• Student ID number: 33636339

• Program of study: BSc Economics with Econometrics

• Module: Econometrics

• Date: 14/12/2021

• Word count: 2196

EXECUTIVE SUMMARY

This paper explores the argument of reducing the gender wage gap to boost economic growth. The hypothesis is that reducing the gender wage gap would lead to increase the rate of GDP growth. Standard econometrics techniques are employed to determine the effects of a reduction in gender wage gap on GDP and obtain quantitative data to quantify the effects of the reduction in wage gap through an increase in average hourly wage rate of female workers. First, standard OLS estimates are obtained through the classic linear multiple regression model. After conducting test for autocorrelation, unit root test, and cointegration, a regression model in first differences and an autoregressive distributed lag model are produced to test the hypothesis. Each model confirms the hypothesis that reducing the gender wage gap would lead to an increase in GDP.

INTRODUCTION

The main economic estimate employed by policy makers to evaluate whether a country is growing is gross domestic product, or better the rate of GDP growth. ''A large body of macroeconomics research supports the notion that the distribution of income, assets, and capabilities […] has implications for the rate of economic growth and development'' (Seguino, 2012, p.59), and it is the purpose of this essay to provide empirical answers to theoretical questions on the effects of the distribution of income on GDP growth. In particular, this essay attempts to provide quantitative data on the potential effects of reducing gender wage gap between males and females in the UK, and it will be done so by employing the econometrics techniques of time series multiple regression analysis. feminist economics focuses mainly on conducting economic analysis on the key concepts of gender, the household and unpaid work and caring, providing economic empirical data in support of closing the gender gap and reducing gender inequality (Staveren, 2010, p.1123), so economic models and theories employed in this paper will be drawn from the feminist economics literature. The difference in wages between males and females is the consequence of the cultural and socio-political context embedded in the countries, and the gender inequality have implications on the entire economy. For this reason, it useful to support economic theories that aim to reduce the gender wage gap with valid econometric results to aid policy makers in the formulation of policies devoted to reducing gender inequality without sacrificing economic growth.

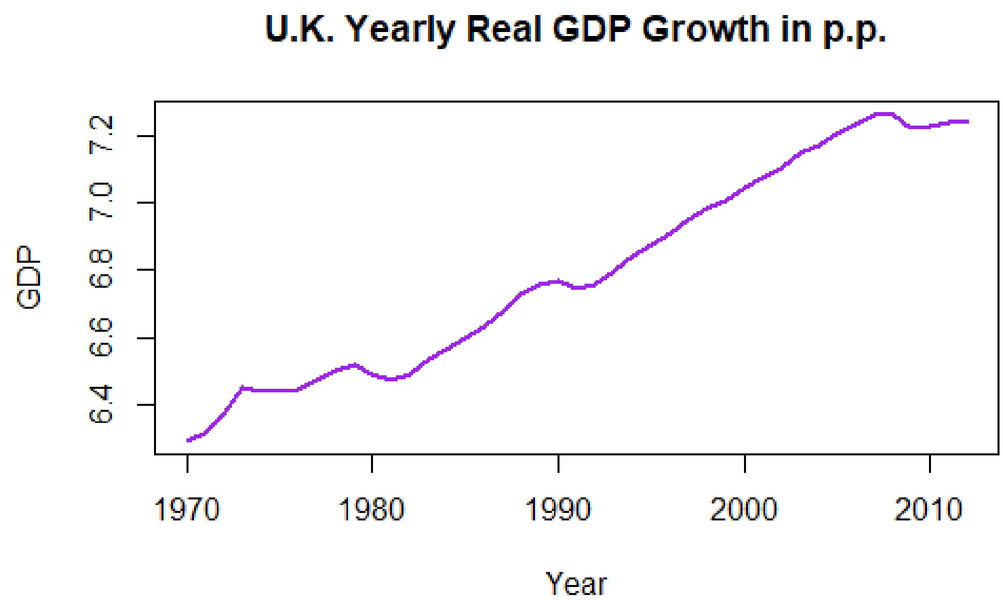

Figure 1

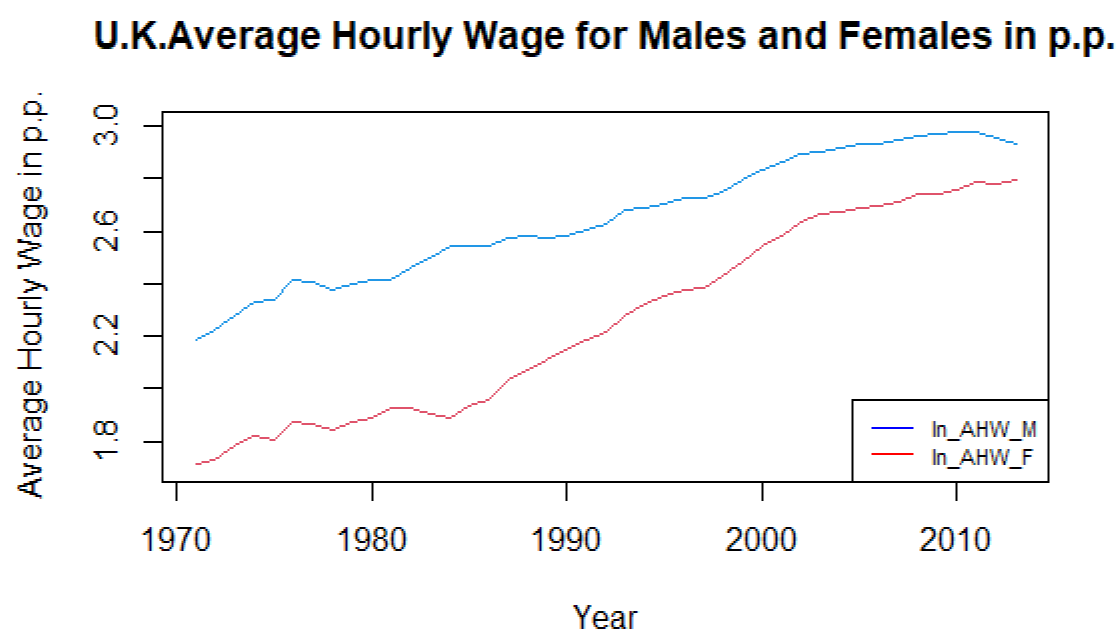

Figure 2

By observing figure 1 and 2, we can inference that since 1970, GDP grew steadily along with a growth in the average hourly wage of both males and females. Figure 2 suggest that from the 2000s the rate of growth of females was higher than males one, which started to decline after 2010, suggesting a reduction in the gender wage gap. This is better depicted in figure 3 below, that shows that gender wage gap growth rate experienced a negative trend between 1985 and 2012, with a significant slump in 2008.

