Chapter 15

Component 1 (Macro): Quantitative (Regression) Analysis

Several analyses were conducted to reach different explanatory models and thereby to determine the most critical variables in the datasets. This quantitative part is correlational, not experimental (causal). Therefore, the importance of a variable is based on its explanatory power, not on causal relations.

The *competitiveness GCI* index contains numerous sub-indices, which are aggregated into 12 different categories (World Economic Forum, 2016). Among the sub-indices, GCI comprises *corruption (CPI)* (or its opposite, *transparency perception*), as well as some of its determinants. However, the interactions between the variables of interest and *corruption*, on the one hand, and between the variables of interest and *competitiveness*, on the other, were analysed and modelled separately.¹

This chapter has three sections. The first overviews the process of modelling competitiveness (GCI) and corruption (CPI). The second corresponds to the methodological protocols and limitations. The third presents the empirical results of the regression analysis for GCI and CPI.

Originally published as “Materials and Methods” in: Garcia Portilla, J. (2019). “Ye Shall Know Them by Their Fruits”: Prosperity and Institutional Religion in Europe and the Americas.

Religions, 10(6), 362. MDPI AG. Retrieved from <https://doi.org/10.3390/rel10060362>

© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Supplementary Information The online version of this chapter (https://doi.org/10.1007/978-3-030-78498-0_15) contains supplementary material, which is available to authorized users.

¹Please refer to Appendices 1–2 and Supplementary Materials for a detailed description of the variables and their interactions in the general research model. Also included in these appendices is a summary of the theories, indicators, and sources utilised in the models.

15.1 Modelling Competitiveness and Corruption

15.1.1 Modelling Competitiveness (GCI) (Stages 1 and 2)

I modelled several variables associated with theories of prosperity (see Chap. 5 and Appendices 1–2, and Supplementary Materials) as possible exogenous determinants of competitiveness in most countries worldwide. For example, the *Environment Performance Index (EPI)* is linked to geographical and environmental determinants, while *religion* (adherents) is related to cultural and Weberian theory. *State religion* and *legal origins* are proxies of the institutional influence of religion. I tested the possible different contributions of those variables to explain the variation in *competitiveness (GCI)*.

The next generation of models focused on countries in Europe and the Americas, i.e. those two continents under study here. Finally, the subsequent generation of models isolated the influence of variables such as *state religion* and *share of population adherents* on different religious affiliations. Such isolation ought to benefit the model fitness as these variables are related (Barro & McCleary, 2005) and might overfit the model.

15.1.2 Modelling Corruption (CPI Stage 3)

Analysing *corruption* as a criterion variable (stage 3) associates several common predictor variables frequently used in the literature to explain corruption: *GDP per capita*, *religion*, or *political liberties* (Rose-Ackerman, 2006) (see Part III, Chap. 6, and Appendix 2). The modelling of *corruption* focused on Europe and the Americas. However, unlike the modelling of *competitiveness*, it bypassed the first generation of models (worldwide).

15.2 Methods of Regression Analysis

15.2.1 Methodological Limitations

15.2.1.1 Latest Available Data Measured at Different Time Periods

To draw the most accurate picture of existing reality, this study is based on the latest data available for all the investigated variables (most data come from 2016 reports). However, 2016 data were not available for all variables (care was taken to ensure that data were not older than 20 years). Saying this, slightly older data would not amount to a significant limitation. The prosperity performance of most Western countries exhibits comparatively long-term persistence (of a century or more) and

thus indicates historical inertia (Michalopoulos & Papaioannou, 2017; Acemoglu & Robinson, 2012; Acemoglu & Robinson, 2008). The same applies to some environmental variables such as geography, latitude (Diamond, 1997), or state religion (Barro & McCleary, 2005).

15.2.1.2 Regression Analysis Is Not a Causal Approach

A significant correlation does not always indicate causation (Arruñada, 2010, p. 908). Moreover, causal approaches to the present dataset and research design (i.e. based on the latest available data) might not be desirable for two reasons: First, this would lead to endogeneity problems. Testing plain variables as evident in the model would render regression results inconsistent due to the likely presence of endogeneity (i.e. because of simultaneous causality). Reverse causality is a widespread issue in regression models. Likewise, testing variable causality blindly (i.e. merely statistically, without any historical or theoretical consideration) would not be appropriate either. Thus, the historical approach to the variables under investigation is explained and addressed in detail (Figs. 1.1 and 8.1; Chaps. 6–12). Second, adopting a causal approach would require Instrumental Variable (IV) estimation, a complex experimental design, and a definition of independent, dependent, and control variables. It would also need instrumental variables, which might not exist in reality (or might be highly challenging to identify and define). While a time series approach would partially resolve the issues arising from a causal quantitative approach, it is not addressed in this study (see below). However, one other method in this research (QCA) uses a causality approach.

15.2.1.3 This Regression Analysis Excludes a Time Series Approach

A time series would be an ideal approach to test whether variables are related over short or rather over long periods. This approach would allow identifying whether relationship patterns between variables exist over time. Conceptually, it would make sense to relate performance in a period t with explanatory factors measured in period $t-1$. One example would be if increasing or decreasing religiosity or the influence of a Christian denomination coincided with a prosperity variable (i.e. whether changes in a population's religious affiliation relate to changes in corruption).

The relationship between prosperity and religion in Protestant versus Catholic societies could be tested in time and space. Regarding time, for example, countries that were Roman Catholic before the Protestant Reformation (e.g. England, Germany or Switzerland) were pre-industrial (mainly primary sector economies) in the Middle Ages. In geographical space, for instance, countries currently under pervasive Roman Catholic influence (e.g. Spain or Latin America) tend to exhibit lower performance than Protestant ones.

