Paweł Kobus[1]
Chair of Agricultural Economics and International Economic Relations
Warsaw University of Life Sciences
Warsaw, Poland

Wheat yields variability in Poland at NUTS 2 level in context of production risk

Abstract. The paper presents an analysis of the wheat yields variability in voivodeships of Poland. The main aim of the study is to present several possible indicators for the crop variability in the context of production risk. It is found that ignoring the long-term yield trends leads to a serious overestimation of production risk.

Key words: wheat yield, variability, production risk.

Introduction

One of the most specific to agriculture risks is the production risk. It is influenced by factors like weather conditions, pests and diseases. Changes of yields along with changes of prices are the main factors of farmers' income variability. The most typical way to stabilize farmers' income is to buy a crop insurance, but according to Bielza et al. [2008] only about 3% of farmers in Poland buy a crop insurance. The rationality of farmers' decisions depends on yields variability. In case of small yield variability and in consequence small probability of exceptional losses a crop insurance is an unreasonable choice, especially when the insurance premium is high. But in the case of high variability of yields the choice of crop insurance as a method of stabilizing income seems very reasonable.

The variability of yields is a key factor for deciding whether to buy a crop insurance. The aim of this paper is to show how to make use of different measures of variability for making such decisions. The data used in this study are aggregated for voivodeships and because of that are used only for illustration purpose.

Basic measures

The statistical data used in this analysis concern the average yields of wheat in Polish voivodeships in years 1995 – 2007 and are available from Eurostat [Eurostat 2009]. The yields troughout the whole paper are expressed in decitons (dt).

One of the simplest ways of measuring crop variability is to use the standard deviation estimator (1):

$$S_x = \sqrt{\tfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{1}$$

where:

[1] DrSc, email: pawel_kobus@sggw.pl

S_x – estimator of the standard deviation σ_x of variable X,

$\bar{x}$ - sample mean,

n – sample size,

x_i – ith observation of variable X.

While it is a very simple measure of crop variability, it does not display expressly the average level of yield. A standard deviation equal to 10 dt with an average yield equal to 30 dt does not carry the same meaning as with an average yield equal to 70 dt.

To overcome this problem the coefficient of variation (2) is usually applied:

$$V_x = \frac{S_x}{\bar{x}} \tag{2}$$

The coefficient of variation, contrary to the standard deviation, is expressed in percents of the average yield. The mentioned above S_x=10 dt for an average yield of 30 dt gives $V_x = 33\%$, while for an average yield of 70 dt gives $V_x = 14\%$.

Values of standard deviations and coefficients of variation of wheat yields in Poland's voivodeships are presented in Table 1. For the sake of comparison the average yields are also given, though they are not measures of variability.

Table 1. Estimates of basic measures of wheat yield variability in Poland's voivodeships (1995-2007).

Voivodeship	Average yield, dt	Standard deviation, dt	Coefficient of variation, %
Łódzkie	30.7	3.4	11.1
Mazowieckie	31.2	2.6	8.3
Małopolskie	30.1	2.7	9.1
Śląskie	34.8	4.3	12.4
Lubelskie	31.4	2.5	8.1
Podkarpackie	30.0	2.1	7.0
Świętokrzyskie	28.6	3.1	10.8
Podlaskie	27.4	3.2	11.6
Wielkopolskie	40.4	4.5	11.3
Zachodniopomorskie	37.8	4.4	11.7
Lubuskie	32.9	5.2	15.7
Dolnośląskie	40.6	4.6	11.3
Opolskie	45.9	5.5	12.0
Kujawsko-Pomorskie	38.4	3.8	9.8
Warmińsko-Mazurskie	36.8	3.3	8.9
Pomorskie	41.6	3.8	9.2

Source: own calculations

The highest standard deviation (5.5 dt) is observed for Opolskie voivodeship but the highest average yield is also observed in the same voivodeship. As a result the coefficient of variation (12%) is lower then for the Lubuskie voivodeship (15.7%), where the standard deviation is equal to 5.2 dt. The lowest standard deviation and coefficient of variation are observed in Podkarpackie voivodeship.

One of conclusions from the analysis of Table 1 is that voivodeships in Western Poland show the highest yield variability measured by coefficient of variation, with exception of Pomorskie voivodeship and, on the other hand, Podlaskie. Another conclusion is not so obvious, there is a strong relationship between the average yield and the standard deviation. Figure 1 illustrates that relation.

