Chapter 4. Instrumental Variables

Joan Llull

Quantitative & Statistical Methods II
Master in Economics of Public Policy
Barcelona School of Economics

I. Identification of causal effects in IV settings

Suppose that $(Y_{1i}, Y_{0i}) \not\perp D_i|X_i$, but we have an exogenous source of variation in D_i so that $(Y_{1i}, Y_{0i}) \perp Z_i|X_i$ that satisfies the relevance condition $Z_i \not\perp D_i|X_i$. In that situation, we can use the variation in Z_i to identify α_{ATE} under certain circumstances. Selection on observables can be regarded as a special case in which $Z_i = D_i$. For simplicity, we do most of the analysis below considering a single binary instrument $Z_i \in \{0, 1\}$, and we abstract from including other covariates. Here, it is crucial to distinguish the two cases we have been discussing so far: homogeneous and heterogeneous treatment effects.

A. Homogeneous treatment effects

Recall that in the homogeneous treatment effects world, the treatment effect is the same for all individuals, $Y_{1i}-Y_{0i} = \beta = \alpha_{ATE} = \alpha_{TT}$ for all individuals. In this case, the availability of an instrumental variable allows us to identify α_{ATE}. This is the traditional situation in econometric models with endogenous explanatory variables (IV regression). In particular, let $Y_i = \beta_0 + \beta D_i + U_i$ as in previous chapters. In this context, $D_i \not\perp U_i$ given that $D_i \not\perp Y_{0i}$. However, we use Z_i as an instrument for D_i in a just-identified fashion. Thus, the IV coefficient is given by:

$$\alpha = \frac{\text{Cov}(Z_i, Y_i)}{\text{Cov}(Z_i, D_i)}. \tag{1}$$

Operating the numerator as in previous chapters we obtain:

$$\begin{aligned}\text{Cov}(Z_i, Y_i) &= \mathbb{E}[Y_i Z_i] - \mathbb{E}[Y_i]\,\mathbb{E}[Z_i] \\ &= \mathbb{E}[Y_i|Z_i = 1]P(Z_i = 1) \\ &\quad - \{\mathbb{E}[Y_i|Z_i = 1]P(Z_i = 1) + \mathbb{E}[Y_i|Z_i = 0](1 - P(Z_i = 1))\}P(Z_i = 1) \\ &= \{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]\}P(Z_i = 1)(1 - P(Z_i = 1)). \end{aligned} \tag{2}$$

Likewise, the denominator is:

$$\text{Cov}(Z_i, D_i) = \{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]\}P(Z_i = 1)(1 - P(Z_i = 1)). \tag{3}$$

Thus, the IV coefficient is:

$$\alpha = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]}, \tag{4}$$

which is known as the ***Wald estimand***. This estimand can also be derived by noting that:

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i = Y_{0i} + \alpha D_i. \tag{5}$$

Since $Y_{0i} \perp\!\!\!\perp Z_i$, then:

$$\left.\begin{array}{l}\mathbb{E}[Y_i|Z_i = 1] = \mathbb{E}[Y_{0i}] + \alpha\,\mathbb{E}[D_i|Z_i = 1] \\ \mathbb{E}[Y_i|Z_i = 0] = \mathbb{E}[Y_{0i}] + \alpha\,\mathbb{E}[D_i|Z_i = 0]\end{array}\right\} \Rightarrow \alpha = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]}. \tag{6}$$

Identification obviously requires that $\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0] \neq 0$, which is the relevance condition. All in all, we are obtaining the effect of D_i on Y_i through the effect of Z_i because Z_i only affects Y_i through D_i (exclusion restriction).

B. Heterogeneous treatment effects

In the heterogeneous case, the availability of instrumental variables is not sufficient to identify a causal effect (e.g. α_{ATE}). An additional assumption that helps in the identification of causal effects is the following ***monotonicity condition***: any person that is willing to treat if assigned to the control group is also willing to treat if assigned to the treatment group. The plausibility of this assumption depends on the context of the application. Under monotonicity, the IV coefficient coincides with the average treatment effect for those individuals whose value of D_i would change when changing the value of Z_i, which is known as the ***local average treatment effect*** (LATE).

