Banco Central de Chile
Documentos de Trabajo

Central Bank of Chile
Working Papers

N° 464

Abril 2008

# INFERENCE USING INSTRUMENTAL VARIABLE ESTIMATORS

Rodrigo Alfaro

**BANCO CENTRAL DE CHILE**

**CENTRAL BANK OF CHILE**



Documentos de Trabajo del Banco Central de Chile
Working Papers of the Central Bank of Chile
Agustinas 1180
Teléfono: (56-2) 6702475; Fax: (56-2) 6702231

**Documento de Trabajo**
**N° 464**

**Working Paper**
**N° 464**

# INFERENCE USING INSTRUMENTAL VARIABLE ESTIMATORS

Rodrigo Alfaro
Gerencia de Estabilidad Financiera
Banco Central de Chile

**Resumen**

Este documento estudia las propiedades de estimadores de Variables Instrumentales, para los casos en que los errores son heterocedásticos y se utilizan muchos instrumentos. En particular, se compara el desempeño del estimador propuesto por Hausman, Newey, Woutersen, Chao, y Swanson (2007) con la versión robusta de JIVE, propuesto por Angrist, Imbens y Krueger (1999). Se presentan resultados teóricos para el test t que considera heterocedasticidad, encontrándose que el mayor efecto está relacionado al sesgo en muestras finitas del estimador.

**Abstract**

This paper studies inference performance of Instrumental Variables Estimators in situations where error terms are heteroskedastic and there are many instruments. In particular, performance of a estimator proposed by Hausman, Newey, Woutersen, Chao, and Swanson (2007) with the robust version of JIVE -proposed by Angrist, Imbens and Krueger (1999)- is analyzed. Theoretical results are presented for the robust t-statistics, which is mostly affected by the finite-sample bias of the estimator.

---

E-mail: ralfaro@bcentral.cl.

# 1 Introduction

In this paper, I discuss feasible inference methods using Instrumental Variables estimators in the presence of heteroskedastic error terms and many instruments. The discussion is based on various instrumental variables (IV) estimators and inference is performed using robust standard errors.

The Two Stage Least Squares (2SLS) estimator has a finite-sample bias that grows with the number of instruments. Under homoskedastic errors, unbiased estimators, such as the Limited Information Maximum Likelihood (LIML), the Bias Corrected 2SLS (B2SLS) and the Fuller (1977) adjusted LIML (LIMLF) are available. However, these estimators are no longer asymptotically unbiased in the presence of heteroskedastic errors and many instruments. Hausman, Newey, Woutersen, Chao, and Swanson (2007) propose a robust version of LIML (RLML) that is asymptotically unbiased under heteroskedasticity and many instruments. This estimator follows the same principle as the Jackknife Instrumental Variable Estimator (JIVE) proposed by Phillips and Hale (1977), and Angrist, Imbens and Krueger (1999).

The inference properties of these estimators have not been studied yet. Hence, I derive an Edgeworth expansion of the robust $t$-statistic computed with robust standard errors based on White's (1980) approach.

The rest of the paper is organized as follows. Section 2 defines the model and estimators along with their asymptotic properties. Section 3 introduces the asymptotic expansion for the estimator of the standard errors and the $t$-statistics. Section 4 presents Monte Carlo simulations to check how the asymptotic results of previous sections work in finite samples. Section 5 concludes.

## 2 Model and Estimators

I consider a linear model with one endogenous explanatory variable ($x$)

$$\begin{aligned} y_i &= \beta x_i + e_i = z_i'\theta + u_i, \\ x_i &= z_i'\pi + v_i, \end{aligned} \tag{1}$$

where $\beta$ is the parameter of interest, and $\theta$ and $\pi$ are the parameters of the reduced form model. Errors terms are correlated ($E(u_i v_i) \neq 0$), and there is a $K \times 1$ set of valid instruments ($Z$), that are used to estimate $\beta$. I consider the following class of Generalized Estimators (GE)

$$\hat{\beta}_{GE} = \frac{x'Sy}{x'Sx}.$$

Depending on the choice of $S$, this estimator corresponds to Two Stage Least Squares (2SLS), Donald and Newey (2001) bias-corrected 2SLS (B2SLS), Limited Information Maximum Likelihood (LIML), Fuller (1977) correction for LIML (LIMLF), robust LIML (RLML) proposed by Hausman, Newey, Woutersen, Chao, and Swanson (2007), robust LIMLF (RFLL), or JIVE2 (hereafter JIVE) proposed by Angrist, Imbens and Krueger (1999). Table 1 shows the choices of $S$ for these estimators, where $W = [y, x]$.

Table 1: Generalized Estimator

| | 2SLS | LIML | LIMLF | B2SLS | RLML | RFLL | JIVE |
|---|---|---|---|---|---|---|---|
| $S$ | $P$ | $P - lI$ | $P - \frac{(n+1)l-1}{(n-1)+l}I$ | $P - \frac{(K-2)}{n}I$ | $R - rI$ | $R - \frac{(n+1)r-1}{(n-1)+l}I$ | $R$ |

Note $P = Z(Z'Z)^{-1}Z'$, $R = P - diag(P)$, and $l$ and $r$ are the minimum eigenvalues of $(W'W)^{-1}W'PW$ and $(W'W)^{-1}W'RW$, respectively.

Additionally, I consider an optimal Generalized Method of Moment (GMM) estimator, defined as follows:

$$\hat{\beta}_{GMM} = \frac{x'ZM^{-1}Z'y}{x'ZM^{-1}Z'x},$$

where $M$ is the optimal weighting matrix

$$M = \frac{1}{n}\sum_{i=1}^{n} e_i^2 z_i z_i'.$$

When the error terms ($e$) are known, the estimator is called the infeasible GMM. In the case of unknown residuals, these can be computed together with $\beta$, then $M = M(\beta)$ and the estimator does not have a closed form solution. This estimator is called Continuous Updating Estimator (CUE) (see Hansen, Heaton and Yaron (1996)). Alternatively, residuals can be estimated using an initial estimator. In that case, the estimator is termed two-step GMM (GMM2) and cannot be represented by $\hat{\beta}_{GE}$. Under the traditional first-order asymptotics, GMM2 is robust to heteroskedastic errors, if the initial estimator is consistent.

To approximate the behavior of the estimators in large samples, I consider an Edgeworth expansion of the generalized estimator. Conditions on the distribution of errors and the set of instruments are presented next.

**Condition 2.1.** *The set of instruments is non-stochastic with the following quadratic properties:*

$$\frac{1}{n}\sum_{i=1}^{n} z_i z_i' \quad = \quad \Delta + o\left(1\right),$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i S_{ij} z_j' \quad = \quad A + o\left(1\right),$$

*where $\Delta$ and $A$ are nonsingular finite positive definite matrices, and $S_{ij}$ denotes the $(i,j)$ element of the matrix $S$.*

This condition follows Newey (2004) and it is used as a normalization. It is possible to extend the analysis to stochastic instruments, but that requires additional assumptions on the joint distribution of the instruments and the error terms of the reduced form model. It should be noted that as $n \to \infty$, for $S_{ij} = P_{ij} = z_i'(Z'Z)^{-1}z_j$

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i P_{ij} z_j' &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i z_i'(Z'Z/n)^{-1} z_j z_j' \\
&= \left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)\left(\frac{1}{n}\sum_{k=1}^{n} z_k z_k'\right)^{-1}\left(\frac{1}{n}\sum_{j=1}^{n} z_j z_j'\right) \\
&\to \Delta\Delta^{-1}\Delta = \Delta.
\end{aligned}$$

This result is important for the asymptotic analysis presented below. Also, for the case of $S_{ij} = R_{ij}$, where the $(i,j)$ component is equal to $P_{ij}$ if $i \neq j$ and zero otherwise, we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i R_{ij} z_j' &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} z_i P_{ij} z_j' = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i} z_i z_i'\left(\frac{Z'Z}{n}\right)^{-1} z_j z_j' \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[z_i z_i'\left(\frac{1}{n}\sum_{k=1}^{n} z_k z_k'\right)^{-1}\left(\frac{1}{n}\sum_{j=1}^{n} z_j z_j' - \frac{1}{n} z_i z_i'\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[z_i z_i' - z_i z_i'\left(\frac{1}{n}\sum_{i=1}^{n} z_i z_i'\right)^{-1}\frac{1}{n} z_i z_i'\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}(1-P_{ii}) z_i z_i' \\
&\to \Delta - \lim_{n\to\infty}\left(\frac{1}{n^2}\sum_{i=1}^{n} z_i z_i'\Delta^{-1} z_i z_i'\right) \equiv \Delta^*.
\end{aligned}$$

$\Delta^*$ will be used in the asymptotic analysis of the jackknife estimators (JIVE, RLML, and RFLL).

**Condition 2.2.** *The error terms $u_i$ and $v_i$ are jointly distributed, with zero mean, possible heteroskedasticity and finite fourth moments. Also, the bivariate distribution is symmetric ($E(u_i^2 v_i) = E(u_i v_i^2) = 0$).*

This condition guarantees the existence of the first two moments for GE and is based on Newey (2004). With these conditions, we will be able to develop inference methods valid for large samples. [1]

**Condition 2.3.** *The number of instruments* ($K$) *increases along with the sample size* ($n$), *but* $K/n$ *converges to a fixed number* $0 \leq \alpha < 1$. *These alternative asymptotic sequences are proposed by Bekker (1994).*

This condition also includes the standard first-order asymptotics (fixed number of instruments) by setting $\alpha = 0$. I will use $\rightarrow$ to denote double asymptotics in $K$ and $n$.

## 2.1 Asymptotic Bias of GE

To analyze the asymptotic properties of the inference procedures, it is useful to derive an approximation of the bias of GE. This bias is obtained using the alternative asymptotic sequences (many instrument asymptotics) described in Condition 2.3.

**Theorem 2.1.** *Under Conditions 2.2 and 2.3, the asymptotic bias for* $\hat{\beta}_{GE}$ *is*

$$ABias(\hat{\beta}_{GE}) \quad = \quad \lim_{K,n\rightarrow\infty} \left[ \frac{1}{n\pi' A\pi} \sum_{i=1}^{n} S_{ii} E(e_i v_i) \right].$$

*Proof.* See Appendix A.4. □

Theorem 2.1 summarizes several results available in the literature of IV estimators. The following examples compute explicitly the bias for the estimators considered in Table 1.

### Example: 2SLS estimator

Under homoskedastic errors and $S = P$, we have that $A = \Delta$; then, the asymptotic bias for 2SLS will be similar to the expression derived in Nagar (1959) or Hahn and Hausman (2002).

$$ABias(\hat{\beta}_{2SLS}) \quad = \quad \lim_{K,n\rightarrow\infty} \left[ \frac{1}{n\pi'\Delta\pi} \sum_{i=1}^{n} P_{ii} E(e_i v_i) \right] = \alpha \frac{\sigma_{ev}}{\pi'\Delta\pi}.$$

[1] It is possible to generalize Condition 2.2 allowing for non-zero third moment. This does not change the main conclusions presented in this chapter, but complicates the notation.

**Example: B2SLS, LIML and LIMLF estimators**

Define $q_{B2SLS} \equiv (K-2)/n$, $q_{LIML} \equiv l$ and $q_{LIMLF} \equiv [(n+1)l-1]/[(n-1)+l]$. Nagar (1959), Rothenberg (1984), and Donald and Newey (2001) show that for all these estimators $q = K/n + o(1)$ as $n \to \infty$. Under Condition 2.3 we have $q \to \alpha$.

Moreover, for these estimators we have $S_{ii} = P_{ii} - q$ and $S_{ij} = P_{ij}$ if $i \neq j$, then

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i S_{ij} z_j' &= \frac{1}{n}\sum_{i=1}^{n} z_i (P_{ii}-q) z_i' + \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} z_i P_{ij} z_j' \\
&= \frac{1}{n}\sum_{i=1}^{n} z_i (P_{ii}-q) z_i' + \frac{1}{n}\sum_{i=1}^{n}(1-P_{ii}) z_i z_i' \\
&\to (1-\alpha)\Delta \equiv A.
\end{aligned}$$

Using this expression, the asymptotic biases for B2SLS, LIML and LIMLF can be computed as follows:

$$ABias(\hat{\beta}) = \lim_{K,n\to\infty} \frac{\sigma_{ev}}{(1-\alpha)\pi'\Delta\pi}\left[\left(\frac{1}{n}\sum_{i=1}^{n} P_{ii}\right) - \frac{K}{n}\right] = 0.$$

The last result is obtained using the fact that $K = \sum_{i=1}^{n} P_{ii}$. When errors are not homoskedastic, the asymptotic bias for B2SLS is not zero.

$$ABias(\hat{\beta}_{B2SLS}) = \lim_{K,n\to\infty}\left[\frac{1}{n(1-\alpha)\pi'\Delta\pi}\sum_{i=1}^{n}\left(P_{ii}-\frac{K}{n}\right) E(e_i v_i)\right].$$

Bekker and van der Ploeg (2005) report that B2SLS is asymptotically unbiased in the case that instruments are group dummies of equal size. In that case, $z_{ki} = 1$ if individual $i$ belongs to group $k$ and zero otherwise. Then $P_{ii}^{k} = z_{ki}(Z_k' Z_k)^{-1} z_{ki} = 1/n_k$, where $n_k$ is the number of individuals in group $k$. Moreover, if the groups have the same size, then $n_k = n/K$, which implies that $P_{ii}^{k} = K/n$.

In the case of LIML, the presence of heteroskedastic errors implies that the minimum eigenvalue satisfies

$$l \xrightarrow{p} \lim_{K,n\to\infty} \left[\frac{1}{n}\sum_{i=1}^{n} E(e_i^2)\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} P_{ii}E(e_i^2)\right] = \lim_{K,n\to\infty} \sum_{i=1}^{n} P_{ii} \underbrace{\left[\frac{E(e_i^2)}{\sum_{i=1}^{n} E(e_i^2)}\right]}_{\lambda_i}.$$

where $\sum_{i=1}^{n} \lambda_i = 1$ (see Lemma A.1 in Appendix A.1). Then, the asymptotic bias for LIML is

$$ABias(\hat{\beta}_{LIML}) \quad = \quad \lim_{K,n\to\infty} \left[\frac{1}{n(1-\alpha)\pi'\Delta\pi} \sum_{i=1}^{n} \left(P_{ii} - \sum_{i=1}^{n} P_{ii}\lambda_i\right) E(e_i v_i)\right].$$

Similarly to B2SLS, LIML is asymptotically unbiased when instruments are group dummies of equal size. But, LIML is also asymptotically unbiased under another condition reported in Hausman, Newey, Woutersen, Chao, and Swanson (2007). Define $\gamma_i \equiv E(e_i v_i)/E(e_i^2)$, and $\Omega_n = \sum_{i=1}^{n} P_{ii}E(e_i^2)$. Then, LIML is asymptotically unbiased when $\gamma_i$ is constant across units. Now, consider $\gamma_i = \gamma$ then $E(e_i v_i) = \gamma E(e_i^2)$, and the asymptotic bias for LIML is

$$\begin{aligned} ABias(\hat{\beta}_{LIML}) \quad &= \quad \lim_{K,n\to\infty} \frac{\gamma}{n(1-\alpha)\pi'\Delta\pi} \sum_{i=1}^{n} \left\{P_{ii} - \sum_{i=1}^{n} P_{ii}\left[\frac{E(e_i^2)}{\sum_{i=1}^{n} E(e_i^2)}\right]\right\} E(e_i^2) \\ &= \quad \lim_{K,n\to\infty} \frac{\gamma}{n(1-\alpha)\pi'\Delta\pi} \left\{\Omega_n - \sum_{i=1}^{n} \left[\frac{\Omega_n}{\sum_{i=1}^{n} E(e_i^2)}\right] E(e_i^2)\right\} \\ &= \quad \lim_{K,n\to\infty} \frac{\gamma}{n(1-\alpha)\pi'\Delta\pi} \left\{\Omega_n - \left[\frac{\Omega_n}{\sum_{i=1}^{n} E(e_i^2)}\right] \sum_{i=1}^{n} E(e_i^2)\right\} = 0. \end{aligned}$$

**Example: JIVE, RLML and RFLL estimators**

Using Lemma A.1 (Appendix A.1), it is easy to show that the minimum eigenvalue for RLML ($r$) converges in probability to zero.

$$r \xrightarrow{p} \lim_{K,n\to\infty} \left[\frac{1}{n}\sum_{i=1}^{n} E(e_i^2)\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} R_{ii}E(e_i^2)\right] = 0.$$

Using this result and Theorem 2.1, the asymptotic biases for jackknife estimators (JIVE, RLML, and RFLL) are the same ($r \xrightarrow{p} 0$) and can be computed as follows:

$$ABias(\hat{\beta}_{RLML}) = \left(\frac{1}{n\pi'\Delta^*\pi}\right) \lim_{K,n\to\infty} \sum_{i=1}^{n} R_{ii} E(e_i v_i) = 0.$$

The expression equals zero because $R_{ii} = 0$, then these estimators are asymptotically unbiased regardless of whether the error terms are homoskedastic or heteroskedastic.

## 2.2 Asymptotic Variance of GE

In order to compute the asymptotic variance of GE, I follow the Laplace approach which is commonly used to approximate the moments of a estimator defined as the ratio of quadratic forms (see Ullah (2004) for details). In particular, we have

$$E[(\hat{\beta}_{GE} - \beta)^2] = \frac{E[(x'Se)^2]}{[E(x'Sx)]^2} + O\left(\frac{1}{n}\right).$$

Using this result we can get the asymptotic variance of $\sqrt{n}(\hat{\beta}_{GE} - \beta)$ as follows

$$\begin{aligned} E[n(\hat{\beta}_{GE} - \beta)^2] &= \lim_{K,n\to\infty} \left[E\left(\frac{x'Sx}{n}\right)\right]^{-2} E\left(\frac{x'See'Sx}{n}\right) \\ &= \frac{B}{(\pi'A\pi)^2}, \end{aligned}$$

where $B = \lim_{K,n\to\infty} B_n$, and

$$B_n = \pi' \left(\frac{1}{n}\sum_{i,j,k=1}^{n} z_i S_{ij} E(e_j^2) S_{jk} z_k'\right)\pi + \frac{1}{n}\sum_{i,j=1}^{n} E[S_{ij}^2(v_i^2 e_j^2 + v_i e_i v_j e_j)].$$

It should be noted that the second argument converges to zero under a fixed number of instruments (standard asymptotics). Also under standard asymptotics, $S$ is replaced by $P$ in the cases of B2SLS, LIML and LIMLF estimators and by $R$ for JIVE, RLML and RFLL estimators. For example, for 2SLS, B2SLS, LIML and LIMLF estimators, the robust asymptotic variance under standard asymptotics (fixed number of instruments) is

$$Var(\hat{\beta}) = n\left[\frac{x'Z(Z'Z)^{-1}}{x'Sx}\right]\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2 z_i z_i'\right)\left[\frac{(Z'Z)^{-1}Z'x}{x'Sx}\right].$$

In the following examples, the asymptotic variance is computed for the estimators under analysis. An alternative approach is presented in the Appendix A.2.

