Milestone III

Sachin Shubham

5/2/2021

```
library(knitr)
library(kableExtra)

## Warning: package 'kableExtra' was built under R version 4.0.5

options(knitr.table.format = "latex")
```

Table S1:

```
library(kableExtra)
setwd("D:/Statistical Methods/Project")
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
df_S1 <- read.csv("S1.csv", header=T, row.names=1)
df_S1[is.na(df_S1)] <- " "

  kbl(df_S1,caption = "Table:S1",
      col.names =c("(1) \n $X_{SA}$","(2) \n $X_{CIP}$","(3) \n
$X_{SA\\&CIP}$","(4) \n $X_{SA}$","(5) \n $X_{CIP}$","(6) \n $X_{SA \\&
CIP}$"),format = "latex", escape = F,booktabs=T)  %>% pack_rows(" ", 1, 7,
hline_before = T) %>%
  pack_rows(" ", 8, 14,hline_before = T) %>%
  pack_rows(" ", 15, 16,hline_before = T)%>%
  column_spec(c(2:7), width = "6em") %>%
  kable_classic(full_width = F, html_font = "Cambria")

  #Table S1
  df_S1

##                                                  XSA                 XCIP
## y                                1.0317 *** (0.0007) 1.0085 *** (0.00110)
## $\\bar{z_j}$                      0.9975   (0.01185) 1.2822 *** (0.02100)
## ln k                             0.8848 *** (0.0099) 1.7527 *** (0.02640)
## ln w                             4.6549 *** (0.0788) 0.9332 *** (0.02255)
## $N_R$                           1.3243 *** (0.02405) 7.8101 *** (0.12555)
## $N_{CIP}$                       1.3073 *** (0.02305)                  NA
## $N_{SA}$                                             1.2158 *** (0.01070)
## $I_{2014+}$
## $I_{R_{NA}}$
## $I_{R_{EU}}$
## $I_{R_{AA}}$
## $I_{R_{NA}} \\times I_{2014+}$
## $I_{R_{EU}} \\times  I_{2014+}$
## $I_{R_{AA}} \\times  I_{2014+}$
## $\\textit{N}$                                 602599               602599
## Pseudo $R^2$                                 0.07253               0.1945
##                                              X.XSACIP                XSA.1
## y                                1.0461 *** (0.00160) 1.0327 *** (0.00085)
## $\\bar{z_j}$                      1.415 *** (0.03325)  0.978 *** (0.01180)
## ln k                             1.5616 *** (0.03400) 0.8968 *** (0.01010)
## ln w                              4.858 *** (0.15630) 4.6775 *** (0.07935)
## $N_R$                           12.0039 *** (0.31475) 1.2108 *** (0.04920)
## $N_{CIP}$                                        NA   1.2944 *** (0.02295)
## $N_{SA}$                                         NA                   NA
## $I_{2014+}$                                              0.9488 * (0.03800)
## $I_{R_{NA}}$                                           0.9131 *** (0.04340)
## $I_{R_{EU}}$                                             0.9422 * (0.04515)
## $I_{R_{AA}}$                                           0.7463 *** (0.03675)
## $I_{R_{NA}} \\times I_{2014+}$                          1.0743 ** (0.04955)
## $I_{R_{EU}} \\times  I_{2014+}$                          0.9553   (0.04450)
## $I_{R_{AA}} \\times  I_{2014+}$                        1.1112 *** (0.05735)
## $\\textit{N}$                                  207281               602599
## Pseudo $R^2$                                   0.3081              0.07346
##                                              XCIP.1              XSA.CIP
## y                               1.0205 *** (0.00150) 1.0609 *** (0.00210)
## $\\bar{z_j}$                    1.2232 *** (0.02040) 1.3077 *** (0.03150)
## ln k                             1.821 *** (0.02765) 1.6387 *** (0.03610)
## ln w                            0.9292 *** (0.02260) 4.8956 *** (0.15910)
## $N_R$                           3.0284 *** (0.11255) 4.5694 *** (0.27500)
## $N_{CIP}$                                      NA                   NA
## $N_{SA}$                         1.206 *** (0.01070)                NA
## $I_{2014+}$                       0.754 *** (0.02430) 0.7163 *** (0.04290)
## $I_{R_{NA}}$                      0.38 *** (0.01655) 0.3974 *** (0.02745)
## $I_{R_{EU}}$                     0.313 *** (0.01405) 0.3448 *** (0.02415)
## $I_{R_{AA}}$                    0.2295 *** (0.01165) 0.1914 *** (0.01465)
## $I_{R_{NA}} \\times I_{2014+}$     1.0065   (0.04460)   1.0115   (0.07175)
## $I_{R_{EU}} \\times  I_{2014+}$    1.0384   (0.04935)   0.9381   (0.06915)
## $I_{R_{AA}} \\times  I_{2014+}$  0.8977 ** (0.06070)   0.9505   (0.08935)
## $\\textit{N}$                                 602599               207281
## Pseudo $R^2$                                  0.2045               0.3191
```

