A comparison of simulation softwares in modelling the crop structure management with a stochastic linear programming model

Sándor Kovács[1]

I created a stochastic linear programming model based on crop structure data. As to determine the optimal structure, I perform an MCMC simulation by using WINBUGS and two other Risk Analysis softwares. Best values and the related coefficients of the goal functions, provided by different softwares, were analysed and compared. A deterministic linear programming solution was also compared to all the result of the stochastic simulations. I also determined and compared the optimal solutions of the different softwares according to András Prékopa's study.

Evaluating the results WINBUGS proved to be the most suitable software for establishing management decisions in crop structure modelling. In my study I also presented the way for implementing stochastic linear programming models in WINBUGS.

Keywords: Winbugs, Bayesian Statistics, crop structure, risk analysis, MCMC simulation

1. Introduction

Important decisions influencing the future of the company have to be made under conditions of risk, when reliable information is available only for the most recent time period. Risks must be considered by every economic agent and they should apply methods that are capable of measuring, monitoring and suggesting responses to risks, provided that the information required for decision-making is current and of sufficient amount and quality. The evaluation of this information should enable decision-makers to formulate and analyze multiple decision alternatives. On one hand, the developments in information technology have facilitated the development of applied risk management tools, which have become affordable for even the smallest of enterprises and easy to use. On the other hand new, complex and wide-ranging types of risk have arisen, the measurement of which requires sophisticated mathematical models (Balogh et al. 1999). Simulation models, whose use in

[1] Sándor Kovács, PhD, assistant lecturer, University of Debrecen, Faculty of Applied Economics and Rural Development, Hungary

agriculture has grown rapidly, attempt to mimic the operation of real systems so as to allow accurate measurement of uncertainty and risk.

One of the simplest ways of optimizing the crop structure is to apply linear programing methods (Csipkés et al. 2008). In cultivation LP methods are one of the most favourable methods. However, these methods are hindered by the great measure of uncertainty. The major reasons for that are changes in the prices and yields of crops. That is the reason why stochastic programming has received much attention from 1950 to the present day in many applications (Dantzig 1955, Prékopa et al. 1980, Williams 1966, Prékopa 2003).

The simulation model is a simplified mathematical realization of a real system, aimed at studying the behavior of the original system when changing various conditions and circumstances. The Monte-Carlo method is a generally accepted method of modelling risks which studies the probable outcome of an event characterized by any input parameters and described by well-known functions. The essence of the Monte-Carlo technique is, on the basis of probability distribution assigned to some uncertain factors, to randomly select values, which are used in each experiment of the simulation (Russel–Taylor 1997). Monte-Carlo methods are the statistical evaluations of numerical methods and their characteristics using the modelling of random quantities of mathematical solutions (Szobol 1981). The method is widely used to simulate the likely outcomes of various events and their probability when input parameters are uncertain. In the model to be analyzed the influencing variables and their possible intervals, their likelihood distribution as well as the connections between the variables are fixed. The distribution values of the variables from the given intervals are developed by a random number generator. In the course of my research I examined the possibilities of planning the crop structure taking advantage of the benefits of simulation methods.

2. Material and methods

Bayesian statistic models are often used to model uncertain future events for example stock prices. In effect, this approach handles the unknown parameters in the given model as random variables and samples them from the distribution function based on my preliminary knowledge. Bayesian statistics were unaccomplishable with computer up till the nineties, but then Markov Chain Monte Carlo (MCMC) simulation methods (Metropolis – hasting and Gibbs sampling) received greater emphasis. Translating the problems into Bayesian statistical language based on MCMC is a very difficult task as it demands significant programing and mathematical knowledge as well as a skill in operating random number generators. For the realization of this process, a so-called BUGS (Bayesian interference Using Gibbs sampler) program was developed by the Biostatistics Sub-department of the Cambridge Medical Research Council in 1995 (Spiegelhalter et al. 1996), the programing language of which is very specifically adjusted to the realization of pure likelihood models based on MCMC. Its version running under Windows operational system is known as WINBUGS.

