



**UNIVERSITA' DEGLI STUDI DI PADOVA**

**DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI  
"M.FANNO"**

**CORSO DI LAUREA IN ECONOMIA E MANAGEMENT**

**PROVA FINALE**

**"IL MERCATO DELL'AUTO: UN'ANALISI LONGITUDINALE DEI  
PRINCIPALI PAESI EUROPEI"**

**RELATORE:**

**CH.MO PROF. NUNZIO CAPPuccio**

**LAUREANDO: ALESSIO BILATO**

**MATRICOLA N. 1065173**

**ANNO ACCADEMICO 2015 – 2016**







# INDICE

<b>Introduzione</b> .....	<b>7</b>
<b>Analisi dei Dati Panel</b> .....	<b>8</b>
Pooled Ordinary Least Squares Method (Pooled OLS) .....	9
Modello ad Effetti Fissi (FE) .....	9
Modello ad Effetti Casuali (RE).....	11
La scelta dello stimatore.....	13
<b>Metodologia operativa</b> .....	<b>15</b>
Interpretazione dell'output .....	15
Test sui modelli .....	16
Costruzione database.....	17
Variabili.....	18
Preparazione dataset.....	32
<b>Stima dei modelli</b> .....	<b>40</b>
Modelli 1-3 .....	40
Sintesi dei risultati.....	44
Revisione dataset.....	45
Modello 4 .....	45
<b>Risultati dell'analisi</b> .....	<b>48</b>
<b>Appendice</b> .....	<b>51</b>
<b>Riferimenti bibliografici</b> .....	<b>55</b>



# INTRODUZIONE

L'analisi riportata in questo elaborato ha l'obiettivo di studiare se esiste un legame tra il numero di automobili vendute in un dato paese ed alcune variabili macroeconomiche come:

- crescita del PIL,
- tasso di disoccupazione,
- tasso di inflazione.

Inoltre, si verificano eventuali legami con altre variabili quali:

- prezzo medio di vendita di un'automobile,
- prezzo della benzina pagato dal consumatore,
- fiducia dei consumatori.

Si vuole inoltre considerare l'eventuale effetto della crisi finanziaria del 2007-2008 sulle vendite di automobili.

Per effettuare tale studio si effettua un'analisi econometrica di un dataset composto da dati panel relativi ai 14 mercati europei più grandi in termini di numero di veicoli immatricolati. Le osservazioni riguardano i periodi dal 2001 al 2014. La raccolta dei dati, la loro pulizia e la preparazione del dataset sono la fase più importante dell'intero studio. Infatti, da dataset solidi si possono costruire modelli validi. Se il dataset difetta di coerenza o di precisione l'intero studio ha delle fondamenta vacillanti. Per questo motivo, i database scelti come fonte dei dati inseriti nel dataset di questa analisi sono stati selezionati tra quelli più affidabili forniti da enti di grandi dimensioni quali OCSE, ICCT, Worldbank. La totalità dei dati utilizzati è reperibile gratuitamente online al momento dello sviluppo dell'analisi. Nelle sezioni seguenti di questo elaborato si affronteranno anche dal punto di vista teorico i differenti modelli di stima dei parametri per dati panel, che verranno poi utilizzati nell'analisi del dataset costruito ad hoc.

# ANALISI DEI DATI PANEL

Generalmente, la maggior parte delle applicazioni di regressioni multiple si effettua su datasets con dati puramente *cross-sectional* o puramente di serie temporali. Nonostante questi due casi siano molto frequenti, un utilizzo congiunto delle dimensioni *cross-sectional* e *time series* è sempre più utilizzato nella ricerca empirica. I datasets di questo tipo si possono costituire in due modi. Da un lato, i *pooled data* sono i dati ottenuti attraverso campionamento casuale effettuato da una popolazione molto ampia studiata in differenti periodi temporali. Essi sono dati riferiti a  $kT$  differenti unità statistiche ( $k$  unità statistiche in ogni periodo  $t$ , che non restano le stesse al variare di  $t$ ). Aspetto chiave delle successive analisi è l'indipendenza tra le osservazioni in quanto essa esclude eventuale correlazione nei termini d'errore delle diverse osservazioni.

Sono definite, invece, dati panel (o *panel data*) le osservazioni ottenute campionando i medesimi  $k$  individui, famiglie, gruppi, aziende, regioni, ecc in differenti  $T$  periodi temporali.

La scelta di effettuare l'analisi tramite dati panel permette di lavorare con maggiori informazioni rispetto a dataset più semplici, includendo quindi più variabilità riducendo di conseguenza la collinearità tra le variabili. Di conseguenza i parametri sono stimati più efficientemente e precisamente. Inoltre, essi permettono di studiare le dinamiche di variazione dei dati, mentre, ad esempio, dati puramente *cross-sectional* permettono solamente di studiare un'immagine istantanea della situazione in un determinato momento. Normalmente i dati panel vengono strutturati e utilizzati in modo da verificare se l'ipotesi di esistenza di una funzione lineare tra  $y_{it}$  (variabile dipendente) e un numero  $k$  di variabili indipendenti è verificata. Il modello generale per l'unità statistica  $i$  al periodo  $t$  è

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \dots + \beta_k x_{itk} + u_{it}$$

$$y_{it} = \beta' X_{it} + u_{it}$$

Con  $i = 1 \rightarrow N$  e  $t = 1 \rightarrow T$ .

Si può scomporre il termine d'errore composito  $u_{it}$  in due parti:  $\alpha_i$  e  $\varepsilon_{it}$ .

$\alpha_i$  è il termine che indica una costante caratteristica di ogni unità statistica  $i$ , che tiene conto dell'effetto sulla variabile dipendente di un insieme di variabili non osservate costanti nel tempo.

Generalmente,  $\alpha_i$  viene anche chiamato *unobserved factors* o *effetto fisso* o *eterogeneità non osservata*. Il fatto stesso che il pedice sia scritto come  $i$  e non  $it$  indica come l'effetto fisso sia time invariant. (Wooldridge, 2006)

Il termine d'errore  $\varepsilon_{it}$  viene chiamato *errore idiosincratice* o *time-varying error*, perché rappresenta l'errore casuale che influenza  $y_{it}$ . L'analogia della natura di questo termine di errore con quella del termine di errore di una semplice regressione su dati storici è molto forte.

Per l'analisi econometrica dei dati panel non è possibile assumere che le osservazioni siano indipendentemente distribuite (iid) nel tempo. Per questo motivo sono stati sviluppati diversi particolari metodi di analisi dei dati panel, quali il modello pooled OLS, il modello ad Effetti Fissi (Fixed Effects, FE) e il modello ad Effetti Casuali (Random Effects, RE).



## POOLED ORDINARY LEAST SQUARES (POOLED OLS)

Il più semplice modello di stima per dati panel è il modello pooled OLS, che sfrutta lo stimatore dei minimi quadrati ordinari.

Nella maggior parte dei casi è improbabile che esso sia adeguato, ma offre una *guideline* per un confronto con modelli più complessi.

La definizione del modello pooled OLS può essere scritta come

$$(1.1) \quad y_{it} = x_{it}\beta + u_{it}$$

laddove  $y_{it}$  è l'osservazione della variabile dipendente per l'unità cross-sectional  $i$  al periodo  $t$ ,  $x_{it}$  è un vettore  $1 \times k$  di variabili indipendenti osservate per l'unità  $i$  al periodo  $t$ ,  $\beta$  è un vettore  $k \times 1$  di parametri, e  $u_{it}$  è un errore o termine di disturbo specifico per l'unità  $i$  al periodo  $t$ .

L'utilizzo del modello pooled OLS presenta uno svantaggio rilevante, ovvero il fatto che, per utilizzare il modello ed ottenere stime consistenti di  $\beta$ , è necessario assumere che l'*unobserved effect*  $\alpha_i$  sia incorrelato con  $x_{it}$ .

$$E(x_{it}\alpha_i) = 0$$

Se si stima il modello pooled OLS, esso è inconsistente e distorto se esiste la suddetta correlazione. La distorsione dovuta a questo problema viene chiamata anche *heterogeneity bias*, ma è semplicemente una distorsione dovuta al fatto che non è stata inclusa alcuna variabile *time-constant*. (Wooldridge, 2006)

La stima tramite pooled OLS necessita della validità di cinque assunzioni (Greene, 2008; Kennedy, 2008):

- **Ass. OLS1:** linearità. La variabile dipendente è formulate come funzione lineare di variabili indipendenti e del termine di disturbo;
- **Ass. OLS2:** esogeneità. Il valore atteso dei disturbi è nullo o gli errori non sono correlate con alcun regressore;
- **Ass. OLS3:** omoschedasticità e non autocorrelazione. I disturbi hanno la stessa varianza e non sono correlate gli uni agli altri;
- **Ass. OLS4:** le osservazioni delle variabili indipendenti non sono stocastiche ma fisse in campioni ripetuti senza errori di misurazione;
- **Ass. OLS5:** rango pieno. Non c'è multicollinearità, ovvero non c'è alcuna relazione lineare perfetta tra le variabili indipendenti.

Se si stima un modello su dati panel usando lo stimatore OLS è possibile effettuare il test di Hausman. Specificando un modello senza alcuna variabile dummy che rappresenti unità cross-sectional, è possibile comparare gli stimatori del modello pooled OLS con le principali alternative, il modello ad effetti fissi (Fixed Effects Model) e il modello ad effetti casuali (Random Effects Model). Tali alternative sono spiegate nella prossima sezione.

## IL MODELLO AD EFFETTI FISSI (FIXED EFFECTS MODEL)

È possibile scomporre il termine  $u_{it}$  (indicante tutte le variabili non osservabili), nelle sue due componenti,  $\varepsilon_{it}$  e  $\alpha_i$ .

$$(1.2) \quad u_{it} = \alpha_i + \varepsilon_{it}$$

fornendo quindi il modello

$$(1.3) \quad y_{it} = x_{it}\beta + \alpha_i + \varepsilon_{it}$$

$u_{it}$  è quindi scomposto nelle due parti  $\alpha_i$ , la parte dell'errore dipendente dall'unità osservata e indipendente rispetto al tempo, comprendente l'effetto di tutte le variabili non osservabili, e  $\varepsilon_{it}$ , la parte dell'errore peculiare dell'osservazione.

Il modello ad effetti fissi si concentra sull'eliminazione dell'intercetta  $\alpha_i$ , costante nel tempo, in quanto essa contiene valori non osservabili e che quindi verrebbero considerati parte integrante dell'errore del modello. Tali valori potrebbero essere correlati con le variabili esplicative  $x_{it}$ , restituendo una stima distorta. L'eliminazione del termine  $\alpha_i$  si basa sul procedimento di *data-demeaning*. Esso consiste nella sottrazione della media di gruppo da ognuna delle variabili e nella stima del modello senza intercetta attraverso lo stimatore pooled OLS. (Wooldridge, 2006)

La variabile dipendente può essere scritta come

$$(1.4) \quad \bar{y}_{it} = y_{it} - \bar{y}_i$$

E la media di gruppo  $\bar{y}_i$  è definita come

$$(1.5) \quad \bar{y}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} y_{it}$$

dove  $T_i$  è il numero di osservazioni per l'unità  $i$ . Una formulazione analoga si applica anche alle variabili indipendenti.

Si consideri un modello semplificato, per chiarezza, ad un modello con una sola variabile esplicativa.

$$(1.6) \quad y_{it} = \beta_1 x_{it} + \alpha_i + \varepsilon_{it}$$

Si vada a riscrivere l'equazione utilizzando le medie delle variabili nel tempo.

$$(1.7) \quad \bar{y}_i = \beta_1 \bar{x}_i + \alpha_i + \bar{\varepsilon}_i$$

Dal momento che  $\alpha_i$  è costante nel tempo esso appare sia nella 1.6 che nella 1.7. Se si sottrae la 1.7 alla 1.6 per ogni  $t$ , si ottiene

$$(1.8) \quad y_{it} - \bar{y}_i = \beta_1(x_{it} - \bar{x}_i) + \varepsilon_{it} - \bar{\varepsilon}_i$$

O anche

$$(1.9) \quad \dot{y}_{it} = \beta_1 \dot{x}_{it} + \dot{\varepsilon}_{it}$$

In questo modo l'effetto fisso  $\alpha_i$  scompare dall'equazione, ed è possibile procedere alla stima tramite pooled OLS sulle variabili *time-demeaned* (Wooldridge, 2006). Lo stimatore pooled OLS basato sulle variabili *time-demeaned* si chiama stimatore *within*,  $\hat{\beta}$ , ed è quello stimatore che tiene conto degli effetti individuali, ma li elimina dal modello utilizzando per ciascun paese (o unità) le informazioni derivanti dalle variazioni nel tempo, attraverso il procedimento di *data de-meaning*. Riscrivere l'equazione 1.9 per un modello con più variabili esplicative è procedimento ora semplice, ottenendo la seguente equazione con variabili *time-demeaned* per ogni  $i$ :

$$(1.10) \quad \dot{y}_{it} = \beta_1 \dot{x}_{it1} + \beta_2 \dot{x}_{it2} + \dots + \beta_k \dot{x}_{itk} + \dot{\varepsilon}_{it}$$

La procedura di *time-demeaning* non è tuttavia implementabile direttamente senza dover tener conto di alcuni dettagli. In particolare, si perde un grado di libertà per lo stimatore *within* a causa della procedura.

Le assunzioni che devono essere valide nel caso di utilizzo dello stimatore *within* sono:

- **Ass. F1:** Per ogni  $i$ , il modello è  $y_{it} = \beta_1 x_{it1} + \dots + \beta_k x_{itk} + \alpha_i + \varepsilon_{it}$  con  $t = 1, \dots, T$  dove  $\beta_k$  sono i parametri da stimare e  $\alpha_i$  è l'effetto fisso.
- **Ass. F2:** si utilizza un campione casuale dalla cross-section.
- **Ass. F3:** ogni variabile esplicativa cambia nel tempo (almeno per qualche  $i$ ) e non esiste alcuna relazione lineare perfetta tra le variabili esplicative.
- **Ass. F4:** per ogni  $t$ ,  $E(\varepsilon_{it} | \mathbf{X}_i, \alpha_i) = 0$
- **Ass. F5:**  $\text{Var}(\varepsilon_{it} | \mathbf{X}_i, \alpha_i) = \text{Var}(\varepsilon_{it}) = \sigma_\varepsilon^2$  per ogni  $t$
- **Ass. F6:** per ogni  $t \neq s$ , gli errori idiosincratichi sono incorrelati.  $\text{Cov}(\varepsilon_{it}, \varepsilon_{is} | \mathbf{X}_i, \alpha_i) = 0$
- **Ass. F7:** gli errori  $\varepsilon_{it}$  sono indipendenti e identicamente distribuiti come una normale  $(0, \sigma_\varepsilon^2)$

In alcuni casi può essere d'interesse conoscere le stime delle intercette  $\hat{\alpha}_i$ . Date le stime dei parametri,  $\hat{\beta}$ , ottenute usando i dati *de-meaned*, è possibile trovare le stime dei vari  $\hat{\alpha}_i$  in questo modo:

$$(1.11) \quad \hat{\alpha}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} (y_{it} - x_{it} \hat{\beta})$$

## IL MODELLO AD EFFETTI CASUALI (RANDOM EFFECTS MODEL)

Anche il modello ad effetti casuali, come il modello ad effetti fissi, scompone il termine  $u_{it}$  nelle sue componenti  $\varepsilon_{it}$  e  $\alpha_i$ . Tuttavia, nel modello si introduce esplicitamente un'intercetta:

$$(1.12) \quad y_{it} = \beta_0 + \beta_1 x_{it1} + \dots + \beta_k x_{itk} + \alpha_i + \varepsilon_{it}$$

In modo da poter effettuare l'assunzione che  $E(\alpha_i) = 0$ . Nel modello ad effetti fissi si cerca di eliminare il termine  $\alpha_i$ , in quanto si suppone esso sia correlato con una o più delle variabili esplicative. Supponendo però che  $\alpha_i$  sia incorrelato ad ogni variabile esplicativa in tutti i periodi  $t$ , ovvero che  $\text{Cov}(x_{itj}, \alpha_i) = 0$  per ogni  $t$  e per ogni  $k$ , ogni trasformazione che rimuova il termine  $\alpha_i$  porterebbe a stimatori inefficienti.

Nell'ambito del modello ad effetti casuali, diversamente dal modello a effetti fissi, gli  $\alpha_i$  non sono trattati come parametri fissi, ma come realizzazioni di una variabile aleatoria (da qui la denominazione effetti casuali), non correlati ai regressori. In tal modo, questi effetti si possono trattare nel modello come se fossero parte del termine d'errore. Si effettua quindi una trasformazione dei dati che produce un dataset con errori non autocorrelati. I dati trasformati soddisfano le assunzioni del teorema di Gauss-Markov, e quindi le stime finali sono efficienti.

Il teorema di Gauss-Markov, secondo il quale l'OLS è lo stimatore lineare corretto più efficiente, dipende dall'assunzione che il termine d'errore sia indipendente e identicamente distribuito (iid). Nel

contesto dei dati panel, l'assunzione iid implica che  $E(u_{it}^2)$  in relazione all'equazione 1.1 sia uguale ad una costante,  $\sigma_u^2$ , per tutti gli  $i$  e  $t$ , mentre la covarianza  $E(u_{is}u_{it})$  sia nulla per  $s \neq t$  e la covarianza  $E(u_{jt}u_{it})$  sia nulla per  $j \neq i$ . Dal momento che tali assunzioni sono difficilmente verificate nel caso dei dati panel, OLS non sarebbe lo stimatore migliore in tal caso. Il problema della correlazione seriale può essere però risolto usando lo stimatore GLS, tenendo in considerazione la struttura della covarianza del termine d'errore.

Si considerino osservazioni su un'unità  $i$  in due differenti periodi  $s$  e  $t$ . Dalle ipotesi sopracitate si può definire che  $\text{Var}(u_{is}) = \text{Var}(u_{it}) = \sigma_\alpha^2 + \sigma_\varepsilon^2$  mentre la covarianza tra  $u_{is}$  e  $u_{it}$  è data da  $E(u_{is}u_{it}) = \sigma_\alpha^2$ . In notazione matriciale, si possono raggruppare le  $T_i$  osservazioni per l'unità  $i$  nel vettore  $y_i$ , scrivendolo come

$$(1.13) \quad \mathbf{y}_i = \mathbf{x}_i\beta + \mathbf{u}_i$$

Il vettore  $\mathbf{u}_i$ , che comprende i disturbi per l'individuo  $i$ , ha una matrice varianza-covarianza data da

$$(1.14) \quad \text{Var}(\mathbf{u}_i) = \sigma_\varepsilon^2 I + \sigma_\alpha^2 J$$

dove  $J$  è una matrice quadrata con tutti gli elementi uguali a 1.

Può essere dimostrato (si veda Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, 2002) che la matrice

$$(1.15) \quad K_i = I - \frac{\theta_i}{T_i} J$$

dove

$$(1.16) \quad \theta_i = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T_i \sigma_\alpha^2}}$$

ha la proprietà

$$(1.17) \quad K_i \Sigma K_i' = \sigma_\varepsilon^2 I$$

Di conseguenza il modello trasformato

$$(1.18) \quad K_i \mathbf{y}_i = K_i \mathbf{x}_i \beta + K_i \mathbf{u}_i$$

soddisfa le condizioni del teorema di Gauss-Markov, e lo stimatore OLS è sufficientemente efficiente. Tuttavia, dal momento che

$$(1.19) \quad K_i \mathbf{y}_i = \mathbf{y}_i - \theta_i \bar{y}_i$$

la stima tramite GLS è equivalente alla stima mediante OLS effettuata variabili “*quasi-demeaned*”, ovvero variabili dalle quali si sottrae una frazione  $\theta$  della loro media. Ponendo attenzione al fatto che per  $\sigma_\varepsilon^2 \rightarrow 0$ ,  $\theta \rightarrow 1$ , mentre per  $\sigma_\alpha^2 \rightarrow 0$ ,  $\theta \rightarrow 0$ ; ciò implica che se tutta la varianza è attribuibile agli effetti individuali, allora lo stimatore ad effetti fissi è ottimale; se, al contrario, gli effetti individuali sono trascurabili, allora lo stimatore pooled OLS è quello ottimale.

