15. INFORMATIONAL CRITERIA FOR THE HOMOSCEDASTICITY OF ERRORS

Daniel CIUIU[1]

Abstract

In this paper we will test the homoscedasticity of errors using the Goldfeld-Quandt test and we will classify the points using the explanatory variable by which we sort them. We will also use the Hartley test for the equality of the class error variances (if we have at least two classes).

For all the points (only one class) and all the possible classifications for which we have homoscedasticity we will compute some informational criteria like Akaike (AIC=Akaike Informational Criterion) and Schwartz (BIC=Bayes Informational Criterion). Of course, from these classifications we will choose that classification with the minimum of the considered criterion.

As application, we have monthly data from November 1990 to November 2008 concerning the price indexes for services, the price indexes for food and for the price indexes of non-food goods.

Keywords: homoscedasticity, classification, informational criteria, price indexes
JEL Classification: C12, C52, L16

1. Introduction

Consider *n* points in $\mathbf{R}^{k+1}$, $X^{(1)},...,X^{(n)}$, where $X^{(i)} = \left(X_1^{(i)}, X_2^{(i)},..., X_k^{(i)}, Y_i\right)$. The regression hyper-plane used in Ciuiu (2007a) to classify patterns has the equation (Saporta, 1990):

$$\mathrm{H} : \mathrm{Y} = \mathrm{A}_0 + \sum_{i=1}^{k} \mathrm{A}_i \mathrm{X}_i \tag{1}$$

such that $\sum_{i=1}^{n} u_i^2$ is minimum, where the residues u_i have the formula:

$$\mathrm{u}_i = \mathrm{Y}_i - \mathrm{A}_0 - \sum_{j=1}^{k} \mathrm{A}_j \mathrm{X}_j^{(i)}. \tag{1'}$$

[1] Technical University of Civil Engineering, Bucharest, dciuiu@yahoo.com; Associate Researcher, Romanian Institute of Economic Forecasting, Romanian Academy.

For the computation of A_i from (1) we have to solve the system (Saporta, 1990)

$$\sum_{j=0}^{k} \overline{X_i \cdot X_j} \cdot A_j = \overline{X_i \cdot Y},\ i = \overline{0,k}\ , \tag{2}$$

where: $\overline{X_0 \cdot X_i} = \overline{X_i}$ and $\overline{X_0^2} = 1$.

The polynomial model (Ciuiu 2007a) is in fact the multi-linear model (1) with the explanatory variables $X_1 = X$, $X_2 = X^2$ and so on.

For the obtained estimators of A_i using (2) and of the residues u_i we have the following hypotheses (Jula, 2003, Voineagu *et al*., 2007):

1) The estimators of A_i are linear.
2) The estimators of u_i have the expectation 0 and the same variance (homoscedasticity).
3) The estimators of u_i are normal.
4) The random variables u_i are independent.

From the above hypotheses and from Gauss-Markov theorem we obtain the following properties (Jula, 2003, Voineagu *et al*., 2007):

1) The estimators of A_i are consistent.
2) The estimators of A_i are unbiased.
3) The estimators of A_i have the minimum variance.
4) The estimators of A_i have the maximum likelihood.

A test for the homoscedasticity of errors is the Goldfeld-Quandt test (Jula 2003, Voineagu *et al*., 2007). For this test we have to find first an explanatory variable X_j positively correlated with the squares of the residues u^2.

After reordering the n points increasing on X_j we divide the n-order sample into three parts: first, n_1 points, next, n_3 points (these points are removed) and last, n_2 points. The number of the removed points from the middle of the sample, n_3, is arbitrary, but in practice it is between $\frac{n}{6}$ and $\frac{n}{5}$.

After we have estimated the coefficients of the regression A_i for both parts of the $n - n_3$ points and we have computed the sum of squares of the residues for these parts:

$$\begin{cases} VTR_1 = \sum\limits_{t=1}^{n_1} u_t^2 \\ VTR_2 = \sum\limits_{t=n-n_2+1}^{n} u_t^2 \end{cases}. \tag{3}$$

Finally we compute the statistics

$$F_c = \frac{VTR_2(n_1 - k - 1)}{VTR_1(n_2 - k - 1)}, \tag{4}$$

and we replace F_c with $\frac{1}{F_c}$ if $F_c < 1$.

