{ij}^2 d_i^2 d_j^2 (1 + \hat{A}^2)^2 \hat{B} = o_p\left(\frac{K}{\mu_n^2}\right).
\]

We also have

\[
\left\|\sum_{i,j} P_{ij}^2(\hat{X}_i \hat{\epsilon}_i \hat{X}_j - \hat{X}_i \epsilon_i \hat{X}_j)\right\|
\leq \sum_{i,j} P_{ij}^2(\|\hat{X}_i \hat{\epsilon}_i\| \|\hat{X}_j \hat{\epsilon}_j\| + \|\hat{X}_j \hat{\epsilon}_j\| \|\hat{X}_i \hat{\epsilon}_i\|) \leq C \mu_n^{-2} \sum_{i,j} P_{ij}^2 d_i^2 d_j^2 (1 + \hat{A}^2)^2 \hat{B} = o_p\left(\frac{K}{\mu_n^2}\right).
\]

The conclusion then follows by the triangle inequality. \(\square\)

**Lemma A12.** If the hypotheses of Theorem 3 are satisfied, then \(\Sigma_1 - \hat{\Sigma}_1 = o_p(K/\mu_n^2)\).

**Proof.** Note first that

\[
\hat{\epsilon}_i - \epsilon_i = -X_i'(\hat{\theta} - \delta_0) = -X_i'S_n^{-1}'S_n(\hat{\theta} - \delta_0) = -(z_i/\sqrt{n} + S_n^{-1}U_i)'\Delta = -D_i'\hat{\Delta},
\]

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

\[
\hat{\epsilon}_i^2 - \epsilon_i^2 = -2\epsilon_i X_i'(\hat{\theta} - \delta_0) + [X_i'(\hat{\theta} - \delta_0)]^2,
\]

\[
\hat{X}_i - \hat{X}_i = -\hat{\gamma}\hat{\epsilon}_i + \gamma_n\epsilon_i = S_n^{-1}\hat{\gamma}D_i'\hat{\Delta} - S_n^{-1}\mu_n(\hat{\gamma} - \gamma_n)e_i/\mu_n.
\]

We now have \(\hat{\Sigma}_1 - \hat{\Sigma}_1 = \sum_{r=1}^s T_r\), where

\[
T_1 = \sum_{i,j \neq k}(\hat{X}_i - \hat{X}_i)P_{ik}(\hat{\epsilon}_k^2 - \epsilon_k^2)P_{kj}(\hat{X}_j - \hat{X}_j)',
\]
\[
T_2 = \sum_{i \neq j \neq k} \dot{X}_i P_{ik}(\hat{\epsilon}_i^2 - \hat{\epsilon}_k^2)P_{kj}(\dot{X}_j - \dot{X}_j)',
\]
\[
T_3 = \sum_{i \neq j \neq k} (\dot{X}_i - \dot{X}_j)P_{ik} \hat{\epsilon}_i^2 P_{kj}(\dot{X}_j - \dot{X}_j)',
\]
\[
T_4 = T_2,
\]
\[
T_5 = \sum_{i \neq j \neq k} (\dot{X}_i - \dot{X}_j)P_{ik} \hat{\epsilon}_i^2 P_{kj}\dot{X}_j',
\]
\[
T_6 = \sum_{i \neq j \neq k} \dot{X}_i P_{ik}(\hat{\epsilon}_i^2 - \hat{\epsilon}_k^2)P_{kj}\dot{X}_j',
\]
\[
T_7 = T_5'.
\]

From the above expression for \(\hat{\epsilon}_i^2 - \hat{\epsilon}_k^2\), we see that \(T_6\) is a sum of terms of the form \(\hat{B} \sum_{i \neq j \neq k} \dot{X}_i P_{ik} \eta_i P_{kj}\dot{X}_j'\), where \(\hat{B} \to 0\) and \(\eta_i\) is a component either of \(-2\epsilon_i X_i\) or of \(X_i X_i\). By Lemma A10, we have \(\sum_{i \neq j \neq k} \dot{X}_i P_{ik} \eta_i P_{kj}\dot{X}_j' = O_p(1)\), so by the triangle inequality, \(T_6 \to \). Also, note that
\[
T_5 = \sum_{i \neq j \neq k} D_i P_{ik} \hat{\epsilon}_i^2 P_{kj}\dot{X}_j' + S_n^{-1} \mu_n (\hat{\gamma} - \gamma_n) \sum_{i \neq j \neq k} (\epsilon_i / \mu_n) P_{ik} \hat{\epsilon}_i^2 P_{kj}\dot{X}_j'.
\]

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

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. \(\square\)

**Lemma A13.** If the hypotheses of Theorem 3 are satisfied, then
\[
\dot{X}_2 = \sum_{i \neq j} P_{ij}^2 z_i z_j \sigma_i^2 / n + S_n^{-1} \sum_{i \neq j} P_{ij}^2 (E[\bar{U}_i \bar{U}_j] \sigma_j^2 + E[\bar{U}_i \epsilon_i] E[\epsilon_j \bar{U}_j]) S_n^{-1}\]
\[
+ o_p(K / \mu_n^2).
\]

