**First-Stage Instrument Hacking and Median Bias**

## 1. Model

Consider the following model

$$
\begin{aligned}
y_i =& \beta x_i + u_i \\
x_i =& \sum_{k=1}^{\bar{K}} \pi_k z_{ki} + \rho u_i
\end{aligned}
\tag{1}
$$

where $u_i \overset{\text{iid}}{\sim} N(0, \sigma_u^2)$; $z_{ki}$ are binary instruments; and $L$ is the total number of instruments in the first-stage equation for $x_i$. Let $K < L$ denote the number of instruments observed by the researcher.

Consider the impact of second-stage $p$-hacking among just-identified specifications on median bias. For this result, we make the following assumptions:

**Assumption 1** (Same Instrument Strength)**.** $\pi_k = \pi_{k'} \equiv \pi \neq 0$ *for all* $k, k' = 1, 2, ..., \bar{K}$.

**Assumption 2** (Same Randomization Procedure)**.** $\sum_i z_{ki} = \sum_i z_{k'i} = N_Z < N$ *for all* $k, k' = 1, 2, ..., \bar{K}$.

**Assumption 3** (Finite Sample Equicorrelation)**.** $\sum_{j \neq k} \sum_i z_{ki} z_{ji} = \sum_{j \neq k'} \sum_i z_{k'i} z_{ji}$ *for all* $k, k' = 1, 2, ..., \bar{K}$.

Assumption 1 assumes all instruments have the same strength, and Assumption 2 imposes identical randomization procedures, in the sense that all instruments make $N_Z$ individuals 'eligible' for treatment. Together, they impose a similarity across instruments in order to isolate the impact of having additional instruments of the same 'type' on median bias.

Imposing some form of restriction on instrument strength across different instruments is necessary for showing that median bias worsens as the number of available instruments increases. To illustrate this, consider the case where a researcher 'draws' a sequence of instruments whereby each additional instrument is ten times stronger than the most recent one drawn (i.e. $\pi_1 = \frac{1}{10}\pi_2 = \frac{1}{100}\pi_3 = ...$). In this example, median bias would likely *decrease* as the number of instruments increases because the most recent instrument is always the strongest, and therefore would deliver the 2SLS smallest standard error and lowest $p$-value in the second stage. Assumption 1 is made to rule out such examples which tell us little about the impact of instrument-hacking.

The correlation between instruments is not the primary interest. Consequently, Assumption 3 imposes equal correlation across pairs of instruments. This will not hold exactly in any given finite sample; however, for statistically independent instruments, for example, it would hold in expectation. The requirement that it holds in finite samples is made to abstract away from sampling variation in the correlation between instruments to iisolate the impact of instrument selection.

## 2. Known $\sigma_u$

The main result assuming $\sigma_u$ is known and not estimated:

**Proposition 1** (Median Bias). *Consider the model in (1) with $\rho \neq 0$, $\beta = 0$ and second-stage instrument-hacking among just-identified specifications. Under Assumptions 1, 2 and 3, the median of the reported 2SLS estimator is biased toward OLS and is strictly increasing in absolute value with the number of available instruments $K$.*

**Proof of Proposition 1:** Define the following terms that are used throughout the proofs for convenience:

$$M_{zu}(k) = \frac{1}{N}\sum_{i=1}^{N} z_{ki}u_i$$

$$C = \frac{1}{N}\sum_i \pi_k z_{ki}^2 + \frac{1}{N}\sum_{l\neq k}\sum_i \pi_j z_{li}z_{ki} = \frac{N_Z}{N}\pi + \frac{1}{N}\pi\sum_{l\neq k}\sum_i z_{li}z_{ki}$$

where the second definition uses Assumptions 1–3). Assume throughout that $C > 0$ and $\rho > 0$. Other cases should have similar logic but with flipped signs.

The proof consists of four steps.

**Step 1.** In the first step, we derive the second-stage $t$-statistic when using the just-identified specification with instrument $k$. First, see that the 2SLS estimate and standard error are given by

$$\hat{\beta}(M_{zu}(k)) = \frac{\sum_i z_{ki}u_i}{\sum_i z_{ki}x_i} = \frac{M_{zu}(k)}{C + \rho M_{zu}(k)}$$

$$se(M_{zu}(k)) = \frac{\sigma_u\sqrt{N_Z}}{|C + \rho M_{zu}(k)|}$$

and thus the absolute value of the $t$-ratio is equal to

$$|t(M_{zu}(k))| = \frac{|\ \hat{\beta}(M_{zu}(k))\ |}{se(M_{zu}(k))} = \frac{|\ M_{zu}(k)\ |}{\sigma_u \sqrt{N_Z}}$$

Note that $M_{zu}(k) \equiv \frac{1}{N} \sum z_{ki} u_i$ is a normalized sum of $N_z$ independent normals following $N(0, \sigma_u^2)$, which implies that $M_{zu}$ follows an $N(0, \frac{q}{N}\sigma_u^2)$ distribution, where $q \equiv N_z/N$.

