Problem Set 2

Econ 4676: Big Data and Machine Learning for Applied Economics

Daniel Gómez Salazar*, Lucas Gómez Tobón†, José Daniel Palacio Murillo ‡

1 Theory Exercises

1. Suppose you have the following spatial model $y = \rho Wy + X\beta + WX\theta + \epsilon$ with $|\rho| < 1$ this is sometimes known as the Spatial Durbin Model

   (a) First consider the following scenario $\beta = \theta = 0$.

      i. **Write the Likelihood function. Can you find a closed form for the parameter estimators? Don't forget to be specific on the assumptions you make.**

$$y = \rho Wy + \epsilon$$

We take into consideration the following asumptions:

- $|\rho| < 1$
- $\epsilon \sim N(0, \sigma^2)$
- w is exogenous

With this being said we can now define y:

$$\begin{aligned} y - \rho Wy &= \epsilon \\ (I_n - \rho W)y &= \epsilon \\ y &= (I_n - \rho W)^{-1}\epsilon \end{aligned}$$

Therefore, the likelihood function can be defined as:

$$L(y) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \cdot |det(v(y))|^{-\frac{1}{2}} exp\left(-\frac{1}{2}(y - E(y))'v(y)^{-1}(y - E(y))\right)$$

*d.gomezs@uniandes.edu.co

†l.gomezt@uniandes.edu.co

‡jd.palacio@uniandes.edu.co

To fin the MLE we need $E[y]$ and $v(y)$:

$$E[y] = E\left[(I_n - \rho W)^{-1} E(\epsilon)\right]$$

$$\text{we assume that } \rho \text{ is given and } \epsilon \overset{i.i.d}{\sim} N(0, \sigma^2)$$

$$\Rightarrow E[y] = 0$$

Now we find $v(y)$

$$v(y) = E[yy'] - E[y]E[y]'$$

$$E[yy'] = E\left[\left((I_n - \rho W)^{-1}\epsilon\right) \cdot \left((I_n - \rho W)^{-1}\epsilon\right)'\right]$$

$$E[yy'] = E\left[(I_n - \rho W)^{-1}\epsilon\epsilon'\left((I_n - \rho W)^{-1}\right)'\right]$$

$$E[yy'] = E\left[(I_n - \rho W)^{-1}\left((I_n - \rho W)^{-1}\right)'\sigma^2\right]$$

$$E[yy'] = \underbrace{[(I_n - \rho W)'(I_n - \rho W)]^{-1}}_{\Omega}\sigma^2$$

$$v(y) = \sigma^2\Omega$$

Now, it is possible to define the likelihood function:

$$L(\rho, \sigma^2, y) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \cdot |\sigma^2\Omega|^{-\frac{1}{2}} \cdot exp\left(-\frac{1}{2}(y-0)'(\sigma^2\Omega)^{-1}(y-0)\right)$$

$$L(\rho, \sigma^2, y) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \cdot |\sigma^2\Omega|^{-\frac{1}{2}} \cdot exp\left(-\frac{1}{2\sigma^2}\left((I_n - \rho W)^{-1}\epsilon\right)'\Omega^{-1}\left((I_n - \rho W)^{-1}\epsilon\right)\right)$$

$$L(\rho, \sigma^2, y) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \cdot |\sigma^2\Omega|^{-\frac{1}{2}} \cdot$$

$$exp\left(-\frac{1}{2\sigma^2}\left(\epsilon'\underbrace{\left((I_n - \rho W)'\right)^{-1}(I_n - \rho W)'}_{\mathrm{I}}\underbrace{(I_n - \rho W)(I_n - \rho W)^{-1}}_{\mathrm{I}}\epsilon\right)\right)$$

$$L(\rho, \sigma^2, y) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \cdot |\sigma^2\Omega|^{-\frac{1}{2}} \cdot exp\left(-\frac{1}{2\sigma^2}\epsilon'\epsilon\right)$$

We use the log function:

$$l(\sigma^2, \rho, y) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\left(|\sigma^2\Omega|\right) - \frac{1}{2\sigma^2}\left((I_n - \rho W)y\right)'(I_n - \rho W)y$$

note that $|\sigma^2\Omega| = \sigma^{2n}|\Omega|$, also $|\Omega| = |(I_n - \rho W)|^{-2}$

$$l(\rho,\sigma^2,y) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}(n)\ln\left(\sigma^2\right) - \frac{n}{2}(-2)\ln(|I-\rho W|) \\ -\frac{1}{2\sigma^2}\left((I_n-\rho W)y\right)'(I_n-\rho W)y$$

$$l(\rho,\sigma^2,y) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) + n\ln(|I-\rho W|) - \frac{1}{2\sigma^2}\left((I_n-\rho W)y\right)'(I_n-\rho W)y$$

