Econ 104 Project 3

Group 12

2024-01-02

Contents

1. Panel Data Model

(a) Briefly discuss your data and the question you are trying to answer with your model.

Ciation

```
## To cite AER, please use:
##
##   Christian Kleiber and Achim Zeileis (2008). Applied Econometrics with
##   R. New York: Springer-Verlag. ISBN 978-0-387-77316-2. URL
##   https://CRAN.R-project.org/package=AER
##
## A BibTeX entry for LaTeX users is
##
##   @Book{,
##     title = {Applied Econometrics with {R}},
##     author = {Christian Kleiber and Achim Zeileis},
##     year = {2008},
##     publisher = {Springer-Verlag},
##     address = {New York},
##     note = {{ISBN} 978-0-387-77316-2},
##     url = {https://CRAN.R-project.org/package=AER},
##   }
```

Description of the data

```
## [1] "N = 18"
```

```
## [1] "T = 19"
```

The dataset we chose, `OECDGas`, is a ***balanced panel data*** on gasoline consumption in 18 OECD countries over 19 years, spanning from 1960-1978.

As the original panel data has 18 individuals (N; countries) observed for 19 years (T), the original panel data is considered *wide and long.* However, for this project purpose, we selected 5 countries with relatively high average gasoline consumption over the years, which include `Spain`, `Italy`, `Japan`, `Netherlands`, `And Greece`. After reducing the number of individuals, we are now dealing with **Long and Narrow** panel data, where $N = 5$ and $T = 19$.

In this first part of the project, the question we attempt to answer is:

"Within the same time frame, how do the impacts of per-capita income, gasoline prices, and the number of cars on motor gasoline consumption differ across countries?"

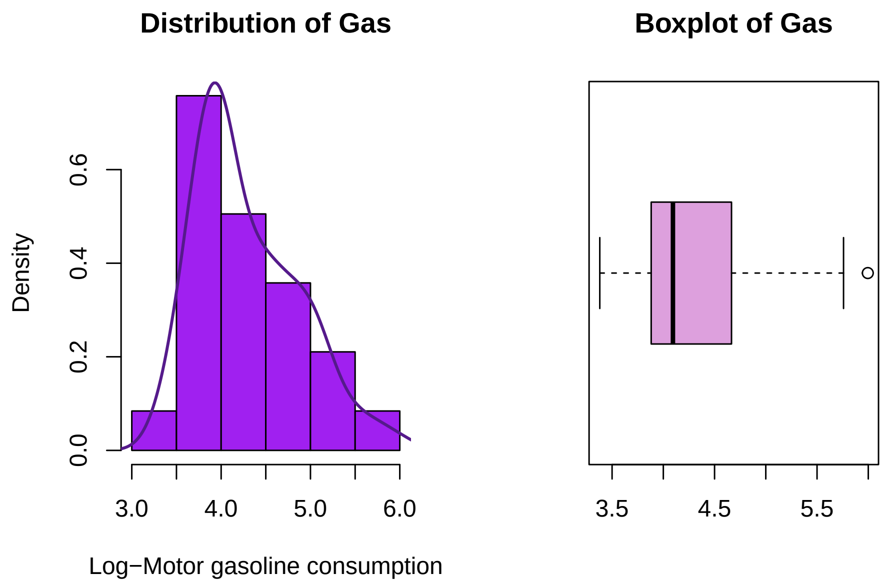
In other words, we are attempting to identify whether or not the impacts of the independent variables are statistically different across countries, and if so, which model is the most appropriate to capture the dynamics, among pooled, fixed effects, and random effects models.

For more detailed descriptions of the variables, refer to part (b).

(b) Provide a descriptive analysis of your variables. This should include relevant figures with comments including some graphical depiction of individual heterogeneity.

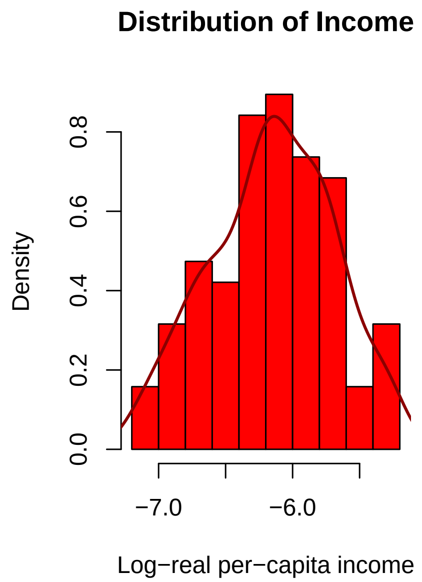
Dependent Variable

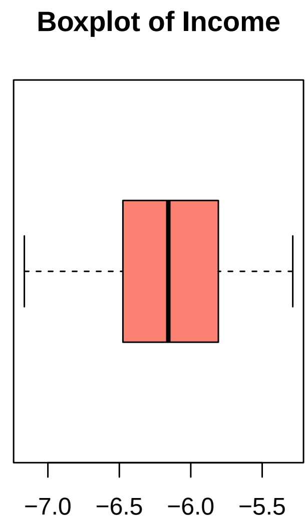
gas: Logarithm of motor gasoline consumption per car. Our dependent variable, `gas`, appears to be right-skewed.



