ACTA UNIVERSITATIS LODZIENSIS

FOLIA OECONOMICA 68, 1987

Iwona Konarzewska*, Władysław Milo**

# ON SOME CONSEQUENCIES OF THE LACK OF INDEPENDENCE BETWEEN NOISE COMPONENT AND EXPLANATORY VARIABLES IN LINEAR REGRESSION MODEL

## 1. Introduction

In this work we consider a following linear model

$$\mathcal{MRL} = (R^k, \mathcal{S}, \mathcal{S}_Y, \mathcal{S}_{\underline{X}}, \mathcal{S}_\Xi, Y = \beta' \underline{X} + \Xi, P_\Xi = \mathcal{N}_\Xi(0, \sigma_\Xi^2)$$
$$P_{\underline{X}} = \mathcal{N}_{\underline{X}}(\mu, \Sigma))$$

where

$$\beta \in R^k, \mu = (\mu_1, \ldots, \mu_k)' \in R^k, \Sigma = (\sigma_{ij})_{i,j=1}^k =$$
$$= \operatorname{cov}(X_i, X_j)_{i,j=1}^k \in R^{k \times k}, \sigma_\Xi^2 \in R_+,$$

$\mathcal{S} = (\mathcal{U}, \mathcal{F}, P)$ - elementary events space,
$\mathcal{F}$ - $\sigma$-field of subsets of $\mathcal{U}$
P - complete probability measure defined on $\mathcal{F}$,

$$Y : (\mathcal{U}, \mathcal{F}, P) \longrightarrow (R, \mathcal{F}_R, P_Y) \equiv \mathcal{S}_Y$$
$$\Xi : (\mathcal{U}, \mathcal{F}, P) \longrightarrow (R, \mathcal{F}_R, P_\Xi) \equiv \mathcal{S}_\Xi$$
$$\underline{X} : (\mathcal{U}, \mathcal{F}, P) \longrightarrow (R^k, \mathcal{F}_{R^k}, P_{\underline{X}}) \equiv \mathcal{S}_{\underline{X}}.$$

* Senior Assistant, Institute of Econometrics and Statistics, University of Łódź.

** Lecturer, Institute of Econometrics and Statistics, University of Łódź.

While analyzing properties of parameter vector $\beta$ estimators one traditionally accepts an assumption about distributional independence between random variables $X_i$ $i = \overline{1, k}$ and random noise component $\Xi$. This is, assuming normal distributions of considered variables, equivalent to the assumption that $\operatorname{cov}(\underline{X}, \Xi) = 0$. We will check the influence of avoiding this assumption on an explanation level of the model measured by the square of correlation coefficient between the variable Y and linear combination of explanatory variables $\beta'\underline{X}$. The assumption of independence between explanatory variables and a random noise component cannot be accepted if, for example due to multicollinearity, one rejects from the model an important explanatory variable (in such a case this rejected variable acts as a part of random noise component $\Xi$). In most cases this variable is not independent from the rest explanatory variables - and consequently $\Xi$ is not independent from them either (and often $E(\Xi) \neq 0$). The existence of such conditions we will call "error in specification of the set of explanatory variables". We will compare explanation levels of models for those well or badly specified (in the above sense).

## 2. Theoretical explanation coefficient of the model $\mathcal{MLRL}$

We define the theoretical explanation coefficient $\varrho^2 = \varrho^2(Y, \beta'\underline{X})$ of the model $\mathcal{MLRL}$ as follows:

$$(1) \qquad \varrho^2 = \frac{\operatorname{cov}(Y, \beta'\underline{X})^2}{\operatorname{var}(Y)\operatorname{var}(\beta'\underline{X})}$$

It is equal to the square of multiple correlation coefficient $\varrho$ between random variable Y and explanatory variables of the model $\mathcal{MLRL}$. Availing definition of variance and the model $\mathcal{MLRL}$ we obtain:

$$\operatorname{var}(\beta'\underline{X}) = \beta'\Sigma\beta ,$$

$$\operatorname{var}(Y) = \sigma_Y^2 = \operatorname{var}(\beta'\underline{X}) + \sigma_\Xi^2 + 2\operatorname{cov}(\beta'\underline{X}, \Xi)$$

$$= \beta'\Sigma\beta + \sigma_\Xi^2 + 2\beta'\operatorname{cov}(\underline{X}, \Xi).$$

