Influence of aggregation level on yield risk measures

Paweł Kobus
e-mail: pawel_kobus@sggw.pl

Paper prepared for presentation at the EAAE 2011 Congress
Change and Uncertainty
Challenges for Agriculture,
Food and Natural Resources

August 30 to September 2, 2011
ETH Zurich, Zurich, Switzerland

Abstract

The paper is dedicated to the relationship of data aggregation level and yield variability. For that purpose yields of the major crop plants in Poland are analysed i.e.: winter wheat, triticale, rye, barley, oat, mixed cereals, rape and sugar beet. The research are based on data from Polish FADN from years 2004 – 2009. The samples' size ranged from 531 to 2893, depending on the plant crop. In the paper six levels of data aggregation are examined, that is: farm, district, powiat[1], voivodship, region and country. It was found out that the degree of yield variability reduction (observed with data aggregation) is crop specific. Nevertheless, the relationship between aggregation level and yield variability can be approximated by the same formula for all of the investigated crop plants: $\frac{1}{S^2} = \beta_0 + \beta_1 \log(MUA)$, where MUA is the average production area in the administrative unite.

Key words: yield risk, aggregation level.

Introduction

Yield risk is one of the risks inherent in the agriculture sector. Factors influencing yield level can be divided into those dependent and those independent of the farmer's decisions. The most important factors from the second group are weather patterns and soil conditions. The factors from the first group can be defined as chosen and implemented production technology. An interaction between the above mentioned factors cause spatial yield variability even when some of the factors affecting yields of crop plants remain on the same level. As a consequence, yields, even on closely situated farms, are not correlated perfectly [Górski and Górska 2003]. It is intuitively obvious, therefore, that a higher level of aggregation results in smaller yield variability. The question remains how much each level of aggregation decreases yield variability observed on the farm level. 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 yields' standard deviations between the farm and the county level ranged from 1.4 to 2.8 [Marra and Shurle 1994]. Authors of the research concerning wheat yield variability in Northwest Mexico [Lobell et al 2007] stated that the field scale yield risk, measured by variance, was 58% higher than the regional scale risk, which is equivalent to the standard deviations' ratio 1.26.

To sum up, extrapolation from the national or even regional yield risk measures to the farm level should be avoided because it is no clear how much the farm level yield risk of a specific crop is higher than the national one. On the other hand, famers usually don't have time series of yields necessary for the estimation of yield risk. Consequently, without research specific to the given crop and the given geographical area it is hard to quantify production risk on the farm level.

The main aim of this paper is to assess the relation between the level of aggregation and yield risk for the major crop plants in Poland.

[1] Powiat is the second level of local government administration in Poland.

Data and applied methods

The main source of data was Polish Farm Accountancy Data Network (FADN), which has the second largest sample in the FADN (over 11 thousand farms). Another source of data was Central Statistical Office of Poland [GUS 2010].

The process of data selection was as follows: the samples for years 2004 – 2009 were screened for farms which were present in the samples for all years, then from that pool a separate selection was carried out for each crop. The criterion for selection was availability of yields for specific crop plant for each year. 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	2748
Triticale	2893
Rye	1627
Barley	2033
Oat	531
Mixed cereals	2318
Rape (with turnip rape)	741
Sugar beet	874

In every sample the same set of variables was observed, that is crop production area in ha and yield in dt/ha.

In this paper six levels of data aggregation are examined, that is: farm, district, powiat[2], voivodship, region and country. The average total area (in thous. ha) of the mentioned data aggregation levels are respectively (starting from district): 14.8, 99.6, 1954.2, 7817.0 and 31267.9. Arable land constituted almost 39% of the total area in year 2008 (GUS 2010). In the same year the average used arable land of the farm in Polish FADN sample was 12.76 ha (FADN 2009), which is much higher than the average arable land of the farm in Poland reported by Central Statistical Office. The reason for this is the fact that the Polish FADN sample is constructed to be representative of farms of economic size of at least 2 ESU.

For all territorial units, on each level of aggregation, yields' standard deviations were calculated according to the following formula:

$$S_{a_l} = \sqrt{\frac{\sum_{i=1}^{6}(Y_{a_l i} - \overline{Y_{a_l}})^2}{6-1}} \quad (1)$$

where $Y_{a_l i}$ is a weighted average yield from all farms in the administrative unit a_l, the subscript l denotes a level of aggregation from the level 0 (no aggregation) to the level 5 with data aggregated for the whole country.