2.1. MCMC (Markov Chain Monte Carlo Simulation)

Let Y be a mass of facts, $\theta = (\theta_1,...,\theta_k)$ vector of random variables (model parameters), with $\pi(\theta, Y)$ joint distribution function. Let us also suppose that $\pi(\theta, Y)$ is complicated and can hardly be given in an analytical form. In terms of the Bayes statistics, $\pi(\theta \mid Y) \propto \pi(\theta, Y)$ and

$$\pi(\theta \mid Y) \alpha \pi(Y \mid \theta)\pi(\theta) \qquad (1)$$

is valid (Congdon 2007), so it is easier to work with this function instead of the joint distribution function. On the other hand, my interest is raised by $\pi(\theta \mid Y)$ distribution because the Y data are given, and I want to use those to estimate the parameters. $\pi(Y \mid \theta)$ is the so-called Likelihood function and $\pi(\theta)$ is the apriori distribution function which I know in advance. Let us also suppose that I'm looking for the expected value of a $h(\theta)$ integrable function, this is given by the following integral based on the $\pi(\theta \mid Y)$ distribution function (Jorgensen 2000):

$$E_\pi(h(\theta)) = \int h(\theta)\pi(\theta \mid Y)d\theta \qquad (2)$$

It is nearly impossible to calculate this integral in an analytical or numeric way. That is why I use the so-called Monte-Carlo integration, the essence of which is that I take a $\theta^{(0)},...,\theta^{(k)}$ sample from $\pi(\theta \mid Y)$ distribution, and so I can estimate the expected result in the following way (David-Scollnik 2001):

$$E_\pi(h(\theta)) \approx \frac{1}{k}\sum_{i=1}^{k} h(\theta^{(i)}) \qquad (3)$$

The combination of the Monte Carlo integration and the Markov Chains is called MCMC simulation, the essence of which is that I simulate possible variations from a Markov chain, the stationary function of which is $\pi(\theta \mid Y)$. It is true that these random samples $\theta^{(0)},...,\theta^{(k)}$ will no longer be independent, but with moderate regularity conditions it is realized that the distribution of $\theta^{(i)}$ converges on $\pi(\theta \mid Y)$ in the case of $i \to \infty$, and as for the expected result it is valid that (David-Scollnik 2001):

$$E_\pi(h(\theta)) \approx \frac{1}{k}\sum_{i=1}^{k} h(\theta^{(i)}), \text{ in the case of } k \to \infty \qquad (4)$$

Now the only question remaining is how I can simulate Markov chain variations, the stationary distribution of which is $\pi(\theta \mid Y)$. Researchers have come up with various methods to accomplish this, one of the most simple techniques is the Gibbs sampling process.

At the beginning of the process I start from the original parameter vector $\theta^{(0)} = (\theta_1^{(0)},...,\theta_k^{(0)})$. I take random samples from the so-called full conditional distribution in the following way (Congdon, 2007):

$$\theta_1^{(1)} \approx \pi(\theta_1 \mid Y, \theta_2^{(0)},...,\theta_k^{(0)})$$
$$\theta_2^{(1)} \approx \pi(\theta_2 \mid Y, \theta_1^{(1)}, \theta_3^{(0)},...,\theta_k^{(0)}) \qquad (5)$$
$$\theta_j^{(1)} \approx \pi(\theta_j \mid Y, \theta_1^{(1)}, \theta_{j-1}^{(1)}, \theta_{j+1}^{(0)},...,\theta_k^{(0)})$$

In fact, this realizes a series of steps by which I go from $\theta^{(0)}$ to $\theta^{(1)}$ parameter vector. After a finite number of iterations I arrive at the Markov chain mentioned above, and the expected result can also be calculated in way (3). Generally speaking I arrive at the desired conditions after a few thousand simulation runs, but often tens of thousands of iterations are needed.

Bayes statistics and simulation methods have a lot in common. Researchers prefer to use these methods in their simulation models primarily because they have an opportunity to build their preliminary knowledge into the model so that this will be able to change during application, influenced by several other factors. David Vose (2006) presents this in his work as a highly efficient analytical tool, based on the Bayes principle and being suitable for estimating parameters on the basis of data more effectively than other methods. The Bayes conclusion theory involves 3 important circumstances (Vose 2006):

- determination of the a priori distribution functions and their parameters,
- determination of suitable likelihood function on the base data,
- determination of posteriori distributions and their parameters

2.2. Stochastic LP model

The general form of the model is as follows:

$$A\overline{x} \le \overline{b} \text{ (restrictive conditions)}$$
$$c\overline{x} -> \max \qquad (6)$$
$$0 \le \overline{x} \text{ consistent distribution on (0,3) interval}$$

In formulae 6 the capacity vector is denoted with „$\overline{b}$” and „A” indicates the technological matrix. The solution vector „$\overline{x}$” consists of variables which are the sowing areas of the different crops in hundred hectares. The goal function coefficients are denoted with „c”, they mean the per unit incomes.

