DEPARTMENT OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

IFRO Working Paper

Bayesian network as a modelling tool for risk management in agriculture

Svend Rasmussen
Anders L. Madsen
Mogens Lund

2013 / 12

IFRO Working Paper 2013 / 12

Bayesian network as a modelling tool for risk management in agriculture

Authors: Svend Rasmussen, Anders L. Madsen, Mogens Lund

Department of Food and Resource Economics (IFRO)
University of Copenhagen
Rolighedsvej 25
DK 1958 Frederiksberg DENMARK
www.ifro.ku.dk

Bayesian network as a modelling tool for risk management in agriculture

Svend Rasmussen[1], Anders L. Madsen[2], and Mogens Lund[3]

[1] University of Copenhagen, Denmark
[2] HUGIN EXPERT A/S and Aalborg University, Denmark
[3] University of Copenhagen, Denmark

Abstract. The importance of risk management increases as farmers become more exposed to risk. But risk management is a difficult topic because income risk is the result of the complex interaction of multiple risk factors combined with the effect of an increasing array of possible risk management tools. In this paper we use Bayesian networks as an integrated modelling approach for representing uncertainty and analysing risk management in agriculture. It is shown how historical farm account data may be efficiently used to estimate conditional probabilities, which are the core elements in Bayesian network models. We further show how the Bayesian network model RiBay is used for stochastic simulation of farm income, and we demonstrate how RiBay can be used to simulate risk management at the farm level. It is concluded that the key strength of a Bayesian network is the transparency of assumptions, and that it has the ability to link uncertainty from different external sources to budget figures and to quantify risk at the farm level.

Keywords: Bayesian network, Risk, Conditional probabilities, Stochastic simulation, Database, Farm account

1. Introduction

Risk management is an important part of business management. The last five years of development with the global financial crisis has increased the focus on corporate risk management. Furthermore, in agriculture, highly volatile prices and a tendency to re-orientate agricultural policy so that farmers to a higher degree bear the consequences of uncertain prices and production conditions have also increased interest in risk management.

Literature reveals that the objective of risk management is difficult to describe in operational terms. Efficiency seems to be the key term, and the objective of risk management is often - implicitly or explicitly - stated as a way to manage the business in an *efficient way*, i.e. to minimise risk for a given level of other objectives (profit/income) or to maximise the achievement of other objectives (profit/income) for a given level of risk. Hubbard[7] (2009) provides a good review of the theoretical and methodological development of risk management since World War II. He also provides a formal definition of risk management as: 'The identification, assessment, and prioritization of risks followed by coordinated and economical application of resources to minimise, monitor, and control the probability and/or impact of unfortunate events' (ibid, p. 10). Chapman[2] (2011) provides an extensive introduction to the state-of-the-art regarding applied business risk management based on the ISO 31000 standards published by the International Organization for Standardization[8] (ISO, 2009). According to the ISO 31000 standards, risk management "…aids decision making by taking account of uncertainty and its effect on achieving objectives and assessing the need for any actions" (p. V), and risk management "… should thus help avoid ineffective and inefficient responses to risk that can unnecessarily prevent legitimate activities and/or distort resource allocation" (p. VI).

In agricultural economics (farm management), risk management has historically been based on the Expected Utility (EU) model and various ad hoc procedures based on stochastic simulation, the concept of stochastic dominance and Value-at-Risk.[1] The major challenge has been (and still is!) that the functional form and the parameters of the utility function are normally unknown and may even vary from one farmer to the other and from one situation to the other.

[1] Hardaker[5] (2006) provides a good review of the approaches to risk management in agriculture and the challenges ahead.

Another challenge is to quantify the risk faced by farmers. The relevant measure is the probability distribution of future income available to meet financial obligations. Such probability distributions may be estimated directly from historical data. However, if the objective is risk management, this is not enough. To control risk, it is necessary to identify the individual risk factors and to model the more or less complex systems in which they interact to generate income. The Bayesian network model described and demonstrated in the following is an extremely suitable modelling approach for modelling complex systems and for the estimation of the relevant parameters (conditional probabilities) based on historical observations when the objective is risk management.

Before the Bayesian network modelling approach is described in Section 3, we shortly describe risk and the general approaches to risk management in agriculture (Section 2). In Section 4 we present the Bayesian network model RiBay which we have developed for risk management in pig and crop production. The use of the model is demonstrated in Section 5, and in Section 6 we discuss the future perspectives and draw some general conclusions.

2. Risk and risk management in agriculture

2.1. Approaches to risk management

Hardaker et al.[6] (2004), EU[4] (2005), Hardaker[5] (2006), OECD[16] (2009) and Schaffnit-Chatterjee[17] (2010) present the mainstream views on risk and contemporary methods for risk management in agriculture. The major sources of risk in agriculture are *production risks*, which are due to unpredictable weather, diseases and other unpredictable production conditions, *price and market risks* associated with the variability in prices and access to markets, *financial risks* resulting from changes in financing conditions and interest rates, *institutional risks*, which embody changes in government policy and regulation and *technological risks,* which are risk factors related to the introduction of new technology.

Farm managers may use four main strategies to manage risk. The first, *retaining the risk*, is a strategy of self-protection where farm managers cope with the consequences of bad outcome by taking ex-post actions based on the use of financial reserves, such as savings, or of credit, which allows them to smooth consumption in the face of varying income. *Avoiding the risk* is a strategy where risk is evaded by not taking (too) risky actions or by eliminating their negative effects (choice of products and technologies). *Reducing the risk* can be achieved through flexibility and through diversification by engaging in various uncorrelated risky activities. Finally, *transferring the risk* is based on classical risk management tools like insurance, but also future contracts, weather derivatives and financial derivatives like options and futures.

