

# GDP growth Prediction with OLS & Penalised Regression

FEM11149 - Introduction to Data Science

Daphne, Riya, Vera, Zsofi

September 29, 2025

## Introduction

This report investigates the factors that drive economic performance across countries. Our research question concerns whether economic and demographic indicators can be used to predict GDP growth, both by explaining variation in the continuous growth rate and by identifying the most relevant predictors. This question is important for policy and decision-making, since governments planning to increase spending often rely on borrowing, which requires sufficient future growth to ensure repayment. Understanding which conditions are associated with stronger growth therefore provides valuable insights for policymakers, international institutions, and investors.

## Data

The data were obtained from the World Bank's Jobs Indicators Database and Balance of Payments (BoP) statistics, which provide internationally comparable measures of economic and demographic conditions. The dataset covers 150 countries and includes 74 variables that capture a broad range of characteristics. These include demographic indicators such as fertility rates, life expectancy, and age dependency ratios; economic variables such as GDP per capita, net trade, and population growth; employment-related indicators such as sectoral employment shares and unemployment rates; and infrastructure or institutional measures such as access to electricity, internet usage, and credit registry coverage. The outcome of interest is GDP growth (annual %), which measures the year-on-year percentage change in a country's gross domestic product. This indicator reflects how fast an economy is expanding or contracting and is a central measure of economic performance. In this report it serves as the dependent variable and is primarily analysed in its continuous form, allowing us to study how variation in demographic, economic, employment, and institutional factors is associated with differences in growth rates across countries. In one part of the analysis, GDP growth is also recoded into a binary measure using a 2.7 percent benchmark to distinguish higher- from lower-growth countries, but the main focus remains on the continuous specification.

## Methodology

### Data Cleaning

The dataset initially contained 74 variables, many of which had long names with special characters such as % which was replaced with "." by using the function `names()` to replace these characters. The dependent variable was renamed to `gdp_annual_growth`, and an important independent variable of interest was renamed to `net_trade_bop_usd` for clarity and ease of reference. Additional cleaning steps included data type conversion for relevant variables to numeric, and the Country variable to a factor; and lastly, missing data or NA values were omitted. After cleaning, the dataset consisted of 150 countries with 74 numeric predictors.

### Model 1: OLS with Stepwise Backward Selection

We began with a multiple linear regression (OLS) framework to model the relationship between GDP growth and the set of significant predictors from the dataset. The general form of the model is:

$$\text{GDP Growth Annual}_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \epsilon_i$$

To identify the most significant predictors, we employed stepwise backward selection using the AIC. The process iteratively removes the least significant variable at each step until the optimal model is reached. To further improve prediction we created a nonlinear term: squared term for fertility rate (fertility\_sq) which would capture potential nonlinear effects of population dynamics. This extended the regression model as follows:

$$\text{GDP Growth Annual}_i = \beta_0 + \beta_1 \cdot \text{Fertility}_i^2 + \beta_2 \cdot \text{Net Trade}_i + \sum_{j=3}^m \beta_j X_{sel,j,i} + \epsilon_i$$

Including this nonlinear term helped improve model fit by allowing for curvature in the relationship between fertility and GDP growth.

### Model 2: LASSO Regression

OLS can perform well in small, clean datasets but becomes unstable and unreliable with high-dimensional or highly correlated predictors, which was the case in our variables in the dataset. Common issues with OLS include overfitting due to too many predictors relative to the sample size, multicollinearity leading to inflated standard errors and unreliable coefficient estimates. Additionally, omitted variable bias is caused by arbitrarily removing variables to fix multicollinearity resulting to biased estimates. To address these limitations, we implemented LASSO (Least Absolute Shrinkage and Selection Operator) regression, which performs automatic variable selection and regularization in a single step.

$$\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

Where:

$\lambda$ : is the penalty parameter controlling the amount of shrinkage.

Larger  $\lambda$ : forces more coefficients toward zero, effectively performing variable selection.

However, we used the same variables identified in the OLS model in order to observe which coefficients were shrunk to zero by LASSO and compare predictive performance against the OLS model.

In order to evaluate out-of-sample performance, we split data into a training set with 110 observations and a test set with 40 observations. Additionally, 10-fold cross-validation was applied to tune the optimal penalty parameter  $\lambda$ . This process ensures that the model generalizes well to unseen data and avoids overfitting.