If the null hypothesis of homoscedasticity is true the statistics F_c is Snedecor-Fisher distributed, with $n_2 - k - 1$ and $n_1 - k - 1$ degrees of freedom (respectively $n_1 - k - 1$ and $n_2 - k - 1$ degrees of freedom if we make the above substitution). Therefore we compare F_c with the centil of the order $1-\varepsilon$ of the Snedecor-Fisher distribution $F_{n_2-k-1, n_1-k-1}(1-\varepsilon)$, and we accept the null hypothesis if and only if (Jula, 2003, Voineagu *et al*., 2007):

$$F_c < F_{n_2-k-1, n_1-k-1}(1-\varepsilon). \tag{5}$$

Suppose we have *m* samples $X_{i,1},...,X_{i,n_i}$ on the random variables $N(m_i, \sigma_i^2)$, with $i = \overline{1,m}$. The Hartley test of variances equality test the null hypothesis $H_0 : \sigma_1^2 = ... = \sigma_m^2$ against the alternative hypothesis H_1 : there exists $i \neq j$ such that $\sigma_i^2 \neq \sigma_j^2$. If $S_i'^2$ are the unbiased estimators of σ_i^2

$$S_i'^2 = \frac{\sum_{j=1}^{n_i}(X_{i,j} - m_i)^2}{n_i - 1} \tag{6}$$

we compute the statistics

$$F_{max} = \max_{i,j} \frac{S_i'^2}{S_j'^2}. \tag{7}$$

We accept the null hypothesis if and only if

$$F_{max} < F_{k_1,k_2}(1-\varepsilon), \tag{8}$$

where the numerator and the denominator in (7) are computed using k_1 and k_2 degrees of freedom.

An informational criterion for (1) is the Akaike criterion, AIC (Jula, 2003):

$$AIC = \frac{VTR}{n} \cdot e^{\frac{2(k+1)}{n}} = \frac{\sum_{t=1}^{n} u_t^2}{n} \cdot e^{\frac{2(k+1)}{n}}, \tag{9}$$

or, in logarithmic expression:

$$\ln AIC = \ln\left(\frac{\sum_{t=1}^{n} u_t^2}{n}\right) + \frac{2(k+1)}{n}. \tag{9'}$$

Another informational criterion for (1) is the Schwartz criterion, BIC:

$$BIC = \frac{VTR}{n} \cdot n^{\frac{k+1}{n}} = \frac{\sum_{t=1}^{n} u_t^2}{n} \cdot n^{\frac{k+1}{n}}, \quad (10)$$

or, in logarithmic expression:

$$\ln BIC = \ln\left(\frac{\sum_{t=1}^{n} u_t^2}{n}\right) + \frac{k+1}{n} \cdot \ln n. \quad (10')$$

For the $ARMA_{p,q}$ models these criteria are as follows (Popescu, 2000). The Akaike criterion is:

$$AIC = -2\ln L_z\left(\hat{\beta}, \hat{\sigma}_a^2\right) + 2(p+q), \quad (11)$$

where: $\hat{\beta}$ is the vector of estimated parameters, $\hat{\sigma}_a^2$ is the estimation of the variance of the corresponding white noise a_t, and $L_z\left(\hat{\beta}, \hat{\sigma}_a^2\right)$ is the maximum likelihood.

The Schwartz criterion is:

$$BIC = (n-p-q)\ln\frac{n\hat{\sigma}_a^2}{n-p-q} + n\ln\left(1+\ln\sqrt{2\pi}\right) + (p+q)\ln\frac{n\left(\hat{\sigma}_z^2 - \hat{\sigma}_a^2\right)}{p+q}, \quad (12)$$

where: $\hat{\sigma}_z^2$ is the sample variance of the initial time series z_t.

Apparently these criteria are different in the time series cases, but we can approximate them by:

$$\begin{cases} AIC \approx n\ln\hat{\sigma}_a^2 + 2(p+q) \\ BIC \approx MDL = n\ln\hat{\sigma}_a^2 + (p+q)\ln n \end{cases}, \quad (13)$$

where: MDL is the Minimum Description Length of Rissanen (Popescu, 2000).

2. The implementation of the tests and classifications

Suppose now we have *n* points $\left(X_{i,1},...,X_{i,k},Y_i\right)_{i=\overline{1,n}}$. When we apply the Goldfeld-Quandt test we must find the explanatory variable X_j positively correlated with the squares of the residues u^2, and we sort the points on X_j.

Between the classification into only one class (all the *n* points) and the classifications with many points we choose only those for which we obtain homoscedasticity by the Goldfeld-Quandt test. If $nrcls > 1$ we use also the Hartley test to verify if the classes have the same errors.

