International Trade using EGM Neural Networks and Random Forests

Base paper:

Neural Network Analysis of International Trade

Author:

Isaac Wohl Jim Kennedy

Group Members

A. Prakhar Pradhan 190618 (CE)(**Group Leader**)
B. Aditya Gupta 190060 (CE)
C. Kaushal Chaudhary 190422 (ECO)
D. Raj Aryan 190673 (CE)
E. Chirag Sharma 190248 (CE)
F. Atul Umak 190202 (CE)
G. Sumit Singh Sheoran 190875 (CE)
H. Saral Verma 190767 (CE)
I. Sandeep Gautam 190754 (ECO)
J. Anurag Yadav 190170 (CE)

Introduction

The Gravity Model is the workhorse of International Trade Literature. It all started with Tinbergen(1962),relating Trade flows with Economic sizes and Geographical Distance between two Trading partners. It is the key tool to find Trade-Related effects and since then there are many versions of Gravity model From Traditional Gravity Theory to Extended Gravity Theory. Recently Gravity Model is not only constrained to Trade in goods but also applied in Trade in Services (Kimura 2006). Now a Days Structural Gravity Models(i.e Theoretically Grounded) are being explored. In this Paper in forthcoming sections we will analyse both intuitive as well as extended gravity model as intuitive model does not account for multilateral resistance terms. We will apply Anderson and Wincoop(2003) "Gravity with Gravitas" model which is Theoretically-Grounded Gravity model. The feature of this is it incorporates both , inward multilateral and outward multilateral resistance terms. To estimate this we move on to Fixed Effects Estimation accounting the multilateral resistance terms. Further to deal with "Heteroskedasticity" and zero trade flow values, we will apply Poisson Pseudo-Maximum Likelihood Estimator because Fixed Effect and Intuitive Model uses OLS which drops zero Trade values and may lead to inappropriate results. Moving on we will also apply negative binomial model to deal with overdispersion. Moreover we will apply extension of PPML which is GPML developed by Anderson(2015).

Moving on, we will deploy Machine Learning Regression Algorithms to analyse the Trade data. we will use Neural Networks and Random Forest Regression. This study uses Random Forest and Neural Networks to analyse the panel data. We will compare the models predictive powers by famous evaluation metric R2 square. As gravity model evaluation are based on hypothesis testing therefore we will use Cross Validation for training the model. There is tendency for the model to get over-fit, which we will deal with regularisation and hyper-tuning. As there are various techniques like Ridge regression, Decision Trees etc. moreover for increasing prediction accuracy there are various tools available like Bagging, Random Forest, Boosting and Extra-trees regression. We will go with Random Forest one and Neural Network analysis. It is expected that random forest will work better than neural network analysis and at par with the PPML as it will provide more information to understand the bilateral trade flow as it understands the complex analysis within the data which may ultimately help the Trade Policy Makers of the country. In the upcoming sections we will analyse the trade panel data from 1948-2019 by various econometric models and Machine Learning Algorithms and compare their predictive powers.

Literature Review

- Tinbergen(1962) gave the Gravity Model followed by Leamer and Levinsohn(1995)'s statement that gravity model gave many robust findings in empirical economics.
- Anderson and Wincoop(2003) gave "gravity with gravitas " model which is demand function which incorporates both outward and inward multilateral resistance terms. It was developed as intuitive model does not account for these terms.Trade cost in intuitive model is function of distance only but in Structural Gravity model we define Trade cost as function of distance , common border, colony , common language and many agreements like EU, WTO and RTA.
- This can be achieved by two ways, one is fixed effect estimation and other is approximation technique developed by Baier and Bergstrand(2009).
- Silva and Tenreyro(2006) developed a way to deal with Zero Trade Flow values and Heteroskedasticity as it runs like non linear squares on the equation. Researchers believed it is better than OLS as it performs better even in dataset with large number of zeros and OLS drops all those values.
- Anderson(2015) developed the extension of PPML which is GPML and many have started using Poisson because it outperforms the OLS.

- Head and mayer(2014) highlighted the importance of fixed effect. By using country specific fixed effects all monadic terms including multilateral terms are captured.
- Wohl and Kennedy (2018) examined that neural networks have nonlinear advantage to analyse complex relation within the dataset . They Pointed that NN are more robust and highly adaptive.
- Baxter and Hersh(2017) ,and Storm,Baylis(2019) offered the study of Time Series projections by applying deep learning approach as Machine Learning Algorithms can be used in studying Trade Disruptions going on in the world like the Brexit and US-China.
- Ke et al.2017 showed R-square as the Statistical Measure among both Supervised and Unsupervised Learnings. Moreover study uses LightGBM and XGboost for boosting. In our study we will proceed with Random Forest Regression.

Dataset Preview

This Study collects panel dataset from the centre d’Etudes et d'Informations Internationales (CEPII), UN-Comtrade database. It includes Trade Flow and GDP between countries which are in real terms not in nominal terms as Multilateral terms affect these. Other variables to account Trade Cost like Geographical distance , Common Language, religion , Border. Many agreement variables like RTA, GATT,EU WTO are also accounted in this study. Our study Consist of Top 18 Countries which are India, China, USA, Japan, Singapore, Belgium, Hongkong, Bangladesh, Pakistan, France, Italy, Korea, SaudiArabia, United Arab, Germany, Myanmar, United Kingdom and Switzerland. We will analyse trade data from 1948-2019 so total of 23328 rows are present in our study. These countries are chosen relevant to India as this study will help Trade Policy Makers to make better policies . Further Section involves implementation of Intuitive and Extended Gravity model followed by PPML,GPML, NLS. Finally implementation of Neural Network and Random Forest and Comparison of predictive powers of these models using R2 Square.

Variables	Description
Tradeflow_comrade_o	Trade flow as reported by the exporter (in thousands current US$) (source: Comtrade)
iso3	ISO3 alphabetic code
Country	Country name
Contig	Dummy equal to 1 if countries are contiguous
Distw	Population-weighted distance between most populated cities (km)
Distwces	Population-weighted distance between most populated cities (km) using CES formulation with $\theta = -1$
Comlang_off	1 if countries share common official or primary language
Comlang_ethno	1 if countries share a common language spoken by at least 9% of the population

Variables	Description
Comcol	1 if countries share a common colonizer post 1945
Comrelig	Religious proximity index
Pop	Population (in thousands)
Gdpcap	GDP per capita (current thousands US$)
Comcur	1 if pair currently shares the same currency
Gatt	1 if country currently is a GATT member
wto	1 if country currently is a WTO member
rta	1 if the pair currently has a RTA (source: WTO)
Rta_coverage	Indicates whether the RTA covers goods only or goods and services (source: WTO)
Rta_type	Indicates the type of RTA (customs union for instance)

Model and Methodology

Writing gravity model in its Basic Form:

$$T_{ij} = \alpha \frac{(GDP_i)^{\beta_1} (GDP_j)^{\beta_2}}{(D_{ij})^{\beta_3}}$$

GDP i and j Refers to GDP(real) of Two Countries and D is the Geographical Distance between the countries. T represents Trade Volume between countries.

Intuitive Gravity Model can appear as Ordinary Least Square(OLS):

$$lnX_{ij,t} = \theta_0 + \theta_1 ln(GDP_{i,t}) + \theta_2 ln(GDP_{j,t}) + \theta_3 ln(Dist_{ij}) + \theta_4 Contig_{ij} + \theta_5 Comlang_{ij} + \theta_6 Col_{ij} + \theta_7 ln(Infra_i) + \theta_8 ln(Infra_j) + \theta_9 Obor_{ij} + \theta_{10} Asean_{ij} + \theta_{11} Eac_{ij} + \theta_{12} Sadc_{ij} + e_{ijt}$$

Adding some fixed country effects:

$$lnX_{ij,t} = \theta_0 + \theta_1 ln(GDP_{i,t}) + \theta_2 ln(GDP_{j,t}) + \theta_3 ln(Dist_{ij}) + \theta_4 Contig_{ij} + \theta_5 Comlang_{ij} + \theta_6 Col_{ij} + \theta_7 ln(Infra_i) + \theta_8 ln(Infra_j) + \theta_9 Obor_{ij} + \theta_{10} Asean_{ij} + \theta_{11} Eac_{ij} + \theta_{12} Sadc_{ij} + \mu_i + \alpha_j + e_{ijt}$$

Neural Network :

$Y = f(w_1x_1 + w_2x_2 + b)$

There are different types of Activation function in the literature defined as follows:

$$\mathrm{sigmoid}(x) = 1/(1 + \exp(-x))$$

$$\tanh(x) = 2*\mathrm{sigmoid}(2x) - 1$$

$$\mathrm{ReLU}(x) = \max(0, x)$$

Random Forest Regression

Supervised learning model which prevents Overfitting of the data. It uses multiple decision Trees which increases accuracy

$$f_{rf}^{B}(x) = (1/B)*(\sum^{B}_{b=1} T_b(x))$$

$$\mathrm{Var}((1/B)*(\sum^{B}_{b=1} T_b(x))) = \rho\sigma^2 + \sigma^2((1 - \rho)/B)$$

Here B is the number of trees.

