

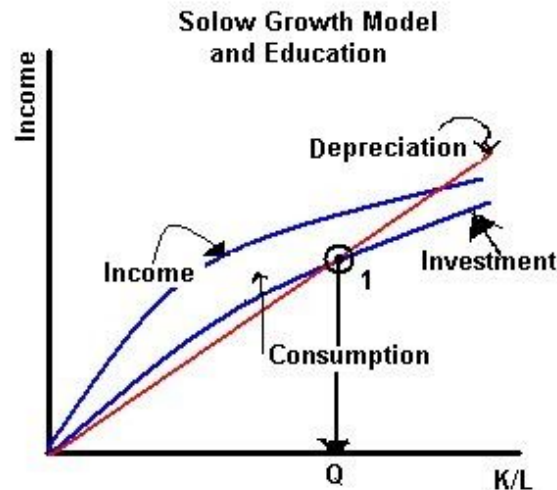


Education in Developing Nations Boosts GDP Growth

Group 16: Harley Ryan (leader), Pallavi Vemuri, Ashley Rommel
DS 4002
October 23rd, 2024

Motivation

- 1) **Motivation:**
 - a) Wanted to test the Solow Growth model
- 2) **Solow Growth Model**
 - a) Model of long run economic growth of a country
 - i) Capital stock (infrastructure)
 - ii) Labor
 - iii) Technology
- 3) **Research Question:**
 - a) What variables are the strongest predictors of GDP growth for developing and developed nations, and are those variable the same?
- 4) **Modeling Approach:**
 - a) Used a SARIMAX model, a subset of ARMA models, to predic future growth based on various factors





Data Acquisition

- 1) Data acquired from Gapminder and merged on year and country
- 2) Used World Population Review list of most developed countries in the world and UN list of least developed countries in the world
 - a) Picked top 10 countries from each list

Column	Description	Potential Responses
country	Name of the country that the data is recorded for. Text data.	"Afghanistan", "Angola"
year	Year that the data was recorded for. Numerical data.	1991, 1992
developed?	Whether the country is developed or not. Numerical data.	"yes", "no"
GDP	Total GDP growth per year. Numerical data.	7.34, 1.22
OWID Education Index	Education index calculated based on average years of schooling, taking values 0 as minimum and 15 as maximum. Numerical data.	10.7, 11.3
Agriculture Workers (%)	Percentage of all employment that works in the agriculture sector. Numerical data.	63.4, 63.7
Industry Workers (%)	Percentage of all employment that works in the industry sector. Numerical data.	12.4, 12.2
Service Workers (%)	Percentage of all employment that works in the service sector. Numerical data.	24.1, 24.0
Exports (%)	Percentage of GDP that is the value of goods and other market services provided to the rest of the world. Numerical data.	57.1, 54.3
Babies per woman (30 years lagged)	Number of children that would be born to a woman if she were to live to the end of her childbearing years and bear children in accordance with age-specific fertility rates. Numerical data.	7.37, 7.4
Child Mortality Rate (30 years lagged)	Number of deaths of children under 5 years of age per 1,000 live births. Numerical data.	230.0, 24.3
Life Expectancy (50 years lagged)	Number of years a newborn infant would live if the current mortality rates at different ages were to stay the same throughout its life. Numerical data.	41.2, 42.9

Motivation

Data Acquisition

Analysis Plan

Tricky Decision

Bias Validation

Results

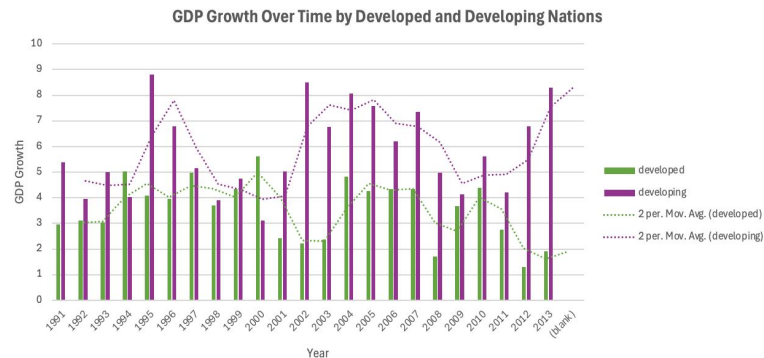
Next Steps

References

Closing

Data Trends

- 1) Afghanistan GDP growth was at its lowest during Taliban control
- 2) GDP growth over time by developed and developing countries fluctuates, but generally developing nations have a higher GDP growth rate than developed nations





Analysis Plan and Justification

- 1) Used a SARIMAX model, a subset of ARMA models, to predict future growth
 - a) Autoregression: Examines past values to predict current values
 - b) Moving Average: The moving average models past errors in predictions
 - c) Integration: Keeps time series stationary
 - d) Seasonal component: Captures variations that occur at regular intervals
 - e) Covariates: Includes external variables that can influence the time series and models them

Motivation	Data Acquisition	Analysis Plan	Tricky Decision	Bias Validation	Results	Next Steps	References	Closing
------------	------------------	---------------	-----------------	-----------------	---------	------------	------------	---------

