Instrument Variable Estimation and Generalized Method of Moments

G.J. LI

1 Method of Moments

Previously we consider the linear regression model in the framework of least squares, that is, we try to minimize the distance between the actual observations and the model fitted values of the dependent variable. In this lecture, we will introduce another method of estimation called method of moments. For the linear regression model

$$y_i = x_i'\beta + \epsilon_i, \quad \text{for } i = 1, 2, \ldots, N, \tag{1}$$

where $x_i = (1, x_{1i}, x_{2i}, \ldots, x_{Ki})'$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_K)'$. One of the assumptions we used to derive the asymptotic properties of OLS estimator is

$$E(x_i\epsilon_i) = 0 \tag{2}$$

Note that the first element of x_i is 1. The assumption implies $E(\epsilon_i) = 0$ and $Cov(x_{ki}, \epsilon_i) = 0$ for $k = 1, 2, \ldots, K$. The sample analog of (2) is

$$\begin{aligned} \frac{\sum_{i=1}^{N} x_i \hat{\epsilon}_i}{N} &= 0 \\ \text{or } \frac{X'\hat{\epsilon}}{N} &= 0 \end{aligned} \tag{3}$$

where $\hat{\epsilon} = (\hat{\epsilon}_1, \hat{\epsilon}_2, \ldots, \hat{\epsilon}_N)'$, $y = (y_1, y_2, \ldots, y_N)'$ and $X = (x_1, x_2, \ldots, x_N)'$. Substituting $\hat{\epsilon} = y - X\hat{\beta}$ into the above second equation gives $X'y - X'X\hat{\beta} = 0$ and hence the solution of $\hat{\beta}$ for (3) is $\hat{\beta}_{MoM} = (X'X)^{-1}X'y$, which is the same as the solution from the least squares method. However, the way we find $\hat{\beta}$ now is motivated by the moment condition of (2), rather than minimizing the distance between the actual value and the model fitted value of the dependent variable.

2 Omitted Variable Bias

In reality, due to data unavailability, it is possible that we are not able to find some explanatory variables to include into the regression. For example, consider the study of a person's wage, although we know a person's ability is important to explain wage, we may not be able to find a good measure of person's ability. Suppose we are interested in finding the effect of education on wage. When we run a regression of wage on education only, we cannot control the effect of person's ability. The question is whether the estimated effect of education is still accurate.

We can formulate the problem as follows. Suppose in the true model, all Gauss-Markov Assumptions to ensure unbiasedness are satisfied both X and w are the relevant independent variables to explain y.

$$y = X\beta + s\gamma + \epsilon, \tag{4}$$

where X is the explanatory variable matrix with 1s in the first column. s is another vector of explanatory variable, which we can not observe and $\gamma \neq 0$. We now can only estimate the equation below,

$$y = X\alpha + u. \tag{5}$$

where u is the new unobserved effect and takes account of the effect of s. If the canonical correlation[1] between s and other explanatory variables is not 0, we would expect u to be correlated with some explanatory variables in X. Hence the assumption (2) is not satisfied, that is, $E(x_i u_i) \neq 0$. When the explanatory variables in a regression are correlated with the error term, the explanatory variables are called **endogenous variables**, otherwise they are called **exogenous variable**. The OLS estimator of α in (5) can be represented as

$$\begin{aligned}\hat{\alpha} &= (X'X)^{-1}X'(X\beta + s\gamma + \epsilon),\\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X's\gamma + (X'X)^{-1}X'\epsilon\\ &= \beta + (X'X)^{-1}X's\gamma + (X'X)^{-1}X'\epsilon\\ &= \beta + (\frac{1}{N}\sum_{i=1}^{N} x_i x_i')^{-1}\frac{1}{N}\sum_{i=1}^{N} x_i s_i \gamma + (\frac{1}{N}\sum_{i=1}^{N} x_i x_i')^{-1}\frac{1}{N}\sum_{i=1}^{N} x_i \epsilon_i.\end{aligned} \tag{6}$$

[1]That is the probability limit of R^2 obtained by regressing s on X.