**Proof.** Note that \(\text{Var}(\epsilon_i^2) \leq C \text{ and } \mu_n^2 \leq Cn\), so that for \(u_{ki} = \epsilon_i^2 S_n^{-1} U_i\),
\[
E[(\dot{X}_{ik} \bar{X}_{ik})^2] \leq CE[\dot{X}_{ik}^4 + \bar{X}_{ik}^4] \leq C(z_{ik}^4 / n^2 + E[u_k^4] + z_{ik}^4 / n^2 + E[u_k^4]) \leq C \mu_n^{-4},
\]
\[
E[(\dot{X}_{ik} \epsilon_i)^2] \leq CE[(z_{ik}^2 \epsilon_i^2 / n + u_{ki}^2 \epsilon_i^2)] \leq Cn^{-1} + C \mu_n^{-2} \leq C \mu_n^{-2}.
\]

Also, we have, for \(\tilde{\Omega}_i = E[\bar{U}_i \bar{U}_j]\),
\[
E[\dot{X}_i \bar{X}_i'] = z_i z_i' / n + S_n^{-1} \tilde{\Omega}_i S_n^{-1}, \quad E[\dot{X}_i \epsilon_i] = S_n^{-1} E[\bar{U}_i \epsilon_i].
\]

Next let \(W_i\) be \(\epsilon_j^2 \dot{X}_j' \epsilon_k\) for some \(j\) and \(k\), so that
\[
E[W_i] = \epsilon_j^2 S_n^{-1} E[\bar{U}_i \bar{U}_j] S_n^{-1} \epsilon_k + z_{ij} z_{ik} / n, \quad |E[W_i]| \leq C \mu_n^{-2}.
\]
Also, let $Y_i = \varepsilon_i^2$. Then \( \sqrt{K}(\bar{\sigma}_W^U + \tilde{\sigma}_W^U \mu_1) \leq CK^{1/2}/\mu_n^2 \), so applying Lemma A9 for this $W_i$ and $Y_i$ gives

\[
\sum_{i \neq j} P_{ij}^2 \hat{X}_i \hat{X}'_j = \sum_{i \neq j} P_{ij}^2 (z_i z_i'/n + S_{ij}^{-1} \hat{\Omega}_i S_{ij}^{-1}) \sigma_j^2 + O_p(\sqrt{K}/\mu_n^2).
\]

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

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

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

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

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

**Proof.** Apply Lemma A10 with $W_i$ equal to an element of $\hat{X}_i$, $Y_j$ equal to an element of $\hat{X}_j$, and $\eta_k$ equal to $\varepsilon_k^2$.

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

\[
\sum_{i=1}^n (X_i X'_i - \hat{X}_i P_{ii} \hat{X}'_i - \hat{X}_i P_{ii} \hat{X}'_i) \varepsilon_i^2
\]

\[
= \sum_{i,j,k=1}^n \hat{X}_i P_{ik} \hat{\varepsilon}_k^2 P_{kj} \hat{X}'_j - \sum_{i,j=1}^n \hat{X}_i P_{ii} \hat{\varepsilon}_i P_{ij} \hat{X}'_j - \sum_{i,j=1}^n \hat{X}_i P_{ij} \hat{\varepsilon}_j P_{ij} \hat{X}'_j
\]

\[
= \sum_{i,j,k=1}^n \hat{X}_i P_{ik} \hat{\varepsilon}_k^2 P_{kj} \hat{X}'_j - \sum_{i,j=1}^n \hat{X}_i P_{ii} \hat{\varepsilon}_i P_{ij} \hat{X}'_j
\]

\[
- \sum_{i \neq j} \hat{X}_i P_{ij} \hat{\varepsilon}_j P_{ij} \hat{X}'_j - 2 \sum_{i=1}^n \hat{X}_i P_{ii}^2 \hat{\varepsilon}_i^2 \hat{X}'_i
\]

\[
= \sum_{i,j,k \neq (i,j)} \hat{X}_i P_{ik} \hat{\varepsilon}_k^2 P_{kj} \hat{X}'_j - \sum_{i=1}^n \hat{X}_i P_{ii}^2 \hat{\varepsilon}_i^2 \hat{X}'_i
\]

\[
= \sum_{i \neq j \neq k} \hat{X}_i P_{ik} \hat{\varepsilon}_k^2 P_{kj} \hat{X}'_j + \sum_{i \neq j} P_{ij}^2 \hat{X}_i \hat{\varepsilon}_j^2 - \sum_{i=1}^n \hat{X}_i P_{ii}^2 \hat{\varepsilon}_i^2 \hat{X}'_i.
\]
Also, for \( Z' \) and \( \hat{Z}' \) equal to the \( i \)th row of \( Z \) and \( \hat{Z} \) = \( Z(Z'Z)^{-1} \), we have

\[
\sum_{k=1}^{K} \sum_{\ell=1}^{K} \left( \sum_{i=1}^{n} \hat{Z}_{ik} \hat{Z}_{\ell i} \hat{X}_{\ell j} \hat{e}_{ij} \right) \left( \sum_{j=1}^{n} Z_{jk} Z_{\ell j} \hat{X}_{\ell j} \hat{e}_{ij} \right)'
\]

\[
= \sum_{i,j=1}^{n} \left( \sum_{k=1}^{K} \sum_{\ell=1}^{K} \hat{Z}_{ik} \hat{Z}_{\ell j} \hat{X}_{\ell j} \hat{e}_{ij} \right) \hat{X}_{i} \hat{e}_{ij} \hat{X}_{j}'
\]

\[
= \sum_{i,j=1}^{n} (\hat{Z}' Z_j)^{2} \hat{X}_{i} \hat{e}_{ij} \hat{X}_{j}' = \sum_{i,j=1}^{n} P_{ij} \hat{X}_{i} \hat{e}_{ij} \hat{X}_{j}'.
\]

