14.382 L2. STRUCTURAL EQUATIONS MODELS AND IV, TAKE 1.

VICTOR CHERNOZHUKOV AND IVÁN FERNÁNDEZ-VAL

ABSTRACT. Here we study a very important linear structural model, with a single endogenous variable and a single instrument, which is frequently used in empirical analysis in economics. We show that we can identify, estimate, and perform inference on the structural parameters of the model using the indirect least squares method. We also provide inferential methods that are robust to weak instruments. The theory relies only upon conventional results for the least squares method. We apply the tools to revisit the analysis of returns-to-schooling and the impact of quality of institutions on economic growth. In the latter case, the use of inference methods robust to weak identification leads to sharper lower bound on the impact of institutions quality than the original empirical results.

NOTATION

Here we use the notation $\perp$ to denote that two random vectors V and W obey

$$\mathrm{E}VW' = 0.$$

In particular, if W includes a constant, this means that V and W are not correlated. The notations $V \perp_S W$ means orthogonality in the sample, namely

$$\mathbb{E}_n VW' = 0.$$

In particular, if W includes a constant, this means that V and W are not correlated in the sample.

1. STRUCTURAL EQUATIONS MODELS

Structural equations models (SEM) specify a collection of functional relations, motivated by economic reasoning, plus shock terms that obey certain orthogonality conditions.

This is an important concept that defines econometrics as a field that is distinct from statistics. Another name for SEMs, frequently used in econometrics, is the simultaneous

equations models. A leading working case, encountered in many modern empirical analyses, is the following linear system of equations:

$$Y = \alpha_1 D + \alpha_2' W + U, \quad U \perp (W', Z')',$$
$$D = \beta_1 Z + \beta_2' W + V, \quad V \perp (W', Z')',$$

(IVM)

where W includes a constant.

We can supply the context to these equations via the Angrist-Krueger analysis of returns-to-schooling: Y is log of wage, D is years of education, W is a set of controls (geographic indicators, year of birth, race, and marital status), V is an education shock, U is a wage shock, and Z is a quarter-of-birth indicator (for whether a person was born in the fourth quarter of the year). The role of Z is unclear at this point, but we will clarify it later. The goal is to learn the parameter α_1, which we interpret as a return-to-schooling parameter. The parameter measures how education influences wages. We shall call such parameters causal or structural.

> We can think of the outcomes (Y, D) as generated in two steps of a Nature's game:
>
>   1. Random variables W, Z, V, U are generated subject to the orthogonality conditions specified above.
>
>   2. The random variables Y and D are jointly determined by the system (IVM).

Accordingly, we call observed variables W and Z the *predetermined* or *exogenous* variables, and we call Y and D as *jointly determined* or *endogenous* variables. We call the unobserved variables U and V as stochastic shocks or error terms. We call Z the instrumental variable or excluded exogenous variable, since it is excluded from one of the equations.

There is a good reason to think that the wage and education shocks U and V are correlated, since, for example, unobserved innate ability could be positively influencing wages and positively influencing the level of educational attainment at the same time. Thus it is reasonable to believe that

$$\mathrm{E}UV \neq 0.$$

This makes the education variable D correlated to U, which in turn implies that we can't identify α_1 as the coefficient of D in the orthogonal projection (i.e. regression) of Y on D and W. Let's analyze this systematically and see if we can still identify α_1 somehow.

In order to focus on the target parameter better, we can get rid of some nuisance parameters first. We do so by partialing out the effect of W, as defined in L1. Application of the partialling out operator to both sides of each of the equations in (IVM) gives us a much

simpler system of equations:

$$\begin{aligned} \tilde{Y} &= \alpha_1 \tilde{D} + U, \quad U \perp \tilde{Z}, \\ \tilde{D} &= \beta_1 \tilde{Z} + V, \quad V \perp \tilde{Z}. \end{aligned} \tag{1.1}$$