[2] Powiat is the second level of local government administration in Poland.

The yields used for calculating the standard deviations were not detrended deliberately. The reason for that were the relatively short time series, which involved a risk of serious overfitting, especially on a low level of aggregation. As a consequence, it could conceal the relation between the level of aggregation and yield risk.
To isolate the above mentioned relation from all factors which influence yield variability, except the level of aggregation, weighted averages of standard deviations were calculated. The areas of administrative units were used as weights.

Results

For all aggregation levels and each of the considered plant crops the following were calculated: the mean area of crop production and the mean standard deviations.

Table 2. Mean standard deviations and variability reduction degree

	Wheat			Triticale		
Aggregation level	Mean unit area, ha	Mean SD	Reduction degree	Mean unit area, ha	Mean SD	Reduction degree
farm	10,4	8,9	0%	6,7	8,2	0%
district	27,7	7,4	17%	18,5	6,6	19%
powiat	101,6	6,3	29%	71,2	5,6	31%
voivodship	1789,8	5,4	40%	1210,2	4,9	40%
region	7159,4	4,9	45%	4841,0	4,6	43%
country	28637,6	4,8	47%	19363,8	4,5	45%
	Rye			Barley		
farm	5,8	6,6	0%	6,7	8,5	0%
district	12,4	5,7	14%	14,8	7,4	13%
powiat	39,7	4,8	27%	49,3	6,5	24%
voivodship	591,1	3,8	42%	847,1	5,4	36%
region	2364,4	3,5	47%	3388,5	4,7	44%
country	9457,8	3,4	48%	13554,1	4,3	50%
	Oat			Mixed cereals		
farm	5,2	7,7	0%	5,8	6,8	0%
district	7,0	7,2	6%	14,1	5,9	14%
powiat	13,3	6,6	14%	51,7	5,1	26%
voivodship	172,6	5,3	31%	837,2	4,4	36%
region	690,5	4,7	38%	3348,9	4,1	40%
country	2762,0	4,4	42%	13395,5	3,8	45%
	Rape (with turnip rape)			Sugar beet		
farm	15,5	6,1	0%	5,2	83,4	0%
district	28,6	5,2	15%	12,2	70,3	16%
powiat	69,9	4,2	32%	35,2	59,6	29%
voivodship	716,0	2,7	55%	300,6	50,8	39%
region	2864,0	2,4	61%	1127,3	46,8	44%
country	11455,8	2,3	63%	4509,3	44,7	46%

To better assess similarities between the various crop plants the degree of reduction of standard deviation values from the farm level was also calculated, according to the following formula:

$$RD_l = \left(1 - \frac{S_l}{S_0}\right) \cdot 100\% \quad (2)$$

where RD_l is the reduction degree on the level l, S_l is the mean standard deviation on that level and S_0 is the mean standard deviation on the farm level of data aggregation.

When looking at the reduction on the country level it may be observed that for all crop plants, except rape, there is similar reduction, i.e.: between 42% and 50%. It means that yield variability measured by standard deviations is almost twice higher at the farm level than at the country level; in case of rape it is three times higher.
There is a interesting coincidence of reduction on the voivodship level. For the winter cereals i.e.: wheat, triticale and rye it is about 40%, while for the spring cereals it is only 30% to 36%. Besides crop effect, the differences in reduction observed on the district level are a result of the sample size. In the case of oat the sample size was only 531 while the number of districts in that sample was 396, and, consequently, in most district there was only one farm. On the higher level of aggregation the results were less prone to the differences in the sample size.

Table 3. Variation coefficients and average yields

aggregation level	Winter wheat	Tritical e	Rye	Barley	Oat	Mixed cereals	Rape	Sugar beet
farm	16,6%	18,8%	22,6%	21,8%	26,1%	20,3%	18,8%	16,8%
district	13,8%	15,2%	19,5%	18,9%	24,4%	17,5%	16,0%	14,1%
powiat	11,7%	13,0%	16,5%	16,6%	22,3%	15,1%	12,8%	12,0%
voivodeship	10,0%	11,2%	13,2%	13,9%	18,0%	13,0%	8,4%	10,2%
region	9,1%	10,7%	12,0%	12,1%	16,1%	12,2%	7,4%	9,4%
country	8,8%	10,4%	11,7%	10,9%	15,1%	11,2%	6,9%	9,0%
Average yield	53,9	43,4	29,2	38,9	29,5	33,5	32,5	498,1

To compare variability between different species variation coefficients were calculated, where average yields were calculated on the country level. Quite surprisingly, the lowest variation coefficients are observed for the rape, wheat and sugar beet. It could be speculated that the reason for that is the quality of land used for production for each of the crop plants. The three mentioned crops are usually cultivated on the best quality land and, therefore, are more resistant to draught, which is the main limiting factor for crops in Poland.

