systems.r

Alexandros

2022-10-21

```
#https://github.com/ccolonescu/PoEdata/tree/master/data
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.2
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.1.3
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(systemfit)
```

```
## Warning: package 'systemfit' was built under R version 4.1.3
```

```
## Loading required package: Matrix
```

```
## Warning: package 'Matrix' was built under R version 4.1.3
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
## Loading required package: car
```

```
## Warning: package 'car' was built under R version 4.1.3

## Loading required package: carData

## Warning: package 'carData' was built under R version 4.1.3

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## The following object is masked from 'package:purrr':
##
##     some

## Loading required package: lmtest

## Warning: package 'lmtest' was built under R version 4.1.1

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

##
## Please cite the 'systemfit' package as:
## Arne Henningsen and Jeff D. Hamann (2007). systemfit: A Package for Estimating
Systems of Simultaneous Equations in R. Journal of Statistical Software 23(4), 1-40.
http://www.jstatsoft.org/v23/i04/.
##
## If you have questions, suggestions, or comments regarding the 'systemfit' package,
please use a forum or 'tracker' at systemfit's R-Forge site:
## https://r-forge.r-project.org/projects/systemfit/

library(broom) #for `glance(`) and `tidy()`

## Warning: package 'broom' was built under R version 4.1.2

library(knitr) #for kable()

load("C:\\Users\\Alexandros\\Downloads\\truffles.rda") #load manually
```

```
truffles
```

```
##         p     q    ps    di    pf
## 1   29.64 19.89 19.97 2.103 10.52
## 2   40.23 13.04 18.04 2.043 19.67
## 3   34.71 19.61 22.36 1.870 13.74
## 4   41.43 17.13 20.87 1.525 17.95
## 5   53.37 22.55 19.79 2.709 13.71
## 6   38.52  6.37 15.98 2.489 24.95
## 7   54.33 15.02 17.94 2.294 24.17
## 8   40.56 10.22 17.09 2.196 23.61
## 9   67.35 23.64 22.72 3.885 19.52
## 10  49.65 16.12 15.74 3.169 20.03
## 11  58.17 24.55 24.64 2.623 15.38
## 12  66.87 18.92 23.70 3.007 22.98
## 13  49.95 11.94 15.93 3.367 25.76
## 14  64.95 18.93 23.34 3.290 25.17
## 15  52.68 12.60 15.21 3.746 25.82
## 16  61.20 20.49 26.04 3.518 19.31
## 17  80.55 22.94 22.95 4.381 26.02
## 18  89.94 21.08 27.10 4.121 29.65
## 19  70.77 16.68 23.65 3.820 27.45
## 20  57.33 17.61 20.06 4.398 18.00
## 21  46.23 16.62 26.38 3.764 18.87
## 22  77.43 20.99 24.28 4.524 24.58
## 23  83.01 24.53 26.64 4.815 25.25
## 24  70.71 19.67 22.65 3.670 24.24
## 25  66.75 23.29 19.68 4.392 22.63
## 26  76.80 16.64 23.82 4.603 27.35
## 27  83.70 20.81 28.98 4.632 27.80
## 28  81.00 14.95 18.52 4.894 30.34
## 29  88.44 26.27 28.16 5.125 24.12
## 30 105.45 20.65 28.43 4.836 34.01
```

```
D <- q~p+ps+di
S <- q~p+pf
sys <- list(D,S)
instr <- ~ps+di+pf
truff.sys <- systemfit(sys, inst=instr,
                       method="2SLS", data=truffles)
```

```
summary(truff.sys)
```

```
##
## systemfit results
```

```
## method: 2SLS
##
##         N DF     SSR detRCov   OLS-R2 McElroy-R2
## system 60 53 692.472 49.8028 0.438964   0.807408
##
##      N DF      SSR      MSE    RMSE        R2    Adj R2
## eq1 30 26 631.9171 24.30450 4.92996 -0.023950 -0.142098
## eq2 30 27  60.5546  2.24276 1.49758  0.901878  0.894610
##
## The covariance matrix of the residuals
##          eq1     eq2
## eq1 24.30451 2.16943
## eq2  2.16943 2.24276
##
## The correlations of the residuals
##         eq1     eq2
## eq1 1.00000 0.29384
## eq2 0.29384 1.00000
##
##
## 2SLS estimates for 'eq1' (equation 1)
## Model Formula: q ~ p + ps + di
## Instruments: ~ps + di + pf
##
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) -4.279471   5.543884 -0.77193 0.4471180
## p           -0.374459   0.164752 -2.27287 0.0315350 *
## ps           1.296033   0.355193  3.64881 0.0011601 **
## di           5.013977   2.283556  2.19569 0.0372352 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.92996 on 26 degrees of freedom
## Number of observations: 30 Degrees of Freedom: 26
## SSR: 631.917143 MSE: 24.304505 Root MSE: 4.92996
## Multiple R-Squared: -0.02395 Adjusted R-Squared: -0.142098
##
##
## 2SLS estimates for 'eq2' (equation 2)
## Model Formula: q ~ p + pf
## Instruments: ~ps + di + pf
##
##               Estimate Std. Error  t value   Pr(>|t|)
## (Intercept) 20.0328022  1.2231148  16.3785 1.5543e-15 ***
## p            0.3379816  0.0249196  13.5629 1.4344e-13 ***
```

```
## pf          -1.0009094  0.0825279 -12.1281 1.9456e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.497585 on 27 degrees of freedom
## Number of observations: 30 Degrees of Freedom: 27
## SSR: 60.554565 MSE: 2.242762 Root MSE: 1.497585
## Multiple R-Squared: 0.901878 Adjusted R-Squared: 0.89461
```

```
Q.red <- lm(q~ps+di+pf, data=truffles)
P.red <- lm(p~ps+di+pf, data=truffles)
kable(tidy(Q.red), digits=4,
      caption="Reduced form for quantity")
```

