# Econometrics with Weak Instruments: Consequences, Detection, and Solutions

Vadim Marmer*
Vancouver School of Economics, UBC

March 17, 2017

*Email: vadim.marmer@ubc.ca. These lecture notes were prepared for the 2017 Lima Summer School in Economics hosted by Lima School of Economics, Universidad de Piura in Lima, Peru. I thank Marcos Agurto and the Lima School of Economics for the invitation and hospitality.

# Contents

## Preface

It is assumed that the reader is familiar with the basics of graduate Econometrics at the level of Davidson and MacKinnon (2004), Chapters 2-4 (OLS-based methods for estimation and inference) and Chapter 8 (IV-based methods). The main concepts and results for IV estimation and inference are also briefly reviewed here in Section 1.

# 1 Overview of standard (strong) IV asymptotic theory

## 1.1 The model

Before we study weak identification, it would be helpful to review the standard theory of instrumental variables (IVs). This will lay out the framework, establish some benchmark results, and provide motivation for investigating weak IVs.

The econometrician observes data $\{(y_{1i}, y_{2i}, Z'_{1i}, Z'_{2i}) : i = 1, \ldots n\}$, where $y_{1i}$ denotes the dependent variable, $y_{2i}$ is the *single* endogenous regressor, $Z_{1i}$ is the $l_1$-vector of instrumental variables, and $Z_{2i}$ is the $l_2$-vector of exogenous regressors. The IV regression model can be stated as

$$y_{1i} = \gamma y_{2i} + Z'_{2i}\beta + u_i, \tag{1.1}$$

$$y_{2i} = Z'_{1i}\pi_1 + Z'_{2i}\pi_2 + v_i, \tag{1.2}$$

where $\gamma \in \mathbb{R}$ is the unknown coefficient on the endogenous regressor. Typically, $\gamma$ is the main object of interest in applied work. The vector $\beta \in \mathbb{R}^{l_2}$ denotes the vector of unknown coefficients on the exogenous regressors. Equation (1.1) is known as the structural equation, and $u_i$ is the unobserved structural error. Equation (1.2) is known as the first stage and connects the endogenous regressor with the IVs; the parameters $\pi_1 \in \mathbb{R}^{l_1}$ and $\pi_2 \in \mathbb{R}^{l_2}$ are unknown, and $v_i$ denotes the unobserved first-stage error.

Exogeneity of $Z_{1i}$ and $Z_{2i}$ is defined as

$$EZ_{1i}u_i = EZ_{1i}v_i = 0, \quad EZ_{2i}u_i = EZ_{2i}v_i = 0. \tag{1.3}$$

The regressor $y_{2i}$ is endogenous in the sense

$$Ey_{2i}u_i \neq 0.$$

In view of exogeneity of $Z_{1i}$ and $Z_{2i}$, the regressor $y_{2i}$ is endogenous if the structural and first-stage errors are correlated, i.e.

$$Eu_iv_i \neq 0.$$

When $Z_{2i}$ includes the intercept, without loss of generality one can assume that $Eu_i = Ev_i = 0$.

Define the $n$-vectors $y_1$, $y_2$, $u$, and $v$ as

$$y_1 = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1n} \end{pmatrix}, \quad y_2 = \begin{pmatrix} y_{21} \\ \vdots \\ y_{2n} \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \quad v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$

Define further the $n \times l_1$ matrix of the observations on the IVs $Z_1$:

$$Z_1 = \begin{pmatrix} Z'_{11} \\ \vdots \\ Z'_{1n} \end{pmatrix}.$$

Let $Z_2$, the $n \times l_2$ matrix of the observations on the exogenous regressors be defined similarly:

$$Z_2 = \begin{pmatrix} Z'_{21} \\ \vdots \\ Z'_{2n} \end{pmatrix}.$$

The model can now be re-written as

$$y_1 = y_2\gamma + Z_2\beta + u, \tag{1.4}$$

$$y_2 = Z_1\pi_1 + Z_2\pi_2 + v. \tag{1.5}$$

### 1.2 IV estimation

The reduced-form equation, which relates the dependent variable $y_1$ with the exogenous variables $Z_1$ and $Z_2$ can be obtained by substitution the first stage (1.5) into the structural equation (1.4):

$$y_1 = Z_1\pi_1\gamma + Z_2(\pi_2\gamma + \beta) + (u + v\gamma), \tag{1.6}$$

which shows that the structural parameter of interest $\gamma$ can be estimated by regressing $y_1$ against $Z_1\pi_1$ and $Z_2$. Since $\pi_1$ is unknown, it must be replaced with its OLS estimator from the first stage.[1]

Define

$$M_2 = I_n - Z_2(Z'_2Z_2)^{-1}Z'_2.$$

It is easy to see that $M_2$ is symmetric and idempotent:

$$M'_2 = M_2 \text{ and } M_2M_2 = M_2.$$

The matrix $M_2$ is called a projection matrix, and $M_2x$ will project any $n$-vector $x$ on the space orthogonal to the span of the columns of $Z_2$. In particular,

$$M_2Z_2 = 0. \tag{1.7}$$

By the Frisch-Waught-Lowell (FWL) theorem (see Davidson and MacKinnon, 2004), the OLS

[1]See, for example, the discussion of different identification strategies in Blundell and Powell (2003).

estimator of $\pi_1$ from the first stage can be written as

$$\hat{\pi}_1 = (Z_1' M_2 Z_1)^{-1} Z_1' M_2 y_2. \tag{1.8}$$

This result is also known as the partitioned regression result.[2] Using the FWL theorem, (1.6), and by replacing $\pi_1$ with $\hat{\pi}_1$, we can now write the IV estimator of $\gamma$ ( it is also known as the 2SLS estimator) as

$$\hat{\gamma} = \frac{\hat{\pi}_1' Z_1' M_2 y_1}{\hat{\pi}_1' Z_1' M_2 Z_1 \hat{\pi}_1}. \tag{1.9}$$

### 1.3 Asymptotics of IV estimation

Assuming that data are iid, we can apply the iid weak law of large numbers (WLLN, Theorem A.1 in Appendix A.1) and the iid central limit theorem (CLT, Theorem A.2) to show consistency and asymptotic normality of the IV estimator provided that the vector of coefficients on the IVs in the first-stage equation is fixed and different from zero: $\pi_1 \neq 0$.

**Assumption 1.1.** *We assume that*

(i) The data $\{(y_{1i}, y_{2i}, Z_{1i}', Z_{2i}') : i = 1, \ldots n\}$ are iid.

(ii) The matrix

$$E\begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix}\begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix}' = \begin{pmatrix} EZ_{1i}Z_{1i}' & EZ_{1i}Z_{2i}' \\ EZ_{2i}Z_{1i}' & EZ_{2i}Z_{2i}' \end{pmatrix} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}' & Q_{22} \end{pmatrix} = Q$$

is finite and positive definite.

(iii) The matrix

$$E\left(\begin{pmatrix} u_i \\ v_i \end{pmatrix}\begin{pmatrix} u_i \\ v_i \end{pmatrix}' \middle| Z_{1i}, Z_{2i}\right) = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} = \Sigma$$

is finite and positive definite.

*Remark.* Assumption 1.1(iii) states that the errors $u$ and $v$ are homoskedastic. The assumption is made for simplicity and convenience only and can be easily relaxed to accommodate heteroskedastic models.

**Theorem 1.2.** *Suppose that $\pi_1 \neq 0$ and fixed. Then,*

**(a)** $\hat{\gamma} \to_p \gamma$.

[2] To see the result, write $y_2 = Z_1\hat{\pi}_1 + Z_2\hat{\pi}_2 + \hat{v}$, where $\hat{\pi}_1$ and $\hat{\pi}_2$ are the OLS estimators of $\pi_1$ and $\pi_2$ respectively, and $\hat{v}$ is the OLS residual satisfying $Z_1\hat{v} = 0$ and $Z_2\hat{v} = 0$. We now have $Z_1' M_2 y_2 = Z_1' M_2 Z_1 \hat{\pi}_1 + Z_1' M_2 \hat{v}$, where we used (1.7). Next, $Z_1' M_2 \hat{v} = Z_1'\hat{v} - Z_1' Z_2 (Z_2' Z_2)^{-1} Z_2' \hat{v} = 0$.

**(b)** $n^{1/2}(\hat{\gamma}-\gamma)\to_d N\left(0,\frac{\sigma_u^2}{\pi_1' Q_{1\cdot 2}\pi_1}\right)$, *where*

$$Q_{1\cdot 2}=Q_{11}-Q_{12}Q_{22}^{-1}Q_{12}'. \tag{1.10}$$

*Proof.* Using the definition of $\hat{\pi}_1$ in (1.8),

$$\hat{\pi}_1' Z_1' M_2 = y_2' M_2 Z_1 (Z_1' M_2 Z_1)^{-1} Z_1' M_2 = y_2' P_{M_2 Z_1},$$
$$\hat{\pi}_1' Z_1' M_2 Z_1 \hat{\pi}_1 = y_2' M_2 Z_1 (Z_1' M_2 Z_1)^{-1} Z_1' M_2 y_2 = y_2' P_{M_2 Z_1} y_2,$$

where we used the following notation. Let $H$ be a full column rank matrix. Then $P_H$ is defined as

$$P_H = H(H'H)^{-1}H'.$$

The matrix $P_H$ is a projection matrix (symmetric and idempotent): $P_H x$ projects a vector $x$ onto the span of $H$. Using the definition of $\hat{\gamma}$ in (1.9), the estimator $\hat{\gamma}$ can now be written as

$$\hat{\gamma}=\frac{y_2' P_{M_2 Z_1} y_1}{y_2' P_{M_2 Z_1} y_2}.$$

Using the model in (1.4),

$$\begin{aligned}\hat{\gamma} &= \gamma + \frac{y_2' P_{M_2 Z_1} u}{y_2' P_{M_2 Z_1} y_2}\\ &= \gamma + \frac{y_2' M_2 Z_1 (Z_1' M_2 Z_1)^{-1} Z_1' M_2 u}{y_2' M_2 Z_1 (Z_1' M_2 Z_1)^{-1} Z_1' M_2 y_2}\\ &= \frac{(Z_1\pi_1+v)' M_2 Z_1 (Z_1' M_2 Z_1)^{-1} Z_1' M_2 u}{(Z_1\pi_1+v)' M_2 Z_1 (Z_1' M_2 Z_1)^{-1} Z_1' M_2 (Z_1\pi_1+v)'}\\ &= \gamma + \frac{(Z_1' M_2 Z_1 \pi_1 + Z_1' M_2 v)' (Z_1' M_2 Z_1)^{-1} Z_1' M_2 u}{(Z_1' M_2 Z_1 \pi_1 + Z_1' M_2 v)' (Z_1' M_2 Z_1)^{-1} (Z_1' M_2 Z_1 \pi_1 + Z_1' M_2 v)}.\end{aligned} \tag{1.11}$$

By the WLLN and Assumption 1.1(i,ii),

$$\frac{Z_1' Z_1}{n} = n^{-1}\sum_{i=1}^n Z_{1i} Z_{1i}' \to_p E Z_{1i} Z_{1i}' = Q_{11}.$$

Similarly:

$$\frac{Z_1' Z_2}{n}\to_p Q_{12} \quad \text{and} \quad \frac{Z_2' Z_2}{n}\to_p Q_{22}.$$

Hence,

$$\frac{Z_1' M_2 Z_1}{n} = \frac{Z_1' Z_1}{n} - \frac{Z_1' Z_2}{n}\left(\frac{Z_2' Z_2}{n}\right)^{-1}\frac{Z_2' Z_1}{n}$$

$$
\begin{aligned}
&\to_p Q_{11} - Q_{12}Q_{22}^{-1}Q_{12}' \\
&= Q_{1\cdot 2},
\end{aligned} \tag{1.12}
$$

where $Q_{1\cdot 2}$ is positive definite. Furthermore, since

$$
\begin{aligned}
&\frac{Z_1'u}{n} = n^{-1}\sum_{i=1}^{n} Z_{1i}u_i \to_p EZ_{1i}u_i = 0, \text{ and similarly} \\
&\frac{Z_2'u}{n} \to_p 0,
\end{aligned}
$$

we have:

$$
\begin{aligned}
\frac{Z_1'M_2u}{n} &= \frac{Z_1'u}{n} - \frac{Z_1'Z_2}{n}\left(\frac{Z_2'Z_2}{n}\right)^{-1}\frac{Z_2'u}{n} \\
&\to_p 0 - Q_{12}Q_{22}^{-1}\cdot 0 \\
&= 0.
\end{aligned} \tag{1.13}
$$

Similarly,

$$
\frac{Z_1'M_2v}{n} \to_p 0. \tag{1.14}
$$

We conclude from (1.11), (1.12), (1.13), and (1.14) that

$$
\hat{\gamma} \to_p \gamma + \frac{(Q_{1\cdot 2}\pi_1 + 0)'\cdot 0}{(Q_{1\cdot 2}\pi_1 + 0)'Q_{1\cdot 2}^{-1}(Q_{1\cdot 2}\pi_1 + 0)'} = \gamma,
$$

which holds since $\pi_1 \neq 0$ and is fixed by assumption. This concludes the proof of part (a).

For part (b), re-write (1.11) as

$$
n^{1/2}(\hat{\gamma} - \gamma) = \frac{n^{-1/2}\pi_1'Z_1'M_2u + o_p(1)}{n^{-1}\pi_1'Z_1'M_2Z_1\pi_1 + o_p(1)}, \tag{1.15}
$$

where we used (1.13) and (1.14). By the CLT,

$$
\begin{aligned}
n^{-1/2}\begin{pmatrix} Z_1'u \\ Z_2'u \end{pmatrix} &= n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix} u_i \\
&\to_d N\left(0, Eu_i^2\begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix}\begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix}'\right) \\
&= N\left(0, \sigma_u^2 Q\right),
\end{aligned} \tag{1.16}
$$

where the equality in the last line holds by the law of iterated expectation (see Davidson and

MacKinnon, 2004, p. 14) and Assumption 1.1(iii):

$$Eu_i^2 \begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix} \begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix}' = E\left\{ E\left(u_i^2 \mid Z_{1i}, Z_{2i}\right) \begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix} \begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix}' \right\} = \sigma_u^2 E \begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix} \begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix}'.$$

Let $\Phi_1$ and $\Phi_2$ be two random vectors jointly distributed as $N(0, \sigma_u^2 Q)$:

$$\begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix} \sim N\left(0, \sigma_u^2 \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{12}' & Q_{22} \end{pmatrix}\right).$$

Write

$$\begin{aligned} \frac{Z_1' M_2 u}{n^{1/2}} &= \frac{Z_1' u}{n^{1/2}} - \frac{Z_1' Z_2}{n} \left(\frac{Z_2' Z_2}{n}\right)^{-1} \frac{Z_2' u}{n^{1/2}} \\ &\to_d \Phi_1 - Q_{12}' Q_{22}^{-1} \Phi_2 \\ &= \Phi_{1\cdot 2}. \end{aligned} \tag{1.17}$$

Note that

$$\begin{aligned} &Var(\Phi_{1\cdot 2}) = \\ &= Var(\Phi_1) + Q_{12}' Q_{22}^{-1} Var(\Phi_2) Q_{22}^{-1} Q_{12} - Cov(\Phi_1 \Phi_2) Q_{22}^{-1} Q_{12} - Q_{12}' Q_{22}^{-1} Cov(\Phi_2, \Phi_1) \\ &= \sigma_u^2 \left(Q_{11} + Q_{12}' Q_{22}^{-1} Q_{12} - 2 Q_{12}' Q_{22}^{-1} Q_{12}\right) \\ &= \sigma_u^2 Q_{1\cdot 2}. \end{aligned}$$

Hence,

$$\frac{Z_1' M_2 u}{n^{1/2}} \to_d \Phi_{1\cdot 2} \sim N(0, \sigma_u^2 Q_{1\cdot 2}). \tag{1.18}$$

It follows from (1.12), (1.15), and (1.18) that

$$n^{1/2}(\hat{\gamma} - \gamma) \to_d \frac{\pi_1' \Phi_{1\cdot 2}}{\pi_1' Q_{1\cdot 2} \pi_1} \sim \frac{N\left(0, \sigma_u^2 \pi_1' Q_{1\cdot 2} \pi_1\right)}{\pi_1' Q_{1\cdot 2} \pi_1} = N\left(0, \frac{\sigma_u^2}{\pi_1' Q_{1\cdot 2} \pi_1}\right),$$

where the last equality follows from the properties of normal distributions, see Theorem A.5 in Section A.2 in the Appendix □

Statistical inference about $\gamma$ (the causal effect of the endogenous regressor on the dependent variable), can be performed using the asymptotic normality result of Theorem 1.2(b) as

$$\frac{n^{1/2}(\hat{\gamma} - \gamma)}{\sqrt{\sigma_u^2 / (\pi_1' Q_{1\cdot 2} \pi_2)}} \to_d N(0, 1).$$

Thus for inference one can use standard normal critical values such as $z_{1-\alpha/2}$, where $\alpha$ denotes the significance level and $z_\tau$ denotes the $\tau$-th quantile (percentile) of the standard normal distribution.

Using the arguments in the proof of Theorem 1.2, $\pi_1' Q_{1\cdot 2}\pi_2$ can be estimated by

$$\frac{y_2' P_{M_2 Z_1} y_2}{n} \to_p \pi_1' Q_{1\cdot 2}\pi_2.$$

To estimate $\sigma_u^2$, consider $y_1 - y_2\hat{\gamma}$:

$$M_2(y_1 - y_2\hat{\gamma}) = M_2 u - M_2 y_2(\hat{\gamma} - \gamma).$$

Define

$$\begin{aligned}\hat{\sigma}_u^2 &= \frac{(y_1 - y_2\hat{\gamma})' M_2 (y_1 - y_2\hat{\gamma})}{n} \\ &= \frac{u' M_2 u}{n} + \frac{(\hat{\gamma}-\gamma)' y_2' M_2 y_2 (\hat{\gamma}-\gamma)}{n} - 2\frac{(\hat{\gamma}-\gamma)' y_2' M_2 u}{n} \\ &= \frac{u'u - u' Z_2 (Z_2' Z_2)^{-1} Z_2' u}{n} + o_p(1) \\ &\to_p \sigma_u^2,\end{aligned} \tag{1.19}$$

where the equality in the third line holds because $\hat{\gamma} - \gamma = o_p(1)$, and the equality in the last line holds since $u'u/n = n^{-1}\sum_{i=1}^n u_i^2 \to_p \sigma_u^2$, and $Z_2' u/n \to_p 0$.

Consider now testing $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$. The $t$-statistic is given by

$$t(\gamma_0) = \frac{n^{1/2}(\hat{\gamma} - \gamma_0)}{\sqrt{\frac{(y_1 - y_2\hat{\gamma})' M_2 (y_1 - y_2\hat{\gamma})}{y_2' P_{M_2 Z_1} y_2}}}. \tag{1.20}$$

A test with asymptotic size $\alpha$ rejects $H_0$ when

$$|t(\gamma_0)| > z_{1-\alpha/2}.$$

The validity of the test holds because when $H_0 : \gamma = \gamma_0$ is true,

$$t(\gamma) = \frac{n^{1/2}(\hat{\gamma} - \gamma)}{\sqrt{\frac{(y_1 - y_2\hat{\gamma})' M_2 (y_1 - y_2\hat{\gamma})/n}{y_2' P_{M_2 Z_1} y_2/n}}} \to_d \frac{N\left(0, \sigma_u^2/(\pi_1' Q_{1\cdot 2}\pi_1)\right)}{\sqrt{\sigma_u^2/(\pi_1' Q_{1\cdot 2}\pi_2)}} = N(0,1).$$

Consequently, for $\mathcal{Z} \sim N(0,1)$

$$P\left(|t(\gamma)| > z_{1-\alpha/2}\right) \to P(|\mathcal{Z}| > z_{1-\alpha/2}) = \alpha. \tag{1.21}$$

The standard error of $\hat{\gamma}$ is given by

$$\text{std.err} = n^{-1/2}\sqrt{\frac{(y_1 - y_2\hat{\gamma})' M_2 (y_1 - y_2\hat{\gamma})}{y_2' P_{M_2 Z_1} y_2}}.$$

The confidence interval for $\gamma$ with asymptotic coverage $1-\alpha$ is constructed as

$$CI_{1-\alpha} = \left\{\gamma_0 \in \mathbb{R} : |t(\gamma_0)| \leq z_{1-\alpha/2}\right\} = [\hat{\gamma} - z_{1-\alpha/2} \times \text{std.err}, \hat{\gamma} + z_{1-\alpha/2} \times \text{std.err}].$$

The asymptotic validity of $CI_{1-\alpha}$ holds as

$$P\left(\gamma \in CI_{1-\alpha}\right) = P\left(|t(\gamma)| \leq z_{1-\alpha/2}\right) \to 1-\alpha,$$

where the last result holds by (1.21).

## 1.4 Motivation for studying weak instruments

The result in Theorem 1.2(b) shows that the variance of the estimator $\hat{\gamma}$ is determined by $\pi_1$, the coefficient on the IVs in the first-stage equation. Moreover, the asymptotic variance of the estimator $\hat{\gamma}$ is inversely related to the weighted Euclidean norm of $\pi_1$:

$$\|\pi_1\|^2_{Q_{1\cdot 2}} = \pi_1' Q_{1\cdot 2} \pi_1.$$

The variance of $\hat{\gamma}$ is larger (and therefore the estimator is less precise and informative) for the values of $\pi_1$ close to zero. Since $\pi_1$ relates the IVs $Z_1$ with the endogenous regressor $y_2$, smaller values of $\|\pi_1\|^2_{Q_{1\cdot 2}}$ correspond to IVs that are less informative about the endogenous regressor or, in other words, "weaker".

Thus, it is important to investigate the behavior of the estimator $\hat{\gamma}$ for the values of $\pi_1$ that are very close to zero (in the sense $\pi_1 \to 0$) as in such situations IVs are relevant, but provide information of poor quality. Unfortunately, Theorem 1.2 breaks down as $\pi_1 \to 0$.

The standard approach adopted in Theorem 1.2 keeps $\pi_1$ fixed as the sample size $n \to \infty$. As a result,

$$\hat{\gamma} = \gamma + O_p\left(\frac{1}{\sqrt{n}}\right).$$

In other words, the estimation error approaches zero at the rate $1/\sqrt{n}$, while $\pi_1$ stays fixed. In view of that, no matter how small it is, any fixed value $\pi_1$ is "large" relatively to the estimation error of order $1/\sqrt{n}$ for all sample sizes large enough. However, in practice the econometrician deals with a fixed sample size, and it is possible that given the actual sample size, the estimation error and $\pi_1$ are comparably small. Hence, the standard asymptotic analysis, which assumes that $\pi_1$ is fixed, can fail to provide accurate approximation to the behavior of $\hat{\gamma}$ in finite samples. One needs a drastically different approach.

Lastly, note that

$$\pi_1' Q_{1\cdot 2}\pi_1 = \pi_1' E Z_{1i} Z_{2i}' (E Z_{2i} Z_{2i}')^{-1} E Z_{2i} Z_{1i}' \pi_1.$$

The quantity can be approximated as

$$\pi_1' Z_1' M_2 Z_1 \pi_1 \sim n\pi_1' Q_{1\cdot 2}\pi_1,$$

where $P_2 = P_{Z_2}$. The expression above is related to the so called the *concentration parameter*:

$$\|\lambda_n\|^2 = \frac{\pi_1' Z_1' M_2 Z_1 \pi_1}{\sigma_v^2},$$

which is used at the measure of the strength of IVs or identification. The concentration parameter $\|\lambda_n\|^2$ captures the strength of the signal from the IVs relatively to the noise in the errors in the first-stage equation. Note that the assumption of fixed $\pi_1 \neq 0$ corresponds to $\|\lambda_n\|^2 \to \infty$: i.e. the signal from the IVs dominates the noise due to the errors. Such cases are referred to as *strong identification* or strong IVs. Weak identification occurs when the signal and noise are of the same magnitude.