The monotonicity condition is well illustrated implementing the potential outcome notation also for the treatment variable. Let D_{0i} denote D_i when $Z_i = 0$, and let D_{1i} denote D_i when $Z_i = 1$ so that $D_{1i}, D_{0i} \perp\!\!\!\perp Z_i$. As we only observe $D_{\ell i}$, for individuals with $Z_i = \ell$, the combination of treatment and instrument define four *observable* groups. However, there are eight *potential* groups, depending on the value of the unobserved treatment status $D_{-\ell}$, which are listed in the following table:

Obs. type	Z	D	D_0	D_1	**Latent type**
Type 1	0	0	0	0	Never-taker
				1	Complier
Type 2	0	1	1	0	Defier
				1	Always-taker

Type 3	1	0	0	0	Never-taker
			1		Defier
Type 4	1	1	0	1	Complier
			1		Always-taker

For example, assume we are interested in the effect of college attendance (treatment) on wages (outcome). Because individuals with higher ability may be more likely to go to college and, for any given educational level, more likely to earn higher wages, independence does not hold neither conditionally nor unconditionally (we do not observe ability). Hence, to be able to identify a causal effect of college attendance on wages, we need an instrument. We consider proximity to a college as an exogenous source of variation: it is associated with the cost of education, but plausibly uncorrelated with later outcomes. To make it dichotomous, we distinguish between being *far* and *close* from school. A complier is an individual that lives close to school and attends, but would not attend if she lived far, or one that does not attend school because she leaves far, but would have attended had she lived close. An individual that goes to school whether she lives close or far is an always-taker, and one that does not go to school whether she lives close or far is a never-taker. Defiers are those individuals that go to school being far, but would not go had they been close, or those who do not go being close, but would have gone had they been far. Monotonicity implies that there are no defiers.

To see that the availability of an instrumental variable is not enough to identify causal effects, consider the second derivation of the treatment effect for the homogeneous effects descried in Equation (6). Now we have:

$$\begin{aligned} \mathbb{E}[Y_i|Z_i = 1] &= \mathbb{E}[Y_{0i}] + \mathbb{E}[(Y_{1i} - Y_{0i})D_{1i}] \\ \mathbb{E}[Y_i|Z_i = 0] &= \mathbb{E}[Y_{0i}] + \mathbb{E}[(Y_{1i} - Y_{0i})D_{0i}], \end{aligned} \tag{7}$$

which implies:

$$\begin{aligned} \mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0] &= \mathbb{E}[(Y_{0i} - Y_{0i})(D_{1i} - D_{0i})] \\ &= \mathbb{E}[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} = 1]P(D_{1i} - D_{0i} = 1) \\ &- \mathbb{E}[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} = -1]P(D_{1i} - D_{0i} = -1). \end{aligned} \tag{8}$$

In this expression, $\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]$ could be negative even if the causal effect is positive for all units, as long as the fraction of defiers, $P(D_{1i} - D_{0i} = -1)$, is sufficiently large. Assuming monotonicity, we avoid this problem: in this case, the second term is zero, and we define $\mathbb{E}[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} = 1]$ as the local average treatment effect (LATE), which is as much as we can identify as we discuss below.

II. Imperfect Compliance and IV

One possibility is to make an even stronger assumption than monotonicity. In particular, assume an *eligibility rule* of the form:

$$P(D_i = 1|Z_i = 0) = 0. \tag{9}$$

This rule implies that individuals with $Z_i = 0$ are denied treatment (observable types 2, 3B, and 4B are ruled out). This situation occurs in the most standard form of imperfect compliance, discussed in Chapter 2: some individuals are assigned to treatment, but they endogenously decide whether to take it or not.

Under this eligibility rule:

$$\begin{aligned}\mathbb{E}[Y_i|Z_i = 1] &= \mathbb{E}[Y_{0i}] + \mathbb{E}[(Y_{1i} - Y_{0i})D_i|Z_i = 1] \\ &= \mathbb{E}[Y_{0i}] + \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1, Z_i = 1]P(D_i = 1|Z_i = 1) \\ &= \mathbb{E}[Y_{0i}] + \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1]P(D_i = 1|Z_i = 1), \end{aligned} \tag{10}$$

where the last equality holds because $D_i = 1$ is a sufficient statistic to indicate that $Z_i = 1$ since $P(D_i = 1|Z_i = 0) = 0$. Likewise:

$$\begin{aligned}\mathbb{E}[Y_i|Z_i = 0] &= \mathbb{E}[Y_{0i}] + \mathbb{E}[(Y_{1i} - Y_{0i})D_i|Z_i = 0] \\ &= \mathbb{E}[Y_{0i}] + \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1, Z_i = 0]P(D_i = 1|Z_i = 0) \\ &= \mathbb{E}[Y_{0i}]. \end{aligned} \tag{11}$$

