Absolute and Conditional Convergence

Taufiqur Rohman

2022-05-25

This is one of the problem that Professor gave me for the final exam of Macroeconomics. He asked me to test the absolute and conditional convergence within several periods. In this project, I will show you the process from loading and extracting the data, cleaning and transforming the data, merging 3 excel files using SQL environment, and conducting regression to test the convergence theory.

Extract, Transform, and Load (ETL) Data

There are several data in the problem, separated into 3 csv or excel files. The "data-wdi.csv" contains all data from World Development Indicators, "barrolee-region.xlsx" contains the geographical data, and "halljones-gadp.xls" contains the GADP. All of this data are needed to answer the problems. Let me load the data first and save it into data frame.

```
df_base <- read.csv("~/Documents/RU/Macroeconomics II/Mid-term Exam/data/data-wdi.csv")
df_region <- read_excel("~/Documents/RU/Macroeconomics II/Mid-term Exam/data/barrolee-region.xlsx")
df_gadp <- read_excel("~/Documents/RU/Macroeconomics II/Mid-term Exam/data/halljones-gadp.xls")

head(df_base)
```

```
##                          Series.Name    Series.Code   Country.Name Country.Code
## 1 GDP per capita (constant 2015 US$) NY.GDP.PCAP.KD    Afghanistan          AFG
## 2 GDP per capita (constant 2015 US$) NY.GDP.PCAP.KD        Albania          ALB
## 3 GDP per capita (constant 2015 US$) NY.GDP.PCAP.KD        Algeria          DZA
## 4 GDP per capita (constant 2015 US$) NY.GDP.PCAP.KD American Samoa          ASM
## 5 GDP per capita (constant 2015 US$) NY.GDP.PCAP.KD        Andorra          AND
## 6 GDP per capita (constant 2015 US$) NY.GDP.PCAP.KD         Angola          AGO
##   X1970..YR1970. X1971..YR1971. X1972..YR1972. X1973..YR1973. X1974..YR1974.
## 1             NA             NA             NA             NA             NA
## 2             NA             NA             NA             NA             NA
## 3       2368.575       2042.663       2532.393        2558.03        2674.97
## 4             NA             NA             NA             NA             NA
## 5      35391.074      35159.467      36166.413       37123.26       37504.74
## 6             NA             NA             NA             NA             NA
##   X1975..YR1975. X1976..YR1976. X1977..YR1977. X1978..YR1978. X1979..YR1979.
## 1             NA             NA             NA             NA             NA
## 2             NA             NA             NA             NA             NA
## 3       2732.324        2878.61       2944.321       3123.377       3258.731
## 4             NA             NA             NA             NA             NA
## 5      36246.683       36175.32      36081.658      35551.736      34462.493
## 6             NA             NA             NA             NA             NA
##   X1980..YR1980. X1981..YR1981. X1982..YR1982. X1983..YR1983. X1984..YR1984.
```

```
## 1            NA            NA            NA            NA            NA
## 2      1740.505      1804.010      1818.367      1799.878      1740.347
## 3      3186.437      3182.260      3281.861      3352.700      3432.806
## 4            NA            NA            NA            NA            NA
## 5     34014.488     32669.462     31709.296     30896.951     30084.567
## 6      3550.083      3276.362      3162.004      3179.348      3252.108
##   X1985..YR1985. X1986..YR1986. X1987..YR1987. X1988..YR1988. X1989..YR1989.
## 1            NA            NA            NA            NA            NA
## 2      1735.290      1798.015      1748.578      1691.531      1808.646
## 3      3453.822      3366.939      3248.881      3128.338      3179.773
## 4            NA            NA            NA            NA            NA
## 5     29480.346     29178.543     29564.821     29854.800     30091.249
## 6      3248.608      3226.817      3242.579      3323.534      3212.663
##   X1990..YR1990. X1991..YR1991. X1992..YR1992. X1993..YR1993. X1994..YR1994.
## 1            NA            NA            NA            NA            NA
## 2      1606.296      1163.491      1086.438      1197.581      1305.001
## 3      3123.998      3011.500      2994.491      2867.196      2783.155
## 4            NA            NA            NA            NA            NA
## 5     30051.178     29642.711     28792.198     27517.549     27408.280
## 6      2998.781      2929.483      2669.435      1964.390      1927.500
##   X1995..YR1995. X1996..YR1996. X1997..YR1997. X1998..YR1998. X1999..YR1999.
## 1            NA            NA            NA            NA            NA
## 2      1488.020      1633.552      1464.298      1603.647      1821.873
## 3      2834.265      2899.193      2884.120      2986.056      3038.273
## 4            NA            NA            NA            NA            NA
## 5     27641.886     28701.095     31325.517     32416.014     33625.244
## 6      2146.456      2360.088      2451.609      2485.067      2458.096
##   X2000..YR2000. X2001..YR2001. X2002..YR2002. X2003..YR2003. X2004..YR2004.
## 1            NA            NA      319.8471       332.220       322.668
## 2      1960.882      2143.526     2247.4975      2380.644      2522.449
## 3      3111.176      3162.812     3297.6128      3490.223      3592.645
## 4            NA            NA    12609.3685     12641.668     12653.282
## 5     34267.582     35974.880    36158.5864     37620.206     39042.959
## 6      2451.510      2471.665     2717.4410      2705.706      2900.167
##   X2005..YR2005. X2006..YR2006. X2007..YR2007. X2008..YR2008. X2009..YR2009.
## 1      345.9258      353.7206      392.7105      398.9711      472.8423
## 2     2675.5079     2851.3670     3044.8958     3298.4780     3432.1700
## 3     3752.0962     3760.1554     3828.1485     3856.4077     3851.2138
## 4    12628.5077    12202.7374    12586.7160    12439.1497    12092.2765
## 5    39782.9288    40602.5376    40391.8715    37610.5609    35362.6966
## 6     3220.0781     3464.2347     3806.8492     4077.7768     3963.2484
##   X2010..YR2010. X2011..YR2011. X2012..YR2012. X2013..YR2013. X2014..YR2014.
## 1      526.1037      511.9985      557.9497      568.9645      565.1793
## 2     3577.1134     3678.0467     3736.3391     3780.6982     3855.7597
## 3     3918.4864     3956.8958     4012.3615     4042.9236     4112.0760
## 4    12256.3429    12328.6653    11812.4591    11507.2257    11694.5454
## 5    34667.1713    34956.5962    33750.0569    33220.5619    34721.8920
## 6     3988.6236     3979.8151     4167.1258     4220.9648     4272.4555
##   X2015..YR2015.
## 1      556.0072
## 2     3952.8025
## 3     4177.8895
## 4    12059.6352
## 5    35770.7767
```

```
## 6      4166.9798
```

```
head(df_region)
```

```
## # A tibble: 6 x 20
##   BLcode country  year sex   agefrom ageto    lu    lp   lpc    ls   lsc    lh
##    <dbl> <chr>   <dbl> <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1      1 Algeria  1950 MF         15   999  80.7  17.6  3.75  1.45  0.46  0.3
## 2      1 Algeria  1955 MF         15   999  81.0  17.0  3.46  1.64  0.5   0.26
## 3      1 Algeria  1960 MF         15   999  82.6  14.3  3.07  2.75  1.05  0.32
## 4      1 Algeria  1965 MF         15   999  80.9  14.4  4.01  4.21  1.79  0.43
## 5      1 Algeria  1970 MF         15   999  73.6  19.2  5.23  6.69  3.26  0.34
## 6      1 Algeria  1975 MF         15   999  64.4  25.2  4.26  9.57  5.1   0.74
## # ... with 8 more variables: lhc <dbl>, yr_sch <dbl>, yr_sch_pri <dbl>,
## #   yr_sch_sec <dbl>, yr_sch_ter <dbl>, pop <dbl>, WBcode <chr>,
## #   region_code <chr>
```

```
head(df_gadp)
```

```
## # A tibble: 6 x 20
##   Code  Country       Sample Imputed logYL logKLa logKY School logHL logA  logL
##   <chr> <chr>          <dbl>   <dbl> <chr> <chr>  <chr> <chr>  <chr> <chr> <chr>
## 1 AGO   Angola             1       1 7.06~ 6.551~ -0.2~ 3.098~ 0.41~ 6.90~ 15.2~
## 2 ARE   United arab ~      0       0 9.98~ NaN    NaN   NaN    NaN   NaN   13.5~
## 3 ARG   Argentina          1       0 9.60~ 10.40~ 0.40~ 6.679~ 0.80~ 8.39~ 16.2~
## 4 AUS   Australia          1       0 10.3~ 11.38~ 0.54~ 10.24  1.09~ 8.67~ 15.8~
## 5 AUT   Austria            1       0 10.1~ 11.17~ 0.52~ 6.639~ 0.80~ 8.80~ 15.0~
## 6 BDI   Burundi            1       1 6.96~ 6.456~ -0.2~ 2.000~ 0.26~ 6.94~ 14.8~
## # ... with 9 more variables: Mining <chr>, EcOrg <dbl>, SocInf <chr>,
## #   YrsOpen <chr>, GADP <chr>, EngFrac <dbl>, EurFrac <dbl>,
## #   `log(FrankRom)` <chr>, Latitude <dbl>
```

