Health Expenditures in Greece: A Multiple Least Squares Regression and Cointegration Analysis Using Bootstrap Simulation in EVIEWS

Giovanis, Eleftherios

2009

Health Expenditures in Greece: A Multiple Least Squares Regression and Cointegration Analysis Using Bootstrap Simulation in EVIEWS

Eleftherios Giovanis

ABSTRACT

This paper examines the factors that are contributing at the most explained and efficient way to health expenditures in Greece. Two methods are applied. Multiple regressions and vector error correction models are estimated, as also unit root tests applied to define in which order variables are stationary. Because the available data are yearly and capture a small period from 1985-2006, so the sample is small, a bootstrap simulation is applied, to improve the estimations.

Keywords: health expenditures, bootstrapping regression, Ordinary Least Squares, Vector Error Correction Model, EVIEWS

INTRODUCTION

Previous studies examined the health expenditures and the factors that explain them. Newhouse (1977) estimated a cross-section regression of per capita medical care expenditure on GDP per capita of 13 countries including Greece. He found that only GDP per capita can explain almost the 92 per cent of regression. The main problem of OLS estimation is that if the residuals of this regression are not stationary, then the OLS estimations are not reliable. Because in the period where Newhouse (1977) wrote his paper cointegration analysis hadn't yet been developed, he estimates with simple OLS regression. Also Newhouse (1977) found the elasticity exceeds unit, including and excluding, Greece and for various levels of GDP per capita. So this good is luxury according to those results. Dreger and Reimers (2005) applied panel unit root tests and cointegration analysis for 21 OECD countries and they found that health expenditures are not determined only by income, but also another driving force can be the medical progress, observed in the evolution of other variables, as the infant mortality and life expectancy. So following this thought we examine more variables to determine the health expenditure in Greece.

DATA

The data are yearly and capture the period from 1985 to 2006. Specifically the variables that examined are total health expenditures per capita in US dollars, which variable is defined as the dependent. The explanatory variables are the Gross Domestic Product expressed in US dollars, the acute care beds, density per 1,000 population, the average length of stay for acute care, all conditions, in days, cancer deaths per 100,000 females and per 100,000 males, life expectancy at 65 years old (in years), for females and males, the hospital beds, density per 1,000 population, infant mortality-deaths per 1,000 live births, practicing physicians, density per 1,000 population and practicing nurses, density per 1,000 population. The data have been obtained by OECD.

METHODOLOGY

The problem which arise is that we have chosen many variables and we have a short sample to estimate, which is it will be unreliable, as most of the estimating parameters are found statistically insignificant. To solve this problem we apply two approaches to reduce the number of variables. The first one is cluster analysis, where first we must find the similarity measures between the variables and this can be done with the commonly correlation coefficient distance measure.

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (1)$$

Ward's cluster method objective is to minimize the sum of squares of the deviations from the mean value as Žiberna *et al*. (2004) have applied.

$$ESS = \sum_i \sum_j \sum_k \left| X_{ijk} - \overline{x}_{ik} \right| \quad (2)$$

The second approach is to apply factor analysis with principal components and Varimax rotation to improve the extractions of factors. These approaches are applied in all variables, except from the GDP per capita, because from the beginning is the most significant explanatory variable. After the selection process of the variables, which will take part in the estimation we apply a multiple OLS regression with bootstrap simulation. The steps for the bootstrap simulation are the following.

First step: A random sample of *n* observation from the following relation is randomly selected with replacement

$$e = Y - X'\beta \tag{3}$$

, where e are the residuals, Y is the dependent variable and $X'$are the various independent variables. This sample is defined as e*

Second Step: The bootstrap sample becomes

$$Y^* = X'\beta + e^* \tag{4}$$

Third step: We estimate with OLS equation (4)

Fourth Step: We repeat steps 1-3 3000 times according to information criteria

A similar procedure is followed for bootstrapping simulation in VECM analysis, where the bootstrapped residuals are more than one, depending on the number of variables examined. For example if have two variables to estimate a VECM, then we need two bootstrapped residual series.

After OLS estimation with bootstrap simulation we apply White's General Heteroskedasticity test (Gujarati, 2004) for the heteroskedasticity presence, Ljung-Box test (Greene, 2002) for autocorrelation presence and ARCH-LM test (Greene, 2002) for testing ARCH effects. As data are yearly we don't expect to find ARCH effects, as usually these effects are present in high-frequency data like daily and weekly data.

