An Analysis of the Impact of EU Membership on the Economic Development of Ireland

Piush Vaish

Abstract— Recent economic recession and Brexit has raised doubts about the relationship with EU among member nations. This paper analyzes the public data from World Bank to determine the impact to the economic development of Ireland since joining the EU on 1st January 1973. The paper shows that EU membership has a positive relationship with the development of Ireland. Economy has transformed from agricultural dependent to industrial focused with annual GDP growth which can be attributed to the population growth as well increase in exports of merchandise, goods and service, inflow of foreign direct investment, trade and gross domestic savings. The number of patents and trademarks has also increased for residents over the years. The life expectancy of people living in Ireland has increased while the death rate and birth rate has decreased. There has been a social shift with more people living in urban areas ,entering third level education and greater participation of females in the labour force. However, there are still some issues which Ireland has to tackle such as unemployment among 15-24 years old males and part-time employment among adult females.

I. INTRODUCTION

Ireland became a member of the European Union (EU) and joined the single market on 1st January 1973. Before accession to the bloc, Ireland had decades of an under-achieving economy which was heavily dependent on the UK. Since then it has transformed into a prosperous and confident country which is a major influence in the global politics. The economy has transformed from agricultural dependent to one driven by the tech industry and global exports. The membership has also affected every part of Irish society from the way the citizens work, travel or even shop[1].However, the recent political turmoil of Brexit and 2008 recession crisis has left certain citizens to wonder the importance of the membership. Sometimes, there is a doubt in EUs ability to provide a good living standard.

Economic development requires economic growth to reflect economic as well as social growth. The common indicators to assess the development includes:

- Economic growth by measuring the annual increase in Gross Domestic Product (GDP), GDP per capita
- Employment and Unemployment figures to determine the number of people able to find work
- Division of the countrys economy into primary, secondary and tertiary industries
- Inflation to determine the increase of price of goods and services.
- Demographic study such as population, growth, literacy and life expectancy[2]

This analysis uses the economic development indicators from world bank to show the benefits of joining the EU.

II. DATASET

The data is a collection of comparable statistics regarding development gathered by the World Bank from different official sources. The data constitutes different national, regional and global indicators such as economic growth, social development, macroeconomic vulnerability and debt for 217 economies. It is a time series data and chronological coverage is 1960-2017. It was released on June 11, 2010 and last updated on November 14, 2018. It is licensed under CC-BY 4.0 and classified as Public under the Access to Information Classification Policy. The users from inside the bank and outside the bank can access the dataset. The data was downloaded in csv file version along with the metadata[3]. There is no literature using the dataset for the analysis of economic development of Ireland.

III. SUITABLE TECHNIQUES

The following are different steps required for a comprehensive analysis:

- Data cleaning deals with missing values, noise, extreme, incorrect or duplicate records. The aim is to have accurate, correct, complete and consistent data while reducing the bias into the dataset. This helps to increase the data consistency, integrity and quality while eliminating meaningless data.
- Data transformation includes transforming features from different formats into formats or units that are more useful for statistical methods. This also helps to process the data easily and gain more information. The common techniques include binning/ bucketing the data to reduce the effect of minor errors, convert categorical variables into Boolean values, centering the values by subtracting the mean and scaling the centered value by dividing with standard deviation.
- Data reduction includes removing records and features which are not required as well as reorganizing the data in an efficient and tidy manner for analysis[4].
- Quantitative methods using statistical, numerical and mathematical techniques to provide quantifiable and objective results. The data is summarized in order to apply generalizations for the population and reproduce result without introducing any bias. The most common quantitative methods include descriptive statistical analyses and inferential statistical analyses. Exploratory data analysis examines the data for distribution, outliers by selecting the most appropriate statistics to describe the data and examine relationship and trend in the data.

Visualization is done to help the assessment of statistical models[5]. It helps to examine the variables before an inference is made from data. The descriptive statistics summarizes the data and help to identify patterns. It is important to identify the scale of measurement of the feature in order to apply appropriate descriptive statistic and produce a meaningful interpretation. The commonly used statistics are frequencies, mean, median, percentiles and mode. Inferential statistics make predictions by examining the difference and relationships between two or more samples of population. It is a complex analysis and is useful to test hypotheses and generalize results to population. Common tests include correlation to describe a mutual relationship between variables, regression to find if one feature is a good predictor of another and Analysis of Variance (ANOVA) to evaluate the statistical significance of means of two sampled groups[6].

