Combining Two Consistent Estimators*

John C. Chao, Department of Economics, University of Maryland, chao@econ.umd.edu

Jerry A Hausman, Department of Economics, MIT, jhausman@mit.edu

Whitney K. Newey, Department of Economics, MIT, wnewey@mit.edu

Norman R. Swanson, Department of Economics, Rutgers University, nswanson@econ.rutgers.edu

Tiemen Woutersen, Department of Economics, University of Arizona, woutersen@email.arizona.edu

April 2012

revised August 2012

JEL classification: C13, C31.

Keywords: endogeneity, instrumental variables, jackknife estimation, many moments, Hausman (1978) test.

----

*Prepared for Advances in Econometrics, Volume 29, Essays in Honor of Jerry Hausman. We all are indebted to Jerry for his advice, insights, wisdom, and wit. We thank Kei Hirano and James Fisher for helpful comments and discussions. Thanks are also owed to the editor, Carter Hill, as well as three anonymous referees for many useful comments on an earlier version of the paper.

Proposed Running Head: Combining Two Consistent Estimators

Corresponding Author:

Whitney K. Newey

Department of Economics

MIT, E52-262D

Cambridge, MA 02142-1347

Abstract

This paper shows how a weighted average of a forward and reverse Jackknife IV estimator (JIVE) yields estimators that are robust against heteroscedasticity and many instruments. These estimators, called HFUL (Heteroscedasticity robust Fuller) and HLIM (Heteroskedasticity robust limited information maximum likelihood (LIML)) were introduced by Hausman et al. (2012), but without derivation. Combining consistent estimators is a theme that is associated with Jerry Hausman and, therefore, we present this derivation in this volume. Additionally, and in order to further understand and interpret HFUL and HLIM in the context of jackknife type variance ratio estimators, we show that a new variant of HLIM, under specific grouped data settings with dummy instruments, simplifies to the Bekker and van der Ploeg (2005) MM (method of moments) estimator.

1 Introduction

One idea that is associated with Jerry Hausman is the idea of combining two estimators. For example, Hausman (1978) takes the difference between an efficient estimator and a robust estimator and derives the now famous Hausman test.[1] Another example is Hahn and Hausman (2002), who examine the difference between a forward and reverse Two Stage Least Squares (TSLS) estimator. This paper shows how a weighted average of a forward and reverse Jackknife IV estimator (JIVE) yields estimators that are robust against heteroscedasticity and many instruments. These estimators, called HFUL and HLIM were introduced by Hausman et al. (2012), but without derivation. Combining consistent estimators is a theme that is associated with Jerry Hausman and, therefore, we present this derivation in this volume.[2]

Jackknife IV estimators were proposed by Phillips and Hale (1977), Blomquist and Dahlberg (1999), Angrist et al. (1999), and Ackerberg and Deveraux (2009). In their paper "The Case Against JIVE," Davidson and MacKinnon (2006) show that JIVE performs poorly compared to LIML, due to the large dispersion of the JIVE. However, Chao et al. (2012) show that Jackknife IV (JIV) is consistent in a heteroscedastic and many instruments framework (unlike LIML). The HLIM and HFUL estimators discussed in this paper have less dispersion than the JIV estimator.

In addition, another result of this paper is that we show how modifying the numerator of the HLIM objective function yields a new estimator. In particular, we replace the numerator of the HLIM objective function with the objective function of so-called JIV1 estimator of Angrist et al. (1999) when forming our new estimator. In the special case that all instruments are dummy variables, this new estimator is the estimator proposed by Bekker and van der Ploeg (2005: BP). In closing, our recommendation is to use heteroskedasticity robust estimators. In this context, HFUL is particularly attractive because it is robust to many weak instruments and heteroskedasticity. In addition to these properties, it also has finite sample moments. Monte Carlo results supporting these findings are reported in Hausman et al. (2012).

In the remainder of this paper we provide a setup (in Section 2.1), interpret HFUL as a combination estimator (in Section 2.2), and compare HFUL with other variance ratio type estimators, such as the MM estimator of BP (2005) in Section 2.3.

[1] Hausman (1978) had received close to 9000 citations at the time of the writing of this paper (according to google scholar).

[2] More specifically, we present a derivation for HLIM. However, similar arguments and interpretations can also be made in the case of HFUL.

2 Deriving a Heteroscedasticity Robust Estimator

2.1 The Model and Previous Estimators

The model that we consider is given by

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

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 U is the matrix of reduced form disturbances. For our 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 an $n \times K$ matrix, Z, of instrumental variable observations. We will assume that $Z_1, ..., Z_n$ are nonrandom and that observations (ε_i, U_i) are independent across i and have mean zero, where the Z_i denote the i^{th} row (observation) of Z, ε_i is the i^{th} element of ε, and U_i is the transpose of the i^{th} row of U.

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 U being zero. The model also allows for Z to approximate the reduced form. For example, let X_i and Υ_i denote the i^{th} row (observation) for X and Υ,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, where $k = 1, ..., K$. In this case linear combinations of Z_i may approximate the unknown reduced form, e.g. as in Donald and Newey (2001).

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

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

The Fuller (1977) estimator (FULL) is obtained as

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

for $\breve{\alpha}^* = [\tilde{\alpha}^* - (1 - \tilde{\alpha}^*)C/n]/[1 - (1 - \tilde{\alpha}^*)C/n]$ and $\tilde{\alpha}^* = \hat{Q}(\tilde{\delta}^*)$, for C a positive constant. Under homoscedasticity, 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.[3] Both LIML and FULL are members of a class of estimators of the form

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

For example, LIML has this form for $\hat{\alpha}^* = \tilde{\alpha}^*$, FULL for $\hat{\alpha}^* = \breve{\alpha}^*$, and 2SLS for $\hat{\alpha}^* = 0$.

