Gaspare Mattarella N° 916974

QUANTITATIVE METHODS FOR ECONOMIC AND SOCIAL RESEARCH

PART I

The paper by Konings and Lehmann *Marshall and Labor Demand in Russia: Going Back to Basics* represents the starting point for the first part of the present paper. By applying the same dataset the scholars used, we will try to estimate the Russian labour demand function *ln = f(lw, ly)* for the years 1996 and 1997 and assess which model better fits our data.

The first part is dedicated to the preliminary analysis of each variable. Here we will check for the presence of outliers, analyse the distributions of the variables and run the normality test.

We will then move on, taking into consideration just the data available for the year 1997 looking at the OLS regression of the labour demand function in that period. We will later run the tests for heteroskedasticity, normality, and the RESET test.

At the end of the first part, some comparison will be made between different models in order to find the best one for our data. We will start by looking at the results obtained from the OLS regression and the robust regression. Then, we will look at the differences between the robust OLS regression and the IV regression. Lastly, moving from a cross-sectional approach to a panel data setting, we will look at the differences between the pooled OLS regression and the fixed effects regression.

The second part of the paper will focus on the analysis of the Austrian system over the last 4 decades, taking into consideration specific variables such as Gross Domestic Product at constant prices, inflation rate, measured through the Gross Domestic Product deflator and the Consumer Price Index, and the unemployment rate.

1.1 Preliminary analysis:

Table 1.1. *Preliminary analysis.*

	ln[1]	**lw**[2]	**ly**[3]
N° observations	1420	1420	1420
Median (50%)	4.941	4.470	8.353
Mean (10% trim)	4.881	4.467	8.285
Mean	4.941	4.450	8.312
Std. Deviation	1.424	0.549	2.041
N mild outliers	31	14	11
N severe outliers	0	1	0
N total outliers	31	15	11
normality (Jarque-Bera test)	2.2e-16	2.675e-07	0.006182

In *Table 1.1* we can see summarised the characteristics for each variable of the labour demand function *ln = f (lw, ly).* As we can the data set is very rich with 1420 observations for each of our variable: ln (dependent variable) lw, and ly (independent variables). In order to better understand their features, we start checking for potential outliers. An outlier is an observation whose values are far outside the usual range of data and can either be mild or severe. A mild outlier has values which exceed the inner fences but are lower than the outer ones, whereas a severe outlier has values which exceed the outer fences.

Our variables present respectively a total of 31 (ln), 14 (lw) and 11 (lw) mild outliers and just the variable lw has a severe one. It is important to be aware of these extreme values as they have an impact on the distribution function and can affect the estimators of our regression models.

A useful way to exclude the extreme values from our mean is by computing the trimmed mean (10%) which disregards the 5% of the highest values and the lowest 5% in the computation.

For the variable ln we can see that the value of the trimmed mean is slightly lower (4.881) compared to the arithmetic mean, implying that in our sample there are some outliers which are to be found mostly beyond the higher inner fence. The same situation occurs for the variable ly whose trimmed mean of 8.285 is just a little lower than the mean implying the presence of some outliers mainly in the higher inner fence. On the other hand, the trimmed mean for the variable lw (4.467) is slightly higher compared to its arithmetic mean showing the presence of some outliers, but not too many, to be found mostly beyond the lower inner fence.

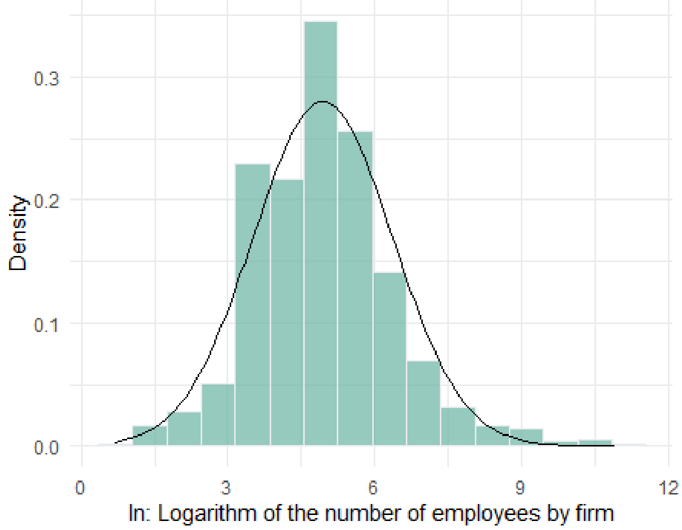
[1] Logarithm of the employment
[2] Logarithm of the ratio between wage and employment
[3] Logarithm of the output

The standard deviation measures the dispersion of a dataset relative to its mean and we can see from our results that the variable presenting the tighter distribution around its mean is lw (0. 549) whereas ly has the largest one (2.041).

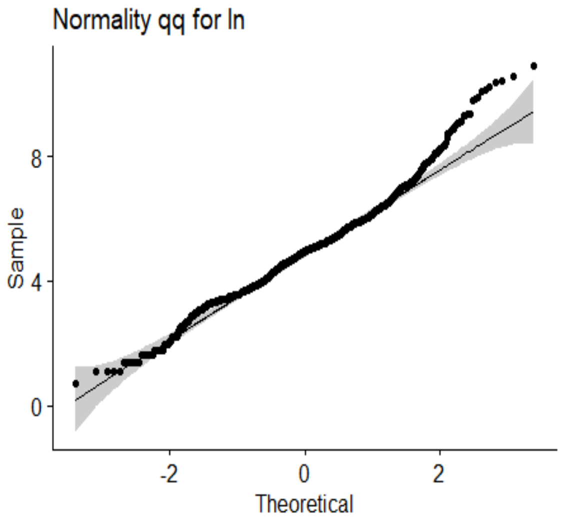
To conclude, according to the results of our Jarque-Bera test for *ln* and *lw*, and *ly*, one can say that the null hypothesis that all the variables *ln*, *lw* and *ly* are normally distributed is rejected.