Analysing Christian religion reveals that the Reformation initiated gradual changes, including its slow institutionalisation and influence on democracy, literacy,

and the state (Becker et al., 2016; Woodberry, 2012; Witte, 2002). Consequently, a time series approach ought to cover longer periods (at least several centuries) based on the assumption that the prosperity of most Western countries maintains almost the same relative position over long-term time periods (a century or more) (Michalopoulos & Papaioannou, 2017; Acemoglu & Robinson, 2008).

A time series approach might also alleviate endogeneity issues, as current performance cannot (directly) influence past conditions. Still, relevant reverse causality might exist via indirect correlation. Nevertheless, a time series approach also has the disadvantage that the same available data does not exist for all variables. Some variables are measured yearly, some every 5 years, and some every 50 years (most have only been measured during the last 50 years or less). For some essential variables, no information and proxies exist, thus making it necessary to create new databases to cover particular periods (e.g. Pre-Reformation and Reformation). Undoubtedly, the time series approach is promising, yet requires considerable time and resources (which, however, exceed the scope of the present book). Notwithstanding this reservation, future research might consider applying this approach.

The current approach uses “legal origin” (La Porta et al., 1999) and “state religion” (Barro & McCleary, 2005) to capture some variable relationships over long periods. Barro and McCleary (2005) proved that a growing fraction of adherents to a country’s main religion increases the probability of a “state religion”. This gradual process means that few Western countries changed their status of having a “state religion” in the twentieth century.

15.2.2 *Data and Empirical Strategy*

Population

The countries studied here were selected based on data availability (the same consideration applies to QCA, see Chap. 16). Data stem from constructed and from secondary (existing) data available in public databases. Further, the data consist of censuses of currently available indicators for the following cross-country analyses:

Stage 1. Worldwide (107 countries): The database comprises data series available on competitiveness, corruption, social inequality, as well as other social, environmental, and economic indicators. It also includes indicators of denominational tradition and religious background for most of the surveyed countries.

Stages 2 and 3. Europe and the Americas (66 countries—competitiveness; 61 countries—corruption): Cases with straightforward data access for most variables in Europe and the Americas.² Unlike in other regions, Christianity has been the

²Due to data availability issues, some countries were excluded (e.g. most of the small island states in the Caribbean). However, most countries in Europe and the Americas provide rich data. See Supplementary Materials for detailed information on countries, sources and main variables.

dominant religion for at least the past two centuries in Europe and the Americas. The populations of these continents currently adhere largely to Roman Catholic, Orthodox, or Protestant churches of various denominations (Johnson & Zurlo, 2016).

15.2.3 Protocol

I followed an 11-step protocol (summarised below):

Stage 1: Competitiveness in the World

Step 1 aimed to include as many countries as possible in the database. However, not all the variables had the same data available for all the observations. Variables such as the *GLOBE* cultural index (House et al., 2004) were excluded from the analysis as the corresponding data were only available for a sample of three dozen countries. If less than 25% of values were still missing in the remaining dataset, they were imputed using Multivariate Imputation by Chained Equations in R (mice).

Step 2 involved using automated functions to help select from a high number of variables (see diagram above) from different theoretical backgrounds (see Appendix 1). Although the results of one automated model search may be arbitrary, about 40 of these models were created using different approaches to eliminate path-dependence and bias.

Step 3 concentrated on meeting linear regression assumptions in all the models. Assumptions were tested in a continuous process whenever a new model was created. Among other assumptions, I tested for multicollinearity by checking the

Variance Inflation Factor (VIF) and the correlation value. VIF results refused multicollinearity and, along with low correlation value, suggest that these variables are orthogonal. Likewise, all significant correlations among the predictor and criterion variables were controlled for.

Step 4 involved using cross-validation based on the prediction quality of shuffled groups of observations. Those model and variables exhibiting the best prediction ability were selected. “Step” is an automatic method based on the R function step (), which is also used to perform variable selection. These methods are useful when the number of explanatory variables is large, and when fitting all possible models proves unfeasible (University of Columbia, 2017).

Step 5 used Beta standardisation, which is necessary for comparing influence among variables. This indicates the significance of each variable in the context of others (and thus facilitates comparison). The goal, however, is not to test significance. Nor can Beta standardisation be computed for interaction variables (see second column “Standardised” in the Results).

In Step 4, Model 1 (see Results), two selected variables that might potentially outshine each other were analysed: religious denomination (proportion in population) and state religion. Consequently, I isolated these variables and ran separate models (see Step 6).

Stage 2: Modelling for Europe and the Americas (*competitiveness*).

Europe and the Americas exhibited higher data quality and availability. Consequently, there was no need to impute data as in the world models (Stage 1: Models 1, 2, and 3). Therefore, stable regional models with higher reliability (i.e. avoiding data imputation and thus reducing bias) were created (Stage 2: Models 4 and 5).

Step 7. Due to data availability issues, some countries were excluded (e.g. most of the small island states in the Caribbean). Step 8 was similar to steps 2, 3, and 4, while step 9 was similar to step 5.

Stage 3: Modelling for Europe and the Americas (*corruption*).

The competitiveness index (GCI) includes corruption (or its opposite, transparency) perception (CPI). However, to analyse whether transparency has the same direction as competitiveness, I ran a model that considers corruption as a criterion variable applicable to both Europe and the Americas.

Step 10. The corruption index (*CPI*) lacks data in several observations; therefore, fewer countries were selected from the dataset than in stage 2 (Step 7: Competitiveness).

Step 11. Regression trees are used for classification purposes or to predict outcomes when the response variable is numerical or continuous as in this situation (Morgan, 2014).

