PRZEGLĄD STATYSTYCZNY
R. LVI – ZESZYT 2 – 2009

JANUSZ WYWIAŁ

TESTING OF THE PREDICTION UNBIASEDNESS ON THE BASIS OF JANUS QUOTIENT

1. INTRODUCTION

The problem of choice the appropriate predictor is being considered. In this paper the analysis is generally focused on the problem of unbiasedness of the predictors. Several tests attempting to verify the unbiasedness of three predictors of the linear trend are proposed. They are based on some modifications of the well known Janus quotient. In some sense the construction of the test statistics is close to those considered [1] which were used to test the linearity of the trend. It is suggested that the proposed tests can be useful in more general problems like testing hypotheses referring to non-linear trends.

Let us consider the following model: $(Y_t) = (Y_1, Y_2, \ldots, Y_n, \ldots, Y_{n+m})$, $E(Y_t) = \mu_t$, $D^2(Y_t) = \sigma^2$, $Cov(Y_t, Y_k) = 0$, for $t \neq$ k and $t = 1, \ldots, n+m$, $k = 1, \ldots, n+m$. Moreover, let $\mathbf{Y}^T = \left[\mathbf{Y}_n^T \; \mathbf{Y}_m^T\right]$.

Let Y_{tp} be a predictor of Y_t, where $t = n+1, \ldots, n+m$ and let Y_{tp} be a function of the variables $(Y_1 \ldots Y_n \ldots Y_{t-1})$. The vector of prediction errors is denoted by $\mathbf{U}^T = [U_1 \ldots U_m]$ where $U_t = Y_t - Y_{tp}$. The vector of prediction bias is $\delta^T = [\delta_1 \ldots \delta_m]$ where $\delta_t = E(U_t)$. A predictor is unbiased when $\delta^T = \mathbf{0}$. The prediction variance is as follows:

$$D^2(U_t) = D^2(Y_t) + 2Cov(Y_t, Y_{tp}) + D^2(Y_{tp}).$$

The mean square prediction error is as follows:

$$MSE(U_t) = E(U_t)^2 = D^2(U_t) + \delta_t^2.$$

The prediction ex-post variance is defined in the following way:

$$S_U^2 = \frac{1}{m}\sum_{t=n+1}^{n+m} U_t^2 = \frac{1}{m}\mathbf{U}^T\mathbf{U}. \tag{1}$$

Let $\hat{Y}_t$ be an estimator of the t-th value of the trend. The vector of residuals is denoted by $\mathbf{e}^T = [e_1 \ldots e_n]$ where $e_t = Y_t - \hat{Y}_t$. The residual variance is as follows.

$$S_e^2 = \frac{1}{m}\sum_{t=1}^{n} e_t^2 = \frac{1}{m}\mathbf{e}^T\mathbf{e}. \tag{2}$$

2. THE JANUS QUOTIENT

In [2] the following Janus coefficient is proposed.

$$J = \frac{S_U^2}{S_e^2} \tag{3}$$

Let us assume that $\mathbf{Y} \sim N(\mu,\ \sigma^2\ \mathbf{I}_{n+m})$ where $\mu^T = \left[\mu_n^T\ \mu_m^T\right]$, $E(\mathbf{Y}_n) = \mu_n$, $E(\mathbf{Y}_m) = \mu_m$,

Let $\mathbf{Z}^T = [\mathbf{e}^T\ \mathbf{U}^T]$ and

$$\mathbf{Z} = \mathbf{CY} \tag{4}$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{O}_{n\times m} \\ \mathbf{B}_{m\times n} & \mathbf{B}_{m\times m} \end{bmatrix} \tag{5}$$

Particularly we have

$$\mathbf{e} = \mathbf{AY}_n, \qquad \mathbf{U} = \mathbf{BY}. \tag{6}$$

where $\mathbf{B} = [\mathbf{B}_{m\times n}\ \mathbf{B}_{m\times m}]$.