IV. TECHNIQUE(S) EMPLOYED

The original data contains 63 columns and 4200024 rows. The columns include country name, indicator name, indicator code and observations from 1960-2017. The dataset is reduced by selecting the data from 1973 (the year when Ireland joined the EU) to 2017. The columns with country code and indicator codes are also deleted. Ireland is selected as the country name from the dataset to extract data relating to Ireland only. The data is transposed so that the dataset has year as index and indicator name as columns to help in formatting the dataset. The values for year are transformed into datetime and the values for indicator name are converted in format which suitable for the statistical and numerical techniques. The columns with missing values for 2015 are removed to ensure the recent data is available. The columns with more than 30 % missing values are also dropped. The new dataset has 45 rows and 374 columns. The data is further reduced by selecting 101 features which are important according to the literature. The techniques used include:

- Descriptive summary statistics
- Determine the normal distribution of a variable by calculating the skewness in the data and shapiro -wilk test
- Visualization to discover the trend over time and linear relationships. This helps to explore quick and easy exploration of the relationship the variables to each other
- Data chunking or binning is done by grouping a few continuous values in smaller number of bins
- Correlation analysis is a technique to determine the statistical measure of mutual association between random variables. The analysis helps to understand relationship between variables and provide foundation for building business and statistical models. The correlation coefficient measure strength of association.
- Simple Regression analysis is the quantitative prediction of a dependent variable for a given value of independent

variable. The relationship between the two variables is calculated using a straight line named regression line
- T-Test (Students T-Test) determines if the average of two means is statistically different to each other after comparing them. It tells the difference is significant or due to randomness. "T-score is the difference between the two groups and the difference within the groups. The larger t-score the more difference between groups and smaller t-score the groups are similar."
- Linear transformation technique such as Principal Component Analysis (PCA) determine the relationship using co-variance instead of correlation. It reduces data in meaningful way, explains the variance and generates meaningful output. The data reduction also helps with the reduction of dimensionality and hence helps to avoid overfitting.[7]

V. FINDINGS

Fig. 1 is a summary statistics in percentage. Since joining the EU, GDP growth has varied a lot because of the standard deviation of 4.60. The minimum value is -4.62 while the maximum value is 25.5. The mean annual GDP growth is 5.05. The population growth does not suffer from the same variability because the standard deviation is only 0.72. The mean of the population growth is 1.02. The distribution is

	GDP growth (annual %)	GDP per capita growth (annual %)	Population growth (annual %)
count	45.000000	45.000000	45.000000
mean	5.058516	3.990037	1.023614
std	4.609385	4.596338	0.721684
min	-4.626772	-5.874566	-0.427686
25%	2.984908	1.708913	0.526556
50%	5.141460	3.709681	1.015663
75%	7.186729	6.234040	1.509146
max	25.557269	24.376534	2.890960

Fig. 1. Summary Statistics for Annual GDP, GDP per capita and Population Growth

further confirmed by the Table I. The skewness for GDP is over 1 and hence the distribution has a longer right tail than the left and is highly skewed on the right. This shows that the growth has been more in the last few years than during the first few years of joining the EU. The skewness of population growth is 0.32. Therefore the distribution is approximately symmetric but the growth has increased over the years.

TABLE I
SKEWNESS TABLE FOR ANNUAL GDP, GDP PER CAPITA AND POPULATION GROWTH

Feature	Skewness
GDP growth (annual %)	1.70
GDP per capita growth (annual %)	1.68
Population growth (annual %)	0.32

Fig. 2 shows the annual growth for GDP and population over the years. It shows that GDP growth has mostly stayed over 0% except during the recession when it fell to the minimum value at -4.6%. The maximum annual population

growth was 2.89% right before the recession started and dropped significantly to 0.5% in 2008. Further, the population growth has an effect on GDP and vice versa. There is an outlier GDP growth of 26.3% for 2015. The reason for such a high rate of increase from 2014 is the relocation of multiple, large multinational companies in Ireland. These companies brought their intellectual property (IP) from other countries. The use of intellectual property to generate sales and the size of the companies contributed to such an increase[8].

