

Proiect Econometrie

**Analiza prețului mașinilor electrice în funcție de specificații,
utilizându-se modele de regresie simplă și multiplă**

Profesori coordonatori:

Alexandru Adriana Ana Maria

Manta Eduard Mihai

Membrii echipă:

Dumitriu Ana Maria

Dumitrescu Teodora

Florean Gabriela

Introducere

Modelul econometric este o reprezentare matematică a relațiilor din economie, ce ajută la definirea legăturilor complexe dintre variabile și studiază un ansamblu de factori ce acționează asupra acestora.

Scopul analizei de regresie este estimarea și previzionarea valorii medii a unei variabile dependente, ce se află în legătură cu una sau mai multe variabile independente (explicative).

Sursa datelor <https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset>

Informatii despre date

Setul de date conține 103 înregistrari (mașini); nu au fost șterse înregistrări; au fost create variabile dummy pe baza variabilelor existente.

Variabila dependentă: **PriceEuro** - prețul mașinii în euro

Variabile independente:

Brand, Model

Accel (s) - timpul în care ajunge la 100km/h

TopSpeed (km/h) - viteza maximă

Range (km) - autonomie

Efficiency (watt-ora/km) - câți Wh din baterie consumă într-un km

FastCharge - încarcare rapidă

PlugType - tipul de încărcător

Tipul de tracțiune (integrală AWD, față FWD, spate RWD)

Caroserie - forma mașinii (SUV, Hatchback, Sedan, Liftback, Pickup)

Segment - categoria (A-minicompactă, F-de lux)

Numar de locuri - în intervalul 2-7

Variabile Dummy:

1 = FC, 0 restul

1 = Type 2 CCS

1 = FWD, 0 restul

1 = HB, 0 restul

1 = B, 0 restul

1 = 5 locuri, 0 restul

Studiul cunoasterii

1. Articolul **Electric Car Purchase Price as a Factor Determining Consumers' Choice and their Views on Incentives in Europe** prezinta preferințele cumpărătorilor din 6 țări europene în funcție de venit anual, gen, educație, zona de reședință și modalitatea de plată. Ajutoarele guvernamentale sau scutirile de taxe influențează decizia de a alege autovehicule electrificate ce au un preț ridicat. Factori care influențează decizia: prețul, consumul de combustibili și costul acestuia, costurile de întreținere, siguranța și confort. (<https://www.mdpi.com/2071-1050/11/22/6357>)

2. Conform articolului **“Who goes electric? The anatomy of electric car ownership in Norway”** publicat de ScienceDirect în martie 2021, Norvegia are cea mai mare flotă de mașini electrice din Europa. În 2020 52,2% dintre mașinile înregistrate au fost electrice(BEV) și 20,4% Plug-In(PHEV). Acest fapt se datorează scutirii de taxe, parcărilor și încărcării gratuite, reducerea taxelor în porturi și posibilitatea acestora de a utiliza benzile speciale pentru transportul în comun, influențând decizia populației de a achiziționa aceste autovehicule.

(<https://www.sciencedirect.com/science/article/abs/pii/S1361920921000316>)

3. Lasse Fridstrom și Vegard Ostli prezintă în lucrarea **“Direct and cross price elasticities of demand for gasoline, diesel, hybrid and battery electric cars: the case of Norway”** elasticitatea prețului autoturismelor pe bază de motor termic sau electric, de-a lungul anilor 2002-2016, în care s-au achiziționat peste 1,8 milioane de autoturisme. Ultimii 6 ani au reprezentat o creștere a vânzării autoturismelor electrificate, ce au câștigat o poziție majoritară pe piața din Norvegia.

(<https://etrr.springeropen.com/articles/10.1186/s12544-020-00454-2>)

4. În cadrul articolului **“The Effect of Reducing Electric Car Purchase Incentives in the European Union”**, Jonatan J. G. V. și Christian T. au arătat că trendul mașinilor electrificate este unul care a înregistrat o creștere puternică datorită subvențiilor oferite de țările Uniunii Europene, care ajung la sume de 3000-4000 de euro pentru modelele plug-in sau full electric, dar și datorită dezvoltării infrastructurii pentru încărcarea lor. Achiziția acestor mașini duce la reducerea cererii de motoare termice, a poluării și a emisiilor de CO₂, precum și la extinderea cererii de bateri bazate pe lithium-ion. (<https://www.mdpi.com/2032-6653/10/4/64>)

5. **Global electric car market deployment considering endogenous battery price development** - Costul de producție al bateriilor pentru mașinile electrice este cel care are rol esențial în influențarea prețului final de vânzare. Astfel în funcție de autonomia celulelor electrice și capacitatea acestora de încărcare se remarcă diferențe în piața mașinilor electrificate, în funcție de modelul lor sau tipul de autovehicul- Bev(electric), PHEV(Plug in), MHEV(mild hybrid) sau HEV(full hybrid).

(<https://www.sciencedirect.com/science/article/pii/B9780128187623000157>)

Regresia simplă

Prețul în funcție de viteza maximă

Ipoteze

Forma funcțională este liniară
 Erorile au media 0
 Homoschedasticitatea erorilor aleatoare.
 Erorile nu sunt autocorelate.
 Necorelare între regresor și erorile aleatoare
 Erorile au distribuție normală

```
rs_topspeed <- lm(PriceEuro ~ TopSpeed, price1)
mean(rs_topspeed$residuals) ->0
white_test(rs_topspeed) # p-value = 0.34 > 0.1
dwtest(rs_topspeed) # p-value= 0.59 > 0.1
cor.test(price1$TopSpeed, rs_topspeed$residuals)
jarque.bera.test(model_speed$residuals)
-> p-value < 0.05 => eliminarea valorilor extreme
```