### Model 3: Ridge and Elastic Net Regression

We further extended the analysis using Ridge regression and Elastic Net, both of which are regularization techniques designed to handle high-dimensional datasets. Ridge Regression: Adds an L2 penalty term that shrinks coefficients closer to zero but does not set them exactly to zero:

$$\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\}$$

Elastic Net: Combines both L1 (LASSO) and L2 (Ridge) penalties:

$$\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \left[ \alpha \sum_{j=1}^p |\beta_j| + (1 - \alpha) \sum_{j=1}^p \beta_j^2 \right] \right\}$$

Where:

Gamma ( $\alpha$ ) controls the mix between LASSO and Ridge (0 = pure Ridge, 1 = pure LASSO).

### Models with Lambda Selection: $\lambda_{\min}$ vs $\lambda_{1se}$

The `cv.glmnet()` function was used to determine two key lambda values:

$\lambda_{\min}$ : The penalty value that produces the lowest cross-validation error. This is referred to as the best lambda due to the lowest MSE. However, this leads to an overfitted predictive model.

$\lambda_{1se}$ : The largest  $\lambda$  within one standard error of the minimum error. Produces a simpler, more regularized model, which is less prone to overfitting and hence easier to interpret.

*Evaluation Metrics* To compare models, we used both regression and classification performance metrics:

Regression Metrics:

- Root Mean Squared Error (RMSE):

Classification Metrics (for Logistic Ridge Regression):

- Accuracy: shows the proportion of correctly classified observations.
- Confusion Matrix: Breakdown of true vs. predicted classifications.
- Area Under the Curve (AUC): Measures the model's ability to discriminate between classes.
- ROC Curves: Visualization of classification performance across thresholds.

## Results

Our initial attempt at a standard linear regression encountered perfect multicollinearity, meaning some explanatory variables were exact combinations of others, preventing the model from estimating unique coefficients. To address this, we removed redundant variables while still capturing the sectoral structure of the economy and retained Net Trade in goods and services (BoP, US\$). This regression ran successfully but yielded a very low  $R^2$  of 3.87, indicating almost no variation in GDP growth was explained. To improve performance, we applied stepwise regression, which systematically selects variables contributing meaningfully to explaining GDP growth while balancing model complexity and fit. Including Net Trade in the regression did not improve the overall fit, suggesting that trade balances, while theoretically important, added little predictive power beyond other correlated indicators. The baseline regression explained a meaningful but incomplete share of growth dynamics, achieving an adjusted  $R^2$  of 0.44 and a residual standard error of 3.3. We tested for nonlinear fertility effects by including a squared fertility term, motivated by the hypothesis that very low fertility can limit labor supply, very high fertility can strain resources, and moderate fertility may be most favorable for growth. The adjusted  $R^2$  barely changed (0.4385 vs. 0.439), and the residual standard error remained virtually identical (3.327), indicating this nonlinear term did not provide additional explanatory value.

Recognizing the limitations of OLS, we next applied LASSO regression. This stabilized coefficient estimates by shrinking less important predictors toward zero. Standardizing predictors was essential, as variables measured on different scales would otherwise be unfairly weighted. Standardization altered which predictors were selected, highlighting its effect on interpretability and variable retention. Compared to OLS, LASSO provided a clearer view of the variables most strongly associated with GDP growth. We extended the penalized approach using ridge and elastic net regression. Cross-validation selected optimal penalty parameters, and the best elastic net solution converged to  $\alpha \approx 0$ , effectively reducing to ridge regression. Both methods produced nearly identical results:  $RMSE \approx 3.98$  and  $R^2 \approx 0.19$  on the training set, and  $RMSE \approx 2.54$  on the test set. These models predict GDP growth with an error margin of 2.5–4.5 percentage points. While explanatory power is modest, the models consistently identified relevant predictors: labour market indicators (employment shares and unemployment, negatively associated), trade performance (export value and volume, positive), and business environment measures (credit depth positive; time to start a business and pay taxes negative). Elastic Net with  $\lambda_{1se}$  highlighted fewer, stronger predictors, producing a parsimonious model, whereas ridge regression distributed influence across many variables, retaining all with small but nonzero weights.