Tb = Random Forest Tree

Evaluation Metric and Results

To evaluate our econometric models we use Hypothesis testing. Generally Distance has negative Coefficient with great negative t-score and null hypothesis is rejected. In OLS estimation we use Robust Standard error and distance is used as the cluster. R2 Square value which is Coefficient of Determination and RMSE is Root mean square error. Generally value of R2 Square is between 0 and 1. Value of 0 predicts that model is failed and value of 1 predicts model perfectly fits Data. RMSE is square root of differences between predicted and actual values. Moreover coefficients of different variables in OLS, PPML and others model gives their effect on Trade. To check the statistical significance we use p value as our measure. In random Forest to improve the accuracy boosting parameters like learning rate , number of trees is being used. In Neural Network Analysis which is Feed Forward Algorithm we have used Rprop Algorithm to minimise loss function and other hyper tuning parameters like learning rate , number of epochs and number of hidden layers.

```
> fit <- ols(dependent_variable = 'tradeflow_comtrade_o',distance = 'distwces',additional_regressors = c('rta','rta_coverag
e','contig','comlang_off','rta_type','comlang_ethno','comcol','comrelig','gatt_o','gatt_d','wto_o','wto_d','eu_o','eu_d','p
_o','pop_d','entry_cost_o','entry_cost_d'),income_origin = 'gdp_o',income_destination='gdp_d',code_origin = 'iso3_o',code_d
tination = 'iso3_d',data = df_1)
> summary(fit)

Call:
y_log_ols ~ dist_log + inc_o_log + inc_d_log + rta + rta_coverage +
    contig + comlang_off + rta_type + comlang_ethno + comcol +
    comrelig + gatt_o + gatt_d + wto_o + wto_d + eu_o + eu_d +
    pop_o + pop_d + entry_cost_o + entry_cost_d

Residuals:
    Min      1Q  Median      3Q     Max
-8.3668 -0.6282  0.0714  0.7727  4.4335

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   -6.208e+00  6.332e-01  -9.803  < 2e-16 ***
dist_log      -5.480e-01  2.920e-02 -18.766  < 2e-16 ***
inc_o_log      6.924e-01  1.591e-02  43.514  < 2e-16 ***
inc_d_log      6.066e-01  1.598e-02  37.953  < 2e-16 ***
rta           -8.091e-01  1.062e-01  -7.615 3.18e-14 ***
rta_coverage   4.293e-01  4.002e-02  10.727  < 2e-16 ***
contig        -8.673e-02  7.434e-02  -1.167 0.243412
comlang_off   -5.297e-03  8.802e-02  -0.060 0.952017
rta_type      -1.495e-02  2.686e-02  -0.556 0.577937
comlang_ethno  6.618e-01  8.047e-02   8.224 2.52e-16 ***
comcol         2.812e-01  6.943e-02   4.050 5.20e-05 ***
comrelig       7.488e-02  9.980e-02   0.750 0.453137
gatt_o         3.903e-01  6.390e-02   6.109 1.08e-09 ***
gatt_d        -1.355e-01  6.339e-02  -2.137 0.032654 *
wto_o         -3.073e-01  2.108e-01  -1.458 0.144905
wto_d         -7.550e-01  2.168e-01  -3.483 0.000501 ***
eu_o           2.229e-02  4.632e-02   0.481 0.630347
eu_d          -6.321e-02  4.726e-02  -1.338 0.181116
pop_o          5.687e-07  5.910e-08   9.623  < 2e-16 ***
pop_d          8.402e-08  5.954e-08   1.411 0.158255
entry_cost_o  -1.928e-02  9.168e-04 -21.036  < 2e-16 ***
entry_cost_d  -1.392e-02  9.264e-04 -15.023  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.107 on 4645 degrees of freedom
  (10301 observations deleted due to missingness)
Multiple R-squared:  0.7135,	Adjusted R-squared:  0.7122
F-statistic: 550.7 on 21 and 4645 DF,  p-value: < 2.2e-16

>
```

OLS Estimate of Intuitive Gravity Model

R square : 0.71

P values are significant Null Hypothesis is rejected

Coefficients of dist , gdp_o , gdp_d etc can be used for correlation. For example 1% increase in dist would decrease trade by 0.55 percent.

```
> fit <- fixed_effects(dependent_variable = 'tradeflow_comtrade_o',distance = 'distwces',additional_regressors = c('rta','rta
_coverage','contig','comlang_off','rta_type','comlang_ethno','comcol','comrelig','gatt_o','gatt_d','wto_o','wto_d','eu_o','eu
_d','pop_o','pop_d','entry_cost_o','entry_cost_d'),code_origin = 'iso3_o',code_destination = 'iso3_d',data = df_1)
> summary(fit)

Call:
y_log_fe ~ dist_log + rta + rta_coverage + contig + comlang_off +
    rta_type + comlang_ethno + comcol + comrelig + gatt_o + gatt_d +
    wto_o + wto_d + eu_o + eu_d + pop_o + pop_d + entry_cost_o +
    entry_cost_d + iso3_o + iso3_d

Residuals:
    Min      1Q  Median      3Q     Max
-6.2622 -0.3490  0.0385  0.3731  3.3099

Coefficients: (4 not defined because of singularities)
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.709e+01  2.709e-01  63.093  < 2e-16 ***
dist_log      -6.002e-01  2.039e-02 -29.437  < 2e-16 ***
rta           -1.103e-01  7.288e-02  -1.513 0.130313
rta_coverage   1.302e-01  2.777e-02   4.690 2.81e-06 ***
contig         1.960e-01  5.228e-02   3.750 0.000179 ***
comlang_off   -1.010e-01  6.536e-02  -1.545 0.122482
rta_type      -5.446e-02  1.833e-02  -2.970 0.002989 **
comlang_ethno  1.252e-01  6.334e-02   1.977 0.048141 *
comcol        -2.004e-02  5.103e-02  -0.393 0.694579
comrelig       8.538e-01  7.253e-02  11.772  < 2e-16 ***
gatt_o         6.361e-01  7.638e-02   8.329  < 2e-16 ***
gatt_d         6.106e-01  6.727e-02   9.077  < 2e-16 ***
wto_o          4.705e-01  1.420e-01   3.313 0.000930 ***
wto_d          4.868e-01  1.459e-01   3.336 0.000858 ***
eu_o           1.324e+00  7.901e-02  16.760  < 2e-16 ***
eu_d          -2.625e-01  7.156e-02  -3.668 0.000248 ***
pop_o          2.806e-06  5.136e-07   5.464 4.91e-08 ***
pop_d          4.447e-06  5.274e-07   8.433  < 2e-16 ***
entry_cost_o  -9.152e-03  9.092e-04 -10.066  < 2e-16 ***
entry_cost_d  -5.225e-03  9.237e-04  -5.656 1.64e-08 ***
iso3_oBEL     -4.664e-01  6.827e-02  -6.831 9.53e-12 ***
iso3_oBGD     -1.764e+00  1.136e-01 -15.520  < 2e-16 ***
iso3_oCHE      5.884e-01  7.507e-02   7.838 5.65e-15 ***
iso3_oCHN      1.504e-01  6.735e-01   0.223 0.823348
iso3_oDEU      7.546e-01  6.233e-02  12.107  < 2e-16 ***
iso3_oFRA     -1.281e-01  6.275e-02  -2.041 0.041261 *
iso3_oGBR      4.958e-02  6.519e-02   0.760 0.447000
iso3_oHKG      1.493e+00  7.392e-02  20.193  < 2e-16 ***
iso3_oIND     -1.709e+00  6.435e-01  -2.656 0.007925 **
iso3_oITA             NA         NA      NA       NA
iso3_oJPN      2.133e+00  9.439e-02  22.594  < 2e-16 ***
iso3_oKOR      1.610e+00  8.217e-02  19.598  < 2e-16 ***
iso3_oMMR     -2.023e+00  1.125e-01 -17.977  < 2e-16 ***
iso3_oPAK     -1.748e+00  1.149e-01 -15.213  < 2e-16 ***
iso3_oSAU             NA         NA      NA       NA
iso3_oSGP      1.466e+00  7.619e-02  19.241  < 2e-16 ***
iso3_oUSA      2.349e+00  1.683e-01  13.959  < 2e-16 ***
iso3_dBEL      1.681e-02  7.004e-02   0.240 0.810331
iso3_dBGD     -2.736e+00  1.050e-01 -26.054  < 2e-16 ***
iso3_dCHE     -1.290e+00  6.710e-02 -19.219  < 2e-16 ***
```

