Metrics Assignment 3

Bobing, Cassandra, Max, Prema, Rajdev, and Yazen

May 2023

1 Exercises (GMM)

1.1 Optimally Weighted GMM

Set-up and notation

GMM is a method that uses moment conditions to construct estimators. The method finds values of parameters that are the closest to satisfying the sample moment conditions. Specifically, we start with a set of equations that are implied by a parametric model of behavior and hold for each observation j, namely: $\mathbb{E}g_j(\beta) = \mathbf{0}$.

The weighting method in GMM uses different weights for different moment conditions such that for weighting matrix $\mathbf{A}$ (which is positive definite), $\hat{\beta}_{GMM} = \min_{b \in B} N \; \mathbf{g}_N(b)^T \mathbf{A} \mathbf{g}_N(b)$ (this is the GMM criterion that is associated with a weighting matrix $\mathbf{A}$).

The optimally weighted GMM estimator is given by:

$$b_N = \underset{b \in B}{\arg\min}\, N \mathbf{g}_N(b)^T \Omega^{-1} \mathbf{g}_N(b) \tag{1}$$

- $g_j(b)$ can be computed for each observation i, where g_i is a known function for each b in parameter space B. B is a compact set and $\beta \in B$, where β is a k-vector of parameters.
- Ω^{-1} provides the optimal weights for the moment conditions, where $\Omega = \mathbb{E}g_j(\beta)g_j(\beta)^T$
- If $\mathbb{E}g_j(\beta) = 0$ and observations are independent, the analogy principle tells us that $\mathbb{E}\mathbf{g}_N(\beta) = \mathbb{E}g_j(\beta) = 0$ where $\mathbf{g}_N(b) = \frac{1}{N}\Sigma_{i=1}^N g_j(b)$

- $\mathbf{Q} = \mathbb{E}\frac{dg_j(b)}{db^T}$.
- $\mathbf{V}_\beta = (\mathbf{Q}^T \mathbf{\Omega}^{-1}\mathbf{Q})^{-1}$ = asymptotic distribution of efficient GMM estimator

The following steps can be used to construct an optimally weighted GMM estimator:
Step 1: Describe the parameter space B;

When describing the parameter space B, we require that β is in the interior of B which will give us $\mathbb{E}\mathbf{Q}_N(\beta)^T\mathbf{A}\mathbf{g}_N(\beta) = \mathbf{0}$

Step 2: Describe a function $g_j(b)$ such that $\mathbb{E}g_j(\beta) = 0$;
We require this function $g_j(b)$ to be continuously differentiable (as $\mathbf{Q}_N(b)$ relies on this). As stated above, $g_j(b)$ can be computed for each observation i, where g_i is a known function for each b in parameter space B.We use this function to calculate the estimator for the covariance matrix, whose inverse is the optimal weighting matrix.

Step 3: Describe an estimator for the covariance matrix $\mathbb{E}g_j(\beta)g_j(\beta)^T$
Denote the covariance matrix as Ω. The inverse of this covariance matrix gives the optimal weights for the moment conditions. The estimator of this covariance matrix is given by applying the analogy principle: $\hat{\Omega} = \frac{1}{N}\Sigma_{i=1}^N(g_j(b)g_j(b))$

1.2 Causal Diagrams and GMM Estimators

1.2-a

$$\mathbb{E}y = \mu; \mathbb{E}(y-\mu)^2 = \sigma^2; \mathbb{E}(y-\mu)^3 = 0 \tag{2}$$

In this case, there are 2 parameters (μ, σ) and 3 moment conditions (k=2, l=3); when k>l, we are overidentified/uberidentified (most desirable case as we have more information to tell us about the values of parameters than is required). The causal diagram for this would be just "y" by itself (no arrows or other elements).

$$g_j(\beta) = \begin{pmatrix} y_j - m \\ (y_j - m)^2 - s^2 \\ (y_i - m)^3 \end{pmatrix}$$

$$g_N(b) = \frac{1}{N}\sum_j \begin{pmatrix} y_j - \mu \\ (y_j-\mu)^2 - \sigma^2 \\ (y_j-\mu)^3 \end{pmatrix}$$

$$Q_N(b) = (\frac{\partial g_N(b)}{\partial b^T}) = \frac{1}{N}\sum_j \begin{pmatrix} -1 & 0 \\ 2(y_j-\mu) & -2\sigma \\ 2(y_j-\mu) & 0 \end{pmatrix}$$

$$\Omega = \mathbb{E}\begin{pmatrix} (y_j-m)^2 & (y_j-m)(y_j-m)^2 - s^2 & (y_j-m)^4 \\ (y_j-m)[y_j-m)^2 - s^2] & [(y_j-m)^2 - s^2]^2 & (y_j m)^3[y_j-m)^2 - s^2] \\ (y_j-m)^4 & (y_j m)^3[y_j-m)^2 - s^2] & [(y_j-m)^3]^2 \end{pmatrix}$$

