Paper prepared for the 123rd EAAE Seminar

PRICE VOLATILITY AND FARM INCOME STABILISATION

Modelling Outcomes and Assessing Market and Policy Based Responses

Dublin, February 23-24, 2012

Modelling yield risk measures of major crop plants in Poland

Kobus P.

Warsaw University of Life Sciences, Faculty of Economics, Warsaw, Poland

pawel_kobus@sggw.pl

Modelling yield risk measures of major crop plants in Poland

Kobus P.

Abstract

The paper deals with the problem of modelling yield risk measures for major crop plants in Poland. Hence, in some cases the gamma distribution offers a better fit to the data than normal distribution, and in addition to linear models, generalized linear models were also used. The research was based on data from Polish FADN, with sample sizes ranging from 416 up to 2300, depending on the crop plant. It was found that models based on the farm level data, can explain on average 20% of variation coefficient unevenness. The most important variables were average yield, type of farming, arable area and land quality. The elimination of the average yield from the models reduced the average determination coefficient to about 9%.

Keywords: production risk, risk measures, yield distribution

JEL classification: Q10, C46

1. INTRODUCTION

Yield risk is one of the prevailing risks in the agriculture sector. The most important factors influencing yield levels are weather conditions, pests and diseases. While Poland is a country relatively moderate in size, the weather patterns and soil conditions are quite diverse. The same can be said about the interaction between weather and soil conditions. And, on top of that, there are differences in production technology used by farmers. Consequently, the yield variability is rather specific to a farm or even a field. Due to the lack of necessary data, analyses of yields variability in Poland were carried out only on the aggregated data, and this was mostly the national or voivodeship level of aggregation (Kobus, 2010).

It is evident, that higher level of aggregation results in smaller yields variability although a decrease may depend on various factors, for example Marra and Schurle reported that the effect of aggregation on yield risk measures may depend on the crop, the geographic area and even on the time period. The ratio of yield standard deviations between the farm and the county level ranged from 1.4 to 2.8 (Marra and Shurle, 1994). Consequently, extrapolation of yield risk from the national to the farm level, even if this is conducted paying close attention to the specificity of crops, may result in dangerous errors in the farmer’s assessment of production risk.

In the author's previous work (Kobus, 2010), it was shown that on NUTS 2 level, the variability of winter wheat yield is influenced by the following factors: land quality, natural logarithm of arable area, average yield and wheat production area, which explain 75% of standard deviations unevenness between voivodeships. Other researchers (Grønlund et al., 2006) found that 19% of wheat yield variability in Norway (farm level) can be explained by 15 significant predictors, the highest ranking predictors being irrigation, percentage of winter wheat, pH and the natural logarithm of farm area.

This paper deals with the modelling of yield risk measures in Poland on farm level of data aggregation. For that purpose, major crop plants yields in Poland have been analysed, i.e.: winter wheat, triticale, rye, barley, oat, mixed cereals, rape and sugar beet.

2. DATA AND METHODS

The main source of data was the Polish Farm Accountancy Data Network (FADN), which has one of the largest samples in the FADN (over 11 thousand farms).

The process of data selection was as follows: the samples for years 2004 – 2010 were screened for farms which were present in the samples throughout all the years, then next, from that pool a separate selection was carried out for each crop. The selection criterion was availability of yields for a specific crop plant for each year. The length of yield time series was regarded as priority. The sizes of samples for each plant are presented in Table 1.

Table 1. The sizes of samples for each plant researched

Crop plant	Sample size
Winter wheat	2300
Triticale	2379
Rye	1286
Barley	1596
Oat	416
Mixed cereals	1848
Rape (with turnip rape)	638
Sugar beet	741

Source: own calculations

In the analysis the following variables were used:

SD – yield standard deviation

VC – yield variation coefficient

AY – average yield

T - type of farming, with the following levels assigned according the FADN methodology: 1 - specialist field crops, 2 - specialist horticulture, 3 - specialist permanent crops, 4 - specialist grazing livestock, 5 - specialist granivore, 6 - mixed cropping, 7 - mixed livestock, 8 - mixed crops-livestock