Hence, $\mathbf{Z} \sim N(\mu_z, \Sigma_z)$ where $\mu_z^T = \left[\mu_e^T\ \mu_U^T\right]$, $\mu_e^T = \mathbf{A}\mu_n$, $\mu_U^T = \mathbf{B}\mu$,

$$\Sigma_z = \begin{bmatrix} \Sigma_{ee} & \Sigma_{eU} \\ \Sigma_{Ue} & \Sigma_{UU} \end{bmatrix} \tag{7}$$

where

$$\textstyle\sum_{ee} = \sigma^2\mathbf{AA}^T,\ \sum_{UU} = \sigma^2\mathbf{BB}^T,\ \sum_{eU} = \sigma^2\mathbf{AB}_{m\times n}^T = \sum_{Ue}^T. \tag{8}$$

The Janus coefficient can be rewritten in the following way.

$$J = \frac{n}{m}\frac{\mathbf{Y}^T\mathbf{B}^T\mathbf{BY}}{\mathbf{Y}^T\mathbf{A}^T\mathbf{AY}}. \tag{9}$$

The Janus quotient is a function of two quadratic forms of the normal vector **Y**. The distributions of the quadratic forms are not necessarily chi-square ones and, moreover, their distributions may not be independent. That distribution function of the J statistic can be transformed in the following way:

$$F_J(J < x) = P\left(\mathbf{Y}^T(n\mathbf{B}^T\mathbf{B} - mx\mathbf{A}^T\mathbf{A})\mathbf{Y} < 0\right) = P\left(\mathbf{Y}^T\mathbf{G}\mathbf{Y} < 0\right) = F(0) \tag{10}$$

where

$$\mathbf{G} = n\mathbf{B}^T\mathbf{B} - \begin{bmatrix} xm\mathbf{A}^T\mathbf{A} & \mathbf{O}_{n\times m} \\ \mathbf{O}_{m\times n} & \mathbf{O}_{m\times m} \end{bmatrix}. \tag{11}$$

The probability $F(0)$ can be evaluated on the basis of the Appendix 6.1.

3. THE MODIFIED JANUS QUOTIENTS

In [6] there are some modifications of the Janus quotient proposed. The first of them is as follows:

$$J_* = \frac{n}{m}\frac{Q_U}{Q_e} \tag{12}$$

where

$$Q_U = \mathbf{U}^T \sum\nolimits_{UU}^{-} \mathbf{U}, \tag{13}$$

$$Q_e = \mathbf{e}^T \sum\nolimits_{ee}^{-} \mathbf{e}, \tag{14}$$

Σ_{UU}^- and Σ_{ee}^- are the pseudo-inverses of the matrices Σ_{UU} and Σ_{ee}, respectively. Hence, $\Sigma_{UU}\Sigma_{UU}^-\Sigma_{UU} = \Sigma_{UU}$ and $\Sigma_{ee}\Sigma_{ee}^-\Sigma_{ee} = \Sigma_{ee}$.

On the basis of general results presented in [5] the random variable Q_U has noncentral chi-square distribution $\chi_k(\kappa_U)$ with $k = R(\Sigma_{UU})$ degree of freedom and noncentrality parameter $\kappa_U = \mu_U^T\Sigma_{UU}^-\mu_U$. Similarly, the quadratic form Q_e has a noncentral chi-square distribution $\chi_h(\kappa_e)$ with $h = R(\Sigma_{ee})$ degrees of freedom and non-centrality parameter $\kappa_e = \mu_e^T\Sigma_{ee}^-\mu_e$.