Asymptotic variance of optimal GMM estimator is given by $V_b = (Q^T\Omega^{-1}Q)^{-1}$

1.2b to 1.2h

Please refer to Appendix A in the end for part 1.2b- 1.2h.

1.3 Data Generating Processes

1.3a to 1.3h, excluding 1.3b

Please refer to Problem Set 3, Question 1.3a-1.3h.

1.4 Finite Sample Performance

For simulation results, please refer to Problem Set 3, Question 1.4. Here we explore finite sample properties for the OLS case using the GMM approach.

2 Breusch-Pagan Extended

Set-up:

$$y = \alpha + \beta x + u \tag{1}$$

In the above regression, y and x are scalar random variables. We are also given:

$$\mathbb{E}(u \cdot x) = \mathbb{E}(u) = 0 \tag{a.i}$$

$$\mathbb{E}(u^2|x) = \sigma^2 \tag{a.ii}$$

2.1

Assumption (a.i) states that the scalar random variable x is uncorrelated with the error term u, and that the expected value of the error term is 0. This is an untestable assumption since the error term, u, is supposed to capture any unobserved variable that affects the outcome y. Since it is not possible to fathom/enlist a universe of unobserved variables, it is not possible to test this assumption.

2.2

Breusch-Pagan (1979) propose an auxiliary equation, as follows, to test (aii).

$$\hat{u}^2 = c + dx + e \tag{2}$$

Since we assume that u is independent of x in (a.ii), it also implies that the variance of u should not depend on x. One way to measure the variance of u in a given sample is to consider the average of the squared values of the residuals, which is $\hat{u}$ in the auxiliary regression (2). The auxiliary regression tests if the *estimated* variance of u is linearly related to the regressor x by testing the hypothesis H0 : d = 0.

2.3

A GMM version of the BP test will have 3 moment conditions. The parameters of interest are α, β and σ^2. Assumption (a.i) gives us our first and second moment conditions and Assumption (a.ii) gives us our third moment condition.

$$\mathbb{E}[\mathbf{g}_1(\alpha, \beta, \sigma^2)] = \mathbb{E}[u] = \mathbb{E}[(y - \alpha - \beta x)] = 0 \tag{m.i}$$

$$\mathbb{E}[\mathbf{g}_2(\alpha,\beta,\sigma^2)] = \mathbb{E}[u.x] = \mathbb{E}[(y-\alpha-\beta x).x] = 0 \tag{m.ii}$$

$$\mathbb{E}[\mathbf{g}_3(\alpha,\beta,\sigma^2)] = \mathbb{E}[(u^2-\sigma^2).x] = \mathbb{E}[((y-\hat{\alpha}-\hat{\beta}x)^2-\sigma^2).x] = 0 \tag{m.iii}$$

Using the analogy principle, our sample moment conditions are as follows:

$$g_N(x,\alpha,\beta,\sigma^2) = \frac{1}{n}\Sigma_i \begin{pmatrix} y_i-\alpha-x_i^{'}\beta \\ x_i(y_i-\alpha-x_i^{'}\beta) \\ x_i^{'}[(y_i-\alpha-x_i^{'}\beta)^2-\sigma^2] \end{pmatrix}$$

Let π represent the matrix of parameters, p, and O be the parameter space. The GMM criterion uses an arbitrary positive definite weighting matrix $\mathbf{A}$ to estimate the above model and is given by

$$\hat{\pi}_{GMM} = \min_{p\in O} N\mathbf{g}_N(p)^T\mathbf{A}\mathbf{g}_N(p). \tag{3}$$

Once we estimate the above model, we can use a J-test to test the null hypothesis that our moment conditions equal 0 (reference: GMM lecture). Rejecting the null hypothesis is an indication for the presence of heteroskedasticity. The J statistic is

$$J = \mathbf{g}_N(x,\alpha,\beta,\sigma^2)^{'}\mathbf{A}\mathbf{g}_N(x,\alpha,\beta,\sigma^2) \tag{3}$$

2.4

First, while we are assured of getting BLUE estimates through OLS estimation in the BP test, we are not similarly assured of getting the most efficient estimates in our alternative GMM version of the BP test. Efficiency of our GMM estimates depend on the choice of the optimal weighting matrix $\mathbf{A}$.