Each of these strategies is based on a number of *risk management tools*. The central question to risk management is to choose between the risk management tools so as to obtain the highest risk reduction at the lowest cost. The Bayesian network model presented in the following focuses on precisely this question, i.e. to investigate the consequences of alternative risk management tools.

2.2. Quantifying risk – estimation of conditional probabilities

The essential data input for Bayesian network models are the conditional probabilities for the individual stochastic variables (risk factors) in the network. Probabilities concerning future events are almost by definition subjective, and the scientific challenge is therefore to choose the best way in which to estimate these conditional probabilities.

Individual farmers have farm accounts including historical data on production and resource use. These data may be used to estimate conditional probabilities and, consequently, conditional probability distributions (CPDs – see next Section) concerning production risk. The advantage of using individual farm account data to estimate CPDs is that the estimated CPDs are to the full extent conditional on all the conditions on the specific farm (manager, production system, soil quality, region, climate, etc.). However, the number of observations for each condition is typically very low, and for some conditions may even be non-existing.

Estimation of CPDs describing production risk must therefore be based on larger data sets. All EU countries are part of the FADN (Farmers Agricultural Data Network), and therefore have national databases of individual farm accounts, which may be used to estimate CPDs concerning production.

Concerning price risk, the individual farmers in a country or region often face the same prices, and therefore more observations on individual farms does not increase the number of (independent) observations. This means that CPDs concerning (yearly) prices typically have to be based on a limited number of observations.

The limited amount of (relevant) historical observations that one may have available means that other sources of information become relevant. Prior knowledge and input from experts may be used to supplement historical observations. In this context the Bayesian network model is suitable, because it is "designed" to being able to combine prior knowledge and historical observations when estimating CPDs.

3. Bayesian networks

3.1. Introduction

A Bayesian network is an efficient, compact and intuitive knowledge representation for handling uncertainty[10]. It consists of two main parts; (1) a graphical structure that defines a set of dependence and independence statements over a set of random variables representing entities of a problem domain and (2) a set of CPDs specifying the strengths of the dependence relations encoded in the graphical structure. A Bayesian network $N=(X, G, \mathrm{P})$ over variables $X=\{X_1, \ldots, X_n\}$, consists of a directed acyclic graph G and a set of CPDs P. It encodes a decomposition of a joint probability distribution as:

$$P(\mathrm{X}) = \prod_{i=1}^{n} P(X_i \mid pa(X_i)) ,$$

where $pa(X_i)$ are the parents of X_i in G.

A Bayesian network can serve as a knowledge integration and representation tool for supporting decision making under uncertainty. The information captured by a Bayesian network may originate from a range of different sources as is the case for the model RiBay presented later. For instance, the CPDs may be defined based on expert knowledge assessment or subjective estimates, mathematical expressions relating parent configurations to states of the child, or estimated from data. This makes a Bayesian network an excellent knowledge integration framework. Probabilistic inference (also known as belief update) is the task of computing posterior probability distributions given evidence, i.e., computing the posterior marginal $P(X_i \mid \varepsilon)$ for X_i given a set of observations $\varepsilon = \{Y_{i_1} = y_{j_1}, \ldots, Y_{i_r} = y_{j_r}\}$, where Y_i is a variable describing evidence of x_i and y is a state of variable Y .

A Bayesian network is an intuitive, graphical representation of a joint probability distribution of a set of random variables. In order to support efficient parameterization of a Bayesian network, the HUGIN tool[12] allows the user to annotate a Bayesian network with a set of function nodes $F=\{F_1, \ldots, F_m\}$. A function node (in this case) represents a numerical value either computed after a belief update operation or a numerical value specified prior to belief update and (usually) used to parameterize a CPD.

3.2. Examples from RiBay

Figure 1 shows an example of a Bayesian network (structure) of four variables $X_1, \ldots, X_4$ representing *Soil type*, *Climate*, *Yield of rape seed* and *Yield of grain*, respectively. That is, the oval vertices as illustrated in the graph of Figure 1 represent random variables. The intuitive meaning of a directed arc is that the parent node has a direct impact on the state of the child node. A missing arc specifies a conditional independence statement.

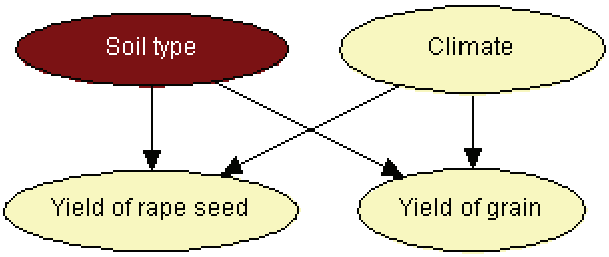


Figure 1. **Bayesian network of four random variables**

The variables *Soil type* and *Climate* are parent variables of the two *Yield* variables. *Soil type* and *Climate* are marginally independent and dependent given any of the *Yield* variables. The variables have a set of mutually exclusive and exhaustive states, e.g., *Soil type* has three states "*clay*", "*sand*" and "*mixed*" while *Climate* has four states "bad", "below average", "above average" and "good". Similarly, *Yield of rape seed* and *Yield of grain* each have seven states (three states representing negative deviation, one state representing neutral and three states representing positive deviations). The quantification of the model amounts to specifying prior probability distributions for *Climate* and *Soil type* and one CPD for each *Yield* variable given *Soil type* and *Climate*. The priors are *P*(Climate) = (0.25, 0.25, 0.25, 0.25) and *P*(Soil type) = (0.5, 0.4, 0.1) where *Soil type* is a system variable. The CPD for variable *Yield of rape seed* given *Climate* and *Soil type* is shown in Figure 2.The CPD for variable *Yield of grain* has different state values, but the same CPT as *Yield of rape seed*.