Table S2:

```
df_S2 <- read.csv("S2.csv", header=T, row.names=1)
df_S2[is.na(df_S2)] <- " "
  kbl(df_S2,caption = "Table:S2",
      col.names =c("(1) \n $\\tiny X_{Distant,SA}$","(2) \n $\\tiny
X_{Distant,CIP}$","(3) \n $\\tiny X_{Neibor,SA\\&CIP}$","(4) \n $\\tiny
X_{Distant,SA}$","(5) \n $\\tiny X_{Distant,CIP}$","(6) \n $\\tiny
X_{Distant,SA \\& CIP}$"),format = "latex", escape = F,booktabs=T)  %>%
pack_rows(" ", 1, 7, hline_before = T) %>%
  pack_rows(" ", 8, 14,hline_before = T) %>%
  pack_rows(" ", 15, 16,hline_before = T)%>%
```

```r
  column_spec(c(0:7), width = "6em") %>%
  kable_classic(full_width = F, html_font = "Cambria")

  #Table S2
  df_S2
```

```
##                                                  XSA                 XCIP
## y                                  1.03 *** (0.0008)  1.0019 ** (0.00130)
## $\\bar{z_j}$                   1.4876 *** (0.01745) 1.3439 *** (0.02630)
## ln k                                0.5304 *** (0.0058)  1.755 *** (0.03110)
## ln w                                1.756 *** (0.02865) 0.8891 *** (0.02570)
## $N_R$                           1.7625 *** (0.02645) 6.2972 *** (0.11395)
## $N_{CIP}$                       1.4289 *** (0.01975)
## $N_{SA}$                                             1.2301 *** (0.01285)
## $I_{2014+}$
## $I_{R_{NA}}$
## $I_{R_{EU}}$
## $I_{R_{AA}}$
## $I_{R_{NA}} \\times I_{2014+}$
## $I_{R_{EU}} \\times  I_{2014+}$
## $I_{R_{AA}} \\times  I_{2014+}$
## $\\textit{N}$                                 602599               602599
## Pseudo $R^2$                                  0.0496              0.17157
##                                             X.XSACIP                XSA.1
## y                               1.0255 *** (0.00210) 1.0275 *** (0.00095)
## $\\bar{z_j}$                   1.7655 *** (0.05025) 1.4278 *** (0.01705)
## ln k                                1.132 *** (0.02920) 0.5432 *** (0.00600)
## ln w                            1.8158 *** (0.07365) 1.7879 *** (0.02930)
## $N_R$                           8.4242 *** (0.22290) 1.8532 *** (0.06005)
## $N_{CIP}$                                            1.4154 *** (0.01965)
## $N_{SA}$
## $I_{2014+}$                                            1.0292   (0.03160)
## $I_{R_{NA}}$                                         1.1223 *** (0.04335)
## $I_{R_{EU}}$                                         1.1918 *** (0.04650)
## $I_{R_{AA}}$                                         0.6259 *** (0.02655)
## $I_{R_{NA}} \\times I_{2014+}$                        1.0528 ** (0.03850)
## $I_{R_{EU}} \\times  I_{2014+}$                        1.0435 * (0.03900)
## $I_{R_{AA}} \\times  I_{2014+}$                      1.2744 *** (0.05930)
## $\\textit{N}$                                 430801               602599
## Pseudo $R^2$                                 0.19193              0.05545
##                                               XCIP.1              XSA.CIP
## y                               1.0117 *** (0.00175) 1.0357 *** (0.00280)
## $\\bar{z_j}$                    1.266 *** (0.02515) 1.6456 *** (0.04755)
## ln k                            1.8324 *** (0.03260)  1.188 *** (0.03070)
## ln w                            0.8886 *** (0.02575) 1.7995 *** (0.07290)
## $N_R$                           2.6172 *** (0.10510) 4.0712 *** (0.22645)
## $N_{CIP}$
## $N_{SA}$                        1.2154 *** (0.01275)
## $I_{2014+}$                     0.7704 *** (0.02690) 0.7949 *** (0.04055)
## $I_{R_{NA}}$                    0.4052 *** (0.01970) 0.5113 *** (0.03520)
## $I_{R_{EU}}$                    0.3269 *** (0.01660)  0.404 *** (0.02965)
## $I_{R_{AA}}$                    0.1718 *** (0.01100) 0.1557 *** (0.01560)
## $I_{R_{NA}} \\times I_{2014+}$    1.0398    (0.05420)   1.0094    (0.07635)
## $I_{R_{EU}} \\times  I_{2014+}$ 1.1137 *** (0.06315)   1.0349    (0.08885)
## $I_{R_{AA}} \\times  I_{2014+}$   1.0806    (0.10095)   1.2098 * (0.17825)
## $\\textit{N}$                                 602599               430801
## Pseudo $R^2$                                  0.1837              0.20409
```

Table S3:

```
df_S3 <- read.csv("S3.csv", header=T, row.names=1)
df_S3[is.na(df_S3)] <- " "
  kbl(df_S3,caption = "Table:S3",
      col.names =c("(1) \n $\\tiny X_{Distant,SA}$","(2) \n $\\tiny
X_{Distant,CIP}$","(3) \n $\\tiny X_{Distant,SA\\&CIP}$","(4) \n $\\tiny
X_{Distant,SA}$","(5) \n $\\tiny X_{Distant,CIP}$","(6) \n $\\tiny
X_{Distant,SA \\& CIP}$"),format = "latex", escape = F,booktabs=T)  %>%
pack_rows(" ", 1, 7, hline_before = T) %>%
  pack_rows(" ", 8, 14,hline_before = T) %>%
  pack_rows(" ", 15, 16,hline_before = T)%>%
  column_spec(c(0:7), width = "6em") %>%
  kable_classic(full_width = F, html_font = "Cambria")

  #Table S3
  df_S3
```

```
##                                                  XSA                 XCIP
## y                                  1.033 *** (7e-04) 1.0175 *** (0.00275)
## $\\bar{z_j}$                    0.6345 *** (0.00705) 1.2104 *** (0.04430)
## ln k                             0.8673 *** (0.0088)   1.74 *** (0.05365)
## ln w                             2.2584 *** (0.0348)  0.9184 ** (0.04975)
## $N_R$                            1.1073 *** (0.0164) 4.5944 *** (0.13670)
## $N_{CIP}$                       1.1813 *** (0.01605)
## $N_{SA}$                                              1.183 *** (0.02300)
## $I_{2014+}$
## $I_{R_{NA}}$
## $I_{R_{EU}}$
## $I_{R_{AA}}$
## $I_{R_{NA}} \\times I_{2014+}$
## $I_{R_{EU}} \\times  I_{2014+}$
## $I_{R_{AA}} \\times  I_{2014+}$
## $\\textit{N}$                                 602599               602599
## Pseudo $R^2$                                 0.03657              0.12735
##                                             X.XSACIP                XSA.1
## y                               1.0435 *** (0.00470) 1.0364 *** (0.00090)
## $\\bar{z_j}$                     0.838 *** (0.04970) 0.6236 *** (0.00705)
## ln k                            1.2889 *** (0.06515)  0.879 *** (0.00895)
## ln w                             2.584 *** (0.20975) 2.2607 *** (0.03485)
## $N_R$                           5.0941 *** (0.23390)   0.9855   (0.03235)
## $N_{CIP}$                                            1.1686 *** (0.01595)
## $N_{SA}$
## $I_{2014+}$                                         0.8708 *** (0.02665)
## $I_{R_{NA}}$                                        0.8721 *** (0.03400)
## $I_{R_{EU}}$                                        0.8941 *** (0.03510)
## $I_{R_{AA}}$                                        0.7247 *** (0.02970)
## $I_{R_{NA}} \\times I_{2014+}$                       1.0632 *** (0.03830)
## $I_{R_{EU}} \\times  I_{2014+}$                        1.0314   (0.03780)
## $I_{R_{AA}} \\times  I_{2014+}$                      1.1177 *** (0.04675)
## $\\textit{N}$                                 396471               602599
## Pseudo $R^2$                                0.12162              0.03753
##                                              XCIP.1              XSA.CIP
## y                              1.0354 *** (0.00385) 1.0714 *** (0.00675)
## $\\bar{z_j}$                   1.1268 *** (0.04165) 0.7503 *** (0.04510)
## ln k                           1.8614 *** (0.05785)  1.403 *** (0.07170)
## ln w                           0.8943 *** (0.04840) 2.4957 *** (0.20290)
## $N_R$                          1.6518 *** (0.11140) 1.9236 *** (0.19885)
## $N_{CIP}$
## $N_{SA}$                       1.1706 *** (0.02270)
## $I_{2014+}$                    0.7353 *** (0.04330) 0.6479 *** (0.06035)
## $I_{R_{NA}}$                   0.3784 *** (0.03255) 0.4495 *** (0.05905)
## $I_{R_{EU}}$                   0.1227 *** (0.01325) 0.1325 *** (0.02220)
## $I_{R_{AA}}$                   0.1884 *** (0.02100) 0.1302 *** (0.02465)
## $I_{R_{NA}} \\times I_{2014+}$   0.8598 ** (0.08035)   0.8419 * (0.11885)
## $I_{R_{EU}} \\times  I_{2014+}$  1.2275 ** (0.16760)   1.0393   (0.23190)
## $I_{R_{AA}} \\times  I_{2014+}$ 0.6456 *** (0.11735)   0.7108 * (0.22600)
## $\\textit{N}$                                 602599               396471
## Pseudo $R^2$                                0.14919              0.14739
```

