Spring
2024

Advanced Econometrics

Final Project

Roei Avraham

Tal Peer

Technion - Israel Institute of Technology

Table of Contents

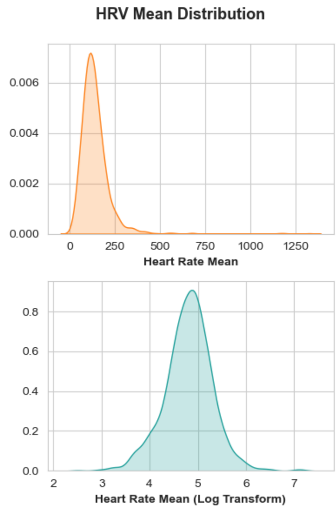
Task - Panel Data

Longitudinal Data - The Effect of Medication on HRV as an Indicator of Concentration in Children Diagnosed with ADHD

We utilized movement and heart rate data collected in a controlled experiment using a smartwatch worn by children diagnosed with ADHD.

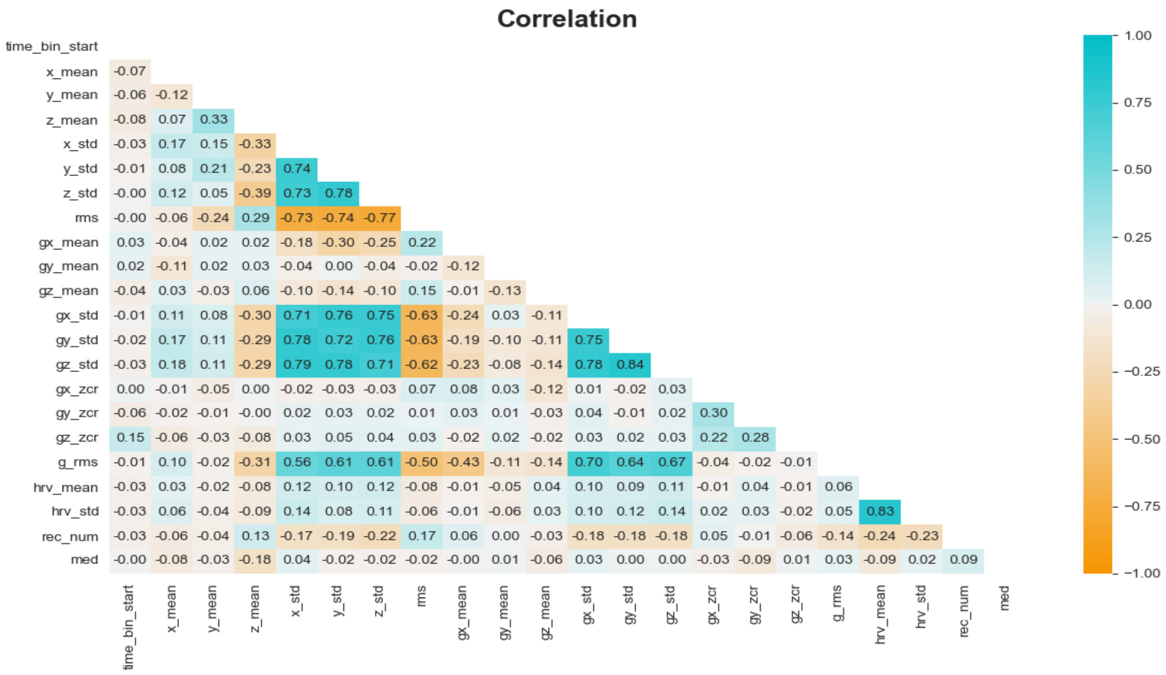
The data was gathered from six children. For each child, we have data collected over a period when the child was not on medication (Time Period 0) and data when the child was under the influence of medication (Time Period 1). For two of the children, we also have data collected under a high dosage of medication (Time Period 2). The following analysis excludes Time Period 2.

In the preliminary work, we aggregated the records into 30-second time windows and filtered out time windows containing fewer than 10 records.
Initially, we refined this filtering by retaining only time windows composed of 25-32 records. This adjustment follows the reasoning of *Agniel et al. (1),* which suggests that while it's important to filter out windows with too few records, the number of records collected within a certain time frame is a parameter that shouldn't be entirely dismissed for devices of this type. Thus, we added the number of records within each time window as a variable.



Next, we examined the distribution of our dependent variable, $\boldsymbol{HRV}$ (Heart Rate Variability), and found that it did not follow a normal distribution (A). However, after applying a log transformation, the variable approximated a normal distribution (B).
Therefore, the dependent variable used in the analysis is the log-transformed HRV.

We assessed the correlations among the independent variables (C) and removed those with high correlations. Based on prior knowledge, we retained the $std(y)$ variable, which represents the standard deviation of the y-axis in gyroscope measurements within each time window (D).



To identify the best Random Effects (RE) and Fixed Effects (FE) models, we coded a function for each model. This function finds the model containing only statistically significant variables with the lowest Mean Squared Error (MSE), enforcing the inclusion of the hrv_std and rec_num variables. The function takes as input the remaining variables and various transformations. This process is documented in the code.

We ran the functions, identified the best **RE** and **FE** models, and conducted the Hausman test (E).
The R^2 values ranged between 0.1-0.13, and this result did not change when we tried to swap variables or transformations.

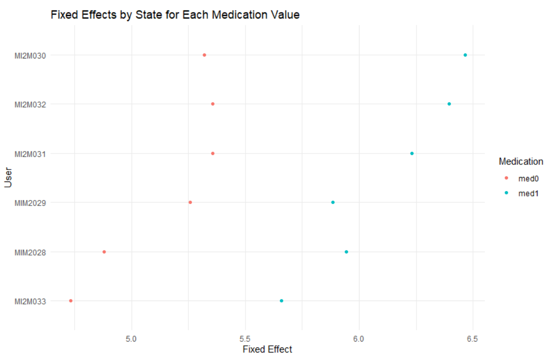
When considering the data under high-dose medication, we achieved R^2 values of 0.15-0.2, which can be attributed to the complexity of the chosen dataset.

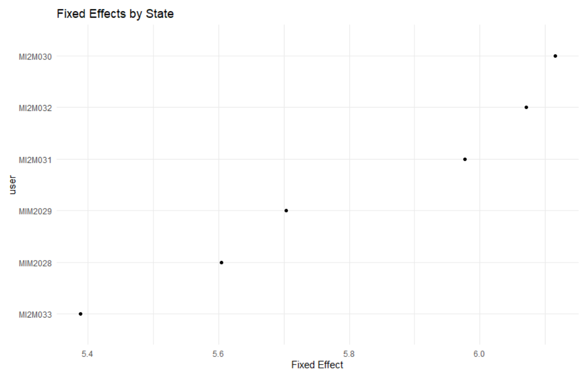
It is worth noting that we also attempted the First Differencing (FD) model, but the results were poor, indicating that the FD transformation might distort our data.

The p_{value} of the Hausman test was very low, suggesting the rejection of the null hypothesis, which leads us to conclude that the **FE** model is better suited to our data.

```
        Hausman Test

data:  formula_best_2501
chisq = 12.482, df = 4, p-value = 0.01411
alternative hypothesis: one model is inconsistent
```

We plotted the Fixed Effects graphs for each child, showing differences in HRV values (F).
However, we are more interested in these graphs under the influence of medication, so we extracted them by splitting the data (G).

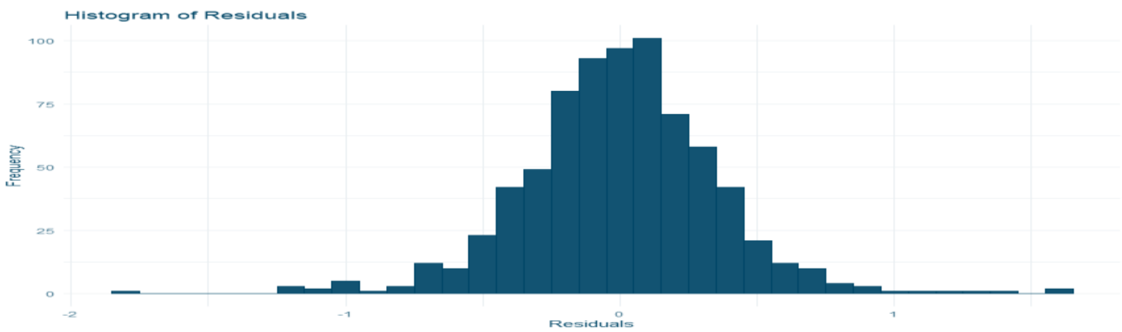




A result that we couldn't explain was the discrepancy between the medication's coefficient in the FE model, which is negative (-0.1894) and indicates 19% (approx.) decrease in HRV under medication, and Graph G, which shows higher log-transformed HRV values for all of the children.

It is worth noting that in the literature (2), attributes such as *prolonged focus* and *calmness* are associated with higher HRV, aligning with the findings in Graph G.

Finally, we checked the distribution of the residuals and confirmed they are normal (H).



Task - Quantiles Regression

Introduction – The World Values Survey (WVS) Dataset

The World Values Survey (WVS) is an international research program devoted to the scientific and academic study of social, political, economic, religious, and cultural values of people in the world. The project's goal is to assess which impact values stability or change over time has on the social, political, and economic development of countries and societies. The main research instrument of the project is a representative comparative social survey which is conducted globally every 5 years.

Attribute	Description	Type
B_COUNTRY_ALPHA	Participant's country (categorial)	Nominal
AGE	Participant's age	Discrete
EDUCATION	Participant's education level	Ordinal
INCOME_LEVEL	Participant's income level	Ordinal
DEFIANCE	Average of Inverses values for respect for authority, national pride, and devoutness	Ratio
DISBELIEF	Average of Inverses indices for Religiousness	Ratio
RELATIVISM	Average of several inverse indices for Participant's **Conformity**	Ratio
SCEPTICISM	Average of the Inverse trust in – Police, Army and Courts	Ratio
SACSECVAL	Weighted average of – DISBELIEF, RELATIVISM and SCEPTICISM	Interval
AUTONOMY	Average of Independence as kid quality, Imagination as kid quality, Obedience	Ratio
EQUALITY	Average of Gender Equalities importance -Jobs, Politics, Education	Ratio
CHOICE	Average of Acceptance – Homosexuality, Abortions, Divorce	Ratio
VOICE	Average of several indices for Participant's Emancipative	Interval
RESEMAVAL	Weighted Average of – AUTONOMY , EQUALITY , CHOICE, VOICE	Interval
ReligCountry	Dummy Variable (top 10 & top bottom countries with participants having RESEMEVAL value lower than the median (~0.413))	Binary

Data Preprocessing

Initially, all observations containing missing values (NA) were removed. Then, we check for not applicable records in the qualtiative data types (for example, remove values lower then 0 for Education which is an ordinal data type with range of integers in [1-9]). A subsequent filtering step was performed to include only the top ten most religious and the top ten most secular countries, as measured by the average *DISBELIEF* score. This score is a composite measure that captures the importance an individual places on religion, as well as their participation in religious practices, among other factors. The binary variable *ReligCountry* was created to distinguish between individuals from religious and secular countries. After taking the 10 most religious countries and 10 most secular countries, we combined them into one dataset, referred as WVS_20.

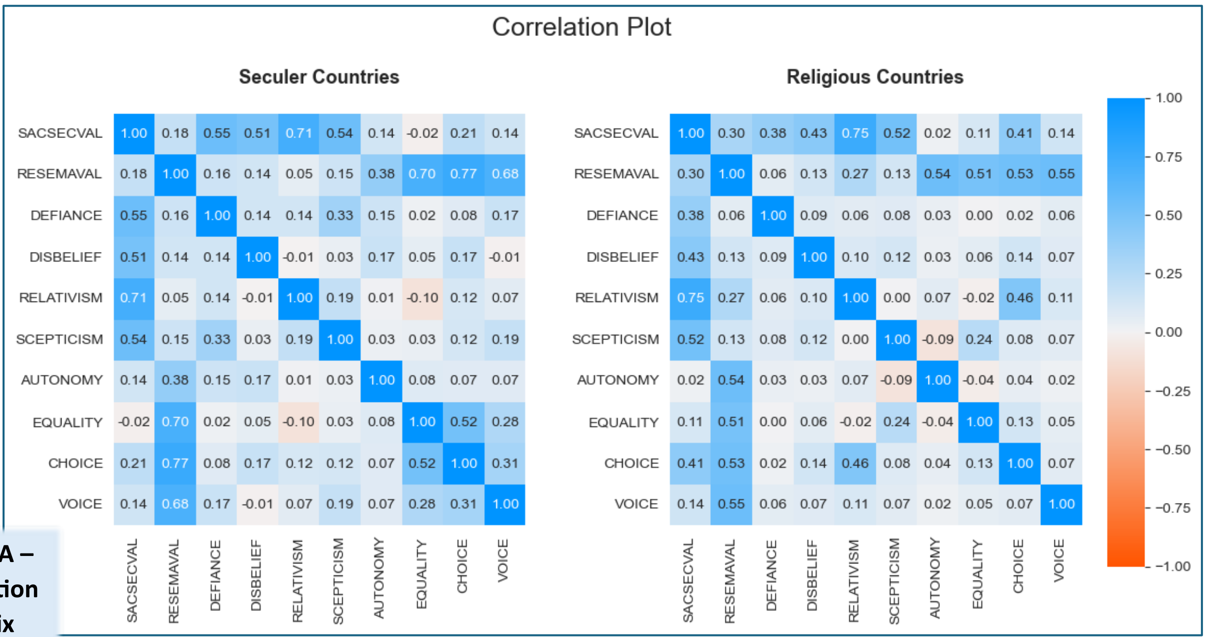


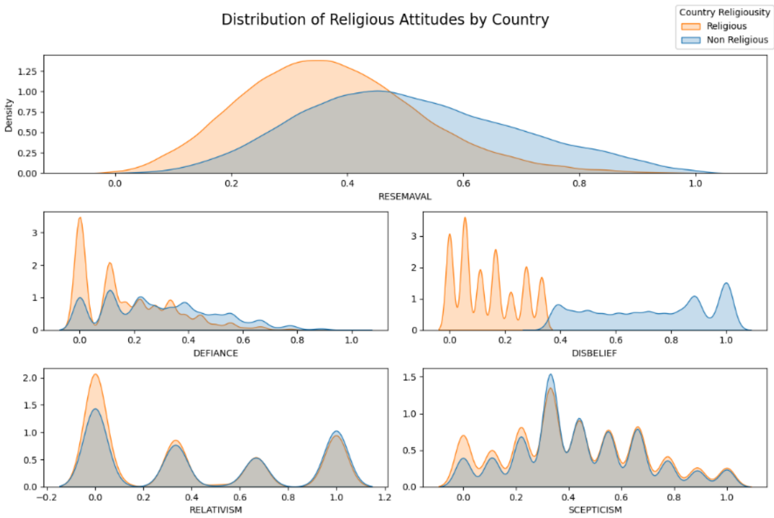
Figure A – correlation matrix

Normality Check

To derive OLS estimates, we need to relize our main assumption holds –

$$E[e_i|x_i] = E[e_i] = 0$$
$$\rightarrow cov[x,e] = E[x \cdot e] = 0$$

We conducted a normality check for the dependent variable, *RESEMAVAL*, and observed a close approximation to a standard normal distribution. This variable serves as the collectivism indicator, and its normality is important for linear modeling assumptions.



