A Stochastic Programming Based Analysis of the Field Use in a Farm

Béla Vizvári [a]   Zsolt Csizmadia[b]   Gergely Kovács[c]

[a]Visitor of RUTCOR, bvizvari@rutcor.rutgers.edu, and Dept. of Operations Research, Eötvös Loránd University of Budapest, H-1117 Budapest, Pázmány Péter sétány 1/c, vizvari@math.elte.hu, and Dept. of Industrial Engineering, Eastern Mediterranean University, TRNC Mersin 10, Turkey

[b]Visitor of RUTCOR, zsolt@rutcor.rutgers.edu, and Dept. of Operations Research, Eötvös Loránd University of Budapest, H-1117 Budapest, Pázmány Péter sétány 1/c, csisza@math.elte.hu

[c]Dept. of Economics, College for Modern Business Studies, Tatabánya, Stúdium tér, H-2800 Hungary, kovacs.gergely@mutf.hu

A Stochastic Programming Based Analysis of the Field Use in a Farm

Béla Vizvári   Zsolt Csizmadia   Gergely Kovács

Abstract. In a previous report (RRR 16-2006) the Hungarian structure of the use of arable land has been analyzed. This paper is devoted to application of a similar approach to farms. The main difference is that somewhat surprisingly the underlying mathematics, i.e. the stochastic programming model, is more complicated in the case of farm level than that one on the national level.

Crop rotation determines that which crops can be produced in a field. The farm want to have certain amount of productions from some crops. The aim is to determine that use of the fields, which guarantees the highest probability of the satisfaction of demand. The results are illustrated via an example.

1   Introduction

In a previous report (RRR 16-2006) the structure of the use of the arable land has been developed in an evolutionary manner in Hungary. Seven main crops were selected for the study. They are: potato, maize, wheat, sugar been, rye, barley, and sunflower. Approximately two third of the Hungarian arable land is used for the production of these crops. The country has 19 counties. In spite of the fact that the conditions of agricultural production are very different in the counties, all seven crops are produced in all counties with a few exceptions. The main result of [4] is that the same level of supply is achievable with a higher probability on a significantly smaller area.

As the agriculture is affected by stochastic factors via the weather, no supply can be guaranteed up to 100%. Thus each production structure provides the required supply only with a certain probability. Thus the underlying mathematical problem of the determination of an optimal structure is a stochastic programming problem.

This approach was developed further on to apply it to farms. Here are the main differences between the national level and farm level. Each county can be considered as a part of a farm having a field for each corn. The basic properties of the field with the exception of its size do not change. The sizes are depending on each other and must obey some rules, e.g. their sum in each county cannot exceed the arable land in the county. On the other hand the fields of a farm (i) can be used for several purposes and the key decision is that for what in the next campaign, and (ii) have fixed sizes.

The paper is organized as follows. In Section 2 some relation of the weather and the yield of crops are discussed. In Section 3 the underlying stochastic programming model is discussed. An algorithm to solve it is proposed in Section 4. The method uses a discretization technique and leads to a large scale multi-dimensional knapsack problem with Generalized Upper Bound (GUB) constraints. The production plans are evaluated in Section 5. Optimal solutions have been determined for both to maximize the probability of the supply and to minimize the area of the used arable land if the requested probability is given. Numerical results for the optimal structure of the production are presented in Section 6. The paper is closed by some conclusions.

2   What data can be used by a farm?

One important characteristics of crop production on arable land is the uncertainty caused by the stochastic nature of weather. On the other hand this is not a completely stochastic relation as the crops have a certain "memory" and the yield is affected by previous years, which are known, of course. In Hungary the weather of a year (from September to the next August) is aggregated into a single quantity called *drought index* defined by Pálfai and denoted by PDI. It is a positive value. The higher the value is, the greater the drought is.

In the case of small values there was no drought in that year at all. Although it was defined later [2] this drought index has been determined for each year since 1930. It is different from the also well-known Palmers drought index. In [4] it was shown that not only the drought index of the current year bur even the last previous ones may affect the yield. It is important to emphasize that the mechanism of the relation of the current yield with an earlier drought index is not explored in [4], it was pointed out only that such relation may exist. Therefore they also must be considered as independent variables in a regression analysis.