Yield of rape

Climate	bad			below average			above average			good		
Soil type	sand	clay	mixed	sand	clay	mixed	sand	clay	mixed	sand	clay	mixed
0.55	0.1	0.05	0.075	0.08	0.035	0.055	0.025	0.01	0.018	0.01	0.01	0.01
0.7	0.125	0.06	0.09	0.1	0.07	0.087	0.1	0.08	0.09	0.08	0.08	0.08
0.85	0.175	0.175	0.175	0.185	0.175	0.18	0.2	0.18	0.19	0.18	0.18	0.18
1	0.2	0.43	0.32	0.27	0.44	0.356	0.35	0.46	0.404	0.46	0.46	0.46
1.15	0.175	0.175	0.175	0.185	0.175	0.18	0.2	0.18	0.19	0.18	0.18	0.18
1.3	0.125	0.06	0.09	0.1	0.07	0.087	0.1	0.08	0.09	0.08	0.08	0.08
1.45	0.1	0.05	0.075	0.08	0.035	0.055	0.025	0.01	0.018	0.01	0.01	0.01

Figure 2. **CPD for variable *Yield of rape seed* given *Climate* and *Soil type***

The Bayesian network model is an efficient representation of a joint probability distribution of $X_1,\ldots,X_4$ allowing us to compute the posterior marginal of any variable given any subset of evidence. For instance, we can compute P(Yield of grain | Climate = bad) = (0.078, 0.096, 0.175, 0.304, 0.175, 0.096, 0.096). Figure 3 shows the marginal distributions for all four variables in this network fragment.

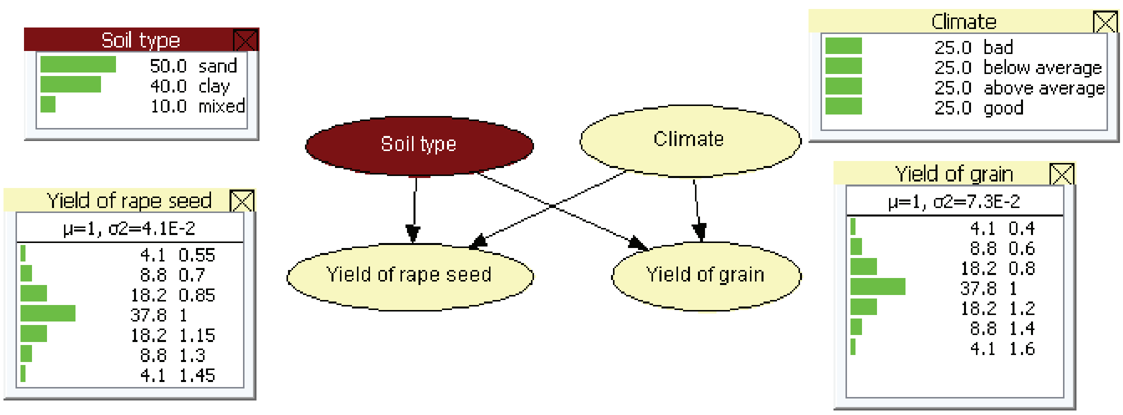


Figure 3. **Marginal distributions for *Soil type*, *Climate*, *Yield of rape seed* and *Yield of grain***

The structure of the model has been developed by the assessment of expert domain knowledge, whereas the (conditional) probability distributions have been estimated from data. The CPD defines a probability distribution of the states of the child variable for each configuration of the parent variables. This means that the size of the CPD grows exponentially with the number of parents.

In the HUGIN tool[12] it is possible to annotate a Bayesian network with a set of function nodes *F*. A function node $F_i \in F$, drawn as a hexagon, may, for instance, represent either a numerical value computed from a mathematical expression, possibly including posterior probabilities after a successful belief update operation. A function node can also be parent of a random variable, in which case the value of the function node can be used to define the (conditional) probability distribution of the random variable from a mathematical expression. Figure 4 shows an example where the function node *Number of piglets purchased* is parent of random variable *Number of piglets* and its value used in the mathematical expression for the CPD of the random variable.

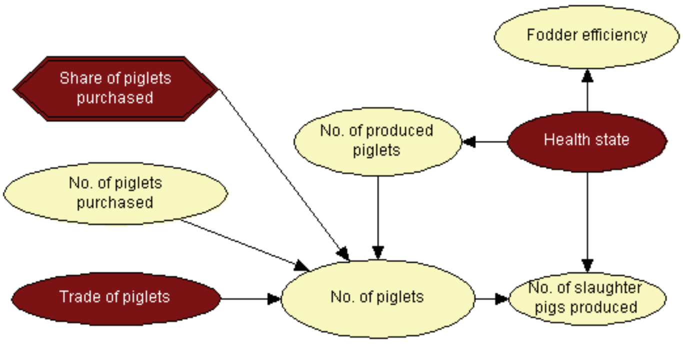


Figure 4. Number of produced piglets and slaughter pigs

The parameters of the CPD of each random variable can be defined independently of the parameters of other CPDs in the network. This means that a Bayesian network can be considered as a model of statistical models, i.e., each variable and its CPD is a model and the models are connected as defined by the graphical structure *G*. The CPDs define the uncertainty of the problem domain encoded in the Bayesian network model. As the size of the CPD grows exponentially with the number of parents, it is desirable to construct a sparse graph of the variables, as this will produce the most efficient knowledge representation and help to ensure efficient probabilistic inference. Using sparse graphs it is possible to construct efficient knowledge representation of even a large number of variables.