Correlation Analysis

Next, we checked for multicollinearity among the independent variables and removed those with high correlations to *RESEMAVAL* in order to avoid issues with multicollinearity which we know could distort the analysis results. (figure A)

Modeling Strategy

To choose the most appropriate formula for the regression models for best-capture the assumed relations, we defined a custom function that enforces the inclusion of some form of *DISBELIEF* in the model—either as the raw indicator value or a polynomial transformation. This function optimally adds other independent variables and/or their transformations while ensuring that all included variables maintain statistical significance. The function was designed to return the model with the lowest AIC and highest Mean Squared Error (MSE), to ensure a good balance between model fit and complexity.

The resulting model[1] –

$$\begin{aligned} RESEMAVAL = \; & DISBELIEF + 10 \cdot (DISBELIEF - 0.5)^2 + RELATIVISM + 10 \cdot (RELATIVISM - 0.5)^2 \\ & + SCEPTICISM + 10 \cdot (SCEPTICISM - 0.5)^2 + AGE + ReligCountry((AGE * EDUCATION) \\ & + DEFIANCE) \end{aligned}$$

Where non-linearity is explored by the **Quadratic Terms**, which take the raw terms and shift them by 0.5, scale them by 10, and then square them, in order to allow the model capture curvilinear effects. This transformation are useful in understanding the non-linear effects of *DISBELIEF*, *RELATIVISM*, and *SKEPTICISM*—where extreme levels of these variables may have stronger or weaker effects on *RESEMAVAL*.

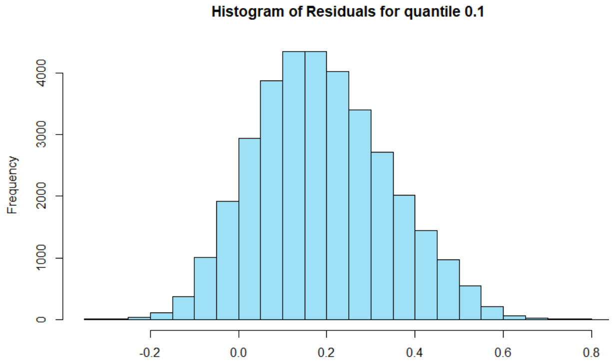
Additionally, we included **Interaction Terms**, specifically the interaction between *AGE* and *EDUCATION* and other attributes with the countries religiosity. The term I(AGE * EDUCATION) suggests that the effect of age on *RESEMAVAL* is contingent upon the respondent's level of education, highlighting a more nuanced relationship between these variables.

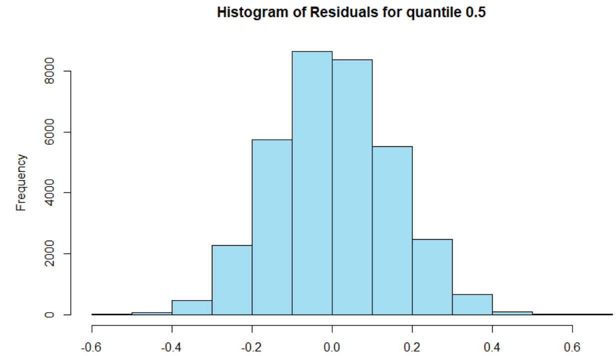
We then validated by checking the normality of its residuals. The residuals were distributed symmetrically around zero, confirming that the model met the assumption of normally distributed errors.

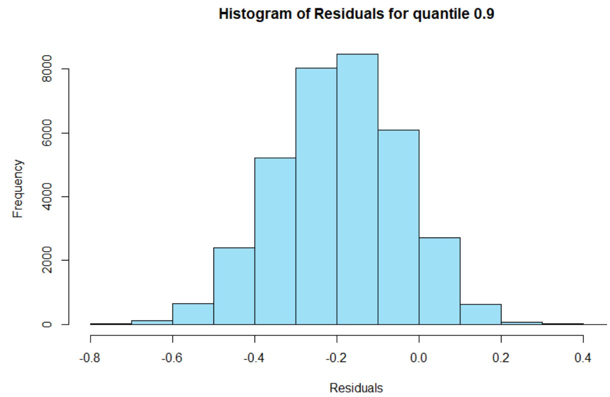
Model Evaluation

[1] Expanded model summary in the appendix section.

For evaluating the performance of the Quantile Regression model, we plotted the residuals and compared them to those of the OLS model. The residuals for the Quantile Regression model showed two distinct clusters, potentially reflecting the dataset's structure, which consists of two distinct groups of countries (religious and secular). A closer examination of the Y-axis values revealed that the dispersion of residuals is indeed tighter around zero compared to the OLS model, which might point on a better fit of the Quantile Regression for the data in the given quantiles.





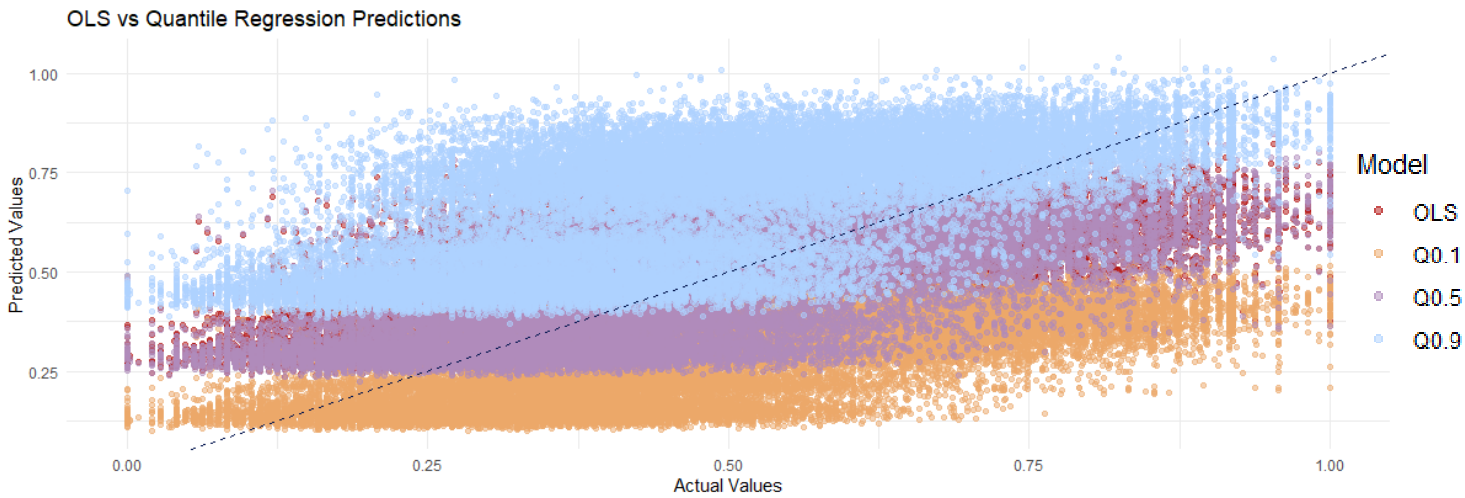


We also compared the predictions from both the OLS and Quantile Regression models to the actual values of *RESEMAVAL*. The analysis revealed that predictions from the Quantile Regression model align more closely with the observed values, especially when compared to the OLS model. This indicates that each model fits best within its respective quantile and may not generalize well across all quantiles.

Significance Testing Conclusions

Significance testing for the independent variables across different quantiles arise different behaviour in different quentiles. For instance, in the upper quantiles for secular countries, there was **no significant relationship between skepticism and respect for authority**. In the median quantile, **three additional variables were found to be insignificant**. In the lower quantile, which represents individuals who exhibit the least respect for authority, the relationships between variables **were much weaker**. This suggests that the OLS model is heavily influenced by the upper quantiles and may require alternative modeling approaches to better capture the relationships in the lower quantiles.

The analysis has shown that while OLS provides a broad overview of the relationships, Quantile Regression model does offers a more detailed perspective by capturing varying effects across different quantiles. The residual analysis and significance testing underscore the limitations of OLS in fully capturing the dynamics within this dataset, particularly in the presence of non-linearity and quantile-specific effects.



Task 3 - Censored and Truncated Data

In this section, we use the same dataset of WSV from the Quantile Regression Task and analyze the relationship between the same regressors (in particular, the same regression formula) and the dependent variable, RESEMAVAL.

We compare between the performance of Ordinary Least Squares (OLS) regression and Tobit regression, where the OLS model serves as our benchmark for comparison, while the Tobit model addresses the left and right censoring present in our dataset.

Reminder – we defined the "Religious" records based on the DISBELIEF attribute, classifying individuals as "Religious" if their DISBELIEF value is higher than the median of the attribute across the entire dataset. This allowed us to calculate the percentages of "Religious" and "Secular" individuals within each country, resulting in a refined dataset consisting of the top 10 most religious countries and the top 10 most secular countries, referred to as WVS_20.

The analysis identified specific quantile values for ***RESEMAVAL*** in both groups, with notable differences indicating that the secular group had higher lower bounds for the variable.

We experimented with different censoring thresholds for the RESEMAVAL variable, specifically:

Censoring Threshold	**RESEMAVAL τ_L 10 Most Religious**	**RESEMAVAL τ_L 10 Most Secular**	**RESEMAVAL τ_L Whole dataset**
$\tau_L = 0.16$	$P(y_{res} \leq 0.16) \cong 0.1$	$P(y_{res} \leq 0.16) < 0.005$	$P(y_{res} \leq 0.16) \cong 0.05$
$\tau_L = 0.32$	$P(y_{res} \leq 0.32) \cong 0.5$	$P(y_{res} \leq 0.32) \cong 0.1$	$P(y_{res} \leq 0.32) \cong 0.3$
$\tau_R = 0.7$	$P(y_{res} \geq 0.7) \cong 0.05$	$P(y_{res} \geq 0.7) \cong 0.75$	$P(y_{res} \geq 0.7) \leq 0.08$
$\tau_R = 0.8$	$P(y_{res} \geq 0.8) \leq 0.001$	$P(y_{res} \geq 0.8) \cong 0.1$	$P(y_{res} \geq 0.8) \cong 0.01$

For example, the table suggests that left-censoring on **0.16** means **10%** of the values in the **10 most religious** countries will be censored, and less than 0.005 for the 10 most secular countries. Considering the whole dataset, approximately **0.05** of the records have RESEMAVAL value of **at most 0.16.**

In order to observe the differences between the models' goodness of fit, especially the biased estimators of OLS with censored data, we will adhere to Left censoring for $\boldsymbol{\tau_L = 0.32}$, but the conclusions for the comparisons between the models remain the same for the other thresholds[2].

Regarding the context of censored data, we know that OLS can produce biased estimates. For instance, if values of ***RESEMAVAL*** below a certain threshold (e.g., 0.32) are treated as valid observations, the censored y^* is not representative of the population and the OLS estimates using censored data leads to inconsistent estimating $\frac{\partial E[y^*|x]}{\partial x} \neq \beta$.

On the other hand, the Tobit model is designed to handle censored data making the estimation of the conditional means consistent (under strong distribution assumptions), obtaining the censored mean by first conditioning the observable y on the binary indicator d and then unconditioning.

Results

While the direction and significance of key predictors (e.g., ***DISBELIEF***, ***RELATIVISM***, ***DEFIANCE***) are similar, the Tobit model tends to have slightly larger coefficient estimates, reflecting the adjustment for censoring. Also, a difference was detected for the interaction terms, especially the interaction term of ***DISBELIEF*** and

[2] Plotting and results for the other thresholds are given in the appendix section

secular countries is stronger in the Tobit model, which captures the reduction in disbelief's effect in secular contexts more prominently.

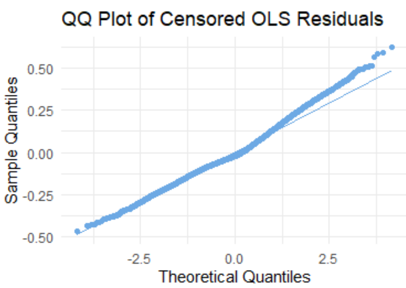
The coefficients in the Tobit model represent the expected change in the latent variable for a unit change in the independent variable. In contrast, the OLS coefficients reflect changes in the observed variable, which can lead to misleading interpretations when the data is censored. The observed significance of the interaction terms in OLS, which are nonlinear in the regressors, and their lack of significance in the Tobit model, highlights that OLS is less accurate in capturing relationships in the censored data.

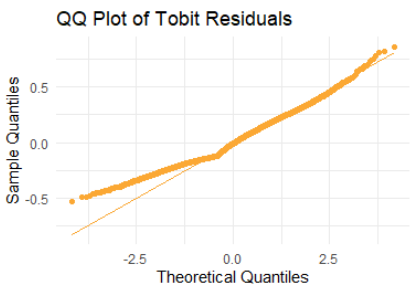
Few differences in the summaries of both models (the whole summary can be found in the appendix section):

Model Term	OLS	Tobit	Interpretation
Intercept	0.2137 (p~0)	0.1063 (p<0.001)	The OLS intercept estimate is higher, likely because it does not account for the right-censoring, thus overestimating the base level of *RESAMAVAL*.
SCEPTICISM	0.0037 (p>0.1)	0.0183 (p<0.01)	OLS fails to detect a significant relationship for *SCEPTICISM*, but the Tobit model finds a positive and significant effect.
AGE	-0.000128 (p ~ 0.07)	-0.0002697 (p~0.006)	OLS underestimates the effect of *AGE* compared to Tobit. The censoring means that the OLS model dilutes the negative effect of *AGE*.
Interaction Terms	The model detects significant relations for the interaction terms besides DEFIANCE and ReligCountrySecular, which the model detects no significancy at all.	The model detects no significancy for SCEPTICISM and ReligCountrySecular, but detects a significance relationship (p~0.01) for DEFIANCE and ReligCountrySecular.	OLS underestimates the interactions between belief-related variables and whether the country is secular or religious. Tobit estimates a stronger negative interaction effect.