Many products of industry are changing fast, e.g. the cars of 1930's are significantly different from the cars of 2000's. Moreover there are industrial products, which did not exist some decades ago. At the same time the agricultural products as wheat, milk, and meat remain more or less the same. Therefore the productivity of agricultural production can be measured in natural quantities even in a long time interval, too. Another important characteristics of crop production on arable land is the constant improvement of the technology proven by the increase of yields. Here we use only the fact that yields have a trend, i.e. time also emerge in the regression equations as an independent variable.

The drought index is a random variable, too. Its distribution function $D(\cdot)$ was estimated in [4] as

$$D(x) \;=\; e^{-\left(1+\rho\frac{x-\nu}{\sigma}\right)^{-\frac{1}{\rho}}}, \tag{1}$$

where the values of the parameters are: $\nu = 4.12678063$, $\sigma = 1.43656559$, $\rho = 0.01546327$.

Each field has a so-called field-register in Hungary. It contains the most important data of the production in every year. Assume that the expected value of the yields is estimated by regression analysis. Then in each year there is an equation for every field. According to the crop rotation the equations of a field belong to different crops in different years. The number of equation may be enough to a good regression, especially if the overall effect of the drought indices is well-approximated.

3 The Stochastic Programming Model

The agricultural problem to be modeled is to determine an optimal production structure such that

- a prescribed supply is provided with at least a fixed probability,

- further technical constraint are satisfied and

- the probability of the fulfillment of the demand is maximal.

The technical constraints include (a) the expected value of the harvested amount must be at least as great as the demand, and (b) a field can be used only for those crops, which are allowed by the current state of the crop rotation.

The following notations will be used:

m　　the number of crops
n　　the number of fields
$\mathbf{e}$　　the n-dimensional vector such that all of its components are equal to 1
$\mathbf{a}$　　the vector of the sizes of the fields
$\mathbf{d}$　　the vector of the demands
$\mathbf{b}_i$　　the characteristic vector of the fields, which can be used for crop i according to
　　　　the current state of the crop rotation
$\underline{\xi}_i$　　the n-dimensional random vector of the yields from crop i
$\underline{\mu}_i$　　the vector of the expected values of the yields from crop i
$\mathbf{C}_i$　　the covariance matrix of the yields of crop i
Φ　　the distribution function of the $N(0,1)$ standard normal distribution
$\mathbf{x}_i$　　the characteristic vector of the fields used for crop i, i.e.
　　　　a vector of binary decision variables, where $x_{ij} = 1$ iff crop i is produced in field j

Moreover if $\mathbf{g}$ and $\mathbf{h}$ are two arbitrary vectors of equal dimension then $\mathbf{g} \cdot \mathbf{h}$ is the product vector having the components $g_k h_k$.

The mathematical model has four sets of constraints.

(i) The expected value of the harvested amount must be at least as great as the demand for each crop:

$$\mathbf{a}^T (\mathbf{x}_i \cdot \underline{\mu}_i) \geq d_i \quad i = 1, ..., m. \tag{2}$$

(ii) Each field must be used for exactly one crop:

$$\sum_{i=1}^{m} \mathbf{x}_i = \mathbf{e}. \tag{3}$$

(iv) Each field can be used only for those crops, which are allowed by the crop rotation:

$$x_{ij} \leq b_{ij} \quad i = 1, ..., m, \ j = 1, ..., n. \tag{4}$$

(iii) All of the variables are binary:

$$\mathbf{x}_i \in \{0, 1\}^n \quad i = 1, ..., m. \tag{5}$$

The objective function is the overall probability of the fulfillment of the demands:

$$\max P \left(\mathbf{a}^T (\mathbf{x}_i \cdot \underline{\xi}_i) \geq d_i \mid i = 1, ..., m \right). \tag{6}$$

Problem (2)-(6) has deterministic constraints. The stochastic component appears only in the objective function. It can be solved if the common distribution function of (6)

or its approximation can be formulated as a function of the decision variables. [1] solved a similar problem with a single inequality in the probabilistic constraint. For a good summary see [3] (Section 8.5). Based on the computational experiences obtained from Kataoka's method [4] an approximate model is suggested.

