Metrics Assignment 1

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

March 2023

1 Admin

2 Exercises

2.1 Inverse Jacobian Rule

We use mathematical rather than econometric notation to answer this question. Let (X, Y) be IID random variables and suppose $Z = X + Y$. Following Hansen (2022b) Theorem 4.25, let $W = Y$ and define a one-to-one function mapping $(X, Y) \mapsto (Z, W)$ as

$$g(x, y) = \begin{cases} x + y & = z \\ y & = w \end{cases}$$

such that

$$g^{-1}(z, w) = \begin{cases} z - y = z - w & = x \\ w & = y. \end{cases}$$

By Theorem 4.24 (Hansen, 2022b),

$$\begin{aligned} f_{Z,W}(z, w) &= f_{X,Y}(g_1^{-1}(z, w), g_2^{-1}(z, w))|J(z, w)| \\ &= f_X(g_1^{-1}(z, w))f_Y(g_2^{-1}(z, w)) \\ &= f_X(z - w)f_Y(w) \\ &= f_X(z - y)f_Y(y) \end{aligned}$$

where the second equality holds because (X,Y) is IID and

$$
\begin{aligned}
J(z,w) &= \begin{vmatrix} \frac{\partial g_1^{-1}(z,w)}{\partial z} & \frac{\partial g_2^{-1}(z,w)}{\partial z} \\ \frac{\partial g_1^{-1}(z,w)}{\partial w} & \frac{\partial g_2^{-1}(z,w)}{\partial w} \end{vmatrix} \\
&= \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} \\
&= 1.
\end{aligned}
$$

The distribution of Z therefore is

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_{Z,W}(z,w)dw \\
&= \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy.
\end{aligned}
$$

2.2 Convolution of Continuous and Discrete Variables

See Question 3.1 for proof.

2.3 Generalized Inverse

If $A \in \mathbb{R}^{m\times n}$ is a matrix of zeros, then there exists a generalized inverse A^- such that $A = AA^-A$ where $A^- \in \mathbb{R}^{n\times m}$ with arbitrary elements $a_{i,j}$.

2.4 Variable Notation

Following (Hansen, 2022a, p. 4), we use y_i to denote a realization of a scalar random variable y. We do not follow mathematics notation (where y is a realization of scalar Y) because we reserve capital letters for matrices as we show below. We denote random vectors as bold scalars.

2.4.1 Regress Scalar on Vector

$$y = \mathbf{x}'\beta + \mathbf{e}$$

2.4.2 Regress Realization of Scalar on Realization of Vector

$$y_i = x_i'\beta + e_i$$

2.4.3 Regress Vector of Realizations on Matrix of Realizations

$$\mathbf{y} = \mathbf{X}\beta + e$$

2.5 Moore-Penrose Inverse

(1)

If $\mathbf{A}$ is a $n \times m$ matrix of zeros, then by the second property of pseudoinverse, $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$, $\mathbf{A}^+$ must also be a matrix of zeros, with dimensions $m \times n$.

(2)

If $\mathbf{A}$ has full column rank, then $\mathbf{A}^\mathbf{T}\mathbf{A}$ has full rank and is thus invertible, so we have

$$(\mathbf{X}^\mathbf{T}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^\mathbf{T}\mathbf{X} = \mathbf{I}.$$

By the first property of pseudoinverse, we know

$$\mathbf{X}\mathbf{X}^+\mathbf{X} = \mathbf{X},$$

left multiplying this equation by the left inverse of $\mathbf{X}$, that is, $(\mathbf{X}^\mathbf{T}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^\mathbf{T}$, we have

$$(\mathbf{X}^\mathbf{T}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^\mathbf{T}\mathbf{X}\mathbf{X}^+\mathbf{X} = (\mathbf{X}^\mathbf{T}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^\mathbf{T}\mathbf{X},$$
$$\mathbf{X}^+\mathbf{X} = \mathbf{I}.$$

Thus $\mathbf{X}^+ = (\mathbf{X}^\mathbf{T}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^\mathbf{T}$, the pseudoinverse is equivalent to left inverse. ■

(3)