Because

$$\begin{aligned}\operatorname{cov}(Y, \beta'\underline{X}) &= \mathcal{E}(Y - \mathcal{E}(Y))\,(\beta'\underline{X} - \beta'\mu)' = \\ &= \mathcal{E}(\beta'\underline{X} + \Xi - \beta'\mu - \mu_{\Xi})\,(\underline{X} - \mu)'\beta = \\ &= \mathcal{E}[\beta'(\underline{X} - \mu) + (\Xi - \mu_{\Xi})]\,(\underline{X} - \mu)'\beta = \\ &= \beta'\Sigma\beta + [\operatorname{cov}(\underline{X}, \Xi)]'\beta\end{aligned}$$

then (1) is of the form

$$(2) \qquad \varrho^2 = \frac{(\beta'\Sigma\beta + [\operatorname{cov}(\underline{X}, \Xi)]'\beta)^2}{(\beta'\Sigma\beta + \sigma_{\Xi}^2 + 2\beta'\operatorname{cov}(\underline{X}, \Xi))\beta'\Sigma\beta}$$

The formula (2) can be introduced also in the form

$$(3) \qquad \varrho^2 = 1 - \frac{\sigma_{\Xi}^2}{\sigma_Y^2} + \frac{\beta'\operatorname{cov}(\underline{X}, \Xi)\ \operatorname{cov}(\underline{X}, \Xi)\beta}{\sigma_Y^2\beta'\Sigma\beta}.$$

In the case when $\operatorname{cov}(\underline{X}, \Xi) : 0 = 0$, $\varrho^2$ takes the form

$$(4) \qquad \varrho_{(o)}^2 = \frac{\beta'\Sigma\beta}{\beta'\Sigma\beta + \sigma_{\Xi}^2}, \quad \varrho_{(o)}^2 = \varrho_{\operatorname{cov}(\underline{X}, \Xi) = 0}^2$$

or

$$(4') \qquad \varrho_{(o)}^2 = 1 - \frac{\sigma_{\Xi}^2}{\sigma_{Y(o)}^2} \quad \sigma_{Y(o)}^2 = \sigma_{Y, \operatorname{cov}(\underline{X}, \Xi) = 0}^2$$

The difference between $\varrho^2$ and $\varrho_{(o)}^2$ is expressed by a value

$$\varrho^2 - \varrho_{(o)}^2 = \frac{\beta' c c' \beta}{\sigma_Y^2\beta'\Sigma\beta} - \frac{\sigma_{\Xi}^2}{\sigma_Y^2} + \frac{\sigma_{\Xi}^2}{\sigma_{Y(o)}^2}$$

$$= \frac{\beta' c c' \beta}{(\beta'\Sigma\beta + \sigma_{\Xi}^2 + 2\beta' c)\beta'\Sigma\beta} + \frac{\sigma_{\Xi}^2\ 2\beta' c}{\sigma_Y^2\,(\sigma_Y^2 - 2\beta' c)}.$$

Denoting $\beta' c = \alpha$, $\beta'\Sigma\beta = \gamma$ we can rewrite the above difference in a following way

$$\varrho^2 - \varrho^2_{(o)} = \frac{\alpha^2}{\gamma(\gamma + \sigma^2_\Xi + 2\alpha)} + \frac{2\sigma^2_\Xi \alpha}{(\gamma + \sigma^2_\Xi + 2\alpha)(\gamma + \sigma^2_\Xi)},$$

$$\varrho^2 - \varrho^2_{(o)} = \frac{\alpha^2(\gamma + \sigma^2_\Xi) + 2\sigma^2_\Xi \alpha\gamma}{\gamma(\gamma + \sigma^2_\Xi)(\gamma + \sigma^2_\Xi + 2\alpha)}.$$

Because of the fact that factors $\gamma$, $\gamma + \sigma^2_\Xi$, $\gamma + \sigma^2_\Xi + 2\alpha$ denote variances of variables $\beta'\underline{X}$, $Y_{cov(\underline{X}, \Xi) = 0}$, Y - the denominator of the above fraction is nonnegative (to secure this condition it is necessary that $\alpha > \frac{-\gamma - \sigma^2_\Xi}{2}$). We assume that this denominator is positive. Now we check for which values of $\alpha$ the considered difference takes the nonnegative values.