Assuming that the yields have normal distribution the expected value of the harvested amount from crop i is

$$\varepsilon_i = \mathbf{a}^T(\mathbf{x}_i \cdot \underline{\mu}_i)$$

and its standard deviation is

$$\sigma_i \sqrt{(\mathbf{a} \cdot \mathbf{x}_i)^T \mathbf{C}_i (\mathbf{a} \cdot \mathbf{x}_i) + 1}.$$

The harvested amount is itself a random variable and has normal distribution determined by the expected value and standard deviation. Thus the probability that the demand will be satisfied from crop i is

$$1 - \Phi\left(\frac{d_i - \varepsilon_i}{\sigma_i}\right). \tag{7}$$

It is clear that the higher the value of ε_i the higher the probability is. If $\varepsilon_i = d_i$ then this probability is 0.5. The probability of the event that all of the demands will be satisfied is high only if all individual probabilities (7) are high, i.e. the expected values ε_i $(i = 1, ..., m)$ are simultaneously high. In the approximate model a common factor, denoted by y, is introduced. It is required that the expected value of the harvested amount exceeds the demand by this common factor. Thus the model is rewritten in the following way:

$$\max y \tag{8}$$

$$\mathbf{a}(\mathbf{x}_i \cdot \underline{\mu}_i) \geq d_i y \quad i = 1, ..., m \tag{9}$$

$$\sum_{i=1}^{m} \mathbf{x}_i = \mathbf{e} \tag{10}$$

$$x_{ij} = b_{ij} \quad i = 1, ..., m, \ j = 1, ..., n \tag{11}$$

$$\mathbf{x}_i \in \{0, 1\}^n \quad i = 1, ..., m. \tag{12}$$

(The pairs of constraints (3) and (10), (4) and (11), (5) and (12) are, of course just the same.) Problem (8)-(12) is a mixed 0-1 programming problem with a single continuous variable.

4   Simulation of Existing and Suggested Production Structure

Similarly to [4] the final evaluation of the production plans is done by simulation. The aim of the simulation is to obtain a good estimation of the probability of the fulfillment of the plan.

In the model there are two types of random elements: **(a)** the drought index and **(b)** the bias of the yields from the values given by the regression equations. The deterministic element of the equation is: **(c)** the year. These three elements are treated in the simulation as follows.

(a) The simulation starts with the drawing of 100005 values of the drought index from the distribution (1). During the simulation a time window of length 6 goes along the series of the drought indices. Thus 100000 event can be simulated for the fulfillment of the plan.

(b) In general the statement that the technology is improving is true to farms, too, as some exogenous components of the production, e.g. used chemicals and seeds, are improving in general. This fact implies that the variances of the biases have increasing trends. Therefore the basic assumption is that *normalized* biases have a multi-dimensional normal distribution. It is well-known that the multi-dimensional normal distribution is uniquely determined the expected values and the covariance matrix. The expected value is zero. The covariance matrix has been determined. Thus for each of the 100000 positions of the time window the normalized biases is determined from this distribution. Then one can obtain the biases by multiplying the normalized biases by the value of the regression function.

(c) The results concern to the same time period, therefore the variable of the year is kept constant.

5   An example

We consider a small farm of 6 fields denoted by A, B, C, D, E and F. The most important properties of fields are summarized in Table 1.

Field	area (ha)	expected value of yield		
		maize	potato	wheat
A	3.2	5.1	20.5	4.2
B	6.8	5.0	21.0	4.1
C	5.3	6.0	22.0	3.7
D	4.7	5.9	21.3	3.9
E	10.5	6.2	23.5	4.0
F	11.6	6.3	22.2	4.1

Table 1. The sizes and expected yields of fields.