Adding this equation to the previous one then gives

\[
\hat{S} = \sum_{i \neq j \neq k} \hat{X}_{i} P_{ik} \hat{e}_{k} \hat{P}_{kj} \hat{X}_{j} + \sum_{i \neq j} P_{ij}^{2} \hat{X}_{i} \hat{e}_{ij} \hat{e}_{ij}' - \sum_{i=1}^{n} \hat{X}_{i} P_{ij}^{2} \hat{e}_{ij} \hat{e}_{ij}' + \sum_{i,j=1}^{n} P_{ij}^{2} \hat{X}_{i} \hat{e}_{ij} \hat{X}_{j}'
\]

\[
= \sum_{i \neq j \neq k} \hat{X}_{i} P_{ik} \hat{e}_{k} \hat{P}_{kj} \hat{X}_{j} + \sum_{i \neq j} P_{ij}^{2} (\hat{X}_{i} \hat{X}_{i} \hat{e}_{ij} \hat{e}_{ij} + \hat{X}_{i} \hat{e}_{ij} \hat{X}_{j}').
\]

It then follows that \( S_{n}^{-1} \hat{S} S_{n}^{-1} = \hat{S}_{1} + \hat{S}_{2} \), so that

\[
S_{n}^\prime \hat{S} S_{n} = (S_{n}^{-1} \hat{H} S_{n}^{-1})^{-1} S_{n}^{-1} \hat{S} S_{n}^{-1} (S_{n}^{-1} \hat{H} S_{n}^{-1})^{-1}
\]

\[
= (S_{n}^{-1} \hat{H} S_{n}^{-1})^{-1} (\hat{S}_{1} + \hat{S}_{2}) (S_{n}^{-1} \hat{H} S_{n}^{-1})^{-1}.
\]

By Lemma A4 we have \( S_{n}^{-1} \hat{H} S_{n}^{-1} \xrightarrow{p} HP \). Also, note that for \( \tilde{z}_i = \sum_j P_{ij} z_i = c_i P z \),

\[
\sum_{i \neq j \neq k} z_i P_{ik} \sigma_{k}^{2} P_{kj} z_j' / n
\]

\[
= \sum_{i \neq j \neq k} z_i P_{ik} \sigma_{k}^{2} P_{kj} / n
\]

\[
= \sum_{i \neq j} \left( \sum_{k} z_i P_{ik} \sigma_{k}^{2} P_{kj} z_j' - z_i P_{ii} \sigma_{i}^{2} P_{ij} z_j' - z_i P_{ij} \sigma_{j}^{2} P_{ij} z_j' \right) / n
\]

\[
= \left( \sum_{k} \tilde{z}_k \sigma_{k}^{2} z_k' - \sum_{i,k} \sigma_{k}^{2} z_i z_k' - \sum_{i} z_i P_{ii} \sigma_{i}^{2} z_i' \right) + \sum_{i} \tilde{z}_i \sigma_{i}^{2} P_{ii} z_i' / n
\]

\[
= \sum_{i} (\tilde{z}_i' P_{ii} z_i' - P_{ii} \tilde{z}_i z_i' + P_{ii} \tilde{z}_i z_i') / n
\]

\[
= \sum_{i} \sigma_{i}^{2} (\tilde{z}_i' z_i' + z_i' \tilde{z}_i' + P_{ii} \tilde{z}_i z_i') / n - \sum_{i \neq j} P_{ij} z_i z_j / n.
\]
Also, it follows similarly to the proof of Lemma A8 that \( \sum_i \|z_i - \tilde{z}_i\|^2/n \leq z'(I - P)z/n \to 0 \). Then by \( \sigma_i^2 \) and \( P_{ii} \) bounded, we have

\[
\left\| \sum_i \sigma_i^2 (z_i \tilde{z}_i' - z_i z_i')/n \right\| \\
\leq \sum_i \sigma_i^2 (2\|z_i\|\|z_i - \tilde{z}_i\| + \|z_i - \tilde{z}_i\|^2)/n \\
\leq C \left( \sum_i \|z_i\|^2/n \right)^{1/2} \left( \sum_i \|z_i - \tilde{z}_i\|^2/n \right)^{1/2} + C \sum_i \|z_i - \tilde{z}_i\|^2/n \\
\to 0,
\]

\[
\left\| \sum_i \sigma_i^2 P_{ii}(z_i \tilde{z}_i' - z_i z_i')/n \right\| \leq \left( \sum_i \sigma_i^4 P_{ii}^2 \|z_i\|^2/n \right)^{1/2} \left( \sum_i \|z_i - \tilde{z}_i\|^2/n \right)^{1/2} \to 0.
\]

It follows that

\[
\sum_{i \neq j \neq k} z_i P_{ik} \sigma_k^2 P_{kj} z_j'/n = \sum_i \sigma_i^2 (1 - P_{ii})^2 z_i z_i'/n + o(1) - \sum_{i \neq j} P_{ij}^2 z_i z_i' \sigma_j^2/n \\
= \Sigma_p - \sum_{i \neq j} P_{ij}^2 z_i z_i' \sigma_j^2/n + o(1).
\]