Fig. 2. GDP Vs Population Growth

Fig. 3 shows that both birth rate and death have a negative linear relationship with the total population while life expectancy has a positive linear relationship. Death rate's R-value of -0.97 and Life expectancy R-value of 0.99 indicates that total population is heavily related to these variables. The relationship between death rate and life expectancy is negative with R-value of -0.97. This is because the life expectancy depends on people to live longer.

Fig. 3. Correlation Analysis of Total Population, Birth Rate, Death Rate and Life Expectancy

Table II shows that from 1973 - 1983, the rural population had a near perfect positive relationship with the total population with R-value of 0.998 although from 2007-2017 the R-value fell to 0.93. During this time period, the relationship between the urban and rural population also fell from R-value of 0.997 to 0.9141. Hence, the number of people living in urban areas has increased over the course of time. Fig. 4

TABLE II

CORRELATION ANALYSIS OF RURAL AND URBAN POPULATION

Population	R-value (1973-1983)	R-value (2007-2017)
Total,Rural	0.998	0.935
Total,Urban	0.999	0.998
Rural,Urban	0.997	0.914

shows that the simple linear equation is GDP = (-6.447e+11) + (1.973e+05) * Total Population with a standard error of 6640.5 (a possible uncertainty between the true GDP and the total population).F-statistics is 883.0 and Prob(F-statistics) is very small at 2.69e -30. Hence, there is a very small chance

that regression parameters are zero. Regression equation has validity in fitting the data.The coefficient of determination (R-squared value) determines the proportion of variance in GDP that can be accounted for by knowing total population and vice versa. The regression model explains 95.4% of the variation in GDP due to total population. Hence,regression model accounts for most of the variance and points of data fall close to the fitted regression line. Consequently, it is a very good model fit to the data.

Fig. 4. Linear Regression Analysis between GDP and Total Population

Fig. 5 shows that exports/imports of goods and service, inflow of foreign direct investment, trade and gross domestic savings have a positive association with GDP. On the other hand, gross national expenditure, expenditure by accounts relating to the consumer spending such as households and non-profit institutions serving households have a strong negative relationship. Similarly, the household's expenditure and domestic savings have a high negative association with the R-value of 0.99 while the national expenditure and household expenditure have a positive correlation with the R-value of 0.92.

Fig. 5. Correlation Analysis of Major Components of GDP

Fig. 6 shows the principal component analysis of annual GDP growth of EU from 1997 to 2017. 2015 is dropped

from the data because of the unusual GDP growth of Ireland. $PC1 = 0.40, PC2 = 0.18$ means that the first principal component explains 0.44 of the variance and the second principal component explains 0.18 of the variance.0.58 of the variance is explained by the two components. The plot shows that Ireland holds a unique position among the EU. The other countries show the same variance if they share a border with each other than far from other countries.Hence, annual GDP growth affects the countries in the same region e.g. France and Belgium are bordering countries and have almost the same annual GDP growth.
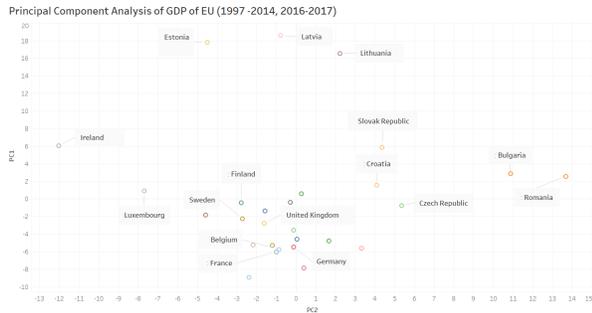

