Linear model with instrumental variables

Usually economic theory prescribes random vectors whose (conditional on some information set, or unconditional) expectation is zero iff it is formed with respect to the relevant unknown distribution. (Identification Condition). This, defines a semi-parametric statistical model with enough structure so as at least one criterion function can be constructed, upon the optimization of which, procedures of statistical inference are designed.

A simple example, is the Instrumental Variables estimator in the context of the semiparametric linear model, when the orthogonality conditions involve the residuals and a matrix of instrumental variables.

Consider the process $(y_t)_{t\in\mathbb{Z}}$ defined by $y_t = \underset{(1\times q)}{x_t}\theta_0 + \epsilon_t$, $\theta_0 \in \mathbb{R}^q$, and a vector $\underset{(1\times \ell)}{z_t}$ for which we know that $E(\epsilon/z) = 0$, $V(\epsilon/z) = I$, where $(\epsilon_t)_{t\in\mathbb{Z}}$ is a strictly stationary and ergodic white noise process.

Given a sample $(y_t, x_t, z_t)_{t=1,\dots,T}$, suppose that our interest centers on estimating the unknown vector of parameters θ_0.

(In many time series models the set of valid instruments to choose from is typically infinite spreading into an infinite past of an instrumental variable and possibly embracing their functions)

Given that $E(\epsilon/z) = 0$ and $\epsilon = Y - X\theta_0$, from the structure of our model we have that:

$M_T : \Omega \times \Theta \to \mathbb{R}^\ell$, $M_T(\omega,\theta) = \frac{1}{T} z'(Y - X\theta)$

and

$$E_{\theta_0}(M_T(\omega,\theta_0)/z) = E\left[\frac{1}{T} z'(Y - X\theta_0)/z\right] = \frac{1}{T} z' \overset{0}{E(\epsilon/z)} = 0_{\ell\times 1}$$

Also,

$$E_{\theta_0}(M_T(\omega,\theta)/z) = E_{\theta_0}\left(\frac{1}{T} z'(Y - X\theta)/z\right) = E\left(\frac{1}{T} z'(X\theta_0 + \epsilon - X\theta)/z\right)$$

$$= \frac{1}{T} Z' E(X(b_0 - b)/Z) + \frac{1}{T} Z' E(\epsilon/Z)$$

$$= \frac{1}{T} \underset{\ell\times T}{Z'} \underset{T\times q}{E(X/Z)} \underset{q\times 1}{(b_0 - b)} : \Theta \to \mathbb{R}^{\ell}$$

which is equal to $0_{\ell\times 1}$ if and only if $b = b_0$.

In order to ensure that the above linear relation is 1-1, it must hold:

$\operatorname{rank} Z' E(X/Z) = q \Leftrightarrow \{ \ell \geq q, \operatorname{rank} Z \geq q, \operatorname{rank} E(X/Z) = q \}$

Therefore,

Identifinication Condition: $b = b_0 \Leftrightarrow E_{b_0}(M_T(\omega, b)) = 0_{\ell\times 1}$

The aforementioned is equivallent to $\| E_{b_0} M_T(\omega, b) \| = 0$ iff $b = b_0$.

Thus, given the IC along with the analogy principle and the properties of norms on vector spaces, we can define the objective function:

$$Q_T(\omega, b) = \| M_T(\omega, b) \|^2_{W_T} = \left(\frac{1}{T} Z'(Y - Xb)\right)' W_T \left(\frac{1}{n} Z'(Y - Xb)\right)$$

where $W_T : \Omega \times \Theta \to \mathbb{R}^{\ell\times\ell}$ is a weighting matrix which is appropriately measurable and $\forall b$ $W_T(\omega, b) : \Omega \to \mathbb{R}^{\ell\times\ell}$ is a P_{b_0} - almost surely $\forall \omega$ positive definite

Hence, the IV estimator is denoted by:

$$b_T = \arg\min_b \left[\left(\frac{1}{T} Z'(Y - Xb)\right)' W_T \left(\frac{1}{T} Z'(Y - Xb)\right) \right]$$

Assumption 1: $\Theta = \mathbb{R}^q$ ($\Rightarrow$ we can obtain an analytical expression of the estimator)

