

**RELATIONSHIP BETWEEN THE HUMAN FERTILITY RATE IN GERMANY AND THE LEVELS OF INCOME OF THE POPULATION**

**Agon Baftijari <sup>1</sup> and Igballe Ferati Preshova <sup>2</sup>**

<sup>1</sup>Department of Business Administration, Faculty of Business Administration,

<sup>2</sup>Department of Finance OSCE Macedonia

<http://doi.org/10.35409/IJBMER.2021.3343>

**ABSTRACT**

This study will focus on the inverse relationship between human fertility rates and the country's Gross domestic product per capita. Many studies have tried connecting the levels of the country's development with the drop-in fertility rates. In this study we will choose Germany as a sample due to its current aging population issues that it has been facing recently and rated as one of the top European countries that is faced with the problem.

The model used will be the simple regression function and we will try to prove the assumption of the classical linear model and spot if there are any violations of the model.

**Keyword:** Human fertility, Gross domestic product, Development, European countries, violations.

**1. INTRODUCTION**

Several researchers suggest that the development of a country has brought series of advancements and improvement in living conditions for its population. However, on the other hand it is believed that as the development increases within a country, its human fertility and population growth rates drop. Germany is a country that has historically been associated with the fertility rates below the globally accepted fertility rate, or the replacement rate of 2.1. It is also believed that the rise of immigration in Germany and the change of its policies towards accepting as many immigrants as possible lately, has to do with its policies and attempts of boosting human population and fertility within the country.

The opinions on whether migration works in a long term are divided. While in the short-term cycles it does boost the figures, and this will be evident in the changing figures of fertility rates, however the long-term benefits of it are still unclear.

The model will be using data employed on testing the GDP per capita and the human fertility rates in Germany. The model should explain the extent that the levels of income in an economy have to do with the decision of having more children.

**2.LITARATURE REVIEW**

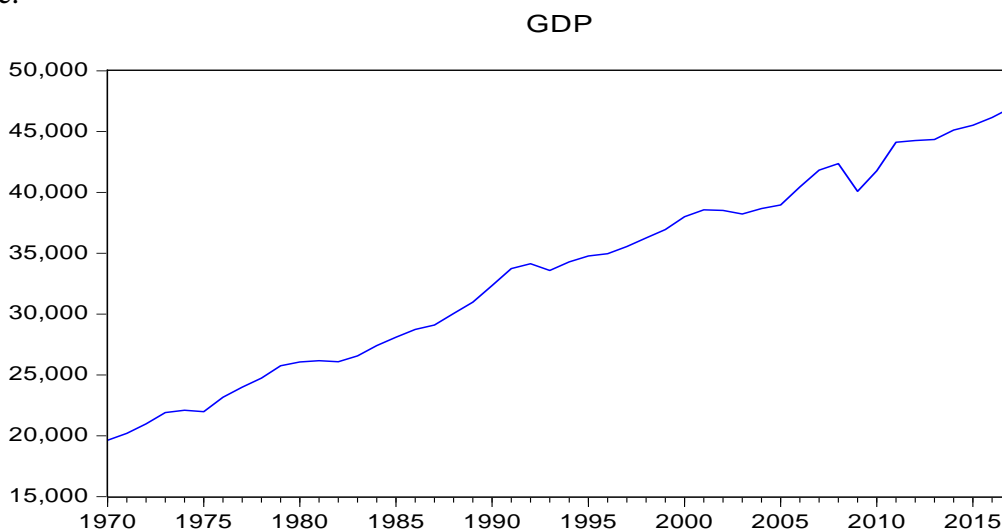
Sutrisno and Handel (2011) outline that the urbanization, housing policies and the improvement of living standards are the main triggers which cause a drop-in fertility rates in Germany. Moreover, Gomez and De Cos (2008) argue that even the historical rates of the economic

development affect the current fertility rates in several countries, meaning that short-term shocks or drops in economic development will not contribute to increasing fertility rates. The authors argue that Germany and Japan in contrast even in the post-war periods when the economic conditions were poor in the countries, the fertility rates were still dropping.