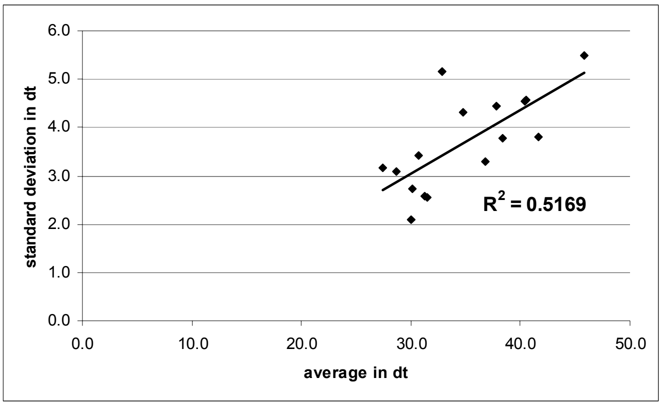

Figure 1. Relationship between the average yield and the standard deviation of yield (1995-2007).
Source: own calculations.

The corresponding determination coefficient is above 51%. It means that the value of standard deviation is explained to a great extent by the average yield alone. It is not an aim of this paper to explain this relation, but it is probable that the intensive production is more vulnerable than the extensive one and that this results in a higher variability of yields.

Proposed measures

The standard deviation or the coefficient of variation are acceptable measures of variability but as production risk measures they are insufficient.

The desirable measures should give an answer to the following questions:

1. What is the probability of yield reduction greater then γ %?
2. What is the expected yield reduction if reduction is greater then γ %?
3. What is the probability of yield reduction greater then δ dt?
4. What is the expected yield reduction if reduction is greater then δ dt?

The values of γ and δ depend on each farmer's aversion to risk or on crop insurance contract conditions. For example in crop insurance contracts proposed by the PZU the value of γ is equal to 10% [Ogólne … 2007].

An answer to the question number 1 comes down to calculation of the following probability[2]:

$$P_\gamma = P(X \le (1-\gamma)EX) \tag{3}$$

[2] EX is expected value of variable X. In the case of a continuous variable EX is defined as $\int_{-\infty}^{\infty} xf(x)dx$, where *f(x)* is the density function of variable *X*.

and an answer to the third question to:

$$P_\delta = P(X \le EX - \delta)\,. \tag{4}$$

To answer the second and fourth questions we must calculate conditional expected values:

for the second question[3]

$$EX_y = E\big(X \mid X <= (1-\gamma)EX\big) - EX = \tfrac{1}{P_\gamma}\int_0^{(1-\gamma)EX} xf(x)dx - EX \tag{5}$$

and for the fourth question

$$EX_\delta = E\big(X \mid X <= EX - \delta\big) - EX = \tfrac{1}{P_\delta}\int_0^{EX-\delta} xf(x)dx - EX \tag{6}$$

The answers to the questions third and fourth have a distant similarity with the value at risk and the conditional value at risk, but VaR gives an answer to the question about what is the value of threshold with a given probability, while formula (4) gives an answer to the question about what the probability with a given threshold is.

In order to calculate the exact values with equations (3) to (6) one would have to know the distribution of variable X which represents yield. In this paper, for the sake of simplicity, it is assumed that the variable X follows the normal distribution although with different parameters in each voivodeship. The normal distribution is fully defined by two parameters: expected value EX usually denoted by μ and variation D^2X denoted by σ^2.

Table 2. Estimates of proposed measures of wheat yield variability in Poland's voivodeships (1995-2007).

Voivodeship	$\breve{P}_{\gamma=0.1}$, %	$\bar{x}_{\gamma=0.1}$, dt	$\breve{P}_{\delta=5dt}$, %	$\bar{x}_{\delta=5dt}$, dt
Łódzkie	18.5	-4.9	7.2	-6.6
Mazowieckie	11.3	-4.6	2.6	-6.1
Małopolskie	13.6	-4.4	3.4	-6.3
Śląskie	21.0	-6.3	12.4	-7.3
Lubelskie	10.8	-4.5	2.5	-6.0
Podkarpackie	7.6	-4.1	0.9	-5.7
Świętokrzyskie	17.6	-4.7	5.2	-6.5
Podlaskie	19.4	-4.6	5.8	-6.4
Wielkopolskie	18.7	-6.6	13.6	-7.3
Zachodniopomorskie	19.7	-6.4	13.0	-7.5
Lubuskie	26.2	-6.7	16.7	-7.9
Dolnośląskie	18.8	-6.6	13.7	-7.4
Opolskie	20.2	-7.9	18.1	-8.4
Kujawsko-Pomorskie	15.4	-6.1	9.3	-7.0
Warmińsko-Mazurskie	13.2	-5.6	6.4	-6.7
Pomorskie	13.7	-6.1	9.5	-7.1

Source: own calculations.

[3] Notation *E(X|X<a)* is for conditional expected value with condition *X<a*. In the case of nonnegative variables $E(X \mid X < a) = \frac{1}{P(X<a)}\int_0^a xf(x)dx$, where *f(x)* is the probability distribution function for variable *X*.