X_1 – natural logarithm of crop plant production area

X_2 – land quality

X_3 – natural logarithm of arable area

X_4 – standard output, per hectare (of arable land)

X_5 – total inputs, per hectare

X_6 – total specific costs, per hectare

X_7,..., X_{10} – seeds cost in the first,...,fourth quarter, per hectare

X_{11},..., X_{14} – fertilisers cost in the first,...,fourth quarter, per hectare

X_{15},..., X_{18} – crop protection cost in the first,...,fourth quarter, per hectare

X_{19} – labour input, per hectare

X_{20} – value of machinery used in crop plants cultivation, per hectare

All the listed above variables were measured or calculated on the farm level.

2.1. Distribution of the measures considered

For all crop plants, two measures of yield variability were calculated, i.e., standard deviation (SD) and variation coefficient (VC), and additionally the average yield (AY).

Table 2. Basic characteristics of considered measures distributions

	Winter wheat			Triticale		
Characteristic	SD	VC	AY	SD	VC	AY
Average	8.43	0.17	49.90	7.97	0.20	42.16
Stand. deviation	3.40	0.08	10.00	3.30	0.09	9.11
Skewness	0.70	1.52	0.09	0.80	1.19	0.39
Kurtosis	4.12	9.42	2.99	4.28	6.47	3.13
Normal v. Gamma Z (Voung test)	-3.55	-5.55	2.25	0.29	-0.55	-3.68
	Rye			Barley		
Average	5.93	0.22	28.12	7.95	0.21	39.20
Stand. deviation	2.67	0.10	6.59	3.31	0.10	7.34
Skewness	1.09	0.75	0.70	0.78	1.36	0.05
Kurtosis	5.67	3.70	4.45	4.05	7.50	3.23
Normal v. Gamma Z (Voung test)	0.21	0.50	-3.82	-5.63	-6.00	2.30
	Oat			Mixed cereals		
Average	6.84	0.23	30.23	6.68	0.20	33.71
Stand. deviation	2.91	0.10	6.02	2.92	0.09	5.87
Skewness	0.77	0.55	0.48	0.77	0.89	0.21
Kurtosis	3.91	2.83	3.20	3.98	4.35	3.41
Normal v. Gamma Z (Voung test)	-3.00	-3.31	-2.95	0.40	0.05	0.41
	Rape (with turnip rape)			Sugar beet		
Average	6.18	0.20	31.75	86.16	0.18	491.29
Stand. deviation	2.23	0.09	5.35	33.46	0.08	60.80
Skewness	0.54	1.05	-0.45	0.56	1.43	0.08
Kurtosis	3.45	4.71	3.19	3.89	8.64	3.16
Normal v. Gamma Z (Voung test)	-1.67	-5.77	5.82	0.05	-2.39	0.58

Source: own calculations

The reason for including the average values in the analysis of risk measures lies in the possible application, for example when a farmer is analysing whether to buy crop insurance or not, he should be aware that the probability of yield reduction, for example, by 20% from the average of his yields, is not the same as the probability of a yield that is 20% lower than the reference yield used by insurance companies. To calculate the probability of qualifying for compensation, he needs not only standard deviation but also the average value of the yield, assuming that the distribution family of the crop yield is known.

Basically, there were two distribution families considered for the measures of interest, i.e., normal distribution and gamma distribution. Table 2 contains values of basic characteristics of the considered measures. Values of Voung closeness test Z statistics were used for deciding which distribution to use. Because the comparison was normal distribution versus gamma distribution, negative values point to gamma distribution while positive to normal distribution. The Z statistic follows asymptotically standard normal distribution, and consequently, the quantiles of standard normal distribution can be used for deciding if the first distribution is significantly closer to the true one than the second one. It is important to distinguish between being closer to the true distribution and *being* the true distribution. Having said that, in the farther analysis, the Voung test will be used for choosing better distribution for a specific measure. The gamma distribution will be preferred only in the case of being significantly better than the normal distribution.

