Life Expectancy Analysis with World Bank Indicators

Gauransh Tandon (gtandon2)

March 31, 2020

Introduction

Life in its various forms has always been a significant factor on Earth for billions of years. Humans have established themselves as one of the most influential life form on Earth. Life expectancy has expanded quickly since the beginning of human race. Appraisals recommend that in a pre-present day, poor world, life expectancy was around 30 years in all areas of the world. In the mid nineteenth century, life expectancy began to increment in the early industrialized nations while it remained low in the remainder of the world. This prompted a high disparity in how wellbeing was disseminated over the world. Great wellbeing in the rich nations and perseveringly terrible wellbeing in those nations that stayed poor. Throughout the most recent decades this worldwide disparity diminished. Nations that in the relatively recent past were experiencing awful wellbeing are making up for lost time quickly. Since 1900 the worldwide normal life expectancy has dramatically increased and is currently moving toward 70 years. No nation on the planet has a lower life expectancy than the nations with the most noteworthy life expectancy in 1800.

What might be the reason for this increase?

In this study, I will carry out some analysis about the life expectancy for humans. Obviously, this is a very, very large endeavor - to ensure that we identify a manageable scope for this project - I deliberated over a bunch of different approaches and finally settled on an Analysis of Life Expectancy based on Socio-Economic and other indicators available from the World Bank's World indicators dataset. Therefore, I will assert certain Socio-Economic Factors I feel are relavant then see whether I was correct.

I suspect the reasons for higher Life Expectancy to be: Lower Birth Rate; Higher GDP and lower Country's debts; Higher Health Care Facilities;Higher Urban Population; Lower Infant Mortality Rates

Now, in order to be more relevant we will use certain assumptions such as, looking into only data from Asia, Europe and Oceania regions, then dividing the data based on Male and Female Life Expectancy. This will help us have conclusions for both Males and females in different forms. I suspect both Male and Females to have different reasons to live more. Since, traditionally, males have been in the workforce more, therefore, I suspect Males to have more impact due to how Economy is doing globally(basically higher GDP and lower debts implying more Job creation further implying less stress and more life on earth). For females, I suspect healthcare to be the major factor as well as how many kids are being born(since more kids => chances of females to die early increases).

After rigourous analysis done throughout this project, the best model for predicting Life expectancy for males and females are somewhat similar to my hypothesis. For males, high GDP was very essential, for females - better health care was essential. Moreso, we found that mobile usage in men can actually lower life expectancy (which is very logical but interesting to see with data). For males, lower birth rate was also essential in them living longer (which again makes sense cause the lowert the birth rate, the lower their responsibility to take care of more people which further implies lower stress and more chances of them living longer). One very important reason for higher life expectancy for both male and female is Urban life. Urbanization has actually been very useful for increasing life expectancy (which again makes sense cause urban life comes with better health care and more jobs).

Data Description

Source

Databank (databank.worldbank.org) is an online web resource that provides simple and quick access to collections of time series data.Tableau has exposed a portion of this data consisting of Socio-economic, Health and Development related information, as part of its built-in datasets. The dataset used for this project is a full CSV export of this Tableau dataset.

Description of Dataset

The dataset contains Socio-economic and Life expectancy data for each of the world countries from 2000-2012. There are 2705 observations in the dataset. Each observation consists of 27 variables. These 27 variables represent various kinds of data points across different topics like Economic, Social, Health, industrial Development, etc. The variables consist of mix of continuous, Nominal, etc.

Economic indicator data : GDP, Health expenditure as % of GDP, Health Expenditure/Capita, Business Tax Rate, Lending Interest Rate, Tourism Inbound, Tourism Outbound, etc.

Social indicator data : total population, population in the various age groups (0-14,15-64, 65+), Urban Population, Internet Usage, Ease of doing business, etc.

Health Indicator Data : Birth Rate, Infant Mortality Rate, Life Expectancy, etc.

Industrial Indicator Data : CO_2 Emissions, Days to Start Business, etc.

It is important to note that there are instances of missing data, and appropriate data cleansing activities will be performed as part of the analysis.

Preparation of Dataset

Data Cleansing

Step 1: Elimination of columns with more than 33% NA only data. Subsequently, removed all null values from dataset, and identified a dataset with 1401 Observations for the purposes of this Analysis. The dataset has data for various regions of the world. From our practical understanding of the world regions, I realize that different regions have different characteristics. So, for the purpose of this Analysis Project (and to keep the scope manageable) I chose to work with the data from Asia, Europe and Oceania regions (code in Appendix)

Inventory of the variables

BR: Birth Rate
CO2: CO2 Emissions
Cntry: Country
GDP: GDP
HCS_GDP: Health Expense % GDP
HCPC: Health Expense per Capita
IMR: Infant Mortality Rate
IU: Internet Usage
LI: Lending Interest
LE_F: Life Expectancy Female
LE_M: Life Expectancy Male
Mobile: Mobile Phone Usage
NR: Number of Record
P_0_14: Population 0-14
P_15_64: Population 15-64
P_65_plus: Population 65+
Pop: Population Total
PU: Population Urban
Region: Region
TI: Tourism Inbound
TO: Tourism Outbound
Year: Year

Data Analysis

Step 2:

As part of the initial Data Analysis, I am using the pairs plot to understand the high level structure of the data. The pairs plot is a scatter plot matrix which is a table of scatter plots. Scatterplot matrices are good for determining rough linear correlations of metadata that contain continuous variables.

In addition, as seen below, the correlation matrix and correlation plot are also used to help gain some added high-level insights into the data. At this point, we only take this information under advise, and do not use it directly in building models.

The scatterplot, correlation matrix, and correlation plot can be found in Appendix.

After looking, I can see that Population based variables include a full population count variable as well as several other variables which represent specific population characteristics. So I want to analyze how much of its variation is explained by a linear relationship with other population based predictors.

After experimenting with that variation, I see how almost all of the variation in the linear regression model with depenedent variable(Pop) and independent variables(BR, CO2, Cntry, HCSPC, IMR, IU ,Mobile, P_15_64, PU) has a linear relationship with the other variables. This can be confirmed by looking at the R squared for this model which comes out to be 0.994 (code in the Appendix). Therefore, I am deciding to not use it in the model creation.

Regression Analysis

Model Building

After having looked through and analyzed the dataset, I have decided to build seperate models for Life Expectancy of Males and Life Expectancy of Females. I am doing this with an intent of identifying the best predictor variables which help with better explaining the chosen response variables. I aim to create the best model using our analaysis

Full Model with all the variables

I create a Comparing the full additive model, with a random smaller additive model (as seen in appendix) - I find that for further analysis, in case of the Life Expectancy of Males the bigger model is to be preferred, while for the Life Expectancy of Females the smaller model is to be preferred (using an $\alpha = 0.05$). The summaries of the Anova models can be found in appendix for both Male And Females.

Exploratory Interaction Model

I a create exploratory interaction models for again both Males and Females seprately. For the male life expectancy regression model, I choose to use independent variables as Country, and square of (Birth Rate, logrithmic GDP, Health Expense per Capita, Lending Interest, Mobile Phone Usage, Population 15-64, Population 64+, Urban Population and Tourism Inbound). Similarly, for the female life expectancy regression model, I choose to use independent variables as Country, and square of (Birth Rate, CO2 Consumption, Health Expense per Capita, Infant Mortality Rate, Internet Usage, Mobile Usage, Population 15-64, and Urban Population). The summaries of these models can be found in the Appendix at the end of this report.

Polynomial Model

Based on the Data analysis, I saw that maybe there are nonlinear phenomenas within the data which require better incoporporation within the modeling. Therefore, I go deeper with understanding Polynomial Multiple Regression and using that for both Male and Female Life expectancy. Again, for the male model, I choose, the Exploratory model's dependent variables but this time I add special cubic functions for all those dependent variables as well. Basically this new model contains squared variables as well as cubic variables for - Birth Rate, logrithmic GDP, Health Expense per Capita, Lending Interest, Mobile Phone Usage, Population 15-64, Population 64+, Urban Population and Tourism Inbound. Similarly, for females, it includes squared variables as well as cubic variables for - Birth Rate, CO2 Consumption, Health Expense per Capita, Infant Mortality Rate, Internet Usage, Mobile Usage, Population 15-64, and Urban Population. The summaries of these models can be found in the Appendix at the end of this report.

Identifying the best models from the starter models

At this point, we will not be looking to eliminate any variables or interaction terms from the model based on VIF. But will proceed further to see if alternate selection of variables is warranted. We will be first pursuing the Modeling of Life Expectancy of Females, thereby we will do the same for Males. We will do the model analysis based on three well known model selection techniques:

AIC Selection Technique

The Akaike information criterion (AIC) is an estimator of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection.

SIC Selection Technique

The Schwarz information criterion is a criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. It is based, in part, on the likelihood function and it is closely related to the Akaike information criterion

All the Rsquared and Adjusted Rsquared for different models can be found in a tabular form in Appendix under Model Analysis.

Validating LINE on the selected models

Now, I am interested in checking out the basic assumptions of Constant Variance and Normality. Based on the experiment(refer to Validating LINE on the selected models within Appendix) have sufficient evidence to believe that our assumptions are valid on the chosen models.

Method for selecting the one best suited model

In order to finally find the best suited model, I will use Adjusted R-squared as well as LOOCV-RMSE.

> Adjusted R-Squared The adjusted R-squared compares the explanatory power of regression models that contain different numbers of predictors. This is helpful because by just using R-squared, every time I add a predictor to a model, the R-squared increases, even if due to chance alone. It never decreases. Consequently, a model with more terms may appear to have a better fit simply because it has more terms. Therefore, Adjusted R-squared helps in that case.
>
> LOOCV-RMSE Another method for finalizing our best model is using Leave One Out Cross Validation - Root Mean Square Error Method. Basically, as the name says, we leave out one variable and calculate the test error on the held out point and also caluclate the average test error and then compare for every term. This is a very effective method.

The analysis can be found under Selecting Best Suited Model in Appendix.

Empirical Results

Final Model

After my analysis, the polynomial based models for Females and Males stand out with both the LOOCV-RMSE as well as the Adjusted R-Squared. However, I also see that the polynomial models have third order terms. The rest of the models (additive & interaction) are also very close in their abilities (to predict as well as to explain the response variable with the predictors).

My goal, as learnt from the course - is to be able to find a model that is simple and, yet, effective in being able to explain the Response variable (Life Expectancy of Females) with the predictors. With this in mind I believe that the model "LE_F ~ Cntry + PU + IMR + HCSPC + PU:IMR" (forward interaction SIC) does an admirably good job given its simplicity and effectiveness. This model is my chosen model.