$$\varrho^2 - \varrho^2_{(o)} \geq 0 \Longleftrightarrow \alpha^2(\gamma + \sigma^2_\Xi) +$$

$$+ 2\sigma^2_\Xi \alpha\gamma \geq 0 \Longleftrightarrow \alpha\left[\alpha(\gamma + \sigma^2_\Xi) + 2\sigma^2_\Xi \gamma\right] \geq 0$$

The above inequality occurs for

$$\alpha \in \left\{\left(-\infty, \frac{-2\sigma^2_\Xi \gamma}{\gamma + \sigma^2_\Xi}\right\rangle \vee \left\langle 0, +\infty\right)\right\}.$$

Including condition of $\sigma^2_Y$ to be positive, that is $\alpha > \frac{-\gamma - \sigma^2_\Xi}{2}$ and noticing that $\forall \gamma, \sigma^2_\Xi$ : $\frac{-\gamma - \sigma^2_\Xi}{2} \leq \frac{-2\sigma^2_\Xi \gamma}{\gamma + \sigma^2_\Xi}$, we obtain

$$(6) \qquad \varrho^2 - \varrho^2_{(o)} \geq 0 \Longleftrightarrow \alpha \in \left\{\left(\frac{-(\gamma + \sigma^2_\Xi)}{2}, \frac{-2\sigma^2_\Xi \gamma}{\gamma + \sigma^2_\Xi}\right\rangle \vee \langle 0, +\infty)\right\}$$

The equality $\varrho^2 = \varrho^2_{(o)}$ occurs in the case $\alpha = \frac{-2\sigma^2_\Xi \gamma}{\gamma + \sigma^2_\Xi}$ or $\alpha = 0$ (which, for $\beta \neq 0$, means that $c = 0$). Among models where identical linear dependencies exist (that is $\mathcal{S}(\underline{X})$ for these models

is constant and fixed and $\sigma_\Sigma^2$ is constant and fixed) the model in which $c \neq 0$ has greater value of theoretical explanation level (in the sense $\varrho^2 - \varrho_{(o)}^2 \geq 0$), except the case when $\beta' c \in (\alpha^*, 0)$ where $\alpha^* = \frac{-2\sigma_\Sigma^2 \gamma}{\gamma + \sigma_\Sigma^2}$. The following example illustrates this conclusion.

Example. Let $\Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$, $\beta' = [1 \ -2]$, $\sigma_\Sigma^2 = 1$. From this we obtain: $\gamma = \beta' \Sigma \beta = 1{,}4$,

$$\sigma_{Y(o)}^2 = \gamma + \sigma_\Sigma^2 = 2.4,$$

$$\varrho_{(o)}^2 = 1 - \frac{1}{2.4} \approx 0.5833.$$

1. Let $c^{(1)} = [0.5 \ \ 0.2]$. Then $\alpha^{(1)} = \beta' c^{(1)} = 0.1$,

$$\sigma_Y^{2(1)} = \gamma + \sigma_\Sigma^2 + 2\alpha^{(1)} = 2.6,$$

$$\varrho^{2(1)} = 1 - \frac{1}{2.6} + \frac{0{,}01}{2.6 \cdot 1.4} \approx 0{,}6177.$$

Consequently $\varrho^{2(1)} > \varrho_{(o)}^2$.

2. Let $c^{(2)} = [-0.2 \ \ 0.5]$. Then $\alpha^{(2)} = -1.2$,

$$\sigma_Y^{2(2)} = 1.4 + 1 - 2.4 = 0 \Longrightarrow \varrho^{2(2)} \text{ does not exist.}$$

3. Let $c^{(3)} = [0.2 \ \ 0.5]$. Then $\alpha^{(3)} = -0.8$,

$$\sigma_Y^{2(3)} = 1.4 + 1 - 1.6 = 0.8,$$

$$\varrho^{2(3)} = 1 - \frac{1}{0.8} + \frac{0{,}64}{0.8 \cdot 1.4} \approx 1 - 1.25 + 0.57 = 0.32.$$

Consequently $\varrho^{2(3)} < \varrho_{(o)}^2$.

Assuming that $\frac{\sigma_\Xi^2}{\sigma_Y^2} < 1$, the measure $\varrho^2_{(o)}$ is normalized in the range (0, 1). Now, we find the adequate assumptions for $\varrho^2$ when $\mathrm{cov}(\underline{X}, \Xi) \neq 0$. We check[1] whether there exist such vectors $c$ for which $\varrho^2 < 0$ or $\varrho^2 > 1$.