It can be expected that the preferred distribution will not vary across the crops, but as it may be observed in table 2, this is not the case. For winter wheat, gamma is a better distribution for the standard deviation and variation coefficient, while for the average yield, it is better to use normal distribution. In the case of rye, it is just the opposite, consequently in each case separate analysis of distribution should be carried out.

2.2. Models

Two models were used for describing behaviour of the measures of interest. The first model was the classical linear model (LM), and the second was the generalized linear model (GLM). Both models shared the same systematic part:

$$E(Y_i) = \mu_i = \alpha_0 + \sum_{j=2}^{k} \alpha_j I_{\{j\}}(T_i) + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}\,, \tag{1}$$

where: Y - dependent variable, T- type of farming, α_0 - coefficient for reference level of farming type, that is specialist field crops, α_j - coefficient for *j*-th type of farming, I – indicator function equal 1 if T variable value is *j*, $\beta_1, \ldots, \beta_p$ coefficients for continuous variables $X_1, \ldots, X_p$.

The models LM and GLM differ in the random part, which for the linear model takes the form:

$$Y_i \sim N(\mu_i, \sigma^2)\,, \tag{2}$$

while for the chosen type of GLM model it is:

$$Y_i \sim Gamma(a, \theta_i)\,, \text{ with } \theta_i = \frac{\mu_i}{a}\,. \tag{3}$$

In order to select the valuable predictors from the set of all considered independent variables, a stepwise procedure has been applied, using the Akaike information criterion (AIC) as the selection criterion. For each step, this criterion was calculated for all models with one variable removed, and that variable, the removal of which offered the highest improvement of AIC was eliminated from the model. The process of variable elimination was stopped when elimination of any additional variable resulted in an increase of the AIC. Hence, there are variables in the model, which by using t-test would be considered insignificant, although usually there are only a few such variables.

When comparing the quality of fit for models from different families, there is a problem of measuring this quality. In the linear model, the standard choice is the determination coefficientR-squared, but this measure is not available for the generalized linear model. There are several quasi R-squared measures listed in the literature of the subject of GLM and they are usually based on the log likelihood function or deviance, but their values are not comparable across different models. The author proposes to use the following measure for all models:

$$R^2_{PRED} = \left(cor(Y,\hat{Y})\right)^2, \qquad (4)$$

Where $cor(Y,\hat{Y})$ is the Pearson correlation between observed values of dependent variables and predicted values from the estimated model. The R^2_{PRED} measure will be used for a direct comparison of the model's predictive quality.

Estimation of model parameters were conducted in R, an environment for statistical computing [R Development … 2011].

3. Results

The results of the models fit are presented in Tables 3, 4, 5 and 6, and in all cases only the final model estimates of parameters are given. Adash, in place of variable coefficient estimates implies that during the stepwise procedure of model selection, those variables had been removed from the model.

To save space, only estimates of model parameters were given without the traditional addition of standard errors and p-values. The "*" mark is used to indicate independent variable significance at the standard significance level, that is 5%. Since, the assumption of dependent variables distribution is not always true, this said significance of independent variables should be treated as a sign of the variables predictive usefulness rather than a formal rejection of the null hypothesis $H_0 : parameter = 0$.

In the considered models, there is the variable *T*, which is a factor variable. In the formulation of the systematic part of the linear and generalized linear models, the parameter α_0 was used as reference level, and it could also be treated like an intercept for farms from the first type of farming, that is, specialist field crops. To get the intercept for other types of farming, the value of related α_j should be added to the value of α_0, , for example the intercept for type 2 is the sum of α_0 and α_2. Hence, an asterisk next to the level *j* of variable T indicates if this type of farming has a different intercept than farms specializing in field crops.

In 5 out of the 8 cases presented in Table 3, there are no estimates for variable T levels except for level 1. This means that in those cases, the same intercept should be used for all farming types.