Fixed Effect Estimation incorporating Multilateral resistance terms

R2 Square : 0.86

P values are significant ,i.e Null Hypothesis Rejected.

```
iso3_oCHE      5.884e-01  7.507e-02   7.838 5.65e-15 ***
iso3_oCHN      1.504e-01  6.735e-01   0.223 0.823348
iso3_oDEU      7.546e-01  6.233e-02  12.107  < 2e-16 ***
iso3_oFRA     -1.281e-01  6.275e-02  -2.041 0.041261 *
iso3_oGBR      4.958e-02  6.519e-02   0.760 0.447000
iso3_oHKG      1.493e+00  7.392e-02  20.193  < 2e-16 ***
iso3_oIND     -1.709e+00  6.435e-01  -2.656 0.007925 **
iso3_oITA             NA         NA      NA       NA
iso3_oJPN      2.133e+00  9.439e-02  22.594  < 2e-16 ***
iso3_oKOR      1.610e+00  8.217e-02  19.598  < 2e-16 ***
iso3_oMMR     -2.023e+00  1.125e-01 -17.977  < 2e-16 ***
iso3_oPAK     -1.748e+00  1.149e-01 -15.213  < 2e-16 ***
iso3_oSAU             NA         NA      NA       NA
iso3_oSGP      1.466e+00  7.619e-02  19.241  < 2e-16 ***
iso3_oUSA      2.349e+00  1.683e-01  13.959  < 2e-16 ***
iso3_dBEL      1.681e-02  7.004e-02   0.240 0.810331
iso3_dBGD     -2.736e+00  1.050e-01 -26.054  < 2e-16 ***
iso3_dCHE     -1.290e+00  6.710e-02 -19.219  < 2e-16 ***
iso3_dCHN     -3.888e+00  6.931e-01  -5.610 2.15e-08 ***
iso3_dDEU      6.489e-01  6.397e-02  10.143  < 2e-16 ***
iso3_dFRA      7.956e-02  6.440e-02   1.235 0.216710
iso3_dGBR      5.943e-01  6.698e-02   8.872  < 2e-16 ***
iso3_dHKG      5.237e-01  6.608e-02   7.925 2.85e-15 ***
iso3_dIND     -5.093e+00  6.590e-01  -7.729 1.32e-14 ***
iso3_dITA             NA         NA      NA       NA
iso3_dJPN      2.294e-01  8.904e-02   2.576 0.010025 *
iso3_dKOR     -7.907e-02  7.568e-02  -1.045 0.296172
iso3_dMMR     -3.172e+00  1.105e-01 -28.695  < 2e-16 ***
iso3_dPAK     -2.630e+00  1.118e-01 -23.527  < 2e-16 ***
iso3_dSAU             NA         NA      NA       NA
iso3_dSGP      7.278e-02  6.894e-02   1.056 0.291147
iso3_dUSA      9.937e-01  1.685e-01   5.897 3.96e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7196 on 4617 degrees of freedom
  (10301 observations deleted due to missingness)
Multiple R-squared:  0.8795,	Adjusted R-squared:  0.8783 
F-statistic:   688 on 49 and 4617 DF,  p-value: < 2.2e-16

>
```

Continued Fixed Effect Estimation

Better than OLS.

Coefficients interpretation :
1% increase in following lead to % change in trade

```
> fit <- ppml(dependent_variable = 'tradeflow_comtrade_o',distance = 'distwces',additional_regressors = c('rta','rta_coverag
e','contig','comlang_off','rta_type','comlang_ethno','comcol','comrelig','gatt_o','gatt_d','wto_o','wto_d','eu_o','eu_d','pop
_o','pop_d','entry_cost_o','entry_cost_d','iso3_o','iso3_d'),data = df_1)
> summary(fit)

Call:
y_ppml ~ dist_log + rta + rta_coverage + contig + comlang_off +
    rta_type + comlang_ethno + comcol + comrelig + gatt_o + gatt_d +
    wto_o + wto_d + eu_o + eu_d + pop_o + pop_d + entry_cost_o +
    entry_cost_d + iso3_o + iso3_d

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-8161.5   -836.8   -130.9    642.2  12544.0

Coefficients: (4 not defined because of singularities)
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.669e+01  5.317e-01  31.382  < 2e-16 ***
dist_log      -5.827e-01  1.328e-02 -43.873  < 2e-16 ***
rta           -1.232e-02  5.491e-02  -0.224 0.822451
rta_coverage   2.948e-02  1.910e-02   1.544 0.122751
contig         3.393e-01  2.928e-02  11.588  < 2e-16 ***
comlang_off   -1.131e-01  4.393e-02  -2.574 0.010094 *
rta_type      -1.261e-01  1.069e-02 -11.796  < 2e-16 ***
comlang_ethno  1.438e-01  4.228e-02   3.401 0.000678 ***
comcol         4.701e-01  4.682e-02  10.040  < 2e-16 ***
comrelig       4.439e-01  6.884e-02   6.448 1.25e-10 ***
gatt_o         8.544e-01  1.201e-01   7.117 1.28e-12 ***
gatt_d         4.751e-01  6.114e-02   7.771 9.53e-15 ***
wto_o          1.016e+00  4.620e-01   2.198 0.027986 *
wto_d          6.528e-01  2.347e-01   2.782 0.005423 **
eu_o           9.015e-01  8.467e-02  10.647  < 2e-16 ***
eu_d          -2.651e-01  5.400e-02  -4.910 9.44e-07 ***
pop_o          6.278e-06  4.540e-07  13.827  < 2e-16 ***
pop_d          5.007e-06  4.085e-07  12.257  < 2e-16 ***
entry_cost_o  -1.426e-03  1.765e-03  -0.808 0.419363
entry_cost_d  -3.093e-03  1.420e-03  -2.178 0.029489 *
iso3_oBEL     -1.438e-01  5.313e-02  -2.706 0.006838 **
iso3_oBGD     -2.210e+00  1.799e-01 -12.289  < 2e-16 ***
iso3_oCHE      4.285e-01  8.645e-02   4.956 7.46e-07 ***
iso3_oCHN     -4.321e+00  5.936e-01  -7.278 3.96e-13 ***
iso3_oDEU      7.899e-01  3.432e-02  23.015  < 2e-16 ***
iso3_oFRA     -8.435e-02  4.405e-02  -1.915 0.055571 .
iso3_oGBR      8.215e-02  4.907e-02   1.674 0.094208 .
iso3_oHKG      1.619e+00  8.390e-02  19.291  < 2e-16 ***
iso3_oIND     -6.740e+00  6.044e-01 -11.150  < 2e-16 ***
iso3_oITA             NA         NA      NA       NA
iso3_oJPN      1.617e+00  9.242e-02  17.495  < 2e-16 ***
iso3_oKOR      1.368e+00  8.796e-02  15.552  < 2e-16 ***
iso3_oMMR     -2.227e+00  2.426e-01  -9.178  < 2e-16 ***
iso3_oPAK     -2.642e+00  1.574e-01 -16.784  < 2e-16 ***
iso3_oSAU             NA         NA      NA       NA
iso3_oSGP      1.342e+00  8.811e-02  15.229  < 2e-16 ***
iso3_oUSA      1.275e+00  1.456e-01   8.760  < 2e-16 ***
iso3_dBEL     -4.111e-01  5.173e-02  -7.947 2.39e-15 ***
iso3_dBGD     -2.674e+00  1.201e-01 -22.259  < 2e-16 ***
```


```
iso3_oBEL     -1.438e-01  5.313e-02  -2.706 0.006838 **
iso3_oBGD     -2.210e+00  1.799e-01 -12.289  < 2e-16 ***
iso3_oCHE      4.285e-01  8.645e-02   4.956 7.46e-07 ***
iso3_oCHN     -4.321e+00  5.936e-01  -7.278 3.96e-13 ***
iso3_oDEU      7.899e-01  3.432e-02  23.015  < 2e-16 ***
iso3_oFRA     -8.435e-02  4.405e-02  -1.915 0.055571 .
iso3_oGBR      8.215e-02  4.907e-02   1.674 0.094208 .
iso3_oHKG      1.619e+00  8.390e-02  19.291  < 2e-16 ***
iso3_oIND     -6.740e+00  6.044e-01 -11.150  < 2e-16 ***
iso3_oITA             NA         NA      NA       NA
iso3_oJPN      1.617e+00  9.242e-02  17.495  < 2e-16 ***
iso3_oKOR      1.368e+00  8.796e-02  15.552  < 2e-16 ***
iso3_oMMR     -2.227e+00  2.426e-01  -9.178  < 2e-16 ***
iso3_oPAK     -2.642e+00  1.574e-01 -16.784  < 2e-16 ***
iso3_oSAU             NA         NA      NA       NA
iso3_oSGP      1.342e+00  8.811e-02  15.229  < 2e-16 ***
iso3_oUSA      1.275e+00  1.456e-01   8.760  < 2e-16 ***
iso3_dBEL     -4.111e-01  5.173e-02  -7.947 2.39e-15 ***
iso3_dBGD     -2.674e+00  1.201e-01 -22.259  < 2e-16 ***
```


```
iso3_dCHE     -8.161e-01  5.790e-02 -14.094  < 2e-16 ***
iso3_dCHN     -4.488e+00  5.348e-01  -8.392  < 2e-16 ***
iso3_dDEU      6.094e-01  3.663e-02  16.636  < 2e-16 ***
iso3_dFRA      1.220e-01  4.154e-02   2.936 0.003337 **
iso3_dGBR      3.976e-01  4.524e-02   8.788  < 2e-16 ***
iso3_dHKG      5.310e-01  4.911e-02  10.811  < 2e-16 ***
iso3_dIND     -6.126e+00  5.353e-01 -11.445  < 2e-16 ***
iso3_dITA             NA         NA      NA       NA
iso3_dJPN      2.342e-01  6.277e-02   3.732 0.000192 ***
iso3_dKOR     -8.359e-03  5.650e-02  -0.148 0.882388
iso3_dMMR     -2.738e+00  1.860e-01 -14.715  < 2e-16 ***
iso3_dPAK     -2.935e+00  1.188e-01 -24.717  < 2e-16 ***
iso3_dSAU             NA         NA      NA       NA
iso3_dSGP      4.098e-02  5.795e-02   0.707 0.479451
iso3_dUSA      1.002e+00  1.203e-01   8.329  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 2564595)

    Null deviance: 1.6843e+11  on 4666  degrees of freedom
Residual deviance: 9.9044e+09  on 4617  degrees of freedom
  (17375 observations deleted due to missingness)
AIC: NA

Number of Fisher Scoring iterations: 6
```

