DE GRUYTER OPEN

Folia Oeconomica Stetinensia
DOI: 10.1515/foli-2017-0002

WYDZIAŁ NAUK EKONOMICZNYCH I ZARZĄDZANIA
UNIWERSYTET SZCZECIŃSKI

THE ECONOMETRIC PROCEDURES OF SPECIFIC TRANSACTION IDENTIFICATION

Mariusz Doszyń, Ph.D., Associate Prof.[1]
Sebastian Gnat, Ph.D.[2]

University of Szczecin
Faculty of Economics and Management
Institute of Econometrics and Statistics
Mickiewicza 64, 71-101 Szczecin, Poland

[1] *e-mail: mariusz.doszyn@usz.edu.pl*
[2] *e-mail: sebastian.gnat@usz.edu.pl*

Marcin Bas, Msc, Eng.

Independent researcher
e-mail: m.bas@estit.pl

Received 9 January 2017, Accepted 22 March 2017

Abstract

The paper presents the econometric procedures of identifying specific transactions, in which atypical conditions or attributes may occur. These procedures are based on studentized and predictive residuals of the accordingly specified econometric models. The dependent variable is a unit transactional price, and explanatory variables are both the real properties' attributes and accordingly defined artificial binary variables. The utility of the proposed method has been verified by means of a real market data base. The proposed procedures can be helpful during the property valuation process, making it possible to reject real properties that are specific (both from the point of view of the transaction conditions and the properties' attributes) and, consequently, to select an appropriate set of similar attributes that are essential for the valuation process.

Keywords: econometric analysis of real estate market, predictive and studentized residuals, identification of specific transactions

JEL classification: C01, R30

Introduction

The analysis of a real estate market is an indispensable element of each real property valuation. According to § 3.2 of the Regulation of the Council of Ministers, 'the valuation of a property is preceded by the analysis of a real estate market, particularly as regards prices, rents and terms of transactions'. In this analysis the realtor applies a comparative approach to assess the usefulness of information about market transactions concluded on the local market. The applicability of the collected data refers to such aspects as transaction prices that, in compliance with the above mentioned Regulation, cannot grossly deviate from the local market average. Such gross deviation from the average prices can result from unusual conditions in which a given transaction is concluded. As stated in the Regulation, such conditions include in particular, sales made in the course of enforcement proceedings, sales of council property at a preferential price, sales with deferred payment or sales with deferred real property handover to the buyer (§ 5.3). The sales prices can also grossly deviate from the market average due to other factors that are specific for a given property. Then, the price of a real property thus vitiated differs significantly from the prices of real properties with the same variants of characteristics specified by the realtor. In the Polish legislature as well as in the National Rules of Real Property Valuation the term of gross deviation from the average price and the term of a specific real property are not clearly defined. The definition of a similar real property provided by the Act on real property management is controversial and is too much open to interpretation, which can lead to contradictory conclusions. The only provision of law where the frames of the "real property specificity" are defined is an indication in § 5.2. of the aforementioned Regulation that in the process of evaluating a real property for the purpose of the sales by tender the prices deviating by more than 20% from the average prices for similar properties sold on the local real estate market should be rejected. Due to the imprecise legal and market context it is essential to introduce new solutions that will improve the objectivity of judgment and clearly define the specific factors which considerably affect the valuer's choice of a given transaction price for the purpose of another property appraisal. In this study a transaction is considered to be specific when the impact of undisclosed or hard to observe attributes of a valuated property has been identified. The paper proposes a method to determine the impact of such attributes (or non-typical conditions of the transaction) basing it on the residuals of the accordingly specified econometric model. The model used in the paper contains independent variables whose values are easily measurable. Then, a hypothesis is verified that the econometric model residuals carry information about the transaction's specificity, thus facilitating the process

of a real estate market analysis. The element of econometric modelling that will be used to assess the specific character of individual transaction prices will be predictive and studentized residuals. The applicability of such residuals in the real estate market analysis was verified in (Doszyń, Gnat, 2016).