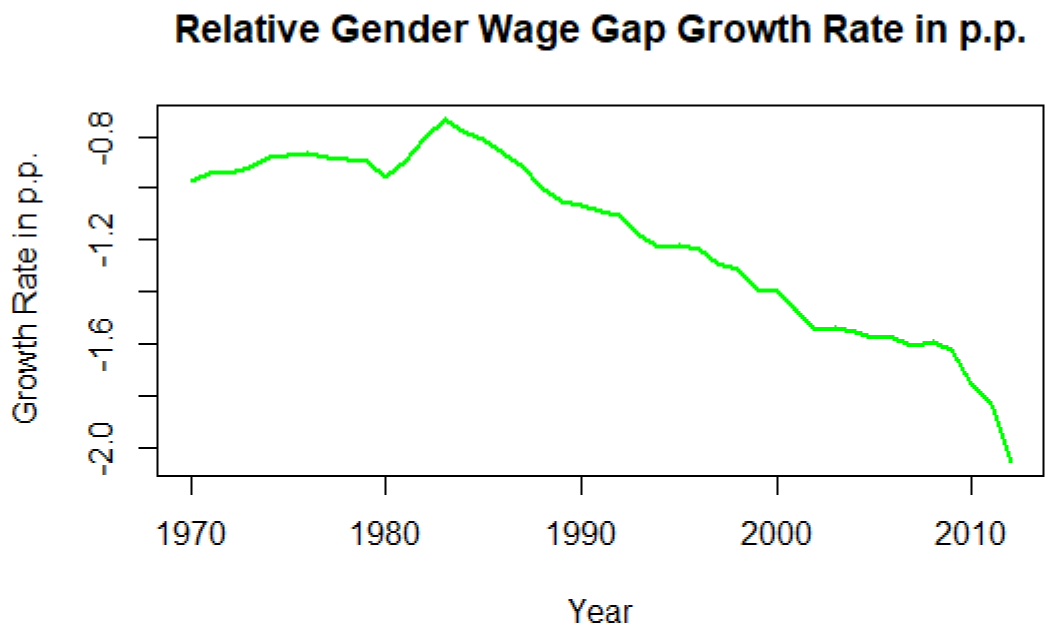

Figure 3

The trends depicted by figure 1,2, and 3 suggest that there might be a relation between a reduction in gender wage gap and increase in GDP, and it is the aim of this paper to explore the topic by obtaining quantitative data to provide more solid arguments in support of the thesis.

METHODOLOGY

Time series multiple regression analysis is the technique utilised to analyse the data provided, and the coefficients of the regression are estimated by ordinary least squares. A regression in first differences and an autoregressive distributed lag model is included to provide better estimates of the coefficient of each independent variables of the econometric model. The starting point of this paper is the following identity:

$$GDP = \omega + \pi$$

In which ω stands for total wage share and π for profit share. The identity is then adapted as follow:

$$GDP = \omega_M + \omega_F + \pi$$

ω_M depicts the wage share of males while ω_F depicts the wage share of females. The hypothesis to be tested is whether if $\Delta\omega_F > 0$ then $\Delta GDP > 0$, and it is tested through the analysis of the variable gender wage gap. Gender wage gap is calculated as the relative male-female wage ratio, namely:

$$G\_W\_Gap = \frac{AHW\_M - AHW\_F}{AHW\_M}$$

in which AHW_M represents average hourly wage of males (calculated by dividing WBR_M by E_M) and AHW_F represents average hourly wage of females (calculated by dividing WBR_M by E_M). Plugging the variables provided by the source into the output identity, we get:

$$GDP = (AHW\_M * E\_M) + (AHW\_F * E\_F) + PPR$$

To include G_W_Gap into the equation and calculate the effects of a reduction in gender wage gap, rules of algebra are applied, so by solving for AHW_F, we obtain:

$$AHW\_F = AHW\_M(1 - G\_W\_Gap)$$

Finally, the following model is constructed:

$$lnGDP = lnAHW_M - \ln G\_W\_Gap + lnE_F + ln\mathrm{WBR\_M} + lnPPR$$

The multiple regression analysis of the model allows to produce the estimated coefficients of the variables through OLS, aiding in the hypothesis testing of whether if $\Delta\omega_F > 0$ then $\Delta GDP > 0$. As previously mentioned, this could be expected to happen through a reduction of the gender wage gap caused by either an increase of the average hourly wage share of females or by a higher participation in the production of output through higher employment given by an increase in the average hours of employment by female workers. Mathematically:

$$\Delta GDP > 0 \; if \; \Delta G\_W\_Gap < 0 \; and \; \Delta\omega_F > 0$$

Notably, $\Delta G\_W\_Gap < 0$ could be achieved also if $\Delta\omega_F > \Delta\omega_M$. The econometrics model employed in this analysis is then derived from the mathematical model above, and is stated as such:

$$lnGDP = \beta_0 + \beta_1 lnAHW\_M + \beta_2 \ln G\_W\_Gap + \beta_3 lnE\_F + \beta_4 ln\mathrm{WBR\_M} + \beta_5 lnPPR$$

Logs are taken to linearise exponential trends of the variables. Standard statistical tests to assess the presence of unit root in the regression model are conducted, with tests for autocorrelation, heteroskedasticity, and cointegration, assessing whether the OLS estimates are BLUE, and if not conduct required adjustments to obtain a valid econometric model to be employed for forecasting. Required adjustments are made applying an ARDL model that allows to obtain coefficients coherent with the assumption of efficiency. The coefficients obtained running the multiple regression analysis above are shown in the regression table found in the appendix (Figure 4).