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

\[
\tilde{x}_1 + \tilde{x}_2 = \sum_{i \neq j \neq k} z_i P_{ik} \sigma_k^2 P_{kj} z_j'/n + \sum_{i \neq j} P_{ij}^2 z_i z_i' \sigma_j^2/n \\
+ S_n^{-1} \sum_{i \neq j} P_{ij}^2 (E[\tilde{U}_i \tilde{U}_j'] \sigma_j^2 + E[\tilde{U}_i \varepsilon_i] E[\varepsilon_j \tilde{U}_j']) S_n^{-1} + o_p(1) + o_p(K/\mu_n^2) \\
= \Sigma_p + K S_n^{-1} (\Psi + o(1)) S_n^{-1} + o_p(1) + o_p(K/\mu_n^2) \\
= \Sigma_p + K S_n^{-1} \Psi S_n^{-1} + o_p(1) + o_p(K/\mu_n^2).
\]

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

\[
S_n' \hat{V} S_n = H^{-1}(\Sigma_p + K S_n^{-1} \Psi S_n^{-1}) H^{-1} + o_p(1) = A I + o_p(1).
\]

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

\[
(\mu_n^2/K) S_n' \hat{V} S_n = H^{-1}((\mu_n^2/K) \Sigma_p + \mu_n^2 S_n^{-1} \Psi S_n^{-1}) H^{-1} + o_p(1) = A II + o_p(1).
\]

Next, consider Case I and note that \( S_n' (\hat{\delta} - \delta_0) \overset{d}{\to} Y \sim N(0, A I) \), \( S_n' \hat{V} S_n \overset{p}{\to} A I \), \( c' \mu_n S_n^{-1} \sim c' S_0' \), and \( c' \mu_n S_n^{-1} \sim c' S_0' \). Then by the continuous mapping and Slutsky theorems,

\[
\frac{c' (\hat{\delta} - \delta_0)}{\sqrt{c' \hat{V} c}} \overset{d}{\to} \frac{c' \mu_n S_n^{-1} S_n' (\hat{\delta} - \delta_0)}{\sqrt{c' \mu_n S_n^{-1} S_n' \hat{V} S_n S_n^{-1} \mu_n c}} \overset{d}{\to} \frac{c' S_0' Y}{\sqrt{c' S_0' A I S_0 c}} \sim N(0, 1).
\]
For Case II, \((\mu_n/\sqrt{K})S_n' (\hat{\delta} - \delta_0) \xrightarrow{d} \tilde{Y} \sim N(0, \Lambda_{II})\) and \((\mu_n^2/K)S_n' \hat{Y} S_n \xrightarrow{p} \Lambda_{II}\). Then by \(c' S_0^\prime A_{II} S_0 c \neq 0\),

\[
\frac{c' (\hat{\delta} - \delta_0)}{\sqrt{c' \hat{V} c}} = \frac{c' S_n^{-1} (\mu_n/\sqrt{K})S_n' (\hat{\delta} - \delta_0)}{\sqrt{c' S_n^{-1} (\mu_n^2/K)S_n' \hat{Y} S_n S_n^{-1} c}} = \frac{c' \mu_n S_n^{-1} (\mu_n/\sqrt{K})S_n' (\hat{\delta} - \delta_0)}{\sqrt{c' \mu_n S_n^{-1} (\mu_n^2/K)S_n' \hat{V} S_n S_n^{-1} \mu_n c}} \xrightarrow{d} \frac{c' S_0' \tilde{Y}}{\sqrt{c' S_0' A_{II} S_0 c}} \sim N(0, 1) \quad \square
\]

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

\[
\mu_n S_n^{-1} = \text{diag}(\mu_n/\sqrt{n}, \ldots, \mu_n/\sqrt{n}, 1) \tilde{S}^{-1}, \quad \tilde{S}^{-1} = \begin{bmatrix} I & 0 \\ -\pi_1 & 1 \end{bmatrix}.
\]

By Assumption 2, \(\gamma = \lim_{n \to \infty} (\mu_n^2/n)\) exists and is either 0 or 1. By \(\sigma_i^2 \geq C > 0\), \(H_p^{-1} \Sigma_p H_p^{-1}\) is positive definite (p.d.). Also \(\Psi = \lim_{n \to \infty} E[(\sum_{i \neq j} P_{ij} \tilde{U}_{ij}) (\sum_{i \neq j} P_{ij} \tilde{U}_{ij} \times \epsilon_j)/K] \) is positive semidefinite, implying \(A_1\) is p.d.

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

Hence

\[
c' S_0' A_1 S_0 c \neq 0
\]

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 \(S_0 = \text{diag}(0, \ldots, 0, 1) \tilde{S}^{-1} = e_G b'\) for the \(G\)th unit vector \(e_G\).

Then under (b), where \(K/\mu_n^2\) is bounded,

\[
S_0 c = e_G b' c \neq 0.
\]

Therefore, \(c' S_0' A_1 S_0 c \neq 0\) and the conclusion follows from Theorem 3.

Finally, if (c) holds and \(\mu_n^2/n \to 0\), note that because \(E[\epsilon_i \tilde{U}_{ij}]\) has the same sign for all \(i\), that \(E[\epsilon_i \tilde{U}_{ij}] E[\epsilon_j \tilde{U}_{ij}] \geq 0\) for all \(i\) and \(j\), and hence

\[
\sum_{i \neq j} P_{ij}^2 (\sigma_i^2 E[\tilde{U}_{ij}^2] + E[\epsilon_i \tilde{U}_{ij}] E[\epsilon_j \tilde{U}_{ij}])/K \geq C \sum_{i \neq j} P_{ij}^2 /K \geq C \left(1 - \sum_i P_{ii}^2 /K\right) \geq C.
\]