Similarly, for the Life Expectancy of Males - for its ability to explain/simplicity/predict - the chosen model will be "LE_M ~ BR + Cntry + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI" (backward additive AIC). I also looked at the backward additive SIC model, however it failed the shapiro-wilk test, and I was comfortable with the backward additive AIC model for its abilities and simplicity.

Summary & Discussion

Does model analysis show that the selected Model has predictors that help explain the response variable?

Yes, I believe that the selected model does a good job of explaining the response variable with the predictors.

What does the model mean in real life?

The model for Life Expectacy for Males of a particular region shows us how average birth rate of that region, country of birth, how well the country is doing Economically (GDP), how expensive is healthcare in that country, how much in debt is the country, how many people are from 15 to 65 plus within the country, how big is the urban population, how much mobile usage is present and finally, how is the tourism for the country.

The model for Life Expectacy for Females of a particular region shows us how country of birth, how well the country is doing Economically (GDP), how expensive is healthcare in that country, how big is the urban population, how bad is the infant mortality rate and how bad is the infant mortality rate in urban areas.

These factors gives us very very good exposure to the economic condition of the country to life expectancy relation.

For males, high GDP was very essential, for females - better health care was essential. Moreso, we found that mobile usage in men can actually lower life expectancy (which is very logical but interesting to see with data). For males, lower birth rate was also essential in them living longer (which again makes sense cause the lowert the birth rate, the lower their responsibility to take care of more people which further implies lower stress and more chances of them living longer). One very important reason for higher life expectancy for both male and female is Urban life. Urbanization has actually been very useful for increasing life expectancy (which again makes sense cause urban life comes with better health care and more jobs).

What about other regions?

I found that there are subtle differences in the models across the different regions (of the world) for the Life Expectancy of both Males and Females. And this is consistent with what I expected. This is reflected in how the models chose key items like industrial, Social and economic factors - which are indicative of access to health care/standard of living/etc. This analysis also helped validate our line of thinking that it is better to create models at the region (roughly equal to Continent) level as opposed to one model for the entire world.

How does the selected model do with predictions?

The selected model is very useful in this context as well.

As seen from the analysis above, the model for Life Expectancy of Males is different than the model arrived at for Life Expectancy of Females. This stands to reason, as there are different factors that affect Males and Females. Given the traditional roles of Males as bread winners and Females as homemakers across various cultures, even as of the current times, we can see that the Males have more of the Economis + Industrial indicators at play compared to the women. So, I believe the models are consistent with our general understanding of the underlying factors.

What are the limitations of this model? What can be done in future for expanding this research?

The limitation of this model is that I am not using all the regions for our analysis. That could have given more diverse reasons which might have been a little different from what I obtain right now. That itself turns to be the future for this research. I encourage the reader to use the dataset and explore different regions and confirm if the model still comes out to be what we selected or if it could have been something else. Another limitation is that we use polynomial model based on only cubic version which could further be explored with quadrilateral or further higher n-terms. Though the reader needs to keep in mind that it’s important to create simple yet effective models so that inferences could be easier.

What are my final thoughts?

```
All models are wrong, but some are useful.

      - George Box ("Robustness in the strategy of scientific model building")
```

“All models are wrong” that is, every model is wrong because it is a simplification of reality. “But some are useful” - simplifications of reality can be quite useful. They can help us explain, predict and understand the universe and all its various components.

Appendix

Data Prepration

Data Cleaning Code

```r
raw_data = read.csv('World Indicators.csv')
raw_data = raw_data[,colSums(is.na(raw_data))<nrow(raw_data)/3]
raw_data = na.omit(raw_data)
colnames(raw_data) = c("BR", "CO2", "Cntry", "GDP", "HCS_GDP", "HCSPC" , "IMR","IU","LI", "LE_F","LE_M","Mobi
le", "NR","P_0_14","P_15_64","P_65_plus", "Pop","PU","Region", "TI", "TO","Year" )

raw_data_copy = raw_data

raw_data_Asia = raw_data[which(raw_data$Region == 'Asia'),]
raw_data_Europe = raw_data[which(raw_data$Region == 'Europe'),]
raw_data_Oceania = raw_data[which(raw_data$Region == 'Oceania'),]

raw_data = raw_data_Oceania
```

Scatterplot

```r
cols <- character(nrow(iris))
cols[] <- "black"

cols[iris$Species %in% c("setosa","versicolor")] <- "blue"
cols[iris$Species == "virginica"] <- "red"
```

```
pairs(raw_data[,c(1,2,3,4,5,6,7,8,9,10,11,16,18,22)], col = cols)
```

Correlation Matrix

```
kable(cor(raw_data[,c(1,2,4,5,6,7,8,9,10,11,14,15,16,18)]))
```

	BR	CO2	GDP	HCS_GDP	HCSPC	IMR	IU	LI	LE_F	LE_M	P_0_14	P_15_64	P_65_plus	PU
BR	1.0000	-0.6103	-0.5979	-0.4695	-0.7960	0.6860	-0.9084	0.5764	-0.8857	-0.8825	0.9622	-0.9116	-0.9049	-0.9536
CO2	-0.6103	1.0000	0.9588	0.3076	0.8106	-0.4150	0.6631	-0.2666	0.6101	0.6090	-0.6692	0.5828	0.6926	0.6415
GDP	-0.5979	0.9588	1.0000	0.3180	0.8763	-0.4155	0.6921	-0.2826	0.6131	0.6169	-0.6669	0.5798	0.6912	0.6358
HCS_GDP	-0.4695	0.3076	0.3180	1.0000	0.4825	-0.1727	0.5347	0.2973	0.3726	0.5129	-0.3698	0.2537	0.4683	0.3478
HCSPC	-0.7960	0.8106	0.8763	0.4825	1.0000	-0.5578	0.9324	-0.3882	0.7949	0.8182	-0.8567	0.7387	0.8950	0.8362
IMR	0.6860	-0.4150	-0.4155	-0.1727	-0.5578	1.0000	-0.6240	0.5903	-0.9028	-0.8692	0.6543	-0.5300	-0.7246	-0.7089
IU	-0.9084	0.6631	0.6921	0.5347	0.9324	-0.6240	1.0000	-0.4591	0.8619	0.8880	-0.9399	0.8284	0.9604	0.9307
LI	0.5764	-0.2666	-0.2826	0.2973	-0.3882	0.5903	-0.4591	1.0000	-0.5811	-0.4977	0.5767	-0.6116	-0.4610	-0.5969
LE_F	-0.8857	0.6101	0.6131	0.3726	0.7949	-0.9028	0.8619	-0.5811	1.0000	0.9766	-0.8623	0.7190	0.9305	0.8775
LE_M	-0.8825	0.6090	0.6169	0.5129	0.8182	-0.8692	0.8880	-0.4977	0.9766	1.0000	-0.8426	0.6857	0.9305	0.8635
P_0_14	0.9622	-0.6692	-0.6669	-0.3698	-0.8567	0.6543	-0.9399	0.5767	-0.8623	-0.8426	1.0000	-0.9549	-0.9310	-0.9877
P_15_64	-0.9116	0.5828	0.5798	0.2537	0.7387	-0.5300	0.8284	-0.6116	0.7190	0.6857	-0.9549	1.0000	0.7808	0.9370
P_65_plus	-0.9049	0.6926	0.6912	0.4683	0.8950	-0.7246	0.9604	-0.4610	0.9305	0.9305	-0.9310	0.7808	1.0000	0.9273
PU	-0.9536	0.6415	0.6358	0.3478	0.8362	-0.7089	0.9307	-0.5969	0.8775	0.8635	-0.9877	0.9370	0.9273	1.0000

Correlation Plot

```
par(mar=c(2,2,2,2))
corrplot::corrplot(cor(raw_data[,c(1,2,4,5,6,7,8,9,10,11,14,15,16,18)]), method = "circle",
    tl.cex = 0.5, tl.pos = "lt",
    tl.col = "dodgerblue")
```

Rsquare for Linear model with depedent variable as Pooulation

```
pop_fit = lm(Pop ~ BR+CO2+Cntry+HCSPC+IMR+IU+Mobile+P_15_64+PU, data = raw_data)
summary(pop_fit)$r.squared
```

```
## [1] 0.9994
```

Model Building

Life Expectancy for Males ANOVA

```
model_add_M = lm(LE_M~BR+CO2+Cntry+HCSPC+IMR+IU+Mobile+P_15_64+PU, data = raw_data)

model_add_all_M = lm(LE_M~BR+CO2+Cntry+log(GDP)+HCS_GDP+HCSPC+IMR+IU+LI+Mobile+P_15_64+P_65_plus+PU+TI+TO,data=
raw_data)

anova(model_add_M, model_add_all_M)
```

```
## Analysis of Variance Table
##
## Model 1: LE_M ~ BR + CO2 + Cntry + HCSPC + IMR + IU + Mobile + P_15_64 +
##     PU
## Model 2: LE_M ~ BR + CO2 + Cntry + log(GDP) + HCS_GDP + HCSPC + IMR +
##     IU + LI + Mobile + P_15_64 + P_65_plus + PU + TI + TO
##   Res.Df   RSS Df Sum of Sq    F   Pr(>F)
## 1     68 12.79
## 2     62  7.73  6      5.06 6.77 0.000015 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Life Expectancy for Females ANOVA