Let us assume that $\kappa_e = \mathbf{0} = E(\mathbf{e})$, which means that the trend function is well fitted to the data in the period $t = 1, \ldots, n$. In this case, the quadratic form Q_e has central chi-square distribution $\chi_k(0)$. When a predictor gives the unbiased forecasts in the period $t = n + 1, \ldots, n + m$, then $\kappa_U = \mathbf{0} = E(\mathbf{U})$ and Q_U has a central chi-square distribution $\chi_k(0)$. In the case, when at least one forecast is biased in the period $t = n + 1, \ldots, n + m$, the quadratic form has a noncentral distribution. Hence, a significantly large value of the J_* statistic leads to rejecting the hypothesis that $\kappa_U = 0$, which means that the considered predictor is unbiased in the period $t = n + 1, \ldots, n + m$. In order to evaluate the p-value of the test, we have to know the distribution of the J_* test statistic. It is an F distribution with k and h degrees of freedom under the additional assumption that quadratic forms Q_U and Q_e are independent. In a more general case, when they are not independent, the distribution function is approximated on the basis of the Appendix 6.1. and the expression (10), where the $\mathbf{G}$ matrix should be defined as follows.

$$\mathbf{G} = n\mathbf{B}^T \Sigma_{UU}^{-} \mathbf{B} - \begin{bmatrix} xm\mathbf{A}^T \Sigma_{ee}^{-} \mathbf{A} & \mathbf{O}_{n \times m} \\ \mathbf{O}_{m \times n} & \mathbf{O}_{m \times m} \end{bmatrix}. \tag{15}$$

The lack of the independence of the quadratic forms leads to numerous difficulties in evaluating the p-value of the test statistic. In order to omit this problem, we introduce a test statistic which can be treated as the second modification of the Janus quotient. Firstly, let us define the vector of the following conditional prediction errors:

$$\mathbf{U}_{U/e} = \mathbf{U} - \Sigma_{Ue}\Sigma_{ee}^{-}\mathbf{e}. \tag{16}$$

So, $\mathbf{U}_{U/e} \sim N(\mu_{U/e}, \Sigma_{UU/e})$ where

$$\mu_{U/e} = \mu_U - \Sigma_{Ue}\Sigma_{ee}^{-}\mu_e, \tag{17}$$

$$\Sigma_{UU/e} = \Sigma_{UU} - \Sigma_{Ue}\Sigma_{ee}^{-}\Sigma_{eU}\,. \tag{18}$$

The distribution of the vector $\mathbf{U}_{U/e}$ does not dependent on the vector $\mathbf{e}$. Moreover, if $\mu_e = E(\mathbf{e}) = \mathbf{0}$, then $\mu_{U/e} = E(\mathbf{U}_{U/e}) = E(\mathbf{U}) = \mu_U$. Hence, distribution of the square form

$$Q_{U/e} = \mathbf{U}_{U/e}^T \Sigma_{UU/e}^{-} \mathbf{U}_{U/e}, \tag{19}$$

has the noncentral chi-square distribution $\chi_k(\kappa_{U/e})$ with $k = R(\Sigma_{UU/e})$ degrees of freedom and the non-centrality parameter $\kappa_{U/e} = \mu_{U/e}^T \Sigma_{UU/e}^{-} \mu_{U/e}$ and it is not dependent on the quadratic form Q_e.

Under the assumption that $\mu_e = E(\mathbf{e}) = \mathbf{0}$, the following statistic

$$J_{U/e} = \frac{h}{r}\frac{Q_{U/e}}{Q_e}, \tag{20}$$

has the F noncentral distribution with r and h degrees of freedom where $r = R(\Sigma_{UU/e})$. The $J_{U/e}$ coefficient can be treated as a next modified Janus quotient. When the prediction is unbiased, then $\mu_{U/e} = E(\mathbf{U}_{U/e}) = E(\mathbf{U}) = \mu_U = \mathbf{0}$ and the test statistic $J_{U/e}$ has the central F distribution. So, a significantly large value of the test statistic leads to the rejection of the hypothesis that $E(\mathbf{U}) = \mu_U = \mathbf{0}$, which means that the prediction is unbiased in the considered time period $t = n + 1, \ldots, n + m$.