# 2 Weak IV asymptotics

## 2.1 Modeling weak IVs

In this section, we introduce the weak IV model using a simplified framework (for clarity).

Consider the case of a simple IV regression with a single endogenous regressor, no exogenous regressors, and a single instrument:

$$\begin{aligned} y_1 &= \gamma y_2 + u, \\ y_2 &= \pi_1 Z_1 + v. \end{aligned}$$

In this case, the IV estimator in (1.9) becomes

$$\begin{aligned} \hat{\gamma} &= \frac{Z_1' y_1}{Z_1' y_2} \\ &= \gamma + \frac{\sum_{i=1}^n Z_{1i} u_i}{\pi_1 \sum_{i=1}^n Z_{1i}^2 + \sum_{i=1}^n Z_{1i} v_i} \\ &= \gamma + \frac{n^{-1/2} \sum_{i=1}^n Z_{1i} u_i}{\pi_1 n^{-1/2} \sum_{i=1}^n Z_{1i}^2 + n^{-1/2} \sum_{i=1}^n Z_{1i} v_i} \\ &= \gamma + \frac{O_p(1)}{\pi_1 n^{-1/2} \sum_{i=1}^n Z_{1i}^2 + O_p(1)}, \end{aligned} \tag{2.1}$$

where $O_p(1)$ stands for bounded in probability, i.e. with a probability arbitrary close to one it can be bounded by a constant for all sample sizes $n$large enough (see Definition A.6 in the Appendix). The claims

$$\begin{aligned} n^{-1/2} \sum_{i=1}^n Z_{1i} u_i &= O_p(1), \\ n^{-1/2} \sum_{i=1}^n Z_{1i} v_i &= O_p(1), \end{aligned}$$

can be justified by the CLT as it implies that the distribution of the random variables on the left-hand side can be approximated by normal distributions. On the other hand,

$$n^{-1/2} \sum_{i=1}^n Z_{1i}^2 = n^{1/2} \left( n^{-1} \sum_{i=1}^n Z_{1i}^2 \right) \to \infty \tag{2.2}$$

due to the WLLN.

The $O_p(1)$ terms in equation (2.1) represent the *noise* due to estimation as they are determined by the errors $u$ and $v$. On the other hand, the $\pi_1 n^{-1/2} \sum_{i=1}^n Z_{1i}^2$ term in the denominator in (2.1) represents the *signal* contained in random data about the true parameter $\gamma$. As long as $\pi_1 \neq 0$ (no matter how small), the signal component diverges to infinity

$$\pi_1 n^{-1/2} \sum_{i=1}^{n} Z_{1i}^2 \to \infty$$

and the signal component will dominate the noise resulting in consistent estimation of $\gamma$, i.e. $\hat{\gamma} \to_p \gamma$.

If $\pi_1 = 0$, the data contains no information (signal) about $\gamma$. One can say that identification of $\gamma$ is weak when data does contain some information about $\gamma$, but the signal is weak in the sense that it can be easily obscured by the noise for any value of $n$. In other words, identification is weak when the signal and the noise are of the same order. In view of (2.2), we can achieve that (for any $n$ and as $n \to \infty$) by modeling $\pi_1$ as a sequence that depends on $n$ and approaches zero at the rate that balances out the terms in the signal component, so that

$$\pi_1 n^{-1/2} \sum_{i=1}^{n} Z_{1i}^2 \to_p \text{constant}.$$

To achieve that, one has to assume that

$$\pi_1 = \frac{C}{n^{1/2}} \tag{2.3}$$

for some unknown constant $C$. In this case, the signal component no longer diverges to infinity:

$$\pi_1 n^{-1/2} \sum_{i=1}^{n} Z_{1i}^2 = C n^{-1} \sum_{i=1}^{n} Z_{1i}^2 \to_p C \cdot E Z_{1i}^2.$$

Now the signal no longer dominates the noise component (they are of the same magnitude), and as a result, consistency of $\hat{\gamma}$ fails!

Note that since $n$ is large, by adopting (2.3), we effectively modeled the first stage coefficient as a small number. Thus, since we are relying on asymptotics, "small" must be modeled in relationship to $n$ as "small" means different things in samples of different sizes! The strength of the weak signal (i.e. within the weak identification framework) is controlled by the constant $C$.

We say that instruments are weak when the coefficient on the instruments in the first stage is of the local-to-zero form in (2.3).

Such a local-to-zero framework for formalizing weak instruments and weak identification was first proposed in Staiger and Stock (1997). Moreover, using uniform validity arguments, it was shown later that the local-to-zero framework (with $n^{-1/2}$ rates) is unavoidable when studying local identification failures and relying on asymptotic arguments (see for example Andrews et al., 2011).

In the following sections, we will investigate the effect of weak IVs on the properties of IV-based estimation and inference.

### 2.2 The distribution of the IV estimator under weak IVs

In this section we derive the asymptotic distribution of the IV estimator $\hat{\gamma}$ in (1.9) for the general model (1.4)-(1.5). We assume that the IVs are weak:

**Assumption 2.1** (Weak IVs). *$\pi_1 = n^{-1/2}C$ for some fixed unknown $l_1$-vector $C$.*

Define the following two random $l_1$-vectors that have a joint Normal distribution:

$$\begin{pmatrix} \mathcal{Z}_u \\ \mathcal{Z}_v \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho_{uv} \\ \rho_{uv} & 1 \end{pmatrix} \otimes I_{l_1}\right), \tag{2.4}$$

where

$$\begin{pmatrix} 1 & \rho_{uv} \\ \rho_{uv} & 1 \end{pmatrix} \otimes I_{l_1} = \begin{pmatrix} I_{l_1} & \rho_{uv} I_{l_1} \\ \rho_{uv} I_{l_1} & I_{l_1} \end{pmatrix},$$

the correlation coefficient

$$\rho_{uv} = \frac{\sigma_{uv}}{\sigma_u \sigma_v},$$

and $\otimes$ denotes the Kronecker product.[3] Note that $\rho_{uv}$ measures the correlation between the structural and first-stage errors, i.e. $\rho_{uv}$ measures the amount of endogeneity in the model.

**Theorem 2.2.** *Suppose that Assumptions 1.1 and 2.1 hold.*

$$\hat{\gamma} \to_d \gamma + \frac{\sigma_u}{\sigma_v} \frac{(\lambda + \mathcal{Z}_v)' \mathcal{Z}_u}{\|\lambda + \mathcal{Z}_v\|^2},$$

*where $\mathcal{Z}_u$ and $\mathcal{Z}_v$ are defined in (2.4), and*

$$\lambda = \frac{Q_{1\cdot 2}^{1/2} C}{\sigma_v}. \tag{2.5}$$

*Remark.*

1. The estimator $\hat{\gamma}$ is inconsistent as $\hat{\gamma} - \gamma$ converges in distribution to a non-degenerate random variable.

2. When $\rho_{uv} = 0$, $\mathcal{Z}_u$ and $\mathcal{Z}_v$ are independent. Let

$$\Delta = \frac{(\lambda + \mathcal{Z}_v)' \mathcal{Z}_u}{\|\lambda + \mathcal{Z}_v\|^2}. \tag{2.6}$$

[3] See Section A.4 in the Appendix.

When $\mathcal{Z}_u$ and $\mathcal{Z}_v$ are independent, the properties of normal distributions (see Section A.2 in the Appendix) imply that

$$\frac{\sigma_u}{\sigma_v}\Delta \mid \mathcal{Z}_v \sim N\left(0, \frac{\sigma_u^2}{\|\sigma_v\lambda + \sigma_v\mathcal{Z}_v\|^2}\right).$$

This distribution is analogous to $N\left(0, \sigma_u^2/\|Q_{1\cdot 2}^{1/2}\pi_1\|^2\right)$ distribution that we have in the strong IVs case in Theorem 1.2(b). In particular, the IV estimator appears to be asymptotically unbiased as the conditional mean of $\Delta \mid \mathcal{Z}_v$ distribution is zero for any value of $\mathcal{Z}_v$.

3. When $\rho_{uv} \neq 0$,

$$\mathcal{Z}_u \mid \mathcal{Z}_v \sim N\left(\rho_{uv}\mathcal{Z}_v, (1-\rho_{uv}^2)I_{l_1}\right),$$

which follows from the properties of multivariate normal distributions, see Theorem A.4 in Section A.2 in the Appendix. In this case,

$$\frac{\sigma_u}{\sigma_v}\Delta \mid \mathcal{Z}_v \sim N\left(\frac{\sigma_{uv}}{\sigma_v^2}\frac{(\lambda+\mathcal{Z}_v)'\mathcal{Z}_v}{\|\lambda+\mathcal{Z}_v\|^2}, \frac{(1-\rho_{uv}^2)\sigma_u^2}{\sigma_v^2\,\|\lambda+\mathcal{Z}_v\|^2}\right).$$

Now the IV estimator is not only inconsistent but also asymptotically biased.

4. The parameter $\lambda{=}\sigma_v^{-1}Q_{1.2}^{1/2}C$ is related to the concentration parameter, as in this case

$$\|\lambda_n\|^2 = \frac{\pi_1' Z_1' P_2 Z_1 \pi_1}{\sigma_v^2} = \frac{n\pi_1'\left(Z_1'P_2Z_1/n\right)\pi_1}{\sigma_v^2} \to_p \frac{C'Q_{1\cdot 2}C}{\sigma_v^2} = \|\lambda\|^2.$$

Since the concentration parameter is finite in the limit, the signal from the IVs does not dominate the noise, which is the source of inconsistency of the IV estimator.

5. The estimator $\hat{\gamma}$ is consistent when the concentration parameter $\|\lambda\| \to \infty$, which corresponds to the strong IVs case:

$$\begin{aligned}
\frac{(\lambda+\mathcal{Z}_v)'\mathcal{Z}_v}{\|\lambda+\mathcal{Z}_v\|^2} &= \frac{(\lambda/\|\lambda\| + o_p(1))'\mathcal{Z}_v}{\|\lambda/\|\lambda\| + o_p(1)\|^2}\frac{1}{\|\lambda\|} \\
&= \frac{(\ell + o_p(1))'\mathcal{Z}_v}{\|\ell + o_p(1)\|^2}\frac{1}{\|\lambda\|}, \quad \text{where } \ell \in \mathbb{R}^{l_1} \text{ and } \|\ell\| = 1 \\
&\to_p 0 \quad \text{as } \|\lambda\| \to \infty.
\end{aligned}$$

*Proof of Theorem 2.2.* Define

$$\begin{pmatrix} \Phi_{1\cdot 2} \\ \Psi_{1\cdot 2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 Q_{1\cdot 2} & \sigma_{uv} Q_{1\cdot 2} \\ \sigma_{uv} Q_{1\cdot 2} & \sigma_v^2 Q_{1\cdot 2} \end{pmatrix}\right), \tag{2.7}$$

where $Q_{1\cdot 2}$ has been defined in (1.10). Note that the variance can be also written as

$$\begin{pmatrix} \sigma_u^2 Q_{1\cdot 2} & \sigma_{uv} Q_{1\cdot 2} \\ \sigma_{uv} Q_{1\cdot 2} & \sigma_v^2 Q_{1\cdot 2} \end{pmatrix} = \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \otimes Q_{1\cdot 2} = \Sigma \otimes Q_{1\cdot 2}.$$

By the CLT,

$$\begin{aligned} n^{-1/2} \begin{pmatrix} Z'u \\ Z'v \end{pmatrix} &= \begin{pmatrix} \frac{Z_1'u}{\sqrt{n}} \\ \frac{Z_2'u}{\sqrt{n}} \\ \frac{Z_1'v}{\sqrt{n}} \\ \frac{Z_2'v}{\sqrt{n}} \end{pmatrix} \\ &= n^{-1/2} \sum_{i=1}^{n} \begin{pmatrix} u_i \\ v_i \end{pmatrix} \otimes \begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix} \\ &\to_d N(0, \Sigma \otimes Q) \\ &= \begin{pmatrix} \Phi_1 \\ \Phi_2 \\ \Psi_1 \\ \Psi_2 \end{pmatrix}. \end{aligned}$$

One can extend the same arguments as in the proof of Theorem 1.2(b), equation (1.17), to show that

$$\begin{pmatrix} \frac{Z_1' M_2 u}{n^{1/2}} \\ \frac{Z_1' M_2 v}{n^{1/2}} \end{pmatrix} \to_d \begin{pmatrix} \Phi_{1\cdot 2} \\ \Psi_{1\cdot 2} \end{pmatrix},$$

where $\Psi_{1\cdot 2}$ satisfies

$$\Psi_{1\cdot 2} = \Psi_1 - Q_{12}' Q_{22}^{-1} \Phi_2.$$

The denominator in the second term on the right-hand side of (1.11) can be re-written as

$$\begin{aligned} & y_2' P_{M_2 Z_1} y_2 = \\ &= \left(Z_1' M_2 Z_1 \pi_1 + Z_1' M_2 v\right)' \left(Z_1' M_2 Z_1\right)^{-1} \left(Z_1' M_2 Z_1 \pi_1 + Z_1' M_2 v\right) \\ &= \left(\left(Z_1' M_2 Z_1\right)^{1/2} \pi_1 + \left(Z_1' M_2 Z_1\right)^{-1/2} Z_1' M_2 v\right)' \left(\left(Z_1' M_2 Z_1\right)^{1/2} \pi_1 + \left(Z_1' M_2 Z_1\right)^{-1/2} Z_1' M_2 v\right) \\ &= \left\|\left(Z_1' M_2 Z_1\right)^{1/2} \pi_1 + \left(Z_1' M_2 Z_1\right)^{-1/2} Z_1' M_2 v\right\|^2, \end{aligned} \tag{2.8}$$

and therefore from (1.11),

$$
\begin{aligned}
\hat{\gamma} &= \gamma + \frac{\left(\pi_1 + (Z_1'M_2Z_1)^{-1}Z_1'M_2v\right)' Z_1'M_2u}{\left\|(Z_1'M_2Z_1)^{1/2}\,\pi_1 + (Z_1'M_2Z_1)^{-1/2}Z_1'M_2v\right\|^2} && (2.9)\\
&= \gamma + \frac{\left(C/\sqrt{n} + (Z_1'M_2Z_1)^{-1}Z_1'M_2v\right)' Z_1'M_2u}{\left\|(Z_1'M_2Z_1)^{1/2}\,C/\sqrt{n} + (Z_1'M_2Z_1)^{-1/2}Z_1'M_2v\right\|^2}\\
&= \gamma + \frac{\left(C + (Z_1'M_2Z_1/n)^{-1}Z_1'M_2v/\sqrt{n}\right)' Z_1'M_2u/\sqrt{n}}{\left\|(Z_1'M_2Z_1/n)^{1/2}\,C + (Z_1'M_2Z_1/n)^{-1/2}Z_1'M_2v/\sqrt{n}\right\|^2}\\
&\to_d \gamma + \frac{\left(C + Q_{1\cdot 2}^{-1}\Psi_{1\cdot 2}\right)' \Phi_{1\cdot 2}}{\|Q_{1\cdot 2}^{1/2}C + Q_{1\cdot 2}^{-1/2}\Psi_{1\cdot 2}\|^2} && (2.10)\\
&= \gamma + \frac{\left(C + Q_{1\cdot 2}^{-1}\sigma_v Q_{1\cdot 2}^{1/2}\mathcal{Z}_v\right)' \sigma_u Q_{1\cdot 2}^{1/2}\mathcal{Z}_u}{\|Q_{1\cdot 2}^{1/2}C + Q_{1\cdot 2}^{-1/2}\sigma_v Q_{1\cdot 2}^{1/2}\mathcal{Z}_v\|^2}\\
&= \gamma + \frac{\sigma_u\sigma_v}{\sigma_v^2}\frac{\left(\sigma_v^{-1}Q_{1\cdot 2}^{1/2}C + \mathcal{Z}_v\right)' \mathcal{Z}_u}{\|\sigma_v^{-1}Q_{1\cdot 2}^{1/2}C + \mathcal{Z}_v\|^2},
\end{aligned}
$$

where the equality in the third line holds by multiplying and dividing by $\sqrt{n}$ in the numerator.

□

### 2.3 The null distribution of the $t$-statistic under weak IVs

We now consider testing hypotheses about $\gamma$ using its IV estimator and the $t$-statistic

$$
t(\gamma_0) = \frac{n^{1/2}(\hat{\gamma} - \gamma_0)}{\sqrt{\frac{(y_1 - y_2\hat{\gamma})'M_2(y_1 - y_2\hat{\gamma})}{y_2'P_{M_2Z_1}y_2}}}.
$$

Recall that the $t$-test rejects $H_0 : \gamma = \gamma_0$ in favor of $H_1 : \gamma \neq \gamma_0$ when

$$
|t(\gamma_0)| > z_{1-\alpha/2}.
$$

The null asymptotic distribution of the $t$-statistic, i.e. when $\gamma = \gamma_0$, is given in the following theorem

**Theorem 2.3.** *Suppose that Assumptions 1.1 and 2.1 hold.*

$$
t(\gamma) \to_d \frac{\|\lambda + \mathcal{Z}_v\|(\lambda + \mathcal{Z}_v)'\mathcal{Z}_u}{\left(\left((\lambda + \mathcal{Z}_v)'\mathcal{Z}_u\right)^2 + \|\lambda + \mathcal{Z}_v\|^4 - 2\rho_{uv}\left((\lambda + \mathcal{Z}_v)'\mathcal{Z}_u\right)\|\lambda + \mathcal{Z}_v\|^2\right)^{1/2}}, \tag{2.11}
$$

*where $\mathcal{Z}_u$ and $\mathcal{Z}_v$ are defined in (2.4).*

*Remark.* 1. As one an see from the statement of the theorem, since $\mathcal{Z}_u$ appears both in the numerator and the denominator of the expression in (2.11), the null distribution of the $t$-statistic is different from $N(0,1)$ even when $\rho_{uv} = 0$. As a result, using standard normal critical values for test decision rules may lead to invalid tests: the null rejection probabilities (when $\gamma = \gamma_0$) may exceed $\alpha$, where $\alpha$ is the null rejection probability one expects to see when using critical values $z_{1-\alpha/2}$. See the discussion below and Figure 1.

2. The limiting distribution in (2.11) depends on $\rho_{uv}$ and two random variables $\|\lambda + \mathcal{Z}_v\|$ and $(\lambda + \mathcal{Z}_v)'\mathcal{Z}_u$. It is show in Lemma A.10 in the Appendix that

$$\|\lambda + \mathcal{Z}_v\|^2 =^d (\|\lambda\| + \mathcal{Z}_{v,1})^2 + \sum_{j=2}^{l_1} \mathcal{Z}_{v,j}^2,$$

$$(\lambda + \mathcal{Z}_v)' \mathcal{Z}_u =^d (\|\lambda\| + \mathcal{Z}_{v,1})\mathcal{Z}_{u,1} + \sum_{j=2}^{l_1} \mathcal{Z}_{v,j}\mathcal{Z}_{u,j},$$

where the notation "$=^d$" stands for equal in distribution, and means that the random variables on the left- and right-hand sides have the same distribution. Thus, while $\lambda$ is a vector, it affects the null distribution of the $t$-statistic only through its norm or the concentration parameter:

$$\|\lambda\|^2 = \frac{C'Q_{1\cdot 2}C}{\sigma_v^2}.$$

We conclude that the limiting distribution in (2.11) is completely determined by three scalar parameters: the concentration parameter $\|\lambda\|^2$, the number of IVs $l_1$, and the endogeneity parameter $\rho_{uv}$.

3. Since $\hat{\gamma}$ is inconsistent when IVs are weak, the endogeneity parameter $\rho_{uv}$ cannot be estimated consistently. Hence, the distribution in (2.11) is unknown as $\rho_{uv}$ remains unknown.

4. Weak IVs asymptotics nests strong IVs asymptotics as a limiting case. One can show that as $\|\lambda\| \to \infty$, the distribution on the right-hand side of (2.11) becomes standard normal:

$$\begin{aligned}
&\frac{((\lambda + \mathcal{Z}_v)'\mathcal{Z}_u)\,\|\lambda + \mathcal{Z}_v\|/\|\lambda\|^2}{\left(((\lambda + \mathcal{Z}_v)'\mathcal{Z}_u)^2 + \|\lambda + \mathcal{Z}_v\|^4 - 2\rho_{uv}\,((\lambda + \mathcal{Z}_v)'\mathcal{Z}_u)\,\|\lambda + \mathcal{Z}_v\|\right)^{1/2}/\|\lambda\|^2} \\
&= \frac{(\lambda/\|\lambda\| + o_p(1))'\mathcal{Z}_u(1 + o_p(1))}{(\|\lambda/\|\lambda\| + o_p(1)\|^4 + o_p(1))^{1/2}} \\
&\to_p \ell'\mathcal{Z}_u, \quad \text{where } \|\ell\| = 1, \\
&= N(0,1).
\end{aligned}$$

*Proof of Theorem 2.3.* By the definition of the $t$-statistic and since

$$\hat{\gamma} = \frac{y_2' P_{M_2 Z_1} y_1}{y_2' P_{M_2 Z_1} y_2} = \gamma + \frac{y_2' P_{M_2 Z_1} u}{y_2' P_{M_2 Z_1} y_2},$$

we have:

$$\begin{aligned} t(\gamma) &= \frac{\hat{\gamma} - \gamma}{\sqrt{\frac{\hat{\sigma}_u^2}{y_2' P_{M_2 Z_1} y_2}}} \\ &= \frac{y_2' P_{M_2 Z_1} u}{\hat{\sigma}_u \sqrt{y_2' P_{M_2 Z_1} y_2}} \\ &= \frac{(Z_1 \pi_1 + v)' \left( M_2 Z_1 (Z_1' M_2 Z_1)^{-1} Z_1' M_2 \right) u}{\hat{\sigma}_u \sqrt{y_2' P_{M_2 Z_1} y_2}} \\ &= \frac{\left( \pi_1 + (Z_1' M_2 Z_1)^{-1} Z_1' M_2 v \right)' Z_1' M_2 u}{\hat{\sigma}_u \left\| (Z_1' M_2 Z_1)^{1/2} \pi_1 + (Z_1' M_2 Z_1)^{-1/2} Z_1' M_2 v \right\|} \end{aligned} \tag{2.12}$$

where in the second line we used the arguments from (1.11), and in the last line we used the arguments from (2.8) and (2.9).

By (1.19),

$$\hat{\sigma}_u^2 = \frac{u' M_2 u}{n} + \frac{(\hat{\gamma} - \gamma)^2 y_2' M_2 y_2}{n} - 2 \frac{(\hat{\gamma} - \gamma) y_2' M_2 u}{n}.$$

Recall that the first term converges in probability to $\sigma_u^2$. In the case of strong IVs, the last two terms were negligible as $\hat{\gamma}$ was consistent for $\gamma$. Now since $\hat{\gamma} - \gamma \to_d \Delta \sigma_u / \sigma_v$, they have to be taken into account as $\Delta \neq 0$ with probability one.

$$\begin{aligned} \frac{y_2' M_2 y_2}{n} &= \frac{v' M_2 v + \pi_1' Z_1' M_2 Z_1 \pi_1 + 2 \pi_1' Z_1' M_2 v}{n} \\ &= \frac{v' M_2 v}{n} + \frac{C' Z_1' M_2 Z_1 C}{n^2} + 2 \frac{C' Z_1' M_2 v}{n^{3/2}} \\ &\to_p \sigma_v^2, \\ \frac{y_2' M_2 u}{n} &= \frac{v' M_2 u + \pi_1' Z_1 M_2 u}{n} \\ &= \frac{v' M_2 u}{n} + \frac{C' Z_1' M_2 u}{n^{3/2}} \\ &\to_p \sigma_{uv}. \end{aligned}$$

Hence,

$$\hat{\sigma}_u^2 \to_d \sigma_u^2 (1 + \Delta^2 - 2 \Delta \rho_{uv}). \tag{2.13}$$

One can see that inconsistency of $\hat{\gamma}$ causes inconsistency of $\hat{\sigma}_u^2$.

