IV Estimation with Heteroskedasticity and Many Instruments

Jerry A. Hausman, Whitney K. Newey, and Tiemen M. Woutersen*

November, 2006

Abstract. In an effort to improve precision, applications often use many instrumental variables. The Fuller (1977, FULL) estimator has good properties under homoskedasticity but it and the limited information maximum likelihood estimator (LIML) are inconsistent with heteroskedasticity and many instruments. This paper proposes a jackknife version of FULL and LIML that are robust to heteroskedasticity and many instruments. Heteroskedasticity consistent standard errors are given. We also give an IV estimator that is efficient under heteroskedasticity. We find that RFLL performs nearly as well as FULL under homoskedasticity.

Keywords: Instrumental variables, heteroskedasticity, many instruments.

*MIT, MIT and Johns Hopkins University

1. Introduction

In an effort to improve precision, applications often use many instrumental variables. Under homoskedasticity inference methods have been developed that account well for use of many instruments. The Fuller (1977, FULL) estimator has good efficiency properties and good accurate inference methods exist. However, with heteroskedasticity it turns out that FULL and the limited information maximum likelihood (LIML) estimators under many instruments, as shown by Bekker and Van der Ploeg (2005) for LIML. Heteroskedasticity seems common in applications, suggesting the need for estimators that are robust to heteroskedasticity but retain good efficiency.

This paper proposes an adjustment to FULL and LIML that make them robust to heteroskedasticity and many instruments, RFLL and RLML respectively. RLML is a jackknife version of LIML that deletes own observation terms in the numerator of the variance ratio. It can be interpreted as an optimal linear combination of forward and reverse jackknife instrumental variable (JIV) estimators, analogous to Hahn and Hausman's (2002) interpretation of LIML as an optimal linear combination of forward and reverse two-stage least squares. We show that these estimators are as efficient as FULL and LIML under homoskedasticity and the many weak instruments sequence of Chao and Swanson (2005) but have a different limiting distribution under the many instrument sequence of Kunitomo (1980) and Bekker (1994). We also give heteroskedasticity consistent standard errors under many weak instruments. We find in Monte Carlo experiments that this estimator and its associated standard errors perform well in a variety of settings.

The RFLL and RLML estimators are not generally efficient under heteroskedasticity and many weak instruments. We give a linear IV estimator that is efficient, that uses a heteroskedasticity consistent weighting matrix and projection residual in the linear combination of instrumental variables. This estimator is as efficient as the continuous updating estimator of Hansen, Heaton, and Yaron (1996), that was shown to be efficient relative to JIV by Newey and Windmeijer (2005), under many weak instruments. Thus we provide a complete solution to the problem of IV estimation in linear models with heteroskedasticity, including a simple, robust estimator, and an efficient estimator that is relatively simple to compute.

We also propose a Hausman (1978) specification test for heteroskedasticity and many instruments. This test is based on comparing FULL (or LIML) with few instruments with FULL with many instruments. The variance estimator associated with this test is the difference of the variance for the few and many instrument estimators plus the many instrument correction.

In comparison with previous work, Hahn and Hausman (2002) had previously considered combining forward and reverse IV estimators. JIV estimators were proposed by Phillips and Hale (1977), Blomquist and Dahlberg (1999), and Angrist and Imbens and Krueger (1999), and Ackerberg and Deveraux (2003). The inconsistency of LIML and bias corrected two stage least squares has been pointed out by Bekker and van der Ploeg (2005), and also by Ackerberg and Deveraux (2003) and Chao and Swanson (2003). Here we add focusing on FULL and by giving a precise characterization of the inconsistency. Chao and Swanson (2004) have previously given heteroskedasticity consistent standard errors and shown asymptotic normality for JIV under many weak instruments. Newey and Windmeijer (2005) have shown that with heteroskedasticity the generalized empirical likelihood estimators are efficient. Our heteroskedasticity efficient estimator is much simpler.

In Monte Carlo results we show that optimally combining forward and reverse JIV overcomes the problems with JIV pointed out by Davidson and MacKinnon (2005). The median bias and dispersion of RFLL is nearly that of FULL except in very weakly identified cases. These results suggest that the estimator is a promising heteroskedasticity consistent and efficient alternative to FULL, LIML, and other estimators with many instruments.

2. The Model and Estimators

The model we consider is given by

$$\begin{aligned} \underset{n\times 1}{y} &= \underset{n\times G}{X}\,\underset{G\times 1}{\delta_0} + \underset{n\times 1}{u}, \\ X &= \Upsilon + V, \end{aligned}$$

where n is the number of observations, G the number of right-hand side variables, Υ is a matrix of observations on the reduced form, and V is the matrix of reduced form disturbances. For the asymptotic approximations, the elements of Υ will be implicitly allowed to depend on n, although we suppress dependence of Υ on n for notational convenience. Estimation of δ_0 will be based on a $n \times K$ matrix Z of instrumental variable observations. We will assume that $Z_1, ..., Z_n$ are nonrandom and that observations (u_i, V_i) are independent of each other.

This model allows for Υ to be a linear combination of Z, i.e. $\Upsilon = Z\pi$ for some $K \times G$ matrix π. Furthermore, columns of X may be exogenous, with the corresponding column of V being zero. The model also allows for Z to be functions meant to approximate the reduced form. For example, let X_i, Υ_i, and Z_i denote the i^{th} row (observation) for X, Υ, and Z respectively. We could have $\Upsilon_i = f_0(w_i)$ be an unknown function of a vector w_i of underlying instruments and $Z_i = (p_{1K}(w_i), ..., p_{KK}(w_i))'$ for 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 Donald and Newey (2001).

To describe the estimators let $P = Z(Z'Z)^{-1}Z'$. The LIML estimator $\tilde{\delta}$ is given by

$$\tilde{\delta} = \arg\min_{\delta} \tilde{Q}(\delta), \tilde{Q}(\delta) = \frac{(y - X\delta)'P(y - X\delta)}{(y - X\delta)'(y - X\delta)}.$$

FULL is obtained as

$$\breve{\delta} = (X'PX - \breve{\alpha}X'X)^{-1}(X'Py - \breve{\alpha}X'y).$$

for $\breve{\alpha} = [\tilde{\alpha} - (1 - \tilde{\alpha})C/T]/[1 - (1 - \tilde{\alpha})C/T]$ and $\tilde{\alpha} = \tilde{u}P\tilde{u}/\tilde{u}'\tilde{u}$, where $\tilde{u} = y - X\tilde{\delta}$. FULL has moments of all orders, is approximately mean unbiased for $C = 1$, and is second order admissible for $C \geq 4$ under standard large sample asymptotics.

The heteroskedasticity robust LIML estimator (RLML) is obtained by dropping $i = j$ terms from the numerator. It takes the form

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

This estimator is the same as LIML except that $\sum_i (y_i - X_i'\delta)^2 P_{ii}$ has been subtracted from the numerator.

This estimator is invariant to normalization. Let $\bar{X} = [y, X]$. Then $\hat{d} = (1, -\hat{\delta}')'$ solves

$$\min_{d:d_1=-1} \frac{d'\left(\sum_{i\neq j} \bar{X}_i P_{ij} \bar{X}_j'\right) d}{d'\bar{X}'\bar{X}d}.$$

Another normalization, such as imposing that another d is equal to 1 would produce the same estimator, up to the normalization. LIML also has this invariance property.

This estimator is as simple to compute as is LIML. Similarly to LIML, $\tilde{\alpha} = \hat{Q}(\hat{\delta})$ is the smallest eigenvalue of $(\bar{X}'\bar{X})^{-1}\sum_{i\neq j} \bar{X}_i P_{ij} \bar{X}_j'$. Also, first order conditions for $\hat{\delta}$ are

$$0 = \sum_{i\neq j} X_i P_{ij}\left(y_j - X_j'\hat{\delta}\right) - \tilde{\alpha}\sum_i X_i(y_i - X_i'\hat{\delta}).$$

Solving gives

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

This estimator has a similar form to LIML except that the terms involving P_{ii} have been deleted.