The true values of μ and σ^2 are unknown and their estimators must be used instead. As a result it is not possible to calculate exact values of expressions given by equations (3) to (6). But it is possible, with application of estimators $\bar{x}$ and S_x, to calculate the maximum likelihood estimators (MLE) of those values.

For instance formula (3) would be transformed to:

$$\tilde{P}_\gamma = F\left(\frac{(1-\gamma)\bar{x}-\bar{x}}{S_x}\right) = F\left(\frac{-\gamma\bar{x}}{S_x}\right) \tag{7}$$

where:

$\tilde{P}_\gamma$ – MLE of P_γ,

F(z) – value of cumulative distribution function of variable *Z* in the case of standard normal distribution.

Estimates of expressions given by formulas (3) to (6) for all voivodeships of Poland are presented in Table 2.

The values of γ and δ in the Table 2 were chosen arbitrarily for the purpose of method illustration, although $\gamma = 0.1$ (10%) agrees with that proposed by the PZU crop insurance contracts [Ogólne … 2009].

The order of voivodeships with respect to the wheat production risk given by $\tilde{P}_\gamma$ agrees with the order given by the coefficient of variation while the order given by $\tilde{P}_\delta$ with the order given by the standard deviation. Hence, if all that is needed is to put regions in ascending or descending order according to production risk there is no difference between the methods and use of simpler measures is suggested. But if measures of variability should be used for decision by an individual farmer whether to buy an insurance contract or how big reserve should he keep to be insured in case of a significant yield reduction the proposed measures seem to be much more informative and, what is even more important, they have very simple interpretations. For example $\tilde{P}_\gamma$ tells what is the chance of a significant crop reduction, with an option to specify by the farmer what value of γ is for him significant. $\bar{x}_\gamma$ tells what would be a typical yield reduction if a significant reduction occurs.

Meaning of yield trend in estimation of yield variability

All the measures of yield variability and the production risk discussed in the previous section assume that each observation of yield is independent. This is not necessarily true. A typical situation is rather that an ascending trend could be observed. This trend can be explained by the biological progress and the technological advancement. In such cases measures calculated with ignoring the existing trend will be biased upward and they will suggest a higher risk than it is in reality. The amount of the bias is proportional to the strength of the trend. If a trend is week it can be treated as negligible but in case of a clear trend it should be taken into account. There are two issues to be addressed: the standard deviation and the expected value.

Let us define the trend as a function of time which explains the conditional expected value:

$$E(X \mid T=t)=g(t) \tag{8}$$

where:

T - time variable,

t – time moment (year),

$E(X \mid T=t)$ – expected yield in year t.

The function $g(t)$ could take any form but in a short time series it is usually safe to use the simplest linear form:

$$g(t)=\beta_0+\beta_1 t \tag{9}$$

The formula for calculating the standard deviation estimator changes to:

$$S_x=\sqrt{\frac{1}{n-2}\sum_{t=1}^{n}(x_t-\hat{x}_{(t)})^2} \tag{10}$$

where:

x_t – observation of yield in year t,

$\hat{x}_{(t)}$ – estimate of $g(t)$.

The issue of expected value is more complicated. If measures which use the expected value or its estimate are thought of as indicators of production risk in the next year, i.e. in year $n+1$, the estimated trend function must be used for calculation of the $E(X \mid T=n+1)$ estimator instead of a simple average:

$$\hat{x}_{(n+1)}=\hat{\beta}_0+\hat{\beta}_1(n+1) \tag{11}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are least squares estimators of β_0 and β_1.

Table 3. Estimates of basic measures of wheat yield variability in Poland's voivodeships based on ascending trend assumption (1995-2007).

Voivodeship	Average yield, dt	Standard deviation, dt	Coefficient of variation, %
Łódzkie	32.6	3.4	10.4
Mazowieckie	32.6	2.6	7.9
Małopolskie	31.6	2.7	8.7
Śląskie	38.1	4.1	10.7
Lubelskie	32.7	2.5	7.8
Podkarpackie	31.4	2.0	6.5
Świętokrzyskie	28.8	3.2	11.2
Podlaskie	28.4	3.3	11.5
Wielkopolskie	40.7	4.7	11.7
Zachodniopomorskie	40.8	4.3	10.5
Lubuskie	32.7	5.4	16.5
Dolnośląskie	43.2	4.5	10.5
Opolskie	51.6	4.7	9.0
Kujawsko-Pomorskie	40.9	3.7	9.0
Warmińsko-Mazurskie	41.2	2.3	5.6
Pomorskie	44.6	3.6	8.1

Source: own calculations.

