# ECON6027
# 7A

## Introduction to Spatial Econometrics

## Solutions to exercises

# PACKAGES YOU NEED

spdep

# SOLUTIONS A - DATA PREP

```
UK_nb = read.gal("UK_nb.gal")
```

#read GAL from the last chapter

```
(W_list = nb2listw(UK_nb))
uk_data =
st_read("uk_data.shp")
st_is_valid(uk_data)
options(scipen = 999)
```

# to remove “e” notation from the summary outputs

```
options(scipen = 0)
```

# to resore “e” notation

# The Variables

- GVA: gross value added
- LP: labour productivity
- PBB: percentage of new business births

```
Call:
lm(formula = gross_va ~ L_prod + pct_bsbirth, data = uk_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.1398 -0.9172 -0.4388  1.0958  2.5365 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -22.31118    3.38594  -6.589 0.000100 ***
L_prod        0.27750    0.05346   5.191 0.000571 ***
pct_bsbirth   0.42239    0.47243   0.894 0.394567    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.791 on 9 degrees of freedom
Multiple R-squared:  0.9072,	Adjusted R-squared:  0.8866 
F-statistic: 43.99 on 2 and 9 DF,  p-value: 0.00002259
```

# Model 1

$$GVA_i = \alpha + \beta_1 LP_i + \beta_2 PBB_i + \epsilon_i$$

```
> model1 = lm(formula = gross_v ~ L_prod + pct_bsb, data=uk_data)
> summary(model1)
```

PBB is not significant, but LP is significant.

```
Call:
lm(formula = gross_va ~ L_prod, data = uk_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.5300 -0.8375 -0.4193  1.0386  3.0149 

Coefficients:
             Estimate Std. Error t value   Pr(>|t|)    
(Intercept) -21.38866    3.19237  -6.700 0.00005365 ***
L_prod        0.31460    0.03335   9.432 0.00000271 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.773 on 10 degrees of freedom
Multiple R-squared:  0.899,	Adjusted R-squared:  0.8889 
F-statistic: 88.97 on 1 and 10 DF,  p-value: 0.000002708
```

# Model 2

$$GVA_i = \alpha + \beta_1 LP_i + \epsilon_i$$

```
> model2 = lm(formula = gross_v ~ L_prod, data=uk_data)
> summary(model2)
```

LP is significant.

```
Call:
lm(formula = gross_va ~ pct_bsbirth, data = uk_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.8006 -1.5308 -0.6353  1.8746  5.6300 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) -16.0547     5.9992  -2.676  0.02325 * 
pct_bsbirth   2.3264     0.5646   4.121  0.00208 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.397 on 10 degrees of freedom
Multiple R-squared:  0.6293,	Adjusted R-squared:  0.5923 
F-statistic: 16.98 on 1 and 10 DF,  p-value: 0.002075
```