If we take into account the logarithmic expressions (9') and (10') and we multiply by *n*, and the approximations (13) we can notice that in both cases (regression and time series) we can write:

$$\begin{cases} \mathrm{AIC} = \alpha + 2 \cdot \mathrm{n}_{\mathrm{par}} \\ \mathrm{BIC} = \alpha + \mathrm{n}_{\mathrm{par}} \cdot \ln \mathrm{m} \end{cases}, \tag{14}$$

where: n_{par} is the number of the estimated parameters and *m* is the number of degrees of freedom if we do not take into account the constraints (in the case of regression the number of degrees of freedom is $n-k-1$, and in the case of time series this number is $n-p-q$).

In our case of the Goldfeld-Quandt test we have homoscedasticity if and only if $F_c < F_{n_1,n_2}(1-\varepsilon)$. In the case of more classes we compute first *frstatfmax*$= \max_{\text{class } i} F_i(F_c)$, where F_i is the Snedecor-Fisher cdf of the class *i*. Of course, if *nrcls*=1 we have *frstatfmax*=F_c, and we accept the homoscedasticity for all classes if *frstatfmax* $< F_{n_1,n_2}(1-\varepsilon)$. We consider only these classifications. We take now into account (14) and the *AIC* formula in the case of time series and we define the Akaike criterion *AIC:*

$$AIC = -2\ln(1-\varepsilon - frstatf\max) + 2 \cdot nrcls \cdot (k+1), \tag{15}$$

and the Schwartz criterion *BIC:*

$$BIC = -2\ln(1-\varepsilon - frstatf\max) + nrcls \cdot (k+1) \cdot \ln(n_1+n_2), \tag{16}$$

where: n_1 and n_2 are from the class for which we have obtained $frstatf\max$.

When we find a classification with all the classes homoscedastic we can compute the maximum number of classes:

$$\max \mathrm{nrcls} = \frac{\mathrm{AIC} + 2\ln(1-\varepsilon)}{2 \cdot \mathrm{k} + 2} \tag{17}$$

in the case of *AIC* criterion, and

$$\max \mathrm{nrcls} = \frac{\mathrm{BIC} + 2\ln(1-\varepsilon)}{(\mathrm{k}+1)\ln(2 \cdot \mathrm{k} + 4)} \tag{17'}$$

in the case of *BIC* criterion. Using this estimation we can decrease the maximum number of classifications, and we reduce the computations.

Taking into account the formulae (15) and (16) we can also reduce the value of the limit of *frstatfmax*: if we consider the same number of classes and we increase the value of *frstatfmax* we obtain a higher value for *AIC* and for *BIC,* respectively. When the number of classes increases this limit (initially $1-\varepsilon$) decreases more: the limit *lim* changes by the formula

$$\lim' = 1 - \varepsilon - (1 - \varepsilon - \lim) \cdot e^{k+1} \tag{18}$$

if we consider the Akaike criterion and

$$\lim' = 1 - \varepsilon - (1 - \varepsilon - \lim) \cdot (2k + 4)^{\frac{k+1}{2}} \qquad (18')$$

if we consider the Schwartz criterion.

3. Applications

Example 1. Consider the price indexes for services (resulting variable Y), for food (explanatory variable X_1) and non-food goods (explanatory variable X_2) from November 1990 to November 2008 (Buletin statistic de prețuri, Nr. 11, 2008).

The above price indexes are presented in the table in Appendix *A*.

For this application we will refer only to the Simpson method as a numerical one, even we can also use in the $C++$ program the rectangles ant trapezes method, because it is the most precise (Păltineanu *et al.*, 1998). As a Monte Carlo method we will refer only to the Box-Muler method, because it is the most rapid (Văduva, 2004). For both tests (Goldfeld-Quandt and Hartley) we take the first degree error $\varepsilon = 0.025$.

The regression plane for all the $n = 217$ points is $Y = 13.56971 + 0.32625X_1 + 0.54608X_2$. To find X_j positive correlated to u^2 we compute first the correlation coefficients r_j between X_j and u^2, and we obtain $r_1 = 0.30971$ and $r_2 = 0.53864$. Therefore we can take $X_j = X_2$.