Multiplying the regression equation by the pseudoinverse $\mathbf{X}^+$ and utilizing the results from (2), we have

$$\mathbf{X}^+\mathbf{y} = \mathbf{X}^+\mathbf{X}\mathbf{b} + (\mathbf{X}^\mathbf{T}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^\mathbf{T}\mathbf{u}.$$

Since $\mathbf{X}^\mathbf{T}\mathbf{u} = 0$, then

$$\mathbf{X}^+\mathbf{y} = \mathbf{X}^+\mathbf{X}\mathbf{b} \implies \mathbf{b} = \mathbf{X}^+\mathbf{y},$$

as $\mathbf{X}^+\mathbf{X} = \mathbf{I}$.

3 Convolutions

Please refer to Assignment 1 Section 3.ipynb where we adapt code from section to create convolutions of discrete and continuous random variables.

4 General Weighted Linear Regressions

In the first half of the class, we considered:

1. Linear Regression
2. GLS
3. IV + 2SLS

Which of these belong to the class of general weighted linear regressions which can be expressed as $T^{'}Y = T^{'}X\beta + T^{'}u$?

Trivially, linear regression belongs to the class if we set $T = I$. T is a constant here, so T is not random.

GLS belongs to the class. Following Greene (2012) 9.3, we can write the GLS estimator as:

$$P^{'}Y = P^{'}X\beta + P^{'}u$$

Where $E[\epsilon\epsilon^{'}|X] = \sigma^2\Omega$ and P is defined such that $\Omega^{-1} = P^{'}P$. Here, then, $T = P$. T is a matrix of constants and is not random.

IV and 2SLS do belong to the class. Recall that the IV estimator can be written as:

$$Z^{'}Y = Z^{'}X\beta_{IV} + Z^{'}u$$

Where Z is our matrix of instruments. In this case, $T = Z$. As Greene (2012) 8.3.4 shows, the 2SLS estimator and the IV estimator are equivalent, meaning that 2SLS must also belong to the class. Since Z is a matrix containing data which we use as an instrument, it may or not be random depending on the data generating process.

5 Simultaneous Equations

5.1

Let y be a $N \times k$ matrix, with N observations of y and k dependent variables. Then the dimension of $X\beta + u$ must be $(N \times k)$. Let p be the number of independent/predictor variables. [Note: $(p+1)$ is equivalent to l from lecture].

This implies that the dimension of X is $(N \times (p+1))$, the dimension of β is $(p+1) \times k$, and the dimension of u is $N \times k$. With these dimensions, the stated objects are 'conformable', allowing for the operations of matrix multiplication and matrix addition.

5.2

In weighted_regression.ipynb, to estimate β, we solve the following equation: $T'y = T'Xb + T'u$, assuming $E(T'u) = 0$. The estimate of β, b^*, is then given by $(T'X)^+(T'y)$. In weighted_regression.ipynb, $X = T^3D + u$, where T is a sample drawn from a normal distribution with mean 0 and a given variance σ^2, and u is a sample drawn from a normal distribution with covariance 0.2.

No, we cannot use the estimator from weighted_regression.ipynb to estimate a system of simultaneous equations. The code in weighted_regression.ipynb requires that k is 1, so we will not be able to use it when k is greater than 1 unless we make changes to ensure conformability (as stated in section 5 , question 1). Specifically, if $k = 3$, then we need to make the following changes to ensure conformability:

- Change Sigma to a diagonal matrix
- Change D to a 3x3 matrix
- Change u so it is drawn from a multivariate normal distrubtion.
- Change β so it is a 3x3 matrix

These changes are made and executed in Section 5, question 3.

5.3

Please refer to bottom section of assignment1_sec5.ipynb.

5.4

Linear regression assumes homoscedasticity, meaning there is a constant variance in errors. When there is a violation of this (ie. there is heteroskedasticity), one can use the weighted least squares method (because the weighted least squares method uses different weights for each observation based on the variance).
In addition to the assumptions required for OLS regression, one of the necessary assumptions for the weighted least squares regression (which is a consequence of the presence of weights) is that the weights applied to different observations are indicative of the actual variance of errors.