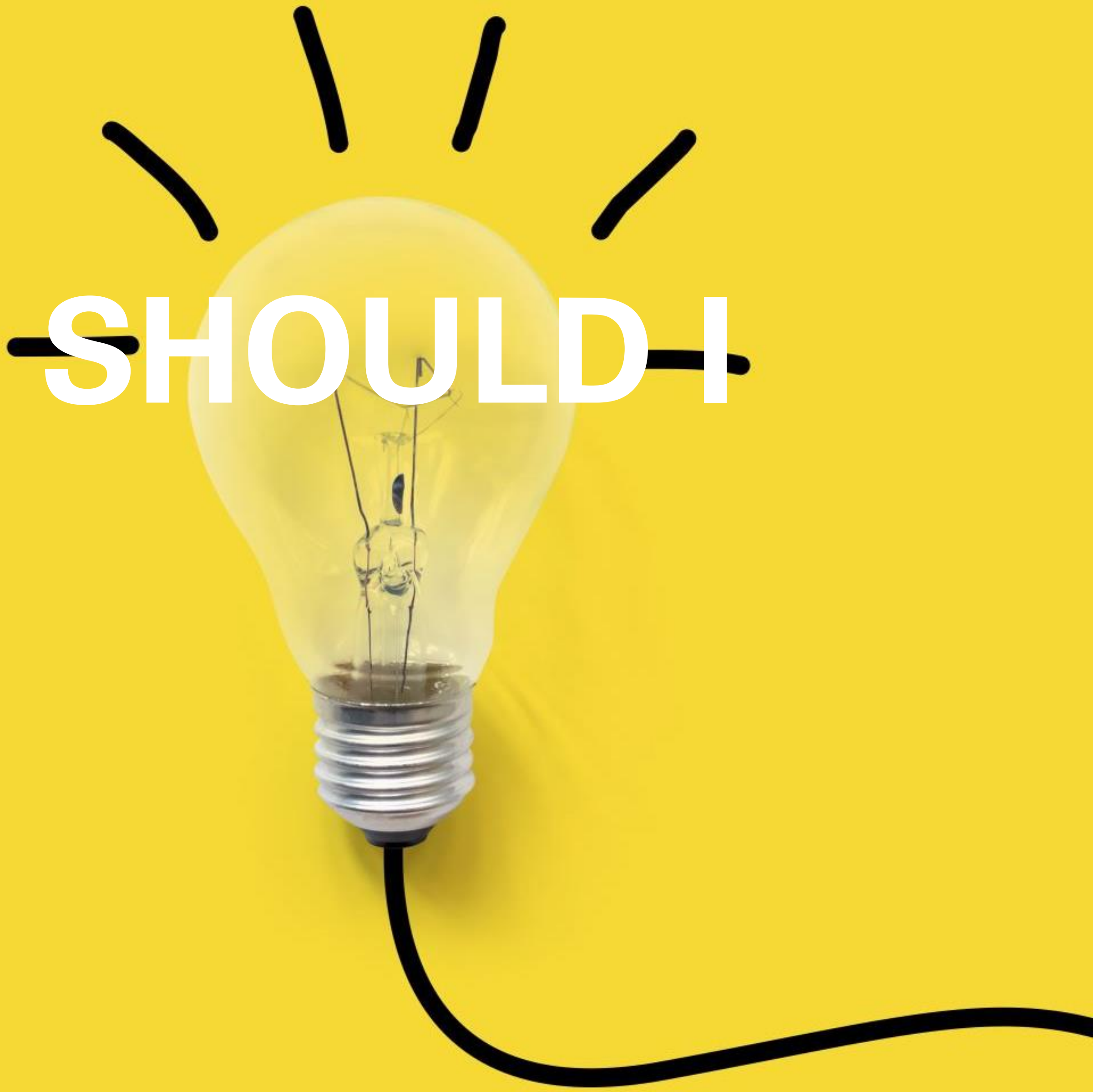
# Model 3

$$GVA_i = \alpha + \beta_2 PBB_i + \epsilon_i$$

```
> model3 = lm(formula = gross_v ~ pct_bsb, data=uk_data)
> summary(model3)
```

PBB is significant, possibly due to omitted variable bias.

# WHICH MODEL SHOULD I CHOOSE?



# Model 1 vs. Model 2 vs. Model 3

I prefer **model 2** as it has the largest adjusted R-square statistic (0.8899) and the most significant F-statistic (p-value: 0.000002708).

Alternatively, you can also consider AIC and BIC to compare.

```
> AIC(model1);AIC(model2);AIC(model3)
```

[1] 52.59553

[1] 51.61663

[1] 67.2135

```
> BIC(model1);BIC(model2);BIC(model3)
```

[1] 54.53516

[1] 53.07135

[1] 68.66822

```
Call:
lm(formula = L_prod ~ pct_bsbirth, data = uk_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-13.192  -6.589  -2.225   4.571  16.980 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)   22.546     18.718   1.205  0.25613   
pct_bsbirth    6.861      1.761   3.895  0.00298 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.6 on 10 degrees of freedom
Multiple R-squared:  0.6027,	Adjusted R-squared:  0.563 
F-statistic: 15.17 on 1 and 10 DF,  p-value: 0.002984
```

# Model 4

$$LP_i = \alpha + \beta PBB_i + \epsilon_i$$

```
> model4 = lm(formula = L_prod ~ pct_bsb, data=uk_data)
> summary(model4)
```

Given the significance of model 4, it maybe that model 1 suffers from multicollinearity.