Second, GMM estimation calls for a more intensive computation compared to the BP test.

2.5

Given, x is distributed uniformly over the interval $[0, 2\pi]$ and $\mathbb{E}(u^2|x) = \sigma^2(x) = \sigma^2 sin(2x)$. Note that the expected value of the squared residuals is a nonlinear function of x. Hence we cannot use the BP test here as the auxiliary regression is constructed such that the variance of error is *linearly* related to x.

But yes, we can use our GMM version to test for heteroskedasticity in this case as GMM uses a more general set of moment conditions. We could also potentially construct a superior alternative to our GMM version of the BP test by:

1. Adding an additional moment condition to over-identify our model as follows:

$$\mathbb{E}[\mathbf{g}_4(\alpha, \beta, \sigma^2)] = \mathbb{E}[u^2 - \sigma^2] = \mathbb{E}[(y - \hat{\alpha} - \hat{\beta}x)^2 - \sigma^2] = 0 \qquad \text{(m.iv)}$$

2. Choosing the weighting matrix $\mathbf{A}$ to be an optimal weighting matrix. We know from Hansen (2018) that the optimal weighting matrix is the inverse of the covariance matrix. Therefore we can choose A $= \Omega^{-1}$, where $\Omega = \mathbb{E}\mathbf{g}_j(p)^T\mathbf{g}_j(p)$. Our GMM criterion using Ω^{-1} as the weighting matrix is:

$$\hat{\pi}_{GMM} = \min_{p \in O} N\mathbf{g}_N(p)^T\Omega^{-1}\mathbf{g}_N(p). \qquad (4)$$

2.6

Given, $\mathbb{E}(u^2|x) = f(x)$. We can extend the BP test to specific an auxiliary *nonlinear* regression that tests the relationship between estimated squared residuals ($\hat{u}^2$) and f(x). We could run this auxiliary regression by taking f(x) to be a nonlinear function of x, such as, x^2. This is not a bulletproof test anymore, but could give us some evidence for/against heteroskedasticity in the model.

2.7

We know from the lecture that the efficiency of GMM estimation can improve by adding more moment conditions. In this case too, we can add more moment conditions (i.e.,overidentify the original model). Since we do not know the exact functional form of f(x), the GMM estimation would still not be as efficient as the optimal GLS estimator.

3 Tests of Normality

3.1

Under the null hypothesis of normality, the first k moments of the the distribution of x can be estimated as

$$m(x,k) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^k,$$

where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$. The corresponding moment conditions for GMM estimation are

$$g_N(\mu,\sigma) = \begin{cases} \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^k, & \text{k odd} \\ \frac{1}{N}\sum_{i=1}^{N}[(x_i - \mu)^k - \sigma^k(k-1)!!], & \text{k even} \end{cases}$$

3.2

The set of k equations that hold for all N observations are

$$g_j(\mu,\sigma) = \begin{cases} (x_j - \mu)^k, & \text{k odd} \\ (x_j - \mu)^k - \sigma^k(k-1)!!, & \text{k even} \end{cases}$$

$$\mathbb{E}g_j(\mu,\sigma) = 0.$$

The covariance matrix of the sample moment restrictions can be written as

$$\Omega = \mathbb{E}g_j(\mu,\sigma)g_j(\mu,\sigma)^T - \mathbb{E}g_j(\mu,\sigma)\mathbb{E}g_j(\mu,\sigma)^T.$$

Again, under the null hypothesis of normality, $\mathbb{E}g_j(\mu,\sigma) = 0$, thus

$$\Omega = \mathbb{E}g_j(\mu,\sigma)g_j(\mu,\sigma)^T,$$

which is positive definite.