Table S4:

```
df_S4 <- read.csv("S4.csv", header=T, row.names=1)
df_S4[is.na(df_S4)] <- " "

  kbl(df_S4,caption = "Table:S4",
      col.names =c("(1) \n $ z_{p}$","(2) \n $ z_{p}$","(3) \n $ z_{p}$","(4)
\n $ z_{p}$","(5) \n $ z_{p}$","(6) \n $ z_{p}$"),format = "latex", escape =
F,booktabs=T)  %>% pack_rows(" ", 1, 3, hline_before = T) %>%
  pack_rows(" ", 4, 5,hline_before = T) %>%
    pack_rows(" ", 6, 7,hline_before = T) %>%
    pack_rows(" ", 8, 9,hline_before = T) %>%
  pack_rows(" ", 10, 12,hline_before = T)%>%
    pack_rows(" ", 13, 17,hline_before = T)%>%
    pack_rows(" ", 18, 19,hline_before = T)%>%
  column_spec(c(1:7), width = "6em") %>%
  kable_classic(full_width = F, html_font = "Cambria")

  #Table S4
  df_S4
```

```
##                                          model1_full
model2_full
## ln k                            0.4041*** (0.001803)    0.4051***
```

```
(0.0018)
## ln w                             0.03258*** (0.00293) 0.03959***
(0.002895)
## t                              -0.0008777  (0.004357) -0.001156
(0.004356)
## $I_{XSA}$                       0.04563*** (0.002659)
## $I_{XCIP}$                       0.07556*** (0.00294)
## $I_{X_{Neighboring,SA}}$                              0.08808***
(0.003108)
## $I_{X_{Neighboring,CIP}}$                             0.07455***
(0.003169)
## $I_{X_{Distant,SA}}$
## $I_{X_{Distant,CIP}}$
## $I_{X_{SA\\&CIP}}$
## $I_{X_{Neighboring,SA\\&CIP}}$
## $I_{X_{Distant,SA\\&CIP}}$
## constant                            -0.2905 (0.07396)     -0.2899
(0.07394)
## year dummy                                          Y
Y
## topic category dummy                                Y
Y
## department category dummy                           Y
Y
## Region dummy                                        Y
Y
## N                                              864590
864590
## adj.$R^2$                                      0.1002
0.1007
## F                                                1484
1491
## \\# researcher profiles                          8988
8988
##                                           model3_full
model4_full
## ln k                             0.4105*** (0.001797)  0.4128***
(0.002701)
## ln w                            0.04211*** (0.002905) 0.03077***
(0.004177)
## t                               -0.0009875  (0.00436)    0.0023
(0.007007)
## $I_{XSA}$
## $I_{XCIP}$
## $I_{X_{Neighboring,SA}}$
## $I_{X_{Neighboring,CIP}}$
## $I_{X_{Distant,SA}}$           -0.01132*** (0.002849)
## $I_{X_{Distant,CIP}}$             0.01055  (0.006261)
## $I_{X_{SA\\&CIP}}$                                     0.1405***
(0.004956)
```

```
## $I_{X_{Neighboring,SA\\&CIP}}$
## $I_{X_{Distant,SA\\&CIP}}$
## constant                            -0.2533 (0.07399)      -0.2161
(0.1201)
## year dummy                                          Y
Y
## topic category dummy                                Y
Y
## department category dummy                           Y
Y
## Region dummy                                        Y
Y
## N                                              864590
378779
## adj.$R^2$                                     0.09922
0.1137
## F                                                1469
821.2
## \\# researcher profiles                          8988
8839
##                                           model5_full           model6_full
## ln k                            0.4155*** (0.002209)  0.3983*** (0.002226)
## ln w                            0.0513*** (0.003429) 0.04094*** (0.003377)
## t                              -0.001008  (0.004693) -0.001355  (0.004983)
## $I_{XSA}$
## $I_{XCIP}$
## $I_{X_{Neighboring,SA}}$
## $I_{X_{Neighboring,CIP}}$
## $I_{X_{Distant,SA}}$
## $I_{X_{Distant,CIP}}$
## $I_{X_{SA\\&CIP}}$
## $I_{X_{Neighboring,SA\\&CIP}}$  0.1321*** (0.005652)
## $I_{X_{Distant,SA\\&CIP}}$                            0.03831*** (0.01053)
## constant                           -0.3635 (0.07806)      -0.241 (0.07939)
## year dummy                                         Y                     Y
## topic category dummy                               Y                     Y
## department category dummy                          Y                     Y
## Region dummy                                       Y                     Y
## N                                             581100                558821
## adj.$R^2$                                    0.08548               0.08612
## F                                              904.6                 880.6
## \\# researcher profiles                         8933                  8912
```

