SPECIAL ARTICLE

Testing the effect of specific socioeconomic factors on the ischemic mortality rate. The case of Greece

Mouza A M

Technological Educational Institute of Serres, Department of Business Administration, Terma Magnisias, 621 24, Serres, Greece

Abstract

In this paper we present a model to evaluate the effect of certain majors socioeconomic factors (such as alcohol and fat consumption, cigarettes smoking, unemployment rate as a proxy for uncertainty which results frustration, number of passenger cars as a proxy for physical exercise and per capita GDP as a proxy for nutrition quality), to the ischemic mortality rate. Since the existing research works on this field, suffer from the proper model testing, we analytically present all the tests necessary to justify the reliability of the result obtained. For this purpose, after specifying and estimating the model, we applied the specification error test, the linearity, multicollinearity and heteroscedasticity tests, the autocorrelation and stability tests and the ARCH effect test.

Finally, we present the aggregate efect of the above socioeconomic factors. In brief, we found that an increase of cigarettes smoked, of fat and alcohol consumption and the number of passenger cars will result to a relevant increase regarding mortality. The latter one is also affected by the changes in unemployment rate. On the other hand, an increase of personal disposable income may negatively affect mortality, by almost the same portion. Hippokratia 2008; 12 (1): 59-64

For terminology explanation see endnotes in page 63

Keywords: model estimation, model testing, ischemic mortality rate, socioeconomic factors

Corresponding author: Mouza A M, Technological Educational Institute of Serres, Department of Business Administration, Terma Magnisias, 621 24, Serres, Greece, e-mail: mouza@teiser.gr, tel: +323210 49178, fax number: +323210 49178

There is a fundamental need to explain changes in morbidity and mortality as they are related to changes in socioeconomic factors, given that at national level, the socioeconomic status of the population is usually the most powerful health predictor. With this in mind, the main aim of this paper is to present a detail methodology as a guideline for the purpose of specification, estimation and testing a model, which enables us to determine the effect of certain specific social economic factors, on the ischemic mortality rate. In this context, in the model presented below we do not consider any medicine oriented or individual health status effects, or any medical treatment effects, but only certain major socioeconomic factors, to explain ischemic mortality rate.

Given that till now, there is no any detailed quantitative study for the case of Greece on this particular field, we decided to present this model which gives a clear picture regarding the extend that each of the main factors considered, influences the above mortality rate. The methodology of model testing, integrates this attempt to correctly assess the significance of specific etiological factors on the population health status, as far as the ishaemic disease is concerned.

On theoretical grounds, the choice of the particular factors considered here is based upon relevant research works discussing the effect of fat and alcohol consumption, cigarette smoking, quality of nutrition, physical exercise and frustration, on various heart diseases[1-5].

Model specification

The model specification is based on similar works[6-12], where the great majority of risk factors biochemical, behavioral stress or physical and health care utilization, are directly or indirectly related to economic growth factors and socioeconomic status. There is an indirect effect of many factors of this kind on the mortality changes. Which factors are to be considered in model specification, depends on their importance as well as the available data. In our case, this combination produced the following structural form, which is presented after estimation by OLS (Ordinary Least Squares) from annual data obtained from various sources as indicated in table 1.

$$\hat{MORTALITY}_i = \underset{(2.961)}{11.732} + \underset{(0.241)}{0.5216}\ ALCOHOL_i + \underset{(0.0036)}{0.03743}\ CARS_{i-2}$$

p values: 1.89E-3 5.1E-2 2.35E-7

$$+ \underset{(0.30013)}{0.7772}\ CIGARS_{i-3} + \underset{(0.258)}{0.59838}\ \Delta UNR_i + \underset{(0.18899)}{0.35747}\ FATS_{i-3}$$