Table 1: Performance Comparison

Model	lambda	alpha	RMSE
Ridge $\lambda$ .min	23.1013	0	2.662008
Ridge $\lambda$ .1se	100000.0000	0	2.591956
ElasticNet $\lambda$ .min	23.1013	0	2.662008
ElasticNet $\lambda$ .1se	100000.0000	0	2.591956

To complement these regressions, we recast the task as a binary classification, defining countries as “high-growth” if GDP growth exceeded 2.7%. Ridge logistic regression with  $\lambda$  min achieved 62.5% accuracy and AUC 0.81, correctly identifying most high-growth countries but showing lower specificity. In contrast,  $\lambda$  1se performed worse, with accuracy dropping to 47.5% and AUC of 0.5, effectively random. Using a smaller training set further reduced predictive performance (45% accuracy), demonstrating the importance of sufficient data. Taken together, ridge regression with  $\lambda$  min provides the most reliable predictions of GDP growth. While explanatory power is modest, it avoids OLS pitfalls such as multicollinearity and overfitting. Expanding the dataset by adding more countries (rows) would likely improve predictive reliability, whereas adding more variables risks redundancy. Predicting continuous GDP growth is valuable for economic planning, market analysis, and investment outlooks, while the binary “Growing More” classification is more actionable for policy targeting, as it identifies countries likely to achieve sustained growth above a key threshold.

## Conclusion and Discussion

In our analysis we considered various economic, demographic and developmental factors of countries with the aim of predicting GDP growth across countries. The methods we used to investigate this were multiple linear regression and penalized regression techniques. Across all models, the ridge regression with  $\lambda$  1SE consistently was the most reliable model for prediction. While the different models all had moderate results when it comes to explanatory power, the ridge regression was able to mitigate the multicollinearity as well as the overfitting issues the OLS regression models faced. As a result, the ridge regression provided us with a more stable and interpretable set of predictor variables. Moreover, elastic net produced very similar results, as the optimal solution for the hyper parameter converged to a ridge regression with ( $\alpha \approx 0$ ). This shows that combining the lasso with the ridge penalties did not offer any additional advantage for interpreting our dataset. The stepwise OLS regression only captured part of the variation in GDP growth, with the addition of the nonlinear squared fertility term not aiding in that.

The penalized methods more clearly defined the key drivers of growth and thereby allowed us to answer our research question, namely: labor market indicators (employment shares and unemployment, negatively associated), trade performance (export value and volume, positive), and business environment measures (credit depth positive; time to start a business and pay taxes negative). Additionally, using ridge logistic regression to predict either high or low growth in GDP for countries had a reasonable accuracy (62.5%) and good discrimination (AUC = 0.81), though the  $\lambda$  1SE model performed poorly due to oversimplification. Predicting continuous GDP growth is valuable for economic planning, market analysis and investment decisions, while the binary classification for growing ‘more’ or ‘less’ is more relevant for policy making as it allows governments to identify if they are meeting economic targets or if their economies demand closer attention. In this research there were some limitations, as there might be omitted variable bias, potential nonlinear interactions that were not captured in our models and the data is only cross-sectional. In future research longitudinal data could be included to help capture growth dynamics over time and non-linear research methods could be utilized. In conclusion, the predictors mentioned above influence GDP growth as shown by penalized regression methods, specifically ridge regression. Overall ridge regression is a robust and interpretable method for predicting GDP growth in high-dimensional datasets.