15.3 Empirical Results of Regression Analysis³

Appendices 1–2 provide a detailed description of the variables, theories, categories, and sources used here, as well as of the interactions between the variables in the general research model. Presented below are the results for competitiveness (Models 1–5 and conclusions) and for corruption (Model 6 and conclusions).

15.3.1 Competitiveness

15.3.1.1 Stage 1: Competitiveness in the World

The analysis of competitiveness as a criterion variable associates different predictor variables usually corresponding to theories of prosperity. Model 1 was chosen from more than 30 other models as it has the highest cross-validations and satisfactorily explains GCI variability.

³Some parts of this section were originally published as “Empirical Results (Correlational Analysis)” in: Garcia Portilla, J. (2019). “Ye Shall Know Them by Their Fruits”: Prosperity and Institutional Religion in Europe and the Americas. *Religions*, 10(6), 362. MDPI AG. Retrieved from <https://doi.org/10.3390/rel10060362>

© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

15.3.1.1.1 Model 1

$$GCI_i = \beta_0 + \beta_1 EPI_i + \beta_2 Mulatto_i + \beta_3 Asian_i + \beta_4 Protestant_St.Rel_i + \beta_5 Catholic_pop_i + \beta_6 Orthodox_pop_i + \beta_7 Protestant_pop_i + \beta_8 Muslim_pop_i + \beta_9 German_LO_i + \epsilon_i$$

Positive Correlations

The most significant variable is the *Environment Performance Index (EPI)*, which is highly positively correlated with competitiveness *GCI* (0.72). If *EPI* was removed from the model, *R*-squared would drop by 17%. However, the same model without *EPI* exhibits similar results to Model 1 (Table 15.1) for the remaining variables. Here, an increase of *EPI* by one point is associated with a *GCI* growth of 0.026 points. *EPI* occupies larger units (mean value 70) than *GCI* (mean value 4) (The scale is standardised, meaning that these values refer to the non-standardised ranges; see Appendix 1 and Supplementary Materials).

A second important variable positively related to *competitiveness (GCI)* is the German legal origin: “*GERMAN (LO)*”. Such countries have on average a higher *GCI* by 0.229.

Third, if Protestantism was the state religion “*PROTESTANT (S.R)*” (e.g. UK, Sweden, Denmark), then the overall effect would be positive on *GCI*. Establishing Protestantism as a state religion leads to a *GCI* increase of 0.223, which is far greater (almost double) than for a highly Protestant population. This confirms the higher

Table 15.1 Model 1. Competitiveness in the world

Coefficients:	Estimate	Standardised	Std. error	t value	Pr(> t)	
(Intercept)	2.74106	0	0.274644	9.98	2.25×10^{-16}	***
EPI	0.026048	0.586	0.002671	9.752	6.86×10^{-16}	***
Mulatto (ethn)	−0.443397	−0.117	0.172934	−2.564	0.011951	*
Asian (ethn)	−1.880652	−0.194	0.434192	−4.331	3.74×10^{-5}	***
PROTESTANT (S. R)	0.577261	0.223	0.168104	3.434	0.00089	***
Catholics (%)	−0.574504	−0.273	0.220611	−2.604	0.010722	*
Orthodox (%)	−0.78216	−0.27	0.23189	−3.373	0.001085	**
Protestants (%)	−0.541959	−0.146	0.310636	−1.745	0.084344	.
Muslims (%)	−0.472164	−0.203	0.234843	−2.011	0.047269	*
GERMAN (L.O.)	0.942412	0.229	0.1817	5.187	1.25×10^{-6}	***
–						

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘.’ 1. Residual standard error: 0.2961 on 97 degrees of freedom. Multiple *R*-squared: 0.8355, Adjusted *R*-squared: 0.8142. *F*-statistic: 39.35 on 12 and 93 DF, *p*-value: $<2.2 \times 10^{-16}$

Source: Author’s collection

importance of the institutional influence of religion compared with that of the proportion of adherents (in line with Sect. 7.1.)⁴

Negative Correlations

All variables related to *religion distribution in a population*—*Catholics (%)*, *Orthodox (%)*, *Protestants (%)*, *Muslims (%)*)—are negatively correlated with *GCI* although these correlations are only marginally significant. Nonetheless, the *Orthodox population* causes the most substantial negative effect. An increase by one percentage point of the *Orthodox population* leads to a decrease of 0.0078 units in *GCI*. The same goes for changes in the *Roman Catholic population*, where an increase in one percentage point of their population would mean a *GCI* decrease by 0.0057. Likewise, if the share of the *Protestant population* increased by one percentage point, *GCI* would decrease by 0.0054 units. Finally, if the share of the *Muslim population* increased by one percentage point, *GCI* would decrease by 0.0047 units.

On the other hand, *Mulatto* and *Asian ethnical* values both have a negative effect. A growth of one percentage point in the mulatto/Asian population means a 0.00117/0.00194 decrease in *GCI*, respectively.

Models Analysing Major Religious Population Groups or State Religions Separately

Model 1 (Table 15.1) has shown a differential influence of the *religion distribution in a population* and of *state religion* on *GCI*. Models 2 (Table 15.2) and 3 (Table 15.3) were run to separate any potential differential influence. Separate analysis of the two variables eliminates the possibility that a religious population distribution and a state religion would “overfit” the model.

15.3.1.1.2 Model 2 with Population Percentage

Model 2 shows the relation between the *percentage of religious adherents* and *competitiveness* while excluding *state religion* variables.

$$\begin{aligned} GCI_i = & \beta_0 + \beta_1 EPI_i + \beta_2 Caucasian_i + \beta_3 Mulatto_i + \beta_4 Asian_i + \beta_5 Dogmas_i \\ & + \beta_6 Socialist_i + \beta_7 German_i + \beta_8 Catholic_pop_i + \beta_9 Muslim_pop_i \\ & + \beta_{10} Orthodox_pop_i + \epsilon_i \end{aligned}$$

Model 2 (Table 15.2) confirms the findings of Model 1 (Table 15.1) in the same order:

⁴Exceptionally, while “*Agnostic*” and “*Independent*” variables of religion distribution in a population positively affected *GCI*, they were not significant enough to be included in the model.