```
model_add_F = lm(LE_F~BR+CO2+Cntry+HCSPC+IMR+IU+Mobile+P_15_64+PU, data = raw_data)

model_add_all_F = lm(LE_F~BR+CO2+Cntry+log(GDP)+HCS_GDP+log(HCSPC)+IMR+IU+LI+Mobile+P_15_64+P_65_plus+PU+TI+TO,
data=raw_data)

anova(model_add_F, model_add_all_F)
```

```
## Analysis of Variance Table
##
## Model 1: LE_F ~ BR + CO2 + Cntry + HCSPC + IMR + IU + Mobile + P_15_64 +
##     PU
## Model 2: LE_F ~ BR + CO2 + Cntry + log(GDP) + HCS_GDP + log(HCSPC) + IMR +
##     IU + LI + Mobile + P_15_64 + P_65_plus + PU + TI + TO
##   Res.Df RSS Df Sum of Sq     F Pr(>F)
## 1     68 14.2
## 2     62 11.8  6      2.41 2.12  0.064 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Summary of Expolatory Interaction Models

```
model_int_M = lm(LE_M ~ Cntry + (BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI)^2 , data
 = raw_data)

model_int_F = lm(LE_F~Cntry+(BR+CO2+HCSPC+IMR+IU+Mobile+P_15_64+PU)^2 , data = raw_data)

summary(model_int_M)
```

```
##
## Call:
## lm(formula = LE_M ~ Cntry + (BR + log(GDP) + HCSPC + LI + Mobile +
##     P_15_64 + P_65_plus + PU + TI)^2, data = raw_data)
##
## Residuals:
##     Min      1Q Median      3Q     Max
## -0.4779 -0.1091  0.0160  0.0889  0.4753
##
## Coefficients:
```

```
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 3.82e+02   4.49e+02    0.85     0.40
## CntryFiji                  -5.42e+01   5.49e+01   -0.99     0.33
## CntryMicronesia, Fed. Sts. -3.79e+01   5.66e+01   -0.67     0.51
## CntryNew Zealand            1.43e+00   1.05e+01    0.14     0.89
## CntryPapua New Guinea      -4.28e+01   5.65e+01   -0.76     0.45
## CntrySamoa                 -3.50e+01   5.66e+01   -0.62     0.54
## CntrySolomon Islands       -3.67e+01   5.71e+01   -0.64     0.52
## CntryTonga                 -3.52e+01   5.55e+01   -0.63     0.53
## CntryVanuatu               -3.86e+01   5.69e+01   -0.68     0.50
## BR                          7.29e+02   4.84e+03    0.15     0.88
## log(GDP)                   -2.20e+01   2.28e+01   -0.96     0.34
## HCSPC                       3.78e-02   1.38e-01    0.27     0.79
## LI                          3.51e+02   6.39e+02    0.55     0.59
## Mobile                      1.96e+01   8.36e+01    0.23     0.82
## P_15_64                    -5.10e+02   8.31e+02   -0.61     0.54
## P_65_plus                  -1.16e+03   2.72e+03   -0.43     0.67
## PU                          4.42e+02   3.14e+02    1.41     0.17
## TI                          2.33e-08   3.88e-08    0.60     0.55
## BR:log(GDP)                 8.67e+01   1.63e+02    0.53     0.60
## BR:HCSPC                   -7.71e-01   1.13e+00   -0.68     0.50
## BR:LI                      -3.50e+03   3.81e+03   -0.92     0.37
## BR:Mobile                   6.36e+01   7.35e+02    0.09     0.93
## BR:P_15_64                 -4.60e+03   1.15e+04   -0.40     0.69
## BR:P_65_plus                9.29e+03   1.74e+04    0.53     0.60
## BR:PU                       1.75e+03   3.85e+03    0.46     0.65
## BR:TI                      -1.26e-08   3.43e-07   -0.04     0.97
## log(GDP):HCSPC              1.89e-04   2.91e-03    0.07     0.95
## log(GDP):LI                 6.88e-01   1.33e+01    0.05     0.96
## log(GDP):Mobile             3.70e-01   2.01e+00    0.18     0.86
## log(GDP):P_15_64            3.86e+01   3.77e+01    1.02     0.31
## log(GDP):P_65_plus         -1.72e+01   6.98e+01   -0.25     0.81
## log(GDP):PU                -9.98e+00   1.22e+01   -0.82     0.42
## log(GDP):TI                -2.81e-10   6.56e-10   -0.43     0.67
## HCSPC:LI                   -3.48e-02   6.74e-02   -0.52     0.61
## HCSPC:Mobile               -1.97e-03   8.95e-03   -0.22     0.83
## HCSPC:P_15_64              -2.65e-02   2.51e-01   -0.11     0.92
## HCSPC:P_65_plus             1.32e-01   2.53e-01    0.52     0.60
## HCSPC:PU                   -3.00e-02   6.64e-02   -0.45     0.65
## HCSPC:TI                   -1.03e-13   2.80e-13   -0.37     0.72
## LI:Mobile                  -2.60e+01   3.61e+01   -0.72     0.48
## LI:P_15_64                 -4.74e+02   1.35e+03   -0.35     0.73
## LI:P_65_plus                4.22e+01   2.39e+03    0.02     0.99
## LI:PU                       5.11e+01   2.73e+02    0.19     0.85
## LI:TI                       7.55e-09   7.25e-09    1.04     0.31
## Mobile:P_15_64             -5.60e+01   1.72e+02   -0.33     0.75
## Mobile:P_65_plus            4.11e+01   1.67e+02    0.25     0.81
## Mobile:PU                   1.76e+01   4.01e+01    0.44     0.66
## Mobile:TI                  -6.32e-10   1.36e-09   -0.47     0.64
## P_15_64:P_65_plus           2.31e+03   4.95e+03    0.47     0.64
## P_15_64:PU                 -3.62e+02   4.87e+02   -0.74     0.46
## P_15_64:TI                 -2.35e-08   8.98e-08   -0.26     0.80
## P_65_plus:PU               -6.03e+02   1.13e+03   -0.53     0.60
## P_65_plus:TI                5.22e-08   1.16e-07    0.45     0.66
## PU:TI                      -5.84e-09   2.33e-08   -0.25     0.80
##
## Residual standard error: 0.312 on 31 degrees of freedom
## Multiple R-squared:  0.999,  Adjusted R-squared:   0.997
## F-statistic:  623 on 53 and 31 DF,  p-value: <2e-16
```

```
summary(model_int_F)
```

```
##
## Call:
## lm(formula = LE_F ~ Cntry + (BR + CO2 + HCSPC + IMR + IU + Mobile +
##     P_15_64 + PU)^2, data = raw_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7005 -0.1393 -0.0135  0.1404  0.5853
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -1.90e+02   2.58e+02   -0.74   0.4662
## CntryFiji                  -7.80e+01   7.11e+01   -1.10   0.2787
## CntryMicronesia, Fed. Sts. -7.38e+01   7.71e+01   -0.96   0.3442
## CntryNew Zealand           -2.95e+01   4.05e+01   -0.73   0.4697
## CntryPapua New Guinea      -6.61e+01   7.85e+01   -0.84   0.4050
## CntrySamoa                 -6.30e+01   7.68e+01   -0.82   0.4174
## CntrySolomon Islands       -7.02e+01   7.76e+01   -0.90   0.3710
## CntryTonga                 -6.41e+01   7.66e+01   -0.84   0.4076
## CntryVanuatu               -6.91e+01   7.69e+01   -0.90   0.3738
## BR                          2.09e+03   6.95e+03    0.30   0.7655
## CO2                        -4.25e-03   2.99e-03   -1.42   0.1636
```

```
## HCSPC                       2.46e-01   1.27e-01    1.93   0.0602
## IMR                         7.33e+03   3.51e+03    2.09   0.0430 *
## IU                         -6.18e+02   2.20e+02   -2.82   0.0075 **
## Mobile                      2.21e+02   7.48e+01    2.96   0.0052 **
## P_15_64                     5.78e+02   6.00e+02    0.96   0.3412
## PU                          1.17e+02   1.47e+02    0.80   0.4302
## BR:CO2                     -5.26e-03   9.50e-03   -0.55   0.5831
## BR:HCSPC                   -1.69e+00   6.37e-01   -2.66   0.0113 *
## BR:IMR                     -1.81e+04   1.48e+04   -1.22   0.2280
## BR:IU                       2.54e+03   2.79e+03    0.91   0.3672
## BR:Mobile                  -1.86e+03   6.05e+02   -3.08   0.0038 **
## BR:P_15_64                 -4.21e+03   1.42e+04   -0.30   0.7679
## BR:PU                       5.13e+03   4.34e+03    1.18   0.2449
## CO2:HCSPC                   9.03e-09   1.01e-08    0.90   0.3751
## CO2:IMR                    -2.28e-03   3.81e-03   -0.60   0.5534
## CO2:IU                      3.61e-05   1.05e-04    0.34   0.7321
## CO2:Mobile                 -8.31e-05   4.41e-05   -1.88   0.0669 .
## CO2:P_15_64                 8.29e-03   5.79e-03    1.43   0.1600
## CO2:PU                     -1.48e-03   1.26e-03   -1.17   0.2474
## HCSPC:IMR                   9.82e-01   5.04e-01    1.95   0.0585 .
## HCSPC:IU                    2.27e-02   8.85e-03    2.57   0.0140 *
## HCSPC:Mobile                6.66e-03   7.11e-03    0.94   0.3550
## HCSPC:P_15_64              -3.99e-01   2.36e-01   -1.69   0.0990 .
## HCSPC:PU                    1.94e-02   3.74e-02    0.52   0.6060
## IMR:IU                     -1.97e+03   8.65e+02   -2.27   0.0284 *
## IMR:Mobile                  3.00e+02   2.22e+02    1.35   0.1845
## IMR:P_15_64                -1.23e+04   6.66e+03   -1.84   0.0730 .
## IMR:PU                      9.63e+02   1.37e+03    0.70   0.4855
## IU:Mobile                  -7.53e+01   2.74e+01   -2.75   0.0089 **
## IU:P_15_64                  1.12e+03   3.78e+02    2.97   0.0050 **
## IU:PU                      -1.71e+02   9.13e+01   -1.87   0.0682 .
## Mobile:P_15_64             -3.26e+02   1.31e+02   -2.49   0.0172 *
## Mobile:PU                   7.76e+01   3.94e+01    1.97   0.0559 .
## P_15_64:PU                 -4.02e+02   3.07e+02   -1.31   0.1978
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.346 on 40 degrees of freedom
## Multiple R-squared:  0.999, Adjusted R-squared:   0.997
## F-statistic:  682 on 44 and 40 DF, p-value: <2e-16
```