In the *Graphs 1 to 6* we can see the comparison between the distribution functions for each variable and their standard normal distribution.

Graph 1. *Distribution of ln with an over imposed Standard Normal Distribution.*



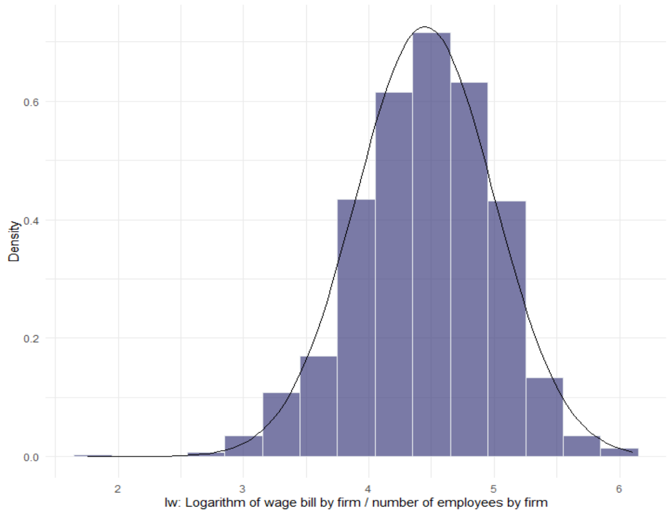
Graph 2*. Q-Q probability plot confronting the ln distribution to a Normal Distribution.*



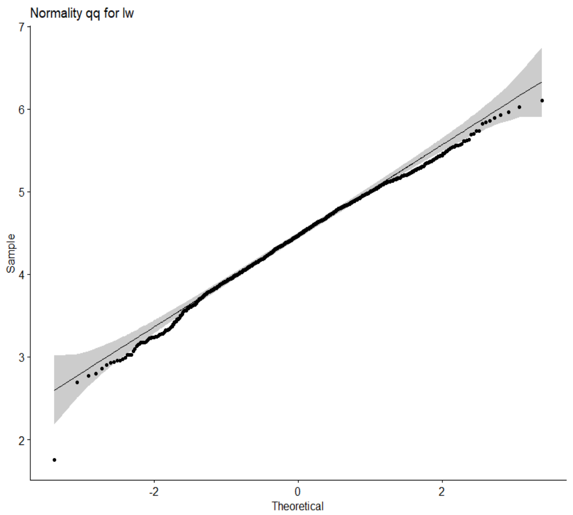
Source: Konings, J., & Lehmann, H. "Marshall and labor demand in Russia: going back to basis". Journal of Comparative Economics 30, p. 134–159 (2002).

The *ln* distribution, as it is possible to note in *Graph 1. And 2.*, is a little skewed to the right and in its wing, it is possible to clearly see a group of outliers that compromise the distribution. Nevertheless, our sample is big enough to be approximated to normal thanks to the CLT.

Graph 3. *Distribution of lw with an over imposed Standard Normal Distribution.*



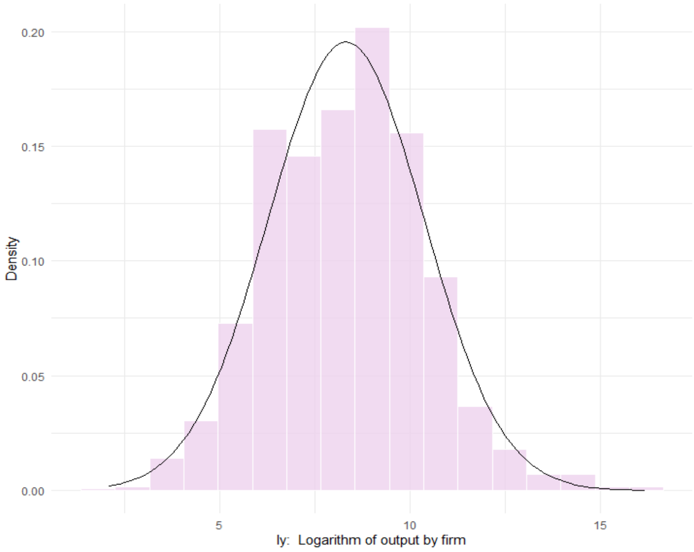
Graph 4*. Q-Q probability plot confronting the lw distribution to a Normal Distribution.*



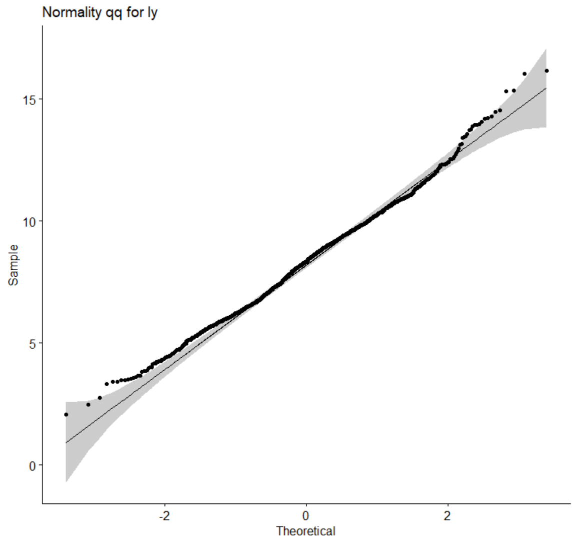
Source: Konings, J., & Lehmann, H. "Marshall and labor demand in Russia: going back to basis". Journal of Comparative Economics 30, p. 134–159 (2002).

The lw distribution is maybe, among the three, the one that better fits a Normal distribution but as the Jarque-Bera test already pointed out, and as we can see in *Graph 4.* its distribution is not Normal. Again, we can rely on the CLT.

