INFORMATION
SYSTEMS IN
MANAGEMENT

Information Systems in Management (2016) Vol. 5 (3) 307–317

ROUNDWOOD PRODUCTION FORECASTING IN POLAND, ON THE BASIS OF THE DATA OF THE CENTRAL STATISTICAL OFFICE IN THE YEARS 2000 - 2012

LYUBOV ANDRUSHKO

Kielce University of Technology,
Faculty of Management and Computer Modelling,
Chair of Management and Marketing

The paper presents the results of research on the construction and evaluation of the forecast of the economic phenomenon in the future time period on the example of roundwood production in Poland.

The process of building an econometric model as a linear function is illustrated in this paper. The trend function was verified and the convergence coefficient was calculated. The point and the interval forecast was constructed. On the basis of the results of studies the forecast accuracy was evaluated and a number of recommendations were presented. The recommendations regard the use of research results for decision-making and the needs of the material economy.

Keywords: forestry and wood industry, roundwood production, econometric model, point forecast, interval forecast, forecast evaluation

1. Introduction

On the basis of the analysis of economic phenomena and processes in the forestry and timber sector one can assume that in modern environmental issues related to the use of forest resources the relative and absolute indicators characterizing their economic effectiveness will be essential in the evaluation of ecological effectiveness of using natural resources. The condition of the natural environment will be affected by its monitoring, proper determination of developmental objectives in the forest management and related sectors.

A systemic approach to the rational use of natural resources nowadays becomes a guarantee to reduce the economic, social and ecological risk. Therefore it is necessary that the executives and business operators have the ability to anticipate future economic events and estimate in their developmental, economic and financial plans the necessary amounts of roundwood production, its price, production revenues and costs.

The aim of the paper is to present the individual stages of constructing the econometric model for the needs of the executives involved in the planning of economic activity in the General Directorate of the State Forests in Poland.

Source data were obtained from the yearbook of the CSO ”Forestry” from the years 2020-2012. The stages of constructing the model include: a preliminary exploration (data preparation), model building, evaluation and verification of the forecasting model, drawing conclusions on the basis of the model verification results and making recommendations for the practical use of the results of the forecast amounts of roundwood.

2. Data exploration (preparation of data for the model)

Data on the amount of roundwood production in the years 2000 – 2012 is presented in Table 1.

Table 1. Roundwood production in the years 2000-2012, thousand m^3

T	Year	Roundwood production in thousand m^3 - y_t
1	2000	27659
2	2001	26671
3	2002	28957
4	2003	30836
5	2004	32733
6	2005	31945
7	2006	32384
8	2007	35935
9	2008	34273
10	2009	34629
11	2010	35467
12	2011	37180
13	2012	37045

Source: based on the data of the Central Statistical Office

3. Construction of the econometric model

Figure 1 shows the scattering diagram, that is, the data from Table 1. The distribution of the empirical points forming a time series indicates that a trend function with accuracy to the random component in the form of linear function (1) can be adopted to describe this phenomenon.

$$\hat{y}_t = a_1 t + a_0 \tag{1}$$

where

y_t – the amount of roundwood produced in the consecutive years , in thousands m^3,
t = 1, 2, ..., 13 – years corresponding to the calendar years 2000, 2001, ..., 2012,
a_1, a_0 – structural parameters of the model (in other words numbers),
n = 13 – number of observations – number of data.

It is called building an econometric model of roundwood production, as a linear function.

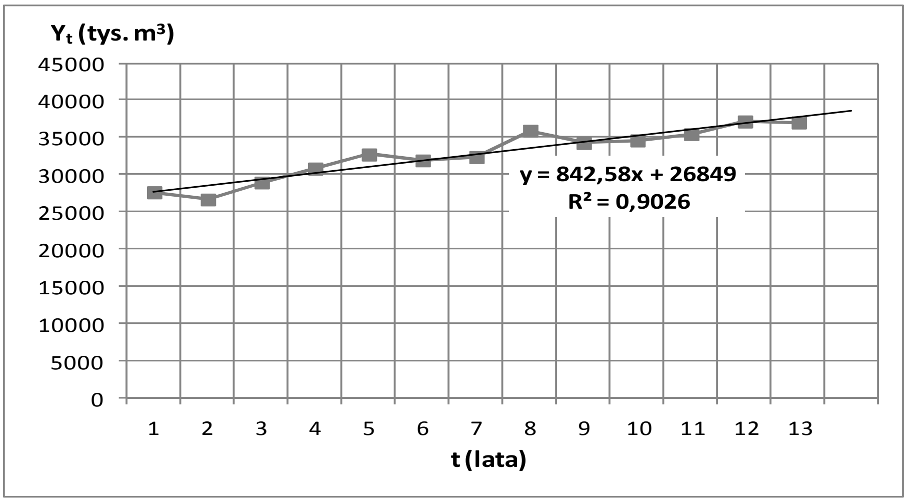

Figure 1. Scatter diagram – data from Table 1

The aim of the econometric analysis is to determine the structural parameters a_1, a_0 in the model (1). In order to do that auxiliary calculations were performed in Table 2.