Ord (1975) showed that

$$|I-\rho W| = \prod_{i=1}^{n}(1-\rho W_i), \text{where } W_i \text{ is the eigenvalue of i.}$$

$$l(\rho,\sigma^2,y) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) + n\Sigma\ln(1-\rho W_i) - \frac{1}{2\sigma^2}\left((I_n-\rho W)y\right)'(I_n-\rho W)y$$

$$l(\sigma^2,y|\rho) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) + n\Sigma\ln(1-\rho W_i) - \frac{1}{2\sigma^2}\left((I_n-\rho W)y\right)'(I_n-\rho W)y$$

- $$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{1}{\sigma^4}\left((I_n-\rho W)y\right)'(I_n-\rho W)y = 0$$

$$\frac{1}{2}\frac{1}{\sigma^4}\left((I_n-\rho W)y\right)'(I_n-\rho W)y = \frac{n}{2}\frac{1}{\sigma^2}$$

$$\boxed{\sigma^2(\rho) = \frac{1}{n}\left((I_n-\rho W)y\right)'(I_n-\rho W)y} = \frac{\epsilon'\epsilon}{n}$$

- $$\frac{\partial l}{\partial \rho} = n\Sigma\frac{-W_i}{1-\rho W_i} + \frac{1}{\sigma^2}\left(y'(I-\rho W)'Wy\right) = 0$$

Since ρ cannot be derived analytically, ρ mmust be obtained from an explicit maximization of a concentrated log-likelihood function using numerical optimization:

$$l(\rho) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2(\rho)) + n\Sigma\ln(1-\rho W_i) - \frac{1}{2\sigma^2(\rho)}\left((I_n-\rho W)y\right)'(I_n-\rho W)y$$

$$\sigma^2_{ML}(\hat{\rho}) = \frac{1}{n}\left((I_n-\hat{\rho}W)y\right)'(I_n-\hat{\rho}W)y$$

ii. **Suppose instead you use MCO, would you obtain the same estimates?**

$$y = (I_n-\rho W)^{-1}\epsilon$$

We now minimize the squared error:

$$\min_{\rho} \quad e'e = \left((I_n-\rho W)y\right)'(I_n-\rho W)y$$

Is not possible to fin a closed form for the estimates $\hat{\rho}_{OLS}$ and $\hat{\sigma}^2_{OLS}$. Therefore:

$$\hat{\sigma}^2_{OLS} \neq \hat{\sigma}^2_{ML} \quad ; \quad \hat{\rho}_{OLS} \neq \hat{\rho}_{ML}$$

$$[\rho]: \quad -2y'(I-\hat{\rho}W)'Wy = 0$$
$$y'(I-\hat{\rho}W')Wy = 0$$
$$(y'-\hat{\rho}y'W')Wy = 0$$
$$y'Wy - \hat{\rho}y'W'Wy = 0$$
$$\hat{\rho}y'W'Wy = y'Wy$$
$$\hat{\rho}_{OLS} = y'Wy(y'W'Wy)^{-1}$$

We also have:

$$\hat{\sigma}^2_{OLS} = \frac{e'e}{n-k} = \frac{((I_n-\rho W)y)'\,(I_n-\rho W)y}{n-k} \neq \hat{\sigma}^2_{ML}$$

(b) Now consider that $\rho = 0$, and let's proceed as before:

i. **Write the Likelihood function. Can you find a closed form for the parameter estimators? Don't forget to be specific on the assumptions you make.**

$$y = X\beta + WX\theta + \epsilon; \quad \epsilon \overset{i.i.d}{\sim} N(0,\sigma^2)$$
$$y = z\gamma + \epsilon; \quad z = [X, WX] \text{and} \gamma = [\beta, \theta]$$

Therefore:

$$L(y) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \cdot |det(v(y))|^{-\frac{1}{2}} exp\left(-\frac{1}{2}(y-E(y))'v(y)^{-1}(y-E(y))\right)$$

$$L(y) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \cdot |det(v(y))|^{-\frac{1}{2}} exp\left(-\frac{1}{2}(y-z\gamma)'v(y)^{-1}(y-z\gamma)\right)$$

$$\Rightarrow l(y) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\big(\sigma^2\big)\big) - \frac{1}{2\sigma^2}(y-z\gamma)'(y-z\gamma)$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{1}{\sigma^4}(y-z\gamma)'(y-z\gamma) = 0$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n}(y-z\gamma)'(y-z\gamma)$$