Table 15.2 Model 2: Competitiveness in the world including the percentage of *religious adherents* and excluding *State religion* variables

Coefficients:	Estimate	Standardised	Std. error	t value	Pr(> t)	
(Intercept)	2.68713	0	0.411929	6.523	3.22×10^{-9}	***
EPI	0.031948	0.718076	0.003588	8.904	3.35×10^{-14}	***
Caucasian (ethn)	0.3062	0.182286	0.137576	2.226	2.84×10^{-2}	*
Mulatto (ethn)	-0.49164	-0.129827	0.212469	-2.314	0.022804	*
Asian (ethn)	-2.262534	-0.233818	0.607969	-3.721	0.000334	***
Dogmas	-0.06555	-0.161069	0.033539	-1.954	5.36×10^{-2}	.
SOCIALIST (L.O.)	-0.223811	-0.125531	0.111884	-2	0.048283	*
GERMAN (L.O.)	0.608521	0.147437	0.221393	2.749	0.007151	**
Catholics (%)	-0.435477	-0.217661	0.131934	-3.301	0.001355	**
Muslims (%)	-0.349915	-0.155187	0.191688	-1.825	0.071044	.
Orthodox (%)	-0.742669	-0.235604	0.192192	-3.864	0.000203	***
–						

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Residual standard error: 0.3554 on 96 degrees of freedom. Multiple *R*-squared: 0.7559, Adjusted *R*-squared: 0.7304. *F*-statistic: 29.72 on 10 and 96 DF, *p*-value: $<2.2 \times 10^{-16}$

Source: Author's collection

1. *EPI* is the most significant variable.
2. Increasing *Orthodox*, *Catholic*, and *Muslim* populations negatively influence *GCI*. However, the *Protestant population* and *other Christian adherents* are not significant in Model 2.
3. Model 2 confirms the direction and influence of most other variables in Model 1. However, in Model 2, *Caucasian* ethnic values positively influence *GCI*. Likewise, *Dogmas* and *Socialist* legal origin negatively impact *GCI*.

15.3.1.1.3 Model 3 (Including State Religion)

Model 3 (Table 15.3) excludes the *percentage of religion adherents* variables. *State religion* variables alone explain most of the variability otherwise explained by religious population.

$$GCI_i = \beta_0 + \beta_1 EPI_i + \beta_2 Caucasian_i + \beta_3 Mulatto_i + \beta_4 Asian_i + \beta_5 Catholic_i + \beta_6 Protestant_i + \beta_7 Socialist_i + \beta_8 German_i + \epsilon_i$$

The results of Model 3 (Table 15.3) ratify those of the previous models. The most significant positive influence on *GCI* comes from *EPI*, *Protestant state religion*, *Caucasian ethnicity*, and *German legal origin*. In contrast, *Asian* and *Mulatto ethnicities*, *Roman Catholic state religion*, and *Socialist legal origin* would negatively affect *GCI*.

Table 15.3 Model 3: Competitiveness in the world including *state religion* and excluding the percentage of *religion adherents*

Coefficients:	Estimate	Standardised	Std. error	<i>t</i> value	Pr(> <i>t</i>)	
(Intercept)	1.94889	0	0.205025	9.506	1.57×10^{-15}	***
EPI	0.032097	0.721417	0.003403	9.433	2.25×10^{-15}	***
Caucasian (ethn)	0.233095	0.138765	0.137925	1.69	9.42×10^{-2}	.
Mulatto (ethn)	−0.432751	−0.114276	0.197873	−2.187	0.03115	*
Asian (ethn)	−1.41671	−0.146408	0.503625	−2.813	0.00594	**
CATHOLIC (S.R)	−0.245126	−0.156268	0.091698	−2.673	8.82×10^{-3}	**
PROTESTANT (S. R)	0.307335	0.118639	0.147376	2.085	0.03966	*
SOCIALIST (L.O.)	−0.232813	−0.130581	0.113454	−2.052	0.04286	*
GERMAN (L.O.)	0.60705	0.147081	0.231532	2.622	0.01015	*
—						

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘.’ 1. Residual standard error: 0.3552 on 98 degrees of freedom. Multiple *R*-squared: 0.7536, Adjusted *R*-squared: 0.7307. *F*-statistic: 32.96 on 9 and 97 DF, *p*-value: $<2.2 \times 10^{-16}$

Source: Author’s collection

15.3.1.2 Stage 2: Modelling Competitiveness (Europe and the Americas)

The following models were produced explicitly for Europe and the Americas, excluding the noise existing in the world database. Likewise, Models 4 and 5 serve to compare whether the same variables chosen in the previous “world” models are still significant in Europe and the Americas.

15.3.1.2.1 Model 4: Results of Cross-Validation

$$GCI_i = \beta_0 + \beta_1 Hostilities_i + \beta_2 German_lan_i + \beta_3 Catholic_pop_i + \beta_4 Orthodox_pop_i + \beta_5 EPI_i + \epsilon_i$$

This result shows the high robustness of this model (Table 15.4), in that it explains almost 66% of *GCI* variability with four significant variables at a 99% confidence level:

1. *EPI* is the most important variable because it accounts for most of the *GCI* variability. The increase in *EPI* by one percentage point is related to a “*GCI*” growth of approximately 0.038 percentage points.
2. The second most important variable is the *Orthodox population*, which exerts the most substantial negative effect in the model. If the *Orthodox population* increased by one percentage point, *GCI* would decrease by approximately 0.010 percentage points.

