Econometrics project

The U.S. Gasoline Market

Céline ADOUSSINGANDE

2024-12-05

Filière : Master 1 Econométrie-statistique

Introduction

Since the 1950, per capita gasoline consumption has followed a trajectory marked by technological change, oil shocks and environmental measures. Gasoline consumption per capita in the United States is therefore a key indicator for understanding the dynamics of the gasoline market and consumer behaviors in an ever-changing economic context. This market is influenced by many factors such as the year, the evolution of the population, the price of gasoline, the income of consumers and many other economic factors.

Indeed, the economic analysis of the US gasoline market offers an opportunity to study the impact of individuals and macroeconomic factors on gasoline consumption per capita. It also identifies the effects of linear and nonlinear constraints in econometric models, in order to better understand how consumers respond to price changes.

In this project, we will examine the log of gasoline consumption per capita by the log of explanatory variables through an econometric model based on historical data. We will then test the economic assumptions on price elasticities and finally analyze the impact of restrictions on coefficient estimates.

Defining Variables

Here are below the different variables used in our study. The conversion of these variables into a log will give the same names preceded by "log".

- YEAR = Year, 1953-2004,

- GASEXP = Total U.S. gasoline expenditure,
- GASCONS = Consumption of gasoline per capita
- POP = U.S. total population in thousands
- GASP = Price index for gasoline,
- INCOME = Per capita disposable income,
- PNC = Price index for new cars,
- PUC = Price index for used cars,
- PPT = Price index for public transportation,
- PD = Aggregate price index for consumer durables,
- PN = Aggregate price index for consumer nondurables,
- PS = Aggregate price index for consumer services.

Question 1

1- Compute of the per capita consumption of gasoline

The calculation of the gasoline consumption per capita has been applied to each of the observations in our dataframe. Then the calculation is done with the following code:

```
Gasoline_cons$GASCONS <-
  Gasoline_cons$GASEXP/(Gasoline_cons$GASP * Gasoline_cons$POP)
```

YEAR	GASEXP	POP	GASP	INCOME	PNC	PUC	PPT	PD	PN	PS	GASCONS
1953	7.4	159565	16.668	8883	47.2	26.7	16.8	37.7	29.7	19.4	2.782343e-06
1954	7.8	162391	17.029	8685	46.5	22.7	18.0	36.8	29.7	20.0	2.820613e-06

The table above shows the first two lines of our dataframe.

2- The multiple regression model of ln per capita consumption of gasoline on ln per capita income, the ln price of gasoline, all the ln of other prices and a time trend.

```r
#The multiple regression model with Gasoline_cons_log
model <- lm(logGASCONS ~ logGASP +logINCOME+logPNC +logPUC +
              logPPT +YEAR +logPD +logPN +logPS, data=Gasoline_cons_log)
```

```
Résult of the regression model
=================================================
                         Dependent variable:
                    -----------------------------
                             logGASCONS
-------------------------------------------------
logGASP                        0.061
                              (0.054)

logINCOME                     0.993***
                              (0.250)

logPNC                         -0.155
                              (0.267)

logPUC                       -0.489***
                              (0.085)

logPPT                         0.019
                              (0.136)

YEAR                          0.038***
                              (0.008)

logPD                         1.732***
                              (0.260)

logPN                        -0.730**
                              (0.265)

logPS                         -0.868*
                              (0.353)

Constant                    -95.224***
                             (13.119)
```

```
-------------------------------------------------
Observations                   52
R2                           0.987
Adjusted R2                  0.984
Residual Std. Error      0.030 (df = 42)
F Statistic          351.326*** (df = 9; 42)
=================================================
Note:               *p<0.05; **p<0.01; ***p<0.001
```

The signs of the estimates and explaination

Based on the regression result of the model, the sign of the estimates does not all match our expectations.

So, the estimated coefficients for variables such as logGASP, logINCOME, logPPT, YEAR, and logPD are positive. This indicates that these variables positively influence per capita gasoline consumption, implying a positive correlation between them and the dependent variable. However, the sign of the estimated coefficient for the variable logGASP does not align with expectations because an increase in the price of petrol logGASP must lead to a reduction in gasoline consumption per capita. But we remark that the coeficient of logGASP and of logPPT are not significant.

On the other hand, the estimated coefficients for variables such as logPNC, logPUC, logPN, and logPS are preceded by a negative sign. This indicates that these variables negatively influence per capita gasoline consumption, implying a negative correlation between them and the dependent variable. This aligns with expectations.