We can see that the estimator of $\hat{\alpha}$ is biased: $E(\hat{\alpha}|X) = \beta + (X'X)^{-1}X's\gamma \neq \beta$ and the slope estimators are not consistent if
$plim\left[(X'X)^{-1}X's\right] = \left[plim(\frac{1}{N}\sum_{i=1}^{N} x_i x_i')\right]^{-1} plim(\frac{1}{N}\sum_{i=1}^{N} x_i s_i) \neq (plim\bar{s}, \underbrace{0, \ldots, 0}_{1\times K})'$
where $\bar{s}$ is the sample average of s.

3 Instrument Variable Estimation

One possible remedy for omitted variable bias is to look for **instrument variables** (IV), which are uncorrelated with the error term, but correlated with the endogenous variables in the regression. Labor economists have used family background variables as IVs for education. For example, mother's education is positively correlated with child's education and may not be correlated with child's ability after s/he grows up. Another IV choice is number of siblings while growing up. Typically, more siblings is associated with lower average levels of education. Note that the choice of IVs is usually through economic reasoning.

Let us formulate the estimation problem as follows. The model we want to estimate is

$$y_i = \alpha_0 + \alpha_1 z_{1i} + \cdots + \alpha_m z_{mi} + \gamma_1 w_{1i} + \cdots + \gamma_n w_{ni} + \epsilon_i, \quad i = 1, \ldots, N, \tag{7}$$

where $w_i = (w_{1i}, w_{2i}, \ldots, w_{ni})'$ are the endogenous variables and $z_i = (z_{1i}, z_{2i}, \ldots, z_{mi})'$ are the exogenous variables. Denote $\underset{(K_x+1)\times 1}{x_i} = (1, z_{1i}, z_{2i}, \ldots, z_{mi}, w_{1i}, w_{2i}, \ldots, w_{ni})'$ with $K_x = m+n$, $\underset{N\times(K_x+1)}{X} = (x_1, x_2, \ldots, x_N)'$ and $\beta = (\alpha_0, \alpha_1, \ldots, \alpha_m, \gamma_1, \ldots, \gamma_n)$.

We can rewrite our model as

$$y_i = x_i'\beta + \epsilon_i, \text{ or } y = X\beta + \epsilon. \tag{8}$$

To estimate β consistently, we have to look for additional IVs apart from the m exogenous regressors already included in the model (which could be IVs as well). The number of additional IVs must be greater than or equal to the number of endogenous variables in the regression. This is called the **order condition**, which is necessary but not sufficient for consistent estimation. Suppose $z_{m+1}, z_{m+2}, \ldots, z_{m+q}$ are the additional IVs. Order condition requires that $q \geq n$. Denote $\underset{(K_z+1)\times 1}{z_i} = (1, z_{1i}, z_{2i}, \ldots, z_{m+q,i})'$ and $\underset{N\times(K_z+1)}{Z} = (z_1, z_2, \ldots, z_N)'$, where $K_z = m + q$. Since instrument variables must be uncorrelated with the error term in (8), we should have

$$E(z_i\epsilon_i) = 0. \tag{9}$$

The sample average is

$$\frac{\sum_{i=1}^N z_i\epsilon_i}{N} = 0, \text{ or } \frac{1}{N}Z'\epsilon = 0 \tag{10}$$

Multiplying both sides by N and substituting $\epsilon = y - X\beta$ yields

$$Z'(y - X\beta) = 0, \text{ or } Z'X\beta = Z'y. \tag{11}$$

The instrument variables should also be closely enough correlated with the endogenous regressors.[2] That is $Z'X$ should have full column rank (K_x+1). In the case of $K_z = K_x$ $(q = n)$, β will be **just identified** and the IV estimator is simply $(Z'X)^{-1}Z'y$.[3] If $K_z > K_x$ $(q > n)$, we will have more equations $(K_z + 1)$ than the number of unknowns (number of elements in β, i.e. $K_x + 1$). We say β is now **over identified**. If all the instruments are valid, we could pick any n out of the q additional instruments to just identify β. But then we are not making full use of all the moment conditions by ignoring the other $q - n$ moment conditions.