After checking the df_base, the column name still include "." in it. Also, the column names for years are not clear. I should rename all the column name first with proper format.

```
year_colnames <- seq(1970, 2015, 1)
colnames(df_base) <- c('series', 'series_code', 'country', 'country_code', year_colnames)

head(df_base)
```

```
##                                series    series_code        country country_code
## 1 GDP per capita (constant 2015 US$) NY.GDP.PCAP.KD    Afghanistan          AFG
## 2 GDP per capita (constant 2015 US$) NY.GDP.PCAP.KD        Albania          ALB
## 3 GDP per capita (constant 2015 US$) NY.GDP.PCAP.KD        Algeria          DZA
## 4 GDP per capita (constant 2015 US$) NY.GDP.PCAP.KD American Samoa          ASM
## 5 GDP per capita (constant 2015 US$) NY.GDP.PCAP.KD        Andorra          AND
## 6 GDP per capita (constant 2015 US$) NY.GDP.PCAP.KD         Angola          AGO
##       1970     1971     1972    1973    1974     1975    1976     1977
## 1       NA       NA       NA      NA      NA       NA      NA       NA
## 2       NA       NA       NA      NA      NA       NA      NA       NA
## 3 2368.575 2042.663 2532.393 2558.03 2674.97 2732.324 2878.61 2944.321
## 4       NA       NA       NA      NA      NA       NA      NA       NA
```

```
## 5 35391.074 35159.467 36166.413 37123.26 37504.74 36246.683 36175.32 36081.658
## 6        NA        NA        NA       NA       NA        NA       NA        NA
##        1978      1979      1980      1981      1982      1983      1984
## 1        NA        NA        NA        NA        NA        NA        NA
## 2        NA        NA  1740.505  1804.010  1818.367  1799.878  1740.347
## 3  3123.377  3258.731  3186.437  3182.260  3281.861  3352.700  3432.806
## 4        NA        NA        NA        NA        NA        NA        NA
## 5 35551.736 34462.493 34014.488 32669.462 31709.296 30896.951 30084.567
## 6        NA        NA  3550.083  3276.362  3162.004  3179.348  3252.108
##        1985      1986      1987      1988      1989      1990      1991
## 1        NA        NA        NA        NA        NA        NA        NA
## 2  1735.290  1798.015  1748.578  1691.531  1808.646  1606.296  1163.491
## 3  3453.822  3366.939  3248.881  3128.338  3179.773  3123.998  3011.500
## 4        NA        NA        NA        NA        NA        NA        NA
## 5 29480.346 29178.543 29564.821 29854.800 30091.249 30051.178 29642.711
## 6  3248.608  3226.817  3242.579  3323.534  3212.663  2998.781  2929.483
##        1992      1993      1994      1995      1996      1997      1998
## 1        NA        NA        NA        NA        NA        NA        NA
## 2  1086.438  1197.581  1305.001  1488.020  1633.552  1464.298  1603.647
## 3  2994.491  2867.196  2783.155  2834.265  2899.193  2884.120  2986.056
## 4        NA        NA        NA        NA        NA        NA        NA
## 5 28792.198 27517.549 27408.280 27641.886 28701.095 31325.517 32416.014
## 6  2669.435  1964.390  1927.500  2146.456  2360.088  2451.609  2485.067
##        1999      2000      2001       2002      2003      2004       2005
## 1        NA        NA        NA   319.8471   332.220   322.668   345.9258
## 2  1821.873  1960.882  2143.526  2247.4975  2380.644  2522.449  2675.5079
## 3  3038.273  3111.176  3162.812  3297.6128  3490.223  3592.645  3752.0962
## 4        NA        NA        NA 12609.3685 12641.668 12653.282 12628.5077
## 5 33625.244 34267.582 35974.880 36158.5864 37620.206 39042.959 39782.9288
## 6  2458.096  2451.510  2471.665  2717.4410  2705.706  2900.167  3220.0781
##         2006       2007       2008       2009       2010       2011       2012
## 1   353.7206   392.7105   398.9711   472.8423   526.1037   511.9985   557.9497
## 2  2851.3670  3044.8958  3298.4780  3432.1700  3577.1134  3678.0467  3736.3391
## 3  3760.1554  3828.1485  3856.4077  3851.2138  3918.4864  3956.8958  4012.3615
## 4 12202.7374 12586.7160 12439.1497 12092.2765 12256.3429 12328.6653 11812.4591
## 5 40602.5376 40391.8715 37610.5609 35362.6966 34667.1713 34956.5962 33750.0569
## 6  3464.2347  3806.8492  4077.7768  3963.2484  3988.6236  3979.8151  4167.1258
##         2013       2014       2015
## 1   568.9645   565.1793   556.0072
## 2  3780.6982  3855.7597  3952.8025
## 3  4042.9236  4112.0760  4177.8895
## 4 11507.2257 11694.5454 12059.6352
## 5 33220.5619 34721.8920 35770.7767
## 6  4220.9648  4272.4555  4166.9798
```

Looking to the df_gadp, the data that I need, the GADP, is in string data type. I need it in dbl data type to make it calculate-able for regression.

```
df_gadp$GADP <- as.double(df_gadp$GADP)
str(df_gadp)
```

```
## tibble [152 x 20] (S3: tbl_df/tbl/data.frame)
##  $ Code         : chr [1:152] "AGO" "ARE" "ARG" "AUS" ...
```

```
##  $ Country      : chr [1:152] "Angola" "United arab e." "Argentina" "Australia" ...
##  $ Sample       : num [1:152] 1 0 1 1 1 1 1 1 1 1 ...
##  $ Imputed      : num [1:152] 1 0 0 0 0 1 0 1 1 0 ...
##  $ logYL        : chr [1:152] "7.0625200000000001" "9.9811599999999991" "9.6027000000000005" "10.304
##  $ logKLa       : chr [1:152] "6.5513599999999999" "NaN" "10.40884" "11.385949999999999" ...
##  $ logKY        : chr [1:152] "-0.25557999999999997" "NaN" "0.40306999999999998" "0.5408699999999999
##  $ School       : chr [1:152] "3.0987399999999998" "NaN" "6.6799999999999997" "10.24" ...
##  $ logHL        : chr [1:152] "0.41522999999999999" "NaN" "0.80667999999999995" "1.09232" ...
##  $ logA         : chr [1:152] "6.9028700000000001" "NaN" "8.3929500000000008" "8.67103" ...
##  $ logL         : chr [1:152] "15.272320000000001" "13.53248" "16.225349999999999" "15.87232" ...
##  $ Mining       : chr [1:152] "0.26800000000000002" "0.34300000000000003" "0.023" "0.03799999999999
##  $ EcOrg        : num [1:152] 0 3 5 5 4 4 5 0 1 3 ...
##  $ SocInf       : chr [1:152] "0.21332999999999999" "NaN" "0.33411000000000002" "0.80978000000000006
##  $ YrsOpen      : chr [1:152] "0" "NaN" "0.088999999999999996" "0.68899999999999995" ...
##  $ GADP         : num [1:152] 0.427 0.556 0.579 0.931 0.949 0.528 0.954 0.376 0.498 0.313 ...
##  $ EngFrac      : num [1:152] 0 0 0 0.95 0 0 0 0 0 0 ...
##  $ EurFrac      : num [1:152] 0 0 0.836 0.95 0.98 0 0.345 0 0 0 ...
##  $ log(FrankRom): chr [1:152] "2.4430000000000001" "3.5089999999999999" "1.7230000000000001" "1.4039
##  $ Latitude     : num [1:152] -8.84 23.39 -36.68 -32.22 48.23 ...
```

The next part is the region. I will need the region data for regression. The region data is included in df_region. All of these 3 data frames contain the country code. I can make this country code as a key for doing the inner join. For this, I use the sqldf package to write SQL in R environment.

```r
df_base_region <- sqldf("
SELECT
    df_region.region_code AS region,
    df_base.*,
    df_gadp.GADP
FROM
    df_base
INNER JOIN df_region
    ON df_base.country_code = df_region.WBcode
INNER JOIN df_gadp
    ON df_base.country_code = df_gadp.Code
")

head(df_base_region)
```

```
##                       region                                 series
## 1 Middle East and North Africa GDP per capita (constant 2015 US$)
## 2 Middle East and North Africa GDP per capita (constant 2015 US$)
## 3 Middle East and North Africa GDP per capita (constant 2015 US$)
## 4 Middle East and North Africa GDP per capita (constant 2015 US$)
## 5 Middle East and North Africa GDP per capita (constant 2015 US$)
## 6 Middle East and North Africa GDP per capita (constant 2015 US$)
##      series_code country country_code     1970     1971     1972    1973
## 1 NY.GDP.PCAP.KD Algeria          DZA 2368.575 2042.663 2532.393 2558.03
## 2 NY.GDP.PCAP.KD Algeria          DZA 2368.575 2042.663 2532.393 2558.03
## 3 NY.GDP.PCAP.KD Algeria          DZA 2368.575 2042.663 2532.393 2558.03
## 4 NY.GDP.PCAP.KD Algeria          DZA 2368.575 2042.663 2532.393 2558.03
## 5 NY.GDP.PCAP.KD Algeria          DZA 2368.575 2042.663 2532.393 2558.03
## 6 NY.GDP.PCAP.KD Algeria          DZA 2368.575 2042.663 2532.393 2558.03
```

```
##      1974     1975    1976     1977     1978     1979     1980    1981     1982
## 1 2674.97 2732.324 2878.61 2944.321 3123.377 3258.731 3186.437 3182.26 3281.861
## 2 2674.97 2732.324 2878.61 2944.321 3123.377 3258.731 3186.437 3182.26 3281.861
## 3 2674.97 2732.324 2878.61 2944.321 3123.377 3258.731 3186.437 3182.26 3281.861
## 4 2674.97 2732.324 2878.61 2944.321 3123.377 3258.731 3186.437 3182.26 3281.861
## 5 2674.97 2732.324 2878.61 2944.321 3123.377 3258.731 3186.437 3182.26 3281.861
## 6 2674.97 2732.324 2878.61 2944.321 3123.377 3258.731 3186.437 3182.26 3281.861
##     1983     1984     1985     1986     1987     1988     1989     1990   1991
## 1 3352.7 3432.806 3453.822 3366.939 3248.881 3128.338 3179.773 3123.998 3011.5
## 2 3352.7 3432.806 3453.822 3366.939 3248.881 3128.338 3179.773 3123.998 3011.5
## 3 3352.7 3432.806 3453.822 3366.939 3248.881 3128.338 3179.773 3123.998 3011.5
## 4 3352.7 3432.806 3453.822 3366.939 3248.881 3128.338 3179.773 3123.998 3011.5
## 5 3352.7 3432.806 3453.822 3366.939 3248.881 3128.338 3179.773 3123.998 3011.5
## 6 3352.7 3432.806 3453.822 3366.939 3248.881 3128.338 3179.773 3123.998 3011.5
##       1992     1993     1994     1995     1996    1997     1998     1999
## 1 2994.491 2867.196 2783.155 2834.265 2899.193 2884.12 2986.056 3038.273
## 2 2994.491 2867.196 2783.155 2834.265 2899.193 2884.12 2986.056 3038.273
## 3 2994.491 2867.196 2783.155 2834.265 2899.193 2884.12 2986.056 3038.273
## 4 2994.491 2867.196 2783.155 2834.265 2899.193 2884.12 2986.056 3038.273
## 5 2994.491 2867.196 2783.155 2834.265 2899.193 2884.12 2986.056 3038.273
## 6 2994.491 2867.196 2783.155 2834.265 2899.193 2884.12 2986.056 3038.273
##       2000     2001     2002     2003     2004     2005     2006     2007
## 1 3111.176 3162.812 3297.613 3490.223 3592.645 3752.096 3760.155 3828.148
## 2 3111.176 3162.812 3297.613 3490.223 3592.645 3752.096 3760.155 3828.148
## 3 3111.176 3162.812 3297.613 3490.223 3592.645 3752.096 3760.155 3828.148
## 4 3111.176 3162.812 3297.613 3490.223 3592.645 3752.096 3760.155 3828.148
## 5 3111.176 3162.812 3297.613 3490.223 3592.645 3752.096 3760.155 3828.148
## 6 3111.176 3162.812 3297.613 3490.223 3592.645 3752.096 3760.155 3828.148
##       2008     2009     2010     2011     2012     2013     2014    2015  GADP
## 1 3856.408 3851.214 3918.486 3956.896 4012.362 4042.924 4112.076 4177.89 0.529
## 2 3856.408 3851.214 3918.486 3956.896 4012.362 4042.924 4112.076 4177.89 0.529
## 3 3856.408 3851.214 3918.486 3956.896 4012.362 4042.924 4112.076 4177.89 0.529
## 4 3856.408 3851.214 3918.486 3956.896 4012.362 4042.924 4112.076 4177.89 0.529
## 5 3856.408 3851.214 3918.486 3956.896 4012.362 4042.924 4112.076 4177.89 0.529
## 6 3856.408 3851.214 3918.486 3956.896 4012.362 4042.924 4112.076 4177.89 0.529
```