Table 15.4 Model 4: Competitiveness in Europe and the Americas (cross-validation method)

Coefficients:	Estimate	Standardised	Std. error	<i>t</i> value	Pr(> <i>t</i>)	
(Intercept)	1.583	0	0.48642	3.254	0.00188	**
Hostilities	0.07461	0.18955	0.03368	2.215	0.03063	*
German (lang)	0.82223	0.24086	0.26625	3.088	0.00307	**
Catholics (%)	−0.49853	−0.29507	0.15608	−3.194	0.00225	**
Orthodox (%)	−1.04274	−0.44616	0.21537	−4.842	9.66×10^{-6}	***
EPI	0.03829	0.52268	0.00596	6.424	2.53×10^{-8}	***
–						

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘.’ 1. Residual standard error: 0.3771 on 59 degrees of freedom. Multiple *R*-squared: 0.6556, Adjusted *R*-squared: 0.6264. *F*-statistic: 22.46 on 5 and 59 DF, *p*-value: 1.54×10^{-12}

Source: Author’s collection

- Similarly, the effect of the *Roman Catholic population* also negatively influences competitiveness. If the *Roman Catholic population* increased by one percentage point, *GCI* would drop by approximately 0.005 percentage points. This finding, along with the previous one (2), is consistent with La Porta et al.’s (1999) conclusions about the negative influence of hierarchical religions on prosperity.
- Finally, the proportion of the *German-speaking population* positively affects *GCI*. An increase of 1 percentage point in the *German-speaking population* would mean a *GCI* increase by approximately 0.008 percentage points. This finding would confirm a positive relationship (as outlined in Sect. 11.1).

15.3.1.2.2 Model 5: Results with Step

$$GCI_i = \beta_0 + \beta_1 Hostilities_i + \beta_2 GERMAN_LEGAL_i + \beta_3 Catholic_pop_i + \beta_4 Orthodox_pop_i + \beta_5 EPI_i + \epsilon_i$$

“Step” is a fundamentally different method, yet mostly exhibits the same results (Table 15.5) as those of cross-validation (Table 15.4). Step analysis further confirms the choice of the right variables (based on reality, not on random data effects). Here, the only difference is that Step analysis chose *GERMAN (legal origin)* instead of *German language* (percentage of German-speaking population) with 99% confidence. The chapters that discuss the theoretical framework and conclusions further discuss these variables and findings.

GERMAN (legal origin) is a binomial variable, meaning that if a country is of *German legal origin*, then its *GCI* is 0.759 higher. If it is not of *German legal origin*, then the variable does not affect *GCI*. Only these two extremes exist.

On the other hand, the variable *Social Hostilities due to religion* also appears in Models 4 (Table 15.4) and 5 (Table 15.5). It exhibits the lowest value of standardised

Table 15.5 Model 5: Competitiveness in Europe and the Americas (Step method)

Coefficients:	Estimate	Standardised	Std. error	<i>t</i> value	Pr(> <i>t</i>)	
(Intercept)	1.606434	0	0.479997	3.347	0.00143	**
Hostilities	0.075303	0.191317	0.033201	2.268	0.027	*
GERMAN (L.O.)	0.759773	0.260367	0.224315	3.387	0.00126	**
Catholic_pop	−0.499403	−0.295584	0.153919	−3.245	0.00194	**
Orthodox_pop	−1.04258	−0.446087	0.212236	−4.912	7.49×10^{-6}	***
EPI	0.037974	0.518342	0.005882	6.456	2.24×10^{-8}	***
—						

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Residual standard error: 0.3719 on 59 degrees of freedom. Multiple *R*-squared: 0.6651, Adjusted *R*-squared: 0.6367. *F*-statistic: 23.43 on 5 and 59 DF, *p*-value: 6.91×10^{-13}

Source: Author's collection

beta and also low values. More importantly, this variable appeared only at 90% confidence value (all other results had a 99% confidence value). Therefore, this variable is not sufficiently significant.

Finally, standard deviation confirmed the results of both cross-validation (Table 15.4) and Step (Table 15.5) and found the same, most robust coefficients. Consequently, only four variables in the models had statistical significance.

15.3.1.3 Conclusions for Competitiveness

15.3.1.3.1 Conclusions for Competitiveness in the World (All Models)

The consistent results of the five previous models are:

1. A positive influence of *EPI* on *GCI*;
2. A positive influence of a *German legal origin* (or *German language*) on *GCI*;
3. A negative influence of an *Orthodox population* on *GCI*;
4. A negative influence of a *Roman Catholic population* (or *Roman Catholic State Religion*) on *GCI*.

These results are valid for the world—Models 1 (Table 15.1), 2 (Table 15.2), and 3 (Table 15.3)—, as well as for Europe and the Americas—Models 4 (Table 15.4) and 5 (Table 15.5)—.

Partially conclusive findings:

Ethnic influence appeared with some degree of importance in the world models but disappeared in Europe and the Americas.

Socialist legal origin negatively influenced *GCI*—Models 2 (Table 15.2), and 3 (Table 15.3)—.

The influence of the share of *Protestants in the population* is not conclusive. In Model 1 (Table 15.1), *Protestants* negatively affected *GCI*, after *Orthodox* and *Roman Catholics*. However, in Model 2 (Table 15.2), which analyses religious population, the influence of *Protestants* disappeared.