The next model which is estimated is a Vector Error Correction model (VECM), to examine if there is a long run relationship and equilibrium between the health expenditure and the other selected variables. Before that a unit root test must be applied to examine in what order variables are stationary. The most common unit root test is the Dickey-Fuller GLS test (Elliot *et al*., 1996), which we apply in our case. After defining the stationary order of the variables a cointegration test and Full Information Maximum Likelihood (FIML) estimation based on Johansen procedure is followed. We examine two tests (Johansen, 1995) to determine the number of co-integrating vectors. The first is the Johansen trace statistic. We test the null hypothesis

$H_0(r): r = r_0$ against the alternative hypothesis $H_1(r): r > r_0$

The trace statistic is defined as:

$$LR_{trace}(r_0) = -T \sum_{i=r_0+1}^{n} \ln(1-\hat{\lambda}_i) \quad (5)$$

The second LR statistic is known as the maximum eigenvalue statistic and is defined as:

$$LR_{max.eigen}(r_0) = -T \sum_{i=r_0+1}^{n} \ln(1-\hat{\lambda}_{r_0+1}) \quad (6)$$

, and we test the null hypothesis $H_0(r_0): r = r_0$ against the alternative hypothesis $H_1(r_0): r_0 > r_0+1$.

RESULTS

In figure 1 the results of the cluster analysis with Ward linkage method are presented, while in table 1 we report the results from the factor analysis. From both methods we conclude that acute care beds average length of stay, infant and hospital beds can be treated as one variable, with negative response to health expenditures. This can be shown in figure 1 and the factor 1 column of table 1. On the other hand life expectancy at 65 old years, for both males and females as also the number of practicing physicians can be treated as the second variable, as it can be shown again in

factor 1 column of table 1. From figure 1 and factor 3 and 4 columns in table 1 variables of cancer deaths for males and females are chosen as the third variable. Finally practicing nurses is considered as a separate variable. So with this procedure the number of variables is reduced in a significant degree, which means for example from the variables of cancer death for males and females we need to take only one variable of them.

The final selection is based on the regression *t-statistics* to test about the significance to choose a variable or not. For the computations we obtain the logarithmic values for all variables to get directly the elasticities. Another conclusion based on the variance explained by the factor is that only first factor could be chosen as it explains 61.7 per cent of the total variance. So with this logic two variables can be accepted –except GDP per capita- rather four according to the previous approach. We estimate two regressions. One with only GDP per capita as explanatory variable and one with life expectancy at 65 years old, infant and cancer deaths as explanatory variables, because if we obtain simultaneously all variables then GDP elasticity becomes much more lower than 0.88, which can be unreliable. In table 2 we present the bootstrap simulated OLS regression results with GDP per capita as explanatory. From a first aspect one could conclude that health expenditure is not a luxury good, but if we consider also the confidence intervals, which are (0.73, 1.03) at 5 per cent significance level, then health expenditure can be a luxury good. Constant is statistically insignificant. In table 3 we present the bootstrap simulated regression results with Life expectancy at 65 years old, infant and cancer deaths as explanatory variables. The specific variables explain 97.5 per cent of the total regression, according to R^2 adjusted. All coefficients have the expected signs, while as the cancer deaths and life expectancy at 65 years old increase then health expenditure increase too, in a way to finance the life extension and cancer therapy. Also the constant, in this case too, is not statistically significant. In figure 2 we present the line graphs for variables examined to investigate if there is a trend or not. In table 4 ADF test results are presented. All variables are stationary in first differences, except from cancer deaths, which is *I(2)*, so all variables are *I(1)*. Life expectancy variable is *I(0)* at 10 per cent, but we prefer to be stationary at least at 5 per cent significance level.