Table S5:

```
df_S5 <- read.csv("S5.csv", header=T, row.names=1)
df_S5[is.na(df_S5)] <- " "

  kbl(df_S5,caption = "Table:S5",
      col.names =c("(1) \n $ z_{p}$","(2) \n $ z_{p}$","(3) \n $
z_{p}$"),format = "latex", escape = F,booktabs=T)  %>% pack_rows(" ", 1, 4,
hline_before = T) %>%
  pack_rows(" ", 5, 6,hline_before = T) %>%
    pack_rows(" ", 7, 8,hline_before = T) %>%
    pack_rows(" ", 9, 10,hline_before = T) %>%
  pack_rows(" ", 11, 15,hline_before = T)%>%
    pack_rows(" ", 16, 19,hline_before = T)%>%
    #column_spec(c(0:2), width = "5em") %>%
  column_spec(0, width = "10em") %>%
  kable_classic(full_width = F, html_font = "Cambria")

  #Table S5
  df_S5
```

```
##                                                             model1_full
## ln k                                              0.4382*** (0.002953)
## ln w                                             0.02491*** (0.004838)
## t                                                 -0.01838* (0.009258)
## $I_{2014+}$                                         0.04437  (0.04239)
## $I_{X_{SA\\&CIP}}$                                0.1554*** (0.005306)
## $I_{X_{SA\\&CIP}}\\times I_{2014+}$             -0.08808*** (0.007352)
## $I_{X_{Neighboring,SA\\&CIP}}$
## $I_{X_{Neighboring,SA\\&CIP}}\\times I_{2014+}$
## $I_{X_{Distant,SA\\&CIP}}$
## $I_{X_{Distant,SA\\&CIP}}\\times I_{2014+} $
## constant                                              -0.2714 (0.2348)
## year dummy                                                            Y
## topic category dummy                                                  Y
## department category dummy                                             Y
## Region dummy                                                          Y
## N                                                                357859
## adj.$R^2$                                                        0.1104
## F                                                                 713.7
## \\# researcher profiles                                            8345
##                                                            model2_full
## ln k                                             0.4247*** (0.002445)
## ln w                                            0.05446*** (0.003888)
## t                                                -0.01392  (0.007633)
## $I_{2014+}$                                        0.02929  (0.03478)
## $I_{X_{SA\\&CIP}}$
## $I_{X_{SA\\&CIP}}\\times I_{2014+}$
## $I_{X_{Neighboring,SA\\&CIP}}$                   0.1822*** (0.006124)
## $I_{X_{Neighboring,SA\\&CIP}}\\times I_{2014+}$    -0.16*** (0.007772)
## $I_{X_{Distant,SA\\&CIP}}$
## $I_{X_{Distant,SA\\&CIP}}\\times I_{2014+} $
## constant                                             -0.3779 (0.2014)
## year dummy                                                           Y
## topic category dummy                                                 Y
## department category dummy                                            Y
## Region dummy                                                         Y
## N                                                              551771
## adj.$R^2$                                                     0.07718
## F                                                               737.5
## \\# researcher profiles                                          8358
##                                                          model3_full
## ln k                                             0.4058*** (0.002468)
## ln w                                            0.03841*** (0.003888)
## t                                                -0.01751* (0.007844)
## $I_{2014+}$                                       -0.02203  (0.03568)
## $I_{X_{SA\\&CIP}}$
## $I_{X_{SA\\&CIP}}\\times I_{2014+}$
## $I_{X_{Neighboring,SA\\&CIP}}$
## $I_{X_{Neighboring,SA\\&CIP}}\\times I_{2014+}$
## $I_{X_{Distant,SA\\&CIP}}$                         0.02795* (0.01153)
## $I_{X_{Distant,SA\\&CIP}}\\times I_{2014+} $      0.04333** (0.01389)
## constant                                             -0.1029 (0.2089)
## year dummy                                                          Y
## topic category dummy                                                Y
## department category dummy                                           Y
## Region dummy                                                        Y
## N                                                              526904
## adj.$R^2$                                                     0.07817
## F                                                               717.8
## \\# researcher profiles                                          8364
```

Figure 5A:

```
library(car)

## Warning: package 'car' was built under R version 4.0.5

## Loading required package: carData

library(stats)
library(ggplot2)
library(readr)
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse
1.3.0 --

## v tibble  3.0.5     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v purrr   0.3.4     v forcats 0.5.1

## -- Conflicts ------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter()     masks stats::filter()
## x dplyr::group_rows() masks kableExtra::group_rows()
## x dplyr::lag()        masks stats::lag()
## x dplyr::recode()     masks car::recode()
## x purrr::some()       masks car::some()
```

```r
A_data <- data.frame(df_S1[1,1],df_S1[1,2],df_S1[1,3],
                      df_S2[1,1],df_S2[1,2],df_S2[1,3],
                      df_S3[1,1], df_S3[1,2],df_S3[1,3])
A_data_split <- str_split_fixed(A_data, " ",3)
A_data_coef <- as.numeric(A_data_split[,1])
A_data_SE <- parse_number(A_data_split[,3])

values <- data.frame(Beta=100*(A_data_coef-1),
               se=100*c(A_data_SE)*1.96,

Type=c("Broad","Broad","Broad","Neighboring","Neighboring","Neighboring","Dis
tant","Distant","Distant"),
               Domain=c("SA","CIP","SA and CIP","SA","CIP","SA and
CIP","SA","CIP","SA and CIP"))

values$Type<-factor(values$Type,levels=c("Broad","Neighboring","Distant"))
values$Domain<-factor(values$Domain,levels = c("SA","CIP","SA and CIP"))

pd = position_dodge(2)

plot1 <- ggplot(values, aes(
    x = Type,
    y = Beta,
    color = factor(Domain)
)) +
    geom_errorbar(
        aes(ymin = Beta-se,
            ymax = Beta+se),
        width = .2,
        size = .7,
        position = position_dodge(.5)
    ) +
    geom_point(shape = 15,
               size = 3,
               position = position_dodge(.5)) +
    theme_bw() +
    theme(
        legend.position = "top",
        legend.direction = "horizontal",
        legend.box = "horizontal",
        panel.grid = element_blank(),
        axis.title.y = element_text(vjust = 1.8),
        axis.title.x = element_text(vjust = -0.5),
        axis.title = element_text(face = "bold")
    ) +
    scale_color_manual(label=c("SA","CIP","SA and CIP"),values = c("light
grey", "dark grey", "black")) +
    geom_hline(yintercept=0, linetype="dashed", color = "black") +
    ylab("Percent increase in Odds")
```

```
#Plot Figure 5A
plot1
```