**Step 2.** In this step, we derive the cdf for the random variable $M_{zu}(k^*)$, where $k^* = \arg\max_{k \geqslant 1} |t(M_{zu}(k))|$. Let $F_{M_{zu}^*}(\cdot)$ denote the cdf. Note that $M_{zu}(k^*)$ is based on choosing the instrument with the largest $M_{zu}(k)$ in absolute terms, but that it retains its original sign.

First, see that $M_{zu}(k^*)$ must be symmetric about zero and hence satisfy $F_{M_{zu}^*}(0) = \frac{1}{2}$. Next, define the absolute value of the chosen instrument as $M = \max |M_{zu}(k)|$. Then, for any $x > 0$,

$$\begin{aligned} F_{M_{zu}^*}(x) &= \mathbb{P}\big(M_{zu}(k^*) \leqslant 0\big) + \mathbb{P}\big(0 < M_{zu}(k^*) < x\big) \\ &= \frac{1}{2} + \mathbb{P}\big(M_{zu}(k^*) > 0 | M < x\big)\mathbb{P}(M < x\big) \\ &= \frac{1}{2} + \frac{1}{2}\big[F_{M_{zu}}(x) - F_{M_{zu}}(-x)\big]^K \end{aligned}$$

where the final line follows from the fact that the sign of $M_{zu}(k^*)$ is independent of its magnitude once we condition on that magnitude being below any positive threshold $x$.

In the case where $x < 0$, we can obtain the cdf by using the fact that the cdf is symmetric: $F_{M_{zu}^*}(x) = 1 - F_{M_{zu}^*}(-x)$, which gives

$$F_{M_{zu}^*}(x) = \begin{cases} \frac{1}{2}\big(1 - [F_{M_{zu}}(-x) - F_{M_{zu}}(x)]^K\big) & \text{if } x \leqslant 0 \\ \frac{1}{2}\big(1 + [F_{M_{zu}}(x) - F_{M_{zu}}(-x)]^K\big) & \text{if } x > 0 \end{cases} \tag{2}$$

**Step 3.** Let $m_K$ denote median bias when $K$ instruments are available. In this step, we show that $m_K \in (0, \frac{1}{\rho})$.

Let $F_{\hat{\beta}*,K}(\cdot)$ denote the cdf of $\hat{\beta}(M_{zu}(k^*))$. Note that $\hat{\beta}(\cdot)$ is invertible at every point on the real line except when $M_{zu}(k) = -C/\rho$, where it is not well-defined, and that the inverse function is $\hat{\beta}^{-1}(y) = \frac{yC}{1-y\rho}$.

Consider the following two inequalities

$$\begin{aligned}
F_{\hat{\beta}*,K}(0) &= \mathbb{P}\big(-\infty < \hat{\beta}(M_{zu}(k^*)) < 0\big) \\
&= \mathbb{P}\bigg(-\frac{C}{\rho} < M_{zu}(k^*) < 0\bigg) < \mathbb{P}\bigg(M_{zu}(k^*) < 0\bigg) = \frac{1}{2} \\
F_{\hat{\beta}*,K}\bigg(\frac{1}{\rho}\bigg) &= \mathbb{P}\bigg(-\infty < \hat{\beta}(M_{zu}(k^*)) < \frac{1}{\rho}\bigg) \\
&= \mathbb{P}\bigg(-\frac{C}{\rho} < M_{zu}(k^*) < \infty\bigg) \\
&= 1 - \mathbb{P}\bigg(M_{zu}(k^*) < -\frac{C}{\rho}\bigg) > 1 - \mathbb{P}\bigg(M_{zu}(k^*) < 0\bigg) = \frac{1}{2}
\end{aligned}$$

Combining these, we get

$$F_{\hat{\beta}*,K}(0) < \frac{1}{2} < F_{\hat{\beta}*,K}\bigg(\frac{1}{\rho}\bigg)$$

By definition, median bias, $m_K$, must satisfy $F_{\hat{\beta}*,K}(m_K) = \frac{1}{2}$. It follows that $m_K \in \big(0, \frac{1}{\rho}\big)$, which is what we wanted to show.