**Example: B2SLS, LIML, and LIMLF estimator**

For the case of these estimators under homoskedastic errors the denominator $\pi' A\pi$ converges to $(1-\alpha)\pi'\Delta\pi$. The numerator requires considering the following facts:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} z_i P_{ij} P_{jk} z_k' &= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} z_i z_i'(Z'Z/n)^{-1} z_j z_j'(Z'Z/n)^{-1} z_k z_k' \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i z_i'(Z'Z/n)^{-1} z_j z_j' = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i P_{ij} z_j' \rightarrow \Delta.\end{aligned}$$

Also, we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} P_{ij}^2 &= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i'(Z'Z/n)^{-1} z_j z_j'(Z'Z/n)^{-1} z_i \\ &= \frac{1}{n^2}\sum_{i=1}^{n} z_i'(Z'Z/n)^{-1} z_i = \frac{1}{n}\sum_{i=1}^{n} P_{ii} = \frac{K}{n} \rightarrow \alpha.\end{aligned}$$

To simplify the notation, consider $q_{ij} = q \times 1_{ij}$, where $1_{ij}$ is the indicator function that is 1 when $i = j$ and zero otherwise. Then, the following relationships hold:

$$\begin{aligned}\frac{1}{n}\sum_{i,j,k=1}^{n} z_i(P_{ij}-q_{ij})(P_{jk}-q_{jk})z_k' &= \frac{1}{n}\sum_{i,j,k=1}^{n} z_i P_{ij} P_{jk} z_k' - \frac{2q}{n}\sum_{i,j=1}^{n} z_i P_{ij} z_j' + \frac{q^2}{n}\sum_{i=1}^{n} z_i z_i' \\ &\rightarrow (1-\alpha)\pi'\Delta\pi,\end{aligned}$$

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(P_{ij}-q_{ij})^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} P_{ij}^2 - 2q\frac{1}{n}\sum_{i=1}^{n} P_{ii} + q^2 \rightarrow \alpha(1-\alpha).$$

For B2SLS, $q$ is not stochastic, then $E[S_{ij}^2(v_i^2 e_j^2 + v_i e_i v_j e_j)] = (P_{ij} - q_{ij})^2[E(v_i^2)E(e_j^2) + E(v_i e_i)E(v_j e_j)] = (P_{ij} - q_{ij})^2(\sigma_v^2\sigma_e^2 + \sigma_{ev}^2)$; using that, we have

$$\begin{aligned} B_n &= \sigma_e^2\pi'\left(\frac{1}{n}\sum_{i,j,k=1}^{n} z_i P_{ij} P_{jk} z_k'\right)\pi + \left(\sigma_v^2\sigma_e^2 + \sigma_{ev}^2\right)\left(\frac{1}{n}\sum_{i,j=1}^{n} P_{ij}^2\right) \\ &\to [\sigma_e^2\pi'\Delta\pi + \alpha(\sigma_v^2\sigma_e^2 + \sigma_{ev}^2)](1-\alpha) \equiv B. \end{aligned}$$

Then, the asymptotic variance for B2SLS under homoscedastic errors is

$$V_{B2SLS} = \frac{\sigma_e^2}{\pi'\Delta\pi} + \frac{\alpha}{1-\alpha}\left[\frac{\sigma_v^2\sigma_e^2 + \sigma_{ev}^2}{(\pi'\Delta\pi)^2}\right].$$

This expression is also obtained by Hahn and Hausman (2002).

For LIML and LIMLF, $q$ is stochastic; then, we can compute $E[S_{ij}^2(v_i^2 e_j^2 + v_i e_i v_j e_j)]$ using conditional expectations (see Appendix A.2). In particular, it is convenient to use $\hat{\gamma} \equiv x'e/e'e$ to build an uncorrelated residual $w_i = v_i - \gamma e_i$. Noting that $\hat{\gamma} \xrightarrow{p} \sigma_{ev}/\sigma_e^2$, and $E(w_i e_i) = E[(v_i - \gamma e_i)e_i] = \sigma_{ev} - \gamma\sigma_e^2 = \sigma_{ev} - (\sigma_{ev}/\sigma_e^2)\sigma_e^2 = 0$, then the asymptotic variance for LIML and LIMLF is

$$V_{LIML} = \frac{\sigma_e^2}{\pi'\Delta\pi} + \frac{\alpha}{1-\alpha}\left[\frac{\sigma_v^2\sigma_e^2 - \sigma_{ev}^2}{(\pi'\Delta\pi)^2}\right].$$

Note that $V_{LIML} \leq V_{B2SLS}$.

**Example: JIVE, RLML and RFLL estimators**

Now we consider the cases of jackknife estimators under heteroskedastic errors. Theorem A.2 in Appendix A.3 shows for JIVE that

$$B_n = \frac{1}{n}\sum_{\substack{i=1\\ j\neq i\\ k\neq j}}^{n} \pi' z_i z_k' \pi P_{ij} P_{jk} E(e_j^2) + \frac{1}{n}\sum_{\substack{i=1\\ j\neq i}}^{n} P_{ij}^2[E(v_i^2)E(e_j^2) + E(v_i e_i)E(v_j e_j)].$$

Defining the first and second terms in parentheses as $V_0$ and $V_1$, respectively, the asymptotic variance of JIVE becomes

$$V_{JIVE} = \lim_{K,n\to\infty} \frac{V_0 + V_1}{(\pi'\Delta^*\pi)^2}.$$

For the case of RLML and RFLL, we can apply Lemma A.5 in Appendix A.3 to show that

$$V_{RLML} = \lim_{K,n\to\infty} \frac{V_0 + V_1'}{(\pi'\Delta^*\pi)^2},$$

where $V_1' = \frac{1}{n}\sum_{i=1}^n \sum_{j\neq i} P_{ij}^2 [E(w_i^2)E(e_j^2) + E(w_i e_i)E(w_j e_j)]$.

Now, we consider the asymptotic variance of these estimators under homoskedastic errors. In this case, $V_0$ can be written as follows:

$$V_0 = \sigma_e^2 \pi' \left( \frac{1}{n} \sum_{j=1}^n \sum_{i\neq j} \sum_{k\neq j} z_i P_{ij} P_{jk} z_k' \right) \pi = \sigma_e^2 \pi' \left[ \frac{1}{n} \sum_{i=1}^n z_i (1 - P_{ii})^2 z_i' \right] \pi.$$

Also, $V_1$ and $V_1'$ become

$$\begin{aligned} V_1 &= (\sigma_v^2\sigma_e^2 + \sigma_{ev}^2) \left[ \frac{1}{n} \sum_{i=1}^n \sum_{j\neq i} P_{ij}^2 \right] = (\sigma_v^2\sigma_e^2 + \sigma_{ev}^2) \left[ \frac{K}{n} - \frac{1}{n} \sum_{j=1}^n P_{jj}^2 \right], \\ V_1' &= (\sigma_v^2\sigma_e^2 - \sigma_{ev}^2) \left[ \frac{1}{n} \sum_{i=1}^n \sum_{j\neq i} P_{ij}^2 \right] = (\sigma_v^2\sigma_e^2 - \sigma_{ev}^2) \left[ \frac{K}{n} - \frac{1}{n} \sum_{j=1}^n P_{jj}^2 \right]. \end{aligned}$$

With these, the asymptotic variances for JIVE and RLML under homoscedastic errors are

$$\begin{aligned} V_{JIVE} &= \lim_{K,n\to\infty} \frac{\sigma_e^2 \left[\frac{1}{n}\sum_{i=1}^n \pi'(1-P_{ii})^2 z_i z_i' \pi\right] + (\sigma_v^2\sigma_e^2 + \sigma_{ev}^2)\left(\frac{K}{n} - \frac{1}{n}\sum_{i=1}^n P_{ii}^2\right)}{(\pi'\Delta^*\pi)^2}, \\ V_{RLML} &= \lim_{K,n\to\infty} \frac{\sigma_e^2 \left[\frac{1}{n}\sum_{i=1}^n \pi'(1-P_{ii})^2 z_i z_i' \pi\right] + (\sigma_v^2\sigma_e^2 - \sigma_{ev}^2)\left(\frac{K}{n} - \frac{1}{n}\sum_{i=1}^n P_{ii}^2\right)}{(\pi'\Delta^*\pi)^2}. \end{aligned}$$

When instruments are group dummies and the groups have equal size, we showed that $P_{ii} = K/n$. Using this and Condition 2.3 ($K/n \to \alpha$), we have $\Delta^* = (1-\alpha)\Delta$, $V_0 \to (1-\alpha)^2\sigma_e^2\pi'\Delta\pi$, $V_1 \to \alpha(1-\alpha)(\sigma_v^2\sigma_e^2 + \sigma_{ev}^2)$, and $V_1' \to \alpha(1-\alpha)(\sigma_v^2\sigma_e^2 - \sigma_{ev}^2)$. With these, $V_{JIVE} = V_{B2SLS}$ and $V_{RLML} = V_{LIML}$.

# 3 Robust Inference

In the previous section, I discussed the asymptotic biases and variances for IV estimators under many instruments (Condition 2.3). Most of them are asymptotically unbiased under homoskedastic errors, but only jackknife estimators (JIVE, RLML, and RFLL) remain unbiased in the presence of heteroskedasticity in the error terms. Also, estimates of the asymptotic variance are available for these estimators.

In practice, however, empirical researchers usually rely on standard fixed number of instruments asymptotics to compute standard errors. This approach may be problematic when the number of instruments is large. For example, consider the 2SLS estimator with homoskedastic errors. Under many instruments, the estimator is asymptotically biased. Moreover, using the standard asymptotic variance, only the first argument of the asymptotic robust variance will be computed using a biased estimator. In particular, if $\sigma_{ev} > 0$ then the estimator of the parameter will be upward biased, while the estimated robust variance will be downward biased, leading to upward biased $t$-statistics. In other words, the null will be rejected more frequently than the nominal level for any critical value.

In this section, I compute the probability limit of the $t$-statistics for GE with White-type robust standard errors based on a fixed number of instruments asymptotic variances. Clearly, this $t$-test can be improved by considering unbiased estimators and using many-instruments robust variance. The formulae can be used for cases where inference matters more than the actual point estimate and unbiased estimators are not available, or for cases where the bias is expected to be small (for example with few, but strong instruments) and a further correction on inference is required.

## 3.1 Expansion for Robust Estimator of the Asymptotic Variance

Consider a expansion of the center of the sandwich of the asymptotic variance (or $\hat{M}$). In particular, since $\hat{e}_i^2 = e_i^2 - 2e_i x_i(\hat{\beta} - \beta) + x_i^2(\hat{\beta} - \beta)^2$, we have

$$
\begin{aligned}
\hat{M} &\equiv \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^2 z_i z_i' = \frac{1}{n}\sum_{i=1}^{n} e_i^2 z_i z_i' - 2(\hat{\beta}-\beta)\frac{1}{n}\sum_{i=1}^{n} e_i x_i z_i z_i' + (\hat{\beta}-\beta)^2\frac{1}{n}\sum_{i=1}^{n} x_i^2 z_i z_i' \\
&= \frac{1}{n}\sum_{i=1}^{n} e_i^2 z_i z_i' - 2(\hat{\beta}-\beta)\frac{1}{n}\sum_{i=1}^{n} e_i x_i z_i z_i' + O_p\left(\frac{1}{n}\right).
\end{aligned}
$$

The last term is small relative to the first two, even in the case where $\hat{\beta}$ is asymptotically biased. For the argument $(\hat{\beta}-\beta)$, the standard asymptotics can be used. The following theorem gives an approximation for the robust asymptotic variance.

**Theorem 3.1.** *Based on Conditions 2.1 and 2.2 and fixed number of instruments, the robust asymptotic variance can be approximated by*

$$
\hat{V}(\hat{\beta}) = \frac{1}{(\pi'\Delta\pi)^2}\frac{1}{n}\sum_{i=1}^{n} e_i^2 \pi' z_i z_i' \pi \left[1 - \frac{2\pi' z_i z_i' \pi}{n(\pi'\Delta\pi)}\right] + O_p\left(\frac{1}{n}\right),
$$

*Proof.* See Appendix A.4. □

However, if a jackknife estimator is used (JIVE, RLML or RFLL), the second argument includes the factor $R_{ii}$, which is zero by definition. This increases the asymptotic variance.

### 3.2 Expansion of $t$-statistics

The setup follows Ullah (2004) for the case of the GLS estimator. I consider the null hypothesis as $H_0 : \beta = b$, and the alternative as $H_1 : \beta \neq b$. Also, let $\overline{\beta}$ be a consistent and asymptotically unbiased estimator, and $\sigma^2_{\overline{\beta}}$ its asymptotic variance evaluated at the true error terms, then the robust $T_n$-statistic is

$$
T_n = \left(\frac{\hat{\beta}-b}{\hat{\sigma}_{\hat{\beta}}}\right),
$$

where $\hat{\beta}$ is the estimator of interest and $\hat{\sigma}^2_{\hat{\beta}}$ is an estimator of the robust variance under standard asymptotics.

It should be noted that $\overline{\beta}$ can be JIVE, RLML or RLLF. As we showed in the previous section, these jackknife estimators are asymptotically unbiased under heteroskedasticity and many instruments.

**Theorem 3.2.** *The $T_n$-statistic can be approximated as follows*

$$T_n = Z_n + \frac{A_n}{\sqrt{n}} + \frac{B_n}{n} + o_p\left(\frac{1}{n}\right),$$

*where $Z_n \equiv (\overline{\beta} - b)/\sigma_{\overline{\beta}} \sim N(\lambda, 1)$ with $\lambda \equiv (\beta - b)/\sigma_{\overline{\beta}}$, and expressions $A_n$ and $B_n$ are asymptotic expansions of $A$ and $B$ up to order $1/n$.*

$$\begin{aligned} A &= \sqrt{n}\left[\left(\frac{\hat{\beta} - \overline{\beta}}{\sigma_{\overline{\beta}}}\right) - \frac{Z_n}{2}\left(\frac{\hat{\sigma}^2_{\hat{\beta}} - \sigma^2_{\overline{\beta}}}{\sigma^2_{\overline{\beta}}}\right)\right], \\ B &= n\left[\frac{3Z_n}{8}\left(\frac{\hat{\sigma}^2_{\hat{\beta}} - \sigma^2_{\overline{\beta}}}{\sigma^2_{\overline{\beta}}}\right)^2 - \frac{1}{2}\left(\frac{\hat{\beta} - \overline{\beta}}{\sigma_{\overline{\beta}}}\right)\left(\frac{\hat{\sigma}^2_{\hat{\beta}} - \sigma^2_{\overline{\beta}}}{\sigma^2_{\overline{\beta}}}\right)\right]. \end{aligned}$$

*Proof.* See Appendix A.4. □

In practice, $(\hat{\beta} - \overline{\beta})/\sigma_{\overline{\beta}}$ is a measure of asymptotic bias of $\hat{\beta}$, which is approximated by Theorem 2.1. Also, $(\hat{\sigma}^2_{\hat{\beta}} - \sigma^2_{\overline{\beta}})/\sigma^2_{\overline{\beta}}$ is approximated using the expansion proposed in section 3.1. Theorem 3.2 can be used to construct an approximation of the finite-sample probability distribution of $T_n$. In particular, under the null $\lambda = 0$, the asymptotic distribution of $T_n$ is standard normal. The expansion decomposes the distribution of $T_n$ into the first order asymptotic distribution ($Z_n$) -which is normal- and additional terms of higher order. For that reason, the proposed expansion is expected to be closer to the finite-sample distribution. One possible application of this is the following Edgeworth approximation for the distribution function of $T_n$

$$P(T_n \leq t) \approx P\left(Z_n \leq t - \frac{A_n}{\sqrt{n}} - \frac{B_n}{n}\right) = E\left[\Phi\left(t - \frac{A_n}{\sqrt{n}} - \frac{B_n}{n}\right)\right] + O\left(\frac{1}{n}\right),$$

where $\Phi(x)$ is the normal probability function. The expectation can be approximated by

Taylor series around $t$ using the Hermite polynomials (see, e.g., Ullah (2004)). For order up to $1/n$, it requires the first moments of $A_n$ and $B_n$ and the second moment of $A_n$.

$$E\left[\Phi\left(t-\frac{A_n}{\sqrt{n}}-\frac{B_n}{n}\right)\right] \;=\; E\left[\Phi\left(t\right)+\phi(t)\left(\frac{A_n}{\sqrt{n}}+\frac{B_n}{n}\right)-\frac{1}{2}\phi(t)t\left(\frac{A_n}{\sqrt{n}}\right)^2\right],$$

where $\phi(t)$ is the normal density function.

On alternative application for the expansion is to normalize $T_n$ using the asymptotic moments up to order $1/n$. In that case, the statistic of interest is

$$T_n^* \;=\; \frac{T_n-E(T_n)}{\sqrt{Var(T_n)}}=\left[1+\frac{Var(A_n)}{n}\right]^{-1/2}\left[T_n-\frac{E(A_n)}{\sqrt{n}}-\frac{E(B_n)}{n}\right]+O_p\left(\frac{1}{n}\right).$$

We will use this formula to identify the size distortion of the inference based on a robust $t$-statistics. The following Lemma summarizes the main results.

**Theorem 3.3.** *The corrections for $T_n$ are*

$$\begin{aligned} E\left(\frac{A_n}{\sqrt{n}}+\frac{B_n}{n}\right) &= \xi\left[1+\frac{\alpha}{2(1-\alpha)}\varsigma\right]+O\left(\frac{1}{n}\right), \\ Var\left(\frac{A_n}{\sqrt{n}}\right) &= 1+\left(\frac{\kappa-1}{4}\right)\left(1-\frac{2\alpha}{1-\alpha}\varsigma\right)+O\left(\frac{1}{n}\right), \end{aligned}$$

*where $\xi\equiv E(\hat{\beta}-\overline{\beta})/\sigma_{\overline{\beta}}$ (normalized bias), $\varsigma\equiv(\sigma_v^2\sigma_e^2-\sigma_{ev}^2)/(\sigma_e^2\pi'\Delta\pi)$, and $\kappa$ is the excess kurtosis of the error term $e$.*

*Proof.* See Appendix A.4. □

Note that $\varsigma>0$, then many instruments ($\alpha>0$) increases the distortion of the $t$-statistic in the same direction as the asymptotic bias of the estimator. For that reason a biased estimator should not be used for inference.

## 4 Monte Carlo Experiment

In order to analyze the properties of the estimators presented in the previous section, I consider 4 designs for the scalar model.[2]

The general setting follows Newey and Windmeijer (2005) and Hahn, Newey, Woutersen, Chao, and Swanson (2007). The key parameters are (1) the correlation between the error terms of the reduced form ($\rho$) as a measure of endogeneity, (2) the concentration parameter ($\delta^2$) measuring the quality of the instruments, and (3) the number of instruments ($K$). Moreover, $z_i \sim N(0, I_K)$, where $1_K$ is a $K$ vector of ones, $u_i \sim N(0,1)$, $v_i \sim N(0,1)$, and $w_i \sim N(0, z_{1i}^2)$. The data generating process is $y_i = x_i\beta + e_i$, $x_i = z_i'\pi + v_i$, with $e_i = \rho v_i + \sqrt{1-\rho^2}(\phi w_i + \theta u_i)/(\phi^2 + \theta^2)$ and $\pi = 1_K\delta^2/(Kn)$. Finally, I will focus on $\delta^2 = 35$ and $n = 200$, and I set $\beta = 1$ and $\theta = 0.74$. This implies a theoretical $R^2$ of the first stage regression of 15%.