By (2.12) and (2.13),

$$t(\gamma)\to_d \frac{1}{\sigma_u\left(1+\Delta^2-2\Delta\rho_{uv}\right)^{1/2}}\frac{\left(C+Q_{1\cdot2}^{-1}\Psi_{1\cdot2}\right)'\Phi_{1\cdot2}}{\|Q_{1\cdot2}^{1/2}C+Q_{1\cdot2}^{-1/2}\Psi_{1\cdot2}\|}$$
$$=\frac{1}{\left(1+\Delta^2-2\Delta\rho_{uv}\right)^{1/2}}\frac{\left(\lambda+\mathcal{Z}_v\right)'\mathcal{Z}_u}{\|\lambda+\mathcal{Z}_v\|}.$$

where we used (2.10) in the first line. The result follows by plugging in the definition of $\Delta$ from equation (2.6):

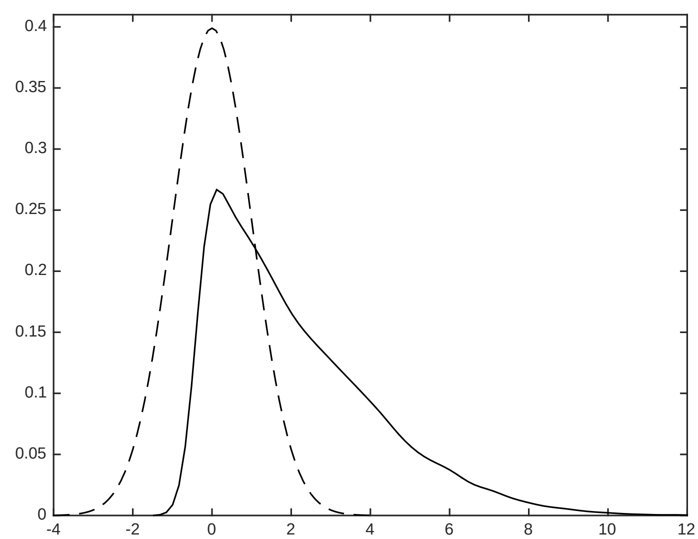
$$1+\Delta^2-2\Delta\rho_{uv}=1+\frac{\left(\left(\lambda+\mathcal{Z}_v\right)'\mathcal{Z}_u\right)^2}{\|\lambda+\mathcal{Z}_v\|^4}-2\rho_{uv}\frac{\left(\lambda+\mathcal{Z}_v\right)'\mathcal{Z}_u}{\|\lambda+\mathcal{Z}_v\|^2}$$
$$=\frac{1}{\|\lambda+\mathcal{Z}_v\|^4}\left(\|\lambda+\mathcal{Z}_v\|^4+\left(\left(\lambda+\mathcal{Z}_v\right)'\mathcal{Z}_u\right)^2-2\rho_{uv}\left(\lambda+\mathcal{Z}_v\right)'\mathcal{Z}_u\|\lambda+\mathcal{Z}_v\|^2\right).$$

□

Figure 1 plots the density of the simulated asymptotic null distribution of the $t$-statistic under weak IVs against the standard normal density.[4] The number of instruments $l_1=2$, the concentration parameter $\|\lambda\|^2=1$, and the correlation coefficient $\rho_{uv}=0.95$. One can see that the distribution of the $t$-statistics substantially deviates from $N(0,1)$. Moreover, the distribution is skewed to the right and puts a lot of mass in what would be the rejection region if one uses standard normal critical values. For example, if the significance level $\alpha=0.05$, $z_{1-\alpha/2}\approx1.96$, the probability of rejecting the null hypothesis is about 43.3% instead of the expected 5%! Thus, weak instruments can cause substantial distortions to inferential procedures (tests and confidence intervals).

[4]The Matlab code used to generate the graph and its data appears in Section B.1 in the Appendix.

Figure 1: The standard normal distribution (dashed line) and the asymptotic null distribution of the $t$-statistic under weak instruments (solid line) for 2 IVs, $\rho_{uv} = 0.95$, and the concentration parameter $\|\lambda\|^2 = 1$



## 3 Detection of weak IVs

In this section, we discuss whether it is possible and how to detect weak IVs. We will see that while it is not possible to formally test if instruments are weak, one can test hypotheses concerning the distortions to the size of $t$-tests due to the presence of weak IVs. The discussion follows Stock and Yogo (2005).

In the previous section, we saw that IVs are said to be strong when the population concentration parameter is infinite:

$$\|\lambda\|^2 = \frac{C' Q_{1\cdot 2} C}{\sigma_v^2} = \infty,$$

i.e. the sample concentration parameter diverges to infinity:

$$\|\lambda_n\|^2 = \frac{\pi_1' Z_1' M_2 Z_1 \pi_1}{\sigma_v^2} \to \infty.$$

While it is possible to approximate $\|\lambda_n\|^2$ using data by replacing $\pi_1$ and $\sigma_v^2$ with their estimators[5], the resulting estimator for $\|\lambda_n\|^2$ is always finite for any value of $n$. Thus, one cannot formally test $H_0 : \|\lambda\|^2 < \infty$ against $H_1 : \|\lambda\|^2 = \infty$. We need a different working definition of weak IVs.

### 3.1 Stock and Yogo's (2005) quantitative characterization of weak IVs

We saw in Section 2.3, that the asymptotic probability of Type I error, i.e. the probability of rejecting $H_0$ when it is true, can substantially exceed assumed significance levels (size) of tests when IVs are weak. Since such distortions are the main concern of having weak instruments, it is reasonable to use the magnitude of such distortions as a practical measure of weakness of instruments.

Theorem 2.3 shows that when IVs are weak in the sense of $\|\lambda\|^2 < \infty$ and the null hypothesis about the structural parameter $\gamma$ is true, the usual $t$-statistic for $\gamma$ converges to the following distribution:

$$\begin{aligned}\mathcal{T}_{\|\lambda\|^2,\rho_{uv},l_1} &= \varphi(\mathcal{Z}_u, \mathcal{Z}_v, \|\lambda\|^2, \rho_{uv}, l_1) \\ &\equiv \frac{\mathcal{X}^{1/2}_{\|\lambda\|^2,\rho_{uv},l_1} \mathcal{Y}_{\|\lambda\|^2,\rho_{uv},l_1}}{\left(\mathcal{Y}^2_{\|\lambda\|^2,\rho_{uv},l_1} + \mathcal{X}^2_{\|\lambda\|^2,\rho_{uv},l_1} - 2\rho_{uv}\mathcal{X}_{\|\lambda\|^2,\rho_{uv},l_1}\mathcal{Y}_{\|\lambda\|^2,\rho_{uv},l_1}\right)^{1/2}},\end{aligned}$$

[5] $\pi_1$ and $\sigma_v^2$ are parameters in the first-stage equation and therefore can be consistently estimated using the usual OLS techniques.

where

$$\mathcal{X}_{\|\lambda\|^2,\rho_{uv},l_1} = \|\lambda + \mathcal{Z}_v\|^2 =^d (\|\lambda\| + \mathcal{Z}_{v,1})^2 + \sum_{j=2}^{l_1} \mathcal{Z}_{v,j}^2,$$

$$\mathcal{Y}_{\|\lambda\|^2,\rho_{uv},l_1} = (\lambda + \mathcal{Z}_v)'\mathcal{Z}_u =^d (\|\lambda\| + \mathcal{Z}_{v,1})\mathcal{Z}_{u,1} + \sum_{j=2}^{l_1} \mathcal{Z}_{v,j}\mathcal{Z}_{u,j},$$

$$\begin{pmatrix} \mathcal{Z}_u \\ \mathcal{Z}_v \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{uv} \\ \rho_{uv} & 1 \end{pmatrix} \otimes I_{l_1}\right).$$

Hence, the asymptotic size of a two-sided $t$-test with nominal size $\alpha$ is given by

$$R_{\alpha,l_1}(\|\lambda\|^2, \rho_{uv}) = P\left(\left|\mathcal{T}_{\|\lambda\|^2,\rho_{uv},l_1}\right| > z_{1-\alpha/2}\right),$$

and the corresponding size distortion is given by

$$R_{\alpha,l_1}(\|\lambda\|^2, \rho_{uv}) - \alpha.$$

As one can see from the above expression, size distortions depend on the chosen significance level $\alpha$, the number of instruments $l_1$, the concentration parameter $\|\lambda^2\|$, and the endogeneity parameter $\rho_{uv}$. While $\alpha$ is chosen by the econometrician, $l_1$ is known, and we can extract some information about $\|\lambda^2\|$ from data, the endogeneity parameter $\rho_{uv}$ is unknown and cannot be estimated since $\hat{\gamma}$ is inconsistent. We therefore have to look at the worst case scenario with respect to $\rho_{uv}$, i.e. the maximum size distortion of a $t$-test:

$$R_{\alpha,l_1}^{\max}(\|\lambda\|^2) - \alpha,$$

where

$$R_{\alpha,l_1}^{\max}(\|\lambda\|^2) = \max_{-1\leq\rho_{uv}\leq1} P\left(\left|\mathcal{T}_{\|\lambda\|^2,\rho_{uv},l_1}\right| > z_{1-\alpha/2}\right)$$

is the maximum with respect to $\rho_{uv}$ probability of Type I error.

Stock and Yogo (2005) propose to measure weakness of instruments in terms of the discrepancy between the nominal size $\alpha$ and the actual size $R_{\alpha,l_1}^{\max}(\|\lambda\|^2)$. More specifically, they suggest to call instruments weak if the difference $R_{\alpha,l_1}^{\max}(\|\lambda\|^2) - \alpha$ exceeds a certain threshold chosen by the econometrician.

**Definition** (Stock and Yogo, 2005)**.** Instrumental variables with concentration parameter $\|\lambda\|^2$ are weak if $R_{\alpha,l_1}^{\max}(\|\lambda\|^2) \geq r$ for some chosen $0 < \alpha \leq r < 1$.

For example, with $\alpha = 0.05$ and $r = 0.10$, we say that instruments are *weak* if the null rejection probability of a nominal 5% $t$-test exceeds 10%. In other words, size distortions exceed 5%. According to this approach one would treat IVs as sufficiently strong if by using 5% standard normal critical values, he obtains a test with actual significance level of 10%. While strictly speaking such IVs are still weak given our earlier definitions, nevertheless when $R^{\max}_{\alpha,l_1}(\|\lambda\|^2) < r$ one can have a valid testing procedure with significance level equal to $r$ by choosing standard normal critical values corresponding to significance level $\alpha$.

In practice to test whether IVs are strong according to definition of Stock and Yogo (2005), one would consider the following testing problem:

$$H_0 : R^{\max}_{\alpha,l_1}(\|\lambda\|^2) \geq r \quad \text{vs.} \quad H_1 : R^{\max}_{\alpha,l_1}(\|\lambda\|^2) < r. \tag{3.1}$$

Note that under $H_1$, instruments are sufficiently strong according to the above definition of Stock and Yogo (2005). Hence, when $H_1$ is true, one can design a valid testing procedure for $\gamma$ despite the instruments being actually weak. One simply has to use larger critical values corresponding to $\alpha$ to obtain a valid level $r$ test.

Since the maximum rejection probability depends on the concentration parameter, one has to calculate the function $R^{\max}_{\alpha,l_1}(\|\lambda\|^2)$ for different values of $\|\lambda\|^2$ to perform the test in (3.1). While it is difficult to obtain an analytical closed-form expression for $R^{\max}_{\alpha,l_1}(\|\lambda\|^2)$ as a function of $\alpha$, $\|\lambda\|^2$ , and $l_1$, one can easily simulate it by taking many draws from the joint distribution of $\mathcal{Z}_u$ and $\mathcal{Z}_v$ and computing the number of average rejections over the draws. It can also be computed by numerical integration of $1\left\{\left|\varphi(\mathcal{Z}_u, \mathcal{Z}_v, \|\lambda\|^2, \rho_{uv}, l_1)\right| > z_{1-\alpha/2}\right\}$ with respect to the joint density of $\mathcal{Z}_u$ and $\mathcal{Z}_v$.

It turns out that the maximum rejection probability and, therefore, the maximum distortion is a non-negative and decreasing function of the concentration parameter. This is related to our previous observation that

$$\mathcal{T}_{\|\lambda\|^2,\rho_{uv},l_1} \to_p N(0,1) \quad \text{as } \|\lambda\|^2 \to \infty,$$

see Remark 4 following Theorem 2.3.

Since $R^{\max}_{\alpha,l_1}(\cdot)$ is a decreasing function (monotone), it can be inverted. Let $L_{\alpha,l_1}(\cdot)$ denote the inverse function of $R^{\max}_{\alpha,l_1}(\cdot)$:

$$L_{\alpha,l_1}(r) \equiv \left(R^{\max}_{\alpha,l_1}\right)^{-1}(r).$$

Thus, for a given value $r$, $L_\alpha(r)$ is the smallest value of the concentration parameter $\|\lambda\|^2$ needed so that the nominal size-$\alpha$ $t$-test would have the maximum significance level (rejection probability) not exceeding $r$.

Once $L_{\alpha,l_1}(r)$ is computed, the testing problem in (3.1) can be re-formulated in terms of hypotheses about the concentration parameter $\|\lambda\|^2$:

$$H_0 : \|\lambda\|^2 \leq L_{\alpha,l_1}(r) \quad \text{vs.} \quad H_1 : \|\lambda\|^2 > L_{\alpha,l_1}(r). \tag{3.2}$$

Again under $H_1$, the instruments can be viewed as sufficiently strong so that one can design a valid testing procedure about $\gamma$ with significance level $r$. The reversion of inequalities in (3.2) relatively to (3.1) is due to the fact that $R^{\max}_{\alpha,l_1}(\|\lambda\|^2)$ is a decreasing function of $\|\lambda\|^2$.

### 3.2 Testing hypotheses about the concentration parameter

Recall that the finite-sample version of the concentration parameter is given by

$$\|\lambda_n\|^2 = \frac{\pi_1' Z_1' M_2 Z_1 \pi_1}{\sigma_v^2}.$$

Using consistent estimators of $\pi_1$ and $\sigma_v^2$, we can estimate the above expression by

$$\|\hat{\lambda}_n\|^2 = \frac{\hat{\pi}_1' Z_1' M_2 Z_1 \hat{\pi}_1}{\hat{\sigma}_v^2}, \tag{3.3}$$

where $\hat{\pi}_1$ was defined earlier in (1.8), and $\hat{\sigma}_v^2$ is a consistent estimator for the variance of the first-stage error $v$. The latter can be constructed as

$$\hat{\sigma}_v^2 = \frac{y_2' M y_2}{n - l_1 - l_2}, \tag{3.4}$$

where for $Z = [Z_1 \ Z_2]$, $M = I_n - Z(Z'Z)^{-1}Z'$. Consistency of $\hat{\sigma}_v^2$ can be shown in the same manner as that of $\hat{\sigma}_u^2$ in (1.19). With those definitions, $\|\hat{\lambda}_n\|^2$ in (3.3) becomes

$$\|\hat{\lambda}_n\|^2 = \frac{y_2' M_2 Z_1 \left(Z_1' M_2 Z_1\right)^{-1} Z_1' M_2 y_2}{y_2' M y_2/(n - l_1 - l_2)} = \frac{y_2' P_{M_2 Z_1} y_2}{y_2' M y_2/(n - l_1 - l_2)},$$

which is related to the usual $F$-statistic for testing the hypothesis that $\pi_1 = 0$:

$$F = \frac{y_2' P_{M_2 Z_1} y_2 / l_1}{y_2' M y_2/(n - l_1 - l_2)}.$$

In finite samples with an additional assumption of normality of the first-stage errors $v$, and when $\pi_1 = 0$, the $F$-statistic has the $F_{l_1, n-l_1-l_2}$ distribution. Without the normality assumption and when $\pi_1 = 0$, the $F$-statistic converges in distribution to $\chi^2_{l_1}/l_1$. Note also that

$$F_{l_1, n-l_1-l_2} \to_p \chi^2_{l_1}/l_1$$

as $n \to \infty$.

While $F_{l_1,n-l_1-l_2}$ (or $\chi^2_{l_1}/l_1$) critical values are appropriate for testing the hypothesis that $\pi_1 = 0$, they cannot be used for testing weakness of IVs as specified in (3.2), since in the latter case $\pi_1 \neq 0$ but only local-to-zero as specified by Assumption 2.1:

$$\pi_1 = n^{-1/2}C.$$

Hence, to use the $F$-statistic for testing the problem in (3.2), we need to obtain its distribution under the local-to-zero specification for $\pi_1$. The distribution turns out to be non-central $\chi^2$ with the non-centrality parameter conveniently given by $\|\lambda\|^2$.

**Theorem 3.1.** *Suppose that Assumptions 1.1 and 2.1 hold. Then,*

$$F_{l_1,n-l_1-l_2} \to_p \chi^2_{l_1}(\|\lambda\|^2)/l_1.$$

*Proof.* As we argued above,

$$\frac{y_2' M y_2}{n - l_1 - l_2} \to_p \sigma_v^2$$

by the same arguments as in the proof of consistency of $\hat{\sigma}_u^2$ in (1.19). Next,

$$\begin{aligned}
\frac{Z_1' M_2 y_2}{n^{1/2}} &= \frac{Z_1' M_2 Z_1 C}{n} + \frac{Z_1' M_2 v}{n^{1/2}} \\
&\to_d Q_{1\cdot 2} C + \Psi_{1\cdot 2} \\
&= \sigma_v Q_{1\cdot 2}^{1/2} \left( Q_{1\cdot 2}^{1/2} C/\sigma_v + \mathcal{Z}_v \right) \\
&= \sigma_v Q_{1\cdot 2}^{1/2} \left( \lambda + \mathcal{Z}_v \right)
\end{aligned}$$

where the convergence in the second line holds by the same arguments as in the proof of Theorem 1.2, equations (1.12) and (1.17), the equality in the third line holds by the definitions of $\Psi_{1\cdot 2}$ and $\mathcal{Z}_v$ in (2.7) and (2.4) respectively, and the equality in the last line holds by the definition of $\lambda$ in (2.5). Hence,

$$\begin{aligned}
F &\to_d \frac{\sigma_v^2 (\lambda + \mathcal{Z}_v)' Q_{1\cdot 2}^{1/2} Q_{1\cdot 2}^{-1} Q_{1\cdot 2}^{1/2} \left( \lambda + \mathcal{Z}_v \right)/l_1}{\sigma_v^2} \\
&= \|\lambda + \mathcal{Z}_v\|^2 / l_1 \\
&\sim \chi^2_{l_1}(\|\lambda\|^2)/l_1,
\end{aligned}$$

where the result in the last line holds by (A.2) in Lemma A.10. □

To test whether IVs are sufficiently strong in the sense of the testing problems in (3.1) and (3.2), the econometrician can proceed as follows. Let $\tau$ denote the significance level of the test for weak IVs.

1. Select $\alpha$ (which determines the normal critical value $z_{1-\alpha}$ to be used for the $t$-test for $\gamma$) and $r$ (the actual desired significance level of the $t$-test).

2. Compute $L_{\alpha,l_1}(r)$ (the smallest concentration parameter needed so that the $t$-test with the critical value $z_{1-\alpha/2}$ has significance level $r$).

3. Compute $F$ (the first-stage $F$-statistic).

4. Select $\tau$ (the desired significance level for the test of weak IVs), and compute $\chi^2_{l_1,1-\tau}(L_{\alpha,l_1}(r))/l_1$ (the non-central $\chi^2$ critical value).

5. Reject $H_0$ of "weak" IVs when $F > \chi^2_{l_1,1-\tau}(L_{\alpha,l_1}(r))/l_1$.

Non-central $\chi^2_{l_1}(\delta)$ critical values can be tabulated and are available with various statistical packages. For example, in Matlab they can computed using the function "`ncx2inv(P,V,DELTA)`", where "`P`" is the quantile order, "V" is the number of degrees of freedom, and "DELTA" is the non-centrality parameter $\delta$.

The most demanding part of performing the test for weak IVs is determining the function $L_{\alpha,l_1}(r)$. The simplest way to obtaining the values $L_{\alpha,l_1}(r)$ is by simulating the function $R^{\max}_{\alpha,l_1}(\|\lambda\|^2)$ for different values of $\|\lambda\|^2$ as $L_{\alpha,l_1}$ is the inverse function of $R^{\max}_{\alpha,l_1}$: $L_{\alpha,l_1}(r) = (R^{\max}_{\alpha,l_1})^{-1}(r)$. Once the values of $R^{\max}_{\alpha,l_1}(\|\lambda\|^2)$ are computed, $L_{\alpha,l_1}(r)$ can be obtained by selecting the smallest value $\|\lambda\|^2$ so that $R^{\max}_{\alpha,l_1}(\|\lambda\|^2) \leq r$.

Table 1 below[6] reports such values of $R^{\max}_{\alpha,l_1}(\|\lambda\|^2)$ for $l_1 = 1, 2, 5$, $\alpha = 0.01, 0.05$, and a range of values for $\|\lambda^2\|$ from 0.01 to $10^3$. The table can be used as follows. Suppose one has two IVs ($l_1 = 2$) and would like to use 1% two-sided standard normal critical values $z_{0.995}$ ($\alpha = 0.01$) to construct a valid two-sided test for $\gamma$ using the $t$-statistic and significance level 5% ($r = 0.05$). According to the data in Table 1, he needs the concentration parameter value approximately equal to 25 ($L_{0.01,2}(0.05) \approx 25$) as the nearest to $r = 0.05$ value for the maximum rejection probability is $R^{\max}_{0.01,2}(25) = 0.045$. The corresponding critical value for the 5% $F$-test for weak IVs is $\chi^2_{2,0.95}(25)/2 \approx 22.66$. Note that this critical value substantially exceeds the threshold of 10 on the first-stage $F$-statistic that has been widely used in the empirical literature. The critical value for the first-stage $F$-statistic remains similarly high if we instead consider $l_1 = 1$ (crit.val $\approx 21.57$) or $l_1 = 5$ (crit.val $\approx 23.53$) while keeping $\alpha = 0.01$ and $r = 0.05$.

Table 1 also shows that to reduce size distortions of the two-sided $t$-test to nearly zero in the case of one instrument, one would need the values of the concentration parameter exceeding 100, which corresponds to the threshold value of 135.60 for the first-stage $F$-statistic. Even large values would be required when there are multiple IVs.

[6]The Matlab code used to generate data for Table 1 appears in Section B.2 in the Appendix.