Neither *Protestant population* (1) nor *Protestant State Religion* (2) are significant variables for competitiveness in Europe and the Americas—Models 4 (Table 15.4) and 5 (Table 15.5)—. Hypothetical reasons for such findings include, respectively, (1) the high influence of Pentecostalism in the Protestant population today might neutralise the possible positive effect of historical Protestantism. Pentecostalism impacts little on human capital and institutions (Becker et al., 2016; McCleary, 2013; Woodberry, 2012) and has often fallen into established practices “of corporatism and clientage” (Martin, 1999, p. 40; Schäfer, 1997) (Sect. 10.4.3). (2) While Switzerland and the USA are the most competitive (*GCI*) countries, they do not have Protestantism as their state religion, despite being historically Protestant (Barro & McCleary, 2005; Inglehart & Baker, 2000).

Nonetheless, *Protestant State Religion* exhibited a positive significance on Model 3 (State Religion). This suggests that *Protestant State Religion* is more important for *GCI* than the proportion of *Protestants in the population*. The effect of *Protestant State Religion* on higher *GCI* might also be related to its influence in diminishing the institutional power of the Roman Church (Sect. 8.3.4). The latter conclusions both confirm Fanfani’s claim as early as 1936 (as cited in Grier, 1997) that the separation of state and church is the critical prosperity trigger. Such a separation occurred mainly in Protestant countries for anti-clerical reasons. Fanfani argued that religion per se harms prosperity unless it leads to the separation of ecclesiastical and political/economic powers in a country (as historical Protestantism did).

15.3.1.3.2 Conclusions for Competitiveness in Europe and the Americas

In Europe and the Americas, Models 4 and 5 discover a combination of variables that largely explains *GCI* variability in the following order of importance:

Environmental Performance Index (EPI)

I found a high correlation of EPI with competitiveness (*GCI*), as suggested by the environment-and-geography theory of prosperity (Diamond, 1997; Sachs, 2001; Brown & Lall, 2006). This index has by far the highest positive influence on *GCI* of all the variables considered and explains most of the *GCI* diversity. Higher *EPI* strongly implies higher *GCI*.

Legal Origin

As predicted in the variables description sections (Chap. 3–12), I found a highly positive relation of *German legal origin* and *German language* with *GCI*. The influence of the Reformation on the German legal system has been widely discussed (Witte, 2002; Berman, 2003) (Sect. 8.3.4). Likewise, several studies (Besch, 1999; Greenslade, 1963) have discussed the influence of the Reformation on the dissemination and standardisation of the German language (Sect. 11.1).

15.3.2 Corruption

15.3.2.1 Proxy Predictor Variables for the Regression Model of Corruption

The quantitative model used here tested some proxy variables of selected main determinants of corruption (Rose-Ackerman, 2006; Alesina et al., 2003). Such predictor variables are:

1. Proxies of economic prosperity: the *log GDP* per capita (*GDPPC*) and the *Index of Economic Freedom* (*IEF*);
2. Proxies of democracy: the *Political Rights and Civil Liberties Index* (*PRCL*) and the *Political Stability Index* (*STABIL*).
3. Proxies of culture: the *percentage of Roman Catholics* (*CATH*), and *Protestants* (*PROT*) in the population. This also includes *ethnic, linguistic, and religious fractionalisation* measures (*ETHF*, *LING* and *RELF*, respectively).

For each variable, the latest available data were used (one observation per country). Please refer to Annex 1 for details on the variables and sources.⁵

$$CPI_i = \beta_0 + \beta_1 GDPPC_i + \beta_2 IEF_i + \beta_3 PRCL_i + \beta_4 STABIL_i + \beta_5 PROT_i + \beta_6 CATH_i + \beta_7 ETHF_i + \beta_8 LINGF_i + \beta_9 RELF_i + \varepsilon_i$$

15.3.2.2 Regression Results for Corruption in Europe and the Americas

Data analysis includes a regression tree (see Fig. 15.1). The *Political Rights and Civil Liberties Index* (*PRCL*) is the most critical factor in determining the *Corruption Perception Index* (*CPI*). Lower corruption levels ($CPI = 81.5$) are evident in countries with a ($PRCL \geq 44$) and that also have a higher *log GDP* per capita ($GDPPC > 10.6$). Such results are entirely consistent with the respective theoretical predictions (Sects. 3.1, 3.3 and 8.1 indicate that these variables are part of the same phenomenon). In line with theory (Sect. 6.1), transparency increases even more if the country in question also has a high proportion of *Protestants* ($PROT > 0.24$) in its population ($CPI = 84.75$).

This regression tree confirms that *PRCL*, *GDPPC*, and *PROT* are all positively related to transparency (or vice versa, negatively related to corruption). *GDPPC* and

⁵Note: I have previously explained the essential theoretical variables and clarified that prosperity, democracy, and transparency are all part of the same outcome in terms of my research aims. Therefore, the variables researched in this regression model do not show causal effects. Finding causality between corruption and GDP, *PRCL*, *IEF*, or *STABIL* is problematic due to endogeneity issues (Lambsdorff, 2006) (i.e. in reality, they might all form part of the same phenomenon or syndrome and all belong to the GCI index) (World Economic Forum, 2016). Further, correlations only exhibit relations but no causalities (Sect. 15.2.1).

Fig. 15.1 Regression tree, with the mean value of *CPI*, the number of observations, and percentage deviance. The minimum number of observations per node for a split to be attempted is set to 15. Any split not decreasing the overall lack of fit by a factor of 0.005 is not attempted. (Source: Author’s collection)

PROT do not seem to play an essential role in countries with low *PRCL* (democracy). This result is also in line with theory. For instance, Treisman (2000) found that only stable democracy (before 1950) reduces corruption. Most of the Protestant upsurge in the developing world has stemmed from Pentecostalism after 1950 (Johnson & Zurlo, 2016), which only marginally impacts human capital and institutions (Becker et al., 2016; McCleary, 2013; Woodberry, 2012).