Thus, we can identify the average treatment effect on the treated in this case, as:

$$\begin{aligned}\alpha_{TT} &= \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1] \\ &= \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{P(D_i = 1|Z_i = 1)} \\ &= \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1]}, \end{aligned} \tag{12}$$

which is the Wald estimand, as $\mathbb{E}[D_i|Z_i = 0] = 0$ by the assumption in (9).

III. Local Average Treatment Effects (LATE)

As discussed above, under monotonicity, Equation (8) reduces to:

$$\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0] = \mathbb{E}[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} = 1]P(D_{1i} - D_{0i} = 1). \tag{13}$$

Also, assuming $D_{1i}, D_{0i} \perp\!\!\!\perp Z_i$ (implying the proportions of compliers, always-takers, and never-takers in the subsample with $Z_i = 1$ coincides with the one in the subsample with $Z_i = 0$) along with monotonicity, we have:

$$\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0] = \mathbb{E}[D_{1i} - D_{0i}] = P(D_{1i} - D_{0i} = 1). \tag{14}$$

Thus, the causal effect that we can identify is:

$$\alpha_{LATE} \equiv \mathbb{E}[Y_{1i} - Y_{0i}|D_{1i} - D_{0i} = 1] = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]}, \tag{15}$$

which is given again by the Wald estimand. Imbens and Angrist (1994) called this parameter a local average treatment effect, because averages treatment effects on the subsample of compliers. Importantly, different instrumental variables lead to different parameters, even under instrument validity, which is counter to standard GMM thinking. This concept changed radically the way we think of and understand IV. As noted, the identified coefficient is the average treatment effect for compliers. Thus, when selecting an instrument, on top of thinking about relevance and orthogonality conditions, the researcher needs to think about the potential group of compliers selected by the instrument.

The most relevant LATEs are those based on instruments that are policy variables. For example, in the college attendance example before, the identified LATE (the effect of schooling for those individuals changing their enrollment based on distance to college) is very relevant for a subsidy policy, even if it is not a good measurement of the average return to education in the whole population.

As a final remark, what happens if there are no compliers? In the absence of defiers, the probability of compliers satisfies $P(D_{1i} - D_{0i} = 1) = \mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]$, so the lack of compliers implies lack of instrument relevance, and, hence, underidentification. This is natural, because if the population is formed of never-takers and always-takers, there is no role to be played by the instrument.

IV. Conditional Estimation with Instrumental Variables

So far we abstracted from the fact that the validity of the instrument may only be conditional on X_i: it may be that $(Y_{1i}, Y_{0i}) \not\perp\!\!\!\perp Z_i$, but the following does:

$$\begin{array}{rl} (Y_{1i}, Y_{0i}) \perp\!\!\!\perp Z_i|X_i & \text{(conditional independence)} \\ Z_i \not\perp\!\!\!\perp D_i|X_i & \text{(conditional relevance)}\ . \end{array} \tag{16}$$

For example, in the analysis of the returns to college, Z_i is an indicator of proximity to college. The problem is that Z_i is not randomly assigned but chosen by parents, and this choice may depend on characteristics that subsequently affect wages. The validity of Z_i may be more credible if we can condition on family background, X_i.

In the linear version of the problem we can estimate using a two-stage procedure: first regress D_i on Z_i and X_i, so that we get $\widehat{D}_i$, and in the second stage we regress Y_i on $\widehat{D}_i$ and X_i. In general, we now have a conditional LATE given X_i:

$$\gamma(X_i) \equiv \mathbb{E}[Y_{1i} - Y_{0i} | D_{1i} - D_{0i} = 1, X_i], \tag{17}$$

and a conditional IV estimator:

$$\beta(X_i) \equiv \frac{\mathbb{E}[Y_i|Z_i = 1, X_i] - \mathbb{E}[Y_i|Z_i = 0, X_i]}{\mathbb{E}[D_i|Z_i = 1, X_i] - \mathbb{E}[D_i|Z_i = 0, X_i]}. \tag{18}$$

