Discussion of “Sensitivity and Informativeness under Local Misspecification”

Jinyong Hahn

April 4, 2019

Sensitivity

- In Hahn and Hausman (2005), we consider a linear IV model

$$y_i = x_i\theta + \varepsilon_i^*$$
$$x_i = z_i'\pi + v_i$$

with local violation of the exclusion restriction

$$\varepsilon_i^* = \frac{1}{\sqrt{n}} z_i' \gamma + \varepsilon_i$$

- The example is (a violation of) the moment problem for $0 = E\left[z_i\left(y_i - x_i\theta\right)\right]$:

$$E\left[z_i\left(y_i - x_i\theta\right)\right] \propto E\left[z_i z_i'\right] \cdot \gamma$$

(I will use 3 different interpretations of 2SLS in this discussion. The GMM interpretation is the first.)

- Linearization aside, it is an identical problem. (You may want to adopt the normalization $E\left[z_i z_i'\right] = I$ if you want to see a 1-1 mapping with the Andrews et al paper.)

Sensitivity

- The asymptotic distribution of the IV estimator using $A' z_i$ as an instrument, i.e.,

$$\widehat{\theta}_A = \left[\frac{1}{n} \sum_{i=1}^{n} \left(A' z_i \right) x_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} \left(A' z_i \right) y_i \right]$$

is asymptotically biased

$$N \left(\underbrace{\left(A' \Phi \pi \right)^{-1} A' \Phi \gamma}_{\text{Bias}}, \sigma_\varepsilon^2 \left(A' \Phi \pi \right)^{-1} \left(A' \Phi A \right) \left(\pi' \Phi A \right)^{-1} \right)$$

where $\Phi = E \left[z_i z_i' \right]$.

Sensitivity

- 2SLS is the special case where $A = \pi$: If $\gamma = 0$, we usually want to minimize $\left(A'\Phi\pi\right)^{-1}\left(A'\Phi A\right)\left(\pi'\Phi A\right)^{-1}$ by choosing $A = \Phi^{-1}\pi$, i.e., 2SLS.
- Hahn and Hausman (2005) consider sensitivity analysis based on

$$\frac{\partial \text{ bias}}{\partial\gamma} = \frac{\partial\left(A'\Phi\pi\right)^{-1}A'\Phi\gamma}{\partial\gamma} = \left(A'\Phi\pi\right)^{-1}A'\Phi,$$

which is in fact identical to the sensitivity in the current paper.
- Hahn and Hausman (2005) go further and propose to minimize $\left\|\partial \text{ bias}/\,\partial\gamma\right\|^2$. Under the normalization $\Phi = I$, it is minimized when $A \propto \pi$, i.e., 2SLS.

Sensitivity

- Although not explicitly stated in Hahn and Hausman (2005), this exercise can be (should have been) based on some (asymptotic) MSE minimization problem. With the normalization $\Phi = I$, the asymptotic distribution is

$$N\left(\left(A'\pi\right)^{-1} A'\gamma, \sigma_\varepsilon^2 \left(A'\pi\right)^{-1} \left(A'A\right) \left(A'\pi\right)^{-1}\right)$$

with the MSE equal to

$$\left(A'\pi\right)^{-1} A'\gamma\gamma' A \left(\pi' A\right)^{-1} + \sigma_\varepsilon^2 \left(A'\pi\right)^{-1} \left(A'A\right) \left(A'\pi\right)^{-1}$$

- We can put weights (prior) on γ such that $E\left[\gamma\gamma'\right] \propto I$, i.e., a spherical distribution, and we would minimize

$$\left(A'\pi\right)^{-1} A'A \left(\pi' A\right)^{-1} + \sigma_\varepsilon^2 \left(A'\pi\right)^{-1} \left(A'A\right) \left(\pi' A\right)^{-1} = C \left(A'\pi\right)^{-1} \left(A'A\right) \left(\pi' A\right)^{-1}$$

with solution equal to $A \propto \pi$.

Sensitivity

- When we depart from the normalization, then we would want to use a different prior (proportional to the inverse of Φ), but it is similar to the various tricks used to 'justify' the χ^2-statistic in multiple hypothesis testing. Prior/weight of convenience.
- In order to relate the asymptotic risk to the limit of finite sample risk, one may want to use the truncation ($\lim_{\zeta\to\infty} \liminf_n E\left[\min\left(n\ell, \zeta\right)\right]$) discussed in Lehmann and Casella (1998), and Hansen (2016, 2017).
- This sort of idea would extend to the heteroscedastic model, i.e., we should expect the usual GMM to be optimal for this alternative purpose. Because everything is based on linearization, nonlinearity is just a minor complication of notation.