Figure 1: Dendrogram with Ward Linkage and Correlation Coefficient Distance

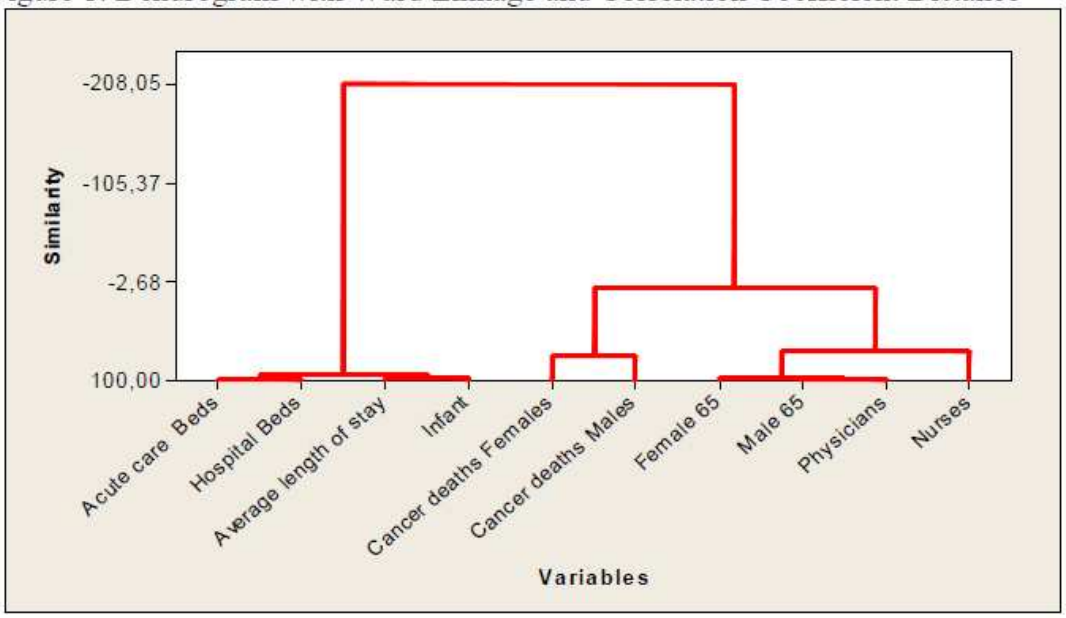

Table 1. Factor analysis with principal components

Variable	Factor1	Factor2	Factor3	Factor4	Factor5
ACUTE CARE BEDS	-0.835	-0.174	-0.28	-0.099	0.429
AVERAGE LENGTH OF STAY	-0.916	-0.305	-0.194	0.088	0.078
CANCER DEATHS (FEMALES)	-0.139	0.084	0.253	0.954	-0.011
CANCER DEATHS (MALES)	0.155	0.087	0.947	0.265	-0.040
LIFE EXPECTANCY (F65)	0.901	0.399	0.055	-0.041	0.067
LIFE EXPECTANCY (M65)	0.956	0.248	0.010	-0.068	0.098
HOSPITAL BEDS	-0.908	-0.179	-0.180	0.040	0.295
INFANT	-0.982	-0.084	-0.052	0.117	0.006
PHYSICIANS	0.980	0.146	0.062	-0.092	-0.052
NURSES	0.345	0.927	0.096	0.105	-0.041
Variance	6.173	1.278	1.1277	1.0379	0.2973
%	0.617	0.128	0.113	0.104	0.030

Table 2. OLS bootstrap simulation results with GDP as explanatory variable

Constant	GDP per capita		
-1.098	0.881	$R^2_{adjusted}$ = 0.858	LBQ^2 (12) = 4.935
(0.721)	(0.077)	AIC = -0.725	Pr =0.960
[-1.521]	[11.337]*	SBC = -0.626	F-statistic for
		LL = 9.977	White test = 0.0816
		F-statistic of	Pr = 0.9219
		regression=128.535	F-statistic for
		Pr =0.000	ARCH-LM (1) test=0.040
			Pr =0.8432

Standard errors in parentheses, t-statistics in brackets.,* denotes significance in 1% level, AIC and SBC refer to Akaike and Schwarz information criteria, LL is the Log Likelihood, LBQ^2 is the Ljung-Box test on squared standardized residuals with 12 lags, ARCH-LM test for ARCH effects with 1 lag

Table 3. OLS bootstrap simulation results with life expectancy at 65 years old, infant and cancer deaths as explanatory variables

Constant	Life expectancy at 65 years old	Infant	Cancer deaths
-14.282	4.573	-0.650	1.859
(8.773)	(1.931)	(0.187)	(1.005)
[-1.627]	[2.368]**	[-3.475]*	[1.848]***
$R^2_{adjusted}$ = 0.975	LBQ^2 (12) = 10.98	F-statistic for	F-statistic for
AIC = -2.200	Pr =0.530	White test = 1.480	ARCH-LM (1)
SBC = -2.002		Pr = 0.2515	test = 1.232
LL = 28.209			Pr = 0.2808

Standard errors in parentheses, t-statistics in brackets.,* denotes significance in 1% level, ** denotes significance in 5% level , *** denotes significance in 10% level AIC and SBC refer to Akaike and Schwarz information criteria, LL is the Log Likelihood, LBQ2 is the Ljung-Box test on squared standardized residuals with 12 lags, ARCH-LM test for ARCH effects with 1 lag