Then \(\Psi \geq C e_G c' e_G\) in the positive semidefinite sense, and hence by \(b' e_G = 1\),

\[
A_{II} \geq C H_p^{-1} S_0 e_G c' e_G S_0 H_p^{-1} = C H_p^{-1} e_G b' e_G c' e_G b' e_G H_p^{-1} = C H_p^{-1} e_G c' e_G H_p^{-1}.
\]

It then follows that

\[
c' S_0' A_{II} S_0 c = c' b' e_G A_{II} e_G b' c \geq C (c'b)^2 (e_G H_p^{-1} e_G)^2 \neq 0,
\]
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 \( \rho \), there exists a constant \( \overline{C} \) such that

\[
E \left[ \frac{1}{n-K} \right]^{-\rho} \leq \overline{C} < \infty.
\]

To proceed, note that it is convenient here to change variables in the following way: define

\[
H_Z = \begin{pmatrix} Z_1 (Z'Z)^{-1/2} \\ Z_2 (Z'Z)^{-1/2} \end{pmatrix} = \begin{pmatrix} H_{Z,1} \\ H_{Z,2} \end{pmatrix}
\]

(say).

Now define

\[
H_{Z}^\perp = \begin{pmatrix} (Z_1')^{-1}Z_2' \\ I_{n-K} \end{pmatrix} [I_{n-K} + H_{Z,2}(H_{Z,1}'H_{Z,1})^{-1}H_{Z,2}']^{-1/2} = \begin{pmatrix} -(Z_1')^{-1}Z_2' \\ I_{n-K} \end{pmatrix} [I_{n-K} + Z_2(Z_1')^{-1}Z_2']^{-1/2} \in V_{n-K,n}.
\]

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,

\[
P = H_Z H_Z', \quad M = H_Z^\perp H_Z'^\perp, \quad \text{and} \quad (H_Z \ H_Z^\perp) \in O(n),
\]

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

\[
W_n = \frac{1}{\sqrt{n-K}} H_{Z}^\perp [\varepsilon X] = \frac{1}{\sqrt{n-K}} H_{Z}^\perp X,
\]

so that

\[
W_n' W_n = \frac{X_1' H_{Z}^\perp H_{Z}' X}{n-K} = \frac{X_1' M X}{n-K}.
\]

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

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

where

\[
\Xi_W = (I_L \otimes H_{Z}^\perp) K_{nl} D_{nl} K_{nl} (I_L \otimes H_{Z}^\perp),
\]
\[
D_{\psi} = \begin{pmatrix}
\Omega_{1}^* & 0 & \cdots & 0 \\
0 & \Omega_{2}^* & \cdots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & \Omega_{n}^*
\end{pmatrix}, \quad \Omega_{i}^* = E[U_i'U_i'] \quad \text{for } i = 1, \ldots, n,
\]

\[
\varphi_n = \text{vec}(\Phi_n), \quad \Phi_n = \begin{pmatrix} 0_{(n-K)\times 1} \frac{1}{\sqrt{n(n-K)}} H_Z^{1/2} D_{\psi}^{1/2} \tilde{\Sigma}_n \end{pmatrix},
\]

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

\[
\lambda_{\min}(\Xi W) = (I_L \otimes H_Z^{1/2}) K'_{nL} K_{nL} (I_L \otimes H_Z^{1/2}) \lambda_{\min}(D_{\psi})
\]

\[
= \lambda_{\min}(D_{\psi}) \quad \text{(since } H_Z^{1/2} H_Z^{1/2} = I_{n-K} \text{ and } K'_{nL} K_{nL} = I_{nL})
\]

\[
= \min_i \lambda_{\min}(\Omega_{i}^*)
\]

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

\[
f_n(W_n)
\]

\[
\leq \lambda_*^{-(n-K)L/2} (2\pi)^{-(n-K)L/2} (n-K)^{(n-K)L/2}
\]

\[
\times \exp \left\{ -\frac{(n-K)}{2\lambda_*} (\text{vec}(W_n)' - \varphi_n)'(\text{vec}(W_n) - \varphi_n) \right\}
\]

\[
\leq \lambda_*^{-(n-K)L/2} (2\pi)^{-(n-K)L/2} (n-K)^{(n-K)L/2}
\]

\[
\times \exp \left\{ -\frac{(n-K)}{2\lambda_*} \text{tr}(W_n - \Phi_n)'(W_n - \Phi_n) \right\}
\]

\[
= f_n^*(W) \quad \text{(say)}.
\]

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 \(\rho > 0\),

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

Write

\[
f_n^*(W_n) = \lambda_*^{-(n-K)L/2} (2\pi)^{-(n-K)L/2} (n-K)^{(n-K)L/2}
\]

\[
\times \exp \left\{ -\frac{(n-K)}{2\lambda_*} \text{tr}(W_n - \Phi_n)'(W_n - \Phi_n) \right\}
\]

\[
= \lambda_*^{-(n-K)L/2} (2\pi)^{-(n-K)L/2} (n-K)^{(n-K)L/2}
\]

\[
\times \exp \left\{ -\frac{(n-K)}{2\lambda_*} \text{tr}(W_n'W_n) \right\} \exp \left\{ -\frac{(n-K)}{2\lambda_*} \text{tr}(\Phi_n'\Phi_n) \right\}
\]

\[
\times \exp \left\{ \frac{(n-K)}{\lambda_*} \text{tr}(\Phi_n'W_n) \right\}.
\]
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 \( W_n \rightarrow (Q_1, T_n) \) as