p values: 2.37E-2 3.9E-2 8.29E-2

$$-\underset{(0.4332)}{1.41337}\ PerCaGDP_{i-2} \tag{1}$$

p value: 6.79E-3

$\bar{R}^2 = 0.97$, ESS =0.66229, s^2= 0.05519, d =2.665, F =101.8, PC = 0.0755, BIC =-2.272

where $MORTALITY_i$ *= Ischemic mortality rate at year i.*

$ALCOCHOL_i$ = Per capita average annual consumption of alcohol at year i.
$CARS_{i-2}$ = Total number of passenger cars (national level), two years earlier.
$CIGARS_{i-3}$ = Per capita average annual consumption of cigarettes, three years earlier.
$\Delta UNR_i = UNR_i - UNR_{i-1}$ (UNR_i = unemployment rate at year i).
$FATS_{i-3}$ = Per capita fat and land animal consumption, three years earlier.
$PerCaGDP_{i-2}$ = Per capita Gross Domestic Product (in real terms), with two years time lag.

It is recalled that in (1) the numbers in brackets are standard errors of the corresponding estimated coefficients, $\overline{R}^2$ the adjusted coefficient of determination, d is the Durbin-Watson statistic, PC is the Amemiya criterion and BIC is the Bayesian Information Criterion, known also as Schwarz Bayesian Criterion. Note also that p values refer to the lowest level of significance (α), for the corresponding coefficient. ESS $= \sum \hat{u}_i^2$ is the error sum of squares, where $\hat{u}_i$ are the residuals from (1). The OLS estimate of the disturbance's variance, is denoted by s^2.

Although the d statistic reveals the possible existence of a slight negative first order autocorrelation, which will be discussed later, we may conclude from (1), since all coefficients seem to be significant at an acceptable α, that at national (aggregate) level, the influence of the major socio-economic factors considered here on mortality rate, can be summarized in what follows.

- We observe a positive effect of the quantity of alcohol consumed during the current year.
- There is a cumulative effect of the cigarettes smoked and the fat consumed during the last three year.
- Also there is a cumulative effect of the increase in the number of passenger cars, during the last two years, since the more intensive use of private cars, restrains physical exercise and increases stress.
- The current change in unemployment rate seems to have a positive effect too, through increasing uncertainty in many individuals, who in turn feel more frustrated.
- The increase of per capita GDP in the last two years seems to have a negative effect, mainly due to the improvements in nutrition quality, the better exercise and health care too.

More details regarding the model variables and further calculations to facilitate the computation of individual elasticities, are presented in table 1.

Table 1. Main characteristics of the variables used

Variables	Mean value	Ratios of means	Units of measurement	Source of data
$MORTALITY_i$ (dependent)	10.7844	1.0	(Deaths/Total population)*10000	National Statistical Service of Greece
$ALCOHOL_i$	2.0632	0.1913	Per capita annual consumption in kgr	National Statistical Service of Greece
$CARS_i$	135.5263	12.5668	Number of passenger cars per 1000 inhabitants	National Statistical Service of Greece
$CIGARS_i$	2.4778	0.2297	Average annual consumption of cigarettes (in '000) per capital, for all ages	National Statistical Service of Greece
ΔUNR_i	0.1737	0.0161	UNR=Total unemployment, as percent of total population	OECD Health Data 99
$FATS_i$	3.0684	0.2845	Average annual consumption o food fats & land animal (kgr per capita)	National Statistical Service of Greece. OECD Health Data 99
$PerCaGDP_i$	7.2332	0.6707	Constant 1995 prices in '000 Euro (market prices)	Ministry of National Economy of Greece.

Model testing

The majority of the empirical works on this field suffer from the lack of further tests[13-16], which establish the grounds to conclude that the estimates we obtained have the desired statistical properties and the specified model is not going to lead us towards misleading inference. With this in mind, we suggest the following tests.

Test for specification error

It has been pointed out, that the number of explanatory variables selected, heavily depends on available data. This implies that we may exclude some other variables, which play an important role in explaining the variation of the dependent variable. In this case we have a specification error, resulting to biased end inconsistent estimates. We have tested this kind of specification error, with the following two ways.