Table 3. Results of model fit, dependent variable Y=SD

	Estimator of model parameters							
Variable	Winter wheat	Triticale	Rye	Barley	Oat	Mixed cereals	Rape	Sugar beet
T:1 (intercept)	6.634 *	5.556 *	5.185 *	5.818 *	5.943 *	4.877 *	5.379 *	108.533 *
T:2	-0.717	-	-1.475 *	-	0.178	-	-	-
T:3	2.203	-	-2.240	-	16.472 *	-	-	-
T:4	-0.743 *	-	-0.865 *	-	-1.199 *	-	-	-
T:5	-0.410	-	-1.146 *	-	-1.659 *	-	-	-
T:6	-0.699	-	-0.574	-	-2.531 *	-	-	-
T:7	-0.555 *	-	-1.044 *	-	-0.393	-	-	-
T:8	-0.572 *	-	-0.418	-	-0.757	-	-	-
X_1	-0.309 *	-0.340 *	-	-	-	-0.481 *	-	-7.857 *
X_2	-	1.252 *	0.668 *	-	-	-	-	-6.862
X_3	0.807 *	0.721 *	0.476 *	0.622 *	0.561 *	0.944 *	0.311 *	-
X_4	-	-	-	-	-0.367 *	-	-	-
X_5	-	-	-0.143 *	-	-	-0.074	-	-
X_6	-	-	0.168 *	-	0.439 *	0.095	-	-
X_7	-1.321 *	-	-	-	3.211	-	-	-
X_8	-	-	-0.990	-2.121 *	-	-	-	49.257 *
X_9	0.892	-	-	6.923 *	-	2.332 *	-2.348	-
X_{10}	-	-	2.623	-	-	2.632 *	-	-
X_{11}	-	-0.801 *	-0.685	-0.853 *	-	-0.944 *	-	-8.166
X_{12}	-1.259 *	-1.292 *	-0.685	-0.961 *	-	-1.122 *	-	-
X_{13}	-	-	-	-1.022 *	-2.069 *	-	-0.636	-
X_{14}	1.038 *	0.797 *	0.579	-	1.462	-	-	-
X_{15}	-1.487 *	-2.129	-3.293 *	-1.561	-	-2.200 *	-	-38.598 *
X_{16}	-1.171 *	-1.034	2.002 *	-	-4.985 *	-	-	-
X_{17}	1.641	-	-	2.829 *	-	-	-	22.218
X_{18}	-	2.608	5.656 *	-	7.958	-	-	-
X_{19}	-	-	-	-	-	0.915 *	-	-40.896 *
X_{20}	0.085 *	-	-	0.077 *	-0.089 *	-	-	1.561 *
R^2_{PRED}	0.054	0.044	0.113	0.060	0.146	0.047	0.023	0.070
Model	GLM	LM	LM	GLM	GLM	LM	LM	LM

Source: own calculations
* denotes significance of a parameter at 5% significance level, verified with t – test

Looking at the determination coefficient values, it must be admitted that the chosen set of descriptive variables does not give models with strong predictive power. Nonetheless, in case of rye and oat, the values of the determination coefficient are similar to results obtained by other researchers who deal with the problem of modelling yield standard deviation on the farm level.

The most surprising result in Table 3 is the absence of any mitigating influence of land quality on yield variability, in all cases except sugar beet. It also worth noting that as in the author's previous works and in those of other researchers, the arable area size reduces the production risk.