Prețul în funcție de viteza maximă

`rs_topspeed <- lm(PriceEuro ~ TopSpeed, price1)`

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -60570.50    8035.95   -7.537 2.12e-11 ***
TopSpeed      649.47     43.59   14.901 < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19180 on 101 degrees of freedom
Multiple R-squared:  0.6873,    Adjusted R-squared:  0.6842 
F-statistic: 222 on 1 and 101 DF,  p-value: < 2.2e-16
```

R-squared: 68.73%

Modelul este semnificativ la 99%

Variabila independentă explică **68.73%** din variabilitatea prețului.

La o creștere cu 1 km/h a vitezei maxime, prețul mașinii crește cu **649.47** euro.

$$\text{PriceEuro} = -60\,570.5 + 649.47 * \text{TopSpeed}$$

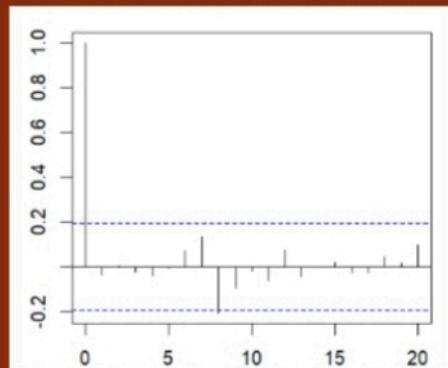
Homoschedasticitatea

-> este prezentă atunci când varianța erorilor este constantă pentru fiecare x

H0: erorile sunt **homoscedastice** (rez. distribuite cu varianță egală)

H1: erorile sunt **heteroscedastice**

```
# Ipoteza - Reziduurile sunt homoscedastice
bptest(rs_topspeed)      # p-value = 0.16 > 0.1 => rez. homoscedastice
white_test(rs_topspeed)   # p-value = 0.34 > 0.1 => homoscedastice
```



Există două laguri care depășesc liniile punctate, nu avem autocorelare

Autocorelare

H0: reziduurile nu sunt autocorelate

H1: reziduurile sunt autocorelate

Testul Durbin-Watson (ordinul 1)

`dwtest(rs_topspeed)`

(p-value= 0.59 > 0.1 => rez. nu sunt autocorelate)

Testul Breusch-Godfrey (order superior)

`bgtest(rs_topspeed)` (p-value = 0.7)

`bgtest(rs_topspeed, order = 2)` (p-value = 0.93)

`bgtest(rs_topspeed, order = 3)` (p-value = 0.97)

Normalitatea

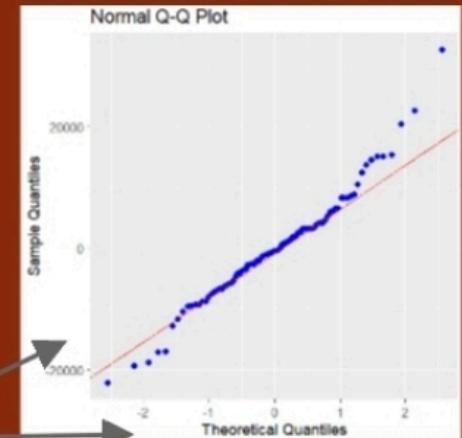
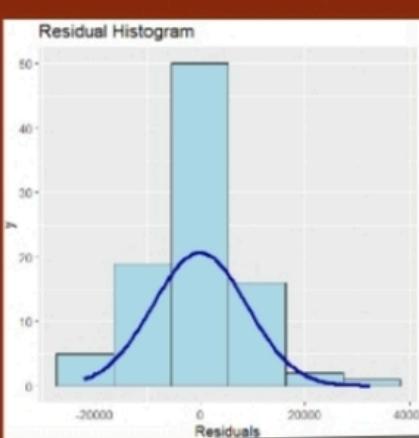
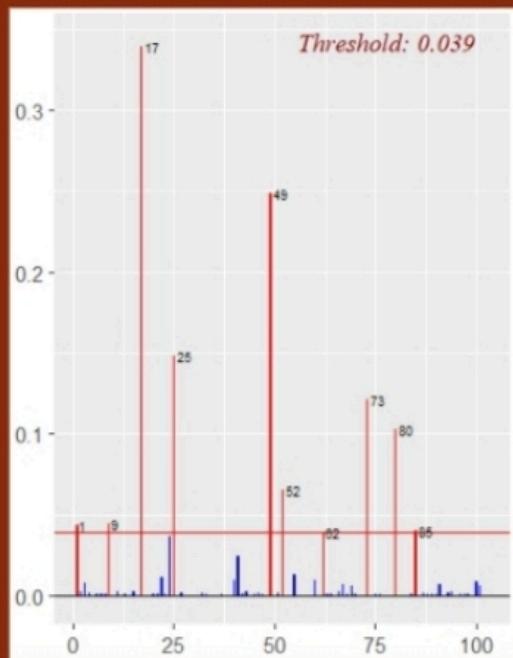
```
jarque.bera.test(model_speed$residuals) ->0
```

```
ols_test_normality(model_speed) ->0
```

p-value < 0.05 => reziduurile nu sunt normal distribuite

Corecție: eliminăm punctele indentificate cu **distanță Cook**
hprice_cook <- price1[-c(1,9,17,25,49,52,62,73,80,85),]

Dupa corecție, reziduurile nu sunt normal distribuite, este o limitare a datelor.



Prognoze

Impărțim setul de date în set de antrenare (80%) și în set de testare, și testăm acuratețea pe setul de testare. MAPE < 1

```
> out_of_sample <- data.frame(TopSpeed = c(200,250,270))  
> # Prognoza  
> y_pred_outsample <- predict(model_2, newdata = out_of_sample)  
> y_pred_outsample  
1 2 3  
67076.92 98795.46 111482.88
```

	fit	lwr	upr
1	67076.92	65042.81	69111.03
2	98795.46	94798.81	102792.11
3	111482.88	106598.86	116366.89
4	117826.58	112491.04	123162.12
5	130514.00	124265.39	136762.61
6	143201.41	136030.63	150372.19

Prețul din **prognoza** pentru o mașină cu viteza maximă **250 km/h** (2) este **98 795.46** euro.

Pentru un **interval de încredere de 90%**, putem estima următoarele valori:

```
new.speeds <- data.frame(TopSpeed = c(200, 250, 270, 280, 300, 320))
```

```
predict(model_2, newdata = new.speeds, se.fit=TRUE,interval="confidence", level=0.90)
```

Prețul pentru o mașină cu viteza maximă de **300 km/h** (5) este cuprins între 136 030 și 150 372 euro, media fiind **143 201** euro.

Regresia multiplă

Prețul în funcție de viteza maximă, numărul de locuri, autonomia

Ipoteze

Forma funcțională este liniară

Variabilitatea în x pozitivă:

Erorile au media 0

Homoschedasticitatea erorilor aleatoare.

Erorile nu sunt autocorelate.

Necorelare între regresor și erorile aleatoare

Erorile au distribuție normală

Multicolinearitate