Design I is homoskedastic ($\phi = 0$) and follows Newey and Windmeijer (2005). Under this setting we note that $E(e_i v_i) = \rho$ and $E(e_i^2) = \rho^2 Var(v_i) + (1-\rho^2)Var(u_i) = 1$, then $\gamma_i = \rho$, which is constant across units. Design II follows the heteroskedastic design of Hahn, Hausman, and Kuersteiner (2004), where $E(e_i^2|z_i) = z_i'z_i/K$ and $\phi = 0$. In this case $\gamma_i = \rho\sqrt{K/z_i'z_i}$, then LIML is biased.

In previous designs the reduced-form is correctly specified. Following Hausman, Newey, Woutersen, Chao, and Swanson (2007), I set $\pi_1 = \sqrt{35/200}$ and $\pi_k = 0$ for $k = 2, \ldots, K$. However, the model is estimated with a set of redundant instruments, defined as $z_{2i} = z_{1i}^2$, $z_{3i} = z_{1i}^3$, $z_{4i} = z_{1i}^4$, and $z_{qi} = z_{1i}D_{qi}$, where $D_{qi} \in \{0,1\}$, $Pr(D_{qi} = 0) = 0.5$, and $q = 5, \ldots, K$. [3] With these specifications two designs are generated: Design III is homoskedastic ($\phi = 0$), and Design IV is heteroskedastic ($\phi = 1$). Note that $E(e_i^2|z_i) = \rho^2 + (1-\rho^2)(0.646 z_{1i}^2 + 0.478)$, and $E(e_i v_i) = \rho$, then $\gamma_i$ depends on $z_{1i}$ for which LIML is no longer asymptotically unbiased.[4]

[2] All the estimators were computed along with the constant, although I set its true value at zero.

[3] The authors note that this design follows the asymptotics presented in Donald and Newey (2001), where the set of instruments includes approximating functions for the optimal set of instruments.

[4] For all these four designs, the parameter sequences are $K = \{5, 10, 15\}$, and $\rho = \{0.3, 0.5, 0.7\}$, with

## 4.1 Distribution of Estimators

Table 2 shows descriptive statistics of the bias of $\beta$ under Design I (homoskedastic errors). As expected, the 2SLS estimator is biased, and GMM2 based on initial 2SLS is also biased. The bias increases with the number of instruments and also with the degree of endogeneity of the model ($\rho$). All other estimators are median unbiased with no particular effect of the number of instruments and $\rho$. Also, the Interquartile Range (IQR) is higher for B2SLS and JIVE relative to LIML and RLML, respectively. It should be noted that the dispersion in terms of the difference between 90th and 10th percentiles ($P_{90} - P_{10}$) is considerably higher in the cases of B2SLS and JIVE. Finally, RLML behaves similarly to LIML with slightly more dispersion in the tails. Table 3 shows the distribution of the bias of $\beta$ under Design II (heteroskedastic errors). Only the second moment of $e_i$ is different under this design; therefore, it is expected that B2SLS remains unbiased (note that $E(e_i v_i) = \rho$). For the case of LIML, the bias originates in the fact that the minimum eigenvalue does not converge at the usual rate ($K/n$). As was expected by the theory, RLML is unbiased. In terms of dispersion, the conclusion is similar to the homoskedastic case.

In Table 5, the results for Design III (homoskedastic errors) is presented. Relative to Design I, the following conclusions remain valid: (1) 2SLS and GMM are biased, and (2) LIML and RLML are unbiased. However, a big difference arises in terms of the bias of B2SLS, which is increasing in the number of redundant instruments. Finally, Table 4 shows the results for Design IV (heteroskedastic errors). Relative to Design II, the conclusions remain for almost all estimators. RLML remains unbiased with the smallest dispersion among unbiased estimators.

In conclusion, RLML seems to be unbiased regardless of the structure of error terms, the number of instruments, the degree of endogeneity of the model, and the specification of the reduced form. Its dispersion is not bigger than LIML, which implies that RLML is a good alternative to LIML. These conclusions agree with the findings reported by Hausman, Newey, Woutersen, Chao, and Swanson (2007).

---

10000 replications.

Table 2: Distribution of the Bias($\hat{\beta}$): Design I

| $K$ | 2SLS | GMM | LIML | LIMLF | B2SLS | RLML | RFLL | JIVE |
|---|---|---|---|---|---|---|---|---|
| | | | | $\rho = 0.3$ | | | | |
| 5 | 0.032 | 0.032 | 0.001 | 0.010 | 0.001 | 0.001 | 0.011 | -0.018 |
| | (0.209) | (0.214) | (0.241) | (0.232) | (0.242) | (0.241) | (0.232) | (0.261) |
| | [0.406] | [0.413] | [0.467] | [0.448] | [0.472] | [0.466] | [0.446] | [0.516] |
| 10 | 0.064 | 0.066 | 0.003 | 0.011 | 0.002 | 0.003 | 0.012 | -0.016 |
| | (0.195) | (0.203) | (0.253) | (0.243) | (0.254) | (0.254) | (0.244) | (0.276) |
| | [0.377] | [0.396] | [0.508] | [0.486] | [0.515] | [0.509] | [0.482] | [0.573] |
| 15 | 0.088 | 0.090 | 0.000 | 0.008 | 0.002 | 0.000 | 0.010 | -0.020 |
| | (0.180) | (0.197) | (0.273) | (0.264) | (0.272) | (0.273) | (0.262) | (0.298) |
| | [0.359] | [0.380] | [0.543] | [0.522] | [0.548] | [0.552] | [0.523] | [0.610] |
| | | | | $\rho = 0.5$ | | | | |
| 5 | 0.051 | 0.051 | -0.003 | 0.012 | -0.002 | -0.002 | 0.013 | -0.033 |
| | (0.210) | (0.213) | (0.247) | (0.235) | (0.246) | (0.247) | (0.233) | (0.275) |
| | [0.406] | [0.409] | [0.475] | [0.449] | [0.489] | [0.474] | [0.449] | [0.550] |
| 10 | 0.103 | 0.104 | -0.004 | 0.011 | -0.003 | -0.003 | 0.014 | -0.034 |
| | (0.195) | (0.202) | (0.263) | (0.251) | (0.269) | (0.263) | (0.250) | (0.301) |
| | [0.378] | [0.388] | [0.512] | [0.487] | [0.549] | [0.511] | [0.482] | [0.617] |
| 15 | 0.144 | 0.145 | -0.005 | 0.010 | 0.000 | -0.005 | 0.011 | -0.036 |
| | (0.183) | (0.191) | (0.278) | (0.263) | (0.293) | (0.278) | (0.262) | (0.328) |
| | [0.344] | [0.369] | [0.540] | [0.510] | [0.578] | [0.542] | [0.504] | [0.660] |
| | | | | $\rho = 0.7$ | | | | |
| 5 | 0.072 | 0.071 | -0.002 | 0.018 | -0.002 | -0.001 | 0.020 | -0.045 |
| | (0.200) | (0.203) | (0.237) | (0.223) | (0.247) | (0.237) | (0.222) | (0.278) |
| | [0.392] | [0.401] | [0.465] | [0.436] | [0.500] | [0.467] | [0.435] | [0.576] |
| 10 | 0.148 | 0.149 | 0.001 | 0.021 | 0.006 | 0.002 | 0.025 | -0.040 |
| | (0.176) | (0.185) | (0.248) | (0.234) | (0.263) | (0.248) | (0.232) | (0.302) |
| | [0.341] | [0.354] | [0.482] | [0.451] | [0.533] | [0.485] | [0.448] | [0.621] |
| 15 | 0.206 | 0.203 | 0.003 | 0.023 | 0.009 | 0.002 | 0.025 | -0.038 |
| | (0.162) | (0.168) | (0.255) | (0.241) | (0.291) | (0.259) | (0.238) | (0.328) |
| | [0.311] | [0.330] | [0.513] | [0.478] | [0.603] | [0.517] | [0.474] | [0.697] |

Median, IQR (in parentheses) and $P_{90} - P_{10}$ (in brackets).

Table 3: Distribution of the Bias($\hat{\beta}$): Design II

| $K$ | 2SLS | GMM | LIML | LIMLF | B2SLS | RLML | RFLL | JIVE |
|---|---|---|---|---|---|---|---|---|
| | | | | $\rho = 0.3$ | | | | |
| 5 | 0.036 | 0.037 | -0.006 | 0.003 | 0.008 | 0.002 | 0.010 | -0.019 |
| | (0.252) | (0.256) | (0.301) | (0.290) | (0.287) | (0.282) | (0.272) | (0.308) |
| | [0.480] | [0.486] | [0.586] | [0.563] | [0.555] | [0.548] | [0.525] | [0.607] |
| 10 | 0.069 | 0.069 | -0.003 | 0.005 | 0.010 | 0.003 | 0.012 | -0.018 |
| | (0.210) | (0.218) | (0.289) | (0.278) | (0.278) | (0.271) | (0.261) | (0.301) |
| | [0.416] | [0.427] | [0.587] | [0.562] | [0.559] | [0.550] | [0.526] | [0.614] |
| 15 | 0.094 | 0.094 | -0.003 | 0.005 | 0.011 | 0.004 | 0.013 | -0.017 |
| | (0.195) | (0.206) | (0.308) | (0.296) | (0.287) | (0.286) | (0.274) | (0.314) |
| | [0.382] | [0.399] | [0.622] | [0.590] | [0.574] | [0.577] | [0.546] | [0.638] |
| | | | | $\rho = 0.5$ | | | | |
| 5 | 0.057 | 0.056 | -0.015 | 0.000 | 0.009 | -0.002 | 0.012 | -0.034 |
| | (0.250) | (0.247) | (0.308) | (0.294) | (0.287) | (0.284) | (0.272) | (0.319) |
| | [0.484] | [0.482] | [0.589] | [0.559] | [0.573] | [0.550] | [0.523] | [0.641] |
| 10 | 0.110 | 0.106 | -0.016 | 0.000 | 0.011 | -0.003 | 0.014 | -0.035 |
| | (0.213) | (0.217) | (0.304) | (0.289) | (0.293) | (0.282) | (0.268) | (0.328) |
| | [0.418] | [0.423] | [0.595] | [0.565] | [0.592] | [0.557] | [0.527] | [0.670] |
| 15 | 0.151 | 0.149 | -0.020 | -0.005 | 0.012 | -0.006 | 0.012 | -0.038 |
| | (0.194) | (0.200) | (0.315) | (0.296) | (0.307) | (0.292) | (0.275) | (0.342) |
| | [0.368] | [0.389] | [0.626] | [0.585] | [0.603] | [0.573] | [0.538] | [0.689] |
| | | | | $\rho = 0.7$ | | | | |
| 5 | 0.083 | 0.083 | -0.019 | 0.001 | 0.015 | 0.000 | 0.020 | -0.046 |
| | (0.235) | (0.238) | (0.300) | (0.283) | (0.283) | (0.283) | (0.265) | (0.324) |
| | [0.470] | [0.470] | [0.584] | [0.550] | [0.575] | [0.549] | [0.513] | [0.664] |
| 10 | 0.158 | 0.155 | -0.017 | 0.004 | 0.025 | 0.003 | 0.025 | -0.040 |
| | (0.198) | (0.203) | (0.288) | (0.272) | (0.289) | (0.270) | (0.253) | (0.332) |
| | [0.380] | [0.384] | [0.572] | [0.528] | [0.578] | [0.530] | [0.492] | [0.673] |
| 15 | 0.214 | 0.209 | -0.017 | 0.005 | 0.027 | 0.004 | 0.026 | -0.037 |
| | (0.174) | (0.178) | (0.291) | (0.272) | (0.301) | (0.271) | (0.252) | (0.348) |
| | [0.334] | [0.341] | [0.579] | [0.537] | [0.621] | [0.543] | [0.499] | [0.733] |

Median, IQR (in parentheses) and $P_{90} - P_{10}$ (in brackets).

Table 4: Distribution of the Bias($\hat{\beta}$): Design III

| $K$ | 2SLS | GMM | LIML | LIMLF | B2SLS | RLML | RFLL | JIVE |
|---|---|---|---|---|---|---|---|---|
| | | | | $\rho = 0.3$ | | | | |
| 5 | 0.048 | 0.040 | 0.013 | 0.022 | 0.016 | 0.013 | 0.025 | -0.011 |
| | (0.276) | (0.292) | (0.315) | (0.301) | (0.312) | (0.317) | (0.302) | (0.351) |
| | [0.504] | [0.580] | [0.580] | [0.556] | [0.574] | [0.594] | [0.562] | [0.656] |
| 10 | 0.081 | 0.069 | 0.016 | 0.022 | 0.018 | 0.012 | 0.023 | -0.005 |
| | (0.272) | (0.285) | (0.350) | (0.335) | (0.349) | (0.351) | (0.333) | (0.389) |
| | [0.487] | [0.550] | [0.644] | [0.623] | [0.650] | [0.659] | [0.634] | [0.744] |
| 15 | 0.104 | 0.099 | 0.024 | 0.032 | 0.031 | 0.026 | 0.042 | 0.005 |
| | (0.264) | (0.278) | (0.360) | (0.346) | (0.375) | (0.369) | (0.354) | (0.430) |
| | [0.470] | [0.505] | [0.686] | [0.662] | [0.718] | [0.734] | [0.696] | [0.832] |
| | | | | $\rho = 0.5$ | | | | |
| 5 | 0.064 | 0.062 | 0.012 | 0.027 | 0.015 | 0.008 | 0.024 | -0.022 |
| | (0.247) | (0.258) | (0.280) | (0.272) | (0.285) | (0.283) | (0.268) | (0.316) |
| | [0.455] | [0.478] | [0.531] | [0.506] | [0.522] | [0.539] | [0.515] | [0.635] |
| 10 | 0.118 | 0.114 | 0.008 | 0.024 | 0.018 | 0.008 | 0.025 | -0.016 |
| | (0.230) | (0.245) | (0.309) | (0.297) | (0.316) | (0.309) | (0.290) | (0.358) |
| | [0.420] | [0.463] | [0.549] | [0.530] | [0.594] | [0.564] | [0.532] | [0.694] |
| 15 | 0.159 | 0.160 | 0.009 | 0.022 | 0.024 | 0.017 | 0.033 | -0.003 |
| | (0.216) | (0.240) | (0.312) | (0.303) | (0.335) | (0.307) | (0.287) | (0.405) |
| | [0.420] | [0.424] | [0.602] | [0.564] | [0.640] | [0.643] | [0.602] | [0.787] |
| | | | | $\rho = 0.7$ | | | | |
| 5 | 0.081 | 0.081 | 0.003 | 0.023 | 0.009 | -0.002 | 0.021 | -0.039 |
| | (0.202) | (0.216) | (0.236) | (0.227) | (0.259) | (0.236) | (0.222) | (0.291) |
| | [0.380] | [0.391] | [0.442] | [0.415] | [0.472] | [0.458] | [0.430] | [0.594] |
| 10 | 0.156 | 0.148 | 0.003 | 0.022 | 0.016 | 0.003 | 0.028 | -0.033 |
| | (0.180) | (0.192) | (0.245) | (0.230) | (0.269) | (0.249) | (0.239) | (0.318) |
| | [0.343] | [0.365] | [0.473] | [0.446] | [0.534] | [0.489] | [0.450] | [0.683] |
| 15 | 0.214 | 0.210 | 0.007 | 0.026 | 0.023 | -0.001 | 0.025 | -0.025 |
| | (0.163) | (0.186) | (0.245) | (0.231) | (0.305) | (0.259) | (0.235) | (0.360) |
| | [0.310] | [0.342] | [0.484] | [0.458] | [0.572] | [0.498] | [0.456] | [0.758] |

Median, IQR (in parentheses) and $P_{90} - P_{10}$ (in brackets).

Table 5: Distribution of the Bias($\hat{\beta}$): Design IV

| $K$ | 2SLS | GMM | LIML | LIMLF | B2SLS | RLML | RFLL | JIVE |
|---|---|---|---|---|---|---|---|---|
| | | | | $\rho = 0.3$ | | | | |
| 5 | 0.044 | 0.028 | -0.052 | -0.040 | 0.015 | 0.016 | 0.025 | 0.001 |
| | (0.244) | (0.222) | (0.366) | (0.348) | (0.286) | (0.258) | (0.246) | (0.281) |
| | [0.464] | [0.421] | [0.768] | [0.696] | [0.536] | [0.508] | [0.490] | [0.552] |
| 10 | 0.083 | 0.073 | -0.127 | -0.104 | 0.028 | 0.022 | 0.033 | 0.000 |
| | (0.232) | (0.205) | (0.442) | (0.424) | (0.313) | (0.282) | (0.267) | (0.309) |
| | [0.431] | [0.395] | [0.973] | [0.863] | [0.578] | [0.555] | [0.526] | [0.602] |
| 15 | 0.109 | 0.101 | -0.206 | -0.180 | 0.029 | 0.024 | 0.036 | 0.007 |
| | (0.216) | (0.194) | (0.597) | (0.553) | (0.304) | (0.315) | (0.292) | (0.313) |
| | [0.418] | [0.390] | [1.285] | [1.100] | [0.615] | [0.599] | [0.564] | [0.660] |
| | | | | $\rho = 0.5$ | | | | |
| 5 | 0.068 | 0.051 | -0.074 | -0.052 | 0.021 | 0.012 | 0.027 | -0.020 |
| | (0.229) | (0.202) | (0.313) | (0.294) | (0.266) | (0.237) | (0.221) | (0.271) |
| | [0.413] | [0.387] | [0.579] | [0.536] | [0.487] | [0.460] | [0.430] | [0.532] |
| 10 | 0.122 | 0.110 | -0.138 | -0.114 | 0.021 | 0.007 | 0.024 | -0.014 |
| | (0.209) | (0.183) | (0.365) | (0.341) | (0.290) | (0.265) | (0.242) | (0.308) |
| | [0.372] | [0.345] | [0.708] | [0.645] | [0.528] | [0.490] | [0.463] | [0.575] |
| 15 | 0.160 | 0.159 | -0.212 | -0.187 | 0.021 | 0.003 | 0.022 | -0.011 |
| | (0.185) | (0.172) | (0.425) | (0.384) | (0.286) | (0.285) | (0.269) | (0.306) |
| | [0.350] | [0.339] | [0.816] | [0.744] | [0.566] | [0.560] | [0.505] | [0.634] |
| | | | | $\rho = 0.7$ | | | | |
| 5 | 0.085 | 0.077 | -0.042 | -0.019 | 0.014 | 0.006 | 0.027 | -0.038 |
| | (0.196) | (0.188) | (0.237) | (0.217) | (0.240) | (0.209) | (0.197) | (0.267) |
| | [0.362] | [0.347] | [0.484] | [0.444] | [0.467] | [0.437] | [0.395] | [0.526] |
| 10 | 0.158 | 0.156 | -0.079 | -0.054 | 0.022 | 0.001 | 0.027 | -0.033 |
| | (0.180) | (0.176) | (0.258) | (0.238) | (0.266) | (0.230) | (0.210) | (0.302) |
| | [0.309] | [0.314] | [0.513] | [0.470] | [0.497] | [0.445] | [0.409] | [0.588] |
| 15 | 0.212 | 0.212 | -0.115 | -0.090 | 0.017 | -0.008 | 0.020 | -0.029 |
| | (0.151) | (0.154) | (0.282) | (0.258) | (0.256) | (0.235) | (0.209) | (0.304) |
| | [0.271] | [0.287] | [0.565] | [0.506] | [0.544] | [0.483] | [0.433] | [0.639] |

Median, IQR (in parentheses) and $P_{90} - P_{10}$ (in brackets).