By replacing $\tilde{\alpha}$ with some other value $\hat{\alpha}$ we can form a k-class version of a jackknife estimator, having the form

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

We can form a heterskedasticity consistent version of FULL by replacing $\tilde{\alpha}$ with $\hat{\alpha} = [\tilde{\alpha} - (1-\tilde{\alpha})C/T]/[1-(1-\tilde{\alpha})C/T]$ for some constant C. The small sample poperties of this estimator are unknown, but we expect its performance relative to RLML to be similar to that of FULL relative to LIML. As pointed out by Hahn, Hausman, and Kuersteiner (2004), FULL has much smaller dispersion than LIML with weak instruments, so we expect that same for RFLL. Monte Carlo results given below confirm these properties.

An asymptotic variance estimator is useful for constructing large sample confidence intervals and tests. To describe it, let $\hat{u}_i = y_i - X_i'\hat{\delta}$, $\hat{\gamma} = X'\hat{u}/\hat{u}'\hat{u}$, $\hat{V} = (I-P)X - \hat{u}\hat{\gamma}'$,

$$\begin{aligned} \hat{H} &= \sum_{i\neq j} X_i P_{ij} X_j' - \hat{\alpha}X'X, \hat{\Sigma} = \sum_{i,j\neq k} X_j P_{ji}\hat{u}_i^2 P_{ik} X_k', \\ \hat{\Sigma}_2 &= \sum_{i\neq j} P_{ij}^2\left(\hat{u}_i^2 \hat{V}_j \hat{V}_j' + \hat{V}_i \hat{u}_i \hat{u}_j \hat{V}_j'\right). \end{aligned}$$

The variance estimator is

$$\hat{V} = \hat{H}^{-1}\left(\hat{\Sigma} + \hat{\Sigma}_2\right)\hat{H}^{-1}.$$

We can interpret $\hat{\delta}$ as a combination of forward and reverse jackknife IV (JIV) estimators. For simplicity, we give this interpretation in the scalar δ case. Let $u_i(\delta) = y_i - X_i'\delta$, $\hat{\gamma}(\delta) = \sum_i X_i u_i(\delta)/\sum_i u_i(\delta)^2$, and $\hat{\gamma} = \hat{\gamma}(\hat{\delta})$. First-order conditions for $\hat{\delta}$ are

$$0 = \frac{\partial \hat{Q}(\hat{\delta})}{\partial \delta}\left(-\frac{1}{2}\right)\sum_i u_i(\hat{\delta})^2 = \sum_{i \neq j}[X_i - \hat{\gamma} u_i(\hat{\delta})]P_{ij}(y_j - X_j'\hat{\delta}).$$

The forward JIV estimator $\bar{\delta}$ is JIVE2 from Angrist, Imbens, and Krueger (1994), being

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

The reverse JIV is obtained as follows. Dividing the structural equation by δ_0 gives

$$X_i = y_i/\delta_0 - u_i/\delta_0.$$

Applying JIV to this equation to estimate $1/\delta_0$ and then inverting gives the reverse JIV

$$\bar{\delta}^r = \left(\sum_{i \neq j} y_i P_{ij} X_j\right)^{-1} \sum_{i \neq j} y_i P_{ij} y_j.$$

Then collecting terms in the first-order conditions for RLML gives

$$\begin{aligned} 0 &= (1 + \hat{\gamma}\hat{\delta})\sum_{i \neq j} X_i P_{ij}(y_j - X_j'\hat{\delta}) - \hat{\gamma}\sum_{i \neq j} y_i P_{ij}(y_j - X_j'\hat{\delta}) \\ &= (1 + \hat{\gamma}\hat{\delta})\sum_{i \neq j} X_i P_{ij} X_j(\bar{\delta} - \hat{\delta}) - \hat{\gamma}\sum_{i \neq j} y_i P_{ij} X_j(\bar{\delta}^r - \hat{\delta}). \end{aligned}$$

Dividing through by $\sum_{i \neq j} X_i P_{ij} X_j$ gives

$$0 = (1 + \hat{\gamma}\hat{\delta})(\bar{\delta} - \hat{\delta}) - \hat{\gamma}\bar{\delta}(\bar{\delta}^r - \hat{\delta}).$$

If we replace $\hat{\gamma}$ by some other estimator $\bar{\gamma}$, such as $\bar{\gamma} = \hat{\gamma}(\bar{\delta})$, and the $\hat{\gamma}\bar{\delta}$ coefficient following the minus sign by $\bar{\gamma}\hat{\delta}$ we obtain a linearized version of this equation that can be solved for $\hat{\delta}$ to obtain

$$\dot{\delta} = \frac{\bar{\delta}}{1 - \bar{\gamma}(\bar{\delta} - \bar{\delta}_r)}.$$

This estimator will be asymptotically equivalent to the RML estimator.

This result is analogous to that of Hahn and Hausman (2002) that under homoskedasticity LIML is an optimal combination of forward and reverse bias corrected two stage least squares estimators. We find a similar result, that RLML is a function of forward and reverse heteroskedasticity robust bias corrected estimators.

3. The Problem with LIML and a Specification Test

As shown by Bekker and van der Ploeg (2005), LIML is inconsistent under many instruments and heteroskedasticity. Some straightforward calculations can be used to show why this is the case and pinpoint the precise condition that would be needed for consistency. The Fisher condition for consistency is that the first derivative of the objective function converges to zero when it is evaluated at the truth. Thus a necessary condition for LIML is that $\partial \tilde{Q}(\beta_0)/\partial\beta \xrightarrow{p} 0$. Let $\sigma_i^2 = E[u_i^2]$, $\gamma_i = E[X_i u_i]/\sigma_i^2 = E[V_i u_i]/\sigma_i^2$, and $\gamma_n = \sum_i E[X_i u_i]/\sum_i \sigma_i^2 = \sum_i \sigma_i^2 \gamma_i / \sum_i \sigma_i^2$. It can be shown that

$$(-2u'u/n)\partial \tilde{Q}(\beta_0)/\partial\beta = \left(X - \frac{X'u}{u'u}u\right)' Pu/n = E[(X - \gamma_n u)' Pu/n] + o_p(1).$$

where the last equality will follow by $X'u/u'u - \gamma_n \xrightarrow{p} 0$ and convergence of quadratic forms to their expectations. Then by independence of the observations,

$$E[(X - \gamma_n u)' Pu/n] = \sum_i E[(X_i - \gamma_n u_i) P_{ii} u_i]/n = \sum_i (\gamma_i - \gamma_n) P_{ii} \sigma_i^2 / n.$$

Under many instrument asymptotics we will not have $P_{ii} \longrightarrow 0$, and hence if $\gamma_i \neq \gamma_n$ the expectation will generally not converge to zero.

There are two basic conditions under which $\sum_i (\gamma_i - \gamma_n) P_{ii} \sigma_i^2 / n = 0$ and LIML should be consistent. They are

1) P_{ii} is constant;

2) $E[V_i u_i]/\sigma_i^2$ does not vary with i.

The first condition will be satisfied if the instruments are dummy variables with equal numbers of ones. This condition is given by Bekker and van der Berg (2005), and is fererred to as equal group sizes. The second condition restricts the joint variance matrix of the structure and reduced form to satisfy

$$Var(u_i, V_i') = \begin{bmatrix} \sigma_i^2 & \sigma_i^2\gamma' \\ \sigma_i^2\gamma & Var(V_i) \end{bmatrix}.$$

Of course, if either 1) or 2) are close to being satisfied, then the LIML estimator will be close to being consistent.