4. The Bayesian network RiBay

The Bayesian network RiBay is a knowledge representation of elements of the farm budget combined with external and stochastic variables (risk factors) that have a strong influence on farm level risk. The scale of the model is a one year budget with a focus on risk at the gross margin level. RiBay provides a holistic view of the major factors impacting farm level risk at the level of a one year budget with a focus on farm income.

4.1. The network structure

The RiBay model is a Bayesian network annotated with a set of function nodes. In this section, the Bayesian network RiBay annotated with function nodes F_{RiBay} is denoted $N_{RiBay} = (X_{RiBay}, G_{RiBay}, P_{RiBay}, F_{RiBay})$.

RiBay consists of three types of stochastic network fragments representing uncertain prices (both selling and buying), uncertain production of slaughter pigs (i.e., number of pigs produced and the weight gain), and uncertain yields of rape seed and grain. The model has a total of 196 nodes in G_{RiBay} of which 28 represent random variables. The complete structure of RiBay is shown in Figure 5. The color coding in the model is useful for documentation purposes. For instance, the blue function nodes are risk mitigation measures, dark yellow function nodes are numbers from the budget, dark red function nodes are system variables and green function nodes are forecast variables, e.g., prices on futures.

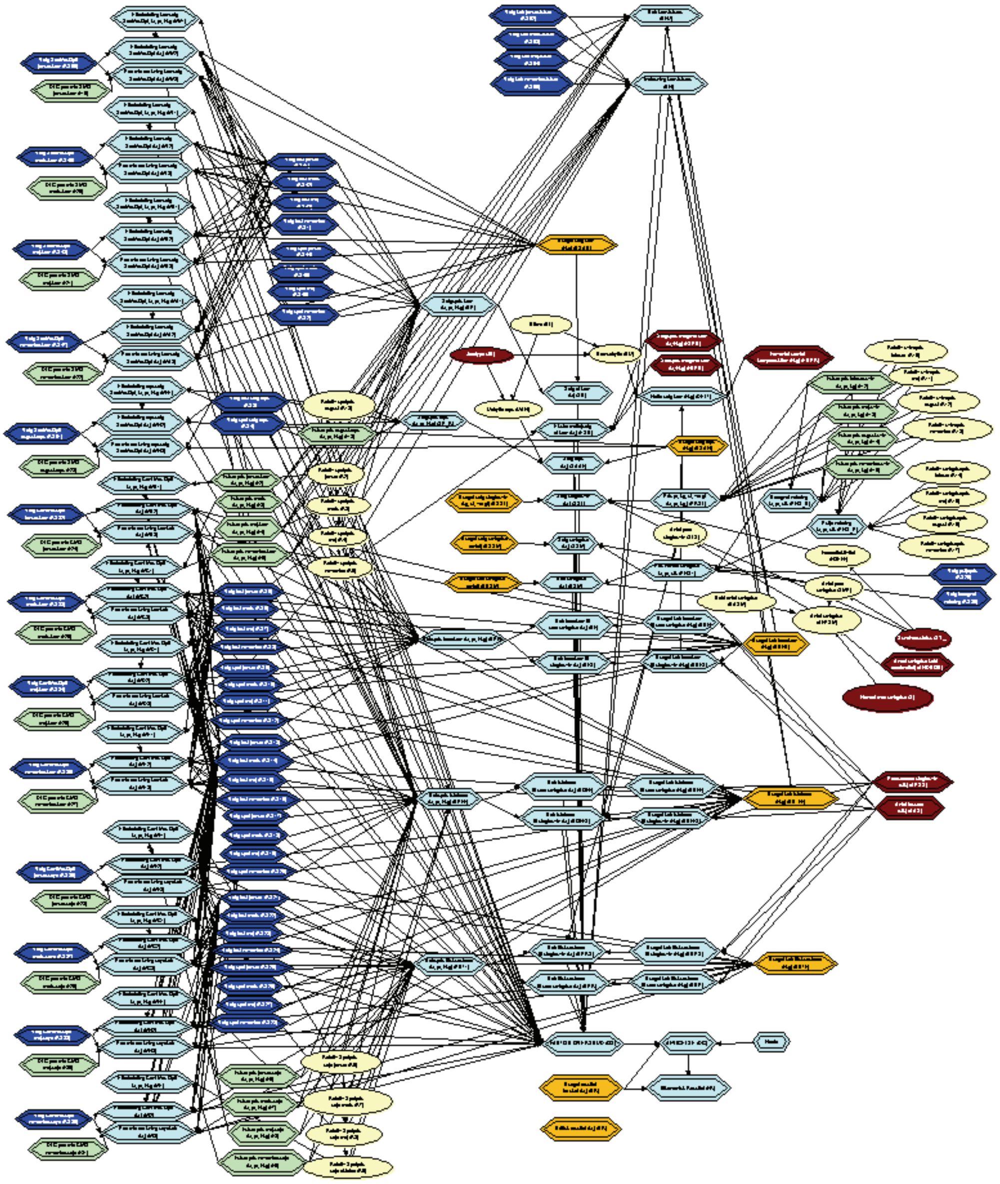


Figure 5. The complete structure of the Bayesian network RiBay

The three key stochastic network fragments of RiBay are illustrated in Figure 1, Figure 4 and Figure 6. Figure 6 shows two different encodings of price dependency in RiBay. The four nodes in the graph on the left hand side encode expected relative spot prices of grain in the future months of January, March, May and November. The dependency is encoded as a first-order Markov chain, i.e., the future is independent of the past given the present. On the other hand, the four nodes on the right hand side encode relative pig prices to be independent. These are simplifying assumptions. A main challenge in the construction of the model is the collection of high quality data available for quantification of the model.