3.3

Take $k > 2$, which in this case is overidentifying, thus we need to minimize an objective function, defined as

$$J(b) = N g_N(b)^T \Omega^{-1} g_N(b),$$

where $b = (\hat{\mu}, \hat{\sigma})$ is the estimator, and Ω is the covariance matrix. We can follow a two-step GMM procedure to estimate the parameters: first, use identity matrix as the weighting matrix, estimate the parameters as b_1 and get the estimated covariance matrix $\hat{\Omega}$; second, use the inverse of $\hat{\Omega}$ as the weighting matrix and estimate the parameters again to yield an efficient GMM estimator. Under the null hypothesis, $J(b)$ should follow a χ_{k-2} distribution, which gives us the critical value for the test of normality.

3.4

Please refer to Problem Set 3, Question 3.

3.5

For simulation results, please refer to Problem Set 3, Question 3. In general, within the same sample, the value of J, the objective function, increases with the number of moment conditions, as does critical value of the test. The precision of estimated parameters seems to improve a little bit with more moment conditions, but using too many conditions (in our case, using more than 8 conditions to estimate 2 parameters) that can be weak will result in less precise estimators, much larger value of J, and thus false rejection of null hypothesis of normality.

3.6

For estimation results using MLE, please refer to Problem Set 3, Question 3. The GMM estimates and MLE estimates are almost identical, because they use essentially the same information from the data. MLE can be viewed as a just-identified case of GMM.

4 Logit

4.1

$$E[Y|X] = \sigma(\beta^T X)$$

$$E[Y|X] - \sigma(\beta^T X) = 0$$

$$E[Y - \sigma(\beta^T X)|X] = 0$$

Which gives us our moment condition:

$$E[X(Y - \sigma(\beta^T X)] = 0$$

So we take g_N:

$$g_N = \frac{1}{N}\sum_{j=1}^{N} x_j(y_j - \sigma(\beta^T x_j))$$

Which is just identified under $k = l$.
and

$$\hat{\beta} = arg\min_b g_N^T \hat{W} g_N$$

4.2

By the independence of draws, we see that the probability $\beta = \beta$ given some realization of data (y, X), is equal to the probability that we drew y_i N times given $\beta = \beta$. Because each draw is independent, we have:

$$L(\beta|y, X) = \prod_{j=1}^{N} (\sigma(\beta^T x_j))^{y_j}(1 - \sigma(\beta^T x_j))^{1-y_j}$$

This is because the probability of drawing (y, X) under β is equal to the product of the probability of drawing (y_j, X_j) under β for all j

4.3

$$log(L(b|y,X)) = \sum_{j=1}^{N} log((\sigma(\beta^T x_j))^{y_j}(1-\sigma(\beta^T x_j))^{1-y_j})$$

$$= \sum_{j=1}^{N} y_j log(\sigma(\beta^T x_j)) + (1-y_j) log(1-\sigma(\beta^T x_j))$$

$$= \sum_{j=1}^{N} y_j log(\sigma(\beta^T x_j)) + log(1-\sigma(\beta^T x_j) - y_j log(1-\sigma(\beta^T x_j))$$

$$= \sum_{j=1}^{N} y_j log(\frac{1}{1+e^{-(\beta^T x_j)}}) + log(1-(\frac{1}{1+e^{-(\beta^T x_j)}}) - y_j log(1-\frac{1}{1+e^{-(\beta^T x_j)}}))$$

FOC:

$$\frac{d}{db} = \sum_{j=1}^{N} y_j \frac{x_j}{1+e^{b^T x_j}} + \frac{x_j e^{b^T x_j}}{1+e^{b^T x_j}} - y_j \frac{x_j e^{b^T x_j}}{1+e^{b^T x_j}}$$

$$= \sum_{j=1}^{N} \frac{x_j}{1+e^{b^T x_j}}(y_j + e^{b^T x_j} - y_j e^{b^T x_j})$$

$$= \sum_{j=1}^{N} \frac{x_j}{1+e^{b^T x_j}}(y_j - e^{b^T x_j} + y_j e^{b^T x_j})$$

$$= \sum_{j=1}^{N} \frac{x_j}{1+e^{b^T x_j}}(-e^{b^T x_j} + y_j(1+e^{b^T x_j}))$$

$$= \sum_{j=1}^{N} x_j(y_j - \frac{e^{b^T x_j}}{1+e^{b^T x_j}})$$

$$= \sum_{j=1}^{N} x_j(y_j - \sigma(\beta^T x_j)) = 0$$

which implies that the sample average of the score is equivalent to zero, or:

$$\frac{1}{N}\sum_{j=1}^{N} x_j(y_j - \sigma(\beta^T x_j)) = 0$$

This gives us an identical basis for our GMM estimator! Because our g_Ns are equivalent, we have that the two must be equally efficient.