Analogous arguments can also be used to show that, with heteroskedasticity, FULL is inconsistent under many instruments and LIML is inconsistent under many weak instruments.

4. Optimal Estimation with Heteroskedasticty

RLML is not asymptotically efficient under heteroskedasticity and many instruments. In GMM terminology, it uses a nonoptimal weighting matrix, one that is not consistent under heteroskedasticity. In addition, it does not use a heteroskedasticity consistent projection of the endogenous variables on the disturbance, which leads to inefficiency in the many instruments correction term. Efficiency can be obtained by modifying the estimator so that the weight matrix and the projection are heterskedasticity consistent. Let

$$\begin{aligned} \hat{\Omega}(\delta) &= \sum_{i=1}^{n} Z_i Z_i' u_i(\delta)^2/n, \hat{B}_k(\delta) = \left(\sum_i Z_i Z_i' u_i(\delta) X_{ik}/n\right)\hat{\Omega}(\delta)^{-1}, \\ \hat{D}_{ik}(\delta) &= Z_i X_{ik} - \hat{B}_k(\delta) Z_i u_i(\delta), \hat{D}_i(\delta) = \left[\hat{D}_{i1}(\delta), ..., \hat{D}_{ip}(\delta)\right], \end{aligned}$$

Also let $\bar{\delta}$ be a preliminary estimator (such as RLML). An IV estimator that is efficient under heteroskedasticity of unknown form and many instruments is

$$\hat{\delta} = \left(\sum_{i \neq j} \hat{D}_i(\bar{\delta})' \hat{\Omega}(\bar{\delta})^{-1} Z_j X_j'\right)^{-1} \sum_{i \neq j} \hat{D}_i(\bar{\delta})' \hat{\Omega}(\bar{\delta})^{-1} Z_j y_j.$$

This is a jackknife IV estimator with an optimal weigthing matrix $\hat{\Omega}(\bar{\delta})^{-1}$ and $\hat{D}_i(\bar{\delta})$ replacing $X_i Z_i'$. The use of $\hat{D}_i(\bar{\delta})$ makes the estimator as efficient as the CUE under many weak instruments.

The asymptotic variance can be estimated by

$$\hat{V} = \hat{H}^{-1}\hat{\Sigma}\hat{H}^{-1}, \hat{H} = \sum_{i \neq j} X_i Z_i' \hat{\Omega}(\tilde{\delta})^{-1} Z_j X_j', \hat{\Sigma} = \sum_{i,j=1}^{n} \hat{D}_i(\tilde{\delta})' \hat{\Omega}(\tilde{\delta})^{-1} \hat{D}_j(\tilde{\delta}).$$

This estimator has a sandwich form similar to that given in Newey and Windmeijer (2005).

5. The Robust, Restricted CUE

The RLML has been made robust to heteroskedasticity by jackknifing where own observation terms are removed. In general this same approach can be used to make the continuous updating estimator (CUE) robust to restrictions on the weighting matrix, such as homoskedasticity. For example, LIML is a CUE where homoskedasticity is imposed on the weighting matrix and RLML is its robust version.

For expository purposes consider a general GMM setup where δ denotes a $G \times 1$ parameter vector and $g_i(\delta)$ an $K \times 1$ vector of functions of the data and parameter satisfying $E[g_i(\delta_0)] = 0$. For example, in the linear IV environment, $g_i(\delta) = Z_i(y_i - X_i'\delta)$. Let $\tilde{\Omega}(\delta)$ denote an estimator of $\Omega(\delta) = \sum_{i=1}^n E[g_i(\delta)g_i(\delta)']/n$, where an n subscript on $\Omega(\delta)$ is suppressed for notational convenience. A CUE is given by

$$\hat{\delta} = \arg\min_{\delta} \hat{g}(\delta)'\tilde{\Omega}(\delta)^{-1}\hat{g}(\delta).$$

When $\tilde{\Omega}(\delta) = \sum_{i=1}^n g_i(\delta)g_i(\delta)'/n$ this estimator is the CUE given by Hansen, Heaton, and Yaron (1996), that places no restrictions on the estimator of the second moment matrices. In general, restrictions may be imposed on the second moment matrix. For example, in the IV setting where $g_i(\delta) = Z_i(y_i - X_i'\delta)$, we may specify $\tilde{\Omega}(\delta)$ to be only consistent under homoskedasticity,

$$\tilde{\Omega}(\delta) = (y - X\delta)'(y - X\delta)\, Z'Z/n^2.$$

In this case the CUE objective function is

$$\hat{g}(\delta)'\tilde{\Omega}(\delta)^{-1}\hat{g}(\delta) = \frac{(y - X\delta)'P(y - X\delta)}{(y - X\delta)'(y - X\delta)},$$

which is the LIML objective function (as is well known; see Hansen, Heaton, and Yaron, 1996).

A CUE will tend to have low bias when the restrictions imposed on $\tilde{\Omega}(\delta)$ are satisfied but may be more biased otherwise. A simple calculation can be used to explain this bias. Consider a CUE where $\tilde{\Omega}(\delta)$ is replaced by its expectation $\bar{\Omega}(\delta) = E[\tilde{\Omega}(\delta)]$. This replacement is justified under many weak moment asymptotics. The expectation of the CUE objective function is then

$$E[\hat{g}(\delta)'\bar{\Omega}(\delta)^{-1}\hat{g}(\delta)] = (1 - n^{-1})\bar{g}(\delta)'\bar{\Omega}(\delta)^{-1}\bar{g}(\delta) + tr(\bar{\Omega}(\delta)^{-1}\Omega(\delta))/n,$$

where $\bar{g}(\delta) = E[g_i(\delta)]$ and $\Omega(\delta) = E[g_i(\delta)g_i(\delta)']$. The first term is minimized at δ_0 where $\bar{g}(\delta_0) = 0$. When $\bar{\Omega}(\delta) = \Omega(\delta)$ then

$$tr(\bar{\Omega}(\delta)^{-1}\Omega(\delta))/n = m/n,$$

so the second term does not depend on δ. In this case the expected value of the CUE objective function is minimized at δ_0. When $\bar{\Omega}(\delta) \neq \Omega(\delta)$ the second term will depend on δ, and so the expected value of the CUE objective will not be minimized at δ_0. This effect will lead to bias in the CUE, because the estimator will be minimizing an objective function with expectation that is not minimized at the truth. It is also interesting to note that this bias effect will tend to increase with n. This bias was noted by Han and Phillips (2005) for two-stage GMM, who referred to the bias term as a "noise" term, and the other term as a "signal."

We robustify the CUE by jackknifing, i.e. deleting the own observation terms in the CUE quadratic form. Note that

$$E[\sum_{i\neq j} g_i(\delta)'\bar{\Omega}(\delta)^{-1}g_j(\delta)/n^2] = (1-n^{-1})\bar{g}(\delta)'\bar{\Omega}(\delta)^{-1}\bar{g}(\delta),$$

which is always minimized at δ_0, not matter what $\bar{\Omega}(\delta)$ is. A corresponding estimator is obtained by replacing $\bar{\Omega}(\delta)$ by $\tilde{\Omega}(\delta)$ and minimizing, i.e.

$$\hat{\delta} = \arg\min_{\delta} \sum_{i\neq j} g_i(\delta)'\tilde{\Omega}(\delta)^{-1}g_j(\delta)/n^2.$$

This is a robust CUE (RCUE), that should have small bias by virtue of the jackknife form of the objective function. The RLML estimator is precisely of this form, for $\tilde{\Omega}(\delta) = (y - X\delta)'(y - X\delta)Z'Z/n^2$.