Table 1: Maximum rejection probabilities of the two-sided $t$-test $R^{\max}_{\alpha,l_1}(\|\lambda\|^2)$ and critical values $\chi^2_{l_1,.95}(\|\lambda\|^2)/l_1$ for testing for weak IVs at the 5% significance level ($\tau = .05$), for different values of the concetration paramer ($\|\lambda\|^2$), number of IVs ($l_1$), and nominal significance level of the $t$-test ($\alpha$ in $z_{1-\alpha/2}$). The number of simulations used to generate each rejection probability is 100,000

| | $l_1 = 1$ | | | $l_1 = 2$ | | | $l_1 = 5$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\|\lambda\|^2$ | $\alpha = .01$ | $\alpha = .05$ | $\chi^2_{1,.95}(\|\lambda\|^2)$ | $\alpha = .01$ | $\alpha = .05$ | $\chi^2_{2,.95}(\|\lambda\|^2)/2$ | $\alpha = .01$ | $\alpha = .05$ | $\chi^2_{5,.95}(\|\lambda\|^2)/5$ |
| 0.01 | 0.530 | 0.593 | 3.88 | 0.817 | 0.863 | 3.01 | 0.992 | 0.995 | 2.22 |
| 0.10 | 0.365 | 0.431 | 4.22 | 0.724 | 0.783 | 3.15 | 0.984 | 0.990 | 2.26 |
| 0.25 | 0.258 | 0.322 | 4.76 | 0.627 | 0.699 | 3.36 | 0.970 | 0.981 | 2.32 |
| 1 | 0.132 | 0.185 | 7.00 | 0.412 | 0.497 | 4.32 | 0.888 | 0.925 | 2.63 |
| 4 | 0.070 | 0.112 | 13.28 | 0.175 | 0.251 | 7.32 | 0.585 | 0.684 | 3.73 |
| 9 | 0.048 | 0.088 | 21.57 | 0.090 | 0.154 | 11.41 | 0.324 | 0.440 | 5.31 |
| 16 | 0.037 | 0.075 | 31.86 | 0.060 | 0.115 | 16.53 | 0.190 | 0.296 | 7.32 |
| 25 | 0.031 | 0.066 | 44.15 | 0.045 | 0.095 | 22.66 | 0.124 | 0.218 | 9.75 |
| 36 | 0.026 | 0.059 | 58.44 | 0.037 | 0.082 | 29.79 | 0.089 | 0.171 | 12.59 |
| 49 | 0.023 | 0.055 | 74.73 | 0.031 | 0.073 | 37.93 | 0.068 | 0.141 | 15.84 |
| 64 | 0.020 | 0.052 | 93.02 | 0.026 | 0.067 | 47.06 | 0.055 | 0.121 | 19.48 |
| 81 | 0.019 | 0.051 | 113.31 | 0.023 | 0.063 | 57.20 | 0.046 | 0.107 | 23.53 |
| 100 | 0.017 | 0.051 | 135.60 | 0.021 | 0.061 | 68.34 | 0.040 | 0.097 | 27.98 |
| 1000 | 0.011 | 0.050 | 1106.74 | 0.011 | 0.052 | 553.88 | 0.013 | 0.055 | 222.17 |

# 4 Robust inference in presence of potentially weak IVs: Anderson-Rubin (AR) approach

One can use the methods described in the previous section to design valid tests on the structural parameter $\gamma$ by choosing higher critical values and checking that the first-stage $F$-statistic is sufficiently large. Nevertheless, this approach cannot give a useful test when the first-stage $F$-statistic is not large enough. Moreover, while by choosing larger critical values, one can somewhat protect himself from size distortions, this leads to the loss of power if IVs are in fact strong.

In this section, we discuss an alternative approach to dealing with weak IVs that produces valid tests regardless of the strength of IVs. Such a robust approach has been advocated by most papers in the econometric literature on weak IVs. The Anderson-Rubin approach discussed in this section was proposed as a solution to the weak IV problem in Staiger and Stock (1997).

## 4.1 AR statistic

The idea of the AR test and other robust tests is based on the fact that, whether IVs are weak or not, they are uncorrelated with the errors:

$$0 = EZ_{1i}u_i = EZ_{1i}(y_{1i} - y_{2i}\gamma - Z_{2i}'\beta).$$

In other words, if one imposes right parameter values on the structural coefficients, the resulting residuals must by uncorrelated with the IVs. However, when wrong a parameter value is used for $\gamma$, the resulting residuals will contain the endogenous regressor $y_{2i}$, which is correlated with the IVs. Hence, one should be able to test hypotheses about the structural coefficients without estimating them first.

Suppose the econometrician is interested in testing

$$H_0 : \gamma = \gamma_0 \quad \text{against} \quad H_1 : \gamma \neq \gamma_0.$$

Note that here the focus is only on $\gamma$, while the coefficients on the exogenous regressors $Z_{2i}$ are unrestricted under the null or alternative hypotheses. To eliminate $\beta$, consider the following *null-restricted residuals*:

$$M_2(y_1 - y_2\gamma_0) = M_2(u + y_2(\gamma - \gamma_0)),$$

where again $M_2 = I_n - Z_2(Z_2'Z_2)^{-1}Z_2'$ so that $M_2Z_2 = 0$. One can see that the null-restricted residuals depend on the true errors $u$ and the distance between the true $\gamma$ and $\gamma_0$ (its value under $H_0$). Moreover, the sample covariance between the null-restricted residuals and the IVs is given by

$$Z_1'M_2(y_1 - y_2\gamma_0) = Z_1'M_2u + (\gamma - \gamma_0)Z_1'M_2y_2. \tag{4.1}$$

When $H_0$ is true and $\gamma = \gamma_0$, the null-restricted residuals capture only the true errors $u$ and therefore,

$$\begin{aligned}\frac{Z_1'M_2(y_1 - y_2\gamma_0)}{n^{1/2}} &= \frac{Z_1'M_2u}{n^{1/2}} \\ &\to_d N(0, \sigma_u^2 Q_{1\cdot 2}).\end{aligned} \tag{4.2}$$

Moreover, the variance parameter $\sigma_u^2$ can be also consistently estimated by using the null-restricted residuals under $H_0$. For $Z = [Z_1\ Z_2]$, let again

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

and note that

$$0 = MZ = M[Z_1\ Z_2]. \tag{4.3}$$

When $\gamma_0 = \gamma$,

$$\begin{aligned}\hat{\sigma}_u^2(\gamma_0) &= \frac{(y_1 - y_2\gamma_0)'M(y_1 - y_2\gamma_0)}{n} && (4.4)\\ &= \frac{u'Mu}{n} && (4.5)\\ &= \frac{u'u}{n} - \frac{u'Z}{n}\left(\frac{Z'Z}{n}\right)^{-1}\frac{Z'u}{n} \\ &\to_p \sigma_u^2 - 0 \cdot Q^{-1} \cdot 0 \\ &= \sigma_u^2,\end{aligned}$$

where $\hat{\sigma}_u^2(\gamma_0)$ is the null-restricted estimator of $\sigma_u^2$.[7, 8]

The AR statistic is constructed by utilizing the null-restricted residuals and the fact that under $H_0$, the econometrician knows the true value of $\gamma$.

$$AR(\gamma_0) = \frac{(y_1 - y_2\gamma_0)'M_2Z_1\left(Z_1'M_2Z_1\right)^{-1}Z_1'M_2(y_1 - y_2\gamma_0)}{(y_1 - y_2\gamma_0)'M(y_1 - y_2\gamma_0)/n}$$

[7]Alternatively, on could use $M_2$ matrix in place of $M$ to construct a null-restricted estimator of $\sigma_u^2$; such an estimator would still be consistent under $H_0$. However, using $M$ in place of $M_2$ is preferable when $H_0$ does not hold and $\gamma \neq \gamma_0$. When $H_0$ is false, $y_1 - y_2\gamma_0 = u + (\gamma - \gamma_0)(Z_1\pi_1 + Z_2\pi_2 + v)$ . Using the $M$ matrix, will annihilate in such cases both $Z_1$ and $Z_2$ in view of (4.3), while $M_2$ annihilates only $Z_2$.

[8]In finite samples, one also might want to substitute $n - l_1 - l_2$ for $n$ in the definition of the null-restricted estimator $\hat{\sigma}_u^2(\gamma_0)$. Note that when $\gamma = \gamma_0$, $(y_1 - y_2\gamma_0)'M_2(y_1 - y_2\gamma_0)/(n - l_1 - l_2)$ is an unbiased estimator of $\sigma_u^2$ since $\text{rank}(M) = n - l_1 - l_2$.

$$= \frac{(y_1 - y_2\gamma_0)' P_{M_2 Z_1} (y_1 - y_2\gamma_0)}{(y_1 - y_2\gamma_0)' M (y_1 - y_2\gamma_0)/n}, \tag{4.6}$$

where $(Z_1' M_2 Z_1)^{-1}$ in the definition of the AR statistic is used to estimate $Q_{1\cdot 2}$ that appears in the limiting distribution in (4.2). The fact that the first-stage parameter $\pi_1$ does not appear on the right-hand side of (4.2), implies that the null distribution of the AR statistic does not depend on the strength of IVs.

**Theorem 4.1.** *Suppose that Assumption 1.1 holds. Then,*

$$AR(\gamma) \to_d \chi^2_{l_1}.$$

*Remark.*

1. Note the use of the true value $\gamma$ in $AR(\gamma)$, which implies that the result applies when $H_0 : \gamma = \gamma_0$ is true.
2. Note also that no assumptions were imposed on $\pi_1$. The result remains valid when $\pi_1$ is fixed and the concentration parameter $\|\lambda\|^2 = \infty$, when $\pi_1 = n^{-1/2}C$ and $\|\lambda\|^2 < \infty$, and when $\pi_1 = 0$ so that IVs are *irrelevant* and the concentration parameter is exactly zero.
3. While $\gamma$ is a scalar parameter, the null distribution of the AR statistic has $l_1 \geq 1$ degrees of freedom. This is because AR tests a single restriction on $\gamma$ by testing $l_1$ restrictions on the covariances of the IVs with the null-restricted residuals.

*Proof.* Since $y_1 - y_2\gamma = Z_2\beta + u$, and $M_2 Z_2 = 0$,

$$\begin{aligned} AR(\gamma) &= \frac{u' M_2 Z_1 \left(Z_1' M_2 Z_1\right)^{-1} Z_1' M_2 u}{u' M u/n} \\ &= \frac{u' M_2 Z_1/n^{1/2} \left(Z_1' M_2 Z_1/n\right)^{-1} Z_1' M_2 u/n^{1/2}}{u' M u/n} \\ &\to_d \frac{\Phi_{1\cdot 2}' Q_{1\cdot 2}^{-1} \Phi_{1\cdot 2}}{\sigma_u^2} \quad \text{for } \Phi_{1\cdot 2} \sim N(0, \sigma_u^2 Q_{1\cdot 2}) \\ &=^d \mathcal{Z}_u' \mathcal{Z}_u \quad \text{for } \mathcal{Z}_u \sim N(0, I_{l_1}) \\ &\sim \chi^2_{l_1}, \end{aligned}$$

where $\Phi_2$ is defined in (1.18), and $\mathcal{Z}_u$ is defined in (2.4). □

In view of the result of Theorem 4.1, the size $\alpha$ AR test should reject $H_0 : \gamma = \gamma_0$ in favor of $H_1 : \gamma \neq \gamma_0$ when

$$AR(\gamma_0) > \chi^2_{l_1, 1-\alpha}.$$

While the strength of IVs ($\pi_1$ or $\|\lambda\|^2$) does not affect the distribution of the AR statistic under the null, it matters under the alternative when $\gamma \neq \gamma_0$ as one can see from the second term on the right-hand side of (4.1): $y_2$ (and therefore $\pi_1$) contribute to the distribution when $\gamma - \gamma_0 \neq 0$.

**Theorem 4.2.** *Suppose that Assumptions 1.1 and 2.1 hold: i.e. the IVs are weak in the sense that $\pi_1 = n^{-1/2}C$ and $\|\lambda\|^2 < \infty$. Suppose further that $\gamma - \gamma_0$ is a fixed number. Then,*

$$AR(\gamma_0) \to_d \chi^2_{l_1}\left(\frac{(\gamma-\gamma_0)^2\sigma_v^2}{\sigma_u^2 + (\gamma-\gamma_0)^2\sigma_v^2 + 2(\gamma-\gamma_0)\sigma_{uv}}\|\lambda\|^2\right).$$

*Remark.*

1. The power properties of the AR test can be analyzed by studying the non-centrality parameter

$$\frac{(\gamma-\gamma_0)^2\sigma_v^2}{\sigma_u^2 + (\gamma-\gamma_0)^2\sigma_v^2 + 2(\gamma-\gamma_0)\sigma_{uv}}\|\lambda\|^2.$$

   For example, when $\gamma = \gamma_0$ (i.e. $H_0$ is true), the non-centrality parameter becomes zero and one obtains the result of Theorem 4.1.

2. The non-centrality parameter is also zero when $\gamma \neq \gamma_0$ but $\|\lambda\|^2 = 0$, which occurs when $\pi_1 = 0$, i.e. IVs are irrelevant. In this case, the distribution of the AR statistic is central $\chi^2$ for any value of $\gamma - \gamma_0$, and the probability of rejecting $H_0$ always equal to $\alpha$. Thus, when IVs are irrelevant, one can have a valid tests for $\gamma$ (in the sense of its size properties), however, the test has no power: it detects deviations from $H_0$ with a trivial probability $\alpha$.

3. Since $P(\chi^2_{l_1}(\delta^2) > \chi^2_{l_1,1-\alpha})$ is an increasing function of the non-centrality parameter $\delta^2$ as is apparent from (A.2), the probability that the AR would detect a deviation from $H_0$ is an increasing function of the concentration parameter $\|\lambda\|^2$. Since the concentration parameter captures the strength of IVs, the AR becomes more powerful as $\|\lambda^2\|$ increases.

4. An unusual feature of the AR test under weak IVs is that the non-centrality parameter does not diverge to $\infty$ as the distance from the null $\gamma - \gamma_0$ increases. Instead,

$$\lim_{\gamma-\gamma_0\to\infty} \frac{(\gamma-\gamma_0)^2\sigma_v^2}{\sigma_u^2 + (\gamma-\gamma_0)^2\sigma_v^2 + 2(\gamma-\gamma_0)\sigma_{uv}}\|\lambda\|^2 \to \|\lambda\|^2.$$

   Hence, even for very large deviations from the null hypothesis, the non-centrality parameter can be small (if $\|\lambda\|^2$ is small), and consequently the probability of detecting very large deviations from $H_0$ would be finite (and can be arbitrary small). Moreover, the non-centrality parameter can be non-monotone in $\gamma - \gamma_0$ and, as a result, the probability of detecting certain small deviations from the null hypothesis can exceed that for large deviations. Unfortunately, these unusual and undesirable features cannot be avoided when there are weak IVs as follows from the results of Dufour (1997).

*Proof of Theorem 4.2.* In the numerator of the AR statistic, we have

$$\begin{aligned}
\frac{Z_1' M_2 (y_1 - y_2\gamma_0)}{n^{1/2}} &= \frac{Z_1' M_2 u}{n^{1/2}} + (\gamma - \gamma_0)\frac{Z_1' M_2 y_2}{n^{1/2}} \\
&= \frac{Z_1' M_2 u}{n^{1/2}} + (\gamma - \gamma_0)\frac{Z_1' M_2 (v + Z_1\pi_1)}{n^{1/2}} \\
&= \frac{Z_1' M_2 u}{n^{1/2}} + (\gamma - \gamma_0)\frac{Z_1' M_2 v}{n^{1/2}} + (\gamma - \gamma_0)\frac{Z_1' M_2 Z_1 C}{n} \\
&\to_d \Phi_{1\cdot 2} + (\gamma - \gamma_0)\Psi_{1\cdot 2} + (\gamma - \gamma_0) Q_{1\cdot 2} C \\
&=^d Q_{1\cdot 2}^{1/2}\left(\sigma_u \mathcal{Z}_u + (\gamma - \gamma_0)\sigma_v \mathcal{Z}_v + (\gamma - \gamma_0) Q_{1\cdot 2}^{1/2} C\right), \qquad (4.7)
\end{aligned}$$

where $\Phi_{1\cdot 2}$ and $\Psi_{1\cdot 2}$ are defined in (2.7), and $\mathcal{Z}_u$ and $\mathcal{Z}_v$ are defined in (2.4). Note also that by the definition in (2.4) and due to the properties of multivariate normal distributions (see Theorem A.5 in the Appendix),

$$\sigma_u \mathcal{Z}_u + (\gamma - \gamma_0)\sigma_v \mathcal{Z}_v \sim N\left(0, (\sigma_u^2 + (\gamma - \gamma_0)^2\sigma_v^2 + 2(\gamma - \gamma_0)\sigma_{uv}) I_{l_1}\right) \qquad (4.8)$$

as $\sigma_{uv} = \rho_{uv}\sigma_u\sigma_v$.

In the denominator of the AR statistic, we have:

$$\begin{aligned}
&\frac{(y_1 - y_2\gamma_0)' M (y_1 - y_2\gamma_0)}{n} = \\
&\quad = \frac{\left(u + (\gamma - \gamma_0)(Z_1 C/n^{1/2} + v)\right)' M \left(u + (\gamma - \gamma_0)(Z_1 C/n^{1/2} + v)\right)}{n} \\
&\quad = \frac{u' M u}{n} + (\gamma - \gamma_0)^2 \frac{v' M v}{n} + 2(\gamma - \gamma_0)\frac{u' M v}{n} + o_p(1) \\
&\quad \to_p \sigma_u^2 + (\gamma - \gamma_0)^2 \sigma_v^2 + 2(\gamma - \gamma_0)\sigma_{uv}, \qquad (4.9)
\end{aligned}$$

which is the same as the variance in (4.8).

Putting (4.7)-(4.9) together, we obtain:

$$\begin{aligned}
AR(\gamma_0) &\to_d \frac{\left\|\sigma_u \mathcal{Z}_u + (\gamma - \gamma_0)\sigma_v \mathcal{Z}_v + (\gamma - \gamma_0) Q_{1\cdot 2}^{1/2} C\right\|^2}{\sigma_u^2 + (\gamma - \gamma_0)^2\sigma_v^2 + 2(\gamma - \gamma_0)\sigma_{uv}} \\
&\sim \left\| N(0, I_{l_1}) + \frac{\gamma - \gamma_0}{(\sigma_u^2 + (\gamma - \gamma_0)^2\sigma_v^2 + 2(\gamma - \gamma_0)\sigma_{uv})^{1/2}} Q_{1\cdot 2}^{1/2} C \right\|^2 \\
&= \left\| N(0, I_{l_1}) + \frac{(\gamma - \gamma_0)\sigma_v}{(\sigma_u^2 + (\gamma - \gamma_0)^2\sigma_v^2 + 2(\gamma - \gamma_0)\sigma_{uv})^{1/2}} \lambda \right\|^2 \\
&=^d \chi^2_{l_1}\left(\frac{(\gamma - \gamma_0)^2\sigma_v^2}{\sigma_u^2 + (\gamma - \gamma_0)^2\sigma_v^2 + 2(\gamma - \gamma_0)\sigma_{uv}} \|\lambda\|^2\right),
\end{aligned}$$

where recall $\lambda = Q_{1\cdot 2}^{1/2} C/\sigma_v$ and the result in the last line holds by the definition of the non-central $\chi^2$ distribution in Lemma A.10, equation (A.2). □

### 4.2 Robust confidence sets (CSs) based on the AR statistic

The AR test can be used to construct CSs for $\gamma$ that remain valid whether IVs are strong or weak. Instead of using the standard approach to building confidence intervals (CIs), i.e. instead of using the formula

$$CI_{1-\alpha} = \text{estimate} \pm \text{std.err} \times z_{1-\alpha/2},$$

robust CSs can be constructed by inverting any test robust to weak IVs, including the AR test. Inversion of a test means that we are going to include in the CS all values $\gamma_0$ that could not be rejected when testing $H_0 : \gamma = \gamma_0$ against $H_1 : \gamma \neq \gamma_0$. In the case of the size $\alpha$ AR test, the corresponding CS with coverage probability $1 - \alpha$ is

$$CS_{1-\alpha}^{AR} = \{\gamma_0 \in \mathbb{R} : \ AR(\gamma_0) \leq \chi^2_{l_1, 1-\alpha}\}. \tag{4.10}$$

Note that the significance level of the AR test ($\alpha$) should match the coverage probability of its CS ($1 - \alpha$). The validity of the AR CS constructed by inversion is follows immediately from the validity of the AR test in Theorem 4.1.

**Theorem 4.3.** *Suppose that Assumption 1.1 holds. Then,*

$$P(\gamma \in CS_{1-\alpha}^{AR}) \to 1 - \alpha.$$

*Proof.* By the definition of the AR CS in (4.10),

$$P\left(\gamma \in CS_{1-\alpha}^{AR}\right) = P\left(AR(\gamma) \leq \chi^2_{l_1, 1-\alpha}\right) \to 1 - \alpha,$$

where the last result holds by Theorem 4.1. □

The definition of the AR CS in (4.10) means that one has to perform a sequence of tests for different values of $\gamma_0$ and then collect all the values of $\gamma_0$ that could not be rejected. This may seem as a computationally intensive/cumbersome procedure. In fact however, when the model has only one endogenous regressor as in our case, the AR CS can be constructed simply by *solving a single quadratic equation.*

**Theorem 4.4.** $\gamma_0 \in CS_{1-\alpha}^{AR}$ *if and only if it solves the following quadratic equation:*

$$a_n \gamma_0^2 - b_n \gamma_0 + c_n \leq 0, \tag{4.11}$$

*where*

$$
\begin{aligned}
a_n &= y_2' \left( P_{M_2 Z_1} - n^{-1} \chi^2_{l_1, 1-\alpha} M \right) y_2, \\
b_n &= 2 y_2' \left( P_{M_2 Z_1} - n^{-1} \chi^2_{l_1, 1-\alpha} M \right) y_1, \\
c_n &= y_1' \left( P_{M_2 Z_1} - n^{-1} \chi^2_{l_1, 1-\alpha} M \right) y_1,
\end{aligned}
$$

The proof of Theorem 4.4 follows immediately from the definitions of the AR statistic in (4.6) and AR CS in (4.10).

When there is a single IV, i.e. $l_1 = 1$ and the model is exactly identified, it is easy to show that $CS^{AR}_{1-\alpha/2}$ cannot be empty: there are always values $\gamma_0$ that satisfy the inequality in (4.11). In particular, one such value is the IV estimator $\hat{\gamma}$.

**Lemma 4.5.** *When* $l_1 = 1$, $\hat{\gamma} \in CS^{AR}_{1-\alpha}$.

*Proof.* When $l_1 = 1$,

$$
Z_1' M_2 (y_1 - y_2 \gamma_0) = 0 \text{ for } \gamma_0 = \frac{Z_1' M_2 y_1}{Z_1' M_2 y_2},
$$

in which case the first line in (4.6) implies that the AR statistic is equal to zero for such a value of $\gamma_0$. However, when $l_1 = 1$, (1.9) implies that

$$
\begin{aligned}
\hat{\gamma} &= \frac{\hat{\pi}_1' Z_1' M_2 y_1}{\hat{\pi}_1' Z_1' M_2 Z_1 \hat{\pi}_1} \\
&= \frac{Z_1' M_2 y_1}{Z_1' M_2 Z_1 \hat{\pi}_1} \\
&= \frac{Z_1' M_2 y_1}{Z_1' M_2 Z_1 \frac{Z_1' M_2 y_2}{Z_1' M_2 Z_1}} \\
&= \frac{Z_1' M_2 y_1}{Z_1' M_2 y_2},
\end{aligned}
$$

where the equality in the second line holds because $\hat{\pi}_1$ is $1 \times 1$ when $l_1 = 1$, and the equality in the third line holds by the definition of $\hat{\pi}_1$ in (1.8). □

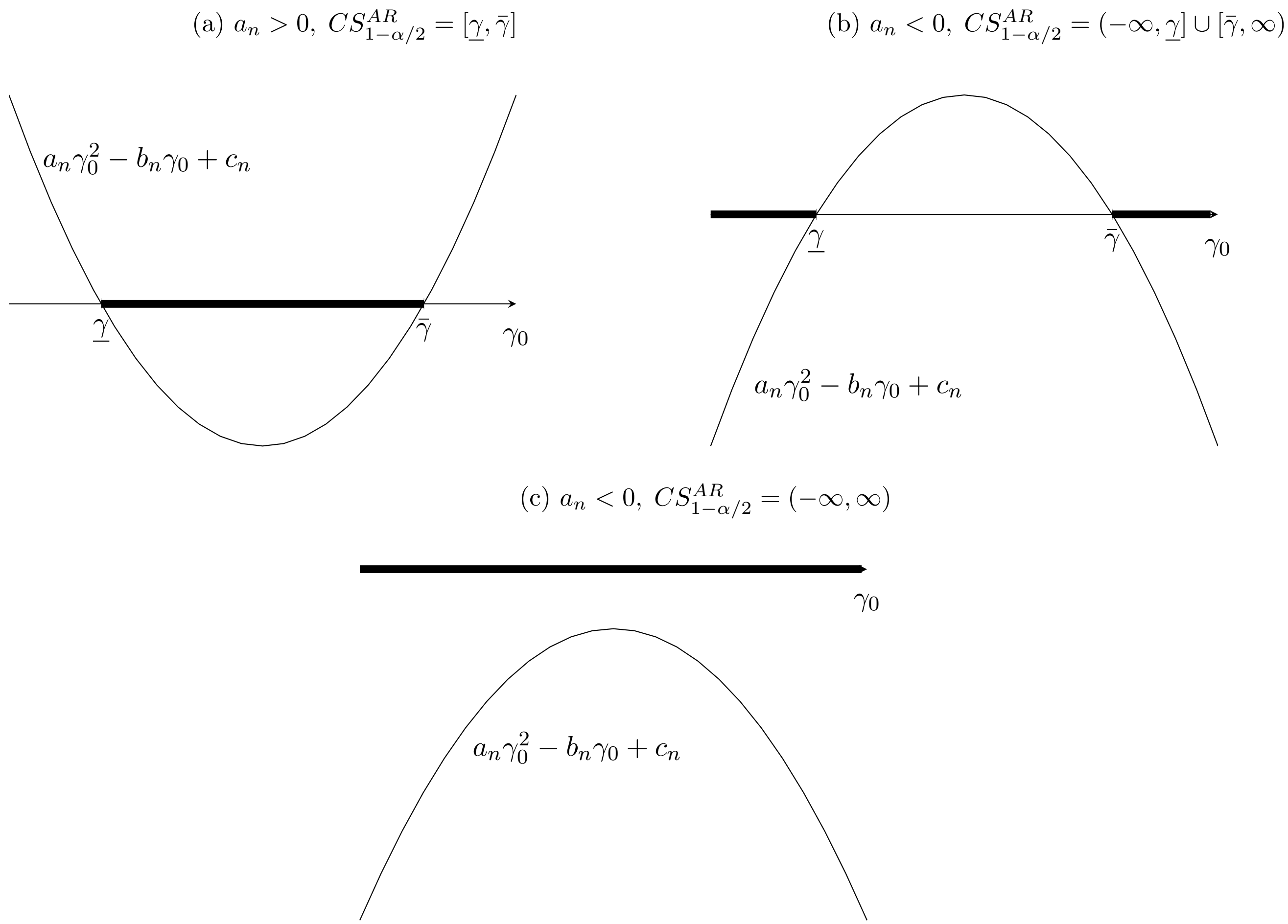
Since the AR CS is defined by a quadratic inequality and cannot be empty when $l_1 = 1$, in that case it can take one of the three forms, as illustrated in Figure 2:

(a) A compact interval of the form $[\underline{\gamma}, \bar{\gamma}]$ if $a_n > 0$.