Milewski (2010) believes that migration is not expected to bring in many results due to evidence from research saying that immigrants tend to adjust over time to the host countries fertility patterns. This again goes in the direction of connecting the fertility rates to the levels of development in the country and maybe other indicators that this model will not test.

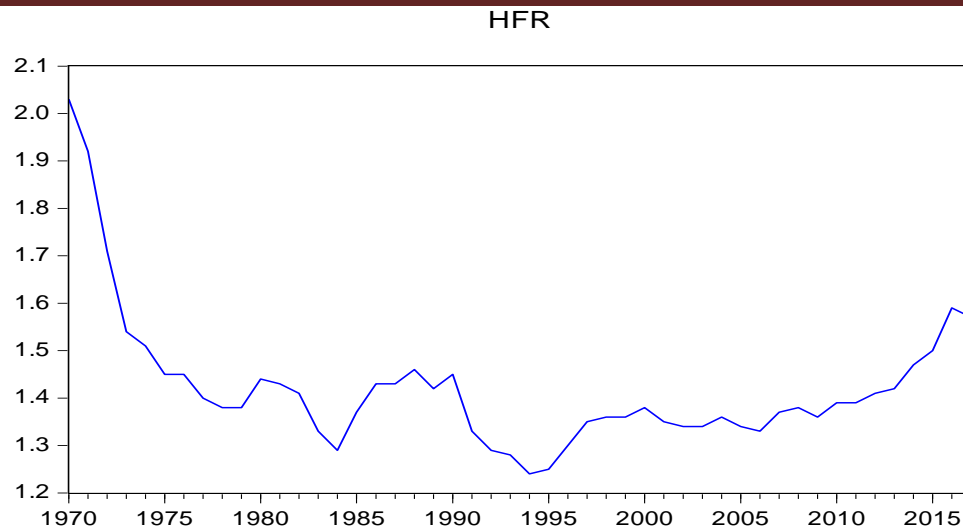
### 3.MODEL SPECIFICATION

In this study we will conduct a regression analysis using the OLS model for two variables: The economic development of Germany measured by its GDP per capita (expressed in US dollars) and its fertility rates. The frequency of the observations is on annual level starting from 1970 all the way to 2017. The model has a total of 48 observations due to the data available on Germany's GDP that is deriving from 1970s onwards and no previous data could be found in order to use them for comparison purposes. The data is extracted from the World Bank Data portals. The study will focus on proving is there is a relationship that can be considered as a statistically significant one between variables by using the regression analysis with the Ordinary Least Squares (OLS) technique.



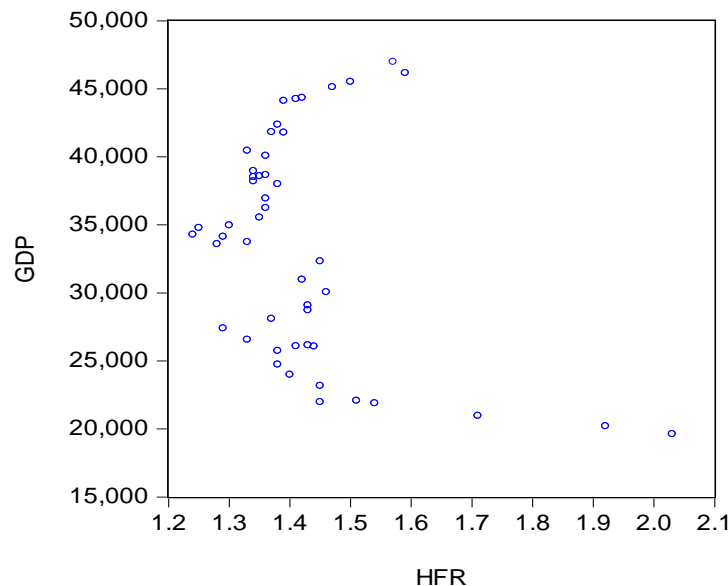
**Figure 1:** GDP per capita in Germany, 1970-2017

Source: The World Bank, <https://data.worldbank.org>



**Figure 2:** Fertility rates in Germany, 1970-2017  
Source: The World Bank, <https://data.worldbank.org>

While viewing at the both linear graphs one can see the inverse tendencies between the two variables that will be at the center of this study. This was also tested in a scattered diagram in order to test for the approximate inverse linear relationship between the variables. This diagram helps in identifying pattern in their natural settings (Touchette *et al*, 1985). The diagram in Figure 3 shows a moderate inverse correlation that will need to be further tested in order to be proven.