PPML estimation

```
> fit <- gpml(dependent_variable = 'tradeflow_comtrade_o',distance = 'distwces',additional_regressors = c('rta','rta_coverag
e','contig','comlang_off','rta_type','comlang_ethno','comcol','comrelig','gatt_o','gatt_d','wto_o','wto_d','eu_o','eu_d','pop
_o','pop_d','entry_cost_o','entry_cost_d','iso3_o','iso3_d'),data = df_1)
> summary(fit)

Call:
y_gpml ~ dist_log + rta + rta_coverage + contig + comlang_off +
    rta_type + comlang_ethno + comcol + comrelig + gatt_o + gatt_d +
    wto_o + wto_d + eu_o + eu_d + pop_o + pop_d + entry_cost_o +
    entry_cost_d + iso3_o + iso3_d

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4777  -0.4852  -0.0916   0.2423   4.2251

Coefficients: (4 not defined because of singularities)
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.739e+01  2.623e-01  66.287  < 2e-16 ***
dist_log      -6.215e-01  1.974e-02 -31.482  < 2e-16 ***
rta            4.146e-02  7.057e-02   0.588 0.556893
rta_coverage   8.253e-02  2.689e-02   3.070 0.002155 **
contig         2.024e-01  5.062e-02   3.999 6.46e-05 ***
comlang_off   -7.743e-02  6.328e-02  -1.224 0.221167
rta_type      -4.962e-02  1.775e-02  -2.795 0.005209 **
comlang_ethno  3.939e-02  6.133e-02   0.642 0.520759
comcol        -5.793e-02  4.941e-02  -1.172 0.241130
comrelig       7.500e-01  7.023e-02  10.680  < 2e-16 ***
gatt_o         9.611e-01  7.395e-02  12.996  < 2e-16 ***
gatt_d         5.964e-01  6.513e-02   9.156  < 2e-16 ***
wto_o          7.027e-01  1.375e-01   5.110 3.35e-07 ***
wto_d          4.804e-01  1.413e-01   3.400 0.000681 ***
eu_o           5.979e-01  7.650e-02   7.816 6.73e-15 ***
eu_d          -3.481e-01  6.929e-02  -5.024 5.26e-07 ***
pop_o          3.917e-06  4.973e-07   7.876 4.17e-15 ***
pop_d          4.395e-06  5.106e-07   8.608  < 2e-16 ***
entry_cost_o  -6.180e-03  8.803e-04  -7.021 2.53e-12 ***
entry_cost_d  -4.491e-03  8.943e-04  -5.022 5.32e-07 ***
iso3_oBEL     -3.061e-01  6.610e-02  -4.631 3.74e-06 ***
iso3_oBGD     -1.843e+00  1.100e-01 -16.753  < 2e-16 ***
iso3_oCHE      2.147e-02  7.268e-02   0.295 0.767706
iso3_oCHN     -1.611e+00  6.522e-01  -2.470 0.013555 *
iso3_oDEU      8.125e-01  6.035e-02  13.464  < 2e-16 ***
iso3_oFRA     -7.461e-02  6.076e-02  -1.228 0.219507
iso3_oGBR      1.631e-01  6.312e-02   2.584 0.009784 **
iso3_oHKG      9.805e-01  7.158e-02  13.699  < 2e-16 ***
iso3_oIND     -3.658e+00  6.231e-01  -5.871 4.63e-09 ***
iso3_oITA             NA         NA      NA       NA
iso3_oJPN      1.359e+00  9.139e-02  14.872  < 2e-16 ***
iso3_oKOR      9.669e-01  7.956e-02  12.152  < 2e-16 ***
iso3_oMMR     -2.370e+00  1.090e-01 -21.754  < 2e-16 ***
iso3_oPAK     -2.207e+00  1.112e-01 -19.843  < 2e-16 ***
iso3_oSAU             NA         NA      NA       NA
iso3_oSGP      1.095e+00  7.377e-02  14.848  < 2e-16 ***
iso3_oUSA      1.471e+00  1.630e-01   9.028  < 2e-16 ***
iso3_dBEL      6.496e-02  6.781e-02   0.958 0.338156
```

```
iso3_dBEL      6.496e-02  6.781e-02   0.958 0.338156
iso3_dBGD     -2.696e+00  1.017e-01 -26.507  < 2e-16 ***
iso3_dCHE     -1.201e+00  6.497e-02 -18.487  < 2e-16 ***
iso3_dCHN     -3.839e+00  6.711e-01  -5.721 1.12e-08 ***
iso3_dDEU      7.033e-01  6.194e-02  11.354  < 2e-16 ***
iso3_dFRA      6.846e-02  6.235e-02   1.098 0.272265
iso3_dGBR      6.059e-01  6.485e-02   9.343  < 2e-16 ***
iso3_dHKG      5.368e-01  6.399e-02   8.389  < 2e-16 ***
iso3_dIND     -4.967e+00  6.381e-01  -7.784 8.62e-15 ***
iso3_dITA             NA         NA      NA       NA
iso3_dJPN      2.482e-01  8.621e-02   2.879 0.004012 **
iso3_dKOR     -1.606e-01  7.328e-02  -2.191 0.028471 *
iso3_dMMR     -2.954e+00  1.070e-01 -27.604  < 2e-16 ***
iso3_dPAK     -2.711e+00  1.082e-01 -25.048  < 2e-16 ***
iso3_dSAU             NA         NA      NA       NA
iso3_dSGP      1.299e-01  6.675e-02   1.946 0.051703 .
iso3_dUSA      9.793e-01  1.632e-01   6.002 2.10e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.4854276)

    Null deviance: 14156.4  on 4666  degrees of freedom
Residual deviance:  2116.1  on 4617  degrees of freedom
  (10301 observations deleted due to missingness)
AIC: 151184

Number of Fisher Scoring iterations: 12

>
```