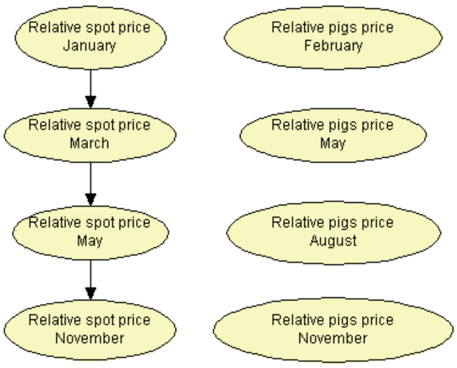


Figure 6. Representations of uncertainty in prices

The three types of stochastic network fragments are the main components of the model and responsible for encoding uncertainty in the risk factors represented as random variables in RiBay. The random variables are linked to a large set of function nodes F_{RiBay}. The set F_{RiBay} includes both nodes representing parameter values in CPDs and nodes representing values computed after belief update. For instance, the function node "*Share of piglets purchased*" represents a system variable entered by the user that parameterizes the mathematical expression for the CPD of "*Number of piglets*" while "*Economic income*" represents the economic values computed by the model.

By construction N_{RiBay} is a representation of a yearly budget at farm level. In RiBay the individual farm is characterized by the value of nine so-called *system variables* which are used to instantiate the model and to reflect the properties of the individual farm. Before applying RiBay, the farmer must prepare the farm budget for the coming year based on *expected prices and expected yields and production*. The expected prices of the individual inputs (feed) and outputs (grain, rape seed and pigs), and the expected production and input use of eight items of feed, pigs, grain and rape seed, are the input variables to be entered into RiBay by the user. Thus, the budget items form the baseline, and the state spaces of random variables therefore represent deviations from the figures used in the budget. The state values of the discrete random variables therefore represent deviations from the budget baseline, and determine the impact variations from the budget baseline have on one or more economic factors, such as revenue on sales of slaughter pigs. For instance, the variable *Yield of grain* has seven states 0.4, 0.6, 0.8, 1, 1.2, 1.4 and 1.6, where a value below 1 (one) means a negative deviation from the budget baseline and a value larger than one means a positive deviation from the budget baseline.

In the graph in Figure 5 expected prices are represented by the light green nodes and expected production and input use are represented by the dark yellow nodes.. In the present version of the model we use the assumption of rational price expectation, and therefore the expected prices are assumed to be the same for all farmers with values equal to the prices on the futures markets. This means that price risk is quantified by the deviation of spot prices from future prices, and production risk is quantified by the deviation of actual production from expected production.

4.2. Quantification

The quantification of the stochastic network fragments amounts to specifying the prior and conditional probability distributions P_{RiBay} as defined by the structure G_{RiBay} of N_{RiBay}. A Bayesian network is an excellent knowledge integration tool, as the elements of P_{RiBay} are specified independently and therefore may originate from completely different information sources. A Bayesian network can combine estimation from historical data with subject matter expert knowledge. In the cases where historical data is unavailable, subject matter expert knowledge is required to quantify the uncertainty represented by the CPDs.

The function nodes F_{RiBay} can be partitioned into four different subsets 1) risk mitigation measures, 2) figures from the budget (can be either input values or factors contributing directly or indirectly to farm income), 3) system variables and 4) price forecast values. Only forecast variables need to be updated on a regular basis.

The quantification of the CPDs of RiBay is based on both estimation from historical data and subject matter expert knowledge. The possible states of the random production and price variables are defined based on subject matter expert knowledge (data from experimental farms, farm accounts and price statistics). As the random variation refers to deviation from expected values applied in the budget, the state values of the random variables refer to deviations around the expected value of 1 (one). Thus, a state value of 0.84 indicates a value which is 16 % (1-0.84) lower the expected value applied in the budget.

Concerning feed and grain prices, the state values are based on statistics of the relation between future prices in October and the actual future price at the maturity data for the future in question. October is chosen to mimic the time where budgeting the income for the next year typically takes place. Thirteen years of price observations (2000-2012) were available, and the random intervals were covered using eight possible state values, using the smallest and the largest observation as the end points. This means that the model assumes that the future cannot be any worse than observed in the past 13 years.

The variability of pig prices were quantified based on the deviation of actual average prices from one year price forecasts. As with other random variables, eight state values were used with the highest and the lowest observation being used as end points.

Physical production data were quantified in a similar way to price data with six or eight possible state values for each variable. The state values were determined by using the local database of farm account data for an 18 year period (1990-2007). The panel data nature of these data made it possible to quantify the variability as the within-farm deviation from expected values. The farm account data used correspond to the farm account data available from the Farm Account Data Network (FADN) at the EU-level.

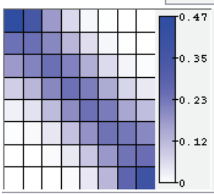
The estimation of the CPDs from data is based on maximum likelihood estimation. The discrete random variables representing *Relative spot prices* for grain have eight possible state values. This means that the CPD of *Relative spot prices* for one year given the previous year *Relative spot prices* has 64 entries and 56 free parameters. A reliable estimate of this CPD cannot be made from the historical data available. As the historical data sets are sparse compared to the number of parameters being estimated, we assumed a linear Gaussian distribution. The price data for grain is complete whereas the soya bean data have missing values.