**Figure 3:** Scattered plot of GDP epr capita and fertility rates in Germany, 1970-2017  
Source: The World Bank, <https://data.worldbank.org>

**4.ECONOMETRIC MODEL**

In order to test the relationship between the given variables we need to employ regression models. As we are dealing with two variables, we will be working on a simple linear regression line. Regression models will help us explain the relationship between the dependent and the independent variables. The most used technique to fit the data into a line is the ordinary least square method (OLS) (Brooks, 2019).

The equation of our model under the simple regression formula would be:

$$HFR = \beta_0 + \beta_1 \times GDP_{pct} + u$$

, where:

HFR is the human fertility rates, or the dependent variable

GDPpc is the dependent variable

$\beta_0$ - is the constant or intercept

$\beta_1$ - is the coefficient

U is the error term.

In order to find the constants and the regression line slope we run the equation in EViews and get the following result.

**Table 1: OLS Table**

Dependent Variable: HFR

Method: Least Squares

Date: 10/15/19 Time: 22:45

Sample: 1970 2017

Included observations: 48

Variable	Coefficien		t-Statistic	Prob.
	t	Std. Error		
GDP	-6.44E-06	2.50E-06	-2.576577	0.0133
C	1.637780	0.085899	19.06645	0.0000
R-squared	0.126119	Mean dependent var	1.422500	
Adjusted R-squared	0.107122	S.D. dependent var	0.146178	
S.E. of regression	0.138127	Akaike info criterion	1.080511	
Sum squared resid	0.877639	Schwarz criterion	1.002544	
Log likelihood	27.93226	Hannan-Quinn criter.	1.051047	
F-statistic	6.638749	Durbin-Watson stat	0.178416	
Prob(F-statistic)	0.013253			

---

After running the function in Eview we are getting the following results and building on our sample regression function. The regression line as expected to be a negative slope due to the inverse relationship between the variables.

$$\text{HFRt} = 1.63778040765 - 6.44096637055 \times \text{GDPpc}$$

As we will be using OLS, linear models are required and need to be expressed in a straight line. More specifically, the model must be linear in the parameters ( $\alpha$  and  $\beta$ ), but it does not necessarily have to be linear in the variables ( $y$  and  $x$ ) (Brooks, 2019).

Interpretation of the parameters

$\beta_0 = 1.63778040765$ . This means that if  $\text{GDP} = 0$  then  $\text{HFR}^{\wedge}$  will be 1.63778040765 billion dollars. In the course of explaining the variables this does not make much sense.

$$\beta_1 = d\text{HFR}^{\wedge}/d\text{GDP} = -6.44096637055$$

If GDP increases by 1 measured unit then  $\text{HFR}^{\wedge}$  will decrease by 6.44096637055 trillion dollars

### **Hypothesis testing:**

Testing the hypothesis is of utmost significance in statistics. Salvatore and Reagle (2002) explain that the first steps would be testing the null hypothesis and move to alternative hypothesis and decide on significance levels.

The first variable to be tested is the independent variable, the GDP per capita which represents the slope coefficient. We are therefore testing the null and alternative hypothesis in this case.

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

In statistics the hypothesis test: can be conducted through the test of significance approach or through the confidence interval approach.

### **T-Test**

A t-test is used to determine if there is a significant difference between the means of two variables. In our model, the t-statistic for the GDPpc is -2.5765 as generated in Eview (see Table 1). The confidence interval will be taken as 95%, which represents standard range of confidence intervals across many samples.