To get an aggregate effect, we proceed differently depending on whether the effects are homogeneous or heterogeneous. In the homogeneous case:

$$Y_{1i} - Y_{0i} = \beta(X_i) \quad \forall i. \tag{19}$$

In the heterogeneous case, it makes sense to consider an average treatment effect for the overall subpopulation of compliers:

$$\begin{aligned} \beta_C &\equiv \int \beta(X_i) \frac{P(\mathit{compliers}|X_i)}{P(\mathit{compliers})} dF(X_i) \\ &= \int \left\{ \mathbb{E}[Y_i|Z_i = 1, X_i] - \mathbb{E}[Y_i|Z_i = 0, X_i] \right\} \frac{1}{P(\mathit{compliers})} dF(X_i), \end{aligned} \tag{20}$$

where:

$$P(\mathit{compliers}) = \int \left\{ \mathbb{E}[D_i|Z_i = 1, X_i] - \mathbb{E}[D_i|Z_i = 0, X_i] \right\} dF(X_i). \tag{21}$$

Intuitively, in the top row we use the Bayes' Theorem to rewrite the density of X_i conditional on being a complier. Replacing (21) into (20) yields:

$$\beta_C = \frac{\int \left\{ \mathbb{E}[Y_i|Z_i = 1, X_i] - \mathbb{E}[Y_i|Z_i = 0, X_i] \right\} dF(X_i)}{\int \left\{ \mathbb{E}[D_i|Z_i = 1, X_i] - \mathbb{E}[D_i|Z_i = 0, X_i] \right\} dF(X_i)}, \tag{22}$$

which can be estimated as a ratio of matching estimators (Frölich, 2003).

V. Continuous Instruments: Marginal Treatment Effects (MTE)

When the support of Z_i is not binary, there is a multiplicity of causal effects. Then, the question is which of these causal effects are relevant for evaluating a given policy. The natural experiments literature has been satisfied with identifying "causal effects" in a a broad sense, without paying much attention to their relevance. But the reality is that some causal effects are more informative than others.

If Z_i is continuous, we can define a different LATE parameter for every pair (z, z'):

$$\alpha_{LATE}(z, z') \equiv \frac{\mathbb{E}[Y_i|Z_i = z] - \mathbb{E}[Y_i|Z = z']}{\mathbb{E}[D_i|Z_i = z] - \mathbb{E}[D_i|Z_i = z']}. \tag{23}$$

The multiplicity is even higher when there is more than one instrument. For a general instrument vector Z_i, there are as many potential treatment status indicators D_{zi} as possible values z of the instrument. The IV assumptions become:

$$\begin{array}{ll} (Y_{1i}, Y_{0i}, D_{zi}) \perp\!\!\!\perp Z_i & \text{(independence)} \\ P(D_i = 1 | Z_i = z) \equiv P(z) \text{ is a nontrivial function of } z & \text{(relevance)} \; . \end{array} \tag{24}$$

The monotonicity assumption for general Z_i can be expressed as follows. For any pair of values (z, z'), all units in the population satisfy either:

$$D_{zi} \geq D_{z'i} \text{ or } D_{zi} \leq D_{z'i}. \tag{25}$$

Alternatively, we can use the propensity score $P(Z_i) \equiv P(D_i = 1|Z_i)$ as the instrument (Heckman and Vytlacil, 2005), which is a completely different role for the propensity score than in matching. This is analogous to using $\widehat{D}_i$ as an instrument for D_i in the regression context. Then, the LATE is:

$$\alpha_{LATE}(P(z), P(z')) = \frac{\mathbb{E}[Y_i | P(Z_i) = P(z)] - \mathbb{E}[Y_i | P(Z_i) = P(z')]}{P(z) - P(z')}. \tag{26}$$

If Z_i is binary, this is equivalent to what we had in the first place, but if Z_i is continuous, taking limits as $z \to z'$, we get a limiting form of LATE, which we refer to as ***marginal treatment effect*** (MTE):

$$\alpha_{MTE}(P(z)) = \frac{\partial \, \mathbb{E}[Y_i | P(Z_i) = P(z)]}{\partial P(z)}. \tag{27}$$

Intuitively, $\alpha_{LATE}(P(z), P(z'))$ gives the ATE for individuals who would change schooling status from changing $P(Z_i)$ from $P(z)$ to $P(z')$. In the presence of covariates X_i, Heckman and Vytlacil (2005) suggest to estimate MTE by estimating the derivative of the conditional mean $\mathbb{E}[Y_i | P(Z_i) = P(z), X_i]$ using kernel-based local linear regression techniques.