6. Asymptotic Theory

Theoretical justification of the estimators proposed here is provided by asymptotic theory where the number of instruments grows with the sample size. Some regularity conditions are important for the results. Let Z_i', u_i, V_i', and Υ_i' denote the i^{th} row of Z, u, V, and Υ respectively. Here we will consider the case where Z is constant, leaving the treatment of random Z to future research.

Assumption 1: Z includes among its columns a vector of ones, $rank(Z) = K$, $P_{ii} \leq C < 1, (i = 1, ..., n)$.

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

Assumption 2: There is a $G \times G$ matrix $S_n = \tilde{S}_n \operatorname{diag}(\mu_{1n}, ..., \mu_{Gn})$ and z_i such that $\Upsilon_i = S_n z_i/\sqrt{n}$, $\tilde{S}_n$ is bounded and the smallest eigenvalue of $\tilde{S}_n \tilde{S}_n'$ is bounded away from zero, for each j either $\mu_{jn} = \sqrt{n}$ or $\mu_{jn}/\sqrt{n} \longrightarrow 0$, $\mu_n = \min_{1 \leq j \leq G} \mu_{jn} \longrightarrow \infty$, and $\sqrt{K}/\mu_n^2 \longrightarrow 0$. Also, $\sum_{i=1}^n \|z_i\|^4 / n^2 \longrightarrow 0$, and $\sum_{i=1}^n z_i z_i'/n$ is bounded and uniformly nonsingular.

Setting $\mu_{jn} = \sqrt{n}$ leads to asymptotic theory like Kunitomo (1980), Morimune (1984), and Bekker (1994), where the number of instruments K can grow as fast as the sample size. In that case the condition $\sqrt{K}/\mu_n^2 \longrightarrow 0$ would be automatically satisfied. Allowing for K to grow and for μ_n to grow slower than $\sqrt{n}$ models having many instruments without strong identification. This condition then allows for some components of the reduced form to give only weak identification (corresponding to $\mu_{jn}/\sqrt{n} \longrightarrow 0$) and other components (corresponding to $\mu_{jn} = \sqrt{n}$) to give strong identification. In particular, this condition allows for fixed constant coefficients in the reduced form.

Assumption 3: $(u_1, V_1), ..., (u_n, V_n)$ are independent with $E[u_i] = 0$, $E[V_i] = 0$, $E[u_i^8]$ and $E[\|V_i\|^8]$ are bounded in i, $Var((u_i, V_i')') = diag(\Omega_i^*, 0)$, and $\sum_{i=1}^n \Omega_i^*/n$ is uniformly nonsingular.

This hypothesis includes moment existence assumptions. It also requres that the average variance of the nonzero reduced form disturbances be nonsingular, as is useful for the proof of consistency.

[1] The observations $w_1, ..., w_T$ are distinct with probability one and therefore, by $K < T$, cannot all be roots of a K^{th} degree polynomial. It follows that for any nonzero a there must be some t with $a'Z_t = a'p^K(w_t) \neq 0$, implying $a'Z'Za > 0$.

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

This condition allows an unknown reduced form that is approximated by a linear combination of the instrumental variables. It is possible to replace this assumption with the condition that $\sum_{i \neq j} z_i P_{ij} z_j'/n$ is uniformly nonsingular.

We can easily interpret all these conditions in an important example of a linear model with exogenous covariates and a possibly unknown reduced form. This example is given by

$$X_i = \begin{pmatrix} \pi_{11} Z_{1i} + \mu_n f_0(w_i)/\sqrt{n} \\ Z_{1i} \end{pmatrix} + \begin{pmatrix} v_i \\ 0 \end{pmatrix}, Z_i = \begin{pmatrix} Z_{1i} \\ p^K(w_i) \end{pmatrix},$$

where Z_{1i} is a $G_2 \times 1$ vector of included exogenous variables, $f_0(w)$ is a $G - G_2$ dimensional vector function of a fixed dimensional vector of exogenous variables w and $p^K(w) \overset{def}{=} (p_{1K}(w), ..., p_{K-G_2,K}(w))'$. The variables in X_i other than Z_{1i} are endogenous with reduced form $\pi_{11} Z_{1i} + \mu_n f_0(w_i)/\sqrt{n}$. The function $f_0(w)$ may be a linear combination of a subvector of $p^K(w)$, in which case $\Delta_n = 0$ in Assumption 4 or it may be an unknown function that can be approximated by a linear combination of $p^K(w)$. For $\mu_n = \sqrt{n}$ this example is like the model in Donald and Newey (2001) where Z_i includes approximating functions for the optimal (asymptotic variance minimizing) instruments Υ_i, but the number of instruments can grow as fast as the sample size. When $\mu_n^2/n \longrightarrow 0$, it is a modified version where the model is more weakly identified.

To see precise conditions under which the assumptions are satisfied, let

$$z_i = \begin{pmatrix} f_0(w_i) \\ Z_{1i} \end{pmatrix}, S_n = \tilde{S}_n diag\left(\mu_n, ..., \mu_n, \sqrt{n}, ..., \sqrt{n}\right), \tilde{S}_n = \begin{pmatrix} I & \pi_{11} \\ 0 & I \end{pmatrix}.$$

By construction we have $\Upsilon_i = S_n z_i/\sqrt{n}$. Assumption 2 imposes the requirements that

$$\sum_{i=1}^n \|z_i\|^4 / n^2 \longrightarrow 0, \sum_{i=1}^n z_i z_i'/n \text{ is bounded and uniformly nonsingular.}$$

The other requirements of Assumption 2 are satisfied by construction. Turning to Assumption 3, we require that $\sum_{i=1}^n Var(u_i, v_i')/n$ is uniformly nonsingular. For Assumption 4, let $\pi_{Kn} = [\tilde{\pi}_{Kn}', [I_{G_2}, 0]']'$. Then Assumption 4 will be satisfied if for each n there exists $\tilde{\pi}_{Kn}$ with

$$\Delta_n^2 = \sum_{i=1}^n \|z_i - \pi_{Kn}' Z_i\|^2/n = \sum_{i=1}^n \|f_0(w_i) - \tilde{\pi}_{Kn}' Z_i\|^2/n \longrightarrow 0.$$

Theorem 1: *If Assumptions 1-4 are satisfied and* $\hat{\alpha} = o_p(\mu_n^2/n)$ *or* $\hat{\delta}$ *is LIML or FULL then* $\mu_n^{-1} S_n'(\hat{\delta} - \delta_0) \xrightarrow{p} 0$ *and* $\hat{\delta} \xrightarrow{p} \delta_0$.

This result gives convergence rates for linear combinations of $\hat{\delta}$. For instance, in the linear model example set up above, it implies that $\hat{\delta}_1$ is consistent and and that $\pi_{11}'\hat{\delta}_1 + \hat{\delta}_2 = o_p(\mu_n/\sqrt{n})$.

The asymptotic variance of the estimator will depend on the growth rate of K relative to μ_n^2. The following condition allows for two cases.

Assumption 5: Either I) K/μ_n^2 is bounded and $\sqrt{K} S_n^{-1} \longrightarrow S_0$ or; II) $K/\mu_n^2 \longrightarrow \infty$ and $\mu_n S_n^{-1} \longrightarrow \bar{S}_0$.

To state a limiting distribution result it is helpful to also assume that certain objects converge. Let $\sigma_i^2 = E[u_i^2]$, $\gamma_n = \sum_{i=1}^n E[V_i u_i] / \sum_{i=1}^n \sigma_i^2$, $\tilde{V} = V - u\gamma_n'$, having i^{th} row $\tilde{V}_i'$; and let $\tilde{\Omega}_i = E[\tilde{V}_i \tilde{V}_i']$.