Summary of Polynomial Models

```
model_poly_M = lm(LE_M~ Cntry + (BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI)^2 +
I(BR^3)+I(log(GDP)^3)+I(HCSPC^3)+I(LI^3)+I(Mobile^3)+I(P_15_64^3)+I(P_65_plus^3)+I(PU^3) + I(TI^3) , data = raw
_data)

model_poly_F = lm(LE_F~ Cntry+(BR+CO2+HCSPC+IMR+IU+Mobile+P_15_64+PU)^2 +
I(BR^3)+I(CO2^3)+I(HCSPC^3)+I(IMR^3)+I(IU^3)+I(Mobile^3)+I(P_15_64^3) + I(PU^3) , data = raw_data)

summary(model_poly_M)
```

```
##
## Call:
## lm(formula = LE_M ~ Cntry + (BR + log(GDP) + HCSPC + LI + Mobile +
##     P_15_64 + P_65_plus + PU + TI)^2 + I(BR^3) + I(log(GDP)^3) +
##     I(HCSPC^3) + I(LI^3) + I(Mobile^3) + I(P_15_64^3) + I(P_65_plus^3) +
##     I(PU^3) + I(TI^3), data = raw_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4371 -0.0693  0.0024  0.0871  0.4033
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -7.89e+02   9.04e+02   -0.87    0.392
## CntryFiji                   -7.72e+01   1.39e+02   -0.56    0.584
## CntryMicronesia, Fed. Sts. -4.50e+01    1.50e+02   -0.30    0.767
## CntryNew Zealand             3.48e+00   3.32e+01    0.10    0.917
## CntryPapua New Guinea       -4.35e+01   1.50e+02   -0.29    0.774
## CntrySamoa                  -4.39e+01   1.51e+02   -0.29    0.774
## CntrySolomon Islands        -4.32e+01   1.50e+02   -0.29    0.777
## CntryTonga                  -4.32e+01   1.50e+02   -0.29    0.777
## CntryVanuatu                -4.59e+01   1.50e+02   -0.31    0.763
## BR                          -1.57e+04   1.28e+04   -1.23    0.233
## log(GDP)                     8.92e+01   7.65e+01    1.17    0.256
## HCSPC                       -1.75e-01   2.41e-01   -0.73    0.474
## LI                           9.74e+02   9.99e+02    0.98    0.340
## Mobile                      -2.01e+02   1.94e+02   -1.04    0.311
## P_15_64                      1.65e+03   1.81e+03    0.91    0.370
## P_65_plus                    1.17e+03   3.98e+03    0.29    0.772
## PU                          -6.32e+02   7.24e+02   -0.87    0.392
## TI                           6.39e-08   6.15e-08    1.04    0.310
## I(BR^3)                      7.35e+05   6.66e+05    1.10    0.282
## I(log(GDP)^3)                3.13e-02   1.56e-02    2.01    0.057 .
```

```
## I(HCSPC^3)                  9.45e-11   8.60e-10    0.11    0.913
## I(LI^3)                    -5.94e+02   6.94e+02   -0.86    0.401
## I(Mobile^3)                -2.50e+00   5.48e+00   -0.46    0.653
## I(P_15_64^3)                9.57e+02   1.82e+03    0.53    0.604
## I(P_65_plus^3)             -1.15e+04   3.32e+04   -0.35    0.733
## I(PU^3)                    -1.58e+02   3.70e+02   -0.43    0.673
## I(TI^3)                    -3.74e-31   2.47e-30   -0.15    0.881
## BR:log(GDP)                -5.20e+02   3.83e+02   -1.36    0.189
## BR:HCSPC                   -8.84e-01   2.08e+00   -0.42    0.676
## BR:LI                      -8.62e+03   6.00e+03   -1.43    0.165
## BR:Mobile                   6.35e+02   1.56e+03    0.41    0.689
## BR:P_15_64                  4.51e+04   3.10e+04    1.45    0.160
## BR:P_65_plus                1.79e+04   2.40e+04    0.75    0.462
## BR:PU                      -4.83e+03   5.44e+03   -0.89    0.384
## BR:TI                       1.40e-07   5.15e-07    0.27    0.787
## log(GDP):HCSPC             -2.97e-03   9.10e-03   -0.33    0.747
## log(GDP):LI                 1.89e+01   2.17e+01    0.87    0.393
## log(GDP):Mobile            -1.04e+00   3.95e+00   -0.26    0.795
## log(GDP):P_15_64           -2.02e+02   1.55e+02   -1.31    0.205
## log(GDP):P_65_plus         -3.07e+01   8.48e+01   -0.36    0.721
## log(GDP):PU                 9.40e+00   2.14e+01    0.44    0.665
## log(GDP):TI                -4.58e-11   1.87e-09   -0.02    0.981
## HCSPC:LI                   -1.22e-02   1.08e-01   -0.11    0.911
## HCSPC:Mobile                5.57e-03   1.57e-02    0.36    0.726
## HCSPC:P_15_64               4.75e-01   3.91e-01    1.22    0.237
## HCSPC:P_65_plus             3.83e-01   4.92e-01    0.78    0.444
## HCSPC:PU                   -1.22e-01   1.30e-01   -0.94    0.358
## HCSPC:TI                   -5.14e-14   1.23e-12   -0.04    0.967
## LI:Mobile                  -3.96e+01   5.91e+01   -0.67    0.510
## LI:P_15_64                 -1.85e+03   2.25e+03   -0.82    0.420
## LI:P_65_plus               -8.78e+02   3.44e+03   -0.25    0.801
## LI:PU                       9.77e+01   5.04e+02    0.19    0.848
## LI:TI                      -3.72e-09   1.89e-08   -0.20    0.846
## Mobile:P_15_64              3.74e+02   3.89e+02    0.96    0.347
## Mobile:P_65_plus            3.32e+02   2.50e+02    1.33    0.198
## Mobile:PU                  -9.09e+01   7.59e+01   -1.20    0.243
## Mobile:TI                   1.79e-09   2.89e-09    0.62    0.542
## P_15_64:P_65_plus          -1.32e+03   6.74e+03   -0.20    0.846
## P_15_64:PU                  1.33e+03   1.17e+03    1.14    0.268
## P_15_64:TI                 -7.52e-08   1.29e-07   -0.58    0.565
## P_65_plus:PU               -1.07e+03   1.73e+03   -0.62    0.544
## P_65_plus:TI                5.35e-08   1.89e-07    0.28    0.780
## PU:TI                      -2.50e-08   3.99e-08   -0.63    0.538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.32 on 22 degrees of freedom
## Multiple R-squared:  0.999, Adjusted R-squared:   0.997
## F-statistic:  507 on 62 and 22 DF, p-value: <2e-16
```

```
summary(model_poly_F)
```

```
##
## Call:
## lm(formula = LE_F ~ Cntry + (BR + CO2 + HCSPC + IMR + IU + Mobile +
##     P_15_64 + PU)^2 + I(BR^3) + I(CO2^3) + I(HCSPC^3) + I(IMR^3) +
##     I(IU^3) + I(Mobile^3) + I(P_15_64^3) + I(PU^3), data = raw_data)
##
## Residuals:
##     Min      1Q Median       3Q     Max
## -0.6090 -0.1060 -0.0016  0.1293  0.5457
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -1.27e+02   5.55e+02   -0.23   0.8207
## CntryFiji                  -1.85e+01   2.01e+02   -0.09   0.9273
## CntryMicronesia, Fed. Sts.  9.81e+00   2.07e+02    0.05   0.9625
## CntryNew Zealand           -8.10e+01   1.72e+02   -0.47   0.6415
## CntryPapua New Guinea       2.46e+01   2.07e+02    0.12   0.9062
## CntrySamoa                  2.06e+01   2.08e+02    0.10   0.9218
## CntrySolomon Islands        1.47e+01   2.07e+02    0.07   0.9440
## CntryTonga                  1.90e+01   2.08e+02    0.09   0.9278
## CntryVanuatu                1.43e+01   2.07e+02    0.07   0.9456
## BR                         -4.65e+03   9.26e+03   -0.50   0.6186
## CO2                        -2.70e-03   5.04e-03   -0.54   0.5956
## HCSPC                       1.90e-01   1.69e-01    1.12   0.2698
## IMR                         6.77e+03   4.95e+03    1.37   0.1811
## IU                         -6.27e+02   4.25e+02   -1.48   0.1498
## Mobile                      1.08e+02   1.20e+02    0.90   0.3767
## P_15_64                     3.63e+02   1.14e+03    0.32   0.7518
## PU                          7.22e+01   5.70e+02    0.13   0.9000
## I(BR^3)                    -1.66e+05   4.38e+05   -0.38   0.7074
## I(CO2^3)                   -2.31e-16   1.40e-15   -0.17   0.8695
## I(HCSPC^3)                 -4.94e-10   6.81e-10   -0.73   0.4730
```

```
## I(IMR^3)                    -2.53e+04   7.34e+04   -0.34   0.7328
## I(IU^3)                     -2.06e+01   1.31e+02   -0.16   0.8758
## I(Mobile^3)                 -1.42e+01   5.67e+00   -2.50   0.0177 *
## I(P_15_64^3)                -3.63e+02   1.21e+03   -0.30   0.7660
## I(PU^3)                      2.07e+02   3.38e+02    0.61   0.5454
## BR:CO2                      -1.10e-02   1.87e-02   -0.59   0.5616
## BR:HCSPC                    -1.16e+00   1.02e+00   -1.13   0.2650
## BR:IMR                      -2.14e+04   2.03e+04   -1.05   0.3006
## BR:IU                        2.09e+03   4.60e+03    0.45   0.6531
## BR:Mobile                   -2.55e+03   8.17e+02   -3.13   0.0037 **
## BR:P_15_64                   1.03e+04   1.88e+04    0.55   0.5856
## BR:PU                        7.49e+02   6.09e+03    0.12   0.9028
## CO2:HCSPC                    5.07e-08   6.57e-08    0.77   0.4463
## CO2:IMR                     -1.13e-02   6.86e-03   -1.65   0.1095
## CO2:IU                       1.74e-04   2.17e-04    0.80   0.4288
## CO2:Mobile                  -2.15e-04   1.63e-04   -1.31   0.1984
## CO2:P_15_64                  6.64e-03   1.04e-02    0.64   0.5281
## CO2:PU                      -2.12e-03   2.89e-03   -0.73   0.4689
## HCSPC:IMR                    1.40e+00   5.52e-01    2.53   0.0164 *
## HCSPC:IU                     2.52e-02   5.96e-02    0.42   0.6752
## HCSPC:Mobile                 2.98e-02   1.12e-02    2.65   0.0123 *
## HCSPC:P_15_64               -3.54e-01   3.00e-01   -1.18   0.2465
## HCSPC:PU                     2.12e-02   5.08e-02    0.42   0.6790
## IMR:IU                      -1.09e+03   1.14e+03   -0.95   0.3485
## IMR:Mobile                  -1.75e+02   3.19e+02   -0.55   0.5874
## IMR:P_15_64                 -1.13e+04   9.37e+03   -1.21   0.2367
## IMR:PU                       1.37e+03   2.51e+03    0.55   0.5886
## IU:Mobile                   -6.15e+01   4.45e+01   -1.38   0.1767
## IU:P_15_64                   1.10e+03   7.05e+02    1.56   0.1286
## IU:PU                       -1.42e+02   1.57e+02   -0.90   0.3733
## Mobile:P_15_64              -3.12e+01   2.26e+02   -0.14   0.8910
## Mobile:PU                   -2.84e+01   6.76e+01   -0.42   0.6769
## P_15_64:PU                  -9.42e+01   1.19e+03   -0.08   0.9376
## ---
## Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.344 on 32 degrees of freedom
## Multiple R-squared:   0.999, Adjusted R-squared:    0.997
## F-statistic:   584 on 52 and 32 DF, p-value: <2e-16
```