GPML Estimation

```
~/R/
> fit <- nbpml(dependent_variable = 'tradeflow_comtrade_o',distance = 'distwces',additional_regressors = c('rta','rta_coverag
e','contig','comlang_off','rta_type','comlang_ethno','comcol','comrelig','gatt_o','gatt_d','wto_o','wto_d','eu_o','eu_d','pop
_o','pop_d','entry_cost_o','entry_cost_d','iso3_o','iso3_d'),data = df_1)
> summary(fit)

Call:
y_nbpml ~ dist_log + rta + rta_coverage + contig + comlang_off +
    rta_type + comlang_ethno + comcol + comrelig + gatt_o + gatt_d +
    wto_o + wto_d + eu_o + eu_d + pop_o + pop_d + entry_cost_o +
    entry_cost_d + iso3_o + iso3_d

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.3373  -0.7452  -0.1406   0.3721   6.4891

Coefficients: (4 not defined because of singularities)
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.739e+01  2.452e-01  70.931  < 2e-16 ***
dist_log      -6.215e-01  1.845e-02 -33.688  < 2e-16 ***
rta            4.149e-02  6.595e-02   0.629 0.529244
rta_coverage   8.252e-02  2.512e-02   3.284 0.001022 **
contig         2.024e-01  4.730e-02   4.279 1.88e-05 ***
comlang_off   -7.745e-02  5.914e-02  -1.310 0.190333
rta_type      -4.962e-02  1.659e-02  -2.991 0.002780 **
comlang_ethno  3.938e-02  5.731e-02   0.687 0.491986
comcol        -5.790e-02  4.618e-02  -1.254 0.209895
comrelig       7.500e-01  6.563e-02  11.428  < 2e-16 ***
gatt_o         9.611e-01  6.911e-02  13.907  < 2e-16 ***
gatt_d         5.964e-01  6.087e-02   9.798  < 2e-16 ***
wto_o          7.027e-01  1.285e-01   5.468 4.54e-08 ***
wto_d          4.804e-01  1.321e-01   3.638 0.000275 ***
eu_o           5.979e-01  7.149e-02   8.363  < 2e-16 ***
eu_d          -3.481e-01  6.475e-02  -5.376 7.63e-08 ***
pop_o          3.917e-06  4.648e-07   8.429  < 2e-16 ***
pop_d          4.395e-06  4.772e-07   9.211  < 2e-16 ***
entry_cost_o  -6.180e-03  8.227e-04  -7.512 5.82e-14 ***
entry_cost_d  -4.491e-03  8.358e-04  -5.373 7.73e-08 ***
iso3_oBEL     -3.061e-01  6.177e-02  -4.955 7.22e-07 ***
iso3_oBGD     -1.844e+00  1.028e-01 -17.927  < 2e-16 ***
iso3_oCHE      2.148e-02  6.793e-02   0.316 0.751824
iso3_oCHN     -1.611e+00  6.095e-01  -2.643 0.008217 **
iso3_oDEU      8.125e-01  5.640e-02  14.407  < 2e-16 ***
iso3_oFRA     -7.460e-02  5.678e-02  -1.314 0.188891
iso3_oGBR      1.631e-01  5.899e-02   2.766 0.005679 **
iso3_oHKG      9.805e-01  6.689e-02  14.659  < 2e-16 ***
iso3_oIND     -3.658e+00  5.823e-01  -6.283 3.33e-10 ***
iso3_oITA             NA         NA      NA       NA
iso3_oJPN      1.359e+00  8.541e-02  15.914  < 2e-16 ***
iso3_oKOR      9.669e-01  7.435e-02  13.004  < 2e-16 ***
iso3_oMMR     -2.370e+00  1.018e-01 -23.278  < 2e-16 ***
iso3_oPAK     -2.207e+00  1.040e-01 -21.233  < 2e-16 ***
iso3_oSAU             NA         NA      NA       NA
iso3_oSGP      1.095e+00  6.894e-02  15.888  < 2e-16 ***
iso3_oUSA      1.471e+00  1.523e-01   9.661  < 2e-16 ***
iso3_dBEL      6.496e-02  6.337e-02   1.025 0.305340
iso3_dBGD     -2.696e+00  9.504e-02 -28.364  < 2e-16 ***
```

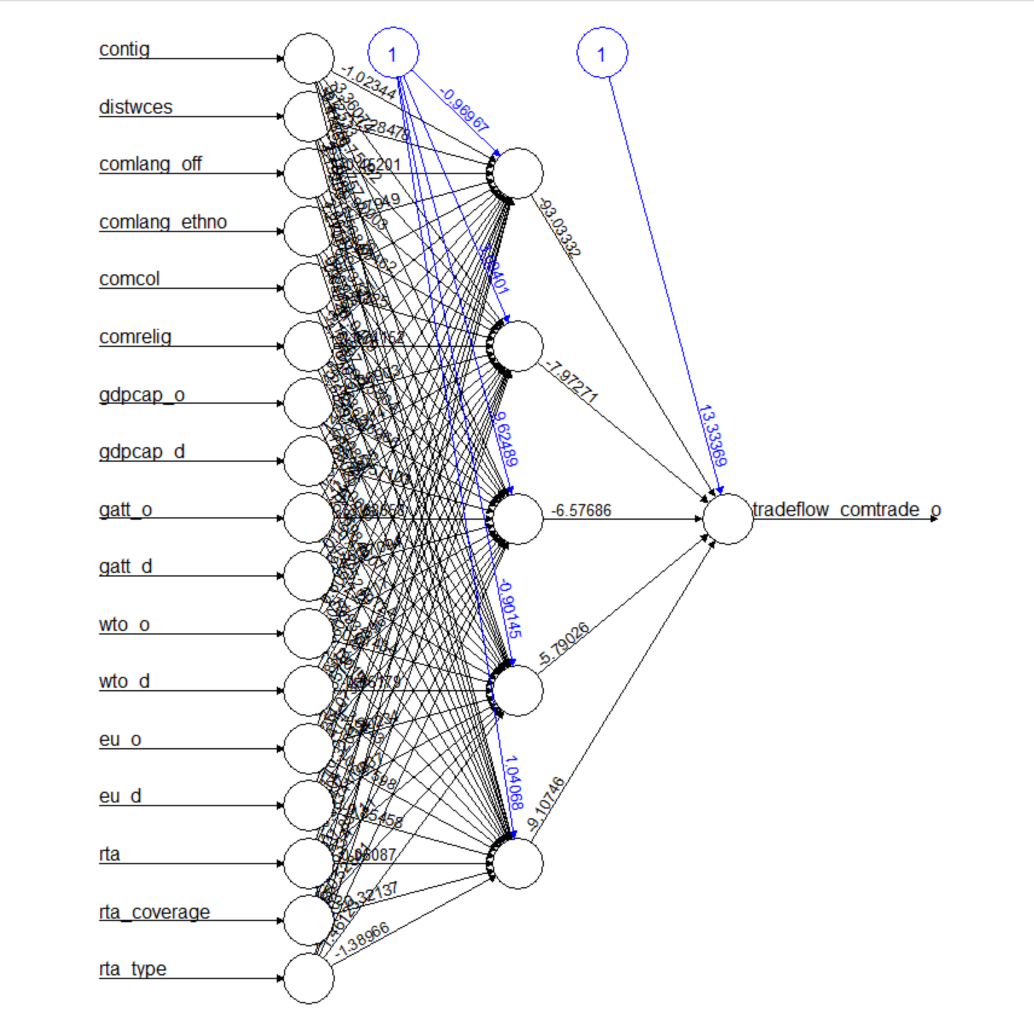

```
iso3_oBGD     -1.844e+00  1.028e-01 -17.927  < 2e-16 ***
iso3_oCHE      2.148e-02  6.793e-02   0.316 0.751824
iso3_oCHN     -1.611e+00  6.095e-01  -2.643 0.008217 **
iso3_oDEU      8.125e-01  5.640e-02  14.407  < 2e-16 ***
iso3_oFRA     -7.460e-02  5.678e-02  -1.314 0.188891
iso3_oGBR      1.631e-01  5.899e-02   2.766 0.005679 **
iso3_oHKG      9.805e-01  6.689e-02  14.659  < 2e-16 ***
iso3_oIND     -3.658e+00  5.823e-01  -6.283 3.33e-10 ***
iso3_oITA             NA         NA      NA       NA
iso3_oJPN      1.359e+00  8.541e-02  15.914  < 2e-16 ***
iso3_oKOR      9.669e-01  7.435e-02  13.004  < 2e-16 ***
iso3_oMMR     -2.370e+00  1.018e-01 -23.278  < 2e-16 ***
iso3_oPAK     -2.207e+00  1.040e-01 -21.233  < 2e-16 ***
iso3_oSAU             NA         NA      NA       NA
iso3_oSGP      1.095e+00  6.894e-02  15.888  < 2e-16 ***
iso3_oUSA      1.471e+00  1.523e-01   9.661  < 2e-16 ***
iso3_dBEL      6.496e-02  6.337e-02   1.025 0.305340
iso3_dBGD     -2.696e+00  9.504e-02 -28.364  < 2e-16 ***
```


```
iso3_dCHE     -1.201e+00  6.071e-02 -19.782  < 2e-16 ***
iso3_dCHN     -3.840e+00  6.271e-01  -6.122 9.23e-10 ***
iso3_dDEU      7.033e-01  5.789e-02  12.150  < 2e-16 ***
iso3_dFRA      6.846e-02  5.827e-02   1.175 0.240016
iso3_dGBR      6.059e-01  6.061e-02   9.998  < 2e-16 ***
iso3_dHKG      5.368e-01  5.980e-02   8.977  < 2e-16 ***
iso3_dIND     -4.967e+00  5.963e-01  -8.329  < 2e-16 ***
iso3_dITA             NA         NA      NA       NA
iso3_dJPN      2.482e-01  8.057e-02   3.081 0.002066 **
iso3_dKOR     -1.606e-01  6.848e-02  -2.345 0.019038 *
iso3_dMMR     -2.954e+00  1.000e-01 -29.538  < 2e-16 ***
iso3_dPAK     -2.711e+00  1.011e-01 -26.803  < 2e-16 ***
iso3_dSAU             NA         NA      NA       NA
iso3_dSGP      1.299e-01  6.238e-02   2.083 0.037281 *
iso3_dUSA      9.793e-01  1.525e-01   6.423 1.34e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(2.3588) family taken to be 1)

    Null deviance: 33392.3  on 4666  degrees of freedom
Residual deviance:  4991.3  on 4617  degrees of freedom
  (17375 observations deleted due to missingness)
AIC: 151172

Number of Fisher Scoring iterations: 1


              Theta:  2.3588
          Std. Err.:  0.0458

 2 x log-likelihood:  -151069.9090
```