## Appendix

```
#stepwise backward selection
base_model <- lm(gdp_growth_annual ~ . - Country, data = df)
step_model <- stepAIC(base_model, direction = "backward")
final_formula <- update(formula(step_model), . ~ . + net_trade_bop_usd)

#OLS MODEL
reg_final <- lm(final_formula, data = df, singular.ok = FALSE)

#NON-LINEAR VARIABLE MODEL: Squared fertility
df <- df %>% mutate(fertility_sq = Fertility.rate..total..births.per.woman.^2)
nonlinear_formula <- update(formula(step_model), . ~ . + fertility_sq +
                             net_trade_bop_usd)
reg_nonlinear <- lm(nonlinear_formula, data = df, singular.ok = FALSE) #Fit the model

#LASSO MODEL with same vars as OLS model
X <- model.matrix(nonlinear_formula, data = df)[, -1]
y <- df$gdp_growth_annual

set.seed(100) #set seed for consistent results
lasso_cv <- cv.glmnet(X, y, alpha = 1, maxit = 1e6)
coef(lasso_cv, s = "lambda.min")

#TESTING AND TRAINING SPLIT
set.seed(100)
train_idx <- sample(1:nrow(df), 110)
train_df <- df[train_idx, ]
test_df <- df[-train_idx, ]

X_train <- model.matrix(gdp_growth_annual ~ . - Country - fertility_sq, data = train_df)[, -1]
y_train <- train_df$gdp_growth_annual
X_test <- model.matrix(gdp_growth_annual ~ . - Country - fertility_sq, data = test_df)[, -1]
y_test <- test_df$gdp_growth_annual

#CROSS VALIDATION
fitControl <- trainControl(method="repeatedcv", number=10, repeats=5)

#RIDGE REG MODEL
ridge_model <- train(x=X_train, y=y_train, method="glmnet",
  tuneGrid=expand.grid(alpha=0, lambda=seq(0.0001, 1000, length=100)),
  trControl=fitControl, preProcess=c("center", "scale"))

#ELASTIC NET REG MODEL
elastic_model <- train(x=X_train, y=y_train, method="glmnet",
  tuneGrid=expand.grid(alpha=seq(0,1,length=10), lambda=seq(0.0001, 1000, length=50)),
  trControl=fitControl, preProcess=c("center", "scale"))

#RESULT COMPARISION: RMSE
res <- resamples(list(Ridge=ridge_model, ElasticNet=elastic_model))

#Ridge using cv.glmnet to get lambda.min and lambda.1se
set.seed(100)
```

```

r_cv <- cv.glmnet(x = X_train, y = y_train, alpha = 0, #Ridge
  lambda = 10^seq(-3, 5, length.out = 100), #Search over a wide range of lambdas
  nfolds = 10) #10-fold cross-validation

r_min_model <- glmnet(X_train, y_train, alpha = 0, lambda = r_cv$lambda.min) #Fit Ridge at lambda.min
r_1se_model <- glmnet(X_train, y_train, alpha = 0, lambda = r_cv$lambda.1se) #Fit Ridge at lambda.1se

best_alpha <- elastic_model$bestTune$alpha #best alpha from caret Elastic Net model

#Elastic Net using cv.glmnet with optimal alpha
set.seed(100)
en_cv <- cv.glmnet(x = X_train, y = y_train, alpha = best_alpha, # Optimal mix between Lasso and Ridge
  lambda = 10^seq(-3, 5, length.out = 100), nfolds = 10)

#Fit Elastic Net at lambda.min and lambda.1se
en_min_model <- glmnet(X_train, y_train, alpha = best_alpha, lambda = en_cv$lambda.min) #Fit EN at lambda.min
en_1se_model <- glmnet(X_train, y_train, alpha = best_alpha, lambda = en_cv$lambda.1se) #Fit EN at lambda.1se

#Predictions and Performance
r_min_pred <- predict(r_min_model, newx = X_test) # Ridge $$min
ridge_1se_pred <- predict(r_1se_model, newx = X_test) # Ridge $$1se
en_min_pred <- predict(en_min_model, newx = X_test) # Elastic Net $$min
en_1se_pred <- predict(en_1se_model, newx = X_test) # Elastic Net $$1se

compare_results <- data.frame( #table
  Model = c("Ridge $$min", "Ridge .1se", "ElasticNet $$min", "ElasticNet $$1se"),
  lambda = c(r_cv$lambda.min, r_cv$lambda.1se, en_cv$lambda.min, en_cv$lambda.1se),
  alpha = c(0, 0, best_alpha, best_alpha),
  RMSE = c(rmse(y_test, r_min_pred), rmse(y_test, ridge_1se_pred), rmse(y_test, en_min_pred), rmse(y_test, en_1se_pred)))

#Logistic Model: dependent variable
df <- df %>% mutate(growing_more = factor(ifelse(gdp_growth_annual > 2.7, 1, 0)))

set.seed(1) # Train-test split USING LOG MODEL
log_train_idx <- sample(1:nrow(df), 110)
train_log <- df[log_train_idx, ]
test_log <- df[-log_train_idx, ]

x_train_log <- model.matrix(growing_more ~ . - Country - gdp_growth_annual,
  data = train_log)[, -1]
y_train_log <- train_log$growing_more
x_test_log <- model.matrix(growing_more ~ . - Country - gdp_growth_annual,
  data = test_log)[, -1]
y_test_log <- test_log$growing_more
cv_ridge_log <- cv.glmnet(x_train_log, y_train_log, family="binomial", alpha=0, nfolds=10)

# ROC & AUC
pred_prob <- predict(cv_ridge_log, newx=x_test_log, s="lambda.min", type="response")
roc_curve <- roc(y_test_log, as.numeric(pred_prob))
plot(roc_curve, col="blue", lwd=2)
auc(roc_curve)