KEY FINDINGS

Autocorrelation test

''OLS to estimate a regression model leads us to BLUE estimates of the parameters only when all the assumptions of the CLRM are satisfied'' (Asterious, 2015, p.157). If the assumption of *Cov(ut, us) = 0 for all t = s* is no longer true, then ''an error occurring at period t may be correlated with one at period s'' (Asterious, 2015, p.157). To test for autocorrelation, the Durbin-Watson test is employed, but the test is only valid when the regression model includes a constant; serial correlation is assumed to be of first-order only; and the equation does not include a lagged dependent variable as an explanatory variable. Under the null hypothesis H0: ρ = 0, then to test for positive serial correlation, the hypotheses are: H0: ρ = 0 no autocorrelation and Ha: ρ > 0 positive autocorrelation. After conducting the test on the model, it is obtained a DW Statistic of 0.5889792, suggesting a positive serial correlation. Moreover, the R2 of the regression are higher than the DW Statistic (1.000>0.5889792), suggesting that the results are spurious regression.

Unit root test and cointegration

Economists working with time-series might face two issues: one time series variable can influence another with a time lag and if the variables are nonstationary, a problem known as spurious regression may arise, which is the case in the regression model of this paper. If the time-series are non-stationary, then those variables should not be included in the regression, but there is a case in which the regression can be ran, that is when the variables are cointegrated (Koop, 2013, p.131-132). In time series a stochastic process is said to be stationary if the following conditions are satisfied:

1. Mean: $E\ Yt$ = μ for all t

2. Variance: $E\ Yt$ − μ 2 = σ 2 for all t

3. Covariance: γk = E[Yt − μ Yt+k − μ] for all t and k ≠ 0

If the opposite does not hold true, then the variable is said to be non-stationary. To test for unit-root, it is employed the Augmented Dickey Fuller Test with time-trend and Akaike Information Criteria to select lags for each variable independently, as this allows to calculate the correct statistical tables from which to take critical values (Koop, 2013, p.164). Results of the tests of each variable can be found in the appendix (*3.2.1*), and overall, we fail to reject the null hypothesis of non-stationarity in every variable at 1% significance level, while for the variable E_F the null hypothesis of non-stationarity can be rejected at the 5pct. The results are then confirmed by computing the Phillips-Perron test, a more powerful test that allows to avoid the limitations of committing type II errors typical of the DF test. If Y and X contain unit roots, then OLS estimation of this regression can yield results which are completely wrong, but if Y and X are cointegrated we do not have to worry about the spurious regression problem (Koop, 2013, p.175). Now, the following step is to test for cointegration, as this not only allows to not account for spurious regression, but also provides important economic intuition. In case of cointegration, the trends of the error terms of the two variables cancel each other, leading to a common trend. Put in another words, the spread between the trends the error term of the two variables over time tends to be constant in average. The tests employed to search for cointegration pattern are the Engle-Granger test, which allows to regress Y on X, and the more sophisticated Johansen test, that allows to evaluate the number of cointegration relationships among the variables of the entire model. By computing the Johansen test, it shows that there are at least four cointegration relationships. Next, by conducting the Engle-Granger test of ln_GDP on ln_G_W_Gap, we fail to reject the null hypothesis of no cointegration, suggesting that GDP and gender wage gap are ‘’trending together or bearing an equilibrium relationship to each other’’ (Koop, 2013, p.181).

Regression in first differences and ADRL model

After running all the tests above, it is suggested to adjust the regression model to account for unit roots and to better calculate the effects of the changes in the independent variables on the dependent variable. The new regression model is then calculated in first differences, but it still might present issues of autocorrelation, and heteroskedasticity, so required tests on the residuals need to be conducted. The test on the autocorrelation of the residuals is the ADF employed above, while the Breusch-Pagan test is employed to test for heteroskedasticity. Then, by applying the Newey West test, a final regression model in which the coefficients are adjusted for the issues just mentioned is obtained. Finally, since one of the main targets of running regressions on time-series data is to construct models that can be used for forecasting and for estimating dynamic casual effects, it can be ran a regression which control for autocorrelation and that include autoregressive components into the model to account for lags of the variables. This is obtained by running the autoregressive distributed lag model, and the final coefficients are included in the regression table in the appendix (Figure 29). Furthermore, this model permits to stationarize the variables that presented non-stationary trends when the ADF test was computed.

Interpretation of the DLRM regression table

	Dependent variable:	
	ln_GDP	L(GDP_Diff)
	OLS	*dynamic linear*
	(1)	(2)
ln_AHW_M	0.316***	
	(0.026)	
ln_G_W_Gap	-0.082***	
	(0.007)	
ln_E_F	0.348***	
	(0.029)	
ln_WBR_M	0.373***	
	(0.023)	
ln_PPR.after.tax	0.305***	
	(0.009)	
L(AHW_M_Diff)		0.284***
		(0.052)
L(G_W_Gap_Diff)		-0.095***
		(0.013)
L(E_F_Diff)		0.269***
		(0.047)
L(WBR_M_Diff)		0.439***
		(0.044)
L(PPR_Diff)		0.301***
		(0.009)
Constant	0.833***	-0.001
	(0.087)	(0.001)
Observations	43	42
R^2	1.000	0.978
Adjusted R^2	1.000	0.975
Residual Std. Error	0.005 (df = 37)	0.004 (df = 36)
F Statistic	32,769.540*** (df = 5; 37)	321.712*** (df = 5; 36)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure 4

To recap, the hypothesis to be tested is whether if $\Delta\omega_F > 0$ then $\Delta GDP > 0$, and to do so it is suggested to observe how the gender wage gap affects the GDP. Mathematically:

$$\Delta GDP > 0 \; if \; \Delta G\_W\_Gap < 0 \; and \; \Delta\omega_F > 0$$

By analysing the coefficients of the DLRM obtained by computing the ADLM, it can be stated that by increasing the gender wage gap of 1% in time t, the GDP in time t+1 would decrease of -0.095%., or, on the opposite, by decreasing the gender wage gap of -1%, GDP would increase of 0.095%. This fails to reject the null hypothesis set at the beginning of the paper supporting the argument that the economy would benefit by closing the gender wage gap.