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

where \( t_{gg} \) is the \( g \)th diagonal element of \( T_n \), so that

\[
\int_{\mathbb{R}^{(n-K)L}} \left[ \det(W_n'W_n) \right]^{-\rho} f_n^*(W_n)(dW_n) = \lambda_*^{-\frac{(n-K)L}{2}} (2\pi)^{-\frac{(n-K)L}{2}(n-K)} \left[ \det(T_n'T_n) \right]^{-\rho} \prod_{g=1}^{L} t_{gg}^{(n-K-g)} (dT_n)
\]

\[
= \lambda_*^{-\frac{(n-K)L}{2}} (2\pi)^{-\frac{(n-K)L}{2}(n-K)} \left[ \det(T_n'T_n) \right]^{-\rho} \prod_{g=1}^{L} t_{gg}^{(n-K-g)} \exp\left\{ -\frac{(n-K)}{2\lambda_*} \text{tr}[\Phi_n'\Phi_n] \right\} 
\]

\[
= \lambda_*^{-\frac{(n-K)L}{2}} (2\pi)^{-\frac{(n-K)L}{2}(n-K)} \left[ \det(T_n'T_n) \right]^{-\rho} \prod_{g=1}^{L} t_{gg}^{(n-K-g)} \exp\left\{ -\frac{(n-K)}{2\lambda_*} \text{tr}[\Phi_n'\Phi_n] \right\} \prod_{i=1}^{L} \prod_{j=1}^{L} \exp\left\{ \frac{(n-K)}{\lambda_*} \text{tr}[\Phi_nQ_1T_n] \right\} (dQ_1)
\]

\[
= \lambda_*^{-\frac{(n-K)L}{2}} (2\pi)^{-\frac{(n-K)L}{2}(n-K)} \left[ \det(T_n'T_n) \right]^{-\rho} \prod_{g=1}^{L} t_{gg}^{(n-K-g)} \exp\left\{ -\frac{(n-K)}{2\lambda_*} \text{tr}[\Phi_n'\Phi_n] \right\} \prod_{i=1}^{L} \prod_{j=1}^{L} \exp\left\{ \frac{(n-K)}{\lambda_*} \text{tr}[\Phi_nQ_1T_n] \right\} (dQ_1)
\]

\[
= \lambda_*^{-\frac{(n-K)L}{2}} (2\pi)^{-\frac{(n-K)L}{2}(n-K)} \left[ \det(T_n'T_n) \right]^{-\rho} \prod_{g=1}^{L} t_{gg}^{(n-K-g)} \exp\left\{ -\frac{(n-K)}{2\lambda_*} \text{tr}[\Phi_n'\Phi_n] \right\} \prod_{i=1}^{L} \prod_{j=1}^{L} \exp\left\{ \frac{(n-K)}{\lambda_*} \text{tr}[\Phi_nQ_1T_n] \right\} (dQ_1)
\]

\[
= \lambda_*^{-\frac{(n-K)L}{2}} (2\pi)^{-\frac{(n-K)L}{2}(n-K)} \left[ \det(T_n'T_n) \right]^{-\rho} \prod_{g=1}^{L} t_{gg}^{(n-K-g)} \exp\left\{ -\frac{(n-K)}{2\lambda_*} \text{tr}[\Phi_n'\Phi_n] \right\} \prod_{i=1}^{L} \prod_{j=1}^{L} \exp\left\{ \frac{(n-K)}{\lambda_*} \text{tr}[\Phi_nQ_1T_n] \right\} (dQ_1)
\]

\[
= \lambda_*^{-\frac{(n-K)L}{2}} (2\pi)^{-\frac{(n-K)L}{2}(n-K)} \left[ \det(T_n'T_n) \right]^{-\rho} \prod_{g=1}^{L} t_{gg}^{(n-K-g)} \exp\left\{ -\frac{(n-K)}{2\lambda_*} \text{tr}[\Phi_n'\Phi_n] \right\} \prod_{i=1}^{L} \prod_{j=1}^{L} \exp\left\{ \frac{(n-K)}{\lambda_*} \text{tr}[\Phi_nQ_1T_n] \right\} (dQ_1)
\]

\[
= \lambda_*^{-\frac{(n-K)L}{2}} (2\pi)^{-\frac{(n-K)L}{2}(n-K)} \left[ \det(T_n'T_n) \right]^{-\rho} \prod_{g=1}^{L} t_{gg}^{(n-K-g)} \exp\left\{ -\frac{(n-K)}{2\lambda_*} \text{tr}[\Phi_n'\Phi_n] \right\} \prod_{i=1}^{L} \prod_{j=1}^{L} \exp\left\{ \frac{(n-K)}{\lambda_*} \text{tr}[\Phi_nQ_1T_n] \right\} (dQ_1)
\]
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

\[
\int_{Q_1 \in VL_{n\times K}} \exp\left\{ \frac{(n-K)}{\lambda_s} \text{tr}[\Phi_n Q_1 T_n] \right\} (dQ_1)
\]

\[
= \frac{I_{n-K-L}}{2(n-K-L)\pi^{(n-K-L)/2}} \times \int_{Q_1 \in VL_{n\times K}} \int_{H \in O(n-K-L)} \exp\left\{ \frac{(n-K)}{\lambda_s} \text{tr}[\Phi_n Q_1 T_n] \right\} (dH)(dQ_1)
\]

\[
= \frac{2L \pi^{(n-K)L/2}}{I_L \left[ \frac{1}{2} (n-K) \right]} \int_{Q \in O(n-K)} \exp\left\{ \frac{(n-K)}{\lambda_s} \text{tr}[\Phi_n Q_1 T_n] \right\} (dQ)
\]