Quindi, per implementare l'approccio GLS, è necessario calcolare  $\theta$ , che a sua volta necessita del calcolo delle stime delle varianze  $\sigma_\varepsilon^2$  e  $\sigma_\alpha^2$ . Queste varianze sono spesso chiamate, rispettivamente, varianza “*within*” e “*between*” in quanto la prima si riferisce alla variazione all'interno di ogni unità cross-sectional e la seconda alle variazioni tra le unità.

Il metodo utilizzato nell'analisi presentata in questo elaborato per stimare tali grandezze è quello di Swamy-Arora (1970).  $\sigma_\varepsilon^2$  è stimato come varianza residua dal modello ad effetti fissi, e la somma  $\sigma_\varepsilon^2 + T_i \sigma_\alpha^2$  è stimata come  $T_i$  volte la varianza residua dallo stimatore "between",

$$(1.20) \bar{y}_i = \bar{X}_i \beta + e_i$$

La trasformazione del modello che deriva dai passaggi precedenti si può riassumere nella seguente equazione

$$(1.21) y_{it} - \theta \bar{y}_i = \beta_0(1 - \theta) + \beta_1(x_{it1} - \theta \bar{x}_{i1}) + \dots + \beta_k(x_{itk} - \theta \bar{x}_{ik}) + (u_{it} - \theta \bar{u}_i)$$

Laddove l'*overbar* indica le medie calcolate nel tempo. La trasformazione RE (random effects, effetti casuali) sottrae quindi alle variabili una frazione della loro media nel tempo la cui entità è legata al numero di periodi T,  $\sigma_\varepsilon^2$  e  $\sigma_\alpha^2$ . Lo stimatore GLS è semplicemente lo stimatore pooled OLS applicato all'equazione 1.17.

Dal momento che  $\theta$  può assumere valori compresi tra 0 e 1, dall'equazione 1.17 si evince come nel caso in cui  $\theta = 0$  ci si trovi nella situazione del modello pooled OLS, mentre nel caso in cui  $\theta = 1$  ci si trovi nel caso del modello ad effetti fissi. Il modello ad effetti casuali si inserisce quindi tra i due modelli nei casi in cui  $\theta \neq 0$  e  $\theta \neq 1$ . Se  $\theta$  si avvicina a 0 si sta lasciando una parte maggiore dell'effetto fisso nel termine d'errore rispetto alle situazioni in cui  $\theta$  si avvicina a 1 e, di conseguenza, la distorsione asintotica dello stimatore RE sarà maggiore.

La comparazione tra le stime fornite dai tre modelli è solitamente utile in quanto permette di comprendere la natura delle distorsioni causate dal fatto di lasciare  $\alpha_i$  interamente (come nel caso pooled OLS) o parzialmente (come nel caso random effects) nel termine d'errore (Wooldridge, 2006). Le assunzioni che devono essere valide per il modello ad effetti casuali sono:

- **Ass. F1, F2, F4, F5 e F6**
- **Ass. R3**, sostituisce la F3: non ci sono relazioni perfettamente lineari tra le variabili esplicative.
- **Ass. R4**, in aggiunta alla F4:  $E(\alpha_i | X_i) = \beta_0$
- **Ass. R5**, in aggiunta alla F5:  $\text{Var}(\alpha_i | X_i) = \sigma_\alpha^2$

## LA SCELTA DELLO STIMATORE

In prima approssimazione, la scelta dello stimatore può essere effettuata in relazione alla natura del dataset. Se il panel comprende osservazioni di un insieme limitato e fisso di unità d'interesse, è corretto presupporre che il modello ad effetti fissi sia il più efficiente. Se il panel include osservazioni su un alto numero di individui selezionati casualmente, si presuppone che sia più adatto il modello a effetti casuali.

Questo criterio decisionale molto semplificato non tiene conto tuttavia di numerose questioni statistiche.

In particolare, alcuni set di dati panel contengono variabili i cui valori sono specifici delle unità *cross-sectional* ma i quali non variano nel tempo. Se si vogliono inserire tali variabili nel modello, l'opzione ad effetti fissi non è percorribile. Quando il modello ad effetti fissi è implementato usando variabili dummy, il problema che si pone è la perfetta collinearità tra variabili *time-invariant* e dummy *per-unit*. Effettuando il processo di *de-meaning*, ovvero la sottrazione delle medie, su variabili con valori costanti, ovviamente il risultato è una variabile con valori nulli.

Restrizioni si riscontrano anche nell'utilizzo del modello a effetti casuali. In particolare, questo stimatore è una combinazione di pooled OLS e stimatore "between". Supponendo di disporre di osservazioni su n unità o individui e di avere k variabili indipendenti di interesse, se  $k > n$  lo stimatore "between" sarebbe indefinito, e di conseguenza sarebbe indefinito anche lo stimatore ad effetti

casuali.

Tenendo conto delle restrizioni sopra presentate, la scelta tra effetti fissi ed effetti casuali si può effettuare a seconda della preferenza tra efficienza e consistenza. Dal punto di vista meramente statistico, esiste un tradeoff tra *robustness* ed *efficiency*. Nel modello ad effetti fissi non si fanno ipotesi sugli effetti “di gruppo” (ovvero le differenze time-invariant nelle medie tra i gruppi) oltre al fatto che esistano e possano essere successivamente testate. Pertanto, una volta che questi effetti sono esclusi con il processo *time-demeaning*, i parametri restanti possono essere stimati. Al contrario, lo stimatore ad effetti casuali non elimina gli effetti di gruppo ma li tratta come estrazioni da una data distribuzione di probabilità. Ciò richiede che gli effetti individuali siano rappresentabili come una parte rilevante del termine d’errore, e che siano incorrelati con ogni variabile esplicativa (Cottrell, 2008).

La conseguenza è che lo stimatore ad effetti fissi è sempre valido in quanto ha un set di ipotesi meno restrittivo del modello ad effetti casuali, ma non riesce a stimare i parametri relativi a variabili che non dipendono dal tempo. Il modello ad effetti casuali, oltre a stimare i parametri sopracitati, permette di effettuare una stima più efficiente dei parametri relativi alle variabili che cambiano nel tempo. Tuttavia, le ipotesi aggiuntive devono essere valide. Se, ad esempio, c’è motivo per pensare che gli effetti individuali siano correlati ad alcune delle variabili dipendenti, il modello random effects sarebbe incoerente, mentre il modello fixed effects sarebbe comunque valido.

# METODOLOGIA OPERATIVA

Lo strumento utilizzato per l'analisi statistica ed econometrica dei dati è il software open-source Gretl. Esso è scritto nel linguaggio di programmazione C ed utilizza il toolkit GTK+ per l'interfaccia grafica (si veda [gretl.sourceforge.net](http://gretl.sourceforge.net) per ulteriori dettagli). L'utilizzo di gnuplot per la generazione dei grafici rende questi ultimi particolarmente chiari ed intuitivi. Il programma, sviluppato da A. Cottrell (Wake Forest University) e R. Lucchetti (Università Politecnica delle Marche) è quindi un adeguato strumento per effettuare l'analisi dei dati presentata in questo elaborato.

## INTERPRETAZIONE DELL'OUTPUT

L'output contenente la stima dei parametri del modello si presenta in maniera tabulare, mostrando in diverse colonne le stime relative ai *coefficienti*, agli *errori standard*, ai *rapporti t* ed ai *p-value*. L'ultima colonna è riservata ad un indicatore visivo relativo alla significatività osservata della stima del parametro. Esso restituisce un numero di asterischi che varia tra 0 e 3. La seguente tabella riassume i livelli di significatività indicati:

***	$p\text{-value} < 0,01$
**	$p\text{-value} < 0,05$
*	$p\text{-value} < 0,10$

Pertanto, a colpo d'occhio, si possono individuare immediatamente i parametri la cui stima sia significativa, una volta fissato il livello di significatività. Inoltre, è d'immediata comprensione anche il livello di significatività osservato. Generalmente, il livello di significatività osservato è la probabilità che, quando l'ipotesi  $H_0$  è vera, la funzione test  $T$  assuma un valore contrario ad  $H_0$  (nel senso indicato da  $H_1$ ) almeno tanto quanto  $t$ , laddove si voglia verificare una certa ipotesi  $H_0$  usando una funzione test  $T$ , e si trovi, in base ai dati, un valore osservato  $t$  per  $T$ .

$\alpha_{oss} = P(T \in D; H_0)$  se  $D$  rappresenta l'insieme di possibili valori di  $T$  che portano prove contro  $H_0$  almeno con la stessa forza con cui lo fa  $t$ .

L'utilizzo del p-value è di immediata interpretazione, e permette di cogliere il grado di coerenza delle ipotesi con  $H_0$ .

Operativamente, in questo elaborato si utilizzerà un semplice schema di analisi dell'output che si può riassumere nelle seguenti fasi:

- *Controllo del livello di significatività*: si discernono le variabili significative da quelle non significative per dare conferma o meno della validità della supposizione di una loro influenza sulla variabile dipendente *registrations*. Nel caso di  $p\text{-value} < 0,01$  (\*\*\*) si ha la risposta ottimale in quanto i valori delle stime sono significativi e i dati supportano il rifiuto dell'ipotesi nulla che il coefficiente sia uguale a zero per valori di significatività  $\alpha$  bassi. Tuttavia, anche livelli del p-value più alti (fino a 0,1, ovvero \*) sono accettabili in quanto la natura del dataset, con un numero di osservazioni contenuto, ha portato alla scelta arbitraria di un livello di significatività di 0,1. Pertanto, le variabili il cui associato p-value sia inferiore a 0,1 saranno considerate significative nell'analisi riportata in questo elaborato.
- *Valutazione del coefficiente*: ogni coefficiente della regressione indica la variazione media nella variabile dipendente al variare di una unità della variabile indipendente, mantenendo le altre variabili indipendenti costanti. Nella fase di interpretazione dell'output si valuta, in primis, il segno della stima del coefficiente, e si valuta se esso sia sensato dal punto di vista economico. Ad esempio, è corretto aspettarsi che all'aumentare del valore della variabile

*GDPgrowth* si verifichi un aumento, in qualche misura, di *registrations*. Se l'output del modello presenta un coefficiente stimato per la variabile *GDPgrowth* positivo, allora la supposizione si può considerare preliminarmente accettabile. Il valore va poi interpretato anche nella sua dimensione in valore assoluto, per cui valori assoluti maggiori delle stime indicano influenze maggiori sul valore della variabile dipendente.

## TEST SUI MODELLI

Saranno effettuati sia test sul singolo modello, sia test tra i differenti modelli.

I singoli modelli stimati saranno testati per verificare che non siano violate le ipotesi di: omoschedasticità, stabilità strutturale, specificazione.

Saranno inoltre effettuati test tra i differenti modelli per capire quale tra essi sia il più adatto ad una stima adeguata dei parametri.

I test sui singoli modelli sono:

### Test di White – test non parametrico di Wald

Il test utilizzato per verificare l'assunzione di omoschedasticità è il *test di White* (White, 1980) per l'eteroschedasticità nel caso del modello pooled OLS, mentre nel caso del modello fixed effects il software Gretl sfrutta il *test non parametrico di Wald* (per il modello ad effetto casuali Gretl non prevede test sull'eteroschedasticità). Il primo ha come ipotesi nulla la non presenza di eteroschedasticità, e il secondo ha come ipotesi nulla che le unità abbiano in comune la varianza dell'errore (omoschedasticità). Pertanto, in sede di output nel software Gretl, entrambi i test sono strutturati nelle ipotesi in modo da fornire la stessa metodologia di interpretazione dei risultati: laddove il p-value della statistica test sia inferiore al livello critico 0,05 si rifiuta  $H_0$ .

### Test Chow

Un test effettuabile per verificare la stabilità strutturale (ovvero la stabilità dei parametri nel tempo) è il *test Chow* (Chow, 1960). Esso ricerca se esiste ed è significativa una data di rottura strutturale, ovvero un momento  $t$  dove si verifica un cambiamento netto dei parametri della regressione. Per fare ciò si modifica la regressione inserendo una variabile dummy che assuma valore 0 in un certo intervallo di tempo e 1 in un altro. Nel caso del dataset analizzato, la variabile  $d$  assume valore 0 dal 2001 al 2007 e 1 dal 2008 al 2014. Si vuole così verificare se, in concomitanza con la crisi economica del 2008, i parametri stimati del modello subiscono variazioni significative. Se il valore del test è superiore a 2, si è verificato un cambiamento strutturale (si veda Wooldridge, 2010). Il p-value inferiore a 0,05 implica quindi una rottura strutturale.

### Test RESET di Ramsey

L'ipotesi che la forma funzionale sia lineare o log-lineare è molto diffusa, tuttavia se si impiega una forma funzionale errata si possono ottenere stime distorte e inconsistenti, e risultati scorretti in fase di inferenza statistica. Per verificare che tale ipotesi sia plausibile il software permette di utilizzare il *test RESET di Ramsey* (Ramsey, 1969). L'acronimo RESET sta per *Regression Equation Specification Error Test*. L'intuizione che sta dietro al test è che se combinazioni non lineari delle variabili esplicative hanno una qualche influenza sulla variabile indipendente, la bontà di adattamento del modello ne risente. Per testare ciò, al modello stimato si aggiunge una nuova variabile esplicativa,  $\hat{y}_i^2$ . Se il coefficiente di questa ulteriore variabile è significativamente diverso da zero, il modello iniziale è inadeguato.

Esistono diversi test che si possono effettuare sui modelli stimati per valutare quale sia quello con la miglior bontà di adattamento. Si possono testare gli effetti fissi mediante un F-test, oppure si possono



testare gli effetti casuali tramite il test LM di Breusch e Pagan (1980). Il primo compara un modello ad effetti fissi e uno pooled OLS per valutare quanto il modello ad effetti fissi migliora la bontà di adattamento, mentre il secondo effettua lo stesso tra pooled OLS e modello ad effetti casuali. Infine, il test Hausman testa la somiglianza tra effetti fissi e casuali.

### **F-Test**

L'*F-test* per gli effetti fissi testa l'ipotesi nulla che gli effetti fissi  $\alpha_i$  siano uguali in tutte le unità. Rifiutare tale ipotesi significa che gli effetti fissi sono non nulli. Pertanto, nel caso in cui l'ipotesi nulla sia rifiutata, si può concludere che c'è un significativo effetto fisso o un significativo aumento della bontà di adattamento del modello nel modello ad effetti fissi, e il modello ad effetti fissi sia migliore del modello pooled OLS.

### **Test LM di Breusch-Pagan**

Il *test LM* di Breusch-Pagan testa l'ipotesi nulla che la varianza di  $\alpha_i$  sia nulla. Se l'ipotesi nulla è rifiutata si può concludere che c'è un significativo effetto casuale nel dataset e che il modello ad effetti casuali riesca a conciliarsi meglio con l'eterogeneità rispetto al modello pooled OLS. (Breusch, Pagan, 1980)

### **Test Hausman**

Il *test Hausman* testa l'ipotesi nulla che gli effetti individuali non siano correlati con alcun regressore nel modello. Se l'ipotesi nulla è rifiutata, si può concludere che gli effetti individuali  $\alpha_i$  siano significativamente correlati con almeno un regressore nel modello e che quindi il modello ad effetti casuali sia problematico. Perciò sarà preferibile l'utilizzo di un modello ad effetti fissi. (Hausman, 1978)

## **COSTRUZIONE DATABASE**

Per studiare un'eventuale correlazione tra le variabili scelte e le vendite di automobili è necessario partire da un database contenente un numero sufficiente di dati rilevanti. In particolare, nell'analisi sviluppata in questo elaborato si è scelto di analizzare i dati provenienti dai 14 mercati europei più grandi in termini di numero di veicoli immatricolati, nel periodo temporale che va dal 2001 al 2014. I 14 mercati selezionati sono, in ordine decrescente del numero medio annuale di veicoli immatricolati: Germania, Francia, Regno Unito, Italia, Spagna, Belgio, Olanda, Austria, Grecia, Portogallo, Svezia, Irlanda, Finlandia, Danimarca. Per motivi di praticità, si assegna ad ogni elemento della lista dei paesi un numero indice, che va da 1 a 14. Pertanto si costituisce un set di dati panel il cui indice per le unità è *Country* ( $n = 1 \rightarrow N=14$ ) e la cui variabile indice temporale è *Year* ( $t = 1 \rightarrow T=14$ ).

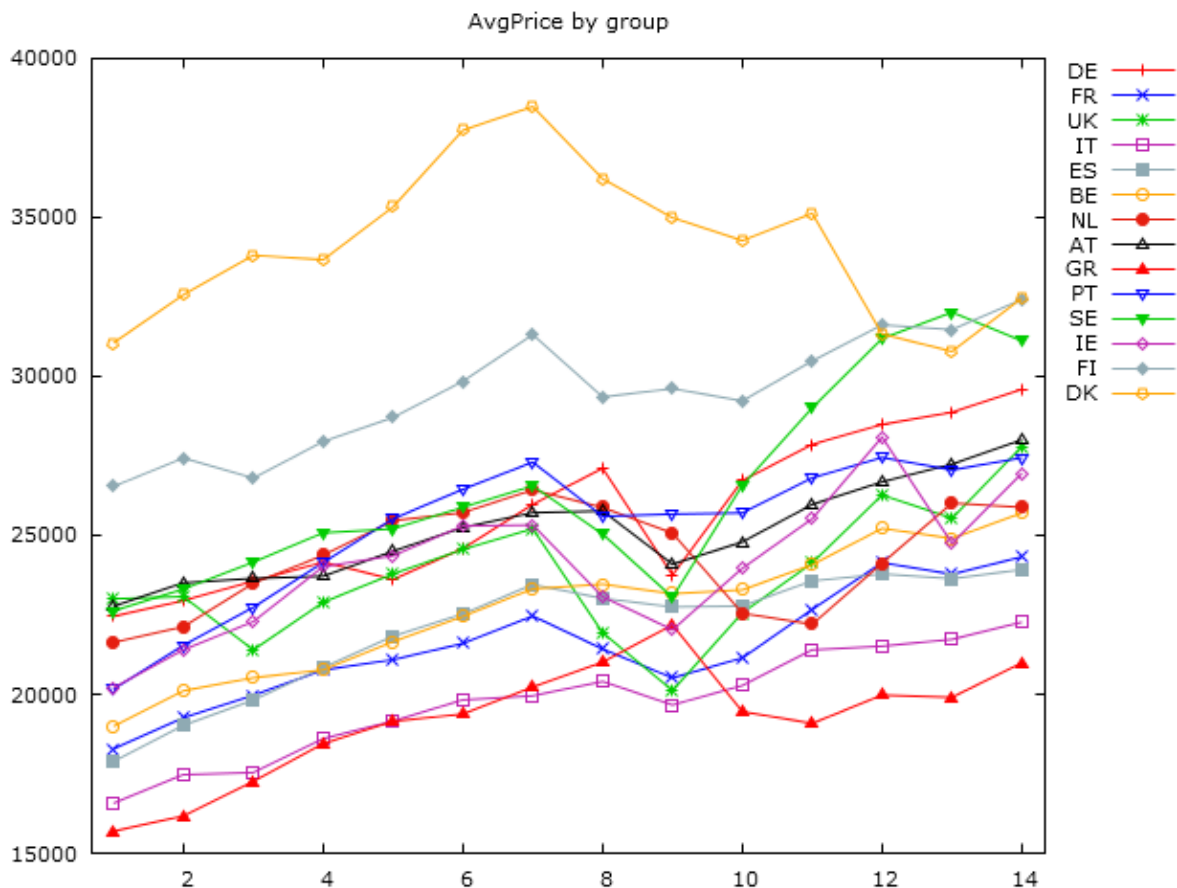
# VARIABILI

Le variabili indipendenti scelte per lo studio sono:

## Prezzo medio di vendita di un'automobile

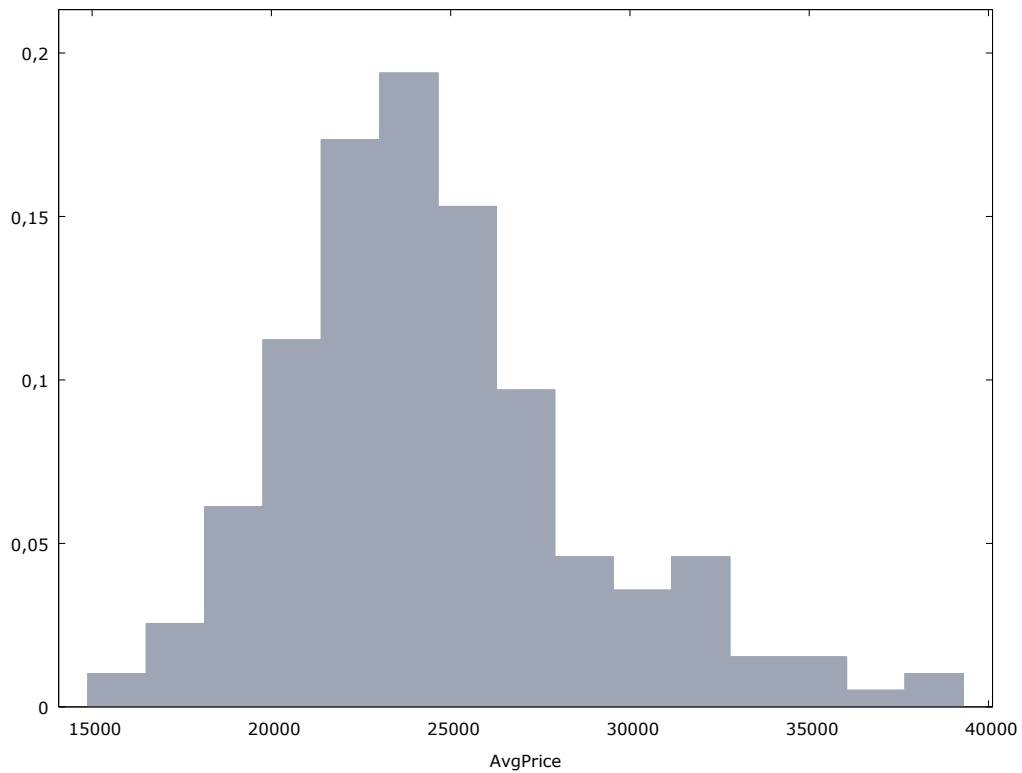
*Avg\_Price<sub>it</sub>*: il prezzo medio di vendita di un'automobile nel paese *i* al periodo *t*. È una variabile interessante in quanto ci si può aspettare che Stati in cui si riscontrano prezzi di vendita più alti presentino redditi medi più elevati, e quindi che le vendite risentano meno di leggere fluttuazioni nei dati relativi alle altre variabili macroeconomiche. Elemento di cui tener conto nell'interpretazione dei dati è che il prezzo medio di vendita delle automobili dipende anche dalle peculiarità dei singoli mercati. Ad esempio, ci si potrebbe aspettare che i cittadini dei paesi nordeuropei paghino prezzi medi di vendita elevati in quanto potrebbero aver bisogno di automobili più prestanti per far fronte alle intemperie tipiche del loro territorio. Al contrario, nei paesi del Sud dell'Europa, ci si potrebbe aspettare prezzi medi di vendita più bassi dovuti ad un mercato automobilistico più imperniato su piccole utilitarie da utilizzare nei centri cittadini.

I dati sui prezzi medi di vendita utilizzati nell'analisi derivano dai database ICCT, compilati secondo i dati forniti dai produttori, dalle associazioni ed enti automobilistici dei vari paesi, dalle registration authorities, dalla UKVCA (United Kingdom Vehicle Certification Agency), dalla KBA (Kraftfahrtbundesamt). (Dati raccolti il 24 Maggio 2016).



Il grafico mostra come sia presente una effettiva separazione tra i prezzi medi pagati dai cittadini nordeuropei e quelli pagati dai cittadini dei paesi del Sud. Si notano inoltre importanti variazioni nel

biennio 2008-2010.



Il grafico della distribuzione di frequenza mostra come la maggior parte dei valori sia compresa tra 18000 e 30000. In particolare, il 83,67% dei valori si concentra nell'intervallo 18137 – 29531.

*Statistiche descrittive, usando le osservazioni 1:01 - 14:14  
per la variabile AvgPrice (196 osservazioni valide)*

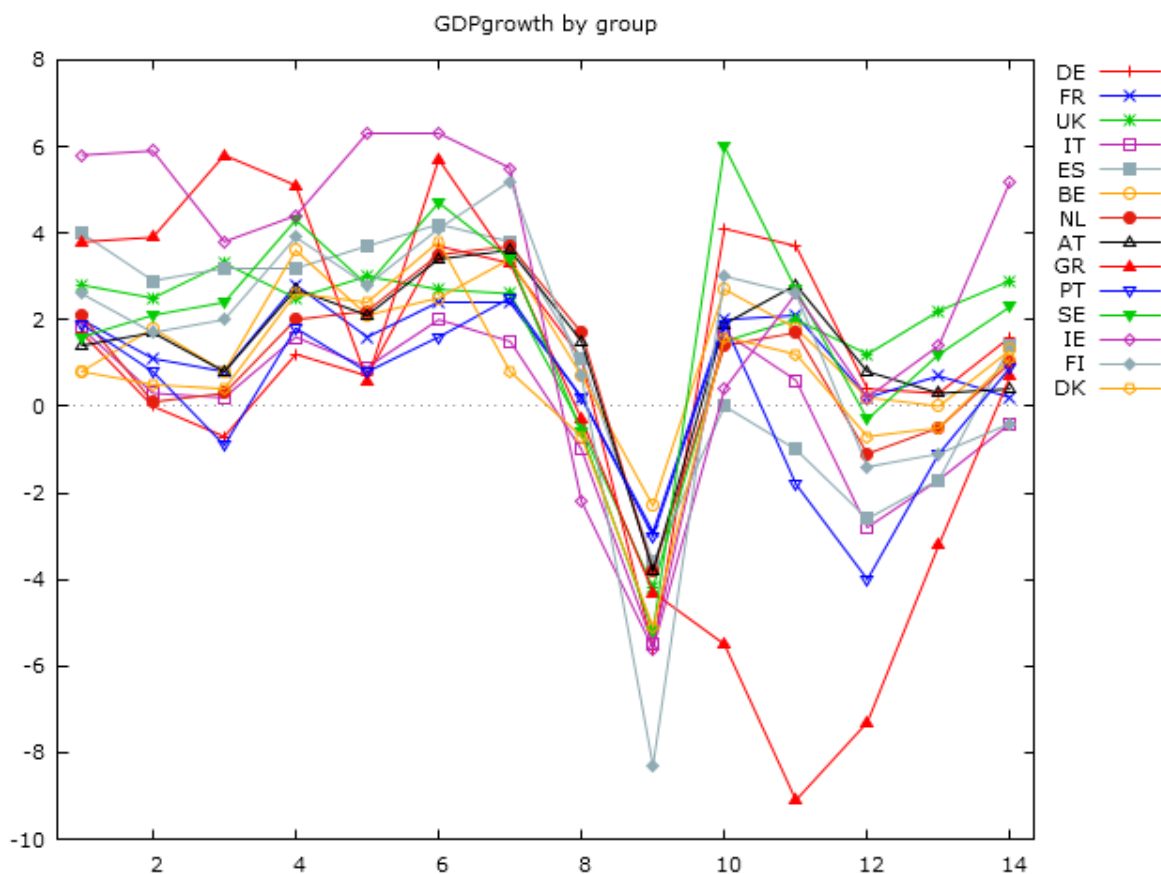
Media	Mediana	Minimo	Massimo
24539,7	23988,5	15695,0	38484,0
Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
4239,55	0,172763	0,760534	0,678302
5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
18579,6	32728,6	4931,75	0

D.s. nei gruppi                    2095,0  
D.s. tra i gruppi                    3856,0

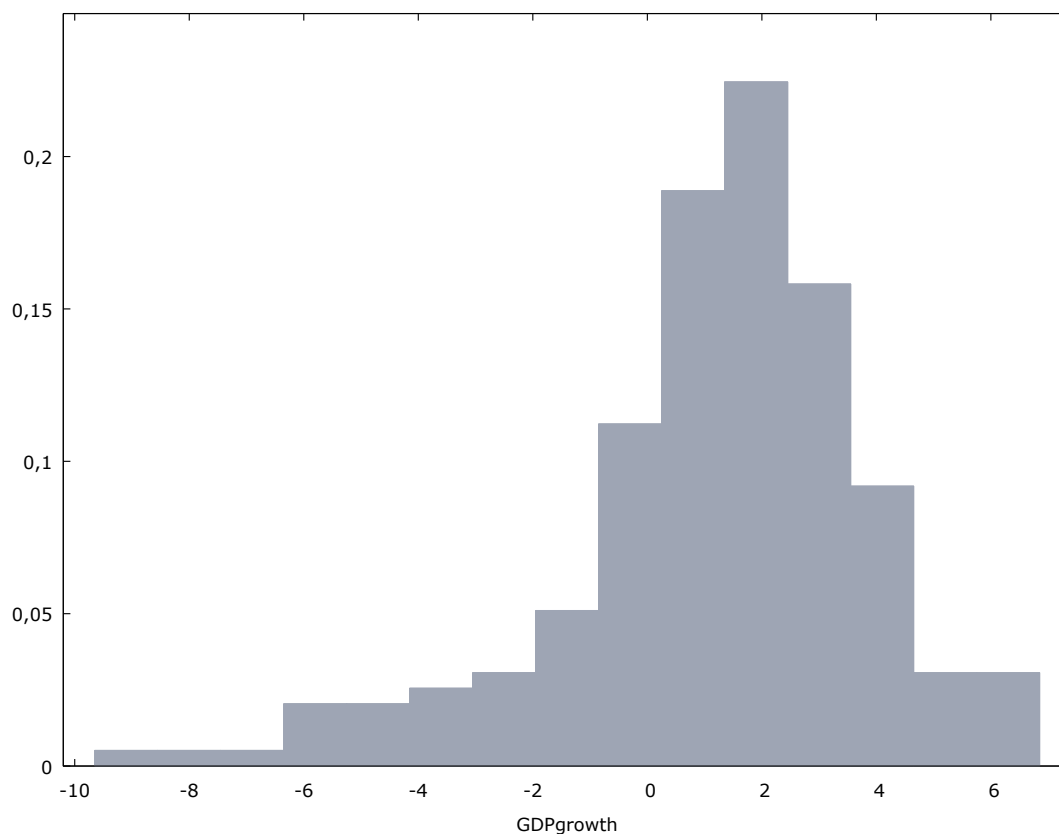
## Tasso di crescita del PIL

$GDP\_growth_{it}$ : la variazione annuale del PIL nel paese  $i$  al periodo  $t$ . Tale variabile dovrebbe essere esplicativa del fenomeno studiato in quanto ci si può ragionevolmente aspettare che in periodi di espansione economica il numero di automobili vendute cresca, mentre in periodi di contrazione economica o recessione il numero di automobili vendute diminuisca. La scelta di utilizzare le variazioni annuali del PIL a livello aggregato e non procapite forse riduce la precisione dell'analisi, ma evita di dover modificare i valori delle altre variabili in modo sostanziale. Ad esempio, se si fossero inseriti nel database i dati sulle variazioni del PIL procapite, anche il numero di veicoli venduti avrebbe dovuto essere rapportato alla popolazione. Questo avrebbe probabilmente permesso un'analisi più accurata, ma un confronto dei dati mi ha permesso di definire i dati sulla variazione del PIL a livello aggregato come sufficienti per l'analisi da effettuare.

I dati sulle variazioni annuali del PIL utilizzati nell'analisi derivano dai database WDI, visualizzati tramite l'interfaccia virtuale del sito web al seguente indirizzo: <http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG> (Dati raccolti il 24 Maggio 2016).



Il grafico mostra come in tutti i paesi in analisi vi sia un trend del tasso di crescita del PIL analogo dal 2001 al 2010. Dal 2001 al 2007 il tasso è stato prevalentemente positivo, mentre nel 2008 vi è stato un crollo causato dagli effetti della crisi economico-finanziaria del 2007-2008. Dal 2010 in poi i trend dei vari paesi non sono più simili a causa delle differenti velocità di reazione delle singole economie nazionali alla recessione. Nel 2014 alcuni paesi presentano ancora un tasso di crescita negativo, mentre altri sono ben oltre il 2%.



Il grafico della distribuzione di frequenza mostra una situazione in cui i valori sono prevalentemente compresi tra -2% e 4%. In particolare, il 73,47% dei valori è nell'intervallo -1,95% – 3,55%.

*Statistiche descrittive, usando le osservazioni 1:01 - 14:14  
per la variabile GDPgrowth (196 osservazioni valide)*

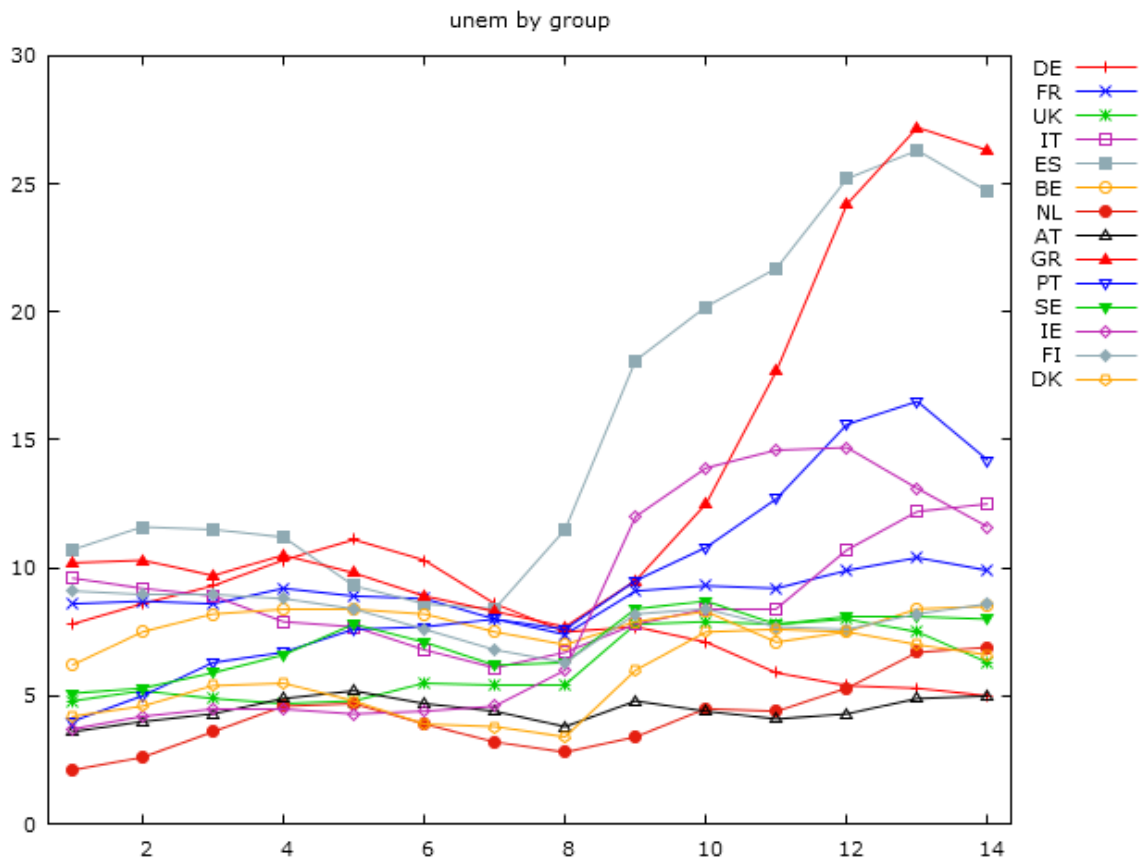
Media	Mediana	Minimo	Massimo
1,12500	1,55000	-9,10000	6,30000
Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
2,67849	2,38088	-1,03327	1,78082
5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
-4,42000	5,20000	2,50000	0

D.s. nei gruppi                    2,6534  
D.s. tra i gruppi                    0,80387

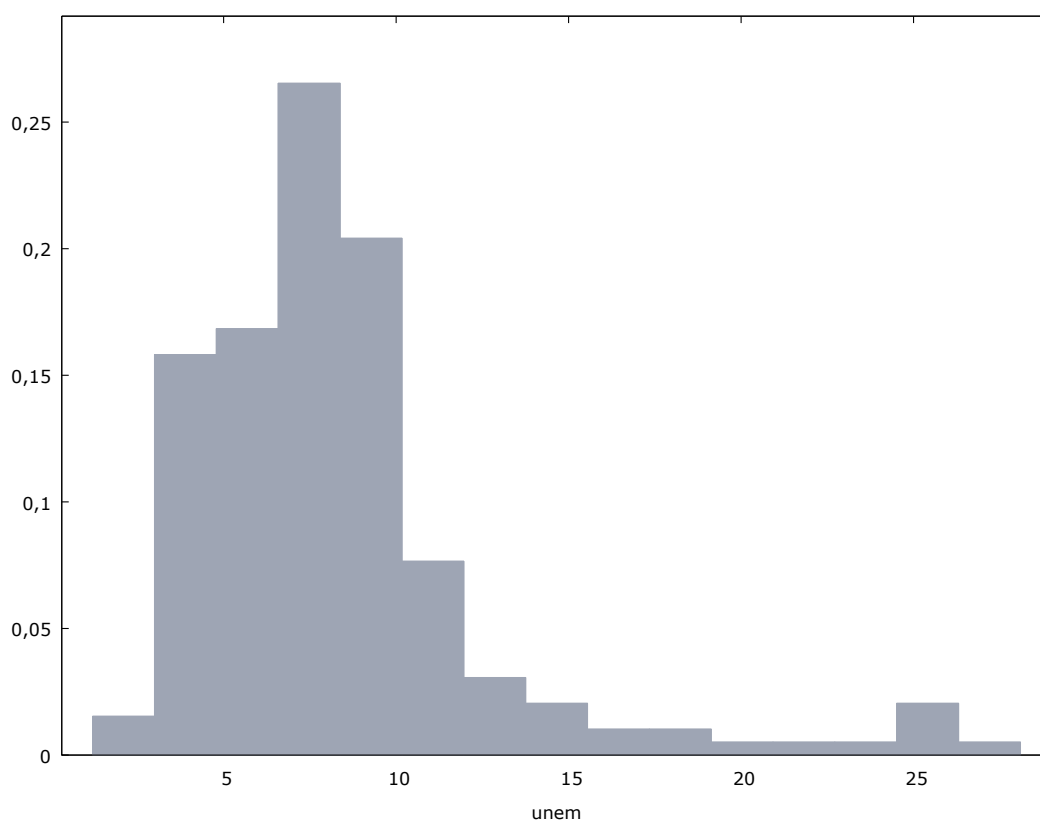
## Tasso di disoccupazione

$Unem_{it}$ : il tasso di disoccupazione nel paese  $i$  al periodo  $t$ . Questa potrebbe essere UNA variabile d'interesse in quanto l'aspettativa è che laddove vi sia un tasso di disoccupazione più elevato vi siano vendite di automobili più ridotte, in quanto il mercato di destinazione viene privato del numero di nuovi disoccupati. Inoltre, sembrerebbe perfettamente sensato affermare che all'aumentare del tasso di disoccupazione il numero di automobili diminuisca, in quanto, semplicemente, i cittadini che non lavorano non possono comprare nuove automobili.