CONCLUSIONS

This essay aimed at providing empirical evidence to support the argument of closing the gender wage gap to increase GDP. The econometrics models constructed throughout the paper, after having tested for autocorrelation, heteroskedasticity, unit root and cointegration, produced valid estimate of the coefficients of the variables included in the model. Particularly, the coefficient of the variable of interest lnG_W_Gap obtained through the ADLM shows a negative sign, suggesting that an increase in the wage gap of 1% leads to a reduction in GDP of 0.095%, failing to reject the null hypothesis that $\Delta GDP > 0 \; if \; \Delta G\_W\_Gap < 0$. Notably, since $\Delta G\_W\_Gap < 0$ could be achieved also if $\Delta\omega_F > \Delta\omega_M$, it can be suggested that the reduction of wage gap should be achieved by increasing the average hourly wage of female, ceteris paribus, or by making sure that the rate of growth of the wage of females should be higher than the growth rate of the wage of males. Although the econometric results of the regression analysis are valid and provide empirical evidence, it might encounter some limitations due to its simplicity and requires further developments that include more complex models and that take into account a broader range of variables.

BIBLIOGRAPHY

- Asteriou, D., and Hall, S. G. (2015). *Applied Econometrics*. Macmillan Education UK.
- Koop, G. (2013). *Analysis of Economic Data*, Wiley Textbooks. *ProQuest Ebook Central*, http://ebookcentral.proquest.com/lib/goldsmiths/detail.action?docID=1935745.
- Stock, J. and Watson, M. (2019), *Introduction to Econometrics*.
- Seguino, S. (2012). "Macroeconomics, Human Development, and Distribution" *in Journal of Human Development and Capabilities*. 13:1, 59-81. DOI: 10.1080/19452829.2011.637376
- Van Staveren, I. (2010). "Post-Keynesianism meets feminist economics" *In Cambridge Journal of Economics*, *34*(6), 1123–1144. http://www.jstor.org/stable/24232059

APPENDIX

1.REGRESSION TABLE

	Dependent variable:	
	ln_GDP *OLS*	L(GDP_Diff) *dynamic linear*
	(1)	(2)
ln_AHW_M	0.316***	
	(0.026)	
ln_G_W_Gap	-0.082***	
	(0.007)	
ln_E_F	0.348***	
	(0.029)	
ln_WBR_M	0.373***	
	(0.023)	
ln_PPR.after.tax	0.305***	
	(0.009)	
L(AHW_M_Diff)		0.284***
		(0.052)
L(G_W_Gap_Diff)		-0.095***
		(0.013)
L(E_F_Diff)		0.269***
		(0.047)
L(WBR_M_Diff)		0.439***
		(0.044)
L(PPR_Diff)		0.301***
		(0.009)
Constant	0.833***	-0.001
	(0.087)	(0.001)
Observations	43	42
R^2	1.000	0.978
Adjusted R^2	1.000	0.975
Residual Std. Error	0.005 (df = 37)	0.004 (df = 36)
F Statistic	32,769.540*** (df = 5; 37)	321.712*** (df = 5; 36)
Note:		*p<0.1; **p<0.05; ***p<0.01