As the table suggests, the OLS model fitted with data with left-censored on 0.32 is biased downward for most variables, particularly for ***RELATIVISM, SCEPTICISM, AGE,*** and the interaction terms involving ReligCountrySecular. This bias is reasonable since we know OLS fails to account for the left-censoring for the lower limit values, and resulting in systematically higher estimates compared to Tobit.

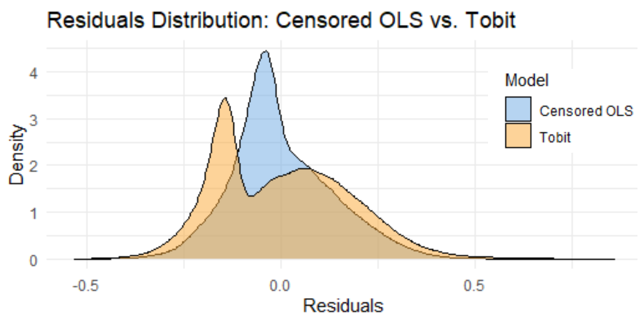
For the OLS model, the Q-Q plot of the residuals captures normality, but it shows a deviation at higher values. This deviation arises because OLS assumes normally distributed and homoscedastic residuals, but its performance diverges for higher values. In contrast, the Tobit residuals show deviations from normality for the censored values, while





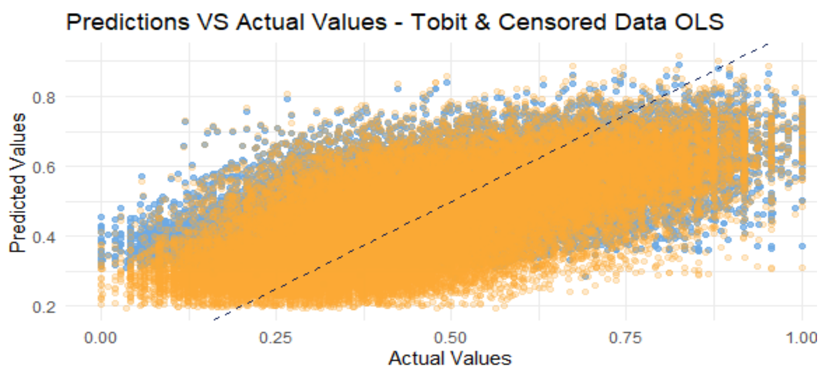
maintaining normality for the uncensored values. Due to the censoring, the observed residuals will deviate from normality, particularly near the censoring thresholds, reflecting the censored nature of the data.

The distribution analysis in the table above highlights that for each group of secular/religious countries there are different values ξ_q. This indicates that censoring the dataset will censor different number of records for each group, for some threshold sometimes a significant difference. This emphasizes the need for careful sample selection in order to avoid misrepresentations in model fitting . These results are correlated with the objectives of Tobit regression, particularly for addressing sample selection issues.

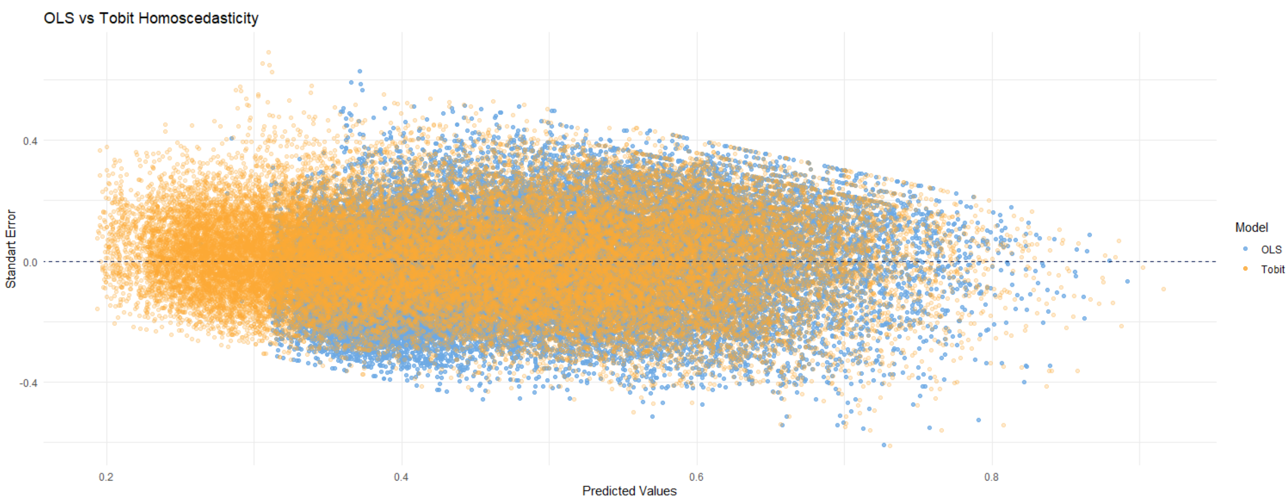


The censored OLS model assumes that the dependent variable follows a continuous and normal distribution, leading to residuals that are symmetrically distributed around zero.

However, the Tobit model creates “two peaks” in the residual’s distribution. This occurs due to discontinuities introduced by the censoring points. This is aligned with the formulation of the Tobit model as shown in class, as the density for left-censored maximum likelihood estimation (MLE) is a hybrid of the PDF and CDF of $y*$ at the censoring thresholds.



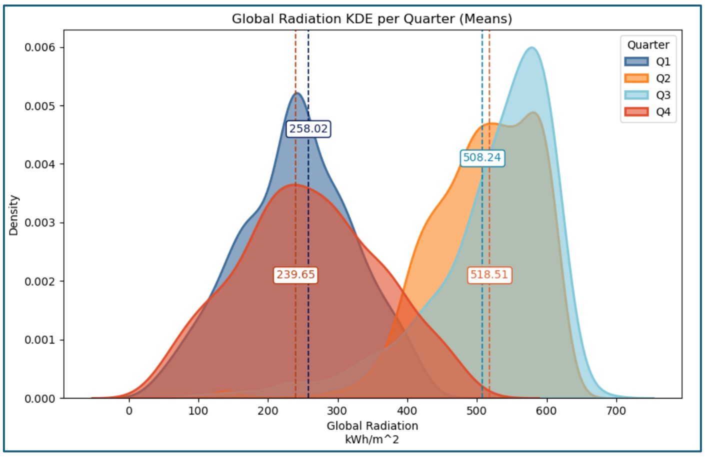
we can observe the OLS model didn’t predict any values under the 0.32 threshold, but the Tobit regression model does. It is also influencing the homoscedasticity of residuals in the Tobit model, which are much more concentrated around zero and the mass area is much more centered around the 0-line comparing to the OLS model.



Task 4 - Time Series Data

We submitted a Jupiter Notebook for this task section. The next report is a concise one. Link to notebook

We will focus on the analysis of Global Radiation measurements between 2017-2023.

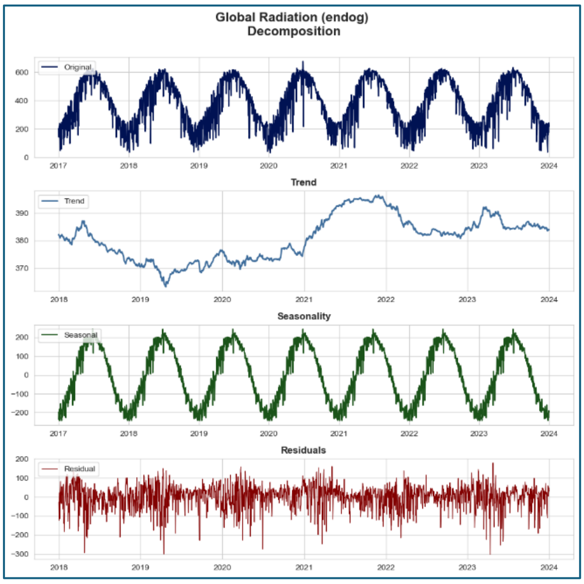


Global radiation refers to the total solar radiation received on a horizontal surface, encompassing both direct and diffuse solar radiation. This radiation is a key component of Earth's energy balance and plays a crucial role in various natural processes.

We selected a dataset spanning hourly measurements of global radiation recorded at the Haifa Technion station, covering the years from 2017 to 2023.

Originally, the dataset provided measurements for each hour from 8:00 to 20:00. However, we opted to aggregate the data to compute daily averages, along with additional temporal information.

The dataset was sourced from the Israeli Meteorological Service website, under the governance of the Ministry of Transport and Road Safety. It comprises 2546 records and the average daily global radiation measured on each date (series index).



Decomposing the dataset, we can observe that seasonality manifests as fluctuations in global radiation levels throughout the year, with higher levels during certain seasons (e.g., Q3) and lower levels during others (e.g., Q1 and Q4). We can observe an overall trend by increasing radiation levels over the mentioned years. Also, we can observe from the boxplot (appendix) and residuals plot a discernible regularities or structures by short-term fluctuations outliers in Q2.

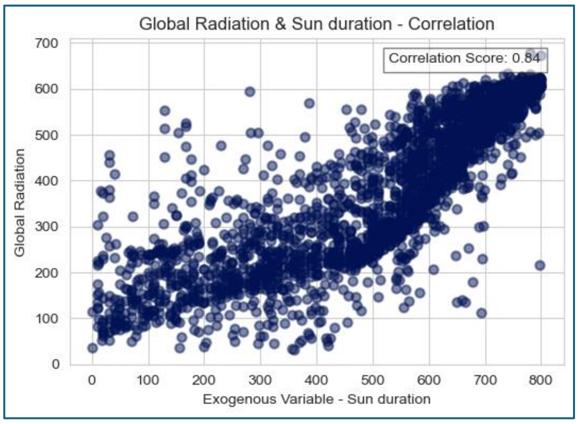
While initially observing the data, we notice significant yearly seasonality. This observation is not surprising as radiation depends on the sun which revolves around the earth based on a roughly estimated 365-days period. Thus, the first model we considered was the SARIMA model, but the results of the fit weren't very informative. Therefore, in this task, we decided to compare multiple models to gain a better understanding of the differences between them.

Incorporating an Exogenous Variable

Inclusion of exogenous variables enables the model to better capture the complex relationships and dependencies present in the data, resulting in more accurate predictions. Moreover, Exogenous variables that exhibit consistent trends or patterns over time are more likely to align closely with the fluctuations in global radiation. We checked 3 exogenous variables – **Rainfall**, **Temperature**, and **Sun duration**. The measures were available in the meteorological database and were all recorded at the Haifa Technion station, covering the years from 2017 to 2023. Preprocessing steps include converting dates to datetime format, calculating daily averages as aggregation, and handling with missing values missing values for Rain (MAR) and Sun duration (MCAR). A more comprehensive details in the appendix section.

Statistical Analysis and Modelling

To determine the most suitable exogenous variable for inclusion we first calculate the correlation score between each exogenous variable (rainfall, sun duration, and temperature) and the target variable, global radiation. The correlation score helps assess the strength and direction of the relationship between each exogenous variable and global radiation. As demonstrated in the correlation plots in the appendix(1+2), the most correlated variable is Sun duration, with 0.84 correlation score.



Both models require some strong assumptions. we checked for the assumptions for models to exist:

<u>Stationarity of time series</u> - The statistical properties of the series, such as mean and variance, should not change over time.

<u>No multicollinearity</u> - The independent variables, including their lags, should not be highly correlated with each other, as this can lead to unstable coefficient estimates

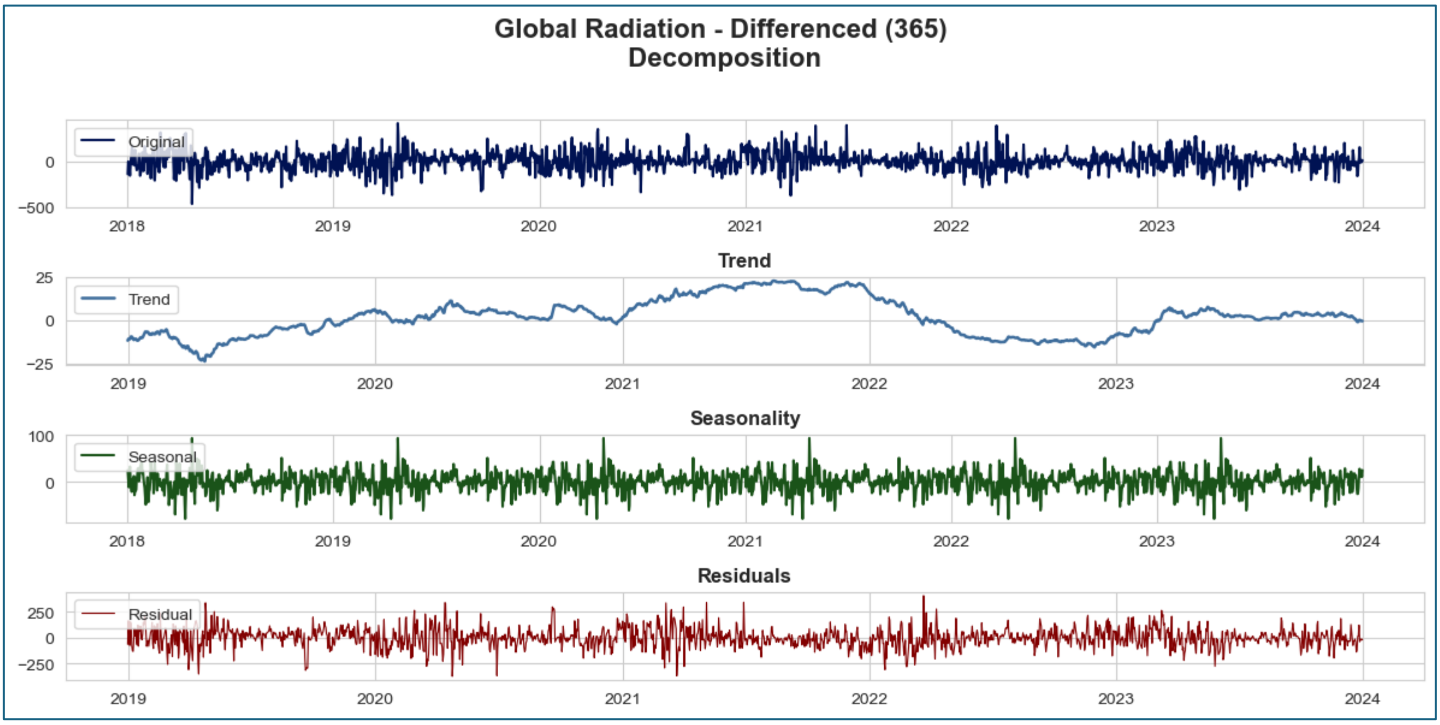
<u>Homoscedasticity</u> - The variance of the error terms should be constant over time.

```
=====================================
Dickey-Fuller test for stationary - Global Radiation (endog)


ADF Statistic: -2.7485778047420575
p-value: 0.06600275063669639
=====================================
```

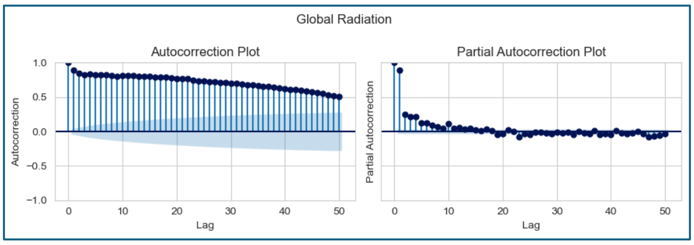
```
ADF Statistic: -10.579944104439367
p-value: 6.947280628653694e-19
=====================================
```

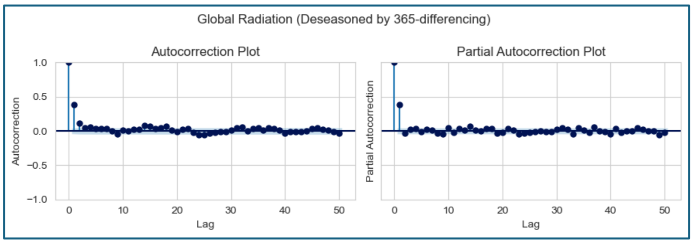
The raw series wasn't stationary, which required us to apply transformation for the series.