4 Generalized Method of Moments

Generalized method of moments (GMM) is an extension of method of moments given the moment condition $E\left[g_i(\theta)\right] = 0$ and the sample average $\bar{g}(\theta) = \frac{1}{N}\sum_{i=1}^{N} g_i(\theta)$, where θ is the collection of parameters in the model (β in the linear regression model). Here $g_i(\theta)$ is a vector function and the subscript i indicates that the function involves the data and its value could change over observations. For (9), $g_i(\theta) = z_i(y_i - x_i'\beta) = z_i\epsilon_i$ is a linear function of β. More generally, $g_i(\theta)$ does not have to be a linear function of θ. Suppose now the dimension of $g(\cdot)$ is greater than the dimension of θ. In other words, θ is over identified just like β in (11) when $K_z > K_x$. Instead of finding the solution that exactly solves the sample average equation, GMM

[2] Further details of the requirements on IV are given in later section.

[3] $X(Z'X)^{-1}Z'$ is called the oblique projector, which is idempotent but not symmetrical. It projects y onto the column space of X along the column space of Z. Note that $I - X(Z'X)^{-1}Z' = Z_\perp(X_\perp' Z_\perp)^{-1}X_\perp'$.

tries to solve the following problem,

$$\min_{\hat{\theta}} \bar{g}(\hat{\theta})' W \bar{g}(\hat{\theta}), \tag{12}$$

where W is some known symmetrical positive semi-definite matrix. For the case of IV, the problem is

$$\min_{\hat{\beta}} \frac{1}{N^2}(y - X\hat{\beta})' ZWZ'(y - X\hat{\beta}), \tag{13}$$

If we want to pick some n out of the q additional instruments to just identify β, we can set W to be a diagonal matrix with relevant ones on the diagonal line. Solving the first order condition gives

$$\hat{\beta}_{GMM} = (X'ZWZ'X)^{-1}X'ZWZ'y. \tag{14}$$

If W is any symmetrical positive semi-definite matrix such that $X'ZWZ'X$ has full rank, $\hat{\beta}$ will be consistent. Now the question is how to choose W. We can choose W such that $\hat{\beta}$ has the smallest asymptotic variance. Suppose Central Limit Theorem works for $\bar{g}(\theta)$: $\sqrt{N}\bar{g}(\theta) \overset{A}{\sim} N(0, V_N)$. If the moment conditions are valid, the following will be true

$$\sqrt{N}(\hat{\theta}_{GMM} - \theta) \overset{A}{\sim} N\left(0, (G_N'WG_N)^{-1}G_N'WV_NWG_N(G_N'WG_N)^{-1}\right). \tag{15}$$

where $G_N = \frac{\partial \bar{g}(\theta)}{\partial \theta'}$. For the IV case, we have $\bar{g}(\theta) = \frac{1}{N}Z'\epsilon = \frac{1}{N}Z'(y - X\beta)$, $G_N = -\frac{1}{N}Z'X$ (does not depend on β) and $V_N = Var(\frac{Z'\epsilon}{\sqrt{N}})$.

Proposition 4.1. *Given*

1. $\{(z_i', x_i', \epsilon_i)\}$ *is an independent sequence;*

2. $E(z_i\epsilon_i) = 0$, *for* $i = 1, 2, \ldots, N$;

3. $E(|z_{hi}\epsilon_i|^{2+\delta}) < \infty$ *for some* $\delta > 0$, *all* $i = 1, 2, \ldots, N$ *and* $h = 1, 2, \ldots, m+q$;

4. $V_N = Var\left(\frac{Z'\epsilon}{\sqrt{N}}\right) = E\left(\frac{\sum_{i=1}^{N} \epsilon_i^2 z_i z_i'}{N}\right)$ *is uniformly positive definite;*

5. $E(|z_{hi}x_{hi}|^{1+\delta}) < \infty$ *for some* $\delta > 0$, *all* $i = 1, 2, \ldots, N$ *and* $h = 1, 2, \ldots, m+q$;

6. $Q_N = E(\frac{1}{N}Z'X)$ *has uniformly full column rank;*

7. $W = O(1)$ *is symmetrical and uniformly positive definite.*

Then $\sqrt{N}(\hat{\beta}_{GMM} - \beta) \overset{A}{\sim} N\left(0, (Q_N'WQ_N)^{-1}Q_N'WV_NWQ_N(Q_N'WQ_N)^{-1}\right)$.