F.O.C.

$$\frac{\partial Q_T(\omega, b)}{\partial b} = 0_{q\times 1} \Rightarrow -2 \frac{X'Z}{T} W_T \frac{1}{T}(Z'Y - Z'Xb) = 0 \Leftrightarrow X'Z W_T (Z'Y - Z'Xb) = 0 \Leftrightarrow$$

$$(X'Z) W_T (Z'Y) = \underbrace{(X'Z) W_T (Z'X)}_{A} b \Leftrightarrow b_T = \left[(X'Z) W_T (Z'X) \right]^{-1} (X'Z) W_T (Z'Y)$$

since A is non-singular given the identification condition (rank $X = q$, rank $Z \geq q$) and the fact that W_T is positive definite.

S.O.C. :

$$\frac{\partial^2 Q_T(w,\beta)}{\partial\beta\partial\beta'} = 2\frac{X'Z}{T} W_T \frac{Z'X}{T} = \frac{2}{T^2}(X'Z) W_T (Z'X)$$

which is a quadratic form and since rank $Z \geq q$, rank $X = q$ and W_T is P.D., it is P.D.

Asymptotic Properties

Consistency

$$\beta_T = \left[(X'Z) W_T (Z'X)\right]^{-1} (X'Z) W_T (Z'Y)$$
$$= \left[(X'Z) W_T (Z'X)\right]^{-1} (X'Z) W_T Z'(X\beta_0 + \epsilon)$$
$$= \beta_0 + \left[(X'Z) W_T (Z'X)\right]^{-1} (X'Z) W_T Z'\epsilon$$

$$= \beta_0 + \left[\left(\frac{1}{T}\sum_{t=1}^{T} x_t' z_t\right) W_T \left(\frac{1}{T}\sum_{t=1}^{T} z_t' x_t\right)\right]^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} x_t' z_t\right) W_T \left(\frac{1}{T}\sum_{t=1}^{T} z_t' \epsilon_t\right)$$

Assumption 2 : $\{z_t, x_t\}$ is jointly strictly stationary and ergodic where (z_t), and (x_t) are vector processes.

Thus, the above implies strictly stationarity and ergodicity for every of the involved vector processes (then each element of the vector forms a univariate strictly stationary and ergodic process - the converse, however is not true)

Assumption 3 : $E|x_t' z_t| < +\infty$

$E|z_t'\epsilon_t| < +\infty$

Hence, by Birkhoff's LLN :

· $\frac{1}{T} \sum_{t=1}^{T} x_t' z_t \to E(x_t' z_t)$ $\mathbb{P}$ a.s.

· $\frac{1}{T} \sum_{t=1}^{T} z_t' x_t = \frac{1}{T} \sum_{t=1}^{T} (x_t' z_t)' \to E\left[(x_t' z_t)'\right] = E(z_t' x_t)$ $\mathbb{P}$ a.s.

· $\frac{1}{T} \sum_{t=1}^{T} z_t' \epsilon_t \to E(z_t' \epsilon_t) \overset{LIE}{=} E\left[E(z_t' \epsilon_t)/z\right] = E\left[z_t' \overset{0}{\cancel{E(\epsilon_t/z)}}\right] = 0$ $\mathbb{P}$ a.s.

<u>Assumption 4</u> : $W_T \to \underset{\ell\times\ell}{W}$ $\mathbb{P}$ a.s.

Thus, due to the CMT :

G $\left[\left(\frac{1}{T} \sum_{t=1}^{T} x_t' z_t\right) W_T \left(\frac{1}{T} \sum_{t=1}^{T} z_t' x_t\right)\right] \to \left[E(x_t' z_t)\, W\, E(z_t' x_t)\right]^{-1}$ $\mathbb{P}$ a.s.

G $\left(\frac{1}{T} \sum_{t=1}^{T} x_t' z_t\right) W_T \left(\frac{1}{T} \sum_{t=1}^{T} z_t' \epsilon_t\right) \to 0$ $\mathbb{P}$ a.s.