The estimates in Table 15.6 corroborate the relation between lower *corruption* (or higher transparency) and a high proportion of Protestants in the population. Similarly, other predictor variables, above all *GDPPC*, *PRCL*, *LING* correlate with lower corruption levels (i.e. higher *CPI*). The negative impact of the share of Catholics in the population is less visible, with a lower coefficient in magnitude. These results are entirely compatible with theory, in particular with Treisman (2000).

As indicated by the positive coefficients for *LING*, linguistic fractionalisation tends to imply lower corruption. This trend is illustrated by countries (e.g. Luxembourg, Canada, or Switzerland) facing low corruption and featuring in the top ten countries of the present sample regarding language fractionalisation. Conversely, a higher level of ethnic fractionalisation (*ETHF*) tends to be associated with higher levels of corruption, as established by other empirical studies (Sect. 10.1). This observation confirms that linguistic heterogeneity does not necessarily coincide with ethnic heterogeneity. For instance, most Latin American countries are relatively homogenous concerning language but less so regarding “ethnicity” or “race”.

Table 15.6 Model 6: Corruption in Europe and the Americas

Coefficients:	Estimate	Std. error	<i>t</i> value	Pr(> <i>t</i>)	
(Intercept)	−95.7267	22.615	−4.233	0.0000962	***
GDPPC	10.5639	2.1555	4.901	0.0000101	***
IEF	0.2921	0.1499	1.949	0.05686	.
PRCL	0.546	0.1721	3.173	0.00256	**
STABIL	4.0372	2.7122	1.489	0.14276	
PROT	17.9297	6.2714	2.859	0.00614	**
CATH	−2.242	4.2163	−0.532	0.59721	
ETHF	−10.4336	6.4496	−1.618	0.11189	
LING	13.9763	6.6231	2.11	0.03976	*
RELF	−1.1927	5.9927	−0.199	0.84303	
–					

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘.’ 1 Residual standard error: 7.503 on 51 degrees of freedom. Multiple *R*-squared: 0.8848, Adjusted *R*-squared: 0.8645. *F*-statistic: 43.54 on 9 and 51 DF, *p*-value: $<2.2 \times 10^{-16}$

Source: Author’s collection

The regressions do not display any significant relation between religion fractionalisation *RELF* and corruption. Consequently, low corruption is present in both multicultural, free societies with high measured religious fractionalisation (e.g. USA, United Kingdom, Canada), and countries with low religious fractionalisation (e.g. Luxembourg or Belgium). A “packed” index of fractionalisation measures works like a “black box” as analysing interactions between the components (i.e. each language or ethnicity) is impossible.

15.3.2.3 Conclusions for Corruption in Europe and the Americas

Generally speaking, the conclusions reached here about the selected variables resemble those of previous research (Treisman, 2000; Paldam, 2001; Arruñada, 2010; Chase, 2010; and to a less extent La Porta et al., 1999). Results were validated using different databases and methods.

15.3.2.3.1 Proportion of Protestants

The findings confirm my hypothesis that corruption levels are directly (i.e. negatively) related to the proportion of Protestants in countries in Europe and the Americas. Thus, the share of the Protestant population might play an essential role in reducing corruption. This result is consistent throughout the regressions and the regression tree and confirms previous empirical studies (e.g. La Porta et al., 1997; Treisman, 2000; and Chase, 2010). The influence of Protestantism is further explored using the QCA method, which enables one to consider other variables also playing an essential role besides population share.

15.3.2.3.2 Proportion of Roman Catholics

Much of the previous empirical literature report that Roman Catholicism, Orthodox, and Islam religions correlate negatively with transparency (high corruption or low *CPI*) (Paldam, 2001; La Porta et al., 1997; Chase, 2010). However, the negative relation between *Roman Catholics adherents* in the population and corruption (*CPI*) is not conclusive in my results. While *Roman Catholics adherents* are present with a negative yet not particularly significant coefficient in the regression model, they are absent in the regression tree. Treisman (2000) reported comparable results when including proxy measurements for democracy and economic development in the model similar to mine. Those factors are, as discussed, interrelated rather than competing explanations.

As stated in the theoretical part (Sects. 3.3 and 8.1), prosperity, democracy, and transparency might all be part of the same phenomenon. When regressing all those variables together, the model might dull the influence of other variables (e.g. *Proportion of Roman Catholics*). QCA further analyses this issue along with different variables.

15.3.2.3.3 Fractionalisation Measurements

No clear pattern exists in ethnic, linguistic, or religious fractionalisation. Only linguistic fractionalisation is significant in the regression analysis, but no fractionalisation measurement appears in the regression tree analysis. Fractionalisation indices express “how diverse” countries are regarding language, religion, and ethnicities. Further, the indices aggregate and “pack” such diversity into one figure (Alesina et al., 2003). Consequently, fractionalisation measurements prevent analysing the interactions of each specific religion, language, or ethnicity units using criterion (prosperity) variables. Such aggregation constitutes the main limitation of fractionalisation measurements. Therefore, the QCA section uses a disaggregated value to reflect the shares of the main ethnicities and languages spoken in a given country or region (Central Intelligence Agency of the United States [CIA], 2016).

15.3.2.3.4 Political Rights and Civil Liberties

As observed in the regression tree analysis, the political rights and civil liberties (*PRCL*) variable is the most significant factor relative to *CPI*. This finding was expected, on account of its theoretical relation with Protestantism (Sect. 8.2.2). Woodberry (2012) and Becker et al. (2016) have provided ample empirical and historical evidence for the essential role of Protestantism (i.e. Protestant missionaries) in the rise and spread of stable democracies around the world.