1. Econometric Identification of Specific Transactions

The proposed procedure was applied on the agricultural real estate market. The database of transaction prices contained 22 items. The study covered data about transactions concluded in 2014 and 2015. Since there were no clear trends in the price dynamics, the price correction was refuted due to the passage of time.[1]

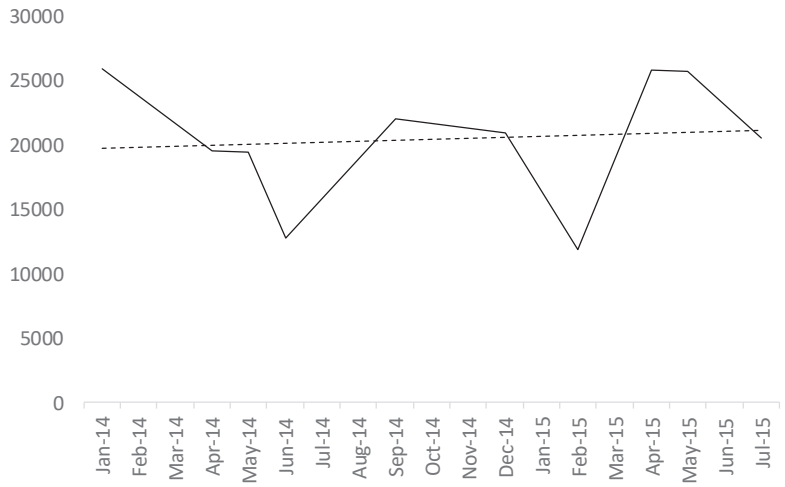

Figure 1. Dynamics of the unit prices of agricultural real properties on a monthly basis (2014–2015) (price PLN/ha)

Source: a realtor's transaction database.

In the first stage we estimated the linear model of the unit prices of agricultural real properties:

$$c_i = \beta_0 + \beta_1 lo_i + \beta_2 ls_i + \beta_3 p_i + \beta_4 k_i + u_i \tag{1}$$

[1] With significance level at $\alpha = 0.05$ there was no basis for hypothesizing that the slope of the trend linear function did not differ significantly from zero. The presented price dynamics was observed on a monthly basis. If there were several transactions in a given month, the prices were averaged. The prices in the months with no recorded transactions were found by linear interpolation.

where:

c_i – unit price of agricultural real properties (PLN/ha),

lo_i – general location,

ls_i – exact location and accessibility,

p_i – size,

k_i – agricultural environment,

u_i – random factor.

Additionally, the difficulties associated with the utilisation of an agricultural property were taken into consideration, despite the impact of that particular attribute was not significant.

The realtor applied the following definitions of market prices:

Table 1. Description of real property attributes

Attribute	Attribute grading	Grading criterion of market attribute
General location	1 – worse (over 12 km) 2 – average (8–12 km>
3 – better (4–8 km>
4 – best (less than 4 km)	Distance in km from the municipal centre (being a local trade and supply centre). The distances are shown on a scale from 1 to 4. Originally, this attribute is a destimulant, therefore, the scale is reversed (the smallest distances in km adopt the largest values and vice versa). The attribute can be described as: worse, average, better, best.
Exact location and accessibility	1 – worse 2 – average 3 – better	The distance from the centre of the nearest village, the type of access and the quality of access roads. The attribute can be described as: worse, average or better.
Size	1 – small (less than 15 ha) 2 – below average (15–30 ha>
3 – average (30–45 ha>
4 – above average (over 45 ha)	The size of a property. As there is larger demand for big agricultural properties, it was observed that prices paid for big single layout plots of land are higher. The attribute can be described as: small, below average, average or above average.
Handicaps	1 – big 2 – small 3 – none	Minuses include: the plot layout, land configuration, the shape of a plot, overhead power lines or power posts, wasteland, buffer strips, etc. The attribute can be described as: big, small or none.
Agricultural environment	1 – poor 2 – average 3 – good	The attribute includes: the level of agricultural measures, the stage of crop development, stones, weeds, soil structure, the level of fertilisation, the condition of drainage systems, etc. Additionally, it comprises the soil quality index according to the utilized agricultural land converters (Act on agricultural tax of 15 November 1984) encompassing soil quality, agricultural utility complex, etc. The attribute can be described as: low, average or good.

Source: own study.