**Step 4.** In this step, we prove the main claim, namely, that median bias is increasing in the number of available instruments: $m_{K+1} > m_K$.

The first step is to show that $F_{\hat{\beta}*,K+1}(x) < F_{\hat{\beta}*,K}(x)$ for any $x \in \big(0, \frac{1}{\rho}\big)$. Using the cdf in (2),

$$\begin{aligned}
F_{\hat{\beta}*,K}(x) &= \mathbb{P}\big(-\infty < \hat{\beta}(M_{zu}(k^*)) \leqslant x\big) \\
&= \mathbb{P}\bigg(-\frac{C}{\rho} < M_{k*,K} \leqslant \hat{\beta}^{-1}(x)\bigg) \\
&= \frac{1}{2}\bigg[F_{M_k}\bigg(\frac{C}{\rho}\bigg) - F_{M_k}\bigg(-\frac{C}{\rho}\bigg)\bigg]^K + \frac{1}{2}\bigg[F_{M_k}\bigg(\hat{\beta}^{-1}(x)\bigg) - F_{M_k}\bigg(-\hat{\beta}^{-1}(x)\bigg)\bigg]^K
\end{aligned}$$

Both expressions in square brackets are positive and strictly less than one: this is immediately clear in the first expression; and holds for the second expression because $\hat{\beta}^{-1}(x) = \frac{xC}{1-x\rho} > 0$ when $x \in \big(0, \frac{1}{\rho}\big)$. From this, it follows that $F_{\hat{\beta}*,K+1}(x) < F_{\hat{\beta}*,K}(x)$ for any $x \in \big(0, \frac{1}{\rho}\big)$.

Finally, note again that median bias $m_K$ by definition satisfies

$$F_{\hat{\beta}*,K}(m_K) = \frac{1}{2} = 1 - F_{\hat{\beta}*,K}(m_K)$$

and the derived inequality implies

$$F_{\hat{\beta}*,K+1}(m_K) < \frac{1}{2} < 1 - F_{\hat{\beta}*,K+1}(m_K)$$

Thus, median bias with $K+1$ available instruments, which by definition equalizes these inequalities, must satisfy $m_{K+1} > m_K$, which completes the proof.

## 3. Generalization to Estimated $\sigma_u$

Now assume that $\sigma_u$ is not known but estimated. Specifically, following Phillips's (1989) result, we assume that for large (but finite) samples that it follows a quadratic distribution which is minimized at population OLS:

$$\begin{aligned}\sigma^2_{u*}(M_{zu}(k)) =& a\hat{\beta}\big(M_{zu}(k)\big)^2 + b \cdot \hat{\beta}\big(M_{zu}(k)\big) + c \\ =& a\left(\frac{M_{zu}(k)}{C + \rho M_{zu}(k)}\right)^2 + b \cdot \left(\frac{M_{zu}(k)}{C + \rho M_{zu}(k)}\right) + c\end{aligned}$$

where $a > 0$ because it is concave up; and $b < 0$ since it is minimized at the known value of population OLS, $\bar{\beta}_{OLS} > 0$. The top panel in Figure 1 plots $\sigma^2_{u*}(M_{zu}(k))$. Note that the variance is quadratic in $\hat{\beta}(M_{zu}(k))$, not in $M_{zu}(k)$, which is what is on the $x$-axis.

With this, the function for the absolute $t$-ratio changes to

$$|t(M_{zu}(k))| = \frac{|\ \hat{\beta}(M_{zu}(k))\ |}{se(M_{zu}(k))} = \frac{|M_{zu}(k)|}{\sqrt{\sigma^2_u\big(M_{zu}(k)\big) \cdot N_Z}}$$

This is plotted in the second panel. This expression is challenging to work with because the problem of instrument selection is no longer equivalent to choosing the $M_{zu}(k)$ with the largest absolute value, as it was in the case where $\sigma_u$ was known. This makes it intractable to find a closed-form expression for the cdf of $F_{M^*_{zu}}(\cdot)$ and therefore also for $F_{\hat{\beta}*,K}(\cdot)$, which the proof above used to make claims about median bias. Idea on how to proceed below.

### 3.1. Idea

This is probably the most straightforward option. In the identified case ($\pi > 0$), which is what weŕe interested in, $\hat{\beta}_{2sls} \to \beta$ and therefore $\hat{\sigma}^2_u \to \sigma^2_u$. Hence, from an asymptotic point of view, we can more or less use the earlier result where $\sigma^2_u$ is assumed to be fixed. This would change the result to a *probabilistic* one (and the proof would have to be adapted accordingly; see below). That is, for any $K \geqslant 1$, as $N \to \infty$, $\mathbb{P}[m_{K+1} > m_K] \to 1$.