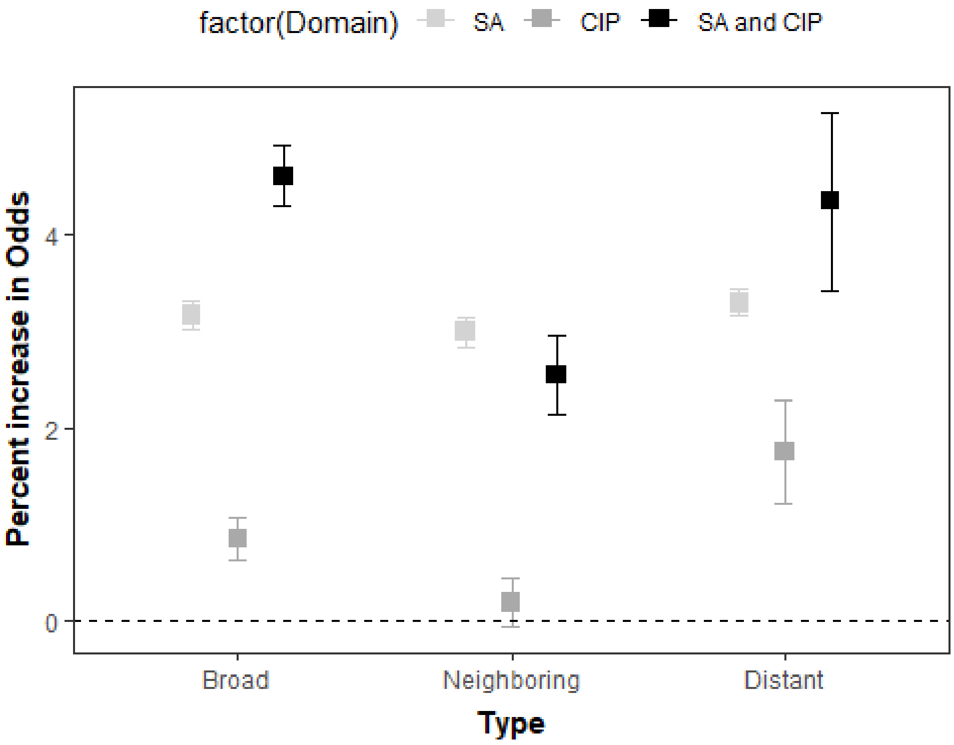


Figure 5B:

```
C_data <- data.frame(df_S1[8,4],df_S1[8,5],df_S1[8,6],
                     df_S2[8,4],df_S2[8,5],df_S2[8,6],
                     df_S3[8,4], df_S3[8,5],df_S3[8,6])
C_data_split <- str_split_fixed(C_data, " ",3)
C_data_coef <- as.numeric(C_data_split[,1])
C_data_SE <- parse_number(C_data_split[,3])

values <- data.frame(Beta=100*(C_data_coef-1),
               se=100*c(C_data_SE)*1.96,

Type=c("Broad","Broad","Broad","Neighboring","Neighboring","Neighboring","Dis
tant","Distant","Distant"),
               Domain=c("SA","CIP","SA and CIP","SA","CIP","SA and
CIP","SA","CIP","SA and CIP"))

values$Type<-factor(values$Type,levels=c("Broad","Neighboring","Distant"))
values$Domain<-factor(values$Domain,levels = c("SA","CIP","SA and CIP"))

pd = position_dodge(2)
```

```
plot2 <- ggplot(values, aes(
    x = Type,
    y = Beta,
    color = factor(Domain)
)) +
    geom_errorbar(
        aes(ymin = Beta-se,
            ymax = Beta+se),
        width = .2,
        size = .7,
        position = position_dodge(.5)
    ) +
    geom_point(shape = 15,
               size = 3,
               position = position_dodge(.5)) +
    theme_bw() +
    theme(
        legend.position = "top",
        legend.direction = "horizontal",
        legend.box = "horizontal",
        panel.grid = element_blank(),
        axis.title.y = element_text(vjust = 1.8),
        axis.title.x = element_text(vjust = -0.5),
        axis.title = element_text(face = "bold")
    ) +
    scale_color_manual(label=c("SA","CIP","SA and CIP"),values = c("light
grey", "dark grey", "black")) +
    geom_hline(yintercept=0, linetype="dashed", color = "black") +
    ylab("Percent increase in Odds")

#Plot Figure 5B
plot2
```