1. By computing the recursive residuals, denoted by $\hat{u}_j^0$ (j = m+1,.....T), with mean $\overline{u}^0$. It is recalled, that m denotes the number of coefficients in (1), and T the number of observations actually used in the estimation process. From these residuals we calculate the statistic ψ from

$$\psi = \left[[T-(m+1)]^{-1} \times \sum_{j=m+1}^{T} (\hat{u}_j^0 - \overline{u}^0)^2 \right]^{-1/2} \times (T-m)^{-1/2} \times \sum_{j=m+1}^{T} \hat{u}_j^0 \quad (2)$$

which follows the Student's t distribution, with T-(m+1) degrees of freedom, under the null hypothesis of the correct model specification. It is clear that a large value of $|\psi|$, indicates the existence of specification error[17,18]. In our case we found that $\overline{u}^0 = -0.8199$, wiht standard error = 0.684 and ψ = -0.1186, indicating that no specification error exists.

2. By applying the RESET (Regression Specification Error Test) of Ramsey[19]. For simplicity, denote the estimated values of the dependent variable $\hat{y}_i$. Then, consider $\hat{y}_i$ as the dependent variable in new (auxiliary) regressions, with explanatory variables $\hat{y}_i^2$, or $(\hat{y}_i^2, \hat{y}_i^3)$, or $(\hat{y}_i^2, \hat{y}_i^3, \hat{y}_i^4)$ and all the independent variables in (1) plus a constant term. From the ESS^* of these regressions we consider the largest F^* which is computed from

$$F^* = \frac{(ESS - ESS^*)/k}{ESS^*/(T-m)} \quad \text{[where } k=1\ (\hat{y}_i^2)\text{, or } k=2\ (\hat{y}_i^2, \hat{y}_i^3)\text{, or } k=3 \text{ (i.e. } \hat{y}_i^2, \hat{y}_i^3, \text{and } \hat{y}_i^4)]$$

and follows the F distribution with [k, T-(m+k-1)] degrees of freedom, where m denotes the number of coefficients of the regression with the largest F^*. In our case the largest F^* corresponds to the auxiliary regression with $\hat{y}_i^2, \hat{y}_i^3$ and $\hat{y}_i^4$ as explanatory variables. Given that F^* is much smaller than the tables F, for α=0.05 and 0.01, we reach to the same conclusion, that no specification error exists.

Linearity test

It is recalled that linearity is one of the basic assumptions, in order to apply OLS for model estimation. This test presupposes the estimation of the regression

$$\hat{u}_i = \underset{(378.62)}{-93.8555} + \underset{(144.92)}{37.2965}\ \hat{y}_i - \underset{(20.672)}{5.51265}\ \hat{y}_i^2 + \underset{(1.3024)}{0.359166}\ \hat{y}_i^3 - \underset{(0.035854)}{0.0087043}\ \hat{y}_i^4$$

to obtain $R^2 = 0.0165753$ and LM = TXR^2, where LM is the Lagrange multiplier, which follows the X^2 distribution with k-1 (k=3) degrees of freedom. Given that LM is much less than the tables X^2 for any standard significance level, we may conclude, that no problem regarding linearity exists.

Multicollinearity test

Given that some of the explanatory variables have a very small total variation (0.12E-2 and 0.86F-2) the use of condition number (which is the largest condition index) will produce misleading results[20]. In such cases it is preferable to employ the Variance Inflation Factor (VIF), for testing multicollinearity. VIF has to be calculated for each explanatory variable i (i = 1, 2,...., m-1) taken as dependent and the remaining as independent, in order to get the corresponding R_i^2. For variable i, the VIF_i is computed from $VIF_i = \dfrac{1}{1-R_i^2}$

The results obtained are presented in table 2.