(b) The union of two half-lines: $(-\infty, \underline{\gamma}] \cup [\bar{\gamma}, \infty)$ when $a_n < 0$.

(c) The entire real line: $(-\infty, \infty)$ when $a_n < 0$.

Figure 2: Possible forms of AR CSs when $l_1 = 1$ for different values of $a_n$. AR CSs are shown by thick lines



The forms in (b) and (c) are unusual and the reason we use the terminology "confidence sets" (CSs) instead of confidence intervals (CIs). In the case of (b) and (c), CSs are unbounded and, therefore, a less informative than the usual interval form in (a). Moreover, in case (c) CSs are completely uninformative.

While the situations described by (b) and (c) thus may seem undesirable, they cannot be avoided if one wants to construct valid confidence sets. This has been shown in Dufour (1997). Moreover, one can show that if IVs are in fact strong, AR CSs are bounded with probability approaching one, as is implied the theorem below.

**Theorem 4.6.** *Suppose that Assumption 1.1 holds.*

*(a)* $P(a_n > 0) \to 1$*, when* $\pi_1$ *is fixed (i.e. IVs are strong).*

*(b)* $\lim_{n\to\infty} P(a_n > 0) = P\left(\|\lambda + \mathcal{Z}_v\|^2 > \chi^2_{l_1,1-\alpha}\right) < 1$*, when* $\pi_1 = n^{-1/2}C$ *(and therefore* $\|\lambda\|^2 < \infty$*, i.e. IVs are weak).*

*Remark.* Recall that in part (b), $\|\lambda + \mathcal{Z}_v\|^2 \sim \chi^2_{l_1}(\|\lambda\|^2)$, which is an increasing function of the concentration parameter $\|\lambda\|^2$. Hence, the probability of the AR CS taking the form of a compact interval depends on the strength of identification. However, for every finite value of the concentration parameter, the probability that a $\chi^2_{l_1}(\|\lambda\|^2)$-distributed random variable takes a value greater than $\chi^2_{l_1,1-\alpha}$ is strictly less than 1.

*Proof.* Write:

$$a_n = y_2' P_{M_2 Z_1} y_2 - \chi^2_{l_1,1-\alpha} \frac{y_2' M y_2}{n}. \tag{4.12}$$

Since both terms on the right-hand side are quadratic forms constructed using positive-semidefinite matrices, $a_n > 0$ when the first term dominates the second. For the first term, we have:

$$\begin{aligned}\frac{y_2' P_{M_2 Z_1} y_2}{n} &= \frac{(Z_1\pi_1 + v)' M_2 Z_1 (Z_1' M_2 Z_1)^{-1} Z_1' M_2 (Z_1 \pi_1 + v)}{n} \\ &= \frac{\pi_1' Z_1' M_2 Z_1 \pi_1}{n} + \frac{v' M_2 Z_1}{n} \left(\frac{Z_1' M_2 Z_1}{n}\right)^{-1} \frac{Z_1' M_2 v}{n} + 2\frac{\pi_1' Z_1' M_2 v}{n} \\ &\to_p \pi_1' Q_{1\cdot 2} \pi_1. \end{aligned} \tag{4.13}$$

Hence,

$$y_2' P_{M_2 Z_1} y_2 = O_p(n),$$

i.e. it diverges to $+\infty$ as $n \to \infty$. For the second term, similar calculations show that

$$\frac{y_2' M y_2}{n} = \frac{v' M_2 v}{n} \to_p \sigma_v^2.$$

Hence

$$a_n = O_p(n) - \chi^2_{l_1,1-\alpha} O_p(1),$$

and the result in part (a) follows.

For part (b), one can show that the two terms in (4.12) are of the same magnitude. For the first term, we have:

$$\begin{aligned}
y_2' P_{M_2 Z_1} y_2 &= \left(Z_1 C/n^{1/2} + v\right)' P_{M_2 Z_1} \left(Z_1 C/n^{1/2} + v\right) \\
&= \frac{C' Z_1' M_2 Z_1 C}{n} + \frac{v' M_2 Z_1}{n^{1/2}} \left(\frac{Z_1' M_2 Z_1}{n}\right)^{-1} \frac{Z_1' M_2 v}{n^{1/2}} + 2 \frac{C' Z_1 M_2 v}{n^{1/2}} \\
&\to_d C' Q_{1\cdot 2} C + \Psi_{1\cdot 2}' Q_{1\cdot 2}^{-1} \Psi_{1\cdot 2} + 2 C' \Psi_{1\cdot 2} \\
&= \left\| Q_{1\cdot 2}^{1/2} C + Q_{1\cdot 2}^{-1/2} \Psi_{1\cdot 2} \right\|^2 \\
&=^d \sigma_v^2 \left\| \lambda + \mathcal{Z}_v \right\|^2,
\end{aligned}$$

where $\Psi_{1\cdot 2}$ and $\mathcal{Z}_v$ are defined in (2.7) and (2.4) respectively. Therefore,

$$\begin{aligned}
P\left(a_n > 0\right) &\to P\left( \left\| \lambda + \mathcal{Z}_v \right\|^2 - \chi^2_{l_1,1-\alpha} > 0 \right) \\
&= P\left( \chi^2_{l_1}(\|\lambda\|^2) > \chi^2_{l_1,1-\alpha} \right) \\
&< 1,
\end{aligned}$$

where the inequality in the last line holds for every $\|\lambda\|^2 < \infty$. □

In the case when the model is over-identified ($l_1 > 1$), there is one additional possibility for form of the AR CS: it can be empty. AR-based CSs would be empty when there is no value $\gamma_0$ that makes the null-restricted residuals $y_1 - y_2\gamma_0$ uncorrelated with the IVs $Z_1$, i.e. even for the true value $\gamma$, $y_1 - y_2\gamma = u$ remain correlated with the IVs $Z_1$ . This would imply that the IVs are invalid in the sense that the restriction $E u_i Z_{1i} = 0$ does not hold. Another possibility is that the model is grossly misspecified and the relationship between $y_1$ and $y_2$ is non-linear. In either case, one should not use IVs in such situation. Thus, AR CSs have a built-in model specification test with empty CSs implying that the model ($y_{1i} = y_2\gamma + Z_{2i}\beta + u_i$ with $E Z_{1i} u_i = 0$ and $E Z_{2i} u_i = 0$) is rejected.

To conclude this section, we will show that AR CSs will correctly capture the true parameter value when IVs are strong and $n \to \infty$.

**Theorem 4.7.** *Suppose that Assumption 1.1 holds and $\pi_1$ is fixed (i.e. IVs are strong). Then,*

$$P\left(CS^{AR}_{1-\alpha} = \{\gamma\}\right) \to 1.$$

*Remark.* The theorem indicates that in the case of strong IVs, AR CSs will collapse to a single point, which is the true value of the coefficient $\gamma$, as $n \to \infty$. This is the behavior one would expect from a good CS/CI when identification is strong as usual CIs based on the IV estimator and its standard error have this property.

*Proof.* First note that

$$a_n \gamma_0^2 - b_n \gamma_0 + c_n \leq 0$$

if and only if

$$\frac{a_n}{n}\gamma_0^2 - \frac{b_n}{n}\gamma_0 + \frac{c_n}{n} \leq 0.$$

As is show in the proof of Theorem 4.6(a),

$$\frac{a_n}{n} \to \pi_1' Q_{1\cdot 2} \pi_1.$$

Next,

$$\begin{aligned}
\frac{b_n}{n} &= \frac{2y_2'\left(P_{M_2 Z_1} - n^{-1}\chi^2_{l_1,1-\alpha} M\right) y_1}{n} \\
&= 2\frac{y_2' P_{M_2 Z_1} y_1}{n} + o_p(1) \\
&= 2\frac{(Z_1\pi_1 + v)' M_2 Z_1 \left(Z_1' M_2 Z_1\right)^{-1} Z_1' M_2 \left(y_2\gamma + u\right)}{n} + o_p(1) \\
&= 2\frac{(Z_1\pi_1 + v)' M_2 Z_1}{n} \left(\frac{Z_1' M_2 Z_1}{n}\right)^{-1} \frac{Z_1' M_2 \left(Z_1\pi_1\gamma + v\gamma + u\right)}{n} + o_p(1) \\
&\to_p 2\pi_1' Q_{1\cdot 2}\pi_1\gamma, \\
\frac{c_n}{n} &= \frac{y_1'\left(P_{M_2 Z_1} - n^{-1}\chi^2_{l_1,1-\alpha} M\right) y_1}{n} \\
&= \frac{\left(Z_1\pi_1\gamma + v\gamma + u\right)' M_2 Z_1}{n} \left(\frac{Z_1' M_2 Z_1}{n}\right)^{-1} \frac{Z_1' M_2 \left(Z_1\pi_1\gamma + v\gamma + u\right)}{n} + o_p(1) \\
&\to_p \pi_1' Q_{1\cdot 2}\pi_1\gamma^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{a_n}{n}\gamma_0^2 - \frac{b_n}{n}\gamma_0 + \frac{c_n}{n} &\to_p \pi_1' Q_{1\cdot 2}\pi_1 \left(\gamma_0^2 - 2\gamma_0\gamma + \gamma^2\right) \\
&= \pi_1' Q_{1\cdot 2}\pi_1 \left(\gamma_0 - \gamma\right)^2.
\end{aligned}$$

Hence, as $n \to \infty$, $\gamma_0 = \gamma$ is the only value that satisfies the limiting version of the inequality defining the AR CS. □

## 5 Improving the power of robust inference

The AR approach described in Section 4 allows one to test hypotheses about the structural coefficient $\gamma$ and construct confidence sets for $\gamma$ despite having potentially weak IVs. The approach "works" in the sense that it remains valid whether IVs are weak or strong. Nevertheless, it turns out that when IVs are strong, the AR approach is not as powerful as the traditional approach based on the IV estimator of $\gamma$ and its standard error. This section discusses the power issues of the AR approach and how they can be alleviated. We will discuss alternative weak-IV-robust test statistics that can be used to design inference procedures that remain valid when IVs are weak, and are as powerful as the traditional $t$-statistic-based inference when IVs are strong.

### 5.1 Power of the $t$-test under strong IVs

Consider again testing $H_0 : \gamma = \gamma_0$ against a two-sided alternative $H_1 : \gamma \neq \gamma_0$. Recall from Section 1.3 equation (1.20), that the $t$-statistic is defined as

$$t(\gamma_0) = \frac{n^{1/2}(\hat{\gamma} - \gamma_0)}{\sqrt{\frac{\hat{\sigma}_u^2}{y_2' P_{M_2 Z_1} y_2 / n}}}, \tag{5.1}$$

and the two-sided $t$-test rejects $H_0$ when

$$|t(\gamma_0)| > z_{1-\alpha/2}.$$

The following theorem shows that the power of the $t$-test can be described using a non-central $\chi_1^2$ random variable, and thus is completely characterized by a scalar non-centrality parameter.

In the case of strong IVs to describe the power of a test, we should consider small deviations of the truth from the null hypothesis:

$$\gamma = \gamma_0 + n^{-1/2}\delta$$

for some unknown $\delta \in \mathbb{R}$.

**Theorem 5.1.** *Suppose that Assumption 1.1 holds, $\gamma = \gamma_0 + n^{-1/2}\delta$, and $\pi_1$ is fixed, i.e IVs are strong. Then,*

$$P\left(|t(\gamma_0)| > z_{1-\alpha/2}\right) \to P\left(\left(\delta\left(\frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma_u^2}\right)^{1/2} + \mathcal{Z}\right)^2 > \chi^2_{1,1-\alpha}\right),$$

*where $\mathcal{Z} \sim N(0,1)$.*

*Remark.* By the definition of the non-central $\chi^2$ distribution,

$$\left(\delta\left(\frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma_u^2}\right)^{1/2} + \mathcal{Z}\right)^2 \sim \chi_1^2\left(\delta^2 \frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma_u^2}\right). \tag{5.2}$$

The non-centrality parameter on the right-hand side of equation (5.2) describes the power of the $t$-test in the case of strong IVs. It depends on two components: $\delta^2$ measures the distance from $H_0$, and $\pi_1' Q_{1\cdot 2}\pi_1/\sigma_u^2$ measures the strength of IVs relatively to the noise of the errors in the structural equation. Thus, the power of the $t$-test increases with the distance between $\gamma$ and $\gamma_0$; it also increases with the strength of IVs.

*Proof of Theorem 5.1.* Recall that from (1.19) that

$$\hat{\sigma}_u^2 \to \sigma_u^2, \tag{5.3}$$

and, when IVs are strong,

$$\frac{y_2' P_{M_2 Z_1} y_2}{n} \to_p \pi_1' Q_{1\cdot 2}\pi_1, \tag{5.4}$$

as shown in (4.13) and earlier in the proof of Theorem 1.2. Similarly to the proof of Theorem 1.2,

$$\begin{aligned}\hat{\gamma} - \gamma_0 &= \gamma - \gamma_0 + \frac{y_2' P_{M_2 Z_1} u}{y_2' P_{M_2 Z_1} y_2} \\ &= n^{-1/2}\delta + \frac{y_2' P_{M_2 Z_1} u}{y_2' P_{M_2 Z_1} y_2},\end{aligned}$$

so that

$$\begin{aligned} n^{1/2}(\hat{\gamma} - \gamma_0) &= \delta + \frac{\pi_1' Z_1' M_2 u/n^{1/2} + (v' M_2 Z_1/n)(Z_1' M_2 Z_1/n)^{-1} Z_1' M_2 u/n^{1/2}}{y_2' P_{M_2 Z_1} y_2} \\ &\to_d \delta + \frac{\pi_1' \Phi_{1\cdot 2}}{\pi_1' Q_{1\cdot 2}\pi_1} \\ &=^d N\left(\delta, \frac{\sigma_u^2}{\pi_1' Q_{1\cdot 2}\pi_1}\right), \end{aligned} \tag{5.5}$$

since $\Phi_{1\cdot 2} \sim N(0, \sigma_u^2 Q_{1\cdot 2})$. The definition of the $t$-statistic in (5.1) and the results in (5.3)-(5.5) imply that

$$t(\gamma_0) \to_d \frac{N\left(\delta, \frac{\sigma_u^2}{\pi_1' Q_{1\cdot 2}\pi_1}\right)}{\sqrt{\frac{\sigma_u^2}{\pi_1' Q_{1\cdot 2}\pi_1}}}$$

$$=^d N\left(\delta\left(\frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma_u^2}\right)^{1/2}, 1\right). \tag{5.6}$$

Next, note that

$$|t(\gamma_0)| > z_{1-\alpha/2}$$

holds if and only if

$$\begin{aligned}(t(\gamma_0))^2 &> z_{1-\alpha/2}^2 \\ &= \chi_{1,1-\alpha}^2,\end{aligned}$$

where the result in the second line holds by the definition of the $\chi^2$ distribution in equation (A.1), see Appendix A.5. Hence,

$$\begin{aligned}P\left(|t(\gamma_0)| > z_{1-\alpha/2}\right) &= P\left((t(\gamma_0))^2 > \chi_{1,1-\alpha}^2\right) \\ &\to P\left(\left(\delta\left(\frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma_u^2}\right)^{1/2} + \mathcal{Z}\right)^2 > \chi_{1,1-\alpha}^2\right),\end{aligned}$$

where the result in the last line holds by (5.6) and the definition of the non-central $\chi_1^2$ distribution in (A.2). □

### 5.2 Power of the AR test under strong IVs

In this section, we derive the power of the AR weak-IV-robust test from Section 4 when IVs are strong. Note that Theorem 4.2 in Section 4.1 analyzes the power of the AR test when IVs are weak ($\pi_1 = n^{-1/2}C$ and $\|\lambda\|^2 < \infty$), while the distance between the truth and the null hypothesis $\gamma - \gamma_0$ is fixed. Here, we proceed under the assumption that $\pi_1$ is fixed while $\gamma = \gamma_0 + n^{-1/2}\delta$, i.e. the distance between the truth and the null hypothesis is small in the local-to-zero sense.

**Theorem 5.2.** *Suppose that Assumption 1.1 holds, $\gamma = \gamma_0 + n^{-1/2}\delta$, and $\pi_1$ is fixed, i.e IVs are strong. Then,*

$$AR(\gamma_0) \to_d \chi_{l_1}^2\left(\delta^2 \frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma_u^2}\right).$$

*Proof.* From (4.6),

$$AR(\gamma_0) = \frac{\left((y_1 - y_2\gamma_0)' M_2 Z_1/n^{1/2}\right)\left(Z_1' M_2 Z_1/n\right)^{-1} Z_1' M_2 (y_1 - y_2\gamma_0)/n^{1/2}}{(y_1 - y_2\gamma_0)' M (y_1 - y_2\gamma_0)/n}.$$

We have:

$$
\begin{aligned}
\left(\frac{Z_1'M_2Z_1}{n}\right)^{-1} &\to_p Q_{1\cdot2}^{-1}, \\
\frac{Z_1'M_2(y_1-y_2\gamma_0)}{n^{1/2}} &= \frac{Z_1'M_2(y_2(\gamma-\gamma_0)+u)}{n^{1/2}} \\
&= \frac{Z_1'M_2\left((Z_1\pi_1+v)\delta/n^{1/2}+u\right)}{n^{1/2}} \\
&= \frac{Z_1'M_2Z_1}{n}\pi_1\delta + \frac{Z_1'M_2u}{n^{1/2}} + \frac{Z_1'M_2v}{n}\delta \\
&\to_d Q_{1\cdot2}\pi_1\delta + \Phi_{1\cdot2} \\
&= \sigma_u Q_{1\cdot2}^{1/2}\left(Q_{1\cdot2}^{1/2}\pi_1\delta/\sigma_u + \mathcal{Z}_u\right),
\end{aligned}
$$

where recall that $\mathcal{Z}_u \sim N(0, I_{l_1})$ as defined in (2.4). In the denominator of the AR statistic, we have a null-restricted estimator of $\sigma_u^2$:

$$
\begin{aligned}
\hat{\sigma}_u^2(\gamma_0) &= \frac{(y_1-y_2\gamma_0)'M(y_1-y_2\gamma_0)}{n} \\
&= \frac{\left(v\delta/n^{1/2}+u\right)'M_2\left(v\delta/n^{1/2}+u\right)}{n} \\
&= \frac{u'Mu}{n} + o_p(1) \\
&\to_p \sigma_u^2,
\end{aligned} \tag{5.7}
$$

where the $o_p(1)$ term is due to $\delta/n^{1/2}$. We now have:

$$
\begin{aligned}
AR(\gamma_0) &\to_d \left\| Q_{1\cdot2}^{1/2}\pi_1\delta/\sigma_u + \mathcal{Z}_u \right\|^2 \\
&\sim \chi^2_{l_1}\left(\delta^2\frac{\pi_1'Q_{1\cdot2}\pi_1}{\sigma_u^2}\right).
\end{aligned}
$$

□

By comparing Theorems 5.1 and 5.2, one can see that when IVs are strong, the power of the $t$- and AR tests depends on the same non-centrality parameter:

$$
\begin{aligned}
P\left(|t(\gamma_0)| > z_{1-\alpha/2}\right) &\to P\left(\chi^2_1\left(\frac{\pi_1'Q_{1\cdot2}\pi_1}{\sigma_u^2}\right) > \chi^2_{1,1-\alpha}\right), \\
P\left(AR(\gamma_0) > \chi^2_{l_1,1-\alpha}\right) &\to P\left(\chi^2_{l_1}\left(\frac{\pi_1'Q_{1\cdot2}\pi_1}{\sigma_u^2}\right) > \chi^2_{l_1,1-\alpha}\right).
\end{aligned}
$$

The only difference between the two expressions for the power is the number of degrees of freedom: one in the case of the $t$-test and $l_1$ in the case of the AR test. Thus, the power of the two tests under strong IVs is identical only when the number of IVs $l_1 = 1$ (the exactly identified model). When the model is over-identified (i.e. $l_1 > 1$), recall that the non-central $\chi^2$ distribution with $l_1$ degrees of freedom can be written as

$$\chi^2_{l_1}\left(\frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma_u^2}\right) =^d \left(\left(\frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma_u^2}\right)^{1/2} + \mathcal{Z}_1\right)^2 + \mathcal{Z}_2^2 + \ldots + \mathcal{Z}_{l_1}^2,$$

where $\mathcal{Z}_1, \mathcal{Z}_2, \ldots \mathcal{Z}_{l_1}$ are iid $N(0,1)$ random variables. Thus, in the case of the AR test we have extra noise components $\mathcal{Z}_2, \ldots \mathcal{Z}_{l_1}$, which appear because the AR test tests $l_1$ restrictions, while the $t$-test tests only one restriction. At the same time, the signal component: $\pi_1' Q_{1\cdot 2}\pi_1/\sigma_u^2$ is the same for both tests. As a result, the AR test is not going to be as powerful as the $t$-test (under strong IVs) if $l_1 > 1$, i.e. the model is over-identified.

### 5.3 LM test

We saw in the previous section that, when IVs are strong and the model is over-identified ($l_1 > 1$), the AR test is not going to be as powerful as the $t$-test. Although the AR test still has the advantage of remaining valid when IVs are weak, it is important to see if it is possible to improve on the power of the AR test when IVs are strong while preserving the validity when IVs are weak.