Informativeness

- Lemma 1 is Le Cam's Third Lemma. Letting $\widehat{b} = (\widehat{c}, \widehat{\gamma})$, which is assumed to be asymptotically linear with influence function ϕ, it shows what will happen under local deviation characterized by the score s.
- Under the local alternative, the asymptotic bias is $c = E[\phi s]$, and the asymptotic variance remains to be $\Sigma = E[\phi \phi']$.
- Lemma 2 uses the fact that $R^2 \leq 1$ and the assumption $E[s^2] = 1$: because

$$R^2 = \frac{c' \Sigma^{-1} c}{E[s^2]} = c' \Sigma^{-1} c \leq 1,$$

Lemma 2 establishes that Condition 2 is satisfied.

Informativeness

Lemma 2 uses the idea that

$$\begin{bmatrix} \overline{c} \\ \overline{\gamma} \end{bmatrix} = \begin{bmatrix} E\left[\phi_c s_\varphi\right] \\ E\left[\phi_\gamma s_\varphi\right] \end{bmatrix}$$

so the problem

$$\max \overline{c}^2$$

s.t.

$$\begin{bmatrix} \overline{c} & \overline{\gamma} \end{bmatrix} \Sigma^{-1} \begin{bmatrix} \overline{c} \\ \overline{\gamma} \end{bmatrix} \leq 1,$$

can be reinterpreted

$$\max_s \left(E\left[\phi_c s_\varphi\right]\right)^2$$

s.t.

$$E\left[s_\varphi^2\right] = 1, \qquad (\spadesuit)$$

where s denotes the score representing the misspecification. The imposition of $\overline{\gamma} = 0$ is equivalent to the additional constraint

$$E\left[\phi_\gamma s_\varphi\right] = 0. \qquad (\heartsuit)$$

Informativeness

It is proposed to measure the "benefit" of the additional constraint (♡) in terms of the "risk" $\left(E\left[\phi_c s_\varphi\right]\right)^2$, more precisely, on *comparison of maximum risks* under two different sets of constraints.

Informativeness - Observation 1

- The meaning of the constraint/normalization (♠) is unclear to me. Consider the model

$$y_i = x_i c_0 + \varepsilon_i$$
$$x_i = z_i' \gamma_0 + v_i$$

where γ_0 is estimated by the first stage OLS, and c_0 is estimated by the moment

$$E\left[\left(z_i'\gamma_0\right)\left(y_i - x_i c_0\right)\right] = 0 \qquad (\blacktriangle)$$

  - Here, I am viewing c_0 as a parameter identified by the second step moment (▲) using the first step estimate of the reduced form parameter γ_0.
  - In other words, I am now viewing 2SLS as a plug-in estimator.

Informativeness - Observation 1

- The informativeness notion, i.e., Δ, seems to be based on the notion of some sort of trade-off *(for lack of better words, I am just referring to the ellipsoid)* *between the violations* of this second stage moment (▲), i.e., exclusion restriction, and the first stage moment

$$E\left[z_i\left(x_i - z_i'\gamma_0\right)\right] = 0, \qquad (\blacktriangledown)$$

  e.g., measurement error in z_i.
- Why should the degree of violation of exclusion restriction have anything to do with the amount of measurement error? (Why should it depend on the covariance between the two errors?) It does not seem "natural"... If you are like me, you may like a rectangle, not an ellipse, to represent the violations/uncertainty...
- In any case, it is based on the normalization (♠) that $E\left[s_\varphi^2\right] = 1$.

Informativeness - Observation 2

- The paper is a (local) partial identification and comparison of maximum risks, and some care is needed for interpretation.
- It may give the impression that if $\Delta = 1$, and if there is misspecification in $\widehat{\gamma}$, then $\widehat{c}$ should be misspecified.
- For this purpose, it may be useful to consider an alternative representation of the LSEM

$$y_i = z_i \gamma_{20} + u_i$$
$$x_i = z_i \gamma_{10} + v_i$$

with the understanding that

$$c_0 = \frac{\gamma_{20}}{\gamma_{10}}$$

so that $\Delta = 1$ now. (I am now viewing the IV as an indirect least squares estimator.)
- Now, let's introduce a twist and assume that z_i is subject to a (classical) measurement error, under which $\widehat{c}$ is still fine.
- Note that $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ are biased toward zero, so one may be worried because $\Delta = 1$. (To be fair, the paper states that they do not exclude the possibility that $c_0 = c(\eta')$ for some $\eta' \neq \eta_0$.)

Informativeness - Observation 2

Another Analogy

- Consider

$$y_i = x_i\theta + \varepsilon_i$$
$$x_i = z_i'\pi + v_i$$

with

$$0 = E\left[z_i\left(y_i - x_i\theta\right)\right]$$

and the normalization $E\left[z_i z_i'\right] = I_2$, $E\left[\varepsilon_i^2\right] = 1$

- With the local violation $E\left[z_i\left(y_i - x_i\theta\right)\right] \propto \gamma$, the asymptotic bias for the 2SLS is $\left(\pi'\pi\right)^{-1}\pi'\gamma$.
- If we impose the restriction that the second component γ_2 of γ is zero, the asymptotic bias changes to $\left(\pi'\pi\right)^{-1}\pi_1\gamma_1$, so the square of the ratio is

$$\frac{\left(\pi_1\gamma_1\right)^2}{\left(\pi'\pi\right)^2} \Bigg/ \frac{\left(\pi'\gamma\right)^2}{\left(\pi'\pi\right)^2} = \frac{\left(\pi_1\gamma_1\right)^2}{\left(\pi_1\gamma_1 + \pi_2\gamma_2\right)^2}$$

It may be sensible to take the average of the above ratio with respect to some weight on γ to summarize the benefit of the restriction $\gamma_2 = 0$?