The above variables are qualitative and could be measured only on an ordinary (or nominal) scale. All of the presented variables (attributes) have some categories. General location has

four categories, exact location and accessibility has three categories and so on. Each category should be provided to the model as a separate dummy (0–1) variable. If there is a constant term, the last categories (for example) in the case of all variables should be omitted, because of the strict collinearity. Such a model would have thirteen parameters (twelve dummy variables and a constant term). This would strongly decrease the number of degrees of freedom because there are only twenty two observations. This is why the variables were taken in the form presented in Table 1, what could be disputable but should be treated as a (rather necessary) simplification.

After estimation the model (1) takes the form:

$$\hat{c}_i = -1,578.48 + 1,749.54lo_i + 4,786.88ls_i + 1,313.49p_i + 2,465.53k_i \quad (2)$$

Being based on the Ramsey RESET (*Regression Equation Specification Error Test*), there were no grounds to reject the hypothesis about the linear form of the model.[2] The empirical significance level of this test, in its variant with the square and cube of the theoretical values of the dependent variable, was at $p_{emp} = 0.978$ and justified adopting the linear form.[3]

Apart from the absolute term, in all the remaining cases we should reject the hypothesis that parameters do not differ significantly from zero. The impact of all the attributes of the property was positive, which resulted from the fact that all the variables had been defined in such a way so as to make them stimulants. The principle of coincidence was met because the signs of the coefficients of the correlation of individual attributes with the dependent variable were also positive (they had the same signs as the corresponding grades allocated to the parameters). Notably, the specific location and accessibility (ls_i) also had a considerable effect on property price. Along with the passage to the next variant of that attribute, unit property price rose by the average of 4,786.88 PLN/ha, *ceteris paribus*. The size of the plot (p_i) and the general location (lo_i) turned out to have the least significant impact. The grading of those parameters was: 1,313.49 PLN/ha and 1,749.54 PLN/ha, respectively.

The model fits well into empirical prices. The corrected coefficient of determination $\overline{R}^2 = 0.746$, and the standard error of residuals S_e = 2,264.18 PLN/ha, which makes 10.3% of the average unit price of the examined agricultural real properties. The correlogram (Figure 2) below shows the empirical values as compared with theoretical values. Apart from the good model fit to the real-life values, it indicates that there some properties are specific in terms of

[2] In the test, which is based on F-statistics, the set of independent variables of a linear model is supplemented by the squares and/or cubes of the dependent variable's theoretical values that have been obtained from the linear model. Then, the increase of the coefficient of determination is tested for its significance. If it is significant, the linear form is not correct.

[3] All the hypotheses made in this paper are verified at the significance level of $\alpha = 0.05$.

the aforementioned attributes. The real properties whose prices considerably depart from the line with a 45° slope are likely to have some specific attributes. The attempt to identify these attributes will be made in one of the stages of this study. One of the reasons may be specific conditions in which the transaction was concluded. The data shown in Figure 2 would indicate that theoretical values are identical with the real ones if the all the points were located on a line sloping at 45°, which is not the case. The proposed procedure will allow to answer the question of which real properties the difference between the real and the theoretical value is an effect of some attributes that make us regard a given transaction price as specific?

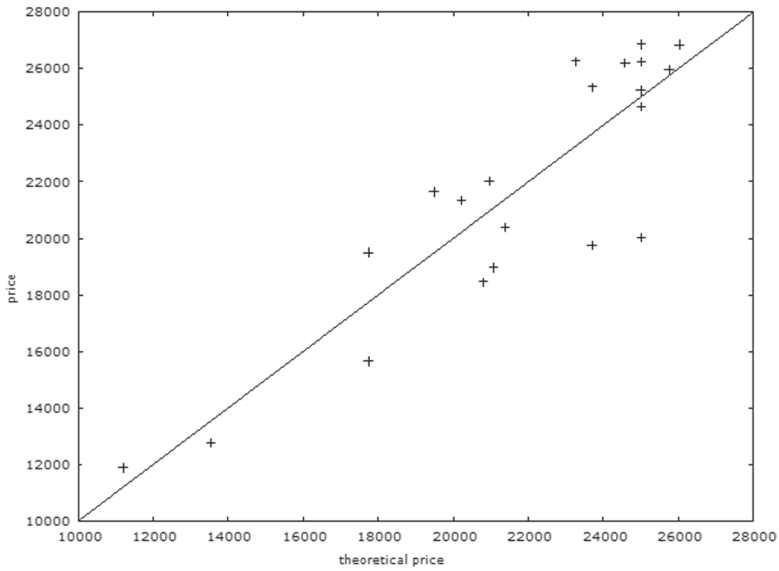

Figure 2. Empirical price values vs. theoretical price values determined with model (2) (PLN/ha)

Source: own study.