Reduced form for quantity

term	estimate	std.error	statistic	p.value
(Intercept)	7.8951	3.2434	2.4342	0.0221
ps	0.6564	0.1425	4.6051	0.0001
di	2.1672	0.7005	3.0938	0.0047
pf	-0.5070	0.1213	-4.1809	0.0003

```
kable(tidy(P.red), digits=4,
      caption="Reduced form for price")
```

Reduced form for price

term	estimate	std.error	statistic	p.value
(Intercept)	-32.5124	7.9842	-4.0721	4e-04
ps	1.7081	0.3509	4.8682	0e+00
di	7.6025	1.7243	4.4089	2e-04
pf	1.3539	0.2985	4.5356	1e-04

```
truffles
```

```
##         p     q    ps    di    pf
## 1   29.64 19.89 19.97 2.103 10.52
## 2   40.23 13.04 18.04 2.043 19.67
## 3   34.71 19.61 22.36 1.870 13.74
## 4   41.43 17.13 20.87 1.525 17.95
## 5   53.37 22.55 19.79 2.709 13.71
## 6   38.52  6.37 15.98 2.489 24.95
## 7   54.33 15.02 17.94 2.294 24.17
```

```
## 8   40.56 10.22 17.09 2.196 23.61
## 9   67.35 23.64 22.72 3.885 19.52
## 10  49.65 16.12 15.74 3.169 20.03
## 11  58.17 24.55 24.64 2.623 15.38
## 12  66.87 18.92 23.70 3.007 22.98
## 13  49.95 11.94 15.93 3.367 25.76
## 14  64.95 18.93 23.34 3.290 25.17
## 15  52.68 12.60 15.21 3.746 25.82
## 16  61.20 20.49 26.04 3.518 19.31
## 17  80.55 22.94 22.95 4.381 26.02
## 18  89.94 21.08 27.10 4.121 29.65
## 19  70.77 16.68 23.65 3.820 27.45
## 20  57.33 17.61 20.06 4.398 18.00
## 21  46.23 16.62 26.38 3.764 18.87
## 22  77.43 20.99 24.28 4.524 24.58
## 23  83.01 24.53 26.64 4.815 25.25
## 24  70.71 19.67 22.65 3.670 24.24
## 25  66.75 23.29 19.68 4.392 22.63
## 26  76.80 16.64 23.82 4.603 27.35
## 27  83.70 20.81 28.98 4.632 27.80
## 28  81.00 14.95 18.52 4.894 30.34
## 29  88.44 26.27 28.16 5.125 24.12
## 30 105.45 20.65 28.43 4.836 34.01

# step by step ----------------------------------------------------------------

#https://www.fsb.miamioh.edu/lij14/411_note_2sls.pdf
#We hope the good part is big, i.e., the IV and x1 are not weakly related
#It is a good idea to use more IV (over-identification) to isolate bigger exogenous
part of
#the apple



#You can think of x1
#as a partially rotten apple consisting of two parts: the bad
#endogenous part (correlated with u) and the good exogenous part (uncorrelated with u)

#. Endogeneity issue arises when the key regressor is correlated with the error term.
#This can happen when (i) there are omitted variables; (ii) there is reverse causation
or
#simultaneity; (iii) there is measurement error
#. In the presence of endogeneity, OLS estimator is biased

modelD=lm(data=truffles,q~p+ps+di)
modelS=lm(data=truffles,q~p+pf)
```

```
cor(modelD$residuals,truffles$p)
```

```
## [1] 1.114444e-16
```

```
#https://www.eco.uc3m.es/~ricmora/MEI/materials/Session_09_2SLS.pdf


#In the rst stage, we regress each endogenous regressor on all
#exogenous variables and compute the predictions y^j


IVQmodel=lm(data=truffles,q~ps+di+pf) ### regress only on variables that affect D, thus
isolate the good part
IVQ=IVQmodel$fitted.values

IVpmodel=lm(data=truffles,p~ps+di+pf)
IVp=IVpmodel$fitted.values



#In the second stage, we regress the dependent variable on all
#exogenous regressors and the predictions y^j
smr=summary(truff.sys)

modelD=lm(data=truffles,q~IVp+ps+di)

modelD %>% summary

##
## Call:
## lm(formula = q ~ IVp + ps + di, data = truffles)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1814 -1.1390  0.2765  1.4595  4.4318
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.27947    3.01383  -1.420 0.167505
## IVp         -0.37446    0.08956  -4.181 0.000291 ***
## ps           1.29603    0.19309   6.712 4.03e-07 ***
## di           5.01398    1.24141   4.039 0.000422 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 2.68 on 26 degrees of freedom
## Multiple R-squared:  0.6974, Adjusted R-squared:  0.6625
## F-statistic: 19.97 on 3 and 26 DF,  p-value: 6.332e-07
```