## 4.2 Robust Inference

In the previous discussion, I checked that 2SLS and GMM (based on 2SLS) are biased, regardless of the structure of the error terms. For the case of homoskedastic errors, alternative estimators such as LIML, LIMLF, B2SLS, RLML, RFLL and JIVE are unbiased but more disperse. Under heteroskedastic errors, only RLML and RFLL remain unbiased. Here, I analyze inference procecures based on two-sided $t$-test, using these estimators.

Table 6 presents the results for homoskedastic errors and known reduced form (Design I). The first three columns show the rejection frequencies for 2SLS, GMM (based on 2SLS) and GMM3 (GMM2 but using Windmeijer's (2005) corrected standard errors). It should be noted that the rejection frequencies are close to the nominal size only when both the degree of endogeneity ($\rho = 0.3$) and the number of instruments ($K = 5$) are small. Other combinations lead to over-rejection of the null. The result was expected because the bias of 2SLS (same for GMM2 and GMM3) should be positive ($\sigma_{ev} = \rho > 0$); then $T_n$ will be upward biased. For example, under $\rho = 0.7$ with 10 instruments, the null would be rejected 20 times out of 100 in cases where it should be only rejected only 5 times out of 100. Unbiased estimators have rejection frequencies closer to the nominal size, as was expected. It is interesting to note that Fuller's corrections to LIML (LIMLF), and RLML (RFLL) estimators improve the inference under high degree of endogeneity.

Table 7 shows the results for heteroskedastic errors and known reduced form (Design II). Here the conclusions are similar to Design I. It seems that the main source of size distortion is due to the bias of the estimator. In other words, the biases for LIML, LIMLF and B2SLS are small in this design then the size distortion is also small if the reduced-form is known.

In Table 8 the rejection frequencies for homoskedastic and unknown reduced-form are presented. The conclusions are similar to Design I, but with higher rejection frequencies for all the estimators. Finally, in Table 9 rejection frequencies are presented for heteroskedastic errors and unknown reduced form. Thus, LIML and LIMLF have size distortion; whereas, B2SLS, RLML and RFLL have rejection frequencies closer to the nominal sizes.[5]

[5] It is important to remark that the $t$-statistics are computed using the standard asymptotics (fixed-

Table 6: Rejection Frequencies: Design I

| $K$ | 2SLS | GMM | GMM3 | LIML | LIMLF | B2SLS | RLML | RFLL | JIVE |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\rho = 0.3$ | | | | |
| 5 | 0.009 | 0.011 | 0.007 | 0.005 | 0.005 | 0.005 | 0.005 | 0.005 | 0.003 |
| | 0.051 | 0.062 | 0.055 | 0.044 | 0.043 | 0.043 | 0.042 | 0.045 | 0.034 |
| | 0.111 | 0.127 | 0.108 | 0.089 | 0.092 | 0.088 | 0.086 | 0.088 | 0.083 |
| 10 | 0.027 | 0.036 | 0.029 | 0.014 | 0.016 | 0.013 | 0.014 | 0.018 | 0.011 |
| | 0.074 | 0.090 | 0.078 | 0.046 | 0.049 | 0.042 | 0.044 | 0.048 | 0.038 |
| | 0.121 | 0.154 | 0.126 | 0.080 | 0.080 | 0.084 | 0.078 | 0.084 | 0.078 |
| 15 | 0.040 | 0.057 | 0.039 | 0.015 | 0.015 | 0.011 | 0.014 | 0.015 | 0.008 |
| | 0.099 | 0.147 | 0.103 | 0.051 | 0.054 | 0.054 | 0.050 | 0.057 | 0.042 |
| | 0.182 | 0.223 | 0.168 | 0.086 | 0.089 | 0.085 | 0.083 | 0.087 | 0.082 |
| | | | | | $\rho = 0.5$ | | | | |
| 5 | 0.026 | 0.032 | 0.025 | 0.014 | 0.018 | 0.014 | 0.012 | 0.017 | 0.009 |
| | 0.080 | 0.086 | 0.076 | 0.041 | 0.044 | 0.044 | 0.042 | 0.047 | 0.038 |
| | 0.139 | 0.150 | 0.137 | 0.083 | 0.088 | 0.091 | 0.075 | 0.092 | 0.077 |
| 10 | 0.065 | 0.084 | 0.064 | 0.015 | 0.019 | 0.018 | 0.014 | 0.018 | 0.011 |
| | 0.144 | 0.168 | 0.140 | 0.055 | 0.060 | 0.057 | 0.054 | 0.058 | 0.043 |
| | 0.213 | 0.238 | 0.195 | 0.094 | 0.098 | 0.093 | 0.093 | 0.096 | 0.077 |
| 15 | 0.097 | 0.136 | 0.099 | 0.019 | 0.019 | 0.023 | 0.016 | 0.022 | 0.016 |
| | 0.217 | 0.260 | 0.190 | 0.049 | 0.051 | 0.049 | 0.047 | 0.050 | 0.034 |
| | 0.291 | 0.341 | 0.295 | 0.072 | 0.083 | 0.091 | 0.070 | 0.080 | 0.081 |
| | | | | | $\rho = 0.7$ | | | | |
| 5 | 0.038 | 0.047 | 0.042 | 0.016 | 0.019 | 0.018 | 0.017 | 0.020 | 0.012 |
| | 0.088 | 0.099 | 0.085 | 0.037 | 0.047 | 0.039 | 0.037 | 0.047 | 0.030 |
| | 0.140 | 0.158 | 0.142 | 0.065 | 0.087 | 0.078 | 0.065 | 0.086 | 0.054 |
| 10 | 0.105 | 0.138 | 0.098 | 0.016 | 0.019 | 0.018 | 0.018 | 0.022 | 0.012 |
| | 0.227 | 0.250 | 0.204 | 0.043 | 0.054 | 0.054 | 0.044 | 0.052 | 0.030 |
| | 0.313 | 0.334 | 0.296 | 0.075 | 0.089 | 0.084 | 0.075 | 0.088 | 0.063 |
| 15 | 0.218 | 0.281 | 0.199 | 0.016 | 0.020 | 0.019 | 0.014 | 0.019 | 0.010 |
| | 0.388 | 0.422 | 0.347 | 0.053 | 0.060 | 0.064 | 0.054 | 0.062 | 0.039 |
| | 0.505 | 0.522 | 0.445 | 0.078 | 0.093 | 0.099 | 0.076 | 0.095 | 0.075 |

Rejection Frequencies associated to nominal sizes of 1%, 5%, and 10%.

Table 7: Rejection Frequencies: Design II

| $K$ | 2SLS | GMM | GMM3 | LIML | LIMLF | B2SLS | RLML | RFLL | JIVE |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\rho = 0.3$ | | | | |
| 5 | 0.015 | 0.015 | 0.013 | 0.008 | 0.009 | 0.009 | 0.008 | 0.008 | 0.006 |
| | 0.051 | 0.061 | 0.054 | 0.039 | 0.042 | 0.038 | 0.041 | 0.044 | 0.036 |
| | 0.113 | 0.126 | 0.105 | 0.088 | 0.087 | 0.094 | 0.091 | 0.092 | 0.086 |
| 10 | 0.025 | 0.034 | 0.027 | 0.010 | 0.012 | 0.013 | 0.012 | 0.013 | 0.009 |
| | 0.080 | 0.095 | 0.079 | 0.044 | 0.047 | 0.044 | 0.046 | 0.050 | 0.039 |
| | 0.121 | 0.154 | 0.130 | 0.075 | 0.080 | 0.087 | 0.078 | 0.083 | 0.077 |
| 15 | 0.042 | 0.058 | 0.036 | 0.015 | 0.017 | 0.016 | 0.015 | 0.017 | 0.011 |
| | 0.104 | 0.144 | 0.102 | 0.048 | 0.050 | 0.053 | 0.048 | 0.050 | 0.047 |
| | 0.180 | 0.226 | 0.169 | 0.085 | 0.085 | 0.093 | 0.082 | 0.084 | 0.083 |
| | | | | | $\rho = 0.5$ | | | | |
| 5 | 0.027 | 0.032 | 0.025 | 0.012 | 0.015 | 0.014 | 0.012 | 0.016 | 0.010 |
| | 0.083 | 0.095 | 0.078 | 0.039 | 0.045 | 0.047 | 0.044 | 0.052 | 0.041 |
| | 0.132 | 0.155 | 0.143 | 0.076 | 0.082 | 0.098 | 0.087 | 0.095 | 0.078 |
| 10 | 0.068 | 0.082 | 0.059 | 0.013 | 0.015 | 0.015 | 0.016 | 0.019 | 0.007 |
| | 0.145 | 0.168 | 0.138 | 0.050 | 0.056 | 0.063 | 0.056 | 0.064 | 0.049 |
| | 0.214 | 0.238 | 0.203 | 0.081 | 0.084 | 0.097 | 0.085 | 0.097 | 0.083 |
| 15 | 0.094 | 0.141 | 0.093 | 0.015 | 0.017 | 0.023 | 0.016 | 0.017 | 0.015 |
| | 0.211 | 0.252 | 0.199 | 0.045 | 0.050 | 0.055 | 0.048 | 0.053 | 0.042 |
| | 0.294 | 0.340 | 0.280 | 0.071 | 0.076 | 0.092 | 0.068 | 0.077 | 0.075 |
| | | | | | $\rho = 0.7$ | | | | |
| 5 | 0.036 | 0.046 | 0.042 | 0.013 | 0.014 | 0.020 | 0.014 | 0.019 | 0.009 |
| | 0.091 | 0.105 | 0.090 | 0.033 | 0.041 | 0.047 | 0.040 | 0.053 | 0.032 |
| | 0.147 | 0.167 | 0.157 | 0.069 | 0.078 | 0.084 | 0.079 | 0.090 | 0.063 |
| 10 | 0.109 | 0.128 | 0.093 | 0.012 | 0.019 | 0.023 | 0.016 | 0.022 | 0.011 |
| | 0.237 | 0.259 | 0.203 | 0.036 | 0.039 | 0.061 | 0.040 | 0.049 | 0.034 |
| | 0.318 | 0.351 | 0.295 | 0.054 | 0.068 | 0.098 | 0.070 | 0.090 | 0.072 |
| 15 | 0.222 | 0.264 | 0.199 | 0.015 | 0.019 | 0.025 | 0.020 | 0.022 | 0.011 |
| | 0.389 | 0.429 | 0.353 | 0.045 | 0.052 | 0.070 | 0.056 | 0.060 | 0.042 |
| | 0.495 | 0.524 | 0.446 | 0.070 | 0.075 | 0.111 | 0.077 | 0.086 | 0.078 |

Rejection Frequencies associated to nominal sizes of 1%, 5%, and 10%.

Table 8: Rejection Frequencies: Design III

| $K$ | 2SLS | GMM | GMM3 | LIML | LIMLF | B2SLS | RLML | RFLL | JIVE |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\rho = 0.3$ | | | | |
| 5 | 0.010 | 0.030 | 0.022 | 0.008 | 0.008 | 0.007 | 0.003 | 0.003 | 0.003 |
| | 0.048 | 0.095 | 0.077 | 0.038 | 0.037 | 0.043 | 0.033 | 0.035 | 0.030 |
| | 0.115 | 0.170 | 0.128 | 0.098 | 0.097 | 0.093 | 0.092 | 0.098 | 0.073 |
| 10 | 0.022 | 0.058 | 0.037 | 0.008 | 0.008 | 0.007 | 0.003 | 0.005 | 0.003 |
| | 0.067 | 0.143 | 0.105 | 0.040 | 0.045 | 0.042 | 0.035 | 0.035 | 0.027 |
| | 0.132 | 0.227 | 0.157 | 0.085 | 0.085 | 0.087 | 0.080 | 0.085 | 0.073 |
| 15 | 0.023 | 0.095 | 0.050 | 0.012 | 0.013 | 0.010 | 0.008 | 0.010 | 0.008 |
| | 0.108 | 0.205 | 0.113 | 0.052 | 0.053 | 0.048 | 0.035 | 0.043 | 0.032 |
| | 0.160 | 0.282 | 0.173 | 0.098 | 0.105 | 0.093 | 0.082 | 0.088 | 0.073 |
| | | | | | $\rho = 0.5$ | | | | |
| 5 | 0.020 | 0.045 | 0.032 | 0.010 | 0.010 | 0.013 | 0.007 | 0.007 | 0.005 |
| | 0.080 | 0.125 | 0.095 | 0.043 | 0.047 | 0.048 | 0.042 | 0.045 | 0.035 |
| | 0.143 | 0.183 | 0.148 | 0.087 | 0.097 | 0.098 | 0.088 | 0.103 | 0.073 |
| 10 | 0.053 | 0.113 | 0.070 | 0.018 | 0.018 | 0.018 | 0.012 | 0.015 | 0.007 |
| | 0.143 | 0.228 | 0.148 | 0.053 | 0.055 | 0.053 | 0.042 | 0.048 | 0.032 |
| | 0.227 | 0.302 | 0.218 | 0.082 | 0.092 | 0.095 | 0.088 | 0.103 | 0.087 |
| 15 | 0.093 | 0.195 | 0.098 | 0.015 | 0.018 | 0.022 | 0.013 | 0.013 | 0.012 |
| | 0.218 | 0.340 | 0.220 | 0.052 | 0.057 | 0.062 | 0.043 | 0.057 | 0.038 |
| | 0.318 | 0.415 | 0.295 | 0.092 | 0.093 | 0.093 | 0.087 | 0.098 | 0.073 |
| | | | | | $\rho = 0.7$ | | | | |
| 5 | 0.053 | 0.087 | 0.060 | 0.020 | 0.025 | 0.022 | 0.018 | 0.023 | 0.008 |
| | 0.142 | 0.168 | 0.132 | 0.057 | 0.072 | 0.067 | 0.062 | 0.082 | 0.043 |
| | 0.210 | 0.250 | 0.195 | 0.098 | 0.118 | 0.107 | 0.100 | 0.123 | 0.080 |
| 10 | 0.157 | 0.215 | 0.143 | 0.022 | 0.027 | 0.022 | 0.022 | 0.032 | 0.015 |
| | 0.280 | 0.368 | 0.252 | 0.062 | 0.082 | 0.077 | 0.068 | 0.085 | 0.057 |
| | 0.370 | 0.430 | 0.335 | 0.102 | 0.117 | 0.105 | 0.103 | 0.125 | 0.098 |
| 15 | 0.295 | 0.400 | 0.242 | 0.025 | 0.030 | 0.032 | 0.020 | 0.030 | 0.025 |
| | 0.475 | 0.547 | 0.423 | 0.058 | 0.067 | 0.075 | 0.058 | 0.075 | 0.055 |
| | 0.558 | 0.610 | 0.523 | 0.082 | 0.097 | 0.102 | 0.097 | 0.110 | 0.082 |

Rejection Frequencies associated to nominal sizes of 1%, 5%, and 10%.

Table 9: Rejection Frequencies: Design IV

| $K$ | 2SLS | GMM | GMM3 | LIML | LIMLF | B2SLS | RLML | RFLL | JIVE |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\rho = 0.3$ | | | | |
| 5 | 0.018 | 0.037 | 0.028 | 0.007 | 0.007 | 0.012 | 0.008 | 0.013 | 0.005 |
| | 0.065 | 0.083 | 0.060 | 0.038 | 0.042 | 0.048 | 0.050 | 0.052 | 0.037 |
| | 0.107 | 0.142 | 0.108 | 0.067 | 0.070 | 0.095 | 0.088 | 0.095 | 0.082 |
| 10 | 0.023 | 0.058 | 0.038 | 0.003 | 0.005 | 0.010 | 0.008 | 0.012 | 0.000 |
| | 0.093 | 0.133 | 0.082 | 0.018 | 0.022 | 0.047 | 0.047 | 0.045 | 0.037 |
| | 0.145 | 0.202 | 0.137 | 0.057 | 0.058 | 0.093 | 0.083 | 0.093 | 0.080 |
| 15 | 0.042 | 0.125 | 0.053 | 0.005 | 0.005 | 0.008 | 0.013 | 0.017 | 0.005 |
| | 0.132 | 0.223 | 0.128 | 0.015 | 0.015 | 0.040 | 0.045 | 0.050 | 0.033 |
| | 0.200 | 0.282 | 0.187 | 0.045 | 0.050 | 0.100 | 0.088 | 0.097 | 0.075 |
| | | | | | $\rho = 0.5$ | | | | |
| 5 | 0.030 | 0.058 | 0.037 | 0.007 | 0.007 | 0.013 | 0.017 | 0.020 | 0.007 |
| | 0.083 | 0.113 | 0.093 | 0.023 | 0.025 | 0.053 | 0.052 | 0.070 | 0.037 |
| | 0.152 | 0.178 | 0.147 | 0.058 | 0.067 | 0.090 | 0.105 | 0.112 | 0.085 |
| 10 | 0.065 | 0.120 | 0.078 | 0.003 | 0.003 | 0.008 | 0.015 | 0.018 | 0.008 |
| | 0.167 | 0.225 | 0.155 | 0.012 | 0.012 | 0.043 | 0.053 | 0.065 | 0.033 |
| | 0.247 | 0.308 | 0.223 | 0.038 | 0.047 | 0.103 | 0.090 | 0.110 | 0.082 |
| 15 | 0.128 | 0.248 | 0.143 | 0.002 | 0.003 | 0.013 | 0.020 | 0.028 | 0.015 |
| | 0.277 | 0.385 | 0.260 | 0.008 | 0.008 | 0.048 | 0.048 | 0.062 | 0.040 |
| | 0.365 | 0.480 | 0.360 | 0.020 | 0.022 | 0.095 | 0.098 | 0.113 | 0.080 |
| | | | | | $\rho = 0.7$ | | | | |
| 5 | 0.068 | 0.088 | 0.063 | 0.013 | 0.018 | 0.017 | 0.023 | 0.035 | 0.017 |
| | 0.137 | 0.170 | 0.148 | 0.037 | 0.045 | 0.062 | 0.062 | 0.088 | 0.037 |
| | 0.202 | 0.237 | 0.192 | 0.055 | 0.073 | 0.100 | 0.100 | 0.120 | 0.057 |
| 10 | 0.172 | 0.240 | 0.168 | 0.010 | 0.012 | 0.027 | 0.022 | 0.027 | 0.023 |
| | 0.313 | 0.390 | 0.305 | 0.022 | 0.025 | 0.055 | 0.053 | 0.063 | 0.045 |
| | 0.413 | 0.462 | 0.390 | 0.033 | 0.045 | 0.107 | 0.078 | 0.110 | 0.065 |
| 15 | 0.332 | 0.462 | 0.325 | 0.003 | 0.003 | 0.025 | 0.023 | 0.028 | 0.030 |
| | 0.508 | 0.595 | 0.468 | 0.008 | 0.012 | 0.073 | 0.048 | 0.063 | 0.053 |
| | 0.597 | 0.668 | 0.578 | 0.023 | 0.028 | 0.105 | 0.073 | 0.100 | 0.082 |

Rejection Frequencies associated to nominal sizes of 1%, 5%, and 10%.