Table 2. Estimated Variance Inflation Factors (VIF_i) and the corresponding R^2.

Variable (as dependent)	R^2	VIF
$ALCOHOL_i$	0.668	3.012
$CARS_{i-2}$	0.878	8.196
$CIGARS_{i-3}$	0.399	1.664
ΔUNR_i	0.516	2.066
$FATS_{i-3}$	0.537	2.159
$PerCaGDP_{i-2}$	0.764	4.237

Since none of the VIF values exceeds 10, we conclude that there is no any severe collinearity problem, regarding the columns of data matrix X of the explanatory variables, which implies that X has full column rank.

Heteroscedasticity test

To apply this test, we computed the Spearmans's rank correlation coefficient r_s, considering each explanatory variable and the (absolute values of) residuals. The obtained coefficients, together with the computed t^* values (given that the standard error is 0.2357), are presented in table 3.

Table 3. The computed rank correlation coefficients r_s and the corresponding t^* values

Independent variables	r_s	t^* statistic (abs. value)
$ALCOHOL_i$	0.214	0.908
$CARS_{i-2}$	-0.268	1.137
$CIGARS_{i-3}$	0.202	0.857
ΔUNR_i	-0.311	1.319
$FATS_{i-3}$	0.116	0.492
$PerCaGDP_{i-2}$	-0.514	2.180

Since the last t* statistic is greater than the tables t, for (T-m) degrees of freedom at α=0.05, we see that there is a problem of heteroscedastisity of the form:

$$\sigma_i^2 = (PerCaGDPi)^2,, \text{ where } \sigma_i^2 = Var(ui)$$

In such a case we have to transform the initial model, dividing throughout by $PerCaGDP_i$ and to re-estimate it. The results are presented bellow.

$$\hat{YSTAR}_i = \underset{(3.0259)}{11.8692}\ ONEovGDP_i + \underset{(0.2508)}{0.522075}\ ALCovGDP_i +$$

$$+ \underset{(0.0037)}{0.037549}\ CARovGDP_i + \underset{(0.3122)}{0.75834}\ CIGARovGDP_i +$$

$$+ \underset{(0.2668)}{0.606567}\ \Delta UNRovGDP_i + \underset{(0.19847)}{0.35267}\ FATovGDP_i - \underset{(0.4447)}{1.42638} \quad (3)$$

$\bar{R}^2 = 0.958$, d =2.68

where

$YSTAR_i = MORTALITY_i/PerCaGDP_i$
$ONEovGDP_i = 1/PerCaGDP_i$
$ALCovGDP_i = ALCOHOL_i/PerCaGDP_i$
$CARovGDP_i = CARS_{i-2}/PerCaGDP_i$
$CIGARovGDP_i = CIGARS_{i-3}/PerCaGDP_i$
$\Delta UNRovGDPi = \Delta UNR_i/PerCaGDP_i$
$FATovGDP_i = FATS_{i-3}/PerCaGDP_i$

It should be noted that we avoided applying the general test of White[21], to test for heteroscedasticity, since there are too many explanatory variables in the model.

Before going back to the original form of the model seeing in (1), by multiplying (3) by $PerCaGDP_i$, it is advisable to perform one more test, regarding autocorrelation.

Autocorrelation and stability test

Given the value of d statistic in (3), one may suspect that there is a problem of first order (negative) autocorrelation. Applying ordinary Durbin-Watson test, we found $d_L<d<d_U$, so that the test was inconclu-

sive. However, according to other tests (Brausch-Godfrey), it seems that the problem of autocorrelation is present. At this point it is useful to recall that we may have some indications for autocorrelation, although the real problem is an ARCH (AutoRegressive Conditional Heteroscedasticity) effect. To be sure that we have to face the problem as being a pure autocorrelation problem, we should trace any ARCH effect, by performing the proper test. With this in mind, and considering that the order h (maximum lag) is 3, we get the following estimates.