There's a problem in the data. The data contains too many duplicate data. Let's just trim and remove the duplicate data.

```
df_base_region <- df_base_region %>%
  distinct()
```

For the final, I create the dummy variables for region to conduct the regression.

```
df_base_region_dummy <- dummy_cols(df_base_region, select_columns = 'region')

head(df_base_region_dummy)
```

```
##                            region                               series
## 1    Middle East and North Africa GDP per capita (constant 2015 US$)
## 2 Latin America and the Caribbean GDP per capita (constant 2015 US$)
## 3              Advanced Economies GDP per capita (constant 2015 US$)
## 4              Advanced Economies GDP per capita (constant 2015 US$)
```

```
## 5    Middle East and North Africa GDP per capita (constant 2015 US$)
## 6                      South Asia GDP per capita (constant 2015 US$)
##      series_code    country country_code       1970       1971       1972
## 1 NY.GDP.PCAP.KD    Algeria          DZA  2368.5747  2042.6634  2532.3928
## 2 NY.GDP.PCAP.KD  Argentina          ARG  9243.2566  9613.6769  9614.2379
## 3 NY.GDP.PCAP.KD  Australia          AUS 26894.5065 27040.5103 27585.7828
## 4 NY.GDP.PCAP.KD    Austria          AUT 18101.9919 18943.1811 20002.5561
## 5 NY.GDP.PCAP.KD    Bahrain          BHR         NA         NA         NA
## 6 NY.GDP.PCAP.KD Bangladesh          BGD   512.1212   474.4632   401.4605
##         1973       1974       1975       1976      1977       1978       1979
## 1  2558.0302  2674.9704  2732.3244  2878.6100  2944.321  3123.3766  3258.7312
## 2  9725.4670 10100.0685  9939.6966  9591.1720 10103.627  9505.8562 10321.2391
## 3 27878.8169 28297.9448 28325.0161 28769.1030 29468.842 29388.5744 30250.3701
## 4 20864.8475 21650.4868 21629.0463 22659.3178 23801.268 23770.3662 25086.4002
## 5         NA         NA         NA         NA        NA         NA         NA
## 6   408.6063   440.6036   414.6048   428.3799   428.987   447.2395   456.1454
##         1980       1981       1982       1983       1984      1985       1986
## 1  3186.4372  3182.2599  3281.8609  3352.6997  3432.8064  3453.822  3366.9389
## 2 10318.1830  9630.1152  9407.8746  9660.1921  9655.1836  9009.002  9412.9948
## 3 30790.9685 31318.0837 31823.6227 30729.8076 31780.2585 32995.138 33767.1612
## 4 25520.7406 25419.0096 25911.6609 26725.1907 26740.5772 27395.873 28008.5546
## 5 21450.5528 19537.9587 17507.8734 18131.0725 18534.2249 17139.070 16779.7366
## 6   447.6949   467.5867   465.2501   470.9053   480.8311   484.048   491.1358
##         1987       1988       1989       1990       1991      1992       1993
## 1  3248.8812  3128.3375  3179.7733  3123.9977  3011.4999  2994.491  2867.1959
## 2  9517.1086  9269.3558  8477.6729  8149.2406  8769.7504  9338.700  9974.0044
## 3 34104.8306 35478.4268 36231.0162 36974.4509 36361.9423 36072.040 37161.6791
## 4 28370.7179 29264.3236 30265.1932 31340.6813 32097.2411 32410.540 32313.7119
## 5 17885.0193 18469.8463 17925.0720 18160.1694 19651.3379 20432.124 22497.0364
## 6   496.4329   495.4038   496.7616   512.0966   517.7637   533.888   547.0514
##         1994       1995       1996       1997       1998       1999       2000
## 1  2783.1550  2834.2652  2899.1932  2884.1202  2986.0557  3038.2729  3111.1762
## 2 10423.3789 10003.0917 10430.6825 11146.7210 11445.5649 10935.6446 10730.6082
## 3 38234.0299 39223.1608 40203.3023 41309.8260 42767.7940 44369.2728 45558.7531
## 4 32962.8174 33790.4850 34537.7396 35220.8877 36442.2882 37664.9433 38842.8905
## 5 21894.5445 22185.4956 22500.0710 22561.7924 22919.3940 23047.0707 23243.5909
## 6   556.2989   572.4385   585.7169   599.2275   617.2923   633.1827   653.8086
##        2001       2002       2003       2004      2005       2006       2007
## 1  3162.812  3297.6128  3490.2232  3592.6453  3752.096  3760.1554  3828.1485
## 2 10146.107  8943.3080  9629.8441 10389.1513 11192.180 11970.6554 12919.2354
## 3 45864.768 47127.2781 47997.1371 49439.1223 50332.875 50948.2250 52538.9756
## 4 39184.809 39636.4826 39815.2220 40651.2266 41281.271 42496.3511 43937.7129
## 5 22697.617 22258.1949 22335.9324 22422.6899 22343.537 22069.2804 22111.6801
## 6   674.167   687.3833   707.6051   732.7488   769.136   809.4736   856.0459
##        2008       2009       2010      2011      2012      2013      2014
## 1  3856.408  3851.2138  3918.4864  3956.896  4012.362  4042.924  4112.076
## 2 13310.624 12398.2836 13551.3392 14200.270 13895.634 14071.509 13567.948
## 3 53338.603 53225.9510 53542.8337 54108.111 55254.606 55723.948 56305.979
## 4 44440.056 42655.1921 43334.5090 44451.000 44549.882 44299.378 44245.169
## 5 21833.412 21057.4208 20982.3405 20774.245 21187.542 22078.990 22676.626
## 6   897.189   931.9865   972.9097  1024.022  1078.288  1129.994  1184.863
##        2015  GADP region_Advanced Economies region_East Asia and the Pacific
## 1  4177.890 0.529                         0                                0
## 2 13789.060 0.579                         0                                0
```

```
## 3 56707.022 0.931                        1                                 0
## 4 44195.818 0.949                        1                                 0
## 5 22634.086 0.707                        0                                 0
## 6  1248.453 0.313                        0                                 0
##   region_Europe and Central Asia region_Latin America and the Caribbean
## 1                              0                                      0
## 2                              0                                      1
## 3                              0                                      0
## 4                              0                                      0
## 5                              0                                      0
## 6                              0                                      0
##   region_Middle East and North Africa region_South Asia
## 1                                   1                 0
## 2                                   0                 0
## 3                                   0                 0
## 4                                   0                 0
## 5                                   1                 0
## 6                                   0                 1
##   region_Sub-Saharan Africa
## 1                         0
## 2                         0
## 3                         0
## 4                         0
## 5                         0
## 6                         0
```

The dataframe still contains four series of variables. GDP per capita, Consumer Price Index (CPI), Secondary School enrollment, and Trade Openess. I will separate the data frame based on the series to make it neater. First, I will check the unique value for the series_code.

```
unique(df_base_region_dummy['series'])
```

```
##                                     series
## 1       GDP per capita (constant 2015 US$)
## 118      Consumer price index (2010 = 100)
## 235 School enrollment, secondary (% gross)
## 352                       Trade (% of GDP)
```

```
unique(df_base_region_dummy['series_code'])
```

```
##        series_code
## 1   NY.GDP.PCAP.KD
## 118    FP.CPI.TOTL
## 235    SE.SEC.ENRR
## 352 NE.TRD.GNFS.ZS
```

Then, I pass the series_code to the SQL code and put the series_code to the WHERE clause. In total, there are 4 data frames. Also, I create another separated data frame for GADP.

```
df_gdp <- sqldf("
SELECT
    *
```

```
FROM
    df_base_region_dummy
WHERE
    series_code = 'NY.GDP.PCAP.KD'
")

df_cpi <- sqldf("
SELECT
    *
FROM
    df_base_region_dummy
WHERE
    series_code = 'FP.CPI.TOTL'
")

df_school <- sqldf("
SELECT
    *
FROM
    df_base_region_dummy
WHERE
    series_code = 'SE.SEC.ENRR'
")

df_trade <- sqldf("
SELECT
    *
FROM
    df_base_region_dummy
WHERE
    series_code = 'NE.TRD.GNFS.ZS'
")

df_gdp <- subset(df_gdp, select = -c(series, series_code))
df_cpi <- subset(df_cpi, select = -c(series, series_code))
df_school <- subset(df_school, select = -c(series, series_code))
df_trade <- subset(df_trade, select = -c(series, series_code))
df_gadp <- distinct(df_base_region[c('country', 'country_code', 'GADP')])
```

ETL is done, let's move to the question.

Problem 1: Absolute Convergence

Test the absolute convergence theory for 1. 1970-2015 period, 2. 1970-1990 period, 3. 1990-2015 period, 4. 1995-2015 period, and 5. 2000-2015 period Present discussions of your results.