Functional form of relation between yields variability and the level of aggregation

Because of the differences in sample size it was judged that the best proxy for aggregation level would be average area of the administrative unit. Figure 1. presents values for wheat; the figure is typical and for all other considered crop plants it would look similarly.

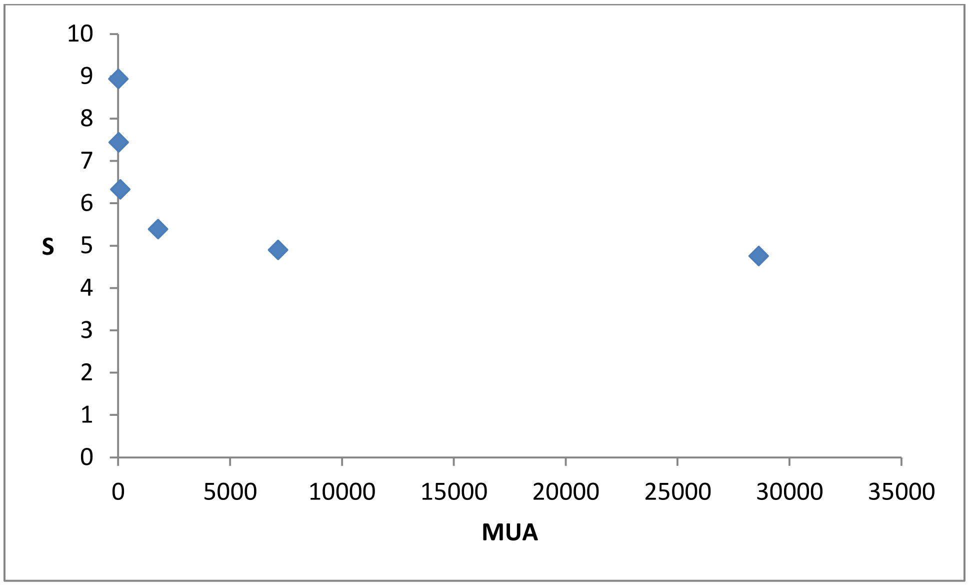

Figure 1. Relation between the mean area of the administrative unit (MUA) and wheat yield variability (without variable transformation) expressed in standard deviations

It is clear that there is no linear relation between variables in the Figure 1. From the considered set of variable transformations, the following were chosen: logarithmic transformation for the mean area of administrative unit and reciprocal for variation. Figure 2. presents the same values as the Figure 1. But on transformed scales.