We use Fisher consistency to characterize the problem with heteroskedasticity and many instruments. Fisher consistency means that the derivative of the objective function at the truth converges to zero when normalized correctly. This condition is necessary for consistency. For expository purposes, consider first 2SLS, having objective function $(y - X\delta)'P(y - X\delta)$. The derivative of this objective function, times $-1/2n$, is equal to $X'P\varepsilon/n$. Like means, quadratic forms converge to their expectations under appropriate conditions. By virtue of independence and $E[\varepsilon_i] = 0$ we have

$$E[X_i P_{ij} \varepsilon_j] = E[X_i] P_{ij} E[\varepsilon_j] = 0, i \neq j,$$

where P_{ij} is the ij^{th} element of the projection matrix P. Thus,

$$\begin{aligned}
\frac{1}{n} X'P\varepsilon &= \frac{1}{n} E[X'P\varepsilon] + o_p(1) = \frac{1}{n} \sum_{i,j=1}^{n} E[X_i P_{ij} \varepsilon_j] + o_p(1) \\
&= \frac{1}{n} \sum_{i=1}^{n} E[X_i P_{ii} \varepsilon_i] + o_p(1) = \frac{1}{n} \sum_{i=1}^{n} E[U_i \varepsilon_i] P_{ii} + o_p(1),
\end{aligned}$$

so the Fisher consistency condition is $\sum_{i=1}^{n} E[U_i \varepsilon_i] P_{ii}/n \longrightarrow 0$. Because of many instruments,

$$P_{ii} \nrightarrow 0,$$

and hence 2SLS is not consistent, even under homoscedasticity, where $E[U_i \varepsilon_i]$ is constant over i. (Refer to Bekker (1994) for further details.)

For LIML, with objective function $\hat{Q}^*(\delta)$ given above, we have

$$(-\varepsilon'\varepsilon/2n) \partial \hat{Q}^*(\delta_0)/\partial \delta = \frac{1}{n} \left(X - \frac{X'\varepsilon}{\varepsilon'\varepsilon} \varepsilon \right)' P\varepsilon = \frac{1}{n} \left(X - \varepsilon \hat{\gamma}' \right)' P\varepsilon, \hat{\gamma} = X'\varepsilon/\varepsilon'\varepsilon.$$

Let $\sigma_i^2 = E[\varepsilon_i^2]$, $\gamma_i = E[X_i \varepsilon_i]/\sigma_i^2 = E[U_i \varepsilon_i]/\sigma_i^2$, and $\gamma^{(n)} = \sum_i E[X_i \varepsilon_i] / \sum_i \sigma_i^2 = \sum_i \sigma_i^2 \gamma_i / \sum_i \sigma_i^2$. By standard arguments $X'\varepsilon/\varepsilon'\varepsilon - \gamma^{(n)} \xrightarrow{p} 0$ and $\varepsilon'P\varepsilon/n$ is bounded in probability. Thus, similar to 2SLS,

$$\begin{aligned}
\frac{1}{n} \left(X - \varepsilon \hat{\gamma}' \right)' P\varepsilon &= \frac{1}{n} \left(X - \varepsilon \gamma^{(n)\prime} \right)' P\varepsilon + o_p(1) = \frac{1}{n} E[\left(X - \varepsilon \gamma^{(n)} \right)' P\varepsilon] + o_p(1) \\
&= \sum_{i=1}^{n} E[\left(X_i - \gamma^{(n)} \varepsilon_i \right) P_{ii} \varepsilon_i]/n + o_p(1) = \sum_{i=1}^{n} (\gamma_i - \gamma^{(n)}) P_{ii} \sigma_i^2/n + o_p(1).
\end{aligned}$$

[3] See Fuller (1977) for more detailed discussion.

It follows that, for LIML, the Fisher consistency condition is

$$\frac{1}{n}\sum_{i=1}^{n}(\gamma_i - \gamma^{(n)})P_{ii}\sigma_i^2 \longrightarrow 0.$$

There are two interesting cases where this condition holds.

A) γ_i does not vary with i : In this case, $\gamma_i = \gamma^{(n)}$ so that the term on the left is identically zero. Thus, homoscedasticity in the coefficient $\gamma_i = E[X_i\varepsilon_i]/E[\varepsilon_i^2]$ of the regression of X_i on ε_i leads to Fisher consistency of LIML.

B) P_{ii} does not vary with i : In this case $\sum_i(\gamma_i - \gamma^{(n)})P_{ii}\sigma_i^2 = P_{11}\sum_i(\gamma_i - \gamma^{(n)})\sigma_i^2 = 0$. When the instruments are dummy variables, sometimes referred to as grouping instruments, this condition is satisfied if all of the columns of Z have the same number of ones, i.e. the group sizes are equal. Bekker and van der Ploeg (2005) showed that this condition gives consistency of LIML for grouping instruments.[4]

In the general heteroscedastic case where γ_i and $P_{ii}\sigma_i^2$ are correlated across i, the Fisher consistency condition will not be satisfied and so LIML will not be consistent. Furthermore, due to weak instruments the bias may be large even when $\sum_i(\gamma_i - \gamma^{(n)})P_{ii}\sigma_i^2/n$ is small. Analogous arguments can also be used to show that, with heteroskedasticity, FULL and LIML are inconsistent under many instruments. BP (2005) and Hausman et al. (2012) point out that LIML can be inconsistent with heteroskedasticity but this appears to be the first characterization of Fisher consistency of LIML.

The lack of consistency of these estimators under many instruments and heteroskedasticity can be attributed to the presence of the $i = j$ terms in the double sums in their first order conditions. One way to make the estimators robust to heteroskedasticity is to remove these terms. A version, $\bar{\delta}$, of 2SLS without the $i = j$ terms solves the normal equations

$$0 = X'P(y - X\bar{\delta}) - \sum_{i=1}^{n} P_{ii}X_i(y_i - X_i'\bar{\delta}) = \sum_{i\neq j} X_iP_{ij}(y_j - X_j'\bar{\delta}).$$

[4]Note that P_{ii} does not converge to 0 under many instruments because the trace of $P = \sum_i P_{ii} = k$. This means that if k and n are approximately of the same order of magnitude, say under a reasonably balanced design (say, $k/n \to C$, for C a constant) then P_{ii} does not go to zero, as it is approximately of order k/n (assuming a reasonably balanced design).

Solving for $\bar{\delta}$ gives

$$\begin{aligned}\bar{\delta} &= \left(\sum_{i\neq j} X_i P_{ij} X_j'\right)^{-1} \sum_{i\neq j} X_i P_{ij} y_j \\ &= \left(X'PX - \sum_{i=1}^{n} P_{ii} X_i X_i'\right)^{-1} \left(X'Py - \sum_{i=1}^{n} P_{ii} X_i y_i\right).\end{aligned}$$

This is the second JIV estimator (JIV2) of Angrist et al. (1999). Because the normal equations remove the $i = j$ terms, this estimator is Fisher consistent. It was pointed out by Ackerberg and Deveraux (2009) and Chao et al. (2012) that it is consistent under many weak instruments and heteroskedasticity.

Under homoscedasticity and many weak instruments this estimator turns out to not be efficient. Also, Davidson and MacKinnon (2005) argue that it has inferior small sample properties under homoscedasticity, when compared with LIML. Using the weighted average of forward and reverse JIVE overcomes these problems.