Concentrated likelihood

$$l^c(y|\gamma, z) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}ln\left(\frac{1}{n}(y-z\gamma)'(y-z\gamma)\right) - \frac{n}{2}$$

$$[\gamma]: \quad \frac{n}{(y-z\hat{\gamma})'(y-z\hat{\gamma})}\cdot z'(y-z\hat{\gamma})=0$$

$$z'y=z'z\hat{\gamma}$$

$$\boxed{\hat{\gamma_{ML}}=(z'z)^{-1}(z'y)}$$

$$\boxed{\hat{\sigma_{ML}}^2=\frac{1}{n}\left(y-z(z'z)^{-1}(z'y)\right)'\left(y-z(z'z)^{-1}(z'y)\right)}=\frac{\epsilon'\epsilon}{n}$$

ii. **Suppose instead you use MCO, would you obtain the same estimates?**

In that case we have:

$$\min_{\gamma} \quad e'e=(y-z\gamma)'(y-z\gamma)$$

$$[\gamma]: \quad -2z'(y-z\gamma)=0$$

$$\boxed{\hat{\gamma}_{OLS}=(z'z)^{-1}(z'y)} \Rightarrow \quad \hat{\gamma}_{ML}=\hat{\gamma}_{OLS}$$

$$\boxed{\hat{\sigma}^2_{OLS}=\frac{\epsilon'\epsilon}{n-k-1}} \Rightarrow \quad \hat{\sigma}^2_{ML}\neq\hat{\sigma}^2_{OLS}$$

And it can be proved as it follows:

$$y=z\gamma+\epsilon \Rightarrow \quad \epsilon=y-z\gamma$$

$$\hat{\epsilon}=y-z\hat{\gamma} \Rightarrow \quad \hat{\epsilon}=y-z(z'z)^{-1}z'y \Rightarrow \quad \hat{\epsilon}=(I-z(z'z^{-1}z)y)$$

$$\hat{\epsilon}=My, \text{where } M \text{ is an idempotent matrix}$$

$$var(\epsilon)=E\left[(\epsilon-E[\epsilon])'(\epsilon-E[\epsilon])\right]=E[\epsilon'\epsilon], \text{since } E(\epsilon)=0$$

$$E[\epsilon'\epsilon|z]=E[y'M'My|z]=E[y'My|z]$$

The scalar $\epsilon'M\epsilon$ is a $1x1$ matrix, so its equal to its trace. By using the result on cyclic perutations

$$E[tr(\epsilon'M\epsilon)|z]=E[tr(M\epsilon\epsilon')|z]$$

Since M is function of z:

$$tr\left(E[\epsilon\epsilon'|z]\right)=tr(M\sigma^2I)=\sigma^2tr(M)$$

$$tr\left((I-z(z'z)^{-1}z)\right)=tr(I)-tr(z(z'z)^{-1}z)=n-k-1$$

$$E[\epsilon'\epsilon|z]=(n-k-1)\sigma^2$$

$$\Rightarrow \quad \hat{\sigma}^2=\frac{\epsilon'\epsilon}{n-k-1}$$

2. Consider the regression model $y = X\beta + \epsilon$ with $\epsilon \sim N(0, \sigma^2 I)$ furthermore assume that β has a normal prior, i.e. $\beta \sim N(0, \tau^2 I)$.

   (a) Find the posterior distribution.

   (b) Compare it with the ridge formula we saw in class.

   (c) What is the relationship between λ in the ridge model and σ^2 and τ^2?

3. Centered Ridge. Suppose that $\bar{x} = 0$, i.e. the data has been centered. Show that the parameters that minimize $R(\beta, \beta_0) = (y - X\beta - \beta_0\iota)'(y - X\beta - \beta_0\iota) + \lambda\beta'\beta$ are $\beta_0 = \bar{y}$ and $\beta = (X'X + \lambda I)^{-1}X'y$

4. Suppose that we have the following regression model $y = X\beta + \epsilon$, and decide to do the following: Augment the centered matrix X with p additional rows with $\sqrt{\lambda}$, and augment y with zeros. Show that this procedures renders the ridge regression estimates, is there a link to the leverage statistic?

5. Reducing elastic net to lasso. Suppose that you have the following functions $EL(\beta) = (y - X\beta)^2 + \lambda_2\beta^2 + \lambda_1|\beta|$ and $L(\beta) = (\tilde{y} - \tilde{X}\beta)^2 + c\lambda_1|\beta|$ where $c = (1 + \lambda_2)^{\frac{-1}{2}}$ show that these two problems are equivalent when $\tilde{y}$ and $\tilde{X}$ are the augmented data versions of the previous exercise.