$$\hat{u}_i^2 = 0.00088 + 0.089\hat{u}_{i-1}^2 - 0.114\hat{u}_{i-2}^2 - 0.059\hat{u}_{i-3}^2 \quad R^2 = 0.0252$$

We further compute LM = (T-h)xR^2, which is much smaller than the tables X^2, for h =3 degrees of freedom, so that we may conclude that the error terms do not follow an ARCH scheme, and thus we are facing a pure autocorrelation problem. The application of Hildreth-Lu iterative procedure to get the minimum ¦d-2¦, produces the following results.

$$YS\hat{T}AR_i = \underset{(2.36169)}{12.02903}\, ONEovGDP_i + \underset{(0.21868)}{0.6300074}\, ALCovGDP_i$$

$$+ \underset{(0.0029399)}{0.0387744}\, CARovGDP_i + \underset{(0.24064)}{0.769779}\, CIGARovGDP_i$$

$$+ \underset{(0.22749)}{0.699451}\, \Delta UNRovGDP_i + \underset{(0.1629)}{0.36783}\, FATovGDP_i - \underset{(0.352828)}{1.515414} \quad (4)$$

$$\overline{R}^2 = 0.98,\ d = 2.01,\ \hat{\rho} = \underset{(0.2283)}{-0.59}$$

To test the coefficients stability, we consider the recursive residuals and compute the CUSUM (cumulative sums) and CUSUMSQ (cumulative sums of squared residuals) statistics, presented in 1 and 2. From these graphs one may conclude that no problem of coefficients stability exists.

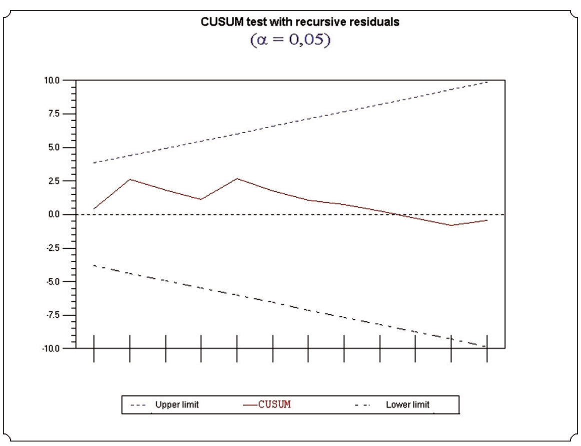

Figure 1.

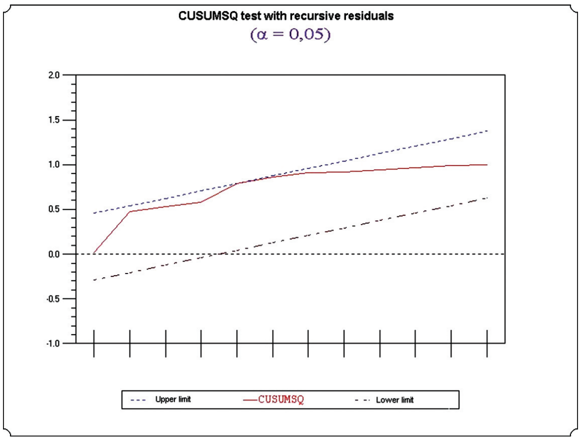

Figure 2.

To transform model (4) to its original form, which is model (1), we multiply (4) throughout by $PerCaGDP_i$ to obtain

$$MORT\hat{A}LITY_i = \underset{(2.36169)}{12.02903} + \underset{(0.21868)}{0.6300074}\, ALCOHOL_i + \underset{(0.0029399)}{0.0387744}\, CARS_{i-2}$$

p values: 2.65E-4 1.38E-2 1.68E-8

$$+ \underset{(0.24064)}{0.769779}\, CIGARS_{i-3} + \underset{(0.22749)}{0.699451}\, \Delta UNR_i$$

p values: 7.65E-3 9.63E-3

$$+ \underset{(0.1629)}{0.36783}\, FATS_{i-3} - \underset{(0.352828)}{1.515414}\, PerCaGDP_{i-2} \quad (5)$$

p values: 4.34E-2 1.041E-3

With the new estimates we get 3, which refers to the observed and estimated values of the dependent variable.