The EM algorithm[11] as implemented in HUGIN software[13] has been applied to perform parameter estimation from historical data. The estimation of probability parameters for prices in RiBay thus proceeds as follows:

1. Raw historical data for each maturity date of the future is collected for the period from 2000-2012 covering a total 13 years.
2. States values are computed from the original data as described above.
3. Estimation of parameters in the networks representing the stochastic network fragments proceeds as follows:
    a. An initial network where variable state values are defined based on subject matter expert knowledge.
    b. Parameter estimation in a linear Gaussian model using EM algorithm.
    c. The estimated linear Gaussian distributions are transferred as mathematical expressions to a discrete Bayesian network model with discretized variables where the center of each interval equals one of the original state values (possibly extended to infinity).
    d. The CPDs generated in the previous step are transferred to the initial model with crisp state values (i.e., not intervals).

The process was iterated adjusting both the number of states and the state values.

Relative spot price March



Relative sp...	0.85	0.895	0.94	0.985	1.04	1.095	1.14	1.185
0.85	0.4656	0.3528	0.1729	0.0658	0.0191	0.0042	0.0007	0.0002
0.9	0.2446	0.2475	0.205	0.1248	0.0557	0.0183	0.0044	0.0016
0.945	0.1736	0.2119	0.2479	0.2131	0.1345	0.0623	0.0212	0.0099
0.995	0.0829	0.1233	0.2038	0.2473	0.2204	0.1443	0.0694	0.0404
1.045	0.0266	0.0487	0.1138	0.1951	0.2456	0.2271	0.1542	0.1104
1.09	0.0057	0.0131	0.0432	0.1046	0.186	0.2429	0.2329	0.203
1.14	0.0008	0.0024	0.0111	0.0381	0.0958	0.1766	0.2392	0.2511
1.19	8.6231E-5	0.0003	0.0022	0.0112	0.0428	0.1244	0.2781	0.3834

Figure 7. CPD for *Relative spot price March* given *Relative spot price January*

Figure 7 shows the resulting CPD for *Relative spot price March* given *Relative spot price January*. This CPD has been estimated from historical data as described above. The relative spot price is defined as the spot price of grain in March divided by the future price for the same commodity in October the previous year.

The model is parameterized by the prices valid at the time the budget is created. This includes future prices on, e.g., grain, rape seed and soybeans. This information should be updated each year. This calibration of the model is performed once before the risk management and analysis is performed.

As mentioned before, the model is quantified by numbers from the farmer's budget, system variables and forecast variables. Budget numbers and system variables are entered each time the system is used to perform a risk analysis. Forecast variables are updated, for instance, each year prior to budget preparation by the knowledge engineer responsible for maintenance of the Bayesian network RiBay.

4.3. Validation of Bayesian network

An important aspect of constructing a Bayesian network is validation. Validation is key to ensuring the high quality of the model, as well as ensuring continuous commitment from stakeholders and engaging key stakeholders in the process of implementing the use of a risk management system based on the Bayesian network RiBay. Validation was performed as described below.

The model was constructed based on subject matter expert knowledge. The three types of stochastic network fragments, i.e., the networks for yields, pig production and prices shown in Figures 1, 4 and 6 respectively, are fairly simple and intuitive. For the price networks assumptions on the correlation of prices at different time points were made. The remaining part of the structure consists of function nodes and their connections. The set of function nodes includes system variables, budget numbers, buying and selling strategy and the use of risk measures. These relations and the expressions associated with function nodes correspond to spreadsheet formulas.

The quantification of the model, including both state values and probability distributions, was validated through a combination of model review and the application of artificial as well as real world data to the model. The artificial data consisted of a set of six different types of example farms and corresponding budget numbers. Model review was performed by subject matter experts interacting with the model, while the validation of the model as well as its interface was performed by subject matter and technology experts using artificial data and through a series of on-site visits with real farmers. Specifically, validation has included up to four visits to a selected set of farms where the system was applied to farm-specific data. The expert defined example farms and the selected set of real-world farms have been developed and identified to assess coverage of the model as well as to validate model performance. The series of multiple on-site visits was a key element in the validation of the model and system.

The validation process also served to improve the quality of the model. A parameter sensitivity analysis has been used to assess the impact of parameter values and state values, i.e., states of discrete random variables. The results of the sensitivity analysis have been exploited to target further knowledge assessment on the elements that have the highest potential of improving the quality and usefulness of the model for risk management.

4.4. Bayesian network used for stochastic simulation

A Bayesian network $N=(X, G, P)$ is a representation of a joint probability distribution over a set of variables. Given that N_{RiBay} represents a joint probability distribution over random variables X_{RiBay}, it can be used to perform a Monte Carlo simulation over X_{RiBay} from P_{RiBay} given a set of evidence ε assuming a belief update operation has been successfully completed. For each simulation operation a configuration over X_{RiBay} is generated at random from P_{RiBay} given evidence ε , and subsequently the values of function nodes downstream from random variables are computed. This means that by sampling N times it is possible to generate an empirical probability distribution for each function node in F_{RiBay}. Based on the sample of N cases it is possible to compute a number of different statistics to support risk management at farm level. This means that RiBay enables us to perform stochastic simulation of risk management at the farm level. This includes simulating the consequences of implementing specific risk measures to control risk on farm income.

A user interface system has been developed on top of the Bayesian network RiBay to help the farmer understand his risk and to investigate the impact of different risk controlling initiatives. This includes assessment of risk and quantification of risk based on farm income.