Graph 5. *Distribution of ly with an over imposed Standard Normal Distribution.*



Graph 6*. Q-Q probability plot confronting the ly distribution to a Normal Distribution.*



Source: Konings, J., & Lehmann, H. "Marshall and labor demand in Russia: going back to basis". Journal of Comparative Economics 30, p. 134–159 (2002).

The ly distribution, as it is possible to note in *Graph 5. And 6.*, is a little skewed to the left presenting a group of outliers on both sides of the distribution.

According to the data analysed so far, none of our variables is distributed like a normal distribution but thanks to the Central Limit Theorem, arising in case of big samples as in our case, we can say that they well approximate the standard normal distribution.

1.2. OLS Regression Model

Using the OLS regression we estimate the labour demand function for the year 1997. The regression function we obtain running the OLS model is the following:

Table 1.2. *OLS Equation*

ln =	**1.902**	**-**	**0.707lw**	**+ 0.746ly**	**(1.1)**
	(0.156)		**(0.042)**	**(0.011)**	

Note. Standard Errors in parenthesis

The estimates of the parameters lw and ly represent the impact on ln of a unit change of the estimates. As we can clearly see in *Table 1.3*, the model is significant and the coefficients lw and ly are significant in predicting the dependent variable ln with an high R2 (0.8847).

If we look at the adjusted R-squared we can see that the value is slightly smaller (0.8844) compared to the R-squared which indicates that the addition of a new regressor reduced the sum of squared residuals of the regression.

Furthermore, the lw negative coefficient indicates that the independent variable has a negative relationship with the dependent variable and that the increase of a unit in lw corresponds to a decrease of about -0.71 in ln. Conversely, the coefficient of the independent variable (ly) is positively correlated with the dependent one (ln) which means that a one-unit increase in ly corresponds to an increase of about 0.75 of the dependent variable.

The p-values of the coefficients all are < 0.01 which means that they are statistically significant. Moreover, the standard errors of the estimated coefficients, which is way of measuring the accuracy of predictions made with a regression line, of lw and ly are respectively 0.0424009 and 0.0110596. These small values point out that the estimated parameters are highly effective in describing the dependent variable ln.

Table 1.3. *OLS Model, a summary.*

	ln	
ly	0.746***	(0.011)
lw	-0.707***	(0.042)
Constant	1.902***	(0.156)
Observations	710	
R2	0.881	
Adjusted R2	0.881	
Residual Std. Error	0.497 (df = 707)	
F Statistic	2,619.294*** (df = 2; 707)	

Note. Standard Errors in parenthesis and '***' means p<0.01

After having run the OLS regression it is very important to see whether the errors are heteroskedastic or homoscedastic. The *Breusch-Pagan test* for heteroskedasticity[4] tests the null hypothesis that the error variance is constant versus the alternative hypothesis that the error variance is not constant. As

[4] When the conditional distribution of the error term has a variance that depends on the independent variables, then the error is *heteroskedastic*. Therefore, if the error terms do not have constant variance, there is heteroscedasticity. Heteroscedasticity does not cause bias or inconsistency in the OLS estimators but causes inefficiency. In fact, if there is heteroskedasticity the OLS estimates are no longer BLUE.

we can see below from *Table 1.4*, we can reject the null that the variance of the residuals is constant, thus heteroskedacity is present.

Table 1.4. *Breusch-Pagan Test for Heteroskedasticity.*

```
-------------------------------
 DF             =    1
 Chi2           =    12.02243
 Prob > Chi2    =    0.0005256425
-------------------------------
```

Having found out that our errors are heteroskedastic we further calculate the heteroskedastic robust errors and as it is possible to note in *Table 1.5* how they are now slightly bigger than they were before in *Table 1.3.*

Table 1.5 *Heteroskedasticity Robust Errors vs. Simple OLS Std. Err.*

```
-----------------------------------------
Robust Std.Err.  OLS Std.Err.
     0.168          0.156
     0.016          0.011
     0.050          0.042
------------------------------------------
```

Running a Robust Regression, we can now obtain corrected coefficients for HC and as we can see in *Table 1.6* now the new coefficients have higher values while the std. errors are lower. The p-values, all way below the threshold of 0.05, tell us the coefficients are all statistically significant.

Table 1.6. *OLS Robust Equation*

$$\begin{aligned} \mathbf{ln} = \; & \mathbf{1.952} \; - \; \mathbf{0.756lw} \; + \mathbf{0.764ly} \\ & \mathbf{(0.136)} \quad \mathbf{(0.036)} \qquad \mathbf{(0.009)} \end{aligned} \tag{1.2}$$

Notes. Robust Standard Errors in Parenthesis.

Furthermore, I'm going to check for the potential presence of an omitted variable bias running the Ramsey Regression Equation Specification Error Test (RESET) on the linear regression.

Given the results in *Table 1.7* it is possible to state that we don't have enough evidence to claim that the model has no omitted variables. So, it may suggest the inclusion of non-linear effect in the model's specification but, we can also assume that it can arise from the presence of outlier observations.

Table 1.7. *RESET Test for omitted variables.*

```
------------------------------
RESET test
data:  lb_lm_1
RESET = 8.4046, df1 = 2, df2 = 705,
p-value = 0.000247
```

The intuition behind the test is that if non-linear combinations of the explanatory variables have any power in explaining the dependent variable, the model is mis specified in the sense that the data generating process might be better approximated by a polynomial or another non-linear functional form.