The T-critical is measured by the formula:  $t_{c} = t_{\alpha/2; T-m}$ , where  $\alpha$ -significance level, set at 0.05

T- number of observations, in our case 48

m-number of parameters.

We then get  $t_c = 0.025; 46$ , where according to the t-tables we get a value of  $t_c = 2$

$$t = \frac{\beta - b_0}{se(\beta)} = -2.5765$$

As  $T > T_c$ , the null hypothesis is rejected, implying the coefficient has statistical significance.

### **Testing the constant:**

As the t-statistic of the constant is 19.06645 and  $19.06645 > 2$ , we are free to reject the null hypothesis for the constant as well.

### **Confidence intervals**

Moreover, we construct the confidence interval of the coefficient to be between the range 163,828 and 164,172, which explains that 0 is not within this range and therefore we are free to reject the null hypothesis and conclude that there is a statistically significant relationship between the variables. Even if we look at it from the p-value side we see that the probability is 0 for us rejecting wrongly the null hypothesis, which indicates again the statistical relationship between the variables used in the model.

**R-Square**

In order to further test or explain how much the model we are using is fitted into explaining the variables, we may use the R squared. The R<sup>2</sup> will explain the proportion of variation in Y or the dependent variable that can be explained by the systematic proportion of the model.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

The R<sup>2</sup> figures are expected to be between 0 and 1 in order to prove for the model's significance level, or  $0 < R^2 < 1$ . In our case demonstrates the R-Square of 0.126119 from the table 1 has been drawn out. This means that the dependent variable is explaining at 12.6% the independent variable.

**F test**

The F-test is used to explain if independent variables in a linear regression model are significant.

$$PRF: HRFi = \alpha_0 - \alpha_1 * GDP_{pc} + U_i$$

$$SRF = HRF_t^{\wedge} = \alpha^{\wedge} + \beta^{\wedge} * GDP_{ct} \Rightarrow 1.63778040765 - 6.44096637 \times GDP_{pc}$$

$$H_0: \alpha_1 = 0$$

$$H_1: \alpha_1 \neq 0$$

The F taken generated from table 1 is 6.638749.

The F-critical can be measured as follows:

**Assumptions of the Classical Linear Regression Model**

We proceed in testing the five assumptions of the Classical Linear Regression Model (CLRM)

1. Linearity: The model specifies a linear relationship between y and x;
2. Exogeneity: The expected value of the error term is expected to be 0, or  $E(u_t) = 0$
3. Homoscedasticity:  $Var(u_t) = \sigma^2 < \infty$ ; variance for the error term is the same for all observations
4. Zero covariance  $Cov(u_i, u_j) = 0$ ; Covariance instead of following how residuals dance around if residuals are a function of x,  $U = f(x)$
5. The model has no autocorrelation between error terms; which means that the X matrix is non-stochastic or fixed in repeated samples
6. Normal distribution between the variables  $\epsilon|X \sim N(0, \sigma^2)$  (Zeng, 2016; Greene, 2003)

**Autocorrelation**

The extent of correlation between the values of the same variables across different observations in the data is called as autocorrelation. In statistics it shows the correlation between the variable and residual variables. Two are the most famous tests which test for autocorrelation in models; the Breusch Godfrey test and Durbin Watson.

**Durbin Watson (DW)**

Durbin-Watson (DW) as a test is more widely used for preliminary autocorrelation, focusing on a relationship between an error and its immediately previous value (Brooks, 2019). The values tested for the DW can be drawn by the Durbin Watson table, using the following parameters; n=48 observations, k'- number of explanatory variables without the constant, meaning 1. From the table we can see that the dL=1.48 and dU=1.57.

DUpper= 1.48

DLower= 1.57

From Table 1 the DW figure of our model as provided by Eview is 0.178416. As  $DW < dL \rightarrow 0.178416 < 1.48$ , the null hypothesis can be rejected.