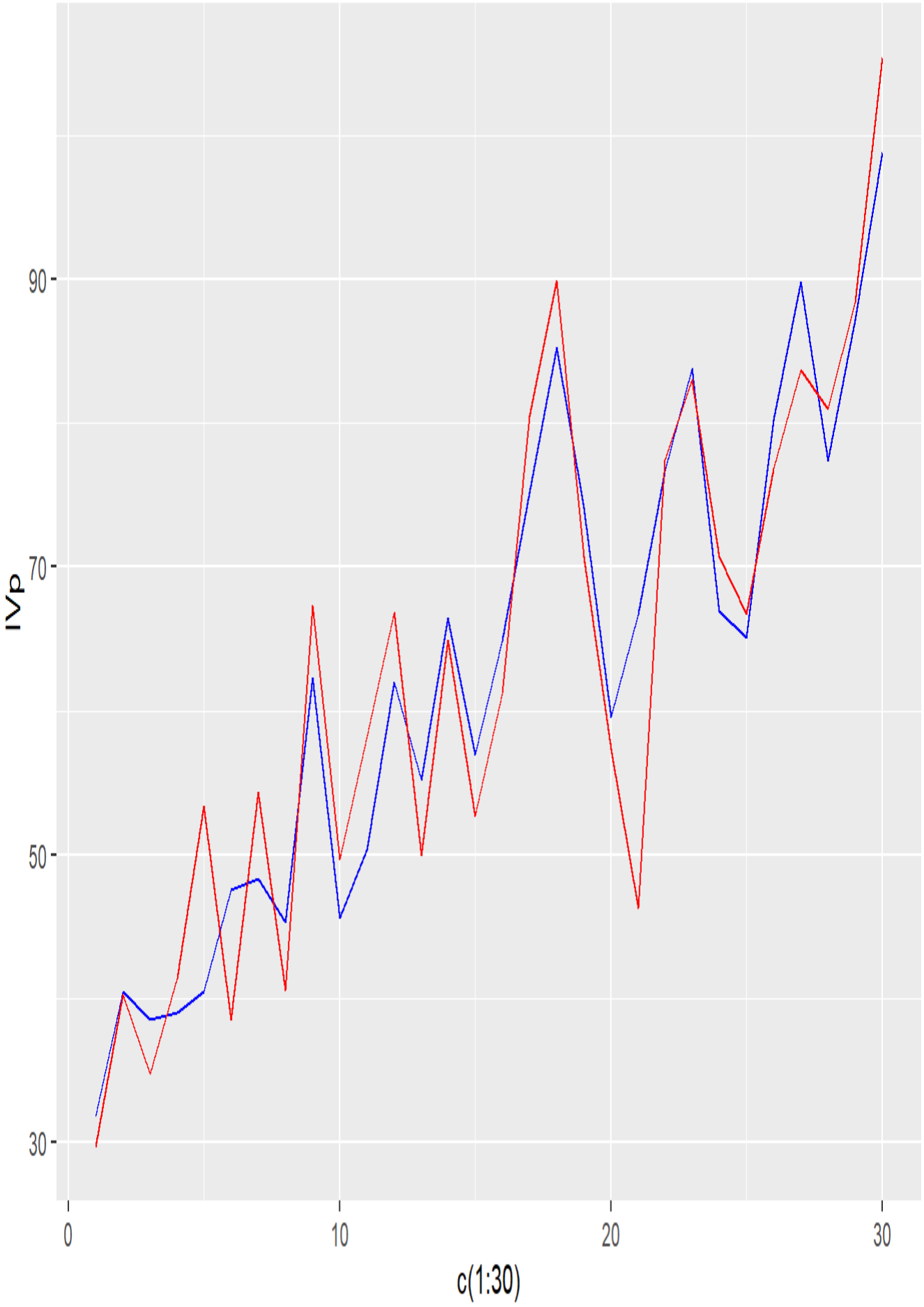
```
smr$eq[1]
```

```
## [[1]]
##
## 2SLS estimates for 'eq1' (equation 1)
## Model Formula: q ~ p + ps + di
## Instruments: ~ps + di + pf
##
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) -4.279471   5.543884 -0.77193 0.4471180
## p           -0.374459   0.164752 -2.27287 0.0315350 *
## ps           1.296033   0.355193  3.64881 0.0011601 **
## di           5.013977   2.283556  2.19569 0.0372352 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.92996 on 26 degrees of freedom
## Number of observations: 30 Degrees of Freedom: 26
## SSR: 631.917143 MSE: 24.304505 Root MSE: 4.92996
## Multiple R-Squared: -0.02395 Adjusted R-Squared: -0.142098
```