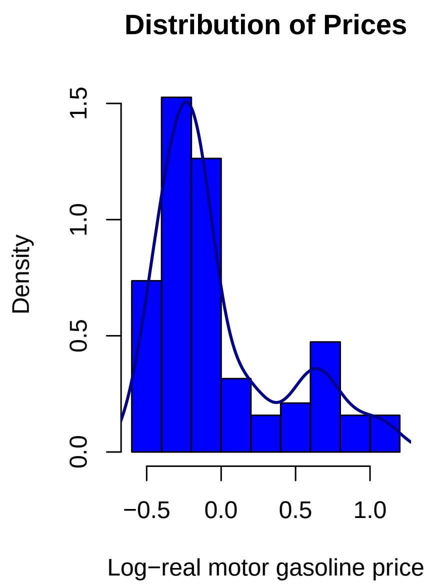
Independent Variable

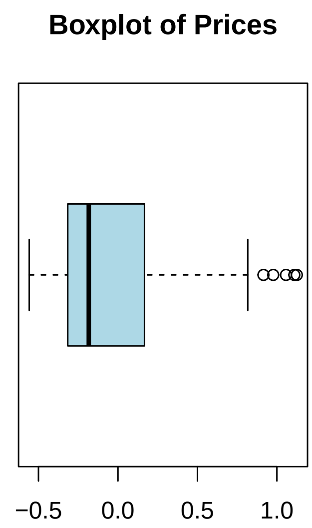
income: Logarithm of real per-capita income. `income` variable is slightly left-skewed.



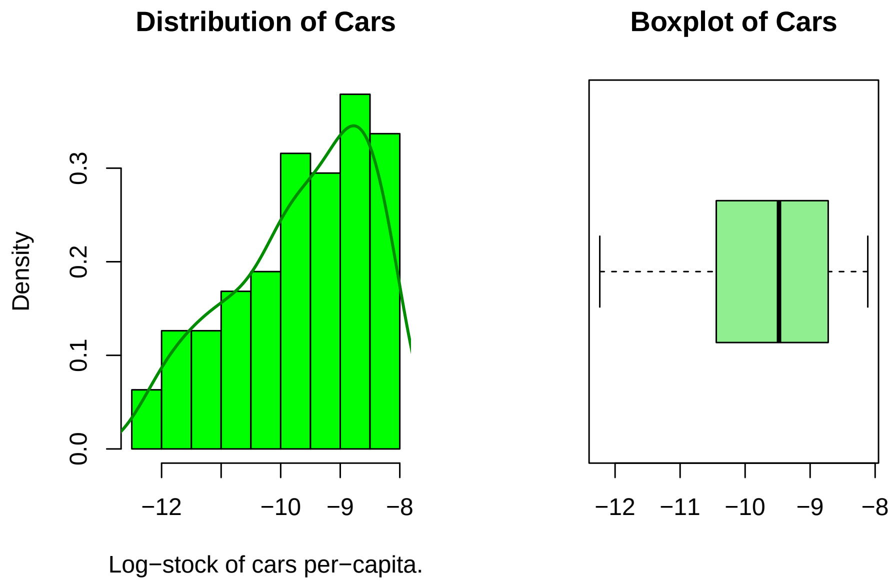


`price`: Logarithm of real motor gasoline price. `price` is right-skewed.



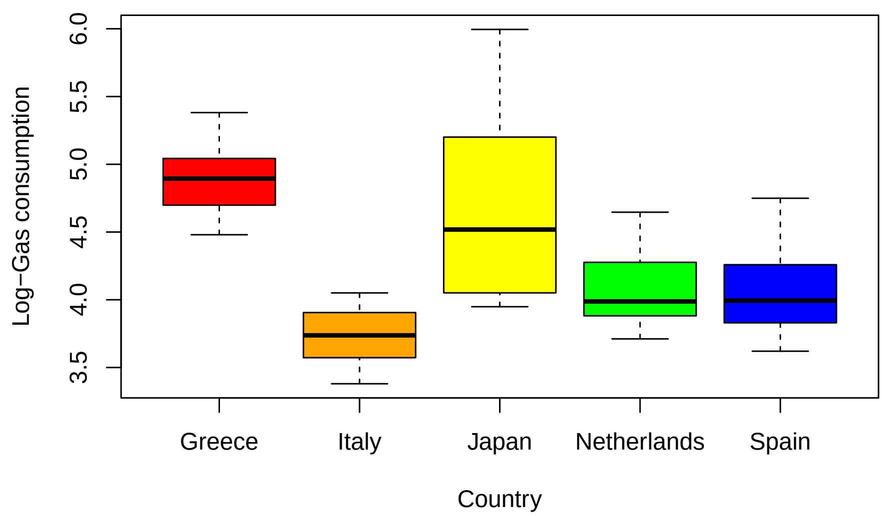


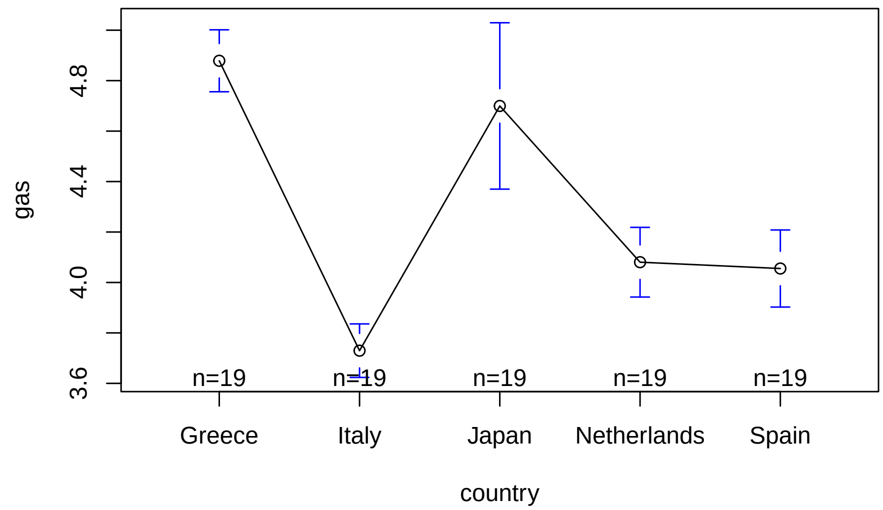
`cars` : Logarithm of the stock of cars per-capita. `cars` is left-skewed.