To calculate it, first we derive and calculate the following variables:

- Per capita real GDP growth rate,is natural log of ((GDP per capita (constant 2010 US dollars) in last year – GDP per capita in initial year) / number of years)
- Initial real per capita GDP in natural log expression

```r
itn_year <- c("1970", "1990", "1995", "2000")
itn_year_log <- c("ln1970", "ln1990", "ln1995", "ln2000")

df_prob1[itn_year_log] <- df_prob1[itn_year] %>%
  log()
```

Then, using these two variables, I can examine the 'absolute convergence' hypothesis by running cross-country OLS regressions and report the result.

```r
prd_1_mdl <- lm(gr_1_prd ~ ln1970, data = df_prob1)
prd_2_mdl <- lm(gr_2_prd ~ ln1970, data = df_prob1)
prd_3_mdl <- lm(gr_3_prd ~ ln1990, data = df_prob1)
prd_4_mdl <- lm(gr_4_prd ~ ln1995, data = df_prob1)
prd_5_mdl <- lm(gr_5_prd ~ ln2000, data = df_prob1)

prd_1_mdl
```

```
##
## Call:
## lm(formula = gr_1_prd ~ ln1970, data = df_prob1)
##
## Coefficients:
## (Intercept)       ln1970
##      -3.900        1.029
```

```r
prd_2_mdl
```

```
##
## Call:
## lm(formula = gr_2_prd ~ ln1970, data = df_prob1)
##
## Coefficients:
## (Intercept)       ln1970
##      -4.347        1.084
```

```r
prd_3_mdl
```

```
##
## Call:
## lm(formula = gr_3_prd ~ ln1990, data = df_prob1)
##
## Coefficients:
## (Intercept)       ln1990
##     -3.3520       0.9354
```

```r
prd_4_mdl
```

```
##
## Call:
## lm(formula = gr_4_prd ~ ln1995, data = df_prob1)
```

```
##
## Coefficients:
## (Intercept)       ln1995
##     -2.4486       0.8411
```

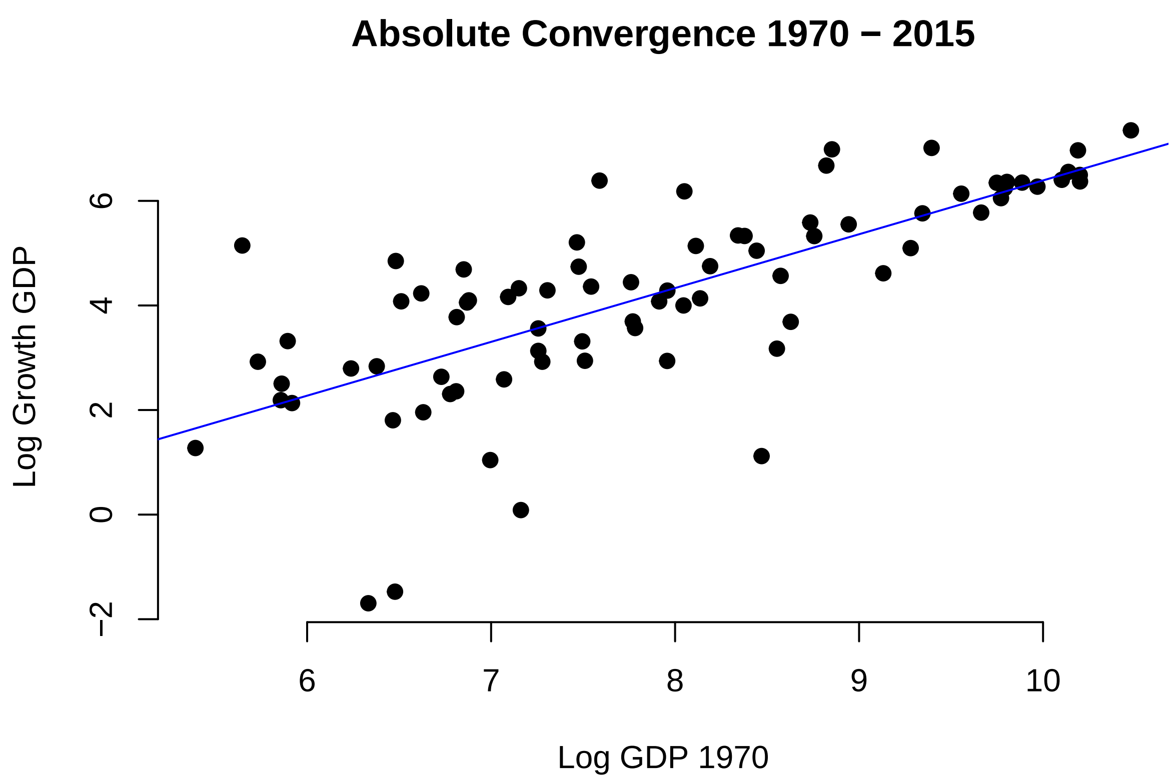
```
prd_5_mdl
```

```
##
## Call:
## lm(formula = gr_5_prd ~ ln2000, data = df_prob1)
##
## Coefficients:
## (Intercept)       ln2000
##     -1.8273       0.7544
```

Let's plot the data and regression line to see the pattern.

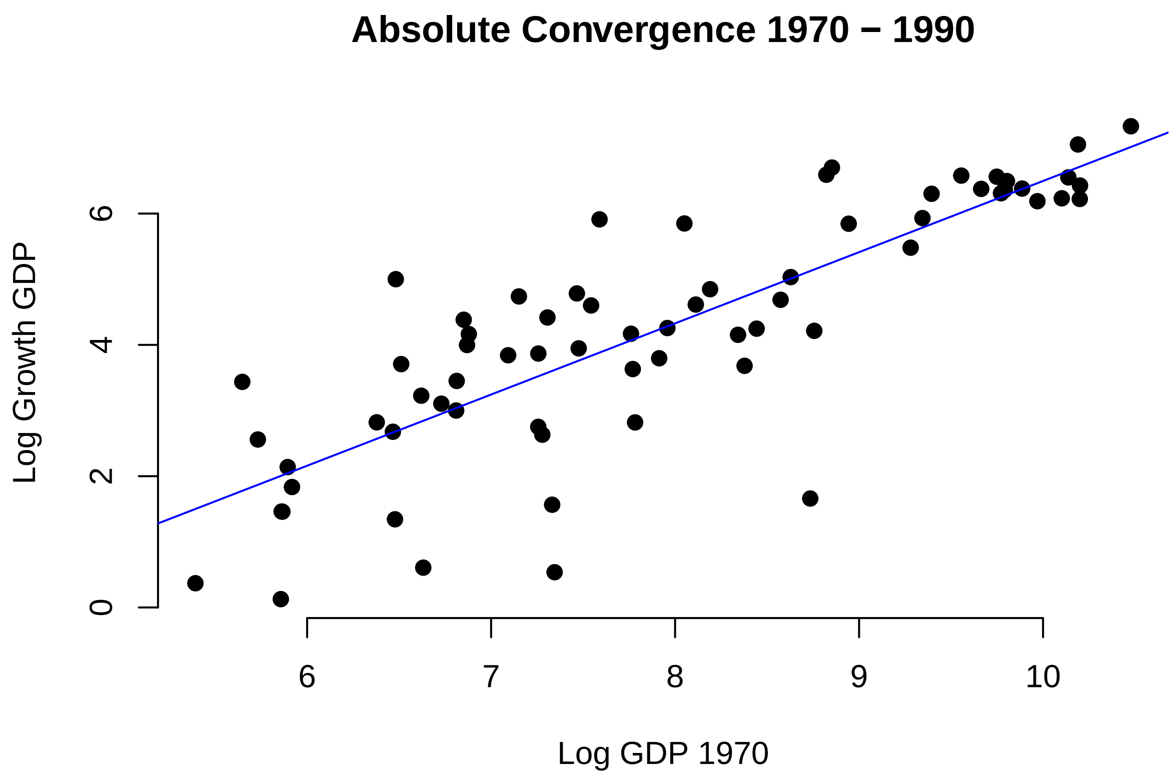
```
plot_1_mdl <- plot(df_prob1$ln1970, df_prob1$gr_1_prd, main = "Absolute Convergence 1970 - 2015",
     xlab = "Log GDP 1970", ylab = "Log Growth GDP",
     pch = 19, frame = FALSE)
abline(prd_1_mdl, data = df_prob1, col = "blue")
```

```
## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...): "data" is
## not a graphical parameter
```



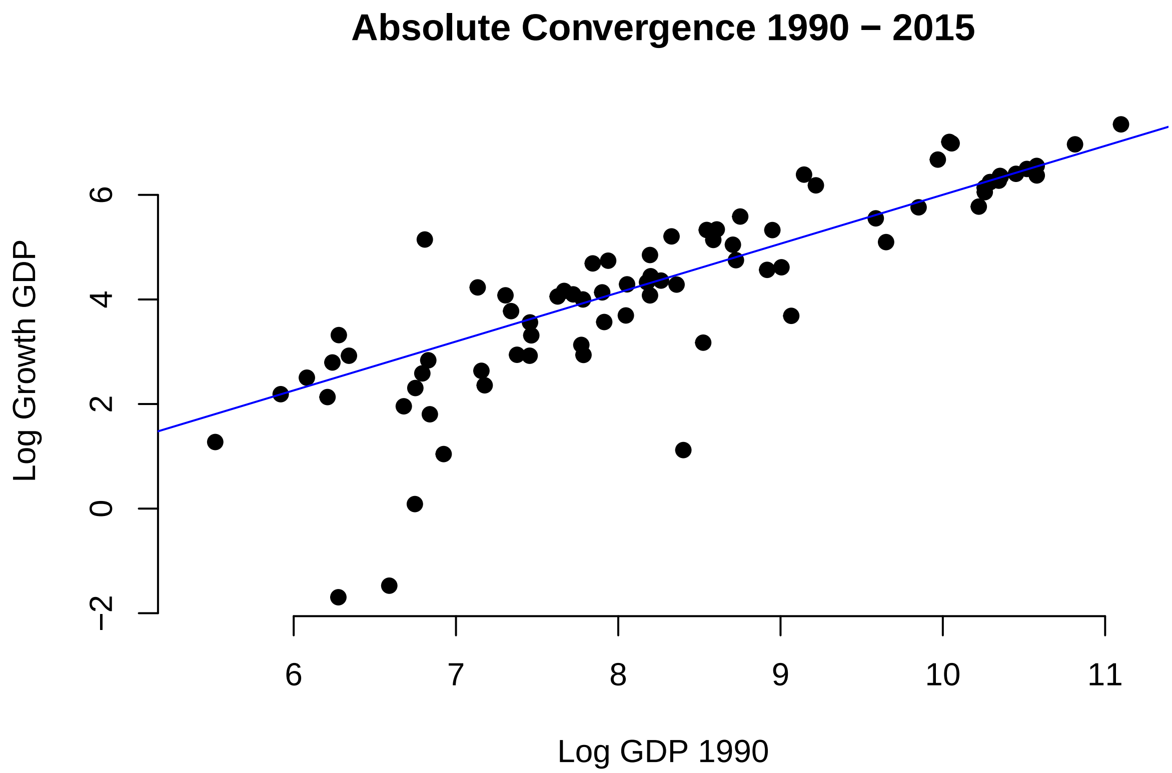
```
plot_2_mdl <- plot(df_prob1$ln1970, df_prob1$gr_2_prd, main = "Absolute Convergence 1970 - 1990",
     xlab = "Log GDP 1970", ylab = "Log Growth GDP",
     pch = 19, frame = FALSE)
abline(prd_2_mdl, data = df_prob1, col = "blue")
```

```
## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...): "data" is
## not a graphical parameter
```