Recall that the partialling-out operator is linear, and therefore combines well with orthogonality properties. In particular, the orthogonality conditions of (IVM) still hold in (1.1) after we apply tildes to all variables. The simplicity of the final equations illustrates the beauty of the partialling-out trick: The model we started with is a fully practical (non-toy) model, but after partialling-out, it really becomes a toy model. We can therefore gain some useful insights, which are also fully practical. See also Figure 1 for a graph-theoretic depiction of the model. In modern analysis of SEMs graphs are often used to communicate visually the structure of the equations (albeit it is not clear whether graphs have to offer more than merely an artistic appeal).
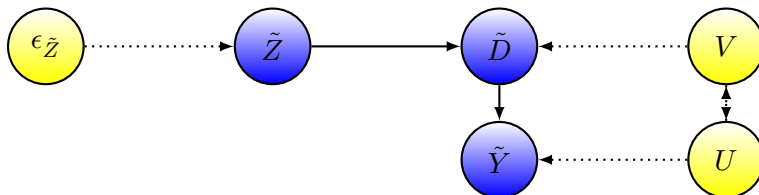

FIGURE 1. Graph-theoretic representation of the IVM: $\tilde{Y} = \alpha_1 \tilde{D} + U$, $\tilde{D} = \beta_1 \tilde{Z} + V$, $\tilde{Z} = \epsilon_{\tilde{Z}}$, after partialling out the effect of W. Random variables $(\tilde{Y}, \tilde{D}, \tilde{Z}, \epsilon_{\tilde{Z}}, U, V)$ are given as nodes or vertices of the graph. Observed nodes are shaded and latent nodes are not. Directed edges represent causal channels. The absence of links between latent nodes signifies the lack of correlation among nodes: $\epsilon_{\tilde{Z}}$ is uncorrelated with U and V.

We begin the analysis with an important negative result, namely that we can't identify α_1 as a projection coefficient of $\tilde{Y}$ on $\tilde{D}$.

Indeed, the projection coefficient is given by:

$$\begin{aligned} \delta_1 = (\mathrm{E}\tilde{D}^2)^{-1}\mathrm{E}\tilde{D}\tilde{Y} &= (\mathrm{E}\tilde{D}^2)^{-1}\mathrm{E}\tilde{D}(\alpha_1\tilde{D} + U) \\ &= \alpha_1 + (\mathrm{E}\tilde{D}^2)^{-1}\mathrm{E}(\beta_1\tilde{Z} + V)U \\ &= \alpha_1 + (\mathrm{E}\tilde{D}^2)^{-1}\mathrm{E}VU \\ &\neq \alpha_1 \quad \text{if } \mathrm{E}VU \neq 0. \end{aligned}$$

Thus, unless $\mathrm{E}VU = 0$, the target parameter α_1 can not be identified from the projection coefficient δ_1. This means we can't use direct least squares to identify and estimate α_1.

So what can we do instead? We now remember that Z creates some fluctuations in D that are uncorrelated to U. We can think of them as *exogenous* or *quasi-experimental* fluctuations. In the AK example, Z is an instrumental variable that is related to compulsory schooling laws: being born in the last quarter, $Z = 1$, forces some students to stay in high-school longer, leading them to acquire more years of schooling; at the same time whether $Z = 1$ or 0 does not seem to affect the wage schedules. Hence, it is reasonable to think that Z appears in the second equation but not in the first equation. In short, there must be some information about α_1 in this set-up. Let's see if we can identify α_1 using somehow this information.

2. Indirect Least Squares and Wright's IV Method

2.1. **Identification.** Here the main insight is to solve for $(\tilde{Y}, \tilde{D})$ in terms of $\tilde{Z}$:

$$\tilde{Y} = \underbrace{\alpha_1\beta_1}_{=:\gamma_1}\tilde{Z} + \underbrace{\alpha_1 V + U}_{=:\varepsilon} =: \gamma_1\tilde{Z} + \varepsilon, \quad \varepsilon \perp \tilde{Z},$$
$$\tilde{D} = \beta_1\tilde{Z} + V, \quad V \perp \tilde{Z}. \tag{2.1}$$