NBPML Estimation

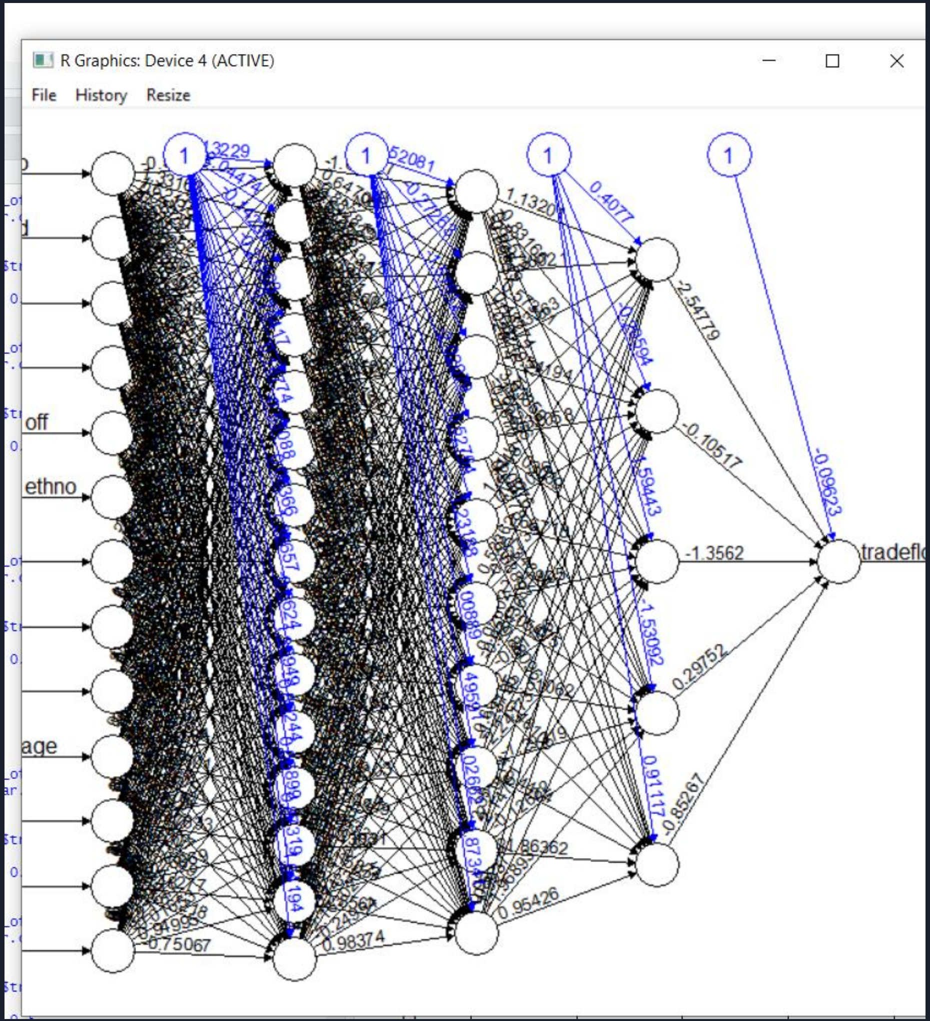
```
> fit <- nls(dependent_variable = 'tradeflow_comtrade_o',distance = 'distwces',additional_regressors = c('contig','comcol','i
so3_o','iso3_d','rta','rta_type','comrelig','comlang_ethno','comlang_off','gatt_o','eu_o','wto_o'),data = df_1)
> summary(fit)

Call:
tradeflow_comtrade_o ~ dist_log + contig + comcol + iso3_o +
    iso3_d + rta + rta_type + comrelig + comlang_ethno + comlang_off +
    gatt_o + eu_o + wto_o

Deviance Residuals:
       Min         1Q      Median         3Q        Max
-201412076    -573398      -12573     709310  223932940

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  13.104649   0.828244  15.822  < 2e-16 ***
dist_log     -0.363313   0.014862 -24.446  < 2e-16 ***
contig        0.327455   0.027289  11.999  < 2e-16 ***
comcol        1.186352   0.083208  14.258  < 2e-16 ***
iso3_oBEL    -0.381245   1.309393  -0.291 0.770932
iso3_oBGD    -1.375026   0.937983  -1.466 0.142686
iso3_oCHE     0.604142   0.292566   2.065 0.038942 *
iso3_oCHN     5.809836   0.815797   7.122 1.12e-12 ***
iso3_oDEU     1.039316   1.309319   0.794 0.427334
iso3_oFRA     0.131549   1.309292   0.100 0.919970
iso3_oGBR     0.246223   1.309337   0.188 0.850839
iso3_oHKG     2.145023   0.290497   7.384 1.62e-13 ***
iso3_oIND     0.978316   0.297338   3.290 0.001003 **
iso3_oITA     0.135193   1.309276   0.103 0.917760
iso3_oJPN     3.136435   0.290691  10.790  < 2e-16 ***
iso3_oKOR     2.190287   0.292556   7.487 7.46e-14 ***
iso3_oMMR    -1.496184   0.861641  -1.736 0.082507 .
iso3_oPAK    -1.612848   0.709103  -2.274 0.022951 *
iso3_oSAU     1.599454   0.879366   1.819 0.068951 .
iso3_oSGP     0.958991   0.293616   3.266 0.001093 **
iso3_oUSA     3.643828   0.290030  12.564  < 2e-16 ***
iso3_dBEL    -0.656924   0.086765  -7.571 3.91e-14 ***
iso3_dBGD    -1.823772   0.279241  -6.531 6.74e-11 ***
iso3_dCHE    -0.630432   0.088118  -7.154 8.80e-13 ***
iso3_dCHN     1.941321   0.077322  25.107  < 2e-16 ***
iso3_dDEU     0.962686   0.078242  12.304  < 2e-16 ***
iso3_dFRA     0.446849   0.079941   5.590 2.31e-08 ***
iso3_dGBR     0.663228   0.080583   8.230  < 2e-16 ***
iso3_dHKG     0.425844   0.081030   5.255 1.50e-07 ***
iso3_dIND    -0.321303   0.091982  -3.493 0.000479 ***
iso3_dITA     0.267727   0.081878   3.270 0.001079 **
iso3_dJPN     1.439700   0.078288  18.390  < 2e-16 ***
iso3_dKOR     0.245736   0.084945   2.893 0.003823 **
iso3_dMMR    -2.887753   0.461087  -6.263 3.88e-10 ***
iso3_dPAK    -2.125393   0.253184  -8.395  < 2e-16 ***
iso3_dSAU    -0.198887   0.127061  -1.565 0.117536
iso3_dSGP    -0.420216   0.091185  -4.608 4.09e-06 ***
iso3_dUSA     2.895071   0.076186  38.000  < 2e-16 ***
rta           0.809098   0.023027  35.137  < 2e-16 ***
rta_type     -0.137023   0.008216 -16.678  < 2e-16 ***
comrelig      0.544000   0.072356   7.518 5.86e-14 ***
comlang_ethno -0.421443   0.043827  -9.616  < 2e-16 ***
```