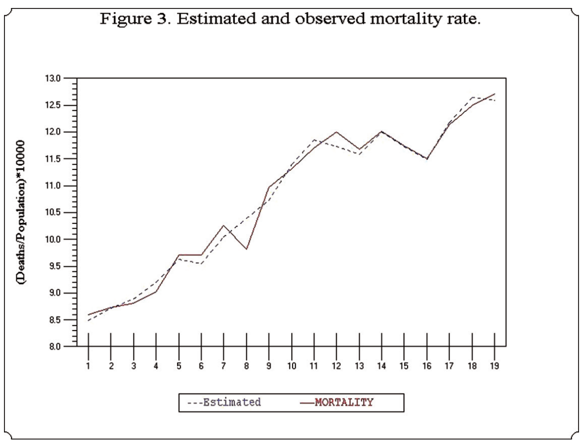

Figure 3.

Discussion

In this paper a relevant model has been specified, estimated and properly tested, in order to investigate the quantitative effect of specific socioeconomic factors on the ischemic mortality rate. We presented step by step all the tests required in order to obtain correct and unbiased results which reflect the real world in the best possible way. We also discussed the effect of each factor on the dependent variable in a rather general manner. Further we may obtain some more detailed results, if we combine the ratios of means of the explanatory variables over the mean value of the dependent variable presented in table 1, together with the estimated coefficients in (5), in order to get estimates of the elasticities. Thus we may determine the extend that each explanatory variable affects mortality, which is the dependent variable in our model. These elasticities are presented in table 4, bellow.

Table 4. Estimated elasticities

Variables	Ratios of means	Estimated coefficient	Estimated elasticity
$ALCOHOL_i$	0.1913	0.6300074	0.12052
$CARS_{i-2}$	12.5668	0.0387744	0.48727
$CIGARS_{i-3}$	0.2297	0.769779	0.17682
ΔUNR	0.0161	0.699451	0.01126
$FATS_{i-3}$	0.2845	0.36783	0.10465
$PerCaGDP_i$	0.6707	-1.515414	-1.01638

It is noted, that all positive elasticities are less than one and only the (absolute) value of the last one is slightly greater than unity. According to the last column of table 4, we can say that a 10% increase of cigarettes smoked (i.e. 100 cigarettes), will result to a 1.77% increase in mortality. This harmful effect of smoking will be realized with a delay of three years. Also an increase of fat consumption by 10%, will gradually increase mortality by 1%, in the following three years An increase in alcohol consumption by 10% (i.e. 100 kgr) will increase mortality rate by 1.2% without any delay, i.e. in the same time period. A similar interpretation holds for the other socioeconomic factors considered, which constitute the set of the explanatory variables. In other words, a 1% increase of the number of passenger cars, will result to a gradual increase in mortality by 0.48%, which is to be realized two years later. Changes in unemployment rate by 1%, will affect mortality by 0.1%, at the same time period, through increasing frustration. On the other hand, an increase of personal disposable income by 1%, may contribute towards reducing mortality, by almost the same portion. This effect is due to the expected improvements in nutrition quality, the better exercise and health care too.

Finally it is worthy to mention that we haven't met in the literature an equivalent quantitative analysis on this particular field.

Endnotes

1. From the mathematical point of view, a model is a mathematical expression to describe a real phenomenon. In this context, the flight of an air plain can be simulated by a single differential equation. On the other hand, to model a national economy one may need more than a hundred of difference equations. A model is used for simulation purposes, and mainly to investigate the effect of some predetermined factors on the characteristic we are interested in.