\[
= \frac{2L \pi^{(n-K)L/2}}{I_L \left[ \frac{1}{2} (n-K) \right]} {}_0 F_1\left( \frac{1}{2} (n-K); \frac{1}{4} \left( \frac{n-K}{\lambda_s} \right)^2 T_n \Phi_n' \Phi_n T_n \right)
\]

where \(R_n = T_n T_n = W_n' W_n\) and where \(_0 F_1\left( \frac{1}{2} (n-K); \frac{1}{4} \left( \frac{n-K}{\lambda_s} \right)^2 \Phi_n' \Phi_n R_n \right)\) 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

\[
(dR_n) = 2^L \prod_{g=1}^{L} t_g^{(L+1-g)} (dT_n),
\]

so that the Jacobian of the transformation \(T_n \rightarrow R_n\) is given by

\[
(dT_n) = 2^{-L} \prod_{g=1}^{L} t_g^{(-L+1+g)} (dR_n).
\]

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

\[
\int_{[L(n-K)]} [\text{det}(W_n' W_n)]^{-\rho} f_n^\rho(W_n) (dW_n)
\]

\[
= 2^{-L} \lambda_s^{- (n-K)L/2} (2\pi)^{- (n-K)L/2} (n-K)^{(n-K)L/2} \times \exp\left\{ \frac{(n-K)}{2\lambda_s} \text{tr}[\Phi_n' \Phi_n] \right\} \frac{2L \pi^{(n-K)L/2}}{I_L \left[ \frac{1}{2} (n-K) \right]}
\]
\[
\times \int_{R_n > 0} F_1 \left( \frac{1}{2} (n - K), \frac{1}{4} \left( \frac{n - K}{\lambda_*} \right)^2 \Phi_n' \Phi_n R_n \right)
\times (\det R_n)^{(n - K - 2\rho - L - 1)/2} \exp \left\{ - \frac{(n - K)}{2 \lambda_*} \text{tr}(R_n) \right\} (dR_n)
\]

(using the fact that \( \prod_{g=1}^L g = (\det T_n^t T_n)^{1/2} = (\det R_n)^{1/2} \))

\[= \lambda_*^{-(n - K)/2} (2\pi)^{-(n - K)/2} (n - K)^{(n - K)/2} \]

\[\times \exp \left\{ - \frac{(n - K)}{2 \lambda_*} \text{tr}(R_n) \right\} (\det R_n)^{(n - K - 2\rho - L - 1)/2}
\]

\[\times F_1 \left( \frac{1}{2} (n - K), \frac{1}{4} \left( \frac{n - K}{\lambda_*} \right)^2 \Phi_n' \Phi_n R_n \right) (dR_n), \]

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

\[0F_1 \left( \frac{1}{2} (n - K), (n - K)^2 A_n R_n \right) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{1}{((n - K)/2)_\kappa} C_\kappa((n - K)^2 A_n R_n) \]

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 \( \kappa \), 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 \( \left( \frac{n - K - g + 1}{2} \right)_{k_g} \) being Pochhammer's symbol or forward factorial, so that

\[\left( \frac{n - K - g + 1}{2} \right)_{k_g} = \left( \frac{n - K - g + 1}{2} \right) \left( \frac{n - K - g + 2}{2} \right) \times \cdots \times \left( \frac{n - K - g + 2k_g - 1}{2} \right)\]
\[
\int_{R_n > 0} \mathbf{0} F_1 \left( \frac{1}{2} (n - K); (n - K)^2 A_n R_n \right) \left( \det R_n \right)^{(n - K - 2\rho - L - 1)/2} \times \exp \left\{ -\frac{(n - K)^2}{2\lambda_s} \right\} (dR_n)
\]
\[
= \sum_{k=0}^{\infty} \sum_{\kappa} \frac{1}{((n - K)/2)_\kappa} \times \int_{R_n > 0} \frac{C\kappa((n - K)^2 A_n R_n)}{k!} \left( \det \left\{ \frac{(n - K)}{2\lambda_s}, I_L \right\} \right)^{-(n - K - 2\rho)/2} \times \exp \left\{ -\frac{(n - K)^2}{2\lambda_s} \right\} (dR_n)
\]
\[
= I_L \left( \frac{n - K - 2\rho}{2} \right) \left( \det \left\{ \frac{(n - K)^2}{2\lambda_s}, I_L \right\} \right)^{-(n - K - 2\rho)/2} \times \mathbf{1} F_1 \left( \frac{n - K - 2\rho}{2}; \frac{n - K}{2}; \left( \frac{n - K}{2\lambda_s} \right) \Phi' \Phi_n \right).
\]

Next, we analyze the asymptotic behavior of the hypergeometric function \( \mathbf{1} F_1 \left( \frac{n - K - 2\rho}{2}; \frac{n - K}{2}; \left( \frac{n - K}{2\lambda_s} \right) \Phi' \Phi_n \right) \). Using (generalized) Kummer's relation, we can write

\[
\mathbf{1} F_1 \left( \frac{n - K - 2\rho}{2}; \frac{n - K}{2}; \left( \frac{n - K}{2\lambda_s} \right) \Phi' \Phi_n \right) = \exp \left( \text{tr} \left[ \left( \frac{n - K}{2\lambda_s} \right) \Phi' \Phi_n \right] \right) \mathbf{1} F_1 \left( \rho; \frac{n - K}{2}; -\left( \frac{n - K}{2\lambda_s} \right) \Phi' \Phi_n \right).
\]