The number of removed points in the Goldfeld-Quandt test must be between $\frac{n}{6}$ and $\frac{n}{5}$, and we take this number as $n_3 = \frac{n}{12} + \frac{n}{10} = \frac{11n}{60}$ (obviously the integer part if the value is not integer). The maximum number of classes is such that after removing the points from each class and the division of the remained points into two subclasses for the Goldfeld-Quandt test we must have in each subclass at least $\dim + 1 = 4$ points. This maximum number of classes is obtained as 22, hence we cannot classify the 217 points into 23 classes or more.

The method used in the $C++$ program is the backtracking method: we start with the classifications into two classes, next we continue with the classifications into three classes and so on. We have to allocate for each class at least $2 \cdot k + 5$ points (9 points in our case), and we do not test the homoscedasticity for the class $cl + 1$ (we do not even allocate points for the class) if the class cl is heteroscedastic. From the order on X_2, the classes have successive points each.

In the following table, we have in the first column the considered number of classes, in the second one the regression planes, in the third one the Goldfeld-Quandt statistics, in the fourth one the c.d.f. of this statistics, in the fifths one whether we accept the homoscedasticity of all classes, in the sixth one the AIC criterion value, and in the last one the BIC criterion value.

Table 1

No. of Classes	Regression planes	Goldfeld-Quandt			AIC	BIC
		GQ	F(GQ)	Accepting homosceda-sticity		
1	Y=13.56971+0.32625 X_1+ 0.54608X_2 (217)	24.59494	0.99999 0.998 0.9971	No		
2 (first)	Y=-78.03959+1.41667 X_1+ 0.3674 X_2 (9)	6.77169	0.73998 0.7677 0.7669	Yes		
	Y=13.54615+0.32601 X_1+ 0.54653 X_2 (208)	26.83126	0.99999 0.9975 0.99725	No		
2	First class ≤110 points, except 23 or 24 points			Yes No		
2	First class >110 points, 23 or 24 points			No Does not matter		
3 (first)	Y=29.38961+0.1628 X_1+ 0.54838 X_2 (83)	1.75032	0.93774 0.9336 0.939	Yes	28.59766 30.14969	38.81711 37.59655
	Y=7.23348+0.23186 X_1+ 0.70516 X_2 (106)	1.56621	0.92078 0.92 0.9212	Yes		
	Y=17.72303+0.62749 X_1+ 0.21271 X_2 (28)	4.08696	**0.97** **0.9727** **0.9658**	Yes		
3 (min AIC, Simpson)	Y=9.30803+0.13661 X_1+ 0.77452 X_2 (104)	1.09037	0.60592	Yes	**20.88016**	22.94869
	Y=17.58411+0.2191 X_1+ 0.61932 X_2 (102)	1.22832	**0.73809**	Yes		
	Y=36.00538+1.19725 X_1- 0.48176 X_2 (11)	1.98242	0.67415	Yes		
3 (min AIC, Box-Muler)	Y=5.76933+0.14521 X_1+ 0.80108 X_2 (105)	1.07003	0.5885	Yes	**20.84924**	22.7401
	Y=17.52307+0.2191 X_1+ 0.61987 X_2 (101)	1.23505	**0.7344**	Yes		
	Y=36.00538+1.19725 X_1- 0.48176 X_2 (11)	1.98242	0.7029	Yes		
3 (min BIC)	Y=9.09023+0.15893 X_1+ 0.75432 X_2 (107)	1.23715	**0.75065** **0.748**	Yes	20.9891 20.96561	**22.76415** **22.74063**
	Y=18.60039+0.21799 X_1+ 0.61104 X_2 (99)	1.19699	0.70706 0.7059	Yes		
	Y=36.00538+1.19725 X_1- 0.48176 X_2 (11)	1.98242	0.67415 0.7073	Yes		

In the above table in the cells with 3 numbers (one class or first classification with two classes, column F(GQ)) the first value is obtained by the Simpson method even we use the AIC criterion or BIC criterion. The other two numbers are obtained using the Monte Carlo method generating the normal variables by the Box-Muler method: the first of them is for the AIC criterion, and the last is for the BIC criterion. The differences for the Box-Muler method are not due to the chosen criterion. They appear because we do not obtain the same value at a different time of running the program: we generate a random variable having the expectation equal to the solution and a small variance (the variance is smaller for more points due to the law of large numbers).

If we have chosen a criterion we have only two numbers in the corresponding cell: the first is obtained by the Simpson method, and the second by the Monte Carlo method generating the normal variables by the Box-Muler method as above.