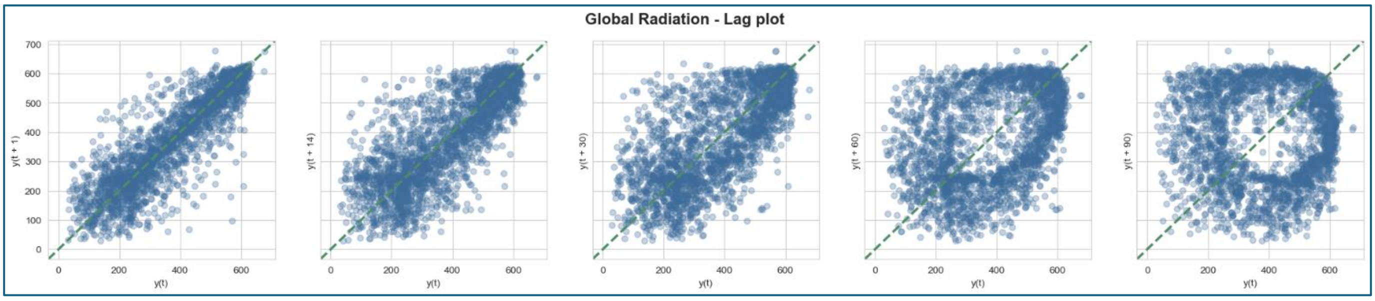
Based on the decomposition plot, ACF plot, and PACF plot, we decided to apply differencing of 365 days. This is a reasonable heuristic due to the seasonality of the variables.



We performed a grid search on model parameters for ARMA(p,q) and ADL(lags) by testing different parameters. A reasonable starting point for the parameters was chosen based on the ACF plots and lag plot of both the endogenous and exogenous variables. Since the model complexity was high, the AIC and BIC scores were relatively high. Therefore, we decided to choose the best model based on the Mean Absolute Error (MAE) score. We also calculated the RMSE scores for the model, but due to fluctuations in the data, RMSE might place more emphasis on shifts in the data. In general, all models exhibited similar values for these scores, and we did not observe large differences between different model orders (ARMA) or lags (ADL).

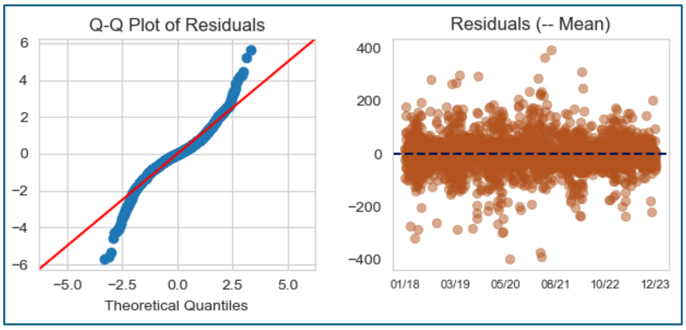


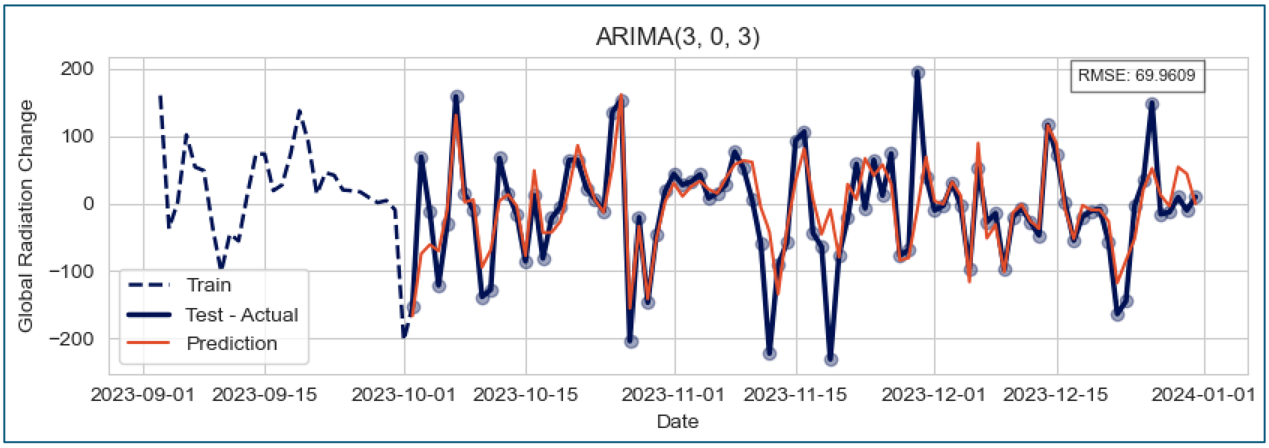




Autoregressive Moving Average - ARMA

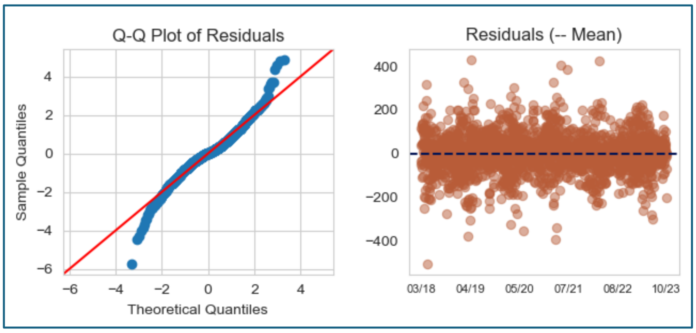
Overall, ARMA predictions provided a good fit with the differenced series. We tried grid search by testing $p, q \in \{0,1,2,3\}$. Each of the params performed quite a good fit. The QQ-plot of residuals appears to be close to normality, and the residuals exhibit homoscedasticity, though with high variance. The variance of the error terms remains constant over time, ensuring that the model assumptions hold.

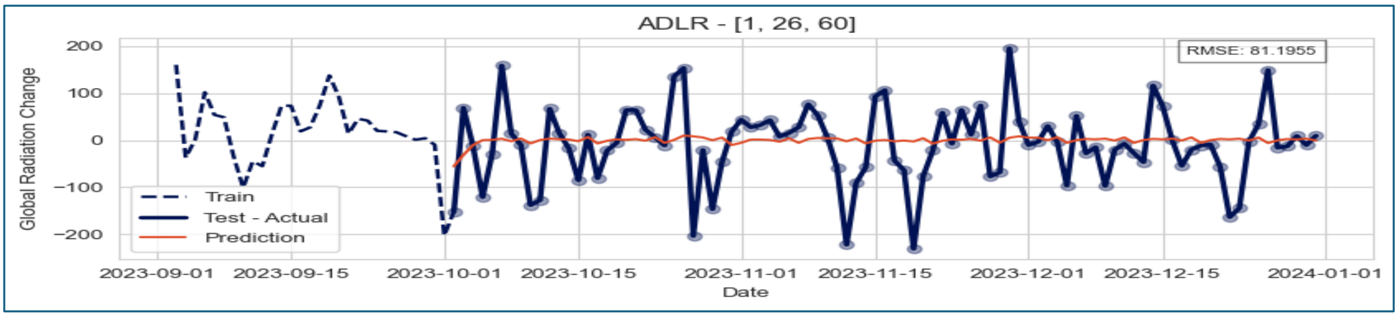




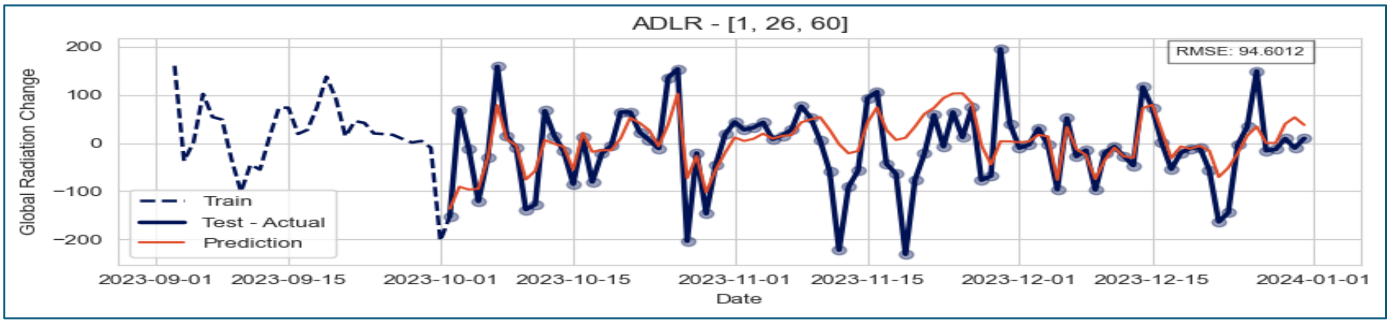
Autoregressive Distributed Lag (ADL)

In the ADL model, we observed some unusual results when testing the model without an exogenous variable. The model fit was not accurate, and despite trying different lags, all model predictions were flat. However, the QQ-plot for residuals and the homoscedasticity check indicated that the model assumptions does align.





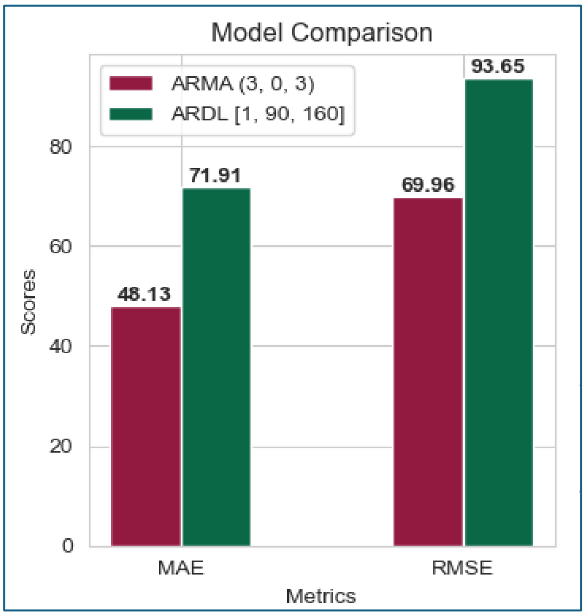
Only when we incorporated the Sun duration variable as exogenous variable, the ADL model performance showed much better fit to the data.



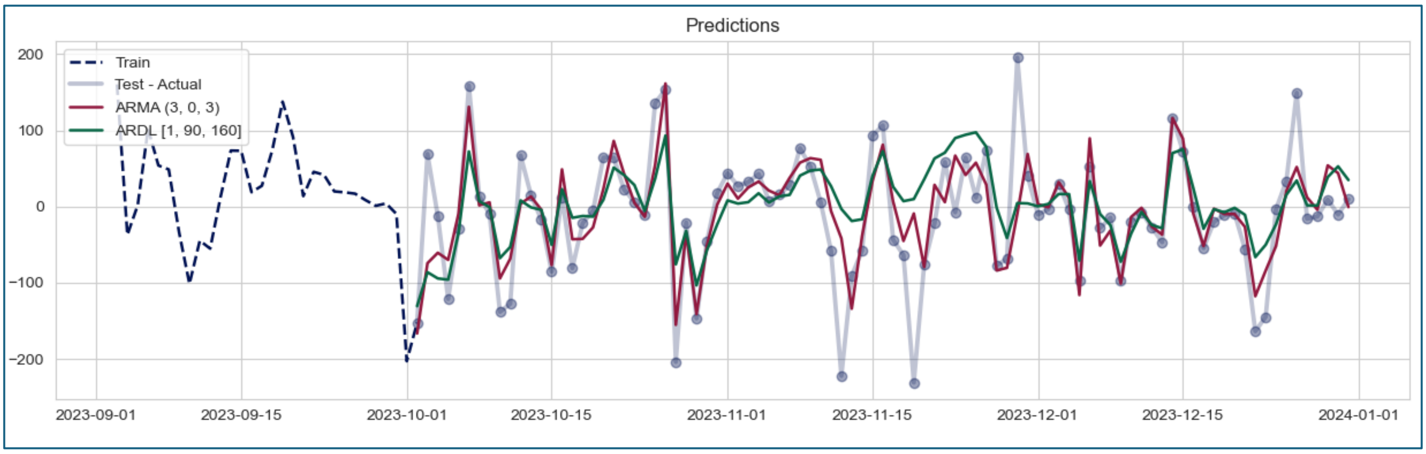
Models Comparison

We can observe that ARDL has both MAE & RMSE lower then ARMA, but ARMA seems to have a better capture of the fluctuations.