It could be shown that if W is chosen to be cV_N^{-1}, where c is some constant, the asymptotic variance of $\hat{\beta}_{GMM}$ will be the smallest among the estimators in the form of (14) and will be equal to $(Q_N'V_N^{-1}Q_N)^{-1}$.

Lemma 4.2. *Let* A *and* B *be positive definite matrices. Then* $A - B$ *is positive semidefinite if and only if* $B^{-1} - A^{-1}$ *is positive semidefinite.*

Using the above Lemma, we can see that

$$
\begin{aligned}
& Q_N'V_N^{-1}Q_N - [(Q_N'WQ_N)^{-1}Q_N'WV_NWQ_N(Q_N'WQ_N)^{-1}]^{-1} \\
&= Q_N'V_N^{-\frac{1}{2}}\left[I - V_N^{\frac{1}{2}}WQ_N(Q_N'WV_NWQ_N)^{-1}Q_N'WV_N^{\frac{1}{2}}\right]V_N^{-\frac{1}{2}}Q_N \\
&= Q_N'V_N^{-\frac{1}{2}}\left[I - D_N(D_N'D_N)^{-1}D_N'\right]V_N^{-\frac{1}{2}}Q_N
\end{aligned}
$$

where $D_N = V_N^{\frac{1}{2}} W Q_N$. The above is a quadratic form in an idempotent matrix and is therefore positive semidefinite.

Since the opimum choice of W is cV_N^{-1}, in the following discussion, we will choose W to be $\hat{V}_N^{-1}$, which a consistent estimator of V_N.

5 Two and Three Stage Least Squares

Under homoskedasticity, we can estimate V_N by $\hat{\sigma^2}\frac{Z'Z}{N}$ and replace W in (14) by $(\frac{Z'Z}{N})^{-1}$ to obtain $\hat{\beta}_{2SLS} = (X'P_Z X)^{-1}X'P_Z y$,[4] which is the two stage least squares estimator (2SLS) obtained by first regressing X on Z to get the fitted value $P_Z X$ and then regressing y on $P_Z X$ to get the estimator for β. From Proposition 4.1, $\sqrt{N}(\hat{\beta}_{2SLS} - \beta) \overset{A}{\sim} N\left(0, \sigma^2(Q_N' S_N^{-1} Q_N)^{-1}\right)$, where $S_N = E(\frac{1}{N}\sum_{i=1}^N z_i z_i')$. In practice, the asymptotic variance of $\hat{\beta}_{2SLS}$ is estimated by $\hat{\sigma^2}\left[\frac{X'Z(Z'Z)^{-1}Z'X}{N}\right]^{-1}$. In comparison to the estimated asymptotic variance of $\hat{\beta}_{OLS}$, it is no smaller than that of OLS. (Why?) Quite often the standard errors associated with 2SLS are bigger than those of OLS, making the t statistics less significant under homoskedasticity and independent data. In general, one important rule of thumb in econometrics is that there is a trade-off between consistency (unbiasedness) and efficiency for many situations. Though 2SLS is consistent, it may suffer from loss of efficiency comparing to OLS. As we see later, it is possible to reduce the asymptotic variance of 2SLS by including additional valid IVs, but at the same time we could increase the finite sample bias of 2SLS. If heteroskedasticity exists,

[4]Since we require $W = O(1)$ to be uniformly positive definite, we at the same time require $E(|z_{hi}^2|^{1+\delta}) < \infty$, for some $\delta > 0$, $h = 1, \ldots, m+q$ and $i = 1, 2, \ldots, N$; and $E(\frac{1}{N}\sum_{i=1}^N z_i z_i')$ to be uniformly positive definite.

we could improve the efficiency by using three stage least squares (3SLS). That is we first estimate β by 2SLS and then estimate V_N from the 2SLS residuals before estimating β again by 14.