```

Results from Ridge and Elastic Net Models

```
##
## Call:
## summary.resamples(object = res)
##
## Models: Ridge, ElasticNet
## Number of resamples: 50
##
## MAE
##           Min. 1st Qu.  Median    Mean 3rd Qu.    Max. NA's
## Ridge      1.22531 1.952380 2.346699 2.553673 2.723158 5.506525    0
## ElasticNet 1.30341 1.784603 2.365549 2.571429 2.903550 5.546618    0
##
## RMSE
##           Min. 1st Qu.  Median    Mean 3rd Qu.    Max. NA's
## Ridge      1.534019 2.671543 3.251213 4.011678 4.032289 11.76438    0
## ElasticNet 1.595691 2.368541 3.121268 3.986613 4.266178 11.58051    0
##
## Rsquared
##           Min. 1st Qu.  Median    Mean 3rd Qu.    Max. NA's
## Ridge      7.504098e-05 0.06041189 0.1641724 0.2089616 0.3349000 0.6617281    0
## ElasticNet 8.798955e-05 0.03084125 0.1093559 0.1736591 0.3015192 0.5740265    0
```

Table 2: Ridge and Elastic Net Coefficients

	Coefficient_EN	Coefficient_R
(Intercept)	2.5013600	2.5013600
Access.to.electricity....of.population.	0.0701572	0.0557938
Adolescent.fertility.rate..births.per.1.000.women.ages.15.19.	-0.0428681	-0.0359289
Age.dependency.ratio....of.working.age.population.	-0.0647116	-0.0537030
Contributing.family.workers..female....of.female.employment...modeled.ILO.estimate.	-0.0115788	-0.0066636
Contributing.family.workers..male....of.male.employment...modeled.ILO.estimate.	0.0643089	0.0504977
Contributing.family.workers..total....of.total.employment...modeled.ILO.estimate.	0.0507890	0.0406164
Depth.of.credit.information.index..0.low.to.8.high.	0.1133024	0.0891822
Employers..female....of.female.employment...modeled.ILO.estimate.	-0.2271122	-0.1835858
Employers..male....of.male.employment...modeled.ILO.estimate.	-0.1757794	-0.1447510
Employers..total....of.total.employment...modeled.ILO.estimate.	-0.2049322	-0.1671473
Employment.in.agriculture....of.total.employment...modeled.ILO.estimate.	0.0005907	0.0014696
Employment.in.agriculture..female....of.female.employment...modeled.ILO.estimate.	-0.0235874	-0.0165128
Employment.in.agriculture..male....of.male.employment...modeled.ILO.estimate.	-0.0054401	-0.0031550
Employment.in.industry....of.total.employment...modeled.ILO.estimate.	0.0696806	0.0550962
Employment.in.industry..female....of.female.employment...modeled.ILO.estimate.	0.0178803	0.0147684
Employment.in.industry..male....of.male.employment...modeled.ILO.estimate.	0.0694410	0.0550566
Employment.in.services....of.total.employment...modeled.ILO.estimate.	-0.0310144	-0.0257322
Employment.in.services..female....of.female.employment...modeled.ILO.estimate.	0.0181899	0.0120275
Employment.in.services..male....of.male.employment...modeled.ILO.estimate.	-0.0380059	-0.0317827
Export.value.index..2000...100.	0.1150765	0.0953473
Export.volume.index..2000...100.	0.1050490	0.0881072
Fertility.rate..total..births.per.woman.	-0.0670382	-0.0542433
Fixed.broadband.Internet.subscribers..per.100.people.	-0.0191313	-0.0143411
GDP.per.capita..constant.2005.US..	-0.0254849	-0.0209715
GDP.per.capita..PPP..constant.2011.international...	-0.0364230	-0.0274300
Individuals.using.the.Internet....of.population.	0.0146824	0.0121660
Labor.force.participation.rate..female....of.female.population.ages.15....modeled.ILO.estimate.	0.0451133	0.0379611