When the Goldfeld-Quandt test fails for at least one class we do not compute the informational criteria AIC and BIC values. The word "first" between parentheses after the number *2* means the first classification: each class must have at least 9 points. By backtracking first classification into *2* classes, the first class has 9 points and the second class 208 points. Because we have at least one classification in three classes with all of them homoscedastic, "first" after the number *3* means the first such classification obtained by backtracking. The other classifications considered for three classes are for minimum AIC using Simpson method, minimum AIC using Monte Carlo method and minimum BIC. The bolded values are optimal (maximum for F(GQ) and minimum for informational criteria values).

The numbers between parentheses after each regression planes mean the number of points contained by the considered class. In the column GQ we find the results of the Goldfeld-Quandt statistics, and in F(GQ) we compute the c.d.f. of this statistics.

After we have found this first classification with all the classes homoscedastic we do not compare any more *frstatfmax* with $1-\varepsilon=0.975$: the new limit becomes 0.97 in the case of the Simpson method (for both informational criteria), 0.9727 in the case of Box-Muler method and the Akaike criterion, respectively 0.9658 in the case of Box-Muler method and the Schwartz criterion.

When all the classes are homoscedastic we apply the Hartley test to check if all the classes have the same variance. In the next table we present the results of this test for the cases with 3 homoscedastic classes.

Table 2

Case	F_{max}	c.d,f. of F_{max}	Accepting the same variances of classes
First classification	58.12196 (25,80)	0.99999 0.9993 0.9992	No
Min AIC, Simpson method	90.30542 (8,101)	0.989	No
Min AIC, Monte Carlo method, Box-Muler	90.20993 (8,102)	0.9893	No
Min BIC	88.51452 (8,104)	0.98878 0.9875	No

In the above table between parentheses after F_{max} we have the numbers of degrees of freedom. The signification of two or three numbers in the cells of "c.d,f. of F_{max}" is the same as for "F(GQ)" in the table with the results of the Goldfeld-Quandt test.

When we want to increase the number of classes to 4 (the maximum number of classes computed by using (17) or (17')) we have to apply (18) or (18'). We obtain the new limit for *frstatfmax* -3.78344 for the Akaike criterion and the Simpson method, -3.85758 for the Akaike criterion and the Box-Muler method, -4.10138 for the Schwartz criterion and the Simpson method, and finally -4.16146 for the Schwartz criterion and the Box-Muler method. Because all these values are negative we do not consider classifications with 4 classes or more: the optimal classification is that with 3 classes.

Example 2. Consider the polynomial model with $X = X_2$ from the previous example (the price indexes for the non-food goods) and the same resulting variable Y (the price indexes for services). The polynomial degree is considered to be 2 (regression parable).

The regression parable for all 217 points is Y=-95.1175+2.89504 X-0.00938 X^2. The correlation coefficients of X and X^2 with u^2 are r_1=0.55688 and r_2=0.56398. Therefore we sort the points on X^2 (in fact on X because the values are positives).

In the following two tables we have the analogues results to example 1, the only difference being that in the second column there are presented the regression parables instead of regression planes.

Table 3

No. of Classes	Regression Parables	Goldfeld-Quandt			AIC	BIC
		GQ	F(GQ)	Accepting homosceda-sticity		
1	Y=-95.1175+2.89504 X-0.00938 X^2 (217)	32.16203	0.99999 0.9977 0.9963	No		
2	First class ≤108 points except 17, 22, 23 or 24 classes			Yes No		
2	First class>108 points, 17,22,23 or 24 points			No Does not matter		
3 (first)	Y=4253.03729-83.10662 X+0.41583 X^2 (103)	1.28667	0.78354 0.7879 0.788	Yes	26.45167 26.20879	48.43553 48.07497
	Y=-280.5683+6.43061 X -0.02619 X^2 (103)	1.76284	**0.96039** **0.9585** **0.9575**	Yes		