6 SUR

6.1

Let the model of simultaneous equations be

$$y = X\beta + u \tag{1}$$

where the dimension of y is $N \times k$, X is $N \times l$, β is $l \times k$ and u is $N \times k$. Therefore, $cov(u|X)$ is a $kN \times kN$ matrix.

The SUR model assumes that for each observation i, there are k dependent variables y_{i1}, y_{i2}, ...,y_{ik} s.t.

$$y_{ij} = X_{ij}\beta_{ij} + u_{ij} \qquad j = 1, 2...k \tag{2}$$

$$E(y_{ij}) = X_{ij}\beta_{ij} \tag{3}$$

$$V(y_{ij}) = \sigma_{jj} I_N \tag{4}$$

(assuming homoskedasticity). Therefore, Ω is a generalization of $V(y_j)$ to k dependent variables.
However, if Ω is not a diagonal matrix, this implies that the system of equations are dependent through the channel of covariance between error terms.

For example, $cov(e_{is}, e_{it}) \neq 0$ for some j = s,t. Additionally, $cov(e_{is}, e_{i't}) = 0\ \forall i \neq i'$. Therefore $\mathbf{C}(e_j, e_k) = \sigma_{jk} I_N$. and observing a realization of y_j may change our prediction of y_k since $\mathbf{C}(e_j, e_k) \neq \sigma_j j I_N$.

More formally,

$$cov(u|X) = \Omega = \begin{bmatrix} \sigma_{11}I_N & \sigma_{12}I_N & . & . & .\sigma_{1k}I_N \\ \sigma_{21}I_N & \sigma_{22}I_N & . & . & . \\ . & & & & \\ . & & & & \\ . & & & & \\ \sigma_{k1}I_N & \sigma_{k2}I_N & . & . & .\sigma_{kk}I_N \end{bmatrix} \tag{5}$$

Note that if Ω was a diagonal matrix, we would have had all the non-diagonal elements in Ω equal to 0.

We can also express Ω as

$$\Omega = \Sigma \otimes I_N \tag{6}$$

where $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & . & . & .\sigma_{1k} \\ \sigma_{21} & \sigma_{22} & . & . & . \\ . & & & & \\ . & & & & \\ . & & & & \\ \sigma_{k1} & \sigma_{k2} & . & . & .\sigma_{kk} \end{bmatrix}$ and $\otimes$ is called the Kronecker product (reference: wikipedia).

6.2

Please refer to pset1.problem6.ipynb where we adapt the code in weightedregression.ipynb to accommodate a general covariance matrix for u.

6.3

The estimates obtained from the SUR system appear to have lower variance for the parameters compared to the OLS estimates. In other words, SUR estimates provide an efficiency improvement over the OLS estimates.
Note: Please refer to pset1.problem6.ipynb for the OLS estimates.

7 Food Expenditures in India

Please refer to Assignment 1 Section 7.ipynb where we adapt code from Kernel density.ipynb to find the density of expenditures and log expenditures.

8 "Plug-in" Kernel Bias Estimator

The Oracle estimator we proposed for the bias of kernel estimation is

$$\mathbf{Bias}(x) = \mathbb{E}(\hat{f}(x)) - f(x) = \int k(u)f(x+hu)du - f(x).$$

Since we do not know the true density, $f(x)$, we cannot assess the exact bias (and variance) of any density estimation. Plugging in our estimate $\hat{f}$ for f in this expression can be useful in the sense that it helps us to gauge the relative performance of different estimators (potentially of different kernels or bandwidths).

To achieve this, we need a good "benchmark" density estimator to compare other estimators to, which may require some conditions including large sample size to improve the precision of estimation, and well-behaved true underlying density that is continuous and have bounded derivatives. We also need to choose appropriate kernel functions and bandwidths to balance the trade-off between bias and variance and get desired estimations.

One potential pitfall is that we should always remember this "plug-in" bias is not the exact bias, rather it provides a comparison between two density estimators that we want to evaluate.

References

Greene, W. (2012). *Econometric Analysis*.

Hansen, B. (2022a). *Econometrics*. Princeton University Press.

Hansen, B. (2022b). *Probability and Statistics for Economists*. Princeton University Press.