```
model_0 <- lm(PriceEuro ~ DummySeats5+  
TopSpeed+ Range, price1)  
var(hprice1$Range) #>0  
mean(model_0$residuals) #0  
white_test(model_0) #0.059 (homo. la 95%)  
bptest(model_0) #0.07 (homo. la 95%)  
dwtest(model_0) #p-value= 0.83  
cor.test(hprice1$DummySeats5, model_0$  
residuals) #p-value = 1 > 0.1  
jarque.bera.test(hprice1$uhat) # p-value < 0.05  
corectăm prin eliminarea extremelor.  
vif(model_0) #valori <10 (nu avem multicol.)
```

Prețul în funcție de nr. de locuri, viteza maximă, autonomie

model_0 <- lm(PriceEuro ~ DummySeats5 + TopSpeed + Range, price1)

PriceEuro = -49734.44 + (-9186.03) * DummySeats5 + 538.22 * TopSpeed + 45.55 * Range

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-49734.44	8913.16	-5.580	2.10e-07 ***
DummySeats5	-9186.03	4120.50	-2.229	0.0281 *
TopSpeed	538.22	66.03	8.152	1.13e-12 ***
Range	45.55	22.66	2.010	0.0471 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18690 on 99 degrees of freedom
Multiple R-squared: 0.709, Adjusted R-squared: 0.7002
F-statistic: 80.4 on 3 and 99 DF, p-value: < 2.2e-16

Bonitatea: R-squared = 70%

Modelul este semnificativ la 99%.

Variabilele independente explică

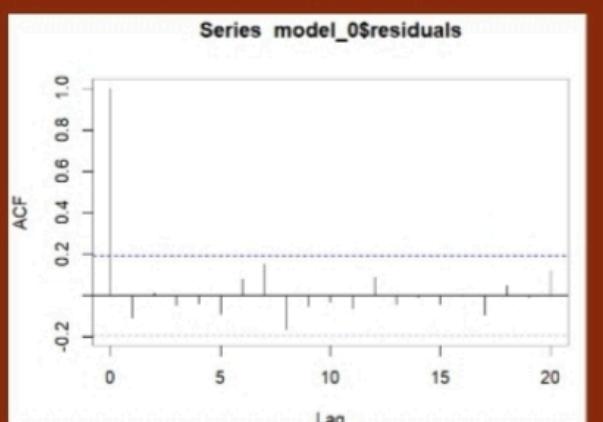
70% din variabilitatea prețului unei mașini și sunt semnificative la 99%.

La o creștere a autonomiei cu 1 km, prețul crește cu 45.55 de euro, la creșterea vitezei maxime cu 1 km/h, prețul crește cu 538.22 euro, iar mașinile care au fix 5 locuri au prețul cu 9 168 euro mai mic.

Homoschedasticitatea

Erorile sunt constante, varianța erorilor este aceeași pentru toate observațiile.

```
# Testul Breusch-Pagan și Testul White  
bptest(model_0)      #p-value = 0.079 > 0.05 => rez homoschedastice la 95%  
white_test(model_0)   #p-value = 0.059 > 0.05 => rez homoschedastice la 95%
```



Nu avem laguri =>
Reziduurile nu sunt autocorelate

Autocorelare

H0: reziduurile **nu sunt autocorelate**

H1: reziduurile sunt autocorelate

Testul Durbin-Watson (ordinul 1)

```
dwtest(model_0)
```

#p-value= 0.83 > 0.1 => rez. nu sunt autocorelate

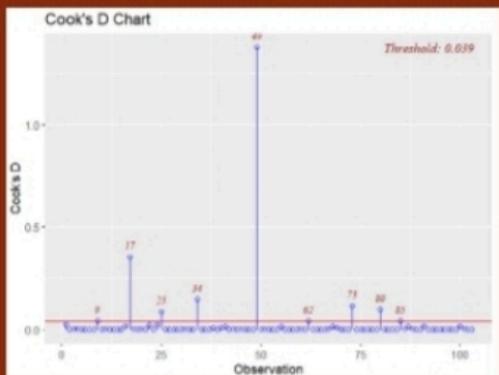
Testul Breusch-Godfrey (ordin superior)

```
bgtest(model_0)           (p-value = 0.28)
```

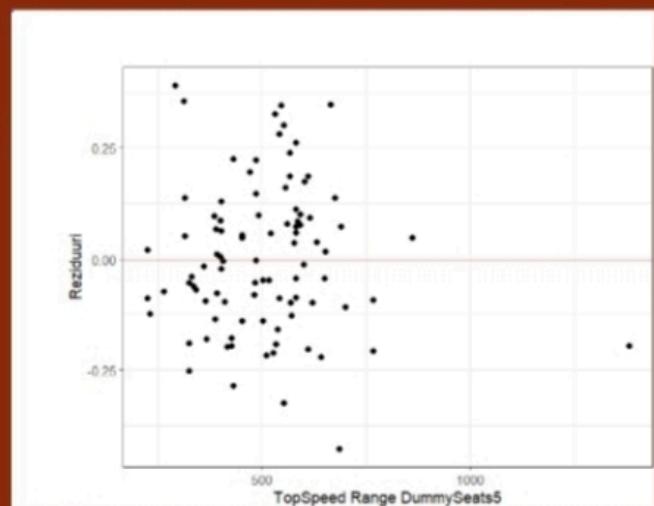
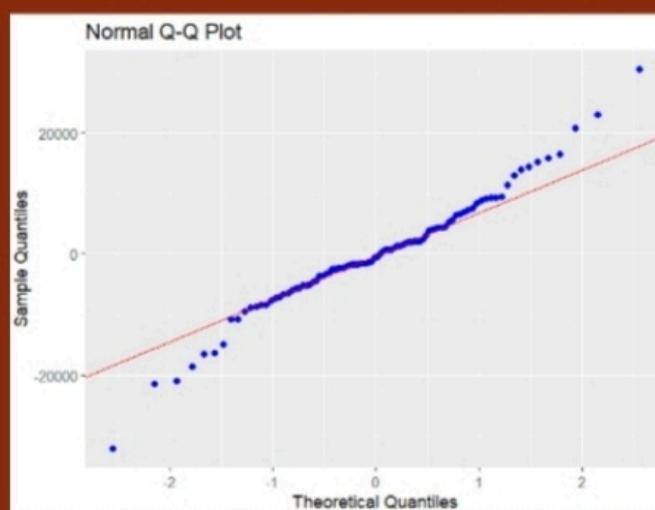
```
bgtest(model_0, order = 2) (p-value = 0.56)
```