Table 1.8 shows the results of a Shapiro – Wilk testing the residuals of our OLS model (1.1) for normality. With a p-value way below the threshold of 0.05 the null hypothesis that the residuals are normally distributed is rejected as we can also see from *Graph 7.*

Table 1.8. *Shapiro - Wilk Test for normality.*

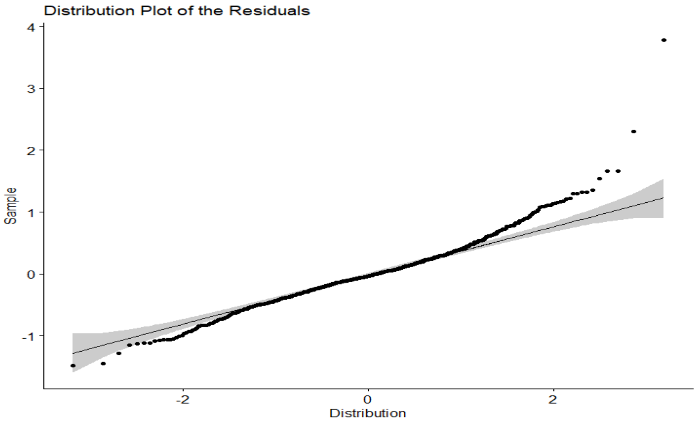
```
-------------------------------
Shapiro - Wilk Normality Test

Test Results:

STATISTIC:
W: 0.9503

P VALUE:
1.003e-14
-------------------------------
```

Graph 7. *Distribution plot of the Residuals of (1.1)*



1.3. Instrumental Variables (IV) Regression

The RESET test (*Table 1.7*) has shown that there are omitted variable bias, meaning that there is a correlation between one of the two regressors *lw* and *ly* (or both) and the error term. One way to overcome this problem is to perform an *Instrumental Variable (IV) regression*. Using this method, it is possible to estimate the effect on Y of a unit change in X using the two stage least squares (TSLS) estimator. In our case, we will use as instruments the lagged data, that is the first lag of our variables for one period (t -1).

The IV equation is the following, reported in *Table 1.9*:

Table 1.9. *IV Regression Equation.*

$$\mathbf{ln} = \underset{(0.206)}{2.192} - \underset{(0.061)}{0.834lw} + \underset{(0.016)}{0.781ly} \quad (1.3)$$

Note. Robust Standard Errors are in parenthesis.

In this model I used as instruments the lags of the variables *lw* and *ly* because any other model failed to pass the Sargan test for overidentifying restrictions.

The diagnosis presented in *Table 2.1.* Shows that it is possible to compare equations (1.2) and (1.3) through a Hausman specification test, that allows us to evaluate the consistency of the estimators by testing the null hypothesis that the IV model is consistent and the OLS regression is efficient. The result shows that the null can be rejected leading us to assume that IV and OLS estimates are significantly different from each other and hence, considering the OLS estimation as inconsistent.

Table 2.1. *Diagnostics of the Wu – Hausman test.*

Diagnostic test	df1	df2	statistic	p-value	
Wu-Hausman	2	705	52.5	<2e-16	***

Note. '***' means p<0.01

A comparison between the two models is provided by the *Table 2.2.*
The coefficients of the IV reg are higher than the OLS and so are the Std. Errors but the model as a whole and the coefficients are statistically significant.

Table 2.2. *Coefficients and Standard errors between the Robust OLS and the IV reg.*

	IV Reg	OLS
ly	0.781***	0.764***
	(0.017)	(0.010)
lw	-0.835***	-0.756***
	(0.061)	(0.037)
Constant	2.192***	1.953***
	(0.207)	(0.136)
Observations	710	710
R2	0.879	
Adjusted R2	0.879	
Residual Std. Error (df = 707)	0.501	0.394

Notes. Robust Standard Errors in Parenthesis. '***' means p-value<0.01

2. Panel Data Approach

In the previous paragraphs we have used a cross-sectional approach but from now on we will adopt a panel approach. Panel data consist in data for multiple entities, where each entity is observed at two or more time periods. We proceed with the calculation of two regressions, estimating the labor demand function using the pooled OLS regression and the Fixed Effects regression. The pooled OLS regression is without effects across entities or across time, but it is potentially biased. The fixed effect is an estimation method that allows "to control *for omitted variables in panel data when the omitted variables vary across entities (states) but do not change over time*"[5]. The fixed effects regression model has n different intercepts, one for each entity.[6] This is the kind of model I am going to use in this paragraph but it's important to understand that there is another equivalent way to write the fixed effects regression model with a common intercept and n - 1 binary regressors. In both formulations, the slope coefficient on X is the same from one state to the next. The state-specific intercepts and the binary regressors have the same source: the unobserved variable *Zi* that varies across entity but not over time[7].

Table 2.3. *Coefficients, robust Standard errors and diagnostics of Fixed and Pooled Effects models.*

Dependent variable: ln	Time & Entity fixed effects (1)	Pooled (2)	Fixed Effects (3)
ly	0.301***	0.754***	0.332***
	(0.031)	(0.015)	(0.029)
lw	-0.319***	-0.754***	-0.405***
	(0.049)	(0.044)	(0.038)
Constant		2.027***	
		(0.147)	
Year1997	-0.043***		
	(0.010)		
Observations	1,420	1,420	1,420
R2	0.996	0.886	0.999
Adjusted R2	0.992	0.886	0.999
Residual Std. Error	0.128 (df = 707)		0.131 (df = 708)
F Statistic	3197***	5,496***	3,084***

*Notes. Robust Standard Errors in Parenthesis. '***' means p<0.01.In this table are omitted the 710 intercepts.*

[5] Stock, J. H. and Watson, M. W., *Introduction to Econometrics.* cap. 10
[6] Stock, J. H. and Watson, M. W., *Introduction to Econometrics.* cap. 10
[7] Stock, J. H. and Watson, M. W., *Introduction to Econometrics.* cap. 10

If some omitted variables are constant over time but vary across states while others are constant across states but vary over time, then it is appropriate to include both entity and time effects. The combined state and time fixed effects regression model eliminates omitted variables bias arising both from unobserved variables that are constant over time and from unobserved variables that are constant across states[8].

Comparing the pooled OLS and the two fixed effects regressions, we can see that there is a big difference between them as shown in *Table 2.3*.