At first we assume $\alpha = \beta' c > \frac{-(\gamma + \sigma_\Xi^2)}{2}$ (necessary condition for positivity of $\sigma_Y^2$).

$$\varrho^2 < 0 \iff 1 - \frac{\sigma_\Xi^2}{\sigma_\Xi^2 + \gamma + 2\alpha} + \frac{\alpha^2}{(\sigma_\Xi^2 + \gamma + 2\alpha)\gamma} < 0 \iff$$

$$\iff \frac{\gamma(\sigma_\Xi^2 + \gamma + 2\alpha) - \sigma_\Xi^2\gamma + \alpha^2}{\gamma(\sigma_\Xi^2 + \gamma + 2\alpha)} < 0 \iff$$

$$\iff \gamma^2 + 2\gamma\alpha + \alpha^2 < 0 \iff (\gamma + \alpha)^2 < 0.$$

Therefore $\forall\, \alpha, \gamma, \beta : \varrho^2 > 0$.

$$\varrho^2 > 1 \iff \frac{-\sigma_\Xi^2\gamma + \alpha^2}{(\sigma_\Xi^2 + \gamma + 2\alpha)\gamma} > 0 \iff$$

$$\iff \alpha^2 \geq \sigma_\Xi^2\gamma.$$

---

[1] From definition of correlation coefficient $\varrho$ it should be that $\varrho^2 \leq 1$. This inequality is obvious for $\varrho^2_o$ (see the relation (4)). However, due to relation (2) and its additive complex form it is not obvious why it should be that $\varrho^2 \leq 1$. From further findings (formulas, relations and examples) it is seen that there are points of the parameter space of the *NMRL* model in which the relation $\varrho^2 \leq 1$ does not hold. Selecting these points is very useful in planning such experiments in which the inequality $\varrho^2 \leq 1$ holds (sensible experiments which are coherent with statistical and econometric theory) and for pre-testing purposes.

Hence for $\alpha \in \{[(-\infty, -\sqrt{\sigma_\Xi^2 \gamma}) \vee (\sqrt{\sigma_\Xi^2 \gamma}, +\infty)] \wedge (\alpha > \frac{-(\gamma + \sigma_\Xi^2)}{2})\}$ the value of $\varrho^2$ exceeds 1.

Summing up, $\varrho^2 \in \langle 0, 1\rangle$ for $\beta' c = \alpha \in \{(-\sqrt{\sigma_\Xi^2 \gamma}, \sqrt{\sigma_\Xi^2 \gamma}) \wedge \alpha > \frac{-(\gamma + \sigma_\Xi^2)}{2}\}$.

In the example from this paragraph the normalizing condition was fullfilled, except $\alpha^{(2)}$.

## 3. Consequencies of wrong specification of explanatory variables

As we noticed in the introduction, the wrong specification of the set of explanatory variables is the one of the causes of the dependence between $X_i$, $i = \overline{1, k}$ and $\Xi$. This dependence is an essential difficulty in an estimation process of the model $\mathcal{MLRL}$ parameters. The least squares estimator is then biased and inconsistent.

To investigate the influence of the wrong specification of the set of explanatory variables we consider two models $\mathcal{M}_1$, $\mathcal{M}_2$. We assume about $\mathcal{M}_1$ and $\mathcal{M}_2$ that

$$
\begin{aligned}
&\mathcal{M}_1: Y = \beta_1' \underline{X}_1 + \Xi_1, \\
&\mathcal{M}_2: Y = \beta_2' \underline{X}_1 + \beta_{2,k+1} X_{k+1} + \Xi_2, \\
&P_{\underline{X}_1} = \mathcal{N}_{\underline{X}_1}(\mu, \Sigma), \quad P_{X_{k+1}} = \mathcal{N}_{X_{k+1}}(\mu_{k+1}, \sigma_{k+1}^2), \\
&\operatorname{cov}(X_{k+1}, \underline{X}_1) = d, \quad P_{\Xi_2} = \mathcal{N}_{\Xi_2}(0, \sigma_{\Xi_2}^2), \\
&\beta_1, \beta_2 \in R^k, \ \beta_{2,k+1} \in R^1, \ \operatorname{cov}(\Xi_2, \underline{X}_1) = 0, \\
&\operatorname{cov}(\Xi_2, X_{k+1}) = 0.
\end{aligned} \tag{7}
$$

Let us suppose that

1) $\mathcal{M}_1$ is badly specified because the essential explanatory variable $X_{k+1}$ is not included. We analyze the consequencies induced by this fact.