```
bgtest(model_0, order = 3) (p-value = 0.70)
```

Normalitatea



H0: **distribuție normală**, Ha: distribuție nenormală
`shapiro.test(hprice1$uhat)` # p-value < 0.05
`jarque.bera.test(hprice1$uhat)` # p-value < 0.05 (->0)
`ols_test_normality(model_0)` # p-value <0.05
#p-value aproape 0=> rez. **nu sunt normal distribuite**
Eliminăm valorile extreme indentificate cu distanța **Cook**
Dupa corectare, rez. nu sunt normal distribuite (limitare a datelor)
`jarque.bera.test(hprice_cook$uhat)` # p-value = **0.044** < 0.05



Prognoze

Impărțim setul de date în set de **antrenare** (80%) și în set de **testare**, și testăm acuratețea pe setul de testare. MAPE < 1

Valori pentru variabilele independente:

`out_of_sample<-data.frame(DummySeats5=c(1,1,1),TopSpeed=c(200,250,270), Range = c(250,300,400))`

Prețul progozat pentru o mașină(2) cu **5 loc.**, viteza maximă **250 km/h**, autonomie **300 km** -> **97 244**

	fit	lwr	upr
1	65060.27	60029.30	70091.24
2	97244.11	89627.94	104860.28
3	110590.26	103723.64	117456.87
4	122706.76	116745.81	128667.71
5	130904.85	125008.53	136801.17
6	147922.15	138545.37	157298.93

Pentru o mașină cu 5 locuri, 270 km/h viteza maximă și 400 km autonomie (3) prețul este cuprins între 103 723 și 117 456 euro, media fiind 110 590 euro.

Pentru un interval de încredere de **90%** , putem estima valorile:

`new.specificatii <- data.frame(DummySeats5 = c(1,1,1,0,1,0),TopSpeed = c(200, 250, 270, 280, 300, 320), Range = c(250,300,400,500,600, 450))`

`predict(model_a, newdata = new.specificatii, se.fit=TRUE,interval="confidence", level=0.90)`

Regresia neliniară

Pretul în funcție de viteza maximă, numărul de locuri, autonomia

Ipoteze

Calculăm reziduurile, pătratul reziduurilor și cubul reziduurilor și adăugăm în model.

Verificăm dacă modelul este bine specificat

Calculăm pentru modelul log-lin

```
model2 <- lm(lPriceEuro ~ TopSpeed + Range  
+ DummySeats5, hprice1)
```

Calculăm modelul cu pătratul variabilelor și adăugăm pătratul și cubul lui yhat

```
model3 <- update(model1, ~ . + TopSpeed_sq +  
Range_sq + DummySeats5)
```

```
model1_RESET <- update(model1, ~ . + yhatsq + yhatcube)
```

H0: mod. **e bine specificat**, H1: mod. nu e bine specificat

```
linearHypothesis(model1_RESET, c("yhatsq = 0", "yhatcube=0"))  
# F = 0.3 > 0.1 => modelul este bine specificat
```

```
linearHypothesis(model2_RESET, c("lyhatsq = 0", "lyhatcube=0"))  
# p-value < 0.1 => modelul nu este bine specificat
```

```
linearHypothesis(model3_RESET, c("yhat1sq = 0", "yhat1cube=0"))  
# F = 0.96 < 0.1 => modelul nu este bine specificat
```

Testul Chow

```
sctest(hprice1$PriceEuro ~ hprice1$TopSpeed, type = "Chow", point = 10) # p-value = 0.22 > 0.1  
#=> nu există rupturi pentru variabila TopSpeed, nu trebuie să avem 2 modele de regresie  
sctest(hprice1$PriceEuro ~ hprice1$Range, type = "Chow", point = 10)  
# p-value = 0.89 > 0.1 => nu există rupturi pentru var. Range  
sctest(hprice1$PriceEuro ~ hprice1$DummySeats5, type = "Chow", point = 10)  
# p-value = 0.57 > 0.1 => nu există rupturi
```

Regresie cu variabile indicatori

FWD = tractiune fata

Testul T pentru diferența pe medii a prețului dintre mașinile cu tractiune față și celelalte
t.test(formula = PriceEuro ~ DummyFWD, data = hprice1)

p-value<0.01 => Media prețului între mașinile cu tractiune față și celelalte mașini este semnificativ diferit la 99%.

Diferența dintre regresia cu mașini fără FWD și mașinile cu FWD

lm(PriceEuro ~ **DummyFWD**, hprice1) || lm(PriceEuro ~ **notFWD**, hprice1)

-> Coeficientul are aceeași magnitudine și p-value dar **semn opus** față de regresia cu DummyFWD

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	34416	4185	8.223	1.23e-12 ***
notFWD	27004	5328	5.068	2.06e-06 ***

Interacțiune cu Dummy

hprice1 %<>% mutate(HB_FWD = DummyBodyStyle *

DummyFWD, notHB_FWD = notBodyStyle * Price = 70 232 +(-31940)*x1 + (-33478)*x2 +26681*x1*x2

DummyFWD, HB_notFWD = DummyBodyStyle * notFWD)

Regresia cu termen de interacțiune cu dummy

model_interactiune_dummy <- lm(PriceEuro ~ DummyBodyStyle + DummyFWD +
DummyBodyStyleXDummyFWD, hprice1)

Coeficienții sunt semnificativi la 99%

R-squared = 36%

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	70232	3486	20.145	< 2e-16 ***
DummyBodyStyle	-31940	6638	-4.812	5.99e-06 ***
DummyFWD	-33478	6138	-5.454	4.29e-07 ***
DummyBodyStyleXDummyFWD	26681	10074	2.648	0.00955 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22590 on 90 degrees of freedom
Multiple R-squared: 0.3809, Adjusted R-squared: 0.3602
F-statistic: 18.46 on 3 and 90 DF, p-value: 2.045e-09

x1 = HB = hatchback x2= FWD = tractiune fata

Interpretare: Efectul lui x1 (HB) în preț este 70232 (intercept) + 26681*x2.

Daca x2 (tractiune față) este 0, efectul lui x1 (HB) în preț este 70232 (intercept). Pentru două mașini cu x2=0 (nu sunt tractiune față), o mașină Hatchback are prețul mai mic cu 31940 euro decât o mașină cu altă caroserie.

Daca x2 = 1, efectul lui x1 este -31940 + 26681 = -5259 euro => Pentru două mașini cu tractiune față, o mașină HB are prețul cu 5259 euro mai mic decât o mașină care nu e HB.

Testul F

F-test pentru semnificația coeficientului comun este utilizat pentru a testa ca mai mulți coeficienți simultan să fie semnificativi diferenți de zero

Generăm termeni de interacțiune