2.2 Combining Forward and Reverse JIVE

The heteroskedasticity robust LIML estimator (HLIM) is obtained by dropping $i = j$ terms from the numerator of the LIML objective function,

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

Similar to JIV, $\tilde{\delta}$ will be consistent under heteroskedasticity because the $i = j$ terms have been removed from the normal equations. Here we will show consistency, asymptotic normality, and consistency of an asymptotic variance estimator.

As is the case with LIML, this estimator is invariant to normalization. Let $\bar{X} = [y, X]$. Then $\tilde{d} = (1, -\tilde{\delta}')'$ 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.

Also, computation of this estimator is straightforward. Similarly to LIML, $\tilde{\alpha} = \hat{Q}(\tilde{\delta})$ 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 $\tilde{\delta}$ are

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

Solving gives

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

This HLIM estimator has a similar form to LIML except that the $i=j$ terms have been deleted from the double sums.

It is interesting to note that LIML and HLIM coincide when P_{ii} is constant. In that case,

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

so that the LIML objective function equals the HLIM objective function plus a constant. This explains why constant P_{ii} will lead to LIML being consistent under heteroskedasticity.

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

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

The JIV2 estimator of Angrist et al. (1999) is obtained by setting $\hat{\alpha}=0$. Now, as shown in Hausman et al. (2012), under homoscedasticity and many weak instruments, HLIM is more efficient than JIV2. Moreover, we conjecture that HLIM is more efficient than any other estimator in this class, under homoscedasticity and many weak instruments. A heteroskedasticity consistent version of FULL is obtained by replacing $\tilde{\alpha}$ with $\hat{\alpha}=[\tilde{\alpha}-(1-\tilde{\alpha})C/n]/[1-(1-\tilde{\alpha})C/n]$, where C is a positive constant. The small sample properties of this estimator are unknown, but we expect its performance relative to HLIM to be similar to that of FULL relative to LIML. As pointed out by Hahn et al. (2004), FULL has much smaller dispersion than LIML with weak instruments, so we expect the same for HFUL. Monte Carlo results given in Hausman et al. (2012) confirm these properties.

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

$$\hat{H}=\sum_{i\neq j}X_iP_{ij}X_j'-\hat{\alpha}X'X,\hat{\Sigma}=\sum_{i,j=1}^{n}\sum_{k\notin\{i,j\}}\hat{X}_iP_{ik}\hat{\varepsilon}_k^2P_{kj}\hat{X}_j'+\sum_{i\neq j}P_{ij}^2\hat{X}_i\hat{\varepsilon}_i\hat{\varepsilon}_j\hat{X}_j'.$$

The variance estimator is

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

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

$$0 = -\frac{\partial \hat{Q}(\tilde{\delta})}{\partial \delta} \sum_i \tilde{\varepsilon}_i^2/2 = \sum_{i \neq j}(X_i - \tilde{\gamma}\tilde{\varepsilon}_i)P_{ij}(y_j - X_j'\tilde{\delta}) = \sum_{i \neq j}[(1+\tilde{\gamma}\tilde{\delta})X_i - \tilde{\gamma}y_i]P_{ij}(y_j - X_j'\tilde{\delta}).$$

The forward JIV estimator $\bar{\delta}$ is

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

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

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

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

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

Collecting terms in the first-order conditions for HLIM gives

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

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

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

Now, let δ_0 be the true value of δ, and define

$$\gamma^* = \sum_{i=1}^{n} E\left[X_i\varepsilon_i\right] / \sum_i E\left[\varepsilon_i^2\right],$$

Note that as $n \to \infty$, $\tilde{\delta} \xrightarrow{p} \delta_0$ and $\tilde{\gamma} - \gamma^* \xrightarrow{p} 0$; and, hence, expression (1) implies that

$$o_p(1) = (1+\gamma^*\delta_0)(\bar{\delta} - \tilde{\delta}) - \gamma^*\delta_0(\bar{\delta}^r - \tilde{\delta})$$

Rewriting the above equation, we obtain

$$\tilde{\delta} = (1 + \gamma^* \delta_0)\bar{\delta} - \gamma^* \delta_0 \bar{\delta}^r + o_p(1), \tag{2}$$

which shows that, at least for n large, $\tilde{\delta}$ can be written as a linear combination of forward and reverse JIV estimators. Moreover, if we were to assume error homoskedasticity, then it can be shown that under many weak instrument asymptotics that for $\delta_0 \neq 0$,

$$Var\left(\bar{\delta}^r\right) - Var\left(\bar{\delta}\right) = 2\frac{\sigma_{\varepsilon\varepsilon}^2}{\delta_0^2}\left(1 + 2\frac{\sigma_{X\varepsilon}}{\sigma_{\varepsilon\varepsilon}}\delta_0\right) = \frac{\sigma_{\varepsilon\varepsilon}^2}{\delta_0^2}\left(\frac{1}{2} + \gamma^* \delta_0\right),$$

where $Var\left(\bar{\delta}\right)$ and $Var\left(\bar{\delta}^r\right)$ denote the variances of the (many-weak-instrument) limiting distribution of $\bar{\delta}$ and of $\bar{\delta}^r$, respectively, and where, to simplify notations, we have let $\sigma_{\varepsilon\varepsilon} = E\left[\varepsilon_i^2\right]$ and $\sigma_{X\varepsilon} = E\left[X_i \varepsilon_i\right]$. It follows that $\bar{\delta}^r$ and $\bar{\delta}$ are equally efficient, i.e., $Var\left(\bar{\delta}^r\right) = Var\left(\bar{\delta}\right)$, if and only if

$$\gamma^* \delta_0 + \frac{1}{2} = 0.$$

Now, define $\theta = \gamma^* \delta_0 + (1/2)$, and we can rewrite (1) as

$$\tilde{\delta} = \left(\theta + \frac{1}{2}\right)\bar{\delta} - \left(\theta - \frac{1}{2}\right)\bar{\delta}^r + o_p(1). \tag{3}$$

Note that expression (3) shows that $\tilde{\delta}$ puts equal weight of 1/2 on both $\bar{\delta}$ and $\bar{\delta}^r$ when they are equally efficient but puts more weight on $\bar{\delta}$ when $\theta > 0$ (i.e., when the forward JIV is more efficient) and put more weight on $\bar{\delta}^r$ when $\theta < 0$ (i.e., when the reverse JIV is more efficient). This result is analogous to that of Hahn and Hausman (2002) where under homoscedasticity LIML is shown to be an optimal combination of forward and reverse bias corrected two stage least squares estimators.