(c) Fit the three models below, and identify which model is your preferred one and why. Make sure to include your statistical diagnostics to support your conclusion, and to comment on your findings.

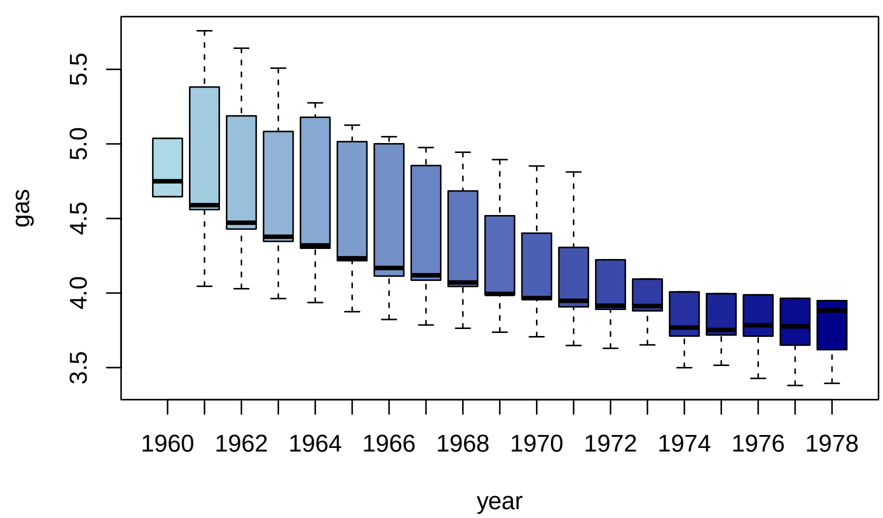
Exploring Heterogeneity across Countries

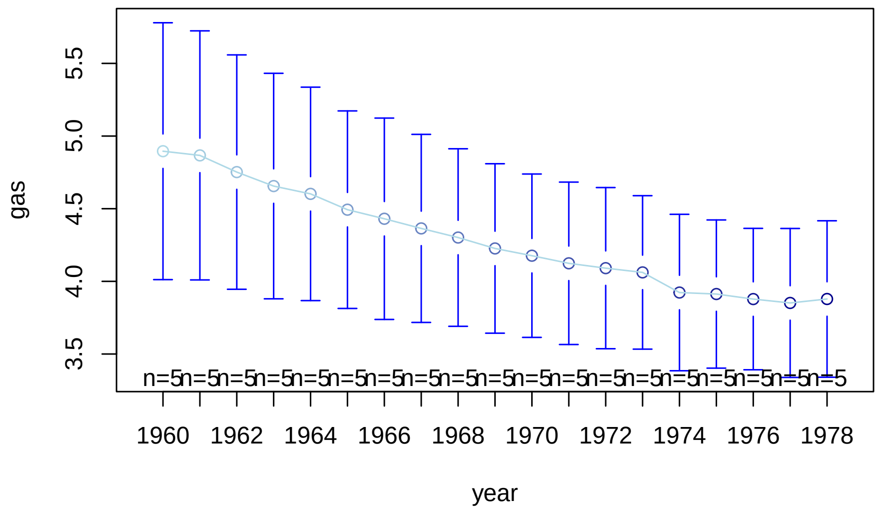




As shown above, heterogeneity across individuals exists. `Greece`, `Italy`, `Netherlands`, and `Spain` seem to be similarly spread out, however, `Japan` seems to have a higher variance compared to the others. While we can't conclude anything statistically at this point, it still provides a good intuitive sense that each country may not be constant.