According to the crop rotation the fields can be used as follows:

crop	fields
maize	A, B, D, E
potato	B, C, D, F
wheat	A, C, D, F

Table 2. The crop rotation constraints

The farm has three main products: maize, potato and wheat. The demands are in metric ton: 80, 220, 50. Then constraints (2) and (4) can be written in the following compact form, where the decision variables are indexed by the first letter of the crop and the name of the field:

$$16.32x_{mA} + 34.00x_{mB} + 27.73x_{mD} + 65.10x_{mE} \geq 80. \tag{13}$$

$$142.8x_{pB} + 116.6x_{pC} + 100.11x_{pD} + 257.52x_{pF} \geq 220 \tag{14}$$

$$13.44x_{wA} + 19.61x_{wC} + 18.33x_{wD} + 47.56x_{wF} \geq 50. \tag{15}$$

Here it is easy to get some consequences. E.g. if fields B and E are assigned to potato then potato must be produced on F. With an implicit enumeration we obtain that there are only four feasible solutions of constraints (2) and (4):

	maize	potato	wheat
1	B, E	F	A, C, D
2	A, E	B, D	C, F
3	A, E	B, C	D, F
4	D, E	B, C	A, F

If model (8)-(12) is applied then the value of the objective-function variable y is 1.028, 1.018, 1.018, 1.160 for the four feasible solutions, respectively. Thus solution No. 4 is the optimal one. The simulated probabilities of the fulfillment of demand up to 100, 95 and 90 percent are these:

No.	100%	95%	90%
1	0.3111	0.3486	0.5093
2	0.4025	0. 4244	0.5105
3	0.4435	0.4639	0.5437
4	0.6781	0.6993	0.7740

Solution No. 4 is again the best one. The objective-function-variable y takes its value at maize in the case of solutions both No. 2 and 3. The quality of the third solution is better according to the simulation because expected value of the harvested amount of potato and wheat is smoother at it than at the second solution.

6   Concluding Remarks

A stochastic programming model has been developed for determining the optimal use of the fields of a farm if the demand is a priori known. The model maximizes the probability of the satisfaction of all demands under deterministic constraints. It is hard to formulate the objective function even in the case if the random variables have normal distribution. Therefore an approximate model has been suggested based on the computational experiences of a similar problem analyzing the use of arable land on national level. It is a completely deterministic one having one continuous and several binary variables. The application of the model has been discussed on an example. The quality of the solutions has been also evaluated by simulation. The two different methods gave the same quality order of them.

References

[1] Kataoka 1963 Kataoka, S., A Stochastic Programming Model, *Econometrica*, 31(1963), 181-196.

[2] Dr. Pálfai, I., Boga, T.L., Sebesvári, J., The experiences of the drought forecasting system helping in the preparation for irrigation (in Hungarian), presented on the XIX-th annual meeting of MHT.

[3] Prékopa, A., Stochastic Programming, Kluwer, 1995.

[4] Vizvari, B., Lakner Z., Kovacs G., Csizmadia, Z., A Stochastic Programming and Simulation Based Analysis of the Structure of the Production on the Arable Land in Hungary, RRR 16-2006.

Appendix A

The regression equations of the yields of the three crops are the followings.

$$YIELD_{A,maize} = -267,7325 + 0,138 * year - 0,353 * PDI_5$$
$$YIELD_{B,maize} = -311,5455 + 0,16 * year - 0,347 * PDI_5$$
$$YIELD_{C,maize} = -271,13 + 0,14 * year - 0,26 * PDI_5$$
$$YIELD_{D,maize} = -394,6175 + 0,202 * year - 0,187 * PDI_1$$
$$YIELD_{E,maize} = -334,4785 + 0,172 * year - 0,261 * PDI_5$$
$$YIELD_{F,maize} = -276,816 + 0,143 * year - 0,252 * PDI_5$$
$$YIELD_{A,potato} = -858,712 + 0,442 * year - 0,495 * PDI_4 + 0,456 * PDI_3 + 0,439 * PDI_0$$
$$YIELD_{B,potato} = -1037,7675 + 0,532 * year + 0,633 * PDI_2$$
$$YIELD_{C,potato} = -829,6615 + 0,429 * year - 0,522 * PDI_4 + 0,246 * PDI_3 + 0,181 * PDI_1$$