In the Fixed Effects, the coefficients of *lw* and *ly* are way lower than those of the pooled OLS. We can also see that *lw* maintains its negative correlation with *ln*, and *ly* the positive correlation with *ln* but they can differ a lot across the models. The Entity and Time Fixed Effects model has the lowest coefficients, meaning that, after having controlled for possible unobserved variables and consequent bias, the actual effects of the variables result limited. While the coefficients of every model are all statistically significant with a p-value below the threshold of 0.05, the R2 reaches a value of .99 in both the fixed effects against the 88 of the Pooled effects model. Furthermore, controlling for the consistencies of the models with an Hausman test and for the significance of the individual effects with an F-test for individual effects as shown in *Table 2.4*, we can clearly affirm that the Pooling model is inconsistent and that both the "individual" and the "twoways" effects are significant compared to the Pooled one and thus, the Fixed Effects model are better suited for the correct estimation of the coefficients and for further analysis.

Table 2.4. *Hausman test and F-test for individual effects comparing the FE model and the Pooled one.*

```
=========================================================================
Hausman Test, data:  ln ~ ly + lw

chisq = 114.37, df = 2, p-value < 2.2e-16
alternative hypothesis: one model is inconsistent

F test for individual effects, Data:  ln ~ ly + lw

F = 26.12, df1 = 709, df2 = 708, p-value < 2.2e-16
alternative hypothesis: significant effects
=========================================================================
```

Final Notes:
All the Tables and the Graphics, the models and the calculations in this paper so far have been produced on R thanks to the same date base used by Konings and Lehmann. They all are fully reproducible thanks to the script attached to the present paper.

[8] Stock, J. H. and Watson, M. W., *Introduction to Econometrics.* cap. 10

PART II

3. TIME SERIES

In the second part of our paper we will be working in a time series setting, analysing the Austrian system between 1980 and 2024. We will be looking at the Austrian Gross Domestic Product at constant prices (NGDP_R), the inflation, measured through the Gross Domestic Product deflator (NGDP_D) and the consumer price index (PCPI), and the unemployment rate (LUR).

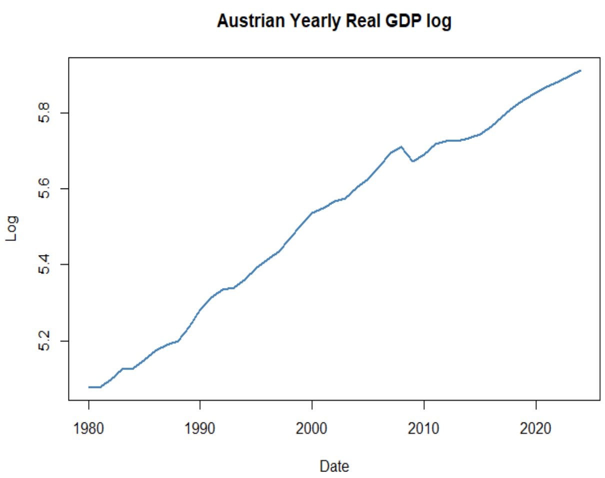
Firstly, we will analyse the data of these variables representing their logged values and growth rate graphically; secondly, we are going to represent them by using a correlogram and finally we are going to check whether the variable's logged levels and first difference of the log are stationary or not.

We will then further our analysis by computing the inflation rate using the CPI and the GDP deflator and comparing their results. At the end of the paper we will estimate the "Phillips curve".

3.1 Data Representation

Firstly, we will look at the GDP at constant prices, an inflation-adjusted measure that reflects the value of all goods and services produced by an economy each year, in Austria analysing the data from 1980 to 2024. As we can see in *Graph 8* the logged GDP presents very mild fluctuation but the overall trend from 1980 until 2008 is positive. In the year of the financial crisis in fact the constant growth came to an end the Austrian GDP experienced a downturn after which slowly kept growing. This can be seen in *Graph 8*. Despite the constant fluctuation the GDP growth rate stayed positive, plummeting just in 2008 but recovering straight afterwards.

Graph 8. *Austrian Gross Domestic Product (GDP) at constant prices*

8.a) Gross Domestic Product growth rate	*8.b) Gross Domestic Product growth rate*
	

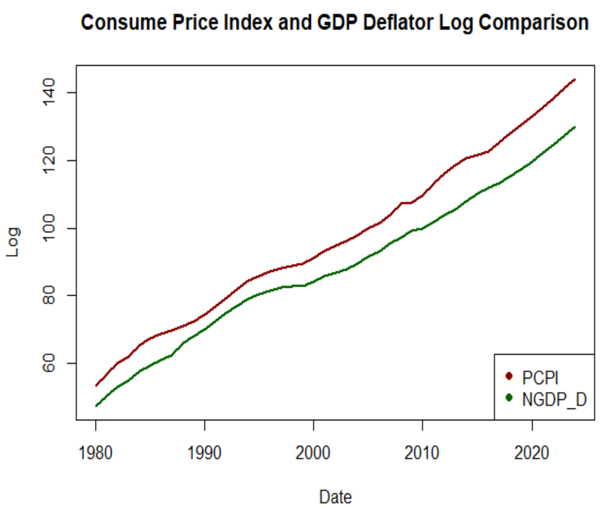
Source: *World Economic Outlook (International Monetary Fund)*

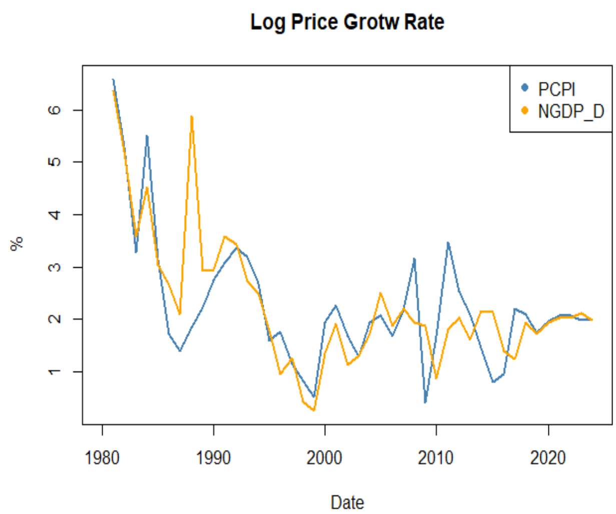
The second important aspect to look at is the price inflation over the same period. Two ways to measure it is through the GDP deflator, a measure of the level of prices of all new, domestically produced, final goods and services in an economy in a year, and the CPI (Consumer Price Index), which measure of the overall prices of the goods and services bought by a typical consumer.