PRF:  $ut = \rho \times ut-1 + \omega t$

SRF:  $\hat{u}t = \hat{\rho} \times ut-1$

$$DW = \frac{\sum_{i=2}^{T=778} (ui-ui-1)^2}{ESS}$$

**Breusch-Godfrey test**

**Table 2- Breusch-Godfrey Serial Correlation LM Test:**

F-statistic	30.55167	Prob. F(2,44)	0.0000
Obs*R-squared	27.90549	Prob. Chi-Square(2)	0.0000

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Date: 10/21/19 Time: 22:17

Sample: 1970 2017

Included observations: 48

Presample missing value lagged residuals set to zero.

Variable	Coefficien		t-Statistic	Prob.
	t	Std. Error		
GDP	5.04E-07	1.67E-06	0.301404	0.7645
C	-0.014370	0.057258	-0.250970	0.8030

RESID(-1)	0.918481	0.148149	6.199712	0.0000
RESID(-2)	-0.200479	0.155627	-1.288200	0.2044
R-squared	0.581364	Mean dependent var	2.78E-17	
Adjusted R-squared	0.552821	S.D. dependent var	0.136650	
S.E. of regression	0.091380	Akaike info criterion	1.867932	
Sum squared resid	0.367411	Schwarz criterion	1.711999	
Log likelihood	48.83037	Hannan-Quinn criter.	1.809005	
F-statistic	20.36778	Durbin-Watson stat	1.260610	
Prob(F-statistic)	0.000000			

PRF:  $U_i = \alpha + \beta GDP_{pc} + \rho_1 U_{i-1} + \rho_2 U_{i-1} + \epsilon_i$

SRF:  $U_i^{\wedge} = 0.037079 - 0.65 GDP_{pc} + 0.218003 U_{i-1} + 0.17338 U_{i-1}$

Hypothesis testing

Ho:  $\rho_1 = \rho_2 = 0$  (No autocorrelation)

H1:  $\rho_1 \neq 0$  and/or  $\rho_2 \neq 0$  (There is autocorrelation)

$LM = Taux \times Raux; Taux = T - r = 48 - 2 = 46$

$LM = 46 * 0.581364 = 26.742744$

Then using the chi-square if is lower than  $\alpha$  then we reject the hypothesis. The Chi-square critical value at 0.05 significance from the Chi-distribution tables for two parameters is 5.99.

If  $LM > Chi\text{-Square critical}$ ,  $H_0$  is rejected, meaning  $26.742744 > 5.99$ , therefore we can assume that the series are not autocorrelated.

**Heteroscedicity test**

One of the assumptions of the CLRM model is homoscedicity, meaning if the variance of errors is same across all distributions. The notion of heteroscedicity describes the case where the variance of errors or the model is not the same for all observations. In order to prove the model right we need to employ the tests and establish if the errors are identically distributed across the model.

Two are tests for heteroscedicity; the Breusch-Pagan test and the White test and modified White test. In the tables below we will see the tests generated in Eview

*Breusch-Pagan Godfrey*

**Table 3- Heterosceditiy**

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	3.904580	Prob. F(1,46)	0.0542
-------------	----------	---------------	--------



---

Obs*R-squared	3.755564	Prob. Chi-Square(1)	0.0526
Scaled explained SS	10.63599	Prob. Chi-Square(1)	0.0011

---

Test Equation:  
 Dependent Variable: RESID^2  
 Method: Least Squares  
 Date: 10/21/19 Time: 22:21  
 Sample: 1970 2017  
 Included observations: 48

---

Variable	Coefficien t	Std. Error	t-Statistic	Prob.
C	0.071513	0.027694	2.582267	0.0131
GDP	-1.59E-06	8.06E-07	-1.976001	0.0542

---

R-squared	0.078241	Mean dependent var	0.018284
Adjusted R-squared	0.058203	S.D. dependent var	0.045888
S.E. of regression	0.044532	Akaike info criterion	3.344433
Sum squared resid	0.091224	Schwarz criterion	3.266466
Log likelihood	82.26638	Hannan-Quinn criter.	3.314969
F-statistic	3.904580	Durbin-Watson stat	0.334137
Prob(F-statistic)	0.054170		