This is the so called *reduced form*: This is a system of linear equations, where only exogenous variables appear on the right side and where the shocks are orthogonal to these variables. Hence the parameters γ_1 and β_1 are simply the regression coefficients: γ_1 is the regression coefficient in the regression of $\tilde{Y}$ on $\tilde{Z}$,

$$\gamma_1 = (\mathrm{E}\tilde{Z}^2)^{-1}\mathrm{E}\tilde{Z}\tilde{Y},$$

and β_1 is the regression coefficient in the regression of $\tilde{D}$ on $\tilde{Z}$,

$$\beta_1 = (\mathrm{E}\tilde{Z}^2)^{-1}\mathrm{E}\tilde{Z}\tilde{D}.$$

Of course, we are talking about population regressions, and we use regression as a synonym for linear projection. We now see a remarkable fact.

Theorem 1 (Identification of α_1 using instrumental variables). *In the IVM model, we can identify the structural parameter α_1 as a ratio of two projection coefficients:*

$$\alpha_1 = \frac{\gamma_1}{\beta_1} = \frac{(\mathrm{E}\tilde{Z}^2)^{-1}\mathrm{E}\tilde{Z}\tilde{Y}}{(\mathrm{E}\tilde{Z}^2)^{-1}\mathrm{E}\tilde{Z}\tilde{D}} = \frac{\mathrm{E}\tilde{Z}\tilde{Y}}{\mathrm{E}\tilde{Z}\tilde{D}}, \quad \textit{if } \beta_1 \neq 0.$$

Thus if $\beta_1 \neq 0$, that is, if "there is a first stage", meaning that $\tilde{Z}$ indeed predicts $\tilde{D}$, then α_1 is identified. Sometimes this method is also called the *indirect least squares method* for identification of α_1. It also leads to an obvious estimation and inference strategy, which we elaborate below. Interestingly, the very last bit of the formula above is the Wright's *instrumental variable (IV) method* for identification of α_1 [8] introduced in 1928. The last bit

of the formula is sometime also called *Wald's IV formula*, which was re-introduced later in 1944, in a different context.

Finally, note also that we can recover α_1 as a regression coefficient in the linear projection of $\gamma_1 \tilde{Z}$, the predicted value of $\tilde{Y}$, on $\beta_1 \tilde{Z}$, the predicted value of $\tilde{D}$:

$$\alpha_1 = (\mathrm{E}(\beta_1 \tilde{Z})^2)^{-1} \mathrm{E}(\gamma_1 \tilde{Z} \beta_1 \tilde{Z}) = \frac{\gamma_1}{\beta_1}.$$

This is the so-called *two-stage least squares* method for identification of α_1. We shall not emphasize this method too much, since it does not extend easily to nonlinear models.

2.2. **Estimation and Inference.** For estimation purposes we assume that we have a sample $\{(Y_i, D_i, Z_i, W_i)\}_{i=1}^n$ of identical copies of (Y, D, Z, W). We assume also that the copies are independent, although the principles outlined below apply more generally to weakly dependent or clustered data.

The above suggests that we can use the following analog estimator for α_1:

$$\hat{\alpha}_1 := \frac{\hat{\gamma}_1}{\hat{\beta}_1}.$$

This is the *instrumental variables (IV) estimator* of α_1. Here we take the least squares estimators $\hat{\gamma}_1$ and $\hat{\beta}_1$:

$$\hat{\gamma}_1 = (\mathbb{E}_n \check{Z}^2)^{-1} \mathbb{E}_n \check{Z} \check{Y},$$

$$\hat{\beta}_1 = (\mathbb{E}_n \check{Z}^2)^{-1} \mathbb{E}_n \check{Z} \check{D}.$$

Recall that "checks" denote the quantities after partialling out the linear effect of W in *the sample*.