The loss of power of the AR approach occurs because it tests hypotheses about $\gamma \in \mathbb{R}$ by testing $l_1$ restrictions: each of the $l_1$ IVs $Z_1$ is uncorrelated with the error. Thus, we can try to improve the power by transforming $l_1 > 1$ restrictions into a single restriction by taking linear combinations. The transformation will be efficient if the power of the resulting test is determined by the same non-centrality parameter as that of the $t$-test under strong IVs:

$$\delta^2 \frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma_u^2}.$$

Recall that the AR test is driven by $l_1$ sample covariances between $Z_1$ and the null-restricted residuals:

$$Z_1' M_2(y_1 - y_2\gamma_0) \in \mathbb{R}^{l_1}.$$

Consider the following (infeasible) linear transformation:

$$\pi_1' Z_1' M_2(y_1 - y_2\gamma_0) \in \mathbb{R}.$$

When IVs are strong and $\gamma = \gamma_0 + \delta/\sqrt{n}$,

$$\frac{\pi_1' Z_1' M_2(y_1 - y_2\gamma_0)}{n^{1/2}} = \frac{\pi_1' Z_1' M_2(u + y_2(\gamma - \gamma_0))}{n^{1/2}}$$

$$
\begin{aligned}
&= \frac{\pi_1' Z_1' M_2 (u + n^{-1/2}\delta(Z_1\pi_1 + v))}{n^{1/2}} \\
&= \delta \frac{\pi_1' Z_1' M_2 Z_1 \pi_1}{n} + \frac{\pi_1' Z_1' M_2 u}{n^{1/2}} + \delta \frac{\pi_1' Z_1' M_2 v}{n} \\
&\to_d \delta \pi_1' Q_{1\cdot 2} \pi_1 + \pi_1' \Phi_{1\cdot 2} \\
&=^d \delta \pi_1' Q_{1\cdot 2} \pi_1 + \sigma_u \pi_1' Q_{1\cdot 2}^{1/2} \mathcal{Z}_u \\
&\sim N\left(\delta \pi_1' Q_{1\cdot 2}\pi_1, \sigma_u^2 \pi_1' Q_{1\cdot 2} \pi_1\right). \qquad (5.8)
\end{aligned}
$$

Hence, the following (infeasible) statistic will have a non-central $\chi^2$ distribution with *one degree of freedom* and the desired non-centrality parameter:

$$
\begin{aligned}
&\frac{\left(\pi_1' Z_1' M_2 (y_1 - y_2\gamma_0)\right)^2}{\hat{\sigma}_u^2(\gamma_0) \pi_1' Z_1' M_2 Z_1 \pi_1} \qquad (5.9) \\
&\quad = \frac{\left(\pi_1' Z_1' M_2 (y_1 - y_2\gamma_0)/n^{1/2}\right)^2}{\hat{\sigma}_u^2(\gamma_0) \pi_1' Z_1' M_2 Z_1 \pi_1 / n} \\
&\quad \to_d \frac{\left(N\left(\delta \pi_1' Q_{1\cdot 2}\pi_1, \sigma_u^2 \pi_1' Q_{1\cdot 2}\pi_1\right)\right)^2}{\sigma_u^2 \pi_1' Q_{1\cdot 2} \pi_1} \\
&\quad = \left(N\left(\sqrt{\delta^2 \frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma_u^2}}, 1\right)\right)^2 \\
&\quad \sim \chi_1^2\left(\delta^2 \frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma_u^2}\right),
\end{aligned}
$$

where the null-restricted estimator $\hat{\sigma}_u^2(\gamma_0) = (y_1 - y_2\gamma_0)' M (y_1 - y_2\gamma_0)/n$ remains consistent for $\sigma_u^2$ when $\gamma = \gamma_0 + \delta/n^{1/2}$ as we saw in (5.7).

The statistic described above is infeasible because $\pi_1$ is unknown. One could try to replace it with the estimator

$$
\hat{\pi}_1 = \left(Z_1' M_2 Z_1\right)^{-1} Z_1' M_2 y_2.
$$

When IVs are strong, $\hat{\pi}_1 \to_p \pi_1$, which would produce the desired result. However, when IVs are weak and as is shown in the proof of Theorem 3.1,

$$
\begin{aligned}
\hat{\pi}_1' Z_1' M_2 Z_1 \hat{\pi}_1 &= y_2' M_2 Z_1 \left(Z_1' M_2 Z_1\right)^{-1} Z_1' M_2 y_2 \\
&= \left(\frac{Z_1' M_2 Z_1}{n} C + \frac{Z_1' M_2 v}{n^{1/2}}\right)' \left(\frac{Z_1' M_2 Z_1}{n}\right)^{-1} \left(\frac{Z_1' M_2 Z_1}{n} C + \frac{Z_1' M_2 v}{n^{1/2}}\right) \\
&\to_d \sigma_v^2 \|\lambda + \mathcal{Z}_v\|^2.
\end{aligned}
$$

Moreover, when IVs are weak and under the null $\gamma = \gamma_0$

$$
\hat{\pi}_1' Z_1' M_2 (y_1 - y_2\gamma_0) = y_2' M_2 Z_1 \left(Z_1' M_2 Z_1\right)^{-1} Z_1' M_2 u
$$

$$= \left(\frac{Z_1' M_2 Z_1}{n} C + \frac{Z_1' M_2 v}{n^{1/2}}\right)' \left(\frac{Z_1' M_2 Z_1}{n}\right)^{-1} \frac{Z_1' M_2 u}{n^{1/2}}$$
$$\to_d \sigma_u \sigma_v \left(\lambda C + \mathcal{Z}_v\right)' \mathcal{Z}_u.$$

Hence, the null distribution of the statistic in (5.9) with $\pi_1$ replaced by by $\hat{\pi}_1$ would be

$$\frac{\left(\left(\lambda C + \mathcal{Z}_v\right)' \mathcal{Z}_u\right)^2}{\|\lambda C + \mathcal{Z}_v\|^2}.$$

This distribution is nonstandard when $\mathcal{Z}_u$ and $\mathcal{Z}_v$ are correlated, which happens when the regressor $y_2$ is endogenous and $\rho_{uv} \neq 0$ (see the definition of $\mathcal{Z}_u$ and $\mathcal{Z}_v$ in (2.4)). However, if $\rho_{uv} = 0$, then $\mathcal{Z}_u$ and $\mathcal{Z}_v$ are independent by the properties of joint normal distributions, see Section A.2 in the Appendix. In that case,

$$\frac{\left(\lambda C + \mathcal{Z}_v\right)' \mathcal{Z}_u}{\|\lambda C + \mathcal{Z}_v\|} \mid \mathcal{Z}_v \sim N\left(0, \frac{\|\lambda C + \mathcal{Z}_v\|^2}{\|\lambda C + \mathcal{Z}_v\|^2}\right) = N(0,1).$$

Thus, when $(\lambda C + \mathcal{Z}_v)$ and $\mathcal{Z}_u$ are uncorrelated,

$$\frac{\left(\left(\lambda C + \mathcal{Z}_v\right)' \mathcal{Z}_u\right)^2}{\|\lambda C + \mathcal{Z}_v\|^2} \mid \mathcal{Z}_v \sim \chi_1^2. \tag{5.10}$$

Since the conditional distribution in (5.10) does not depend on $\mathcal{Z}_v$, it is the same for all realizations of $\mathcal{Z}_v$, i.e. it is the same as the unconditional distribution: $\chi_1^2$.

The result in (5.10) suggests the following approach for obtaining a test statistic that:

1. has a $\chi_1^2$ distribution under $H_0$ whether IVs are strong or weak.

2. has a non-central $\chi_1^2$ distribution with the non-centrality parameter $\delta^2 \pi_1' Q_{1\cdot 2} \pi_1 / \sigma_u^2$ when IVs are strong and $\gamma = \gamma_0 + n^{-1/2}\delta$.

To obtain such a statistic, one needs to replace $\hat{\pi}_1$ with an alternative estimator that:

1. converges in probability to $\pi_1$ when IVs are strong (so that we have the right non-centrality parameter under the alternative.

2. is asymptotically uncorrelated/independent with $\mathcal{Z}_u$ when IVs are weak and $\gamma = \gamma_0$ (so that we have $\chi_1^2$ distribution under the null).

The approach was suggested and developed in Kleibergen (2002)and Moreira (2001). Let $S(\gamma_0)$ denote the statistic that measures the violation of the null hypothesis:

$$S(\gamma_0) = \frac{Z_1' M_2 (y_1 - y_2 \gamma_0)}{n}.$$

We need to construct an estimator of $\pi_1$ that would be uncorrelated with $S(\gamma_0)$. To construct such an estimator, consider "regressing" $\hat{\pi}_1$ and keeping the residuals from that "regression":

$$\tilde{\pi}_1(\gamma_0) = \hat{\pi}_1 - \widehat{AsyCov}(\hat{\pi}_1, S(\gamma_0)) \left(\widehat{AsyVar}\left(S(\gamma_0)\right)\right)^{-1} S(\gamma_0),$$

where $\widehat{AsyCov}(\hat{\pi}_1, S(\gamma_0))$ denotes an estimator of the asymptotic covariance between $\hat{\pi}_1$ and $S(\gamma_0)$, and $\widehat{AsyVar}\left(S(\gamma_0)\right)$ denotes and estimator of the asymptotic variance of $S(\gamma_0)$.

Recall that

$$n^{1/2}S(\gamma) = \frac{Z_1'M_2(y_1 - y_2\gamma)}{n^{1/2}} = \frac{Z_1'M_2u}{n^{1/2}},$$
$$n^{1/2}(\hat{\pi}_1 - \pi_1) = \left(\frac{Z_1'M_2Z_1}{n}\right)^{-1}\frac{Z_1'M_2v}{n^{1/2}}.$$

Hence, the asymptotic variance of $S(\gamma_0)$ and the asymptotic covariance between $\hat{\pi}_1$ and $S(\gamma_0)$ are given by

$$AsyVar\left(S(\gamma_0)\right. = \sigma_u^2 Q_{1\cdot 2},$$
$$AsyCov(\hat{\pi}_1, S(\gamma_0)) = Q_{1\cdot 2}^{-1}\sigma_{uv}Q_{1\cdot 2} = \sigma_{uv}I_{l_1}.$$

We therefore can define the following null-restricted estimator of $\pi_1$:

$$\tilde{\pi}_1(\gamma_0) = \hat{\pi}_1 - \frac{\hat{\sigma}_{uv}(\gamma_0)}{\hat{\sigma}_u^2(\gamma_0)}\left(\frac{Z_1'M_2Z_1}{n}\right)^{-1} S(\gamma_0), \tag{5.11}$$

where $\hat{\sigma}_{uv}(\gamma_0)$ and $\hat{\sigma}_u^2(\gamma_0)$ denote the null-restricted estimators of $\sigma_{uv}$ and $\sigma_u^2$ respectively:

$$\hat{\sigma}_{uv}(\gamma_0) = \frac{y_2'M(y_1 - y_2\gamma_0)}{n}, \tag{5.12}$$
$$\hat{\sigma}_u^2(\gamma_0) = \frac{(y_1 - y_2\gamma_0)'M(y_1 - y_2\gamma_0)}{n}. \tag{5.13}$$

With those definitions, Kleibergen/Moreira's statistic (referred to as the LM statistic[9]) can be written as

$$LM(\gamma_0) = \frac{\left(\tilde{\pi}_1(\gamma_0)'Z_1'M_2(y_1 - y_2\gamma_0)\right)^2}{\hat{\sigma}_u^2(\gamma_0)\tilde{\pi}_1(\gamma_0)'Z_1'M_2Z_1\tilde{\pi}_1(\gamma_0)}, \tag{5.14}$$
$$= \frac{\left(n\tilde{\pi}_1(\gamma_0)'S(\gamma_0)\right)^2}{\hat{\sigma}_u^2(\gamma_0)\left(\tilde{\pi}_1(\gamma_0)'Z_1'M_2Z_1\tilde{\pi}_1(\gamma_0)\right)}, \tag{5.15}$$

where the expression in the first line matches the infeasible efficient statistic in equation (5.9), but with unknown $\pi_1$ replaced with the estimator $\tilde{\pi}_1(\gamma_0)$. The lemma below gives an

[9] LM stands for Lagrange Multiplier.

alternative and more compact expression for the LM statistic.

**Lemma 5.3.** *The LM statistic in (5.14) can be also written as*

$$LM(\gamma_0) = \frac{(y_1 - y_2\gamma_0)' P_{\tilde{y}_2(\gamma_0)} (y_1 - y_2\gamma_0)}{(y_1 - y_2\gamma_0)' M (y_1 - y_2\gamma_0)/n},$$

*where*

$$\begin{aligned}
\tilde{y}_2(\gamma_0) &= M_2 Z_1 \tilde{\pi}_1(\gamma_0), \text{ and} \\
P_{\tilde{y}_2(\gamma_0)} &= \tilde{y}_2(\gamma_0) \left(\tilde{y}_2(\gamma_0)'\tilde{y}_2(\gamma_0)\right)^{-1} \tilde{y}_2(\gamma_0)'.
\end{aligned}$$

*Proof.* Using the definition of $\tilde{y}_2$,

$$\begin{aligned}
\frac{\left(\tilde{\pi}_1(\gamma_0)' Z_1' M_2 (y_1 - y_2\gamma_0)\right)^2}{\tilde{\pi}_1(\gamma_0)' Z_1' M_2 Z_1 \tilde{\pi}_1(\gamma_0)} &= \frac{\left(\tilde{y}_2(\gamma_0)'(y_1 - y_2\gamma_0)\right)^2}{\tilde{y}_2(\gamma_0)'\tilde{y}_2(\gamma_0)} \\
&= \frac{(y_1 - y_2\gamma_0)\tilde{y}_2(\gamma_0)\tilde{y}_2(\gamma_0)'(y_1 - y_2\gamma_0)}{\tilde{y}_2(\gamma_0)'\tilde{y}_2(\gamma_0)} \\
&= (y_1 - y_2\gamma_0)\tilde{y}_2(\gamma_0) \left(\tilde{y}_2(\gamma_0)'\tilde{y}_2(\gamma_0)\right)^{-1} \tilde{y}_2(\gamma_0)'(y_1 - y_2\gamma_0) \\
&= (y_1 - y_2\gamma_0)' P_{\tilde{y}_2(\gamma_0)} (y_1 - y_2\gamma_0),
\end{aligned}$$

and the result follows from the definition of the LM statistic in (5.14). □

The next theorem shows that the LM statistic has the same $\chi^2_1$ asymptotic null distribution whether IVs are strong of weak, i.e. the LM statistic is robust to weak IVs. Therefore, the size $\alpha$ weak-IV-robust LM test of $H_0 : \gamma = \gamma_0$ against $H_1 : \gamma \neq \gamma_0$ is

$$\text{Reject } H_0 \text{ when } K(\gamma_0) > \chi^2_{1-\alpha}.$$

**Theorem 5.4.** *Suppose that Assumption 1.1 holds.*

*(a) Suppose that $\gamma = \gamma_0 + n^{-1/2}\delta$, and $\pi_1$ is fixed (i.e. IVs are strong). Then,*

$$LM(\gamma_0) \to_d \chi^2_1\left(\delta^2 \frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma^2_u}\right).$$

*In particular when $\gamma = \gamma_0$ (i.e. $\delta = 0$), $LM(\gamma_0) \to_d \chi^2_1$.*

*(b) $LM(\gamma_0) \to_d \chi^2_1$ when $\pi_1 = n^{-1/2}C$ (i.e. IVs weak as in Assumption 2.1 and $\|\lambda\|^2 < \infty$).*

*Proof.* For part (a), first note that when IVs are strong and $\pi_1$ is fixed,

$$\hat{\pi}_1 \to_p \pi_1.$$

Furthermore,

$$\begin{aligned} S(\gamma_0) &= \frac{Z_1' M_2(u + y_2(\gamma - \gamma_0))}{n} \\ &= \frac{Z_1' M_2(u + n^{-1/2}\delta(Z_1\pi_1 + v))}{n} \\ &= \frac{Z_1' M_2 u}{n} + o_p(1) \\ &\to_p 0. \end{aligned}$$

Hence, by the definition of $\tilde{\pi}_1(\gamma_0)$ in (5.11),

$$\tilde{\pi}_1(\gamma_0) \to_p \pi_1.$$

The rest of the proof for part (a) follows the same steps as those for the infeasible statistic in (5.9):

$$\begin{aligned} LM(\gamma_0) &= \frac{\left(\tilde{\pi}_1(\gamma_0)' Z_1' M_2(y_1 - y_2\gamma_0)\right)^2}{\hat{\sigma}_u^2(\gamma_0)\tilde{\pi}_1(\gamma_0)' Z_1' M_2 Z_1 \tilde{\pi}_1(\gamma_0)} \\ &= \frac{\left(\tilde{\pi}_1(\gamma_0)' Z_1' M_2(u + n^{-1/2}\delta(Z_1\pi_1 + v))/\sqrt{n}\right)^2}{\hat{\sigma}_u^2(\gamma_0)\tilde{\pi}_1(\gamma_0)' Z_1' M_2 Z_1 \tilde{\pi}_1(\gamma_0)/n} \\ &= \frac{\left(\tilde{\pi}_1(\gamma_0)' Z_1' M_2 u/\sqrt{n} + \delta\tilde{\pi}_1(\gamma_0)' Z_1' M_2 Z_1 \pi_1/n + o_p(1)\right)^2}{\hat{\sigma}_u^2(\gamma_0)\tilde{\pi}_1(\gamma_0)' Z_1' M_2 Z_1 \tilde{\pi}_1(\gamma_0)/n} \\ &\to_d \frac{\left(\pi_1'\Phi_{1\cdot 2} + \delta\pi_1' Q_{1\cdot 2}\pi_1\right)^2}{\sigma_u^2 \pi_1' Q_{1\cdot 2}\pi_1} \\ &\sim \frac{\left(N(\delta\pi_1' Q_{1\cdot 2}\pi_1, \pi_1' Q_{1\cdot 2}\pi_1)\right)^2}{\sigma_u^2 \pi_1' Q_{1\cdot 2}\pi_1} \\ &=^d \left(N\left(\delta\sqrt{\frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma_u^2}}, 1\right)\right)^2 \\ &=^d \chi_1^2\left(\delta^2 \frac{\pi_1' Q_{1\cdot 2}\pi_1}{\sigma_u^2}\right). \end{aligned}$$

For part (b), suppose that IVs are weak and $\gamma = \gamma_0$, so that

$$M_2(y_1 - y_2\gamma_0) = M_2 u.$$

In that case,

$$\begin{aligned} n^{1/2} S(\gamma_0) &= \frac{Z_1' M_2 u}{n^{1/2}} \to_d \Phi_{1\cdot 2} \\ \hat{\sigma}_u^2(\gamma_0) &= \frac{u' M u}{n} \to_p \sigma_u^2, \end{aligned} \tag{5.16}$$

$$\hat{\sigma}_{uv}(\gamma_0) = \frac{v'Mu}{n} \to_p \sigma_{uv}.$$

Moreover,

$$\begin{aligned} n^{1/2}\hat{\pi}_1 &= n^{1/2}\pi_1 + \left(\frac{Z_1'M_2Z_1}{n}\right)^{-1} \frac{Z_1'M_2v}{n^{1/2}} \\ &= C + \left(\frac{Z_1'M_2Z_1}{n}\right)^{-1} \frac{Z_1'M_2v}{n^{1/2}} \\ &\to_d C + Q_{1\cdot 2}^{-1}\Psi_{1\cdot 2}. \end{aligned}$$

Therefore,

$$n^{1/2}\tilde{\pi}_1(\gamma_0) \to_d \left(C + Q_{1\cdot 2}^{-1}\Psi_{1\cdot 2}\right) - \frac{\sigma_{uv}}{\sigma_u^2} Q_{1\cdot 2}^{-1}\Phi_{1\cdot 2}. \tag{5.17}$$

Denote the expression for the limiting distribution of $n^{1/2}\tilde{\pi}(\gamma_0)$ as $\Upsilon$:

$$\Upsilon = \left(C + Q_{1\cdot 2}^{-1}\Psi_{1\cdot 2}\right) - \frac{\sigma_{uv}}{\sigma_u^2} Q_{1\cdot 2}^{-1}\Phi_{1\cdot 2}, \tag{5.18}$$

and note that $\Upsilon$ is normally distributed. We have

$$\left(n^{1/2}\tilde{\pi}_1(\gamma_0), n^{1/2}S(\gamma_0)\right) \to_d \left(\Upsilon, \Phi_{1\cdot 2}\right). \tag{5.19}$$

Next, the asymptotic covariance of $n^{1/2}\tilde{\pi}_1(\gamma_0)$ and $n^{1/2}S(\gamma_0)$ is zero by construction of $\tilde{\pi}_1(\gamma_0)$:

$$\begin{aligned} Cov\left(\Upsilon, \Phi_{1\cdot 2}\right) &= Q_{1\cdot 2}^{-1} Cov(\Psi_{1\cdot 2}, \Phi_{1\cdot 2}) - \frac{\sigma_{uv}}{\sigma_u^2} Q_{1\cdot 2}^{-1} Cov(\Phi_{1\cdot 2}, \Phi_{1\cdot 2}) \\ &= Q_{1\cdot 2}^{-1}\left(\sigma_{uv}Q_{1\cdot 2}\right) - \frac{\sigma_{uv}}{\sigma_u^2} Q_{1\cdot 2}^{-1}\left(\sigma_u^2 Q_{1\cdot 2}\right) \\ &= \sigma_{uv} I_{l_1} - \sigma_{uv} I_{l_1} \\ &= 0, \end{aligned}$$

where the equality in the second line follows from the definition of $\Phi_{1\cdot 2}$ and $\Psi_{1\cdot 2}$ in (2.7). Hence $\Upsilon$ and $\Phi_{1\cdot 2}$ are uncorrelated. However, since $\Upsilon$ and $\Phi_{1\cdot 2}$ are jointly normal, they are also independent.

The results in (5.16) and (5.19) as well as the representation of the LM statistic in (5.15) imply that

$$LM(\gamma_0) = \frac{\left(n^{1/2}\tilde{\pi}_1(\gamma_0)' n^{1/2} S(\gamma_0)\right)^2}{\hat{\sigma}_u^2(\gamma_0)\left(n^{1/2}\tilde{\pi}_1(\gamma_0)'\left(Z_1'M_2Z_1/n\right) n^{1/2}\tilde{\pi}_1(\gamma_0)\right)}, \tag{5.20}$$

$$\to_d \frac{\left(\Upsilon'\Phi_{1\cdot 2}\right)^2}{\sigma_u^2\left(\Upsilon' Q_{1\cdot 2}\Upsilon\right)}.$$

Since $\Upsilon$ and $\Phi_{1\cdot 2}$ are independent,

$$\Upsilon'\Phi_{1\cdot 2} \mid \Upsilon \sim N\left(0, \sigma_u^2 \Upsilon' Q_{1\cdot 2}\Upsilon\right).$$

Therefore,

$$\begin{aligned}
\frac{\left(\Upsilon'\Phi_{1\cdot 2}\right)^2}{\sigma_u^2\left(\Upsilon' Q_{1\cdot 2}\Upsilon\right)} \mid \Upsilon &\sim \frac{\left(N\left(0, \sigma_u^2 \Upsilon' Q_{1\cdot 2}\Upsilon\right)\right)^2}{\sigma_u^2 \Upsilon' Q_{1\cdot 2}\Upsilon} \\
&=^d (N(0,1))^2 \\
&\sim \chi_1^2.
\end{aligned}$$

Since the conditional distribution given $\Upsilon$ does not depend on $\Upsilon$, it is the same for all values of $\Upsilon$, i.e.

$$\frac{\left(\Upsilon'\Phi_{1\cdot 2}\right)^2}{\sigma_u^2\left(\Upsilon' Q_{1\cdot 2}\Upsilon\right)} \sim \chi_1^2.$$

□

## 5.4 Robust CSs based on the LM test

Similarly to AR CSs, one can construct robust CSs for $\gamma$ by inverting the LM test:

$$CS_{1-\alpha}^{LM} = \{\gamma_0 : LM(\gamma_0) \le \chi_{1,1-\alpha}^2\}. \tag{5.21}$$

Since the LM test is more powerful than the AR test when IVs are strong and the model is over-identified ($l_1 > 1$), LM-based CSs are expected to be more precise than AR-based CSs in such situations. Moreover unlike AR CSs, LM-based CSs cannot be empty even when $l_1 > 1$.