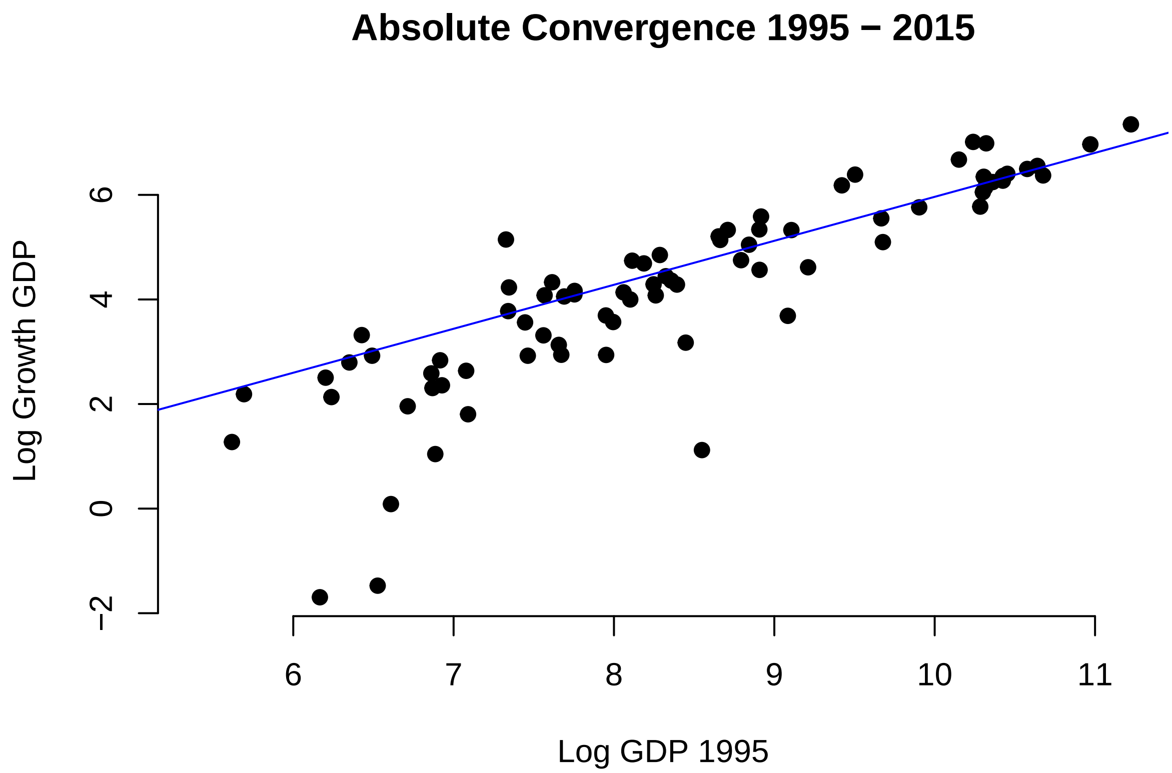

```
plot_3_mdl <- plot(df_prob1$ln1990, df_prob1$gr_1_prd, main = "Absolute Convergence 1990 - 2015",
     xlab = "Log GDP 1990", ylab = "Log Growth GDP",
     pch = 19, frame = FALSE)
abline(prd_3_mdl, data = df_prob1, col = "blue")
```

```
## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...): "data" is
## not a graphical parameter
```



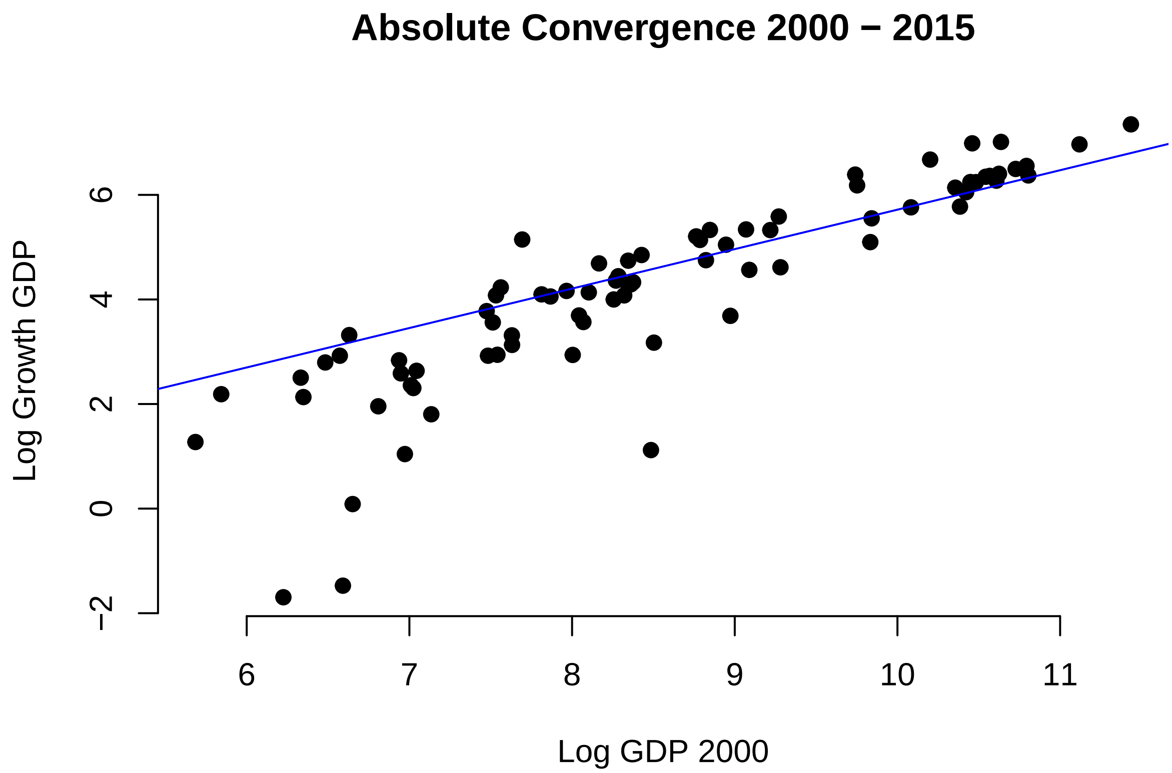

```r
plot_4_mdl <- plot(df_prob1$ln1995, df_prob1$gr_1_prd, main = "Absolute Convergence 1995 - 2015",
     xlab = "Log GDP 1995", ylab = "Log Growth GDP",
     pch = 19, frame = FALSE)
abline(prd_4_mdl, data = df_prob1, col = "blue")
```

```
## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...): "data" is
## not a graphical parameter
```




```
plot_5_mdl <- plot(df_prob1$ln2000, df_prob1$gr_1_prd, main = "Absolute Convergence 2000 - 2015",
     xlab = "Log GDP 2000", ylab = "Log Growth GDP",
     pch = 19, frame = FALSE)
abline(prd_5_mdl, data = df_prob1, col = "blue")
```

```
## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...): "data" is
## not a graphical parameter
```



Looking from the scatter plot with regression line above, the initial GDP has a positive correlation to the GDP growth for all period, either short term or long term. It shows that we countries around the world have not yet reached the convergence of economy, as it's supposed to be negatively correlated if the theory holds.

Problem 2: Conditional Convergence

Test the conditional convergence theory using those control variables (in the growth regressions guide) for the same periods. Present discussions of your results. Report the whole period result.

Now derive these control variables and countries.

1. Initial education variable
2. Institution such as GADP (Government Anti-Diversion Policy)
3. 3 Regional dummies: East Asia and Pacific, Latin America and Caribbean, and Sub-Saharan Africa from World Bank
4. Consumer Price Index growth rate
5. Average of trade openness: (export + import) / GDP from WDI

These variables can be added one by one or together

Using these common control variables, set up the cross-country growth regression, showing 'conditional convergence'. Run the regression using simple OLS, and report the result. Explain the result and theoretical reason for that.

a) Adding initial education

We can use either natural log of secondary school enrollment ratio (from WDI) or average schooling year (from Barro-Lee dataset) for all population in 1980 or 1981 (initial GDP and initial education variables are basic controls). In this problem I will use the natural log of secondary school enrollment ratio (from WDI). As the data in the data set still in the ratio, I should calculate the log natural for the initial education first.

```
var_prob2 <- c("country", "country_code", "1971", "1990", "1995", "2000", "2015")
df_prob2_a <- df_school[var_prob2]
```

```
itn_year_2 <- c("1971", "1990", "1995", "2000")
itn_year_log_2 <- c("ln1971", "ln1990", "ln1995", "ln2000")

df_prob2_a[itn_year_log_2] <- df_prob2_a[itn_year_2] %>%
  log()
```