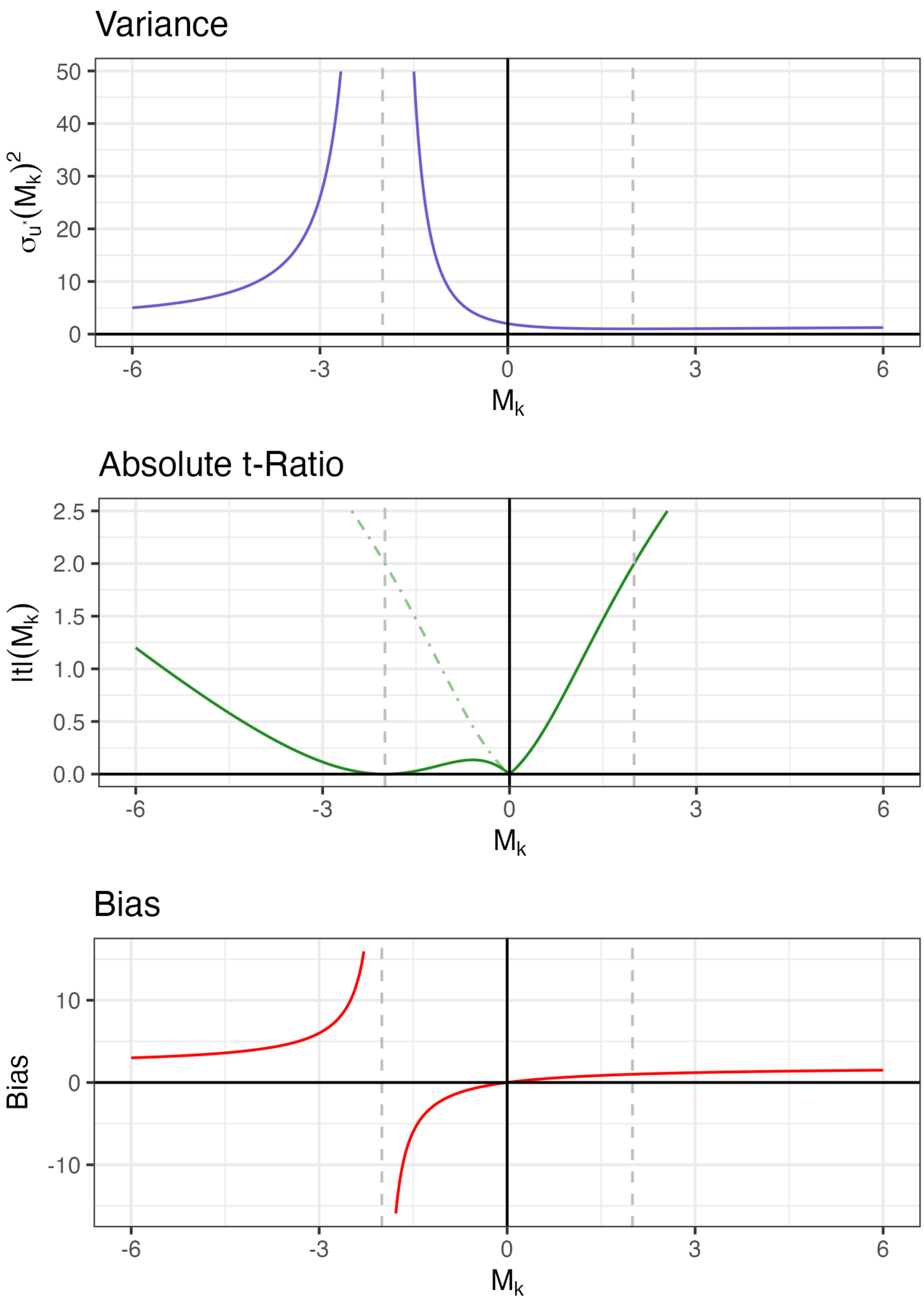


Figure 1: Intuition for Median Bias Proposition

### 3.2. Proof according to ChatGPT

It probably makes sense to discuss if this idea seems sensible before writing out a proof in detail. However, as a sketch, I prompted ChatGPT to get the following proof to get some idea of how it would change.