In addition we present the unit root test results for the residuals of the two multiple regressions and we conclude, as it was expected, that residuals are not stationary in levels and they are *I(1)*. So regression results might be unreliable. So for this reason cointegration analysis will be applied to investigate the long run relationship between health expenditure and the factors that have been selected. In figure 2 (a) we observe that health expenditure follows almost the same trend with GDP, which can be assumed to be linear. On the other hand in figure 1 (b) life expectancy and cancer deaths do not present trend, while with DF-GLS test we found that life expectancy and cancer deaths are *I(1)* and *I(2)* respectively. Infant presents a smoothly negative trend. For the first set of estimation, health expenditure and GDP

variables, the lag order is four, according to information criteria, while for the second set, which includes health expenditure, life expectancy, infant and cancer deaths variables the lag order has been found equal with one. For the first set we chose with intercept and linear trend, according to figure 2 (a) and there is one cointegrating relation. For the second set and based on test results we chose no intercept and no trend. In table 6 are presented the VECM estimations. We observe that there is a cointegration relation [1 -1.44] between health expenditures per capita and per capita GDP, according to panel A, so health is a luxury good. In panel B the VECM estimations for the other variables are presented and they all have the correct signs, and all estimations in both panels, are statistically significant, expect from life expectancy. So we conclude that there is a long run relationship between health expenditure and the variables we examine and so health expenditures are not determined only by per capita GDP.

Figure 2. Line graphs for a) Health expenditure per capita in US dollars and per capita GDP in US dollars

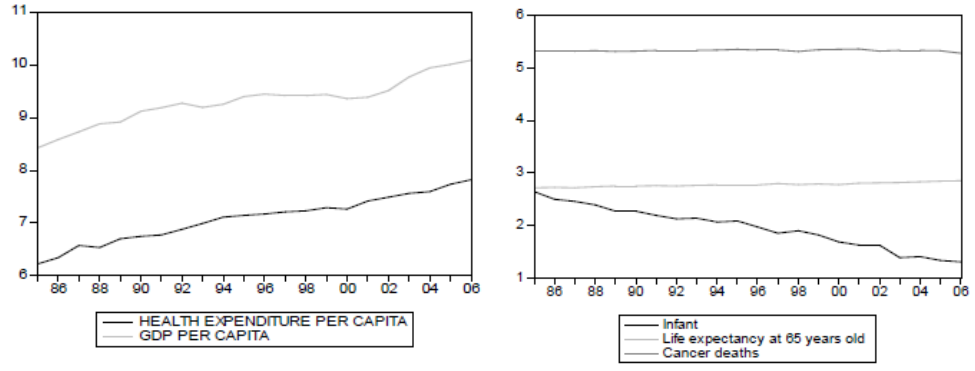

Table 4. DF-GLS unit root tests

Variable	Unit root test-statistic (DF-GLS)	Critical values
Residuals from regression of table 2 in levels	-3.184	
Residuals from regression of table 2 in first differences	-5.821	
Residuals from regression of table 3 in levels	-1.393	
Residuals from regression of table 3 in first differences	-5.919	1% -3.770 5% -3.190 10% -2.890
Health Expenditure per capita in $ in levels	-2.828	
Health Expenditure per capita in $ in first differences	-6.155	
GDP per capita in $ in levels	-2.244	
GDP per capita in $ in first differences	-3.076	
Life expectancy at 65 years old in levels	-3.146	
Life expectancy at 65 years old in first differences	-6.846	
Cancer deaths in levels	-1.561	
Cancer deaths in first differences	-1.740	
Cancer deaths in second differences	-5.440	

Table 5. Cointegration tests

	First Set	Second Set
Data Trend: Test Type	Linear Intercept Trend	None No Intercept No Trend
Trace	1	1
Max-Eig	1	1

Table 6. VECM estimations

A			
Health Exp(-1)	GDP(-1)	Trend	Constant
1.0000	-1.444	0.021	6.122
	(0.084)	(0.005)	
	[-17.179]	[3.999]	
B			
Health Exp(-1)	Life expectancy (-1)	Infant(-1)	Cancer deaths(-1)
1.000	-0.658	0.926	-1.346
	(1.097)	(0.216)	(0.657)
	[-0.599]	[4.283]	[-2.048]

CONCLUSION

We applied a simulated bootstrap multiple regression with OLS and we found that health expenditure is not a luxury good. We mentioned the problems that OLS regression confronts, like residuals are not stationary in levels, so regression results might be unreliable. We solved this problem testing for cointegration and applying a simulated bootstrap VECM to investigate simultaneously the long run relationship between health expenditure and the factors that have been chosen. We concluded that there is actually a long run relationship and in addition health expenditure is a luxury good, according to long run GDP elasticity.