The R squared being equal to 0.987 this means that 98.7% of the log of gasoline consumption per capita is explained by the log of explanatory variables.

3- The change in the gasoline market that occurred in 1973

a- The difference of change of the ln of per capita gasoline consumption using the average values in the years 1953-1973 and 1974-2004

Let **mean_GASCONS_1** represent the logarithm of the average per capita gasoline consumption during the first period (1953–1973), and let **mean_GASCONS_2** represent the logarithm of the average per capita gasoline consumption during the second period (1974–2004). So, we have:

- Calculates averages by period

```r
# The mean of logGASCONS for each period (periode 1: 1953-1973
#and periode 2: 1974-2004)
mean_GASCONS_1 <- mean(GASCONS_1$logGASCONS, na.rm = TRUE)
mean_GASCONS_1
```

```
[1] -12.48074
```

```r
mean_GASCONS_2 <- mean(GASCONS_2$logGASCONS, na.rm = TRUE)
mean_GASCONS_2
```

```
[1] -12.08536
```

- The difference of averages

```r
# Compute the difference
dif_mean_GASCONS <- mean_GASCONS_2 - mean_GASCONS_1
dif_mean_GASCONS
```

```
[1] 0.3953766
```

The average values of the ln of per capita gasoline consumption in the years 1953-1973 (**mean__GASCONS__1**) is **-12.48074** and in the years 1974-2004 (**mean__GASCONS__2**) is **-12.08536**.

So, the difference of the ln of per capita gasoline consumption between the two periods is **0.3953766.**

b- The difference of change of the ln of per capita gasoline consumption using a Oaxaca-Blinder decomposition

- The regression model of the two periods

```r
#Estimate model for the two periodes
#model 1 (1953-1973)
model1 <- lm(logGASCONS ~ logGASP +logINCOME+logPNC +logPUC +
               logPPT +YEAR +logPD +logPN +logPS, data=GASCONS_1)

#model 2 (1974-2004)
model2 <- lm(logGASCONS ~ logGASP +logINCOME+logPNC +logPUC +
               logPPT +YEAR +logPD +logPN +logPS, data=GASCONS_2)
```

- Oaxaca-Blinder Decomposition

This method consists of calculating the difference in the expected value of log_GASCONS between the two periods in two components, namely: the explained component and the unexplained component.

- **Explained Component**

```
# Explained Component
explained_component <- sum((means_predictors_period_2 -
                            means_predictors_period_1) * coef_period_1)
explained_component
```

```
[1] 0.8980851
```

- **Unexplained Component**

```
# Unexplained Component
unexplained_component <- sum(means_predictors_period_2 *
                             (coef_period_2 - coef_period_1))
unexplained_component
```

```
[1] -0.5027085
```

- **Total difference of the log_consumption per capita in the two periodes**

```
#Total difference of the logGASCONS in the two periodes
total_difference <- explained_component + unexplained_component
total_difference
```

```
[1] 0.3953766
```

The value of the explained component is **0.8980851** and the value of the unexplained component is **-0.5027085**. So, the change value is **0.3953766**.

We remark that we have the same result with using the average method.

c- A confidence interval for the part of the change that can be attributed to structural change in the market (change in the regression coefficients)

```
# Confidence interval
critical_value <- 1.96  # For 95% confidence level
lower_value <- unexplained_component - critical_value * sd_error
lower_value
```

```
          [,1]
[1,] -0.984593
```

```
high_value <- unexplained_component + critical_value * sd_error
high_value
```

```
            [,1]
[1,] -0.02082403
```

A confidence interval for the part of the change that can be attributed to structural change in the market is:

Ic (95%) = [-0.984593 ; -0.02082403]

Question 2

1- A test of the hypothesis that the three aggregate price indices are not significant determinants of the demand for gasoline.

- **model**:

logGASCONS = alpha + betha_GASP * logGASP + betha_INCOME * logINCOME + gamma_PNC * logPNC + gamma_PUC * logPUC + gamma_PPT * logPPT + teltha * YEAR + sigma_PD * logPD + sigma_PN * logPN + sigma_PS * logPS

H0: sigma_PD = sigma_PN = sigma_PS = 0

H1: sigma_PD, sigma_PN, sigma_PS =! 0

```
#Unrestricted model
model_unrestricted <- lm(logGASCONS ~ logGASP + logINCOME +
                           logPNC + logPUC + logPPT + YEAR + logPD + logPN +
                           logPS, data=Gasoline_cons_log)

#Restricted model:
model_restricted <- lm(logGASCONS ~ logGASP + logINCOME + logPNC +
                         logPUC + logPPT + YEAR, data=Gasoline_cons_log)
```