Figure 5

2. DATA VISUALITATION

2.1 GDP Plots

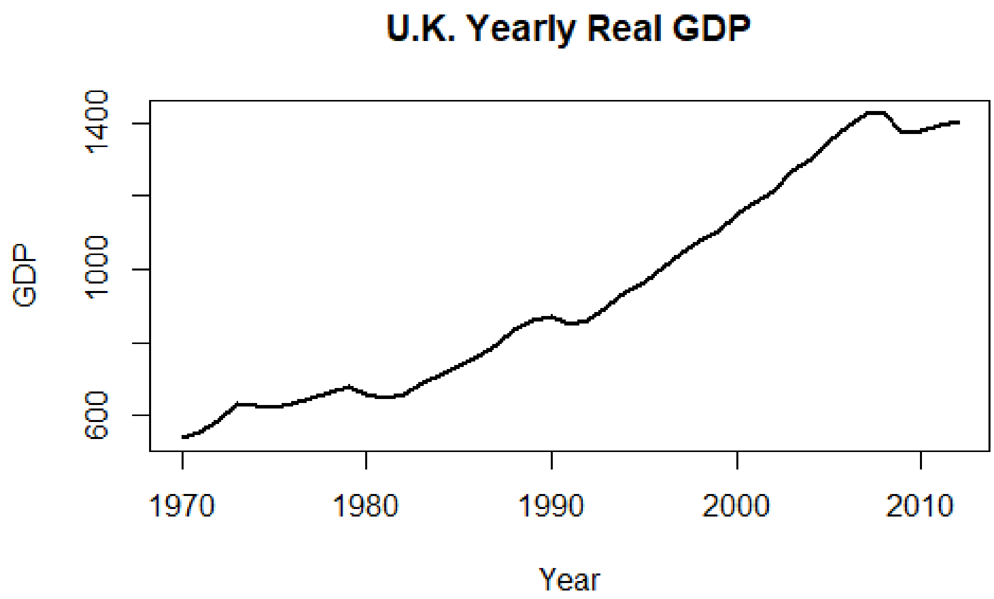

Figure 6

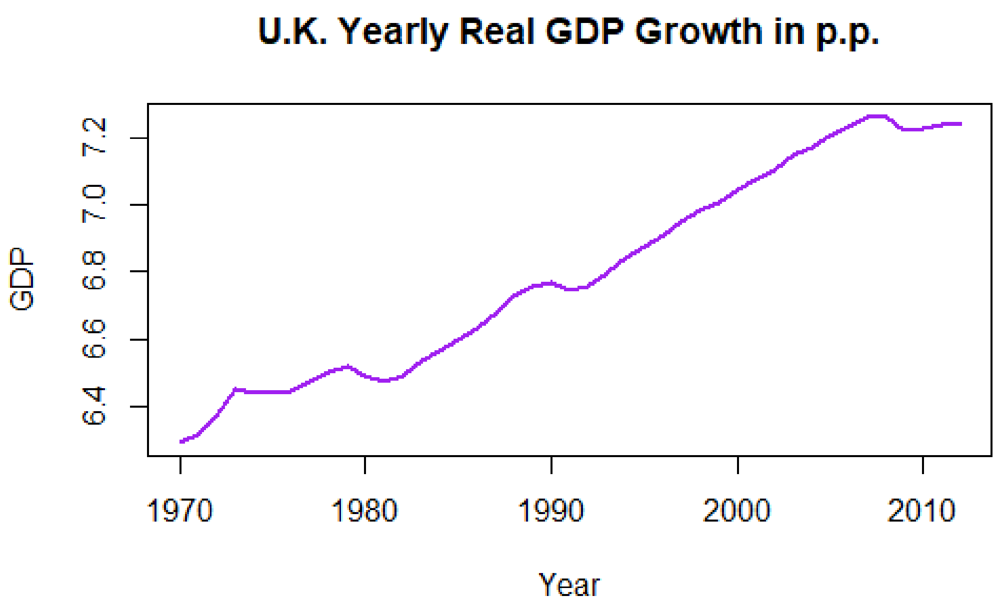

Figure 7

2.2 Wage plots

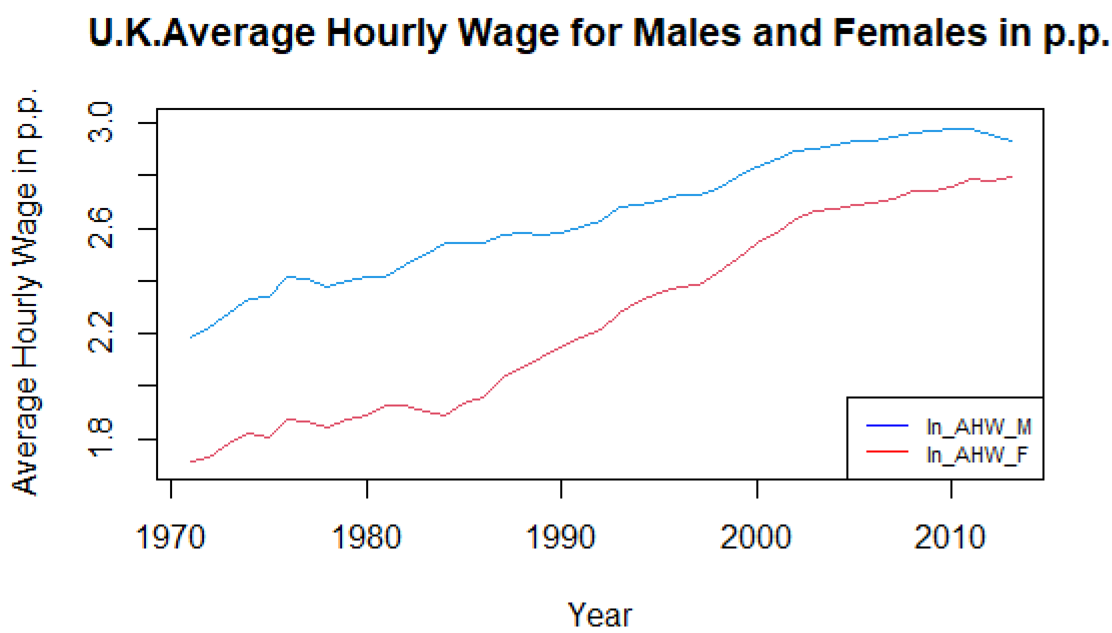

Figure 8

2.3 Gender Wage Gap

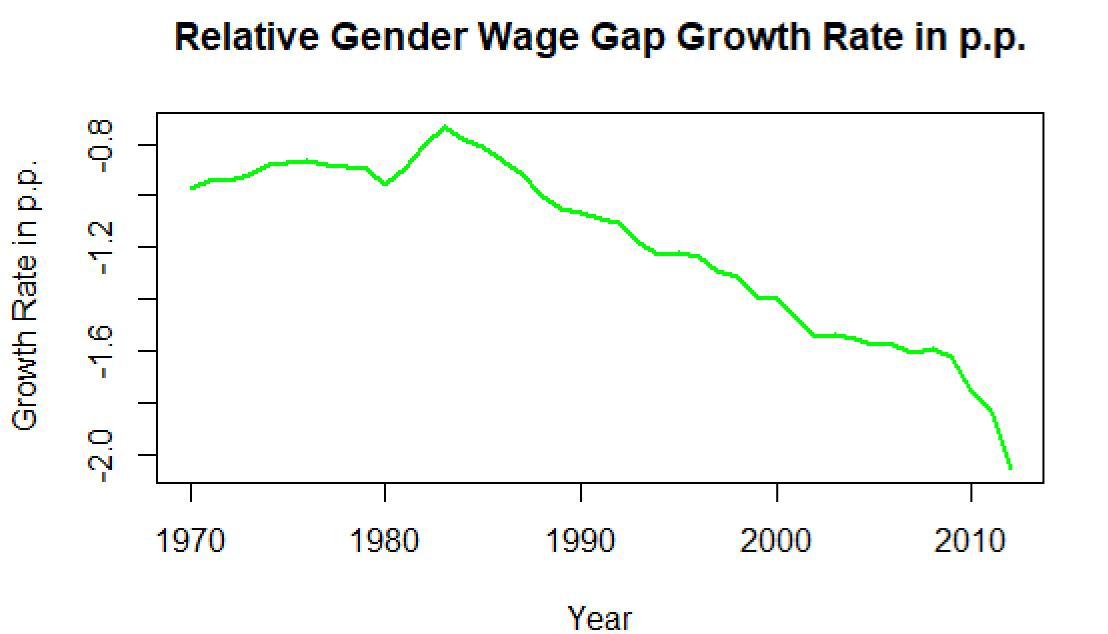

Figure 9

2.4 Hours worked by Gender

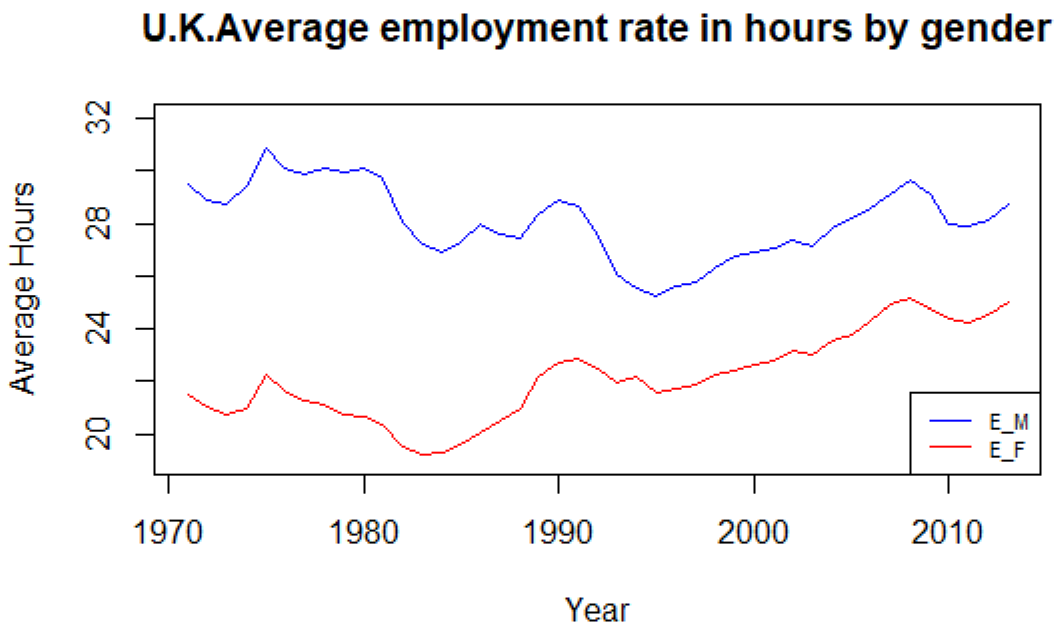

Figure 10

2.5 Plot of residuals Linear Model

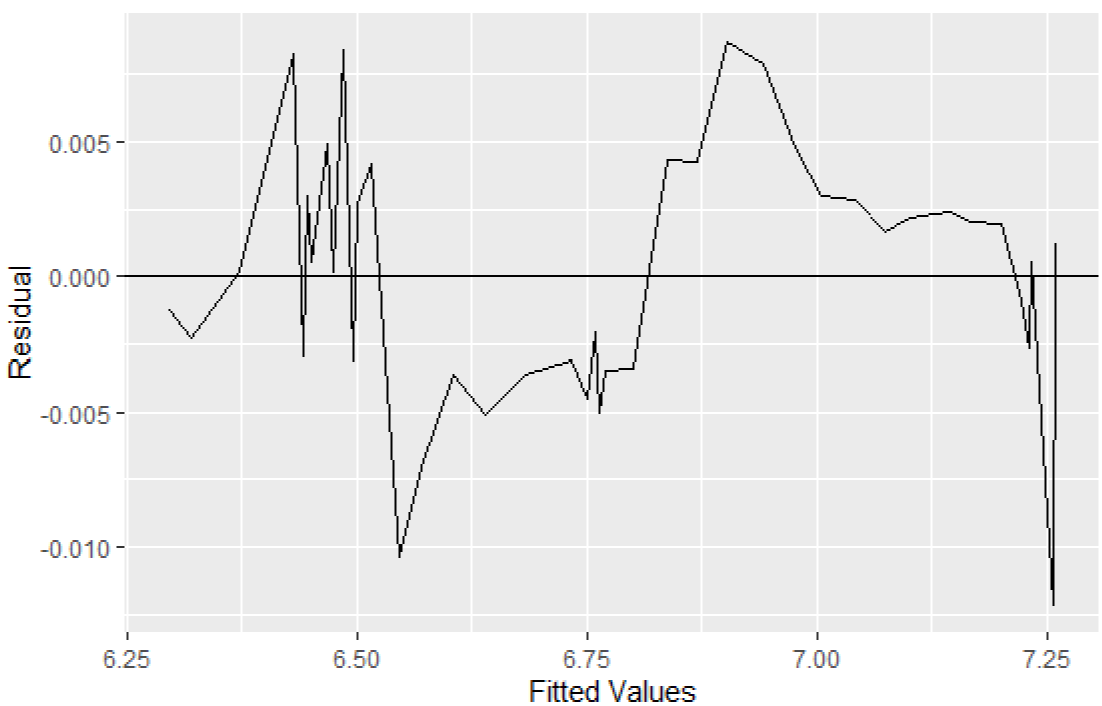

Figure 11