Assumption 6: $H_P = \lim_{n \longrightarrow \infty} \sum_{i=1}^n (1 - P_{ii}) z_i z_i'/n$, $\Sigma_p = \lim_{n \longrightarrow \infty} \sum_{i=1}^n (1 - P_{ii})^2 z_i z_i' \sigma_i^2/n$, $\Psi = \lim_{n \longrightarrow \infty} \sum_{i \neq j} P_{ij}^2 \left(\sigma_i^2 E[\tilde{V}_j \tilde{V}_j'] + E[\tilde{V}_i u_i] E[u_j \tilde{V}_j'] \right) / K$.

We can now state the asymptotic normality results. In case I) we will have

$$S_T'(\hat{\delta} - \delta_0) \xrightarrow{d} N(0, \Lambda_I), \Lambda_I = H_P^{-1} \left(\Sigma_P + S_0 \Psi S_0' \right) H_P^{-1}, \tag{1}$$

In case II) we will have

$$(\mu_T/\sqrt{K}) S_T'(\hat{\delta} - \delta_0) \xrightarrow{d} N(0, \Lambda_{II}), \Lambda_{II} = H_P^{-1} \bar{S}_0 \Psi \bar{S}_0' H_P^{-1}. \tag{2}$$

The asymptotic variance expressions allow for the many instrument sequence of Kunitomo (1980), Morimune (1983), and Bekker (1994) and the many weak instrument sequence of Chao and Swanson (2003, 2005). When K grows faster than μ_n^2 the asymptotic variance of $\hat{\delta}$ may be singular. This occurs because the many instruments adjustment term $\bar{S}_0 \Psi \bar{S}_0'$ will be singular with included exogenous variables and it dominates the matrix Σ_P when K grows that fast.

Theorem 2: *If Assumptions 1-6 are satisfied,* $\hat{\alpha} = \tilde{\alpha} + 0_p(1/T)$ *or* $\hat{\delta}$ *is LIML or FULL, then in case I) equation (1) is satisfied and in case II) equation (2) is satisfied.*

It is interesting to compare the asymptotic variance of the RLML estimator with that of LIML under when the disturbances are homoskedastic. Under homoskedasticity the variance of $Var((u_i, V_i'))$ will not depend on i, e.g. so that $\sigma_i^2 = \sigma^2$. Then $\gamma_n = E[X_i u_i]/\sigma^2 = \gamma$ and $E[\tilde{V}_i u_i] = E[V_i u_i] - \gamma\sigma^2 = 0$, so that

$$\Sigma_p = \sigma^2 \tilde{H}_p, \tilde{H}_P = \lim_{n \longrightarrow \infty} \sum_{i=1}^{n} (1 - P_{ii})^2 z_i z_i'/n, \Psi = \sigma^2 E[\tilde{V}_j \tilde{V}_j'](1 - \lim_{n \longrightarrow \infty} \sum_i P_{ii}^2/K).$$

Focusing on Case I, letting $\Lambda = S_0 \Psi S_0'$, the asymptotic variance of RLML is then

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

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 (2006), for $H = \lim_{n \longrightarrow \infty} \sum_{i=1}^{n} z_i z_i'/n$ and $\tau = \lim_{n \longrightarrow \infty} K/n$, the asymptotic variance of LIML is

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

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

In the many instruments case where K and μ_n^2 grow as fast as n we can compare the leading terms in the asymptotic variances. It follows from asymptotic efficiency of least squares relative to instrumental variables, with regressors z_i and instruments $(1 - P_{ii})z_i$, that

$$H^{-1} \leq H_P^{-1} \tilde{H}_P H_P^{-1}$$

in the positive semi-definite sense. Thus the leading term in the LIML asymptotic variance is smaller than the leading term in the RLML asymptotic variance. Furthermore, it follows from the results of Chioda and Jansson (2006) that the entire LIML asymptotic variance is less than the RLML asymptotic variance, so that there is some efficiency loss in RLML under homoskedasticity. As previously mentioned, this loss does not seem very important for practice where n tends to be small relative to K.

It remains to show consistency of the asymptotic variance estimator. This is straighforward to do under many weak instruments, and will be included in the next version of this paper. For many instruments, it appears that the variance estimator is not consistent. Again, the many weak instrument case seems most relevant for practice.

7. Simulations

A Monte Carlo study provides some evidence about how RLML and RFLL behave. To describe the design, let the first column of Z (denoted $Z_{\cdot 1}$)be drawn from standard normal distribution; the second column to the $(K-1)th$ column of Z are constructed by $Z_{i1} \cdot w_{ir}$, where $w_{ir} = 1$ with probability .5 and $w_{ir} = 0$ with probability .5 for $2 \leq r \leq K-1$ and $w_{i1}, ..., w_{i,K-2}$ are independent, and $Z_{iK} = 1$. The reduced form is given by $X = Z_{\cdot 1} + v$, where $v_i \tilde{} N(0,1)$. The structural disturbance is given by

$$u_i = \rho v_i + \eta_i, \eta_i \tilde{} N(0, Z_{i1}^2).$$

We also set $\rho = .3$ and $n = 800$. This is a design that will lead to LIML being inconsistent with many instruments.

Below we report results on median bias and the range between the .05 and .95 quantiles for LIML, RLML, Jackknife, . Interquartile range results were similar. We find that LIML is biased when there are many instruments, RLML is not, and that RLML also has smaller spread than LIML. Here the performance of all the jackknife estimators is similar, because identification is quite strong.

Median Bias

K	$LIML$	$RFLL$	$RLML$	JK
10	$-.0064$	.0001	.0001	$-.0011$
20	$-.0140$	$-.0010$	$-.0010$	$-.0020$
50	$-.0362$	.0005	.0005	$-.0005$
100	$-.0873$	.0001	.0001	$-.0002$

Nine Decile Range; .05 to .95.

K	$LIML$	$RFLL$	$RLML$	JK
10	$-.2083$	.2000	.2000	.1994
20	$-.2166$	.1991	.1992	.2010
50	$-.2341$	.1931	.1931	.1936
100	$-.0873$	.1935	.1935	.1940

8. Appendix: Proofs of Theorems.

Throughout, let C denote a generic positive constant that may be different in different uses and let M, CS, and T denote the conditional Markov inequality, the Cauchy-Schwartz inequality, and the Triangle inequality respectively. The first Lemma is proved in Hansen, Hausman, and Newey (2006).

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

The next three results are proven in Chao, Newey, and Swanson (2006).

Lemma A1: *If* $(w_i, v_i), (i = 1, ..., n)$ *are independent,* w_i *and* v_i *are scalars, and* P *is symmetric, idempotent of rank* K *then for* $\bar{w} = E[(w_1, ..., w_n)']$ *and* $\bar{v} = E[(v_1, ..., v_n)']$,

$$\begin{aligned}\sum_{i\neq j} P_{ij}w_iv_j &= \sum_{i\neq j} P_{ij}\bar{w}_i\bar{v}_j + O_p(K^{1/2}\max_{i\leq n}[Var(w_i)Var(v_i)]^{1/2}) \\ &\quad + O_p(\max_{i\leq n} Var(v_i)^{1/2}[\bar{w}'\bar{w}]^{1/2} + \max_{i\leq n} Var(w_i)^{1/2}[\bar{v}'\bar{v}]^{1/2}).\end{aligned}$$