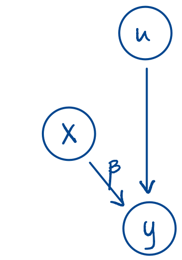
A Question 1.2b-1.2h

1b (b) $y = \alpha + X\beta + u$; with $\mathbb{E}(X^\top u) = \mathbb{E}u = 0$.

Note on notation:

$y = \alpha + X\tilde{\beta} + u$

$\beta = [\alpha \ \hat{\beta}]$ where we add a column of 1s to X



$y = \alpha + X\tilde{\beta} + u$

$E(X^T u) = Eu = 0$

In DAG, let $\beta = [\alpha \ \tilde{\beta}]$ where $y = \alpha + X\tilde{\beta} + u$

$$g_j(b) = \left(X_j\left(y_j - X_j(b)\right)\right)$$

$$g_N(b) = \frac{1}{N}\left(\sum_j \left(X_j\left(y_j - X_j\hat{b}\right)\right)\right) = \frac{X(y - Xb)}{N} = \frac{Xu}{N}$$

$$Q = E\left(-X_j X_j\right) = \frac{X^T X}{N} = Q$$

$$\Omega = E\left(g_j(b)\, g_j(b)^T\right) = E\left(\left(X_j\left(y_j - X_j(b)\right)\right)\left(X_j\left(y_j - X_j(b)\right)\right)^T\right)$$

$$= E\left(\left(X_j(u_j)\right)\left(X_j(u_j)\right)^T\right)$$

$$= \frac{1}{N}\sum_j\left(\left(X_j(u_j)\right)^2\right)$$

$$= \frac{Xu(Xu)^T}{N}$$

$$\Rightarrow \Omega^{-1} = \frac{1}{N}\left(Xu(Xu)^T\right)^{-1}$$

$$V_B = \left(Q^T \Omega^{-1} Q\right)^{-1} = \frac{1}{N}\left(X^T X\left(Xu(Xu)^T\right)^{-1} X^T X\right)^{-1}$$

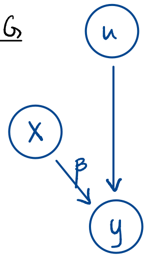
(1c) (c) $y = \alpha + X\beta + u$; with $\mathbb{E}(X^\top u) = \mathbb{E}u = 0$, and $\mathbb{E}(u^2) = \sigma^2$.

Note on notation:

$y = \alpha + X\tilde{\beta} + u$

$\beta = [\alpha \;\; \tilde{\beta}]$ where we add a column of 1s to X

DAG



$y = \alpha + X\tilde{\beta} + u$

$E(X^T u) = Eu = 0$

$E(u^2) = \sigma^2$

In DAG, let $\beta = [\alpha \;\; \tilde{\beta}]$ where $y = \alpha + X\tilde{\beta} + u$

1c is the same as 1b, but with the addition of the 2nd moment condition.
Addition of 2nd moment condition won't change DAG.

$$g_j(b) = \begin{pmatrix} X_j(y_j - X_j(b)) \\ u^2 - s^2 \end{pmatrix} = \begin{pmatrix} X_j(y_j - X_j(b)) \\ (y_j - X_j(b))^2 - s^2 \end{pmatrix}$$

$$g_N(b) = \frac{1}{N}\begin{pmatrix} \sum_j (x_j(y_j - x_j b)) \\ \sum_j ((y_j - x_j b))^2 - s^2 \end{pmatrix}$$

$$\begin{array}{ccc} & \partial/\partial b & \partial/\partial s \\ Q = E & \begin{pmatrix} -X_j \cdot X_j & 0 \\ 2(y_j - X_j b)b & -2s \end{pmatrix} \end{array} = \frac{1}{N}\begin{pmatrix} -X^T X & 0 \\ 2(y - Xb)b & -2s \end{pmatrix} = Q$$

$$\Omega = E\left(g_j(b)\, g_j(b)^T\right) = E\left(\begin{pmatrix} X_j(y_j - X\cdot(b_j)) \\ (y_j - X_j(b))^2 - s^2 \end{pmatrix}\begin{pmatrix} X_j(y_j - X_j(b)) \\ (y_j - X_j(b))^2 - s^2 \end{pmatrix}^T\right)$$

$$= E\begin{pmatrix} (X_j(y_j - X_j(b)))^2 & (X_j(y_j - X_j(b)))((y_j - X_j(b))^2 - s^2) \\ (X_j(y_j - X_j(b)))((y_j - X_j(b))^2 - s^2) & ((y_j - X_j(b))^2 - s^2)^2 \end{pmatrix}$$