Exploring Heterogeneity across Time

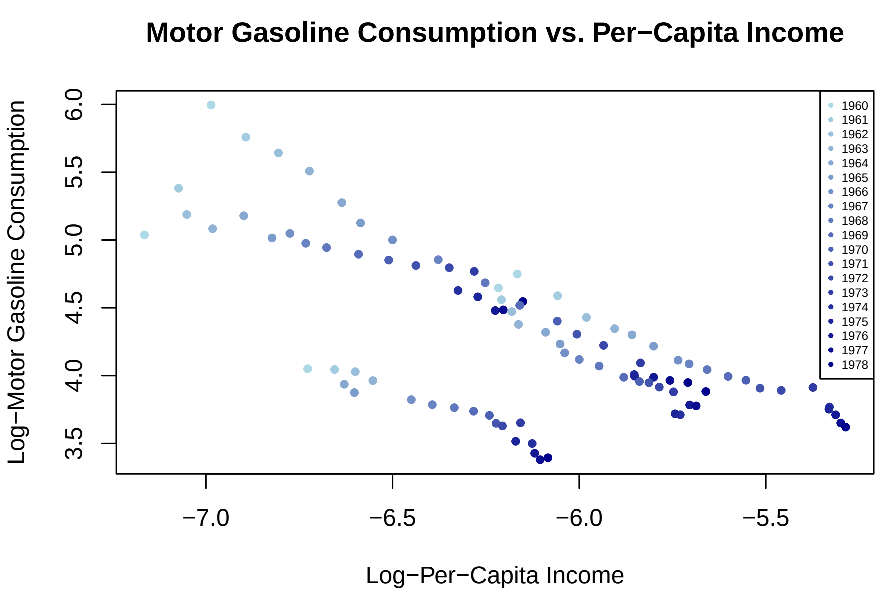


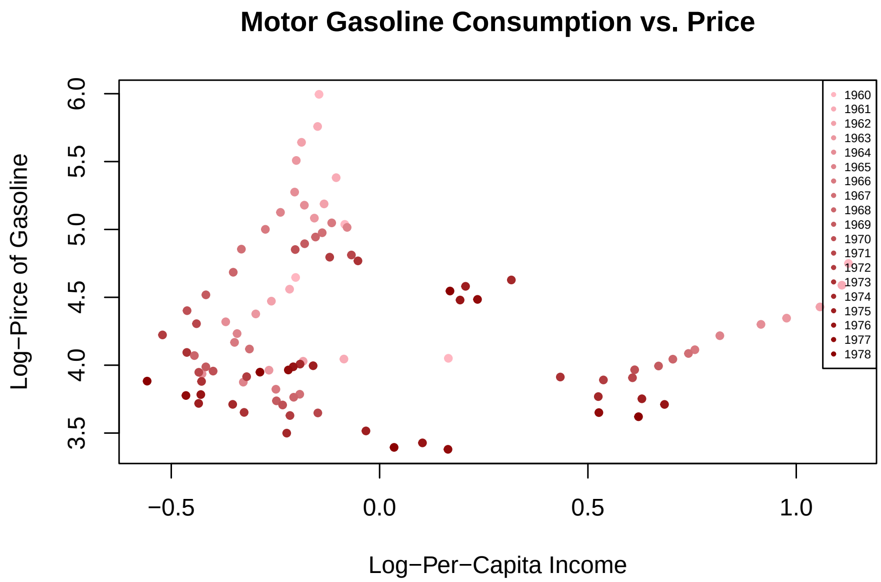


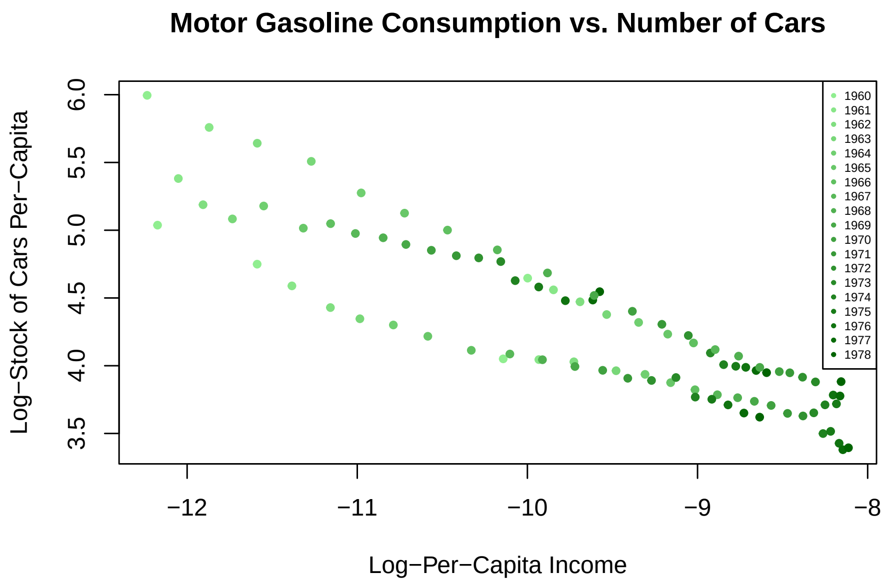
It is observed that heterogeneity also seems to exist across time. The variability tends to decrease for the more recent years, and there is a gradual downward trend.

Visualizing Two-way Heterogeneity:
The three plots below visualize the relationship between the dependent variable and each independent variable. As shown below, `gas` and `income`, and `gas` and `cars` seem to be negatively correlated, and the relationship between `gas` and `price` is not clear based purely on the plot.

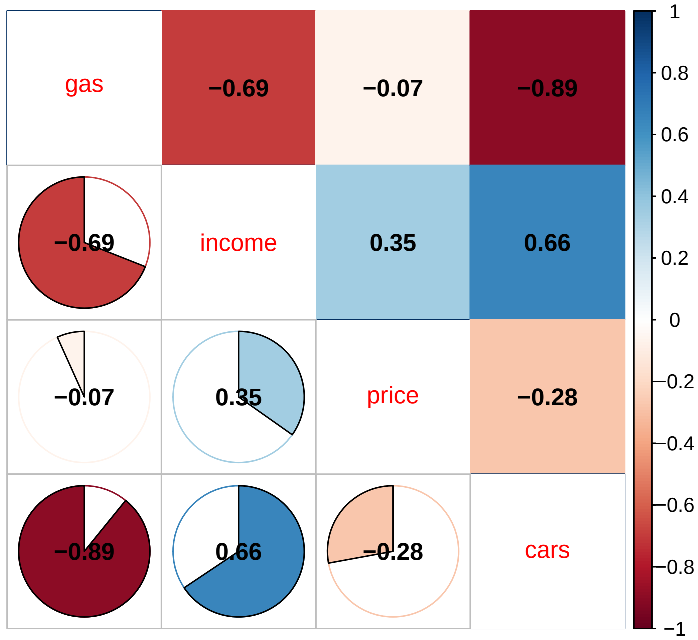
The scatter plots also include color effects where darker colors are more recent years. By visually inspecting the two-way heterogeneity, we can get a big picture that there may exist both individual and time heterogeneity.