5. Using the Bayesian network model for risk management

RiBay is still in the development phase. The intention is to further develop the stochastic specification of agricultural production and prices at the farms level, and to include more risk management tools. The model is currently being refined, and will be offered to farmers as a tool in fall 2013 to support risk management of their 2014 budget.

However, even the relatively simple version developed at this stage can demonstrate the large potential this type of model has for modelling risk management in agriculture. This is done using two cases below.

5.1. Two cases illustrating risk management

The demonstration model developed so far only includes tools for price risk management related to buying pig feed and selling grain. In relation to pig management it would make sense to include tools for pig price management as well. But until now such tools (pig futures) have not been applied in Denmark.

The following two cases demonstrate the use of RiBay for risk management:

Case 1. Slaughter pig production only

Purchase of piglets, number:	11,000
Produced slaughter pigs, number/kg:	10,500/861,000
Purchase of compound feed mixture, kg:	2,200,000
Income before tax, DKK:	1,500,000
Critical income, DKK:	400,000

Case 2. Sow and crop production

Number of sows:	400
Produced slaughter pigs, number/kg:	10,500/861,000
Grain production and sale, kg:	2,400,000
Rape seed production and sale, kg:	300,000
Purchase of compound feed mixture, kg:	3,200,000
Income before tax, DKK:	3,100,000
Critical income, DKK:	-200,000

Case 1 is a slaughter pig production based on piglets purchased at 30 kg, and the production and sale of slaughter pigs with a slaughter weight of 82 kg. Production is based solely on purchased feed (compound feed mixture). The stochastic variables are the number of slaughter pigs delivered to the slaughter house, slaughter weight (feed efficiency), the price of piglets, the price of slaughter pigs, and the price of compound feed, which in turn depends on the price of grain and soybeans. The farm has other income and expenses not considered here (assumed fixed), and the total expected net income before taxes is DKK 1,500,000. Critical income is DKK 400,000 and is defined as the minimum income needed for the farmer to meet his financial obligations.

Case 2 is a farm with both pig and crop production. The farm produces the same amount of slaughter pigs as in case 1, but slaughter pig production in this case is based on piglets produced by the farmer himself from his 400 sows. The farmer also produces 2,400 tons of grain and 300 tons of rape seed. The grain could have been used for pig production, but instead the farmer sells all the grain and buys compound feed for the pigs. The rape seed is also sold. The stochastic variables are the number of slaughter pigs delivered to the slaughter house, slaughter weight (feed efficiency), the price of slaughter pigs, the price of compound feed (which in turn depends on the price of grain and soybeans), the yield of grain, the yield of rape seed, the price of grain and the price of rape seed. The farm has other income and expenses not considered here (fixed), and the total expected net income before taxes is DKK 3,100,000. Critical income in this case is DKK -200,000.

All the stochastic variables mentioned may be controlled by applying appropriate risk management tools. As the Bayesian network model developed so far only allows for the risk management of feed and crop prices, these are the only variables that will be the subject of risk management in the following.

In both cases, the stochastic compound feed prices are controlled by three alternative risk management tools: a) buy grain and soybean futures, b) buy DLG products and c) buy feed on forward contracts. The first tool, futures, is a well-known tool used for hedging. The second tool is a relatively new price risk management tool developed and marketed by DLG[3], a large Danish agricultural company trading grain and feed. This tool combines forward contracts and options so that in effect only the “down-side” risk is removed. The third tool, forward contract, refers to trading at fixed prices.

The results of risk management for a random sample of 40,000 observations are shown in Table 1 (case 1) and Table 2 (case 2). The consequence of risk management is a change in the probability distribution of income before taxes. Instead of presenting the complete probability distribution, we have chosen to illustrate the consequences using a few key numbers, i.e., expected income before taxes, the standard deviation of income before taxes, and the probability of earning an income before taxes (FR) which is less than the critical value (KR).

Table 1. Results. Pig production only (Case 1)

	Expected (DKK)	STD (DKK)	P(FR<KR)	dE/dP
No price risk management	1,498,346	656,424	0.046	-
Grain Futures	1,468,898	589,192	0.030	19,958
ConfMaxOpti	1,423,803	568,108	0.032	55,155
Forward contract	1,411,669	562,937	0.032	66,440

The “No price risk management” row shows the results when no risk management tools are involved. The first column shows an expected income before taxes of DKK 1,498,346 with a standard deviation (STD, second column) of DKK 656,424. The probability of an expected income below the critical level of DKK 400,000 (P(FR<KR)) is 0.046 (4.6 %).

The second row shows the effect of using futures to fully hedge the trade of grain involved when buying compound feed. The uncertainty is reduced in the sense that the standard deviation decreases to DKK 589,192 and the probability of earning less than DKK 400,000 before taxes is reduced from 0.046 to 0.030. However, this risk reduction comes at the cost of DKK 29,448, which is the decrease in expected income before taxes (1,498,346 - 1,468,898).

The ConfMaxOpti product in the third row (the DLG risk management product) reduces the standard deviation even further, namely to DKK 568,108. However, the cost also increases in the sense that expected income decreases further to DKK 1,423,803. The probability of earning less than DKK 400,000 before taxes is, however, only reduced to 0.032, which is higher than hedging with futures.

The last risk management tool in Table 1, forward contracts, is the most expensive measured by the decrease in expected income before taxes. The standard deviation is also the lowest, but the probability of earning less than DKK 400,000 before taxes is still 0.032 as with the DLG product.