↳ inverse of this $= \Omega^{-1}$

$$V_b = (Q^T \Omega^{-1} Q)^{-1}$$

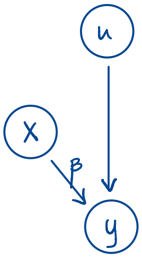
1d

(d) $y = \alpha + X\beta + u$; with $\mathbb{E}(X^\top u) = \mathbb{E}u = 0$, and $\mathbb{E}(u^2) = e^{X\sigma}$.

Note on notation:

$y = \alpha + X\tilde{\beta} + u$

$\beta = [\alpha \ \tilde{\beta}]$ where we add a column of 1s to X



$y = \alpha + X\tilde{\beta} + u$

$E(X^T u) = Eu = 0$

$E(u^2) = e^{X\sigma}$

In DAG, let $\beta = [\alpha \ \tilde{\beta}]$ where $y = \alpha + X\tilde{\beta} + u$

$$g_j(b) = \begin{pmatrix} X_j(y_j - X_j(b)) \\ u^2 - e^{Xs} \end{pmatrix} = \begin{pmatrix} X_j(y_j - X_j b) \\ (y_j - X_j b) - e^{Xs} \end{pmatrix}$$

$$g_N(b) = \frac{1}{N}\begin{pmatrix} \sum_j (X_j(y_j - X_j \hat{b})) \\ \sum_j ((y_j - X_j b))^2 - e^{X_j s} \end{pmatrix}$$

$$\varphi = E\begin{array}{cc} \partial/\partial b & \partial/\partial s \\ \left(\begin{matrix} -X_j \cdot X_j \\ -2(y_j - X_j b)(-b) \end{matrix}\right. & \left.\begin{matrix} 0 \\ -Xe^{X_j s} \end{matrix}\right) \end{array} = \frac{1}{N}\begin{pmatrix} -X'X & 0 \\ 2(y - Xb)b & -Xe^{Xs} \end{pmatrix} = \varphi$$

$$\varphi = E\begin{array}{cc} \partial/\partial b & \partial/\partial s \\ \left(\begin{matrix} -X_j \cdot X_j \\ 2(y_j - X_j b)b \end{matrix}\right. & \left.\begin{matrix} 0 \\ -2s \end{matrix}\right) \end{array} = \frac{1}{N}\begin{pmatrix} -X^T X & 0 \\ 2(y - Xb)b & -2s \end{pmatrix} = \varphi$$

$$\Omega = E(g_j(b)\, g_j(b)^T) = E\left(\begin{pmatrix} X_j(y_j - X \cdot (b_j)) \\ (y_j - X_j(b))^2 - s^2 \end{pmatrix}\begin{pmatrix} X_j(y_j - X_j(b)) \\ (y_j - X_j(b))^2 - s^2 \end{pmatrix}^T\right)$$

$$= E\begin{pmatrix} (X_j(y_j - X_j(b)))^2 & (X_j(y_j - X_j(b)))((y_j - X_j(b))^2 - s^2) \\ (X_j(y_j - X_j(b)))((y_j - X_j(b))^2 - s^2) & ((y_j - X_j(b))^2 - s^2)^2 \end{pmatrix}$$

↳ inverse of this $= \Omega^{-1}$

$$\Omega = E(g_j(b)\, g_j(b)^T) = E\left(\begin{pmatrix} X_j(y_j - X_j(b)) \\ (y_j - X_j(b)) - e^{X_j s} \end{pmatrix}\begin{pmatrix} X_j(y_j - X_j(b)) \\ (y_j - X_j(b)) - e^{X_j s} \end{pmatrix}^T\right)$$

$$= \frac{1}{N}\left(\begin{pmatrix} Xu \\ u - e^{Xs} \end{pmatrix}\begin{pmatrix} Xu \\ u - e^{Xs} \end{pmatrix}^T\right)$$

Ω^{-1} = inverse of ↑

$$V = (\varphi^T \Omega^{-1} \varphi)^{-1}$$

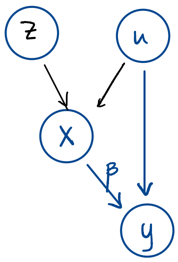
(1e) (e) $y = \alpha + X\beta + u$; with $\mathbb{E}(Z^\top u) = \mathbb{E}u = 0$ and $\mathbb{E}Z^\top X = Q$.