Proposition 5.1. *Partition $Z = (Z_1, Z_2)$ and assume the conditions in Proposition 4.1 hold for both Z_1 and Z_2. Define $V_{1N} = Var\left(\frac{Z_1'\epsilon}{\sqrt{N}}\right)$, $\tilde{\beta} = (X'Z_1\hat{V}_{1N}^{-1}Z_1'X)^{-1}X'Z_1\hat{V}_{1N}^{-1}Z_1'y$, $\hat{\beta} = (X'Z\hat{V}_N^{-1}Z'X)^{-1}X'Z\hat{V}_N^{-1}Z'y$ and $AVar(\cdot)$ as the asymptotic variance of an estimator. Then $AVar(\tilde{\beta}) - AVar(\hat{\beta})$ is a positive semidefinite matrix for all N sufficiently large.*

Proof. By Proposition 4.1, we know $AVar(\hat{\beta}) = (Q_N'V_N^{-1}Q_N)^{-1}$ and $AVar(\tilde{\beta}) = (Q_{1N}'V_{1N}^{-1}Q_{1N})^{-1}$, where $Q_{1N} = E(\frac{Z_1'X}{N})$, $Q_{2N} = E(\frac{Z_2'X}{N})$ and $V_N = \begin{bmatrix} V_{1N} & V_{12N} \\ V_{21N} & V_{2N} \end{bmatrix}$. Using the partitioned inverse formula in footnote 2 of Lecture 2 gives,

$$
\begin{aligned}
& [AVar(\hat{\beta})]^{-1} - [AVar(\tilde{\beta})]^{-1} \\
=& (Q_{1N}', Q_{2N}')V_N^{-1}(Q_{1N}', Q_{2N}')' - Q_{1N}'V_{1N}^{-1}Q_{1N}, \\
=& Q_{1N}'(V_{1N}^{-1} + V_{1N}^{-1}V_{12N}E_N^{-1}V_{21N}V_{1N}^{-1})Q_{1N} \\
& - Q_{2N}'E_N^{-1}V_{21N}V_{1N}^{-1}Q_{1N} - Q_{1N}'V_{1N}^{-1}V_{12N}E_N^{-1}Q_{2N} \\
& + Q_{2N}'E_N^{-1}Q_{2N} - Q_{1N}'V_{1N}^{-1}Q_{1N} \\
=& (V_{21N}V_{1N}^{-1}Q_{1N} - Q_{2N})'E_N^{-1}(V_{21N}V_{1N}^{-1}Q_{1N} - Q_{2N}),
\end{aligned}
\tag{16}
$$

where $E_N = V_{2N} - V_{21N}V_{1N}^{-1}V_{12N}$ is a symmetrical positive definite matrix. Hence $[AVar(\hat{\beta})]^{-1} - [AVar(\tilde{\beta})]^{-1}$ is positive semidefinite and by Lemma 4.2 $AVar(\tilde{\beta}) - AVar(\hat{\beta})$ is a positive semidefinite matrix. $\square$

6 Validation of IV Estimation

We use IV estimation because there are endogenous variables in the regression. We can verify such endogeneity by Hausman specification test. If the regressors are not endogenous, both IV and OLS should be consistent while OLS is more efficient under homoskedasticity. In what follows, we suppose OLS estimator is more efficient. The Hausman specification statistic is

$$N(\hat{\beta}_{IV} - \hat{\beta}_{OLS})' \widehat{AVar}(\hat{\beta}_{IV} - \hat{\beta}_{OLS})^{-1} (\hat{\beta}_{IV} - \hat{\beta}_{OLS}) \tag{17}$$

where $AVar(\hat{\beta}_{IV} - \hat{\beta}_{OLS}) = AVar(\hat{\beta}_{IV}) + AVar(\hat{\beta}_{OLS}) - ACov(\hat{\beta}_{IV}, \hat{\beta}_{OLS}) - ACov(\hat{\beta}_{OLS}, \hat{\beta}_{IV})$. Hausman showed that $ACov(\hat{\beta}_{OLS}, \hat{\beta}_{IV}) = ACov(\hat{\beta}_{IV}, \hat{\beta}_{OLS}) = AVar(\hat{\beta}_{OLS})$. Hence the statistic can be rewritten as

$$N(\hat{\beta}_{IV} - \hat{\beta}_{OLS})' \left[\widehat{AVar(\hat{\beta}_{IV})} - \widehat{Avar(\hat{\beta}_{OLS})} \right]^{-1} (\hat{\beta}_{IV} - \hat{\beta}_{OLS}). \tag{18}$$