Therefore,

$b_T \to b_0 + \left[E(x_t' z_t)\, W\, E(z_t' x_t)\right]^{-1} \cdot 0$ $\mathbb{P}$ a.s.

$b_T \to b_0$ $\mathbb{P}$ a.s.

hence, the GMM (IV) estimator is strongly consistent

Asymptotic Distribution

Given the consistency of the GMME we can now derive its asymptotic distribution.

Remember that,

$$b_T = b_0 + \left[\left(\frac{1}{T}\sum_{t=1}^{T} x_t' z_t\right) W_T \left(\frac{1}{T}\sum_{t=1}^{T} z_t' x_t\right)\right]^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} x_t' z_t\right) W_T \left(\frac{1}{T}\sum_{t=1}^{T} z_t' \epsilon_t\right) \Rightarrow$$

$$\sqrt{T}(b_T - b_0) = \left[\left(\frac{1}{T}\sum_{t=1}^{T} x_t' z_t\right) W_T \left(\frac{1}{T}\sum_{t=1}^{T} z_t' x_t\right)\right]^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} x_t' z_t\right) W_T \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t' \epsilon_t\right)$$

Given Assumptions 1-4 and the Identification condition, in order to find the asymptotic distribution of the estimator, we need to strengthen the orthogonality conditions.

Assumption 5: $\{z_t'\epsilon_t\}$ is a martingale difference vector process wrt. a filtration $f_t = \sigma(z_t, z_{t-1}, z_{t-2}, \dots, \epsilon_{t-1}, \epsilon_{t-2}, \dots)$ with $E|\epsilon_t^2 z_t' z_t| < +\infty$.

(which implies that $\cdot\ (z_t'\epsilon_t)_{t\in \mathbb{Z}}$ is adopted to $(f_t)_{t\in\mathbb{Z}}$ $\cdot\ E(\epsilon_t^2 z_t' z_t) < +\infty \ \forall t \in \mathbb{Z}$ $\cdot\ E(z_t'\epsilon_t / f_t) = 0 \ \forall t \in \mathbb{Z}$, so it is actually a square martingale difference sequence).

A sufficient condition for the aforementioned assumption is that $E(\epsilon_t / f_t) = 0 \ \forall t \in \mathbb{Z}$ which means that besides $(\epsilon_t)_{t\in\mathbb{Z}}$ being a m.d. sequence, it must also be orthogonal not only to the current but also to the past instruments.

Thereby, by the CLT for s.m.d. processes (actually a multivariate version of the CLT):

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t'\epsilon_t \xrightarrow{d} z \sim N(0, S) \quad \text{where} \quad S = E(\epsilon_0^2 z_0' z_0)$$

Thus, by Slutsky's Theorem:

$\sqrt{T}(\delta_T - \delta_0) \xrightarrow{d} z \sim N(0, AV_{\delta_T})$

where $AV_{\delta_T} = [E(x_t' z_t) W E(z_t' x_t)]^{-1} E(x_t' z_t) W E(\epsilon_t^2 z_t' z_t) W E(z_t' x_t) [E(x_t' z_t) W E(z_t' x_t)]^{-1}$

$$= (S_{xz} W S_{zx})^{-1} S_{xz} W S W S_{zx} (S_{xz} W S_{zx})^{-1}$$

As we can see the asymptotic variance of the GMME depends on W, so the question arising is, can it be chosen optimally wrt asymptotic efficiency? Also, the asymptotic variance depends on δ_0 and the limits that construct it may be analytically unknown, so the issue of consistent estimation of the asymptotic variance occurs.

However, in our case we have assumed that the errors are conditionally homoskedastic which implies that:

$S = E(\epsilon_t^2 z_t' z_t) \overset{LIE}{=} E(E(\epsilon_t^2 z_t' z_t / z)) = E(E(\epsilon_t^2 / z) z_t' z_t) = E(z_t' z_t) = S_{zz}$

thus, the asymptotic variance of the GMME is:

$AV_{\delta_T} = (S_{xz} W S_{zx})^{-1} S_{xz} W S_{zz} W S_{zx} (S_{xz} W S_{zx})^{-1}.$

which does not depend on δ so the estimation is much easier.