Lemma A2: *If i)* P *is a symmetric, idempotent matrix with rank*$(P) = K$, $P_{ii} \leq C < 1$; *ii)* (W_{1n}, V_1, u_1), ..., (W_{nn}, V_n, u_n) *are independent and* $D_n = \sum_{i=1}^n E[W_{in}W_{in}']$ is bounded; *iii)* $E[W_{in}'] = 0$, $E[V_i] = 0$, $E[u_i] = 0$ *and there exists a constant* C *such that* $E[\|V_i\|^4] \leq C$, $E[u_i^4] \leq C$; *iv)* $\sum_{i=1}^n E[\|W_{in}\|^4] \longrightarrow 0$; *(v) for* $\Sigma_i = Var((u_i, v_i'))$, $\lambda_{\min}(\Sigma_i) \geq C$; *vi)* $K \longrightarrow \infty$; *then for* $\bar{\Sigma}_n \stackrel{\text{def}}{=} \sum_{i\neq j} P_{ij}^2\left(E[V_iV_i']E[u_j^2] + E[V_iu_i]E[u_jV_j']\right)/K$ *and for* $\Xi_n = c_{1n}'D_nc_{1n} + c_{2n}'\bar{\Sigma}_nc_{2n} > C$ *with* c_{1n} *and* c_{2n} *being any sequence of bounded nonzero vectors, it follows that*

$$\begin{aligned}Y_n &= \Xi_n^{-1/2}(\sum_{i=1}^n c_{1n}'W_{in} + \\ &\quad c_{2n}'\sum_{i\neq j} V_iP_{ij}u_j/\sqrt{K}) \xrightarrow{d} N(0,1).\end{aligned}$$

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

Lemma A3: *If Assumptions 1-4 are satisfied and* $\sqrt{K}/\mu_n^2 \longrightarrow 0$ *then*

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

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

LEMMA A4: *If Assumptions 1-4 are satisfied then* $S_n'(\hat{\delta}-\delta_0)/\mu_n \xrightarrow{p} 0$.

Proof: Let $\bar{\Upsilon}=[0,\Upsilon]$, $\bar{V}=[u,V]$, $\bar{X}=[y,X]$, so that $\bar{X}=(\bar{\Upsilon}+\bar{V})D$ for

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

Let $\hat{B}=\bar{X}'\bar{X}/n$. Note that $\|S_n/\sqrt{n}\|\le C$. Then by $tr(\bar{\Upsilon}'\bar{\Upsilon})=tr(S_n z'zS_n')/n$ and $E[\bar{V}\bar{V}']\le CI_n$,

$$E[\left\|\bar{\Upsilon}'\bar{V}\right\|^2/n^2]=tr(\bar{\Upsilon}'E[\bar{V}'\bar{V}]\bar{\Upsilon})/n^2\le Ctr(S_n z'zS_n')/n^3\longrightarrow 0,$$

so that $\bar{\Upsilon}'\bar{V}/n\xrightarrow{p}0$ by M. Let $\bar{\Omega}_n=\sum_{i=1}^n E[\bar{V}_i\bar{V}_i']/n=diag(\sum_{i=1}^n\Omega_i^*/n,0)\ge Cdiag(I_{G-G_2+1},0)$ by Assumption 3. By M we have $\bar{V}'\bar{V}/n-\bar{\Omega}_n\xrightarrow{p}0$, so it follows that w.p.a.1.

$$\hat{B}=(\bar{V}'\bar{V}+\bar{\Upsilon}'\bar{V}+\bar{V}'\bar{\Upsilon}+\bar{\Upsilon}'\bar{\Upsilon})/n=\bar{\Omega}_n+\bar{\Upsilon}'\bar{\Upsilon}/n+o_p(1)\ge Cdiag(I_{G-G_2+1},0).$$

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

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

Next, as defined preceding Lemma A3 let $\bar{S}_n=diag(\mu_n,S_n)$ and $\tilde{X}=[u,X]\bar{S}_n^{-1\prime}$. Note that by $P_{ii}\le C<1$ and uniform nonsingularity of $\sum_{i=1}^n z_iz_i'/n$ we have $H_n\ge(1-C)\sum_{i=1}^n z_iz_i'/n\ge CI_G$. Then by Lemma A3, w.p.a.1.

$$\hat{A}\overset{def}{=}\sum_{i\ne j}P_{ij}\tilde{X}_i\tilde{X}_j'\ge Cdiag(0,I_G),$$

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

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

Let $\hat{Q}(\delta)=(n/\mu_n^2)\sum_{i\ne j}(y_i-X_i'\delta)P_{ij}(y_j-X_j'\delta)/(y-X\delta)'(y-X\delta)$. Then by the upper left element of the conclusion of Lemma A3 $\mu_n^{-2}\sum_{i\ne j}u_iP_{ij}u_j\xrightarrow{p}0$. Then w.p.a.1

$$\left|\hat{Q}(\delta_0)\right|=\left|\mu_n^{-2}\sum_{i\ne j}u_iP_{ij}u_j/\sum_{i=1}^n u_i^2/n\right|\xrightarrow{p}0.$$

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

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

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

LEMMA A5: *If Assumptions 1-4 are satisfied,* $\hat{\alpha} = o_p(\mu_n^2/n)$, *and* $S_n'(\hat{\delta}-\delta_0)/\mu_n \xrightarrow{p} 0$ *then for* $H_n = \sum_{i=1}^n (1-P_{ii}) z_i z_i'/n$,

$$S_n^{-1}\left(\sum_{i\neq j} X_i P_{ij} X_j' - \hat{\alpha} X'X\right) S_n^{-1\prime} = H_n + o_p(1), S_n^{-1}(\sum_{i\neq j} X_i P_{ij} \hat{u}_j - \hat{\alpha} X'\hat{u})/\mu_n \xrightarrow{p} 0.$$

Proof: By M $X'X = O_p(n)$ and $X'\hat{u} = O_p(n)$. Therefore, by $\left\|S_n^{-1}\right\| = O(\mu_n^{-1})$,

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

Lemma A3 (lower right hand block) and T then give the first conclusion. By Lemma A3 (off diagonal) we have $S_n^{-1}\sum_{i\neq j} X_i P_{ij} u_j/\mu_n \xrightarrow{p} 0$, so that

$$S_n^{-1}\sum_{i\neq j} X_i P_{ij} \hat{u}_j/\mu_n = o_p(1) - \left(S_n^{-1}\sum_{i\neq j} X_i P_{ij} X_j' S_n^{-1\prime}\right) S_n'(\hat{\delta}-\delta_0)/\mu_n \xrightarrow{p} 0. Q.E.D.$$

LEMMA A6: *If Assumptions 1 - 4 are satisfied and* $S_n'(\hat{\delta}-\delta_0)/\mu_n \xrightarrow{p} 0$ *then* $\sum_{i\neq j} \hat{u}_i P_{ij} \hat{u}_j / \hat{u}'\hat{u} = o_p(\mu_n^2/n)$.

Proof: Let $\hat{\beta} = S_n'(\hat{\delta}-\delta_0)/\mu_n$ and $\breve{\alpha} = \sum_{i\neq j} u_i P_{ij} u_j / u'u = o_p(\mu_n^2/n)$. Note that $\hat{\sigma}_u^2 = \hat{u}'\hat{u}/n$ satisfies $1/\hat{\sigma}_u^2 = O_p(1)$ by M. By Lemma A5 with $\hat{\alpha} = \breve{\alpha}$ we have $\tilde{H}_n = S_n^{-1}(\sum_{i\neq j} X_i P_{ij} X_j' - \breve{\alpha} X'X) S_n^{-1\prime} = O_p(1)$ and $W_n = S_n^{-1}(X'Pu - \breve{\alpha} X'u)/\mu_n \xrightarrow{p} 0$, so

$$\begin{aligned} \frac{\sum_{i\neq j} \hat{u}_i P_{ij} \hat{u}_j}{\hat{u}'\hat{u}} - \breve{\alpha} &= \frac{1}{\hat{u}'\hat{u}}\left(\sum_{i\neq j} \hat{u}_i P_{ij} \hat{u}_j - \sum_{i\neq j} u_i P_{ij} u_j - \breve{\alpha}\left(\hat{u}'\hat{u} - u'u\right)\right) \\ &= \frac{\mu_n^2}{n}\frac{1}{\hat{\sigma}_u^2}\left(\hat{\beta}'\tilde{H}_n\hat{\beta} - 2\hat{\beta}'W_n\right) = o_p(\mu_n^2/n), \end{aligned}$$

so the conclusion follows by T. Q.E.D.