NLS estimation



Neural Network

Hidden layer -5
Algorithm : “Rprop”
Learning rate = 0.1



More Dense Network

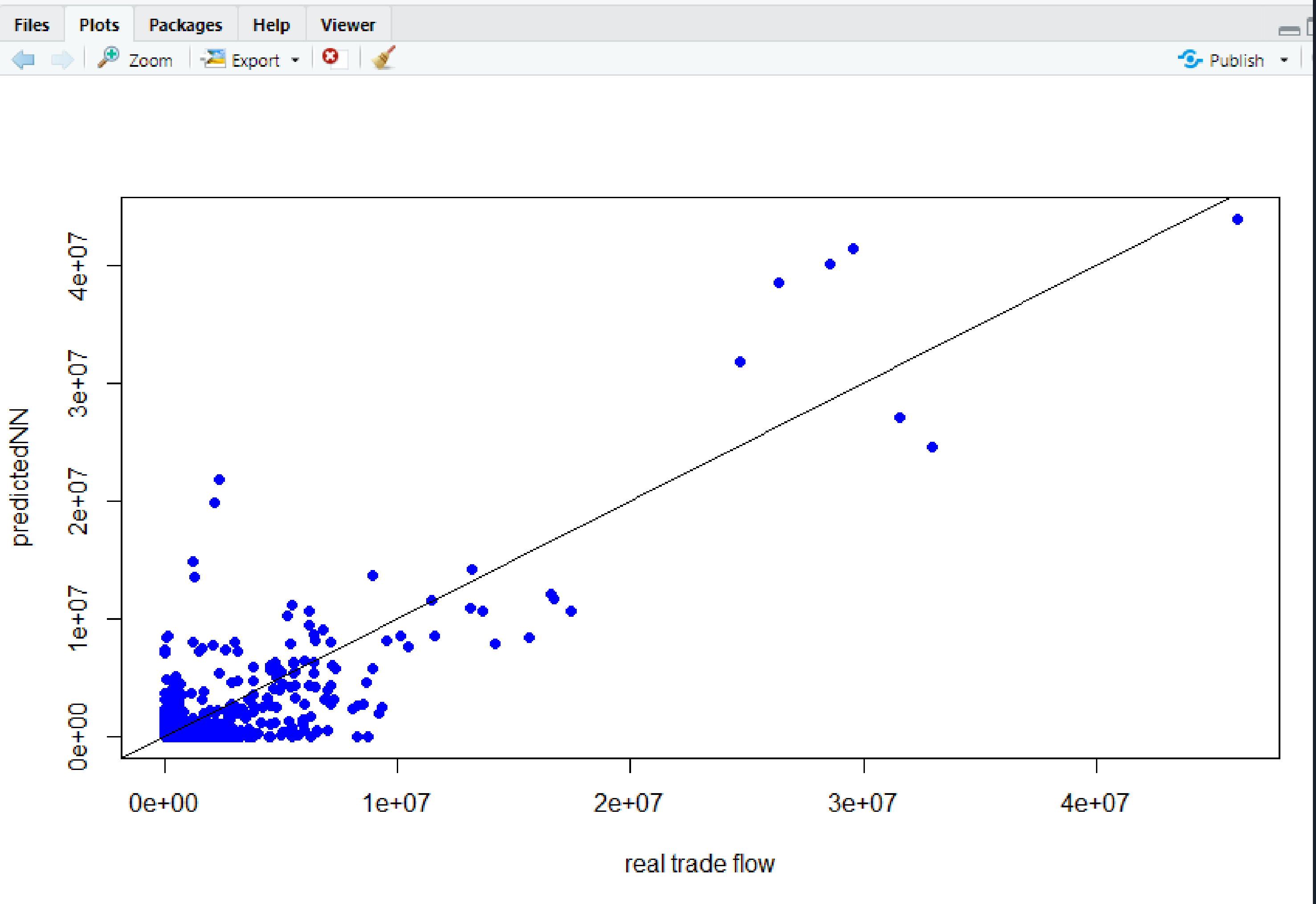
Hidden Layer :- c(15,10,5)
Algorithm :- rprop+
Learning rate :- 0.01

```
> NN = neuralnet(tradeflow_comtrade_o ~ gdpcap_o + gdpcap_d +distwces + contig + comlang_off + comlang_ethno + comcol + rta +
comrelig+ rta_coverage + rta_type + wto_o + wto_d , trainNN, hidden = c(15,10,5) , linear.output = F,algorithm = "rprop+",lea
rningrate=0.001 )
> predict_testNN = compute(NN, testNN)
> predict_testNN = (predict_testNN$net.result * (max(df$tradeflow_comtrade_o) - min(df_1$tradeflow_comtrade_o)) + min(df_1$t
radeflow_comtrade_o)
> RMSE.NN = (sum((datatest$tradeflow_comtrade_o - predict_testNN)^2) / nrow(datatest)) ^ 0.5
> RMSE.NN
[1] 514519.8
> rsq <- cor(predict_testNN, datatest$tradeflow_comtrade_o)^2
> rsq
          [,1]
[1,] 0.8576321
```

R sq value is 0.857 which clearly states that neural network is far more better than intuitive Gravity model

```
> samplesize = 0.60 * nrow(df)
> set.seed(80)
>
> index = sample( seq_len ( nrow ( df ) ), size = samplesize )
> datatrain = df[ index, ]
> datatest = df[ -index, ]
> max = apply(df , 2 , max)
> min = apply(df, 2 , min)
> scaled = as.data.frame(scale(df, center = min, scale = max - min))
> library(neuralnet)
> trainNN = scaled[index , ]
> testNN = scaled[-index , ]
> set.seed(2)
> NN = neuralnet(tradeflow_comtrade_o ~ contig + distwces + comlang_off + comlang_ethno + comcol + comrelig + gdpcap_o + gdpc
ap_d + gatt_o + gatt_d + wto_o + wto_d + eu_o + eu_d + rta + rta_coverage + rta_type , trainNN, hidden = 5 , linear.output =
 F, algorithm = "rprop+" )
> plot(NN)
> predict_testNN = compute(NN, testNN)
> predict_testNN = (predict_testNN$net.result * (max(df$tradeflow_comtrade_o) - min(df$tradeflow_comtrade_o))) + min(df$trade
flow_comtrade_o)
> plot(datatest$tradeflow_comtrade_o, predict_testNN, col='blue', pch=16, ylab = "predicted rating NN", xlab = "real rating")
> abline(0,1)
```

R code for Neural net using neural net package



A line clearly fits the data.

s Help Saving...

Comment Share

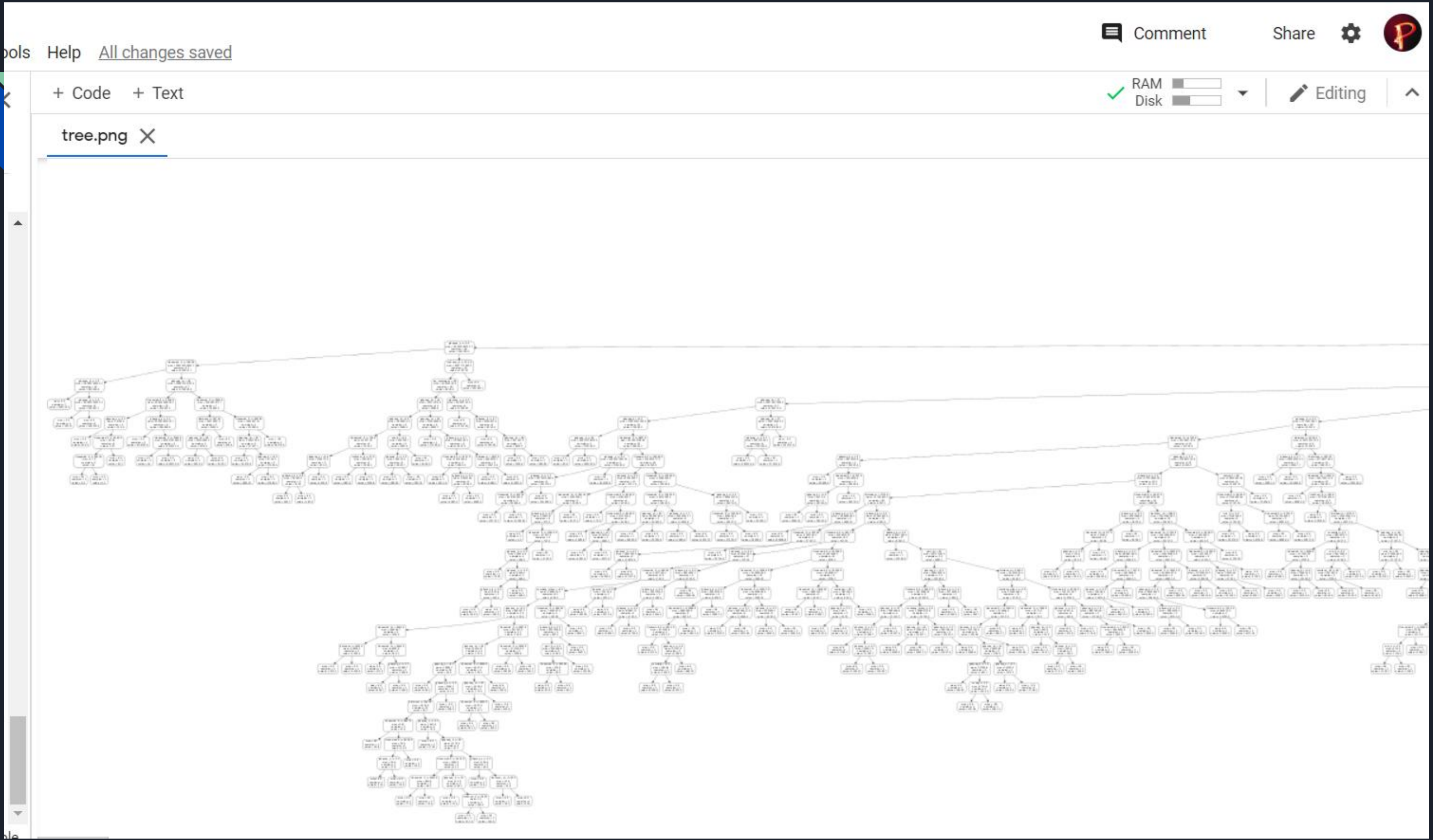
\+ Code + Text

RAM Disk Editing

```
[55] from google.colab import drive
     drive.mount('/content/drive')
     import pandas as pd
     path = '/content/drive/MyDrive/sixth_try.csv'
     df = pd.read_csv(path)
     df.head()
     from sklearn.ensemble import RandomForestRegressor
     x = df.drop(columns=['tradeflow_comtrade_o'])
     y = df['tradeflow_comtrade_o']
     from sklearn.model_selection import train_test_split
     X_train, X_test, y_train, y_test = train_test_split(x, y, test_size = 0.3)
     regressor = RandomForestRegressor(n_estimators = 100, random_state = 0)
     regressor.fit(X_train, y_train)
     y_pred = regressor.predict(X_test)
     from sklearn.metrics import r2_score,mean_squared_error
     mse = mean_squared_error(y_test,y_pred)
     rmse = np.sqrt(mse)
     r2 = r2_score(y_test,y_pred)
     print(r2 , rmse)
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
0.9615314903484582 286753.955574708
```

R square Value :- 0.96 . Far Better than Both Intuitive and Extended model. Moreover better than Neural Network as well.