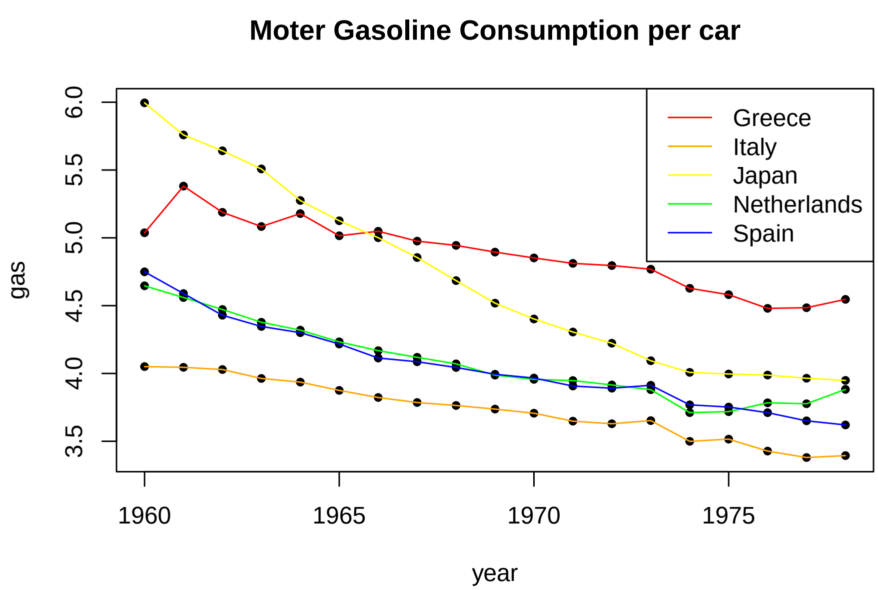




Correlation Matrix:
Based on the correlation matrix, we can see the dependent variable and the independent variables are negatively correlated, which was also expected as we explored the visuals above. Particularly, the `income` and `cars` variables are strongly correlated with the dependent variable.



The Big Picture The plot below visualizes the motor gasoline consumption data for the five countries. We can see that `Japan` has a little steeper slope than other countries, but overall, a slow and steady downward trend is observed, which verifies again that there exists heterogeneity across time.



- ***Pooled Model***

The pooled model is a model that assumes constant marginal effects across individuals. In other words, the model has no provision for individual differences that could potentially result in different coefficients. The mathematical notation is as follows:

$$Gas_{it} = \beta_1 + \beta_2 Income_{it} + \beta_3 Price_{it} + \beta_4 Cars_{it} + \epsilon_{it}$$

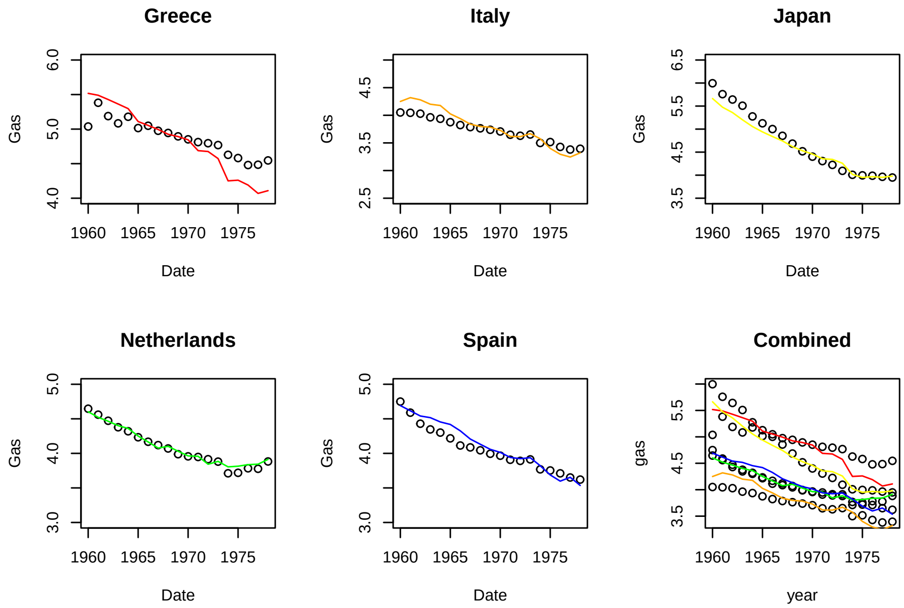
Note that the coefficients do NOT have i or t subscripts, which implies:

1. β does not change over time
2. β does not change across individuals

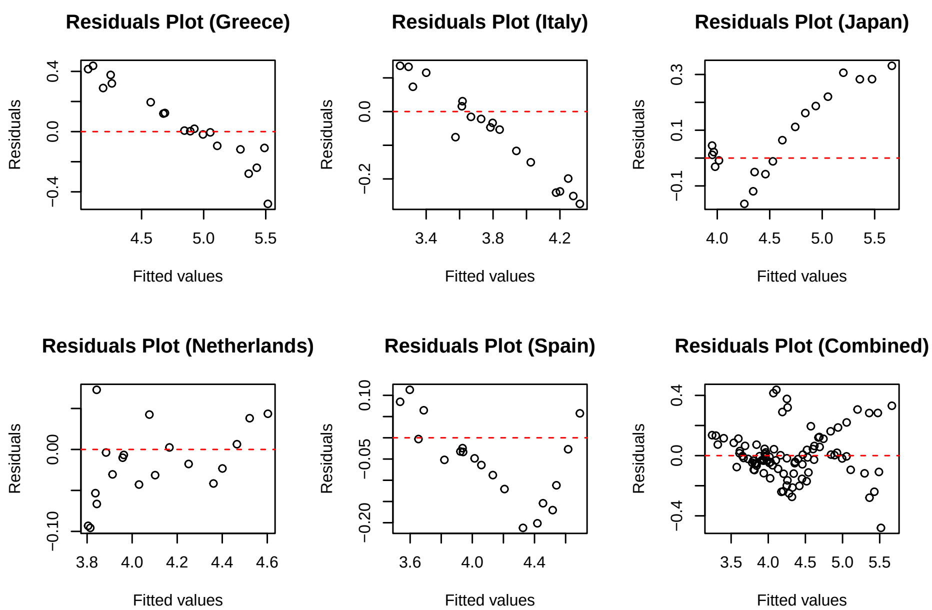
Below are the summary statistics and plots that provide detailed information on our pooled model:

```
## Pooling Model
##
## Call:
## plm(formula = gas ~ income + price + cars, data = pd_gas, model = "pooling")
##
## Balanced Panel: n = 5, T = 19, N = 95
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -0.479934 -0.081904 -0.017717  0.064478  0.437182
##
## Coefficients:
##              Estimate Std. Error  t-value  Pr(>|t|)
## (Intercept)  0.574710   0.260732   2.2042   0.03003 *
## income       0.370765   0.071385   5.1939  1.25e-06 ***
## price       -0.689248   0.059389 -11.6056 < 2.2e-16 ***
## cars        -0.619488   0.027617 -22.4314 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    31.368
## Residual Sum of Squares: 2.3536
## R-Squared:      0.92497
## Adj. R-Squared: 0.92249
## F-statistic: 373.937 on 3 and 91 DF, p-value: < 2.22e-16
```

Visualizing the model fit for each country



Residuals Plot



BP (Breusch-Pagan) test

```
##
##  studentized Breusch-Pagan test
##
## data:  gas_pool
## BP = 23.285, df = 3, p-value = 3.522e-05
```

As also can be anticipated by the model fit, the residuals plot and BP test confirm that there exists heteroskedasticity in the residuals, which means the standard errors output earlier are incorrect. Thus, we correct the standard errors using Cluster-robust standard errors.

Correcting the standard errors with Cluster-robus standard errors.

```
##
## t test of coefficients:
##
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)  0.574710   0.429531   1.3380    0.1842
## income       0.370765   0.088949   4.1683 6.984e-05 ***
## price       -0.689248   0.091085  -7.5671 2.988e-11 ***
## cars        -0.619488   0.053992 -11.4737 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To summarize, we observed that while the fit describes the actual data relatively well for some countries, such as the `Netherlands`, it is also clear that the fit does not apply to all countries as seen in `Greece` or `Italy`. This pooled model will be compared with a fixed effects model in the next section.

- ***Fixed Effects Model***

The fixed effects model relaxes the assumption that was considered in the pooled model, which is that all individuals have the same coefficients. This implies that the coefficients (β) can be different across individuals. The mathematical notation for the fixed effects model is as follows:

$$Gas_{it} = \beta_{1i} + \beta_2 Income_{it} + \beta_3 Price_{it} + \beta_4 Cars_{it} + \epsilon_{it}$$

Note that an i subscipt was added to β_1. It is assumed that ***individual heterogeneity***, all behavioral differences across individuals are captured by the varying intercept.

In this section, two approaches are used to estimate the fixed effects model.

1. The Least Squares Dummy Variable Estimator (LSDV)
2. The Within Estimator

1. LSDV model For LSDV model, a dummy variable is included for each individual:

$$D_{Greece,i} = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{otherwise} \end{cases} \qquad D_{Italy,i} = \begin{cases} 1 & \text{if } i = 2 \\ 0 & \text{otherwise} \end{cases}$$
$$D_{Japan,i} = \begin{cases} 1 & \text{if } i = 3 \\ 0 & \text{otherwise} \end{cases} \quad D_{Netherlands,i} = \begin{cases} 1 & \text{if } i = 4 \\ 0 & \text{otherwise} \end{cases}$$
$$D_{Spain,i} = \begin{cases} 1 & \text{if } i = 5 \\ 0 & \text{otherwise} \end{cases}$$

Thus, the full model would be:

$$\beta_{11} D_{Greece,i} + \beta_{12} D_{Italy,i} + \beta_{13} D_{Japan,i} + \beta_{14} D_{Netherlands,i} + \beta_{15} D_{Spain,i} + \beta_2 Income_{it} + \beta_3 Price_{it} + \beta_4 Cars_{it} + \epsilon_{it}$$

Below are the summary statistics for the Fixed effects model, outputted by the LSDV approach:

Estimating the Fixed Effects model

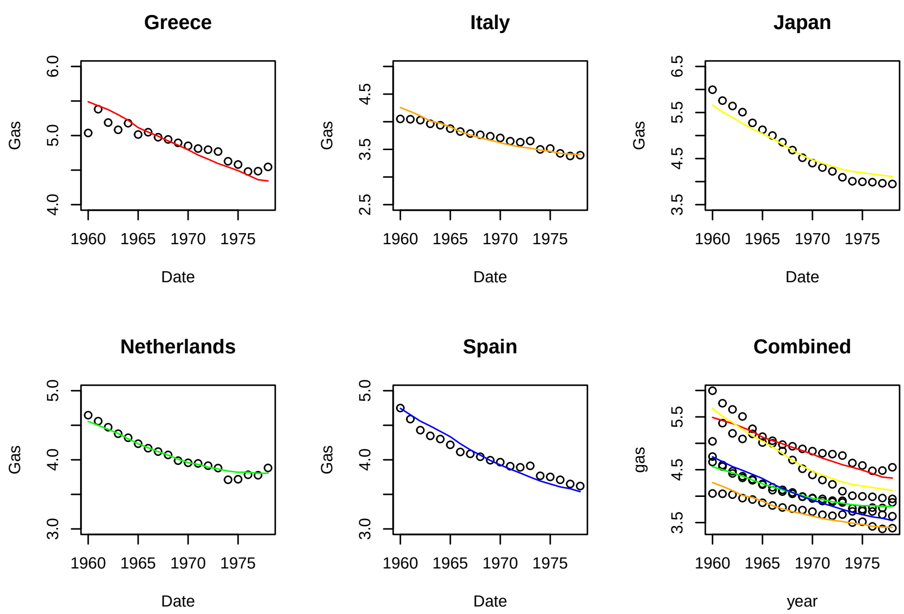
```
##
## Call:
## lm(formula = gas ~ income + price + cars + factor(country) -
##     1, data = pd_gas5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45246 -0.06089  0.00359  0.07347  0.33440
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## income                   -0.15958    0.30274  -0.527 0.599464
## price                    -0.06931    0.10087  -0.687 0.493811
## cars                     -0.37264    0.10599  -3.516 0.000699 ***
## factor(country)Greece    -0.19572    0.88413  -0.221 0.825321
```