Model Analysis

Life Expectancy of Female AIC Track

```
addToDF = function(type, df_original, model){
df_1 = data.frame(type,toString(eval(summary(model)$call[[2]])),summary(model)  $r.squared[1],summary(model)$adj.
r.squared[1])
colnames_df = c("Type","Formula","R-Squared","Adjusted R-Squared")
names(df_1) = colnames_df
df_original = rbind(df_original, df_1)

df_original
}

aic_bwd_selected_F = step(model_add_F,trace=0)
results_DF = data.frame("bwd:AIC",toString(eval(summary(aic_bwd_selected_F)$call[[2]])),summary(aic_bwd_selecte
d_F)$r.squared[1],summary(aic_bwd_selected_F) $adj.r.squared[1])
colnames_df = c("Type","Formula","R-Squared","Adjusted R-Squared")
names(results_DF) = colnames_df

model_none_F = lm(LE_F~1,data=raw_data)
aic_fwd_selected_F = step(model_none_F,scope=LE_F ~ BR + CO2 + Cntry + HCSPC + IMR + IU + Mobile + P_15_64 + P
U, direction = "forward",trace=0)
results_DF = addToDF("fwd:AIC",results_DF,aic_fwd_selected_F)

int_aic_bwd_selected_F = step(model_int_F,trace=0)
results_DF = addToDF("bwd:AIC",results_DF,int_aic_bwd_selected_F)

int_aic_fwd_selected_F = step(model_none_F,scope=LE_F~Cntry+(BR+CO2+HCSPC+IMR+IU+Mobile+P_15_64+PU)^2, directio
n = "forward",trace=0)
results_DF = addToDF("fwd:AIC",results_DF,int_aic_fwd_selected_F)

kable(results_DF)
```

Type	Formula	R-Squared	Adjusted R-Squared
bwd:AIC	~, LE_F, BR + Cntry + HCSPC + IMR + Mobile + P_15_64 + PU	0.9961	0.9953
fwd:AIC	~, LE_F, Cntry + PU + IMR + HCSPC + BR + P_15_64 + Mobile	0.9961	0.9953
bwd:AIC	~, LE_F, Cntry + BR + CO2 + HCSPC + IMR + IU + Mobile + P_15_64 + PU + BR:HCSPC + BR:IMR + BR:IU + BR:Mobile + BR:PU + CO2:HCSPC + CO2:Mobile + CO2:P_15_64 + CO2:PU + HCSPC:IMR + HCSPC:IU + HCSPC:Mobile + HCSPC:P_15_64 + IMR:IU + IMR:Mobile + IMR:P_15_64 + IU:Mobile + IU:P_15_64 + IU:PU	0.9986	0.9975

	+ Mobile:P_15_64 + Mobile:PU + P_15_64:PU		
fwd:AIC	~, LE_F, Cntry + PU + IMR + HCSPC + BR + Mobile + PU:IMR + PU:BR + IMR:Mobile + IMR:BR + HCSPC:BR + PU:Mobile	0.9972	0.9963

Life Expectancy of Female SIC Track

```
sic_bwd_selected_F = step(model_add_F,k=log(nrow(raw_data)),trace=0)
results_DF = addToDF("bwd:BIC",results_DF,sic_bwd_selected_F)

model_none_F = lm(LE_F~1,data=raw_data)
sic_fwd_selected_F = step(model_none_F,scope=LE_F ~ BR + CO2 + Cntry + HCSPC + IMR + IU + Mobile + P_15_64 + P
U, direction = "forward",k=log(nrow(raw_data)),trace=0)
results_DF = addToDF("fwd:BIC",results_DF,sic_fwd_selected_F)

int_sic_bwd_selected_F = step(model_int_F,k=log(nrow(raw_data)),trace=0)
results_DF = addToDF("bwd:BIC",results_DF,int_sic_bwd_selected_F)

int_sic_fwd_selected_F = step(model_none_F,scope=LE_F~Cntry+(BR+CO2+HCSPC+IMR+IU+Mobile+P_15_64+PU)^2, directio
n = "forward",k=log(nrow(raw_data)),trace=0)
results_DF = addToDF("fwd:BIC",results_DF,int_sic_fwd_selected_F)

kable(results_DF)
```

Type	Formula	R-Squared	Adjusted R-Squared
bwd:AIC	~, LE_F, BR + Cntry + HCSPC + IMR + Mobile + P_15_64 + PU	0.9961	0.9953
fwd:AIC	~, LE_F, Cntry + PU + IMR + HCSPC + BR + P_15_64 + Mobile	0.9961	0.9953
bwd:AIC	~, LE_F, Cntry + BR + CO2 + HCSPC + IMR + IU + Mobile + P_15_64 + PU + BR:HCSPC + BR:IMR + BR:IU + BR:Mobile + BR:PU + CO2:HCSPC + CO2:Mobile + CO2:P_15_64 + CO2:PU + HCSPC:IMR + HCSPC:IU + HCSPC:Mobile + HCSPC:P_15_64 + IMR:IU + IMR:Mobile + IMR:P_15_64 + IU:Mobile + IU:P_15_64 + IU:PU + Mobile:P_15_64 + Mobile:PU + P_15_64:PU	0.9986	0.9975
fwd:AIC	~, LE_F, Cntry + PU + IMR + HCSPC + BR + Mobile + PU:IMR + PU:BR + IMR:Mobile + IMR:BR + HCSPC:BR + PU:Mobile	0.9972	0.9963
bwd:BIC	~, LE_F, BR + Cntry + HCSPC + IMR + P_15_64 + PU	0.9959	0.9952
fwd:BIC	~, LE_F, Cntry + PU + IMR + HCSPC	0.9956	0.9949
bwd:BIC	~, LE_F, Cntry + BR + CO2 + HCSPC + IMR + IU + Mobile + P_15_64 + PU + BR:HCSPC + BR:IMR + BR:Mobile + BR:PU + CO2:Mobile + CO2:P_15_64 + HCSPC:IMR + HCSPC:IU + HCSPC:Mobile + HCSPC:P_15_64 + IMR:IU + IMR:P_15_64 + IU:Mobile + IU:P_15_64 + IU:PU + Mobile:P_15_64 + Mobile:PU + P_15_64:PU	0.9984	0.9974
fwd:BIC	~, LE_F, Cntry + PU + IMR + HCSPC + PU:IMR	0.9958	0.9951

Life Expectancy of Male AIC Track

```
addToDF = function(type, df_original, model){
df_1 = data.frame(type,toString(eval(summary(model)$call[[2]])),summary(model)  $r.squared[1],summary(model)$adj.
r.squared[1])
colnames_df = c("Type","Formula","R-Squared","Adjusted R-Squared")
names(df_1) = colnames_df
df_original = rbind(df_original, df_1)

df_original
}

aic_bwd_selected_M = step(model_add_M,trace=0)
results_DF = data.frame("bwd:AIC",toString(eval(summary(aic_bwd_selected_M)$call[[2]])),summary(aic_bwd_selecte
d_M)$r.squared[1],summary(aic_bwd_selected_M) $adj.r.squared[1])
colnames_df = c("Type","Formula","R-Squared","Adjusted R-Squared")
names(results_DF) = colnames_df

model_none_M = lm(LE_M~1,data=raw_data)

aic_fwd_selected_M = step(model_none_M,scope=LE_M ~ BR + CO2 + Cntry +GDP + HCSPC + LI + Mobile + P_15_64 + P_6
5_plus + PU + TI, direction = "forward",trace=0)
results_DF = addToDF("fwd:AIC",results_DF,aic_fwd_selected_M)

int_aic_bwd_selected_M = step(model_int_M,trace=0)
results_DF = addToDF("bwd:AIC",results_DF,int_aic_bwd_selected_M)

int_aic_fwd_selected_M = step(model_none_M,scope=LE_M~Cntry+(BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_
65_plus + PU + TI)^2, direction = "forward",trace=0)
results_DF = addToDF("fwd:AIC",results_DF,int_aic_fwd_selected_M)

kable(results_DF)
```

Type	Formula	R-Squared	Adjusted R-Squared
bwd:AIC	~, LE_M, BR + CO2 + Cntry + HCSPC + IMR + Mobile + P_15_64 + PU	0.9960	0.9951
fwd:AIC	~, LE_M, Cntry + PU + LI + HCSPC + BR + P_15_64 + TI + P_65_plus	0.9971	0.9964
bwd:AIC	~, LE_M, Cntry + BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI + BR:HCSPC + BR:LI + BR:P_65_plus + log(GDP):P_15_64 + log(GDP):PU + HCSPC:LI + HCSPC:P_65_plus + HCSPC:PU + HCSPC:TI + LI:Mobile + LI:TI + Mobile:P_15_64 + Mobile:PU + P_15_64:P_65_plus + P_15_64:PU + P_15_64:TI + P_65_plus:PU	0.9990	0.9983
fwd:AIC	~, LE_M, Cntry + log(GDP) + PU + LI + P_15_64 + HCSPC + BR + TI + P_65_plus + LI:P_15_64 + HCSPC:BR + PU:BR + PU:HCSPC + log(GDP):BR + log(GDP):PU + BR:TI + P_15_64:TI	0.9985	0.9980

Life Expectancy of Male SIC Track

```
sic_bwd_selected_M = step(model_add_M,k=log(nrow(raw_data)),trace=0)
results_DF = addToDF("bwd:BIC",results_DF,sic_bwd_selected_M)

model_none_M = lm(LE_M~1,data=raw_data)
sic_fwd_selected_M = step(model_none_M,scope=LE_M ~ BR + CO2 + Cntry +GDP + HCSPC + LI + Mobile + P_15_64 + P_6
5_plus + PU + TI, direction = "forward",k=log(nrow(raw_data)),trace=0)
results_DF = addToDF("fwd:BIC",results_DF,sic_fwd_selected_M)

int_sic_bwd_selected_M = step(model_int_M,k=log(nrow(raw_data)),trace=0)
results_DF = addToDF("bwd:BIC",results_DF,int_sic_bwd_selected_M)

int_sic_fwd_selected_M = step(model_none_M,scope=LE_M~ Cntry+(BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P
_65_plus + PU + TI)^2, direction = "forward",k=log(nrow(raw_data)),trace=0)
results_DF = addToDF("fwd:BIC",results_DF,int_sic_fwd_selected_M)

kable(results_DF)
```