Similarly, $\alpha_{MTE}(P(z))$ gives the ATE for individuals who would change schooling status following a marginal change in $P(z)$ or, in other words, who are indifferent between schooling choices at $P(Z_i) = P(z)$. Integrating $\alpha_{MTE}(U)$ over different ranges of U we can get other ATE measures. For example:

$$\alpha_{LATE}(P(z), P(z')) = \frac{\int_{P(z')}^{P(z)} \alpha_{MTE}(u) du}{P(z) - P(z')}, \tag{28}$$

and:

$$\alpha_{ATE} = \int_0^1 \alpha_{MTE}(u) du, \tag{29}$$

which makes it clear that to be able to identify α_{ATE} we need identification of $\alpha_{MTE}(u)$ over the entire $(0,1)$ range.

Constructing suitably integrated marginal treatment effects, it may be possible to identify policy relevant treatment effects. If the instrument is an indicator of a policy change, LATE gives the per capita effect of the policy for those induced to change by the policy (e.g. policies that change college fees or distance to school, under the assumption that the policy change affects the probability of participation but not the gain itself).

VI. Some Remarks about Unobserved Heterogeneity in IV Settings

Applied researchers are often concerned about the implications of unobserved heterogeneity. The balance between observed and unobserved heterogeneity depends on how detailed information on agents is available, which ultimately is an empirical issue. The worry for IV-based identification of treatment effects is not heterogeneity *per se*, but the fact that heterogeneous gains may affect program participation. In the absence of an economic model or a clear notional experiment, it is often difficult to interpret IV estimates. Knowing that IV estimates can be interpreted as averages of heterogeneous effects is not very useful if understanding the heterogeneity itself is first order. This is clearly a drawback of the approach.

Heterogeneity of treatments may be also quite important. For example, the literature has found significant differences in returns to different college majors. A problem of aggregating educational categories is that returns are less meaningful. Sometimes education outcomes are aggregated into just two categories, because some techniques are only well developed for binary explanatory variables. A methodological emphasis may offer new opportunities but also impose constraints.

VII. Weak Instruments

Let Z_i be a random variable that satisfies relevance and orthogonality conditions with respect to Y_{1i}, Y_{0i} and D_i in theory, and thus is a candidate for instrument. Assume, however, that the empirical association between Z_i and D_i is weak. While theoretically (and asymptotically) this instrument would eliminate the bias generated by the endogenous assignment of treatment, one should use it with caution, as the weak link between D_i and Z_i introduces a number of complications. In particular, it produces estimates with low precision and it may introduce sample biases in the estimation.

The first problem is apparent in the case in which there is only one regressor D_i and only one instrument Z_i, and errors are homoskedastic (e.g. homogeneous treatment effects). With simple algebra we obtain:

$$\mathrm{Var}(\hat{\beta}^{OLS}) = \frac{1}{N}\frac{\mathrm{Var}(U_i)}{\mathrm{Var}(D_i)} \leq \frac{1}{N}\frac{\mathrm{Var}(U_i)}{\mathrm{Var}(D_i)}\frac{1}{\rho^2_{D_iZ_i}} = \mathrm{Var}(\hat{\beta}^{IV}), \tag{30}$$

where $\rho_{D_iZ_i} \equiv \mathrm{Cov}(D_i, Z_i)/\sqrt{\mathrm{Var}(D_i)\,\mathrm{Var}(Z_i)}$ is the correlation coefficient between D_i and Z_i. This expression indicates that the lower the correlation between Z_i and D_i, the higher is the variance of $\hat{\beta}^{IV}$ (the lower the precision).