Informativeness - Observation 2

Another Analogy

- On the other hand, we can adopt the idea in the paper and consider

$$\max_{\gamma}\left(\left(\pi'\pi\right)^{-1}\pi'\gamma\right)^2$$

s.t.

$$\gamma'\left(E\left[\left(z_i\varepsilon_i\right)\left(z_i\varepsilon_i\right)'\right]\right)^{-1}\gamma=\gamma'\gamma=1$$

where the constrained maximum is

$$\left(\left(\pi'\pi\right)^{-1}\pi'\gamma\right)^2=\frac{\left(\pi'\gamma\right)^2}{\left(\pi'\pi\right)^2}\leq\frac{\left(\pi'\pi\right)\left(\gamma'\gamma\right)}{\left(\pi'\pi\right)^2}=\frac{1}{\pi'\pi}$$

- If we impose an additional constraint that $\gamma_2=0$, while maintaining the normalization $\gamma'\gamma=1$, we get the constrained maximum equal to

$$\frac{\left(\pi_1\gamma_1\right)^2}{\left(\pi'\pi\right)^2}=\frac{\pi_1^2}{\left(\pi'\pi\right)^2}$$

- The ratio of the two maxima is

$$\frac{\pi_1^2}{\left(\pi'\pi\right)^2}\Bigg/\frac{1}{\pi'\pi}=\frac{\pi_1^2}{\pi_1^2+\pi_2^2}$$

Informativeness - Observation 2

Another Analogy

- Is

$$\frac{\pi_1^2}{\pi_1^2 + \pi_2^2}$$

a good summary of

$$\frac{(\pi_1\gamma_1)^2}{(\pi_1\gamma_1 + \pi_2\gamma_2)^2}?$$

Informativeness - Observation 3

- IIA (for lack of better description) may be desired
- Consider for simplicity the case where $\dim(\gamma) = 2$, and we impose the restriction that $\overline{\gamma}_2 = 0$. We can proceed in two different ways (including and excluding γ_1 in the optimization problem), and it would be nice to get the same answer.
- Andrews et al considers an all-or-nothing option on γ by the way, so it only applies to an analysis that does not involve γ_1.

Informativeness - Observation 3

- First, we can recognize the fact that $\dim(\gamma) = 2$, and consider full information calculation: We want to solve

$$\max \overline{c}^2$$

s.t.

$$\begin{bmatrix} \overline{c} & \overline{\gamma}_1 & \overline{\gamma}_2 \end{bmatrix} \Sigma^{-1} \begin{bmatrix} \overline{c} \\ \overline{\gamma}_1 \\ \overline{\gamma}_2 \end{bmatrix} \leq 1,$$

solve the same problem s.t.

$$\begin{bmatrix} \overline{c} & \overline{\gamma}_1 & 0 \end{bmatrix} \Sigma^{-1} \begin{bmatrix} \overline{c} \\ \overline{\gamma}_1 \\ 0 \end{bmatrix} \leq 1, \qquad (\clubsuit)$$

and compare the ratio of the two answers.

- Here, $\overline{\gamma}_1$ plays the role of an "irrelevant alternative", so we can repeat the exercise now deleting $\overline{\gamma}_1$

Informativeness

- Second, we may want to use a limited information calculation: We want to solve

$$\max \overline{c}^2$$

subject to

$$\left[\begin{array}{cc} \overline{c} & \overline{\gamma}_2 \end{array}\right] \widetilde{\Sigma}^{-1} \left[\begin{array}{c} \overline{c} \\ \overline{\gamma}_2 \end{array}\right] \leq 1,$$

solve the same problem subject to

$$\left[\begin{array}{cc} \overline{c} & 0 \end{array}\right] \widetilde{\Sigma}^{-1} \left[\begin{array}{c} \overline{c} \\ 0 \end{array}\right] \leq 1,$$

and compare the ratio of the two answers. Here, $\widetilde{\Sigma}$ denotes the appropriate 2×2 submatrix of Σ.

- The second approach can be solved by the result in the paper, and the ratio is $1 - \Delta$.

Informativeness

- As for the first approach, it seems that the maximum of $\overline{c}^2$ subject to (♣) is

$$\text{Var}\left(\widehat{c}\right) - \frac{\text{Cov}\left(\widehat{c}, \widehat{\gamma}_2\right)^2}{\text{Var}\left(\widehat{\gamma}_2\right)} = \text{Var}\left(\widehat{c}\right) \cdot (1 - \Delta)$$

so the desired IIA holds.