4. TESTING THE UNBIASEDNESS OF SOME PREDICTORS

Let us introduce the following assumptions (apart from those stated in the Introduction): We will consider the linear trend $\mu_t = E(Y_t) = at + b$ for $t = 1, \ldots, n$. It can change in the next period $t = n + 1, \ldots, n + m$, so it is possible that $E(Y_t) \neq at + b$ for at least $t = n + 1, \ldots, n + m$. We are going to consider the three predictors of the $\mathbf{Y}_m$ vector based on an extrapolation of the trend. They will be denoted by $\hat{\mathbf{Y}}_g$, $g = 1, 2, 3$. Let $\mathbf{U}_g = \mathbf{Y}_m - \hat{\mathbf{Y}}_g$ be the vector of prediction errors of the $g - th$ predictor,

g = 1, 2, 3. We will test the hypothesis that $H_0 : \mathbf{U}_g = \mathbf{0}$, against the alternative one that $H_0 : \mathbf{U}_g \neq \mathbf{0}$.

The vector of residuals of the well-known mean square error estimator of the linear trend is as follows:

$$\mathbf{e} = \mathbf{M}\mathbf{Y}_n \tag{21}$$

where

$$\mathbf{M} = \mathbf{I}_n - \mathbf{X}_n \left(\mathbf{X}_n^T \mathbf{X}_n\right)^{-1} \mathbf{X}_n^T, \quad \mathbf{X}_n = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ .. & .. \\ n & 1 \end{bmatrix}. \tag{22}$$

The first predictor is the following:

$$\hat{\mathbf{Y}}_1 = \mathbf{b}_1 \mathbf{Y}_n \tag{23}$$

where

$$\mathbf{b}_1 = \mathbf{X}_m \left(\mathbf{X}_n^T \mathbf{X}_n\right)^{-1} \mathbf{X}_n^T \tag{24}$$

where

$$\mathbf{X}_n = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ .. & .. \\ n & 1 \end{bmatrix}, \quad \mathbf{X}_m = \begin{bmatrix} n+1 & 1 \\ n+2 & 1 \\ .. & .. \\ n+m & 1 \end{bmatrix}.$$

The vector of the prediction error is $\mathbf{U}_1 = \mathbf{B}_1 \mathbf{Y}$ where

$$\mathbf{B}_1 = [-\mathbf{b}_1 \ \ \mathbf{I}_m] \tag{25}$$

The expression (8) leads to the following well-known ones.

$$\Sigma_{UU} = \sigma^2 \left(\mathbf{I}_m + \mathbf{X}_m \left(\mathbf{X}_n^T \mathbf{X}_n\right)^{-1} \mathbf{X}_m^T\right), \ \Sigma_{eU} = \mathbf{O}_{n \times m}. \tag{26}$$

On the basis of the expressions (12)-(14) we have

$$J_{*1} = \frac{n-2}{m} \frac{\mathbf{U}_1^T \left(\mathbf{B}_1 \mathbf{B}_1^T\right)^{-1} \mathbf{U}_1}{\mathbf{e}^T \mathbf{M} \mathbf{e}}. \tag{27}$$

The residual vector $\mathbf{e}$ and the vector of the prediction errors $\mathbf{U}_1$ are independent. So, the J_{*1} statistic has the F distribution with m and $h = n - 2$ degrees of freedom when the hypothesis H_0 is true.

The next vector predictor is as follows:

$$\hat{\mathbf{Y}}_2 = \mathbf{b}_2\mathbf{Y} \tag{28}$$

where

$$\mathbf{b}_2 = \begin{bmatrix} \mathbf{b}_{2,n} & \mathbf{O}_{1\times m} \\ \mathbf{b}_{2,n+1} & \mathbf{O}_{1\times(m-1)} \\ \cdots & \cdots \\ \mathbf{b}_{2,n+m-2} & \mathbf{O}_{1\times 2} \\ \mathbf{b}_{2,n+m-1} & 0 \end{bmatrix}, \tag{29}$$

$$\mathbf{b}_{2,t} = \mathbf{x}_{t+1}\left(\mathrm{X}_t^T\mathbf{X}_t\right)^{-1}\mathbf{X}_t^T,\ t = n+1,\ \ldots,\ n+m-1$$
$$\mathbf{x}_{t+1} = \left[t+1\ \ 1\right],\ \mathbf{X}_t = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ .. & .. \\ t & 1 \end{bmatrix}. \tag{30}$$