No. of Classes	Regression Parables	Goldfeld-Quandt GQ	F(GQ)	Accepting homosceda-sticity	AIC	BIC
	Y=751.25142-11.05237 *X*+0.04792 X^2 (11)	5.13678	0.70141 0.7067 0.7024	Yes		
3 (min AIC and min BIC, Simpson or min BIC, Box-Muler)	*Y*=2500.87556-48.32439 *X*+0.24322 X^2 (108)	1.52995	**0.9112** **0.91212**	Yes	**23.504** 23.53311	**26.25638** **26.22727**
	Y=67.05484-0.10884 *X*+ 0.00455 X^2 (97)	1.53842	0.90257 0.9031	Yes		
	Y=387.70781-5.21554 *X* +0.02454 X^2 (12)	9.71273	0.90514 0.9109	Yes		
3 (min AIC, Box-Muler)	*Y*=3429.5133-66.75957 *X*+0.33471 X^2 (107)	1.45431	0.8811	Yes	**23.51343**	26.31165
	Y=-201.4046+4.96054 *X* -0.01937 X^2 (97)	1.56094	0.9113	Yes		
	Y=623.9845-9.02066 *X*+ 0.03983 X^2 (13)	11.19645	**0.9173**	Yes		

Table 4

Case	F_{max}	c.d,f. of F_{max}	Accepting the same variances of classes
First classification	183.98514 (8,100)	0.99998 0.9785 0.9942	No
Min AIC or min BIC, Simpson method, or min BIC, Monte Carlo method, Box-Muler	160.6187 (9,105)	0.9938 0.993	No
Min AIC, Monte Carlo method, Box-Muler	148.5919 (10,104)	0.9939	No

When we apply the formulae (18) and (18') as in example 1 to increase the number of classes to 4 we obtain the new limit of $\mathrm{frstatf} \quad -0.28794$ if we use the Akaike criterion and the Simpson method, -0.30043 if we use the Akaike criterion and the Box-Muler method, -0.44777 if we use the Schwartz criterion and the Simpson method and -0.46863 if we use the Schwartz criterion and the Box-Muler method. Therefore we do not need to classify the points into four or more classes.

4. Conclusions

When we have more than one class, the Goldfeld-Quandt test checks if the residues are the same intra-classes and the Hartley test checks if the residues are the same inter-classes. When the Goldfeld-Quandt test fails we need to classify data to obtain homoscedasticity. But even the Goldfeld-Quandt test does not fail (not the case of our paper) we can do such classification: the c.d.f. of the Goldfeld-Quandt statistics decrease when we increase the number of parameters. That's why we need an informational criterion (increasing on the error and the number of parameters) to make an equilibrium between these cost measures.

When we classify the 217 points in examples 1 and 2 we notice that for the optimal classification the last class has 11 points (the first example), 12 points (the second example, except the case of the Box-Muler method and the Akaike criterion) or 13 points (the case of the Box-Muler method and the Akaike criterion). Of course, the classification is done after sorting the data on the price indexes for non-food goods. The last 13 values in the above order correspond to February 1997, August 1993, July 1993, February 1992, October 1993, January 1997, January 1993, January 1992, November 1990, January 1991, December 1990, March 1997 and May 1993. We notice first that these months are between 1990 and 1993 (the beginning of the Romanian market economy) and the first three months in 1997 (the liberalization of prices).

We notice also that there are five months in 1993 (January, May, July, August and October). The majority in these 13 months is reached by January: 4 cases from 13 (1991, 1992, 1993 and 1997), and this can be explained by the fact that it is the beginning of the year (we have also two apparitions of February).

The reason why we have chosen the variable X as the price indexes for non-food goods is first that this is the explanatory variable positively correlated with the square of residues in example 1 and another interesting thing: the correlation coefficient between X and u^2 is 0.55688 which is less than the correlation coefficient between X^2 and u^2 (0.56398). In the other cases, the correlation coefficient between X and u^2 is 0.31942 if X is the price index for food and Y is the price index for non-food goods, and 0.34326 if X is the same variable and Y is the price index for services, respectively. The correlation coefficients between X^2 and u^2 are in these cases 0.30477 and 0.32991, respectively. These values are lower than the previous ones, and we can also notice that the previous correlation coefficients are less than 0.4. In the case of the polynomial model as in example 2 we reorder in fact the points on X or on X^2. The last order differs from the order of X only if X can take negative values.

The Goldfeld-Quandt test was chosen from among other tests (like White, for instance) because we use the same type of statistics (Snedecor-Fisher) as for the

Hartley test, and the order of points in the classes is natural: after we reorder the points by the price indexes for non-food goods the first points must be in the first class, the next points must be in the second and so on. Of course, we can define the informational criteria using other homoscedasticity test. These criteria must be increasing by the considered statistics and by the number of estimated parameters, but the main difficulty is in classification: we have no natural order.