Note on notation:

$y = \alpha + X\tilde{\beta} + u$

$\beta = [\alpha \ \tilde{\beta}]$ where we add a column of 1s to X

DAG:



$y = \alpha + X\tilde{\beta} + u$

$E(Z^T u) = Eu = 0$

$E(Z^T X) = Q$

In DAG, let $\beta = [\alpha \ \tilde{\beta}]$ where $y = \alpha + X\tilde{\beta} + u$

Assumption:

- Q has full column rank

$$g_j(b) = \left(Z_j (y_j - X_j(b))\right)$$

$$g_N(b) = \frac{1}{N}\left(\sum_j \left(Z_j (y_j - X_j \hat{b})\right)\right) = \frac{Zu}{N}$$

$$Q = E\left(\frac{\partial g_j(b)}{\partial b^T}\right) = E(-Z_j X_j) = \frac{1}{N}\sum_j -Z_j X_j = \frac{-ZX^T}{N} = Q$$

$$\Omega = E\left(g_j(b)\, g_j(b)^T\right) = E\left(\left(Z_j (y_j - X_j(b))\right)\left(Z_j (y_j - X_j(b))\right)^T\right)$$

$$= E\left((Z_j(u_j))(Z_j(u_j))^T\right)$$

$$= \frac{1}{N}\sum_j \left((Z_j(u_j))^2\right)$$

$$= \frac{1}{N}(Zu)^2$$

$\longrightarrow$ If we assume homoscedasticity:

$$= \frac{1}{N}\sum Z_j^2 = \frac{Z^T Z}{N}$$

$$\Omega^{-1} = \left(\frac{Z^T Z}{N}\right)^{-1}$$

(assuming homoscedasticity)

$$V = (\varphi^T \Omega^{-1} \varphi)^{-1} = \frac{1}{N}\left(-X^T Z \,(Z^T Z)^{-1} (-X^T Z)\right)^{-1}$$

1f

(f) $y = f(X\beta) + u$; with f a known scalar function and with $\mathbb{E}(Z^\top u) = \mathbb{E}u = 0$ and $\mathbb{E}Z^\top X f'(X\beta) = Q(\beta)$. (Bonus question: where does this last restriction come from, and what role does it play?)

Non-linear least squares

$$\begin{tikzcd} y & & X \arrow[ll, "f(\cdot)"'] \\ & u \arrow[ul] \arrow[ur] & \end{tikzcd}$$

$y = f(X\beta) + u$

$f =$ known scalar

$E(Z^\top u) = Eu = 0$

Assumptions:

- f is continuously differentiable
- $Q(\beta)$ has full column rank when $b = \beta$ is in the interior of the parametre space

$$g_j(b) = \left(Z_j\,(y - f(x_j b)\right)$$

$$g_N(b) = \frac{1}{N}\sum_j \left(Z_j\,(y_j - f(x_j b))\right) = \frac{Z^\top (y - f(xb))}{N}$$

$$Q = E\left(\frac{\partial g_N(b)}{\partial b}\right) = E\left(-\frac{f'(X, b)\,Z^\top X}{N}\right) = -\frac{Q(\beta)}{N}$$

$$\Omega = E\left(g_j(b)\,g_j(b)^\top\right) = E\left(\left(Z_j(y - f(x_j b))\right)\left(Z_j(y - f(x_j b))\right)^\top\right)$$

$$= E\left((Z_j u_j)(Z_j u_j)^\top\right) = \frac{1}{N}\sum_j \left((Z_j u_j)(Z_j u_j)^\top\right) = \frac{1}{N}(Zu)(Zu)^\top$$

$$= (Zu)(Zu)^\top \frac{1}{N}$$

$$\Omega^{-1} = \left[\frac{(Zu)(Zu)^\top}{N}\right]^{-1}$$

$$V = \left(Q^\top \Omega^{-1} Q\right)^{-1}$$

1g

(g) $y = f(X, \beta) + u$; with f a known function and with $\mathbb{E}(Z^\top u) = \mathbb{E}u = 0$ and $\mathbb{E}Z^\top \frac{\partial f}{\partial \beta^\top}(X, \beta) = Q(\beta)$.