Type	Formula	R-Squared	Adjusted R-Squared
bwd:AIC	~, LE_M, BR + CO2 + Cntry + HCSPC + IMR + Mobile + P_15_64 + PU	0.9960	0.9951
fwd:AIC	~, LE_M, Cntry + PU + LI + HCSPC + BR + P_15_64 + TI + P_65_plus	0.9971	0.9964
bwd:AIC	~, LE_M, Cntry + BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI + BR:HCSPC + BR:LI + BR:P_65_plus + log(GDP):P_15_64 + log(GDP):PU + HCSPC:LI + HCSPC:P_65_plus + HCSPC:PU + HCSPC:TI + LI:Mobile + LI:TI + Mobile:P_15_64 + Mobile:PU + P_15_64:P_65_plus + P_15_64:PU + P_15_64:TI + P_65_plus:PU	0.9990	0.9983
fwd:AIC	~, LE_M, Cntry + log(GDP) + PU + LI + P_15_64 + HCSPC + BR + TI + P_65_plus + LI:P_15_64 + HCSPC:BR + PU:BR + PU:HCSPC + log(GDP):BR + log(GDP):PU + BR:TI + P_15_64:TI	0.9985	0.9980
bwd:BIC	~, LE_M, BR + Cntry + HCSPC + P_15_64 + PU	0.9955	0.9947
fwd:BIC	~, LE_M, Cntry + PU + LI + HCSPC + BR + P_15_64 + TI + P_65_plus	0.9971	0.9964
bwd:BIC	~, LE_M, Cntry + BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI + BR:HCSPC + BR:LI + BR:P_65_plus + log(GDP):P_15_64 + log(GDP):PU + HCSPC:LI + HCSPC:PU + LI:Mobile + LI:TI + Mobile:P_15_64 + Mobile:PU + P_15_64:P_65_plus + P_15_64:PU + P_15_64:TI	0.9989	0.9983
fwd:BIC	~, LE_M, Cntry + log(GDP) + PU + LI + P_15_64 + HCSPC + BR + TI + P_65_plus + LI:P_15_64	0.9975	0.9968

Polynomial Model Selection based on Life Expectancy of Female

```
poly_aic_bwd_selected_F = step(model_poly_F,trace=0)
results_DF = addToDF("bwd:AIC",results_DF,poly_aic_bwd_selected_F)

poly_aic_fwd_selected_F = step(model_none_F,scope=formula(model_poly_F), direction = "forward",trace=0)
results_DF = addToDF("fwd:AIC",results_DF,poly_aic_fwd_selected_F)

poly_sic_bwd_selected_F = step(model_none_F, scope=formula(model_poly_F),k=log(nrow(raw_data)), trace=0)
results_DF = addToDF("bwd:BIC",results_DF,poly_sic_bwd_selected_F)

poly_sic_fwd_selected_F = step(model_none_F,scope=formula(model_poly_F), direction = "forward",k=log(nrow(raw_d
ata)),trace=0)
results_DF = addToDF("fwd:BIC",results_DF,poly_sic_fwd_selected_F)

kable(results_DF)
```

Type	Formula	R-Squared	Adjusted R-Squared
bwd:AIC	~, LE_M, BR + CO2 + Cntry + HCSPC + IMR + Mobile + P_15_64 + PU	0.9960	0.9951
fwd:AIC	~, LE_M, Cntry + PU + LI + HCSPC + BR + P_15_64 + TI + P_65_plus	0.9971	0.9964

bwd:AIC	~, LE_M, Cntry + BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI + BR:HCSPC + BR:LI + BR:P_65_plus + log(GDP):P_15_64 + log(GDP):PU + HCSPC:LI + HCSPC:P_65_plus + HCSPC:PU + HCSPC:TI + LI:Mobile + LI:TI + Mobile:P_15_64 + Mobile:PU + P_15_64:P_65_plus + P_15_64:PU + P_15_64:TI + P_65_plus:PU	0.9990	0.9983
fwd:AIC	~, LE_M, Cntry + log(GDP) + PU + LI + P_15_64 + HCSPC + BR + TI + P_65_plus + LI:P_15_64 + HCSPC:BR + PU:BR + PU:HCSPC + log(GDP):BR + log(GDP):PU + BR:TI + P_15_64:TI	0.9985	0.9980
bwd:BIC	~, LE_M, BR + Cntry + HCSPC + P_15_64 + PU	0.9955	0.9947
fwd:BIC	~, LE_M, Cntry + PU + LI + HCSPC + BR + P_15_64 + TI + P_65_plus	0.9971	0.9964
bwd:BIC	~, LE_M, Cntry + BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI + BR:HCSPC + BR:LI + BR:P_65_plus + log(GDP):P_15_64 + log(GDP):PU + HCSPC:LI + HCSPC:PU + LI:Mobile + LI:TI + Mobile:P_15_64 + Mobile:PU + P_15_64:P_65_plus + P_15_64:PU + P_15_64:TI	0.9989	0.9983
fwd:BIC	~, LE_M, Cntry + log(GDP) + PU + LI + P_15_64 + HCSPC + BR + TI + P_65_plus + LI:P_15_64	0.9975	0.9968
bwd:AIC	~, LE_F, Cntry + BR + CO2 + HCSPC + IMR + IU + Mobile + P_15_64 + PU + I(HCSPC^3) + I(Mobile^3) + I(P_15_64^3) + BR:HCSPC + BR:IMR + BR:IU + BR:Mobile + BR:P_15_64 + CO2:HCSPC + CO2:IMR + CO2:IU + CO2:Mobile + CO2:PU + HCSPC:IMR + HCSPC:IU + HCSPC:Mobile + HCSPC:P_15_64 + IMR:IU + IMR:P_15_64 + IMR:PU + IU:Mobile + IU:P_15_64 + IU:PU + Mobile:PU	0.9988	0.9978
fwd:AIC	~, LE_F, Cntry + I(BR^3) + IU + PU + BR + Mobile + IMR + P_15_64 + IU:BR + Mobile:IMR + PU:IMR + IU:P_15_64	0.9975	0.9968
bwd:BIC	~, LE_F, Cntry + I(BR^3) + PU + BR + Mobile + I(IU^3)	0.9967	0.9961
fwd:BIC	~, LE_F, Cntry + I(BR^3) + IU + PU + BR + Mobile	0.9967	0.9961

Polynomial Model Selection based on Life Expectancy of Male

```
poly_aic_bwd_selected_M = step(model_poly_M,trace=0)
results_DF = addToDF("bwd:AIC",results_DF,poly_aic_bwd_selected_M)

poly_aic_fwd_selected_M = step(model_none_M,scope=formula(model_poly_M), direction = "forward",trace=0)
results_DF = addToDF("fwd:AIC",results_DF,poly_aic_fwd_selected_M)

poly_sic_bwd_selected_M = step(model_none_M, scope=formula(model_poly_M),k=log(nrow(raw_data)), trace=0)
results_DF = addToDF("bwd:BIC",results_DF,poly_sic_bwd_selected_M)

poly_sic_fwd_selected_M = step(model_none_M,scope=formula(model_poly_M), direction = "forward",k=log(nrow(raw_d
ata)),trace=0)
results_DF = addToDF("fwd:BIC",results_DF,poly_sic_fwd_selected_M)

kable(results_DF)
```

Type	Formula	R-Squared	Adjusted R-Squared
bwd:AIC	~, LE_M, BR + CO2 + Cntry + HCSPC + IMR + Mobile + P_15_64 + PU	0.9960	0.9951
fwd:AIC	~, LE_M, Cntry + PU + LI + HCSPC + BR + P_15_64 + TI + P_65_plus	0.9971	0.9964
bwd:AIC	~, LE_M, Cntry + BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI + BR:HCSPC + BR:LI + BR:P_65_plus + log(GDP):P_15_64 + log(GDP):PU + HCSPC:LI + HCSPC:P_65_plus + HCSPC:PU + HCSPC:TI + LI:Mobile + LI:TI + Mobile:P_15_64 + Mobile:PU + P_15_64:P_65_plus + P_15_64:PU + P_15_64:TI + P_65_plus:PU	0.9990	0.9983
fwd:AIC	~, LE_M, Cntry + log(GDP) + PU + LI + P_15_64 + HCSPC + BR + TI + P_65_plus + LI:P_15_64 + HCSPC:BR + PU:BR + PU:HCSPC + log(GDP):BR + log(GDP):PU + BR:TI + P_15_64:TI	0.9985	0.9980
bwd:BIC	~, LE_M, BR + Cntry + HCSPC + P_15_64 + PU	0.9955	0.9947
fwd:BIC	~, LE_M, Cntry + PU + LI + HCSPC + BR + P_15_64 + TI + P_65_plus	0.9971	0.9964
bwd:BIC	~, LE_M, Cntry + BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI + BR:HCSPC + BR:LI + BR:P_65_plus + log(GDP):P_15_64 + log(GDP):PU + HCSPC:LI + HCSPC:PU + LI:Mobile + LI:TI + Mobile:P_15_64 + Mobile:PU + P_15_64:P_65_plus + P_15_64:PU + P_15_64:TI	0.9989	0.9983
fwd:BIC	~, LE_M, Cntry + log(GDP) + PU + LI + P_15_64 + HCSPC + BR + TI + P_65_plus + LI:P_15_64	0.9975	0.9968
bwd:AIC	~, LE_F, Cntry + BR + CO2 + HCSPC + IMR + IU + Mobile + P_15_64 + PU + I(HCSPC^3) + I(Mobile^3) + I(P_15_64^3) + BR:HCSPC + BR:IMR + BR:IU + BR:Mobile + BR:P_15_64 + CO2:HCSPC + CO2:IMR + CO2:IU + CO2:Mobile + CO2:PU + HCSPC:IMR + HCSPC:IU + HCSPC:Mobile + HCSPC:P_15_64 + IMR:IU + IMR:P_15_64 + IMR:PU + IU:Mobile + IU:P_15_64 + IU:PU + Mobile:PU	0.9988	0.9978
fwd:AIC	~, LE_F, Cntry + I(BR^3) + IU + PU + BR + Mobile + IMR + P_15_64 + IU:BR + Mobile:IMR + PU:IMR + IU:P_15_64	0.9975	0.9968
bwd:BIC	~, LE_F, Cntry + I(BR^3) + PU + BR + Mobile + I(IU^3)	0.9967	0.9961
fwd:BIC	~, LE_F, Cntry + I(BR^3) + IU + PU + BR + Mobile	0.9967	0.9961
bwd:AIC	~, LE_M, Cntry + BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI + I(BR^3) + I(log(GDP)^3) + I(HCSPC^3) + I(LI^3) + I(P_15_64^3) + BR:log(GDP) + BR:HCSPC + BR:LI + BR:P_15_64 + BR:P_65_plus + BR:PU + log(GDP):HCSPC + log(GDP):LI + log(GDP):P_15_64 + HCSPC:P_15_64 +	0.9993	0.9983