# 5 Conclusion

I have reviewed the bias of Instrumental Variables estimators under homoskedastic and heteroskedastic errors. Using Bekker's (1994) alternative sequence, the asymptotic variances for these estimators were also computed. However, the focus of the discussion was based on inference using the standard asymptotics, and robust standard errors, following White (1980).

I derived an Edgeworth expansion of the robust $t$-statistics that can be used for further correction on the inference obtained by the standard asymptotics. In particular, the expansion shows that the distortion of the statistics is given by the normalized bias of the estimator and the missing arguments in the asymptotic variance. If the estimator is asymptotically biased, then the $t$-statistics are biased in the same direction as the estimator. In particular, inference using 2SLS, GMM2 or GMM3 should not be reliable.

Monte Carlo simulations show that asymptotic results are reasonable approximations for the behavior of the estimators in finite samples. In particular, LIML and B2SLS are biased under heteroskedastic errors, whereas RLML and RFLL are not. These biases imply a size distortion on the $t$-statistics. In contrast, asymptotically unbiased estimators such as RLML or RFLL have the lowest size distortion. In particular, RFLL has the closest rejection frequency to the nominal size.

JIVE exhibits finite-sample bias under homoskedastic and heteroskedastic errors, even when it is theoretically unbiased. Also, its dispersion (measured in terms of IQR and $P_{90} - P_{10}$) is the largest among all the estimators in the analysis.

It is recommended to correct the inference obtained by 2SLS or GMM based on 2SLS computing the adjustment proposed in Section 3. In particular, the technique proposed there can be extended to non-linear models for which other unbiased estimators are not available or are difficult to compute (such as the Continuous Updating Estimator).

---

instruments). However, under many-instruments standard errors, Hausman, Newey, Woutersen, Chao, and Swanson (2007) report size distortion for LIML and LIMLF, but not for RLML and RFLL (B2SLS is not included in the analysis).

## Reference


Angrist, J., Imbens, G., and Krueger, A. (1999) "Jackknife Instrumental Variables Estimation" *Journal of Applied Econometrics*, Vol. 14, No. 1, 57-67

Bekker, P. (1994) "Alternative Approximations to the Distribution of Instrumental Variable Estimators," *Econometrica*, 62, 657-681.

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

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

Fuller, W. (1977) "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939-953.

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

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

Hausman, J., Newey, W., Woutersen, T., Chao, J., and Swanson, N. (2007) "Instrumental Variable Estimation with Heteroskedasticity and Many Instruments" *Working Paper*, Massachusetts Institute of Technology.

Nagar, A. (1959) "The Bias and Moment Matrix of the General $k$-class estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 575-595.

Newey, W. (2004) "Many Instruments Asymptotics," *Working Paper*, Massachusetts Institute of Technology.

Newey, W., and Windmeijer, F. (2007) "GMM with Many Weak Moment Conditions" *Working Paper*, Massachusetts Institute of Technology.

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

Poirier, D. (1995) *Intermediate Statistics and Econometrics: A Comparative Approach*, Cambridge, MA: MIT Press.

Rothenberg, T. (1984) "Approximating the Distribution of Econometrics Estimators of Test Statistics," in *Handbook of Econometrics, Vol II,* ed. by Z. Griliches and M. Intriligator. Amsterdam: North-Holland, 881-936.

Ullah, A. (2004) *Finite Sample Econometrics* New York: Oxford University Press.

White, H. (1980) "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.

# A Appendix

## A.1 Lemmas

**Lemma A.1.** *Consider* $\overline{\beta} \xrightarrow{p} \beta$, $\overline{e} = y - \overline{\beta}x$, $P = Z(Z'Z)^{-1}Z'$ *and* $P_{ij}$ *the* $(i,j)$ *element of* $P$ *or* $P_{ij} = z_i'(Z'Z)^{-1}z_j$, *then*

$$\begin{aligned}
\frac{\overline{e}'\overline{e}}{n} &\xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n} E(e_i^2), \quad \frac{x'\overline{e}}{n} \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n} E(v_i e_i), \quad \frac{x'x}{n} \xrightarrow{p} \pi'\Delta\pi + \frac{1}{n}\sum_{i=1}^{n} E(v_i^2), \\
\frac{\overline{e}'P\overline{e}}{n} &\xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n} P_{ii}E(e_i^2), \quad \frac{x'P\overline{e}}{n} \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n} P_{ii}E(v_i e_i), \text{ and } \frac{x'Px}{n} \xrightarrow{p} \pi'\Delta\pi + \frac{1}{n}\sum_{i=1}^{n} P_{ii}E(v_i^2)
\end{aligned}$$

*Proof.* In the first row, the first argument follows by the Law of Large Numbers, for the second term in the same row $E(x_i e_i) = E[(\pi' z_i + v_i)e_i] = E(v_i e_i)$, the last equality is justified by the fact that $Z$ is a valid instrument. With a similar argument and using Condition 2.1 the third argument in that row can be proved.

For the first argument in the second row, the expression $E(e_i P_{ij} e_j) = P_{ii}E(e_i^2)$ when $i = j$ and zero otherwise. For the second term in that row, note that $E(x_i P_{ij} e_j) = E[(\pi' z_i + v_i)P_{ij}e_j] = \pi' E(z_i P_{ij} e_j) + E(P_{ij} v_i e_j) = P_{ij}E(v_i e_j)$, which can be reduced to $E(x_i P_{ii} e_i) = P_{ii}E(v_i e_i)$ and zero when $i \neq j$. Finally for the last term, $E(x_i P_{ij} x_j) = E[(\pi' z_i + v_i)P_{ij}(z_j'\pi + v_j)] = \pi' z_i P_{ij} z_j' \pi + E(\pi' z_i P_{ij} v_j) + E(v_i P_{ij} z_j' \pi) + E(v_i P_{ij} v_j) = \pi' z_i P_{ij} z_j' \pi + P_{ij}E(v_i v_j)$, which yields $E(x_i^2 P_{ii}) = \pi' z_i P_{ii} z_i' \pi + P_{ii}E(v_i^2)$ when $i = j$ and $\pi' z_i P_{ij} z_j' \pi$ for $i \neq j$. Adding both terms,

$$\frac{x'Px}{n} \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \pi' z_i P_{ij} z_j' \pi + \frac{1}{n}\sum_{i=1}^{n} P_{ii}E(v_i^2) = \pi'\Delta\pi + \frac{1}{n}\sum_{i=1}^{n} P_{ii}E(v_i^2)$$

$\square$

**Lemma A.2.** *Under the same conditions as in Lemma A.1, the following relations hold*

$$\frac{x'Py}{n} \xrightarrow{p} \beta\pi'\Delta\pi + \frac{1}{n}\sum_{i=1}^{n} P_{ii}E(u_i v_i)\ ,\ and\ \ \frac{y'Py}{n} \xrightarrow{p} \beta^2\pi'\Delta\pi + \frac{1}{n}\sum_{i=1}^{n} P_{ii}E(u_i^2)$$

*Proof.* Using the previous results I have $E(x_iP_{ij}y_j) = E[x_iP_{ij}(\beta x_j + e_j)] = \beta E(x_iP_{ij}x_j) + E(x_iP_{ij}e_j) = \beta[\pi' z_iP_{ij}z_j'\pi + P_{ij}E(v_iv_j)] + P_{ij}E(v_ie_j)$, which is $\beta\pi' z_iP_{ij}z_j'\pi$ when $i \neq j$, and $E(x_iP_{ii}y_i) = \beta[\pi' z_iP_{ii}z_i'\pi + P_{ii}E(v_i^2)] + P_{ii}E(v_ie_i)$ when $i = j$. Adding both terms,

$$\begin{aligned}\frac{x'Py}{n} &\xrightarrow{p} \beta\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\pi' z_iP_{ij}z_j'\pi + \frac{1}{n}\sum_{i=1}^{n}P_{ii}[\beta E(v_i^2) + E(v_ie_i)]\\ &= \beta\pi'\Delta\pi + \frac{1}{n}\sum_{i=1}^{n}P_{ii}[\beta E(v_i^2) + E(e_iv_i)]\end{aligned}$$

Note that $e_i = u_i - \beta v_i$ then $E(e_iv_i) = E(u_iv_i) - \beta E(v_i^2)$ or $E(u_iv_i) = E(e_iv_i) + \beta E(v_i^2)$.

For the second expression, $E(y_iP_{ij}y_j) = E[(x_i\beta + e_i)P_{ij}(x_j\beta + e_j)] = \beta^2E(x_iP_{ij}x_j) + \beta[E(x_iP_{ij}e_j) + E(x_jP_{ij}e_i)] + E(e_iP_{ij}e_j) = \beta^2[\pi' z_iP_{ij}z_j'\pi + P_{ij}E(v_iv_j)] + \beta[P_{ij}E(v_ie_j) + P_{ji}E(v_je_i)] + P_{ij}E(e_ie_j)$, which yields $E(y_i^2P_{ii}) = \beta^2[\pi' z_iP_{ii}z_i'\pi + P_{ii}E(v_i^2)] + 2\beta P_{ii}E(v_ie_i) + P_{ii}E(e_i^2)$ and $\beta^2\pi' z_iP_{ij}z_j'\pi$ when $i \neq j$, then

$$\begin{aligned}\frac{y'Py}{n} &\xrightarrow{p} \beta^2\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\pi' z_iP_{ij}z_j'\pi + \frac{1}{n}\sum_{i=1}^{n}P_{ii}[\beta^2E(v_i^2) + 2\beta E(v_ie_i) + E(e_i^2)]\\ &= \beta^2\pi'\Delta\pi + \frac{1}{n}\sum_{i=1}^{n}P_{ii}[\beta^2E(v_i^2) + 2\beta E(v_ie_i) + E(e_i^2)]\end{aligned}$$

Finally, note that $u_i^2 = e_i^2 + 2\beta e_iv_i + \beta^2v_i^2$. □

**Lemma A.3** (Wishart Distribution)**.** *If $U \sim N(0,\Omega)$ and $Z = U'PU$ then $Z \sim W(\Omega, K)$ (Wishart) where $K$ is the rank of $P$. Define $\omega_{ij}$ the $(i,j)$ element of $\Omega$. Then the first two moments are defined as follows: $E(z_{ij}) = K\omega_{ij}$, $Var(z_{ij}) = K(\omega_{ij}^2 + \omega_{ii}\omega_{jj})$ and $Cov(z_{ij}, z_{km}) = K(\omega_{ik}\omega_{jm} + \omega_{im}\omega_{jk})$.*

*Proof.* See Becker (1994) or Poirier (1995). □

**Lemma A.4** (Optimal Linear Combination)**.** *Let $b_1$ and $b_2$ two zero mean estimators with variances $\sigma_1^2$ and $\sigma_2^2$ and covariance $\sigma_{12}$, then the variance of $b = ab_1 + (1-a)b_2$ is*

$$
\begin{aligned}
V &\equiv Var(b) = a^2 Var(b_1) + 2a(1-a)Cov(b_1, b_2) + (1-a)^2 Var(b_2) \\
&= a^2\sigma_1^2 + 2a(1-a)\sigma_{12} + (1-a)^2\sigma_2^2
\end{aligned}
$$

*It is minimized when $a = (\sigma_2^2 - \sigma_{12})/(\sigma_1^2 - 2\sigma_{12} + \sigma_2^2)$ then at the minimum value*

$$
V = (\sigma_1^2\sigma_2^2 - \sigma_{12})/(\sigma_1^2 - 2\sigma_{12} + \sigma_2^2)
$$

*Proof.* The First Order Condition (FOC) is

$$
\frac{\partial V}{\partial a} = 2a\sigma_1^2 + (2-4a)\sigma_{12} - 2(1-a)\sigma_2^2 = 0
$$

using that it is easy to get the optimal $a$. Now I will check the Second Order Condition (SOC). First, consider $\rho = \sigma_{12}/\sqrt{\sigma_1^2\sigma_2^2}$ or $\sigma_{12} = \rho\sigma_1\sigma_2$, then the SOC is

$$
\frac{\partial^2 V}{\partial a^2} = 2\sigma_1^2 - 4\sigma_{12} + 2\sigma_2^2 = 2(\sigma_1^2 - 2\sigma_{12} + \sigma_2^2) = 2(\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2)
$$

By Cauchy-Schwarz inequality $-1 \leq \rho \leq 1$, then the minimum value of the right expression occurs when $\rho = 1$, which implies that $\partial^2 V/\partial a^2 > 2(\sigma_1^2 - 2\sigma_1\sigma_2 + \sigma_2^2) = 2(\sigma_1 - \sigma_2)^2 > 0$, then $a$ obtained from the FOC is a maximum.

Taking the optimal value for $a$ into $V$

$$
\begin{aligned}
V &= \left(\frac{\sigma_2^2-\sigma_{12}}{\sigma_1^2-2\sigma_{12}+\sigma_2^2}\right)^2\sigma_1^2+2\left(\frac{\sigma_2^2-\sigma_{12}}{\sigma_1^2-2\sigma_{12}+\sigma_2^2}\right)\left(\frac{\sigma_1^2-\sigma_{12}}{\sigma_1^2-2\sigma_{12}+\sigma_2^2}\right)\sigma_{12} \\
&\quad +\left(\frac{\sigma_1^2-\sigma_{12}}{\sigma_1^2-2\sigma_{12}+\sigma_2^2}\right)^2\sigma_2^2 \\
&= \frac{(\sigma_2^4-2\sigma_{12}\sigma_2^2+\sigma_{12}^2)\sigma_1^2+2\sigma_{12}(\sigma_1^2\sigma_2^2-\sigma_1^2\sigma_{12}-\sigma_2^2\sigma_{12}+\sigma_{12}^2)}{(\sigma_1^2-2\sigma_{12}+\sigma_2^2)^2} \\
&\quad +\frac{(\sigma_1^4-2\sigma_{12}\sigma_1^2+\sigma_{12}^2)\sigma_2^2}{(\sigma_1^2-2\sigma_{12}+\sigma_2^2)^2} \\
&= \frac{\sigma_1^2\sigma_2^4+\sigma_1^4\sigma_2^2-\sigma_{12}^2\sigma_2^2-\sigma_{12}^2\sigma_1^2-2\sigma_{12}\sigma_1^2\sigma_2^2+2\sigma_{12}^3}{(\sigma_1^2-2\sigma_{12}+\sigma_2^2)^2} \\
&= \frac{\sigma_1^2\sigma_2^2(\sigma_2^2+\sigma_1^2-2\sigma_{12})-\sigma_{12}^2(\sigma_2^2+\sigma_1^2-2\sigma_{12})}{(\sigma_1^2-2\sigma_{12}+\sigma_2^2)^2} \\
&= \frac{\sigma_1^2\sigma_2^2-\sigma_{12}^2}{\sigma_1^2-2\sigma_{12}+\sigma_2^2}
\end{aligned}
$$

□

## A.2 B2SLS and LIML: Homoscedastic Errors

In this section consistency and asymptotic normality are proved for B2SLS and LIML under normal homoscedastic errors and Condition 2.3. With the same set of assumptions it is proved that LIML has the minimum variance relative to a linear combination between the forward and reverse B2SLS estimators as it is suggested in Hahn and Hausman (2002).

### Consistency

In this section consistency will prove for B2SLS and LIML under the assumption of homoscedastic errors. Based on Lemma A.3: $E(u'Pu/n) = (K/n)\sigma_u^2$, $E(v'Pu/n) = E(u'Pv/n)$, which is $(K/n)\sigma_{uv}$ and $E(v'Pv/n) = (K/n)\sigma_v^2$. Therefore $E(e'Pe/n) = (K/n)\sigma_e^2$ and $E(x'Pe/n) = E(v'Pe/n) = (K/n)\sigma_{ev}$. Using probability limit ($p\lim$) to stand $K, n \to \infty$ and $K/n \to \alpha$ it is clear that

$$
\begin{aligned}
p\lim\left(\frac{x'Px}{n}\right) &= p\lim\left[\frac{(\pi'z'+v')P(z\pi+v)}{n}\right]\\
&= p\lim\left(\frac{\pi'z'z\pi}{n}\right)+p\lim\left(\frac{v'Pv}{n}\right)\\
&= \pi'\Delta\pi+\alpha\sigma_v^2
\end{aligned}
$$

$$
\begin{aligned}
p\lim\left(\frac{x'Py}{n}\right) &= p\lim\left[\frac{(\pi'z'+v')P(\beta z\pi+u)}{n}\right]\\
&= \beta p\lim\left(\frac{\pi'z'z\pi}{n}\right)+p\lim\left(\frac{v'Pu}{n}\right)\\
&= \beta\pi'\Delta\pi+\alpha\sigma_{uv}
\end{aligned}
$$

$$
\begin{aligned}
p\lim\left(\frac{y'Py}{n}\right) &= p\lim\left(\frac{(\beta\pi'z'+u')P(\beta z\pi+u)}{n}\right)\\
&= \beta^2 p\lim\left(\frac{\pi'z'z\pi}{n}\right)+p\lim\left(\frac{u'Pu}{n}\right)\\
&= \beta^2\pi'\Delta\pi+\alpha\sigma_u^2
\end{aligned}
$$

Also, $p\lim(x'x/n)=\pi'\Delta\pi+\sigma_v^2$, $p\lim(x'y/n)=\beta\pi'\Delta\pi+\sigma_{uv}$ and $p\lim(y'y/n)=\beta^2\pi'\Delta\pi+\sigma_u^2$. Then 2SLS is inconsistent for any $\alpha\neq 0$, but consistent for $\alpha=0$ which is the traditional asymptotics.

$$
p\lim(\hat{\beta}_{2SLS}) \;=\; p\lim\left(\frac{x'Py/n}{x'Px/n}\right)=\frac{\beta\pi'\Delta\pi+\alpha\sigma_{uv}}{\pi'\Delta\pi+\alpha\sigma_v^2}=\beta+\frac{\alpha(\sigma_{uv}-\beta\sigma_v^2)}{\pi'\Delta\pi+\alpha\sigma_v^2}
$$

For LIML the minimized objective function (minimum eigenvalue) $q\approx K/n\xrightarrow{p}\alpha$ and LIMLF is asymptotically equivalent to LIML. For B2SLS the correction factor is $K/n$, then

$$
\begin{aligned}
p\lim\left(\frac{x'Px}{n}-q\frac{x'x}{n}\right) &= p\lim\left(\frac{x'Px}{n}\right)-\alpha p\lim\left(\frac{x'x}{n}\right)\\
&= \pi'\Delta\pi+\alpha\sigma_v^2-\alpha(\pi'\Delta\pi+\sigma_v^2)\\
&= (1-\alpha)\pi'\Delta\pi
\end{aligned}
$$

$$
\begin{aligned}
p\lim\left(\frac{x'Py}{n} - q\frac{x'y}{n}\right) &= p\lim\left(\frac{x'Py}{n}\right) - \alpha p\lim\left(\frac{x'y}{n}\right) \\
&= (\beta\pi'\Delta\pi + \alpha\sigma_{uv}) - \alpha(\beta\pi'\Delta\pi + \sigma_{uv}) \\
&= (1-\alpha)\beta\pi'\Delta\pi
\end{aligned}
$$

$$
\begin{aligned}
p\lim\left(\frac{y'Py}{n} - q\frac{y'y}{n}\right) &= p\lim\left(\frac{y'Py}{n}\right) - \alpha p\lim\left(\frac{y'y}{n}\right) \\
&= \beta^2\pi'\Delta\pi + \alpha\sigma_u^2 - \alpha(\beta^2\pi'\Delta\pi + \sigma_u^2) \\
&= (1-\alpha)\beta^2\pi'\Delta\pi
\end{aligned}
$$

Using these results is clear that the forward and reverse B2SLS as well as LIML and LIMLF are consistent estimators.