Looking at *Graph 9.a)* we can see that the growth is constant and that there is no particular divergence between the two indicators even if the CPI seems to be more sensible, we can thereby note a peak in 2008 that is emphasized more rather than GDP deflator. The steep increase of logs just before 1990 happens while we observe a sharp spike in the growth rates. We can see an increasingly flattening slope when the growth rates slow in 2000, getting close to 0.
The main difference between the two growth rates is that the high peak present in 1989 is not detected by the CPI which registers on the other hand peaks around 2008 and 2011 not detected by GDP deflator.

Graph 9. *Austrian Inflation both measured through GDP Deflator and Consumer Price Index.*

9.a) GDP deflator and CPI logs	*9.b) Log Price Growth Rate*



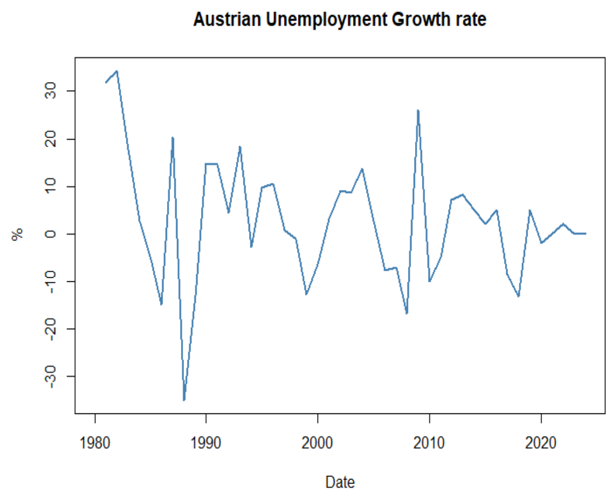


Source: World Economic Outlook (International Monetary Fund)

Lastly, we take into consideration the unemployment rate (the percentage of unemployed workers in the total labour) of these past decades.
As we can see in *Graph 10.* there has been an overall mild positive trend. However, it is important to point out how Austria registered some negative spikes during these years. The first one corresponds to 1989 in which it was registered an unemployment rate of 2.3%. Afterwards we notice another period of growth followed by the second downturn in 2000, probably due to the potential change in currency. Finally, the third pit can be seen in 2008 where unemployment at 4.1% because of the financial crisis.

Graph 10 *Austrian Unemployment.*

10.a) Unemployment Rate	*10.b)* Unemployment Growth Rate

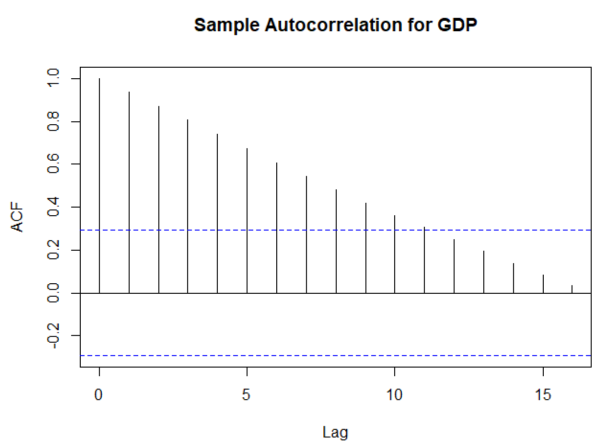


Source: World Economic Outlook (International Monetary Fund)

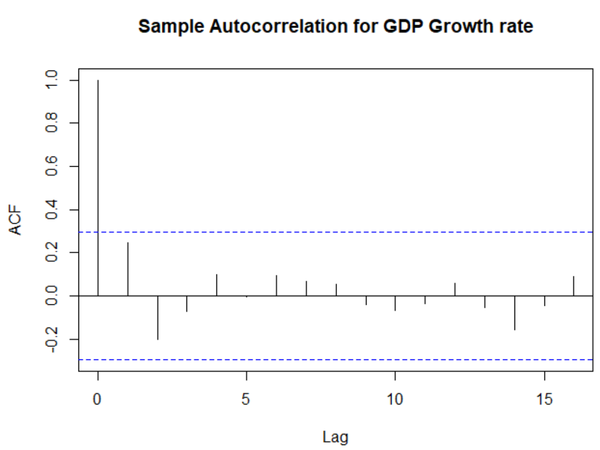
3.2 Correlograms

We will continue our analysis by looking at the correlograms of the logged values of our variables and at their growth rate as well. A correlogram is a visual way to show serial correlation in data that changes over time and it is useful to see whether our data show autocorrelation or not. The autocorrelation is on the y axis whereas the x axis tells you the lag. In the case of logged values of the Austrian GDP at constant prices we can see that there is a high degree of linear association, a deterministic trend, which decreases as the number of lag increases (*Graph 11*). Here we can say that the values at a specific period are influenced by the previous ones. On the other hand, some of lags of the GDP growth rate show positive correlation whereas some others a negative one. In this case no regular trend can be detected, and we can say that here there is no autocorrelation (*Graph 12*)