	HCSPC:P_65_plus + HCSPC:PU + LI:Mobile + LI:P_15_64 + LI:P_65_plus + LI:TI + Mobile:P_15_64 + Mobile:P_65_plus + Mobile:PU + Mobile:TI + P_15_64:PU + P_15_64:TI + P_65_plus:PU + PU:TI		
fwd:AIC	~, LE_M, Cntry + log(GDP) + I(BR^3) + HCSPC + PU + I(PU^3) + I(log(GDP)^3) + LI + I(TI^3) + BR + I(P_15_64^3) + log(GDP):PU + HCSPC:BR + HCSPC:PU + log(GDP):LI	0.9988	0.9983
bwd:BIC	~, LE_M, Cntry + log(GDP) + I(BR^3) + PU + I(PU^3) + I(log(GDP)^3) + LI	0.9977	0.9972
fwd:BIC	~, LE_M, Cntry + log(GDP) + I(BR^3) + HCSPC + PU + I(PU^3) + I(log(GDP)^3) + LI	0.9978	0.9973

Model Selection

Validating LINE on the selected models

```
diagnostics = function(model, pcol = "grey", lcol = "dodgerblue", alpha = 0.05,
                       plotit = TRUE, testit = TRUE) {

  if (plotit == TRUE) {

    # side-by-side plots (one row, two columns)
    par(mfrow = c(1, 2))

    # fitted versus residuals
    plot(fitted(model), resid(model),
         col = pcol, pch = 20, cex = 1.5,
         xlab = "Fitted", ylab = "Residuals",
         main = "Fitted versus Residuals")
    abline(h = 0, col = lcol, lwd = 2)
    grid()

    # qq-plot
    qqnorm(resid(model), col = pcol, pch = 20, cex = 1.5)
    qqline(resid(model), col = lcol, lwd = 2)
    grid()
  }

  if (testit == TRUE) {
    # p-value and decision
    p_val = shapiro.test(resid(model))$p.value
    decision = ifelse(p_val < alpha, "Shapiro-Wilk Test : Reject", "Shapiro-Wilk Test : Fail to Reject")

    list(p_val = p_val, decision = decision)
  }

}

diagnostics(aic_bwd_selected_F,plotit = FALSE)
```

```
## $p_val
## [1] 0.3492
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```
diagnostics(aic_bwd_selected_F,testit = FALSE)
```

```
diagnostics(sic_bwd_selected_F,plotit = FALSE)
```

```
## $p_val
## [1] 0.25
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```
diagnostics(sic_bwd_selected_F,testit = FALSE)
```

```
diagnostics(int_aic_bwd_selected_F,plotit = FALSE)
```

```
## $p_val
## [1] 0.4367
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```
diagnostics(int_aic_bwd_selected_F,testit = FALSE)
```

```
diagnostics(int_sic_fwd_selected_F,plotit = FALSE)
```

```
## $p_val
## [1] 0.3219
##
## $decision
```

```
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```
diagnostics(int_sic_fwd_selected_F,testit = FALSE)
```

```
diagnostics(poly_aic_bwd_selected_F,plotit = FALSE)
```

```
## $p_val
## [1] 0.02304
##
## $decision
## [1] "Shapiro-Wilk Test : Reject"
```

```
diagnostics(poly_aic_bwd_selected_F,testit = FALSE)
```

```
diagnostics(poly_aic_fwd_selected_F,plotit = FALSE)
```

```
## $p_val
## [1] 0.6133
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```
diagnostics(poly_aic_fwd_selected_F,testit = FALSE)
```

```r
diagnostics(poly_sic_bwd_selected_F,plotit = FALSE)
```

```
## $p_val
## [1] 0.9046
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```r
diagnostics(poly_sic_bwd_selected_F,testit = FALSE)
```

```r
diagnostics(poly_sic_fwd_selected_F,plotit = FALSE)
```

```
## $p_val
## [1] 0.8665
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```r
diagnostics(poly_sic_fwd_selected_F,testit = FALSE)
```

```
diagnostics(aic_bwd_selected_M,plotit = FALSE)
```

```
## $p_val
## [1] 0.3129
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```
diagnostics(aic_bwd_selected_M,testit = FALSE)
```

```
diagnostics(sic_bwd_selected_M,plotit = FALSE)
```

```
## $p_val
## [1] 0.3285
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```
diagnostics(sic_bwd_selected_M,testit = FALSE)
```

```r
diagnostics(int_aic_bwd_selected_M,plotit = FALSE)
```

```
## $p_val
## [1] 0.871
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```r
diagnostics(int_aic_bwd_selected_M,testit = FALSE)
```

```r
diagnostics(int_sic_fwd_selected_M,plotit = FALSE)
```

```
## $p_val
## [1] 0.09005
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```r
diagnostics(int_sic_fwd_selected_M,testit = FALSE)
```

```
diagnostics(poly_aic_bwd_selected_M,plotit = FALSE)
```

```
## $p_val
## [1] 0.1632
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```
diagnostics(poly_aic_bwd_selected_M,testit = FALSE)
```

```
diagnostics(poly_aic_fwd_selected_M,plotit = FALSE)
```

```
## $p_val
## [1] 0.2055
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```
diagnostics(poly_aic_fwd_selected_M,testit = FALSE)
```

```
diagnostics(poly_sic_bwd_selected_M,plotit = FALSE)
```

```
## $p_val
## [1] 0.3621
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```
diagnostics(poly_sic_bwd_selected_M,testit = FALSE)
```

```
diagnostics(poly_sic_fwd_selected_M,plotit = FALSE)
```

```
## $p_val
## [1] 0.1786
##
## $decision
## [1] "Shapiro-Wilk Test : Fail to Reject"
```

```
diagnostics(poly_sic_fwd_selected_M,testit = FALSE)
```

LOOCV And Adjusted R^Square for Selecting Best Suited Model

```
get_loocv_rmse = function(model) {
  sqrt(mean((resid(model) / (1 - hatvalues(model))) ^ 2))
}

get_adj_r2 = function(model) {
  summary(model)$adj.r.squared
}
```

```
finalDF = function(type,df_original, model){
df_1 = data.frame(type,toString(eval(summary(model)$call[[2]])),round(get_loocv_rmse(model),5),round(get_adj_r2
(model),5))
colnames_df = c("From","Model","LOOCV-RMSE","Adjusted R-Squared")
names(df_1) = colnames_df
df_original = rbind(df_original, df_1)

df_original
}

final_DF = c("Step:AIC","Model","0.0","0.0")
colnames_df = c("From","Model","LOOCV-RMSE","Adjusted R-Squared")
names(final_DF) = colnames_df

final_DF = finalDF("bwd_add:AIC",final_DF,aic_bwd_selected_F)
final_DF = finalDF("bwd_int:AIC",final_DF,int_aic_bwd_selected_F)
final_DF = finalDF("fwd_add:AIC",final_DF,aic_fwd_selected_F)
final_DF = finalDF("fwd_int:AIC",final_DF,int_aic_fwd_selected_F)
final_DF = finalDF("bwd_add:SIC",final_DF,sic_bwd_selected_F)
final_DF = finalDF("bwd_int:SIC",final_DF,int_sic_bwd_selected_F)
final_DF = finalDF("fwd_add:SIC",final_DF,sic_fwd_selected_F)
final_DF = finalDF("fwd_int:SIC",final_DF,int_sic_fwd_selected_F)
final_DF = finalDF("bwd_poly:AIC",final_DF,poly_aic_bwd_selected_F)
final_DF = finalDF("fwd_poly:AIC",final_DF,poly_aic_fwd_selected_F)
final_DF = finalDF("bwd_poly:SIC",final_DF,poly_sic_bwd_selected_F)
final_DF = finalDF("fwd_poly:SIC",final_DF,poly_sic_fwd_selected_F)

final_DF = final_DF[-1,]

rownames(final_DF) = NULL

kable(final_DF)
```