15.3.2.3.5 Gross Domestic Product (GDP) Per Capita

As explained, *GDP* per head was expected to be a significant predictor variable and is negatively related to corruption in the tree regression method.

15.3.3 General Conclusion: Models of Religious Population (Competitiveness and Corruption)

Generally, the models presented here suggest that Orthodox and Roman Catholic populations correlate negatively with competitiveness in Europe and the Americas. However, their role in corruption appears less evident in the corruption model. In contrast, a Protestant population might play an essential role in reducing corruption. Nevertheless, the role of the Protestant population in competitiveness is less clear. The next component (QCA) expands on the understanding of these variables along with other proxies.

References

- Acemoglu, D., & Robinson, J. A. (2008). Persistence of power, elites, and institutions. *American Economic Review*, 98(1), 267–293.
- Acemoglu, D., & Robinson, J. (2012). *Why nations fail. The origins of power, prosperity and poverty*. Crown Business.
- Alesina, A., Devleeschauwer, A., Easterly, W., Sergio, K., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8(2), 155–194.
- Arruñada, B. (2010). Protestants and catholics: Similar work ethic, Different social ethic. *The Economic Journal*, 120(547), 890–918.
- Barro, R. J., & McCleary, R. M. (2005). Which countries have state religions? *The Quarterly Journal of Economics*, 120(4), 1331–1370.
- Becker, S., Pfaff, S., & Rubin, J. (2016). *Causes and consequences of the protestant reformation*. Warwick Economics Research Paper Series.
- Berman, H. (2003). *Law and revolution II: The impact of the protestant reformations on the western legal tradition*. Harvard University Press, Belknap Press.
- Besch, W. (1999). *Die Rolle Luthers in der deutschen Sprachgeschichte*. Universitätsverlag C. Winter.
- Brown, C., & Lall, U. (2006). Water and economic development: The role of variability and a framework for resilience. *Natural Resources Forum*, 30(4), 306–317.
- Central Intelligence Agency (CIA). (2016). *Religions*. In the world fact book. Retrieved from <https://www.cia.gov/library/publications/resources/the-world-factbook/index.html>
- Chase, G. (2010). Religion's impact on corruption. *Journal of Business and Economics Perspectives*, 5(1), 102–112.
- Diamond, J. (1997). *Guns, germs, and steel. The fates of human societies*. W.W. Norton.
- Fanfani, A. (1936). *Catholicism, protestantism, and capitalism*. IHS Press.
- Garcia Portilla, J. (2019). “Ye shall know them by their fruits”: Prosperity and institutional religion in Europe and the Americas. *Religions*, 10(6), 362. MDPI AG. Retrieved from <https://doi.org/10.3390/rel10060362>

- Greenslade, S. (1963). *The Cambridge history of the Bible: The west from the reformation to the present day*. Cambridge University Press.
- Grier, R. (1997). The effect of religion on economic development: A cross national study of 63 former colonies. *Kyklos*, 50(1), 47–62.
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Sage.
- Inglehart, R., & Baker, W. (2000). Modernization, cultural change, and the persistence of traditional values. *American Sociological Review*, 65(1), 19–51.
- Johnson, T. M., & Zurlo, G. A. (2016). *World Christian database*. Retrieved from: Center for the Study of Global Christianity at Gordon-Conwell Theological Seminary website: www.globalchristianity.org
- La Porta, R., Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (1997). Trust in large organisations. *American Economic Review*, 137(2), 333–338.
- La Porta, R., Lopez-de-Silanes, F., & Vishny, A. S. (1999). The quality of government. *Journal of Law, Economics and Organization*, 15(1), 222–279.
- Lambsdorff, J. G. (2006). Causes and consequences of corruption: What do we know from a cross-section of countries? In S. Rose-Ackerman (Ed.), *International handbook on the economics of corruption* (pp. 3–51). Edward Elgar.
- Martin, D. (1999). The evangelical upsurge and its political implications. In P. Berger (Eds.), *The desecularization of the world* (pp. 37–49). Ethics and Public Policy Center and Wm. B. Eerdmans.
- McCleary, R. M. (2013). *Protestantism and Human Capital in Guatemala and the Republic of Korea*. Asian Development Bank.
- Michalopoulos, S., & Papaioannou, E. (2017). *The long economic and political shadow of history—Volume 3: Europe and the Americas*. CEPR Press Centre for Economic Policy Research.
- Morgan, J. (2014). *Classification and regression tree analysis*. Department of Health Policy and Management, Boston University School of Public Health.
- Paldam, M. (2001). Corruption and religion. Adding to the economic model. *Kyklos*, 54(2-3), 383–414.
- Rose-Ackerman, S. (2006). *International handbook on the economics of corruption*. Edward Elgar.
- Sachs, J. (2001). *Tropical underdevelopment*. National Bureau of Economic Research. Working Paper, 8119 (NBER), 1–37.
- Schäfer, H. (1997). Oh Lord of heaven, give us power on earth! Fundamentalism and charisma: The reconquest of the field of action in Latin America. *Mesoamerica*, 18(33), 125–146.
- Treisman, D. (2000). The causes of corruption: A cross-national study. *Journal of Public Economics*, 76(3), 399–457.
- University of Columbia. (2017). *Variable selection*. Department of Statistics. Retrieved from <http://www.stat.columbia.edu/~martin/W2024/R10.pdf>
- Witte, J. (2002). *Law and protestantism: The legal teachings of the Lutheran reformation*. Cambridge: Cambridge University Press.
- Woodberry, R. (2012). The missionary roots of liberal democracy. *American Political Science Review (APSR)*, 106(2), 244–274.
- World Economic Forum. (2016). *The global competitiveness report 2014–2015* (Full Data Edition). Geneva: The World Economic Forum.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.