2) $\mathcal{M}_2$ is well specified (because variable $X_{k+1}$ is included).

3) $X_1, \ldots, X_k, X_{k+1}$ and $\Xi_2$ are independently distributed.

From the definitions of $\mathcal{M}_1$ and $\mathcal{M}_2$ we obtain the following properties of the noise component $\Xi_1$ of the model $\mathcal{M}_1$.

$$(8) \qquad \Xi_1 = (\beta_2 - \beta_1)' \underline{X}_1 + \beta_{2,k+1} X_{k+1} + \Xi_2$$

$$(9) \qquad \mathcal{E}(\Xi_1) = \mu_{\Xi_1} = (\beta_2 - \beta_1)' \mu + \beta_{2,k+1} \mu_{k+1}$$

$$(10) \qquad \operatorname{var}(\Xi_1) = \sigma^2_{\Xi_1} = (\beta_2 - \beta_1)' \Sigma (\beta_2 - \beta_1) + \beta^2_{2,k+1} \sigma^2_{k+1} + \sigma^2_{\Xi_2} + 2 \beta_{2,k+1} (\beta_2 - \beta_1)' d$$

$$(11) \qquad \operatorname{cov}'(\underline{X}_1, \Xi_1) = c = (\beta_2 - \beta_1)' \Sigma + \beta_{2,k+1} d'$$

It follows that, in general, wrong specification implies dependence (11) between $\underline{X}_1$ and $\Xi_1$, that is, $c \neq 0$ except the case when $d = 0$ and $\beta_1 = \beta_2$ ($X_{k+1}$ is independent from $\underline{X}_1$) - variance-covariance matrix $\Sigma$ is assumed to be nonsingular and positive definite.

## 4. Wrong specification of explanatory variables and theoretical explanation coefficient of a model

Now, we compare the explanation coefficients of $\mathcal{M}_1$ and $\mathcal{M}_2$. From (3) we have:

$$(12) \qquad \varrho^2_1 = \varrho^2(\mathcal{M}_1) = 1 - \frac{\sigma^2_{\Xi_1}}{\sigma^2_Y} + \frac{\beta_1' c c' \beta_1}{\sigma^2_Y \beta_1' \Sigma \beta_1}.$$

Substituting for $\sigma^2_{\Xi_1}$ an expression (10) we obtain

$$\varrho^2_1 = 1 - \frac{\sigma^2_{\Xi_2}}{\sigma^2_Y} -$$

$$- \frac{(\beta_2 - \beta_1)' \mathfrak{A} (\beta_2 - \beta_1) + \sigma^2_{k+1} \beta^2_{2,k+1} + 2\beta_{2,k+1} (\beta_2 - \beta_1)' \mathbf{d}}{\sigma^2_Y} +$$

$$(13) \qquad + \frac{\beta_1' \mathbf{c}\mathbf{c}' \beta_1}{\sigma^2_Y \beta_1' \mathfrak{A} \beta_1}.$$

The explanation coefficient of $\mathcal{M}_2$ can be expressed by the following formula

$$(14) \qquad \varrho^2_2 = 1 - \frac{\sigma^2_{\Xi_2}}{\sigma^2_Y}.$$

From (13) and (14) it follows that the condition for $\varrho^2_2 > \varrho^2_1$ (that is, the explanation level of the well specified model $\mathcal{M}_2$ to be greater than adequate value for badly specified model $\mathcal{M}_1$) is a relation

$$(\beta_2 - \beta_1)' \mathfrak{A} (\beta_2 - \beta_1) + \sigma^2_{k+1} \beta^2_{2,k+1} +$$

$$(15) \qquad + 2 \beta_{2,k+1} (\beta_2 - \beta_1)' \mathbf{d} > \frac{\beta_1' \mathbf{c}\mathbf{c}' \beta_1}{\beta_1' \mathfrak{A} \beta_1}.$$