Optimal (Asymptotic) Weighting

In the way we defined the GMM estimator in our case of a linear model with instrumental variables, as $\hat{\theta}_T = \arg\min_{\theta} \left[\left(\frac{1}{T} Z'(Y - X\theta) \right)' W_T \left(\frac{1}{T} Z'(Y - X\theta) \right) \right]$ we observed that the estimator depends on the norm we have considered (in this case it is the norm that arises from the relevant quadratic form i.e. inner product), thereby by the relevant weighting matrix. (However, when $l=q$ remember that the GMME does not depend on the weighting matrix since $\hat{\theta}_T = (Z'X)^{-1} Z'Y$).

Considering the general case where $l > q$, we observe that in contrast to the issue of consistency, the asymptotic variance depends on the choice of the weighting matrix, hence the question of optimal selection of this matrix arises naturally with a view towards the asymptotic efficiency of the GMME. We will adress this using a Gauss-Markov argument.

Remember that,

$$AV_{\hat{\theta}_T} = \left[E(x_t' z_t) W E(z_t' x_t) \right]^{-1} E(x_t' z_t) W S W E(z_t' x_t) \left[E(x_t' z_t) W E(z_t' x_t) \right]^{-1}$$

$$= (S_{xz} W S_{zx})^{-1} S_{xz} W S W S_{zx} (S_{xz} W S_{zx})^{-1}$$

Lemma: When $l > q$ and $W = S^{-1}$ then $AV_{\hat{\theta}_T}$ is minimal

Proof: First, notice that when $W = S^{-1}$ then

$$AV_{\hat{\theta}_T}(S^{-1}) = (S_{xz} S^{-1} S_{zx})^{-1} S_{xz} S^{-1} S S^{-1} S_{zx} (S_{xz} S^{-1} S_{zx})^{-1}$$
$$= (S_{xz} S^{-1} S_{zx})^{-1} S_{xz} S^{-1} S_{zx} (S_{xz} S^{-1} S_{zx})^{-1}$$
$$= (S_{xz} S^{-1} S_{zx})^{-1}$$

we conclude to an expression similar to the case where $l=q$, which is an indication that the variance is truly the optimal.

Hence, it suffices to show that $(S_{xz} W S_{zx})^{-1} S_{xz} W S W S_{zx} (S_{xz} W S_{zx})^{-1} - (S_{xz} W S_{zx})^{-1}$ is positive semidefinite in order to prove that $(S_{xz} S^{-1} S_{zx})^{-1}$ is the minimal

Let's write the previous difference as :

$$(S_{x2}WS_{2x})^{-1}S_{x2}WSWS_{2x}(S_{x2}WS_{2x})^{-1} - (S_{x2}S^{-1}S_{2x})S_{x2}S^{-1}SS^{-1}S_{2x}(S_{x2}S^{-1}S_{2x})^{-1}$$

Since S is positive semidefinite, by the Cholesky decomposition [decomposition of a positive (semi)definite matrix into the product of a lower triangular matrix and its conjugate transpose] we can write S as $S = LL'$ where L is a square matrix.

Hence, the above difference can be written as:

$$(S_{x2}WS_{2x})^{-1}S_{x2}WLL'WS_{2x}(S_{x2}WS_{2x})^{-1} - (S_{x2}S^{-1}S_{2x})^{-1}S_{x2}S^{-1}LL^{-1}S^{-1}S_{2x}(S_{x2}S^{-1}S_{2x})^{-1}$$

In order to prove that it is positive semidefinite it suffices to prove that it can be written as BB' (since every positive (semi) definite matrix has a Cholesky decomposition).