This is reasonable due to different model structure. ARDL models handle long-term relationships between the variables and lags of both the dependent and independent variables. ARDL might be better at forecasting the average trend or levels of the series but could smooth out short-term fluctuations due to its focus on both explanatory variables and lagged terms. Also, as discussed on class, ADRL is a good solution for time series that do not perform well in cases of instant changes.



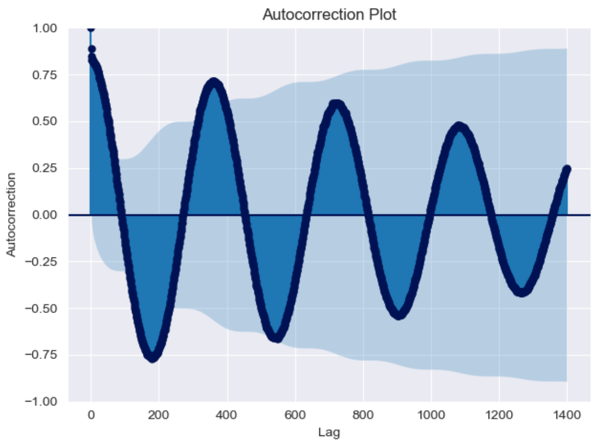
On the other hand, models rely solely on the lags of the dependent variable (AR component) and past errors (MA component). While ARMA might produce higher MAE/RMSE, it can capture short-term dependencies and fluctuations more effectively. The MA part of the ARMA model adjusts based on recent shocks in the data, which can help in fitting rapid changes or short-term volatility better than ARDL.



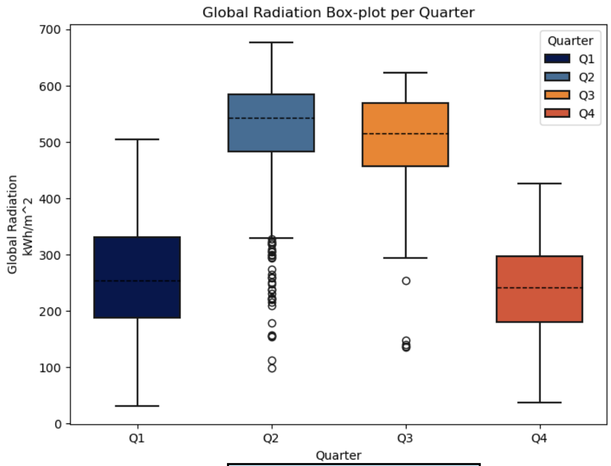
References

1. Israel Meteorological Service databases
2. Detailed guidebook to the user of IMS data (Hebrew)
3. About the radiation database (Hebrew)
4. Israel Meteorological Service Database API
5. Linear Interpolation & Spline linear interpolation for missing data
6. Spline linear interpolation
7. Mean squared error regression loss - Skicit-Learn Documentation
8. Mean Absolute Error regression loss – Skicit-Learn Documentation
9. Denis Agniel and research fellow. (2018). Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. BMJ, 361
10. Roderik J. S. Gerritsen and Guido P. H. Band (2018). Breath of Life: The Respiratory Vagal Stimulation Model of Contemplative Activity

Appendix – Time Series

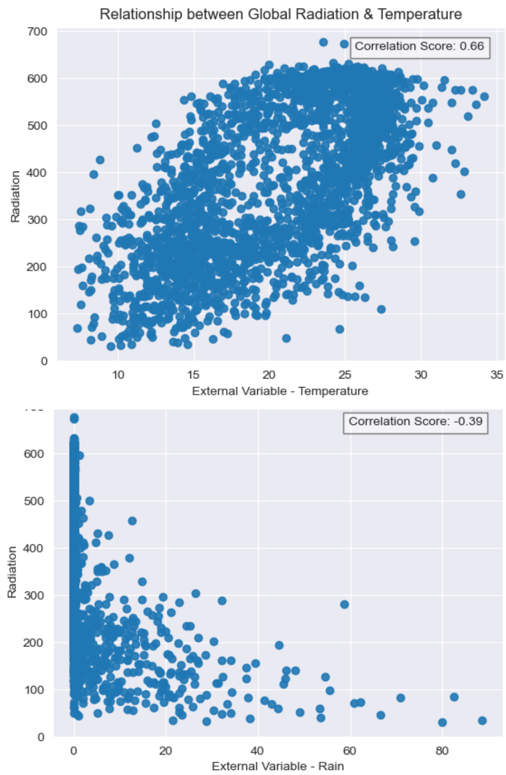


Global radiation Autocorrelation plot (lag = 1400)



Global Radiation box plot by Quarters

Correlations in details



Sun duration represents the amount of time during the day that the sun is visible, which directly impacts the amount of solar radiation reaching the Earth's surface. The duration of the sun's caution refers to the measurement of the flux of direct radiation above a certain threshold Reflects the duration of time, between sunrise and sunset, in which cloudiness, if present, does not interfere with the passage of direct radiation. Since there is more time for solar radiation to penetrate the atmosphere and reach the ground as the sun duration is longer, it is reasonable to check if the variable correlates with higher global radiation levels.

Preprocessing

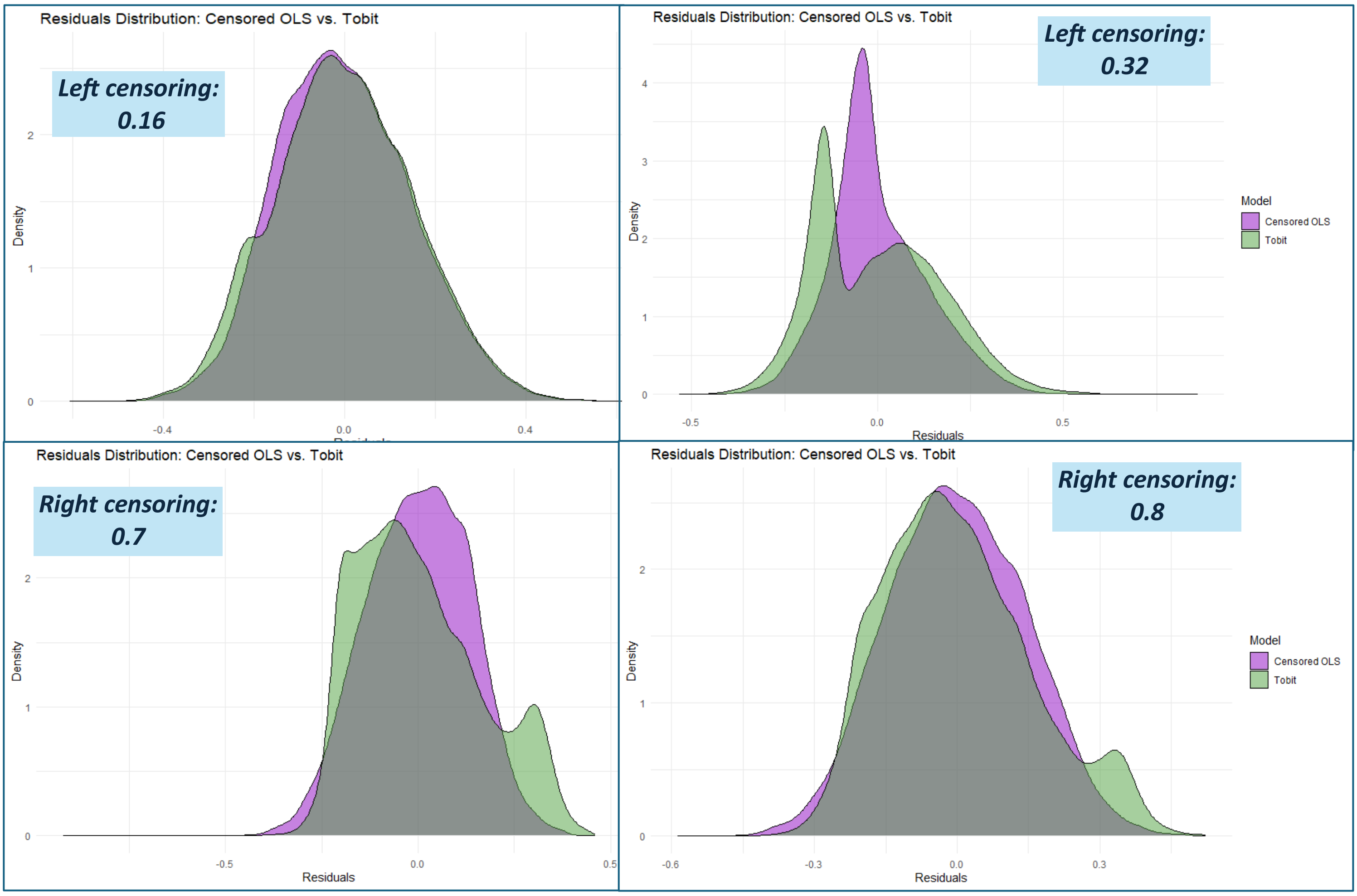
Sun duration – We couldn't figure why specific dates were missing from the original DataFrame, so we assumed missingness of type **MCAR**. Since data wasn't very sufficient with size, we figure that imputation methods might increase the bias. We found a simple yet robust way to deal with the missingness – Linear spline interpolation. Linear interpolation is a suitable choice for completing missing data in time series datasets like sun duration. Its simplicity, based on the assumption of a linear relationship between consecutive data points, makes it computationally efficient and easy to implement. By preserving the overall trend and smoothness of the data, linear interpolation provides conservative estimates without introducing abrupt changes, ensuring that the interpolated values align closely with the underlying pattern of the time series. We chose to use Linear Spline interpolation due to its ability to capture more complex relationships between data points. Unlike linear interpolation, which assumes a straight line between consecutive points, spline interpolation fits piecewise linear segments between neighbouring data points. This flexibility allows spline interpolation to better accommodate nonlinear trends and fluctuations in the data.

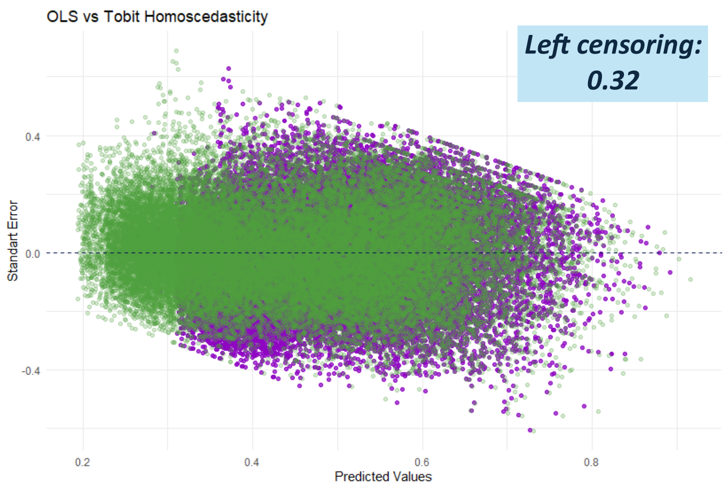
Appendix – Quantile Regression

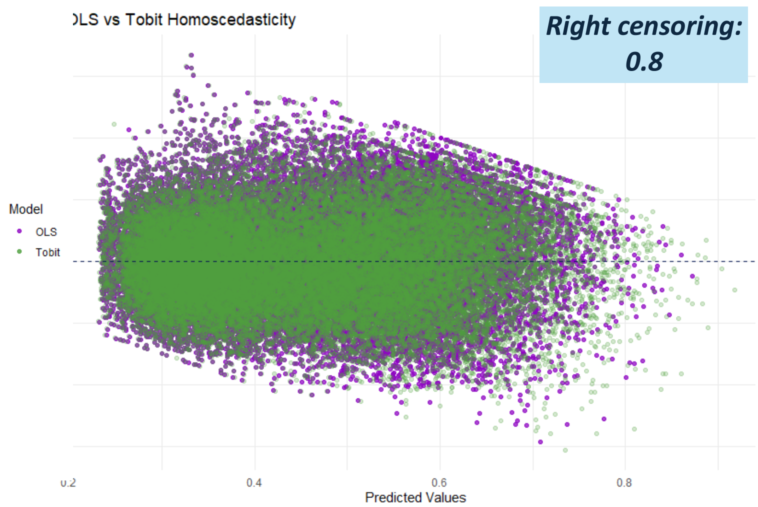
The World Values Survey (WVS) Dataset – Data Scheme

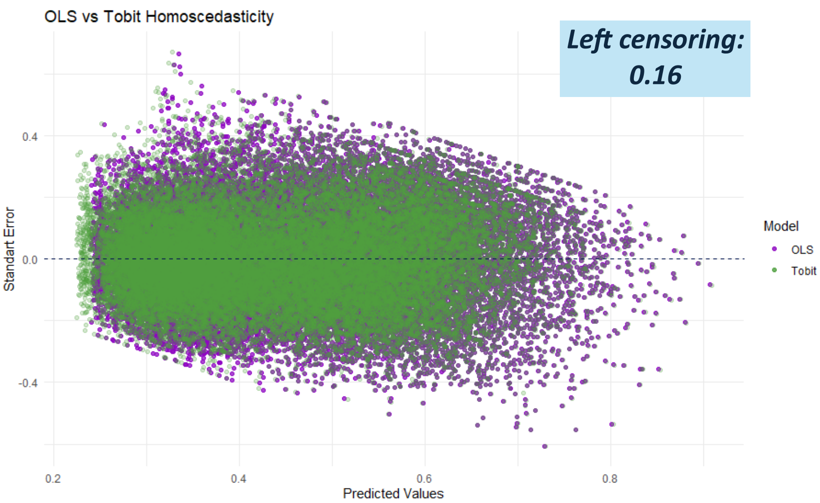
B_COUNTRY_ALPHA	Participant's country (categorial)	Qualitative - Nominal
AGE	Participant's age (discrete numeric)	Quantitative – Discrete
EDUCATION	Participant's education level	Qualitative - Ordinal
INCOME_LEVEL	Participant's income level	Qualitative - Ordinal
DEFIANCE	Average values for – - Inverse respect for authority - Inverse national pride, - Inverse devoutness	Quantitative – Ratio
DISBELIEF	Average of – - Inverse importance of Religiousness - Inverse acknowledgement as religious - Inverse Religiousness practice	Quantitative – Ratio
RELATIVISM	Average of several inverse indices for Participant's **Conformity**	Quantitative – Ratio
SCEPTICISM	Average of the Inverse trust in – Police, Army and Courts	Quantitative – Ratio
SACSECVAL	Weighted average of – - DISBELIEF - RELATIVISM - SCEPTICISM	Quantitative - Interval
AUTONOMY	Average of – Independence as kid quality, Imagination as kid quality, Obedience not kid	Quantitative – Ratio
EQUALITY	Average of Gender Equalities importance -Jobs, Politics, Education	Quantitative – Ratio
CHOICE	Average of Acceptance – Homosexuality, Abortions, Divorce	Quantitative – Ratio
VOICE	Average of several indices for Participant's Emancipative (Autonomy, Equality, Choice)	Quantitative – Interval
RESEMAVAL	Weighted Average of – - AUTONOMY - EQUALITY - CHOICE - VOICE	Quantitative – Interval
ReligCountry	Dummy Variable (countries with participants having RESEMEVAL value lower than the median (~0.41)	Quantitative – Interval