I dati sulla disoccupazione utilizzati nell'analisi derivano dai database della International Labour Organization, visualizzati tramite l'interfaccia grafica del sito web al seguente indirizzo: <http://data.worldbank.org/indicatore/SL.UEM.TOTL.ZS> (Dati raccolti il 30 Maggio 2016)



Il grafico mostra come l'andamento del tasso di disoccupazione sia simile in tutti i paesi, tranne cinque. Italia, Irlanda, Portogallo, Spagna e Grecia infatti, dopo la crisi del 2007-2008 mantengono un tasso di disoccupazione al di sopra del 10%, con picchi oltre il 25% nel caso di Grecia e Spagna.



Il grafico della distribuzione di frequenza mostra come la maggior parte dei valori si concentri tra 3% e 12%. In particolare, il 90,31% dei valori del tasso di disoccupazione si trovano nell'intervallo 2,9964% - 13,754%.

*Statistiche descrittive, usando le osservazioni 1:01 - 14:14  
per la variabile unem (196 osservazioni valide)*

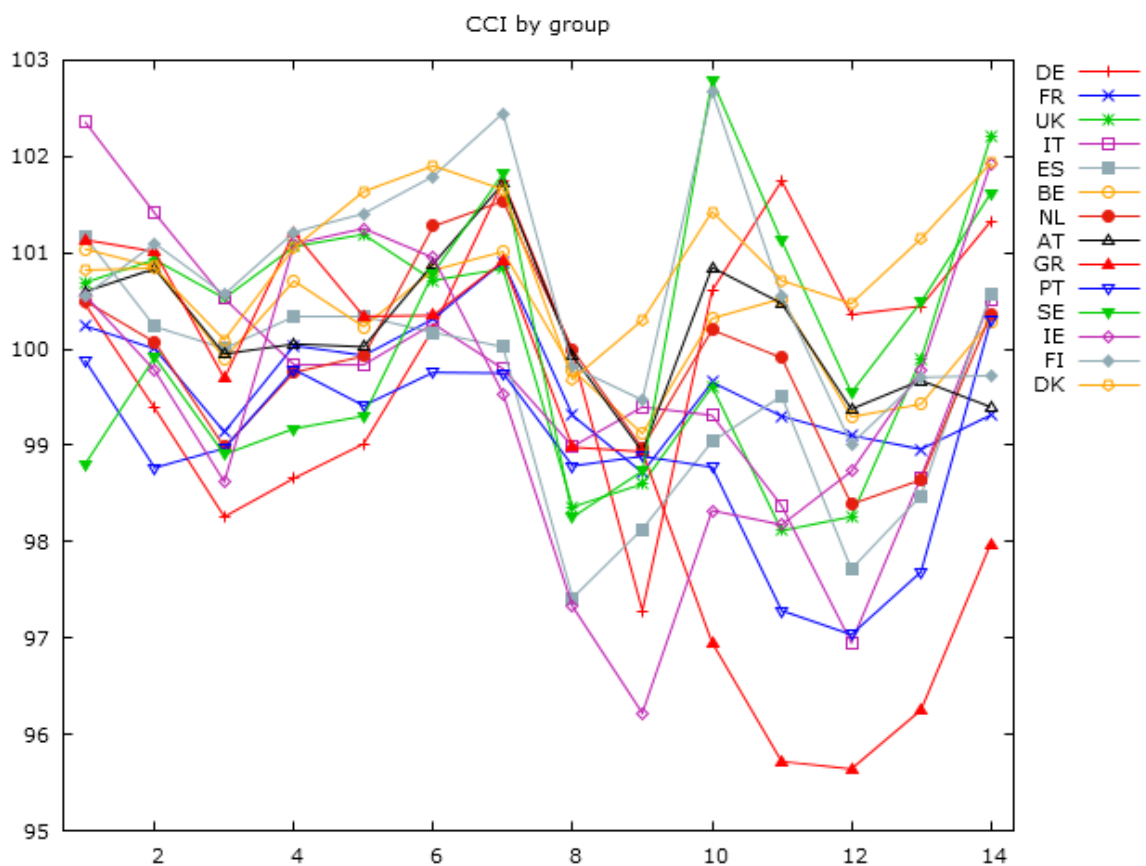
Media	Mediana	Minimo	Massimo
8,29592	7,75000	2,10000	27,2000
Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
4,39756	0,530087	2,20233	6,05128
5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
3,78500	17,7600	3,90000	0

D.s. nei gruppi                    3,2550  
D.s. tra i gruppi                    3,1820

## Consumer Confidence Index

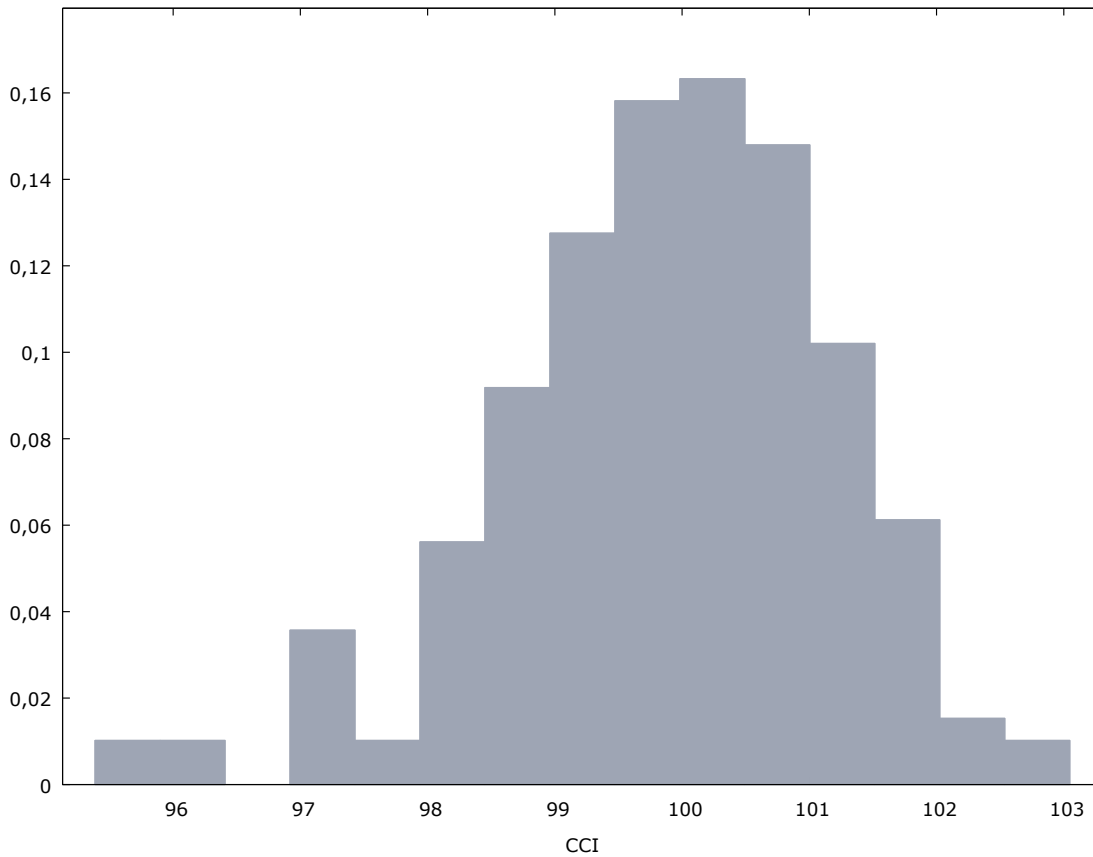
$CCI_{it}$ : il valore dell'indicatore CCI (Consumer Confidence Index) nel paese  $i$  al periodo  $t$ . Il CCI, indice di fiducia dei consumatori, è basato sulle previsioni dei nuclei familiari di effettuare acquisti economicamente importanti, in relazione alla loro situazione economica attuale e alle loro aspettative future. Si raccolgono le opinioni quantificandole rispetto al valore indice di 100. Valori sopra il 100 indicano una fiducia dei consumatori più alta rispetto allo status quo, mentre valori sotto il 100 indicano un livello di fiducia più basso rispetto a quello del valore indice. La differenza tra le risposte positive e negative fornisce un indice qualitativo delle condizioni economiche dei nuclei familiari. Normalmente produttori, commercianti, banche e governi includono i dati relativi al CCI nel loro processo decisionale. Infatti, cambiamenti repentini, in misura maggiore del 5%, storicamente indicano un cambiamento di trend dell'economia. Nell'analisi è stato inserito l'indice CCI per tenere conto di una misura della percezione dello stato di salute dell'economia. Ci si aspetta che la relazione tra CCI e numero di automobili vendute sia positiva.

I dati relativi all'indice CCI utilizzati nell'analisi sono le medie annuali calcolate dai dati mensili ritrovati nei database OCSE nel sito web all'indirizzo: <https://data.oecd.org/leadind/consumer-confidence-index-cci.htm> OECD (2016), Consumer confidence index (CCI) (indicator). doi: 10.1787/46434d78-en (Dati raccolti il 31 Maggio 2016)



Il grafico mostra come non vi sia un trend generale definito per tutti i paesi, in quanto i valori oscillano sia sopra che sotto il valore indice di 100. Tuttavia, nel biennio 2008-2009 si osserva come l'indice di fiducia dei consumatori scenda sensibilmente, per poi assumere trend completamente differenti nei vari paesi.





Il grafico della distribuzione di frequenza mostra come i valori siano prevalentemente compresi tra 98 e 101,5. In particolare, il 84,70% dei valori si trova nell'intervallo 97,940 – 101,51.

*Statistiche descrittive, usando le osservazioni 1:01 - 14:14  
per la variabile CCI (196 osservazioni valide)*

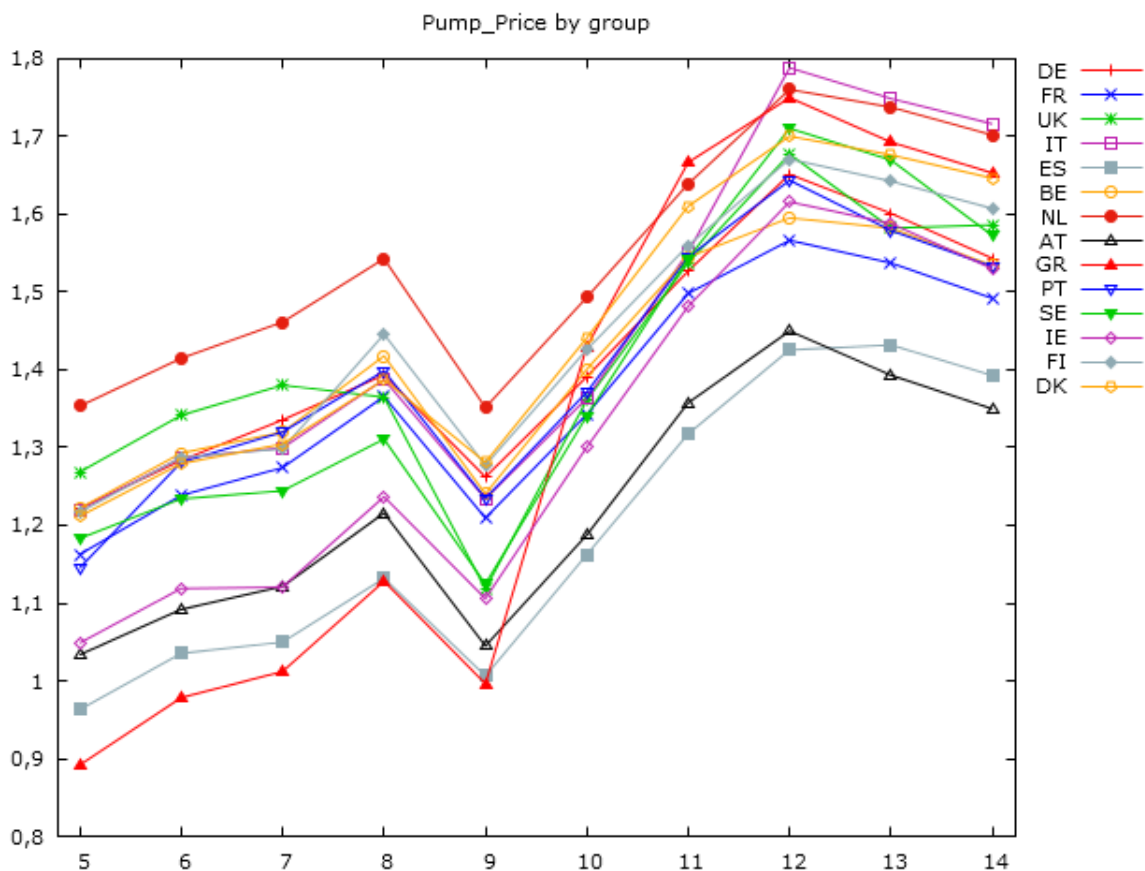
Media	Mediana	Minimo	Massimo
99,8811	99,9782	95,6415	102,792
Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
1,29056	0,0129210	-0,609373	0,710833
5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
97,3211	101,798	1,79059	0

D.s. nei gruppi                    1,2024  
D.s. tra i gruppi                    0,58209

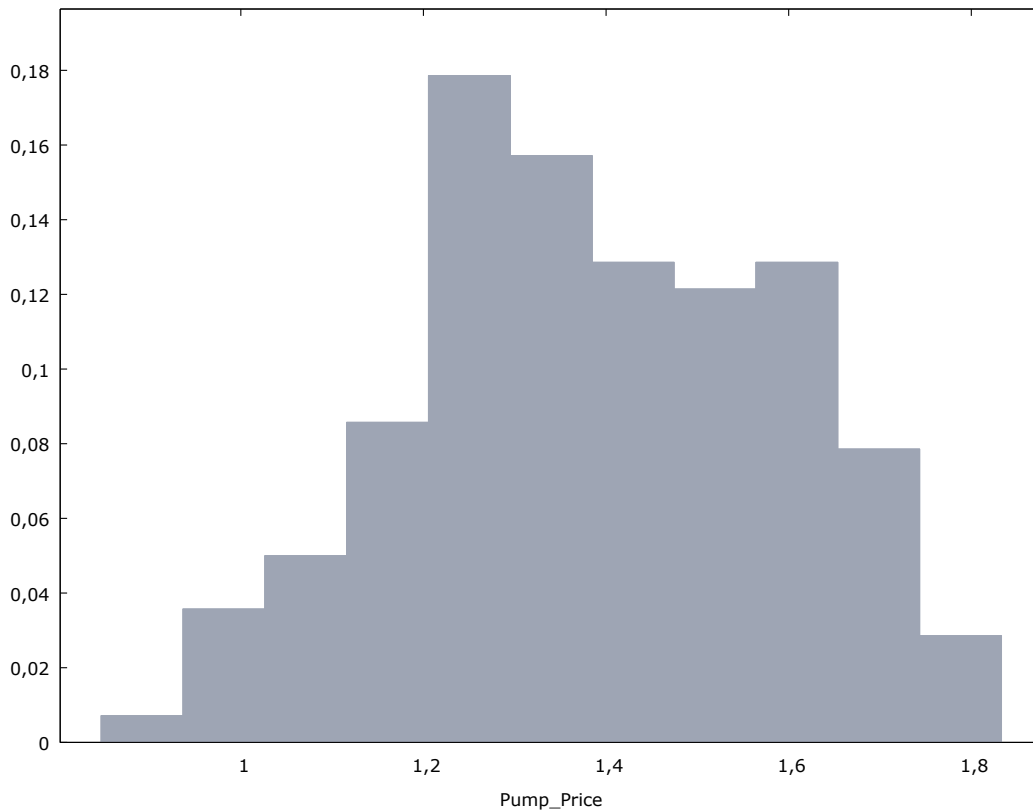
## Prezzo della benzina

$Pump\_Price_{it}$ : il prezzo medio annuale della benzina Super95 nel paese  $i$  al periodo  $t$ . Esso è calcolato come media annuale dei prezzi settimanali rilevati in diverse stazioni di servizio sparse nel paese. Il prezzo si intende comprensivo di tasse e accise, quindi esattamente lo stesso pagato dai consumatori al momento dell'acquisto. L'inserimento nel modello di una variabile che tenga conto del costo dell'utilizzo delle automobili permette di considerare, nel corso dell'analisi, le sensibili variazioni nel costo degli spostamenti avvenute negli anni. Un'alternativa all'inserimento della variabile scelta nell'analisi sarebbe potuta essere l'utilizzo dei dati sul prezzo medio annuale del greggio. Tali dati sarebbero stati di più semplice reperimento e la procedura di pulizia dei dati sarebbe stata quasi nulla in quanto i database contenenti i prezzi del petrolio sono molto più comuni e ben strutturati. Tuttavia, dal momento che si sta studiando il fenomeno delle vendite delle automobili è sembrato più interessante inserire dei valori direttamente a contatto con il consumatore finale, le cui decisioni d'acquisto sono poi oggetto d'analisi. È un indicatore più attinente all'analisi in corso rispetto al prezzo del greggio. Qualche semplice algoritmo implementato tramite il linguaggio di programmazione *Python* ha semplificato la procedura di acquisizione dei dati dai database più complessi.

Dati sul prezzo medio della benzina pagato dal consumatore tratti dai database della Commissione Europea. Medie annuali calcolate a partire dai dati settimanali presentati nei Weekly Oil Bulletins dal 2005 al 2014, rinvenibili nella pagina del sito web all'indirizzo: <https://ec.europa.eu/energy/en/data-analysis/weekly-oil-bulletin> (Dati raccolti il 6 Giugno 2016).



Il grafico mostra come sia presente un andamento uniforme dei dati sul prezzo della benzina.



Il grafico della distribuzione di frequenza mostra come la maggior parte dei valori sia compresa tra 1,10 e 1,60. In particolare, l'80% dei valori si trova nell'intervallo 1,1159 – 1,6537.