```
hprice1 %>% mutate(TopSpeedXHB = DummyBodyStyle * TopSpeed,
```

```
          RangeXHB = DummyBodyStyle * Range, AccelXHB = DummyBodyStyle * Accel)
```

Creăm modelul: model_10 <- lm(PriceEuro ~ TopSpeed + Range + Accel + TopSpeedXHB + RangeXHB + AccelXHB, hprice1)

Calculăm suma reziduurilor

```
> ssr_ur [1] 8213467165
```

```
> ssr_r [1] 8573712728
```

Calculam F-Statistic

```
(F_stat <- ((ssr_r - ssr_ur)/q) / (ssr_ur/df_resid))
```

```
#0.953
```

F stat > F critic (0.95)

=> coeficienții sunt semnificativi simultan

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-65592.36	10462.62	-6.269	1.35e-08	***
TopSpeed	695.93	55.98	12.433	< 2e-16	***
Range	-18.89	19.88	-0.950	0.345	
Accel	250.87	575.26	0.436	0.664	
TopSpeedXHB	-139.82	82.32	-1.698	0.093	.
RangeXHB	44.69	32.17	1.389	0.168	
AccelXHB	749.74	698.40	1.073	0.286	

Signif. codes: 0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 9716 on 87 degrees of freedom					
Multiple R-squared: 0.8893, Adjusted R-squared: 0.8817					
F-statistic: 116.5 on 6 and 87 DF, p-value: < 2.2e-16					

Testul Chow

Creăm modele nerestricționate pentru BodyStyle (HB):

```
model_12 <- update(model_11, subset = (DummyBodyStyle == 1))
```

```
model_13 <- update(model_11, subset = (DummyBodyStyle == 0))
```

Calculăm suma pentru reziduri:

```
ssr1 <- sum(resid(model_12)^2) #482341710 || ssr2 <- sum(resid(model_13)^2) #7516531866
> # Calculam Chow F-statistic
> (F_stat <- ((ssr_r-(ssr1+ssr2))/(k+1)) / ((ssr1+ssr2)/(n-2*(k+1))))
[1] 1.545098
> # Valoarea critica F
> qf(0.99, k+1, n-2*(k+1))
[1] 3.545424
```

=> F-stat = 1.54 < F-critical = 3.54 => acceptăm nula, coeficienții nu sunt semnificativi simultan

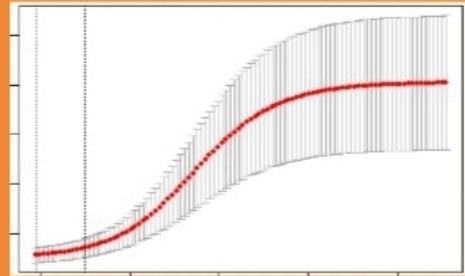
Modele de regularizare

Regresia Ridge, Regresia LASSO, Elastic Net Regression, Boruta

Regresia Ridge

Se folosește când există **multicoliniaritate** în date, **minimizează**
 $SSR = \text{sum}((y - y_{\text{fit}})^2) + \lambda \cdot \text{sum}(\beta^2)$

```
model_log <- lm(lPriceEuro ~ DummySeats5 + TopSpeed  
+ Range, hprice1) # R-sq=88.4%  
Lambda MSE minimizat, utilizând validarea încruțișată  
cv_model <- cv.glmnet(x, y, alpha = 0)  
best_lambda <- cv_model$lambda.min # 2640.67  
Modelul cu valoarea lambda optimă  
best_model <- glmnet(x, y, alpha = 0, lambda = best_lambda)
```



Prognoza: new <- matrix(c(1, 210, 400), nrow=1, ncol=3) # 5 locuri, TopSpeed, Range

predict(best_model, s = best_lambda, newx = new) # pret 70225

(Intercept) -46615.8972
DummySeats5 -3859.0923
TopSpeed 506.8496
Range 35.6543

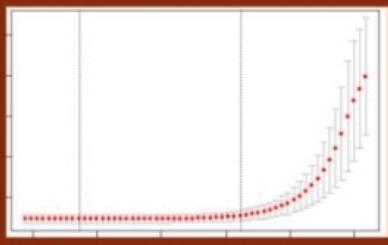
R-squared: 87.3%

Regresia LASSO

Similară cu regresie Ridge,
minimizează
 $SSR + \lambda \cdot \text{sum}(|\beta|)$

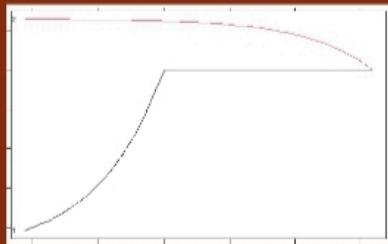
```
cv_model <- cv.glmnet(x, y, alpha = 1)  
best_lambda <- cv_model$lambda.min # 131.42
```

Diagrama Trace - modul în care **estimările coef.** s-au modificat ca urmare a **creșterii valorii lui lambda**



Modelul cu valoarea lambda optimă:

```
best_model <- glmnet(x, y, alpha = 1, lambda = best_lambda)
```



Prognoza: predict(best_model, s = best_lambda, newx = new)

72842 euro R-squared: 88.45%

(Intercept) -59008.3467
DummySeats5 -1669.0442
TopSpeed 635.5819
Range .

Elastic net regression

Similară cu Ridge și LASSO, adaugă ambele penalități $SSR + \lambda \sum(\beta_i^2) + \lambda \sum(|\beta_i|)$

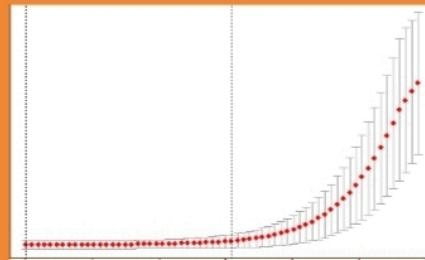
```
cv_model <- cv.glmnet(x, y, alpha = 0.5)
best_lambda <- cv_model$lambda.min #150.41
```

Modelul cu valoarea lambda optimă

```
best_model <- glmnet(x, y, alpha = 0.5, lambda = best_lambda)
coef(best_model)
```

Prognoza:

```
predict(best_model, s = best_lambda, newx = new)
#72773 euro      R-squared: 88.44%
```



(Intercept)	-5.923016e+04
DummySeats5	-2.170033e+03
TopSpeed	6.384709e+02
Range	2.380666e-01

Comparare modele

	Ridge Regression	LASSO Regression	Elastic Net Regression
Alpha	0	1	0.5
Lambda	2640.67	131	150.41
R-squared	87.3%	88.45%	88.45%
Pret	70 519 euro	72 343 euro	72 773 euro
Coef	(Intercept) -46615.8972 DummySeats5 -3859.0923 TopSpeed 506.8496 Range 35.6543	(Intercept) -59008.3467 DummySeats5 -1669.0442 TopSpeed 635.5819 Range .	(Intercept) -5.923016e+04 DummySeats5 -2.170033e+03 TopSpeed 6.384709e+02 Range 2.380666e-01

Modelul care are **bonitatea** maximă este cel determinat utilizând **Elastic Net Regression**

(88.45%), coeficienții variabilelor independente nu sunt semnificativi.

Modelul calculat cu Ridge R. are o bonitate de 87.3%; la o creștere cu un km/h a vitezei maxime, prețul crește cu 506.854 euro, la o creștere cu un km a autonomiei, prețul crește cu 35.65 euro, iar construcția mașinii având fix 5 locuri scade din prețul mașinii 3859.15 euro.

Modelul calculat cu LASSO R. are R-squared 88.45%. La o creștere cu o unitate a vitezei maxime, prețul crește cu 635.58 euro, construcția cu fix 5 locuri scade din prețul mașinii 1669 euro, iar variabila Range nu este semnificativă.

Boruta

Modelul realizat cu toate variabilele importante:

```
model_Boruta1 <- lm(PriceEuro ~ Brand + Accel + TopSpeed + Range + Efficiency + PowerTrain + Segment + Seats, hprice1) # R-squared = 98.82%
```

Model realizat cu variabilele numerice importante:

```
model_Boruta2 <- lm(PriceEuro ~ Accel + TopSpeed + Efficiency + Range, hprice1) #R-squared=90.76%.
```

Model realizat cu var. numerice și cu var. dummy create folosind var. calitative importante

```
model_Boruta3 <- lm(PriceEuro ~ Accel + TopSpeed + Efficiency + Range + DummySegmentB + DummySeats5, hprice1) # Bonitatea modelului = 91.17%.
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.057e+05	1.261e+04	-8.384	8.15e-13 ***
Accel	8.786e+02	4.967e+02	1.769	0.0804 .
TopSpeed	6.307e+02	4.713e+01	13.381	< 2e-16 ***
Efficiency	2.071e+02	4.073e+01	5.083	2.10e-06 ***
Range	6.171e+00	1.445e+01	0.427	0.6704
DummySegmentB	3.352e+03	2.397e+03	1.399	0.1655
DummySeats5	-2.922e+03	2.144e+03	-1.363	0.1765

- + 1 km/h vitezei, prețul + 630.7 euro
- + 1 km autonomiei, prețul - 6.17 euro
- mașina este segment B, prețul + 3353 euro
- mașina are 5 locuri, prețul - 2922 euro

PriceEuro = -105700 + 8786* Accel + 630.7* TopSpeed + 207.1 * Efficiency - 6.17 * Range + 3352 * DummySegmentB -2922 *DummySeats5

Panel

GDP in functie de consumul de energie nucleara, hidroenergie si energie din resurse regenerabile

Obiectiv. Studiul cunoasterii

1. În lucrarea „Replacing hard coal with wind and nuclear power in Finland- impacts on electricity and district heating markets” publicată în ScienceDirect, autorii A.Khosravi, V.Olkkonen, A.Farsaei și S.Syri modelează impactul tranzitiei de la energia pe bază de carbon la energia eoliană și nucleară, adoptat de Finlanda, pe piețele de energie electrică și de termoficare.

(<https://www.sciencedirect.com/science/article/abs/pii/S0360544220309919>)

2. Scopul lucrării “Consequences of the energy policy in member states of the European Union – the renewable energy sources targets” este de a pune în vedere și de a evalua diversificarea consumului de surse de energie regenerabilă în statele membre ale Uniunii Europene.

(<https://bibliotekanauki.pl/articles/283192>)

3. În lucrarea “Renewable energy production capacity and consumption in Europe” se analizează atât capacitatea de producție a energiei pe bază de surse regenerabile, cât și factorii determinanți ai consumului de energie pe bază de surse regenerabile, pe un eșantion de 30 de țări europene.

(<https://www.sciencedirect.com/science/article/abs/pii/S0048969722056911>)

4. Cercetarea “Five years left – How are the EU member states contributing to the 20% target for EU's renewable energy consumption; the role of woody biomass” compară și evaluatează rolul biomasei în energia regenerabilă în UE, concentrându-se pe resursele potențiale ale țărilor și pe sprijinul politiciilor. Studiul arată că toate țările depun eforturi pentru a atinge obiectivul de 20% în 2020 și că prezintă o tendință de creștere a energiei regenerabile.