Proof of Theorem 1: For RLML the conclusion follows from Lemma A4. For RFLL, note that $\tilde{\alpha} = o_p(\mu_n^2/n)$, so by the formula for RFLL, $\hat{\alpha} = \tilde{\alpha} + O_p(1/n) = o_p(\mu_n^2/n)$. Then by Lemma A4 we have

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

Let $\tilde{\alpha}(\bar{\delta}) = \sum_{i\neq j} u_i(\delta)P_{ij}u_j(\delta)/u(\delta)'u(\delta)$ and

$$\hat{D}(\delta) = \partial[\sum_{i\neq j} u_i(\delta)P_{ij}u_j(\delta)/2u(\delta)'u(\delta)]/\partial\delta = \sum_{i\neq j} X_i P_{ij}u_j(\delta) - \tilde{\alpha}(\delta)X'u(\delta).$$

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

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

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

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

where the second equality follows by $\hat{D}(\bar{\delta}) = \sum_{i\neq j} X_iP_{ij}\bar{u}_j - (\bar{u}'\bar{u})\bar{\alpha}\bar{\gamma}$. By Lemma A6 we have $\bar{\alpha} = o_p(\mu_n^2/n)$. By standard arguments, $\bar{\gamma} = O_p(1)$ so that $S_n^{-1}\bar{\gamma} = O_p(1/\mu_n)$. Then by Lemma A5 and $\hat{D}(\bar{\delta}) = \sum_{i\neq j} X_iP_{ij}\bar{u}_j - \bar{\alpha}X'\bar{u}$

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

The conclusion then follows by T. Q.E.D.

LEMMA A7: *If Assumptions 1-4 are satisfied then for* $\gamma_n = \sum_i E[V_iu_i]/\sum_i E[u_i^2]$ *and* $\tilde{V}_i = V_i - \gamma_n u_i$

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

Proof: Note that for $W = z'(P-I)u/\sqrt{n}$ by $I-P$ idempotent and $E[uu'] \leq CI_n$ we have

$$\begin{aligned} E[WW'] &\leq Cz'(I-P)z/n = C(z - Z\pi_n')'(I-P)(z-Z\pi_n')/n \\ &\leq CI\sum_{i=1}^n \|z_i - \pi_nZ_i\|^2/n = O(\Delta_n^2) \longrightarrow 0, \end{aligned}$$

so $z'(P-I)u/\sqrt{n} = o_p(1)$. Also, by M, $\tilde{\gamma} = X'u/u'u = \gamma_n + O_p(1/\sqrt{n})$. Therefore, it follows by Lemma A1 and $\hat{D}(\delta_0) = \sum_{i\neq j} X_iP_{ij}u_j - u'u\tilde{\alpha}(\delta_0)\tilde{\gamma}$

$$\begin{aligned} S_n^{-1}\hat{D}(\delta_0) &= \sum_{i\neq j} z_iP_{ij}u_j/\sqrt{n} + S_n^{-1}\sum_{i\neq j}\tilde{V}_iP_{ij}u_i - S_n^{-1}(\hat{\gamma}-\gamma_n)u'u\tilde{\alpha}(\delta_0) \\ &= z'Pu/\sqrt{n} - \sum_i P_{ii}z_iu_i/\sqrt{n} + S_n^{-1}\sum_{i\neq j}\tilde{V}_iP_{ij}u_j + O_p(1/\sqrt{n}\mu_n)o_p(\mu_n^2/n) \\ &= \sum_{i=1}^n (1-P_{ii})z_iu_i/\sqrt{n} + S_n^{-1}\sum_{i\neq j}\tilde{V}_iP_{ij}u_j + o_p(1). Q.E.D. \end{aligned}$$

Proof of Theorem 2: Consider first the case where $\hat{\delta}$ is RLML. Then $\mu_n^{-1}S_n'(\hat{\delta}-\delta_0) \xrightarrow{p} 0$ by Theorem 1, implying $\hat{\delta} \xrightarrow{p} \delta_0$. The first-order conditions for LIML are $\hat{D}(\hat{\delta}) = 0$. Expanding gives

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

where $\bar{\delta}$ lies on the line joining $\hat{\delta}$ and δ_0 and hence $\bar{\beta} = \mu_n^{-1}S_n'(\bar{\delta} - \delta_0) \xrightarrow{p} 0$. Then by Lemma A6, $\bar{H}_n = S_n^{-1}[\partial \hat{D}(\bar{\delta})/\partial\delta]S_n^{-1\prime} = H_P + o_p(1)$. Then $\partial \hat{D}(\bar{\delta})/\partial\delta$ is nonsingular w.p.a.1 and solving gives

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

Next, apply Lemma A3 with $V_i = \tilde{V}_i$ and

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

By u_i having bounded fourth moment, and $P_{ii} \leq 1$,

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

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

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

Consider c such that $c'\Gamma c > 0$. Then by the conclusion of Lemma A2 we have $c'U_n \xrightarrow{d} N(0, c'\Gamma c)$. Also, if $c'\Gamma c = 0$ then it is straightforward to show that $c'U_n \xrightarrow{p} 0$. Then it follows that

$$U_n = \begin{pmatrix} \sum_{i=1}^n W_{in} \\ \sum_{i\neq j}\tilde{V}_i p_{ij}u_j/\sqrt{K} \end{pmatrix} \xrightarrow{d} N(0,\Gamma), \Gamma = \text{diag}\left(\Sigma_P, \Psi\right).$$

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

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

Then by Lemma A7,

$$\begin{aligned} S_n^{-1}\hat{D}(\delta_0) &= F_n U_n + o_p(1) \xrightarrow{d} N(0, \Sigma_P + S_0\Psi S_0'), \\ S_n'(\hat{\delta} - \delta_0) &= -\bar{H}_n^{-1}S_n^{-1}\hat{D}(\delta_0) \xrightarrow{d} N(0, \Lambda_I). \end{aligned}$$

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

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

and $(\mu_n/\sqrt{K})o_p(1) = o_p(1)$. Then by Lemma A7,

$$\begin{aligned} (\mu_n/\sqrt{K})S_n^{-1}\hat{D}(\delta_0) &= (\mu_n/\sqrt{K})F_n U_n + o_p(1) \xrightarrow{d} N(0, \bar{S}_0 \Psi \bar{S}_0'), \\ (\mu_n/\sqrt{K})S_n'(\hat{\delta} - \delta_0) &= -\bar{H}_n^{-1}(\mu_n/\sqrt{K})S_n^{-1}\hat{D}(\delta_0) \xrightarrow{d} N(0, \Lambda_{II}).Q.E.D. \end{aligned}$$

LEMMA A8: *If $(w_i, v_i), (i = 1, ..., n)$ are independent, w_i and v_i are scalars, and P is symmetric, idempotent of rank K then for $\bar{w} = E[(w_1, ..., w_n)']$ and $\bar{v} = E[(v_1, ..., v_n)']$,*

$$\sum_{i \neq j} P_{ij}^2 w_i v_j = \sum_{i \neq j} P_{ij}^2 \bar{w}_i \bar{v}_j + O_p(K^{1/2} \max_{i \leq n}[Var(w_i)Var(v_i)]^{1/2}).$$