Graph 11. *Correlogram of GDP logs*



Graph 12. *Correlogram of GDP growth rate*

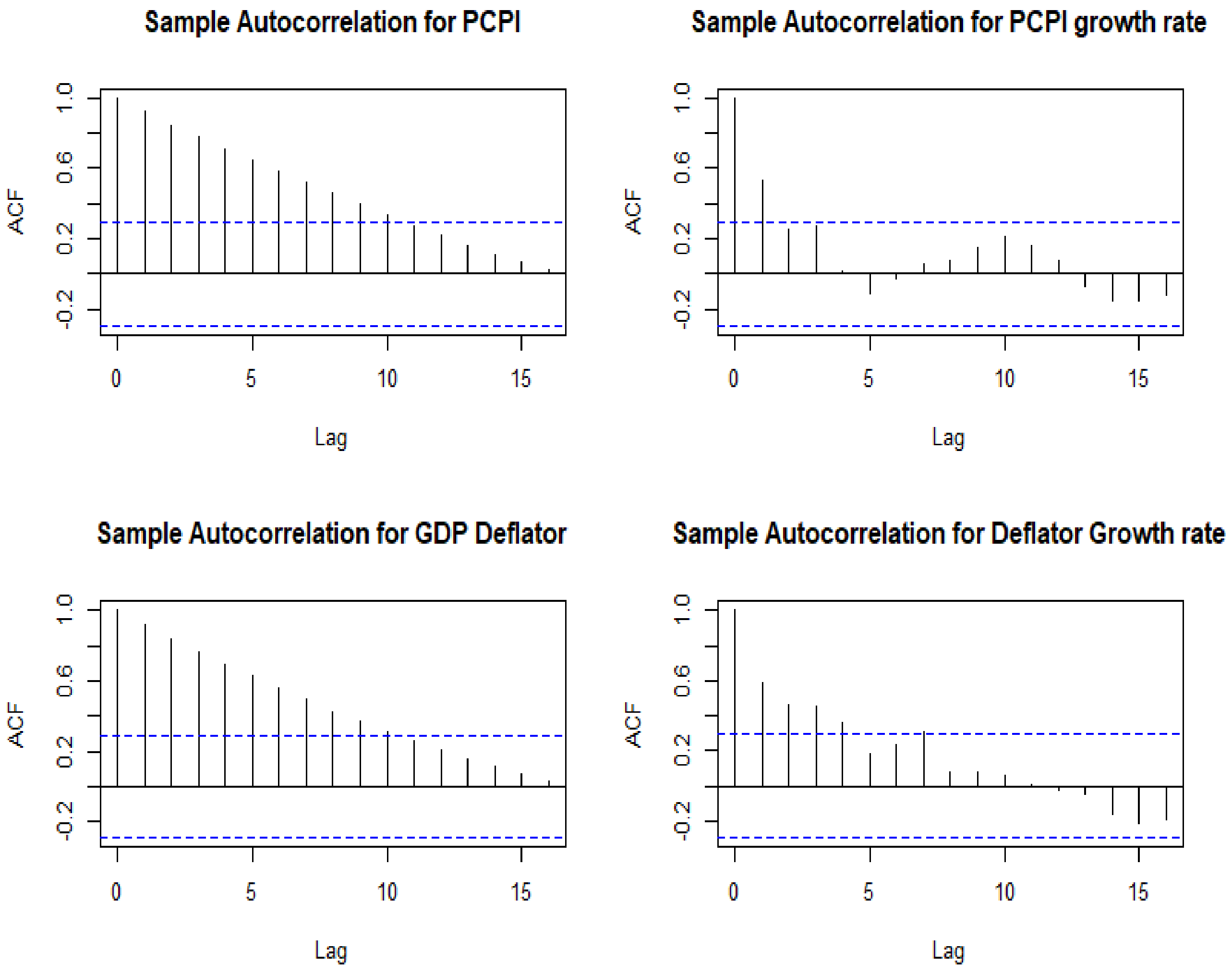


In order to talk about Inflation, we again take in consideration both GDP Deflator and CPI.

We can see here (*Graph 13.)* as well a strong positive autocorrelation decreasing as the number of lags increases in the logged levels for both the indicators. This means that inflation is correlated with its past values that we can consider helpful in order to understand its rate and growth in any given period of time.

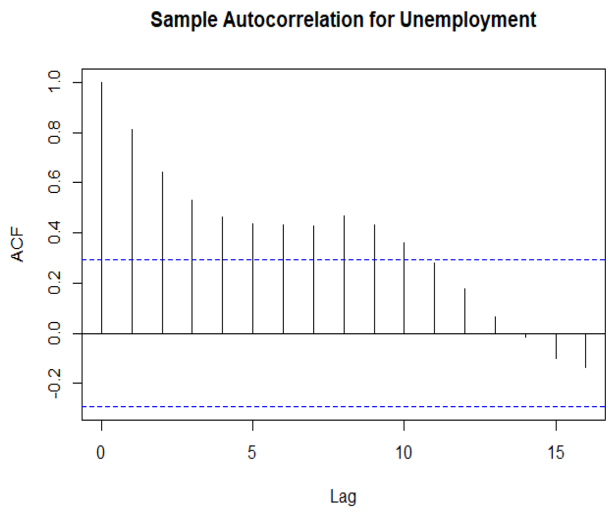
Regarding the respective Growth rates, the trend is not regular and the lags in some cases show weak negative correlation and in some other weak positive one, except for the first lags.