*Statistiche descrittive, usando le osservazioni 1:01 - 14:14  
per la variabile Pump\_Price (140 osservazioni valide)*

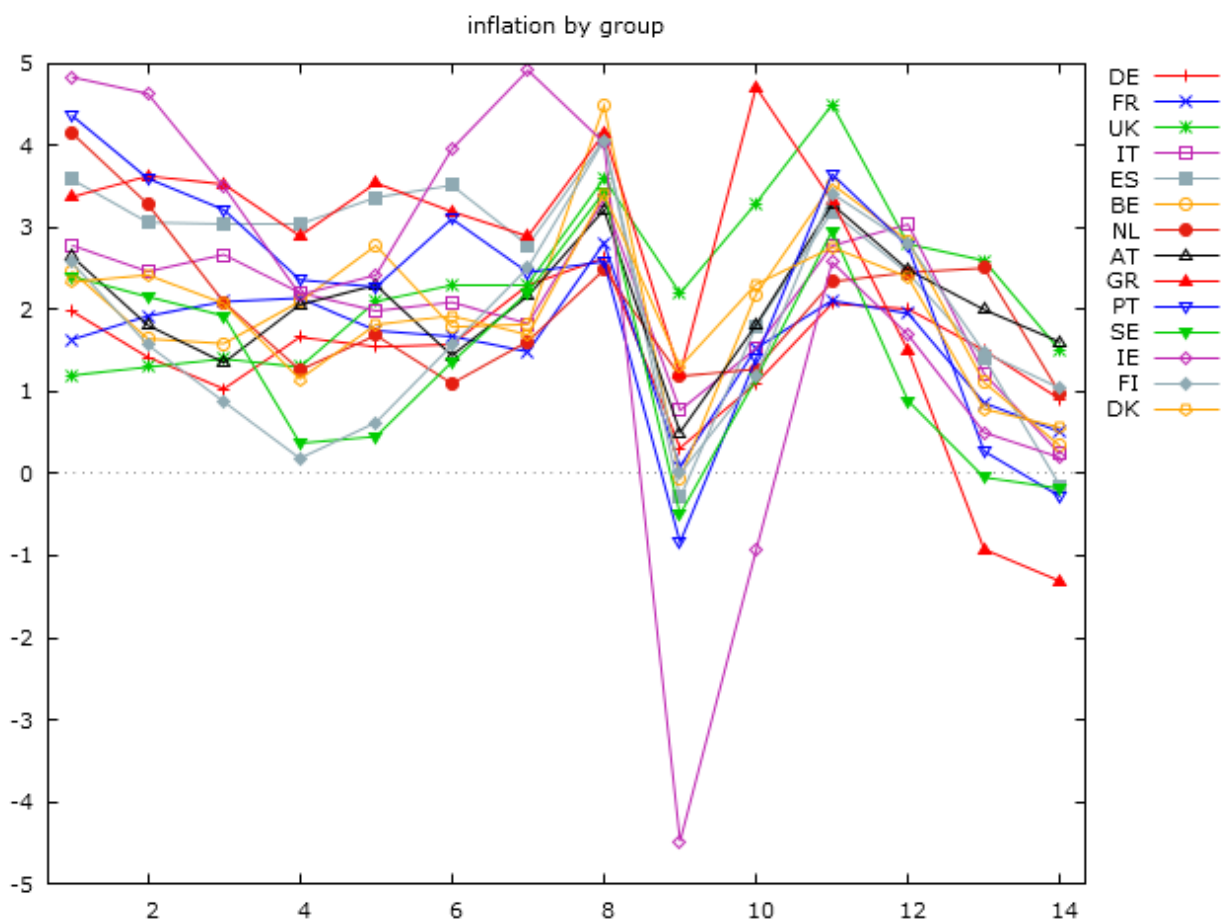
Media	Mediana	Minimo	Massimo
1,38408	1,36830	0,891796	1,78821
Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
0,206802	0,149415	-0,0825727	-0,782809
5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
1,03362	1,70998	0,314556	56

D.s. nei gruppi            0,19485  
D.s. tra i gruppi         0,094505

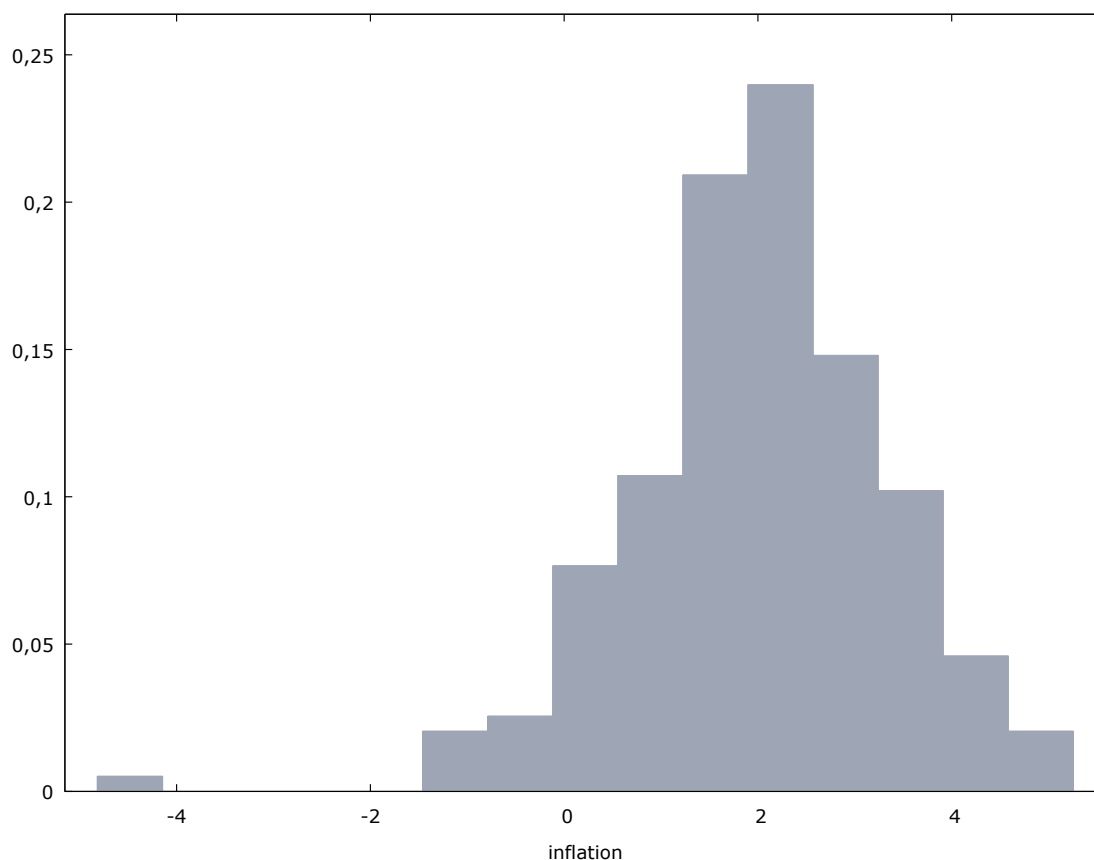
## Tasso di inflazione

$inflation_{it}$ : il tasso di inflazione nel paese  $i$  al periodo  $t$ . L'indicatore inflazionistico scelto è l'indice dei prezzi al consumo. L'inserimento di un indicatore dei prezzi al consumo permette di valutare l'effetto delle variazioni del costo della vita sul numero di automobili vendute. La scelta dell'indice CPI (Consumer Price Index) al posto del deflatore del PIL deriva dalla natura dell'indicatore. Il CPI infatti, computa il prezzo dei beni importati ed esclude il prezzo di quelli riservati all'esportazione. Inoltre, esso considera solamente i beni consumati dai cittadini, attraverso la composizione di un "carrello della spesa" che comprende i beni acquistati ad alta e media frequenza dai cittadini. Dal momento che l'analisi sulle vendite delle automobili è un'analisi sul consumo dei cittadini, sembra più coerente utilizzare il CPI piuttosto che il deflatore del PIL.

I dati relativi al CPI derivano dai database OCSE accessibili tramite l'interfaccia della pagina del sito web all'indirizzo: <https://data.oecd.org/price/inflation-cpi.htm> (Dati raccolti il 12 Luglio 2016).



Il grafico mostra un andamento molto simile dei dati per tutti i paesi considerati. Nel 2009 vi è un crollo generalizzato del tasso di inflazione, mentre in tutti gli altri periodi esso è generalmente positivo.



Il grafico della distribuzione di frequenza mostra come i valori siano distribuiti prevalentemente tra 0 e 4. In particolare, l'88,27% dei valori si trova nell'intervallo -0,11729 – 3,9116.

*Statistiche descrittive, usando le osservazioni 1:01 - 14:14  
per la variabile inflation (196 osservazioni valide)*

Media	Mediana	Minimo	Massimo
2,00476	2,08301	-4,48186	4,91876
Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
1,29271	0,644821	-0,678532	2,52831
5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
-0,155264	4,08728	1,49668	0

D.s. nei gruppi                    1,2915  
D.s. tra i gruppi                    0,35006

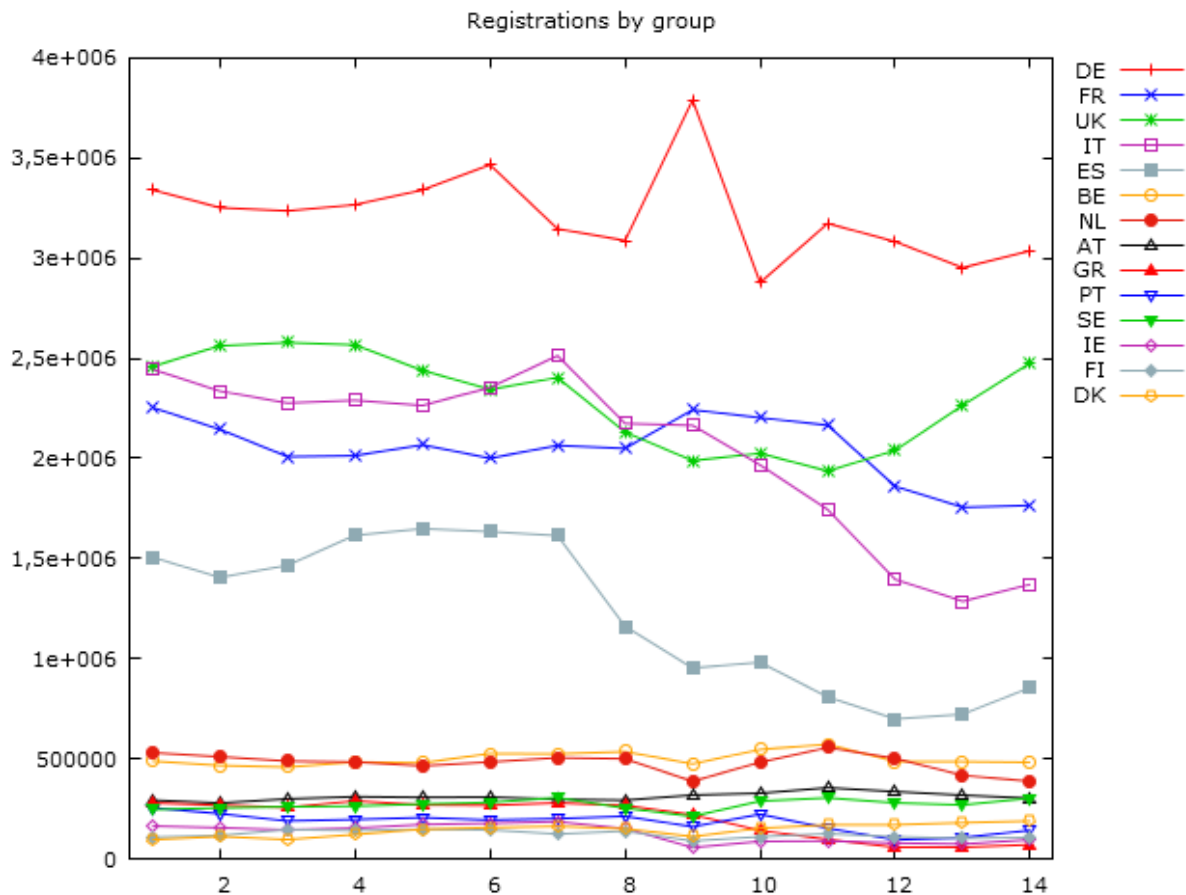
## Variabile dummy – effetto della crisi economica del 2007-2008

$d_{it}$ : variabile dummy che analizza l’impatto della crisi economica del 2007-2008 sul numero di veicoli immatricolati. Essa assume valore 0 se  $Year < 08$ , mentre assume valore 1 se  $Year \geq 08$ . Viene utilizzata nel test Chow per la verifica della presenza di un break strutturale in corrispondenza dell’anno 2008.

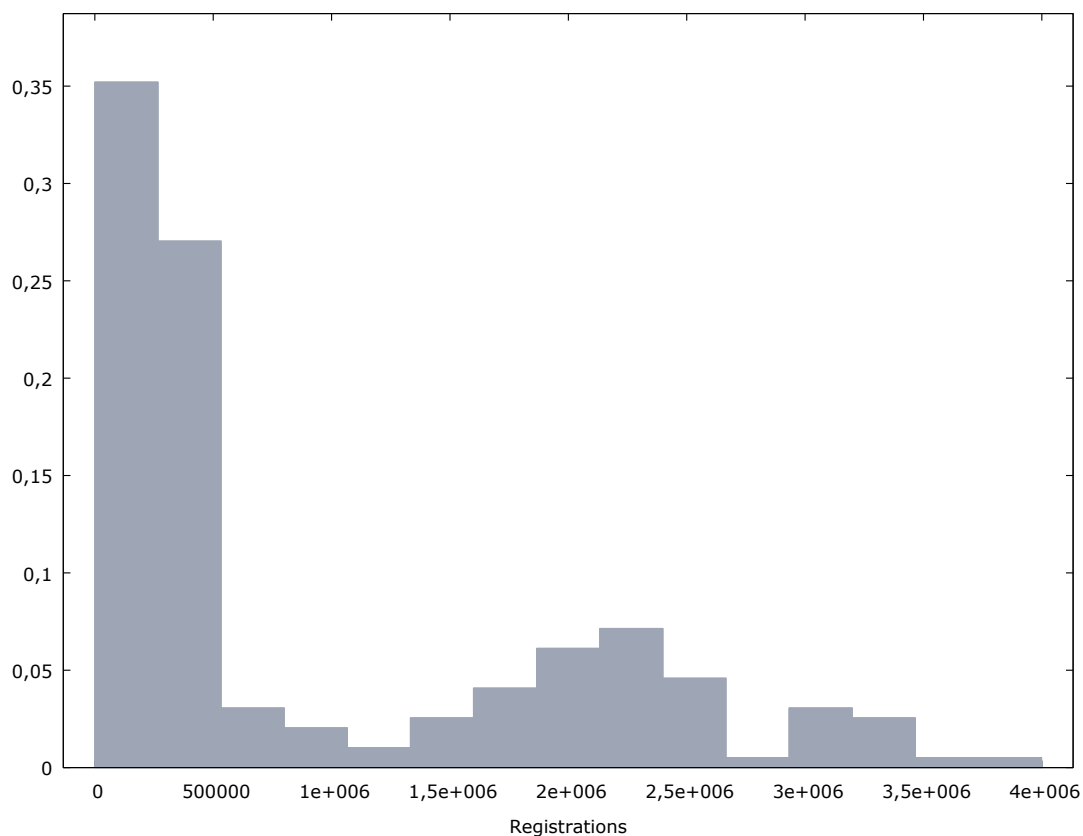
## Variabile dipendente: numero di automobili immatricolate

La variabile dipendente del modello è  $Registrations_{it}$  ovvero il numero di automobili immatricolate nel paese  $i$  al periodo  $t$ .

I dati relativi alle immatricolazioni di nuove automobili sono tratti dai database ICCT, compilati secondo i dati forniti dai produttori, dalle associazioni ed enti automobilistici dei vari paesi, dalle registration authorities, dalla UKVCA (United Kingdom Vehicle Certification Agency), dalla KBA (Kraftfahrtbundesamt). (Dati raccolti il 24 Maggio 2016).



Il grafico non mostra un evidente andamento generale dei dati per i vari paesi.



Il grafico della distribuzione di frequenza dei valori delle immatricolazioni non fornisce indicazioni particolarmente utili in questo momento. Si nota comunque che il 62,24% delle osservazioni si trova nell'intervallo 0 – 533200 e il 17,86% delle osservazioni si trovi nell'intervallo 1599000 – 2666000.

*Statistiche descrittive, usando le osservazioni 1:01 - 14:14  
per la variabile Registrations (196 osservazioni valide)*

Media	Mediana	Minimo	Massimo
940485,	371712,	57455,0	3,78956e+006
Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
1,01353e+006	1,07767	1,04041	-0,325530
5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
94551,8	3,09703e+006	1,73878e+006	0

D.s. nei gruppi            1,8259e+005  
D.s. tra i gruppi            1,0331e+006

## PREPARAZIONE DATASET

Compilato il dataset, è opportuno osservarne nel complesso le statistiche descrittive più comuni per valutare eventuali interventi correttivi da effettuare prima di stimare il modello.

Statistiche descrittive, usando le osservazioni 1:01 - 14:14  
(i valori mancanti sono stati saltati)

Variabile	Media	Mediana	Minimo	Massimo
<i>Registrations<sub>it</sub></i>	940485,	371712,	57455,0	3,78956e+006
<i>AvgPrice<sub>it</sub></i>	24539,7	23988,5	15695,0	38484,0
<i>GDPgrowth<sub>it</sub></i>	1,12500	1,55000	-9,10000	6,30000
<i>Unem<sub>it</sub></i>	8,29592	7,75000	2,10000	27,2000
<i>CCI<sub>it</sub></i>	99,8811	99,9782	95,6415	102,792
<i>Pump_Price<sub>it</sub></i>	1,38410	1,36830	0,89180	1,78821
<i>Inflation<sub>it</sub></i>	2,00476	2,08301	-4,48186	4,91876

Variabile	Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
<i>Registrations<sub>it</sub></i>	1,01353e+006	1,07767	1,04041	-0,325530
<i>AvgPrice<sub>it</sub></i>	4239,55	0,172763	0,760534	0,678302
<i>GDPgrowth<sub>it</sub></i>	2,67849	2,38088	-1,03327	1,78082
<i>Unem<sub>it</sub></i>	4,39756	0,530087	2,20233	6,05128
<i>CCI<sub>it</sub></i>	1,29056	0,0129210	-0,609373	0,710833
<i>Pump_Price<sub>it</sub></i>	0,20680	0,14942	-0,082573	-0,78281
<i>Inflation<sub>it</sub></i>	1,29271	0,644821	-0,678532	2,52831

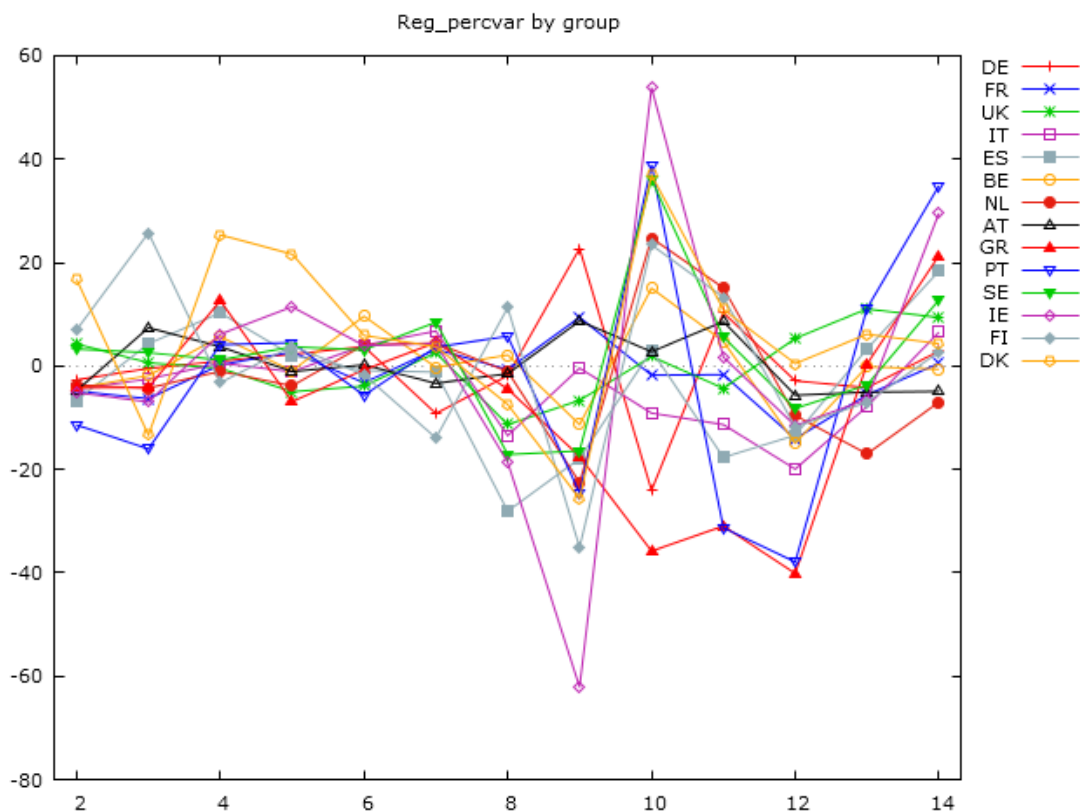
Variabile	5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
<i>Registrations<sub>it</sub></i>	94551,8	3,09703e+006	1,73878e+006	0
<i>AvgPrice<sub>it</sub></i>	18579,6	32728,6	4931,75	0
<i>GDPgrowth<sub>it</sub></i>	-4,42000	5,20000	2,50000	0
<i>Unem<sub>it</sub></i>	3,78500	17,7600	3,90000	0
<i>CCI<sub>it</sub></i>	97,3211	101,798	1,79059	0
<i>Pump_Price<sub>it</sub></i>	1,0336	1,70998	0,314556	56
<i>Inflation<sub>it</sub></i>	-0,155264	4,08728	1,49668	0