The vector of the prediction error is $\mathbf{U}_2 = \mathbf{B}_2\mathbf{Y}$ where

$$\mathbf{B}_2 = \mathbf{g} - \mathbf{b}_2,\ \mathbf{g} = \left[\mathbf{O}_{m\times n}\ \ \mathrm{I}_m\right] \tag{31}$$

The expression (8) leads to the following ones:

$$\Sigma_{UU} = \sigma^2\mathbf{B}_2^T\mathbf{B}_2. \tag{32}$$

$$\Sigma_{eU} = \sigma^2\left[\mathbf{M}\ \ \mathbf{O}_{n\times m}\right]\mathbf{B}_2^T = \mathbf{O}_{n\times m}. \tag{33}$$

On the basis of the expressions (12)-(14) we have

$$J_{*2} = \frac{n-2}{m}\frac{\mathbf{U}_2^T\left(\mathbf{B}_2\mathbf{B}_2^T\right)^{-1}\mathbf{U}_2}{\mathbf{e}^T\mathbf{Me}}. \tag{34}$$

The residual vector **e** and the vector of the prediction errors $\mathbf{U}_1$ are independent. So, the J_{*2} statistic has the F distribution with m and $h = n - 2$ degrees of freedom when the hypothesis H_0 is true.

The third vector predictor is as follows:

$$\hat{\mathbf{Y}}_3 = \mathbf{b}_3\mathbf{Y} \tag{35}$$

where

$$\mathbf{b}_3 = \begin{bmatrix} \mathbf{b}_{3,n} & \mathbf{O}_{1\times m} & \\ 0 & \mathbf{b}_{3,n+1} & \mathbf{O}_{1\times(m-1)} \\ \mathbf{O}_{1\times 2} & \mathbf{b}_{3,n+2} & \mathbf{O}_{1\times(m-2)} \\ \cdots & \cdots & \cdots \\ \mathbf{O}_{1\times(m-2)} & \mathbf{b}_{3,n+m-2} & \mathbf{O}_{1\times 2} \\ \mathbf{O}_{1\times(m-1)} & \mathbf{b}_{3,n+m-1} & 0 \end{bmatrix}, \mathbf{X}_m = \begin{bmatrix} n+1 & 1 \\ n+2 & 1 \\ .. & .. \\ n+m & 1 \end{bmatrix},$$

The $\mathbf{b}_{2,t}$ vector is defined by the right hand of the equation(30), but in this case

$$\mathbf{x}_{t+1} = [t+1 \;\; 1],\ \mathbf{X}_t = \begin{bmatrix} t-n+1 & 1 \\ t-n+2 & 1 \\ \cdots & \cdots \\ t & 1 \end{bmatrix}.$$

The vector of the prediction error is $\mathbf{U}_3 = \mathbf{B}_3\mathbf{Y}$ where

$$\mathbf{B}_3 = \mathbf{g} - \mathbf{b}_3,\ \mathbf{g} = [\mathbf{O}_{m\times n} \;\; \mathbf{I}_m]. \tag{36}$$

The expression (8) leads to the following ones.

$$\Sigma_{UU} = \sigma^2 \mathbf{B}_3^T \mathbf{B}_3,\ \Sigma_{eU} = \sigma^2 [\mathbf{M} \;\; \mathbf{O}_{n\times m}] \mathbf{B}_3^T. \tag{37}$$

On the basis of the expressions (12)-(14) we have

$$J_{*3} = \frac{n-2}{m} \frac{\mathbf{U}_3^T (\mathbf{B}_3\mathbf{B}_3^T)^{-1} \mathbf{U}_3}{\mathrm{e}^T \mathbf{Me}} = \frac{n-2}{m} \frac{\mathbf{Y}^T \mathbf{B}_3^T (\mathbf{B}_3\mathbf{B}_3^T)^{-1} \mathbf{B}_3 \mathbf{Y}}{\mathbf{Y}_n^T \mathbf{M} \mathbf{Y}_n}. \tag{38}$$