## 1. Model and notation

**Data and first–stage covariances.** For each sample size $N$ we observe i.i.d. rows $(u_i, z_{1i}, \ldots, z_{Ki})$ and form

$$M_k = \frac{1}{N}\sum_{i=1}^{N} z_{ki}u_i, \qquad k = 1, \ldots, K,$$

with $\mathrm{Var}(z_{ki}u_i) = \sigma_M^2 > 0$. The number of instruments $K$ is *fixed* throughout.

**Structural constants.** $C > 0,\ \rho > 0$ (pole at $M = -C/\rho < 0$).

**Quadratic variance estimator.**

$$\hat{\beta}(M) = \frac{M}{C + \rho M}, \quad \hat{\sigma}_u^2(M) = a_N\hat{\beta}(M)^2 + b_N\hat{\beta}(M) + c_N.$$

**Absolute $t$-statistic and researcher's pick.**

$$g(M) = |t(M)| = \frac{|M|}{\sqrt{N_z}\,\hat{\sigma}_u(M)}, \quad k^*(K) = \arg\max_{1 \leqslant k \leqslant K} g(M_k).$$

**Published estimator and medians.**

$$\hat{\beta}^*(K) = \hat{\beta}\big(M_{k^*(K)}\big), \qquad m_{K,N} = \mathrm{med}\{\hat{\beta}^*(K)\}.$$

Define the *scaled* medians $\mu_{K,N} := \sqrt{N}\, m_{K,N}$.

## 2. Minimal assumptions

**Assumption 4** (Root-$N$ consistency of the variance)**.** $a_N, b_N = O(N^{-1/2})$ *and* $c_N \to \sigma_u^2 > 0$.

**Assumption 5** (Non-degenerate baseline)**.** *The scaled one-instrument median converges to a positive constant:*

$$\sqrt{N}\, m_{1,N} \longrightarrow \mu_1 > 0.$$

## 3. Main theorem

[Strict ordering with probability → 1] Under Assumptions 4–5, for every fixed integer $K \geqslant 1$,

$$\Pr\bigl[m_{K+1,N} > m_{K,N}\bigr] \xrightarrow[N\to\infty]{} 1.$$

## 4. Proof

*Proof.* **Step 1 (tube around the pole).** Fix $\eta \in (0, \frac{1}{2})$ and set

$$\delta_N := N^{-1/2+\eta}, \qquad \mathcal{T}_N := \{|M + C/\rho| < \delta_N\}.$$

For each $k$, $\Pr(M_k \in \mathcal{T}_N) = O(N^{-1/2+\eta})$; with $K$ fixed,

$$\Pr\bigl[k^*(K) \in \mathcal{T}_N\bigr] = O\bigl(N^{-1/2+\eta}\bigr). \tag{3}$$

**Step 2 (monotonicity outside $\mathcal{T}_N$).** By Assumption 4,

$$\hat{\sigma}_u(M) = \sigma_u\,[1 + O(N^{-1/2})] \quad \text{when } |M + C/\rho| \geqslant \delta_N.$$

Thus $g(M) = |M|/\bigl(\sqrt{N_z}\,\sigma_u\bigr)[1 + O(N^{-1/2})]$ and inherits the ordering of $|M|$.

**Step 3 (limit of the scaled estimator).** On the "good" event $\mathcal{G}_N := \{k^*(K) \notin \mathcal{T}_N\}$,

$$\sqrt{N}\,\hat{\beta}^*(K) = \frac{\sqrt{N}\,M_{k^*}}{C}\,[1 + o_p(1)].$$

Because $\sqrt{N}\,M_k \Rightarrow N(0, \sigma_M^2)$,

$$\mu_{K,N} \;\longrightarrow\; \mu_K := \frac{\sigma_M}{C\sqrt{N_z}}\,\mathrm{med}\Bigl\{\max_{1\leqslant k\leqslant K}|Z_k|\Bigr\}, \quad Z_k \overset{\text{iid}}{\sim} N(0,1).$$

**Step 4 (strict ordering of the limits).** For standard normals, $\mathrm{med}\bigl\{\max_{1\leqslant k\leqslant K}|Z_k|\bigr\}$ is strictly increasing in $K$; thus $\mu_{K+1} > \mu_K$.

**Step 5 (translate back to raw medians).** For large $N$,

$$m_{K+1,N} - m_{K,N} = \frac{\mu_{K+1} - \mu_K}{\sqrt{N}} + o\bigl(N^{-1/2}\bigr) > 0.$$

Combining with (3) yields $\Pr[m_{K+1,N} > m_{K,N}] \to 1$. $\square$