Si notano immediatamente alcune sfumature sulle quali sarebbe opportuno intervenire. In primis, si evidenzia una differenza notevole negli ordini di grandezza delle medie delle diverse variabili. Inoltre, si riscontra come alcuni dati siano relativi a variazioni annuali (*GDPgrowth<sub>it</sub>*, *Inflation<sub>it</sub>*) mentre altri siano relativi a valori assoluti (*Registrations<sub>it</sub>*, *AvgPrice<sub>it</sub>*, *Unem<sub>it</sub>*, *CCI<sub>it</sub>*, *Pump\_Price<sub>it</sub>*). Per ovviare a questi problemi, la metodologia utilizzata è quella di sostituire nel modello le variabili con le rispettive variazioni percentuali. Le nuove variabili sono:

- ***Reg\_percvar<sub>it</sub>***: variazione percentuale del numero di automobili vendute, nel paese *i* al periodo *t*, rispetto al periodo *t-1*. Sostituisce la variabile *Registrations<sub>it</sub>* e viene definita secondo la formula 2.1

$$(2.1) \quad Reg\_percvar_{it} = \left( \frac{Registrations_{it}}{Registrations_{i,t-1}} - 1 \right) * 100$$





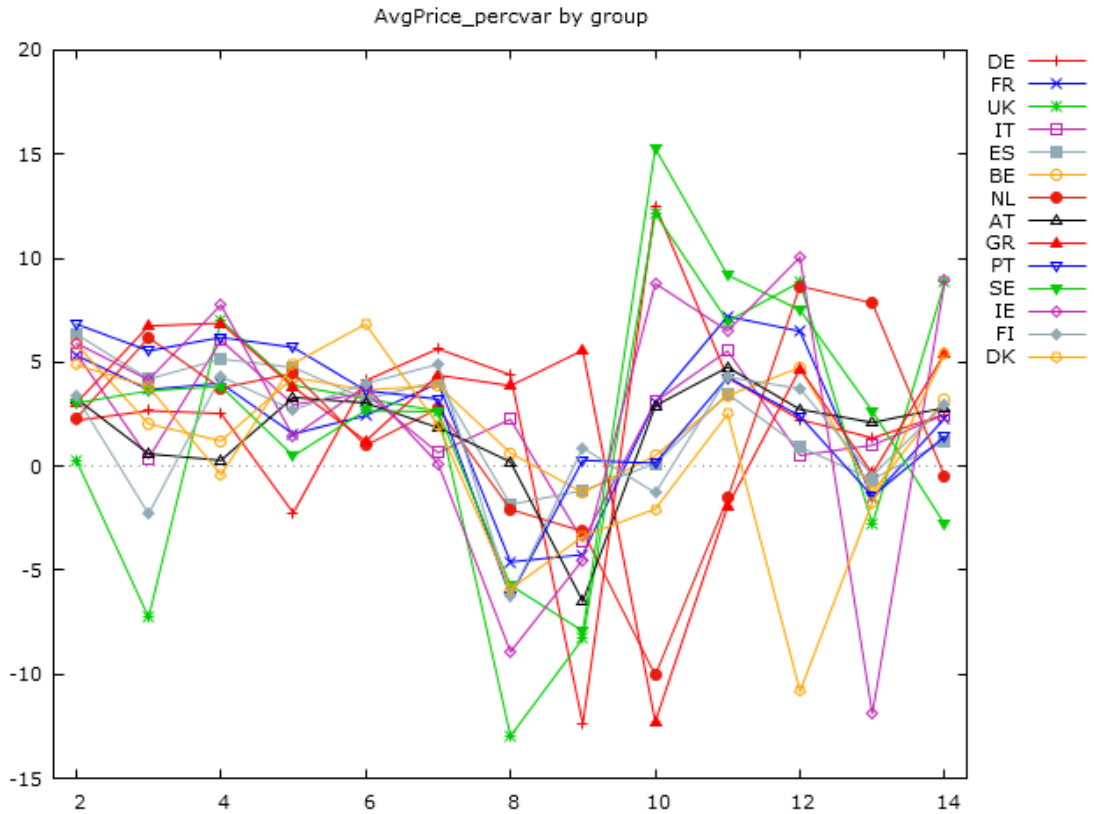
Statistiche descrittive, usando le osservazioni 1:01 - 14:14  
per la variabile Reg\_percvar (182 osservazioni valide)

Media	Mediana	Minimo	Massimo
-0,725500	-0,473981	-62,1049	53,8247
Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
14,3518	19,7820	-0,140729	3,12109
5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
-25,5147	24,5329	11,0825	14

D.s. nei gruppi                    14,510  
D.s. tra i gruppi                    3,3642

- **AvgPrice\_percvar<sub>it</sub>**: variazione percentuale del prezzo medio di vendita delle automobili, nel paese *i* al periodo *t*. Sostituisce la variabile AvgPrice<sub>it</sub> e viene definita secondo la formula 2.2

$$(2.2) \quad AvgPrice\_percvar_{it} = \left( \frac{AvgPrice_{it}}{AvgPrice_{i,t-1}} - 1 \right) * 100$$



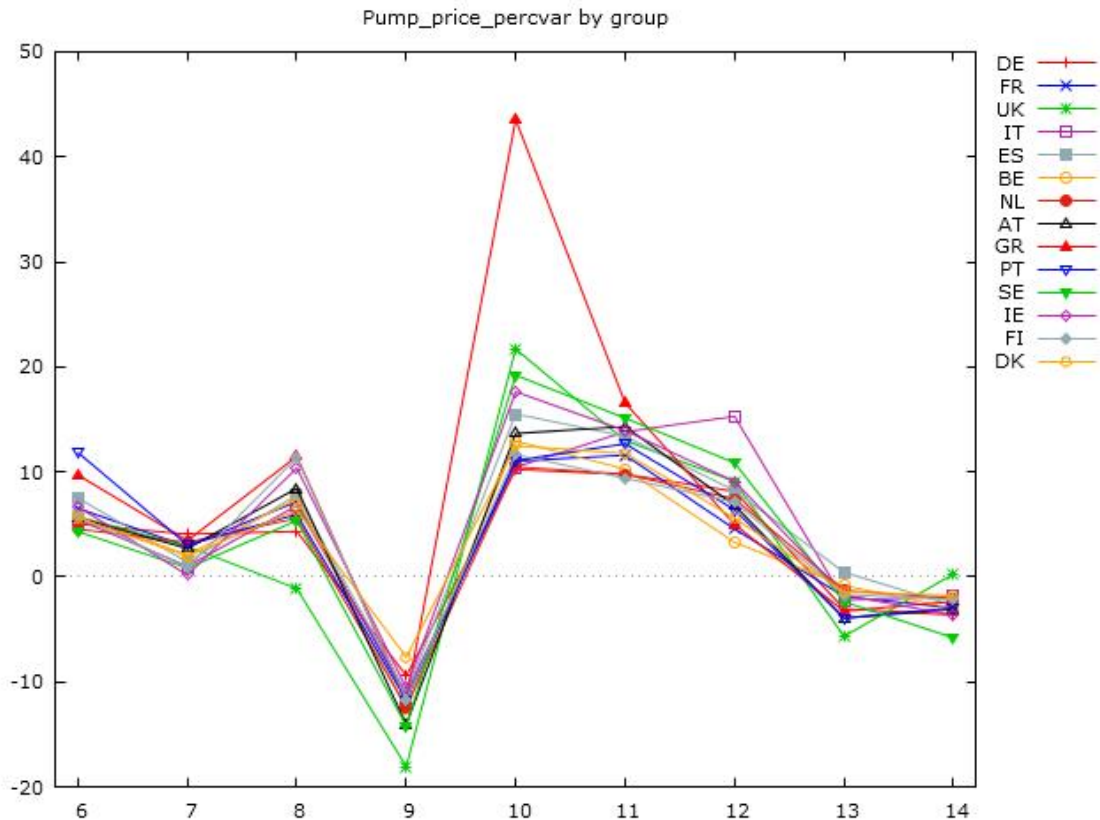
*Statistiche descrittive, usando le osservazioni 1:01 - 14:14 per la variabile AvgPrice\_percvar (182 osservazioni valide)*

Media	Mediana	Minimo	Massimo
2,03557	2,85895	-12,9657	15,2661
Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
4,62424	2,27171	-0,872207	1,64251
5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
-7,82772	8,76029	4,33169	14

D.s. nei gruppi                    4,7639  
D.s. tra i gruppi                    0,58425

- **Pump\_Price\_percvar<sub>it</sub>**: variazione percentuale del prezzo medio al consumatore finale della benzina, nel paese *i* al periodo *t*, rispetto al periodo *t-1*. Sostituisce la variabile *Pump\_Price<sub>it</sub>* e viene definita secondo la formula 2.3

$$(2.3) \text{ Pump\_Price\_percvar}_{it} = \left( \frac{\text{Pump\_Price\_Percvar}_{it}}{\text{Pump\_Price\_percvar}_{i,t-1}} - 1 \right) * 100$$



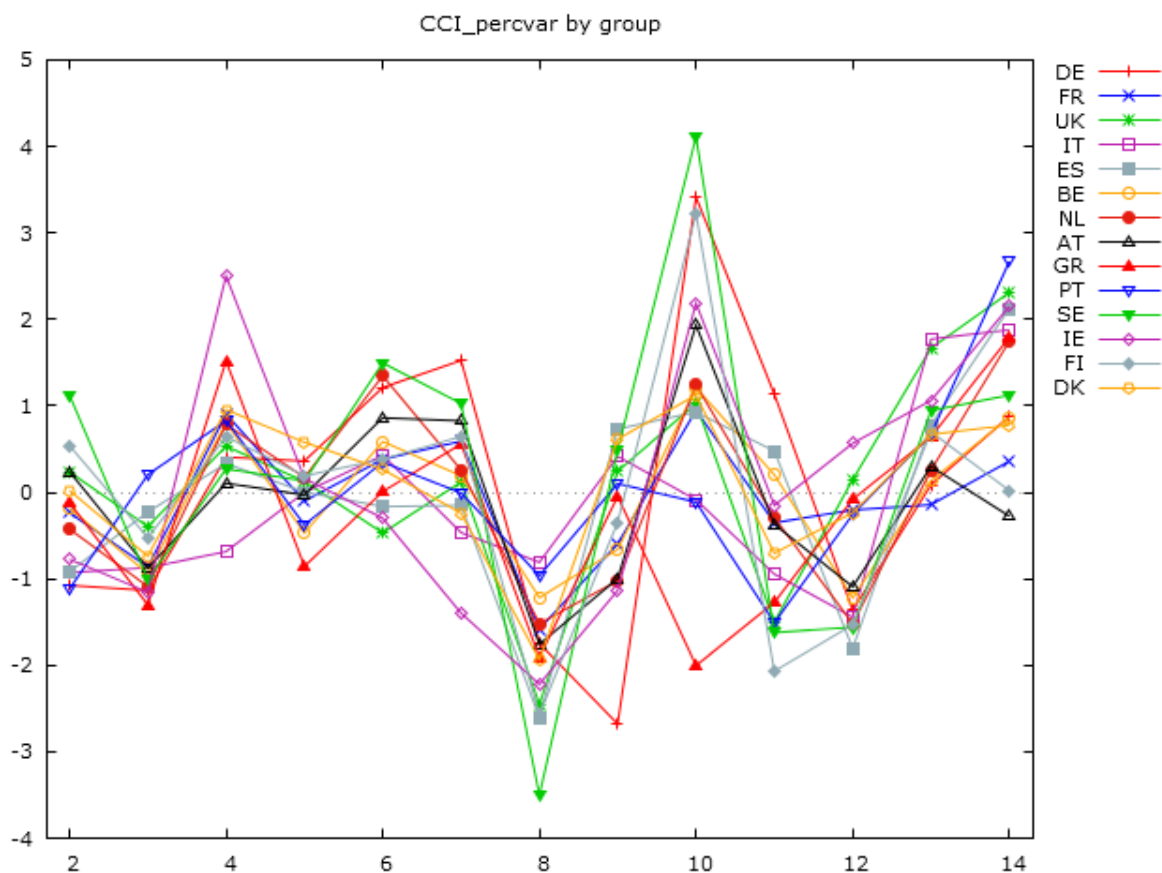
*Statistiche descrittive, usando le osservazioni 1:01 - 14:14 per la variabile Pump\_price\_percvar (126 osservazioni valide)*

Media	Mediana	Minimo	Massimo
3,83627	5,01988	-18,0577	43,5328
Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
8,79485	2,29255	0,336305	2,30535
5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
-11,7051	15,4225	11,9679	70

D.s. nei gruppi                    9,1884  
D.s. tra i gruppi                    1,3491

- ***CCI\_percvar<sub>it</sub>***: variazione percentuale del valore dell'indice CCI, nel paese *i* al periodo *t*. Sostituisce la variabile *CCI<sub>it</sub>* e viene definito secondo la formula 2.4

$$(2.4) \ CCI\_percvar_{it} = \left( \frac{CCI\_Percvar_{it}}{CCI\_percvar_{i,t-1}} - 1 \right) * 100$$



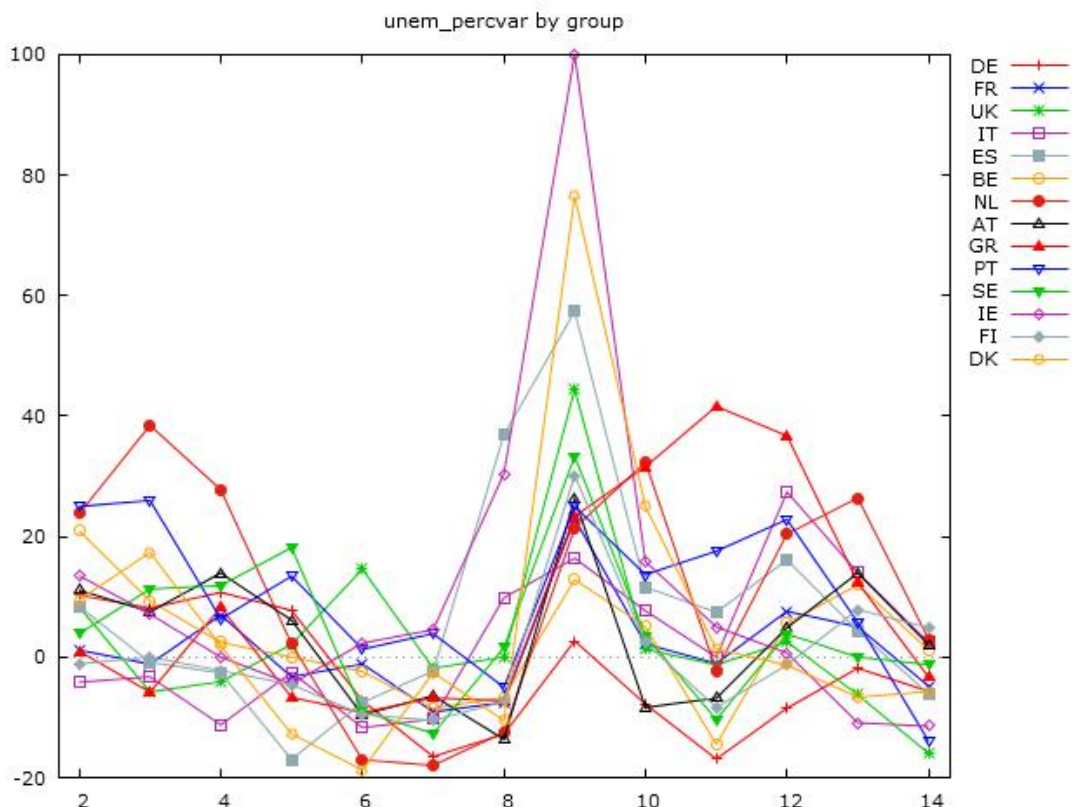
*Statistiche descrittive, usando le osservazioni 1:01 - 14:14  
per la variabile CCI\_percvar (182 osservazioni valide)*

Media	Mediana	Minimo	Massimo
-0,000693658	0,0534745	-3,50360	4,11110
Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
1,19086	1716,78	0,187609	0,742200
5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
-1,93408	2,09727	1,54405	14

D.s. nei gruppi                      1,2301  
D.s. tra i gruppi                      0,12112

- $unem\_percvar_{it}$ : variazione percentuale del valore del tasso di disoccupazione, nel paese  $i$  al periodo  $t$ . Sostituisce la variabile  $unem_{it}$  e viene definito secondo la formula 2.5

$$(2.5) \quad unem\_percvar_{it} = \left( \frac{unem\_percvar_{it}}{unem\_percvar_{i,t-1}} - 1 \right) * 100$$



*Statistiche descrittive, usando le osservazioni 1:01 - 14:14 per la variabile unem\_percvar (182 osservazioni valide)*

Media	Mediana	Minimo	Massimo
5,06876	1,94633	-18,7500	100,000
Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
16,2848	3,21277	2,02126	7,34040
5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
-13,8939	33,1863	17,5576	14

D.s. nei gruppi                    16,289  
D.s. tra i gruppi                    4,5015

Riosservando le statistiche descrittive più comuni, non si notano ulteriori problematiche su cui intervenire o perfezionamenti da apportare.

Statistiche descrittive, usando le osservazioni 1:01 - 14:14  
(i valori mancanti sono stati saltati)

Variabile	Media	Mediana	Minimo	Massimo
<i>Reg_percvar<sub>it</sub></i>	-0,725500	-0,473981	-62,1049	53,8247
<i>AvgPrice_percvar<sub>it</sub></i>	2,03557	2,85895	-12,9657	15,2661
<i>GDPgrowth<sub>it</sub></i>	1,12500	1,55000	-9,10000	6,30000
<i>Unem_percvar<sub>it</sub></i>	5,0688	1,9463	-18,750	100,00
<i>CCI_percvar<sub>it</sub></i>	-0,00069366	0,053475	-3,5036	4,1111
<i>Pump_price_percvar<sub>it</sub></i>	3,85105	5,01988	-18,0577	43,5328
<i>Inflation<sub>it</sub></i>	2,00476	2,08301	-4,48186	4,91876

Variabile	Dev. Std.	Coeff. di variazione	Asimmetria	Curtosi
<i>Reg_percvar<sub>it</sub></i>	14,3518	19,7820	-0,140729	3,12109
<i>AvgPrice_percvar<sub>it</sub></i>	4,62424	2,27171	-0,872207	1,64251
<i>GDPgrowth<sub>it</sub></i>	2,67849	2,38088	-1,03327	1,78082
<i>Unem_percvar<sub>it</sub></i>	16,285	3,2128	2,0213	7,3404
<i>CCI_percvar<sub>it</sub></i>	1,1909	1716,8	0,18761	0,74220
<i>Pump_price_percvar<sub>it</sub></i>	8,95740	2,32596	0,284686	2,15373
<i>Inflation<sub>it</sub></i>	1,29271	0,644821	-0,678532	2,52831

Variabile	5% Perc.	95% Perc.	Range interquartile	Osservazioni mancanti
<i>Reg_percvar<sub>it</sub></i>	-25,5147	24,5329	11,0825	14
<i>AvgPrice_percvar<sub>it</sub></i>	-7,82772	8,76029	4,33169	14
<i>GDPgrowth<sub>it</sub></i>	-4,42000	5,20000	2,50000	0
<i>Unem_percvar<sub>it</sub></i>	-13,894	33,186	17,558	14
<i>CCI_percvar<sub>it</sub></i>	-1,9341	2,0973	1,5441	14
<i>Pump_price_percvar<sub>it</sub></i>	-12,1582	16,1886	11,9679	70
<i>Inflation<sub>it</sub></i>	-0,155264	4,08728	1,49668	0

Si nota dall'ultima colonna che sono assenti alcune delle osservazioni. In particolare, le 14 osservazioni mancanti per le variabili *Reg\_percvar<sub>it</sub>* , *AvgPrice\_percvar<sub>it</sub>* , *Unem\_percvar<sub>it</sub>* , *CCI\_percvar<sub>it</sub>* sono ovviamente dovute alla loro natura, in quanto partendo da T=14 osservazioni ed effettuando il calcolo delle variazioni percentuali, si ottengono 13 valori. Pertanto, per ogni paese, si avrà  $t 1 \rightarrow T=13$  con dati relativi ai periodi dal 2002 al 2014.