Notes in here, for initial education in 1970, the Professor wants me to change it into year 1971 as the there are so many null data in initial education in 1970.

```
cc_a_prd_1_mdl <- lm(df_prob1$gr_1_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971)
cc_a_prd_2_mdl <- lm(df_prob1$gr_2_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971)
cc_a_prd_3_mdl <- lm(df_prob1$gr_3_prd ~ df_prob1$ln1990 + df_prob2_a$ln1990)
cc_a_prd_4_mdl <- lm(df_prob1$gr_4_prd ~ df_prob1$ln1995 + df_prob2_a$ln1995)
cc_a_prd_5_mdl <- lm(df_prob1$gr_5_prd ~ df_prob1$ln2000 + df_prob2_a$ln2000)

summary(cc_a_prd_1_mdl) # 1970 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_1_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7665 -0.3655  0.0774  0.7252  2.4677
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -3.2977     0.8603  -3.833 0.000285 ***
## df_prob1$ln1970     0.5437     0.1607   3.384 0.001205 **
## df_prob2_a$ln1971   0.9935     0.2433   4.083 0.000122 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.144 on 66 degrees of freedom
##   (48 observations deleted due to missingness)
## Multiple R-squared:  0.6343, Adjusted R-squared:  0.6233
## F-statistic: 57.25 on 2 and 66 DF,  p-value: 3.813e-15
```

```
summary(cc_a_prd_2_mdl) # 1970 - 1990
```

```
##
## Call:
```

```
## lm(formula = df_prob1$gr_2_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5216 -0.4499  0.1111  0.4520  2.4543
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -3.3155     0.7091  -4.675 1.80e-05 ***
## df_prob1$ln1970      0.6725     0.1333   5.047 4.75e-06 ***
## df_prob2_a$ln1971    0.6927     0.1958   3.538 0.000802 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9209 on 58 degrees of freedom
##   (56 observations deleted due to missingness)
## Multiple R-squared:  0.7332, Adjusted R-squared:  0.724
## F-statistic: 79.69 on 2 and 58 DF,  p-value: < 2.2e-16
```

```
summary(cc_a_prd_3_mdl) # 1990 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_3_prd ~ df_prob1$ln1990 + df_prob2_a$ln1990)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3155 -0.4383  0.1595  0.5593  2.2305
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -3.6458     0.6718  -5.427 7.34e-07 ***
## df_prob1$ln1990      0.4900     0.1365   3.591 0.000599 ***
## df_prob2_a$ln1990    1.0352     0.2727   3.796 0.000304 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.011 on 72 degrees of freedom
##   (42 observations deleted due to missingness)
## Multiple R-squared:  0.6768, Adjusted R-squared:  0.6678
## F-statistic: 75.37 on 2 and 72 DF,  p-value: < 2.2e-16
```

```
summary(cc_a_prd_4_mdl) # 1995 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_4_prd ~ df_prob1$ln1995 + df_prob2_a$ln1995)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.87650 -0.29436  0.06522  0.38667  1.67470
##
## Coefficients:
```

```
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -2.0705     0.5318  -3.893 0.000228 ***
## df_prob1$ln1995     0.5461     0.1005   5.434 8.03e-07 ***
## df_prob2_a$ln1995   0.5558     0.2140   2.598 0.011496 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7513 on 68 degrees of freedom
##   (46 observations deleted due to missingness)
## Multiple R-squared:  0.725,  Adjusted R-squared:  0.7169
## F-statistic: 89.62 on 2 and 68 DF,  p-value: < 2.2e-16
```

```
summary(cc_a_prd_5_mdl) # 2000 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_5_prd ~ df_prob1$ln2000 + df_prob2_a$ln2000)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2348 -0.4424  0.1191  0.5863  1.8271
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -1.9329     0.6322  -3.057  0.00307 **
## df_prob1$ln2000     0.5530     0.1124   4.922 4.77e-06 ***
## df_prob2_a$ln2000   0.4423     0.2593   1.706  0.09210 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.871 on 77 degrees of freedom
##   (37 observations deleted due to missingness)
## Multiple R-squared:  0.6145, Adjusted R-squared:  0.6044
## F-statistic: 61.36 on 2 and 77 DF,  p-value: < 2.2e-16
```

Take a look at the table of conditional convergence now. Throughout history, the importance of early education cannot be overstated, as it is significantly affect the GDP growth rate. The positive sign indicates that countries with stronger initial education have had greater GDP development. It demonstrates the importance of education for economic development.

b) Adding GADP

```
cc_b_prd_1_mdl <- lm(df_prob1$gr_1_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 + df_gadp$GADP)
cc_b_prd_2_mdl <- lm(df_prob1$gr_2_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 + df_gadp$GADP)
cc_b_prd_3_mdl <- lm(df_prob1$gr_3_prd ~ df_prob1$ln1990 + df_prob2_a$ln1990 + df_gadp$GADP)
cc_b_prd_4_mdl <- lm(df_prob1$gr_4_prd ~ df_prob1$ln1995 + df_prob2_a$ln1995 + df_gadp$GADP)
cc_b_prd_5_mdl <- lm(df_prob1$gr_5_prd ~ df_prob1$ln2000 + df_prob2_a$ln2000 + df_gadp$GADP)

summary(cc_b_prd_1_mdl) # 1970 - 2015
```

```
##
```

```
## Call:
## lm(formula = df_prob1$gr_1_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 +
##     df_gadp$GADP)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0486 -0.4469  0.1105  0.7122  1.7847
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -2.6587     0.8695  -3.058 0.003235 **
## df_prob1$ln1970      0.3095     0.1820   1.700 0.093931 .
## df_prob2_a$ln1971    0.8882     0.2385   3.724 0.000412 ***
## df_gadp$GADP         2.5116     1.0254   2.449 0.017013 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.103 on 65 degrees of freedom
##   (48 observations deleted due to missingness)
## Multiple R-squared:  0.6652, Adjusted R-squared:  0.6498
## F-statistic: 43.06 on 3 and 65 DF,  p-value: 1.913e-15
```

```
summary(cc_b_prd_2_mdl) # 1970 - 1990
```

```
##
## Call:
## lm(formula = df_prob1$gr_2_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 +
##     df_gadp$GADP)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.15797 -0.37448  0.06832  0.40252  2.06775
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -2.6900     0.6848  -3.928 0.000234 ***
## df_prob1$ln1970      0.4049     0.1484   2.728 0.008460 **
## df_prob2_a$ln1971    0.5697     0.1854   3.074 0.003243 **
## df_gadp$GADP         2.9438     0.9067   3.247 0.001957 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8533 on 57 degrees of freedom
##   (56 observations deleted due to missingness)
## Multiple R-squared:  0.7748, Adjusted R-squared:  0.763
## F-statistic: 65.38 on 3 and 57 DF,  p-value: < 2.2e-16
```

```
summary(cc_b_prd_3_mdl) # 1990 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_3_prd ~ df_prob1$ln1990 + df_prob2_a$ln1990 +
##     df_gadp$GADP)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3043 -0.4042  0.1139  0.6342  1.9855
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -3.4072     0.7694  -4.428 3.53e-05 ***
## df_prob1$ln1990     0.3034     0.2153   1.409 0.163249
## df_prob2_a$ln1990   1.1471     0.2893   3.965 0.000179 ***
## df_gadp$GADP        1.3623     1.2329   1.105 0.273066
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.014 on 68 degrees of freedom
##   (45 observations deleted due to missingness)
## Multiple R-squared:  0.6898, Adjusted R-squared:  0.6761
## F-statistic: 50.41 on 3 and 68 DF,  p-value: < 2.2e-16
```

```r
summary(cc_b_prd_4_mdl) # 1995 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_4_prd ~ df_prob1$ln1995 + df_prob2_a$ln1995 +
##     df_gadp$GADP)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.75521 -0.29275  0.08509  0.42391  1.53273
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        -1.8353     0.5895  -3.113  0.00277 **
## df_prob1$ln1995     0.4030     0.1444   2.791  0.00691 **
## df_prob2_a$ln1995   0.6202     0.2168   2.860  0.00571 **
## df_gadp$GADP        1.0967     0.8603   1.275  0.20696
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7467 on 64 degrees of freedom
##   (49 observations deleted due to missingness)
## Multiple R-squared:  0.7344, Adjusted R-squared:  0.7219
## F-statistic: 58.98 on 3 and 64 DF,  p-value: < 2.2e-16
```

```r
summary(cc_b_prd_5_mdl) # 2000 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_5_prd ~ df_prob1$ln2000 + df_prob2_a$ln2000 +
##     df_gadp$GADP)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -3.2198 -0.3661  0.1377  0.5246  1.7294
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -1.8718     0.6794  -2.755  0.00743 **
## df_prob1$ln2000      0.4555     0.1500   3.036  0.00333 **
## df_prob2_a$ln2000    0.5285     0.2673   1.977  0.05182 .
## df_gadp$GADP         0.6758     0.8207   0.823  0.41296
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8737 on 72 degrees of freedom
##   (41 observations deleted due to missingness)
## Multiple R-squared:  0.6259, Adjusted R-squared:  0.6104
## F-statistic: 40.16 on 3 and 72 DF,  p-value: 2.306e-15
```

GADP is a variable that sums up how people feel about structural policies and institutional contexts. After including the institutional variable, the variable is only significant for the 1970 – 1990 period. The variable is also significance in 1970 – 2015, but it is weaker (at 85-percent confidence interval). The institution variable appears to be only important in the past period.