Thus $\hat{\alpha}_1$ is a smooth transformation of the least squares estimators and so its consistency and large sample properties follow directly from the properties of the least squares estimators via continuous mapping theorem and the delta method respectively. Let $\theta = (\gamma_1, \beta_1)'$, $\hat{\theta} = (\hat{\gamma}_1, \hat{\beta}_1)'$, $f(\theta) = \gamma_1/\beta_1$, then $\hat{\alpha}_1 = f(\hat{\theta})$. Let $\nabla f(\theta) = \partial f(\theta)/\partial \theta = (1/\beta_1, -\gamma_1/\beta_1^2)'$ provided that $\beta_1 \neq 0$. Then under general conditions the least squares estimators jointly obey

$$\hat{\theta} \stackrel{\mathrm{a}}{\sim} N(\theta, V_\theta/n),$$

and there is a consistent estimator $\hat{V}_\theta$ of V_θ available (e.g. White's estimator for i.i.d. data, Newey-West for time series data, etc.).

By a mean-value expansion, for some $\bar{\theta}$ on the line connecting $\hat{\theta}$ and θ:

$$
\begin{aligned}
\sqrt{n}(\hat{\alpha}_1 - \alpha) &= \nabla f(\bar{\theta})' \sqrt{n}(\hat{\theta} - \theta) \\
&= [\nabla f(\theta) + o_P(1)]' \sqrt{n}(\hat{\theta} - \theta) \\
&\overset{a}{\sim} \nabla f(\theta)' N(0, V_\theta) \\
&= N(0, V_{\alpha_1}), \quad V_{\alpha_1} = \nabla f(\theta)' V_\theta \nabla f(\theta),
\end{aligned}
$$

where we have used the continuous mapping theorem in the second and third step. We then can estimate the asymptotic variance by:

$$
\hat{V}_{\alpha_1} = \nabla f(\hat{\theta})' \hat{V}_\theta \nabla f(\hat{\theta}).
$$

Using this approach we can proceed to construct standard confidence intervals. However, the delta method works poorly when $\beta_1 \approx 0$, when the first stage is weak.

It is also possible to work out the asymptotic distribution of the IV estimator more directly. Note that

$$
\hat{\alpha}_1 = (\mathbb{E}_n \check{Z} \check{D})^{-1} \mathbb{E}_n \check{Z} \check{Y}.
$$

In what follows we assume that $\mathrm{E}\tilde{Z}\tilde{D}$ is bounded away from zero. First, we consider the following infeasible IV estimator:

$$
\tilde{\alpha}_1 = (\mathbb{E}_n \tilde{Z} \tilde{D})^{-1} \mathbb{E}_n \tilde{Z} \tilde{Y}.
$$

This estimator uses quantities with tildes, which are not available in the sample. Notice that

$$
\tilde{\alpha}_1 = (\mathbb{E}_n \tilde{Z} \tilde{D})^{-1} \mathbb{E}_n \tilde{Z} \tilde{D} \alpha_1 + (\mathbb{E}_n \tilde{Z} \tilde{D})^{-1} \mathbb{E}_n \tilde{Z} U = \alpha_1 + (\mathbb{E}_n \tilde{Z} \tilde{D})^{-1} \mathbb{E}_n \tilde{Z} U.
$$

Under general conditions the law of large numbers and central limit theorem give

$$
\mathbb{E}_n \tilde{Z} \tilde{D} - \mathrm{E} \tilde{Z} \tilde{D} \to_P 0, \quad \mathbb{E}_n \tilde{Z} U \overset{a}{\sim} N(0, \Omega/n),
$$

where $\Omega = \mathrm{Var}(\sqrt{n} \mathbb{E}_n \tilde{Z} U)$. Application of the continuous mapping theorem yields that

$$
\sqrt{n}(\tilde{\alpha}_1 - \alpha_1) \overset{a}{\sim} (\mathrm{E} \tilde{Z} \tilde{D})^{-1} N(0, \Omega).
$$

Second, we can show that under general conditions, the following property holds:

$$
\sqrt{n}(\hat{\alpha}_1 - \tilde{\alpha}_1) \to_P 0. \tag{2.2}
$$

In words, the sample-based partialling-out is asymptotically as good as the population-based partialling out. Thus the feasible and infeasible IV estimators are first-order equivalent, which is a very useful observation.

We thus established the following result.