To estimate the structural parameters of this function the dependencies resulting from the classical method of the least squares were used [1].

$$\begin{cases} y_t = na_0 + a_1 \sum t \\ \sum (y_t t) = a_0 \sum t + \sum t^2 \end{cases} \tag{2}$$

After substituting the appropriate data from table 2 we obtain:

$$\begin{cases} 425714 = 13a_0 + 91a_1 \\ 3133348 = 91a_0 + 819a_1 \end{cases} \tag{3}$$

Table 2. Auxiliary calculations

Year	t	y_t	$y_t t$	t^2	$\hat{y}_t$	$(y_t - \hat{y}_t)^2$	$(y_t - \bar{y}_t)^2$
2000	1	27659	27659	1	27691,732	1071,383824	25890092,36
2001	2	26671	53342	4	28534,314	3471939,063	36920580,36
2002	3	28957	86871	9	29376,896	176312,6508	14365849,28
2003	4	30836	123344	16	30219,478	380099,3765	3652803,053
2004	5	32733	163665	25	31062,06	2792040,484	202,5147929
2005	6	31945	191670	36	31904,642	1628,768164	643574,2071
2006	7	32384	226688	49	32747,224	131931,6742	131936,5917
2007	8	35935	287480	64	33589,806	5499934,898	10161872,67
2008	9	34273	308457	81	34432,388	25404,53454	2327971,746
2009	10	34629	346290	100	35274,97	417277,2409	3541055,438
2010	11	35467	390137	121	36117,552	423217,9047	7397144,669
2011	12	37180	446160	144	36960,134	48341,05796	19649443,05
2012	13	37045	481585	169	37802,716	574133,5367	18470820,36
$\sum$	91	425714	3133348	819	-	13943332,6	143153346,3

By solving the system of equations (3) we obtain the values of parameters a_1, a_0.

$$\begin{cases} a_1 = 842{,}582 \\ a_0 = 26849{,}15 \end{cases} \tag{4}$$

The equation of the linear trend function has then the following form:

$$\hat{y}_t = 842{,}58t + 26849{,}15 + e_t \tag{5}$$

where $e_t = y_t - \hat{y}_t$ – remainder component.

The graphic illustration of the trend function is presented in Figure 1.

The average of the y_t variable is

$$\bar{y}_t = \frac{1}{n}\sum_{t=1}^{13} y_t = \frac{1}{13}425714 = 32747{,}23$$

The average of the t variable is

$$\bar{t} = \frac{1}{n}\sum_{t=1}^{13} t = \frac{1}{13}91 = 7$$

3. Results of the empirical studies

Verification of the trend function. The verification of the trend function (econometric model) consists in checking how well it matches the empirical data. We calculate variation of the random component:

$$S_e^2 = \frac{\sum_{i=1}^{13}(y_t - \hat{y}_t)^2}{n-k} \qquad (6)$$

where n = 13 – the number of data, k = 2 – the number of estimated parameters (a_0, a_1) of the trend function.

From Table 2 we read

$$\sum_{t=1}^{13}(y_t - \hat{y}_t)^2 = 13943332{,}6$$

hence

$$S_e^2 = \frac{13943332{,}6}{13-2} = 1267575{,}7 \cdot$$

The standard deviation of the remainder component – the estimation error of the model is

$$S_e = \sqrt{S_e^2} = \sqrt{1267575{,}7} = 1125{,}86 \text{ thousand m}^3 \qquad (7)$$

The remainder variation coefficient is

$$V_e = \frac{S_e}{\bar{y}_t} * 100\% = \frac{1125{,}86}{32747{,}23} * 100\% = 3{,}44\% \qquad (8)$$

The obtained coefficient V_e < 10%.

It is significantly less than 10%, which indicates a good match of the model to the empirical data.