Table 4. Results of model fit, dependent variable Y=VC

	Estimator of model parameters							
Variable	Winter wheat	Triticale	Rye	Barley	Oat	Mixed cereals	Rape	Sugar beet
T:1 (intercept)	0.263 *	0.272 *	0.265 *	0.263 *	0.240 *	0.186 *	0.302 *	0.262 *
T:2	-0.009	0.061 *	-0.012	0.000	-0.021	-	-	-
T:3	0.153 *	0.132 *	-0.024	0.118 *	0.551 *	-	-	-
T:4	-0.018 *	-0.022 *	-0.018	-0.021 *	-0.051 *	-	-	-
T:5	-0.017 *	-0.024 *	-0.036 *	-0.040 *	-0.081 *	-	-	-
T:6	-0.003	-0.010	0.005	0.014	-0.096 *	-	-	-
T:7	-0.016 *	-0.028 *	-0.031 *	-0.024 *	-0.027	-	-	-
T:8	-0.013 *	-0.023 *	-0.001	-0.025 *	-0.033	-	-	-
X_1	-	-0.011 *	0.008 *	-	-	-0.008	-	-0.020 *
X_2	-0.055 *	-0.025 *	-	-0.048 *	-	-0.032 *	-0.047 *	-0.031 *
X_3	-	0.007	-	0.013 *	0.016 *	0.023 *	-	-
X_4	-0.002	-0.008 *	-	-	-	-	-	-0.005
X_5	-	-	-0.006 *	-	-	-0.002 *	-0.001	-
X_6	-	0.001 *	0.006 *	-	0.007 *	-	-	-
X_7	-	-	-	-	-	-	-	-
X_8	-	-	-	-0.067 *	-	-	-	0.137 *
X_9	-0.032	-0.049 *	-0.098 *	0.086	-	0.075 *	-0.127 *	-
X_{10}	-	-	0.163 *	-	-	0.088 *	-	-
X_{11}	-0.016 *	-0.031 *	-0.043 *	-0.025 *	-	-0.026 *	-	-0.018
X_{12}	-0.042 *	-0.043 *	-0.024	-0.039 *	-	-0.028 *	-	-
X_{13}	-0.024 *	-0.032 *	-0.054 *	-0.042 *	-0.070 *	-0.025	-0.027 *	-
X_{14}	-	-	-	-0.016	0.045	-	-0.016	-
X_{15}	-0.054 *	-0.106 *	-0.177 *	-	-	-0.072 *	-0.078 *	-0.045 *
X_{16}	-0.043 *	-0.103 *	-	-0.066 *	-0.246 *	-0.043	-0.053 *	-
X_{17}	-	-	-0.084	-	-	-	-0.046	-
X_{18}	-	-	-	-	-	-	-	-
X_{19}	0.017 *	-	-	-	-	0.049 *	-	-0.070 *
X_{20}	0.002 *	-	-	-	-0.004 *	-	-	-
R^2_{PRED}	0.091	0.085	0.085	0.087	0.117	0.058	0.094	0.069
Model	GLM	LM	LM	GLM	GLM	LM	GLM	GLM

Source: own calculations

* denotes significance of a parameter at 5% significance level, verified with t – test

In case of the variation coefficient, the majority of the models performed better than in the case of standard deviation with determination coefficient close to 10%. This time, the land quality showed a reducing effect on the risk production measure. But most of this improvement is caused by the better predictive powers of available variables in case of average yields, not the variability itself. The variation coefficient is a function of standard deviation and of the average - the higher the average, the lower the variation coefficient.