Theorem 2. *Under regularity conditions, the IV estimator obeys*
$$\sqrt{n}(\hat{\alpha}_1 - \alpha_1) \overset{a}{\sim} (E\tilde{Z}\tilde{D})^{-1}N(0,\Omega) = N(0, V_{\alpha_1})$$
$$V_{\alpha_1} = (E\tilde{Z}\tilde{D})^{-1}\Omega(E\tilde{Z}\tilde{D})^{-1}, \quad \Omega = \text{Var}(\sqrt{n}\mathbb{E}_n\tilde{Z}U).$$

Using this result we can build conventional confidence intervals for α_1.

The interested reader may supply a set of regularity conditions and derive (2.2). Here we will skip these details, since the asymptotic theory of the IV estimator will be established rigorously as a part of GMM (generalized method of moments) analysis.

2.3. **Estimation and Inference under Weak Identification.** When $\beta_1 \approx 0$, the delta method may provide a poor approximation to the finite-sample distribution of the IV estimator and the estimator itself starts to behave poorly, because we are dividing by a random quantity that fluctuates dangerously close to zero [7].

Formally, we may call the cases where the correlation between $\tilde{D}$ and $\tilde{Z}$ is close to zero as weakly-identified cases or cases with weak instruments. We shall use the terms weak identification and weak instruments interchangeably in what follows.

Econometricians designed various proposals and rules of thumb for detecting cases with weak instruments.

One rule of thumb states the following: if the F statistic for testing the hypothesis that the regression coefficients in the projection of the endogenous variable on the instruments is smaller than 20, then we deem the instruments as weak; if the F statistic for testing the hypothesis is greater than 20, then there is not a problem and we deem the instruments as strong.

This rule is quite practical but it is not ideal. Perhaps the best way to detect the weak instrument problem is to design a computational experiment that mimics the empirical situation the researcher is facing (for example, fit a parametric model to the reduced form (2.1) for the data at hand and conduct experiments using the fitted model). The experiment would examine the finite-sample behavior of the estimator $\hat{\alpha}_1$ and the coverage properties of the conventional confidence intervals.

Recall that F statistic is just the Wald statistic divided by the number of restrictions being tested. In our case, we test a single restriction so the F statistic is:

$$F = \left(\frac{\hat{\beta}_1}{\text{se}(\hat{\beta}_1)} \right)^2,$$

where $\text{se}(\hat{\beta}_1)$ is the standard error of the estimator $\hat{\beta}_1$. Thus if $F < 20$ we should deem the instruments $\tilde{Z}$ as weak. As a preview for the empirical examples to be given later, we note that in the AK problem, $F \approx 50$, and in the Acemoglu et. al problem, introduced later, $F \approx 11$.

When instruments are weak the IV estimator $\hat{\alpha}_1$ exhibits non-standard behavior which is difficult to work with. So we take a completely different approach, which can be traced back to the ideas of Anderson and Rubin who worked in the 40s [2]:

Instead of estimating the structural parameter α_1 directly, we shall test potential values of this parameter using a test statistic that is well-behaved regardless of whether the identification is strong or weak.

Toward this goal, we observe that we have the model

$$\tilde{Y} - \alpha_1 \tilde{D} = U, \quad U \perp \tilde{Z}.$$

This means that the projection coefficient of $\tilde{Y} - \alpha_1 \tilde{D}$ on $\tilde{Z}$ should be 0. That is we have the regression equation,

$$\tilde{Y} - \alpha_1 \tilde{D} = 0 \cdot \tilde{Z} + U, \quad U \perp \tilde{Z}.$$

Thus, in the population, we can conclude that a value a is not equal to α_1 if the projection coefficient of $\tilde{Y} - a\tilde{D}$ on $\tilde{Z}$ is not 0.

This becomes the basis of our test. We set up a parameter space $\mathcal{A}_1$ for α_1 and then test all values a in $\mathcal{A}_1$ by testing using the Wald statistics whether the projection coefficient of $\tilde{Y} - a\tilde{D}$ on $\tilde{Z}$ is zero or not. The collection of all a's that have passed the test becomes our confidence region for α_1.