REFERENCES

Dreger, C. and Reimers, H-E. (2005). Health Care Expenditures in OECD Countries: A Panel Unit Root and Cointegration Analysis, *IZA, Discussion Paper Series*, 1469

Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient Tests for an Autoregressive Unit Root, Econometrica, Vol. 64 No. 4, pp. 813-836

Greene, W.H.(2002). *Econometric Analysis*, New Jersey: Prentice Hall

Gujarati D. (2004). *Basic Econometrics*, New York: McGraw Hill

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, New York: Oxford University Press

Newhouse J.P. (1977). Medical-Care Expenditure: A Cross National Survey. The Journal Of Human Resources, Vol. 12 No. 1, pp.115-125

Žiberna, A., Kejžar, N. and Golob, P. (2004). A Comparison of Different Approaches to Hierarchical Clustering of Ordinal Data. *Metodološki zvezki*, Vol. 1 No. 1, pp.57-73

Appendix

EVIEWS programming routine for OLS and standard bootstrapping in residuals

For table 2

```
'estimate regression

eq1.ls health_exp c gdp

'get residuals to bootstrap
eq1.makeresid res1

'estimate model
eq1.makemodel(mod1)
mod1.solve

mod1.addassign(i) @all

'Monte Carlo parameters
!reps =500
rndseed(type=mt) 1234567

'bootstrap loop
for !i=1 to !reps
'bootstrap residuals
smpl @all if res1<>na
res1.resample health_exp_a
'generate bootstrap data
mod1.solve
'estimate OLS with bootstrap data

eq2.ls  health_exp_0 c gdp_0
eq2.makemodel(mod2)
mod2.solve
next
```

For table 3 exactly the same just we change the independent variables

```
'estimate regression

eq1.ls health_exp c life_exp infant death_can

'get residuals to bootstrap
eq1.makeresid res1

'estimate model
eq1.makemodel(mod1)
mod1.solve

mod1.addassign(i) @all

'Monte Carlo parameters
!reps =500
rndseed(type=mt) 1234567

'bootstrap loop
```

```
for !i=1 to !reps
'bootstrap residuals
smpl @all if res1<>na
res1.resample health_exp_a
'generate bootstrap data
mod1.solve
'estimate OLS with bootstrap data

eq2.ls  health_exp_0 c life_exp_0 infant_0 death_can_0
eq2.makemodel(mod2)
mod2.solve
next
```

EVIEWS programming routine for VECM and standard bootstrapping in residuals

For panel A in table 6

```
var1.ec(c,1) 1 4  health_exp  gdp @ c

'get residuals to bootstrap
var1.makeresid(n=gres) res1 res2

'make model out of estimated VAR
var1.makemodel(mod1)

mod1.solve

'assign add factors for bootstrap residuals
mod1.addassign(i) @all

'set monte carlo parameters
!reps = 500                                          'bootstrap replications


'set random number generator
rndseed(type=mt) 1234567

'bootstrap loop
for !i=1 to !reps
        'bootstrap residuals
        smpl @all if res1<>na

        gres.resample health_exp_a

        'generate bootstrap data
        mod1.solve

        'estimate VAR with bootstrap data
        smpl @all

        var var2.ec(c,1) 1 4  health_exp_0 gdp_0  @ c
```

```
        'make model from bootstrap estimates
        var2.makemodel(mod2)

        smpl @all
        mod2.solve

next
```

For panel B in table 6

```
var1.ec(a,1) 1 1  health_exp  life_exp infant death_can  @ c

'get residuals to bootstrap
var1.makeresid(n=gres) res1 res2 res3 res4

'make model out of estimated VAR
var1.makemodel(mod1)

mod1.solve

'assign add factors for bootstrap residuals
mod1.addassign(i) @all

'set monte carlo parameters
!reps = 500                                          'bootstrap replications


'set random number generator
rndseed(type=mt) 1234567

'bootstrap loop
for !i=1 to !reps
        'bootstrap residuals
        smpl @all if res1<>na

        gres.resample health_exp _a

        'generate bootstrap data
        mod1.solve

        'estimate VAR with bootstrap data
        smpl @all

        var var2.ec(a,1) 1 1  health_exp_0 life_exp_0 infant_0 death_can_0  @ c

        'make model from bootstrap estimates
        var2.makemodel(mod2)

        smpl @all
        mod2.solve

next
```