**Introduction**

I ran a variety of models to test the impact of government education expenditures education, gross fixed capital formation, annual population growth, and trade as a percentage of GDP on annual GDP growth.

**Instrumental and Dummy Variables**

The countries in the dataset represented two regions: Western Europe and Latin America. Consequently, I ran separate regressions for each region. I also created a dummy variable for each decade and one for the Great Recession that included all years between 2008 and 2012. However, I removed this variable from the final model because it was not significant (p > 0.5). It is also worth noting that I ran regressions using both t and t2 as predictor variables; I ultimately included t2 in the final model because it was significant at a lower p value.

**Best-Fit Model**

The model that best fit the data from European countries was:

*log\_annual\_gdp\_growth* = *β0 + β1t2 + β2dum\_80s + β3dum\_90s + β4dum\_00s + β4l.log\_trade\_pct\_gdp + ε*

where:

* *log\_annual\_gdp\_growth* is the log of the per-capita GDP annual growth,
* *t2* is time squared,
* *dum\_80s* is a dummy variable accounting for all observations from 1980 through 1989,
* *dum\_90s* is a dummy variable accounting for all observations from 1990 through 1999, and
* *dum\_00s* is a dummy variable accounting for all observations from 2000 through 2009.
* *l.log\_trade\_pct\_gdp* is the lagged log of trade as a percentage of GDP

The table below lists the coefficients, significance levels, and R2 for the regression using only countries in Europe.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Predictor** | **Coefficient** | **Standard Error** | **t** | **P > |t|** |
| t2 | 0.000000806 | 0.000000185 | 4.37 | 0.000 |
| dum\_80s | 0.2974803 | 0.0695092 | 4.28 | 0.000 |
| dum\_90s | 0.25935988 | 0.0673736 | 4.28 | 0.000 |
| dum\_00s | -0.1974373 | 0.0744229 | -2.65 | 0.008 |
| l.log\_trade\_pct\_gdp | 0.2108382 | 0.0736942 | 2.86 | 0.005 |
| constant | 0.2254107 | 0.3367702 | 0.67 | 0.504 |
|  | | | | |
| **Source** | **Sum of squares** | **df** | **MS** | |
| Model | 11.8424759 | 5 | 2.36849518 | |
| Residual | 56.9687921 | 308 | 0.184963611 | |
| Total | 68.811268 | 313 | 0.219844307 | |
|  | | | | |
| Number of observations | 314 | | | |
| F(4, 313) | 12.81 | | | |
| Prob > F | 0.0000 | | | |
| R-squared | 0.1721 | | | |
| Adjusted R-squared | 0.1587 | | | |
| Root MSE | 0.43007 | | | |

The model that best fit the data from Latin American countries was:

*log\_annual\_gdp\_growth* = *β0 + β1t2 + β2dum\_90s + β3dum\_00s + β4dum\_10s + β5l.log\_pop\_growth\_annual + ε*

where:

* *log\_annual\_gdp\_growth* is the log of the per-capita GDP annual growth,
* *t2* is time squared,
* *dum\_90s* is a dummy variable accounting for all observations from 1990 through 1999,
* *dum\_00s* is a dummy variable accounting for all observations from 2000 through 2009,
* *dum\_10s* is a dummy variable accounting for all observations from 2010 through 2017, and
* *l.log\_pop\_growth\_annual* is a one-year lag of the logged annual percent change in population.

The table below lists the coefficients, significance level, and R2 for the regression using only countries in Latin America.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Predictor** | **Coefficient** | **Standard Error** | **t** | **P > |t|** |
| t2 | 0.000000609 | 0.000000222 | 2.74 | 0.007 |
| dum\_90s | 0.2881554 | 0.1195233 | 2.41 | 0.017 |
| dum\_00s | 0.3097664 | 0.1215607 | 2.55 | 0.011 |
| dum\_10s | 0.2459835 | 0.1287256 | 1.91 | 0.057 |
| l.log\_pop\_growth\_annual | 0.131074 | 0.0732116 | 1.79 | 0.075 |
| constant | 1.375726 | 0.107012 | 12.86 | 0..000 |
|  | | | | |
| **Source** | **Sum of squares** | **df** | **MS** | |
| Model | 6.61300486 | 5 | 1.32260097 | |
| Residual | 106.245296 | 257 | 0.41340582 | |
| Total | 112.858301 | 262 | 0.430756873 | |
|  | | | | |
| Number of observations | 263 | | | |
| F(5, 257) | 3.20 | | | |
| Prob > F | 0.0081 | | | |
| R-squared | 0.0586 | | | |
| Adjusted R-squared | 0.0403 | | | |
| Root MSE | 0.64297 | | | |

It is worth noting that annual population growth was a significant predictor for the Latin America dataset but not for the European one. This makes some intuitive sense given that population growth tended to vary less in Europe and that population growth matters more for developing economies compared with developed ones. It is interesting, however, that population growth did not have an inverse effect on annual per-capita GDP growth, as is often assumed.

Also of note is that the dummy variable for the 1980s was significant for the Europe dataset but not for the Latin America one. Again, this makes intuitive sense if we consider the variety of economic and political conditions in Latin America during the 1980s (e.g., the repressive rule of Augusto Pinochet and other leaders, the fear struck by Peru’s Sendero Luminoso, the growing rural instability wrought by Colombia’s FARC guerillas, etc.), as compared to the relative uniform growth and expanding liberalism that Western Europe experienced leading up to the fall of the Berlin Wall in 1991.

**Model Specification Tests**

Both models pass the Ramsey RESET test, indicating there are no omitted variables [F(3, 310) = 0.70, p > 0.5 and F(3, 254) = 0.95, p > 0.4]. The data also appear to be homoscedastic, as the model passes the Breusch-Pagan and Cook-Weisberg test for heteroscedasticity (χ2 = 1.73, p > 0.18 and χ2 = 0.14, p > 0.7). Furthermore, the residuals appear to follow a normal and independent distribution, as one can see from the plots below and from the fact that the models pass the Shapiro-Wilk test and the skewness/kurtosis normality tests below and on the following page.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Skewness/Kurtosis tests for normality** | | | | | |
| **Variable** | **Observations** | **Pr(skewness)** | **Pr(kurtosis)** | **Adjusted χ2** | **Prob > χ2** |
| rEurope | 318 | 0.1185 | 0.7350 | 2.57 | 0.2773 |
| rLatin America | 263 | 0.5240 | 0.8303 | 0.46 | 0.7961 |
|  | | | | | |
| **Shapiro-Wilk test for normal data** | | | | | |
| **Variable** | **Observations** | **W** | **V** | **Z** | **Prob > z** |
| rEurope | 318 | 0.99367 | 1.360 | 0.725 | 0.23437 |
| rLatin America | 263 | 0.99730 | 0.512 | -1.563 | 0.94092 |

|  |  |
| --- | --- |
| **Europe** | **Latin America** |
|  |  |
|  |  |
|  |  |

**Alternative Model**

I also ran a Prais-Winsten correction to remove serial correlation in the data and to calculate an alternate Durbin-Watson statistic. This model returned a transformed Durbin-Watson statistic of 1.74 and an R2 of 0.0913 for the Europe dataset and a Durbin-Watson statistic of 1.87 and an R2 of 0.1621 for the Latin America dataset. These models explain about the same proportion of the variance in the model, and they transform the data and use robust standard errors to account for the significant autocorrelation in the data.