Some formal details of this procedure are as follows: for each $a \in \mathcal{A}_1$ we obtain the regression decomposition:

$$\check{Y} - a\check{D} = \hat{\delta}_a \check{Z} + U_a, \quad \check{Z} \perp_S U_a,$$

where $\check{Z} \perp_S U_a$ means orthogonality in the sample, i.e. $\mathbb{E}_n \check{Z} U_a = 0$. Thus, by definition, $\hat{\delta}_a$ is the sample projection coefficient of $\check{Y} - a\check{D}$ on $\check{Z}$. Under general conditions we have that

$$\hat{\delta}_a - \delta_a \overset{a}{\sim} N(0, V_{\delta_a}/n),$$

and we let $\text{se}(\hat{\delta}_a) = \sqrt{\hat{V}_{\delta_a}/n}$ be the standard error, where $\hat{V}_{\delta_a}$ denotes a consistent estimator of V_{δ_a}.

We then formulate the Wald statistic for testing the null hypothesis $\delta_a = 0$:

$$W(a) = \left(\frac{\hat{\delta}_a}{\mathrm{se}(\hat{\delta}_a)} \right)^2.$$

Under the null hypothesis we can conclude that $W(a) \sim \chi^2(1)$ under regularity conditions. Given confidence level $1 - p$, our confidence region for α_1 is

$$CR_{1-p}(\alpha_1) = \{a \in \mathcal{A}_1 : W(a) \leq c_{1-p}\},$$

where c_{1-p} is the $(1-p)$-quantile of the $\chi^2(1)$ distribution. Note that the practical implementation of this procedure requires replacing $\mathcal{A}_1$ by a finite grid of potential values of α_1, with mesh size determined by an economically meaningful tolerance.

We make the following conclusion.

Theorem 3 (Weak-Id Robust Inference). *Under general regularity conditions*

$$W(\alpha_1) \overset{a}{\sim} \chi^2(1),$$

so that the robust confidence region covers α_1 with probability approaching $1 - p$:

$$\mathrm{P}(\alpha_1 \in CR_{1-p}(\alpha_1)) = \mathrm{P}(W(\alpha_1) \leq c_{1-p}) \to \mathrm{P}(\chi^2(1) \leq c_{1-p}) = 1 - p.$$

---

3. METHOD OF MOMENTS

The method of moments is another line of attack on the problem, which conveniently moves us closer to the introduction of the generalized method of moments. Observe that equation (1.1) written as

$$\tilde{Y} - \alpha_1 \tilde{D} = U, \quad U \perp \tilde{Z},$$

is equivalent to the following *moment condition*:

$$\mathrm{E}(\tilde{Y} - \alpha_1 \tilde{D})\tilde{Z} = 0.$$

Provided that $\mathrm{E}\tilde{D}\tilde{Z} \neq 0$, this equation has a unique solution:

$$\alpha_1 = (\mathrm{E}\tilde{D}\tilde{Z})^{-1}\mathrm{E}\tilde{Y}\tilde{Z}.$$

This is the IV formula, which is exactly the same as in Theorem 1.

We can formulate a method of moments estimator $\hat{\alpha}_1$ from the empirical analog of the moment condition above:

$$\mathbb{E}_n(\check{Y} - \hat{\alpha}_1 \check{D})\check{Z} = 0.$$

Solving this equation gives the explicit form of the estimator:

$$\hat{\alpha}_1 = (\mathbb{E}_n\check{D}\check{Z})^{-1}\mathbb{E}_n\check{Y}\check{Z}.$$

This is the IV estimator that we have derived before. Hence the previous characterization of the asymptotic distribution applies here as well.