The LP task can be considered stochastic, because the goal function coefficients (the values of the per unit incomes) come from a distribution. In the calculation of their values I take into consideration the selling price and quantity of the main product, which also come from a distribution.

In the marking system of MCMC „Y” is the data set, „A” is the technological matrix and „$\bar{b}$” is the capacity vector. c=$h(\theta)$, where θ contains the variables of prices, average yield and area. The $\pi(\theta \mid Y)$ distribution function takes a special form on account of the restrictive conditions:

$$\pi(\theta \mid Y) = \begin{Bmatrix} 1, if\ A\bar{x} \le \bar{b} \\ 0, if\ \bar{b} < A\bar{x} \end{Bmatrix} \tag{7}$$

Because of this, goal function values are formed during the simulation runs if the restrictive conditions are fulfilled. This way the stochastic linear programing task can be rendered into the language of WINBUGS simulation software. WINBUGS saves the θ values (the values of price, average yield and area variables) as well as the goal function values generated during all the simulation runs. On the basis of this, it is possible to find the maximum value for the goal function.

2.3. Optimality Criterion

Since the maximum value does not equal with the optimal value, I had to select an "Optimality Criterion" for evaluating the performance of the applied softwares as well. I formulated the criterion according to Prekopa's work (Prékopa et al. 1980). Based on this study a solution vector x_k is considered optimal if the difference between the values of the goal function at x_{k+1} and x_k does not exceed 1% of the latter and at the same time each individual component of x_{k+1} - x_k does not exceed 2% of the corresponding component of x_k (Prékopa et al. 1980).

3. Result and discussion

In my analysis I have used the data of an agricultural company farming in Löszhát, Hajdúság. The sowing area of the company is 800 hectares, crops grown are winter wheat, maize, winter coleseed, sunflower and green peas. The technology used by the company was built into the model.

When giving the restrictive conditions, I took into consideration the resources available for the company at the time, as well as the professional rules pertaining crop rotation. These per unit data are deterministic in the model.

In the goal function there appears the profit contribution (per unit income). The per unit changing costs in the individual branches, just like the capacity vector and the per unit demands, may be considered fixed. Within the model, i consider the average yields and the selling prices, that is, the return from sales, probability variables.
I determined the distributions of prices and their parameters taking into consideration the time series sales data of the company being analyzed (Table 1):

Table 1. The distributions of prices applied during the simulation

Crop	Distribution	Parameter 1	Parameter 2
Maize	Gamma	17,91	0.83
Wheat	Gamma	39,3	0.33
Coleseed	Gamma	1,54	6.50
Green peas	Normal	50	20.10
Sunflower	Normal	65	33.3

Source: own calculation

I applied the distributions and parameters of average yields based on the farm data of ARI (Agricultural Research Institute), North Plains Region for the years 2001-2005 (Table 2).

Table 2. The distributions applied to the average yields during simulation

Crop	***Distribution***	***Parameter 1***	***Parameter 2***
Maize	Normal	7.55	1.75
Wheat	Normal	4.54	1.15
Coleseed	Gamma	9.53	0.18
Green peas	Normal	5.56	1.25
Sunflower	Normal	2.18	0.61

Source: own calculation, based on data provided by Agricultural Research Institute

After the simulation runs, performed by 3 different programs, I analyzed the goal function with the help of statistical tools (Table 3). According to the measurement of risk by the decision maker, the appropriate decision alternative can be choosen based on the data set. Risk and Crystal Ball programs use the simpler version of the Monte-Carlo simulation, while the WINBUGS applies the Markov Chain Monte Carlo simulation combined with Bayesian Statistics.