Because of positive definitness of matrix $\mathfrak{A}$ both sides of the inequality (15) can be, without changing the sign, multiplied by $\beta_1' \mathfrak{A} \beta_1$. Substituting for $\mathbf{c}$ the expression (11) we obtain

$$(\beta_2 - \beta_1)' \mathfrak{A} (\beta_2 - \beta_1) \beta_1' \mathfrak{A} \beta_1 + \beta^2_{k+1} \beta^2_{2,k+1} \beta_1' \mathfrak{A} \beta_1 +$$

$$+ 2 \beta_{2,k+1} (\beta_2 - \beta_1)' \mathbf{d} \beta_1' \mathfrak{A} \beta_1 >$$

$$> \beta_1' \mathfrak{A} (\beta_2 - \beta_1)(\beta_2 - \beta_1)' \mathfrak{A} \beta_1 + \beta^2_{2,k+1} \beta_1' \mathbf{d}\mathbf{d}' \beta_1 +$$

$$+ 2 \beta_{2,k+1} \beta_1' \mathfrak{A} (\beta_2 - \beta_1) \mathbf{d}' \beta_1.$$

After simplification and utilization of the following substitutions

$$\beta_2' \overset{*}{\cdot} \beta_2 = \gamma_{22}, \quad \beta_2' \overset{*}{\cdot} \beta_1 = \gamma_{21}, \quad \beta_1' \overset{*}{\cdot} \beta_1 = \gamma_{11}$$

we obtain a relation

$$\gamma_{22} \cdot \gamma_{11} - \gamma_{21}^2 + \beta_{2,k+1}^2 \beta_1' (\sigma_{k+1}^2 \overset{*}{\cdot} - dd') \beta_1 +$$

$$(16) \qquad + 2\beta_{2,k+1} (\beta_2' d \gamma_{11} - \gamma_{21} d' \beta_1) > 0.$$

Fulfilment of the relation (16) depends on concrete values of vector d elements - that is, on the values of covariances between $X_{k+1}$ and other explanatory variables, and on the degree of dependence between $X_{k+1}$ and Y. Therefore it is not obvious whether some well specified models have always greater coefficients (level) of explanation than some badly specified ones.

## 5. Concluding remarks

In the paper we formulated a measure of theoretical explanation level in linear regression model as the multiple correlation coefficient of explained variable with the set of explanatory variables. We showed that a value of this measure is closely connected with a lack of existence of dependencies between explanatory variables $\underline{X}$ and noise component $\Xi$ which expresses the model error. We found the necessary conditions for this measure to be normalized in the range (0, 1) when there exist non-zero covariances between $\underline{X}$ and $\Xi$. In the badly specified model $\mathcal{M}_1$ a set of explanatory variables is such that considered dependence of the form $\operatorname{cov}(\underline{X}_1, \Xi_1) \neq 0$ exists, except the situation when the variables not included are independent from the rest of explanatory variables. This fact is of great importance in the case when we use mechanical techniques for variable selection (for instance rejection of the variable because of strong multicollinearity although from economical theory we know about its important

influence on explained variable). It can imply that least squares estimates are biased and inconsistent.

It is not, however, obvious in what way good specification implies the level of explanation coefficient of the considered model. It was shown that inclusion of all important variables does not necessarily leads to an increase in the value of explanation coefficient of our linear model.

Our results can be especially useful in designing of simulation experiments - they make possible a serious reduction of experimental parameter space. The authors plan to make similar analysis for empirical models substituting theoretical (population) measures by their estimators.

Iwona Konarzewska, Władysław Milo

## O PEWNYCH SKUTKACH BRAKU NIEZALEŻNOŚCI MIĘDZY SKŁADNIKIEM ZABURZAJĄCYM A ZMIENNYMI OBJAŚNIAJĄCYMI W LINIOWYM MODELU REGRESJI

Przy analizie własności estymatorów wektora parametrów $\beta$ liniowego modelu regresji przyjmuje się, tradycyjnie, założenie o niezależności rozkładów zmiennych objaśniających $X_i$ $i = 1, k$ i zaburzenia losowego $\Xi$ (oznacza to, że w przypadku rozkładów normalnych, cov $(\underline{X}, \Xi) = 0$). W pracy badamy wpływ uchylenia tego założenia na stopień objaśnienia modelu, zdefiniowany jako kwadrat współczynnika korelacji między zmienną objaśnianą $Y$ a liniową kombinacją zmiennych objaśniających $\beta' \underline{X}$. Rozważamy także przypadek, nazwany błędem specyfikacji zbioru zmiennych objaśniających, polegający na nieuwzględnieniu w tym zbiorze ważnej zmiennej, ze względu na jej wysokie skorelowanie z pozostałymi zmiennymi objaśniającymi. Pokazano także, że nie jest oczywiste, jak "dobra" specyfikacja wpływa na wartość stopnia objaśnienia modelu.