Fig. 6.   Principal Component Analysis of annual GDP growth rate of EU

Fig. 7 displays annual GDP growth of EU for 2017.Ireland has the highest GDP growth rate among the EU members at 7.80.Eastern European countries who have joined EU recently such as Romania and Bulgaria have the next highest GDP growth rate.Hence, these countries are benefiting from EU membership.
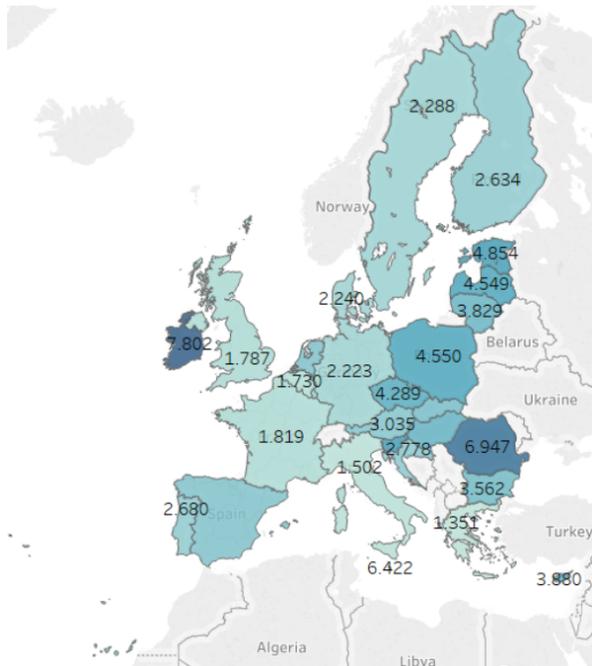

Fig. 7.   Visualization of Annual GDP Growth Rate of EU for 2017

Fig. 8 shows that the mean annual inflation is 0.13 with a standard deviation of 3.107 and a standard error of 0.93. The

annual inflation fell to -5.0 and -3.22 in 2009 and 2010 (years after the recession) with a huge increase (7.28) in 2015. Consumer price index has been positive after the recession and has been increasing annually.

	Variable	N	Mean	SD	SE	95% Conf.	Interval
0	Consumer price index (2010 = 100)	11.0	103.577218	1.958629	0.590549	102.261393	104.893042
1	Inflation, GDP deflator (annual %)	11.0	0.139181	3.107443	0.936929	-1.948427	2.226790

Fig. 8.   Summary Statistics of Consumer Price Index and Annual Inflation (2007 - 2017)

Fig. 9 is the t-test score for the patent's application made by both residents and non-residents.The difference between the patents by non-residents is higher than the residents with a mean of 642.29 and a small t-score of -3.0984. Hence, the two groups are three times as different from each other as they are within each other. It also means that the results are less likely to be repeated. An further analysis of the data shows that the patents for non-residents have been falling over the years with 2088 patents in 1973 to 85 patents in 2016. The patents for residents have been increasing and show that Irish are moving to a knowledge based economy. 2015 saw an increase in the number of patents by non-residents to 190 from 58 in 2014 as pointed out in the OECD report[8]

Independent t-test	results
Difference (Patent applications, residents - Patent applications, nonresidents) =	-642.296
Degrees of freedom =	86
t =	-3.0984
Two side test p value =	0.0026
Mean of Patent applications, residents > mean of Patent applications, nonresidents p value =	0.0013
Mean of Patent applications, residents < mean of Patent applications, nonresidents p value =	0.9987
Cohen's d =	-0.6606
Hedge's g =	-0.6548
Glass's delta =	-2.6205
r =	0.3169

Fig. 9.   T-test Statistics for Patents

Fig. 10 is the t-test score for the trademark applications made by both residents and non-residents.The difference between the trademark by non-residents is higher than the residents with a mean of 2746.52 and t-score of 12.21. Hence, the two groups are almost 12 times as different from each other as they are within each other. The number of trademark applications from non-residents has decreased from 3102 in 1973 to 1952 in 2016. Although the number of trademark application has increased for residents from 431 in 1973 to 1952 in 2016.This means that indigenous companies are growing in Ireland.