The J_{*3} statistic is proportionate to the ratio of two dependent quadratic forms of the normal vector. The value of its distribution can be evaluated on the basis of the appendix and the expression (10) where

$$\mathbf{G} = n\mathbf{B}_3^T (\mathbf{B}_3\mathbf{B}_3^T)^{-1} \mathbf{B}_3 - \begin{bmatrix} xm\mathbf{M} & \mathbf{O}_{n\times m} \\ \mathbf{O}_{m\times n} & \mathbf{O}_{m\times m} \end{bmatrix}. \tag{39}$$

According to the expressions (16)-(20) we define the following vector of the conditional prediction errors and its parameters.

$$\mathbf{U}_{U_3/e} = \mathbf{U}_3 - \mathbf{B}_3 \begin{bmatrix} \mathbf{M} \\ \mathbf{O}_{m\times n} \end{bmatrix} \mathbf{e}. \tag{40}$$

When $E(\mathbf{e}) = \mathbf{0}$ then $\mathbf{U}_{U_3/e} \sim N(\mu_{U_3}, \Sigma_{U_3U_3/e})$ where $\mu_{U_3} = E(\mathbf{U}_3)$ and

$$\Sigma_{U_3U_3/e} = \sigma^2 \mathbf{B}_3^T \mathbf{B}_3 - \sigma^2 \begin{bmatrix} \mathbf{M} & \mathbf{O}_{n\times m} \\ \mathbf{O}_{m\times n} & \mathbf{O}_{m\times m} \end{bmatrix}. \tag{41}$$

The equation (20) leads to the following modified Janus coefficient.

$$J_{U_3/e} = \frac{h}{r}\frac{\mathbf{U}^T_{U_3/e}\Sigma^-_{U_3U_3/e}\mathbf{U}_{U_3/e}}{\mathbf{e}^T\mathbf{Me}}. \tag{42}$$

The $J_{U_3/e}$ statistic has the F distribution with $r = R(\Sigma_{U_3U_3/e}$ and $h = n - 2$ degrees of freedom when the hypothesis H_0 is true.

Example 1. Let us consider the following time series $\{y_t\} = \{10, 12, 13, 12, 16, 15, 17, 18, 20, 19, 21, 23, 26, 29, 29, 31, 34, 36, 39, 40\}$. On the basis of the first $n = 10$ observations, the following trend function was determined by means of the least square error: $\hat{y}_t = 1.07t + 9.33$. Similarly, in the next period $t = 11, \ldots, 20$, the trend was: $\hat{y}_t = 2.12t - 2.08$.

Values of the modified Janus quotients (with the p-values of the test in the brackets) were evaluated be means of the computer programmes are as follows: $J_{*1} = 256.70$ ($p = 0.000$), $J_{*2} = 164.73$ ($p = 0.000$), $J_{*3} = 4.39$ ($p = 0.103$) and $J_{U_3/e} = 5.21$ ($p = 0.014$). Each of the statistics J_{*1}, J_{*2} and $J_{U_3/e}$ has F distribution with 10 and 8 degrees of freedom.

The analysis of the p-values of the tests based on the statistics J_{*1} and J_{*2} leads to rejection of the hypotheses that the predictors $\hat{\mathbf{Y}}_1$ and $\hat{\mathbf{Y}}_2$ are unbiased. The hypothesis that the $\hat{\mathbf{Y}}_3$ predictor is unbiased is not rejected by any tests based on the statistics J_{*3} and $J_{U_3/e}$ under the significance level $\alpha = 0.01$.

5. CONCLUSIONS

Although the proposed tests are constructed under a very strong assumption connected with defining the model of time series, they can support the choice of the most accurate predictor. The analysis of the example 1 leads to conclusion that it is especially possible to assess how useful in practice the adaptive predictors are. It is easy to show that the tests can be useful in the cases of other trend functions as well as in the case of time series with seasonal factors. More complicated models with more explanatory variables can be considered, too. Moreover, it seems that testing a hypothesis referring to the parameters of exponential smoothing methods is possible, too.