**Asymptotic Distribution for B2SLS**

If the error terms are assumed to be jointly normal distributed then the properties of Wishart distribution (see Lemma A.3) can be used to get the asymptotic variance of the B2SLS estimator.

$$
\begin{aligned}
nVar\left(\frac{u'Pu}{n}\right) &= \frac{2K(\sigma_u^2)^2}{n} \to 2\alpha(\sigma_u^2)^2 \\
nVar\left(\frac{v'Pv}{n}\right) &= \frac{2K(\sigma_i^2)^2}{n} \to 2\alpha(\sigma_v^2)^2 \\
nVar\left(\frac{u'Pv}{n}\right) &= \frac{K(\sigma_{uv}^2 + \sigma_u^2\sigma_v^2)}{n} \to \alpha(\sigma_{uv}^2 + \sigma_u^2\sigma_v^2) \\
nCov\left(\frac{u'Pv}{n}, \frac{u'Pu}{n}\right) &= \frac{2K\sigma_{uv}\sigma_i^2}{n} \to 2\alpha\sigma_{uv}\sigma_i^2 \\
nCov\left(\frac{u'Pv}{n}, \frac{v'Pv}{n}\right) &= \frac{2K\sigma_{uv}\sigma_i^2}{n} \to 2\alpha\sigma_{uv}\sigma_i^2
\end{aligned}
$$

Using these results and the fact that $e = u - \beta v$ and $M \equiv I - P$ it is possible to compute

$$
\begin{aligned}
nVar\left(\frac{v'Pe}{n}\right) &= nVar\left(\frac{v'Pu}{n}\right) - 2\beta nCov\left(\frac{v'Pu}{n}, \frac{v'Pv}{n}\right) + \beta^2 nVar\left(\frac{v'Pv}{n}\right) \\
&\rightarrow \alpha(\sigma_{uv}^2 + \sigma_u^2\sigma_v^2) - 4\alpha\beta\sigma_{uv}\sigma_v^2 + 2\alpha\beta^2(\sigma_v^2)^2 \\
&\equiv AVar\left(\frac{v'Pe}{n}\right) \\
nVar\left(\frac{v'Me}{n}\right) &\rightarrow (1-\alpha)(\sigma_{uv}^2 + \sigma_u^2\sigma_v^2) - 4(1-\alpha)\beta\sigma_{uv}\sigma_v^2 + 2(1-\alpha)\beta^2(\sigma_v^2)^2
\end{aligned}
$$

It is clear to see that $Var(v'Pe/n) = [\alpha/(1-\alpha)]Var(v'Me/n)$. Defining $\sigma_{ev} = \sigma_{uv} - \beta\sigma_v^2$, the asymptotic variance $AVar(v'Pe/n)$ can be written as

$$
\begin{aligned}
AVar(v'Pe/n) &= \alpha(\sigma_{uv}^2 + \sigma_u^2\sigma_v^2) - 4\alpha\beta\sigma_{uv}\sigma_v^2 + 2\alpha\beta^2(\sigma_v^2)^2 \\
&= \alpha\left[\sigma_v^2(\sigma_u^2 - 2\beta\sigma_{uv} + \beta^2\sigma_v^2) + \sigma_{uv}^2 - 2\beta\sigma_{uv}\sigma_v^2 + \beta^2(\sigma_v^2)^2\right] \\
&= \alpha\left[\sigma_v^2\sigma_e^2 + (\sigma_{uv} - \beta\sigma_v^2)^2\right] \\
&= \alpha\left(\sigma_v^2\sigma_e^2 + \sigma_{ev}^2\right)
\end{aligned}
$$

Applying some algebra the B2SLS, LIML or LIMLF estimator can be decomposed as follows

$$
\begin{aligned}
\hat{\beta} &= \frac{x'Py - qx'y}{x'Px - qx'x} = \frac{\beta x'Px + x'Pe - q(\beta x'x + x'e)}{x'Px - qx'x} \\
&= \frac{\beta(x'Px - qx'x) + x'Pe - qx'e}{x'Px - qx'x} = \beta + \frac{x'Pe - qx'e}{x'Px - qx'x}
\end{aligned}
$$

Define $N = (x'Pe - qx'e)/n = [\pi'Z'Pe + v'Pe - q(\pi'Z'e + v'e)]/n = (1-q)\pi'Z'e/n + (v'Pe - qv'e)/n$, then $nVar(N) = nVar[(1-q)\pi'Z'e/n + v'Pe/n - qv'e/n] = nVar[(1-q)\pi'Z'e/n + (1-q)v'Pe/n - qv'Me/n] = n[(1-q)^2 Var(\pi'Z'e/n) + (1-q)^2 Var(v'Pe/n) + q^2 Var(v'Me/n)]$ (given that $MP = 0$ and the third moment is zero). With this

$$
\begin{aligned}
nVar(N) &= (1-q)^2\left[nVar(\pi'Z'e/n) + nVar(v'Pe/n) + \frac{q^2}{(1-q)^2}nVar(v'Me/n)\right] \\
&= (1-q)^2\left[\frac{\pi'Z'Z\pi}{n}\sigma_e^2 + nVar(v'Pe/n) + \frac{q^2(1-\alpha)}{(1-q)^2\alpha}nVar(v'Pe/n)\right] \\
&\xrightarrow{p} (1-\alpha)^2\left[\sigma_e^2\pi'\Delta\pi + AVar(v'Pe/n) + \frac{\alpha^2(1-\alpha)}{(1-\alpha)^2\alpha}AVar(v'Pe/n)\right] \\
&= (1-\alpha)^2\left[\sigma_e^2\pi'\Delta\pi + AVar(v'Pe/n) + \frac{\alpha}{1-\alpha}AVar(v'Pe/n)\right] \\
&= (1-\alpha)^2\left[\sigma_e^2\pi'\Delta\pi + \frac{1}{1-\alpha}AVar(v'Pe/n)\right] \\
&= (1-\alpha)^2\left[\sigma_e^2\pi'\Delta\pi + \frac{\alpha}{1-\alpha}\left(\sigma_v^2\sigma_e^2 + \sigma_{ev}^2\right)\right]
\end{aligned}
$$

From the previous results $(x'Px/n - qx'x/n)^2 \xrightarrow{p} [(1-\alpha)\pi'\Delta\pi]^2$, then the asymptotic variance for B2SLS will be

$$
V_{B2SLS} = \frac{\sigma_e^2}{\pi'\Delta\pi} + \frac{\alpha}{1-\alpha}\left[\frac{\sigma_v^2\sigma_e^2 + \sigma_{ev}^2}{(\pi'\Delta\pi)^2}\right]
$$

Finally, the asymptotic distribution for B2SLS will be $\sqrt{n}(\hat{\beta}_{B2SLS} - \beta) \sim N\left(0, V_{B2SLS}\right)$. In a similar way, the variance for the reverse B2SLS (R2SLS) is

$$
V_{R2SLS} = \frac{\sigma_e^2}{\pi'\Delta\pi} + \frac{\alpha}{1-\alpha}\left[\frac{\sigma_u^2\sigma_e^2 + \sigma_{eu}^2}{\beta^2(\pi'\Delta\pi)^2}\right]
$$

and the covariance between these estimators is[6]

$$
C_{B,R} = \frac{\sigma_e^2}{\pi'\Delta\pi} + \frac{\alpha}{1-\alpha}\left[\frac{\sigma_{uv}\sigma_e^2 + \sigma_{eu}\sigma_{ev}}{\beta(\pi'\Delta\pi)^2}\right]
$$

### Asymptotic Distribution for LIML

Following Newey (2004), Theorem 2, it is possible to get the asymptotic distribution of LIML under homoscedastic error using the sandwich variance.

First, consider the FOC for LIML which is $x'Pe(e'e) - x'e(e'Pe) = 0$, where $e = y - \hat{\beta}x$.

[6]See Hahn and Hausman (2002) for details.

Defining $D \equiv x'Pe/n-(e'Pe/e'e)x'e/n$ it is clear that under optimal $\hat{\beta}$, $D$ is zero. Moreover $\partial D/\partial\beta$ can be computed as follows

$$\frac{\partial D}{\partial \beta} = -\frac{x'Px}{n} + \frac{x'x}{n}\left(\frac{e'Pe}{e'e}\right) - \frac{x'e}{n}\left[\frac{\partial}{\partial \beta}\left(\frac{e'Pe}{e'e}\right)\right]$$

The derivative in brackets is the FOC for LIML which is zero under $\hat{\beta}$, then

$$\frac{\partial D}{\partial \beta} = -\frac{x'Px}{n} + \frac{x'x}{n}\left(\frac{e'Pe}{e'e}\right)$$

Using the previous results

$$\frac{\partial D}{\partial \beta} \xrightarrow{p} -(\pi'\Delta\pi + \alpha\sigma_v^2) + \alpha(\pi'\Delta\pi + \sigma_v^2) = (\alpha - 1)\pi'\Delta\pi$$

For valid instruments $p\lim(x'e/n) = p\lim(v'e/n) = \sigma_{ev}$, then $\hat{\gamma} \equiv x'e/e'e \xrightarrow{p} \sigma_{ev}/\sigma_e^2 \equiv \gamma$. Taking the fact that $e'Pe/e'e \xrightarrow{p} \alpha$ and defining $w = v - \gamma e$, then

$$\begin{aligned}
\sqrt{n}(\hat{\gamma} - \gamma)\frac{e'Pe}{n} &= \frac{(x'e - e'e\gamma)}{\sqrt{n}}\frac{e'Pe}{e'e} \\
&= \frac{e'Pe}{e'e}\frac{(x - \gamma e)'e}{\sqrt{n}} \\
&\xrightarrow{d} \alpha\frac{(z\pi + w)'e}{\sqrt{n}}
\end{aligned}$$

These results can be used to compute the distribution of $D$ as follows

$$\begin{aligned}
\sqrt{n}D &= \frac{(x - \gamma e)'Pe}{\sqrt{n}} - \sqrt{n}(\hat{\gamma} - \gamma)\frac{e'Pe}{n} \\
&\xrightarrow{d} \frac{(z\pi + w)'Pe}{\sqrt{n}} - \alpha\frac{(z\pi + w)'e}{\sqrt{n}} \\
&= \frac{(z\pi + w)'(P - \alpha I)e}{\sqrt{n}}
\end{aligned}$$

Using the expression above, it is possible to compute the variance for $\sqrt{n}D$ as follows $\sigma_e^2[(1 - \alpha)^2\pi'\Delta\pi + \alpha(1 - \alpha)E(w^2)]$, where $E(w^2) = \sigma_v^2 - \sigma_{ev}^2/\sigma_e^2$. Using the sandwich theorem, the variance for LIML is

$$
\begin{aligned}
V_{LIML} &= \left(\frac{\partial D}{\partial \beta}\right)^{-2} \sigma_e^2 \left[(1-\alpha)^2 \pi' \Delta \pi + \alpha(1-\alpha)\left(\sigma_v^2 - \frac{\sigma_{ev}^2}{\sigma_e^2}\right)\right] \\
&= \frac{(1-\alpha)^2 \pi' \Delta \pi \sigma_e^2 + \alpha(1-\alpha)\left(\sigma_e^2 \sigma_v^2 - \sigma_{ev}^2\right)}{[(1-\alpha)\pi' \Delta \pi]^2} \\
&= \frac{\sigma_e^2}{\pi' \Delta \pi} + \frac{\alpha}{1-\alpha}\left[\frac{\sigma_v^2 \sigma_e^2 - \sigma_{ev}^2}{(\pi' \Delta \pi)^2}\right]
\end{aligned}
$$

Finally, the asymptotic distribution for LIML is $\sqrt{n}(\hat{\beta}_{LIML} - \beta) \sim N\left(0, V_{LIML}\right)$.

**Optimality**

In this section, I will show that asymptotic variance for LIML ($V_{LIML}$) is the minimum variance that can be obtained from a linear combination between the B2SLS and R2SLS (reverse B2SLS). Define

$$
b = a b_{B2SLS} + (1-a) b_{R2SLS}
$$

and its asymptotic variance $V \equiv Var(b)$. Using the result obtained in Lemma A.4 the optimal weight for $a$ is

$$
a = \frac{V_{R2SLS} - C_{B,R}}{V_{B2SLS} - 2C_{B,R} + V_{R2SLS}}
$$

Using the results obtained for B2LS, the denominator for the minimum variance ($V$) is

$$
\begin{aligned}
V_{B2SLS} - 2C_{B,R} + V_{R2SLS} &= \frac{\sigma_e^2}{\pi' \Delta \pi} + \frac{\alpha}{1-\alpha}\left[\frac{\sigma_v^2 \sigma_e^2 + \sigma_{ev}^2}{(\pi' \Delta \pi)^2}\right] \\
&\quad -2\left\{\frac{\sigma_e^2}{\pi' \Delta \pi} + \frac{\alpha}{1-\alpha}\left[\frac{\sigma_{uv}\sigma_e^2 + \sigma_{eu}\sigma_{ev}}{\beta(\pi' \Delta \pi)^2}\right]\right\} \\
&\quad +\frac{\sigma_e^2}{\pi' \Delta \pi} + \frac{\alpha}{1-\alpha}\left[\frac{\sigma_u^2 \sigma_e^2 + \sigma_{eu}^2}{\beta^2(\pi' \Delta \pi)^2}\right] \\
&= \frac{\alpha}{1-\alpha}\left[\frac{\sigma_v^2 \sigma_e^2 + \sigma_{ev}^2}{(\pi' \Delta \pi)^2} - 2\frac{\sigma_{uv}\sigma_e^2 + \sigma_{eu}\sigma_{ev}}{\beta(\pi' \Delta \pi)^2} + \frac{\sigma_u^2 \sigma_e^2 + \sigma_{eu}^2}{\beta^2(\pi' \Delta \pi)^2}\right]
\end{aligned}
$$

and the numerator for $V$ is

$$
\begin{aligned}
V_{B2SLS}V_{R2SLS} - C_{B,V}^2 &= \left\{\frac{\sigma_e^2}{\pi'\Delta\pi} + \frac{\alpha}{1-\alpha}\left[\frac{\sigma_v^2\sigma_e^2+\sigma_{ev}^2}{(\pi'\Delta\pi)^2}\right]\right\} \\
&\quad \times \left\{\frac{\sigma_e^2}{\pi'\Delta\pi} + \frac{\alpha}{1-\alpha}\left[\frac{\sigma_u^2\sigma_e^2+\sigma_{eu}^2}{\beta^2(\pi'\Delta\pi)^2}\right]\right\} \\
&\quad - \left\{\frac{\sigma_e^2}{\pi'\Delta\pi} + \frac{\alpha}{1-\alpha}\left[\frac{\sigma_{uv}\sigma_e^2+\sigma_{eu}\sigma_{ev}}{\beta(\pi'\Delta\pi)^2}\right]\right\}^2 \\
&= \frac{\alpha}{1-\alpha}\left[\frac{\sigma_v^2\sigma_e^2+\sigma_{ev}^2}{(\pi'\Delta\pi)^2} - 2\frac{\sigma_{uv}\sigma_e^2+\sigma_{eu}\sigma_{ev}}{\beta(\pi'\Delta\pi)^2} + \frac{\sigma_u^2\sigma_e^2+\sigma_{eu}^2}{\beta^2(\pi'\Delta\pi)^2}\right]\frac{\sigma_e^2}{\pi'\Delta\pi} \\
&\quad + \left(\frac{\alpha}{1-\alpha}\right)^2\left[\frac{\sigma_v^2\sigma_e^2+\sigma_{ev}^2}{(\pi'\Delta\pi)^2}\right]\left[\frac{\sigma_u^2\sigma_e^2+\sigma_{eu}^2}{\beta^2(\pi'\Delta\pi)^2}\right] \\
&\quad - \left(\frac{\alpha}{1-\alpha}\right)^2 - \left[\frac{\sigma_{uv}\sigma_e^2+\sigma_{eu}\sigma_{ev}}{\beta(\pi'\Delta\pi)^2}\right]^2
\end{aligned}
$$

then the minimum variance is

$$
\begin{aligned}
V &= \frac{\sigma_e^2}{\pi'\Delta\pi} + \frac{\left(\frac{\alpha}{1-\alpha}\right)^2\left\{\left[\frac{\sigma_v^2\sigma_e^2+\sigma_{ev}^2}{(\pi'\Delta\pi)^2}\right]\left[\frac{\sigma_u^2\sigma_e^2+\sigma_{eu}^2}{\beta^2(\pi'\Delta\pi)^2}\right] - \left[\frac{\sigma_{uv}\sigma_e^2+\sigma_{eu}\sigma_{ev}}{\beta(\pi'\Delta\pi)^2}\right]^2\right\}}{\frac{\alpha}{1-\alpha}\left[\frac{\sigma_v^2\sigma_e^2+\sigma_{ev}^2}{(\pi'\Delta\pi)^2} - 2\frac{\sigma_{uv}\sigma_e^2+\sigma_{eu}\sigma_{ev}}{\beta(\pi'\Delta\pi)^2} + \frac{\sigma_u^2\sigma_e^2+\sigma_{eu}^2}{\beta^2(\pi'\Delta\pi)^2}\right]} \\
&= \frac{\sigma_e^2}{\pi'\Delta\pi} + \left[\frac{\alpha/(1-\alpha)}{(\pi'\Delta\pi)^2}\right]\left[\frac{(\sigma_v^2\sigma_e^2+\sigma_{ev}^2)(\sigma_u^2\sigma_e^2+\sigma_{eu}^2)-(\sigma_{uv}\sigma_e^2+\sigma_{eu}\sigma_{ev})^2}{\beta^2(\sigma_v^2\sigma_e^2+\sigma_{ev}^2)-2\beta(\sigma_{uv}\sigma_e^2+\sigma_{eu}\sigma_{ev})+(\sigma_u^2\sigma_e^2+\sigma_{eu}^2)}\right]
\end{aligned}
$$

Note that $\sigma_{ev} = \sigma_{uv} - \beta\sigma_v^2$ and $\sigma_{eu} = \sigma_u^2 - \beta\sigma_{uv}$, then $\sigma_{eu} - \beta\sigma_{ev} = \sigma_u^2 - 2\beta\sigma_{uv} + \beta^2\sigma_v^2 = \sigma_e^2$. Also, $\sigma_u^2\sigma_{ev}^2 - 2\sigma_{uv}\sigma_{eu}\sigma_{ev} + \sigma_v^2\sigma_{eu}^2 = \sigma_u^2(\sigma_{uv}^2 - 2\beta\sigma_{uv}\sigma_v^2 + \beta^2\sigma_v^4) - 2\sigma_{uv}(\sigma_u^2\sigma_{uv} - \beta\sigma_u^2\sigma_v^2 - \beta\sigma_{uv}^2 + \beta^2\sigma_{uv}\sigma_v^2) + \sigma_v^2(\sigma_u^4 - 2\beta\sigma_{uv}\sigma_u^2 + \beta^2\sigma_{uv}^2) = \sigma_u^4\sigma_v^2 + \beta^2\sigma_u^2\sigma_v^4 - 2\beta\sigma_u^2\sigma_v^2\sigma_{uv} + 2\beta\sigma_{uv}^3 - \sigma_u^2\sigma_{uv}^2 - \beta^2\sigma_v^2\sigma_{uv}^2 = (\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)(\sigma_u^2 - 2\beta\sigma_{uv} + \beta^2\sigma_v^2)$.