From the point of view of the properties of the model parameter estimators, the properties of the random factor are important. White's test proves that it is not justified to reject the hypothesis about the homoskedasticity (the same variance) of the random factor. The empirical level of significance in the applied test $p_{emp} = 0.986$.

The outcome of the Doornik-Hansen test shows that there are no grounds to reject the hypothesis about the conformity of the distribution of residuals to the normal distribution. The empirical significance level $p_{emp} = 0.100$. However, based on the figure below, we can say that there are 'specific' residual deviations that can also confirm the occurrence of 'specific' real properties, which will be verified further on in the paper.

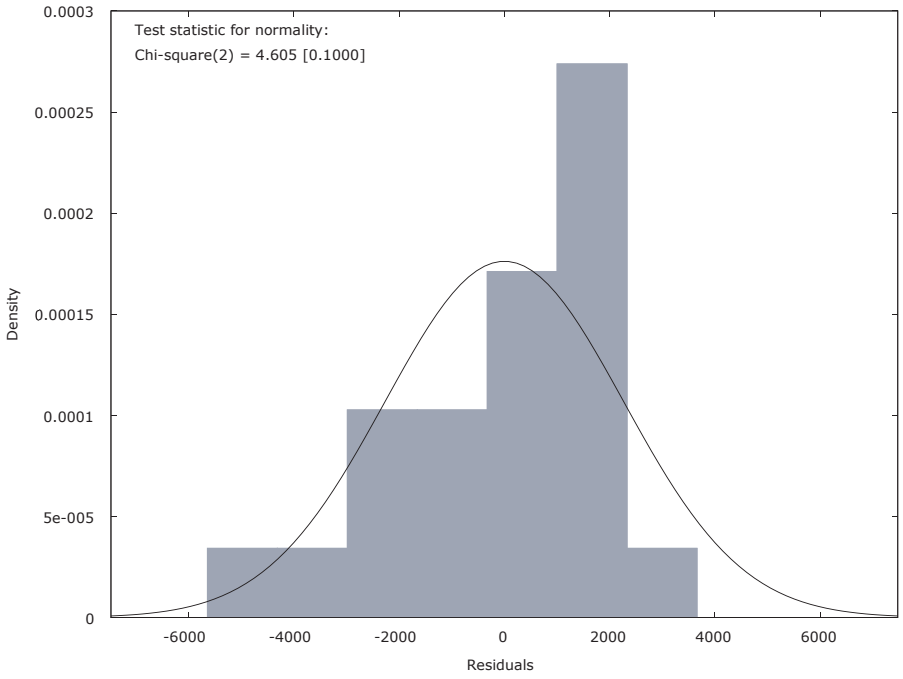

Figure 3. Model (2) residuals distribution

Source: own study.

In the next stage of the study we attempted to identify econometrically these real properties that could have some specific attributes in terms of the conditions in which the transaction was concluded. To this end we used predictive and studentized residuals calculated for model (2). The ways to determine different types of residuals, along with their strong and weak points are outlined in, e.g. (Doszyń, 2013). Residuals of the type mentioned above are described in numerous econometric publications, such as (Greene, 2003). The use of residuals of the least squares method to investigate the behaviour of economic entities (cost-efficiency of businesses) is discussed in such works as (Hozer, 1993; Pawłowski, 1976).

The predictive and studentized residuals were estimated by means of dummy variables z_i, where $i = 1, 2, \ldots, n$, and n is a number of real properties. Therefore, the number of estimated models $n = 22$:

$$\begin{aligned} c_i &= \beta_0 + \beta_1 lo_i + \beta_2 ls_i + \beta_3 p_i + \beta_4 k_i + \beta_{51} z_1 + u_i \\ c_i &= \beta_0 + \beta_1 lo_i + \beta_2 ls_i + \beta_3 p_i + \beta_4 k_i + \beta_{52} z_2 + u_i \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ c_i &= \beta_0 + \beta_1 lo_i + \beta_2 ls_i + \beta_3 p_i + \beta_4 k_i + \beta_{5n} z_n + u_i \end{aligned} \tag{3}$$

The assessments of the parameter β_{5i} are the predictive residuals, while the Student t-statistics determined for them are the studentized residuals. There are as many predictive and studentized residuals as there are real properties (n = 22). When verifying the hypothesis that a given predictive (studentized) residual differs significantly from zero, we adopted the level of significance at $\alpha = 0.05$.