Assertion : $\left[(S_{x2}WS_{2x})^{-1}S_{x2}W - (S_{x2}S^{-1}S_{2x})^{-1}S_{x2}S^{-1}\right]L = B$

$$BB' = \left[(S_{x2}WS_{2x})^{-1}S_{x2}W - (S_{x2}S^{-1}S_{2x})^{-1}S_{x2}S^{-1}\right]LL'\left[W'S_{2x}(S_{x2}WS_{2x})^{-1} - S^{-1}S_{2x}(S_{x2}S^{-1}S_{2x})^{-1}\right]$$

$$= (S_{x2}WS_{2x})^{-1}S_{x2}WLL'WS_{2x}(S_{x2}WS_{2x})^{-1} - (S_{x2}WS_{2x})^{-1}S_{x2}WLL'S^{-1}S_{2x}(S_{x2}S^{-1}S_{2x})^{-1}$$
$$- (S_{x2}S^{-1}S_{2x})^{-1}S_{x2}S^{-1}LL'W'S_{2x}(S_{x2}WS_{2x})^{-1} + (S_{x2}S^{-1}S_{2x})^{-1}S_{x2}S^{-1}LL'S^{-1}S_{2x}(S_{x2}S^{-1}S_{2x})^{-1}$$

$$= AV_{\theta_T}(W) - (S_{x2}S^{-1}S_{2x})^{-1} - \left[(S_{x2}S^{-1}S_{2x})^{-1}\right]' + AV_{\theta_T}(S^{-1})$$
$$= AV_{\theta_T}(W) - 2\,AV_{\theta_T}(S^{-1}) + AV_{\theta_T}(S^{-1})$$
$$= AV_{\theta_T}(W) - AV_{\theta_T}(S^{-1})$$

Therefore, when $l > q$ the optimal weighting matrix is the asymptotic variance of the vector of moments (there might be also other optimal weighting matrices, but the one we found is definitely optimal).

Hence, the efficient GMM estimator results from setting $W_T = \hat{S}^{-1}$ such that $\hat{S} \xrightarrow{P} S$ and the resulting asymptotic variance reduces to $AV_{\theta_T}(S^{-1}) = (S_{x2}S^{-1}S_{2x})^{-1}$, of which a consistent estimator would be

$$\widehat{AV}_{\delta_T} = (\Sigma_{xz}\hat{S}^{-1}\Sigma_{zx})^{-1} = \left[\left(\frac{1}{T}\sum_{t=1}^{T} x_t' z_t\right)\hat{S}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} z_t' x_t\right)\right]^{-1}$$

which requires a consistent estimator of S.

In our case, we have assumed that the errors are conditionally homoskedastic, thereby the optimal $W = S_{z'z}^{-1} = \left(E(z_t' z_t)\right)^{-1}$ (independent of δ) and

$AV_{\delta_T}(S^{-1}) = (S_{xz} S_{zz}^{-1} S_{zx})^{-1}$.

a consistent estimator of which is $\left[\left(\frac{1}{T}\sum_{t=1}^{T} x_t' z_t\right)\left(\frac{1}{T}\sum_{t=1}^{T} z_t' z_t\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} z_t' x_t\right)\right]^{-1}$.

GMM under serial correlation

Remember that in our analysis so far
we considered $(\epsilon_t)_{t\in\mathbb{Z}}$ to be a strictly stationary and ergodic white noise process, thus serially uncorrelated. In this context we proved consistency and asymptotic normality of the GMM estimator, imposing several assumptions.
What if ϵ_t's were serially correlated i.e. error terms ϵ_t and ϵ_s, for $t\neq s$ are correlated?
The GMM estimator remains consistent and asymptotically normal.
However, in order to prove asymptotic normality, we assumed that the vector process $(z_t'\epsilon_t)_{t\in\mathbb{Z}}$ is a s.m.d. process, which does not allow for serial correlation. We need to relax Assumption 5 in order to allow for serial correlation.

Assumption 5*: $\{z_t'\epsilon_t\}$ satisfies Gordin's condition. Its long-run variance matrix is non-singular.