From	Model	LOOCV-RMSE	Adjusted R-Squared
bwd_add:AIC	~, LE_F, BR + Cntry + HCSPC + IMR + Mobile + P_15_64 + PU	0.5148	0.99527
bwd_int:AIC	~, LE_F, Cntry + BR + CO2 + HCSPC + IMR + IU + Mobile + P_15_64 + PU + BR:HCSPC + BR:IMR + BR:IU + BR:Mobile + BR:PU + CO2:HCSPC + CO2:Mobile + CO2:P_15_64 + CO2:PU + HCSPC:IMR + HCSPC:IU + HCSPC:Mobile + HCSPC:P_15_64 + IMR:IU + IMR:Mobile + IMR:P_15_64 + IU:Mobile + IU:P_15_64 + IU:PU + Mobile:P_15_64 + Mobile:PU + P_15_64:PU	0.52391	0.99747
fwd_add:AIC	~, LE_F, Cntry + PU + IMR + HCSPC + BR + P_15_64 + Mobile	0.5148	0.99527
fwd_int:AIC	~, LE_F, Cntry + PU + IMR + HCSPC + BR + Mobile + PU:IMR + PU:BR + IMR:Mobile + IMR:BR + HCSPC:BR + PU:Mobile	0.46352	0.99634
bwd_add:SIC	~, LE_F, BR + Cntry + HCSPC + IMR + P_15_64 + PU	0.5175	0.99518
bwd_int:SIC	~, LE_F, Cntry + BR + CO2 + HCSPC + IMR + IU + Mobile + P_15_64 + PU + BR:HCSPC + BR:IMR + BR:Mobile + BR:PU + CO2:Mobile + CO2:P_15_64 + HCSPC:IMR + HCSPC:IU + HCSPC:Mobile + HCSPC:P_15_64 + IMR:IU + IMR:P_15_64 + IU:Mobile + IU:P_15_64 + IU:PU + Mobile:P_15_64 + Mobile:PU + P_15_64:PU	0.47245	0.99736
fwd_add:SIC	~, LE_F, Cntry + PU + IMR + HCSPC	0.51814	0.99488
fwd_int:SIC	~, LE_F, Cntry + PU + IMR + HCSPC + PU:IMR	0.50697	0.99513
bwd_poly:AIC	~, LE_F, Cntry + BR + CO2 + HCSPC + IMR + IU + Mobile + P_15_64 + PU + I(HCSPC^3) + I(Mobile^3) + I(P_15_64^3) + BR:HCSPC + BR:IMR + BR:IU + BR:Mobile + BR:P_15_64 + CO2:HCSPC + CO2:IMR + CO2:IU + CO2:Mobile + CO2:PU + HCSPC:IMR + HCSPC:IU + HCSPC:Mobile + HCSPC:P_15_64 + IMR:IU + IMR:P_15_64 + IMR:PU + IU:Mobile + IU:P_15_64 + IU:PU + Mobile:PU	0.63131	0.9978
fwd_poly:AIC	~, LE_F, Cntry + I(BR^3) + IU + PU + BR + Mobile + IMR + P_15_64 + IU:BR + Mobile:IMR + PU:IMR + IU:P_15_64	0.42879	0.99676
bwd_poly:SIC	~, LE_F, Cntry + I(BR^3) + PU + BR + Mobile + I(IU^3)	0.44802	0.99614
fwd_poly:SIC	~, LE_F, Cntry + I(BR^3) + IU + PU + BR + Mobile	0.45067	0.99606

```
final_DF_1 = c("Step:AIC","Model","0.0","0.0")
colnames_df_1 = c("From","Model","LOOCV-RMSE","Adjusted R-Squared")
names(final_DF_1) = colnames_df_1

final_DF_1 = finalDF("bwd_add:AIC",final_DF_1,aic_bwd_selected_M)
final_DF_1 = finalDF("bwd_int:AIC",final_DF_1,int_aic_bwd_selected_M)
final_DF_1 = finalDF("fwd_add:AIC",final_DF_1,aic_fwd_selected_M)
final_DF_1 = finalDF("fwd_int:AIC",final_DF_1,int_aic_fwd_selected_M)
final_DF_1 = finalDF("bwd_add:SIC",final_DF_1,sic_bwd_selected_M)
final_DF_1 = finalDF("bwd_int:SIC",final_DF_1,int_sic_bwd_selected_M)
final_DF_1 = finalDF("fwd_add:SIC",final_DF_1,sic_fwd_selected_M)
final_DF_1 = finalDF("fwd_int:SIC",final_DF_1,int_sic_fwd_selected_M)
final_DF_1 = finalDF("bwd_poly:AIC",final_DF_1,poly_aic_bwd_selected_M)
```

```
final_DF_1 = finalDF("fwd_poly:AIC",final_DF_1,poly_aic_fwd_selected_M)
final_DF_1 = finalDF("bwd_poly:SIC",final_DF_1,poly_sic_bwd_selected_M)
final_DF_1 = finalDF("fwd_poly:SIC",final_DF_1,poly_sic_fwd_selected_M)

final_DF_1 = final_DF_1[-1,]

rownames(final_DF_1) = NULL

kable(final_DF_1)
```

From	Model	LOOCV-RMSE	Adjusted R-Squared
bwd_add:AIC	~, LE_M, BR + CO2 + Cntry + HCSPC + IMR + Mobile + P_15_64 + PU	0.48325	0.99514
bwd_int:AIC	~, LE_M, Cntry + BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI + BR:HCSPC + BR:LI + BR:P_65_plus + log(GDP):P_15_64 + log(GDP):PU + HCSPC:LI + HCSPC:P_65_plus + HCSPC:PU + HCSPC:TI + LI:Mobile + LI:TI + Mobile:P_15_64 + Mobile:PU + P_15_64:P_65_plus + P_15_64:PU + P_15_64:TI + P_65_plus:PU	0.35238	0.9983
fwd_add:AIC	~, LE_M, Cntry + PU + LI + HCSPC + BR + P_15_64 + TI + P_65_plus	0.40716	0.99644
fwd_int:AIC	~, LE_M, Cntry + log(GDP) + PU + LI + P_15_64 + HCSPC + BR + TI + P_65_plus + LI:P_15_64 + HCSPC:BR + PU:BR + PU:HCSPC + log(GDP):BR + log(GDP):PU + BR:TI + P_15_64:TI	0.33136	0.99797
bwd_add:SIC	~, LE_M, BR + Cntry + HCSPC + P_15_64 + PU	0.48758	0.9947
bwd_int:SIC	~, LE_M, Cntry + BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI + BR:HCSPC + BR:LI + BR:P_65_plus + log(GDP):P_15_64 + log(GDP):PU + HCSPC:LI + HCSPC:PU + LI:Mobile + LI:TI + Mobile:P_15_64 + Mobile:PU + P_15_64:P_65_plus + P_15_64:PU + P_15_64:TI	0.3621	0.99833
fwd_add:SIC	~, LE_M, Cntry + PU + LI + HCSPC + BR + P_15_64 + TI + P_65_plus	0.40716	0.99644
fwd_int:SIC	~, LE_M, Cntry + log(GDP) + PU + LI + P_15_64 + HCSPC + BR + TI + P_65_plus + LI:P_15_64	0.38637	0.99682
bwd_poly:AIC	~, LE_M, Cntry + BR + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI + I(BR^3) + I(log(GDP)^3) + I(HCSPC^3) + I(LI^3) + I(P_15_64^3) + BR:log(GDP) + BR:HCSPC + BR:LI + BR:P_15_64 + BR:P_65_plus + BR:PU + log(GDP):HCSPC + log(GDP):LI + log(GDP):P_15_64 + HCSPC:P_15_64 + HCSPC:P_65_plus + HCSPC:PU + LI:Mobile + LI:P_15_64 + LI:P_65_plus + LI:TI + Mobile:P_15_64 + Mobile:P_65_plus + Mobile:PU + Mobile:TI + P_15_64:PU + P_15_64:TI + P_65_plus:PU + PU:TI	0.37296	0.99835
fwd_poly:AIC	~, LE_M, Cntry + log(GDP) + I(BR^3) + HCSPC + PU + I(PU^3) + I(log(GDP)^3) + LI + I(TI^3) + BR + I(P_15_64^3) + log(GDP):PU + HCSPC:BR + HCSPC:PU + log(GDP):LI	0.29011	0.99832
bwd_poly:SIC	~, LE_M, Cntry + log(GDP) + I(BR^3) + PU + I(PU^3) + I(log(GDP)^3) + LI	0.35713	0.99725
fwd_poly:SIC	~, LE_M, Cntry + log(GDP) + I(BR^3) + HCSPC + PU + I(PU^3) + I(log(GDP)^3) + LI	0.35669	0.99728

Summary Statistics for Final Model

```
final_model_M = lm(LE_M ~ BR + Cntry + log(GDP) + HCSPC + LI + Mobile + P_15_64 + P_65_plus + PU + TI, data = r
aw_data)
final_model_F = lm(LE_F ~ Cntry + PU + IMR + HCSPC + PU:IMR, data = raw_data)
```

Female

```
summary(final_model_F)
```

```
##
## Call:
## lm(formula = LE_F ~ Cntry + PU + IMR + HCSPC + PU:IMR, data = raw_data)
##
## Residuals:
##     Min      1Q Median      3Q     Max
## -1.3635 -0.2753  0.0522  0.2995  0.8372
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               3.59e+01   8.58e+00    4.18  8.1e-05 ***
## CntryFiji                 1.16e+01   4.87e+00    2.39  0.01945 *
## CntryMicronesia, Fed. Sts. 3.25e+01  6.32e+00    5.14  2.2e-06 ***
## CntryNew Zealand          2.44e-01   3.40e-01    0.72  0.47454
## CntryPapua New Guinea     4.15e+01   6.14e+00    6.77  2.9e-09 ***
## CntrySamoa                3.32e+01   6.23e+00    5.33  1.1e-06 ***
## CntrySolomon Islands      3.18e+01   6.35e+00    5.00  3.9e-06 ***
## CntryTonga                3.06e+01   6.08e+00    5.03  3.4e-06 ***
## CntryVanuatu              2.91e+01   6.10e+00    4.77  9.3e-06 ***
## PU                        4.98e+01   1.13e+01    4.42  3.5e-05 ***
## IMR                      -5.04e+02   8.94e+01   -5.63  3.2e-07 ***
## HCSPC                     6.78e-04   1.82e-04    3.72  0.00039 ***
## PU:IMR                    8.94e+02   4.13e+02    2.17  0.03365 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.458 on 72 degrees of freedom
## Multiple R-squared:  0.996, Adjusted R-squared:   0.995
## F-statistic: 1.43e+03 on 12 and 72 DF, p-value: <2e-16
```

Male

```
summary(final_model_M)
```

```
##
## Call:
## lm(formula = LE_M ~ BR + Cntry + log(GDP) + HCSPC + LI + Mobile +
##     P_15_64 + P_65_plus + PU + TI, data = raw_data)
##
## Residuals:
##     Min      1Q Median      3Q     Max
## -0.8692 -0.1827  0.0004  0.2151  0.6129
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                2.25e+01   9.95e+00    2.26  0.02715 *
## BR                        -2.56e+02   6.77e+01   -3.77  0.00034 ***
## CntryFiji                  1.44e+01   2.99e+00    4.83  8.4e-06 ***
## CntryMicronesia, Fed. Sts.  3.74e+01   4.48e+00    8.34  6.0e-12 ***
## CntryNew Zealand           1.44e-01   8.28e-01    0.17  0.86224
## CntryPapua New Guinea      3.26e+01   5.11e+00    6.38  1.9e-08 ***
## CntrySamoa                 3.94e+01   4.56e+00    8.63  1.7e-12 ***
## CntrySolomon Islands       3.78e+01   4.67e+00    8.09  1.6e-11 ***
## CntryTonga                 4.03e+01   4.52e+00    8.92  5.3e-13 ***
## CntryVanuatu               3.58e+01   4.34e+00    8.25  8.6e-12 ***
## log(GDP)                   9.92e-01   2.91e-01    3.41  0.00110 **
## HCSPC                      1.79e-03   2.70e-04    6.61  7.4e-09 ***
## LI                        -1.73e+01   4.47e+00   -3.87  0.00025 ***
## Mobile                    -6.21e-01   4.57e-01   -1.36  0.17875
## P_15_64                   -2.63e+01   9.16e+00   -2.87  0.00550 **
## P_65_plus                 -9.30e+01   4.08e+01   -2.28  0.02596 *
## PU                         7.13e+01   9.95e+00    7.16  7.8e-10 ***
## TI                        -2.03e-10   4.36e-11   -4.67  1.5e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.345 on 67 degrees of freedom
## Multiple R-squared:  0.998, Adjusted R-squared:   0.997
## F-statistic: 1.58e+03 on 17 and 67 DF, p-value: <2e-16
```