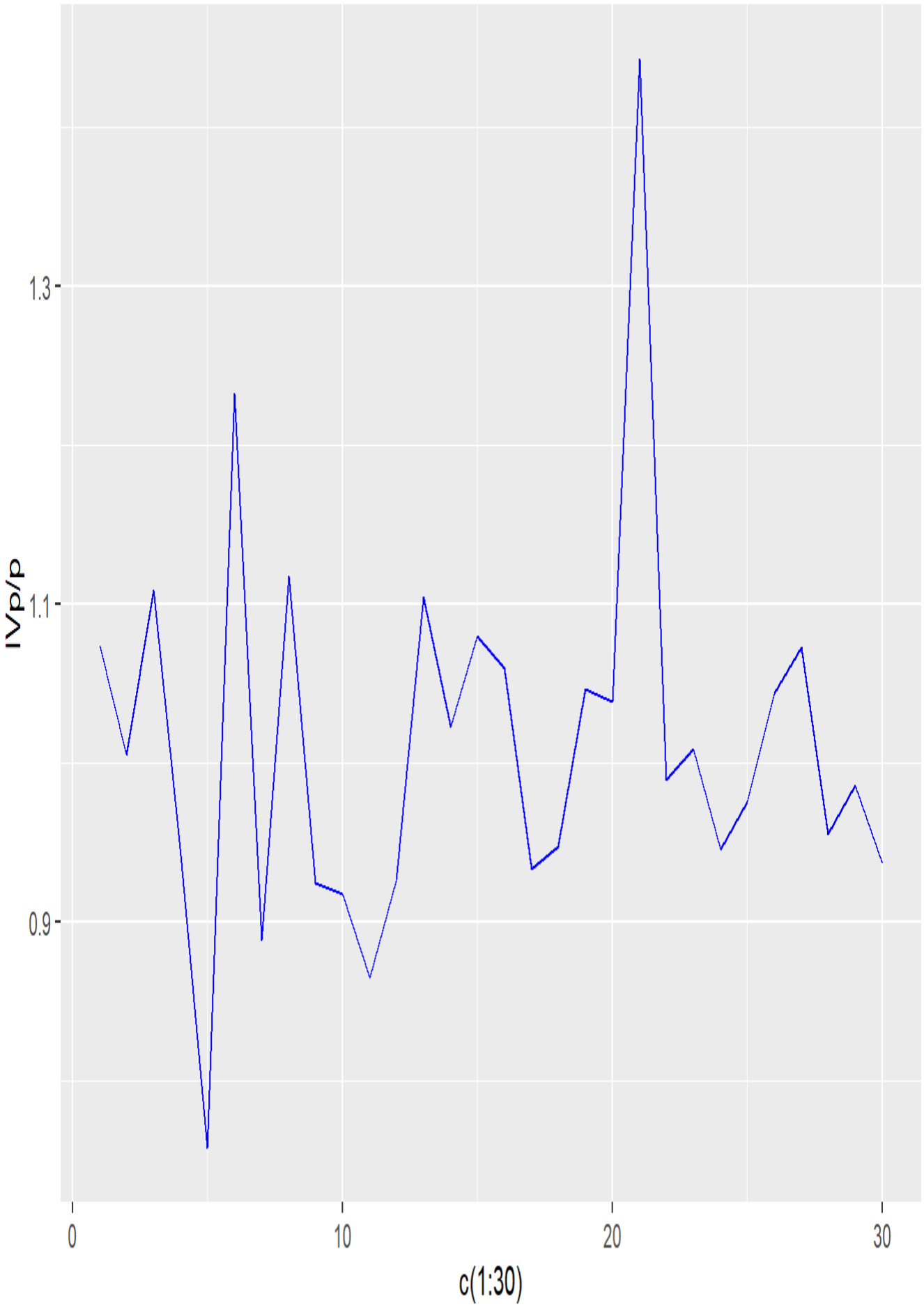
```
modelS=lm(data=truffles,q~IVp+pf)
modelS %>% summary
```

```
##
## Call:
## lm(formula = q ~ IVp + pf, data = truffles)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0732 -0.9754  0.5228  1.8115  3.8940
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 20.03280    2.16570    9.25 7.36e-10 ***
## IVp          0.33798    0.04412    7.66 3.07e-08 ***
## pf          -1.00091    0.14613   -6.85 2.33e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.652 on 27 degrees of freedom
## Multiple R-squared:  0.6924, Adjusted R-squared:  0.6696
```

```
## F-statistic: 30.38 on 2 and 27 DF,  p-value: 1.226e-07

smr$eq[2]

## [[1]]
##
## 2SLS estimates for 'eq2' (equation 2)
## Model Formula: q ~ p + pf
## Instruments: ~ps + di + pf
##
##               Estimate Std. Error  t value   Pr(>|t|)
## (Intercept) 20.0328022  1.2231148  16.3785 1.5543e-15 ***
## p            0.3379816  0.0249196  13.5629 1.4344e-13 ***
## pf          -1.0009094  0.0825279 -12.1281 1.9456e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.497585 on 27 degrees of freedom
## Number of observations: 30 Degrees of Freedom: 27
## SSR: 60.554565 MSE: 2.242762 Root MSE: 1.497585
## Multiple R-Squared: 0.901878 Adjusted R-Squared: 0.89461

library(ggplot2)

ggplot(data=truffles,aes(x=c(1:30)))+
  geom_line(aes(y=IVp),color="blue")+
  geom_line(aes(y=p),color="red")
```


90
70
50
30
IVp
0
10
20
30
c(1:30)

```
##blue line shows the exogenous part of p
##red is total p

ggplot(data=truffles,aes(x=c(1:30)))+
  geom_line(aes(y=IVp/p),color="blue")
```


IVp/p
1.3
1.1
0.9
0
10
20
30
c(1:30)

```
## In summary, regress the exogenous variables on the endogenous to be used in the regression in order
## to isolate the exogenous part, and the run regressions

#Hausman test for endogeinity
```