Instrumental variables

Class Notes

Manuel Arellano

March 8, 2018

1 Introduction

So far we have studied regression models. That is, models for the conditional expectation of one variable given the values of other variables, or linear approximations to those expectations. Now we wish to study relations between random variables that are not regressions. We have already seen some examples: the relationship between y_t and y_{t-1} in an ARMA(1,1) model, or the geometric distributed lag model.

A linear regression model can be seen as a linear relationship between observable and unobservable variables with the property that the regressors are orthogonal to the unobservable term. For example, given two variables (y_i, x_i), the regression of y on x is

$$y_i = \alpha + \beta x_i + u_i \tag{1}$$

where $\beta = Cov\,(y_i, x_i)\,/Var(x_i)$, therefore $Cov\,(x_i, u_i) = 0$.

Similarly, the regression of x on y is:

$$x_i = \gamma + \delta y_i + \varepsilon_i$$

where $\delta = Cov\,(y_i, x_i)\,/Var(y_i)$, and $Cov\,(y_i, \varepsilon_i) = 0$. Solving the latter for y_i we can also write:

$$y_i = \alpha^\dagger + \beta^\dagger x_i + u_i^\dagger \tag{2}$$

with $\alpha^\dagger = -\gamma/\delta$, $\beta^\dagger = 1/\delta$, $u_i^\dagger = -\varepsilon_i/\delta$.

Both (1) and (2) are statistical linear relationships between y and x. If we are interested in some economic relation between y and x, how should we choose between (1) and (2) or none of the two? If the goal is to describe means, clearly we would opt for (1) if interested in the mean of y for given values of x, and we would opt for (2) if interested in the mean of x for given values of y.

In equation (2) $Cov\left(x, u^\dagger\right) \neq 0$ but $Cov\left(y, u^\dagger\right) = 0$ whereas in equation (1) the opposite is true. However, in the ARMA(1,1) model (referred to in the time series class notes) both the left-hand side and the right-hand side variables are correlated with the error term.

To respond a question of this kind we need a prior idea about the nature of the unobservables in the relationship. We first illustrate this situation by considering measurement error models.

2 Measurement error

Consider an exact relationship between the variables y_i^* and x_i^*:

$$y_i^* = \alpha + \beta x_i^*$$

Suppose that we observe x_i^* without error but we observe an error-ridden measure of y_i^* :

$$y_i = y_i^* + v_i$$

where v_i is a zero-mean measurement error independent of x_i^*. Therefore,

$$y_i = \alpha + \beta x_i^* + v_i.$$

In this case β coincides with the slope coefficient in the regression of y_i on x_i^*:

$$\beta = \frac{Cov\left(x_i^*, y_i\right)}{Var\left(x_i^*\right)}$$

Now suppose that we observe y_i^* without error but x_i^* is measured with an error ε_i independent of (y_i^*, x_i^*):

$$x_i = x_i^* + \varepsilon_i.$$

The relation between the observed variables is

$$y_i^* = \alpha + \beta x_i + \zeta_i \tag{3}$$

where $\zeta_i = -\beta\varepsilon_i$. In this case the error is independent of y_i^* but is correlated with x_i. Thus, β coincides with the inverse slope coefficient in the regression of x_i on y_i^*:

$$\beta = \frac{Var\left(y_i^*\right)}{Cov\left(x_i, y_i^*\right)}. \tag{4}$$

In general, inverse regression may make sense if one suspects that the error term in the relationship between y and x is essentially driven by measurement error in x. As it will become clear later (4) can be interpreted as an instrumental-variable parameter in the sense that y_i^* is used as an instrument for x_i in (3). Next, we consider measurement error in regression models as opposed to exact relationships.

2.1 Regression model with measurement error

Measurement error may be the result of conceptual differences between the variable of economic interest and the one available in data, but it could also be the result of rounding errors or misreporting in survey data or administrative records.

Let us consider the regression model

$$y_i^* = \alpha + \beta x_i^* + u_i^*$$

where u_i^* is independent of x_i^*. Below we distinguish two cases: one in wish there is only measurement error in y_i^* and another in which there is only measurement error in x_i^*.