(<https://www.sciencedirect.com/science/article/abs/pii/S0961953416303063>)

5. Lucrarea “On the causal dynamics between emissions, nuclear energy, renewable energy, and economic growth” examinează relația cauzală dintre emisiile de CO₂, consumul de energie nucleară, consumul de energie regenerabilă și creșterea economică.

(<https://www.sciencedirect.com/science/article/abs/pii/S0921800910002399>)

Semnificația datelor

Setul de date conține 13 înregistrări (țări), analizate pe o perioadă de 11 ani (2010-2020).

Variabila dependentă: **GDP (\$)** – Produsul intern brut real total, ajustat în funcție de inflație

Variabile independente:

nuclear_consumption (TW/h) - Consumul de energie primară din energia nucleară

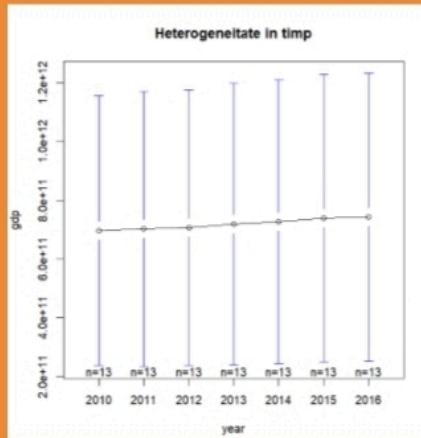
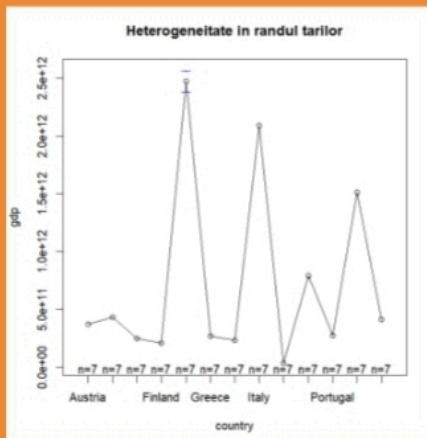
hydro_consumption (TW/h) - Consumul de energie primară din hidroenergie

renewables_consumption - Consumul de energie primară din surse regenerabile

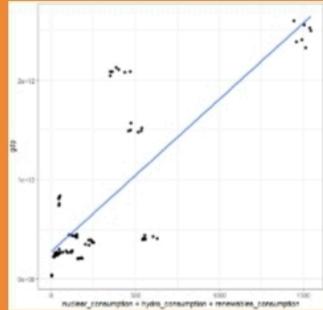
Heterogeneitate. Corelatii

Heterogeneitatea presupune ca există diferențe între unitatile studiate. Graficele trasează un interval de încredere de 95% în jurul mediilor.

Există heterogeneitate în randul tarilor Nu există heterogeneitate în timp



Există **heteroschedasticitate** în model.



Rezultate. Ipoteze

ols GBP este explicată în proporție de 83.27% de variabilitatea variabilelor independente
p-value < 0.05

fe Model cu efecte fixe.
fe <- plm(gdp ~ nuclear_consumption + hydro_consumption + renewables_consumption, data, index = c('country', 'year'), model = 'within')

R-squared = 41.9%.
Modelul este semnificativ

re Model cu efecte random (aleatoare)
re <- plm(gdp ~ nuclear_consumption + hydro_consumption + renewables_consumption, data, index = c('country', 'year'), model = 'between')

R-squared = 85.1%.
Modelul este semnificativ

Testul Hausmann

phtest(fe,ols) Testul Hausmann este utilizat pentru a decide ce model se pretează cel mai bine în rezolvare.
phtest(fe,re) H0: model cu efecte random H1: model cu efecte fixe
p-value < 0.05 => Vom utiliza modelul cu efecte fixe.

Testarea efectelor fixe în timp

pFtest(fixed.time, fe) p-value = 0.27 > 0.05
plmtest(fe, c('time'), type = 'bp') p-value = 0.25

Se verifică dacă este necesară utilizarea efectelor fixe în timp.

H0: nu sunt necesare efectele fixe în timp;

H1: sunt necesare efectele fixe în timp

Ipoteza nulă este respinsă, vom utiliza modelul cu efecte fixe

```
fixed.time <- plm(gdp ~ nuclear_consumption + hydro_consumption + renewables_consumption + factor(year), data = data, index = c("country", "year"), model = "within")
```

Modelul pool se folosește în testarea efectelor aleatorii

Testul ne ajută să decidem între RE și OLS

plmtest(pool, type=c("bp")) # p-value<0.05 H0: variațiile în timp sunt 0 H1: variațiile în timp sunt diferențiate de 0

Respingem ipoteza nulă, există diferențe semnificative între țări

Breusch-Pagan LM și testul Parasan CD

pcdtest(fe, test = 'lm') p-value < 0.05 => există dependență transversală

pcdtest(fe, test = 'cd') p-value = 0.42 > 0.05 => nu există dependență transversală

Breusch-Pagan

bptest(model) p-value < 0.05
=> avem heteroschedasticitate

Breusch-Godfrey/Wooldridge

pbgtest(fe) # p-value < 0.05
=> avem autocorelare

Concluzii

Informații proiect

Modelul analizează modul în care variabilele independente (specificațiile mașinilor) influențează variabila dependentă (prețul) acestora.

Setul de date cuprinde 103 observații - mașini. Despre fiecare mașină se cunosc informații precum: prețul în euro, brand, model, accelerație, viteza maximă, autonomie, eficiență, încarcare rapidă și tipul de încărcător, tipul de tractiune, caroserie, segment și număr de locuri.

Pe baza acestor variabile am creat variabile dummy precum: DummFC (1 = are încărcare rapidă) DummyFWD (1 = tractiune față), DummyHB (1 = caroserie Hatchback), DummySeats5 (1 = are 5 locuri)

Regresia simplă

Prețul în funcție de viteza maximă

R-squared = 68.73%, modelul este semnificativ la 99%, variabila TopSpeed este semnificativă la 99%.

La o creștere cu 1 km/h a vitezei maxime, prețul mașinii crește cu 649.47 euro.