---

*White test*

Table 4- Heteroskedasticity Test: White

---

F-statistic	11.05610	Prob. F(2,45)	0.0001
Obs*R-squared	15.81509	Prob. Chi-Square(2)	0.0004
Scaled explained SS	44.78931	Prob. Chi-Square(2)	0.0000

---

Test Equation:  
 Dependent Variable: RESID^2  
 Method: Least Squares  
 Date: 10/21/19 Time: 22:36  
 Sample: 1970 2017

Included observations: 48

Variable	Coefficien		t-Statistic	Prob.
	t	Std. Error		
C	0.486317	0.103802	4.685035	0.0000
GDP^2	3.99E-10	9.72E-11	4.106249	0.0002
GDP	-2.81E-05	6.49E-06	-4.327894	0.0001
R-squared	0.329481	Mean dependent var	0.018284	
Adjusted R-squared	0.299680	S.D. dependent var	0.045888	
S.E. of regression	0.038401	Akaike info criterion	3.620998	
Sum squared resid	0.066359	Schwarz criterion	3.504048	
Log likelihood	89.90395	Hannan-Quinn criter.	3.576802	
F-statistic	11.05610	Durbin-Watson stat	0.410762	
Prob(F-statistic)	0.000124			

The regression functions are:

$$PRF: ut_2 = \gamma_{2;0} + \gamma_{2;1} \times UNt + \gamma_{2;2} \times UNt^2 + \varepsilon_{2;t}$$

$$SRF: \hat{ut}_2 = 0.486317 - 2,81 \times UNt + 3.99 \times UNt^2$$

$LM=46 \times 0.329481 = 15,156126 > 5.991$  which indicates that the variance in error terms is confirming the assumptions of the model used.

#### 4.CONCLUSION

The OLS model that we employ is testing the linear relationship between in one of the biggest problems that Germany is facing nowadays, the human fertility. It is seen that we cannot expect that this model with one dependent variable will explain the entire issue of human fertility in the country due to the fact that there are a number of other factors that are affecting it.

Our model shows moderate significance in explaining the dependent variable by the independent one that we have chosen. The issue requires further research in employing more variables in order to explain the model. For the purposes of our coursework, we have tested the assumptions of the classical linear regression model and can therefore, conclude that no violations of the assumptions have been spotted.

**REFERENCES**

- [1] Agunbiade, A. and Adeboye, O., 2012. Estimation under Heteroscedasticity: A Comparative Approach Using Cross-Sectional Data. *Mathematical Theory and Modeling*, 2(11), pp.1-8.
- [2] Ayinde, K., 2007. A comparative study of the performances of the OLS and some GLS estimators when stochastic regressors are both collinear and correlated with error terms. *Journal of Mathematics and Statistics*, 3(4), pp.196-200.
- [3] Brooks, C., 2019. Introductory econometrics for finance. *Cambridge university press*.
- [4] Greene, W.H., 2003. Econometric analysis. Pearson Education India.
- [5] Gómez, R. and De Cos, P.H., 2008. The importance of being mature: the effect of demographic maturation on global per capita GDP. *Journal of population economics*, 21(3), pp.589-608.
- [6] Milewski, N., 2007. First child of immigrant workers and their descendants in West Germany: Interrelation of events, disruption, or adaptation?. *Demographic Research*, 17, pp.859-896.
- [7] Salvatore, D. and Reagle, D., 2002. Statistics and Econometrics, *Schaum's Outline Series*.
- [8] Sutrisno, A. and Handel, O., 2011. Dynamic Aging Population in Germany: A case study about demographic change. *University of Bergen*.
- [9] Touchette, P.E., MacDonald, R.F. and Langer, S.N., 1985. A scatter plot for identifying stimulus control of problem behavior. *Journal of applied behavior analysis*, 18(4), pp.343-351.
- [10] Zeng, Y., 2016. Classical Linear Regression Model: Assumptions and Diagnostic Tests. study notes, 3.
- [11] The World Bank <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=DE>