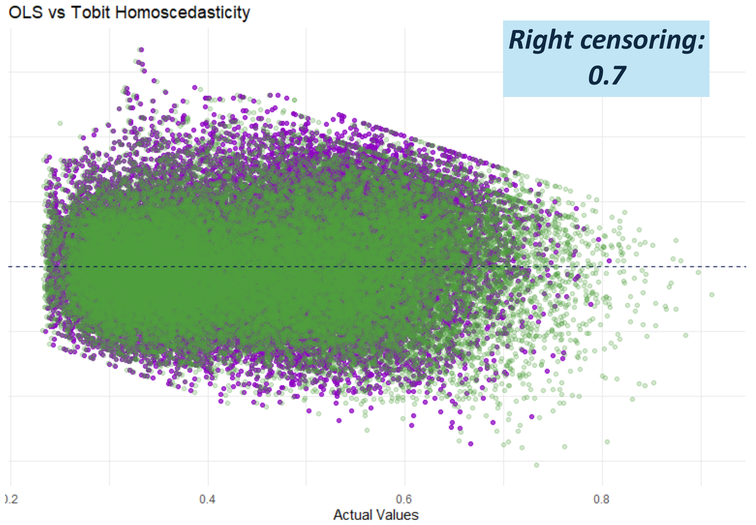
Appendix – Censored Data












```
> tobit_model <- tobit_model_right
> summary(tobit_model)
Call:
vglm(formula = tobit_formula, family = tobit(Upper = right_censor_threshold),
    data = WVS_20)

Coefficients:
                                             Estimate Std. Error  z value Pr(>|z|)    
(Intercept):1                               1.591e-01  3.104e-02    5.124 2.99e-07 ***
(Intercept):2                              -1.885e+00  4.101e-03 -459.494  < 2e-16 ***
DISBELIEF                                   3.846e-01  9.097e-02    4.228 2.36e-05 ***
I((10 * (DISBELIEF - 0.5))^2)               3.113e-03  1.275e-03    2.442  0.01461 *  
RELATIVISM                                  7.784e-02  3.067e-03   25.384  < 2e-16 ***
I((10 * (RELATIVISM - 0.5))^2)              1.335e-04  1.174e-04    1.137  0.25558    
SCEPTICISM                                  1.459e-02  5.016e-03    2.909  0.00362 ** 
I((10 * (SCEPTICISM - 0.5))^2)              2.336e-04  1.568e-04    1.490  0.13632    
AGE                                        -1.771e-04  8.363e-05   -2.118  0.03415 *  
I(AGE * EDUCATION)                          5.524e-04  1.362e-05   40.570  < 2e-16 ***
DEFIANCE                                    1.078e-01  7.292e-03   14.779  < 2e-16 ***
ReligCountrySecular                         3.169e-01  3.298e-02    9.607  < 2e-16 ***
DISBELIEF:ReligCountrySecular              -4.230e-01  9.266e-02   -4.564 5.01e-06 ***
I((10 * (DISBELIEF - 0.5))^2):ReligCountrySecular  1.280e-03  1.326e-03    0.966  0.33428    
RELATIVISM:ReligCountrySecular             -6.780e-02  4.341e-03  -15.620  < 2e-16 ***
I((10 * (RELATIVISM - 0.5))^2):ReligCountrySecular -2.004e-03  1.638e-04  -12.236  < 2e-16 ***
SCEPTICISM:ReligCountrySecular              1.751e-02  8.191e-03    2.137  0.03257 *  
I((10 * (SCEPTICISM - 0.5))^2):ReligCountrySecular -1.637e-03  2.460e-04   -6.657 2.80e-11 ***
AGE:ReligCountrySecular                    -1.787e-03  1.178e-04  -15.169  < 2e-16 ***
I(AGE * EDUCATION):ReligCountrySecular      1.640e-04  1.863e-05    8.803  < 2e-16 ***
DEFIANCE:ReligCountrySecular               -8.650e-03  9.686e-03   -0.893  0.37184    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Names of linear predictors: mu, loglink(sd)

Log-likelihood: 12009.65 on 65247 degrees of freedom

Number of Fisher scoring iterations: 6

No Hauck-Donner effect found in any of the estimates
```

Tobit
Right censoring - 0.8

```
> censored_ols_model <- lm(tobit_formula, data = WVS_20)
> summary(censored_ols_model)

Call:
lm(formula = tobit_formula, data = WVS_20)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.5876 -0.1021 -0.0024  0.1032  0.4855 

Coefficients:
                                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                 1.662e-01  2.946e-02   5.642 1.69e-08 ***
DISBELIEF                                   3.624e-01  8.636e-02   4.197 2.71e-05 ***
I((10 * (DISBELIEF - 0.5))^2)               2.857e-03  1.210e-03   2.361  0.01826 *  
RELATIVISM                                  7.750e-02  2.911e-03  26.619  < 2e-16 ***
I((10 * (RELATIVISM - 0.5))^2)              1.443e-04  1.115e-04   1.294  0.19563    
SCEPTICISM                                  1.513e-02  4.762e-03   3.178  0.00149 ** 
I((10 * (SCEPTICISM - 0.5))^2)              2.384e-04  1.489e-04   1.601  0.10931    
AGE                                        -1.745e-04  7.938e-05  -2.198  0.02796 *  
I(AGE * EDUCATION)                          5.421e-04  1.291e-05  41.996  < 2e-16 ***
DEFIANCE                                    1.049e-01  6.922e-03  15.162  < 2e-16 ***
ReligCountrySecular                         3.038e-01  3.129e-02   9.710  < 2e-16 ***
DISBELIEF:ReligCountrySecular              -3.822e-01  8.795e-02  -4.346 1.39e-05 ***
I((10 * (DISBELIEF - 0.5))^2):ReligCountrySecular  7.829e-04  1.258e-03   0.622  0.53381    
RELATIVISM:ReligCountrySecular             -6.582e-02  4.109e-03 -16.020  < 2e-16 ***
I((10 * (RELATIVISM - 0.5))^2):ReligCountrySecular -1.769e-03  1.548e-04 -11.427  < 2e-16 ***
SCEPTICISM:ReligCountrySecular              1.386e-02  7.743e-03   1.790  0.07347 .  
I((10 * (SCEPTICISM - 0.5))^2):ReligCountrySecular -1.642e-03  2.328e-04  -7.054 1.77e-12 ***
AGE:ReligCountrySecular                    -1.644e-03  1.116e-04 -14.734  < 2e-16 ***
I(AGE * EDUCATION):ReligCountrySecular      1.105e-04  1.752e-05   6.308 2.86e-10 ***
DEFIANCE:ReligCountrySecular               -2.518e-02  9.157e-03  -2.750  0.00597 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1444 on 32614 degrees of freedom
Multiple R-squared:  0.4117,	Adjusted R-squared:  0.4114 
F-statistic:  1201 on 19 and 32614 DF,  p-value: < 2.2e-16
```

OLS
Right censoring - 0.8

```
> tobit_model <- tobit_model_left
> summary(tobit_model)
Call:
vglm(formula = tobit_formula, family = tobit(Lower = left_censor_threshold),
    data = WVS_20)

Coefficients:
                                             Estimate Std. Error  z value Pr(>|z|)    
(Intercept):1                               1.484e-01  3.159e-02    4.698 2.62e-06 ***
(Intercept):2                              -1.876e+00  4.108e-03 -456.784  < 2e-16 ***
DISBELIEF                                   4.024e-01  9.253e-02    4.349 1.37e-05 ***
I((10 * (DISBELIEF - 0.5))^2)               3.302e-03  1.298e-03    2.544  0.01094 *  
RELATIVISM                                  8.095e-02  3.116e-03   25.980  < 2e-16 ***
I((10 * (RELATIVISM - 0.5))^2)              1.341e-04  1.194e-04    1.123  0.26152    
SCEPTICISM                                  1.596e-02  5.110e-03    3.123  0.00179 ** 
I((10 * (SCEPTICISM - 0.5))^2)              2.130e-04  1.599e-04    1.332  0.18298    
AGE                                        -1.889e-04  8.542e-05   -2.212  0.02698 *  
I(AGE * EDUCATION)                          5.653e-04  1.382e-05   40.899  < 2e-16 ***
DEFIANCE                                    1.106e-01  7.410e-03   14.922  < 2e-16 ***
ReligCountrySecular                         3.261e-01  3.351e-02    9.731  < 2e-16 ***
DISBELIEF:ReligCountrySecular              -4.378e-01  9.421e-02   -4.647 3.37e-06 ***
I((10 * (DISBELIEF - 0.5))^2):ReligCountrySecular  9.495e-04  1.348e-03    0.704  0.48128    
RELATIVISM:ReligCountrySecular             -7.076e-02  4.380e-03  -16.153  < 2e-16 ***
I((10 * (RELATIVISM - 0.5))^2):ReligCountrySecular -1.950e-03  1.651e-04  -11.810  < 2e-16 ***
SCEPTICISM:ReligCountrySecular              1.586e-02  8.255e-03    1.921  0.05477 .  
I((10 * (SCEPTICISM - 0.5))^2):ReligCountrySecular -1.581e-03  2.484e-04   -6.363 1.98e-10 ***
AGE:ReligCountrySecular                    -1.759e-03  1.193e-04  -14.738  < 2e-16 ***
I(AGE * EDUCATION):ReligCountrySecular      1.380e-04  1.869e-05    7.383 1.54e-13 ***
DEFIANCE:ReligCountrySecular               -1.179e-02  9.768e-03   -1.207  0.22735    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Names of linear predictors: mu, loglink(sd)

Log-likelihood: 12618.15 on 65247 degrees of freedom

Number of Fisher scoring iterations: 5

No Hauck-Donner effect found in any of the estimates
```

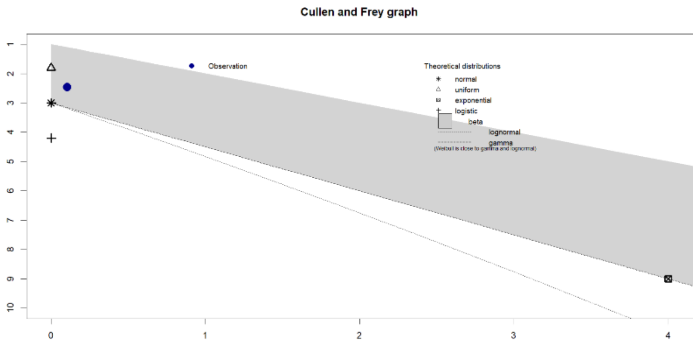
Tobit
Left censoring - 0.16

```
> censored_ols_model <- lm(tobit_formula, data = WVS_20)
> summary(censored_ols_model)

Call:
lm(formula = tobit_formula, data = WVS_20)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.56882 -0.10648 -0.00816  0.10026  0.66558 

Coefficients:
                                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                 1.548e-01  3.014e-02   5.135 2.84e-07 ***
DISBELIEF                                   4.230e-01  8.835e-02   4.788 1.69e-06 ***
I((10 * (DISBELIEF - 0.5))^2)               3.645e-03  1.238e-03   2.944  0.00324 ** 
RELATIVISM                                  7.306e-02  2.979e-03  24.529  < 2e-16 ***
I((10 * (RELATIVISM - 0.5))^2)              1.533e-04  1.140e-04   1.344  0.17889    
SCEPTICISM                                  1.289e-02  4.872e-03   2.646  0.00816 ** 
I((10 * (SCEPTICISM - 0.5))^2)              2.039e-04  1.523e-04   1.339  0.18073    
AGE                                        -1.794e-04  8.121e-05  -2.210  0.02713 *  
I(AGE * EDUCATION)                          5.351e-04  1.321e-05  40.513  < 2e-16 ***
DEFIANCE                                    1.059e-01  7.081e-03  14.948  < 2e-16 ***
ReligCountrySecular                         3.235e-01  3.201e-02  10.105  < 2e-16 ***
DISBELIEF:ReligCountrySecular              -4.625e-01  8.998e-02  -5.140 2.76e-07 ***
I((10 * (DISBELIEF - 0.5))^2):ReligCountrySecular  6.597e-04  1.287e-03   0.512  0.60832    
RELATIVISM:ReligCountrySecular             -6.329e-02  4.203e-03 -15.057  < 2e-16 ***
I((10 * (RELATIVISM - 0.5))^2):ReligCountrySecular -1.969e-03  1.583e-04 -12.434  < 2e-16 ***
SCEPTICISM:ReligCountrySecular              1.875e-02  7.921e-03   2.368  0.01791 *  
I((10 * (SCEPTICISM - 0.5))^2):ReligCountrySecular -1.558e-03  2.381e-04  -6.542 6.16e-11 ***
AGE:ReligCountrySecular                    -1.757e-03  1.141e-04 -15.392  < 2e-16 ***
I(AGE * EDUCATION):ReligCountrySecular      1.641e-04  1.792e-05   9.158  < 2e-16 ***
DEFIANCE:ReligCountrySecular               -7.708e-03  9.367e-03  -0.823  0.41061    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1477 on 32614 degrees of freedom
Multiple R-squared:  0.4143,	Adjusted R-squared:  0.414 
F-statistic:  1214 on 19 and 32614 DF,  p-value: < 2.2e-16
```