Finally, if we replace $\tilde{\gamma}$ in (1) above by some other estimator $\bar{\gamma}$ and the $\tilde{\gamma}\bar{\delta}$ coefficient following the minus sign by $\bar{\gamma}\tilde{\delta}$ 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 HLIM and the HFUL estimator.

2.3 Comparing Variants of HLIM and HFUL with the MM Estimator of Bekker and van der Ploeg

BP (2005) considered estimators that are consistent with dummy instruments and group heteroskedasticity. A particular interesting estimator which BP (2005) propose is referred to as the

MM estimator. It turns out that this MM estimator is a special case of a type of jackknifed LIML estimator, where the numerator quadratic form corresponds to the objective function of JIV1 instead of that of JIV2 (as in HLIM). More specifically, consider an estimator which minimizes the following modified variance ratio

$$\begin{aligned} Q_{JLIM}(\delta) &= \frac{(y - X\delta)'(P - D_P)(I_n - D_P)^{-1}(y - X\delta)}{(y - X\delta)' MD_P (I_n - D_P)^{-1}(y - X\delta)} \\ &= \frac{Q_{JIV1}(\delta)}{(y - X\delta)' MD_P (I_n - D_P)^{-1}(y - X\delta)}, \end{aligned} \tag{4}$$

where $M = I_n - P$ and $D_P = diag(P_{11}, ..., P_{nn})$. Note that the numerator of (4) is simply the objective function of JIV1, since minimizing $Q_{JIV1}(\delta)$ with respect to δ leads to the estimator

$$\begin{aligned} \widehat{\delta}_{JIV1} &= \left(X'(P - D_P)(I_n - D_P)^{-1} X\right)^{-1} X'(P - D_P)(I_n - D_P)^{-1} y \\ &= \left(\sum_{i \neq j} X_i P_{ij} (1 - P_{jj})^{-1} X_j'\right)^{-1} \sum_{i \neq j} X_i P_{ij} (1 - P_{jj})^{-1} y_j, \end{aligned}$$

which is the jackknife IV estimator originally proposed by Phillips and Hale (1977). See Chao et al. (2012) for further discussion.

It is also possible to rewrite the objective function (4) in an alternative form which will be more convenient for the purpose of establishing a correspondence with the results of BP (2005). To proceed, note first that, by elementary algebraic manipulations, it is easy to show that

$$(P - D_P)(I_n - D_P)^{-1} = P - MD_P (I_n - D_P)^{-1}.$$

Hence, we can rewrite (4) as

$$\begin{aligned} Q_{JLIM}(\delta) &= \frac{(y - X\delta)'\left[P - MD_P (I_n - D_P)^{-1}\right](y - X\delta)}{(y - X\delta)' MD_P (I_n - D_P)^{-1}(y - X\delta)} \\ &= \frac{(y - X\delta)' P (y - X\delta)}{(y - X\delta)' MD_P (I_n - D_P)^{-1}(y - X\delta)} - 1 \\ &= Q^*_{JLIM}(\delta) - 1 \end{aligned}$$

so that the estimator which minimizes $Q_{JLIM}(\delta)$ is clearly the same as the one which minimizes

$$Q^*_{JLIM}(\delta) = \frac{(y - X\delta)' P (y - X\delta)}{(y - X\delta)' MD_P (I_n - D_P)^{-1}(y - X\delta)}. \tag{5}$$

To show that the estimator obtained from (5) specializes to the MM estimator of BP (2005) with dummy-variable instruments and across-group heteroskedasticity, consider the grouped data IV regression model studied in their paper, which takes the form

$$\begin{aligned} y_{ij} &= x_{ij}'\delta + \varepsilon_{ij}, \\ x_{ij} &= \pi_j + v_{ij} \end{aligned}$$

for $i = 1, ..., n_j$ and $j = 1, ..., m$. Stacking first the observations within each group, we obtain (for $j = 1, ..., m$)

$$y_j = \begin{pmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{n_j,j} \end{pmatrix}, X_j = \begin{pmatrix} x_{1j}' \\ x_{2j}' \\ \vdots \\ x_{n_j,j}' \end{pmatrix}, \ \varepsilon_j = \begin{pmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \vdots \\ \varepsilon_{n_j,j} \end{pmatrix}, \ V_j = \begin{pmatrix} v_{1j}' \\ v_{2j}' \\ \vdots \\ v_{n_j,j}' \end{pmatrix},$$

and also let

$$Z_j = \iota_{n_j} e_{j,m}'$$

where $\iota_{n_j} = (1, 1, ..., 1)'$ is an $(n_j \times 1)$ vector of ones and $e_{j,m}$ is the j^{th} column of a $m \times m$ identity matrix. Using these notations, it is easily seen that the IV model studied in BP (2005) can be written in our notations as

$$y = X\delta + \varepsilon, \tag{6}$$

$$X = Z\Pi + V, \tag{7}$$

where $y = (y_1', y_2', ..., y_m')'$, $X = (X_1', X_2', ..., X_m')'$, $Z = (Z_1', Z_2', ..., Z_m')'$, $\Pi = (\pi_1, \pi_2, ..., \pi_m)'$, $\varepsilon = (\varepsilon_1', \varepsilon_2', ..., \varepsilon_m')'$, and $V = (V_1', V_2', ..., V_m')'$ with the components of these vectors and matrices being as defined above. Moreover, with dummy-variable instruments as considered in BP (2005), we have

$$\begin{aligned} P &= Z\left(Z'Z\right)^{-1}Z' \\ &= \begin{pmatrix} n_1^{-1}\iota_{n_1}\iota_{n_1}' & 0 & \cdots & 0 \\ 0 & n_2^{-1}\iota_{n_2}\iota_{n_2}' & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & n_m^{-1}\iota_{n_m}\iota_{n_m}' \end{pmatrix} = \begin{pmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & P_m \end{pmatrix}, \end{aligned}$$

$$D_P = \begin{pmatrix} n_1^{-1}I_{n_1} & 0 & \cdots & 0 \\ 0 & n_2^{-1}I_{n_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & n_m^{-1}I_{n_m} \end{pmatrix}$$

and

$$\begin{aligned} M &= I_n - P \\ &= \begin{pmatrix} M_1 & 0 & \cdots & 0 \\ 0 & M_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & M_m \end{pmatrix}, \end{aligned}$$

where $n = \sum_{j=1}^{m} n_j$. Hence, in this setting, (5) specializes to the MM objective function given in equation (13) of BP (2005). Namely,