Table 3. Characteristics of the profit contribution

Statistic	***CB***	***Risk***	***Winbugs***
Median	42250,81	44025,17	43450,04
Average	43653,99	44398,14	44401,75
Deviation	17089,83	14982,86	18760,57
Relative Deviation	39,15	33,75	42,25
Minimum	3173,57	7807,51	-10350,00
Maximum	102583,12	85510,90	103100,00
Range	99409,55	77703,39	113450,00
Skewness	0,34	0,09	0,29
Kurtosis	-0,27	-0,63	-0,30

Source: own calculation

Table 3 clearly suggests that the widest range, the highest deviation and the highest goal function value can be gained by WINBUGS. An other significant difference between WINBUGS and the two other programs is that WINGUGS can produce negative result for the profit contribution.

Table 4. Comparison of the crop structure and goal function values

Factor	***Applied Softwares***						***Deterministic Linear Programming***
	Best result (highest value)			***Best ten results***			
	Crystal Ball	***Risk***	***Winbugs***	***Crystal Ball***	***Risk***	***Winbugs***	
Goal function*	102583,12	85510,90	103100,05	92481,95	81307,90	97164,02	88460,24
Maize**	2,40	2,43	2,03	2,14	2,36	1,98	2,46
Wheat**	1,92	2,33	2,12	1,88	2,18	1,82	2,36
Coleseed**	0,78	0,78	0,22	0,74	0,44	0,49	0,00
Green peas**	0,16	0,51	0,21	0,33	0,41	0,27	0,53
Sunflower**	1,25	1,77	1,56	1,27	1,81	1,09	2,00

Source: own calculation;*: value in Hungarian Forint; **:sowing area in hundred hectares

Evaluating the results of the simulation runs, it can be observed that by using Risk and Crystal Ball the simulation resulted in higher values for profit contribution than the deterministic linear programming optimum in eight cases, while in the case of WINBUGS in seventeen cases. Table 4 shows the best and the average from the best ten results of the applied softwares, and all the relevant values of the decision variables. For the sake of the better interpretation, I calculated the standard deviation from the deterministic linear programming solution (Table 5).

Table 5. Standard deviations from the deterministic solution

Factor	***Applied Softwares***					
	Best result			***Best ten results***		
	Crystal Ball	***Risk***	***Winbugs***	***Crystal Ball***	***Risk***	***Winbugs***
Goal function	0,16	-0,03	0,17	0,05	-0,08	0,10
Maize	-0,06	-0,03	-0,43	-0,32	-0,10	-0,49
Wheat	-0,44	-0,03	-0,24	-0,48	-0,18	-0,54
Coleseed	0,78	0,78	0,22	0,74	0,44	0,49
Green peas	-0,37	-0,02	-0,32	-0,21	-0,13	-0,26
Sunflower	-0,75	-0,23	-0,44	-0,73	-0,19	-0,91

Source: own calculation

Taking the best result in table 5 into consideration, it can be stated that Risk produces the closest result (goal function and sowing areas) to the deterministic linear programming optimum and the solution of WINBUGS and Crystal Ball differs the most from it. Based on the best ten results in table 5, it is obvious that WINBUGS produces the most different result from the deterministic model taking the highest values into consideration and the object function value is the highest in the case of this software.

Table 6. Comparison of the crop structure and goal function values according to Prékopa's optimality criterion***

Factor	***Applied Softwares***						***Deterministic Linear Programming***
	Best optimal solution			***All optimal solutions***			
	Crystal Ball	***Risk***	***Winbugs***	***Crystal Ball***	***Risk***	***Winbugs***	
Goal function*	56940.09	57642.19	73870.00	43978.50	52321.20	47212.86	88460,24
Maize**	1.53	1.40	1.69	1.25	1.56	1.38	2,46
Wheat**	1.74	1.77	1.53	1.18	1.32	1.18	2,36
Coleseed**	0.27	0.24	0.30	0.50	0.45	0.40	0,00
Green peas**	0.31	0.34	0.23	0.24	0.30	0.27	0,53
Sunflower**	1.07	1.60	1.58	1.01	1.07	0.97	2,00

Source: own calculation;*: value in Hungarian Forint; **:sowing area in hundred hectares
****:* Prékopa et al. (1980)

Table 6 shows the best and all optimal solutions of the applied softwares, and all the relevant values of the decision variables. The optimal solution of the Crystal Ball and Risk Softwares are almost the same, while WINBUGS produce the closest value to the deterministic optimum. Regarding the average of all the optimal solutions, the three softwares produced almost the same solution for the crop structure.