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

\[
\mathbf{1} F_1 \left( \rho; \frac{n - K}{2}; -\left( \frac{n - K}{2\lambda_s} \right) \Phi' \Phi_n \right) = \left[ \det \left( I_L + \frac{\Phi' \Phi_n}{\lambda_s} \right) \right]^{-\rho} (1 + O(n^{-1})) = O(1).
\]
It follows that

\[
\int_{R^n>0} 0F_1 \left( \frac{1}{2} (n-K); (n-K)^2 A \right) (\det R_n)^{(n-K-2\rho-L-1)/2} \times \exp \left\{ -\frac{(n-K)}{2\lambda_*} \operatorname{tr}(R_n) \right\} (dR_n) \\
= I_L \left( \frac{n-K-2\rho}{2} \right) (\det R_n)^{(n-K-2\rho)L/2} \times \left[ \det \left( I_L + \frac{\Phi'_n \Phi_n}{\lambda_*} \right) \right]^{-\rho} \left( 1 + O(n^{-1}) \right).
\]

Substituting (12) into (10), we have

\[
\int_{R^{(n-K)L}} [\det(W_n'W_n)]^{-\rho} f^*_n(W_n)(dW_n) \\
= \lambda_*^{-(n-K)L/2} (2\pi)^{-(n-K)L/2} (n-K)^{(n-K)L/2} \times \exp \left\{ -\frac{(n-K)}{2\lambda_*} \operatorname{tr}[\Phi'_n \Phi_n] \right\} \frac{\pi^{(n-K)L/2}}{I_L \left[ \frac{1}{2} (n-K) \right]} \\
\times I_L \left( \frac{n-K-2\rho}{2} \right) (\det R_n)^{(n-K-2\rho)L/2} \times \left[ \det \left( I_L + \frac{\Phi'_n \Phi_n}{\lambda_*} \right) \right]^{-\rho} \left( 1 + O(n^{-1}) \right) \\
= \frac{I_L \left[ \frac{1}{2} (n-K-2\rho) \right]}{I_L \left[ \frac{1}{2} (n-K) \right]} (n-K)^{\rho L}(2\lambda_*)^{-\rho L} \left[ \det \left( I_L + \frac{\Phi'_n \Phi_n}{\lambda_*} \right) \right]^{-\rho} \left( 1 + O(n^{-1}) \right).
\]

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

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

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

\[
\Gamma \left[ \frac{n-K-2\rho-g+1}{2} \right] \\
= \left( \frac{4\pi}{n-K-2\rho-g+1} \right)^{1/2} \left( \frac{n-K-2\rho-g+1}{2e} \right)^{(n-K-2\rho-g+1)/2} \left( 1 + O(n^{-1}) \right) \\
= (4\pi)^{1/2}(2e)^{-(n-K-2\rho-g+1)/2}(n-K-2\rho-g+1)^{(n-K-2\rho-g)/2} \left( 1 + O(n^{-1}) \right) 
\]
Applying (14) and (15) to (13), we obtain

\[
\int \mathbb{E}(n_K-L, \mathbf{W}_n) \frac{1}{\sqrt{n \cdot \lambda}} \mathbf{W}_n \mathbf{W}_n^T \exp \left( - \frac{\mathbf{W}_n^T \mathbf{W}_n}{2} \right) d\mathbf{W}_n \\
= (2\pi)^{L(L+1)/4} (n-K)^{L(L+1)/2} e^{-(n-K)/2} (1 + O(n^{-1}))
\]

so that

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

Similarly, we have

\[
I_L \left[ \frac{1}{2} (n-K) \right] = \pi^{L(L-1)/4} \prod_{g=1}^{L} I \left[ \frac{n-K-g+1}{2} \right]
= (2\pi)^{L(L+1)/4} (n-K)^{L(L+1)/2} e^{-(n-K)/2} (1 + O(n^{-1}))
\]

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

\[
\int \mathbb{E}(n_K-L, \mathbf{W}_n) \frac{1}{\sqrt{n \cdot \lambda}} \mathbf{W}_n \mathbf{W}_n^T \exp \left( - \frac{\mathbf{W}_n^T \mathbf{W}_n}{2} \right) d\mathbf{W}_n \\
= \frac{I_L \left[ \frac{1}{2} (n-K-2\rho) \right]}{I_L \left[ \frac{1}{2} (n-K) \right]} \rho^L (2\lambda)^{-\rho L} \left[ \det \left( I_L + \frac{\Phi'_n \Phi_n}{2\lambda} \right) \right]^{\rho} (1 + O(n^{-1}))
\]

\[
= (2\pi)^{L(L+1)/4} (n-K-2\rho)^{L(L+1)/2} e^{-(n-K)/2} (1 + O(n^{-1}))
\]

\[
\times (n-K)^{L[L-n-K-(L+1)/2]/2} (2\pi)^{L(L+1)/4} (n-K-1)^{L/2}
\]
\[ \times e^{(n-K)2(n-K)(n-K-(L+1)/2)}(n-K)^{\rho L} \]
\[ \times (2\lambda_*)^{-\rho L} \left[ \det \left( I_L + \frac{\Phi_n' \Phi_n}{2\lambda_*} \right) \right]^{-\rho} (1 + O(n^{-1})) \]
\[ = \lambda_*^{-\rho L} \left[ \det \left( I_L + \frac{\Phi_n' \Phi_n}{2\lambda_*} \right) \right]^{-\rho} (1 + O(n^{-1})) \]
\[ \leq \lambda_*^{-\rho L} (1 + O(n^{-1})) \]
\[ = O(1). \]

References


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]