The denominator of the last expression in brackets is $\beta^2(\sigma_v^2\sigma_e^2 + \sigma_{ev}^2) - 2\beta(\sigma_{uv}\sigma_e^2 + \sigma_{eu}\sigma_{ev}) + (\sigma_u^2\sigma_e^2 + \sigma_{eu}^2) = \sigma_e^2(\sigma_u^2 - 2\beta\sigma_{uv} + \beta^2\sigma_v^2) + (\sigma_{eu}^2 - 2\beta\sigma_{eu}\sigma_{ev} + \beta^2\sigma_{ev}^2) = \sigma_e^4 + (\sigma_{eu} - \beta\sigma_{ev})^2 = 2\sigma_e^4$, and its numerator $(\sigma_v^2\sigma_e^2 + \sigma_{ev}^2)(\sigma_u^2\sigma_e^2 + \sigma_{eu}^2) - (\sigma_{uv}\sigma_e^2 + \sigma_{eu}\sigma_{ev})^2 = \sigma_e^4(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2) + \sigma_e^2(\sigma_u^2\sigma_{ev}^2 - 2\sigma_{uv}\sigma_{eu}\sigma_{ev} + \sigma_v^2\sigma_{eu}^2) = 2\sigma_e^4(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)$, then the expression is just $(\sigma_u^2\sigma_v^2 - \sigma_{uv}^2)$. With that the minimum variance is

$$
V = \frac{\sigma_e^2}{\pi'\Delta\pi} + \left(\frac{\alpha}{1-\alpha}\right)\left[\frac{\sigma_u^2\sigma_v^2 - \sigma_{uv}^2}{(\pi'\Delta\pi)^2}\right]
$$

Finally, note that $\sigma_v^2\sigma_e^2 - \sigma_{ev}^2 = \sigma_v^2(\sigma_u^2 - 2\beta\sigma_{uv} + \beta^2\sigma_v^2) - (\sigma_{uv} - \beta\sigma_v^2)^2 = \sigma_u^2\sigma_v^2 - \sigma_{uv}^2$, then $V$ is exactly the same as the asymptotic variance obtained for LIML ($V_{LIML}$).

## A.3 JIVE and RLML

In this section, I show the asymptotic distribution for JIVEs (forward and reverse) and RLML under general errors.

The robustness to heteroskedasticity of RLML is obtained modifying the numerator of the objective function for LIML. Define $R = P - diag(P)$, the new objective function for RLML is $e_c'Re_c/e_c'e_c$. In the similar way as LIML the value which minimizes this function ($r$) is the smallest eigenvalue of $(W'W)^{-1}W'RW$, then the estimator can be written as

$$\hat{\beta}_{RLML} = \frac{x'Ry - rx'y}{x'Rx - rx'x}$$

As the same as LIML and LIMLF, a finite sample modification can be applied to RLML. The modified RLML (called RFLL by Hausman, Newey, Woutersen, Chao, and Swanson (2007)) can be defined as

$$\hat{\beta}_{RFLL} = \frac{x'Ry - sx'y}{x'Rx - sx'x}$$

where $s = [(n+1)r - 1]/[(n-1) + r]$. In addition, the JIVE proposed by Angrist, Imbens and Krueger (1999) is defined as[7]

$$\hat{\beta}_{JIVE} = \frac{x'Ry}{x'Rx}$$

It is clear that RLML is similar to JIVE when $r = 0$. I will show next that $r \xrightarrow{p} 0$, then the relationship between RLML and JIVE is similar to the one between LIML and B2SLS under homoscedastic errors and BAA (many instruments).

For the purpose of efficiency I will consider as well the reverse JIVE, which is defined as $1/\hat{\beta}_{RJIVE} = y'Ry/x'Ry$. Also the regular JIVE will be called forward JIVE following the

[7] Here I am using JIVE2 in nomenclature of Angrist, Imbens and Krueger (1999).

nomenclature of Hahn and Hausman (2002).

In the following, I will consider the model (1), the number of instruments ($K$) grows along with the sample size ($n$) but $K/n \to \alpha$ (BAA) and Condition 2.1.

### Consistency of JIVE and RLML

Using Lemma A.1 it is easy to prove that $\bar{e}'R\bar{e}/\bar{e}'\bar{e} \xrightarrow{p} 0$. Moreover, adding the results proved in Lemma A.2 it is clear that

$$\begin{aligned} p\lim\left(\frac{x'Ry}{n}\right) &= p\lim\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i R_{ij} y_j\right) \\ &= p\lim\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} \beta\pi' z_i P_{ij} z_j'\pi\right) \\ &= \beta\pi'\Delta^*\pi \end{aligned}$$

$$\begin{aligned} p\lim\left(\frac{x'Rx}{n}\right) &= p\lim\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i R_{ij} x_j\right) \\ &= p\lim\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} \pi' z_i P_{ij} z_j'\pi\right) \\ &= \pi'\Delta^*\pi \end{aligned}$$

$$\begin{aligned} p\lim\left(\frac{y'Ry}{n}\right) &= p\lim\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i R_{ij} y_j\right) \\ &= p\lim\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} \beta^2\pi' z_i P_{ij} z_j'\pi\right) \\ &= \beta^2\pi'\Delta^*\pi \end{aligned}$$

**Theorem A.1.** *Under BAA and general errors, JIVE (reverse and forward), and RLML are consistent estimators for $\beta$ in model 1.*

*Proof.* For JIVEs the proof is trivial taking the results presented above. For RLML the consistency holds given that $r \xrightarrow{p} 0$. $\square$

Now, I will compute the Asymptotic Distribution for the forward and reverse JIVEs and for RLML. It will be clear that the asymptotic variances for these estimator are very close to the ones obtained for B2SLS and LIML under homoscedastic errors, then the estimators are not losing efficiency in that case.

**Asymptotic Distribution for JIVE**

For the asymptotic distribution, consider the following relations

$$\sqrt{n}\left(\hat{\beta}_{JIVE} - \beta\right) = \frac{\frac{1}{\sqrt{n}}x'Re}{\frac{1}{n}x'Rx} \quad \text{and} \quad \sqrt{n}\left(\frac{1}{\hat{\beta}_{RJIVE}} - \beta\right) = \frac{\frac{1}{\sqrt{n}}y'Re}{\frac{1}{n}x'Ry}$$

By Lemma A.1 and Lemma A.2, it is clear that the numerators of the expressions above are zero, then $Var(x'Re/\sqrt{n}) = E[(x'Re/\sqrt{n})^2]$, $Var(y'Re/\sqrt{n}) = E[(y'Re/\sqrt{n})^2]$ and $Cov(x'Re/\sqrt{n}, y'Re/\sqrt{n}) = E[(x'Re/\sqrt{n})(y'Re/\sqrt{n})]$.

**Lemma A.5.** *Consider $R = P - diag(P)$.*

$$\begin{aligned}
E\left(\frac{x'Ree'Rx}{n}\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} x_i R_{ij} e_j x_k R_{kl} e_l\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq j} \pi' z_i z_k' \pi P_{ij} P_{jk} E(e_j^2) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} P_{ij}^2 [E(v_i^2)E(e_j^2) + E(v_i e_i)E(v_j e_j)] \\
E\left(\frac{y'Ree'Ry}{n}\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} y_i R_{ij} e_j y_k R_{kl} e_l\right) \\
&= \beta^2 \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq j} \pi' z_i z_k' \pi P_{ij} P_{jk} E(e_j^2) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} P_{ij}^2 [E(u_i^2)E(e_j^2) + E(u_i e_i)E(u_j e_j)] \\
E\left(\frac{y'Ree'Rx}{n}\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} y_i R_{ij} e_j x_k R_{kl} e_l\right) \\
&= \beta \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq j} \pi' z_i z_k' \pi P_{ij} P_{jk} E(e_j^2) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i} P_{ij}^2 [E(u_i v_i)E(e_j^2) + E(u_i e_i)E(v_j e_j)]
\end{aligned}$$

*Proof.* For the first expression, note that $E(x_i R_{ij} e_j x_k R_{kl} e_l) = \pi' z_i z_k' \pi R_{ij} R_{lk} E(e_j e_l) + R_{ij} R_{lk} E(v_i e_j e_l v_k)$ (under normality $R_{ij} R_{lk} z_k' \pi E(v_i e_j e_l) = R_{ij} R_{lk} z_i' \pi E(e_j e_l v_k) = 0$). Also, note that $E(e_j e_l)$ is not zero when $l = j$, which yields $E(e_j^2)$. Moreover, $E(v_i e_j e_l v_k) = E(v_i^2)E(e_j^2)$ when $k = i$ and $l = j$, but also $E(v_i e_j e_l v_k) = E(v_i e_i)E(e_j v_j)$ when $l = i$ and $k = j$, then $E(x_i R_{ij} e_j x_k R_{kl} e_l) = \pi' z_i z_k' \pi P_{ij} P_{kj} E(e_j^2) + P_{ij}^2[E(v_i^2)E(e_j^2) + E(v_i e_i)E(e_j v_j)]$ when $i \neq j$ and $k \neq j$.

For the second expression, $E(y_i R_{ij} e_j y_k R_{kl} e_l) = R_{ij} R_{lk}[\beta^2 E(x_i e_j e_l x_k) + \beta E(e_i e_j e_l x_k) + \beta E(x_i e_j e_l e_k) + E(e_i e_j e_k e_l)$. The first argument is exactly the same as the previous expression, then the new terms are $R_{ij} R_{lk} E(x_i e_j e_l e_k) = R_{ij} R_{lk} E(v_i e_j e_l e_k) = R_{ij}^2 E(v_i e_i) E(e_j^2)$ (using the normality assumption) and $R_{ij} R_{lk} E(e_i e_j e_k e_l) = R_{ij}^2 E(e_i^2) E(e_j^2)$ (the fourth moment is multiplied by zero). Then $E(y_i R_{ij} e_j y_k R_{kl} e_l) = \pi' z_i z_k' \pi P_{ij} P_{jk} E(e_j^2) + P_{ij}^2[E(e_i^2)E(e_j^2) + 2\beta E(v_i e_i)E(e_j^2) + \beta^2 E(v_i^2)E(e_j^2) + \beta^2 E(v_i e_i)E(v_j e_j)]$. The latter expression can be written using the $u$ error instead of $v$ as follow $E(y_i R_{ij} e_j y_k R_{kl} e_l) = \pi' z_i z_k' \pi P_{ij} P_{jk} E(e_j^2) + P_{ij}^2[E(u_i^2)E(e_j^2) + E(u_i e_i)E(u_j e_j)]$, using $e_i = u_i - \beta v_i$.

For the last term note that $E(y_i R_{ij} e_j x_k R_{kl} e_l) = R_{ij} R_{lk}[\beta E(x_i e_j e_l x_k) + E(x_i e_j e_k e_l)] = \pi' z_i z_k' \pi P_{ij} P_{jk} E(e_j^2) + P_{ij}^2[E(v_i e_i)E(e_j^2) + \beta E(v_i^2)E(e_j^2) + \beta E(v_i e_i)E(v_j e_j)]$. This is also equivalent to $E(y_i R_{ij} e_j x_k R_{kl} e_l) = \pi' z_i z_k' \pi P_{ij} P_{jk} E(e_j^2) + P_{ij}^2[E(u_i v_i)E(e_j^2) + E(u_i e_i)E(v_j e_j)]$. □

**Theorem A.2.** *Under BAA and general errors, the asymptotic variance covariance matrix for JIVEs (reverse and forward) is composed by*

$$nVar(\hat{\beta}_{JIVE}) = \frac{V_0 + V_1}{(\pi'\Delta^*\pi)^2}\ ,\ nVar(1/\hat{\beta}_{RJIVE}) = \frac{\beta^2 V_0 + V_2}{\beta^2(\pi'\Delta^*\pi)^2} \quad \textit{and}$$
$$nCov(\hat{\beta}_{JIVE}, 1/\hat{\beta}_{RJIVE}) = \frac{\beta V_0 + V_3}{\beta(\pi'\Delta^*\pi)^2}$$

*where*

$$\begin{aligned}
V_0 &= \lim_{K,n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j\neq i} \sum_{k\neq j} \pi' z_i z_k' \pi P_{ij} P_{jk} E(e_j^2) \right\} \\
V_1 &= \lim_{K,n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j\neq i} P_{ij}^2 [E(v_i^2)E(e_j^2) + E(v_i e_i)E(v_j e_j)] \right\} \\
V_2 &= \lim_{K,n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j\neq i} P_{ij}^2 [E(u_i^2)E(e_j^2) + E(u_i e_i)E(u_j e_j)] \right\} \\
V_3 &= \lim_{K,n\to\infty} \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j\neq i} P_{ij}^2 [E(u_i v_i)E(e_j^2) + E(u_i e_i)E(v_j e_j)] \right\}
\end{aligned}$$

*Proof.* The proof is based on the results presented in Lemma A.5. □

### Asymptotic Distribution for RLML

Following Newey (2004), the asymptotic distribution for RLML can be obtained using the First Order Condition

$$
\begin{aligned}
0 &= -x'R\hat{e}_d(\hat{e}_d'\hat{e}_d) + (\hat{e}_d'R\hat{e}_d)x'\hat{e}_d \\
0 &= -\frac{x'R\hat{e}_d}{n} + \left(\frac{\hat{e}_d'R\hat{e}_d}{\hat{e}_d'\hat{e}_d}\right)\frac{x'\hat{e}_d}{n} \equiv -H
\end{aligned}
$$

Note that by construction $H$ is zero under the optimal choice of $\beta$. In addition,

$$
\begin{aligned}
\frac{\partial H}{\partial \beta} &= -\frac{x'Rx}{n} + \frac{x'x}{n}\left(\frac{e'Re}{e'e}\right) - \frac{x'e}{n}\left[\frac{\partial}{\partial \beta}\left(\frac{e'Re}{e'e}\right)\right] \\
&= -\frac{x'Rx}{n} + \frac{x'x}{n}\left(\frac{e'Re}{e'e}\right) \\
&\xrightarrow{p} -\pi'\Delta^*\pi
\end{aligned}
$$

Last line is obtained using the facts that $e'Re/e'e \xrightarrow{p} 0$ and

$$
\frac{x'x}{n} \xrightarrow{p} \pi'\Delta\pi + \lim_{n\to\infty}\left(\frac{1}{n}\sum_i E(v_i^2)\right).
$$

For the asymptotic variance consider

$$
\begin{aligned}
\sqrt{n}H &= \frac{x'Re}{\sqrt{n}} - \left(\frac{e'Re}{e'e}\right)\frac{x'e}{\sqrt{n}} = \frac{x'Re}{\sqrt{n}} - \left(\frac{x'e}{e'e}\right)\frac{e'Re}{\sqrt{n}} = \frac{x'Re}{\sqrt{n}} - \hat{\gamma}\frac{e'Re}{\sqrt{n}} \\
&= \frac{x'Re}{\sqrt{n}} - \gamma\frac{e'Re}{\sqrt{n}} + O_p(1) = \frac{(x-\gamma e)'Re}{\sqrt{n}} + O_p(1) = \frac{\hat{x}'Re}{\sqrt{n}} + O_p(1)
\end{aligned}
$$

For the variance of $\sqrt{n}H$ consider that $\hat{x}_i \equiv x_i - \gamma e_i = z_i'\pi + w_i$, where $w_i \equiv v_i - \gamma e_i$. From the results obtained for JIVEs the asymptotic variance for RLML is

$$
V_{RLML} = \frac{V_0 + \lim_{K,n\to\infty}\left[\frac{1}{n}\sum_{i=1}^n\sum_{j\neq i}P_{ij}^2[E(w_i^2)E(e_j^2) + E(w_ie_i)E(w_je_j)]\right]}{(\pi'\Delta^*\pi)^2}
$$

The main difference between this result and the one obtained for JIVE is the error $w$. Note that under homoscedastic errors $E(w_ie_i) = E[(v_i - \gamma e_i)e_i] = \sigma_{ev} - \gamma\sigma_e^2 = \sigma_{ev} - (\sigma_{ev}/\sigma_e^2)\sigma_e^2 = 0$ and $E(w_i^2) = \sigma_v^2 - \sigma_{ev}^2/\sigma_e^2$, then the second expression in the numerator reduces to $(\sigma_v^2\sigma_e^2 - \sigma_{ev}^2)[(K/n - (1/n)\sum_{i=1}^n P_{ii}^2]$.

## A.4 Proofs

**Proof of Theorem 2.1**

Note that $\hat{\beta} - \beta = (x'Se)/(x'Sx)$, then

$$\frac{x'Se}{n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i S_{ij} e_j = \pi'\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i S_{ij} e_j\right) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij} v_i e_j$$

and

$$\begin{aligned}
\left(\frac{x'Sx}{n}\right)^{-1} &= \left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i S_{ij} x_j\right]^{-1} \\
&= \left[\pi'\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i S_{ij} z_j'\right)\pi + \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}\pi' z_i v_j + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij} v_i v_j\right]^{-1} \\
&= \left[\pi' A\pi + \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}\pi' z_i v_j + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij} v_i v_j + O\left(\frac{1}{n}\right)\right]^{-1} \\
&\approx \frac{1}{\pi' A\pi} - \left[\frac{2\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}\pi' z_i v_j + \sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij} v_i v_j}{n(\pi' A\pi)^2}\right]
\end{aligned}$$

with the previous results and Condition 2.2

$$\begin{aligned}
\left(\frac{x'Sx}{n}\right)^{-1}\left(\frac{x'Se}{n}\right) &= \frac{1}{\pi' A\pi}\left[\pi'\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} z_i S_{ij} e_j\right) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij} v_i e_j\right] \\
&\quad - \frac{1}{(\pi' A\pi)^2}\left[\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}\pi' z_i v_j\right]\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij} e_i z_j'\pi\right) + O_p\left(\frac{1}{n}\right)
\end{aligned}$$

Taking expectation

$$E(\hat{\beta} - \beta) = \frac{1}{(n\pi' A\pi)}\sum_{i=1}^{n} S_{ii} E(e_i v_i) - \frac{2}{(n\pi' A\pi)^2}\pi'\left[\sum_{i=1}^{n}\sum_{j=1}^{n} z_i S_{ij}^2 z_j' E(e_i v_j)\right]\pi + O\left(\frac{1}{n}\right)$$

Under Condition 2.3, only the first term in the expression does not converge to zero.