OLS
Left censoring - 0.16

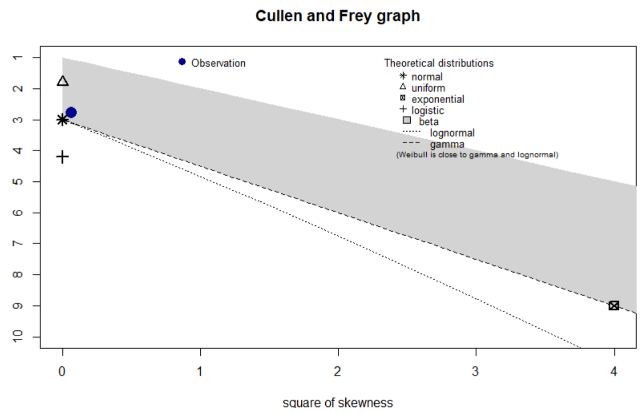
OLS

```
Coefficients:
                                                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                      2.127e-01  9.921e-03  21.443  < 2e-16 ***
I((10 * (DISBELIEF - 0.5))^2)                    1.927e-03  3.835e-04   5.026 5.05e-07 ***
DISBELIEF                                        1.186e-01  2.106e-02   5.633 1.79e-08 ***
RELATIVISM                                       9.436e-02  3.171e-03  29.760  < 2e-16 ***
I(AGE * EDUCATION)                               2.073e-04  1.826e-05  11.355  < 2e-16 ***
SCEPTICISM                                       2.413e-02  5.293e-03   4.559 5.17e-06 ***
AGE                                             -5.733e-04  9.088e-05  -6.309 2.85e-10 ***
I((10 * (RELATIVISM - 0.5))^2)                   7.853e-04  1.227e-04   6.401 1.56e-10 ***
I((10 * (SCEPTICISM - 0.5))^2)                   4.085e-04  1.602e-04   2.549 0.010794 *  
DEFIANCE                                         4.389e-02  7.904e-03   5.552 2.84e-08 ***
ReligCountrySecular                              2.761e-01  1.153e-02  23.955  < 2e-16 ***
I((10 * (DISBELIEF - 0.5))^2):ReligCountrySecular -1.365e-03  4.067e-04  -3.357 0.000788 ***
DISBELIEF:ReligCountrySecular                   -4.555e-02  2.145e-02  -2.123 0.033755 *  
RELATIVISM:ReligCountrySecular                  -1.006e-01  4.171e-03 -24.111  < 2e-16 ***
I(AGE * EDUCATION):ReligCountrySecular           2.667e-04  2.063e-05  12.929  < 2e-16 ***
SCEPTICISM:ReligCountrySecular                   5.703e-02  8.337e-03   6.841 7.97e-12 ***
AGE:ReligCountrySecular                         -1.213e-03  1.121e-04 -10.814  < 2e-16 ***
I((10 * (RELATIVISM - 0.5))^2):ReligCountrySecular -2.959e-03  1.586e-04 -18.654  < 2e-16 ***
I((10 * (SCEPTICISM - 0.5))^2):ReligCountrySecular -1.069e-03  2.395e-04  -4.463 8.11e-06 ***
DEFIANCE:ReligCountrySecular                     2.184e-02  9.649e-03   2.264 0.023592 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1477 on 34360 degrees of freedom
Multiple R-squared:  0.4526,	Adjusted R-squared:  0.4523 
F-statistic:  1496 on 19 and 34360 DF,  p-value: < 2.2e-16

> print(best_result$MSE)
[1] 0.02181231
> print(best_result$AIC)
[1] -33904.94
```



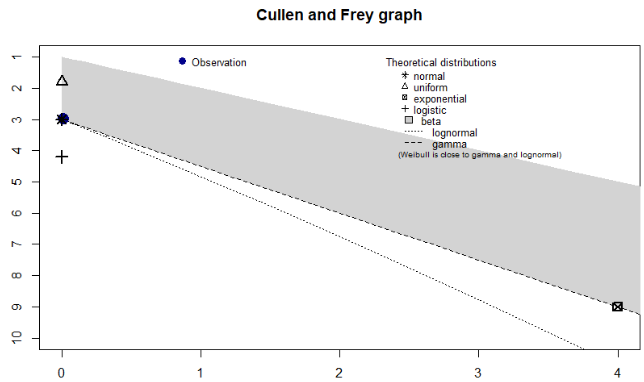
Quantile Regression (Q=0.1)

```
tau: [1] 0.1

Coefficients:
                                                    Value    Std. Error t value   Pr(>|t|)
(Intercept)                                          0.07808  0.01473    5.30105   0.00000
DISBELIEF                                            0.06685  0.03082    2.16919   0.03007
I((10 * (DISBELIEF - 0.5))^2)                        0.00130  0.00058    2.23959   0.02512
RELATIVISM                                           0.07968  0.00424   18.77583   0.00000
I((10 * (RELATIVISM - 0.5))^2)                       0.00065  0.00017    3.77646   0.00016
SCEPTICISM                                           0.00513  0.00615    0.83470   0.40389
I((10 * (SCEPTICISM - 0.5))^2)                       0.00003  0.00020    0.15245   0.87883
AGE                                                 -0.00024  0.00012   -2.02490   0.04289
I(AGE * EDUCATION)                                   0.00018  0.00002    7.64214   0.00000
DEFIANCE                                             0.05118  0.00982    5.20915   0.00000
ReligCountrySecular                                  0.22443  0.01708   13.14117   0.00000
DISBELIEF:ReligCountrySecular                       -0.01657  0.03179   -0.52121   0.60223
I((10 * (DISBELIEF - 0.5))^2):ReligCountrySecular   -0.00144  0.00063   -2.29456   0.02176
RELATIVISM:ReligCountrySecular                      -0.06111  0.00528  -11.56916   0.00000
I((10 * (RELATIVISM - 0.5))^2):ReligCountrySecular  -0.00191  0.00023   -8.25814   0.00000
SCEPTICISM:ReligCountrySecular                       0.04591  0.01394    3.29260   0.00099
I((10 * (SCEPTICISM - 0.5))^2):ReligCountrySecular  -0.00027  0.00034   -0.78195   0.43425
AGE:ReligCountrySecular                             -0.00157  0.00015  -10.52104   0.00000
I(AGE * EDUCATION):ReligCountrySecular               0.00026  0.00003    8.57390   0.00000
DEFIANCE:ReligCountrySecular                        -0.00431  0.01193   -0.36158   0.71767
```



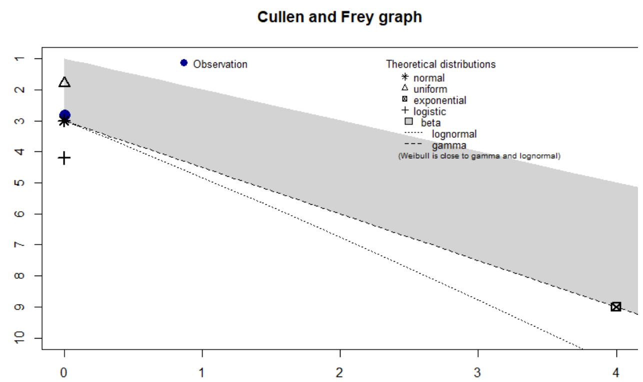
Quantile Regression (Q=0.5)

```
tau: [1] 0.5

Coefficients:
                                                    Value    Std. Error t value   Pr(>|t|)
(Intercept)                                          0.20429  0.01325   15.41495   0.00000
DISBELIEF                                            0.12579  0.02848    4.41595   0.00001
I((10 * (DISBELIEF - 0.5))^2)                        0.00195  0.00045    4.28390   0.00002
RELATIVISM                                           0.09091  0.00408   22.25595   0.00000
I((10 * (RELATIVISM - 0.5))^2)                       0.00072  0.00014    5.22747   0.00000
SCEPTICISM                                           0.02109  0.00585    3.60427   0.00031
I((10 * (SCEPTICISM - 0.5))^2)                       0.00034  0.00018    1.88792   0.05904
AGE                                                 -0.00047  0.00009   -5.17042   0.00000
I(AGE * EDUCATION)                                   0.00024  0.00002   11.92191   0.00000
DEFIANCE                                             0.04643  0.00953    4.87375   0.00000
ReligCountrySecular                                  0.27609  0.01586   17.40793   0.00000
DISBELIEF:ReligCountrySecular                       -0.04758  0.02767   -1.71950   0.08553
I((10 * (DISBELIEF - 0.5))^2):ReligCountrySecular   -0.00131  0.00053   -2.45384   0.01414
RELATIVISM:ReligCountrySecular                      -0.10081  0.00558  -18.04971   0.00000
I((10 * (RELATIVISM - 0.5))^2):ReligCountrySecular  -0.00334  0.00021  -15.55451   0.00000
SCEPTICISM:ReligCountrySecular                       0.08516  0.01055    8.07088   0.00000
I((10 * (SCEPTICISM - 0.5))^2):ReligCountrySecular  -0.00110  0.00031   -3.58118   0.00034
AGE:ReligCountrySecular                             -0.00162  0.00014  -11.19159   0.00000
I(AGE * EDUCATION):ReligCountrySecular               0.00034  0.00003   12.32942   0.00000
DEFIANCE:ReligCountrySecular                         0.01481  0.01268    1.16825   0.24271
```



Quantile Regression (Q=0.9)

```
tau: [1] 0.9

Coefficients:
                                                    Value    Std. Error t value   Pr(>|t|)
(Intercept)                                          0.34752  0.01557   22.32073   0.00000
DISBELIEF                                            0.16581  0.03769    4.39894   0.00001
I((10 * (DISBELIEF - 0.5))^2)                        0.00247  0.00071    3.46101   0.00054
RELATIVISM                                           0.10713  0.00530   20.22875   0.00000
I((10 * (RELATIVISM - 0.5))^2)                       0.00090  0.00019    4.73409   0.00000
SCEPTICISM                                           0.04425  0.00852    5.19370   0.00000
I((10 * (SCEPTICISM - 0.5))^2)                       0.00066  0.00030    2.23943   0.02513
AGE                                                 -0.00073  0.00014   -5.16326   0.00000
I(AGE * EDUCATION)                                   0.00024  0.00003    9.08349   0.00000
DEFIANCE                                             0.03497  0.01026    3.40722   0.00066
ReligCountrySecular                                  0.35357  0.01794   19.70522   0.00000
DISBELIEF:ReligCountrySecular                       -0.07960  0.03858   -2.06342   0.03908
I((10 * (DISBELIEF - 0.5))^2):ReligCountrySecular   -0.00167  0.00077   -2.17247   0.02983
RELATIVISM:ReligCountrySecular                      -0.14133  0.00677  -20.88004   0.00000
I((10 * (RELATIVISM - 0.5))^2):ReligCountrySecular  -0.00299  0.00027  -11.26164   0.00000
SCEPTICISM:ReligCountrySecular                       0.01067  0.01485    0.71819   0.47265
I((10 * (SCEPTICISM - 0.5))^2):ReligCountrySecular  -0.00122  0.00047   -2.57373   0.01007
AGE:ReligCountrySecular                             -0.00072  0.00017   -4.24178   0.00002
I(AGE * EDUCATION):ReligCountrySecular               0.00012  0.00003    4.27680   0.00002
DEFIANCE:ReligCountrySecular                         0.07804  0.01315    5.93317   0.00000
```