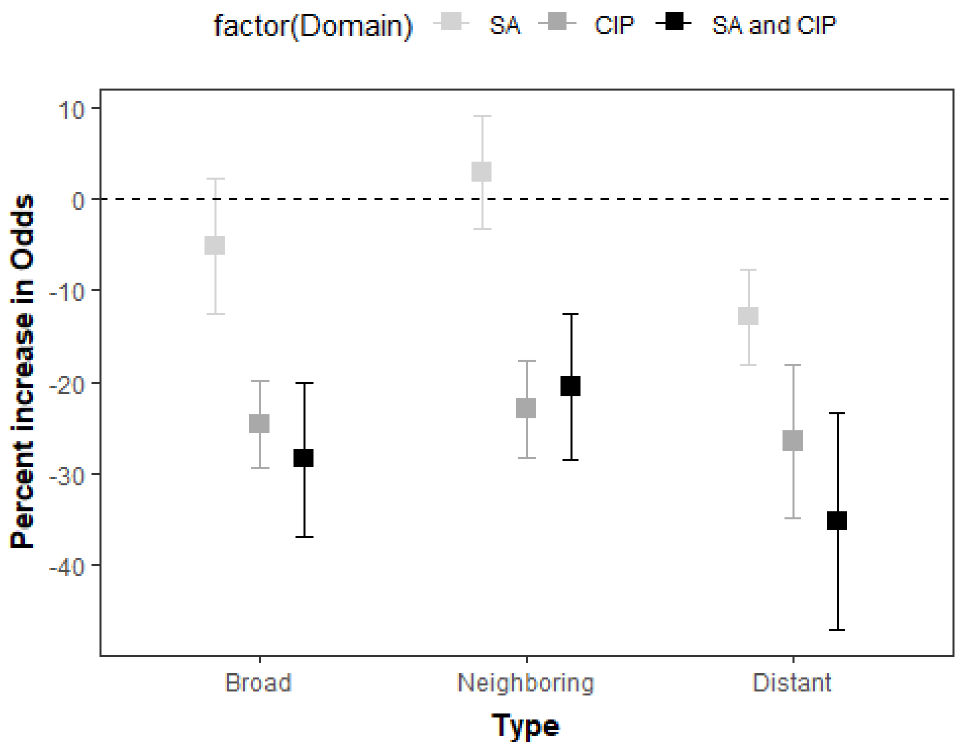


Figure 5C:

```
C_data <- data.frame(df_S4[4,1],df_S4[5,1],df_S4[10,4],
                      df_S4[6,2],df_S4[7,2],df_S4[11,5],
                      df_S4[8,3], df_S4[9,3],df_S4[12,6])
C_data_split <- str_split_fixed(C_data, " ",3)
C_data_coef <- parse_number(C_data_split[,1])
C_data_SE <- parse_number(C_data_split[,2])

values <- data.frame(Beta=100*(C_data_coef*1.24),
               se=100*c(C_data_SE*1.24)*1.96,

Type=c("Broad","Broad","Broad","Neighboring","Neighboring","Neighboring","Dis
tant","Distant","Distant"),
               Domain=c("SA","CIP","SA and CIP","SA","CIP","SA and
CIP","SA","CIP","SA and CIP"))

values$Type<-factor(values$Type,levels=c("Broad","Neighboring","Distant"))
values$Domain<-factor(values$Domain,levels = c("SA","CIP","SA and CIP"))

pd = position_dodge(2)

plot3<- ggplot(values, aes(
    x = Type,
    y = Beta,
    color = factor(Domain)
)) +
    geom_errorbar(
        aes(ymin = Beta-se,
            ymax = Beta+se),
        width = .2,
        size = .7,
        position = position_dodge(.5)
    ) +
    geom_point(shape = 15,
               size = 3,
               position = position_dodge(.5)) +
    theme_bw() +
    theme(
        legend.position = "top",
        legend.direction = "horizontal",
        legend.box = "horizontal",
        panel.grid = element_blank(),
        axis.title.y = element_text(vjust = 1.8),
        axis.title.x = element_text(vjust = -0.5),
        axis.title = element_text(face = "bold")
    ) +
    scale_color_manual(label=c("SA","CIP","SA and CIP"),values = c("light
grey", "dark grey", "black")) +
    geom_hline(yintercept=0, linetype="dashed", color = "black") +
    ylab("Percent increase in citations associated with X")

#Plot Figure 5C
plot3
```

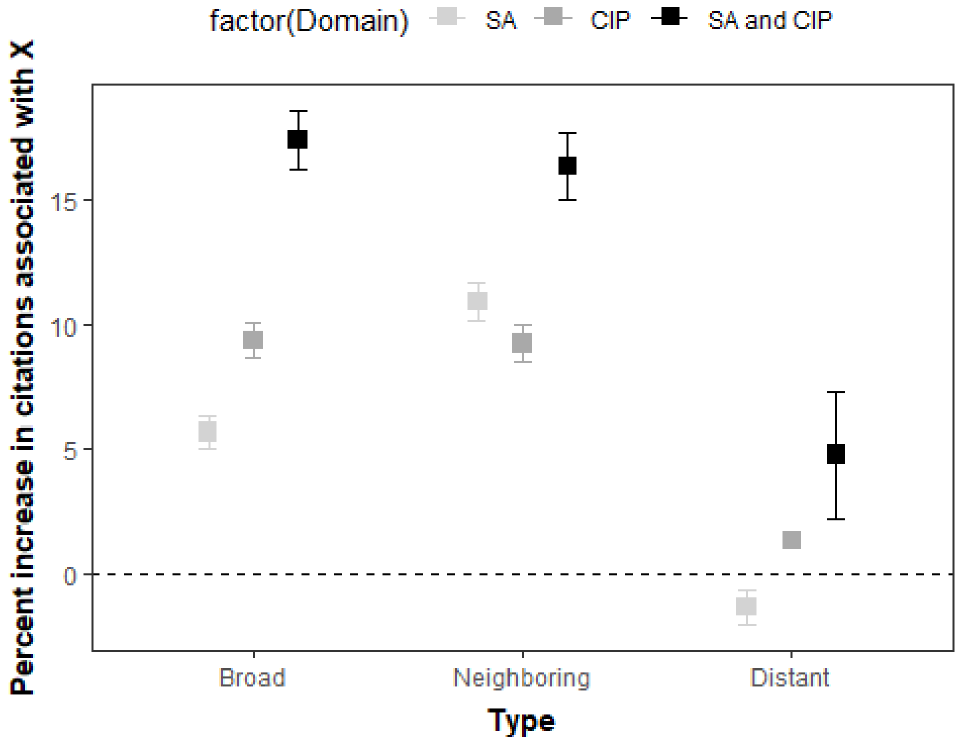
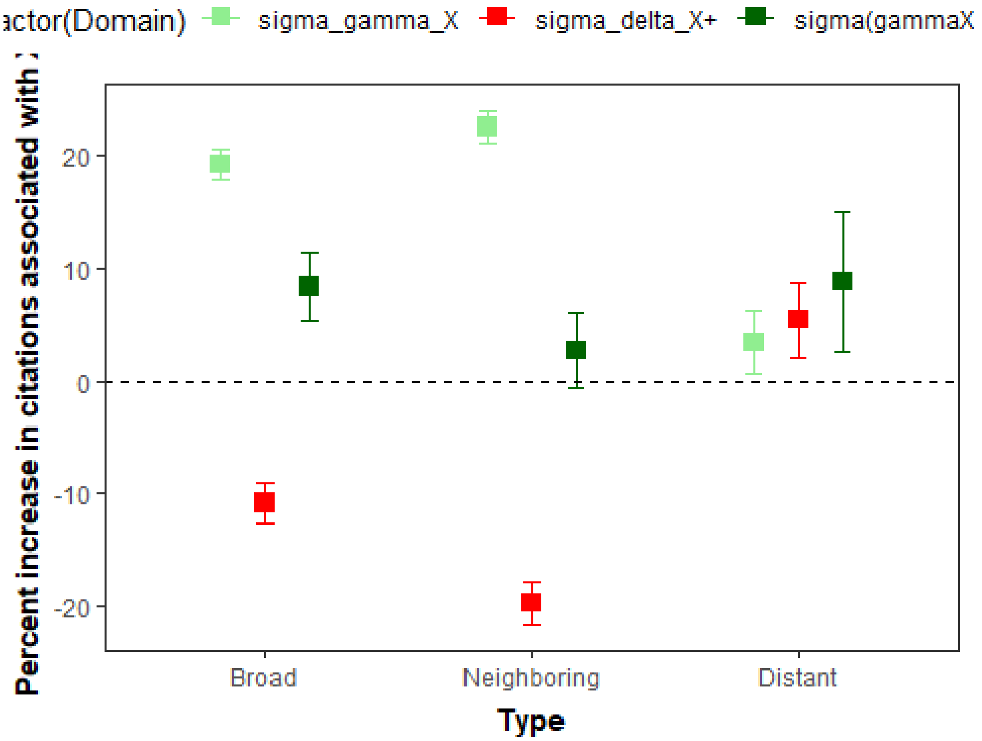