The procedure described here is a multi-stage one. If it turns out that a given predictive residual differs significantly from zero, the binary variable z_i is added to the set of independent variables and the procedure of estimating the predictive (and studentized) residuals is repeated. The whole procedure is complete when in a given stage there are no grounds to reject the hypothesis that any residual is significantly different from zero. If the procedure was single-stage, the share of significant residuals would be similar to the adopted significance level α, i.e. it would roughly correspond to the probability of the type I error (i.e. of rejecting a true null hypothesis).

Particular mention should be made of problems associated with applying the proposed procedure. First of all, we cannot rule out the possibility that what has been regarded as a specific factor can be a matter of chance. Therefore, it seems to be more appropriate to hypothesise about the occurrence of specific factors. What is more, the accompanying data mining is related to the fact that the applied 'nominal' significance levels differ from the actual significance levels, see: (Charemza, Deadman, 1997).

The results of the proposed procedure are shown in Table 2.

Table 2. Predictive and studentized residuals

Property No.	Stage I		Stage II		Stage III	
	predictive residuals	studentized residuals	predictive residuals	studentized residuals	predictive residuals	studentized residuals
1	2	3	4	5	6	7
1	398.32	0.12	396.01	0.14	1,127.58	0.51
2	**–5,505.26**	**–2.71**	–	–	–	–
3	–4,635.45	–1.43	–4,860.25	–1.82	–4,168.87	–1.94
4	–2,482.82	–1.07	–2,616.34	–1.35	–2,806.30	–1.87
5	–1,421.34	–0.52	–1,589.66	–0.68	–1,557.23	–0.83
6	–1,346.42	–0.44	–1,181.24	–0.45	–1,121.08	–0.53
7	1,599.16	0.57	1,565.70	0.65	2,303.22	1.21
8	1,654.27	0.59	847.14	0.35	263.50	0.13
9	–4,572.63	–2.05	**–5,152.84**	**–3.03**	–	–
10	1,445.65	0.47	1,098.16	0.42	–158.65	–0.07
11	–2,634.06	–1.04	–2,495.79	–1.16	–2,827.24	–1.70
12	1,225.62	0.40	1,685.07	0.64	1,720.51	0.81
13	2,007.78	0.79	1,486.09	0.68	1,448.20	0.82

1	2	3	4	5	6	7
14	1,375.46	0.57	797.77	0.38	266.03	0.15
15	–399.50	–0.16	–997.58	–0.48	–1,549.72	–0.92
16	3,304.56	1.43	2,918.24	1.49	2,603.19	1.66
17	2,038.30	0.85	1,468.22	0.71	944.11	0.55
18	250.06	0.10	–340.56	–0.16	–885.23	–0.52
19	230.15	0.09	–360.70	–0.17	–905.60	–0.53
20	1,915.01	0.78	1,393.66	0.66	543.45	0.31
21	2,216.68	0.87	2,357.69	1.09	2,043.76	1.17
22	2,511.54	1.03	2,447.44	1.19	1,956.07	1.16
Critical value	2.12		2.13		2.14	

Source: own study.

In the first stage, it was the predictive (and studentized) residual attributed to property No. 2 that differed significantly from zero, while in the second stage – it was the residual attributed to property No. 9. Understandably, in the second stage the residual ascribed to property No. 2 was not determined, and in the third stage the residuals for the properties No. 2 and 9 were left out. As it has been mentioned before, if a given residual is significantly different from zero, then in the next stage the binary variable z_i corresponding to a given property is added to the set of independent variables. The aim is to eliminate the impact of a significant residual in the subsequent stage of the procedure in order to 'spot' the next specific observations. In the third stage, in the case of all the determined residuals, there was no reason to reject a hypothesis about their insignificance, which meant the end of the procedure.