4. Conclusion

Traditional planning is still the most often applied method in cultivation, which provides adequate planning, but also determines an increasing shortfall in economic competition. Due to price and yield fluctuations a methodologically appropriate optimizing planning is necessary. In optimizing planning, linear programing models are most often used, however, because of their deterministic nature, in choosing from among decision alternatives, risks cannot be taken properly into consideration. Applying simulation models may be a solution, in my work I have presented one such application. Out of the three programs, WINBUGS produces the most different result from the deterministic model taking the highest values into consideration and the object function value is the highest in the case of this software, as the software uses Bayesian Statistics. The range of the object function values produced by WINBUGS is wider than in the case of the two other programs. Risk Software produces the closest result (goal function and sowing areas) to the deterministic linear programming optimum. Regarding the optimal solutions, Crystal Ball and Risk Softwares produced almost the same values, while WINBUGS produce the closest value to the deterministic optimum. Regarding the average of all the optimal solutions, the three softwares produced almost the same solution for the crop structure. To sum up, risks and optimal solutions can be modelled and considered in a more widespread and accurate way by using WINBUGS software.

What the crop structure concerns I can state that maize and wheat are of grater importance as these crops constitute a large part in the crop structure. The growing of sunflower might become more important due to the large increase in biodiesel production.

References

Balogh, P. –Bai, A. –Posta, L. –Búzás, F. 1999: Analysis of environmental effects of cereals and energy plantations in case of unfavourable agricultural conditions, Conference on „The resources of the environment and the sustained development”, University of Oradea, Oradea.

Congdon, P. 2007: Bayesian statistical modelling. Wiley, New York.

Csipkés, M. – Etsey, I. – Nagy, L. 2008: Vetésszerkezeti variánsok összehasonlító elemzése a Hajdúságban. XI. Nemzetközi Tudományos Napok, Gyöngyös, 14-21. p.

Dantzig, G. B. 1955: Linear programming under uncertainty. Management Sci. 1, 197–206. p.

David, P. – Scollnik, M. 2001: Actuarial modeling with MCMC and BUGS. North American Actuarial Journal, 5/2, 96-124. p.

Jorgensen, E. 2000: Monte Carlo simulation models: Sampling from the joint distribution of “State of Nature”-parameters. In: Fels, V.D. – Klerx, I. – Mourits, M. (eds): Proceedings of the Symposium on “Economic modelling of Animal Health and Farm Management”, Farm Management Group, Dept. of Social Sciences. Wageningen University, Wageningen, p. 73-84. http://gbi.agrsci.dk/~ejo/publications/preprints/specprior.pdf.

Prékopa, A. – Ganczer, S. – Deák, I. – Patyi, K. 1980: The STABIL stochastic programming model and its experimental application to the electricity in Hungary In: Dempster, M. A. H. (ed): Stochastic Programming, Academic Press, London, 369-385. p.

Prékopa, A. 2003: Probabilistic programming. In: Ruszczynski, A. – Shapiro, A. (eds): Stochastic Programming Handbooks Operation Research Management Sci. 10, Elsevier, Amsterdam, 267-351. p.

Russel, R. S. – Taylor, B. W. 1997: Operations Management, Focusing on quality and competitiveness. Prentice Hall, New Jersey, 610-613. p.

Spiegelhalter, D. – Thomas, A. – Best, N. – Gilks, W. 1996: Bayesian inference using Gibbs sampling program Version 0.50. MRC Biostatistics Unit of Cambridge, UK, http://www.mrc-bsu.cam.ac.uk/bugs/documentation/Download/manual05.pdf.

Szobol, M. 1981: A Monte-Carlo módszerek alapjai. Műszaki Könyvkiadó, Budapest.

Vose, D. 2006: Risk analysis, a quantitative guide. Wiley, New York.

Williams, A. 1966: Approximation formulas for stochastic linear programming. SIAM J. Appl. Math. 14, 668–877. p.