Independent t-test	results
Difference (Trademark applications, direct resident - Trademark applications, direct nonresident) =	-2746.52
Degrees of freedom =	86
t =	-12.2167
Two side test p value =	0
Mean of Trademark applications, direct resident > mean of Trademark applications, direct nonresident p value =	0
Mean of Trademark applications, direct resident < mean of Trademark applications, direct nonresident p value =	1
Cohen's d =	-2.6046
Hedge's g =	-2.5818
Glass's delta =	-4.8612
r =	0.7965

Fig. 10.   T-test Statistics for Trademark

Fig. 11 show the data chunking technique to understand school enrollment better. It shows that the enrollment for

tertiary and secondary schools have been increasing over the years. There is a decrease in enrollment for primary school which can be attributed to the declining birth rate. More people are going for the third-level education. A correlation analysis shows that there is no relationship between primary school enrollment and secondary school enrollment. However, there is a positive association between the secondary school and tertiary school enrollment with R-value of 0.922. Hence, most of the students progress to tertiary school if enrolled in secondary school. However, there is a small decrease in the relationship between females enrolling for tertiary education than males. The females R-value is 0.89 while for males is 0.91. Therefore, there is a small possibility that females do not complete third-level education than males.

	Year	Primary School Enrollment (% gross)	Secondary School Enrollment (% gross)	Tertiary School Enrollment (% gross)
0	1973-1976	423.704338	334.134544	52.558649
1	1977-1980	425.803055	355.128929	65.420521
2	1981-1984	389.415291	387.544609	76.159569
3	1985-1988	385.026024	390.841545	93.756491
4	1989-1992	395.908684	382.500755	112.410662
5	1993-1996	397.462769	432.541214	137.424765
6	1997-2000	386.473663	409.655342	169.292759
7	2001-2004	423.416916	410.703789	202.208099
8	2005-2008	428.786400	447.293747	222.669662
9	2009-2012	422.917145	489.715492	253.884865
10	2013-2016	406.690071	380.300758	233.962334

Fig. 11.   Data Transformation of School Enrollment

Fig. 12 is a visualization of the data for merchandise being grouped into bins of 4 years. In the first 10 years after joining the EU, Ireland used to import more merchandise than export it. Since then Ireland exports more than it imports. During the recession period of 2009-2012, Ireland's import dropped slightly due to lack of consumer demand. However, the exports have been steady and even increasing slightly since 2005. The R-value of 0.9892 between exports and imports of the merchandise shows that there is a strong positive association between the two.
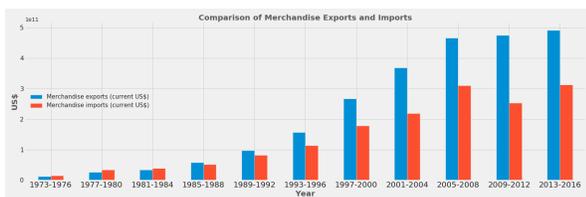

Fig. 12.   Data Transformation of Merchandise

Table III is divided into two equal periods of 20 years. From 1974-1995, Irish exports and imports were mostly dependent on the agricultural, food and ores and metal industry. Hence the economy was mostly dependent on land. However, in the recent 20 years, the economy has transformed towards manufacturing industry with the ratio increasing to 24.76. the food industry has seen a big drop with the ratio between exports and imports falling by almost 50%. This can be due to fewer food exports or more consumption because of the population growth.

Fig. 13 shows a negative association between the exports of manufactured goods and agricultural and food products

Merchandise	Ratio(1974 - 1995)	Ratio(1996 - 2017)
Agricultural raw materials	18.58	15.30
Food	54.74	23.88
Fuel	1.87	2.14
Ores & Metal	20.15	8.92
Manufactures	19.01	24.76

with the R-value of 0.93 and 0.95 respectively. However, there is a positive relationship between food exports and agricultural exports. Hence, agricultural products are mostly food related.

Fig. 13.   Correlation Analysis of Various Components of Merchandise

Fig. 14 shows that Ireland has become more of an exporter than an importer of the goods and services over time. This has helped to have a strong economy. Even during the recession, Ireland was able to export more than it imported. The gap has also increased between the two.
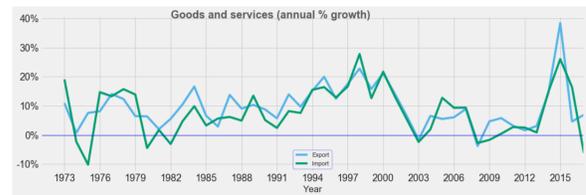