Proof. Let $\tilde{w}_i = w_i - \bar{w}_i$ and $\tilde{v}_i = v_i - \bar{v}_i$. Note that

$$\sum_{i \neq j} P_{ij}^2 w_i v_j - \sum_{i \neq j} P_{ij}^2 \bar{w}_i \bar{v}_j = \sum_{i \neq j} P_{ij}^2 \tilde{w}_i \tilde{v}_j + \sum_{i \neq j} P_{ij}^2 \tilde{w}_i \bar{v}_j + \sum_{i \neq j} P_{ij}^2 \bar{w}_i \tilde{v}_j.$$

Let $D_n = \max_{i \leq n}[Var(w_i)Var(v_i)]$. Note that $\sum_{i \neq j} P_{ij} \tilde{w}_i \tilde{v}_j = \sum_{j<i} P_{ij}(\tilde{w}_i \tilde{v}_j + \tilde{w}_j \tilde{v}_i)$. Also,

$$E\left[\left(\sum_{j<i} P_{ij}^2 \tilde{w}_i \tilde{v}_j\right)^2\right] = \sum_{i=1}^n \sum_{s=1}^n \sum_{j<i} \sum_{t<s} P_{ij}^2 P_{st}^2 E\left[\tilde{w}_i \tilde{v}_j \tilde{w}_s \tilde{v}_t\right].$$

Note that in this sum, that $i \neq j$ and that if $i = t$ then $j \neq s$. Therefore the only nonzero terms are $i = s$ and $j = t$, so by CS and $\sum_i P_{ij}^2 = P_{jj} \leq 1$, implying $P_{ij}^2 \leq 1$,

$$E\left[\left(\sum_{j<i} P_{ij}^2 \tilde{w}_i \tilde{v}_j\right)^2\right] = \sum_{i=2}^n \sum_{j<i} P_{ij}^4 E\left[\tilde{w}_i^2\right] E\left[\tilde{v}_j^2\right] \leq D_n \sum_{i,j} P_{ij}^2 = D_n K.$$

Then by M, $\sum_{i \neq j} P_{ij}^2 \tilde{w}_i \tilde{v}_j = O_p(D_n^{1/2} K^{1/2})$. Let $\tilde{w} = (\tilde{w}_1, ..., \tilde{w}_n)'$. Note that for $\tilde{P} = [P_{ij}^2]_{i,j}$,

$$\sum_{i \neq j} P_{ij}^2 \tilde{w}_i \bar{v}_j = \bar{v}' \tilde{P} \tilde{w} - \sum_i P_{ii}^2 \tilde{w}_i \bar{v}_i.$$

By independence across i we have $E[\tilde{w}\tilde{w}'] \leq \max_{i \leq n} Var(w_i) I_n$, so that by $\sum_i P_{ik}^2 = P_{kk}$

$$\begin{aligned} E[(\bar{v}' \tilde{P} \tilde{w})^2] &= \bar{v}' \tilde{P} E[\tilde{w}\tilde{w}'] \tilde{P} \bar{v} \leq \max_{i \leq n} Var(w_i) \sum_{i,j,k} \bar{v}_i \bar{v}_j P_{ik}^2 P_{jk}^2 \\ &\leq D_n^2 \sum_k (\sum_i P_{ik}^2)(\sum_j P_{jk}^2) = D_n^2 \sum_k P_{kk}^2 \leq D_n^2 K, \\ E[\left(\sum_i P_{ii}^2 \tilde{w}_i \bar{v}_i\right)^2] &= \sum_i P_{ii}^4 E[\tilde{w}_i^2] \bar{v}_i^2 \leq D_n^2 K. \end{aligned}$$

Then by M we have $\bar{v}'\tilde{P}\tilde{w} = O_p(D_n^{1/2}K^{1/2})$ and $\sum_i P_{ii}^2 \tilde{w}_i \bar{v}_i = O_p(D_n^{1/2}K^{1/2})$, so by T it follows that

$$\sum_{i \neq j} P_{ij}^2 \tilde{w}_i \bar{v}_j = O_p(D_n^{1/2}K^{1/2}).$$

It follows similarly that $\sum_{i \neq j} P_{ij}^2 \bar{w}_i \tilde{v}_j = O_p(D_n^{1/2}K^{1/2})$, so the conclusion follows by T. Q.E.D.

References

Ackerberg, D.A. and P. Devereux (2003): "Improved JIVE Estimators for Overidentified Models with and without Heteroskedasticity," working paper, UCLA.

Angrist, J.D., G.W.Imbens, and A.Krueger (1999), "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14, 57-67.

Angrist, J. and A. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings", *Quarterly Journal of Economics 106*, 979–1014.

Bekker, P.A. (1994): "Alternative Approximations to the Distributions of Instrumental Variables Estimators," *Econometrica* 63, 657-681.

Bekker, P.A. and J. van der Ploeg (2005): "Instrumental variable estimation based on grouped data," *Statistica Neerlandica* 59, 239-267.

Blomquist, S. and M. Dahlberg (1999). "Small Sample Properties of LIML and Jackknife IV Estimators: Experiments with Weak Instruments," *Journal of Applied Econometrics* 14, 69-88.

Chao, J. and N. Swanson (2004): "Estimation and Testing Using Jackknife IV in Heteroskedastic Regressions With Many Weak Instruments," working paper.

Chao, J. and N. Swanson (2005): "Consistent Estimation With a Large Number of Weak Instruments," *Econometrica* 73, 1673-1692.

Davidson, R. and J.G. MacKinnon (2006): "The Case Against Jive," (with discussion and reply), *Journal of Applied Econometrics,* forthcoming.

Donald, S. G. and W. K. Newey (2001): "Choosing the Number of Instruments" *Econometrica* 69, 1161-1191.

Dufour, J.M. (1997): "Some Impossibility Theorems in Econometrics, With Applications to Structural and Dynamic Models," *Econometrica* 65, 1365 - 1388.

Fuller, W.A. (1977): "Some Properties of a Modification of the Limited Information Estimator," *Econometrica* 45, 939-954.

Hahn, J. and J. Hausman (2002) "A New Specification Test for the Validity of Instrumental Variables", *Econometrica 70,* 163-189.

Hahn, J., J.A. Hausman, and G.M. Kuersteiner (2004): "Estimation with Weak Instruments: Accuracy of higher-order bias and MSE approximations," *Econometrics Journal, Volume 7.*

Hahn, J. and A. Inoue (2002): "A Monte Carlo Comparison of Various Asymptotic Approximations to the Distribution of Instrumental Variables Estimators," *Econometric Reviews* 21, 309-336.

Han, C. and P.C.B. Phillips (2005). "GMM With Many Moment Conditions," *Econometrica,* forthcoming.

Hansen, C., J.A. Hausman, W.K. Newey (2005): "Estimation with Many Instrumental Variables," working paper, MIT.

Hansen, L.P., J. Heaton and A. Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics* 14, 262-280.

Kunitomo, N. (1980): "Asymptotic Expansions of Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of the American Statistical Association* 75, 693-700.

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.

Newey, W.K. and Windjmeier (2005): "GMM with Many Weak Moment Conditions Many Moments," working paper, MIT.

Phillips, G.D.A. and C. Hale (1977). "The Bias of Instrumental Variable Estimators of Simultaneous Equation Systems," *International Economic Review*, 18, 219-228.

Staiger, D. and J. Stock (1997): "Instrumental Variables Regression with Weak Instruments", *Econometrica 65*, 557-586.

Stock, J. and M. Yogo (2005): "Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments," Chapter 6, in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Andrews, D.W.K. and J.H. Stock eds., Cambridge: Cambride University Press.

White, H. (1982): "Instrumental Variables Estimation with Independent Observations," *Econometrica* 50, 483-500.