Calculation of the convergence coefficient. The convergence coefficient was determined from the dependency according to [1] and data from Table 1.

$$\varphi^2 = \frac{\sum_{t=1}^{13}(y_t - \hat{y}_t)^2}{\sum(y_t - \hat{y}_t)^2} = \frac{13943332{,}6}{143153346{,}3} = 0{,}0974 \qquad (9)$$

The determination coefficient is defined from the dependency

$$R^2 = 1 - \varphi^2 = 1 - 0{,}0974 = 0{,}9026 \qquad (10)$$

The value of the determination coefficient indicates that 90.26% variation of the variable y_t is explained by the model (1).

Verification of the structural parameters. Calculating average estimation error of the structural parameters: for parameter a_1 according to [1] and data from Table 1.

$$D(a_1) = \frac{S_e}{\sqrt{\sum t^2 - n * \bar{t}^2}} = \frac{1125{,}86}{\sqrt{819 - 13 * 7^2}} = \frac{1125{,}86}{13{,}49} = 83{,}46 \qquad (11)$$

for parameter a_0 according to [1] data from Table 1.

$$D(a_1) = S_e\sqrt{\frac{\sum t^2}{n\left(\sum t^2 - n\bar{t}^2\right)}} = 1125{,}86\sqrt{\frac{819}{13(819 - 13 * 7^2)}} = 1125{,}86 * 0{,}58 = 662{,}39 \qquad (12)$$

The relative average estimation errors of parameters a_1 and a_0 are respectively [1]:

$$\begin{cases} V_{a_1} = \dfrac{D(a_1)}{a_1}100\% = \dfrac{83{,}45}{842{,}582}100\% = 9{,}9 < 50\% \\ V_{a_0} = \dfrac{D(a_0)}{a_0} = \dfrac{662{,}398}{26849{,}15}100\% = 2{,}47\% < 50\% \end{cases} \qquad (12a)$$

And are significantly smaller than 50%, which indicates good match of the model to the empirical data.

The econometric model of roundwood production in thousands of m^3 in the consecutive years is noted in the general form

$$\hat{y}_t = a_1 t + a_0 + e_t \qquad (13)$$

$D(a_1),\ D(a_0),\ S_e \quad R^2$

And for this case it is

$$\hat{y}_t = 842{,}58t + 26849{,}15 + e_t \qquad (14)$$

$$(83{,}46) \quad (662{,}39) \quad (1125{,}86) \quad R^2 = 0{,}9026$$

The quality of function (14) is assessed by determining the significance of the trend slope coefficient, the so-called significance of parameter a_1.

We hypothesize, $H_0 : a_1 = 0$ against the alternative hypothesis $H_1 : a_1 \neq 0$.

For verification of H_0 we use the statistics according to [1].

$$T = \frac{a_1}{D(a_1)} = \frac{842{,}58}{83{,}46} = 10{,}09 \qquad (15)$$

From the t-Student distribution table, for the assumed confidence level of $1 - \alpha = 0.95$, and $n - k = 13 - 2 = 11$ degrees of freedom, we read the critical value of the test $t_\alpha = 2{,}201$.

The critical test area for the assumed hypothesis H_1 is:

$$\begin{cases} K = (-\infty, -t_\alpha) \cup (t_{\alpha,} + \infty) \\ K = (-\infty, -2{,}201) \cup (2{,}201, +\infty) \end{cases} \qquad (16)$$

Conclusion: $T \in K,$ (17)

means that the H_0 hypothesis should be rejected in favour of the H_1 hypothesis, this means that the parameter a_1 is significantly different from zero. The value of this parameter $a_1 = 842{,}58$ m^3/year is interpreted as follows: for the years 2000-2012, on average, each year 842,58 thousand m^3 more roundwood was produced in comparison to the previous one.

Summary of the outcomes:

a) statistically significant estimation of the parameter a_1 – dependences (4) and (17);

b) low average estimation errors of parameters $a_1 \, and \, a_0$ – dependencies (11) and (12) and low relative errors (12a);

c) low value of the remainder variation coefficient V_e – dependency (8);

d) determination coefficient R^2 close to unity – (10), indicates that the linear trend function (14) describes well the amount of the roundwood production in the time function and can be used for short period forecasting.