$$\begin{aligned} Q^*_{JLIM}(\delta) &= \frac{(y - X\delta)' P (y - X\delta)}{(y - X\delta)' M D_P (I_n - D_P)^{-1} (y - X\delta)} \\ &= \frac{\sum_{j=1}^{m} (y_j - X_j\delta)' P_j (y_j - X_j\delta)}{\sum_{j=1}^{m} (y_j - X_j\delta)' \left[n_j^{-1} / \left(1 - n_j^{-1}\right)\right] M_j (y_j - X_j\delta)} \\ &= \frac{\sum_{j=1}^{m} (y_j - X_j\delta)' P_j (y_j - X_j\delta)}{\sum_{j=1}^{m} (y_j - X_j\delta)' (n_j - 1)^{-1} M_j (y_j - X_j\delta)} \\ &= \frac{n \sum_{j=1}^{m} \delta'_\Delta w_j \overline{Y}_j \overline{Y}'_j \delta_\Delta}{\sum_{j=1}^{m} \delta'_\Delta S_j \delta_\Delta} \\ &= n Q_{MM}(\delta) \end{aligned} \tag{8}$$

where $\delta_\Delta = (1, -\delta')'$, $w_j = n_j / \sum_{j=1}^{m} n_j = n_j/n$,

$$\begin{aligned} \overline{Y}_j &= \frac{1}{n_j} \sum_{i=1}^{n_j} \left(y_{ij}, x'_{ij}\right)', \\ S_j &= \frac{1}{n_j - 1} \sum_{i=1}^{n_j} \left(\begin{pmatrix} y_{ij} \\ x_{ij} \end{pmatrix} - \overline{Y}_j\right) \left(\begin{pmatrix} y_{ij} \\ x_{ij} \end{pmatrix} - \overline{Y}_j\right)' \end{aligned}$$

as defined in their paper. Note a slight difference in our notations and that of BP (2005) in that we set

$$M_j = I_{n_j} - P_j = I_{n_j} - n_j^{-1} \iota_{n_j} \iota'_{n_j}$$

here, whereas

$$M_j = w_j \overline{Y}_j \overline{Y}'_j$$

in the notations of BP (2005).

Now, given that the numerator of (4) corresponds to the objective function of JIV1, we can show, following arguments similar to that used for proving Theorems 1 and 2 of Hausman et al. (2012), that the estimator which minimizes $Q_{JLIM}(\delta)$ is consistent and asymptotically normal under many instrument and many weak instrument asymptotics. Indeed, in the special case with dummy instruments and across-group heteroskedasticity, BP (2005) have already shown that the MM estimator is consistent and asymptotically normal under group-asymptotics which takes $m \to \infty$ holding each n_j fixed, for $j = 1, ..., m$. However, it should be noted that the MM estimator will not in general be consistent under group asymptotics in the presence of more general instruments which are not group indicators. To analyze more fully situations where the MM estimator may or may not be consistent, we consider the following slight generalization of the setup studied in BP (2005). Namely, consider

$$\begin{aligned} y_{ij} &= x'_{ij}\delta + \varepsilon_{ij}, \\ \underset{G\times 1}{x_{ij}} &= \underset{1\times 1}{z_{ij}}\ \underset{G\times 1}{\pi_j} + \underset{G\times 1}{v_{ij}}, \end{aligned}$$

for $j = 1, ..., m$ and $i = 1, ..., n_j$, so that we allow for instruments which are possibly not group indicators. Now, from the previous discussion, it is apparent that the MM estimator can be equivalently obtained by minimizing the alternative objective function

$$Q^*_{MM}(\delta) = \frac{\sum_{j=1}^{m} (y_j - X_j\delta)' \left[P_j - (n_j - 1)^{-1} M_j\right] (y_j - X_j\delta)}{\sum_{j=1}^{m} (y_j - X_j\delta)' (n_j - 1)^{-1} M_j (y_j - X_j\delta)},$$

where $P_j = z_{\cdot j}\left(z'_{\cdot j} z_{\cdot j}\right)^{-1} z'_{\cdot j}$ with $z_{\cdot j} = \left(z_{1j}, z_{2j}, .., z_{n_j,j}\right)'$ and $M_j = I_{n_j} - P_j$. All other notations are as defined previously. Now, in large samples, this objective function will be close to

$$\overline{Q}_{MM}(\delta) = \frac{\sum_{j=1}^{m} E\left\{(y_j - X_j\delta)' \left[P_j - (n_j - 1)^{-1} M_j\right] (y_j - X_j\delta)\right\}}{\sum_{j=1}^{m} E\left\{(y_j - X_j\delta)' (n_j - 1)^{-1} M_j (y_j - X_j\delta)\right\}}.$$

To gain some insight into conditions under which the limiting objective function $\overline{Q}_{MM}(\delta)$ may be minimized at $\delta = \delta_0$, we define $P_{j,ii}$ to be the i^{th} diagonal element of P_j and let

$$\begin{aligned} \sigma^2(i,j) &= E\left[\varepsilon_{ij}^2\right],\ \vartheta(i,j) = E\left[x_{ij}\varepsilon_{ij}\right], \varpi(i,j) = \frac{1}{n_j - 1}\left(1 - P_{j,ii}\right), \\ W(i,j) &= \frac{P_{j,ii} - \varpi(i,j)}{\varpi(i,j)},\ \pi(i,j) = \frac{\varpi(i,j)}{m}; \end{aligned}$$

Note that, by straightforward calculations, we obtain

$$
\begin{aligned}
& \left.\frac{\partial}{\partial \delta} \overline{Q}_{MM}(\delta)\right|_{\delta=\delta_0} \\
= \; & \frac{-2 \sum_{j=1}^{m} E\left(X_j' \left[P_j - (n_j - 1)^{-1} M_j\right] \varepsilon_j\right)}{\sum_{j=1}^{m} (n_j - 1)^{-1} E\left(\varepsilon_j' M_j \varepsilon_j\right)} \\
& + 2 \frac{\sum_{j=1}^{m} E\left(\varepsilon_j' \left[P_j - (n_j - 1)^{-1} M_j\right] \varepsilon_j\right)}{\left[\sum_{j=1}^{m} (n_j - 1)^{-1} E\left(\varepsilon_j' M_j \varepsilon_j\right)\right]^2} \sum_{j=1}^{m} \left(\frac{1}{n_j - 1}\right) E\left(X_j' M_j \varepsilon_j\right) \\
= \; & \frac{-2}{\sum_{j=1}^{m} \sum_{i=1}^{n_j} \sigma^2(i,j)\, \pi(i,j)} \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n_j} W(i,j) \left(\vartheta(i,j) - \sigma^2(i,j) \frac{\sum_{j=1}^{m} \sum_{i=1}^{n_j} \vartheta(i,j)\, \pi(i,j)}{\sum_{j=1}^{m} \sum_{i=1}^{n_j} \sigma^2(i,j)\, \pi(i,j)} \right) \pi(i,j) \right\} \\
= \; & -2 \left(\widehat{E}\left[\sigma^2(i,j)\right]\right)^{-1} \widehat{E}\left[W(i,j) \left\{ \vartheta(i,j) - \frac{\widehat{E}\left[\vartheta(i,j)\right]}{\widehat{E}\left[\sigma^2(i,j)\right]} \sigma^2(i,j) \right\} \right] \\
= \; & -2 \left(\widehat{E}\left[\sigma^2(i,j)\right]\right)^{-1} \widehat{E}\left[W(i,j)\, \psi(i,j)\right] \qquad (9)
\end{aligned}
$$