This approach also admits a weak-id robust approach to inference. We can check if a value $a = \alpha_1$ by checking whether

$$M(a) = \mathrm{E}(\tilde{Y} - a\tilde{D})\tilde{Z} \neq 0.$$

So we formulate

$$\hat{M}(a) = \mathbb{E}_n(\check{Y} - a\check{D})\check{Z}.$$

Assume i.i.d. sampling in what follows for simplicity. Under mild regularity conditions, these partialled-out moments have the adaptivity property:

$$\sqrt{n}(\hat{M}(a) - \tilde{M}(a)) \to_P 0, \tag{3.1}$$

where

$$\tilde{M}(a) = \mathbb{E}_n(\tilde{Y} - \tilde{D})\tilde{Z}.$$

Under i.i.d. sampling

$$\tilde{M}(a) - M(a) \overset{\mathrm{a}}{\sim} N(0, V_a/n),$$

where $V_a = \mathrm{Var}[(\tilde{Y} - a\tilde{D})\tilde{Z}]$, which together with the adaptivity property implies

$$\hat{M}(a) - M(a) \overset{\mathrm{a}}{\sim} N(0, V_a/n).$$

Suppose we have an estimator $\hat{V}_a$ such that $\hat{V}_a/V_a \to_P 1$, for instance, we can take $\hat{V}_a = \mathbb{E}_n[(\check{Y} - a\check{D})^2 \check{Z}^2] - (\mathbb{E}_n(\check{Y} - a\check{D})\check{Z})^2$. Then the Wald statistic takes the form:

$$W(a) = \frac{(\hat{M}(a))^2}{\hat{V}_a/n}, \text{ and } W(\alpha_1) \overset{\mathrm{a}}{\sim} \chi^2(1).$$

These statistics are asymptotically equivalent to the statistics we have considered in the previous subsection. Hence for $c_{1-p} = (1-p)$-quantile of $\chi^2(1)$,

$$CR_{1-p}(\alpha_1) = \{a \in \mathcal{A}_1 : W(a) \leq c_{1-p}\}$$

covers α_1 with probability approaching $1 - p$:

$$\mathrm{P}(\alpha_1 \in CR_{1-p}(\alpha_1)) = \mathrm{P}(W(\alpha_1) \leq c_{1-p}) \to \mathrm{P}(\chi^2(1) \leq c_{1-p}) = 1 - p.$$

This confidence interval is asymptotically equivalent to the confidence interval in the previous subsection.

4. An IV Analysis of Returns to Schooling

We illustrate the estimation and inference in SEMs using the returns-to-schooling application of [3]. The sample comprises $329,509$ observations including males born in 1930–1939 from the 1980 U.S. Census. The model includes the log of weekly earning as Y; the number of years of education as D; an indicator for being born in the fourth quarter of the year as the instrument Z; and a constant, 9 year of birth indicators, 8 region indicators, and indicators for married, black and SMSA as the controls W. According to the rule of thumb the instrument is strong because the first stage F-statistic is 49.65.

TABLE 1. Returns to Schooling in AK data

	Estimate	Std. Error	95% CI
LS	6.32	0.04	(6.25, 6.40)
IV	7.94	2.80	(2.47, 13.42)
Robust	8.08	2.89	(2.41, 13.74)

All the entries are multiplied by 100. Robust uses a mesh
of 201 equidistant points over $\mathcal{A}_1 = [1.28, 14.61]$.
Robust "Estimate" and "Std Error" are the center and
the rescaled width of the robust CI.

Table 1 reports least squares and instrumental variables estimates of the returns-to-schooling coefficient α_1, together with 95% confidence intervals. We construct conventional confidence intervals based on Theorem 2 and robust confidence intervals based on Theorem 3. In the Robust row, the estimate is the center of the CI, and the standard error is the length of the CI divided by 2×1.96, using that 1.96 is the $(1 - \alpha/2)$-quantile of the standard normal distribution with $\alpha = .05$.[1] The Least squares produces an estimate of the return-to-schooling of 6.32%, which is significantly lower than the IV estimate of almost 8%. We refer to [5] for an explanation of the negative bias of least squares based on heterogeneity in the discount rate. The conventional and robust methods produce similar confidence intervals for the return-to-schooling in this application where the identification is strong.