**Theorem 5.5.** *$CS_{1-\alpha}^{LM}$ defined in (5.21) cannot be empty.*

*Proof.* The result will be established by showing that (i) the LM statistic is proportional to the derivative of $AR(\gamma_0)$ with respect to $\gamma_0$, and that (ii) $AR(\gamma_0)$ has a minimum as a function of $\gamma_0$. This implies that the value $\underline{\gamma}$ that minimizes $AR(\cdot)$ also satisfies $LM(\underline{\gamma}) = 0$ , which in turn implies that $\underline{\gamma} \in CS_{1-\alpha}^{LM}$.

We first show (i).

$$\begin{aligned}
\frac{1}{2}\frac{dAR(\gamma_0)}{d\gamma_0} &= \frac{1}{2}\frac{d}{d\gamma_0}\left(\frac{(y_1 - y_2\gamma_0)' P_{M_2 Z_1}(y_1 - y_2\gamma_0)}{(y_1 - y_2\gamma_0)' M (y_1 - y_2\gamma_0)/n}\right) \\
&= -\frac{y_2' P_{M_2 Z_1}(y_1 - y_2\gamma_0)}{(y_1 - y_2\gamma_0)' M (y_1 - y_2\gamma_0)/n} + \frac{y_2' M (y_1 - y_2\gamma_0)/n\left((y_1 - y_2\gamma_0)' P_{M_2 Z_1}(y_1 - y_2\gamma_0)\right)}{\left((y_1 - y_2\gamma_0)' M (y_1 - y_2\gamma_0)/n\right)^2}
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{(y_1 - y_2\gamma_0)'M(y_1 - y_2\gamma_0)/n} \\
&\quad \times \left( y_2' P_{M_2 Z_1}(y_1 - y_2\gamma_0) - y_2' M(y_1 - y_2\gamma_0) \frac{(y_1 - y_2\gamma_0)' P_{M_2 Z_1}(y_1 - y_2\gamma_0)}{(y_1 - y_2\gamma_0)'M(y_1 - y_2\gamma_0)} \right) \\
&= -\frac{y_2' P_{M_2 Z_1}(y_1 - y_2\gamma_0)}{\hat{\sigma}_u^2} \left( 1 - \frac{y_2' M(y_1 - y_2\gamma_0)}{y_2' P_{M_2 Z_1}(y_1 - y_2\gamma_0)} AR(\gamma_0) \right). \qquad (5.22)
\end{aligned}
$$

At the same time, in the numerator of the LM statistic we have:

$$
\begin{aligned}
&(y_1 - y_2\gamma_0)' M_2 Z_1 \tilde{\pi}_1(\gamma_0) = \\
&\quad = (y_1 - y_2\gamma_0)' M_2 Z_1 \left( \hat{\pi}_1 - \frac{\hat{\sigma}_{uv}(\gamma_0)}{\hat{\sigma}_u^2(\gamma_0)} \left( \frac{Z_1' M_2 Z_1}{n} \right)^{-1} \frac{Z_1' M_2 (y_1 - y_2\gamma_0)}{n} \right) \\
&\quad = (y_1 - y_2\gamma_0)' M_2 Z_1 \\
&\qquad \times \left( (Z_1' M_2 Z_1)^{-1} Z_1' M_2 y_2 - \frac{y_2' M(y_1 - y_2\gamma_0)}{(y_1 - y_2\gamma_0)'M(y_1 - y_2\gamma_0)} (Z_1' M_2 Z_1)^{-1} Z_1' M_2 (y_1 - y_2\gamma_0) \right) \\
&\quad = (y_1 - y_2\gamma_0)' P_{M_2 Z_1} y_2 - y_2' M(y_1 - y_2\gamma_0) \frac{(y_1 - y_2\gamma_0)' P_{M_2 Z_1}(y_1 - y_2\gamma_0)}{(y_1 - y_2\gamma_0)'M(y_1 - y_2\gamma_0)}) \\
&\quad = y_2' P_{M_2 Z_1}(y_1 - y_2\gamma_0) \left( 1 - \frac{y_2' M(y_1 - y_2\gamma_0)}{y_2' P_{M_2 Z_1}(y_1 - y_2\gamma_0)} AR(\gamma_0) \right) \\
&\quad = -\frac{\hat{\sigma}_u^2(\gamma_0)}{2} \frac{dAR(\gamma_0)}{d\gamma_0}, \qquad (5.23)
\end{aligned}
$$

where the equality in the last line holds by (5.22).

Next, we show (ii). Minimization of $AR(\gamma_0)$ is related to the theory of the LIML estimator, see for example p. 549 in Davidson and MacKinnon (2004).

$$
\begin{aligned}
n^{-1} AR(\gamma_0) &= \frac{(y_1 - y_2\gamma_0)' P_{M_2 Z_1}(y_1 - y_2\gamma_0)}{(y_1 - y_2\gamma_0)'M(y_1 - y_2\gamma_0)} \\
&= \frac{a_0' Y' P_{M_2 Z_1} Y a_0}{a_0' Y' M Y a_0},
\end{aligned}
$$

where

$$
a_0 = \begin{pmatrix} 1 \\ -\gamma_0 \end{pmatrix} \quad \text{and} \quad Y = [y_1 \quad y_2],
$$

so that

$$
Y a_0 = y_1 - y_2\gamma_0.
$$

Next,

$$
\frac{1}{2}\frac{d}{da_0}\left( \frac{a_0' Y' P_{M_2 Z_1} Y a_0}{a_0' Y' M Y a_0} \right) = \frac{Y' P_{M_2 Z_1} Y a_0}{a_0' Y' M Y a_0} - Y' M Y a_0 \frac{a_0' Y' P_{M_2 Z_1} Y a_0}{(a_0' Y' M Y a_0)^2}
$$

$$= \frac{1}{a_0' Y' M Y a_0} \left( Y' P_{M_2 Z_1} Y a_0 - Y' M Y a_0 AR(\gamma_0) \right).$$

Setting this to zero and denoting by $\underline{\gamma}$ the minimizer of $AR(\gamma_0)$, $\underline{a} = (1 \quad \underline{\gamma})'$, and $\underline{\kappa} = AR(\underline{\gamma})$, we obtain:

$$Y' P_{M_2 Z_1} Y \underline{a} - \underline{\kappa} Y' M Y \underline{a} = 0,$$

or

$$\begin{aligned} 0 &= \left( Y' P_{M_2 Z_1} Y - \underline{\kappa} Y' M Y \right) \underline{a} \\ &= \left( Y' M Y \right)^{1/2} \left( \left( Y' M Y \right)^{-1/2} Y' P_{M_2 Z_1} Y \left( Y' M Y \right)^{-1/2} - \underline{\kappa} I_2 \right) \left( Y' M Y \right)^{1/2} \underline{a}. \end{aligned}$$

Now, defining

$$a^* = \left( Y' M Y \right)^{1/2} \underline{a},$$

we can re-write the first-order condition for minimization of $AR(\gamma_0)$ as

$$\left( \left( Y' M Y \right)^{-1/2} Y' P_{M_2 Z_1} Y \left( Y' M Y \right)^{-1/2} - \underline{\kappa} I_2 \right) a^* = 0.$$

The last equation means that

$$\underline{\kappa} = \min_{\gamma_0} AR(\gamma_0)$$

is the *smallest* eigenvalue of the $2 \times 2$ matrix $\left( Y' M Y \right)^{-1/2} Y' P_{M_2 Z_1} Y \left( Y' M Y \right)^{-1/2}$.

Using the derivative of $AR(\gamma_0)$ in (5.22), the first order condition for minimization of the AR statistic implies that

$$\begin{aligned} 0 &= 1 - \frac{y_2' M (y_1 - y_2 \underline{\gamma})}{y_2' P_{M_2 Z_1} (y_1 - y_2 \underline{\gamma})} AR(\underline{\gamma}) \\ &= 1 - \frac{y_2' M (y_1 - y_2 \underline{\gamma})}{y_2' P_{M_2 Z_1} (y_1 - y_2 \underline{\gamma})} \underline{\kappa}. \end{aligned}$$

Hence,

$$y_2' P_{M_2 Z_1} (y_1 - y_2 \underline{\gamma}) = y_2' M (y_1 - y_2 \underline{\gamma}) \underline{\kappa},$$

which can be re-written as

$$y_2' \left( P_{M_2 Z_1} - \underline{\kappa} M \right) y_1 = y_2' \left( P_{M_2 Z_1} - \underline{\kappa} M \right) y_2 \underline{\gamma},$$

or

$$\underline{\gamma} = \frac{y_2' \left( P_{M_2 Z_1} - \underline{\kappa} M \right) y_1}{y_2' \left( P_{M_2 Z_1} - \underline{\kappa} M \right) y_2}.$$

We conclude that $AR(\gamma_0)$ is minimized at $\underline{\gamma}$, therefore

$$\frac{dAR(\underline{\gamma})}{d\gamma_0} = 0.$$

Equation (5.23) now implies that

$$(y_1 - y_2\underline{\gamma})' M_2 Z_1 \tilde{\pi}_1(\underline{\gamma}) = \frac{\hat{\sigma}_u^2(\underline{\gamma})}{2} \frac{dAR(\underline{\gamma})}{d\gamma_0} = 0.$$

Hence

$$LM(\underline{\gamma}) = 0,$$

and therefore

$$\underline{\gamma} \in CS_{1-\alpha}^{LM}.$$

□

*Remark.*

1. The minimizer of the AR statistic $\underline{\gamma}$ appearing in the proof of Theorem 5.5 is known as the *limited information maximum likelihood* (LIML) estimator:

$$\hat{\gamma}^{LIML} = \underline{\gamma} = \frac{y_2' \left(P_{M_2 Z_1} - \underline{\kappa} M\right) y_1}{y_2' \left(P_{M_2 Z_1} - \underline{\kappa} M\right) y_2}.$$

   Thus, the LIML estimator is always included in CSs constructed by inverting the LM test.

2. The proof of Theorem 5.5 reveals a peculiar property of CSs based on the LM statistic. Define

$$\bar{\kappa} = \max_{\gamma_0} AR(\gamma_0),$$

   and let $\bar{\gamma}$ denote the maximizer of $AR(\gamma_0)$, i.e.

$$\bar{\gamma} = \arg\max_{\gamma_0} AR(\gamma_0),$$

   and

$$AR(\bar{\gamma}) = \bar{\kappa}.$$

   Since $\bar{\kappa}$ is the largest possible value of the AR statistic, $\bar{\gamma}$ is the value that the AR test would reject as long as $CS_{1-\alpha}^{AR}$ does not contain the entire real line. At the same time,

$\bar{\gamma}$ must satisfy the first-order condition for maximization of $AR(\gamma_0)$, and therefore,

$$\frac{d}{d\gamma_0}AR(\bar{\gamma}) = 0.$$

By the same arguments as in the proof of Theorem 5.5,

$$LM(\bar{\gamma}) = 0$$

and therefore,

$$\bar{\gamma} \in CS^{LM}_{1-\alpha}.$$

This implies, that CSs based on inversion of the LM test are built around two values $\underline{\gamma}$ and $\bar{\gamma}$. Including the first in a CS makes a lot of sense: this is the value of $\gamma_0$ that makes the correlation between the null restricted errors $y_1 - y_2\gamma_0$ and IVs $Z_1$ as small as possible (given the data). On the other hand including the latter, $\bar{\gamma}$, appears to be wrong as this is the value that maximizes the correlation between the null-restricted errors and IVs.

3. Since $LM(\underline{\gamma}) = LM(\bar{\gamma}) = 0$, the LM test can suffer form undesirable loss of power around the values $\bar{\gamma}$. In practice, this issue can be a problem only when IVs are weak since, as we have seen, in the case of strong IVs and the LM test has the same optimal power as usual tests based on the IV estimator. Thus, when IVs are strong, the LM test is expected to outperform the AR test. Nevertheless, the AR test can be more powerful than the LM test when IVs are weak.

4. Computation of LM-test-based CSs is discussed in Mikusheva (2010). She shows that LM test CSs can be constructed by solving three quadratic equations, and as a result LM test CSs can take one of the three possible forms: i) $[\gamma_1, \gamma_2] \cup [\gamma_3, \gamma_4]$, ii) $(-\infty, \gamma_1] \cup [\gamma_2, \gamma_3] \cup [\gamma_4, \infty)$, and iii) $(-\infty, \infty)$. Cases ii) and iii) tend to occur when IVs are weak. Also, when IVs are strong, case i) typically becomes $[\gamma_1, \gamma_2]$.

5. The fast and accurate methods of Mikusheva (2010) for construction of robust CSs have been implemented in Stata (see Mikusheva and Poi (2006)). Stata command is `condivreg`; its option `ar` provides AR-based CSs, and option `lm` provides LM-based CSs.

### 5.5 CLR test and CSs

The LM test discussed in the previous section, delivers weak-IV-robust inference. Moreover when IVs are strong, it attains efficiency in the sense that it has the same power properties of the t-test under strong IVs. Nevertheless when IVs are weak, the LM test might suffer from power loss and be inferior to the AR test. These issues to a great extent can be alleviated by using the conditional likelihood ratio (CLR) test proposed by Moreira (2003) (see also Andrews et al., 2006).

The CLR statistic is derived following the Likelihood Ratio principle and its construction assumes that the errors $u_i$ and $v_i$ are jointly normal.[10] It turns out, however, that the CLR statistic combines the AR and LM statistics. When IVs are strong, it becomes equivalent to the LM statistic in large samples. When IVs are weak it utilizes the information contained in the AR statistic to deliver inference that is always more powerful than that of the LM test and typically more powerful than that of the AR test. However, there are certain scenarios where the AR approach can dominate the CLR approach. This occurs when the correlation between the structural error $u_i$ and the first-stage error $v_i$ is zero or very close to zero (see Andrews et al., 2016). While there is no optimal test when IVs are weak, the CLR test is the preferred test overall.

To introduce the CLR statistic, we need some additional definitions. Recall from (5.17) and (5.18) that when IVs are weak,

$$\begin{aligned} n^{1/2}\tilde{\pi}_1(\gamma_0) &\to_d C + Q_{1\cdot 2}^{-1}\Psi_{1\cdot 2} - \frac{\sigma_{uv}}{\sigma_u^2} Q_{1\cdot 2}^{-1}\Phi_{1\cdot 2} \\ &= C + Q_{1\cdot 2}^{-1/2}\left(\sigma_v \mathcal{Z}_v - \frac{\sigma_{uv}}{\sigma_u}\mathcal{Z}_u\right) \\ &= C + \sigma_v Q_{1\cdot 2}^{-1/2}\left(\mathcal{Z}_v - \rho_{uv}\mathcal{Z}_u\right), \end{aligned}$$

where $\rho_{uv} = \sigma_{uv}/(\sigma_u\sigma_v)$ is the correlation between $u_i$ and $v_i$. Let

$$\mathcal{Z}_{v\cdot u} = \frac{\mathcal{Z}_v - \rho_{uv}\mathcal{Z}_u}{\sqrt{1-\rho_{uv}^2}} \sim \frac{N\left(0, (1-\rho_{uv}^2)I_{l_1}\right)}{\sqrt{1-\rho_{uv}^2}} = N(0, I_{l_1}).$$

Thus,

$$n^{1/2}\tilde{\pi}_1(\gamma_0) \to_d C + \sigma_v\sqrt{1-\rho_{uv}^2}\, Q_{1\cdot 2}^{-1/2}\mathcal{Z}_{v\cdot v},$$

and the limiting variance of $\tilde{\pi}_1(\gamma_0)$ is

$$\sigma_v^2(1-\rho_{uv}^2)Q_{1\cdot 2}^{-1} = \frac{\sigma_v^2\sigma_u^2 - \sigma_{uv}^2}{\sigma_u^2} Q_{1\cdot 2}^{-1}.$$

Let $T^*(\gamma_0)$ denote the standardized version of the estimator $\tilde{\pi}_1(\gamma_0)$ constructed so that the limiting variance of $\tilde{\pi}_1(\gamma_0)$ is $I_{l_1}$:

$$n^{1/2}T^*(\gamma_0) = \sqrt{\frac{\hat{\sigma}_u^2(\gamma_0)}{\hat{\sigma}_v^2\hat{\sigma}_u^2(\gamma_0) - \hat{\sigma}_{uv}^2(\gamma_0)}}\left(\frac{Z_1' M_2 Z_1}{n}\right)^{1/2} n^{1/2}\tilde{\pi}_1(\gamma_0), \tag{5.24}$$

[10] CLR's large sample properties are unaffected by the distributional assumption.

where the estimators $\hat{\sigma}_v^2$, $\hat{\sigma}_{uv}(\gamma_0)$, and $\hat{\sigma}_u^2(\gamma_0)$ were previously defined in (3.4), (5.12), and (5.16). Note that when IVs are weak and $\gamma = \gamma_0$,

$$\begin{aligned} n^{1/2}T^*(\gamma) &\to_d \sqrt{\frac{\sigma_u^2}{\sigma_v^2\sigma_u^2 - \sigma_{uv}^2}} Q_{1\cdot 2}^{1/2} C + \mathcal{Z}_{v\cdot u} \equiv \mathcal{X}_T, \text{ where} \\ \mathcal{X}_T &\sim N\left(\sqrt{\frac{\sigma_v^2\sigma_u^2}{\sigma_v^2\sigma_u^2 - \sigma_{uv}^2}} \frac{Q_{1\cdot 2}^{1/2} C}{\sigma_v}, I_{l_1}\right) = N\left(\frac{\lambda}{\sqrt{1-\rho_{uv}^2}}, I_l\right), \end{aligned} \tag{5.25}$$

$\lambda = Q_{1\cdot 2}^{1/2} C/\sigma_v$, and $\|\lambda\|^2$ is the concentration parameter. Also, note that $\mathcal{Z}_{v\cdot u}$ and $\mathcal{Z}_u$ are independent by construction since

$$Cov(\mathcal{Z}_{v\cdot u}, \mathcal{Z}_u) = \frac{Cov(\mathcal{Z}_v, \mathcal{Z}_u) - \rho_{uv} Var(\mathcal{Z}_u)}{\sqrt{1-\rho_{uv}^2}} = \frac{\rho_{uv} I_{l_1} - \rho_{uv} I_{l_1}}{\sqrt{1-\rho_{uv}^2}} = 0. \tag{5.26}$$

Since

$$n^{1/2}S(\gamma) = \frac{Z_1' M_2 (y_1 - y_2\gamma)}{n^{1/2}} \to_d \sigma_u Q_{1\cdot 2}^{1/2} \mathcal{Z}_u,$$

$T^*(\gamma_0)$ and $S(\gamma_0)$ are asymptotically independent. Let $S^*(\gamma_0)$ denote the standardized version of $S(\gamma_0)$:

$$\begin{aligned} n^{1/2}S^*(\gamma_0) &= \frac{1}{\hat{\sigma}_u(\gamma_0)} \left(\frac{Z_1' M_2 Z_1}{n}\right)^{-1/2} n^{1/2} S(\gamma_0) \\ &\to_d \mathcal{Z}_u, \text{ when } \gamma_0 = \gamma. \end{aligned} \tag{5.27}$$

Lastly, note that

$$AR(\gamma_0) = nS(\gamma_0)' \left(\frac{Z_1' M_2 Z_1}{n}\right)^{-1} S(\gamma_0)/\hat{\sigma}_u^2(\gamma_0) = nS^*(\gamma_0)' S^*(\gamma_0).$$

Recall from (5.20) that

$$\begin{aligned} LM(\gamma_0) &= \frac{\left(n^{1/2}\tilde{\pi}_1(\gamma_0)' n^{1/2} S(\gamma_0)\right)^2}{\hat{\sigma}_u^2(\gamma_0)\left(n^{1/2}\tilde{\pi}_1(\gamma_0)' \left(Z_1' M_2 Z_1/n\right) n^{1/2}\tilde{\pi}_1(\gamma_0)\right)} \\ &\frac{\left(n^{1/2}T^*(\gamma_0)' \left(Z_1' M_2 Z_1/n\right)^{-1/2} n^{1/2} S(\gamma_0)\right)^2}{\hat{\sigma}_u^2(\gamma_0)\left(n^{1/2}T^*(\gamma_0)' n^{1/2} T^*(\gamma_0)\right)} \\ &= \frac{n\left(T^*(\gamma_0)' S^*(\gamma_0)\right)^2}{T^*(\gamma_0)' T^*(\gamma_0)}. \end{aligned}$$

The CLR statistic is defined as

$$CLR(\gamma_0) = 0.5\left(Q_{SS}(\gamma_0) - Q_{TT}(\gamma_0) + \sqrt{(Q_{SS}(\gamma_0) - Q_{TT}(\gamma_0))^2 + 4Q_{ST}^2(\gamma_0)}\right), \tag{5.28}$$

where

$$Q_{SS}(\gamma_0) = nS^*(\gamma_0)'S^*(\gamma_0), \tag{5.29}$$
$$Q_{TT}(\gamma_0) = nT^*(\gamma_0)'T^*(\gamma_0), \tag{5.30}$$
$$Q_{ST}(\gamma_0) = nS^*(\gamma_0)'T^*(\gamma_0). \tag{5.31}$$

Since

$$AR(\gamma_0) = Q_{SS}(\gamma_0),$$
$$LM(\gamma_0) = (Q_{ST}(\gamma_0))^2 / Q_{TT}(\gamma_0),$$

the CLR statistic can also be re-written as

$$CLR(\gamma_0) = 0.5\left(AR(\gamma_0) - Q_{TT}(\gamma_0) + \sqrt{(AR(\gamma_0) - Q_{TT}(\gamma_0))^2 + 4Q_{TT}(\gamma_0)LM(\gamma_0)}\right),$$

which establishes the connection between the AR, LM and CLR statistics.

The null distribution of the CLR statistic under weak IVs is described in the following theorem.

**Theorem 5.6.** *Suppose that Assumptions 1.1 and 2.1 hold. Then,*

$$CLR(\gamma) \to_d 0.5\left(\mathcal{Z}_u'\mathcal{Z}_u - \mathcal{X}_T'\mathcal{X}_T + \sqrt{\left(\mathcal{Z}_u'\mathcal{Z}_u - \mathcal{X}_T'\mathcal{X}_T\right)^2 + 4\left(\mathcal{X}_T'\mathcal{X}_T\right)\left(\mathcal{X}_T'\mathcal{Z}_u\right)}\right),$$

*where $\mathcal{X}_T$ is defined in (5.25) and independent of $\mathcal{Z}_u$.*

*Remark.*

1. The null distribution of the CLR statistic is non-standard and cannot be tabulated as

$$\mathcal{X}_T \sim N\left(\frac{\lambda}{\sqrt{1-\rho_{uv}^2}}, I_l\right)$$

and therefore depends on the strength of IVs through unknown $\lambda$. However under $H_0$, the strength of IVs affects only $T^*(\gamma_0)$ and its limiting counterpart $\mathcal{X}_T$.

2. Since the asymptotic null distribution depends on $\lambda$ only through $\mathcal{X}_T$, and $\mathcal{X}_T$ and $\mathcal{Z}_u$ are *independent*, the null distribution of the CLR statistic can be easily simulated by *conditioning* on $T^*(\gamma_0)$ as described below.

*Proof of Theorem 5.6.* The result follows since by (5.25) and (5.27),

$$n^{1/2}S^*(\gamma) \to_d \mathcal{Z}_u,$$
$$n^{1/2}T^*(\gamma) \to_d \mathcal{X}_T,$$

and $\mathcal{Z}_u$ and $\mathcal{X}_T$ are independent by (5.26). □

One can simulate critical values for the CLR test by following the steps below:

1. Compute $n^{1/2}S^*(\gamma_0)$, $n^{1/2}T^*(\gamma_0)$, and the CLR statistic $CLR(\gamma_0)$ as described in (5.28)-(5.31).

2. Generate $R$ independent standard normal $l_1$-vectors $\{\mathcal{Z}_r : r = 1, \ldots, R\}$, $\mathcal{Z} \sim N(0, I_{l_1})$.