Measurement error in y_i^* We observe $y_i = y_i^* + v_i$ such that $v_i \perp (x_i^*, u_i^*)$. In this case,

$$y_i = \alpha + \beta x_i^* + (u_i^* + v_i),$$

so that

$$\beta = \frac{Cov(x_i^*, y_i^*)}{Var(x_i^*)} = \frac{Cov(x_i^*, y_i)}{Var(x_i^*)}.$$

The only difference with the original regression model is that the variance of the error term is larger due to the measurement error component, which means that the R^2 will be smaller:

$$R_*^2 = \frac{\beta^2 Var(x_i^*)}{\beta^2 Var(x_i^*) + \sigma_u^2}, \qquad R^2 = \frac{\beta^2 Var(x_i^*)}{\beta^2 Var(x_i^*) + \sigma_u^2 + \sigma_v^2},$$

so that the larger σ_v^2 the smaller R^2 will be relative to R_*^2:

$$R^2 = \frac{R_*^2}{1 + \frac{\sigma_v^2}{\beta^2 Var(x_i^*) + \sigma_u^2}}.$$

Measurement error in x_i^* Now $x_i = x_i^* + \varepsilon_i$ such that $\varepsilon_i \perp (x_i^*, u_i^*)$. In this case,

$$y_i^* = \alpha + \beta x_i + (u_i^* - \beta \varepsilon_i).$$

Then

$$\beta = \frac{Cov(x_i, y_i^*)}{Var(x_i)} = \frac{Cov(x_i^*, y_i^*)}{Var(x_i^*) + \sigma_\varepsilon^2} = \frac{\beta}{1 + \frac{\sigma_\varepsilon^2}{Var(x_i^*)}} = \beta - \beta\left(\frac{\lambda}{1+\lambda}\right)$$

where $\lambda = \sigma_\varepsilon^2 / Var(x_i^*)$. Thus, OLS estimates will be biased for β with a bias that depends on the noise to signal ratio λ. For example, if $\lambda = 1$ the regression coefficient will be half the size of the effect of interest.

An example: y_i^* = consumption, x_i^* = permanent income, u_i^* = transitory consumption, ε_i = transitory income.

Identification using λ If we have measurements of λ or σ_ε^2 then consistent estimation may be based on the following expressions:

$$\beta = (1+\lambda)\frac{Cov(x_i, y_i^*)}{Var(x_i)} = \frac{Cov(x_i, y_i^*)}{Var(x_i) - \sigma_\varepsilon^2}. \tag{5}$$

More generally, if x_i is a vector of variables measured with error, so that

$$y_i = x_i'\beta + (u_i - \varepsilon_i'\beta)$$

$$x_i = x_i^* + \varepsilon_i, \qquad E(\varepsilon_i \varepsilon_i') = \Omega,$$

a vector-valued generalization of (5) takes the form:

$$\beta = \left[E(x_i x_i') - \Omega\right]^{-1} E(x_i y_i).$$

3 Instrumental-variable model

3.1 Identification

The set-up is as follows. We observe $\{y_i, x_i, z_i\}_{i=1}^n$ with $\dim(x_i) = k$, $\dim(z_i) = r$ such that

$$y_i = x_i'\beta + u_i \qquad E(z_i u_i) = 0.$$

Typically there will be overlap between variables contained in x_i and z_i, for example a constant term ("control" variables). Variables in x_i that are absent from z_i are endogenous explanatory variables. Variables in z_i that are absent from x_i are external instruments.

The assumption $E(z_i u_i) = 0$ implies that β solves the system of r equations:

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

or

$$E\left(z_i x_i'\right)\beta = E(z_i y_i). \tag{6}$$

If $r < k$, system (6) will have a multiplicity of solutions for β, so that β is not point identified. If $r \geq k$ and $\operatorname{rank} E\left(z_i x_i'\right) = k$ then β is identified. In estimation we will distinguish between the just-identified case ($r = k$) and the over-identified case ($r > k$).