For the Hartley test, when all the classes are homoscedastic, we obtain a classification with different errors. This can be explained by the order by price indexes for non-food goods and the positive correlation between this variable or its square and the square of the residues (greater than 0.5 in both examples). Theoretically we can obtain a classification with the same residues because the square of residues is not increasing by the explanatory variable: it is only positively correlated.

References

Ciuiu, D., (2007a), "Pattern Classification using Polynomial and Linear Regression", *Proceedings of the International Conference Trends and Challenges in Applied Mathematics, 20-23 June 2007*, Technical University of Civil Engineering, Bucharest, Romania, pp. 153-156.

Ciuiu, D., (2007b), " Pattern Classification using Principal Components Regression ", *Proceedings of the International Conference Trends and Challenges in Applied Mathematics, 20-23 June 2007*, Technical University of Civil Engineering, Bucharest, Romania, pp. 149-152.

Ciuiu, D., (2008), "Pattern Classification using Secondary Components Perceptron and Economic Applications", *Romanian Journal of Economic Forecasting*, 9(2): 51-66.

Craiu, V. and Ciucu, G. (1974), "*Inferență statistică"*, Ed. Didactică şi Pedagogică, Bucureşti.

Jula, D. (2003), *Introducere în econometrie*, Ed. Professional Consulting, Bucureşti.

Păltineanu, G., Matei, P. and Trandafir, R. (1998), "*Analiză numerică"*, Bucureşti: Ed. Conspress.

Popescu, Th. (2000), "*Serii de timp. Aplicații în analiza sistemelor"* , Bucureşti : Ed. Tehnică.

Saporta, G. (1990), "*Probabilités, Analyse des Donées et Statistique"*, Paris : Editions Technip.

Văduva, I. (2004), *Modele de simulare*, Ed. Universității Bucureşti.

Voineagu, V. *et al*. (2007), *Teorie şi practică econometrică*, Bucureşti: Meteor Press.

Buletin statistic de prețuri, Nr. 11, 2008, Institutul Național de Statistică, http://www.insse.ro.

Appendix A

The Price Indexes for Services, for Food and for Non-Food Goods

The price indexes for food goods are the following:

Year	Jan.	Feb.	March	April	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
1990											120.4	105.1
1991	108.5	104.1	112.9	158.6	101	98.3	110.8	117.2	106.9	109.6	108.7	116.7
1992	121.7	112.4	107.9	106.2	116.3	104.5	101.1	101.6	112.1	110.8	114.6	115
1993	107	105.7	115.2	111.8	130.3	100.4	110.1	109.7	112	117.1	116.5	105.5
1994	103.9	106.9	111.1	105.9	105.5	101.1	100.6	100.4	105	104.9	103.2	102.6
1995	103	101.5	100.3	101.4	100.7	100.1	103.1	100	101.2	103.5	103.8	104.5
1996	101.1	101.9	101.5	102.4	106.9	100.6	105.4	101.9	102.3	103.6	105.8	111.9
1997	110.6	125.2	131.4	105.5	102.8	101	99.5	103.9	102.1	106	104.1	105.8
1998	104.5	107.7	102.5	102.1	100.7	100	99	99	102.7	101.6	101.8	102.5
1999	102.5	102.5	104.7	105.2	103.1	100.5	98.8	100	103.6	103.4	103.4	104.1
2000	106.8	103.1	102.4	102.3	101.9	103.7	105.2	101.2	103	103.1	102.9	103
2001	103.8	103.1	102.5	103.3	101.9	102	100.1	100.7	101.4	101.8	101.2	102.6
2002	102.5	100.7	100.5	102.3	102.3	101.5	99	100.3	100.1	100.9	102.2	102.7
2003	101.5	101.7	101.3	101.3	100.3	101.2	101.1	99.3	100.2	101.2	102.1	101.8
2004	100.4	100.8	100.7	100.3	100	100.4	100.8	100.2	100.7	100.9	100.9	101.1
2005	100.4	100.5	100.3	100	100.1	100.4	100.2	100	100.2	101.2	101.2	101.1
2006	100.32	100.55	100.42	100.45	100.08	99.55	98.76	99.15	99.34	100.05	101.29	101.15
2007	100.28	99.97	99.95	100.49	100.23	100.32	100.56	100.68	101.94	101.3	101.17	100.92
2008	100.8	100.38	100.55	101.05	100.53	100.32	99.32	100.2	100.33	101.14	100.73	

The price indexes for non-food goods are the following:

Year	Jan.	Feb.	March	April	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
1990											120.2	124
1991	122	108.8	103.4	105.1	108.9	105.1	109.3	106	107.3	111.4	110.5	111.2
1992	119.6	114.1	111.8	103.5	109.6	104.2	104.6	104.9	107.9	108.7	112.3	111.1
1993	117.6	111.9	103.1	108.5	132.5	108.9	114	113	111.1	116.7	111.3	108.7
1994	105.7	104.7	104.9	106.3	104.1	103.2	102.7	102.8	102.7	103.6	102.6	101.4
1995	101.1	101.3	101	102.1	101	101.9	101.9	101.5	101.6	103.3	104.2	102.9
1996	101.5	101.8	101.5	101.3	103.9	101.4	110	105	102.1	103.2	106.8	110.2
1997	117.1	112.3	127.5	107.9	106.4	102.4	101.7	102.7	103.2	106	104.1	103.4
1998	102.8	107	105.2	102.3	104.1	102.6	102.6	101.9	101.8	105.1	101.9	101.9
1999	102.5	101.4	108.8	103.7	103.4	107.3	102.6	102.2	102.9	103.6	105.3	102.5
2000	102.4	101.3	101.2	105.3	101.9	103	103.9	102.1	103.1	102.3	103.5	102.4
2001	102.2	101.3	101.8	102.4	101.8	101.2	102	103.9	102.3	102.6	104.8	101.4
2002	102.4	101.6	100	101.6	101.4	100.8	101.5	101	101.1	102.4	103	100.7
2003	101.1	100.8	101	100.6	100.7	100.6	101.5	100.6	104.5	101	100.7	100.5
2004	101.8	100.5	100.3	100.7	100.5	100.6	102.1	100.6	101	101.5	101	100.4
2005	101.3	100	100.3	103.6	100.2	99.9	101.7	100.3	100.7	100	101.2	100.2
2006	101.88	100.16	100.11	100.38	101.2	100.45	101.15	100.31	100.24	100.07	101.23	100.87
2007	99.8	99.97	100.13	100.84	100.33	100.08	100.44	100.08	100.19	100.69	100.61	100.39
2008	100.37	101.08	100.67	100.46	100.47	100.28	102.31	100	100.19	100.52	99.95	

The price indexes for services are the following:

Year	Jan.	Feb.	March	April	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
1990											137.7	101.3
1991	104.4	108	103.5	118.5	107.6	103.7	106.2	109.2	108.7	110.2	119	112
1992	112.3	107.7	111.7	103.3	105.4	103.8	106.8	104.9	110.7	108.3	112.9	113.5
1993	109.8	105	108.6	107.8	123.8	114.9	121.9	107.5	106.9	111.9	116.1	109.7
1994	106.1	106	108	106.1	105.7	106.6	102	103.8	103.7	104.9	102.3	101.9
1995	101.1	101.6	103.6	100.9	102.6	104	102.7	102.8	103	104.4	104.9	103
1996	100.8	102.1	103.3	101.8	103.9	101.5	108.4	106.8	103.4	103.4	103	105.6
1997	114.4	116.7	138.4	109.3	103.3	106.6	101.6	104.8	107.8	109.4	105.1	103.5
1998	111.6	106.2	103.8	105.7	102.2	101.6	104.5	101.9	105.1	106.5	102.3	102.1
1999	106	102.7	105	106.7	116.2	111.8	106	101.6	102.9	106.8	102.5	101.3
2000	103.3	102	101.7	108.9	101.5	100.8	103.1	102.7	102.1	103	101.5	101.6
2001	107	102.4	101.4	101.5	101.4	101.4	102.8	102.5	102.5	103.7	101.9	102.9
2002	101.7	101.4	101.1	102.5	101.8	101.3	102	102	100.9	101.6	102.5	100.5
2003	101.1	98.7	100.8	101.6	100.7	100.5	100.6	102.1	101.4	103.9	101.6	101.1
2004	101.2	100.5	100.5	101	100.6	101.2	100.3	101.4	101.4	101.4	99.3	99.6
2005	100.6	102.5	100.2	101.5	101	101	100.7	100	101.2	102.2	101.2	100.3
2006	100.59	99.71	99.96	100.42	100.32	100.81	100.55	100.73	101.18	100.92	100.3	99.53
2007	101.04	100.38	100.16	99.8	102.31	99.91	99.33	101.06	101.42	100.98	101.21	100.67
2008	102.12	100.51	100.92	99.56	100.45	100.19	99.7	99.1	101.05	102.23	100.38	