Fig. 14.   Visualization of Imports and Exports of Goods and Services

Table IV presents the mean ratio of female to male labor force participation rate since 1984. It is 65.02% with a standard deviation of 12.3. The minimum value is 44.8% and maximum value is 81.18%. The shapiro-wilks test for normality value is 0.910 and demonstrates that it is not a Gaussian distribution and hence the ratio of female to male in the labour force has increased over the years. Fig. 15 shows

Statistic	Value
mean	65.02
std	12.37
min	44.85
25%	54.27
50%	66.55
75%	77.12
max	81.18

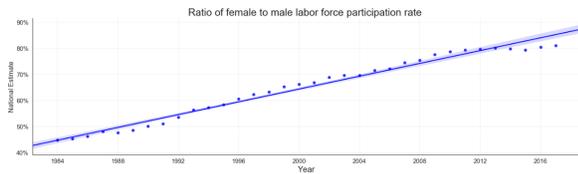relationship over time. It confirms that the participation of female has increased since 1984.

Fig. 15. Regression Plot of Ratio of Female to Male Labor Force Participation Rate

Fig. 17. Visualization of the Youth Unemployment since 1983

Fig. 16 displays the simple linear equation is Youth Employment to Population Ratio = 52.1476 + -0.5371 * Total Youth Unemployment with a standard error of 0.112 measuring the uncertainty (a possible difference between the true youth employment to population ratio and the total youth unemployment).The youth is someone who is 15-24 years old.F-statistics is 22.98 and Prob(F-statistics) is very small at 3.38e-05. Hence,the regression equation has some validity in fitting the data.R-squared value is 0.411. Therefore, the regression model explains 41.1% of the variation in youth employment to population ratio due to total youth unemployment. Hence, most of the variance is not accounted for by the regression model and the data points fall far away to the fitted regression line. Consequently, it is not a good model fit to the data.

most likely to be employed than females. The mean for the 15+ male population to 15+ female population is 62.7 and 40.9 with the standard error of 4.24 and 9.49 respectively.

Fig. 18. Summary Statistics of Employment to Population Ratio since 1983

Fig. 19 shows that females have a positive linear relationship with employment with R-value of 0.89 while males have a less strong relationship with R-value of 0.696. This less positive association between males and employment can be associated with the higher unemployment among the male youth population.

Fig. 19. Correlation Analysis of Employment to Population Ratio since 1983

Fig. 16. Linear Regression Analysis between the Youth Employment to Population Ratio and Total Youth Unemployment since 1983

Fig. 17 shows that young males are more likely to be unemployed than young females. The youth unemployment to its lowest level at almost 5% in 2000. It increased during the recession and reached the highest level in 2010 to almost 30%. As the economy is recovering the youth unemployment is falling again.

Fig. 18 shows that a mean of 51% from the total population are employed who are above the age of 15+. Males are
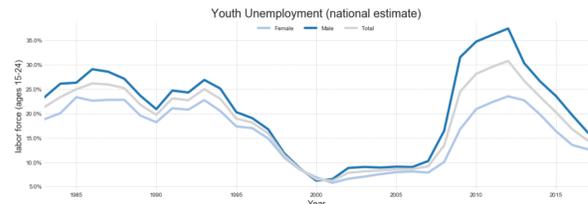
Fig. 20 shows that the simple linear equation is Employment to Population Ratio (15+) = 62.2377 + 0.0228 * Part-time employment among Males(15+). The standard error of 0.103 measuring the uncertainty (a possible difference between the true employment to population ratio and the part-time employment among male).F-statistics is 0.04915 and Prob(F-statistics) is 0.826. Therefore, there is 82.6 chance in 100 that all of the regression parameters are zero. This high value suggests that the regression equation is unable to fit the data. The independent variable is purely random with respect to the dependent variable. The regression model explains 0.1% of the variation in employment to population ratio due to part-time male employment. Consequently, it is not a good model fit to the data.

Fig. 21 shows that the simple linear equation is Employment to Population Ratio (15+) = 13.6868 + 1.1688 * Part-time employment Females (15+). The standard error of 0.108 measures the uncertainty between the true employment to population ratio and the part time employment among females.F-statistics is 118.0 and Prob(F-statistics) is 1.96e-12. This low value indicates that there is a small chance that all of the regression parameters are zero and that the regression equation does have independent variable that is not purely random with respect to the dependent variable. The regression model explains 78.6% of the variation in