3. TESTS

3.1 Autocorrelation

3.1.1 Durbin Watson test and residuals plot

```
 lag Autocorrelation D-W Statistic p-value
   1       0.6248308     0.5889792       0
 Alternative hypothesis: rho != 0
```

Figure 12

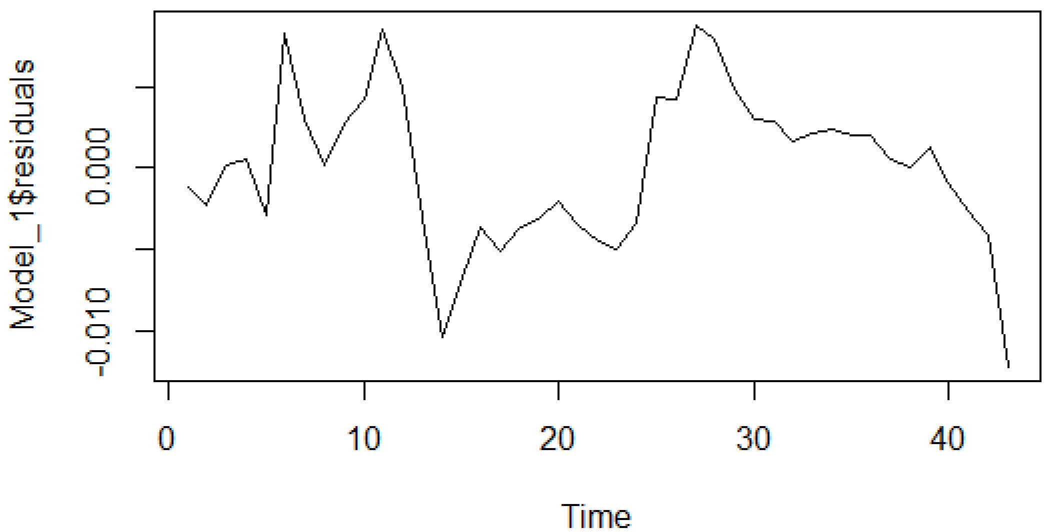

Figure 13

3.2 Stationarity Unit Root

3.2.1 Augmented Dickey Fuller with trend and AIC

#LogGDP

```
Value of test-statistic is: -1.6064 2.5418 2.0025

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
```