```
## factor(country)Italy       -0.58363    0.99947  -0.584 0.560771
## factor(country)Japan       -0.02330    0.85036  -0.027 0.978206
## factor(country)Netherlands -0.17561    0.86217  -0.204 0.839079
## factor(country)Spain       -0.47923    0.71983  -0.666 0.507324
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1235 on 87 degrees of freedom
## Multiple R-squared:  0.9993, Adjusted R-squared:  0.9992
## F-statistic: 1.457e+04 on 8 and 87 DF,  p-value: < 2.2e-16
```

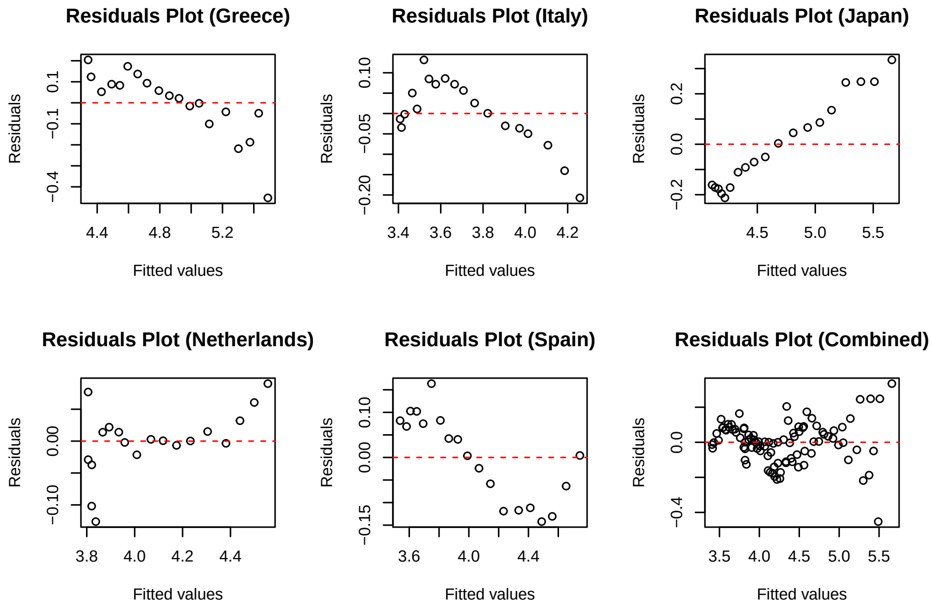
2. Within Approach

```
##      Greece       Italy       Japan Netherlands       Spain
##   -0.195724   -0.583629   -0.023297   -0.175606   -0.479234
```

Visual inspection



Residuals Plot



Hypothesis testing: As seen above, we can see improvement in the fit using the fixed effects model. Now, we verify if the improvement observed is statistically significant. In other words, using diagnostic statistics, we can confirm if the fixed effects are actually needed to model the original panel data.