The analysis of the properties of the tests should be developed in the direction of examining their power as well as their robustness of their assumptions. The author would like to consider those problems in his next papers.

6. APPENDIX

6.1. APPROXIMATION OF THE DISTRIBUTION FUNCTION OF THE QUADRATIC FORMS

In [4] the following procedure for the approximation of the distribution of the form $Q = \mathbf{Y}^T\mathbf{CY} = \sum_{i=1}^{r} d_i Z_i^2$ where $Z_i \sim N(0, 1)$, Z_i, Z_j are independent for $i \neq j$, $i = 1, \ldots, r$, $j = 1, \ldots, r$ is proposed. He showed that

$$Q \approx cU_k + v \tag{43}$$

where $Z_k \sim \chi_k$ and

$$c = \frac{\theta_3}{\theta_2},\ \upsilon = -\frac{\theta_2^2}{\theta_3} + \theta_1,\ k = \frac{\theta_2^3}{\theta_3^2},\ \theta_h = \sum_{i=1}^{k} = tr\mathbf{C}^h,\ h = 1, 2, \ldots \tag{44}$$

Under this notation

$$P(Q < q) \approx P(U_k < u) \tag{45}$$

where

$$u = \frac{q - \theta_1}{\sqrt{\theta_2}}\sqrt{k} + k. \tag{46}$$

Let us note that in [3] there are other approximations of the distribution function of the quadratic forms considered.

6.2. PROOF OF INDEPENDENCE OF THE VECTORS $\mathbf{e}$ AND $\mathbf{U}_2$

Evaluation of the expression (33) is as follows. Let e_i be the $i-th$ residual ($i = 1, \ldots, n$) end U_{t+1} be the $(t + 1) - th$ prediction error ($t = n + 1, \ldots, n + m$).

$$Cov(e_i, U_{t+1}) = E\left(Y_i - \mathbf{x}_i\left(\mathbf{X}_n\mathbf{X}_n^T\right)^{-1}\mathbf{X}_n^T\mathbf{Y}_n\right)\left(Y_{t+1} - \mathbf{x}_{t+1}\left(\mathbf{X}_n\mathbf{X}_n^T\right)^{-1}\mathbf{X}_t^T\mathbf{Y}_t\right),$$

$$Cov(e_i, U_{t+1}) = -\mathbf{x}_{t+1}\left(\mathbf{X}_n\mathbf{X}_n^T\right)^{-1}\mathbf{X}_t^T E(\mathbf{Y}_t y_i) +$$
$$+\mathbf{x}_i\left(\mathbf{X}_n\mathbf{X}_n^T\right)^{-1}\mathbf{X}_n^T E\left(\mathbf{Y}_n\mathbf{Y}_t^T\right)\mathbf{X}_t\left(\mathbf{X}_n\mathbf{X}_n^T\right)^{-1}\mathbf{x}_{t+1}^T,$$

$$Cov(e_i, U_{t+1}) = -\mathbf{x}_{t+1}\left(\mathbf{X}_n\mathbf{X}_n^T\right)^{-1}\mathbf{x}_i^T +$$
$$+\mathbf{x}_i\left(\mathbf{X}_n\mathbf{X}_n^T\right)^{-1}\mathbf{X}_n^T\left[\mathbf{I}_n\ \ \mathbf{O}_{n\times(t-n)}\right]\mathbf{X}_t\left(\mathbf{X}_n\mathbf{X}_n^T\right)^{-1}\mathbf{x}_{t+1}^T,$$

$$Cov(e_i, U_{t+1}) = -\mathbf{x}_{t+1}\left(\mathbf{X}_n\mathbf{X}_n^T\right)^{-1}\mathbf{x}_i^T +$$
$$+\mathbf{x}_i\left(\mathbf{X}_n\mathbf{X}_n^T\right)^{-1}\mathbf{X}_n^T\mathbf{X}_n\left(\mathbf{X}_n\mathbf{X}_n^T\right)^{-1}\mathbf{x}_{t+1}^T,$$

$$Cov(e_i, U_{t+1}) = -\mathbf{x}_{t+1}\left(\mathbf{X}_n\mathbf{X}_n^T\right)^{-1}\mathbf{x}_i^T + \mathbf{x}_i\left(\mathbf{X}_n\mathbf{X}_n^T\right)^{-1}\mathbf{x}_{t+1}^T = 0.$$

This leads straightforwardly to the expression (33).