Graph 13. *Correlograms for CPI and GDP Deflator logs and Growth Rate.*

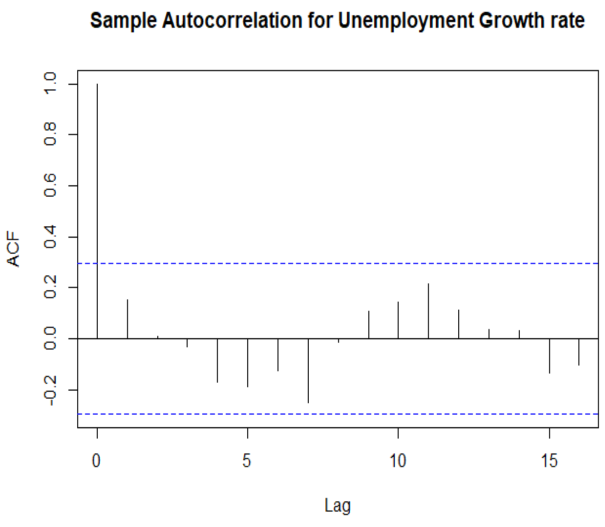


Finally, looking at the logged values of the unemployment rate (*Graph 14*) we can say here as well that there is a high level of autocorrelation. The first 11 lags present in fact a positive decreasing trend, whereas for what concerns the unemployment growth rate correlogram (*Graph 15*) we can see how their pattern is random.

Graph 14. *Correlogram of unemployment logs*



Graph 15. *Correlogram of unemployment growth rate*



3.3 Stationarity and Dickey-Fuller test

After having check the autocorrelation of the variables with their lags we proceed running the Augmented Dickey-Fuller (ADF) test in order to understand if our variables present a stochastic trend or not. The null hypothesis of the ADF test, is that the series has a unit root, which is to say that it has a stochastic trend. In this case the variable of interest is distributed randomly over time. On the other hand, under the alternative hypothesis there is no unit root, hence we have stationarity.

Choosing the lag length to be used in the ADF must be done carefully since a low number of lags could lead to the potential omission of valuable information whereas a high one could cause the introduction of additional estimation errors into the forecast.

With this in mind, before running the ADF test for each variable's logged level and first difference, we have to run a function that allow us to compute and compare the BIC (Bayes Information Criterion) and the AIC (Akaike Information Criterion), two information criterion that suggest the number of lag to be included in autocorrelation. The results of the BIC and AIC tests are here omitted[9] but for all tests the two different criterions gave me the same results, a lag order between 1 and 2.

Here in *Table 3.1.* the results of the ADF tests:

[9] The results are, obviously, reproducible with R.

Table 3.1. *ADF tests*

```
=============================================================================
      Augmented Dickey-Fuller Test // alternative hypothesis: stationary

data:  GDP first difference of the log
Dickey-Fuller = -5.3977, Lag order = 1, p-value = 0.01

data:  GDP Log Levels
Dickey-Fuller = -1.4665, Lag order = 1, p-value = 0.7856

data:  PCPI Log Levels
Dickey-Fuller = -3.2644, Lag order = 2, p-value = 0.08938

data:  PCPI first difference of the log
Dickey-Fuller = -4.2354, Lag order = 1, p-value = 0.01

data:  GDP Deflator log Levels
Dickey-Fuller = -2.4383, Lag order = 2, p-value = 0.4

data:  GDP Deflator first difference of the log
Dickey-Fuller = -3.2291, Lag order = 1, p-value = 0.09513

data:  Unemployment first difference of the log
Dickey-Fuller = -4.5551, Lag order = 2, p-value = 0.01

data:  Unemployment Log Levels
Dickey-Fuller = -3.8968, Lag order = 1, p-value = 0.02253
=================================================================================
```

After running the ADF, as shown in the *Table 3.1*, we can state the following: for the GDP, CPI, and Unemployment first differences of the log we can reject the null hypothesis that there is a unit root. As for the Unemployment log levels. This is to say that they probably present a stationary trend. On the other hand, for the GDP and CPI Log levels and both GDP deflator log levels and first differences of the log we cannot reject the null that there is a Unit Root and therefore they must be considered to have stochastic trends.

3.4 "Phillips Curve"

The intuition behind the Phillips Curve is that within an economy a change in unemployment has a predictable effect on price inflation. The inverse relationship between unemployment and inflation is depicted as a downward sloping, curve. Inflation is shown on the y-axis whereas unemployment on the x-axis. Increasing inflation decreases unemployment, and vice versa.
In our case, to explain further levels of first differences of the logged values of the Consumer Price Index, we will consider its past history and the past levels of the unemployment rate. As in the Augmented Dickey Fuller test analysis, it is important to choose the right number of lags.

In order to do so, I used a function[10] to loop over our model *1:n lags* of our variables and filtered for the model that minimize the BIC and the AIC. In this case, we will use an autoregressive distributed

[10] The details of the functions can be found and reproduced in the R script, as everything else.

lag (ADL) model, autoregressive because lagged values of the dependent variable are included as regressors, as in an autoregression, and distributed lag because the regression also includes multiple lags (a "distributed lag") of an additional predictor. Both BIC and AIC criterions suggest that the best suited model for this task is an ADL(1,1) i.e. a model with just one lag for each variable.

Table 3.2. *ADL(1,1) model*

	Dependent variable:
	ADL(1,1)
L(P)	0.447***
	(0.097)
L(U)	-0.002**
	(0.001)
Constant	0.020***
	(0.005)
Observations	41
R2	0.423
Adjusted R2	0.392
Residual Std. Error	0.008 (df = 38)
F Statistic	13.917*** (df = 2; 38)

*Notes. Robust Standard Errors in Parenthesis. '***' p<0.01, '**' p<0.05.*

Table 3.2. represent the ADL (1,1) and what can be noticed is that the values of the coefficients are very low, particularly the constant and the unemployment but we can still note the "classic" trend of the Phillips Curve: future inflation is negatively influenced by past levels of unemployment along with past levels of inflation itself. Though, even if the model and all the coefficients are statistically significant, the relatively low R^2 (0.40) suggests that the model could forecast further levels of first differences of the logged values of the Consumer Price Index only approximately.

Final Notes:
All the Tables and the Graphics, the models and the calculations in the second part of this paper have been produced on R thanks to the World Economic Outlook database of the International Monetary Fund. They all are fully reproducible thanks to the script attached to the present paper.