Table 5. Results of model fit, dependent variable Y=AY

	Estimator of model parameters							
Variable	Winter wheat	Triticale	Rye	Barley	Oat	Mixed cereals	Rape	Sugar beet
T:1 (intercept)	25.542 *	21.186 *	16.879 *	20.935 *	19.300 *	24.668 *	18.998 *	367.453 *
T:2	-4.639 *	-8.353 *	-8.401 *	-1.297	-2.115	-4.271 *	-9.670 *	27.555
T:3	-15.176 *	-27.393 *	-4.654	-12.813 *	-5.693	-6.607 *	-13.058 *	32.773
T:4	1.500 *	0.913	-1.207	0.096	1.446	1.529	-0.267	-9.927
T:5	2.546 *	3.271 *	-0.468	3.596 *	4.495 *	2.578 *	3.163 *	25.032 *
T:6	-1.936	-0.542	-2.244	-1.329	1.279	-1.287	-3.961 *	-3.365
T:7	1.296 *	1.975 *	-0.307	1.724 *	2.616 *	1.619 *	2.226 *	11.068
T:8	0.775	0.954	-0.710	0.898	0.758	0.164	1.070 *	9.460
X_1	-	-	-0.919 *	-1.372 *	-0.617	-1.422 *	0.613	9.434 *
X_2	12.434 *	11.342 *	4.216 *	8.677 *	4.652 *	7.095 *	5.009 *	46.633 *
X_3	1.527 *	1.251 *	2.138 *	1.742 *	1.218 *	1.160 *	-	-
X_4	0.977 *	1.140 *	0.852 *	-	-	0.251	-	-
X_5	-	0.334 *	-	0.286 *	-	0.177 *	0.800 *	1.791 *
X_6	-0.110 *	-0.589 *	-	-	-	-	-0.899 *	-
X_7	-	-1.998	-1.899	-	-	-	5.438	-
X_8	-3.990 *	-5.936 *	-	-	-	-2.851 *	-	-
X_9	14.851 *	13.095 *	18.054 *	11.921 *	-	3.653	8.634 *	44.867
X_{10}	-0.893	-	-	9.733 *	-4.275	-3.696 *	-	-
X_{11}	2.444 *	1.805 *	1.173	-	3.641 *	-	-	-
X_{12}	3.810 *	-	-1.799	-	-	-	-	-
X_{13}	6.349 *	4.799 *	5.977 *	4.779 *	7.568 *	2.059 *	2.089 *	42.843 *
X_{14}	5.244 *	2.437 *	-	2.702 *	2.820 *	1.506 *	2.788 *	23.505 *
X_{15}	3.613 *	14.711 *	15.659 *	-	-	5.264 *	7.075 *	-45.170
X_{16}	7.633 *	18.520 *	13.545 *	9.452 *	10.551 *	5.338 *	6.019 *	49.323 *
X_{17}	11.757 *	11.382 *	7.020 *	3.565	-	-	5.848 *	40.272
X_{18}	-	23.446 *	17.820 *	-	-	7.989	4.962 *	-
X_{19}	-10.083 *	-3.781 *	-	-1.020	-	-4.167 *	-7.094 *	-
X_{20}	0.155 *	0.187 *	-	-	0.173	0.111 *	-	4.668 *
R^2_{PRED}	0.487	0.496	0.342	0.394	0.296	0.299	0.279	0.237
Model	GLM	LM	LM	GLM	GLM	LM	GLM	GLM

Source: own calculations

* denotes significance of a parameter at 5% significance level, verified with t – test

Table 5 contains models for the average yields. The determination coefficients are between 24% for sugar beet and almost 50% for triticale and winter wheat, and this is 5 times higher than the determination coefficient for the winter wheat model in case of the variation coefficient and 9 times higher than the determination coefficient for the standard deviation model.

Table 6. Results of model fit, dependent variable Y=VC, with additional X_{21}=AY

	Estimator of model parameters							
Variable	Winter wheat	Triticale	Rye	Barley	Oat	Mixed cereals	Rape	Sugar beet
T:1 (intercept)	0.058 *	0.059 *	0.058 *	-0.081 *	0.026	-0.013	-0.076 *	0.112 *
T:2	-0.009	0.023	-0.037	-	0.020	-	-	-
T:3	0.050	0.056	-0.026	-	0.490 *	-	-	-
T:4	-0.012 *	-0.018 *	-0.021	-	-0.038 *	-	-	-
T:5	-0.008	-0.013	-0.032 *	-	-0.057 *	-	-	-
T:6	-0.009	-0.008	-0.007	-	-0.084 *	-	-	-
T:7	-0.010	-0.021 *	-0.029 *	-	-0.012	-	-	-
T:8	-0.010 *	-0.019 *	-0.003	-	-0.028	-	-	-
X_1	-	-0.009 *	-	-	-	-0.013 *	-	-0.019 *
X_2	-0.021 *	0.014 *	0.014	0.000	-	-	-	-0.021 *
X_3	0.005	0.012 *	0.013 *	0.017 *	0.017 *	0.027 *	0.009 *	-
X_4	-	-0.005 *	-	-0.002	-0.015 *	-	-	-
X_5	-	-	-0.005 *	-	-	-0.003	-	-
X_6	-	0.001 *	0.006 *	0.002	0.014 *	0.003	-	-
X_7	-	-	-	-	0.099	-	-	-
X_8	-	-	-	-	-	-	-	0.110 *
X_9	-	-	-	0.176 *	-	0.057 *	-	-
X_{10}	-	-	0.102	-	-	0.086 *	-	-
X_{11}	-	-0.021 *	-0.030 *	-0.024 *	-	-0.027 *	-	-0.015
X_{12}	-0.025 *	-0.035 *	-0.026	-0.027 *	-	-0.034 *	-	-
X_{13}	-	-	-	-0.024	-0.071 *	-	-0.020	-
X_{14}	0.016 *	0.017 *	0.019	-	0.043	-	-	-
X_{15}	-	-0.071 *	-0.128 *	-	-	-0.061 *	-	-0.052 *
X_{16}	-0.021 *	-0.062 *	-	-	-0.178 *	-	-	-
X_{17}	-	-	-	0.052	-	-	-	0.042
X_{18}	-	-	-	-	0.215	-	-	0.048
X_{19}	-0.023 *	-	-	-	-	0.032 *	-	-0.072 *
X_{20}	0.001 *	-	-	0.001	-0.003 *	-	-	0.003 *
X_{21}	6.492 *	5.489 *	4.068 *	8.784 *	5.445 *	5.394 *	7.700 *	56.641 *
R^2_{PRED}	0.192	0.155	0.18	0.255	0.237	0.145	0.309	0.132
Model	GLM	LM	LM	GLM	GLM	LM	GLM	GLM