$$YIELD_{D,potato} = -1177,774 + 0,608 * year - 1,196 * PDI_4 - 0,806 * PDI_3 - 0,402 * PDI_0$$
$$YIELD_{E,potato} = -1098,2655 + 0,568 * year - 0,621 * PDI_5 - 0,44 * PDI_4 - 0,734 * PDI_3$$
$$YIELD_{F,potato} = -857,0865 + 0,445 * year - 0,213 * PDI_2 - 0,388 * PDI_1 - 0,68 * PDI_0$$
$$YIELD_{A,wheat} = -253,3801 + 0,13 * year - 0,0664 * PDI_3 - 0,105 * PDI_2$$
$$YIELD_{B,wheat} = -243,88974 + 0,125 * year - 0,08836 * PDI_5 - 0,112 * PDI_3 + 0,126 * PDI_0$$
$$YIELD_{C,wheat} = -249,787 + 0,128 * year - 0,111 * PDI_3 - 0,095 * PDI_2$$
$$YIELD_{D,wheat} = -279,57986 + 0,143 * year - 0,106 * PDI_3 - 0,119 * PDI_2 + 0,07696 * PDI_0$$
$$YIELD_{E,wheat} = -281,3309 + 0,144 * year - 0,112 * PDI_3 - 0,16 * PDI_2 + 0,0854 * PDI_0$$
$$YIELD_{F,wheat} = -250,805935 + 0,128 * year + 0,08341 * PDI_4 + 0,116 * PDI_0,$$

where PDI_t is the value of the Pálfai's Drought Index t years earlier. E.g. PDI_0 is the index of the current year. If $year = 1986$ and all $PDI_t = 3.5$ then the value of the right-hand sides equal to the expected value of the yield in the example.

Appendix B

The stochastic effect of weather is taken into consideration as the relative error of the regression equations, e.g. if the experienced yield is y_{Am} from maize on field A then the random component is

$$\frac{y_{Am} - YIELD_{A,maize}}{YIELD_{A,maize}}.$$

The hypothetical covariance matrices of the example was taken from the measured values on county level:

field	A	B	C	D	E	F
A	0.022588	0.016509	0.01531	0.009942	0.013648	0.015223
B	0.016509	0.020404	0.007435	0.007955	0.011634	0.010675
C	0.01531	0.007435	0.020542	0.011876	0.014291	0.015633
D	0.009942	0.007955	0.011876	0.019387	0.01134	0.008119
E	0.013648	0.011634	0.014291	0.01134	0.018168	0.014928
F	0.015223	0.010675	0.015633	0.008119	0.014928	0.018566

Table B.1 The covariance matrix of maize.

field	A	B	C	D	E	F
A	0.037569	0.027712	0.026934	0.020785	0.007276	0.01748
B	0.027712	0.062632	0.030246	0.025039	0.022738	0.020064
C	0.026934	0.030246	0.035397	0.012958	0.003286	0.016324
D	0.020785	0.025039	0.012958	0.106984	0.032328	0.030999
E	0.007276	0.022738	0.003286	0.032328	0.045979	0.013813
F	0.01748	0.020064	0.016324	0.030999	0.013813	0.031437

Table B.2 The covariance matrix of potato.

field	A	B	C	D	E	F
A	0.015244	0.012309	0.012881	0.010936	0.009513	0.010406
B	0.012309	0.017721	0.010641	0.010907	0.008462	0.012047
C	0.012881	0.010641	0.01595	0.009209	0.010216	0.009705
D	0.010936	0.010907	0.009209	0.012943	0.010884	0.010554
E	0.009513	0.008462	0.010216	0.010884	0.017191	0.011843
F	0.010406	0.012047	0.009705	0.010554	0.011843	0.013715

Table B.3 The covariance matrix of wheat.