Akademia Ekonomiczna w Katowicach

REFERENCES

[1] Azzalini A., Bowman A., [1993], *On nonparametric regression for checking linear relationships*, „The Journal of the Royal Statistical Society", B55(2), s. 549-557.

[2] Gadd A., Wold H., [1964], *The Janus quotient; a measure for the accuracy of prediction*, In Econometric Model Building, Amsterdam.

[3] Mathai A.M., Provost S.B., [1992], *Quadratic Forms in Random Variables (Theory and Applications)*, Marcel Decker, Inc., New York-Basel-Hong Kong.
[4] Pearson E.S., [1959], *Note on an approximation to the distribution of noncentral* χ^2, „Biometrika", 46, pp. 346-346.
[5] Rao C.R., Mitra S.K., [1971], *Generalized inverse of matrices and its applications*, John Wiley and Sons, New York-London-Sydney-Toronto.
[6] Wywiał J., [1995], *Weryfikacja hipotez o błędach predykcji adaptacyjnej*, Ossolineum, Wrocław-Warszawa--Kraków.

Praca wpłynęła do redakcji we wrześniu 2008 r.

O TESTOWANIU NIEOBCIĄŻONOŚCI PREDYKCJI NA PODSTAWIE WSPÓŁCZYNNIKA JANUSOWEGO

Streszczenie

W pracy jest rozważany problem wyboru odpowiedniego predykatora. Przyjęto, że postulowanym kryterium tego wyboru jest jego nieobciążoność. W pracy są proponowane testy na nieobciążoność predykcji trzech predyktorów trendu liniowego. Punktem wyjścia konstrukcji sprawdzianów testów jest znany współczynnik Janusowy używany do oceny dokładności ciągów wyznaczanych prognoz, który jest ilorazem wariancji predykcji *ex-post* i wariancji resztowej. Z formalnego punktu widzenia każdy z rozważanych sprawdzianów testów jest ilorazem dwóch form kwadratowych wektora zmiennych o rozkładzie normalnym. Te formy kwadratowe mogą być zależne. Dlatego w celu wyznaczenia rozkładu prawdopodobieństwa sprawdzianu testu jest adaptowana jedna ze znanych metod pozwalających na przybliżone wyliczanie wartości jego dystrybuanty. Rozważania zilustrowano przykładem. Otrzymane w pracy wyniki można uogólnić na przypadek predykcji na podstawie modelu regresji.

Słowa kluczowe: Test statystyczny, błąd predykcji, nieobciążoność, współczynnik Janusowy, forma kwadratowa wektora o rozkładzie normalnym, aproksymacja funkcji dystrybuanty, wariancja resztowa, wariancja predykcji.

TESTING OF THE PREDICTION UNBIASEDNESS ON THE BASIS OF JANUS QUOTIENT

Summary

The problem of choosing the appropriate predictor is being considered. Generally, in this paper the analysis is focused on the problem of unbiasedness of the predictors. Several tests attempting to verify the unbiasedness of three predictors of the linear trend are proposed. They are based on some modifications of the well-known Janus quotient being a ratio of the variance of prediction errors and the residual variance. In general each of the considered test statistic can be represented as the ratio of two quadratic forms of normal vectors. These two quadratic forms can be dependent, so its distribution function has to be approximated. An example of testing hypothesis on unbiasedness is presented. The obtained results can be generalized in the case of prediction on the basis of regression models.

Keyword: Test statistic, prediction error, unbiasednees, Janus quotient, quadratic form of normally distributed vector, approximation of distribution function, residual variance, prediction variance.