where, in the expressions above, we have taken

$$
\begin{aligned}
\widehat{E}\left[\vartheta(i,j)\right] &= \sum_{j=1}^{m} \sum_{i=1}^{n_j} \vartheta(i,j)\, \pi(i,j)\,, \;\; \widehat{E}\left[\sigma^2(i,j)\right] = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \sigma^2(i,j)\, \pi(i,j)\,, \\
\psi(i,j) &= \vartheta(i,j) - \frac{\widehat{E}\left[\vartheta(i,j)\right]}{\widehat{E}\left[\sigma^2(i,j)\right]} \sigma^2(i,j)\,, \;\; \widehat{E}\left[W(i,j)\, \psi(i,j)\right] = \sum_{j=1}^{m} \sum_{i=1}^{n_j} W(i,j)\, \psi(i,j)\, \pi(i,j)\,.
\end{aligned}
$$

Now, we can interpret $W(i,j)$, $\sigma^2(i,j)$, and $\psi(i,j)$ as functions of the discrete random variables (indices) i and j, which have joint probability mass distribution (pmf) given by $\pi(i,j)$. Moreover, interpret

$$
\pi(j) = \sum_{i=1}^{n_j} \pi(i,j) = \sum_{i=1}^{n_j} \frac{\varpi(i,j)}{m} = \frac{1}{m} \sum_{i=1}^{n_j} \frac{1}{n_j - 1} \left(1 - P_{j,ii}\right) = \frac{1}{m}
$$

as the marginal pmf of j and define

$$
\pi(i \mid j) = \frac{\pi(i,j)}{\pi(j)} = \varpi(i,j)
$$

to be the conditional pmf of i given j. Furthermore, observe that

$$
\begin{aligned}
\widehat{E}\left[W(i,j) \mid j\right] &= \sum_{i=1}^{n_j} W(i,j)\, \pi(i \mid j) = \sum_{i=1}^{n_j} \left[P_{j,ii} - (n_j - 1)^{-1} \left(1 - P_{j,ii}\right)\right] = 0, \\
\widehat{E}\left[W(i,j)\right] &= \widehat{E}\left[\widehat{E}\left[W(i,j) \mid j\right]\right] = 0 \;\; \text{(by law of iterated expectations)},
\end{aligned}
$$

so that, in particular,

$$\widehat{E}\left[W\left(i,j\right)\psi\left(i,j\right)\right]=\widehat{Cov}\left(W\left(i,j\right),\psi\left(i,j\right)\right)+\widehat{E}\left[W\left(i,j\right)\right]\widehat{E}\left[\psi\left(i,j\right)\right]=\widehat{Cov}\left(W\left(i,j\right),\psi\left(i,j\right)\right).$$

Hence, we can rewrite (9) as

$$\left.\frac{\partial}{\partial\delta}\overline{Q}_{MM}\left(\delta\right)\right|_{\delta=\delta_0}=-2\left(\widehat{E}\left[\sigma^2\left(i,j\right)\right]\right)^{-1}\widehat{Cov}\left(W\left(i,j\right),\psi\left(i,j\right)\right) \tag{10}$$

from which it follows that $\delta=\delta_0$ is a critical point of $\overline{Q}_{MM}\left(\delta\right)$ if and only if

$$\widehat{Cov}\left(W\left(i,j\right),\psi\left(i,j\right)\right)=0,$$

since

$$\widehat{E}\left[\sigma^2\left(i,j\right)\right]=\sum_{j=1}^{m}\sum_{i=1}^{n_j}\sigma^2\left(i,j\right)\pi\left(i,j\right)\leq C\sum_{j=1}^{m}\sum_{i=1}^{n_j}\pi\left(i,j\right)=C<\infty,$$

holds under a condition that the second moments of $\{\varepsilon_{ij}\}$ are uniformly bounded.

Next, consider the situation where error variance is homoskedastic within-group, but there may be heteroskedasticity across groups, i.e., for each j, $\sigma^2\left(i,j\right)=\sigma^2\left(j\right)$ and $\vartheta\left(i,j\right)=\vartheta\left(j\right)$ so that $\psi\left(i,j\right)=\psi\left(j\right)$. Here, by the law of iterated expectations

$$\widehat{Cov}\left(W\left(i,j\right),\psi\left(i,j\right)\right)=\widehat{E}\left[W\left(i,j\right)\psi\left(j\right)\right]=\widehat{E}_j\left[\psi\left(j\right)\widehat{E}\left[W\left(i,j\right)\ \mid j\right]\right]=0,$$

so that $\delta=\delta_0$ is a critical point of $\overline{Q}_{MM}\left(\delta\right)$ in this case. This suggests that, in the absence of within-group heteroskedasticity, the MM estimator will be consistent even in situations where the available instruments are not group indicators. On the other hand, suppose that instruments are group indicators as assumed in BP (2005); then, $P_{j,ii}=1/n_j$ for $i=1,...,n_j$; and it follows that

$$W\left(i,j\right)=P_{j,ii}-\left(n_j-1\right)^{-1}\left(1-P_{j,ii}\right)=\frac{1}{n_j}-\frac{1}{n_j-1}\left(1-\frac{1}{n_j}\right)=0$$

for $i=1,..,n_j$ and for each j, so that, trivially,

$$\widehat{Cov}\left(W\left(i,j\right),\psi\left(i,j\right)\right)=0.$$

Note that this is true even with within-group heteroskedasticity. Finally, given the possibility of within-group heteroskedasticity and instruments which are not group indicators; we obtain in general

$$
\begin{aligned}
& \widehat{Cov}\left(W\left(i,j\right),\psi\left(i,j\right)\right) \\
= \;& \widehat{E}\left[W\left(i,j\right)\psi\left(i,j\right)\right] \\
= \;& \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(\frac{P_{j,ii}-\varpi\left(i,j\right)}{\varpi\left(i,j\right)}\right)\psi\left(i,j\right)\pi\left(i,j\right) \\
= \;& \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left[P_{j,ii}-\left(n_j-1\right)^{-1}\left(1-P_{j,ii}\right)\right]\psi\left(i,j\right) \\
= \;& \frac{1}{m}\sum_{j=1}^{m}\left(\frac{n_j}{n_j-1}\right)\sum_{i=1}^{n_j}\left\{P_{j,ii}-\left(\frac{1}{n_j}\right)\right\}\psi\left(i,j\right)\neq 0
\end{aligned}
$$