2. The right specification is the first step in model building. This implies that we must know in advance the factors (usually called explanatory variables), which actually affect the evolution of the characteristic under consideration (usually called dependent variable).

3. A model should comply with some specific tests, in order to be efficient and reliable. All these tests are described in details and properly applied in this paper, given that there is a lack of such an analysis in the relevant literature.

References

1. Hansagi H, Romelsjö A, Garhardsson de Verdier M et al. Alcohol consumption and stroke mortality, 20-year follow-up of 15,077 men and women. Stroke 1995;26: 1768-1773
2. Shaper AG, Wannamethee SG. Alcohol intake and mortality in middle aged men with diagnosed coronary heart disease. Heart 2000;83: 394-399
3. Mukamal KJ, Ascherio A, Mittleman MA, et al. Alcohol and risk for ischemic stroke in men: the role of drinking patterns and usual beverage. Ann Intern Med 2005;142: 11-19
4. Wannamethee SG, Shaper AG. Patterns of alcohol intake and risk of stroke in middle-aged British men. Stroke 1997;27: 1033-1039
5. Johansson SE, Sundquist K, Ovist J, Sundquist J. Smokeless tobacco and coronary heart disease: a 12-year follow-up study. Eur J Cardiovasc Prev Rehabil 2005;12: 387-392
6. Brenner MH. The economic change model of pathology. Health projections in Europe. Methods and applications. World Health Organization 1986
7. Brenner MH. Economic, alcohol consumption and heart disease mortality in nine industrialized countries. Soc Sci Med 1987;25: 119-132
8. Kauhanen J, Kaplan G, Colberg D, Salonen J. Beer Bingeing and Mortality: Results from Kuopio Ischaemic Heart Disease Risk Factor Study, a Prospective Population Based Study. BMJ 1997;315: 846-851
9. Lozano R, Murray CJ, Lopez AD, Satoh T. Miscoding and misclassification of ischaemic heart disease mortality. World Health Organization September, 2001
10. Murray C, Lopez A. The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries and Risk Factors in 1990 and Projected to 2020. Harvard University Press, 1996;Vol: 1
11. Peto R, Lopez A, Boreham J, Thun M. Heath C. Mortality from Tobacco in Developed Countries: Indirect Estimation from National Vital Statistics. Lancet 1992;339: 1268-1278
12. Zatonski W. Alcohol and Health: What is good for the French may not be for the Russians. J Epidemiol Community Health 1998;52: 766-767
13. Critchley JA, Unal B. Is smokeless tobacco a risk factor for coronary heart desease? A systematic review if epidemiological studies. Eur J Cardiovasc Prev Rehabil. 2004;11: 101-112
14. Asplund K, Nasic S, Janlert U, Stegmayr B. Smokeless tobacco as a possible risk factor for stroke in men: a nested case-control study. Stroke 2003;34: 1754-1759
15. Huhtasaari F, Lundberg V, Eliassin M, et al. Smokeless tobacco as a possible risk factor for myocardial infraction: a population-

based study in middle-aged men. J Am Coll Cardiol 1999;34: 1784-1790
16. Milosević M, Mustajbegović J, Abdivić S, et al. Relationship of regional alcohol drinking habit and cardiovascular risk factors in Croatia. Acta Med Croatica 2007;61: 275-279
17. Harvey AC. The Econometric Analysis of Time Series. Oxford: Philip Allan Publishers, 1981
18. Harvey AC. The Econometric Analysis of Time Series (2nd edition). Massachusetts: The MIT Press: Cambridge, 1990
19. Ramsey JB. Testing for Specification Errors in Classical Linear Least Squares Regression Analysis. J R Stat Soc 1969;31: 350-371
20. Lazaridis A. A Note Regarding the Condition Number: The Case of Spurious and Latent Multicollinearity. Qual Quant 2007; 41: 123-135
21. White H. A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity. Econometrica 1980;48: 817-838