Reziduurile sunt homoschedastice, nu sunt autocorelate și nu sunt normal distribuite (am corectat eliminând punctele extreme).

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -60570.50    8035.95  -7.537 2.12e-11 ***
TopSpeed      649.47     43.59   14.901 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19180 on 101 degrees of freedom
Multiple R-squared:  0.6873,    Adjusted R-squared:  0.6842 
F-statistic: 222 on 1 and 101 DF,  p-value: < 2.2e-16
```

Regresia multiplă

Prețul în funcție de viteza maximă, autonomie și număr de locuri (dummy 1 = 5 locuri, 0 altfel)

R-squared = 70%, modelul este semnificativ la 99%, variabilele TopSpeed, DummySeats5 și Range sunt semnificate la 99%. Reziduurile sunt homoschedastice la 95%, nu sunt autocorelate și nu sunt normal distribuite (am corectat eliminând punctele extreme).

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -49734.44    8913.16   -5.580 2.10e-07 ***
DummySeats5 -9186.03    4120.50   -2.229  0.0281 *  
TopSpeed      538.22     66.03    8.152 1.13e-12 ***
Range         45.55     22.66    2.010  0.0471 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18690 on 99 degrees of freedom
Multiple R-squared:  0.709,    Adjusted R-squared:  0.7002 
F-statistic: 80.4 on 3 and 99 DF,  p-value: < 2.2e-16
```

Regresia neliniară

Am calculat modelul de regresie multiplă în care am adăugat pătratul și cubul reziduurilor.

LinearHypothesis a demonstrat ca modelul este bine specificat ($F=0.3>0.1$)

Testul Chow a demonstrat ca nu avem rupturi în variabile.

Testul T a demonstrat ca media prețului pentru mașini cu tracțiune față și pentru celelalte este același la 99% ($p\text{-value}<1$).

Am realizat modelul de interacțiune a variabilelor DummyFWD (1 = tracțiune față) și DummyHB (1 = caroserie Hatchback).

Testul F a demonstrat ca regresia cu termeni de interacțiune (accelerație, viteza maximă și caroserie HB) are coeficienți semnificativi diferenți de 0 simultan.

Testul Chow a demonstrat ca modelele create utilizându-se cel anterior cu subset pentru variabila dummy legată de caroserie nu au coeficienții semnificativi simultan).

Modele de regularizare

```
(Intercept) -46615.8972
DummySeats5 -3859.0923
TopSpeed      506.8496
Range         35.6543
```

Ridge R.: R-squared = 83.7%

```
(Intercept) -59008.3467
DummySeats5 -1669.0442
TopSpeed      635.5819
Range         .
```

LASSO R.: R-squared = 88.45%

```
(Intercept) -5.923016e+04
DummySeats5 -2.170033e+03
TopSpeed      6.384709e+02
Range         2.380666e-01
```

Elastic Net R.: R-squared = 88.45%

Boruta

```
model_Boruta1 <- lm(PriceEuro ~ Brand + Accel + TopSpeed + Range + Efficiency
+ PowerTrain + Segment + Seats, hprice1) # R-squared = 98.82%
```

```
model_Boruta3 <- lm(PriceEuro ~ Accel + TopSpeed + Efficiency + Range +
DummySegmentB + DummySeats5, hprice1) # Bonitatea modelului = 91.17%.
```

**PriceEuro = -105700 + 8786 * Accel + 630.7 * TopSpeed + 207.1 * Efficiency - 6.17 * Range +
+ 3352 * DummySegmentB - 2922 * DummySeats5**

Panel

```
fe <- plm(gdp ~ nuclear_consumption + hydro_consumption
+ renewables_consumption, data, index = c('country', 'year'), model = 'within')
```

GDP = -960087024 * nuclear_c -1109690438 *hydro_c +
998311614 * renewables_c

Bonitate: 41.9%.

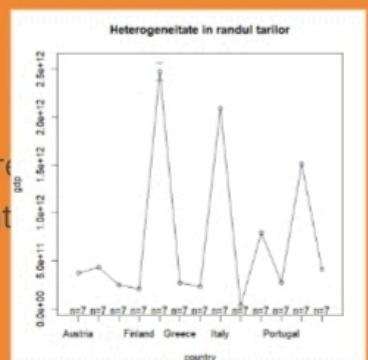
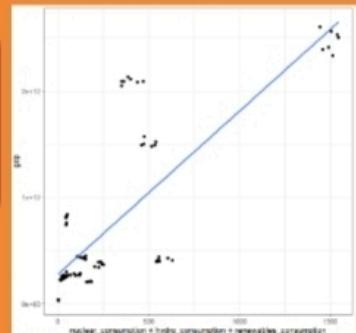
Model semnificativ la 99%. Variabile semnificative la 99%

La o crestere cu o unitate (1 TW/h) a consumului de energie hidroelectrică, GDP crește cu 998.311.614 \$.

Modelul analizează modul în care variabilele independente (consumul energiei produse din diverse resurse) influențează variabila dependentă (GDP).

Conform testului Hausmann, cel mai potrivit model pentru estimare este modelul cu efecte fixe. Conform testului plm, variațiile în timp sunt 0 (există diferențe între țări).

Modelul prezintă dependențe transversale, dar nu și dependențe de timp. Există autocorelare și heteroscedasticitate.



Bibliografie

Sursa seturilor de date:

<https://www.kaggle.com/datasets/pralabhpoudel/world-energy-consumption>
<https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset>

Documentatie R:

<https://www.rdocumentation.org/>

Articole:

<https://www.mdpi.com/2071-1050/11/22/6357>
<https://www.sciencedirect.com/science/article/abs/pii/S1361920921000316>
<https://etrr.springeropen.com/articles/10.1186/s12544-020-00454-2>
<https://www.mdpi.com/2032-6653/10/4/64>
<https://www.sciencedirect.com/science/article/pii/B9780128187623000157>

<https://www.sciencedirect.com/science/article/abs/pii/S0360544220309919>
<https://bibliotekanauki.pl/articles/283192>
<https://www.sciencedirect.com/science/article/abs/pii/S0048969722056911>
<https://www.sciencedirect.com/science/article/abs/pii/S0961953416303063>
<https://www.sciencedirect.com/science/article/abs/pii/S0921800910002399>