Hence, $\delta=\delta_0$ may not be a critical point of $\overline{Q}_{MM}\left(\delta\right)$ in more general settings; and it can be shown that, unlike HLIM, the MM estimator will not be consistent when both non-dummy instruments and within-group heteroskedasticity are present, although a generalization of the MM estimator obtained by minimizing (4) will be.

It is also of interest to compare the asymptotic distribution of the MM estimator to that of HLIM in the setting studied by BP (2005), i.e., in the setting with across-group heteroskedasticity (but within-group homoscedasticity) and dummy-variable instruments. Under group asymptotics, it can be shown, following the same argument as that used to prove Theorem 2 of Hausman et al. (2012), that

$$
V_{MM}^{-1/2}\left(\widehat{\delta}_{MM}-\delta_0\right)\xrightarrow{d}N\left(0,I_G\right),\text{ as } m\rightarrow\infty,
$$

where

$$
V_{MM}=H_{MM}^{-1}\left(\overline{\Omega}_{MM}+\Psi_{MM}\right)H_{MM}^{-1} \tag{11}
$$

and where

$$
\begin{aligned}
H_{MM} &= \frac{1}{m}\sum_{j=1}^{m}n_j\pi_j\pi_j',\ \overline{\Omega}_{MM}=\frac{1}{m}\sum_{j=1}^{m}n_j\sigma_j^2\pi_j\pi_j', \\
\Psi_{MM} &= \frac{1}{m}\sum_{j=1}^{m}\sum_{1\leq\ i\neq k\leq\ n_j}\left(\frac{1}{n_j-1}\right)^2\left\{E\left[\varepsilon_{ij}^2\right]E\left[\overline{V}_{kj}\overline{V}_{kj}'\right]+E\left[\overline{V}_{ij}\varepsilon_{ij}\right]E\left[\varepsilon_{kj}\overline{V}_{kj}'\right]\right\} \\
&= \frac{1}{m}\sum_{j=1}^{m}\left(\frac{n_j}{n_j-1}\right)\Xi_{MM,j},
\end{aligned}
$$

Here, $\overline{V}_{ij} = v_{ij} - \left(\overline{\sigma}_{21}/\overline{\sigma}^2\right)\varepsilon_{ij}$, with $\overline{\sigma}_{21} = m^{-1}\sum_{j=1}^{m} E\left[v_{ij}\varepsilon_{ij}\right]$ and $\overline{\sigma}^2 = m^{-1}\sum_{j=1}^{m} E\left[\varepsilon_{ij}^2\right]$. Also, $\Xi_{MM,j} = E\left[\varepsilon_{ij}^2\right] E\left[\overline{V}_{kj}\overline{V}'_{kj}\right] + E\left[\overline{V}_{ij}\varepsilon_{ij}\right] E\left[\varepsilon_{kj}\overline{V}'_{kj}\right]$, where the dependence of $\Xi_{MM,j}$ on the index j only is due to the fact that we have within-group homoscedasticity. On the other hand, for HLIM, we obtain

$$V_{HLIM}^{-1/2}\left(\widehat{\delta}_{HLIM} - \delta_0\right) \xrightarrow{d} N\left(0, I_G\right), \text{ as } m \to \infty,$$

where

$$V_{HLIM} = H_{HLIM}^{-1}\left(\overline{\Omega}_{HLIM} + \Psi_{HLIM}\right) H_{HLIM}^{-1} \tag{12}$$

and where

$$\begin{aligned}
H_{HLIM} &= \frac{1}{m}\sum_{j=1}^{m}\left(n_j - 1\right)\pi_j\pi_j',\ \overline{\Omega}_{HLIM} = \frac{1}{m}\sum_{j=1}^{m}\frac{\left(n_j-1\right)^2}{n_j}\sigma_j^2\pi_j\pi_j',\\
\Psi_{HLIM} &= \frac{1}{m}\sum_{j=1}^{m}\sum_{1\leq\ i\neq k\leq\ n_j}\left(\frac{1}{n_j}\right)^2\left\{E\left[\varepsilon_{ij}^2\right] E\left[\widetilde{V}_{kj}\widetilde{V}'_{kj}\right] + E\left[\widetilde{V}_{ij}\varepsilon_{ij}\right] E\left[\varepsilon_{kj}\widetilde{V}'_{kj}\right]\right\}\\
&= \frac{1}{m}\sum_{j=1}^{m}\left(\frac{n_j-1}{n_j}\right)\Xi_{HLIM,j},
\end{aligned}$$

with $\widetilde{V}_{ij} = v_{ij} - \left(\overline{\overline{\sigma}}_{21}/\overline{\overline{\sigma}}^2\right)\varepsilon_{ij}$, $\overline{\overline{\sigma}}_{21} = \left(\sum_{j=1}^{m} n_j\right)^{-1}\sum_{j=1}^{m} n_j E\left[v_{ij}\varepsilon_{ij}\right]$ and $\overline{\overline{\sigma}}^2 = \left(\sum_{j=1}^{m} n_j\right)^{-1}\sum_{j=1}^{m} n_j E\left[\varepsilon_{ij}^2\right]$; and $\Xi_{HLIM,j} = E\left[\varepsilon_{ij}^2\right] E\left[\widetilde{V}_{kj}\widetilde{V}'_{kj}\right] + E\left[\widetilde{V}_{ij}\varepsilon_{ij}\right] E\left[\varepsilon_{kj}\widetilde{V}'_{kj}\right]$.