Remember that this implies that the process $(z_t'\epsilon_t)$ satisfies:
Gordin's CLT (multivariate extension of Gordin's CLT we have already stated)
The stochastic vector process $(z_t'\epsilon_t)_{t\in\mathbb{Z}}$ is stationary ergodic. If

1. $E(\epsilon_t^2 z_t' z_t) < +\infty$
2. $\lim_{j\to\infty} E\left(E(z_t'\epsilon_t / f_{t-j})\right)^2 = 0$ for all t where $f_t = \sigma(\epsilon_{t-i}, z_{t-i}, i\geq 0)$
3. for $r_{t,j} = E(z_t'\epsilon_t / f_{t-j}) - E(z_t'\epsilon_t / f_{t-j-1})$, $\sum_{j=0}^{\infty} E(r_{t,j}^2)^{1/2} < +\infty$

then $E(z_0'\epsilon_0) = 0$, $\sum_{j=-\infty}^{\infty} |\Gamma_j| < +\infty$ and

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_t'\epsilon_t \xrightarrow{d} \mathcal{N}\left(0, \sum_{j=-\infty}^{\infty} \Gamma_j\right)$$

Thereby, given the previous asymptotic framework, by Slutsky's thm:
$\sqrt{T}(\beta_T - \beta_0) \xrightarrow{d} z \sim N(0, AV_{\beta_T})$

where $AV_{\beta_T} = \left[E(x_t'z_t)\, W\, E(z_t'x_t)\right]^{-1} E(x_t'z_t)\, W \left(\sum_{j=-\infty}^{\infty} \Gamma_j\right) W\, E(z_t'x_t)\left[E(x_t'z_t)\, W\, E(z_t'x_t)\right]^{-1}$

$$= \left[E(x_t'z_t)\, W\, E(z_t'x_t)\right]^{-1} E(x_t'z_t)\, W \underbrace{\left[\Gamma_0 + 2\sum_{j=1}^{\infty} \Gamma_j\right]}_{S} W\, E(z_t'x_t)\left[E(x_t'z_t)\, W\, E(z_t'x_t)\right]^{-1}$$

where $\Gamma_0 = E(\epsilon_t^2 z_t' z_t)$
$\Gamma_j = E(\epsilon_t \epsilon_{t-j}\, z_t' z_{t-j})$

All the results we obtained in the asymptotic analysis of GMM without serial correlation, concerning the optimal asymptotic weighting and the consistent estimation of the asymptotic variance, carry over, provided that $S = \Gamma_0 + 2\sum_{j=1}^{\infty} \Gamma_j$ is consistently estimated.
Obtaining such an estimate is not easy, especially when autocovariances do not vanish after finite lags.
A natural estimator of the matrix of autocovariances would be
$\hat{\Gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T} \hat{\epsilon}_t \hat{\epsilon}_{t-j}\, z_t' z_{t-j}$, where $\hat{\epsilon}_t = y_t - x_t'\hat{\beta}$, $\hat{\beta} \xrightarrow{p} \beta$

- Parametric estimates requires us to specify a model for ϵ_t. So, if we know that $\Gamma_j = 0$ for $j > q$ where q is known and finite, then
$\hat{S} = \hat{\Gamma}_0 + \sum_{j=1}^{q} (\hat{\Gamma}_j + \hat{\Gamma}_j') = \sum_{j=-q}^{q} \hat{\Gamma}_j$.

- A consistent estimate of S may be obtained using non-parametric methods.
A popular estimator is the Newey-West weighted autocovariance estimator, which is a kernel based estimator which uses a Barlett kernel and has the form:
$$\hat{S}_{NW} = \hat{\Gamma}_0 + \sum_{j=1}^{q(T)-1} \left(1 - \frac{j}{q(T)}\right)(\hat{\Gamma}_j + \hat{\Gamma}_j')$$

where $\hat{\Gamma}_0 = \frac{1}{T}\sum_{t=j+1}^{T} \hat{\epsilon}_t z_t' z_t$, $\hat{\Gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T} \hat{\epsilon}_t \hat{\epsilon}_{t-j}\, z_t' z_{t-j}$, $\hat{\epsilon}_t = y_t - x_t'\hat{\beta}$, $\hat{\beta} \xrightarrow{p} \beta$

Therefore, the Newey-West estimator uses a weighted sum of the autocovariances to estimate S, where the weights are determined by the Barlett kernel $k\left(\frac{j}{q(T)}\right) = \begin{cases} 1 - \left|\frac{j}{q(T)}\right|, & \text{for } |j| \leq q(T) - 1 \\ 0, & \text{for } |j| > q(T) - 1. \end{cases}$

and the bandwidth parameter $q(T)$.

(See Hayashi, Section 6.6)