5. An IV Analysis of Impact of Institutions on Growth

We illustrate the difference between conventional and robust-to-weak-identification confidence intervals using the data of [1]. They examined the effect of institutions on economic

---

[1]The logic behind these calculations is that the standard form of a two-sided $(1 - \alpha)$-CI for a parameter θ is $\hat{\theta} \pm c(1 - \alpha)SE(\hat{\theta})$, where $\hat{\theta}$ is an estimate of θ; $c(1 - \alpha)$ is a critical value, typically the $(1 - \alpha/2)$-quantile of the standard normal distribution, and $SE(\hat{\theta})$ is the standard error of $\hat{\theta}$.

performance using mortality rates among European colonists as an instrument for current institutions. The sample consists of 64 countries. We use the same model specification as in column 2 of Table 2 in [4] that includes the the log of PPP adjusted GDP per capita in 1995 as Y; the average protection against expropriation risk from 1985 to 1995, which provides a measure of institutions and well-enforced property rights, as D; the log of the European settler mortality rates as the instrument Z; and a constant and a normalized measure of distance from the equator (latitude) as W. According to the rule of thumb the instrument is weak because the first stage F-statistic is 10.61. This indicates that the conventional intervals might not be reliable in this application where the identification is weak.

TABLE 2. Effect of Institutions on Growth in AJR data

	Estimate	Std. Error	95% CI
LS	0.487	0.064	(0.362, 0.613)
IV	0.969	0.216	(0.547, 1.392)
Robust	1.323	0.334	(0.668, 1.978)

Robust uses a mesh of 251 equidistant points over
$\mathcal{A}_1 = [0.107, 2.262]$.
Robust "Estimate" and "Std Error" are the center and
the rescaled width of the robust CI.

Table 2 reports least squares and instrumental variables estimates of the effect of institutions α_1, together with 95% confidence intervals. The standard errors use the small sample adjustment HC3 of [6]. We construct conventional confidence intervals based on Theorem 2 and robust confidence intervals based on Theorem 3. In the Robust row, the estimate is the center of the CI and the standard error is computed as the length of the CI divided by 2×1.96. While more than 50% narrower, the conventional confidence interval covers values that are below the lower end point of the robust confidence interval. Therefore these results *improve upon* the original empirical investigation of the problem that relied upon strong instrument assumption. We conclude that there is strong evidence that institutions do matter for GDP and that the lower bound on the effect is substantial.

NOTES

APPENDIX A. PROBLEMS

(1) Explain why in the IVM, the structural parameter can not be consistently estimated by the direct least squares method. Explain how the IVM can be estimated via the indirect least squares or the method of moments.

(2) Provide a Monte-Carlo example that illustrates the break-down of normal approximation to the finite-sample distribution of the IV estimator $\hat{\alpha}_1$ when β_1 gets close to zero.

(3) Supply a set of primitive regularity conditions for Theorem 2, e.g. assuming i.i.d. sampling. Derive the adaptivity property (2.2).

(4) Derive the adaptivity property (3). Supply a set of primitive regularity conditions for section 3 and formalize the assertions there in the form of a theorem.

(5) Obtain point estimates, conventional confidence bands, and weak-id robust confidence bands for the AJR problem. Provide a thorough explanation for what you are doing.

(6) Obtain point estimates, conventional confidence bands and weak-id robust confidence bands for the AK problem. Note that the data set is quite big, so computation takes some time. Provide a thorough explanation for what you are doing.

REFERENCES

[1] Daron Acemoglu, Simon Johnson, and James A Robinson. The colonial origins of comparative development: An empirical investigation. *The American Economic Review*, 91(5):1369–1401, 2001.

[2] Theodore W Anderson and Herman Rubin. Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, pages 46–63, 1949.

[3] Joshua D Angrist and Alan B Krueger. Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014, 1991.

[4] Victor Chernozhukov and Christian Hansen. The reduced form: A simple approach to inference with weak instruments. *Economics Letters*, 100(1):68–71, 2008.

[5] Kevin Lang. Ability bias, discount rate bias and the return to education. Technical report, Boston University, 1993.

[6] James G MacKinnon and Halbert White. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305–325, 1985.

[7] D. Staiger and J. H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65:557–586, 1997.

[8] P. Green Wright. *The tariff on animal and vegetable oils*. New York: The Macmillan company, 1928.

14.382 Econometrics
Spring 2017