Source: own calculations
* denotes significance of a parameter at 5% significance level, verified with t – test

The land quality coefficient estimators (X_2) are significant for all crop plants, and of course they are positive - the better land quality, the higher the average yields. In case of standard deviation models, the estimator for land quality parameters were not significant and for triticale and rye, they did even imply that the better land quality, the higher the yield variability.

If the intended application of variability analysis is to assess the production risk for a crop plant which has not been previously cultivated on the farm, the proposed set of independent variables does not offer a satisfying predictive quality. It is possible that expanding it by including variables describing weather conditions characteristic for the area of interest may improve the predictive quality of the models, but at the moment, the accuracy of available weather data was not satisfactory.

On the other hand, if the purpose of the analysis is purely descriptive, and the average yields of the crop plants are available, they can be used as independent variables to improve the model fit. Because the relation between average yield and variation coefficient is negative, the additional variable used in models presented in Table 6 (X_{21}) is equal to$(AY)^{-1}$. Adding X_{21} to the variation coefficient models allowed for an increase in the average determination coefficient from 8.6% to 20%. In case of rape, the determination coefficient was almost 31%.

4. CONCLUDING REMARKS

While the research suggests that it is possible to construct models for predicting yields variability, it also shows that the predictive quality of such models is unsatisfactory. In case of standard deviation modelling, the average value of the determination coefficient was only 7% and for variation coefficient models, this was 8.6%.

The predictive quality of average yields modelling was much better when the determination coefficient's mean value exceeded 35%.

The addition of the average yield inverse to the models used for variation coefficient allowed for substantial improvement of the determination coefficients, on average up to the level of 20%. However, such models cannot be used for production risk predictions in case of those crop plants which weren't previously cultivated on the farm.

Summarizing, although the investigated models show potential, they need further improvement before they are used for production risk predictions, unless average yields of evaluated crop plants are available.

REFERENCES

[Article in a journal] Monier-Dilhan, S. and Ossard, H. (1998). Producers' loss due to asymmetric information: An application to a specific case. *European Review of Agricultural Economics* 25: 155-169.

Grønlund A., Njøs A., Vestgarden L.S., Lyngstad I. (2006). Utilisation of agricultural databases for statistical evaluation of yields of barley and wheat in relation to soil variables and management practices. *Acta Agriculturæ Scandinavica* B 56: pp. 1-8.

Kobus P. (2010). Modelling wheat yields variability in Polish voivodeships. *Problems of World Agriculture* 10, No. 3: pp. 33-40.

Marra M. C., Schurle B. W. (1994). Kansas wheat yield risk measures and aggregation: a meta-analysis approach. *Journal of Agricultural and Resource Economics* 19(1): pp. 69-77.

R Development Core Team (2011). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, [available at] URL http://www.R-project.org.