Discussing the second problem requires the sampling distributions of $\hat{\beta}^{OLS}$ and $\hat{\beta}^{IV}$, whose derivation is out of the scope of the course. Consider the following model:

$$\begin{aligned} Y_i &= \beta D_i + U_i \\ D_i &= \pi Z_i + V_i. \end{aligned} \tag{31}$$

The IV coefficient is:

$$\hat{\beta}^{IV} = \frac{\sum_{i=1}^N Z_iY_i}{\sum_{i=1}^N Z_iD_i} = \frac{\sum_{i=1}^N Z_i(\beta\pi Z_i + \beta V_i + U_i)}{\sum_{i=1}^N Z_i(\pi Z_i + V_i)} = \beta + \frac{\sum_{i=1}^N Z_iU_i}{\sum_{i=1}^N \pi Z_i^2 + \sum_{i=1}^N Z_iV_i}, \tag{32}$$

and the OLS counterpart is:

$$\hat{\beta}^{OLS} = \frac{\sum_{i=1}^N D_iY_i}{\sum_{i=1}^N D_i^2} = \beta + \frac{\sum_{i=1}^N (\pi Z_i + V_i)U_i}{\sum_{i=1}^N (\pi Z_i + V_i)^2}. \tag{33}$$

The case in which $\pi \to 0$ is illustrative. In that case:

$$(\hat{\beta}^{OLS} - \beta) \to \frac{\sum_{i=1}^N V_iU_i}{\sum_{i=1}^N V_i^2} \equiv \frac{\hat{\sigma}_{UV}}{\hat{\sigma}_V^2}, \tag{34}$$

and:

$$(\hat{\beta}^{IV} - \beta) \to \frac{\sum_{i=1}^N Z_iU_i}{\sum_{i=1}^N Z_iV_i}. \tag{35}$$

Noting that, without loss of generality, we can define $\varepsilon_i \perp\!\!\!\perp V_i$ such that:

$$U_i = \mathbb{E}[U_i|V_i] + \varepsilon_i = \frac{\sigma_{UV}}{\sigma_V^2}V_i + \varepsilon_i, \tag{36}$$

and that both U_i and V_i are zero-mean, Equation (37) can be written as:

$$(\hat{\beta}^{IV} - \beta) \to \frac{\sum_{i=1}^N Z_i\left(\frac{\sigma_{UV}}{\sigma_V^2}V_i + \varepsilon_i\right)}{\sum_{i=1}^N Z_iV_i} = \frac{\sigma_{UV}}{\sigma_V^2} + \frac{\sum_{i=1}^N Z_i\varepsilon_i}{\sum_{i=1}^N Z_iV_i}. \tag{37}$$

Assuming the expectation exists:

$$\mathbb{E}[\hat{\beta}^{IV} - \beta] \to \frac{\sigma_{UV}}{\sigma_V^2} + \mathbb{E}\left[\frac{\sum_{i=1}^N Z_i\varepsilon_i}{\sum_{i=1}^N Z_iV_i}\right] = \frac{\sigma_{UV}}{\sigma_V^2} + \mathbb{E}\left[\frac{\sum_{i=1}^N Z_i\,\mathbb{E}[\varepsilon_i|Z_i, V_i]}{\sum_{i=1}^N Z_iV_i}\right] = \frac{\sigma_{UV}}{\sigma_V^2}. \tag{38}$$

Thus, when the instrument is weak, there is an IV sample bias that tends to the OLS bias as the instrument tends to irrelevant. With more tedious algebra (available in the book) we obtain, for $\pi \neq 0$, the following expression:

$$\mathbb{E}[\hat{\beta}^{IV} - \beta] \approx \frac{\sigma_{UV}}{\sigma_V} \frac{1}{\mathbb{E}[F]}, \tag{39}$$

where $\mathbb{E}[F] \equiv \frac{\mathbb{E}[\pi' Z' Z \pi]/K}{\sigma_V^2} + 1$, which is the expectation of the F-statistic for the test of the hypothesis that the coefficients on the instruments in the first-stage regression are equal to zero (excluding the coefficients of the covariates X_i if they are included in the regression). As the F-statistic is positive by construction, the weak instruments bias is in the same direction as the OLS bias. Also, the bias is smaller than OLS is $\mathbb{E}[F] > 1$. In practice, many researchers use the rule of thumb $F > 10$ to talk about strong instruments, but, in reality, the level of concern about weak instruments depends on numerous factors like the size of the OLS bias to begin with, the number of excluded instruments used in estimation, or the number of observations.