**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 created dummy variables for each of these regions. I also created a dummy variable 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 was:

*log\_annual\_gdp\_growth* = *β0 + β1dum\_eur + β2t2+ β3l.log\_gross\_fixed\_capital + β4l.log\_pop\_growth\_annual+ β5l.log\_gov\_exp\_edu + ε*

where:

* *dum\_eur* is a dummy variable accounting for all observations in Western Europe
* *t2* is time squared,
* *l.log\_gross\_fixed\_capital* is a one-year lag of the gross fixed capital formation as a percentage of GDP,
* *l.log\_pop\_growth\_annual* is a one-year lag of the annual percent change in population, and
* *l.log\_gov\_exp\_edu* is a one-year lag of the government’s total expenditures on education.

The table below lists the coefficients, significance levels, and R2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Predictor** | **Coefficient** | **Standard Error** | **t** | **P > |t|** |
| dum\_eur | -0.4713918 | 0.106394 | -4.43 | 0.000 |
| t2 | 0.000000368 | 0.00000019 | 1.93 | 0.054 |
| l.ln\_gross\_fixed\_capital | 0.101055 | 0.0777307 | 1.30 | 0.194 |
| l.ln\_pop\_growth\_annual | -0.1835892 | 0.1176289 | -1.56 | 0.120 |
| l.ln\_gov\_exp\_edu | -0.1525422 | 0.1261743 | -1.21 | 0.228 |
| constant | 1.754951 | 0.4066743 | 4.32 | 0.000 |
|  | | | | |
| **Source** | **Sum of squares** | **Df** | **MS** | |
| Model | 14.2059018 | 5 | 2.84118036 | |
| Residual | 95.4205018 | 340 | 0.280648535 | |
| Total | 109.626404 | 345 | 0.317757692 | |
|  | | | | |
| Number of observations | 346 | | | |
| F(5, 340) | 10.12 | | | |
| Prob > F | 0.0000 | | | |
| R-squared | 0.1296 | | | |
| Adjusted R-squared | 0.1168 | | | |
| Root MSE | 0.52976 | | | |

While the model has a relatively low R2 value, the transformations I had to make to the data to correct for serial correlation and non-normality of the residuals meant that a model that explained a higher proportion of the variance would have ultimately been mis-specified.

**Model Specification Tests**

The model passes the Ramsey RESET test, indicating there are no omitted variables (F(3, 337) = 0.90, p > 0.4). The data also appear to be homoscedastic, as the model passes the Breusch-Pagan and Cook-Weisberg test for heteroscedasticity (χ2 = 1.18, p > 0.25). 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 model passes the Shapiro-Wilk test and the skewness/kurtosis normality tests.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Skewness/Kurtosis tests for normality** | | | | | |
| **Variable** | **Observations** | **Pr(skewness)** | **Pr(kurtosis)** | **Adjusted χ2** | **Prob > χ2** |
| r | 346 | 0.4455 | 0.6612 | 0.78 | 0.6771 |
|  | | | | | |
| **Shapiro-Wilk test for normal data** | | | | | |
| **Variable** | **Observations** | **W** | **V** | **Z** | **Prob > z** |
| r | 346 | 0.99440 | 1.354 | 0.717 | 0.23675 |

****

****

**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 1.62 and an R2 of 0.0652. Again, this model explains relatively little of the variation but does transform the data and use robust standard errors to account for the significant autocorrelation in the data.