**Proof of Theorem 3.1**

Using the Edgeworth expansion for $M$ presented in the chapter, and the standard asymptotic approximation for 2SLS estimator, which is:

$$\hat{\beta} - \beta = \frac{\pi'}{\pi'\Delta\pi}\left(\frac{1}{n}\sum_{i=1}^{n} z_i e_i\right).$$

We can approximate the robust asymptotic variance as follows:

$$\begin{aligned}
\hat{V}(\hat{\beta}) &= n\left[\frac{x'Z(Z'Z)^{-1}}{x'Px}\right]\left(\sum_{i=1}^{n}\hat{e}_i^2 z_i z_i'\right)\left[\frac{(Z'Z)^{-1}Z'x}{x'Px}\right] \\
&= \left[\frac{(x'Z/n)(Z'Z/n)^{-1}}{(x'Px/n)}\right]\left(\frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^2 z_i z_i'\right)\left[\frac{(Z'Z/n)^{-1}(Z'x/n)}{(x'Px/n)}\right] \\
&= \left[\frac{\pi'\Delta\Delta^{-1}}{\pi'\Delta\pi}\right]\left(\frac{1}{n}\sum_{i=1}^{n}\hat{e}_i^2 z_i z_i'\right)\left[\frac{\Delta^{-1}\Delta\pi}{\pi'\Delta\pi}\right] + O_p\left(\frac{1}{n}\right) \\
&= \frac{1}{(\pi'\Delta\pi)^2}\pi'\left[\frac{1}{n}\sum_{i=1}^{n} e_i^2 z_i z_i' - 2(\hat{\beta}-\beta)\frac{1}{n}\sum_{i=1}^{n} e_i x_i z_i z_i'\right]\pi \\
&\quad + \frac{1}{(\pi'\Delta\pi)^2}\pi'\left[(\hat{\beta}-\beta)^2\frac{1}{n}\sum_{i=1}^{n} x_i^2 z_i z_i'\right]\pi + O_p\left(\frac{1}{n}\right) \\
&= \frac{1}{(\pi'\Delta\pi)^2}\left[\frac{1}{n}\sum_{i=1}^{n} e_i^2\tau_i^2 - 2\frac{\pi'}{\pi'\Delta\pi}\left(\frac{1}{n}\sum_{i=1}^{n} z_i e_i\right)\frac{1}{n}\sum_{i=1}^{n} e_i x_i \tau_i^2\right] + O_p\left(\frac{1}{n}\right) \\
&= \frac{1}{(\pi'\Delta\pi)^2}\left[\frac{1}{n}\sum_{i=1}^{n} e_i^2\tau_i^2 - 2\frac{1}{n^2(\pi'\Delta\pi)}\sum_{i=1}^{n} e_i^2\tau_i^4\right] + O_p\left(\frac{1}{n}\right) \\
&= \frac{1}{(\pi'\Delta\pi)^2}\frac{1}{n}\sum_{i=1}^{n} e_i^2\tau_i^2\left[1 - \frac{2\tau_i^2}{n(\pi'\Delta\pi)}\right] + O_p\left(\frac{1}{n}\right),
\end{aligned}$$

where $\tau_i^2 \equiv \pi' z_i z_i' \pi$.

**Proof of Theorem 3.2**

By Taylor expansion around $z_0 = 0$ it is clear that

$$
\begin{aligned}
\frac{1}{\sqrt{1+z}} &= \frac{1}{\sqrt{1+z_0}} - \frac{1}{2}\frac{1}{\sqrt{(1+z_0)^3}}(z-z_0) + \frac{3}{8}\frac{1}{\sqrt{(1+z_0)^5}}(z-z_0)^2 + O\left((z-z_0)^2\right) \\
&= 1 - \frac{1}{2}z + \frac{3}{8}z^2 + O\left(z^2\right)
\end{aligned}
$$

where the last equality is obtained under $z_0 = 0$. Applying this argument to $T_n$

$$
\begin{aligned}
T_n &= \left[\frac{(\overline{\beta}-b)+(\hat{\beta}-\overline{\beta})}{\sigma_{\overline{\beta}}}\right]\left[1+\left(\frac{\hat{\sigma}^2_{\hat{\beta}}-\sigma^2_{\overline{\beta}}}{\sigma^2_{\overline{\beta}}}\right)\right]^{-1/2} \\
&\approx \left[Z_n + \frac{(\hat{\beta}-\overline{\beta})}{\sigma_{\overline{\beta}}}\right]\left[1-\frac{1}{2}\left(\frac{\hat{\sigma}^2_{\hat{\beta}}-\sigma^2_{\overline{\beta}}}{\sigma^2_{\overline{\beta}}}\right)+\frac{3}{8}\left(\frac{\hat{\sigma}^2_{\hat{\beta}}-\sigma^2_{\overline{\beta}}}{\sigma^2_{\overline{\beta}}}\right)^2\right] \\
&= Z_n - \frac{Z_n}{2}\left(\frac{\hat{\sigma}^2_{\hat{\beta}}-\sigma^2_{\overline{\beta}}}{\sigma^2_{\overline{\beta}}}\right) + \frac{3Z_n}{8}\left(\frac{\hat{\sigma}^2_{\hat{\beta}}-\sigma^2_{\overline{\beta}}}{\sigma^2_{\overline{\beta}}}\right)^2 + \left(\frac{\hat{\beta}-\overline{\beta}}{\sigma_{\overline{\beta}}}\right) - \frac{1}{2}\left(\frac{\hat{\beta}-\overline{\beta}}{\sigma_{\overline{\beta}}}\right)\left(\frac{\hat{\sigma}^2_{\hat{\beta}}-\sigma^2_{\overline{\beta}}}{\sigma^2_{\overline{\beta}}}\right)
\end{aligned}
$$

where higher order terms were discarded. Finally, the result is obtained replacing $A$ and $B$.

**Proof of Theorem 3.3**

First, we consider the case of biased estimators. In particular the moments of $\hat{V}(\hat{\beta})$ for 2SLS under homoskedastic and symmetric errors are

$$
\begin{aligned}
E[\hat{V}(\hat{\beta})] &= \frac{1}{(\pi'\Delta\pi)^2}\frac{1}{n}\sum_{i=1}^{n}E(e_i^2)\tau_i^2\left[1-\frac{2\tau_i^2}{n(\pi'\Delta\pi)}\right] + O\left(\frac{1}{n}\right) \\
&= \frac{\sigma_e^2}{(\pi'\Delta\pi)^2}\left[\frac{1}{n}\sum_{i=1}^{n}\tau_i^2 - \frac{2}{n(\pi'\Delta\pi)}\frac{1}{n}\sum_{i=1}^{n}\tau_i^4\right] + O\left(\frac{1}{n}\right) \\
&= \frac{\sigma_e^2}{\pi'\Delta\pi} - \frac{2\sigma_e^2}{n\pi'\Delta\pi} + O\left(\frac{1}{n}\right) = \frac{\sigma_e^2}{\pi'\Delta\pi} + O\left(\frac{1}{n}\right)
\end{aligned}
$$

$$
\begin{aligned}
E[\hat{V}(\hat{\beta})^2] &= \frac{1}{n^2(\pi'\Delta\pi)^4}\left[\sum_{i=1}^{n} E(e_i^4)\tau_i^4 + 2\sum_{i=1}^{n}\sum_{j\neq i} E(e_i^2)E(e_j^2)\tau_i^2\tau_j^2\right] + O\left(\frac{1}{n}\right) \\
&= \frac{\kappa\sigma_e^4}{n^2(\pi'\Delta\pi)^4}\left[\sum_{i=1}^{n}\tau_i^4 + \sum_{i=1}^{n}\sum_{j\neq i}\tau_i^2\tau_j^2\right] + O\left(\frac{1}{n}\right) \\
&= \frac{\kappa\sigma_e^4}{(\pi'\Delta\pi)^2} + O\left(\frac{1}{n}\right)
\end{aligned}
$$

where $\kappa$ represents the excess of kurtosis of the error terms. For the case of errors normally distributed $\kappa = 3$. From previous results B2SLS, LIML and LIMLF are asymptotically unbiased under homoscedastic errors and many instruments. However, LIML is asymptotically more efficient than B2SLS with asymptotic variance

$$
V_{LIML} = \frac{\sigma_e^2}{\pi'\Delta\pi} + \frac{\alpha}{1-\alpha}\left[\frac{\sigma_v^2\sigma_e^2 - \sigma_{ev}^2}{(\pi'\Delta\pi)^2}\right]
$$

Using previous results, the following relations hold

$$
E[\hat{V}(\hat{\beta}) - V_{LIML}] = -\frac{\alpha}{1-\alpha}\left[\frac{\sigma_v^2\sigma_e^2 - \sigma_{ev}^2}{(\pi'\Delta\pi)^2}\right] + O\left(\frac{1}{n}\right) \leq 0
$$

$$
\begin{aligned}
Var[\hat{V}(\hat{\beta})] &= \frac{\kappa\sigma_e^4}{(\pi'\Delta\pi)^2} - \left[\frac{\sigma_e^2}{\pi'\Delta\pi}\right]^2 + O\left(\frac{1}{n}\right) \\
&= \frac{(\kappa-1)\sigma_e^4}{(\pi'\Delta\pi)^2} + O\left(\frac{1}{n}\right)
\end{aligned}
$$

The first relation shows that the robust variance is expected to be lower than the asymptotic variance with many instruments. Only for large $n$ and a fixed number of instruments ($\alpha = 0$) the inequality holds. In terms of inference a lower standard error implies that the null will be rejected more often than the true nominal size, regardless of the bias of the estimator.

The second relation implies that the effect in the standard error only affects the first moment of the estimated variance. In this case $E(A_n/\sqrt{n})$ represents the normalized bias obtained in Theorem 2.1, then the following expression can be computed

$$E\left(\frac{A_n}{\sqrt{n}}+\frac{B_n}{n}\right) = E\left\{\xi\left[1-\frac{1}{2}\left(\frac{\tilde{\sigma}^2-\sigma^2}{\sigma^2}\right)\right]\right\}$$

The moments of the factor $(\tilde{\sigma}^2-\sigma^2)$ can be computed as follows

$$\begin{aligned}
E\left(\frac{\tilde{\sigma}^2-\sigma^2}{\sigma^2}\right) &= -\left\{\frac{2\sigma_e^2}{n\pi'\Delta\pi}+\left[\frac{\alpha(\sigma_v^2\sigma_e^2-\sigma_{ev}^2)}{(1-\alpha)(\pi'\Delta\pi)^2}\right]\right\}\left\{\left(\frac{\sigma_e^2}{\pi'\Delta\pi}\right)+\left[\frac{\alpha(\sigma_v^2\sigma_e^2-\sigma_{ev}^2)}{(1-\alpha)(\pi'\Delta\pi)^2}\right]\right\}^{-1} \\
&\approx -\left[\frac{2\sigma_e^2}{n\pi'\Delta\pi}+\frac{\alpha}{1-\alpha}\left(\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{(\pi'\Delta\pi)^2}\right)\right]\left(\frac{\sigma_e^2}{\pi'\Delta\pi}\right)^{-1} \\
&\quad +\left[\frac{2\sigma_e^2}{n\pi'\Delta\pi}+\frac{\alpha(\sigma_v^2\sigma_e^2-\sigma_{ev}^2)}{(1-\alpha)(\pi'\Delta\pi)^2}\right]\left[\frac{\alpha}{1-\alpha}\left(\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{(\pi'\Delta\pi)^2}\right)\left(\frac{\sigma_e^2}{\pi'\Delta\pi}\right)^{-2}\right] \\
&= -\left[\frac{2}{n}+\frac{\alpha}{1-\alpha}\left(\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{\sigma_e^2\pi'\Delta\pi}\right)-\frac{2\alpha}{1-\alpha}\left(\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{n\pi'\Delta\pi\sigma_e^2}\right)\right] \\
&= -\frac{2}{n}-\frac{\alpha}{1-\alpha}\left(\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{\sigma_e^2\pi'\Delta\pi}\right)\left(1+\frac{2}{n}\right) \\
&= -\frac{\alpha}{1-\alpha}\underbrace{\left(\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{\sigma_e^2\pi'\Delta\pi}\right)}_{\varsigma}+O\left(\frac{1}{n}\right)
\end{aligned}$$

and

$$\begin{aligned}
E\left[\left(\frac{\tilde{\sigma}^2-\sigma^2}{\sigma^2}\right)^2\right] &= \left\{\left[\frac{(\kappa-1)\sigma_e^4}{(\pi'\Delta\pi)^2}\right]+\left[\frac{2\sigma_e^2}{n\pi'\Delta\pi}+\frac{\alpha}{1-\alpha}\left(\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{(\pi'\Delta\pi)^2}\right)\right]^2\right\} \\
&\quad \times\left[\left(\frac{\sigma_e^2}{\pi'\Delta\pi}\right)+\frac{\alpha}{1-\alpha}\left(\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{(\pi'\Delta\pi)^2}\right)\right]^{-2}+O\left(\frac{1}{n}\right) \\
&\approx \frac{\sigma_e^4}{(\pi'\Delta\pi)^2}\left\{\left(\kappa-1+\frac{4}{n^2}\right)+\frac{4\alpha}{1-\alpha}\left[\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{n\sigma_e^2\pi'\Delta\pi}\right]\right\} \\
&\quad \times\left\{\frac{\sigma_e^4}{(\pi'\Delta\pi)^2}+\frac{2\alpha\sigma_e^2}{1-\alpha}\left[\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{(\pi'\Delta\pi)^3}\right]\right\}^{-1}+O\left(\frac{1}{n}\right) \\
&\approx \left\{(\kappa-1)+\frac{4\alpha}{1-\alpha}\left[\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{n\sigma_e^2\pi'\Delta\pi}\right]\right\}\left\{1+\frac{2\alpha}{1-\alpha}\left(\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{\sigma_e^2\pi'\Delta\pi}\right)\right\}^{-1} \\
&= \left\{(\kappa-1)+\frac{4\alpha}{1-\alpha}\left[\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{n\sigma_e^2\pi'\Delta\pi}\right]-\frac{2(\kappa-1)\alpha}{1-\alpha}\left(\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{\sigma_e^2\pi'\Delta\pi}\right)\right\} \\
&= (\kappa-1)\left\{1-\frac{2\alpha}{1-\alpha}\left(\frac{\sigma_v^2\sigma_e^2-\sigma_{ev}^2}{\sigma_e^2\pi'\Delta\pi}\right)\left(1-\frac{2}{n(\kappa-1)}\right)\right\} \\
&= (\kappa-1)\left[1-\frac{2\alpha}{1-\alpha}\varsigma\right]+O\left(\frac{1}{n}\right)
\end{aligned}$$

Let $\xi \equiv E(\hat{\beta} - \overline{\beta})/\sigma$ be the normalized bias, then

$$
\begin{aligned}
E\left(\frac{A_n}{\sqrt{n}} + \frac{B_n}{n}\right) &= E\left\{\frac{A_n}{\sqrt{n}}\left[1 - \frac{1}{2}\left(\frac{\tilde{\sigma}^2 - \sigma^2}{\sigma^2}\right)\right]\right\} \\
&= \xi\left[1 + \frac{1}{n} + \frac{\alpha}{2(1-\alpha)}\left(\frac{\sigma_v^2\sigma_e^2 - \sigma_{ev}^2}{\sigma_e^2\pi'\Delta\pi}\right)\left(1 + \frac{2}{n}\right)\right] \\
&\approx \xi\left[1 + \frac{\alpha}{2(1-\alpha)}\varsigma\right] \\
Var\left(A_n\right) &= 1 + \frac{1}{4}\left[Var\left(\frac{\tilde{\sigma}^2 - \sigma^2}{\sigma^2}\right)\right] \\
&= 1 + \frac{1}{4}\left\{E\left[\left(\frac{\tilde{\sigma}^2 - \sigma^2}{\sigma^2}\right)^2\right] - \left[E\left(\frac{\tilde{\sigma}^2 - \sigma^2}{\sigma^2}\right)\right]^2\right\} \\
&= 1 + \frac{\kappa - 1}{4}\left(1 - \frac{2\alpha}{1-\alpha}\varsigma\right) > 1
\end{aligned}
$$

For the cases of unbiased estimators, $\hat{\beta}$ can be used as $\overline{\beta}$, then only the adjustment to the variance of $T_n$ must be computed. The original $T_n$ uses a different standard error, that could be higher or lower depending on the combination of the parameter of the model: number of instruments over sample size ($\alpha$), concentration parameter ($\pi'\Delta\pi$), and the correlation between error terms ($\sigma_{ev}$). However, the size distortion is small under homoskedastic errors.

| **Documentos de Trabajo Banco Central de Chile** | **Working Papers Central Bank of Chile** |
|---|---|
| NÚMEROS ANTERIORES | PAST ISSUES |

La serie de Documentos de Trabajo en versión PDF puede obtenerse gratis en la dirección electrónica: www.bcentral.cl/esp/estpub/estudios/dtbc. Existe la posibilidad de solicitar una copia impresa con un costo de $500 si es dentro de Chile y US$12 si es para fuera de Chile. Las solicitudes se pueden hacer por fax: (56-2) 6702231 o a través de correo electrónico: bcch@bcentral.cl.

Working Papers in PDF format can be downloaded free of charge from: www.bcentral.cl/eng/stdpub/studies/workingpaper. Printed versions can be ordered individually for US$12 per copy (for orders inside Chile the charge is Ch$500.) Orders can be placed by fax: (56-2) 6702231 or e-mail: bcch@bcentral.cl.

DTBC-463 Abril 2008
**Nonlinear Dynamic in the Chilean Stock Market: Evidence from Returns and Trading Volume**
Rodrigo Aranda y Patricio Jaramillo

DTBC-462 Abril 2008
**Medidas de Volatilidad de Índices Accionarios: El Caso del IPSA**
Rodrigo Alfaro y Carmen Gloria Silva

DTBC-461 Marzo 2008
**Medidas Extendidas de Restricciones a Los Flujos de Capitales**
Jorge Selaive, Beatriz Velásquez y José Miguel Villena

DTBC-460 Marzo 2008
**External Imbalances, Valuation Adjustments and Real Exchange Rate: Evidence of Predictability in an Emerging Economy**
Pablo Pincheira y Jorge Selaive

DTBC-459 Febrero 2008
**Combining Tests of Predictive Ability Theory and Evidence for Chilean and Canadian Exchange Rates**
Pablo Pincheira

DTBC-458 Diciembre 2007
**Copper Price, Fiscal Policy and Business Cycle in Chile**
Juan Pablo Medina y Claudio Soto

DTBC-457 Diciembre 2007
**The Chilean Business Cycles Through the Lens of a Stochastic General Equilibrium Model**
Juan Pablo Medina y Claudio Soto

DTBC-456 Diciembre 2007
**Is Ownership Structure a Determinant of Bank Efficiency?**
Rodrigo Fuentes y Marcos Vergara

DTBC-455 Diciembre 2007
**Estimating the Output Gap for Chile**
Rodrigo Fuentes, Fabián Gredig y Mauricio Larraín

DTBC-454 Diciembre 2007
**Un Nuevo Marco Para la Elaboración de los Programas de Impresión y Acuñación**
Rómulo Chumacero, Claudio Pardo y David Valdés

DTBC-453 Diciembre 2007
**Development Paths and Dynamic Comparative Advantages: When Leamer Met Solow**
Rodrigo Fuentes y Verónica Mies

DTBC-452 Diciembre 2007
**Experiences With Current Account Deficits in Southeast Asia**
Ramon Moreno

DTBC-451 Diciembre 2007
**Asymmetric Monetary Policy Rules and the Achievement of the Inflation Target: The Case of Chile**
Fabián Gredig

DTBC-450 Diciembre 2007
**Current Account Deficits: The Australian Debate**
Rochelle Belkar, Lynne Cockerell y Christopher Kent

DTBC-449 Diciembre 2007
**International Reserves Management and the Current Account**
Joshua Aizenman

DTBC-448 Diciembre 2007
**Estimating the Chilean Natural Rate of Interest**
Rodrigo Fuentes y Fabián Gredig

DTBC-447 Diciembre 2007
**Valuation Effects and External Adjustment: A Review**
Pierre-Oliver Gourinchas