Per quanto riguarda le 70 osservazioni mancanti per la variabile *Pump\_price\_percvar<sub>it</sub>*, esse sono legate a due situazioni differenti: 56 osservazioni mancano perché i dati relativi ai periodi dal 2001 al 2004 non sono recuperabili agevolmente in alcun database accessibile, mentre 14 ulteriori osservazioni sono state perse effettuando la procedura di calcolo delle variazioni percentuali. La stima dei modelli si può effettuare anche nel caso di dati panel *unbalanced*, ovvero non contenenti tutte le osservazioni per tutte le unità statistiche in tutti i periodi. Tuttavia, nel caso dell'analisi presentata in questo elaborato, la scelta effettuata è quella di preferire l'elaborazione di un panel dati bilanciato. Pertanto, per ogni paese, si avrà  $t 1 \rightarrow T=9$ .

Tenendo conto di queste limitazioni, e della conseguente riduzione dell'utilizzo del dataset ai soli dati

relativi al periodo dal 2005 al 2014 per una stima agevole dei parametri, si può procedere alla stima dei modelli.

# STIMA DEI MODELLI

Si procede alla stima dei modelli mediante stimatore pooled OLS, ad effetti fissi e ad effetti casuali. Le stime sono ora riportate per esteso singolarmente, ma gli elementi essenziali sono riportati poi per comodità in una tabella di sintesi.

## Modello 1: pooled OLS

Modello 1: Pooled OLS, usando 126 osservazioni  
 Incluse 14 unità cross section  
 Lunghezza serie storiche = 9  
 Variabile dipendente: *Reg\_percvar*

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>rapporto t</i>	<i>p-value</i>	
<i>const</i>	3,99253	2,39998	1,6636	0,0988	*
<i>AvgPrice_percvar</i>	-0,765424	0,255127	-3,0002	0,0033	***
<i>GDPgrowth</i>	1,85672	0,602779	3,0803	0,0026	***
<i>CCI_percvar</i>	2,69471	1,16763	2,3078	0,0227	**
<i>Pump_price_percvar</i>	0,272753	0,165591	1,6472	0,1022	
<i>inflation</i>	-2,75287	1,22894	-2,2400	0,0269	**
<i>unem_percvar</i>	-0,302647	0,086689	-3,4912	0,0007	***

Media var. dipendente	-1,544667	SQM var. dipendente	16,32354
Somma quadr. residui	18454,54	E.S. della regressione	12,45312
R-quadro	0,445931	R-quadro corretto	0,417994
F(6, 119)	15,96242	P-value(F)	2,13e-13
Log-verosimiglianza	-492,9536	Criterio di Akaike	999,9073
Criterio di Schwarz	1019,761	Hannan-Quinn	1007,973
rho	-0,176965	Durbin-Watson	2,232968

Si nota immediatamente come alcune variabili non siano significative mentre altre lo siano. Si ricordi la decisione di stabilire il livello di significatività a 0,1. Le variabili significative sono quindi *AvgPrice\_percvar*, *GDPgrowth*, *CCI\_percvar*, *inflation*, *unem\_percvar*. Le variabili non significative sono invece *Pump\_price\_percvar* e *const*.

Concentrando l'attenzione sulle variabili significative, si può vedere come le stime dei parametri sembrano confermare le aspettative iniziali. Infatti, è ragionevole aspettarsi che, all'aumentare di *AvgPrice\_percvar*, *inflation* e *unem\_percvar* l'effetto sulle *registrations* sia negativo (anche l'effetto di *Pump\_price\_percvar* dovrebbe essere analogo). I dati confermano l'aspettativa. Inoltre, è ragionevole aspettarsi che, all'aumentare di *GDPgrowth*, *CCI\_percvar* l'effetto su *registrations* sia positivo. I dati confermano anche questa aspettativa.

Il test di White per l'eteroschedasticità evidenzia tuttavia problemi in tal senso:

Test di White per l'eteroschedasticità -  
 Ipotesi nulla: eteroschedasticità non presente  
 Statistica test: LM = 69,8902  
 con p-value = P(Chi-quadro(27) > 69,8902) = 1,16287e-005



Effettuando il test Chow per verificare la presenza di break strutturali in concomitanza dell'anno 2008 si ottiene il seguente output, che conferma la presenza di un break strutturale:

Test Chow per differenza strutturale rispetto a d -  
 Ipotesi nulla: nessuna differenza strutturale  
 Statistica test:  $F(7, 112) = 2,0121$   
 con p-value =  $P(F(7, 112) > 2,0121) = 0,0596196$

Si valuta infine la forma funzionale tramite il test RESET di Ramsey:

Test RESET di specificazione (solo quadrati) -  
 Ipotesi nulla: la specificazione è adeguata  
 Statistica test:  $F(1, 118) = 8,34275$   
 con p-value =  $P(F(1, 118) > 8,34275) = 0,00460814$

Il p-value della statistica test di Ramsey è molto basso, portando a rigettare l'ipotesi nulla di adeguata specificazione del modello come modello lineare.

## Modello 2: ad effetti fissi

Modello 2: Effetti fissi, usando 126 osservazioni  
 Include 14 unità cross section  
 Lunghezza serie storiche = 9  
 Variabile dipendente: *Reg\_percvar*

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>rapporto t</i>	<i>p-value</i>	
<i>const</i>	5,04255	2,6703	1,8884	0,0617	*
<i>AvgPrice_percvar</i>	-0,708803	0,269166	-2,6333	0,0097	***
<i>GDPgrowth</i>	1,70846	0,71422	2,3921	0,0185	**
<i>CCI_percvar</i>	2,39066	1,27179	1,8798	0,0629	*
<i>Pump_price_percvar</i>	0,316832	0,179883	1,7613	0,0811	*
<i>inflation</i>	-3,30865	1,40532	-2,3544	0,0204	**
<i>unem_percvar</i>	-0,339485	0,0990803	-3,4264	0,0009	***

Media var. dipendente	-1,544667	SQM var. dipendente	16,32354
Somma quadr. residui	17707,27	E.S. della regressione	12,92477
R-quadro LSDV	0,468366	R-quadro intra-gruppi	0,439613
LSDV F(19, 106)	4,915019	P-value(F)	4,30e-08
Log-verosimiglianza	-490,3495	Criterio di Akaike	1020,699
Criterio di Schwarz	1077,425	Hannan-Quinn	1043,745
rho	-0,221128	Durbin-Watson	2,321077

Si ricordi la decisione di stabilire il livello di significatività a 0,1. Le variabili significative sono *AvgPrice\_percvar*, *GDPgrowth*, *inflation*, *unem\_percvar*, *Pump\_price\_percvar*, *CCI\_percvar* e *const*.

Concentrando l'attenzione sul valore dei coefficienti, si può vedere come le stime dei parametri sembrano confermare le aspettative iniziali. Infatti, è ragionevole aspettarsi che, all'aumentare di *AvgPrice\_percvar*, *inflation*, *CCI\_percvar* e *unem\_percvar* l'effetto sulle *registrations* sia negativo (anche l'effetto di *Pump\_price\_percvar* dovrebbe essere analogo). I dati confermano l'aspettativa.

Inoltre, è ragionevole aspettarsi che, all'aumentare di *GDPgrowth* l'effetto su *registrations* sia positivo. I dati confermano anche questa aspettativa.

Test congiunto sui regressori -  
 Statistica test:  $F(6, 106) = 13,8592$   
 con p-value =  $P(F(6, 106) > 13,8592) = 1,40675e-011$

Il test congiunto sui regressori è un F-test che valuta se tutti i coefficienti nel modello sono differenti da 0. In questo caso fornisce un valore del p-value molto basso, indicando quindi che i coefficienti sono differenti da 0.

Test non-parametrico di Wald per l'eteroschedasticità -  
 Ipotesi nulla: le unità hanno in comune la varianza dell'errore  
 Statistica test asintotica:  $\text{Chi-quadro}(14) = 363,148$   
 con p-value =  $7,15995e-069$

Si rifiuta l'ipotesi  $H_0$ , è pertanto presente eteroschedasticità.

### Modello 3: ad effetti casuali

Modello 3: Effetti casuali (GLS), usando 126 osservazioni  
 Include 14 unità cross section  
 Lunghezza serie storiche = 9  
 Variabile dipendente: *Reg\_percvar*

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>z</i>	<i>p-value</i>	
<i>const</i>	3,99253	2,39998	1,6636	0,0962	*
<i>AvgPrice_percvar</i>	-0,765424	0,255127	-3,0002	0,0027	***
<i>GDPgrowth</i>	1,85672	0,602779	3,0803	0,0021	***
<i>CCI_percvar</i>	2,69471	1,16763	2,3078	0,0210	**
<i>Pump_price_percvar</i>	0,272753	0,165591	1,6472	0,0995	*
<i>inflation</i>	-2,75287	1,22894	-2,2400	0,0251	**
<i>unem_percvar</i>	-0,302647	0,086689	-3,4912	0,0005	***

Media var. dipendente	-1,544667	SQM var. dipendente	16,32354
Somma quadr. residui	18454,54	E.S. della regressione	12,40112
Log-verosimiglianza	-492,9536	Criterio di Akaike	999,9073
Criterio di Schwarz	1019,761	Hannan-Quinn	1007,973

Varianza 'between' = 0  
 Varianza 'within' = 167,05  
 Theta usato per la trasformazione = 0

Si ricordi la decisione di stabilire il livello di significatività a 0,1. Le variabili significative sono *AvgPrice\_percvar*, *GDPgrowth*, *CCI\_percvar*, *inflation*, *unem\_percvar*, *Pump\_price\_percvar*, e *const*.

Concentrando l'attenzione sul valore delle stime dei coefficienti, si può vedere come anche in questo caso le stime dei parametri sembrano confermare le aspettative iniziali. Infatti, è ragionevole aspettarsi che, all'aumentare di *AvgPrice\_percvar*, *inflation*, *CCI\_percvar* e *unem\_percvar* l'effetto sulle *registrations* sia negativo (anche l'effetto di *Pump\_price\_percvar* dovrebbe essere analogo). Ed

anche con questo modello i dati confermano l'aspettativa. Inoltre, è ragionevole aspettarsi che, all'aumentare di *GDPgrowth* l'effetto su *registrations* sia positivo. I dati confermano anche questa aspettativa.

Test congiunto sui regressori -  
Statistica test asintotica: Chi-quadro(6) = 95,7745  
con p-value = 1,9071e-018

Il test congiunto sui regressori mostra come anche in questo modello non vi siano problemi relativi a parametri nulli.

Si ricorda che il test Breusch-Pagan testa l'ipotesi nulla che la varianza di  $\alpha_i$  sia nulla. Se l'ipotesi nulla è rifiutata si può concludere che c'è un significativo effetto casuale nel dataset e che il modello ad effetti casuali riesca a conciliarsi meglio con l'eterogeneità rispetto al modello pooled OLS.

Test Breusch-Pagan -  
Ipotesi nulla: varianza dell'errore specifico all'unità = 0  
Statistica test asintotica: Chi-quadro(1) = 3,45289  
con p-value = 0,0631413

Anche se non nettamente, il test favorisce il modello pooled OLS rispetto all'alternativa ad effetti casuali.

Effettuando il test di Hausman per valutare se sia migliore l'adattamento del modello ad effetti fissi o casuali, si ottiene il seguente output:

Test di Hausman -  
Ipotesi nulla: le stime GLS sono consistenti  
Statistica test asintotica: Chi-quadro(6) = 3,31273  
con p-value = 0,768695

L'elevato valore del p-value indica che lo stimatore GLS sia il più indicato tra i due modelli (FE e RE) a descrivere il modello.

## SINTESI DEI RISULTATI

Si valutano ora i risultati ottenuti riordinandoli in una tabella riassuntiva, inserendo le stime dei parametri e i valori dei p-value per poter poi analizzare la situazione:

	Pooled OLS		Fixed Effects		Random Effects	
	Stima	p-value	Stima	p-value	Stima	p-value
<i>const</i>	3,99253	0,0988 *	5,04255	0,0617 *	3,99253	0,0962 *
<i>AvgPrice_percvar</i>	-0,765424	0,0033 ***	-0,70880 3	0,0097 ***	-0,76542	0,0027 ***
<i>GDPgrowth</i>	1,85672	0,0026 ***	1,70846	0,0185 **	1,85672	0,0021 ***
<i>CCI_percvar</i>	2,69471	0,0227 **	2,39066	0,0629 *	2,69471	0,0210 **
<i>Pump_price_percvar</i>	0,272753	0,1022	0,316832	0,0811 *	0,272753	0,0995 *
<i>inflation</i>	-2,75287	0,0269 **	-3,30865	0,0204 **	-2,75287	0,0251 **
<i>unem_percvar</i>	-0,302647	0,0007 ***	-0,33948 5	0,0009 ***	-0,30264	0,0005 ***
<i>Test Reset (p-value)</i>	0,00460814					
<i>Test White (p-value)</i>	1,16287e-005					
<i>Test Wald (p-value)</i>			7,15995e-069			
<i>Test Chow (p-value)</i>	0,0596196					
<i>Test Breusch-Pagan (p-value)</i>					0,0631413	
<i>Test Hausman (p-value)</i>					0,768695	

Dalla tabella riassuntiva si nota come le stime dei parametri siano identiche per il modello pooled OLS e per il modello ad effetti casuali. Ciò è insolito, ma si potrebbe spiegare con i dati relativi alla scomposizione della varianza forniti dal software. In particolare, la varianza “*between*” è nulla, indicando una situazione in cui apparentemente non vi è varianza tra le medie dei gruppi. La conseguenza di questa situazione si sostanzia nel Theta usato nella trasformazione del modello ad effetti casuali che assume valore nullo, annullando tutti gli effetti della trasformazione stessa.

In ogni caso, le stime effettuate sono tutte affette da problematiche di eteroschedasticità (test White e Wald) e non corretta specificazione della forma funzionale (test RESET). Pertanto è necessario intervenire e correggere tali questioni.

I test tra i differenti modelli portano a favorire il modello pooled OLS rispetto alle due alternative FE e RE. Si vedrà in seguito se tale situazione sarà mantenuta anche con il dataset modificato.

## REVISIONE DATASET

Dall'osservazione delle stime riportate nella sezione precedente si possono trarre alcune prime conclusioni preliminari, che possono portare a degli interventi correttivi effettuabili sul dataset nell'ottica di miglioramento dei risultati.

I modelli sopra presentati sono, infatti, da considerarsi *misspecified*, e il basso p-value del test RESET ne fornisce la conferma. Per risolvere tale problema è opportuno effettuare una selezione più accurata delle variabili da inserire. In particolare, il risultato più evidente riscontrato in tutte e tre le tipologie di modelli è la non significatività dei parametri relativi alla variabile *Pump\_Price\_percvar*. Pertanto, può essere interessante effettuare delle nuove stime del modello rimuovendo tale variabile dal dataset. Il procedimento che verrà utilizzato per la selezione delle variabili è il metodo *step-wise* nella variante *backward elimination*. In particolare, si partirà con una prima stima del modello con tutte le variabili possibili, per poi effettuare un test sull'eliminazione di ogni singola variabile usando un criterio di comparazione del modello scelto, ed eliminando successivamente la variabile la cui eliminazione migliora il modello in misura maggiore. Tale processo si ripete fintanto che non sia possibile alcun ulteriore miglioramento. L'intero procedimento è riportato nell'appendice, mentre nella prossima sezione verrà riportato l'output finale della procedura.

Un'ulteriore problematica che si riscontra dalle stime effettuate è l'esistenza di un break strutturale, come confermato dal test Chow. Si inserirà quindi nel modello la variabile dummy *d* utilizzata in precedenza solamente per indicare al software in concomitanza di quale osservazione ricercare il punto di break strutturale. Inserendo nel modello la variabile dummy e i termini di interazione tra essa e le altre variabili esplicative, si dovrebbero ottenere dei risultati migliori.

Infine, dal momento che si evidenziano problemi di eteroschedasticità, è opportuno stimare il modello pooled OLS HAC (Heteroskedasticity and Autocorrelation Consistent). Senza scendere nei dettagli tecnici avanzati, è sufficiente dire che si stima la matrice della covarianza in maniera tale da eliminare autocorrelazione ed eteroschedasticità. Pertanto l'output del modello sarà esente da tali problematiche. (si veda Wooldridge, 2006, p. 429).

Si procederà ora in maniera più spedita nell'esposizione degli output. In particolare, si effettuerà il procedimento step-wise utilizzando lo stimatore pooled OLS HAC, ed infine si effettueranno i test tra modelli (Breusch-Pagan, Hausman) per valutare le alternative ad effetti casuali e ad effetti fissi.

### Modello 4: pooled OLS HAC

La procedura step-wise ha portato al seguente output:

Modello 4: Pooled OLS, usando 182 osservazioni  
Include 14 unità cross section  
Lunghezza serie storiche = 13  
Variabile dipendente: *Reg\_percvar*  
Errori standard robusti (HAC)

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>rapporto t</i>	<i>p-value</i>	
<i>const</i>	1,66312	0,893438	1,8615	0,0854	*
<i>AvgPrice_percvar</i>	-0,735324	0,362677	-2,0275	0,0636	*
<i>unem_percvar</i>	-0,190743	0,0755115	-2,5260	0,0253	**

<i>CCI_percvar</i>	3,5285	1,14835	3,0727	0,0089	***
<i>GDPgrowth</i>	0,47711	0,256712	1,8585	0,0859	*
<i>int_d_GDPgr</i>	2,56161	0,866037	2,9579	0,0111	**

Media var. dipendente	-0,725500	SQM var. dipendente	14,35180
Somma quadr. residui	20477,99	E.S. della regressione	10,78667
R-quadro	0,450718	R-quadro corretto	0,435113
F(5, 13)	23,11759	P-value(F)	4,80e-06
Log-verosimiglianza	-688,0488	Criterio di Akaike	1388,098
Criterio di Schwarz	1407,322	Hannan-Quinn	1395,891
rho	-0,185736	Durbin-Watson	2,268367

Test RESET di specificazione -  
 Ipotesi nulla: la specificazione è adeguata  
 Statistica test:  $F(2, 174) = 3,59762$   
 con p-value =  $P(F(2, 174) > 3,59762) = 0,294452$

Test Chow per break strutturale all'osservazione 7:14 -  
 Ipotesi nulla: nessun break strutturale  
 Statistica test asintotica:  $\text{Chi-quadro}(6) = 22,0917$   
 con p-value = 0,116543

Test RESET e test Chow restituiscono risultati migliori rispetto ai modelli precedenti. Si effettuano dunque i test tra i modelli pooled OLS HAC, ad effetti fissi e ad effetti casuali, per verificare quale di essi sia il più adatto.