4. Roundwood production forecast

Point forecast. The point forecast of the amount of roundwood in the year 2013, that is, for $t = T = 14$ determined by the relation (5) and is

$$y^*_{T=14} = a_1 T + a_0 = 842{,}58 * 14 + 26849{,}15 = 38645{,}298 \text{ thousand m}^3 \quad (18)$$

The point forecast is given with the accuracy to standard deviation of the remainder component S_e dependency (7).

$$\begin{cases} y^*_{T=14} = a_1 T + a_0 \pm S_e \\ y_{T=14} = 38645{,}298 \pm 1125{,}86 = [37519{,}438 \div 39771{,}158] \end{cases} \quad (19)$$

Next, the absolute and relative forecast error is determined according to [1] and data from Table 2.

Absolute forecast error

$$D_T = S_e \sqrt{1 + \frac{1}{n} + \frac{(T - \bar{t})^2}{\sum_{i=1}^{n} (t - \bar{t})^2}} = 1125{,}86 \sqrt{1 + \frac{1}{13} + \frac{(14-7)^2}{182}} = 1306{,}266 \text{ thousand m}^3 \quad (20)$$

The relative error of this forecast is equal to:

$$D_{Trelative} = D'_T = \frac{D_T}{y^*_T} * 100\% = \frac{1306{,}266}{38645{,}298} * 100\% = 3{,}38\% < 10\% \quad (21)$$

It is assumed that the relative forecast error D'_T less than 10% is a small relative error. This means that the resulting forecast (19) may be regarded as acceptable. The estimated linear trend function is a good predictor.

Interval forecast. In order to make the interval forecast one should check earlier the normality of the random component distribution of the trend function. The check will be performed with the Jacque-Berry test. Table 3 contains the appropriate auxiliary calculations.

The null hypothesis is tested:

H_0 - the random component of the trend function has a normal distribution, with the alternative hypothesis

H_1 - the random component of the trend function does not have the normal distribution.

The JB statistics according to [1] and data from Table 2 are used to verify H_0, defined as follows:

$$JB = n\left[\frac{1}{6} * B_1 + \frac{1}{24} * (B_2 - 3)^2\right] \tag{22}$$

where

$$\begin{cases} B_1 = \left(\frac{1}{n}\sum \frac{e_t^3}{S_t^3}\right) = \left(\frac{1}{13}9{,}21277\right)^2 = 0{,}502 \\ S_t = \sqrt{\frac{1}{n}\sum e_t^2} = \sqrt{\frac{1}{13}13943332{,}6} = 1035{,}647 \\ B^2 = \frac{1}{n} * \sum \frac{e_t^4}{S_t^4} = \frac{1}{13}44{,}313 = 3{,}408 \\ e^t = y_t - \hat{y}_t - remainder...component \end{cases} \tag{23}$$

Table 3. Auxiliary calculations cont

Year	t	y_t	$\hat{y}_t$	$y_t - \hat{y}_t$	e_t^2	e_t^3	e_t^4	$\frac{e_t^3}{S_t^4}$	$\frac{e_t^4}{S_t^4}$
2000	1	27659	27691,732	-32,732	1071,384	-35068,53533	1147863,298	-3,1571E-05	9,978E-07
2001	2	26671	28534,314	-1863,314	3471939,063	-6469312662	1,20544E+13	-5,8240258	10,4784663
2002	3	28957	29376,896	-419,896	176312,651	-74032976,83	31086150838	-0,0666485	0,02702219
2003	4	30836	30219,478	616,522	380099,376	234339627,8	1,44476E+11	0,21096523	0,12558791
2004	5	32733	31062,06	1670,94	2792040,484	4665332126	7,79549E+12	4,19998477	6,77636756
2005	6	31945	31904,642	40,358	1628,768	65733,82556	2652885,732	5,9177E-05	2,3061E-06
2006	7	32384	32747,224	-363,224	131931,674	-47920750,42	17405966651	-0,04314086	0,01513044
2007	8	35935	33589,806	2345,194	5499934,898	12898414322	3,02493E+13	11,6118515	26,2947248
2008	9	34273	34432,388	-159,388	25404,535	-4049177,952	645390375,4	-0,00364529	0,00056102
2009	10	34629	35274,97	-645,97	417277,241	-269548579,3	1,7412E+11	-0,24266224	0,15135715
2010	11	35467	36117,552	-650,552	423217,905	-275325254,3	1,79113E+11	-0,24786271	0,15569749
2011	12	37180	36960,134	219,866	48341,058	10628555,05	2336857884	0,0095684	0,00203136
2012	13	37045	37802,716	-757,716	574133,537	-435030166,9	3,29629E+11	-0,39163773	0,28653611
$\sum$	91	425714	425713,91	0,088	13943332,57	10233525728	5,0978E+13	9,21277437	44,3134856