Based on the obtained results a conclusion can be stated that in the case of properties No. 2 and 9 there were significant specific factors (or specific transaction conditions) that caused the lowering of the agricultural property price. The residuals were negative; hence specific factors reduced the average property price. The analysis of the available information about properties No. 2 and 9 did not provide clearly specified factors that would have led to prices lower than the ones that were typical of other properties with identical variants of market attributes. This suggests that in those cases the transactions could have been concluded in non-typical conditions. Consequently, those two transactions should be eliminated from the set of the transactions under this market analysis.

In order to define precisely the impact of specific attributes a model was estimated where, beside the observed property attributes, the set of independent variables contained binary variables for properties No. 2 and 9:

$$\begin{aligned}\hat{c}_i &= -2,796.31 + 2,099.44lo_i + 4,856.27ls_i + 1,119.15p_i + \\ &+ 2,824.23k_i - 5,982.99z_2 - 5,152.84z_9, \overline{R}^2 = 0.878, S_e = 1,571.175\end{aligned} \tag{4}$$

Apart from the absolute term, in all cases a hypothesis should be rejected that the parameters do not differ significantly from zero (the significance level $\alpha = 0.05$). At such significance level there is no reason to refute a hypothesis about the homoskedasticity of the random factor as well as a hypothesis about the conformity of residual distribution to the normal distribution. We can say that the impact of specific factors lowered the unit property price by 5,982.99 PLN/ha (property No. 2) and by 5,152.82 PLN/ha (property No. 9).

Quality of the models (2) and (4) is compared in Table 3.

Table 3. Comparison of models (2) and (4) by selected criteria

Criterion	Model (2)	Model (4)
$\overline{R}^2$	0.746	0.878
AIC	406.660	391.829
BIC	412.115	399.466
HQC	407.945	393.628
White's test – p_{emp}	0.986	0.084
Doornik – Hansen test – p_{emp}	0.100	0.790

Source: own study.

From the point of view of almost every criterion (excluding the empirical significance level in White's test), model (4) is better specified than model (2). The conformity of the distribution of residuals to the normal distribution is evidently improved. In model (4) the values of information criteria *AIC*, *BIC*, *HQC* are smaller (more favourable). Model (4) not only allows us to determine the impact of the specific attributes of a real property, but it is also better specified.

Conclusions

Finding the solution to the problem of the unbiased identification of specific (non-typical) transactions is essential from the point of view of real estate market analysis and real property valuation. In the process of valuation the use of data that grossly deviate from typical data bears the risk of erroneous results, and the inadequate grading of attributes that affect the market value of a given property. The imperfect quality of market information has been addressed by many authors. One of them is Professor E. Kucharska-Stasiak who claims that 'real properties that are available on the market are heterogeneous and have diverse attributes. (…) on the real estate market the distribution of prices of individual real property attributes is not known' (Kucharska-Stasiak, 2010). This paper proposes an econometric procedure that can be used to determine the

impact of the specific attributes of a real property on the basis of the predictive and studentized residuals of a specified econometric model. As it has been shown above, the real properties that were considered non-typical are likely to have been concluded in specific conditions and should be excluded from any further analyses. The presented results show that the proposed procedure can add more objectivity to the real estate market analysis performed for the purpose of real property valuation. The procedure provides realtors with a tool to improve the accuracy of representing regularities on the real estate market by eliminating from the analysed databases those elements that, due to the impact of specific factors, disturb the analysis thus distorting the obtained results as well as increasing the risk of an erroneous market analysis based on data that cannot be legally taken into account.

References

Charemza, W.W., Deadman, D.F. (1997). *Nowa ekonometria*. Warszawa: PWE.

Doszyń, M. (2013). *Skłonności w ekonomii. Ujęcie ilościowe*. Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.

Greene, W.H. (2003). *Econometric analysis*. Fifth edition. Upper Saddle River, New Jersey: Prentice Hall.

Hozer, J. (1993). *Mikroekonometria. Analizy, diagnozy, prognozy*. Warszawa: PWE.

Kucharka-Stasiak, E. (2010). Odwzorowanie cech nieruchomości w cenach i skutki dla procesu wyceny. *Studia i Materiały Towarzystwa Naukowego Nieruchomości*, *18* (3), 7–15.

Pawłowski, Z. (1976). *Ekonometryczna analiza procesu produkcyjnego*. Warszawa: PWN.