If $r = k$ and the rank condition holds we have

$$\beta = \left[E\left(z_i x_i'\right)\right]^{-1} E(z_i y_i). \tag{7}$$

In the simple case where $x_i = (1, x_{oi})'$, $z_i = (1, z_{oi})'$ and $\beta = (\beta_1, \beta_2)'$ we get

$$\beta_2 = \frac{Cov(z_{oi}, y_i)}{Cov(z_{oi}, x_{oi})}$$

and

$$\beta_1 = E(y_i) - \beta_2 E(x_{oi}).$$

In general, the OLS parameters will differ from the parameters in the instrumental-variable model. In the previous simple example we have:

$$\frac{Cov(x_i, y_i)}{Var(x_i)} = \beta_2 + \frac{Cov(x_i, u_i)}{Var(x_i)}. \tag{8}$$

Sometimes the orthogonality between instruments and error term is expressed in the form of a stronger mean independence assumption instead of lack of correlation:

$$E(u_i \mid z_i) = 0.$$

3.2 Examples

Demand equation In this example the units are markets across space or over time, y_i is quantity, the endogenous explanatory variable is price and the external instrument is a supply shifter, such as weather variation in the case of an agricultural product. This is the classic example from the simultaneous equations literature.[1]

Evaluation of a training program Here the units are workers, the endogenous explanatory variable is an indicator of participation in a training program and y_i is some subsequent labor market outcome, such as wages or employment status. The external instrument is an indicator of random assignment to access to the program. In this example we would expect the coefficient in the instrumental-variable line to be positive, whereas the coefficient in the OLS line could be negative.

Measurement error Consider the measurement error regression model:

$$y_i = \beta_1 + \beta_2 x_i^* + v_i$$

where we observe two measurements of x_i^* with independent errors:

$$\begin{aligned} x_{1i} &= x_i^* + \varepsilon_{1i} \\ x_{2i} &= x_i^* + \varepsilon_{2i}. \end{aligned}$$

All unobservables $\{x_i^*, v_i, \varepsilon_{1i}, \varepsilon_{2i}\}$ are mutually independent. In this example, we could have $x_i = (1, x_{1i})'$, $z_i = (1, x_{2i})'$ and $u_i = v_i - \beta_2 \varepsilon_{1i}$; or alternatively $x_i = (1, x_{2i})'$, $z_i = (1, x_{1i})'$ and $u_i = v_i - \beta_2 \varepsilon_{2i}$.

Time series regression with dynamics and serial correlation A simple example is the ARMA(1,1) model:

$$\begin{aligned} y_t &= \beta_1 + \beta_2 y_{t-1} + u_t \\ u_t &= \varepsilon_t + \theta \varepsilon_{t-1} \end{aligned}$$

where ε_t is a white noise error term. Here $x_t = (1, y_{t-1})'$ and $z_t = (1, y_{t-2})'$.

3.3 Estimation

Simple IV estimator When $r = k$ a simple instrumental-variable estimator is the sample counterpart of (7):

$$\widehat{\beta} = \left(\sum_{i=1}^{n} z_i x_i' \right)^{-1} \sum_{i=1}^{n} z_i y_i.$$

[1] Haavelmo, T. (1943): "The statistical implications of a system of simultaneous equations," *Econometrica*, 11, 1–12.

The estimation error is given by

$$\widehat{\beta}-\beta=\left(\frac{1}{n}\sum_{i=1}^{n}z_i x_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}z_i u_i.$$

Thus, $\text{plim}_{n\to\infty}\widehat{\beta}=\beta$ if $\text{plim}\,\frac{1}{n}\sum_{i=1}^{n}z_i x_i'=E\left(z_i x_i'\right)=H$, $\text{rank}\,H=k$, and $\text{plim}\,\frac{1}{n}\sum_{i=1}^{n}z_i u_i=E\left(z_i u_i\right)=0$.

Also,

$$\sqrt{n}\left(\widehat{\beta}-\beta\right)\xrightarrow{d}\mathcal{N}\left(0,H^{-1}WH'^{-1}\right)$$

if $n^{-1/2}\sum_{i=1}^{n}z_i u_i\xrightarrow{d}\mathcal{N}\left(0,W\right)$. When $\{y_i,x_i,z_i\}_{i=1}^{n}$ is a random sample then $W=E\left(u_i^2 z_i z_i'\right)$.