# Hypothesis tests

```
> lmtest::bptest(model1)
```

studentized Breusch-Pagan test

data: model1

BP = 1.5183, df = 2, p-value = 0.4681

Conclusion: Null of homoskedasticity is not rejected. No evidence of heteroskedasticity.

```
>
tseries::jarque.bera.test(model1$residuals)
```

Jarque Bera Test

data: model1$residuals

X-squared = 0.091982, df = 2, p-value = 0.9551

Conclusion: Null of normality is not rejected.

# Solutions B

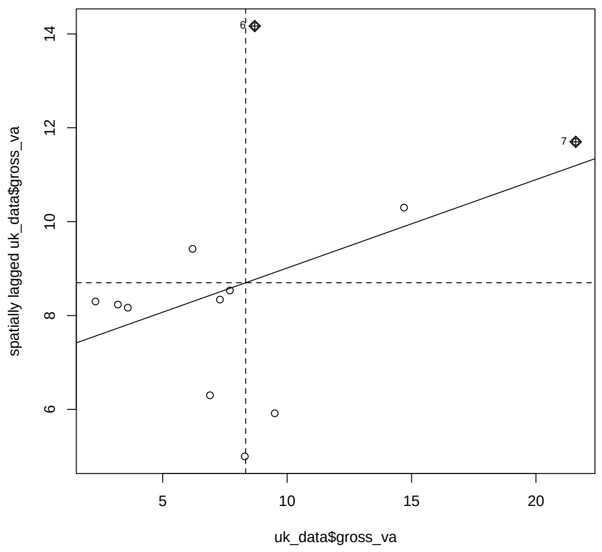
## Lagged labour productivity variable (W*LP)

```
> as.matrix(lag.listw(W_list,uk_data$L_prod))
          [,1]
 [1,]  90.06667
 [2,]  87.93333
 [3,]  88.00000
 [4,]  93.50000
 [5,]  91.48000
 [6,] 112.40000
 [7,] 102.55000
 [8,] 100.92000
 [9,]  92.96667
[10,]  89.16667
[11,]  85.90000
[12,]  96.90000
```

# SOLUTIONS C

We consider model 1 in this exercise

# Lagged dependent variable




```
> (L_GVA =
as.matrix(lag.listw(W_list,uk_data$gross_va)))
         [,1]
 [1,]  8.233333
 [2,]  5.916667
 [3,]  6.300000
 [4,]  9.420000
 [5,]  8.340000
 [6,] 14.166667
 [7,] 11.700000
 [8,] 10.300000
 [9,]  8.533333
[10,]  8.166667
[11,]  5.000000
[12,]  8.300000
> moran.plot(uk_data$gross_va, W_list)
```

There seem to be positive spatial autocorrelation in the dependent variable "GVA" with most points in the H-H and L-L quadrants.