Diagnosi: ipotesi di panel bilanciato con 14 unità cross section  
 osservate per 13 periodi

### Stimatore a effetti fissi

Consente intercette diverse per ogni unità cross section

	<i>Coefficiente</i>	<i>Errore standard</i>	<i>P-value</i>
<i>Const</i>	1,7279	1,1975	0,15096
<i>AvgPrice_percvar</i>	-0,70307	0,20586	0,00080
<i>Unem_percvar</i>	-0,20165	0,072279	0,00590
<i>CCI_percvar</i>	3,4335	0,74262	0,00001
<i>GDPgrowth</i>	0,41713	0,56498	0,46139
<i>Int_d_GDPgr</i>	2,6449	0,7193	0,00032

14 medie di gruppo sono state sottratte dai dati

Varianza dei residui:  $19643/(182 - 19) = 120,509$   
 Significatività congiunta delle differenti medie dei gruppi:  
 $F(13, 163) = 0,533014$  con p-value 0,90151  
 (un basso p-value conta contro l'ipotesi nulla che il modello pooled OLS  
 sia adeguato, in favore del modello alternativo con effetti fissi)

Statistica test di Breusch-Pagan:

LM = 1,73175 con p-value =  $\text{prob}(\text{chi-quadro}(1) > 1,73175) = 0,188188$   
 (un basso p-value conta contro l'ipotesi nulla che il modello pooled OLS  
 sia adeguato, in favore del modello alternativo con effetti casuali)

Variance estimators:  
 between = 0  
 within = 120,509  
 theta used for quasi-demeaning = 0

### Stimatore a effetti casuali

Consente che il termine di errore sia specifico per ogni unità

	<i>Coefficiente</i>	<i>Errore standard</i>	<i>P-value</i>
<i>Const</i>	1,6631	1,1272	0,14189
<i>AvgPrice_percvar</i>	-0,73532	0,19736	0,00026
<i>Unem_percvar</i>	-0,19074	0,064981	0,00378
<i>CCI_percvar</i>	3,5285	0,72723	0,00000
<i>GDPgrowth</i>	0,47711	0,51149	0,35221
<i>Int_d_GDPgr</i>	2,5616	0,65817	0,00014

Statistica test di Hausman:

H = 5,10475 con p-value =  $\text{prob}(\text{chi-quadro}(5) > 5,10475) = 0,403231$   
 (un basso p-value conta contro l'ipotesi nulla che il modello con coefficienti  
 casuali sia adeguato, in favore del modello con effetti fissi)

Dai risultati dei test si evince che il miglior modello sia quello stimato mediante stimatore pooled OLS HAC.

# RISULTATI DELL'ANALISI

È utile confrontare in un'unica tabella i risultati ottenuti:

Variabile dipendente: *Reg\_percvar*

	(1)	(2)	(3)
	Pooled OLS HAC	Effetti Fissi	Effetti casuali
<i>const</i>	1,663* (0,8934)	1,728 (1,197)	1,663 (1,127)
<i>AvgPrice_percvar</i>	-0,7353* (0,3627)	-0,7031** (0,2059)	-0,7353** (0,1974)
<i>unem_percvar</i>	-0,1907** (0,07551)	-0,2017** (0,07228)	-0,1907** (0,06498)
<i>CCI_percvar</i>	3,529** (1,148)	3,434** (0,7426)	3,529** (0,7272)
<i>GDPgrowth</i>	0,4771* (0,2567)	0,4171 (0,5650)	0,4771 (0,5115)
<i>int_d_GDPgr</i>	2,562** (0,8660)	2,645** (0,7193)	2,562** (0,6582)
<i>n</i>	182	182	182

Errori standard tra parentesi

\* indica significatività al livello del 10 per cento

\*\* indica significatività al livello del 5 per cento

Il modello finale da utilizzare nell'interpretazione dei risultati è il modello 4, stimato mediante stimatore pooled OLS HAC, in quanto modello più solido tra le sei alternative proposte. Tale modello non presenta le problematiche relative all'eteroschedasticità degli errori, alla misspecification e alla presenza di break strutturale marcato che sono invece presenti negli altri modelli. L'eteroschedasticità degli errori è sanata dalla scelta stessa dello stimatore HAC, la misspecification e la presenza di break strutturale sono stati contenuti effettuando la procedura di regressione step-wise by backward elimination dopo aver inserito la variabile dummy *d* (valore 0 fino al 2007, valore 1 dal 2008) ed i termini di interazione tra essa e le altre variabili esplicative.



Si valutano quindi i risultati ottenuti e si rapportano alle aspettative iniziali descritte nella sezione introduttiva e nella sezione relativa alla scelta delle variabili.

Le variabili inizialmente inserite nei modelli, che però si sono rivelate non significative, sono:

- *Pump\_price\_percvar*: il prezzo medio della benzina si è rivelato ininfluenza sul numero di automobili immatricolate, nonostante si potesse immaginare che avesse una qualche influenza su di esso. Pertanto, stando ai dati sperimentali, l'elasticità della domanda di automobili nuove rispetto al prezzo della benzina pagato dal consumatore è nulla.
- *Inflation*: l'inflazione era una delle variabili che probabilmente ci si aspettava rimanesse nel modello finale. I dati sperimentali portano, però, alla conclusione contraria.

Le variabili significative sono:

- *AvgPrice\_percvar*: la variazione del prezzo medio di vendita delle automobili è variabile significativa, la cui stima del parametro indica una relazione inversa con il numero di automobili vendute. Pertanto la domanda di automobili rispetto al prezzo delle stesse è elastica. In particolare, un aumento di una unità della variazione percentuale del prezzo medio delle automobili genera una variazione percentuale del -0,7353% nella domanda di automobili.
- *unem\_percvar*: la variazione percentuale del tasso di disoccupazione è variabile significativa. All'aumentare del tasso di disoccupazione si evidenzia una diminuzione della domanda di automobili. Ciò è prevedibile, in quanto laddove vi sia un tasso di disoccupazione più elevato vi dovrebbero essere vendite di automobili più ridotte, in quanto il mercato di destinazione verrebbe ridotto di un numero di consumatori uguale al numero di nuovi disoccupati. Inoltre, sembrerebbe perfettamente sensato affermare che all'aumentare del tasso di disoccupazione il numero di automobili debba diminuire, in quanto, semplicemente, i cittadini che non lavorano non possono comprare nuove automobili. Si è evidenziato quindi che la domanda di automobili è influenzata dalla variazione del tasso di disoccupazione.
- *CCI\_percvar*: la variazione percentuale dell'indice di fiducia dei consumatori è variabile significativa. All'aumentare di una unità della variazione percentuale year-on-year dell'indice CCI si prevede un aumento percentuale del numero di automobili vendute pari al +3,529%. L'evidenza empirica conforta la tesi iniziale di utilità di un indicatore della fiducia dei consumatori nel modello previsionale relativo alle vendite di automobili.
- *GDPgrowth*: il tasso annuale di crescita del PIL è variabile significativa. In particolare, ad un suo aumento di una unità si associa una previsione di crescita percentuale delle vendite di automobili pari al +0,4771%. Questa relazione positiva è prevedibile in quanto un tasso di crescita del PIL positivo di un Paese indica generalmente un miglioramento delle condizioni dell'intero suo sistema economico, portando maggiori opportunità di acquisto ai consumatori.
- *int\_d\_GDPgr*: il termine di interazione tra *d*, variabile dummy relativa all'influenza della crisi economica del 2008 sulle altre variabili, e *GDPgrowth* è significativo. Pertanto si riconosce l'influenza della crisi economica sulle variazioni annuali della crescita del PIL.



# APPENDICE

Sono riportati, in basso, gli output dei singoli passaggi della procedura di regressione step-wise by backward elimination effettuata usando lo stimatore pooled OLS HAC. Per ogni passaggio è stato effettuato il test sull'eliminazione di ogni singola variabile valutando quale variabile fornisce il maggior miglioramento del modello una volta eliminata. Inoltre, in ogni passaggio sono stati effettuati i test RESET e Chow per valutare la presenza di problematiche relative a misspecification o presenza di break strutturali.

Modello A1: Pooled OLS, usando 126 osservazioni

Incluse 14 unità cross section

Lunghezza serie storiche = 9

Variabile dipendente: *Reg\_percvar*

Errori standard robusti (HAC)

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>rapporto t</i>	<i>p-value</i>
<i>const</i>	7,92987	4,065	1,9508	0,0730
<i>AvgPrice_percvar</i>	-0,83675	0,814882	-1,0268	0,3232
<i>Pump_price_percvar</i>	-0,121362	0,429465	-0,2826	0,7819
<i>unem_percvar</i>	-0,153385	0,145514	-1,0541	0,3110
<i>CCI_percvar</i>	-2,32037	2,33586	-0,9934	0,3387
<i>inflation</i>	-1,08817	1,82484	-0,5963	0,5612
<i>d</i>	-2,35048	2,73681	-0,8588	0,4060
<i>GDPgrowth</i>	-0,383759	1,12698	-0,3405	0,7389
<i>int_d_AvgPrice</i>	0,0302953	0,862123	0,0351	0,9725
<i>int_d_unem</i>	-0,152634	0,26856	-0,5683	0,5795
<i>int_d_inflation</i>	-1,50976	1,23093	-1,2265	0,2418
<i>int_d_cci</i>	5,23942	2,49271	2,1019	0,0556
<i>int_d_GDPgr</i>	3,01453	1,84129	1,6372	0,1256
<i>int_d_pump</i>	0,330407	0,460242	0,7179	0,4855

Modello A2: Pooled OLS, usando 126 osservazioni

Incluse 14 unità cross section

Lunghezza serie storiche = 9

Variabile dipendente: *Reg\_percvar*

Errori standard robusti (HAC)

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>rapporto t</i>	<i>p-value</i>
<i>const</i>	7,86633	3,9513	1,9908	0,0679
<i>AvgPrice_percvar</i>	-0,807091	0,40058	-2,0148	0,0651
<i>Pump_price_percvar</i>	-0,123228	0,416701	-0,2957	0,7721
<i>unem_percvar</i>	-0,152844	0,147944	-1,0331	0,3204
<i>CCI_percvar</i>	-2,34267	2,04958	-1,1430	0,2737
<i>inflation</i>	-1,09432	1,75471	-0,6236	0,5436
<i>d</i>	-2,28625	2,30932	-0,9900	0,3402
<i>GDPgrowth</i>	-0,383014	1,11506	-0,3435	0,7367
<i>int_d_unem</i>	-0,153182	0,271531	-0,5641	0,5823

<i>int_d_inflation</i>	-1,50375	1,14789	-1,3100	0,2129
<i>int_d_cci</i>	5,26233	2,21346	2,3774	0,0335
<i>int_d_GDPgr</i>	3,01411	1,83295	1,6444	0,1240
<i>int_d_pump</i>	0,33236	0,45324	0,7333	0,4764

Modello A3: Pooled OLS, usando 182 osservazioni

Incluse 14 unità cross section

Lunghezza serie storiche = 13

Variabile dipendente: *Reg\_percvar*

Errori standard robusti (HAC)

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>rapporto t</i>	<i>p-value</i>
<i>const</i>	5,32323	2,58943	2,0557	0,0605
<i>AvgPrice_percvar</i>	-0,806778	0,347637	-2,3207	0,0372
<i>unem_percvar</i>	-0,0741729	0,0601259	-1,2336	0,2392
<i>CCI_percvar</i>	1,87799	1,3444	1,3969	0,1858
<i>inflation</i>	-1,2299	0,594767	-2,0679	0,0592
<i>d</i>	0,256508	3,31793	0,0773	0,9396
<i>GDPgrowth</i>	0,457178	0,642146	0,7120	0,4891
<i>int_d_unem</i>	-0,23185	0,18089	-1,2817	0,2223
<i>int_d_inflation</i>	-1,3681	1,26168	-1,0843	0,2979
<i>int_d_cci</i>	1,04137	1,80848	0,5758	0,5746
<i>int_d_GDPgr</i>	2,17376	1,46554	1,4832	0,1618
<i>int_d_pump</i>	0,209089	0,206156	1,0142	0,3290

Modello A4: Pooled OLS, usando 182 osservazioni

Incluse 14 unità cross section

Lunghezza serie storiche = 13

Variabile dipendente: *Reg\_percvar*

Errori standard robusti (HAC)

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>rapporto t</i>	<i>p-value</i>
<i>const</i>	5,50511	2,28242	2,4120	0,0314
<i>AvgPrice_percvar</i>	-0,806942	0,346855	-2,3265	0,0368
<i>unem_percvar</i>	-0,0754111	0,0491416	-1,5346	0,1489
<i>CCI_percvar</i>	1,85704	1,46841	1,2647	0,2282
<i>inflation</i>	-1,28635	1,02825	-1,2510	0,2330
<i>GDPgrowth</i>	0,44387	0,528784	0,8394	0,4164
<i>int_d_unem</i>	-0,229507	0,157351	-1,4586	0,1684
<i>int_d_inflation</i>	-1,28038	1,05225	-1,2168	0,2453
<i>int_d_cci</i>	1,07828	1,80649	0,5969	0,5608
<i>int_d_GDPgr</i>	2,18771	1,35887	1,6099	0,1314
<i>int_d_pump</i>	0,207593	0,208439	0,9959	0,3375

Modello A5: Pooled OLS, usando 182 osservazioni

Incluse 14 unità cross section

Lunghezza serie storiche = 13

Variabile dipendente: *Reg\_percvar*

	Errori standard robusti (HAC)			
	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>rapporto t</i>	<i>p-value</i>
<i>const</i>	5,59037	2,26429	2,4689	0,0282
<i>AvgPrice_percvar</i>	-0,806571	0,344248	-2,3430	0,0357
<i>unem_percvar</i>	-0,0688862	0,0498633	-1,3815	0,1904
<i>CCI_percvar</i>	2,70584	1,47134	1,8390	0,0889
<i>inflation</i>	-1,11369	1,04236	-1,0684	0,3048
<i>GDPgrowth</i>	0,261363	0,430281	0,6074	0,5540
<i>int_d_unem</i>	-0,232457	0,161162	-1,4424	0,1729
<i>int_d_inflation</i>	-1,54328	0,882844	-1,7481	0,1040
<i>int_d_GDPgr</i>	2,43356	1,17934	2,0635	0,0596
<i>int_d_pump</i>	0,208802	0,207558	1,0060	0,3328

Modello A6: Pooled OLS, usando 182 osservazioni

Incluse 14 unità cross section

Lunghezza serie storiche = 13

Variabile dipendente: *Reg\_percvar*

Errori standard robusti (HAC)

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>rapporto t</i>	<i>p-value</i>
<i>const</i>	4,90266	2,27073	2,1591	0,0501
<i>AvgPrice_percvar</i>	-0,753358	0,354746	-2,1237	0,0535
<i>unem_percvar</i>	-0,0655761	0,049433	-1,3266	0,2075
<i>CCI_percvar</i>	2,76354	1,43641	1,9239	0,0765
<i>inflation</i>	-0,970131	1,01099	-0,9596	0,3548
<i>GDPgrowth</i>	0,30728	0,410436	0,7487	0,4674
<i>int_d_unem</i>	-0,21991	0,153361	-1,4339	0,1752
<i>int_d_inflation</i>	-0,875918	0,721561	-1,2139	0,2464
<i>int_d_GDPgr</i>	2,58828	1,17876	2,1958	0,0469

Modello A7: Pooled OLS, usando 182 osservazioni

Incluse 14 unità cross section

Lunghezza serie storiche = 13

Variabile dipendente: *Reg\_percvar*

Errori standard robusti (HAC)

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>rapporto t</i>	<i>p-value</i>
<i>const</i>	4,22066	1,75362	2,4068	0,0317
<i>AvgPrice_percvar</i>	-0,788081	0,374355	-2,1052	0,0553
<i>unem_percvar</i>	-0,0902742	0,0513024	-1,7596	0,1020
<i>CCI_percvar</i>	3,03661	1,30336	2,3298	0,0366
<i>GDPgrowth</i>	-0,12031	0,339891	-0,3540	0,7290
<i>int_d_unem</i>	-0,187242	0,137789	-1,3589	0,1973
<i>int_d_inflation</i>	-1,55782	0,863756	-1,8035	0,0945
<i>int_d_GDPgr</i>	3,00107	1,05344	2,8488	0,0137

Modello A8: Pooled OLS, usando 182 osservazioni

Incluse 14 unità cross section

Lunghezza serie storiche = 13  
 Variabile dipendente: *Reg\_percvar*  
 Errori standard robusti (HAC)

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>rapporto t</i>	<i>p-value</i>
<i>const</i>	4,33435	1,93249	2,2429	0,0430
<i>AvgPrice_percvar</i>	-0,757026	0,378053	-2,0024	0,0665
<i>unem_percvar</i>	-0,218103	0,0840635	-2,5945	0,0222
<i>CCI_percvar</i>	2,7089	1,3941	1,9431	0,0740
<i>GDPgrowth</i>	-0,208485	0,39636	-0,5260	0,6077
<i>int_d_inflation</i>	-1,75909	0,959636	-1,8331	0,0898
<i>int_d_GDPgr</i>	3,42616	0,927556	3,6938	0,0027

Modello A9: Pooled OLS, usando 182 osservazioni  
 Includi 14 unità cross section  
 Lunghezza serie storiche = 13  
 Variabile dipendente: *Reg\_percvar*  
 Errori standard robusti (HAC)

	<i>Coefficiente</i>	<i>Errore Std.</i>	<i>rapporto t</i>	<i>p-value</i>
<i>const</i>	1,66312	0,893438	1,8615	0,0854
<i>AvgPrice_percvar</i>	-0,735324	0,362677	-2,0275	0,0636
<i>unem_percvar</i>	-0,190743	0,0755115	-2,5260	0,0253
<i>CCI_percvar</i>	3,5285	1,14835	3,0727	0,0089
<i>GDPgrowth</i>	0,47711	0,256712	1,8585	0,0859
<i>int_d_GDPgr</i>	2,56161	0,866037	2,9579	0,0111

# RIFERIMENTI BIBLIOGRAFICI

## Bibliografia

Wooldridge Jeffrey M. 2006. *Introductory econometrics: a modern approach*. Mason, OH, Thomson/South-Western.

Greene, William H. 2008. *Econometric Analysis*, 6th ed. Upper Saddle River, NJ: Prentice Hall

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Kennedy, Peter. 2008. *A Guide to Econometrics*, 6th ed. Malden, MA: Blackwell Publishing

Breusch, T. S., and A. R. Pagan. 1980. "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics." *Review of Economic Studies*, 47(1):239-253.

Chow, Gregory C. 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica*, 28 (3): 591–605.

Hausman, J. A. 1978. "Specification Tests in Econometrics." *Econometrica*, 46(6):1251- 1271.

Swamy, P. A. V. B. 1970. "Efficient Inference in a Random Coefficient Regression Model." *Econometrica*, 38: 311-323.

White, H. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity". *Econometrica*. 48 (4): 817–838.

Cottrell A., Lucchetti R. 2008. *Gretl's User Guide. Gnu regression, Econometrics and Time-series Library*

## Sitografia

Database WTI su crescita del PIL [online], accesso 24.05.2016  
<http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG>

Database ICCT su vendite di automobili e prezzi medi di vendita [online], accesso 24.05.2016  
<http://eupocketbook.theicct.org/data>

Database ILO su tasso di disoccupazione [online], accesso 30.05.2016  
<http://data.worldbank.org/indicator/SL.UEM.TOTL.ZS>

Database OCSE su indice CCI [online], accesso 31.05.2016  
<https://data.oecd.org/leadind/consumer-confidence-index-cci.htm>

Database Commissione Europea su prezzi benzina [online], accesso 06.06.2016  
<https://ec.europa.eu/energy/en/statistics/weekly-oil-bulletin>

Database OCSE su inflazione (CPI) [online], accesso 12.07.2016  
<https://data.oecd.org/price/inflation-cpi.htm>