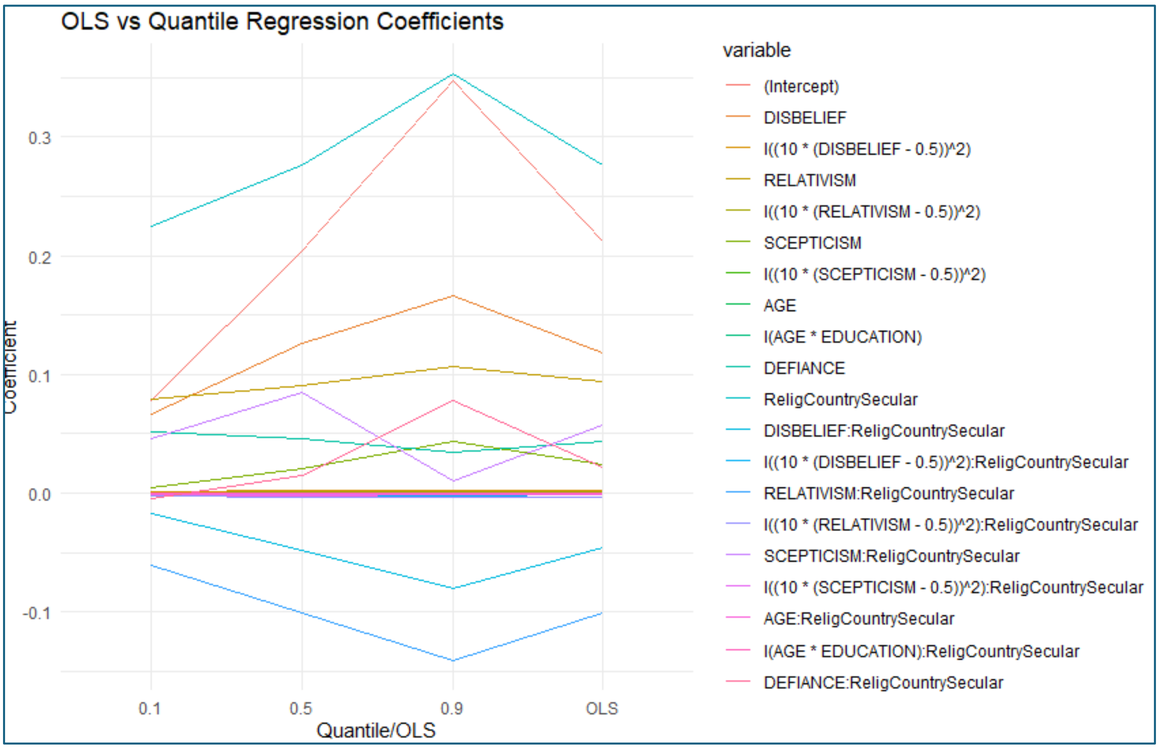


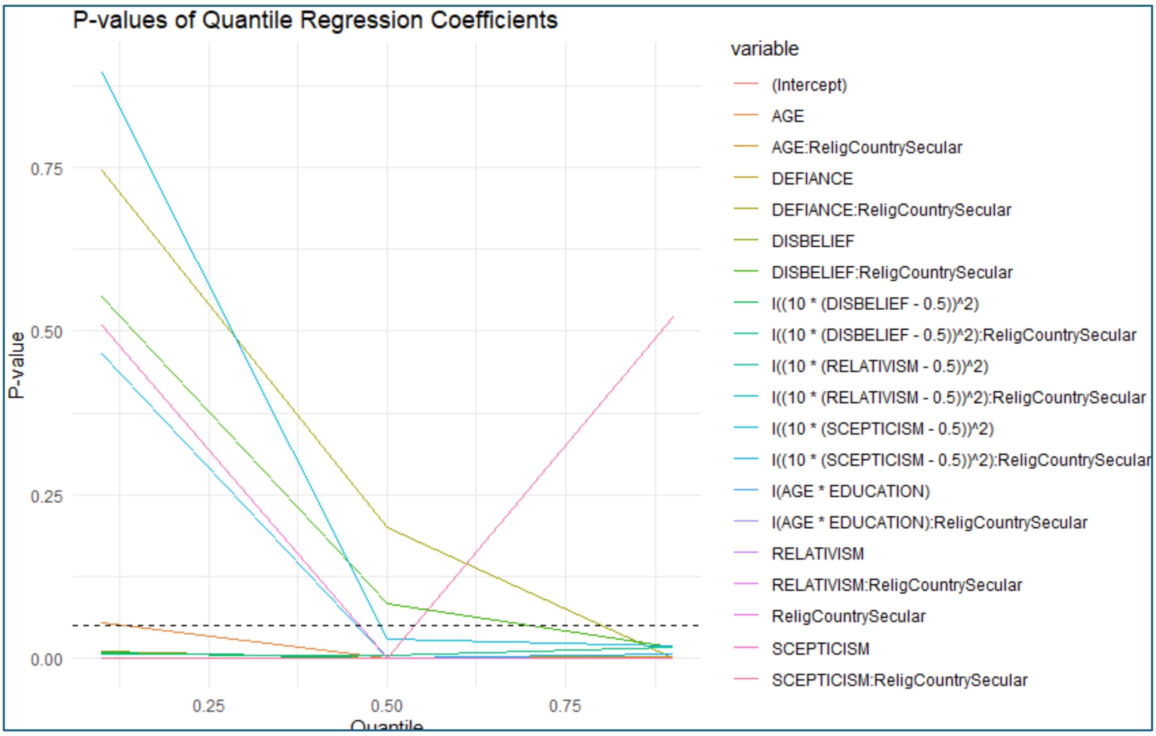
LS
Left censoring
- 0.32

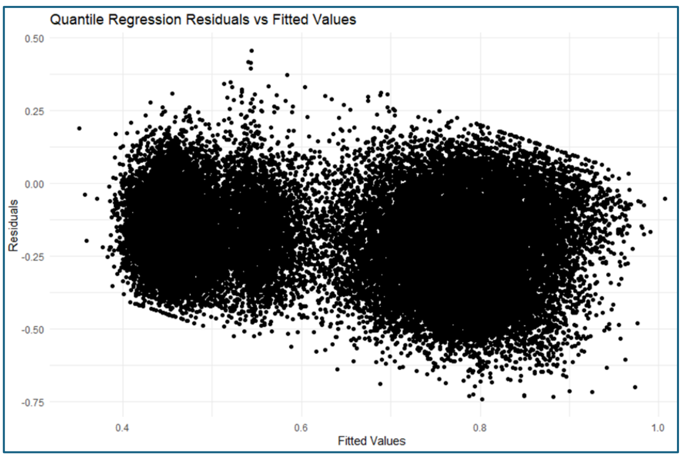
```
> censored_ols_model <- lm(tobit_formula, data = WVS_20)
> summary(censored_ols_model)

Call:
lm(formula = tobit_formula, data = WVS_20)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.47102 -0.08061 -0.02246  0.07777  0.62848 

Coefficients:
                                                Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                    2.137e-01  2.641e-02   8.093 6.01e-16 ***
DISBELIEF                                      4.343e-01  7.739e-02   5.612 2.02e-08 ***
I((10 * (DISBELIEF - 0.5))^2)                  4.119e-03  1.085e-03   3.798 0.000146 ***
RELATIVISM                                     4.475e-02  2.609e-03  17.150  < 2e-16 ***
I((10 * (RELATIVISM - 0.5))^2)                 1.077e-04  9.990e-05   1.078 0.281025    
SCEPTICISM                                     3.681e-03  4.268e-03   0.862 0.388480    
I((10 * (SCEPTICISM - 0.5))^2)                 8.480e-05  1.334e-04   0.636 0.525090    
AGE                                           -1.281e-04  7.114e-05  -1.800 0.071863 .  
I(AGE * EDUCATION)                             4.127e-04  1.157e-05  35.676  < 2e-16 ***
DEFIANCE                                       8.574e-02  6.203e-03  13.821  < 2e-16 ***
ReligCountrySecular                            2.914e-01  2.804e-02  10.392  < 2e-16 ***
DISBELIEF:ReligCountrySecular                 -4.986e-01  7.882e-02  -6.326 2.55e-10 ***
I((10 * (DISBELIEF - 0.5))^2):ReligCountrySecular   4.784e-04  1.128e-03   0.424 0.671387    
RELATIVISM:ReligCountrySecular                -3.904e-02  3.682e-03 -10.603  < 2e-16 ***
I((10 * (RELATIVISM - 0.5))^2):ReligCountrySecular -1.872e-03  1.387e-04 -13.494  < 2e-16 ***
SCEPTICISM:ReligCountrySecular                 2.858e-02  6.940e-03   4.118 3.83e-05 ***
I((10 * (SCEPTICISM - 0.5))^2):ReligCountrySecular -1.277e-03  2.086e-04  -6.120 9.44e-10 ***
AGE:ReligCountrySecular                       -1.670e-03  1.000e-04 -16.700  < 2e-16 ***
I(AGE * EDUCATION):ReligCountrySecular         2.415e-04  1.570e-05  15.382  < 2e-16 ***
DEFIANCE:ReligCountrySecular                   4.490e-03  8.206e-03   0.547 0.584288    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1294 on 32614 degrees of freedom
Multiple R-squared:  0.4018,	Adjusted R-squared:  0.4015 
F-statistic:  1153 on 19 and 32614 DF,  p-value: < 2.2e-16
```

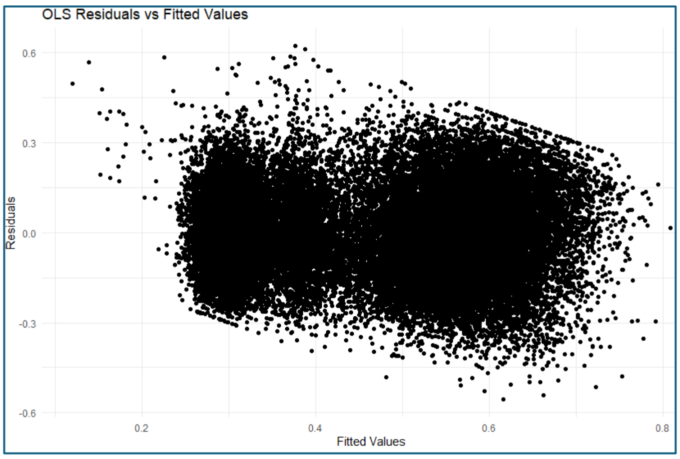
Tobit
Left censoring
- 0.32

```
> tobit_model <- tobit_model_left
> summary(tobit_model)
Call:
vglm(formula = tobit_formula, family = tobit(Lower = left_censor_threshold), 
    data = WVS_20)

Coefficients: 
                                                Estimate Std. Error  z value Pr(>|z|)    
(Intercept):1                                  1.063e-01  3.588e-02    2.961  0.00306 ** 
(Intercept):2                                 -1.821e+00  4.659e-03 -390.822  < 2e-16 ***
DISBELIEF                                      4.831e-01  1.047e-01    4.613 3.97e-06 ***
I((10 * (DISBELIEF - 0.5))^2)                  3.932e-03  1.475e-03    2.665  0.00770 ** 
RELATIVISM                                     8.713e-02  3.517e-03   24.778  < 2e-16 ***
I((10 * (RELATIVISM - 0.5))^2)                 1.922e-04  1.356e-04    1.417  0.15644    
SCEPTICISM                                     1.827e-02  5.835e-03    3.131  0.00174 ** 
I((10 * (SCEPTICISM - 0.5))^2)                 2.663e-04  1.836e-04    1.451  0.14686    
AGE                                           -2.697e-04  9.889e-05   -2.727  0.00639 ** 
I(AGE * EDUCATION)                             6.194e-04  1.552e-05   39.899  < 2e-16 ***
DEFIANCE                                       1.278e-01  8.352e-03   15.306  < 2e-16 ***
ReligCountrySecular                            3.722e-01  3.784e-02    9.837  < 2e-16 ***
DISBELIEF:ReligCountrySecular                 -5.317e-01  1.064e-01   -4.996 5.85e-07 ***
I((10 * (DISBELIEF - 0.5))^2):ReligCountrySecular   6.178e-04  1.526e-03    0.405  0.68567    
RELATIVISM:ReligCountrySecular                -7.616e-02  4.818e-03  -15.806  < 2e-16 ***
I((10 * (RELATIVISM - 0.5))^2):ReligCountrySecular -2.040e-03  1.821e-04  -11.205  < 2e-16 ***
SCEPTICISM:ReligCountrySecular                 1.445e-02  9.061e-03    1.595  0.11067    
I((10 * (SCEPTICISM - 0.5))^2):ReligCountrySecular -1.718e-03  2.745e-04   -6.261 3.83e-10 ***
AGE:ReligCountrySecular                       -1.773e-03  1.335e-04  -13.282  < 2e-16 ***
I(AGE * EDUCATION):ReligCountrySecular         1.078e-04  2.056e-05    5.242 1.58e-07 ***
DEFIANCE:ReligCountrySecular                  -2.722e-02  1.077e-02   -2.529  0.01145 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Names of linear predictors: mu, loglink(sd)

Log-likelihood: 2909.727 on 65247 degrees of freedom

Number of Fisher scoring iterations: 9 
```


OLS vs Quantile Regression Coefficients
variable
(Intercept)
DISBELIEF
I((10 * (DISBELIEF - 0.5))^2)
RELATIVISM
I((10 * (RELATIVISM - 0.5))^2)
SCEPTICISM
I((10 * (SCEPTICISM - 0.5))^2)
AGE
I(AGE * EDUCATION)
DEFIANCE
ReligCountrySecular
DISBELIEF:ReligCountrySecular
I((10 * (DISBELIEF - 0.5))^2):ReligCountrySecular
RELATIVISM:ReligCountrySecular
I((10 * (RELATIVISM - 0.5))^2):ReligCountrySecular
SCEPTICISM:ReligCountrySecular
I((10 * (SCEPTICISM - 0.5))^2):ReligCountrySecular
AGE:ReligCountrySecular
I(AGE * EDUCATION):ReligCountrySecular
DEFIANCE:ReligCountrySecular
Coefficient
0.3
0.2
0.1
0.0
-0.1
0.1
0.5
0.9
OLS
Quantile/OLS


P-values of Quantile Regression Coefficients
variable
(Intercept)
AGE
AGE:ReligCountrySecular
DEFIANCE
DEFIANCE:ReligCountrySecular
DISBELIEF
DISBELIEF:ReligCountrySecular
I((10 * (DISBELIEF - 0.5))^2)
I((10 * (DISBELIEF - 0.5))^2):ReligCountrySecular
I((10 * (RELATIVISM - 0.5))^2)
I((10 * (RELATIVISM - 0.5))^2):ReligCountrySecular
I((10 * (SCEPTICISM - 0.5))^2)
I((10 * (SCEPTICISM - 0.5))^2):ReligCountrySecular
I(AGE * EDUCATION)
I(AGE * EDUCATION):ReligCountrySecular
RELATIVISM
RELATIVISM:ReligCountrySecular
ReligCountrySecular
SCEPTICISM
SCEPTICISM:ReligCountrySecular
P-value
0.75
0.50
0.25
0.00
0.25
0.50
0.75
Quantile


Quantile Regression Residuals vs Fitted Values
Residuals
0.50
0.25
0.00
-0.25
-0.50
-0.75
0.4
0.6
0.8
1.0
Fitted Values


OLS Residuals vs Fitted Values
Residuals
0.6
0.3
0.0
-0.3
-0.6
0.2
0.4
0.6
0.8
Fitted Values

Appendix – Panel Data

Fixed Effects Model

```
> summary(fixed_effects_model)
Oneway (individual) effect Within Model

Call:
plm(formula = formula_best_2501, data = ADHD, model = "within",
    index = c("user", "med"))

Balanced Panel: n = 6, T = 2, N = 750

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-1.7866352 -0.2122037  0.0065847  0.2098726  1.6372718

Coefficients:
                          Estimate  Std. Error t-value  Pr(>|t|)
med1                    -0.18947991  0.02803215 -6.7594 2.824e-11 ***
x_mean                  -0.00953463  0.00426467 -2.2357 0.0256691 *
I(x_mean^2)             -0.00304807  0.00151874 -2.0070 0.0451187 *
I(sqrt(abs(x_mean)))     0.12816080  0.03527912  3.6328 0.0002999 ***
I(y_mean^2)              0.00196937  0.00080588  2.4438 0.0147689 *
I(sqrt(abs(y_std)))      0.06024121  0.02218782  2.7151 0.0067819 **
gz_mean                  0.17026746  0.08414859  2.0234 0.0433919 *
I(log(z_mean + 1e-06))   0.06091412  0.02784111  2.1879 0.0289896 *
rec_num                 -0.04915425  0.00938674 -5.2366 2.138e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    110.85
Residual Sum of Squares: 94.692
R-Squared:      0.14578
Adj. R-Squared: 0.12951
F-statistic: 13.9376 on 9 and 735 DF, p-value: < 2.22e-16
```

Random Effects Model

```
> summary(random_effects_model)
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = formula_compare_to_re_2501, data = ADHD, model = "random",
    index = c("user", "med"))

Balanced Panel: n = 6, T = 2, N = 750

Effects:
                 var std.dev share
idiosyncratic 0.1336  0.3655 0.056
individual    2.2705  1.5068 0.944
theta: 0.8309

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-1.868460 -0.222729  0.010173  0.224944  1.644081

Coefficients:
                         Estimate Std. Error z-value  Pr(>|z|)
(Intercept)             6.1722726  0.2933562 21.0402 < 2.2e-16 ***
med1                   -0.2032444  0.0284376 -7.1470 8.867e-13 ***
rec_num                -0.0510474  0.0095287 -5.3572 8.451e-08 ***
I(sqrt(abs(x_mean)))    0.0650661  0.0212418  3.0631   0.00219 **
I(log(z_mean + 1e-06))  0.0453856  0.0218023  2.0817   0.03737 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    112.17
Residual Sum of Squares: 100.34
R-Squared:      0.10551
Adj. R-Squared: 0.10071
Chisq: 87.8807 on 4 DF, p-value: < 2.22e-16
```