Thus, the empirical value of the JB statistics is:

$$JB = 13 * \left[\frac{1}{6} * 0{,}502 + \frac{1}{24} * (3{,}408 - 3)^2\right] = 13 * [0{,}0837 + 0{,}007] = 1{,}1786 \tag{24}$$

JB statistics has the *chi-square* distribution with two degrees of freedom. For the assumed confidence level of 1 – α = 0,95, we read the critical values of the test from the distribution tables x^2.

$$x_{\alpha}^{2} = 5{,}991 \qquad (25)$$

For this case

$$JB = 1{,}1786 < x_{\alpha}^{2} = 5{,}991 \qquad (26)$$

means that there is no basis for rejecting the H_0 hypothesis declaring that the remainder component of the trend function has normal distribution.
Now the interval forecast can be constructed:

$$T\{y_T^* - t_\alpha D_t < y_t < y_T^* + t_\alpha D_t\} = 1 - \alpha \qquad (27)$$

For the confidence level $1-\alpha = 0{,}95$ and the sample size n = 13, value t_α is read from the distribution table *t-Student* for n – 2 = 13 – 2 = 11 degrees of freedom: $t_\alpha = 2{,}201$. Considering the dependencies (18) and (20), the interval forecast will be noted:

$$38645{,}298 - 2{,}201 * 1306{,}266 < \mathrm{y_T} < 38645{,}298 + 2{,}201 * 1306{,}266$$
$$37770{,}206 < y_T < 41520{,}389 \text{ thousand m}^3 \qquad (28)$$

Absolute error *ex-ante* of this forecast is:

$$V_t = |t_\alpha D_T| = 2{,}201 * 1306{,}266 = 2875{,}09 \text{ thousand m}^3 \qquad (29)$$

Relative error *ex-ante* of this forecast is:

$$V_{Trelative} = V_T' = \frac{V_T}{y_T^*} * 100\% = \frac{2875{,}09}{38645{,}298} * 100\% = 7{,}43 < 10\% \qquad (30)$$

The errors (29) and (30) can be considered low and the forecast can be regarded as acceptable. The forecast value for $T = 14$ is in the range of (28) with a probability of 95%.

Evaluation of the results. From the data of the Central Statistical Office the amount of roundwood production was $y_{14} = 37996\, thousand\ m^3$. The error *ex-post* for our forecast is:

$$V_{post} = y_{T=14}^* - y_{14} = 38645{,}298 - 37996 = 649{,}29 \text{ thousand m}^3$$
$$V_{post} < S_e = 1125{,}86 \text{ thousand m}^3$$

This means that the forecast was fulfilled almost 100%.

5. Discussion and summary

The paper illustrates only one type of forecasting model. The linear trend function describes well the amount of roundwood production in the time function and can be applied to short-term forecasting. However, if the analysis results indicated a low level of matching and a large *ex post* error, it would be advisable to analyze several forms of non-linear econometric models and select the form of model that would match best with the empirical data.

In practical business it is not only forest management employees that are interested in the forecast results but also business operators involved in wood processing including: production of paper, electric energy, furniture or minor wood processing that need this kind of information.

For the purpose of monitoring the natural environment, forecasting the amount of roundwood provides information on the level of roundwood production, which should remain within the tolerance limits of using natural resources.

REFERENCES

[1] Sobczyk M. (2008) *Prognozowanie. Teoria. Przykłady. Zadania,* Wydawnictwo Placet, Warszawa

[2] Kufel T. (2011) *Ekonometria: rozwiązywanie problemów z wykorzystaniem programu GRETL,* Polskie Wydawnictwo Naukowe PWN, Warszawa

[3] Borkowski B., Dudek H., Szczęsny W. (2003) *Ekonometria: wybrane zagadnienia,* Polskie Wydawnictwo Naukowe PWN, Warszawa

[4] Ignatczak W., Chromińska M. (2004) *Statystyka: teoria i zastosowanie,* Wydawnictwo Wyższej Szkoły Bankowej, Poznań

[5] Błaszczuk D. (2014) *Podstawy prognozowania, symulacji i sterowania optymalnego* Polskie Wydawnictwo Naukowe PWN, Warszawa

[6] Lis Ch. (2010) *Modelowanie predyktywne wartości dodanej brutto w Polsce,* Centrum Badań Ekonomicznych, Szczecin