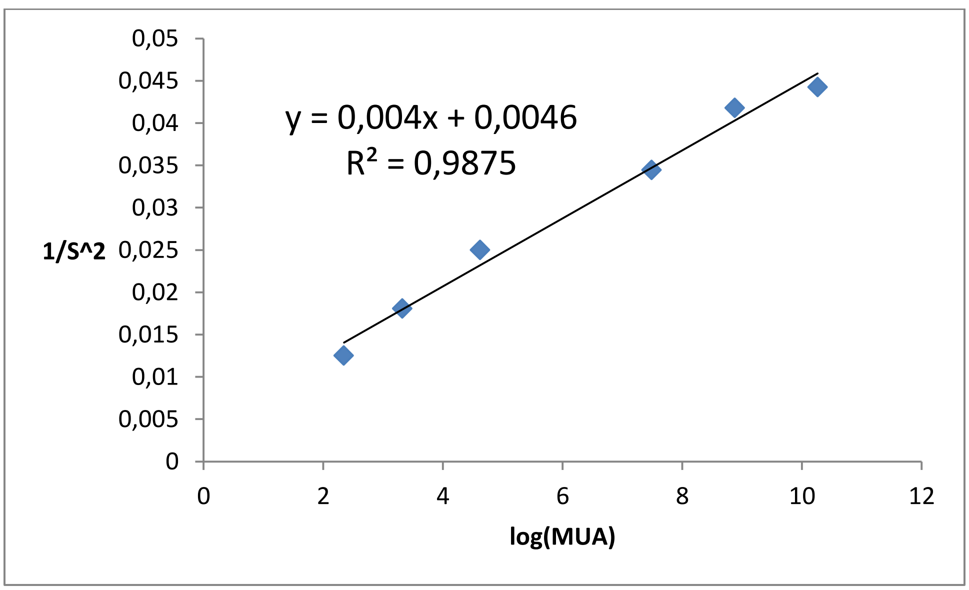

Figure 2. Relation between the mean area of the administrative unit (MUA) and wheat yield variability (transformed variables)

This time all points fits fairly well to the linear function. Moving back from the transformed variables to the original ones would result in the following formula:

$$\frac{1}{S^2} = \beta_0 + \beta_1 \log(MUA) \quad (3)$$

Values for estimators of the given parameters with respective determination coefficient are presented in Table 4.

Table 4. Estimates for parameters of functional form of relation between the mean area of administrative unit and wheat yield, on transformed scales

Crop plant	β_1	β_0	R^2
Winter wheat	0,0040	0,0046	0,988
Triticale	0,0042	0,0103	0,967
Rye	0,0089	0,0091	0,991
Barley	0,0051	0,0037	0,982
Oat	0,0053	0,0087	0,998
Mixed cereals	0,0060	0,0128	0,992
Rape	0,0273	-0,0519	0,993
Sugar beet	0,0000531	0,0000735	0,987

The high values of determination coefficients in the Table4. indicate only good quality of approximation function given by the formula (3). It does not mean that the estimated formulas allow to forecast yield variability on the farm level for a specific farm. It must remembered that by averaging the standard deviations for each level of aggregation we subtracted all the variability of standard deviations around means on each level of aggregation. Nevertheless, The estimated formulas allow to assess the net effect of data aggregation for the investigated crop plants.

Discussion

The results of this research confirm that the ratio of the standard deviations on the farm level and on a high level of aggregation range from 1.4 to 2.7, depending on the level of aggregation and the crop plant. This range is very similar to results reported by Marra and Schurlay [Marra and Shurle 1994]; where for the county level it was 1.4 to 2.8. On the other hand, the numbers from the research concerning wheat yield variability in Northwest Mexico [Lobell et al 2007] were not observed for any of the investigated crop plants. The reason for that may be the fact that in Northwest Mexico the wheat plantations were irrigated, which in general decrease spatial variability of yields.

Górski and Górska proposed in their research concerning the relation between the size of the field and yield variability [Górski and Górska 2003] the power transformation of field size i.e.: field size to the power 0.4. Such transformation did not fit the data presented in this paper. Author propose to use logarithmic transformation for the production area and reciprocal for the variation. The logarithmic transformation for the production area agrees with previous research of the author concerning modelling wheat yields variability in Polish voivodships [Kobus 2010].

Conclusions

The relationship between aggregation level and yield variability can be approximated by the following formula: $\frac{1}{S^2} = \beta_0 + \beta_1 \log(MUA)$, where MUA is the average production area in the administrative unite.
Values of the parameters β_0 and β_1 are dependent on the crop plant, therefore, generalization should be limited to the function formula.
The given formulas allows only general forecast of the analysed crop plants yield variability on the farm level. However, it is clear that, on average, using estimates of yield variability from the country level or even from the voivodship level would result in serious underestimating of the production risk.

Reference

1. Górski T., Górska K. (2003). The Effects of Scale on Crop Yield Variability. *Agricultural Systems* 78 (2003): pp. 425–434.
2. GUS (2010). *Rocznik Statystyczny Rolnictwa 2009*, Zakład Wydawnictw Statystycznych, Warszawa. p. 75.
3. Kobus P. (2010). Modelling Wheat Yields Variability in Polish Voivodeships. Problems of World Agriculture, Vol. 10 (XXV), pp. 33-40.
4. Lobell D. B, Ortiz-Monasterio J. I., Falcon W. P. (2007). Yield uncertainty at the field scale evaluated with multi-year satellite data. *Agricultural Systems* 92 (2007): pp. 76–90.
5. 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.