Results of the regression model

	Dependent variable:	
	logGASCONS	
	model (1)	model (2)
logGASP	0.061	-0.054
	(0.054)	(0.043)
logINCOME	0.993***	1.649***
	(0.250)	(0.203)
logPNC	-0.155	-0.032
	(0.267)	(0.206)
logPUC	-0.489***	-0.074
	(0.085)	(0.105)
logPPT	0.019	-0.062
	(0.136)	(0.123)
YEAR	0.038***	-0.013*
	(0.008)	(0.005)
logPD	1.732***	
	(0.260)	
logPN	-0.730**	
	(0.265)	
logPS	-0.868*	
	(0.353)	
Constant	-95.224***	-1.821
	(13.119)	(8.865)
Observations	52	52
R2	0.987	0.965
Adjusted R2	0.984	0.960

```
Residual Std. Error     0.030 (df = 42)         0.047 (df = 45)
F Statistic         351.326*** (df = 9; 42) 207.553*** (df = 6; 45)
===================================================================
Note:                                    *p<0.05; **p<0.01; ***p<0.001
```

In the restricted model, the variables such as logGASP , logPNC, logPUC and logPPT are non significant. The R squared being equal to 0.965 this means that 96.5% of the log of gasoline consumption per capita is explained by the log of explanatory variables.

- **F-statistic**

```r
#The F_test
F_test<-(((rsqt_unrestricted-rsqt_restricted)/q)/((1-rsqt_unrestricted)/df))
F_test
```

```
[1] 23.24582
```

```r
#The Fcritical value
F_cv<-qf(0.95, df1=3, df2=42)
F_cv
```

```
[1] 2.827049
```

F_test > F_cv => 23.24582 > 2.827049

so we reject H0. Then, the price indices are significant determinants of the log of the consumption per capita.

2- Describe in detail the test of the validity of the restriction.

- The linear restriction

H0 : gamma_PNC = gamma_PUC

H1 : gamma_PNC =! gamma_PUC

logGASCONS = alpha + betha_GASP * logGASP + betha_INCOME * logINCOME +gamma_PNC * logPNC + gamma_PNC * logPUC + gamma_PPT * logPPT + teltha * YEAR + sigma_PD * logPD + sigma_PN * logPN + sigma_PS * logPS

logGASCONS = alpha + betha_GASP * logGASP + betha_INCOME * logINCOME +gamma_PNC * (logPNC + logPUC) + gamma_PPT * logPPT + teltha * YEAR + sigma_PD * logPD + sigma_PN * logPN + sigma_PS * logPS

```r
Gasoline_cons_log$logPNC_PUC <- (Gasoline_cons_log$logPNC +
                                  Gasoline_cons_log$logPUC)

restricted_model_linear <- lm(logGASCONS ~ logGASP + logINCOME +
                               logPNC_PUC + logPPT + YEAR + logPD + logPN
                             + logPS, data=Gasoline_cons_log)
```

- The Nonlinear Restriction

H0: gamma_PNC * sigma_PS = gamma_PPT * sigma_PD

H1: gamma_PNC * sigma_PS =! gamma_PPT * sigma_PD

```r
#The value of gamma_PNC with the restriction
 gamma_PNC <-  gamma_PPT*sigma_PD / sigma_PS
```

gamma_PNC = -0.03845262

```r
#The non linear model restriction
 model_restricted_nonlinear <- lm(logGASCONS ~ logGASP + logINCOME +
                                   offset(gamma_PNC * logPNC) + logPUC +
                                   logPPT + YEAR + logPD + logPN + logPS,
                                 data=Gasoline_cons_log)
```

```
Results of the regression model
===========================================================================================
                                              Dependent variable:
                    -----------------------------------------------------------------------
                                                  logGASCONS
                              model                     model                     model
                               (1)                       (2)                       (3)
-------------------------------------------------------------------------------------------
logGASP                       0.061                     0.032                     0.071
                             (0.054)                   (0.048)                   (0.048)

logINCOME                   0.993***                  0.769***                  1.088***
                             (0.250)                   (0.160)                   (0.123)

logPNC                       -0.155
                             (0.267)
```