c) Adding 3 regional dummies

```r
df_region <- distinct(df_base_region_dummy[c('country', 'country_code', 'region', 'region_East Asia and
```

```r
names(df_region)[4] <- "east_asia"
names(df_region)[5] <- "latin_america"
names(df_region)[6] <- "sub_saharan"
```

```r
cc_c_prd_1_mdl <- lm(df_prob1$gr_1_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan)
cc_c_prd_2_mdl <- lm(df_prob1$gr_2_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan)
cc_c_prd_3_mdl <- lm(df_prob1$gr_3_prd ~ df_prob1$ln1990 + df_prob2_a$ln1990 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan)
cc_c_prd_4_mdl <- lm(df_prob1$gr_4_prd ~ df_prob1$ln1995 + df_prob2_a$ln1995 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan)
cc_c_prd_5_mdl <- lm(df_prob1$gr_5_prd ~ df_prob1$ln2000 + df_prob2_a$ln2000 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan)

summary(cc_c_prd_1_mdl) # 1970 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_1_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
##     df_region$sub_saharan)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5667 -0.3619 -0.0225  0.6279  2.1816
```

```
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -0.9700     1.0812  -0.897  0.37308
## df_prob1$ln1970          0.2460     0.1942   1.267  0.20989
## df_prob2_a$ln1971        0.3803     0.2544   1.495  0.14002
## df_gadp$GADP             3.5944     1.1811   3.043  0.00343 **
## df_region$east_asia      0.3800     0.4188   0.907  0.36776
## df_region$latin_america  0.2721     0.3832   0.710  0.48032
## df_region$sub_saharan   -1.3442     0.4479  -3.001  0.00388 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1 on 62 degrees of freedom
##   (48 observations deleted due to missingness)
## Multiple R-squared:  0.7373, Adjusted R-squared:  0.7119
## F-statistic:    29 on 6 and 62 DF,  p-value: 2.955e-16
```

```
summary(cc_c_prd_2_mdl) # 1970 - 1990
```

```
##
## Call:
## lm(formula = df_prob1$gr_2_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
##     df_region$sub_saharan)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.82002 -0.22513  0.06323  0.31794  2.02159
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -2.5185     0.9260  -2.720  0.00877 **
## df_prob1$ln1970          0.5553     0.1720   3.229  0.00212 **
## df_prob2_a$ln1971        0.3312     0.2521   1.314  0.19447
## df_gadp$GADP             2.3043     1.0539   2.186  0.03314 *
## df_region$east_asia      0.2787     0.3728   0.748  0.45797
## df_region$latin_america -0.5257     0.3421  -1.537  0.13023
## df_region$sub_saharan   -0.4632     0.4419  -1.048  0.29920
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8457 on 54 degrees of freedom
##   (56 observations deleted due to missingness)
## Multiple R-squared:  0.7905, Adjusted R-squared:  0.7672
## F-statistic: 33.96 on 6 and 54 DF,  p-value: < 2.2e-16
```

```
summary(cc_c_prd_3_mdl) # 1990 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_3_prd ~ df_prob1$ln1990 + df_prob2_a$ln1990 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
```

```
##     df_region$sub_saharan)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3875 -0.3780  0.0054  0.5206  1.5315
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -1.7670     1.1588  -1.525   0.1321
## df_prob1$ln1990            0.2257     0.2233   1.011   0.3159
## df_prob2_a$ln1990          0.7749     0.2931   2.644   0.0103 *
## df_gadp$GADP               2.1993     1.2965   1.696   0.0946 .
## df_region$east_asia        0.6069     0.4121   1.473   0.1457
## df_region$latin_america    0.3942     0.3522   1.119   0.2671
## df_region$sub_saharan     -0.8354     0.4061  -2.057   0.0437 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9449 on 65 degrees of freedom
##   (45 observations deleted due to missingness)
## Multiple R-squared:  0.7423, Adjusted R-squared:  0.7186
## F-statistic: 31.21 on 6 and 65 DF,  p-value: < 2.2e-16
```

```
summary(cc_c_prd_4_mdl) # 1995 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_4_prd ~ df_prob1$ln1995 + df_prob2_a$ln1995 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
##     df_region$sub_saharan)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.51219 -0.29218  0.05154  0.36065  1.32109
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -1.3142     0.9873  -1.331   0.1881
## df_prob1$ln1995            0.2977     0.1545   1.927   0.0586 .
## df_prob2_a$ln1995          0.5141     0.2346   2.191   0.0323 *
## df_gadp$GADP               2.2067     0.9827   2.246   0.0284 *
## df_region$east_asia        0.4650     0.3171   1.467   0.1476
## df_region$latin_america    0.5140     0.3212   1.600   0.1147
## df_region$sub_saharan     -0.3213     0.3806  -0.844   0.4020
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7186 on 61 degrees of freedom
##   (49 observations deleted due to missingness)
## Multiple R-squared:  0.7655, Adjusted R-squared:  0.7425
## F-statistic:  33.2 on 6 and 61 DF,  p-value: < 2.2e-16
```

```r
summary(cc_c_prd_5_mdl) # 2000 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_5_prd ~ df_prob1$ln2000 + df_prob2_a$ln2000 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
##     df_region$sub_saharan)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1298 -0.3609  0.1453  0.4783  1.4911
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -1.153454   1.101072  -1.048  0.29849
## df_prob1$ln2000          0.482506   0.158234   3.049  0.00325 **
## df_prob2_a$ln2000        0.275987   0.294779   0.936  0.35241
## df_gadp$GADP             0.843084   0.927411   0.909  0.36648
## df_region$east_asia      0.836216   0.397615   2.103  0.03911 *
## df_region$latin_america -0.005626   0.289608  -0.019  0.98456
## df_region$sub_saharan   -0.331869   0.376568  -0.881  0.38121
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8503 on 69 degrees of freedom
##   (41 observations deleted due to missingness)
## Multiple R-squared:  0.6604, Adjusted R-squared:  0.6309
## F-statistic: 22.37 on 6 and 69 DF,  p-value: 1.91e-14
```

The next analysis is by adding 3 regional dummy variables; East Asia and Pacific, Latin America and Carribean, Sub-Saharan Africa. Looking at the regression result, they all have negative sign to the dependent variable. Even though the significance can only be found in 1970 – 2015 and 1990 – 2015. It shows that African countries have a lower GDP growth than non-African countries.

d) Adding CPI growth rate

Consumer Price Index growth rate, from WDI, similarly calculated by the method in the problem 1, absolute convergence.

```r
cc_d_prd_1_mdl <- lm(df_prob1$gr_1_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan +
                       df_prob2_d$cpi_1_prd)

cc_d_prd_2_mdl <- lm(df_prob1$gr_2_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan +
                       df_prob2_d$cpi_2_prd)

cc_d_prd_3_mdl <- lm(df_prob1$gr_3_prd ~ df_prob1$ln1990 + df_prob2_a$ln1990 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan +
                       df_prob2_d$cpi_3_prd)

cc_d_prd_4_mdl <- lm(df_prob1$gr_4_prd ~ df_prob1$ln1995 + df_prob2_a$ln1995 + df_gadp$GADP +
```

```r
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan +
                       df_prob2_d$cpi_4_prd)

cc_d_prd_5_mdl <- lm(df_prob1$gr_5_prd ~ df_prob1$ln2000 + df_prob2_a$ln2000 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan +
                       df_prob2_d$cpi_5_prd)

summary(cc_d_prd_1_mdl) # 1970 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_1_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
##     df_region$sub_saharan + df_prob2_d$cpi_1_prd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.94036 -0.34274  0.04453  0.45534  1.98714
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -1.06471    1.61767  -0.658 0.513571
## df_prob1$ln1970           0.37474    0.26342   1.423 0.161327
## df_prob2_a$ln1971         0.49056    0.29139   1.684 0.098770 .
## df_gadp$GADP              2.05444    1.40548   1.462 0.150331
## df_region$east_asia       0.27718    0.46116   0.601 0.550631
## df_region$latin_america   0.06916    0.42561   0.162 0.871602
## df_region$sub_saharan    -1.77310    0.47429  -3.738 0.000493 ***
## df_prob2_d$cpi_1_prd     -0.27288    0.68940  -0.396 0.693991
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9061 on 48 degrees of freedom
##   (61 observations deleted due to missingness)
## Multiple R-squared:  0.7895, Adjusted R-squared:  0.7588
## F-statistic: 25.71 on 7 and 48 DF,  p-value: 3.308e-14
```

```r
summary(cc_d_prd_2_mdl) # 1970 - 1990
```

```
##
## Call:
## lm(formula = df_prob1$gr_2_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
##     df_region$sub_saharan + df_prob2_d$cpi_2_prd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.76426 -0.21558  0.01988  0.33925  1.42208
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -2.94979    0.96897  -3.044  0.00397 **
```

```
## df_prob1$ln1970          0.61416    0.20245   3.034  0.00409 **
## df_prob2_a$ln1971        0.18658    0.24326   0.767  0.44727
## df_gadp$GADP             2.85027    1.36479   2.088  0.04272 *
## df_region$east_asia      0.35188    0.38214   0.921  0.36229
## df_region$latin_america -0.43448    0.36017  -1.206  0.23430
## df_region$sub_saharan   -0.51901    0.46201  -1.123  0.26751
## df_prob2_d$cpi_2_prd    -0.06194    0.10761  -0.576  0.56792
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7578 on 43 degrees of freedom
##   (66 observations deleted due to missingness)
## Multiple R-squared:  0.8298, Adjusted R-squared:  0.8021
## F-statistic: 29.94 on 7 and 43 DF,  p-value: 1.475e-14
```

```
summary(cc_d_prd_3_mdl) # 1990 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_3_prd ~ df_prob1$ln1990 + df_prob2_a$ln1990 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
##     df_region$sub_saharan + df_prob2_d$cpi_3_prd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.90077 -0.37725  0.00095  0.40824  1.86595
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -2.9146     1.4888  -1.958   0.0552 .
## df_prob1$ln1990           0.2708     0.2531   1.070   0.2892
## df_prob2_a$ln1990         0.8130     0.3093   2.629   0.0110 *
## df_gadp$GADP              2.5408     1.3720   1.852   0.0692 .
## df_region$east_asia       0.7612     0.4167   1.827   0.0730 .
## df_region$latin_america   0.3594     0.3698   0.972   0.3352
## df_region$sub_saharan    -0.9607     0.4107  -2.339   0.0229 *
## df_prob2_d$cpi_3_prd      0.3614     0.2850   1.268   0.2100
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8986 on 57 degrees of freedom
##   (52 observations deleted due to missingness)
## Multiple R-squared:  0.7699, Adjusted R-squared:  0.7416
## F-statistic: 27.25 on 7 and 57 DF,  p-value: 5.27e-16
```

```
summary(cc_d_prd_4_mdl) # 1995 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_4_prd ~ df_prob1$ln1995 + df_prob2_a$ln1995 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
##     df_region$sub_saharan + df_prob2_d$cpi_4_prd)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -2.38536 -0.26572 -0.04623  0.34654  1.38246 
## 
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)  
## (Intercept)               -1.6153     1.2301  -1.313   0.1944  
## df_prob1$ln1995            0.2902     0.1748   1.660   0.1024  
## df_prob2_a$ln1995          0.4892     0.2476   1.976   0.0530 .
## df_gadp$GADP               2.6549     1.0837   2.450   0.0174 *
## df_region$east_asia        0.4459     0.3186   1.400   0.1671  
## df_region$latin_america    0.5876     0.3375   1.741   0.0871 .
## df_region$sub_saharan     -0.5088     0.4109  -1.238   0.2206  
## df_prob2_d$cpi_4_prd       0.1664     0.1599   1.041   0.3025  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.7124 on 57 degrees of freedom
##   (52 observations deleted due to missingness)
## Multiple R-squared:  0.7746, Adjusted R-squared:  0.7469 
## F-statistic: 27.99 on 7 and 57 DF,  p-value: 2.96e-16
```

```r
summary(cc_d_prd_5_mdl) # 2000 - 2015
```

```
## 
## Call:
## lm(formula = df_prob1$gr_5_prd ~ df_prob1$ln2000 + df_prob2_a$ln2000 + 
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america + 
##     df_region$sub_saharan + df_prob2_d$cpi_5_prd)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -3.1643 -0.3186  0.0849  0.4639  1.5611 
## 
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)   
## (Intercept)              -1.40652    1.30942  -1.074  0.28673   
## df_prob1$ln2000           0.49119    0.17490   2.808  0.00657 **
## df_prob2_a$ln2000         0.26987    0.37626   0.717  0.47580   
## df_gadp$GADP              0.98325    1.03754   0.948  0.34680   
## df_region$east_asia       0.83517    0.40316   2.072  0.04228 * 
## df_region$latin_america -0.03246    0.30029  -0.108  0.91426   
## df_region$sub_saharan   -0.43462    0.41571  -1.045  0.29967   
## df_prob2_d$cpi_5_prd      0.09754    0.21523   0.453  0.65192   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.8607 on 65 degrees of freedom
##   (44 observations deleted due to missingness)
## Multiple R-squared:  0.6647, Adjusted R-squared:  0.6286 
## F-statistic: 18.41 on 7 and 65 DF,  p-value: 2.923e-13
```