Figure 5D:

```
C_data <- data.frame(df_S5[5,1],df_S5[6,1],df_S5[6,1],
                      df_S5[7,2],df_S5[8,2],df_S5[8,2],
                      df_S5[9,3], df_S5[10,3],df_S5[10,3])
C_data_split <- str_split_fixed(C_data, " ",3)
C_data_coef <- parse_number(C_data_split[,1])
C_data_coef[3] <-C_data_coef[1]+C_data_coef[3]
C_data_coef[6] <-C_data_coef[4]+C_data_coef[6]
C_data_coef[9] <-C_data_coef[7]+C_data_coef[9]
C_data_SE <- parse_number(C_data_split[,2])
C_data_SE[3] <-C_data_SE[1]+C_data_SE[3]
C_data_SE[6] <-C_data_SE[4]+C_data_SE[6]
C_data_SE[9] <-C_data_SE[7]+C_data_SE[9]
values <- data.frame(Beta=100*(C_data_coef*1.24),
               se=100*c(C_data_SE*1.24)*1.96,

Type=c("Broad","Broad","Broad","Neighboring","Neighboring","Neighboring","Dis
tant","Distant","Distant"),

Domain=c("sigma_gamma_X","sigma_delta_X+","sigma(gammaX+deltaX+)","sigma_gamm
a_X","sigma_delta_X+","sigma(gammaX+deltaX+)","sigma_gamma_X","sigma_delta_X+
","sigma(gammaX+deltaX+)"))

values$Type<-factor(values$Type,levels=c("Broad","Neighboring","Distant"))
values$Domain<-factor(values$Domain,levels =
c("sigma_gamma_X","sigma_delta_X+","sigma(gammaX+deltaX+)"))
```

```r
pd = position_dodge(2)

plot4<- ggplot(values, aes(
    x = Type,
    y = Beta,
    color = factor(Domain)
)) +
    geom_errorbar(
        aes(ymin = Beta-se,
            ymax = Beta+se),
        width = .2,
        size = .7,
        position = position_dodge(.5)
    ) +
    geom_point(shape = 15,
               size = 3,
               position = position_dodge(.5)) +
    theme_bw() +
    theme(
        legend.position = "top",
        legend.direction = "horizontal",
        legend.box = "horizontal",
        panel.grid = element_blank(),
        axis.title.y = element_text(vjust = 1.8),
        axis.title.x = element_text(vjust = -0.5),
        axis.title = element_text(face = "bold")
    ) +

scale_color_manual(label=c("sigma_gamma_X","sigma_delta_X+","sigma(gammaX+deltaX+)"),values = c("light green", "red", "dark green")) +
    geom_hline(yintercept=0, linetype="dashed", color = "black") +
    ylab("Percent increase in citations associated with X")

#Plot Figure 5D
plot4
```



Conclusion:

As per our study for table S1 to S5, we conclude that:

Table S1:

1. Table S1 shows the results of a logistic regression model of the cross-domain activity at an article level.

2. The researchers from different backgrounds collaborating (SA) by co-authoring published papers in differing categories (CIP).

3. Data was filtered for cross-domain activity by the years (Yp) from 1970 to 2018, the number of co-authors of a paper (Kp) ≥ 2, and the number Medical Subject Heading (MeSH) words ≥ 2 to distinguish mono-domain versus cross-domain activity.

4. The parameters contribute to the total information of predicting papers that have cross domain activity with numbers reported in odds ratios (what are the odds of cross-domain activity based on a high value of certain parameters).

5. The columns are each separate models that show what parameters contribute to SA, CIP, or SACIP in the logistic regressions. Robust error estimations were also made which act as confidence intervals for the estimated contribution of the parameters.

6. The log number of MeSH words (w) relate more to the SA and SACIP. The region the paper was published (NRegp) appears to contribute highly to CIP and SACIP. This

suggests that researchers in certain regions of the world historically publish papers ranging from a wide variety of subject fields from their own.

7. The pseudo R-square value also showed that the interactions of years papers were published contributed very little to the logistic regression model, suggesting that these trends are prevalent from the 1970s to 2018 rather than just 2014.

Table S2:

1. The neighboring and shorter-distance cross-domain combinations. When a cross-domain article is published in a relatively "close" category (for example a neuroscience researcher publishes a paper in Biology), it is considered a "neighbor".

2. Data was filtered for neighborSA, neighborCIP, and neighborSACIP. neighborSA activity was found by finding if any research was in SA category 1 to 4, neighborCIP was found by finding id any research was published in any CIP category of 1 to 7. neighborSACIP was found by checking if the article had both neighborCIP and neighborSA. Data was filtered for cross-domain activity by the years (Yp) from 1970 to 2018, the number of co-authors of a paper (Kp) ≥ 2, and the number Medical Subject Heading (MeSH) words ≥ 2 to distinguish mono-domain versus cross-domain activity.