Overidentified IV If $r>k$ the system (6) contains more equations than unknowns. To determine the population value of β we could solve any rank-preserving k linear combinations for some $k\times r$ matrix G:

$$GE\left(z_i x_i'\right)\beta=GE\left(z_i y_i\right)$$

so that

$$\beta=\left[E\left(Gz_i x_i'\right)\right]^{-1}E\left(Gz_i y_i\right),\tag{9}$$

leading to consistent estimators of the form

$$\widehat{\beta}_G=\left(\sum_{i=1}^{n}Gz_i x_i'\right)^{-1}\sum_{i=1}^{n}Gz_i y_i.\tag{10}$$

Note that while (9) should be invariant to the choice of G if the model is correctly specified, the estimated quantity (10) will differ due to sample error. For example, if $x_i=(1,x_{oi})'$ and $z_i=(1,z_{1i},z_{2i})'$ we will have

$$\frac{Cov\left(z_{1i},y_i\right)}{Cov\left(z_{1i},x_{oi}\right)}=\frac{Cov\left(z_{2i},y_i\right)}{Cov\left(z_{2i},x_{oi}\right)}$$

but

$$\frac{\widehat{Cov}\left(z_{1i},y_i\right)}{\widehat{Cov}\left(z_{1i},x_{oi}\right)}\neq\frac{\widehat{Cov}\left(z_{2i},y_i\right)}{\widehat{Cov}\left(z_{2i},x_{oi}\right)}.$$

Asymptotic normality Turning to large sample properties, repeating the previous asymptotic normality argument for (10), under *iid* sampling we get:

$$\sqrt{n}\left(\widehat{\beta}_G-\beta\right)\xrightarrow{d}\mathcal{N}\left(0,V_G\right)$$

with

$$V_G=\left[GE\left(z_i x_i'\right)\right]^{-1}GE\left(u_i^2 z_i z_i'\right)G'\left[E\left(x_i z_i'\right)G'\right]^{-1}.\tag{11}$$

Thus, the large sample variance depends on the choice of G.

Optimality Let us now consider optimality following Sargan (1958).[2] For $G = E(x_i z_i') \left[E\left(u_i^2 z_i z_i'\right)\right]^{-1}$ the matrix V_G equals

$$V_0 = \left[E\left(x_i z_i'\right)\left[E\left(u_i^2 z_i z_i'\right)\right]^{-1} E\left(z_i x_i'\right)\right]^{-1}.$$

Moreover, it can be shown that for any other choice of G we have:[3]

$$V_G - V_0 \geq 0.$$

Therefore, estimators of the form

$$\widehat{\beta}_{G_n} = \left(\sum_{i=1}^n G_n z_i x_i'\right)^{-1} \sum_{i=1}^n G_n z_i y_i \tag{12}$$

with a possibly stochastic G_n such that $G_n \xrightarrow{p} E(x_i z_i') \left[E\left(u_i^2 z_i z_i'\right)\right]^{-1}$ (up to a multiplicative constant) are optimal in the sense of being minimum asymptotic variance within the class of linear instrumental-variable estimators, which use z_i as instruments.

Under homoskedasticity $E\left(u_i^2 z_i z_i'\right) = \sigma^2 E(z_i z_i')$, therefore a choice of G_n such that

$$G_n \xrightarrow{p} E\left(x_i z_i'\right)\left[E\left(z_i z_i'\right)\right]^{-1} = \Pi$$

is optimal. The matrix Π contains the OLS population coefficients in linear regressions of the x_i variables on z_i.

Two-stage least squares Letting $\widehat{\Pi} = \left(\sum_{i=1}^n x_i z_i'\right)\left(\sum_{i=1}^n z_i z_i'\right)^{-1}$ be the sample counterpart of Π, the two-stage least squares estimator is

$$\widehat{\beta}_{2SLS} = \left(\sum_{i=1}^n \widehat{\Pi} z_i x_i'\right)^{-1} \sum_{i=1}^n \widehat{\Pi} z_i y_i \tag{13}$$

or in short

$$\widehat{\beta}_{2SLS} = \left(\sum_{i=1}^n \widehat{x}_i x_i'\right)^{-1} \sum_{i=1}^n \widehat{x}_i y_i \tag{14}$$

where $\widehat{x}_i = \widehat{\Pi} z_i$ is the vector of fitted values in the ("first-stage") regressions of the x_i variables on z_i:

$$x_i = \Pi z_i + v_i \tag{15}$$

[2] Sargan, J. D. (1958): "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415.