# Moran’s I test for dependent variable

```
> moran.test(uk_data$gross_va, W_list)
```

Moran I test under randomisation

data: uk_data$gross_va

weights: W_list

Moran I statistic standard deviate = 1.6398, p-value = 0.05053

alternative hypothesis: greater

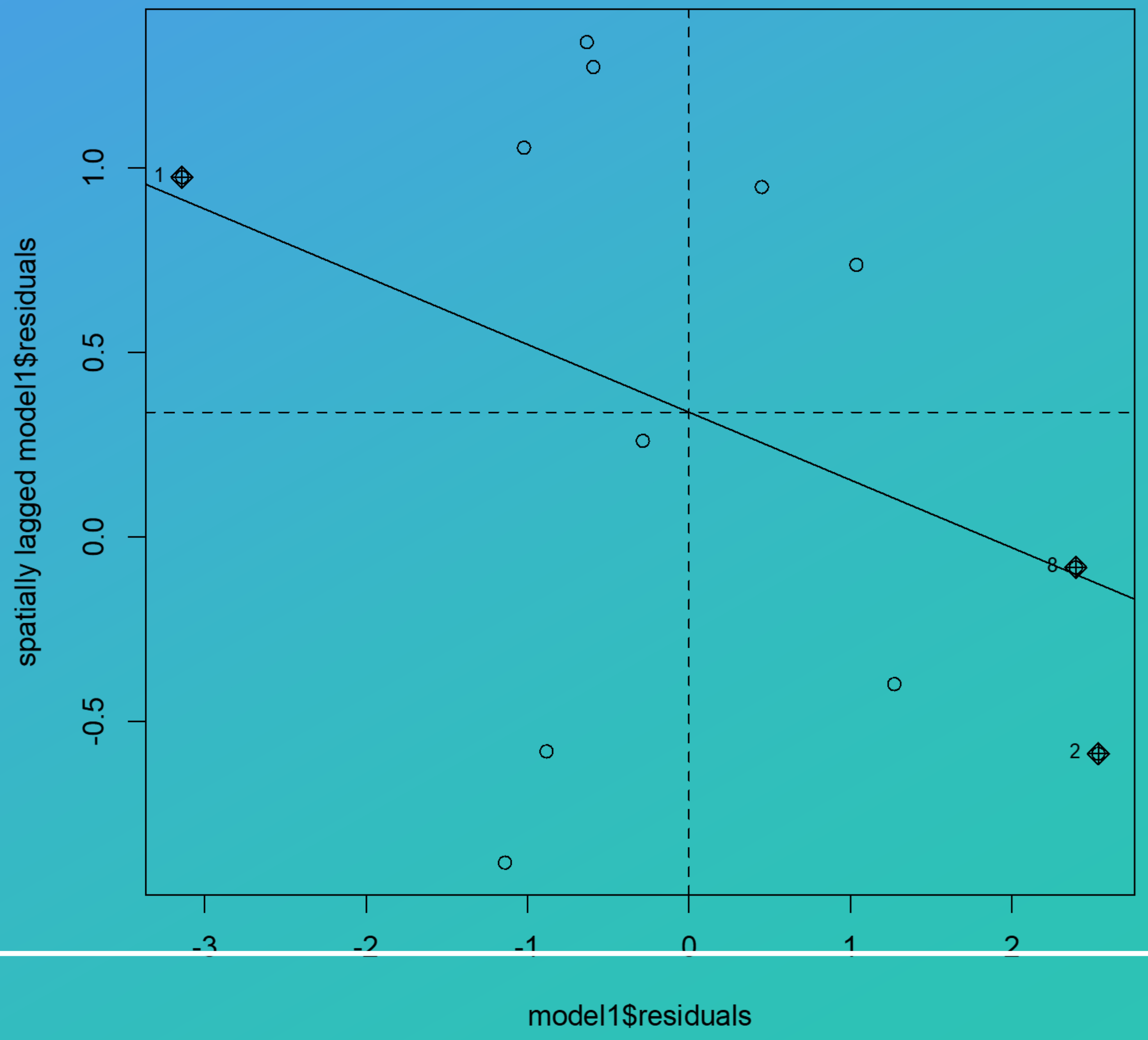
sample estimates:

| Moran I statistic | Expectation | Variance |
|---|---|---|
| 0.18820554 | -0.09090909 | 0.02897315 |

Conclusion: no evidence of spatial autocorrelation in the dependent variable at 5% level of significance. (However, this is very close to a rejection.)

# LAGGED RESIDUALS

> moran.plot(model1$residuals, W_list)



# Moran’s I test for residuals of Model 1

```
> lm.morantest(model1, W_list)
```

Global Moran I for regression residuals

model: lm(formula = gross_v ~ L_prod + pct_bsb, data = uk_data)

weights: W_list

Moran I statistic standard deviate = -0.17214, p-value = 0.8633

alternative hypothesis: two.sided

sample estimates:

| Observed Moran I | Expectation | Variance |
|---|---|---|
| -0.18356947 | -0.15277051 | 0.03201064 |

Conclusion: Do not reject H0. no evidence of spatial autocorrelation in the OLS residuals.

# Model 1:

Disturbances

- Does not suffer from heteroskedasticity (BP test)
- Does not suffer from non-normality (JB test)
- Does not suffer from autocorrelation (Moran's I = -0.1836, p-value = 0.5683)

The dependent variable, GVA also does not display spatial autocorrelation (Moran's I = 0.1882, p-value = 0.0505).

Thus, additional complications need not be considered on Model 1.

However, it is possible that the model 1 suffers from multicollinearity given the significance of model 4 (the alternative model 2 is likely a better candidate).

# Important R functions

- lm()
- AIC()
- BIC()
- bptest()
- jarque.bera()
- lag.listw()