3. Generate $R$ values from the conditional null distribution of the CLR statistic given $T^*(\gamma_0)$ by using (5.28)-(5.31), but with $n^{1/2}S^*(\gamma_0)$ replaced with $\mathcal{Z}_r$, $r = 1, \ldots, R$:

$$\begin{aligned} CLR_r(\gamma_0) = 0.5 \big( & \mathcal{Z}_r'\mathcal{Z}_r - nT^*(\gamma_0)'T^*(\gamma_0) \\ & + \sqrt{\left(\mathcal{Z}_r'\mathcal{Z}_r - nT^*(\gamma_0)'T^*(\gamma_0)\right)^2 + 4n\left(T^*(\gamma_0)'\mathcal{Z}_r\right)^2} \Bigg), \end{aligned} \tag{5.32}$$

   which simulates the null distribution of (5.28). Note that $\mathcal{Z}_r \sim N(0, I_{l_1})$ captures the null distribution of $n^{1/2}S^*(\gamma_0)$, which is free of any parameters.

4. The critical value $cv_{1-\alpha}(\gamma_0)$ is given by the $(1-\alpha)$-th empirical quantile of $\{CLR_r(\gamma_0) : r = 1, \ldots, R\}$.

*Remark.*

1. Note that critical values for the CLR test depend on the value $\gamma_0$ and data: thus, different data sets and different null hypotheses would require different critical values.

2. Instead of simulations, one can compute p-values for the CLR test by numerical integration as shown in Andrews et al. (2007).

3. Mikusheva (2010) shows how to compute CLR CSs $CS_{1-\alpha}^{CLR} = \{\gamma_0 : CLR_{1-\alpha}(\gamma_0) \leq cv_{1-\alpha}(\gamma_0)\}$ fast and accurately using the numerical integration approach of Andrews et al. (2007). She also shows that CLR CSs cannot be empty and include the LIML estimator of $\gamma$. In fact, $CLR(\cdot)$ is minimized at the LIML estimator. Mikusheva (2010) also shows that CLR CSs exclude as undesirable point that maximizes $AR(\cdot)$ (and is always included in LM CSs).

4. Stata command `condivreg` produces CLR-based CSs (see Mikusheva and Poi, 2006).

We conclude this section by showing that the CLR statistic is asymptotically equivalent to the LM statistic when IVs are strong. This would verify that CLR-based inference is efficient when IVs are strong (in the sense that it attains the same power as that of the usual $t$-test-based inference under strong IVs).

**Theorem 5.7.** *Suppose that Assumption 1.1 holds,* $\gamma = \gamma_0 + n^{-1/2}\delta$*, and* $\pi_1$ *is fixed, i.e. IVs are strong. Then,*

$$CLR(\gamma_0) = LM(\gamma_0) + o_p(1). \tag{5.33}$$

$$cv_{1-\alpha}(\gamma_0) = \chi^2_{1,1-\alpha} + o_p(1). \tag{5.34}$$

*Proof of Theorem 5.7.* We show the result in (5.33) first. Recall that when IVs are strong and $\gamma = \gamma_0 + n^{-1/2}\delta$, $\tilde{\pi}_1(\gamma_0) \to_p \pi_1$ as implied by (5.11). It follows then from the definition of $T^*(\gamma_0)$ in (5.24) that

$$n^{1/2}T^*(\gamma_0) = \sqrt{\frac{\hat{\sigma}_u^2(\gamma_0)}{\hat{\sigma}_v^2\hat{\sigma}_u^2(\gamma_0) - \hat{\sigma}_{uv}^2(\gamma_0)}}\left(\frac{Z_1'M_2Z_1}{n}\right)^{1/2} n^{1/2}\tilde{\pi}_1(\gamma_0) \to \pm\infty,$$

or

$$Q_{TT}(\gamma_0) = nT^*(\gamma_0)'T^*(\gamma_0) \to \infty.$$

In what follows, we omit the dependence on $\gamma_0$ for simplicity. Let

$$J = AR - LM,$$

and note that both $AR$ and $LM$ are $O_p(1)$ when IVs are strong and $\gamma = \gamma_0 + n^{-1/2}\delta$. We have

$$\begin{aligned} CLR &= 0.5\left(AR - Q_{TT} + \sqrt{(AR - Q_{TT})^2 + 4Q_{TT}LM}\right) \\ &= 0.5\left(LM + J - Q_{TT} + \sqrt{(LM + J - Q_{TT})^2 + 4Q_{TT}LM}\right) \\ &= 0.5\left(LM + J - Q_{TT} + \sqrt{(LM - J + Q_{TT})^2 + 4J \cdot LM}\right), \end{aligned} \tag{5.35}$$

where the result in the last line follows because

$$\begin{aligned} (LM + J - Q_{TT})^2 + 4Q_{TT}LM &= (J + LM)^2 + Q_{TT}^2 - 2Q_{TT}(J + LM) + 4Q_{TT}LM \\ &= (J + LM)^2 + Q_{TT}^2 - 2Q_{TT}(J - LM) \\ &= (J - LM)^2 + Q_{TT}^2 - 2Q_{TT}(J - LM) + 4J \cdot LM \\ &= (LM - J)^2 + Q_{TT}^2 + 2Q_{TT}(LM - J) + 4J \cdot LM \\ &= (LM - J + Q_{TT})^2 + 4J \cdot LM. \end{aligned}$$

Consider a mean-value expansion of $\sqrt{(LM - J + Q_{TT})^2 + 4J \cdot LM}$ around $(LM - J + Q_{TT})^2$:

$$\begin{aligned}\sqrt{(LM - J + Q_{TT})^2 + 4J \cdot LM} &= \sqrt{(LM - J + Q_{TT})^2} - \frac{4J \cdot LM}{2\sqrt{(J - LM - Q_{TT})^2 + \xi}} \\ &= LM - J + Q_{TT} + o_p(1), \end{aligned} \tag{5.36}$$

where $\xi$ is the mean-value (i.e. $|\xi| < 4|J| \cdot LM$), and the $o_p(1)$ term in the second line is due to the fact that $Q_{TT} \to \infty$ in the denominator of the second term in the first line. By (5.35) and (5.36),

$$\begin{aligned} CLR &= 0.5\left(LM + J - Q_{TT} + (LM - J + Q_{TT} + o_p(1))\right) \\ &= LM + o_p(1). \end{aligned}$$

The result in (5.34) can be shown by re-writing the simulated CLR statistic in (5.32) as

$$\begin{aligned} CLR_r(\gamma_0) &= 0.5\left(\mathcal{Z}_r'\mathcal{Z}_r - nT^*(\gamma_0)'T^*(\gamma_0)\right. \\ &\qquad \left. + \sqrt{(\mathcal{Z}_r'\mathcal{Z}_r - nT^*(\gamma_0)'T^*(\gamma_0))^2 + 4n\left(T^*(\gamma_0)'\mathcal{Z}_r\right)^2}\right) \\ &= 0.5\left(\mathcal{Z}_r'\mathcal{Z}_r - Q_{TT}(\gamma_0) + \sqrt{(\mathcal{Z}_r'\mathcal{Z}_r - Q_{TT}(\gamma_0))^2 + 4Q_{ST,r}^2(\gamma_0)}\right), \end{aligned}$$

where

$$\begin{aligned} Q_{ST,r}^2(\gamma_0) &= \left(n^{1/2}T^*(\gamma_0)'\mathcal{Z}_r\right)^2 \\ &= Q_{TT} \cdot LM_r(\gamma_0), \text{ and} \\ LM_r(\gamma_0) &= \frac{\left(n^{1/2}T^*(\gamma_0)'\mathcal{Z}_r\right)^2}{nT^*(\gamma_0)'T^*(\gamma_0)} \sim \chi_1^2 \text{ for any } n. \end{aligned}$$

The rest of the proof of the result in (5.34) follows the same steps as that of (5.33). One can show that

$$CLR_r(\gamma_0) \to_d \chi_1^2,$$

and therefore the quantiles of the empirical distribution of $\{CLR_r(\gamma_0) : r = 1, \ldots, R\}$ converge to the quantiles of the $\chi_1^2$ distribution, since the $\chi_1^2$ distribution has a continuous CDF. □

*Remark.* Examination of the proof of Theorem 5.7 reveals that the CLR attains efficiency when $n^{1/2}T^*(\gamma_0) \to \infty$. By inspecting the limiting distribution of $n^{1/2}T^*(\gamma_0)$ in (5.25),

$$n^{1/2}T^*(\gamma_0) \to_d N\left(\frac{\lambda}{\sqrt{1 - \rho_{uv}^2}}, I_l\right),$$

one can see that $n^{1/2}T^*(\gamma_0) \to \infty$ when $\|\lambda\|^2 = \infty$, i.e. IVs are strong. However, it is possible that $n^{1/2}T^*(\gamma_0) \to \infty$ even when IVs are weak ($\|\lambda\|^2 < \infty$). With finite values of the concentration parameter, that can happen when $\rho_{uv} \to \pm 1$, i.e. the correlation between the structural error $u_i$ and $v_i$ approaches one. This idea is formalized in Andrews et al. (2016).

## 5.6 Concluding remarks

While weak IVs can pose serious challenges for empirical research, robust and computationally simple methods are available for inference with weak IVs. Robust methods are as reliable with weak IVs as with strong IVs, and when using them the researcher does not need to be concerned with the strength of identification. Moreover, if the regression model contains only one endogenous regressor (which is the main object of interest of the econometrician), weak-IV-robust methods based on the CLR (or LM) approach will be as powerful as the usual $t$-statistic-based inference in case IVs are strong. The form of AR, LM, and CLR-based CSs can also provide information about the strength of IVs. In addition, AR CSs come with built-in model specification diagnostics.

# A Appendix: Results from probability and linear algebra

## A.1 Limit theorems

Large-sample or asymptotic properties of estimators are established using weak laws of large numbers (WLLNs) and central limit theorems (CLTs), see for example White (2001) for details and references.

**Theorem A.1** (WLLN for iid data)**.** *Suppose that* $\{X_i : i = 1, \ldots, n\}$ *are iid random variables, and* $E|X_i| < \infty$*. Then,* $n^{-1}\sum_{i=1}^{n} X_i \to_p EX_i$*.*

**Theorem A.2** (CLT for iid data)**.** *Suppose that* $\{X_i : i = 1, \ldots, n\}$ *are iid random vectors, and* $Var(X_i)$ *is positive definite and finite, Then,* $n^{-1/2}\sum_{i=1}^{n}(X_i - EX_i) \to_d N(0, Var(X_i))$*, where* $N(\mu, \Omega)$ *denotes the multivariate normal distribution with mean* $\mu$ *and variance-covariance matrix* $\Omega$*.*

## A.2 Multivariate normal distributions

Suppose that $X$ and $Y$ are jointly normally distributed:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}\right).$$

Here, $\mu_X$ and $\mu_Y$ denote the means of $X$ and $Y$ respectively, $\Sigma_{XX}$ and $\Sigma_{YY}$ denote their respective variances, and $\Sigma_{XY} = Cov(X, Y)$.

**Theorem A.3.** $X$ *and* $Y$ *are independent if* $\Sigma_{XY} = 0$*, i.e.* $Cov(X, Y) = 0$*.*

**Theorem A.4.** *The conditional distribution of* $Y$ *given* $X$ *is normal:*

$$\begin{aligned} Y|X &\sim N\left(\mu_{Y|X}(X), \Sigma_{Y|X}\right), \text{ where} \\ \mu_{Y|X}(X) &= \mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(X - \mu_X) \text{ (a vector-valued function of } X\text{)}. \\ \Sigma_{Y|X} &= \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} \text{ (a fixed matrix)}. \end{aligned}$$

**Theorem A.5.** *Let* $\Gamma$ *be a fixed matrix.*

$$\Gamma X \sim N\left(\Gamma\mu_X, \Gamma\Sigma_{XX}\Gamma'\right).$$

## A.3 $O$ notation

**Definition A.6.** A sequence of random variables $X_n$ is said to be $O_p(1)$ if for all $\varepsilon > 0$ there are constants $K_\varepsilon > 0$ and $N_\varepsilon > 0$ such that

$$P\left(|X_n| \leq K_\varepsilon\right) \geq 1 - \varepsilon$$

for all $n \geq N_\varepsilon$.

**Definition A.7.** A sequence of random variables $X_n$ is said to be $o_p(1)$ if $X_n \to_p 0$.

## A.4 Kronecker product

**Definition A.8.** Let $A$ be a $k \times l$ matrix, and let $B$ be an $m \times n$ matrix. The Kronecker product of $A$ and $B$, denoted as $A \otimes B$, is defined as

$$A \otimes B = \begin{pmatrix} A_{11}B & \dots & A_{1l}B \\ & \dots & \\ A_{k1}B & \dots & A_{kl}B \end{pmatrix}.$$

The properties of the Kronecker product as as follows:

1. The dimensions of $A \otimes B$ are $km \times nl$.
2. $(A \otimes B)' = A' \otimes B'$.
3. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.
4. $(A \otimes B)(C \otimes D) = AC \otimes BD$.

## A.5 Non-central $\chi^2$ distributions

Let $\mathcal{Z} \sim N(0, I_l)$ be an $l$-vector where each component is a standard normal random variable, and the components are independent of each other. The $\chi^2$ distribution with $l$ degrees of freedom is defined as

$$\mathcal{Z}'\mathcal{Z} = \|\mathcal{Z}\|^2 = \sum_{i=1}^{l} \mathcal{Z}_i^2 \sim \chi_l^2. \tag{A.1}$$

This distribution depends only on the number of components, i.e. the number of degrees of freedom $l$ is the single parameter determining the $\chi^2$ distribution.

This discussion can be extended to the case of independent normal variables with non-zero means and unit variances. For $\lambda = (\lambda_1, \dots, \lambda_l)'$, consider

$$\lambda + \mathcal{Z} \sim N(\lambda, I_l).$$

While the distribution of $\lambda + \mathcal{Z}$ involves $l$ parameters the distribution of $(\lambda + \mathcal{Z})'(\lambda + \mathcal{Z}) = \|\lambda + \mathcal{Z}\|^2$ depends only on two parameters: $l$ and $\|\lambda\|$, the norm of the vector of means $\lambda$.

**Definition A.9** (Equality in distribution)**.** We say that two random $l$-vectors $X$ and $Y$ are equal in distribution, denoted $X =^d Y$ if the CDF of $X$ is equal to the CDF of $Y$: for all $u \in \mathbb{R}^l$,

$$P(X \leq u) = P(Y \leq u).$$

**Lemma A.10.** *Let* $\mathcal{Z}_u = (\mathcal{Z}_{u,1}, \dots, \mathcal{Z}_{u,l})'$ *and* $\mathcal{Z}_v = (\mathcal{Z}_{v,1}, \dots, \mathcal{Z}_{v,l})'$ *be two* $l$*-vectors with a normal joint distribution:*

$$\begin{bmatrix} \mathcal{Z}_u \\ \mathcal{Z}_v \end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \otimes I_l \right).$$

*Then,*

$$\|\lambda + \mathcal{Z}_v\|^2 =^d (\|\lambda\| + \mathcal{Z}_{v,1})^2 + \sum_{j=2}^{l} \mathcal{Z}_{v,j}^2, \tag{A.2}$$

$$(\lambda + \mathcal{Z}_v)' \mathcal{Z}_u =^d (\|\lambda\| + \mathcal{Z}_{v,1})\mathcal{Z}_{u,1} + \sum_{j=2}^{l} \mathcal{Z}_{v,j}\mathcal{Z}_{u,j}. \tag{A.3}$$

*Remark.*
1. Note that the expressions on the right-hand sides of (A.2) and (A.3) depend on a scalar parameter $\|\lambda\|$ and not on the entire vector $\lambda$.
2. The distribution $\|\lambda + \mathcal{Z}_v\|^2$ of is called *non-central* $\chi^2$ with $l$ degrees of freedom and denoted as
$$\|\lambda + \mathcal{Z}_v\|^2 \sim \chi_l^2(\|\lambda\|^2).$$
When $\|\lambda\| = 0$, the non-central $\chi_l^2$ distribution becomes the usual (central) $\chi_l^2$.

*Proof of Lemma A.10.* Let $b_1, \dots, b_l$ be orthonormal $l$-vectors ($b_i'b_i$=1 and $b_i'b_j = 0$ for $i \neq j$) such that $b_1 = \lambda/\|\lambda\|$. Define an $l \times l$ matrix

$$B = \begin{bmatrix} \frac{\lambda'}{\|\lambda\|} \\ b_2' \\ \vdots \\ b_l' \end{bmatrix},$$

and note that $BB' = I_l$, $(B^{-1})'B^{-1} = I_l$, and

$$B\lambda = \begin{bmatrix} \|\lambda\| \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Define

$$\mathcal{X}_u = B\mathcal{Z}_u \text{ and } \mathcal{X}_v = B\mathcal{Z}_v.$$

We have

$$\begin{aligned}
\begin{bmatrix} \mathcal{X}_u \\ \mathcal{X}_v \end{bmatrix} &= \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes B \right) \begin{bmatrix} \mathcal{Z}_u \\ \mathcal{Z}_v \end{bmatrix} \\
&\sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \otimes BB' \right) \\
&= N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \otimes I_l \right),
\end{aligned}$$

where the result in the second line follows by Property (iv) of the Kronecker product in Section A.4. Now,

$$\begin{aligned}
(\lambda + \mathcal{Z}_v)' (\lambda + \mathcal{Z}_v) &= (B(\lambda + \mathcal{Z}_v))' \left(B^{-1}\right)' B^{-1} (B(\lambda + \mathcal{Z}_v)) \\
&= (B\lambda + \mathcal{X}_v)' (B\lambda + \mathcal{X}_v) \\
&= (\|\lambda\| + \mathcal{X}_{v,1})^2 + \sum_{j=2}^{l_1} \mathcal{X}_{v,j}^2. \\
\mathcal{Z}_u' (\lambda + \mathcal{Z}_v) &= (B\mathcal{Z}_u)' \left(B^{-1}\right)' B^{-1} (B(\lambda + \mathcal{Z}_v)) \\
&= \mathcal{X}_u' (B\lambda + \mathcal{X}_v) \\
&= (\|\lambda\| + \mathcal{X}_{v,1})\mathcal{X}_{u,1} + \sum_{j=2}^{l_1} \mathcal{X}_{v,j}\mathcal{X}_{u,j}.
\end{aligned}$$

The result follows since

$$\begin{bmatrix} \mathcal{X}_u \\ \mathcal{X}_v \end{bmatrix} =^d \begin{bmatrix} \mathcal{Z}_u \\ \mathcal{Z}_v \end{bmatrix}.$$

□

# B Matlab programs

## B.1 Graphing the distribution of the $t$-statistic under weak IVs

The code below is used to generate Figure 1. It compares the null distribution (PDF) of the $t$-statistic under weak IVs as described in Theorem 2.3 with the PDF of the standard normal distribution.

```
l=2; %number of IVs
cp=1; %concentration parameter
rho=0.95; %endogeneity parameter
```

```
alpha=0.05; %nominal significance level
R=10000; %the number of simulations

%grid for density plotting
step=0.1; xi=[-4:step:4];
Sigma=[1 rho; rho 1];
V=kron(Sigma,eye(l));
H=chol(V);
T=zeros(R,1);

for i=1:R
   Z=H'*normrnd(0,1,2*l,1);
   Zu=Z(1:l);
   Zv=Z(l+1:2*l)+[sqrt(cp);
   zeros(l-1,1)];
   Ch1=(Zv'*Zu);
   Ch2=(Zv'*Zv);
   Numer=Ch1*sqrt(Ch2);
   Denom=Ch2^2+Ch1^2-2*rho*Ch1*Ch2;
   T(i,1)=Numer/sqrt(Denom);
end

[f,xj]=ksdensity(T); %kernel smoothing of the distribution of T
plot(xj,f,'k',xi,normpdf(xi),'k--');
axis([-4 12 0 0.41])
%legend('T under weak IVs','N(0,1)','location','best')

disp('--------------------')
disp('Null Rejection Rate:')
disp(sum(abs(T)>norminv(1-alpha/2))/R)
disp('--------------------')
```

## B.2 Simulating the maximum rejection probabilities of the $t$-test under weak IVs

The code below is used to generate data for Table 1, which reports the maximum rejection probabilities of the two-sided $t$-test when IVs are weak, i.e. the values of the $R^{\max}_{\alpha,l_1}(\|\lambda\|^2)$

function in Section 3.1. The code somewhat overlaps with the code in Section B.1, however, here it is implemented without loops over simulation iterations to reduce computational time.

```
l=1; %number of IVs
alpha=0.05; %alpha value for choosing normal crit.vals
B=100000; %number of simulations

%grid of values for the concentration parameter
CP=[0.01 0.1 0.25 1 4 9 16 25 36 49 64 81 100 1000]';

%grid of values for rho_uv
Rho=[-.99:0.01:.99];


RAND_N=normrnd(0,1,2*l,B);

Rmax=zeros(length(CP),1);

randseedoffset=81474;

parfor i=1:length(CP);

    rng(randseedoffset+i, 'twister'); %to generate the same random sequences

    cp=CP(i);
    R=zeros(length(Rho),1);

    for j=1:length(Rho)
        rho=Rho(j);

        Sigma=[1 rho; rho 1];
        V=kron(Sigma,eye(l));
        H=chol(V);
        Z=H'*RAND_N;
        Zu=Z(1:l,:);
        Zv=Z(l+1:2*l,:)+repmat([sqrt(cp); zeros(l-1,1)],1,B);
        X=sum(Zv.^2,1);
        Y=sum(Zv.*Zu,1);
```

```
        Numer=(sqrt(X).*Y);
        Denom=Y.^2+X.^2-2*rho*X.*Y;
        T=Numer./sqrt(Denom);

        %rejections: R function for alpha, the concentr.parm, rho, and # of IVs
        R(j)=sum(abs(T)>norminv(1-alpha/2))/B;
        Rmax(i)=max(R);
    end
end
```

## References


Andrews, D. W. K., X. Cheng, and P. Guggenberger (2011): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," Cowles Foundation Discussion Paper 1813. 13

Andrews, D. W. K., V. Marmer, and Z. Yu (2016): "A Note on Optimal Inference in the Linear IV Model," Cowles Foundation Discussion Paper No. 2073 , Yale University. 56, 62

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2006): "Optimal Invariant Similar Tests For Instrumental Variables Regression," *Econometrica*, 74, 715–752. 56

——— (2007): "Performance of conditional Wald tests in IV regression with weak instruments," *Journal of Econometrics*, 139, 116–132. 59

Blundell, R. and J. L. Powell (2003): "Endogeneity in nonparametric and semiparametric regression models," in *Advances in Economics and Econometrics*, ed. by M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, Cambridge, United Kingdom: Cambridge University Press, vol. 36 of *Econometric Society Monographs*, 312–357. 4

Davidson, R. and J. G. MacKinnon (2004): *Econometric Theory and Methods*, Oxford University Press, New York. 2, 4, 7, 52

Dufour, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1387. 32, 37

Kleibergen, F. (2002): "Pivotal Statistics For Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803. 46

Mikusheva, A. (2010): "Robust confidence sets in the presence of weak instruments," *Journal of Econometrics*, 157, 236–247. 55, 59

Mikusheva, A. and B. P. Poi (2006): "Tests and confidence sets with correct size when instruments are potentially weak," *Stata Journal*, 6, 335–347. 55, 59

Moreira, M. J. (2001): "Tests with Correct Size When Instruments Can Be Arbitrarily Weak," Department of Economics, University of California, Berkeley. 46

——— (2003): "A Conditional Likelihood Ratio Test For Structural Models," *Econometrica*, 71, 1027–1048. 56

Staiger, D. and J. H. Stock (1997): "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65, 557–586. 13, 29

Stock, J. H. and M. Yogo (2005): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews and J. H. Stock, New York: Cambridge University Press, chap. 6, 80–108. 22, 23, 24

White, H. (2001): *Asymptotic Theory For Econometricians*, San Diego: Academic Press. 63