[3] To see this, let $W^{-1} = C'C$, $\overline{H} = CH$, $\overline{D} = (GH)^{-1} GC^{-1}$, and note that:

$$V_G - V_0 = (GH)^{-1} GWG' \left(H'G'\right)^{-1} - \left(H'W^{-1}H\right)^{-1} = \overline{D}\left[I - \overline{H}\left(\overline{H}'\overline{H}\right)^{-1}\overline{H}'\right]\overline{D}'.$$

This is a positive semi-definite matrix because $\left[I - \overline{H}\left(\overline{H}'\overline{H}\right)^{-1}\overline{H}'\right]$ is idempotent. This optimality result also applies to clustered and serially dependent data since it does not require that W equals $E\left(u_i^2 z_i z_i'\right)$.

If a variable in x_i is also contained in z_i its fitted value will coincide with the variable itself and the corresponding element of v_i will be equal to zero.

Sometimes it is convenient to use matrix notation as follows:

$$\widehat{\Pi} = \left(X'Z\right)\left(Z'Z\right)^{-1}$$

so that

$$\widehat{\beta}_{2SLS} = \left[\left(X'Z\right)\left(Z'Z\right)^{-1}\left(Z'X\right)\right]^{-1}\left(X'Z\right)\left(Z'Z\right)^{-1}\left(Z'y\right)$$

and

$$\widehat{\beta}_{2SLS} = \left(\widehat{X}'X\right)^{-1}\widehat{X}'y$$

where $\widehat{X} = Z\left(Z'Z\right)^{-1}\left(Z'X\right)$.

Note that $\widehat{\beta}_{2SLS}$ is also the OLS regression of y on $\widehat{X}$:

$$\widehat{\beta}_{2SLS} = \left(\widehat{X}'\widehat{X}\right)^{-1}\widehat{X}'y.$$

This interpretation of the 2SLS estimator is the one that originated its traditional name.

Two-stage least squares estimation relies on a powerful intuition: we use as instrument the linear combination of the instrumental variables that best predicts the endogenous explanatory variables in the linear projection sense.

Consistency of $\widehat{\beta}_{2SLS}$ relies on $n \to \infty$ for fixed r. Note that if $r = n$ then $\widehat{X} = X$ so that 2SLS and OLS coincide. If r is less than n but close to it, one would expect 2SLS to be close to OLS.

Robust standard errors Although its optimality requires homoskedasticity, 2SLS (like OLS) remains a popular estimator under more general conditions. Particularizing expression (11) to $G = \Pi$ we obtain the asymptotic variance of the 2SLS estimator

$$V_{\Pi} = \left[\Pi E\left(z_i z_i'\right)\Pi'\right]^{-1}\Pi E\left(u_i^2 z_i z_i'\right)\Pi'\left[\Pi E\left(z_i z_i'\right)\Pi'\right]^{-1}. \tag{16}$$

Heteroskedasticity-robust standard errors and confidence intervals can be obtained from the estimated variance:

$$\begin{aligned}\widehat{V}_{\Pi} &= \left[\widehat{\Pi}\widehat{E}\left(z_i z_i'\right)\widehat{\Pi}'\right]^{-1}\widehat{\Pi}\widehat{E}\left(\widehat{u}_i^2 z_i z_i'\right)\widehat{\Pi}'\left[\widehat{\Pi}\widehat{E}\left(z_i z_i'\right)\widehat{\Pi}'\right]^{-1} \\ &= n\left(\widehat{X}'\widehat{X}\right)^{-1}\left(\sum_{i=1}^{n}\widehat{u}_i^2\widehat{x}_i\widehat{x}_i'\right)\left(\widehat{X}'\widehat{X}\right)^{-1}\end{aligned}$$

where the $\widehat{u}_i$ are 2SLS residuals $\widehat{u}_i = y_i - x_i'\widehat{\beta}_{2SLS}$.

With homoskedastic errors, (16) boils down to

$$V_{\Pi} = \sigma^2\left[\Pi E\left(z_i z_i'\right)\Pi'\right]^{-1} \tag{17}$$

where $\sigma^2 = E\left(u_i^2\right)$. In this case a consistent estimator of V_{Π} is simply

$$\widetilde{V}_{\Pi} = n\widehat{\sigma}^2 \left(\widehat{X}'\widehat{X}\right)^{-1} \tag{18}$$

where $\widehat{\sigma}^2 = n^{-1}\sum_{i=1}^{n} \widehat{u}_i^2$.