logPUC	-0.489***		-0.492***
	(0.085)		(0.084)
logPNC_PUC		-0.451***	
		(0.079)	
logPPT	0.019	-0.040	0.044
	(0.136)	(0.127)	(0.123)
YEAR	0.038***	0.040***	0.037***
	(0.008)	(0.007)	(0.007)
logPD	1.732***	1.838***	1.669***
	(0.260)	(0.244)	(0.213)
logPN	-0.730**	-0.754**	-0.707**
	(0.265)	(0.265)	(0.257)
logPS	-0.868*	-0.658*	-0.954**
	(0.353)	(0.304)	(0.290)
Constant	-95.224***	-97.264***	-93.695***
	(13.119)	(13.054)	(12.521)
Observations	52	52	52
R2	0.987	0.986	0.988
Adjusted R2	0.984	0.984	0.986
Residual Std. Error	0.030 (df = 42)	0.030 (df = 43)	0.030 (df = 43)
F Statistic	351.326*** (df = 9; 42)	391.855*** (df = 8; 43)	452.777*** (df = 8; 43)
Note:			*p<0.05; **p<0.01; ***p<0.001

Model (1) represents the model without any restriction, model (2) represents the model with linear restriction, and model (3) represents the model with nonlinear restriction. Indeed, we note that the variables logGASP and logPPT are not significant in any of the three models. In addition, the R-squared (the rate of explanation of the log of consumption per capita) of model (2) is 0.986 or 98.6% and 0.988 or 98.8% for model (3).

3- Test of the two restrictions

- The Linear Restriction

- **F-statistic**

```
#The F_test
 F_test1<-(((rsqt_unrestricted-rsqt_restricted_linear)/q)/
              ((1-rsqt_unrestricted)/df))
 F_test1
```

```
[1] 1.353153
```

```
#The F critical value
 F_cv_linear<-qf(0.95, df1=1, df2=42)
 F_cv_linear
```

```
[1] 4.072654
```

F_test < F_cv => 1.353153 < 4.072654

so we fail to reject H0. Then, gamma_PNC = gamma_PUC. The linear model is valid.

- The non linear restriction

- **Log-likelihood**

```
# Likelihood Test Statistics
 LR <- -2 * (LL_restricted - LL_unrestricted)
 LR
```

```
[1] 0.2342942
```

```
# p-value
 p_value <- pchisq(LR, df = k, lower.tail = FALSE)
 p_value
```

```
[1] 0.6283579
```

p__value > 0.05 so we fail to reject H0. The non linear model is valid.

- **F__statistic**

```
#The F_test
F_test2<-(((rsqt_unrestricted-rsqt_restricted_nonlinear)/q)/((1-rsqt_unrestricted)/df))
F_test2
```

```
[1] -4.411717
```

```
#The F critical value
F_cv_nonlinear<-qf(0.95, df1=1, df2=42)
F_cv_nonlinear
```

```
[1] 4.072654
```

We fail to rejet H0. The F-test also confirm that the non linear model is valid.

Conclusion

The econometric analysis of the US gasoline market has deepened our understanding of the factors that influence the log of gasoline consumption per capita. Thus, by modeling the relationships between the log of economic variables and the log of gasoline consumption per capita (logGASCONS), we noticed that the log gasoline price (logGASP), the log of revenue per capita (logINCOME), the year (YEAR), the log of the public transport price index (logPPT) and the log of the aggregate price index for sustainable consumption (logPD) positively influence gasoline consumption unlike other factors such as the log of the index price for news cars (logPNC), the log of the price index for used cars (logPUC), the log of the aggregate price index for consumers non-durable (logPN) and the log of the aggregate price index for consumer services (logPS) which negatively influence the log of gasoline consumption per capita. But the sign of the log coefficient of the price of gasoline does not correspond to our expectations and this variable is not significant according to our analyses.

However using the averaging method or the Oaxaca-Blinder decomposition method, the difference in the log consumption of gasoline per capita between 1953-1973 and 1974-2004 is 0.3953766, or 39.64% with a confidence interval of "[-0.984593; -0.02082403]" at 95%.

In addition, the tests carried out made it possible to reject the hypothesis that the coefficients of the log of "aggregate price index" variables are zero and to validate the hypothesis that the log model of gasoline consumption is linear if the coefficient of the log of index price for news cars (logPNC) is equal to the coefficient of the log of price index for used cars (logPUC) and that the model is non-linear if the coefficient of the log of price index for news cars (logPNC) is equal to the coefficient of the log of public transport index price (logPPT) multiplied by the coefficient of the log of the aggregate index price for sustainable consumption (logPD) divided by the coefficient of the log of the aggregate price index for consumer services (logPS).