e) Adding average trade openness

For the final variable, I will include the average trade openness.

```r
period1 <- as.character(seq(1970, 2015, 1))
period2 <- as.character(seq(1970, 1990, 1))
period3 <- as.character(seq(1990, 2015, 1))
period4 <- as.character(seq(1995, 2015, 1))
period5 <- as.character(seq(2000, 2015, 1))

df_trade['avg_prd_1'] <- rowMeans(df_trade[period1])
df_trade['avg_prd_2'] <- rowMeans(df_trade[period2])
df_trade['avg_prd_3'] <- rowMeans(df_trade[period3])
df_trade['avg_prd_4'] <- rowMeans(df_trade[period4])
df_trade['avg_prd_5'] <- rowMeans(df_trade[period5])
```

```r
cc_e_prd_1_mdl <- lm(df_prob1$gr_1_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan +
                       df_prob2_d$cpi_1_prd + df_trade$avg_prd_1)

cc_e_prd_2_mdl <- lm(df_prob1$gr_2_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan +
                       + df_prob2_d$cpi_2_prd + df_trade$avg_prd_2)

cc_e_prd_3_mdl <- lm(df_prob1$gr_3_prd ~ df_prob1$ln1990 + df_prob2_a$ln1990 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan +
                       + df_prob2_d$cpi_3_prd + df_trade$avg_prd_3)

cc_e_prd_4_mdl <- lm(df_prob1$gr_4_prd ~ df_prob1$ln1995 + df_prob2_a$ln1995 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan +
                       df_prob2_d$cpi_4_prd + df_trade$avg_prd_4)

cc_e_prd_5_mdl <- lm(df_prob1$gr_5_prd ~ df_prob1$ln2000 + df_prob2_a$ln2000 + df_gadp$GADP +
                       df_region$east_asia + df_region$latin_america + df_region$sub_saharan +
                       df_prob2_d$cpi_5_prd + + df_trade$avg_prd_5)

summary(cc_e_prd_1_mdl) # 1970 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_1_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
##     df_region$sub_saharan + df_prob2_d$cpi_1_prd + df_trade$avg_prd_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.87527 -0.28441  0.02199  0.36180  2.09717
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -2.390180   1.673646  -1.428 0.160315
## df_prob1$ln1970          0.529595   0.267169   1.982 0.053719 .
## df_prob2_a$ln1971        0.555932   0.289259   1.922 0.061107 .
```

```
## df_gadp$GADP             0.707071   1.512431   0.468 0.642445
## df_region$east_asia      0.804153   0.548767   1.465 0.149927
## df_region$latin_america -0.201668   0.430644  -0.468 0.641886
## df_region$sub_saharan   -1.798693   0.470745  -3.821 0.000414 ***
## df_prob2_d$cpi_1_prd     0.224732   0.705372   0.319 0.751537
## df_trade$avg_prd_1       0.004982   0.003256   1.530 0.133211
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8781 on 44 degrees of freedom
##   (64 observations deleted due to missingness)
## Multiple R-squared:  0.8166, Adjusted R-squared:  0.7832
## F-statistic: 24.48 on 8 and 44 DF,  p-value: 8.075e-14
```

```
summary(cc_e_prd_2_mdl) # 1970 - 1990
```

```
##
## Call:
## lm(formula = df_prob1$gr_2_prd ~ df_prob1$ln1970 + df_prob2_a$ln1971 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
##     df_region$sub_saharan + +df_prob2_d$cpi_2_prd + df_trade$avg_prd_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.17432 -0.33459  0.00791  0.38757  1.24914
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -3.697993   0.725943  -5.094 9.30e-06 ***
## df_prob1$ln1970          0.737101   0.155858   4.729 2.93e-05 ***
## df_prob2_a$ln1971        0.359980   0.184814   1.948  0.05866 .
## df_gadp$GADP             1.255719   1.135857   1.106  0.27571
## df_region$east_asia      0.917113   0.331557   2.766  0.00863 **
## df_region$latin_america -0.305203   0.274564  -1.112  0.27312
## df_region$sub_saharan   -0.363742   0.358672  -1.014  0.31677
## df_prob2_d$cpi_2_prd    -0.030475   0.080050  -0.381  0.70550
## df_trade$avg_prd_2       0.003824   0.002514   1.521  0.13636
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5464 on 39 degrees of freedom
##   (69 observations deleted due to missingness)
## Multiple R-squared:  0.9123, Adjusted R-squared:  0.8944
## F-statistic: 50.74 on 8 and 39 DF,  p-value: < 2.2e-16
```

```
summary(cc_e_prd_3_mdl) # 1990 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_3_prd ~ df_prob1$ln1990 + df_prob2_a$ln1990 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
##     df_region$sub_saharan + +df_prob2_d$cpi_3_prd + df_trade$avg_prd_3)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.87226 -0.40494  0.00044  0.55903  1.90689
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -2.802146   1.582034  -1.771  0.08262 .
## df_prob1$ln1990          0.263838   0.296885   0.889  0.37843
## df_prob2_a$ln1990        0.587664   0.375866   1.563  0.12424
## df_gadp$GADP             3.029197   1.656043   1.829  0.07334 .
## df_region$east_asia      0.696978   0.457869   1.522  0.13425
## df_region$latin_america  0.313638   0.397027   0.790  0.43328
## df_region$sub_saharan   -1.197468   0.442056  -2.709  0.00922 **
## df_prob2_d$cpi_3_prd     0.478459   0.315817   1.515  0.13607
## df_trade$avg_prd_3       0.005558   0.003201   1.736  0.08869 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9103 on 50 degrees of freedom
##   (58 observations deleted due to missingness)
## Multiple R-squared:  0.7727, Adjusted R-squared:  0.7364
## F-statistic: 21.25 on 8 and 50 DF,  p-value: 1.284e-13
```

```r
summary(cc_e_prd_4_mdl) # 1995 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_4_prd ~ df_prob1$ln1995 + df_prob2_a$ln1995 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
##     df_region$sub_saharan + df_prob2_d$cpi_4_prd + df_trade$avg_prd_4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.36899 -0.30704 -0.01437  0.37604  1.15853
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -0.625215   1.245100  -0.502  0.61765
## df_prob1$ln1995          0.144357   0.189861   0.760  0.45043
## df_prob2_a$ln1995        0.302539   0.290202   1.043  0.30191
## df_gadp$GADP             3.789260   1.132871   3.345  0.00152 **
## df_region$east_asia      0.430428   0.317972   1.354  0.18159
## df_region$latin_america  0.669998   0.333899   2.007  0.04991 *
## df_region$sub_saharan   -0.698850   0.430775  -1.622  0.11067
## df_prob2_d$cpi_4_prd     0.231060   0.157631   1.466  0.14860
## df_trade$avg_prd_4       0.003407   0.001937   1.759  0.08438 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6813 on 53 degrees of freedom
##   (55 observations deleted due to missingness)
## Multiple R-squared:  0.7768, Adjusted R-squared:  0.7431
## F-statistic: 23.05 on 8 and 53 DF,  p-value: 1.027e-14
```

```r
summary(cc_e_prd_5_mdl) # 2000 - 2015
```

```
##
## Call:
## lm(formula = df_prob1$gr_5_prd ~ df_prob1$ln2000 + df_prob2_a$ln2000 +
##     df_gadp$GADP + df_region$east_asia + df_region$latin_america +
##     df_region$sub_saharan + df_prob2_d$cpi_5_prd + +df_trade$avg_prd_5)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2322 -0.3261  0.0866  0.4996  1.5135
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -1.257855   1.339393  -0.939   0.3513
## df_prob1$ln2000          0.426363   0.187667   2.272   0.0266 *
## df_prob2_a$ln2000        0.271971   0.380794   0.714   0.4778
## df_gadp$GADP             1.273750   1.106185   1.151   0.2540
## df_region$east_asia      0.763264   0.412051   1.852   0.0687 .
## df_region$latin_america -0.015982   0.305639  -0.052   0.9585
## df_region$sub_saharan   -0.449604   0.422823  -1.063   0.2918
## df_prob2_d$cpi_5_prd     0.153227   0.222658   0.688   0.4939
## df_trade$avg_prd_5       0.001985   0.002538   0.782   0.4371
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8682 on 62 degrees of freedom
##   (46 observations deleted due to missingness)
## Multiple R-squared:  0.6482, Adjusted R-squared:  0.6028
## F-statistic: 14.28 on 8 and 62 DF,  p-value: 1.483e-11
```

Finally, after accounting for the CPI and trade openness, it seems that neither the CPI nor the trade openness have an impact on GDP growth (column 5a, 5b, 5c, 6a, 6b, 6c). What's notable is that the Sub-Saharan Africa dummy variable continues to have a negative impact on GDP growth, with substantial effects from 1970 to 2015 and 1990 to 2015. Another interesting discovery is that the absolute convergence theory only holds true for the 1970–2015 period. The relevance of the initial GDP, even after accounting for all of the control factors, demonstrates this.

Conclusion

Overall, I discovered that the absolute convergence hypothesis no longer probably applies for the current period. We should consider conditional convergence, which states that economic convergence is also dependent on country similarities. Policymakers in emerging countries should consider improving education levels in order to achieve a better future development rate.