Figure 14

#ln_AHW_M

```
Value of test-statistic is: -2.013 2.3304 2.2199

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
```

Figure 15

#ln_W_Gap

```
Value of test-statistic is: -2.0902 4.6461 6.9203

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
```

Figure 16

#ln_E_F

```
Value of test-statistic is: -4.0724 6.297 8.3671

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
```

Figure 17

#ln_WBR_M

```
Value of test-statistic is: -2.9156 3.2712 4.4415

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
```

Figure 18

#ln_PPR.after.tax

```
Value of test-statistic is: 1.7062 2.8534 4.2654

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
```

Figure 19

<u>3.2.2 Phillips Perron test</u>

#LogGDP

```
        Phillips-Perron Unit Root Test

data:  LogGDP
Dickey-Fuller Z(t_alpha) = -1.7901, Truncation lag parameter = 3, p-value =
0.6566
alternative hypothesis: stationary
```

Figure 20

#ln_AHW_M

```
        Phillips-Perron Unit Root Test

data:  ln_AHW_M
Dickey-Fuller Z(t_alpha) = -1.6723, Truncation lag parameter = 3, p-value =
0.7032
alternative hypothesis: stationary
```

Figure 21

#ln_W_Gap

```
        Phillips-Perron Unit Root Test

data:  ln_G_W_Gap
Dickey-Fuller Z(t_alpha) = -0.99199, Truncation lag parameter = 3, p-value =
0.9279
alternative hypothesis: stationary
```

Figure 22

#ln_E_F

```
        Phillips-Perron Unit Root Test

data:  ln_E_F
Dickey-Fuller Z(t_alpha) = -2.2964, Truncation lag parameter = 3, p-value =
0.4566
alternative hypothesis: stationary
```

Figure 23

#ln_WBR_M

```
        Phillips-Perron Unit Root Test

data:  ln_WBR_M
Dickey-Fuller Z(t_alpha) = -2.0583, Truncation lag parameter = 3, p-value =
0.5507
alternative hypothesis: stationary
```

Figure 24

#ln_PPR.after.tax

```
        Phillips-Perron Unit Root Test

data:  ln_PPR.after.tax
Dickey-Fuller Z(t_alpha) = -2.7393, Truncation lag parameter = 3, p-value =
0.2817
alternative hypothesis: stationary
```

Figure 25

3.3 Cointegration

3.3.1 Johansen Test

```
Values of teststatistic and critical values of test:

           test 10pct  5pct  1pct
r <= 5 |   9.17  7.52  9.24 12.97
r <= 4 |  34.86 13.75 15.67 20.20
r <= 3 |  51.31 19.77 22.00 26.81
r <= 2 |  60.51 25.56 28.14 33.24
r <= 1 | 104.23 31.66 34.40 39.79
r = 0  | 130.54 37.45 40.30 46.82
```

Figure 26

3.3.2 Engle-Granger

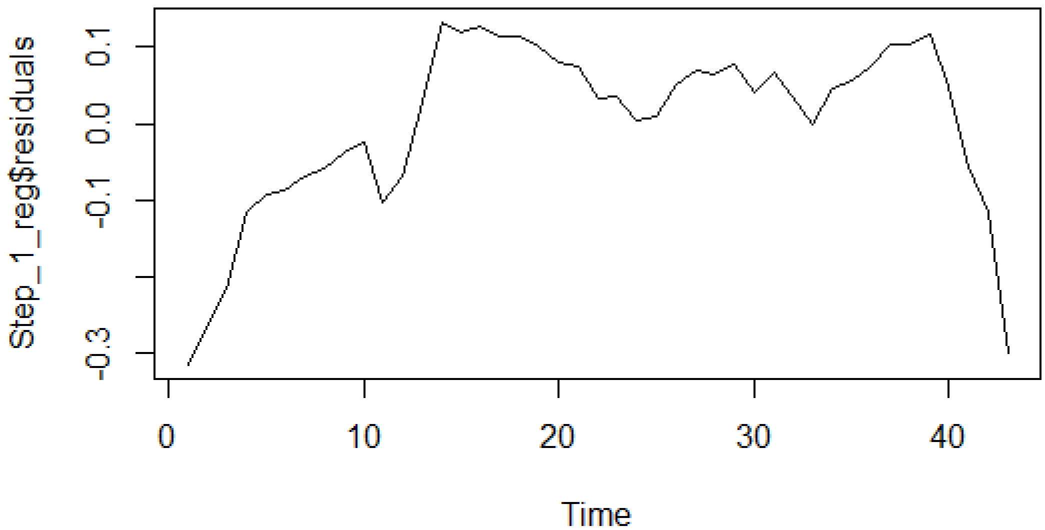

Figure 27

```
Value of test-statistic is: -1.3735

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.62 -1.95 -1.61
```

Figure 28

4 REGRESSION IN FIRST DIFFERENCE

	Dependent variable:
	GDP_Diff
AHW_M_Diff	0.284^{***}
	(0.052)
G_W_Gap_Diff	-0.095^{***}
	(0.013)
E_F_Diff	0.269^{***}
	(0.047)
WBR_M_Diff	0.439^{***}
	(0.044)
PPR_Diff	0.301^{***}
	(0.009)
Constant	-0.001
	(0.001)
Observations	42
R^2	0.978
Adjusted R^2	0.975
Residual Std. Error	0.004 (df = 36)
F Statistic	321.712^{***} (df = 5; 36)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 29