Comparing (11) with (12) for the case where $n_j = n_*$ for all j, i.e., for the case where the size of the sample is the same for all groups; we see that

$$\begin{aligned}
\overline{\overline{\sigma}}_{21} &= \left(\sum_{j=1}^{m} n_j\right)^{-1}\sum_{j=1}^{m} n_j E\left[v_{ij}\varepsilon_{j,i}\right] = \frac{1}{mn_*}\sum_{j=1}^{m} n_*\sigma_{21,j} = \frac{1}{m}\sum_{j=1}^{m}\sigma_{21,j} = \overline{\sigma}_{21},\\
\overline{\overline{\sigma}}^2 &= \left(\sum_{j=1}^{m} n_j\right)^{-1}\sum_{j=1}^{m} n_j E\left[\varepsilon_{j,i}^2\right] = \frac{1}{mn_*}\sum_{j=1}^{m} n_*\sigma_j^2 = \frac{1}{m}\sum_{j=1}^{m}\sigma_j^2 = \overline{\sigma}_{21},
\end{aligned}$$

so that

$$\widetilde{V}_{ij} = v_{ij} - \left(\overline{\overline{\sigma}}_{21}/\overline{\overline{\sigma}}^2\right)\varepsilon_{ij} = v_{ij} - \left(\overline{\sigma}_{21}/\overline{\sigma}^2\right)\varepsilon_{ij} = \overline{V}_{ij},$$

and, thus,

$$\begin{aligned}
\Xi_{MM,j} &= E\left[\varepsilon_{ij}^2\right] E\left[\overline{V}_{kj}\overline{V}'_{kj}\right] + E\left[\overline{V}_{ij}\varepsilon_{ij}\right] E\left[\varepsilon_{kj}\overline{V}'_{kj}\right]\\
&= E\left[\varepsilon_{ij}^2\right] E\left[\widetilde{V}_{kj}\widetilde{V}'_{kj}\right] + E\left[\widetilde{V}_{ij}\varepsilon_{ij}\right] E\left[\varepsilon_{kj}\widetilde{V}'_{kj}\right]\\
&= \Xi_{HLIM,j}\\
&= \Xi_j \quad (say),
\end{aligned}$$

for $j = 1, ..., m$ and $i, k = 1, ..., n_j$. It follows that in this case

$$
\begin{aligned}
& V_{MM} \\
= \; & H_{MM}^{-1}\left(\overline{\Omega}_{MM} + \Psi_{MM}\right) H_{MM}^{-1} \\
= \; & \left(\frac{1}{m}\sum_{j=1}^{m} n_* \pi_j \pi_j'\right)^{-1} \left[\frac{1}{m}\sum_{j=1}^{m} n_* \sigma_j^2 \pi_j \pi_j' + \frac{1}{m}\sum_{j=1}^{m}\left(\frac{n_*}{n_*-1}\right)\Xi_j\right] \left(\frac{1}{m}\sum_{j=1}^{m} n_* \pi_j \pi_j'\right)^{-1} \\
= \; & \left(\frac{1}{m}\sum_{j=1}^{m} \pi_j \pi_j'\right)^{-1} \left[\frac{1}{mn_*}\sum_{j=1}^{m} \sigma_j^2 \pi_j \pi_j' + \frac{1}{mn_*}\sum_{j=1}^{m}\left(\frac{1}{n_*-1}\right)\Xi_j\right] \left(\frac{1}{m}\sum_{j=1}^{m} \pi_j \pi_j'\right)^{-1} \\
= \; & \left(\frac{1}{m}\sum_{j=1}^{m} \left(n_*-1\right) \pi_j \pi_j'\right)^{-1} \left[\frac{1}{m}\sum_{j=1}^{m} \frac{\left(n_*-1\right)^2}{n_*}\sigma_j^2 \pi_j \pi_j' + \frac{1}{m}\sum_{j=1}^{m}\left(\frac{n_*-1}{n_*}\right)\Xi_j\right] \left(\frac{1}{m}\sum_{j=1}^{m} \left(n_*-1\right) \pi_j \pi_j'\right)^{-1} \\
= \; & H_{HLIM}^{-1}\left(\overline{\Omega}_{HLIM} + \Psi_{HLIM}\right) H_{HLIM}^{-1} \\
= \; & V_{HLIM}
\end{aligned}
$$

Hence, in the special case where the number of observations are the same in each group, the asymptotic covariance matrix of the MM estimator is equivalent to that of HLIM. However, in the more general case where n_j varies with j, the covariance matrices are not equal, but it does not appear that there is a uniform ranking of the two estimators in terms of asymptotic efficiency since the relative "size" of the covariance matrices in this case will depend on the values of the underlying parameters. This is analogous to the result obtained in Chao et al. (2012), where JIV1 and JIV2 are not found to dominate each other in terms of asymptotic efficiency under many-instrument and many-weak-instrument asymptotics.

3 References

Ackerberg, D.A. and P. Deveraux, (2009), Improved JIVE estimators for overidentified models with and without heteroskedasticity. Review of Economics and Statistics 91, 351-362.

Angrist, J.D., G.W. Imbens, and A. Krueger, (1999), Jackknife instrumental variables estimation. Journal of Applied Econometrics 14, 57-67.

Bekker, P.A., (1994), Alternative approximations to the distributions of instrumental variable estimators. Econometrica 62, 657-681.

Bekker, P. A. and J. van der Ploeg, (2005), Instrumental variable estimation based on grouped data. Statistica Neerlandica 59, 506-508.

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.C., N.R. Swanson, J.A. Hausman, W.K. Newey, and T. Woutersen, (2012), Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments, Econometric Theory 28, 42-86.

Davidson, R. and J.G. MacKinnon, (2006), The case against JIVE (with discussion and reply). Journal of Applied Econometrics 21, 827-833.

Donald, S.G. and W.K Newey, (2001), Choosing the number of instruments. Econometrica 69, 1161-1191.

Fuller, W.A., (1977), Some properties of a modification of the limited information estimator. Econometrica 45, 939-954.

Hahn, J. and J.A. Hausman, (2002), A new specification test for the validity of instrumental variables. Econometrica 70, 163-189.

Hahn, J., J.A. Hausman and G. Kuersteiner, (2004), Estimation with weak instruments: accuracy of higher order bias and MSE approximations. Econometrics Journal, 7 (2004), 272-306.

Hausman, J.A., (1978), Specification tests in econometrics. Econometrica 46, 1251-71.

Hausman, J.A., W.K. Newey, T. Woutersen, J. Chao, and N.R. Swanson, (2012), IV estimation with heteroskedasticity and many instruments. Quantitative Economics, forthcoming.

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.

Phillips, G.D.A. and C. Hale, (1977), The bias of instrumental variable estimators of simultaneous equation systems. International Economic Review 18, 219-228.