Standard deviations and coefficients of variation when assuming a trend are presented in Table 3.

The most noteworthy changes after trend introduction occurred in the Opolskie and Warmińsko-Mazurskie voivodeships, with a coefficient of variation reduction of about 3%.

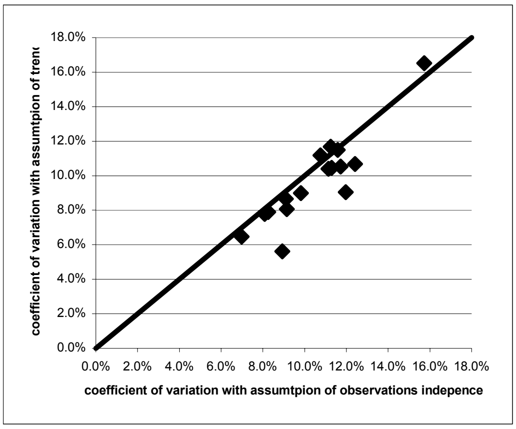

Figure 2. Relation between coefficients of variation in two compared cases of an estimation (1995-2007).

Source: own calculations.

The relation between coefficients of variation in the two above mentioned cases is presented in Figure 2. The diagonal black line is a demarcation line below which are situated the cases where the coefficient assuming existence of a trend is lower than the coefficient assuming observations independence. There are three points which are above that line. It shows that in case of a very weak trend or no trend at all it is better to use formulas which assume independence of observations.

Table 4. Estimates of proposed measures of wheat yield variability in Poland's voivodeships based on ascending trend assumption (1995-2007).

Voivodeship	$\widetilde{P}_{\gamma=0.1}$, %	$\overline{x}_{\gamma=0.1}$, dt	$\widetilde{P}_{\delta=5dt}$, %	$\overline{x}_{\delta=5dt}$, dt
Łódzkie	16.8	-5.2	7.0	-6.6
Mazowieckie	10.3	-4.6	2.6	-6.1
Małopolskie	12.4	-4.6	3.4	-6.3
Śląskie	17.5	-6.2	11.0	-7.2
Lubelskie	9.9	-4.5	2.5	-6.0
Podkarpackie	6.1	-4.1	0.7	-5.7
Świętokrzyskie	18.5	-4.6	6.0	-6.5
Podlaskie	19.2	-4.7	6.3	-6.6
Wielkopolskie	19.6	-6.9	14.6	-7.6
Zachodniopomorskie	17.1	-6.6	12.2	-7.3
Lubuskie	27.2	-7.0	17.7	-8.2
Dolnośląskie	16.9	-6.9	13.4	-7.6
Opolskie	13.4	-7.9	14.2	-7.5
Kujawsko-Pomorskie	13.3	-6.2	8.7	-6.8
Warmińsko-Mazurskie	3.8	-5.1	1.6	-5.9
Pomorskie	10.8	-6.4	8.2	-6.7

Source: own calculations.

The estimates of expressions given by formulas (3) to (6) for all voivodeships in Poland, assuming existence of a trend, are presented for comparison in Table 4.

There are very small changes in conditional expected reductions of yield but the probability of a significant reduction of yield has changed noticeably, with most spectacular case of the Warmińsko-Mazurskie voivodeship, where previously the $\widetilde{P}_{\gamma}$ was about 13% and now it is about 4%. This example shows that a disregard for the long-term trend may lead to a serious overestimation of production risk.

Conclusions

The basic measures of variability are not sufficient for a production risk estimation. But they are sufficient for putting regions in order with accordance to the yield variability.

The proposed measures of variability are easier for interpretation as they answer the natural questions which arise when assessing the production risk. One of such questions is: what is a chance of a significant reduction of yield occurring.

In order to prevent an overestimation of production risk the long-term yield trends should be taken into account.

The voivodeships of Poland are very diverse in aspect of production risk. The probability of a yield reduction by 10% of the expected value ranges from about 4% in the Warmińsko-Mazurskie to about 27% in the Lubuskie voivodeship.

References

Bielza M., Conte C., Dittmann C., Gallego J., Stroblmair J. [2008]: Agricultural Insurance Schemes. Economic analysis and evaluation for European Commission, modified version. [Available at:] http://ec.europa.eu/agriculture/analysis/external/schemes2/index_en.htm. [Accessed June 2009].

Eurostat, Statistical Office of the European Communities. [Available at:] http://ec.europa.eu/eurostat. [Accessed June 2009].

Ogólne warunki ubezpieczenia upraw PZU S.A. [Available at:] http://www.pzu.pl/pub/files/PZU_SA/produkty/rolne/owu_5303_IV_B.pdf. [Accessed June 2009].