3. The parameters contribute to the total information of predicting papers that have cross domain activity with numbers reported in odds ratios (what are the odds of cross-domain activity based on a high value of certain parameters).

4. The columns are each separate models that show what parameters contribute to SA, CIP, or SACIP in the logistic regressions. Robust error estimations were also made which act as confidence intervals for the estimated contribution of the parameters. In SA columns, the w and NRegp seemed to relate weakly to the logistic regression (with a very low pseudo R^2 of ~ 0.05) while in CIP and SACIP, the nRegp appears to be much more strongly related and contributed the largest amount to the regression information (with pseudo-R^2 without year interactions of about 0.17 and 0.19).

5. The pseudo R^2 value also showed that the interactions of years papers were published contributed very little to the logistic regression model (only adding about 0.01 to the value), suggesting that these trends are prevalent from the 1970s to 2018 rather than just 2014.

Table S3:

1. When a cross-domain article is published in a relatively far category (for example a neuroscience research co-author publishes a paper in Engineering), it is considered "distant".

2. Data was filtered for distantSA, distantCIP, and distantSACIP. distantSA activity was found by finding if a research paper was in SA category 1 to 4 AND in SA 5 to 6. distantCIP was found by finding if a research paper was in any CIP category 1,3, or 5 and also in CIP 4 or 8. distantSACIP was found by checking if the article had both distantCIP and distantSA. Data was filtered for cross-domain activity by the years (Yp)

from 1970 to 2018, the number of co-authors of a paper (Kp) ≥ 2, and the number Medical Subject Heading (MeSH) words ≥ 2 to distinguish mono-domain versus cross-domain activity.

3. The parameters contribute to the total information of predicting papers that have cross domain activity with numbers reported in odds ratios (what are the odds of cross-domain activity based on a high value of certain parameters). The columns are each separate models that show what parameters contribute to distantSA, distantCIP, or distantSACIP in the logistic regressions.

4. Robust error estimations were also made which act as confidence intervals for the estimated contribution of the parameters. In distantSA and distantSACIP columns, the log(w) seemed to relate more to the logistic regression information compared to the other parameters (with a very low pseudo R-square of 0.0375 without year interactions and 0.496 with year interactions) while in CIP and SACIP, the nRegp appears to also be strongly related and contributed the largest amount to the regression information (pseudo-R^2 without year interactions of about 0.149 and 0.147).

5. The pseudo R^2 value also showed that the interactions of years papers were published contributed more to the distantCIP and distantSACIP logistic regression models (adding about 0.025 and 0.45 to the odds of predicting the paper being cross-domain), suggesting that there was slight increases in distant cross-domain brain-related research published from 2014 to 2018.

Table S4:

1. The career-level analysis with individual researcher fixed effects shown in this table.

2. Data was filtered by the years (Yp) from 1970 to 2018, the number of co-authors of a paper (Kp) ≥ 2, the number of MeSH words ≥ 2, and researchers with number of articles published (Na) ≥ 10.

3. Robust standard errors are shown in parenthesis below each estimate and Y indicates additional fixed effects in the regression model. Each column is comparing normalized citation measures (Zp) with log(Kp), log(w), and the difference between the year the paper was published and the main authors first publication year (τ). There are additional parameters for each column comparing the “broad” SA and CIP, “neighbor” SA and CIP, “distant” SA and CIP, “broad” SACIP, “neighbor” SACIP, and “distant” SACIP.

4. The log number of coauthors (Kp) appears to be relatively the strongest positive correlation to predicting the number of citations a paper will receive. This is prevalent in all columns of data showing higher regression coefficients than any of the additional parameters as well.

5. The pseudo R^2 values are relatively low, showing a range of adjusted R^2 values of 0.09 to 0.13. The highest adjusted R^2 value of 0.13 is comparing zp with “broad” SACIP with also the lowest number of articles (N=358237) of the 6 columns. All of the

F-statistics were >> 1 ranging from 193 to 262 and every parameter robust standard error was statistically significant.

Table S5:

1. The Flagship Project Effect using career-level analysis with researcher fixed effects shown in this table.

2. Data was filtered by the years (Yp) from 1970 to 2018, Kp ≥ 2, w ≥ 2, and researchers with Na ≥ 10.

3. Robust standard errors are shown in parenthesis below each estimate and Y indicates additional fixed effects in the regression model. Each of the 3 columns are comparing Zp with log(Kp), log(w), and the whether the papers were published from years 2014 to 2018 (I_year). There are additional parameters for each column comparing the "broad" SACIP and "broad" SACIP cross interaction with I_year, "neighbor" SACIP and "neighbor" SACIP cross interaction with I_year, along with "distant" SACIP and "distant" SACIP cross-interaction with I_year.

4. The log number of coauthors (Kp) appears to be relatively the strongest positive correlation to predicting the number of citations a paper will receive. This is prevalent in all columns of data showing higher regression coefficients than any of the additional parameters as well. The adjusted R-square values are relatively low, showing a range of adjusted R-square values of 0.09 to 0.13. The highest adjusted R-square value of 0.13 is comparing zp with "broad" SACIP and the cross-interaction of "broad" SACIP ith I_year. It is also the lowest number of articles (N=358237) of the 3 columns.

5. All of the F-statistics were >> 1 ranging from 191 to 229 and every parameter robust standard error except "distant" SACIP parameter in column 3 was statistically significant.

As per our study for Figure 5, we conclude that:

1. Figure 5A finds annual growth rate likelihood of research having cross domain characteristics. Varying from 0% to 4% growth rate, the likelihood of SA&CIP papers increase the most in "broad" and "distant papers and SA papers have a slightly higher growth rate likelihood in"neighboring" categories.

2. Figure 5B shows the decreased probable likelihood of having cross-domain characteristics from 2014 onward. It appears there is a much more prominent decrease in CIP as well as SA&CIP papers after 2014.

3. Figure 5C analyzes the percent increase in citations relative to mono-domain research articles. Papers that are SA&CIP together show the strongest percentage increase of citations relative to mono domain papers in all 3 categories of broad, neighboring, and distant.

4. Figure 5D displays the Difference-in-Difference of the Flagship project effect on impact on citations when a paper is cross-domain research. For broad"= and neighboring

categories, having an SA paper appeared to give the highest difference-on-difference citation impact for cross-domain research. CIP published data on the other hand appeared to give a decreased impact on the citation impact of cross domain research.