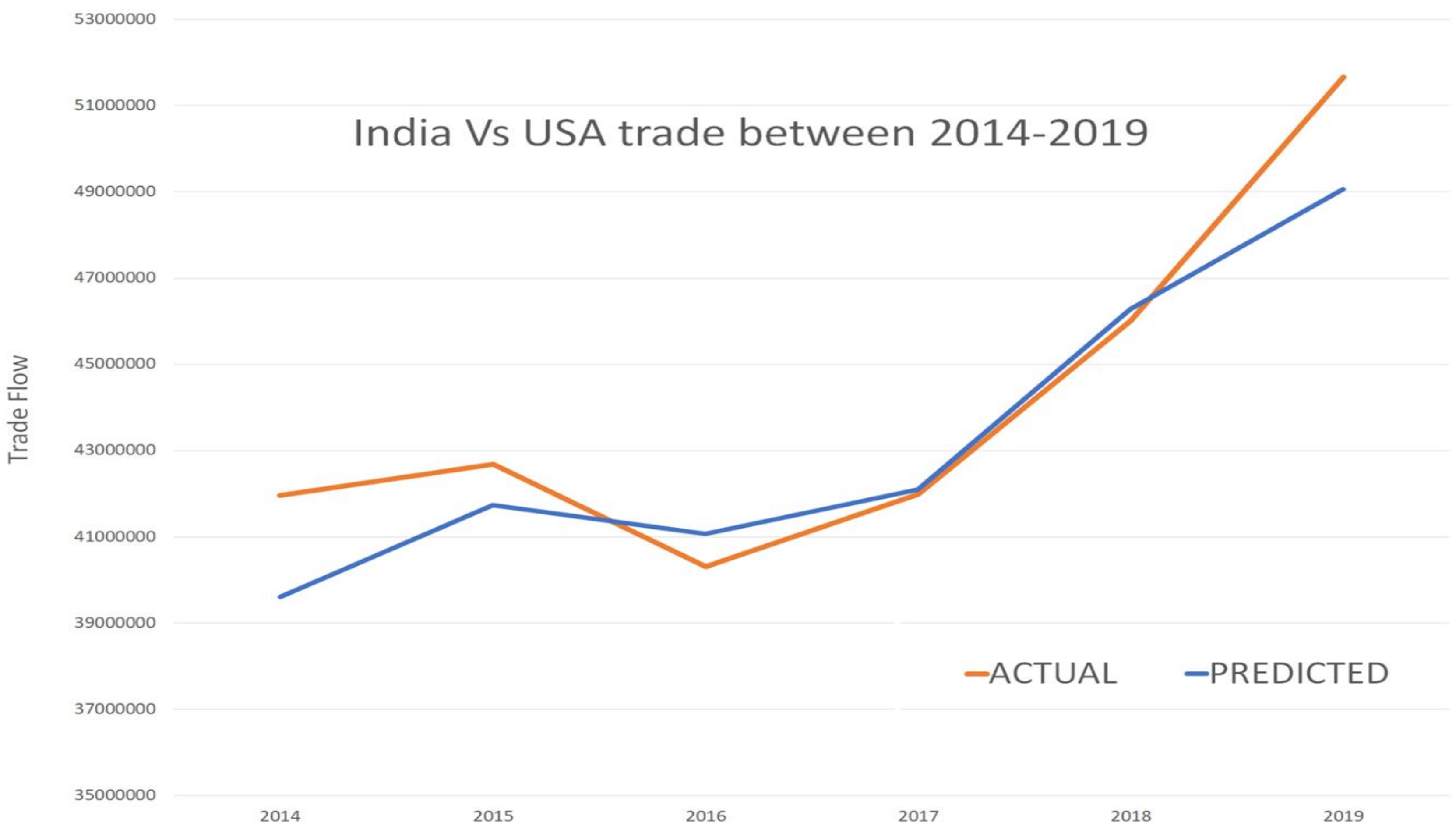
Depiction of the Tree

```
Variable: gdpcap_d                Importance: 0.2
Variable: iso3num_d               Importance: 0.15
Variable: gdpcap_o                Importance: 0.06
Variable: distwces                Importance: 0.05
Variable: comrelig                Importance: 0.05
Variable: rta_coverage            Importance: 0.04
Variable: rta_type                Importance: 0.02
Variable: comlang_off             Importance: 0.01
Variable: comlang_ethno           Importance: 0.01
```

Variable Importance Table

```
from sklearn.tree import export_graphviz
import pydot
tree = regressor.estimators_[5]

[58] feature_list = list(x.columns)
     export_graphviz(tree, out_file = 'tree.dot', feature_names =feature_list , rounded = True, precision = 1)

[59] (graph, ) = pydot.graph_from_dot_file('tree.dot')
     graph.write_png('tree.png')

[61] importances = list(regressor.feature_importances_)
     feature_importances = [(feature, round(importance, 2)) for feature, importance in zip(feature_list, importances)]

[62] feature_importances = sorted(feature_importances, key = lambda x: x[1], reverse = True)

[67] [print('Variable: {:20} Importance: {}'.format(*pair)) for pair in feature_importances];
```

Further Code Of Random Forest Regression


India Vs USA trade between 2014-2019
Trade Flow
53000000
51000000
49000000
47000000
45000000
43000000
41000000
39000000
37000000
35000000
2014
2015
2016
2017
2018
2019
ACTUAL
PREDICTED

Conclusions

- Neural networks have a high degree of accuracy in prediction compared to R-Square within the Gravity model.
- Model’s explanatory power increased by using the Fixed effect model and it clearly underlines the importance of multilateral resistance terms.
- PPML is more preferred to Other econometric methods.
- Neural network RMSE is much lower than gravity ones which clearly indicates that in future we should move onto use the machine learning models for estimating international trade patterns as it helps in dealing with large datasets which have numerous interconnected variables
- We saw that Intuitive Gravity Model is basic one which has R square value of 0.71. By applying fixed effect including multilateral terms we got to see R square value increased to 0.87.
- We also applied new methods like PPML,NBPML,GPML and saw their results.
- Finally we applied neural network and Random Forest Regression and we saw R square value to be more than intuitive Gravity Model which clearly states model is better.

- ❖ This will be a useful aid to policymakers, analysts and firms engaged in the international trade business and provide impetus for all players in the industry to gauge the effects of this trade collaboration
- ❖ This will be useful to policy makers and analysts and firms engaged of country who are involved in international trade to measure the effects of trade variables and improve their policies.
- ❖ Factors like Common Border , language play a major role specifically for india in its trade and estimation of these trades using neural network model offers various ways in which country can measure its impact and possible sustainability of such partnership.
- ❖ Importance of Machine learning algorithms increases as it helps in time of trade disruption for example Brexit case.
- ❖ Both supervised and Unsupervised Algorithm provide useful insights and high degree of accuracy.

References

Base Paper: Neural Network Analysis of International Trade

Isaac Wohl and JIm Kennedy: Neural Network Analysis of International Trade

Bhattacharyya, R., Banerjee, T., 2006. Does the gravity model explain India's direction of trade? A panel data approach. Working Paper No. 2006-09-01, Indian Institute of Management, Ahmadabad.

Anderson, J.E., 2011. The Gravity Model. Annual Review of Economics 3(1), 133-160.

Bhattacharya, S. K., 2004. Does Bangladesh benefit from preferential trade with India? a gravity analysis. Economic and Political Weekly 39 (48), 5152-5162.

Yotov, Y.V. A simple solution to the distance puzzle in international trade. Econ. Lett. 2012, 117, 794–798

Anderson, James and Eric van Wincoop. "Gravity with Gravitas: A Solution to the Border Puzzle." American Economic Review, vol. 93, no. 1, March

Head, Keith and Thierry Mayer. "Gravity Equations: Workhorse, Toolkit, and Cookbook." CEPII Working Paper, September 2013

Anderson, J. 1979. "A Theoretical Foundation for the Gravity Model." American Economic Review, 69(1): 106-116

Datasets

1. https://data.imf.org/?sk=9d6028d4-f14a-464c-a2f2-59b2cd424b85&sId=1390030341854
2. https://data.worldbank.org/
3. https://commerce.gov.in/
4. http://www.cepii.fr/cepii/en/bdd_modele/bdd.asp
5. https://kof.ethz.ch/en/forecasts-and-indicators/indicators/kof-globalisation-index.html

Thank You