```
##
##  F test for individual effects
##
## data:  gas ~ income + price + cars
## F = 16.828, df1 = 4, df2 = 87, p-value = 2.979e-10
## alternative hypothesis: significant effects
```

The p-value is very close to zero, which means we reject the null hypothesis that the heterogeneity across individuals is not statistically significant. **Therefore, we can conclude that the fixed effects are necessary to account for individual-specific effects or time-variant characteristics that significantly contribute to the variation in the panel data.**

- ***Random Effects Model***

In the random effects model, it is assumed that the individual differences are **random** as opposed to the fixed effects model, where we consider them **fixed**. One similarity is that the random effects are also captured in the intercept parameter (β_{1i}), which consists of two parts: fixed and random.

$$\beta_{1i} = \bar{\beta}_1 + u_i$$

where

$\bar{\beta}_1$: population mean (fixed)

u_i: random effects (random)

$$Gas_{it} = \bar{\beta}_1 + \beta_2 Income_{it} + \beta_3 Price_{it} + \beta_4 Cars_{it} + v_{it}$$

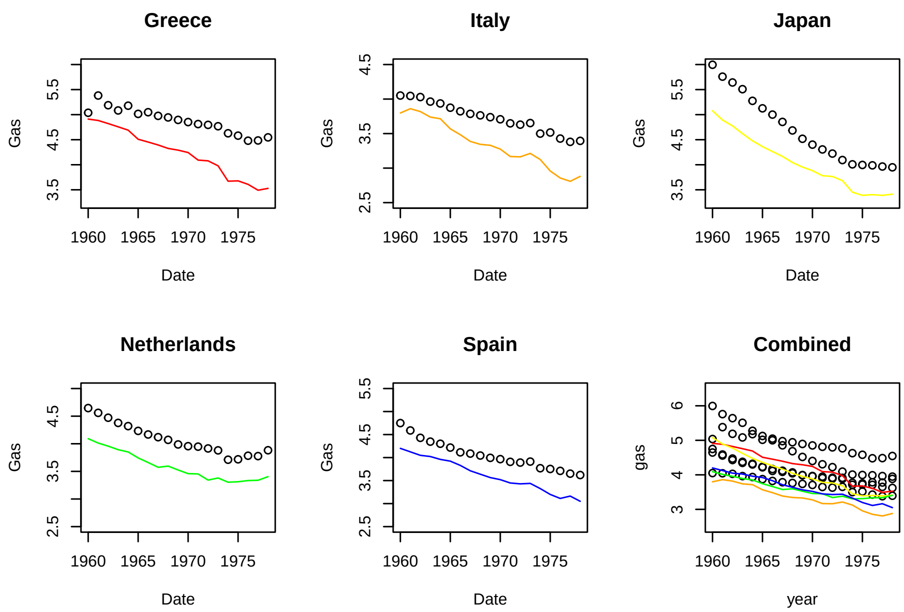
and the error term v_{it} is:

$$v_{it} = u_i + e_{it}$$

Estimating the random effect model

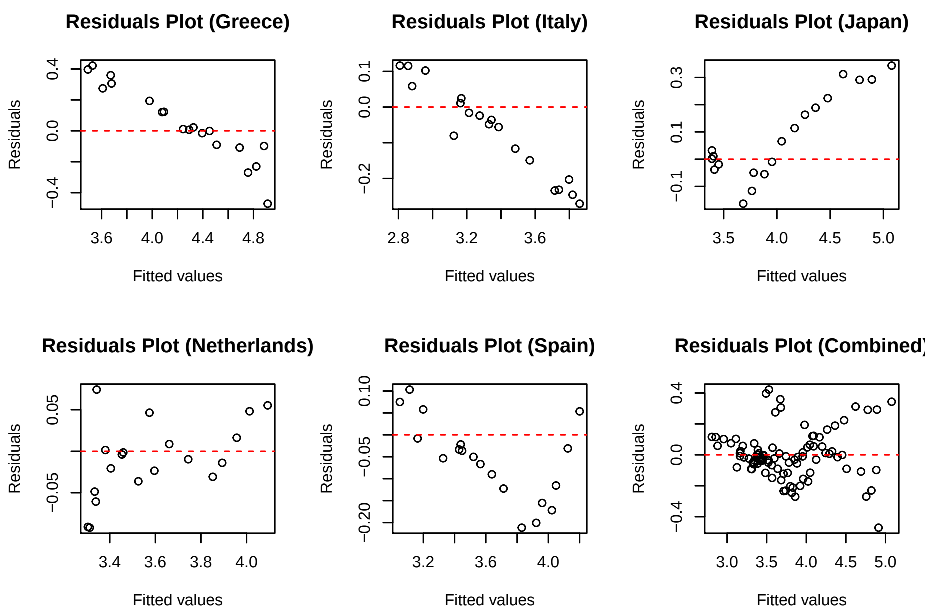
```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = gas ~ income + price + cars, data = pd_gas, model = "random")
##
## Balanced Panel: n = 5, T = 19, N = 95
##
## Effects:
##                    var  std.dev share
## idiosyncratic 0.015252 0.123501 0.985
## individual    0.000239 0.015460 0.015
## theta: 0.1222
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -0.470663 -0.085151 -0.014919  0.058279  0.422010
##
## Coefficients:
##              Estimate Std. Error  z-value  Pr(>|z|)
## (Intercept)  0.571976   0.277746   2.0594   0.03946 *
## income       0.345591   0.078333   4.4118 1.025e-05 ***
## price       -0.652741   0.064369 -10.1405 < 2.2e-16 ***
## cars        -0.603809   0.030125 -20.0432 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    27.326
## Residual Sum of Squares: 2.2691
## R-Squared:      0.91696
## Adj. R-Squared: 0.91422
## Chisq: 1004.87 on 3 DF, p-value: < 2.22e-16
```

Visual Inspection



From the visual inspection, we can see that the model tends to underestimate the true values. Compared to the fixed effects model, the model fit of the random effects model indicates inaccuracy and slight biases.

Residuals Plot



Since we have confirmed earlier that heterogeneity exists across individuals, we decide which model is better by comparing the fixed effects model and the random effects model.

Comparing the `Fixed effects model` and `Random effects model` To decide which model is better, the **Hausman Test** was used, which detects the correlation between the error terms and the regressors (right-hand side of the equation).

```
##
##  Hausman Test
##
## data:  gas ~ income + price + cars
## chisq = 23.22, df = 3, p-value = 3.633e-05
## alternative hypothesis: one model is inconsistent
```

Since the p-value is small, we reject the null hypothesis that the random effects model is preferred to the fixed effects model.

(d) Conclusion

In conclusion, in the first part of this project, we investigated panel data on gasoline consumption in 5 countries: Greece, Italy, Japan, Netherlands, and Spain. We have explored heterogeneity across individuals (countries) and time and concluded that each country's gas consumption is impacted differently by each variable, including `income`, `price`, and `cars`, and the **fixed effects** across individuals are statistically significant. This answers the question that we aimed to answer; ***"Within the same time frame, how do the impacts of per-capita income, gasoline prices, and the number of cars on motor gasoline consumption differ across countries?"***