$$\begin{tikzcd} y & & X \arrow[ll, "f(\cdot)"'] \arrow[dl] \\ & u \arrow[ul] & \end{tikzcd}$$

$$y = f(X, \beta) + u$$

$$E(Z^\top u) = Eu = 0$$

$$E\, Z^\top \frac{\partial f}{\partial \beta^\top}(X, \beta) = Q(\beta)$$

Assumptions:

- f is continuously differentiable
- $Q(\beta)$ has full column rank when $b = \beta$ is in the interior of the parametre space

$$g_j(b) = Z_j\,(y_j - f(x_j, \beta))$$

$$g_N(b) = \frac{1}{N}\sum_j \left(Z_j\,(y_j - f(x_j, \beta))\right)$$

$$= \frac{1}{N}\left(Z(y - f(X, \beta))\right)$$

$$Q = E(Q_N(b)) = E\left(\frac{\partial g_N(b)}{\partial b}\right)$$

$$= E\left(-Z f'(X, \beta)\right)$$

$$= Q(\beta) = Q$$

$$\Omega = E(g_j(b)\, g_j(b)^\top) = E\left[\left(Z_j\,(y_j - f(x_j, \beta))\right)\left(Z_j\,(y_j - f(x_j, \beta))\right)^\top\right]$$

$$= E\left[(Z_j u_j)(Z_j u_j)^\top\right]$$

$$= \frac{1}{N}\sum_j \left[(Z_j u_j)(Z_j u_j)^\top\right]$$

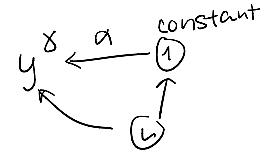
$$= \frac{1}{N}(Zu)(Zu)^\top$$

$$\Omega^{-1} = \left(\frac{(Zu)(Zu)^\top}{N}\right)^{-1}$$

$$V = (Q^\top \Omega^{-1} Q)^{-1}$$

1h

(h) $y^\gamma = \alpha + u$, with $y > 0$ and γ a scalar, and $\mathbb{E}(Z^\top u) = \mathbb{E}u = 0$ and $\mathbb{E}Z^\top \begin{bmatrix} \gamma y^{\gamma-1} \\ -1 \end{bmatrix} = Q(\gamma)$.



Given: $y^\gamma = \alpha + u$ $\quad y > 0$

γ : scalar

$y^\gamma = \alpha + u$ is a power equation

$E(Z^\top u) = Eu = 0$

$$EZ^\top \begin{bmatrix} \gamma y^{\gamma-1} \\ -1 \end{bmatrix} = Q(\gamma)$$

$$g_j(b) = \begin{pmatrix} Z_j^\top (y_j^\gamma - \alpha) \\ y_j^\gamma - \alpha \end{pmatrix}$$

$$g_N(b) = \frac{1}{N} \sum \begin{pmatrix} Z_j^\top y_j^\gamma - Z_j^\top \alpha \\ y_j^\gamma - \alpha \end{pmatrix} = \frac{1}{N} \begin{pmatrix} Z^\top y^\gamma - Z^\top \alpha \\ y^\gamma - \alpha \end{pmatrix}$$

$$\begin{array}{ccc} & \partial/\partial\gamma & \partial/\partial\alpha \end{array}$$

$$\varphi = E\left(\frac{\partial g_j(b)}{\partial b}\right) = E\begin{pmatrix} Z_j^\top \gamma y_j^{\gamma-1} & -Z_j^\top \\ \gamma y^{\gamma-1} & -1 \end{pmatrix}$$

$$\Omega = E\left(g_j(b) g_j(b)^\top\right) = E\left(\begin{pmatrix} Z_j^\top (y_j^\gamma - \alpha) \\ y_j^\gamma - \alpha \end{pmatrix} \begin{pmatrix} Z_j^\top (y_j^\gamma - \alpha) \\ y_j^\gamma - \alpha \end{pmatrix}^\top\right)$$

$$V = (\varphi^\top \Omega^{-1} \varphi)^{-1}$$