	Coefficient_EN	Coefficient_R
Labor.force.participation.rate..male....of.male.population.ages.15....modeled.ILO.estimate.	0.0597413	0.0515819
Labor.force.participation.rate..total....of.total.population.ages.15....modeled.ILO.estimate.	0.0646755	0.0551096
Labor.force..total	0.0385018	0.0356124
Life.expectancy.at.birth..female..years.	0.0594349	0.0471368
Life.expectancy.at.birth..male..years.	0.0731303	0.0569594
Mobile.cellular.subscriptions..per.100.people.	0.0363365	0.0296924
Own.account.workers..female....of.female.employment...modeled.ILO.estimate.	0.0212088	0.0179760
Own.account.workers..male....of.male.employment...modeled.ILO.estimate.	0.0366066	0.0319913
Own.account.workers..total....of.male.employment...modeled.ILO.estimate.	0.0320261	0.0275144
Population.ages.0.14....of.total.	-0.0286885	-0.0252906
Population.ages.0.14..total	0.0420252	0.0376708
Population.ages.15.64....of.total.	0.0634957	0.0532871
Population.ages.15.64..total	0.0407409	0.0371914
Population.ages.65.and.above....of.total.	-0.0168069	-0.0120189
Population.ages.65.and.above..total	0.0239317	0.0244014
Population.density..people.per.sq..km.of.land.area.	-0.0393692	-0.0293202
Population.growth..annual...	-0.0355038	-0.0307759
Population..total	0.0402938	0.0367877
Private.credit.bureau.coverage....of.adults.	0.0093624	0.0084719
Public.credit.registry.coverage....of.adults.	0.0330223	0.0269860
Rural.population	0.0479479	0.0423686
Rural.population....of.total.population.	0.0207739	0.0176353
Self.employed..female....of.female.employment...modeled.ILO.estimate.	-0.0036711	-0.0008946
Self.employed..male....of.male.employment...modeled.ILO.estimate.	0.0293456	0.0247915
Self.employed..total....of.total.employment...modeled.ILO.estimate.	0.0267798	0.0224407
Tax.payments..number.	0.0111800	0.0055072
Telephone.lines..per.100.people.	0.0061783	0.0039921
Time.required.to.enforce.a.contract..days.	0.0279542	0.0208263
Time.required.to.start.a.business..days.	-0.1011281	-0.0794829
Time.to.prepare.and.pay.taxes..hours.	-0.0879252	-0.0679129
Unemployment..female....of.female.labor.force...modeled.ILO.estimate.	-0.0770622	-0.0675918
Unemployment..male....of.male.labor.force...modeled.ILO.estimate.	-0.0776047	-0.0666981
Unemployment..total....of.total.labor.force...modeled.ILO.estimate.	-0.0602497	-0.0544573
Unemployment..youth.female....of.female.labor.force.ages.15.24...modeled.ILO.estimate.	0.0288017	-0.0297182
Unemployment..youth.male....of.male.labor.force.ages.15.24...modeled.ILO.estimate.	-0.0552899	-0.0497392
Unemployment..youth.total....of.total.labor.force.ages.15.24...modeled.ILO.estimate.	-0.0396159	-0.0380645
Urban.population	0.0260422	0.0257748
Urban.population....of.total.	-0.0208563	-0.0176824
Vulnerable.employment..female....of.female.employment...modeled.ILO.estimate.	0.0076312	0.0081969
Vulnerable.employment..male....of.male.employment...modeled.ILO.estimate.	0.0494474	0.0413008
Vulnerable.employment..total....of.total.employment...modeled.ILO.estimate.	0.0447204	0.0370547
Wage.and.salaried.workers..female....of.female.employment...modeled.ILO.estimate.	0.0035685	0.0008137
Wage.and.salaried.workers..male....of.male.employment...modeled.ILO.estimate.	-0.0294141	-0.0248617
Wage.and.salaried.workers..total....of.total.employment...modeled.ILO.estimate.	-0.0268128	-0.0224992
net_trade_bop_usd	0.0090290	0.0084332