The null hypothesis is $H_0 : \hat{\beta}_{IV} = \hat{\beta}_{OLS}$ or $E(x_i \epsilon_i) = 0$. Note that Hausman statistic follows a chi-squared distribution with degrees of freedom equal to the number of endogenous regressors since not all the explanatory variables are suspicious of being endogenous. Therefore, $\widehat{AVar(\hat{\beta}_{IV})} - \widehat{Avar(\hat{\beta}_{OLS})}$ will not have full rank unless the number of endogenous regressors is the total number of regressors. To calculate (18) involves finding Moore-Penrose inverse, see e.g. Magnus and Neudecker (2007), Theorem 1.4. An asymptotically equivalent, but simpler way is to test whether the estimated residuals obtained from regressing the endogenous variables on all the exogenous instruments are jointly significant in the following regression:

$$y = X\beta + M_Z W\gamma + \epsilon, \tag{19}$$

estimated residual

where W is a collection of the suspected endogenous regressors in the model. Hausman specification test is now performed on the joint significance of γ. Note that the OLS estimator for β from (19) is the same as the 2SLS estimator. We could also replace $M_Z W$ by $P_Z W$, which will not affect the test.

One requirement for IV is that IV must be closely enough correlated with the endogenous regressors, which means we need to check whether $Z'X$ indeed has full column rank. In practice, we can first regress each endogenous regressor on Z and check the joint significance of the additional IVs to have a rough idea of whether the requirement could be satisfied.

Another requirement of IV is that IV must not be correlated with the error term. That is we need to test

$$H_0 : E(z_i \epsilon_i) = 0, \tag{20}$$

which is called overidentification test. Under the conditions in Proposition 4.1, $\frac{Z'\epsilon}{\sqrt{N}} \overset{A}{\sim} N(0, V_N)$ and therefore $\frac{\epsilon' Z V_N^{-1} Z'\epsilon}{N} \sim \chi^2_{K_z+1}$. In practice, we have to estimate ϵ by first estimating the $K_x + 1$ parameters in the model. The overidentification statistic below has degrees of freedom $K_z - K_x$ $(q - n)$. Hence (20) is only testatble if $K_z > K_x$ or $q > n$ (numbder of additional IVs greater than number of endogenous regressors).

overidentification test

$$\frac{(y - X\hat{\beta}_{IV})'Z\hat{V}_N^{-1}Z'(y - X\hat{\beta}_{IV})}{N} = \frac{\hat{\epsilon}'_{IV}Z\hat{V}_N^{-1}Z'\hat{\epsilon}_{IV}}{N} \tag{21}$$

In the case of homoskedasticity, we can estimate V_N by $\hat{\sigma^2}_{IV}\frac{Z'Z}{N}$, where $\hat{\sigma^2}_{IV} = \frac{\hat{\epsilon}'_{IV}\hat{\epsilon}_{IV}}{N}$. The overidentification statistic can be calculated as $N\frac{\hat{\epsilon}'_{IV}Z(Z'Z)^{-1}Z'\hat{\epsilon}_{IV}}{\hat{\epsilon}'_{IV}\hat{\epsilon}_{IV}} = NR^2_*$, where R^2_* is obtained by regressing $\hat{\epsilon}_{IV}$ on Z.