4.1 Autocorrelation Regression First Differences Residuals

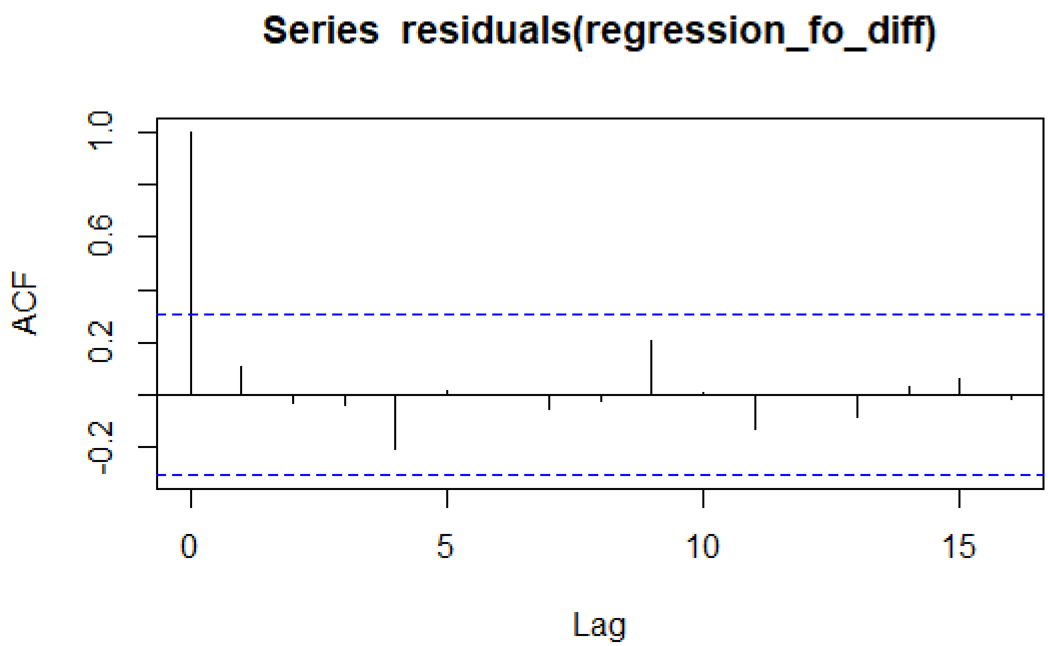

Figure 30

<u>*4.2 Breusch-Pagan test for autocorrelation*</u>

```
	studentized Breusch-Pagan test

data:  regression_fo_diff
BP = 5.4419, df = 5, p-value = 0.3644
```

Figure 31

<u>*4.3 Coefficients adjusted for autocorrelation in first order differences regression*</u>

```
t test of coefficients:

                Estimate  Std. Error t value  Pr(>|t|)    
(Intercept)  -0.00079066  0.00074063 -1.0675 0.2928328    
AHW_M_Diff    0.28426223  0.05561020  5.1117 1.068e-05 ***
G_W_Gap_Diff -0.09478063  0.02409001 -3.9344 0.0003649 ***
E_F_Diff      0.26917612  0.05213678  5.1629 9.125e-06 ***
WBR_M_Diff    0.43909678  0.04566215  9.6162 1.749e-11 ***
PPR_Diff      0.30084843  0.01231649 24.4265 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 32

<u>*4.4 Coefficients adjusted for heteroskedasticity in first order differences regression*</u>

```
t test of coefficients:

              Estimate   Std. Error t value  Pr(>|t|)
(Intercept)  -0.00079066  0.00083288 -0.9493    0.3488
AHW_M_Diff    0.28426223  0.04360635  6.5188 1.416e-07 ***
G_W_Gap_Diff -0.09478063  0.02042081 -4.6414 4.471e-05 ***
E_F_Diff      0.26917612  0.05307230  5.0719 1.206e-05 ***
WBR_M_Diff    0.43909678  0.04170861 10.5277 1.545e-12 ***
PPR_Diff      0.30084843  0.01110283 27.0965 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 33

__4.5 ADRL Model Residuals plot__

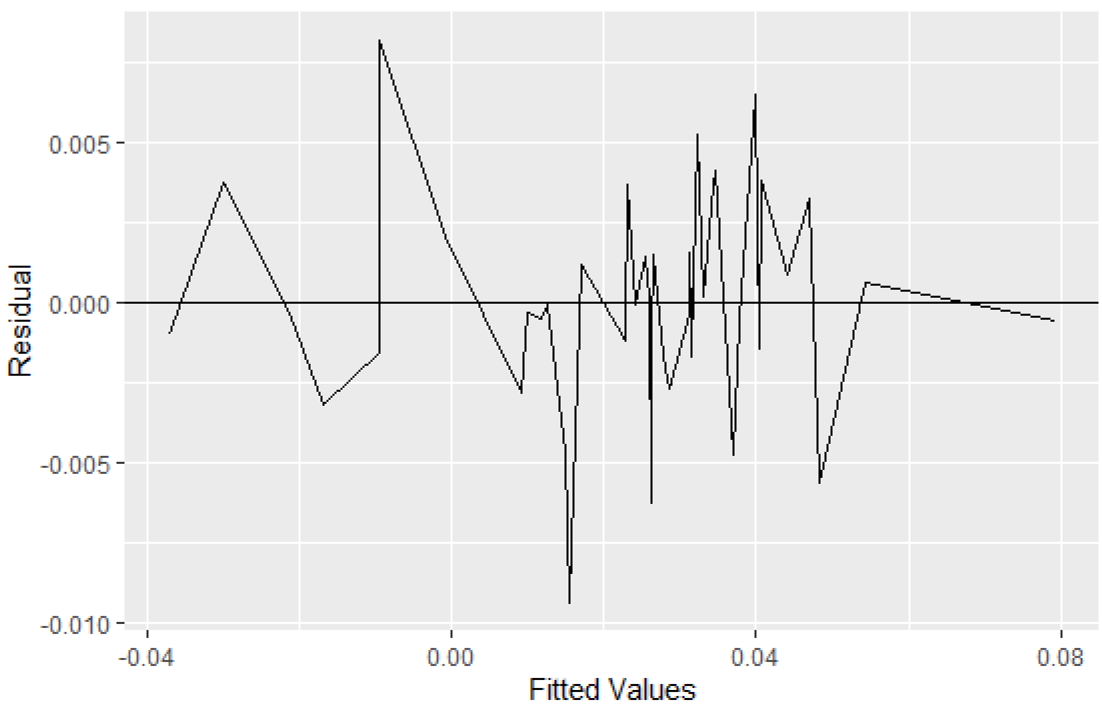

Figure 34