Note that if the residual variance is calculated from fitted-value residuals $y - \widehat{X}\widehat{\beta}_{2SLS}$ instead of $\widehat{u} = y - X\widehat{\beta}_{2SLS}$, we would get an inconsistent estimate of σ^2 and therefore also of V_{Π} in (17).

3.4 Testing overidentifying restrictions

When $r > k$ an IV estimator sets to zero k linear combinations of the moments:

$$GE\left(z_i x_i'\right)\beta = GE\left(z_i y_i\right)$$

Thus, there remains $r - k$ linearly independent combinations that are not set to zero in estimation but should be close to zero under correct specification. A test of overidentifying restrictions or Sargan test is a test of the null hypothesis that the remaining $r - k$ linear combinations are equal to zero.

Under classical errors the form of the statistic is given by

$$S = \frac{\widehat{u}'Z\left(Z'Z\right)^{-1}Z'\widehat{u}}{\widehat{\sigma}^2} \xrightarrow{d} \chi^2_{r-k} \tag{19}$$

It is easy to see that $S = nR^2$ where R^2 is the r-squared in a regression of $\widehat{u}$ on Z.

A sketch of the result in (19) is as follows. With classical errors $n^{-1/2}\sum_{i=1}^{n} z_i u_i \xrightarrow{d} \mathcal{N}\left[0, \sigma^2 E\left(z_i z_i'\right)\right]$ and therefore also

$$\frac{1}{\sqrt{n}}\frac{1}{\widehat{\sigma}}C'Z'u \xrightarrow{d} \mathcal{N}\left(0, I_r\right)$$

where we are using the factorization $\left(Z'Z/n\right)^{-1} = CC'$.

Next, using

$$\widehat{u} = y - X\widehat{\beta}_{2SLS} = u - X\left(\widehat{\beta}_{2SLS} - \beta\right)$$

and

$$\widehat{\beta}_{2SLS} - \beta = \left[\left(X'Z\right)\left(Z'Z\right)^{-1}\left(Z'X\right)\right]^{-1}\left(X'Z\right)\left(Z'Z\right)^{-1}Z'u,$$

we get

$$h = \frac{1}{\sqrt{n}}\frac{1}{\widehat{\sigma}}C'Z'\widehat{u} = \left[I_r - B\left(B'B\right)^{-1}B'\right]\frac{1}{\sqrt{n}}\frac{1}{\widehat{\sigma}}C'Z'u$$

where $B = C'\left(Z'X/n\right)$.

Since the probability limit of $\left[I - B\left(B'B\right)^{-1}B'\right]$ is idempotent with rank $r - k$ it follows that

$$h'h = n\frac{\widehat{u}'Z\left(Z'Z\right)^{-1}Z'\widehat{u}}{\widehat{u}'\widehat{u}} \xrightarrow{d} \chi^2_{r-k}.$$

In the presence of heteroskedasticity, the statistic S in (19) is not asymptotically chi-square, not even under correct specification. An alternative robust Sargan statistic is:

$$S_R = \left(\widetilde{u}'Z\right) \widetilde{W}^{-1} \left(Z'\widetilde{u}\right) \xrightarrow{d} \chi^2_{r-k} \tag{20}$$

where $\widetilde{W} = \left(\sum_{i=1}^{n} \widehat{u}_i^2 z_i z_i'\right)$ and $\widetilde{u} = y - X\widehat{\beta}_{G_n^\dagger}$ with $G_n^\dagger = (X'Z)\,\widetilde{W}^{-1}$.

Contrary to $\widehat{\beta}_{2SLS}$, the IV estimator $\widehat{\beta}_{G_n^\dagger}$ given by

$$\widehat{\beta}_{G_n^\dagger} = \left[\left(X'Z\right) \widetilde{W}^{-1} \left(Z'X\right)\right]^{-1} \left(X'Z\right) \widetilde{W}^{-1} \left(Z'y\right) \tag{21}$$

uses an optimal choice of G_n under heteroskedasticity. This improved IV estimator was studied by Halbert White in 1982 under the name two-stage instrumental variables (2SIV) estimator.[4]

[4] White, H. (1982): "Instrumental Variables Regression with Independent Observations," *Econometrica*, 50, 483–499.