7 Hypothesis Testing

Consider testing linear restrictions: $R\beta = r$, where R is a $l \times (K_x+1)$ matrix with rank equal to $l \leq (K_x + 1)$ and r is a $l \times 1$ vector. We know what R and r are. For the restricted model,

$$\min_{\tilde{\beta}}(y - X\tilde{\beta})'Z\bar{V}_N^{-1}Z'(y - X\tilde{\beta}), \qquad s.t. \quad R\tilde{\beta} = r. \tag{22}$$

where $\bar{V}_N$ is some consistent estimator for V_N. The first order conditions of the Lagrangian are

$$\frac{\partial \mathcal{L}}{\partial \tilde{\beta}} = 2(X'Z\bar{V}_N^{-1}Z'X\tilde{\beta} - X'Z\bar{V}_N^{-1}Z'y) + R'\lambda = 0, \tag{23}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = R\tilde{\beta} - r = 0. \tag{24}$$

We can obtain the solutions as follows,

$$
\begin{aligned}
\tilde{\lambda} &= 2[R(X'Z\bar{V}_N^{-1}Z'X)^{-1}R']^{-1}(R\hat{\beta}-r) \\
&= 2[R(X'Z\bar{V}_N^{-1}Z'X)^{-1}R']^{-1}R(\hat{\beta}-\tilde{\beta}), \qquad (25)
\end{aligned}
$$

tuda->restricted

$$\tilde{\beta} = \hat{\beta}-(X'Z\bar{V}_N^{-1}Z'X)^{-1}R'[R(X'Z\bar{V}_N^{-1}Z'X)^{-1}R']^{-1}(R\hat{\beta}-r), \qquad (26)$$

hat->unrestricted

$$\hat{\beta} = (X'Z\bar{V}_N^{-1}Z'X)^{-1}X'Z\bar{V}_N^{-1}Z'y. \qquad (27)$$

Note that $R\tilde{\beta}=r$. Unlike the unrestricted OLS estimator, $X'(y-X\hat{\beta})\neq 0$ for the IV estimator and the R-squared could be outside $[0,1]$. Moreover, we should not use SSR to construct F statistic and should use Wald and LM for hypothesis tests.

Proposition 7.1. *$AVar(\hat{\beta})-AVar(\tilde{\beta}) = AVar(\hat{\beta})R'(R\cdot AVar(\hat{\beta})\cdot R')^{-1}R\cdot AVar(\hat{\beta})$, which is a positive semindefinite matrix.*

This result guarantees that imposing correct a priori restrictions leads to an efficiency improvement over the efficient IV estimator without these restrictions. Interestingly, imposing the restrictions with an inefficient IV estimator (W is not chosen to be $\bar{V}_N^{-1}$ in (22)) may or may not lead to efficiency gains relative to the inefficient IV estimator.

Proposition 7.2. *Let the conditions in Proposition 4.1 hold. Then under $H_0: R\beta = r$,*

1. $\sqrt{N}(R\hat{\beta}-r)\overset{A}{\sim} N\left(0, R(Q_N'V_N^{-1}Q_N)^{-1}R'\right)$;

2. *The Wald statistic is defined as* $\frac{1}{N}(R\hat{\beta}-r)'[R(X'Z\hat{V}_N^{-1}Z'X)^{-1}R']^{-1}(R\hat{\beta}-r)\overset{A}{\sim}\chi^2_l$, *where $\hat{V}_N$ is based on the estimated residuals from the unrestricted model.*

Proposition 7.3. *Let the conditions in Proposition 4.1 hold. Then under $H_0 : R\beta = r$,*

1. $\frac{1}{\sqrt{N}}\tilde{\lambda} \stackrel{A}{\sim} N\left(0, 4[R(Q_N' V_N^{-1} Q_N)^{-1} R']^{-1}\right)$;

2. *The Lagrange multiplier test statistic is defined as*

$$\frac{1}{N}(R\hat{\beta} - r)'[R(X'Z\tilde{V}_N^{-1}Z'X)^{-1}R']^{-1}(R\hat{\beta} - r),$$

which follows a chi-squared distribution with degrees of freedom l with $\tilde{V}_N$ calculated based on the estimated residuals from the restricted model.

References

[1] W.H. Greene, *Econometric Analysis*, 7th ed., Pearson Education, Chapter 12, 2011.

[2] J.R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3 ed., John Wiley & Sons Ltd., 2007.

[3] H. White, *Asymptotic Theory for Econometricians*, Chapter 4.3, Prentice Hall, 2001.