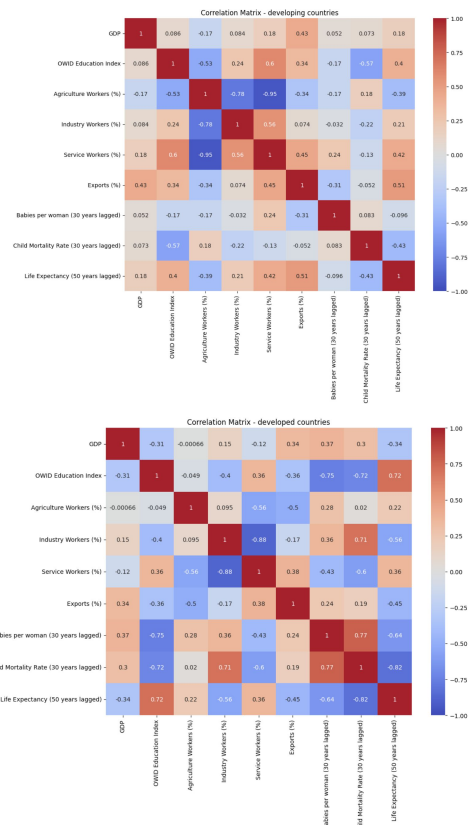


Tricky Analysis Decision

- 1) Decided to focus on education rather than all of the variables we pulled from
 - a) A lot of the variables were lagged; unclear whether it was lagged the right amount
 - b) Most interesting results
- 2) Switched from ARMA to SARIMAX model to capture seasonal component of covariates and to incorporate relevant exogenous variables

Bias and Uncertainty Validation

- 1) Tested for multicollinearity through correlation plots– did not include variables significantly contributing to multicollinearity in model
- 2) Ran a Goldfeld-Quandt test to check for heteroskedasticity– unable to conclude that there is heteroskedasticity within the data at the 95% level of significance



Results

- 1) Found a positive relationship between education scores and GDP growth in developing countries, slight negative relationship between education scores and GDP growth in developed countries.
 - a) Negative relationship is explained by the massive positive relationship between being developed and GDP growth
 - i) Correlation between being developed and GDP growth that we were unable to control for
- 2) Found autoregressive component to GDP growth, where the previous year's GDP growth can be used to predict the next year's GDP growth.

```
Optimization terminated successfully.
Current function value: 3.006117
Iterations: 6
Function evaluations: 419

=====
SARIMAX Results
=====
Dep. Variable:      GDP      No. Observations:      421
Model:              SARIMAX(1, 0, 1)  Log Likelihood        -1265.575
Date:              Fri, 25 Oct 2024    AIC                   2545.150
Time:              09:54:27           BIC                   2573.449
Sample:            12-31-1991         HQIC                  2556.334
                  - 12-31-2013
Covariance Type:    opg
=====
              coef    std err          z      P>|z|    [0.025    0.975]
-----
Intercept      0.1354      0.161      0.840      0.401     -0.181     0.451
OWID Education Index  0.1567      0.046      3.415      0.001      0.067     0.247
dummyDeveloped  13.4923      3.768      3.580      0.000      6.106    20.878
dummy OWID     -0.3209      0.064     -5.036      0.000     -0.446    -0.196
ar.L1           0.9171      0.084     10.959      0.000      0.753     1.081
ma.L1          -0.8478      0.114     -7.444      0.000     -1.071    -0.625
sigma2         23.8700      0.663     36.002      0.000     22.578    25.178
=====
Ljung-Box (L1) (Q):              0.59   Jarque-Bera (JB):              5678.89
...
=====
```



Next Steps

- 1) Autocorrelation in model between being a developed country and having higher education → fixing that in the future would create a better model
- 2) In future models we could create a predictive model to project out GDP growth for the next 5 years rather than a model to understand the contributors of GDP growth themselves.



References

Github: <https://github.com/ashleivr05/DS4002-Project2>

[1]Gapminder, "Gapminder: Unveiling the beauty of statistics for a fact based world view,," Gapminder.org, 2023. <https://www.gapminder.org/>

[2]Scikit-learn, "Scikit-learn: Machine Learning in Python,," Scikit-learn.org, 2023. <https://scikit-learn.org/stable/>

[3]United Nations, "List of LDCs | Office of the High Representative for the Least Developed Countries, Landlocked Developing Countries and Small Island Developing States," www.un.org, 2023. <https://www.un.org/ohrrls/content/list-ldcs>

[4]World Population Review, "Developed Countries List 2020," worldpopulationreview.com, 2022. <https://worldpopulationreview.com/country-rankings/developed-countries>

[5]Minitab, "Modified Box-Pierce (Ljung-Box) Chi-Square Statistic" support.minitab.com,

2023.<https://support.minitab.com/en-us/minitab/help-and-how-to/statistical-modeling/time-series/how-to/arima/interpret-the-results/all-statistics-and-graphs/modified-box-pierce-ljung-box-chi-square-statistics/#:~:text=multiples%20of%2012,Interpretation,for%20each%20chi%20square%20statistic>.

[6]Datascientest, "SARIMAX: What is It?" datascientest.com, 2023. <https://datascientest.com/en/sarimax-model-what-is-it-how-can-it-be-applied-to-time-series>

[7]Statsmodels, "SARIMAXResults.testheteroskedasticity" statsmodels.org,

2021.https://www.statsmodels.org/dev/generated/statsmodels.tsa.statespace.sarimax.SARIMAXResults.test_heteroskedasticity.html#statsmodels.tsa.statespace.sarimax.SARIMAXResults.test_heteroskedasticity

Motivation	Data Acquisition	Analysis Plan	Tricky Decision	Bias Validation	Results	Next Steps	References	Closing
------------	------------------	---------------	-----------------	-----------------	---------	------------	------------	---------

Closing Remarks

- 1) Education positively affects GDP growth in developing countries
- 2) It is demonstrably a good investment for developing countries to invest in education