If critical value is the main focus of risk management, one could calculate the cost of reducing the probability of earning an income below the critical level. The result of this calculation is shown in the last column of Table 1 in which dE is the change of expected income, and dP is the change of probability that income will fall below the critical level KR. The cost is the lowest when using futures (second row), which costs only DKK 19,958 of expected income per unit of risk reduction measured as change in the probability of earning an income below the critical level. ConfMaxOpti has a price of DKK 55,155, and forward contracting is here the most expensive with a price of DKK 66,440.

The results of case 2 are shown in in Table 2.

Table 2. Results. Pig and crop production

	Expected	STD	P(FR<KR)	dE/dP
No price risk mgt.	2,939,640	1,658,612	0.038	
Grain futures	2,893,357	1,681,216	0.038	-
ConfMaxOpti	2,827,312	1,655,304	0.040	-
Forward contract	2,808,876	1,669,393	0.041	-

If we look at the results in the first column, the relationship between the numbers is as in the first case, i.e., the cost of risk management is a reduction in expected income before taxes. However, in this example, risk management does not pay. The probability of earning an income before taxes less than the critical level stays the same or even increases (from 0.038 to 0.041 in the case of forward contracts). Also, the standard deviation increases in two of the three cases.

In Table 2 the cost of risk management in the form of expected income per unit of risk reduction measured as the change in the probability of earning an income below the critical level is not calculated. The reason for this is that risk management simply does not pay. For instance, using forward contracts would cost DKK 130,764 and would not reduce the probability of earning an income before taxes less than the critical level.

The reason for this 'strange' result is that the farmer in case 2 already performs a considerable amount of risk management because he both sells grain and buys grain (part of compound feed). This means that there would be no benefit from controlling grain price variation, because potential losses when selling at lower prices are compensated by lower prices when buying grain.

5.2. Further development of RiBay

The Bayesian network model RiBay developed so far is a model of a pig farm including crop production. The intention is to further develop the model to include more risk factors and their interaction, also over time, and to include additional risk management tools. The intention is also to develop similar models for other farm types, for instance specialized dairy farms and crop farms.

An important part of future work includes extending the model to capture the dynamics within a fiscal year and between fiscal years. In this context it is important to differentiate between risk management at the strategic, tactical and operational levels. The present version of RiBay focuses on risk management at the tactical level, in the sense that the objective refers to the fiscal year income, and the risk management tools refer to the objective of controlling prices and price variability within the year.

The model developed so far only includes tools for price risk management related to buying pig feed and selling grain. One would expect that a pig farm model would also include tools for controlling uncertain pig prices. However, when it comes to prices for purchase and sale of pigs and pork there is no real tradition of risk management in Denmark. It is not possible to trade future contracts, neither on purchase nor sale of pigs, and markets for trade of pig or pork futures and options do not exist. Such exchanges exist in other countries (for instance Amsterdam and Frankfurt), but transaction costs are apparently so high that it is not a realistic option for Danish farmers.

While the obvious focus at the tactical and operational levels are the tools for price risk management, there are other risk management tools available for risk management at the farm level. Choice of products (diversification), choice of production methods, insurance, weather derivatives, level of financial reserves, etc., are all examples of options for risk management - even at the tactical level. The plan is to extend the model by including these risk management tools. In this context time becomes an important parameter, and a dynamic version of the model is an obvious objective for further work.

The challenge does not lie in the construction of the model. The real challenge is to collect data or information of sufficient quality to quantify the model. In this context, the databases of farm account statistics in the EU farm accountancy network (FADN) may prove to be a valuable asset as the basis for further research.

6. Summary and conclusion

In this paper we have demonstrated that the Bayesian network model can be a useful tool for risk management. The value of the Bayesian network model as the platform for building decision support systems for risk management in agriculture is due to the flexibility of the model, including the ability to model individual risk factors and their interaction, but also the way in which historical evidence can be systematically utilized to estimate the model parameters.

In a Bayesian network data gaps and uncertainties are quantified as conditional probability distributions. The flexibility of the approach means that the model can be updated with new information as it becomes available. The model can be updated by performing local changes only. The updating and refinement of the model can be implemented as a continuous process. Due to the intuitive and modular nature of the RiBay model, it is a straightforward process to update the model with new data and information and extend the model with, for instance, new risk mitigation measures.

A decision support system based on the Bayesian network RiBay has the potential to provide consistency in the risk management process at farm level. Its intuitive, graphical and flexible nature is an important property in the implementation of a risk management solution at farm level. A solution based on the use of Bayesian networks has the ability to link uncertainty from different external sources to budget figures and to quantify the risk at farm level.

As with any other model, a Bayesian network is a finite representation of a complex world based on a set of assumptions. A key strength of a Bayesian network is the transparency of assumptions. This allows its users to identify the information included in the model, as well as the information excluded from the model.

The ability to perform belief update and stochastic simulation of a Bayesian network annotated with function nodes makes it an excellent tool for identifying the potential impact a random variable, system variable or risk measure has on the farm income. This makes it possible for the user to interrogate the model to understand risk and identify the most influential risk measure.

Acknowledgment

We want to thank the “Norma and Frode S. Jacobsen Foundation” and “The Nordea Bank Foundation” for their financial support of this research project.

References

1. Bielza, M., A. Garrido and J. M. Sumpsi (2007), "Finding optimal price risk management instruments: the case of the Spanish potato sector", *Agricultural Economics*, Vol . 36, pp. 67-78.
2. Chapman, R. J. (2011), *Simple tools and techniques for enterprise risk management*, 2nd ed., John Wiley and Sons Inc., Chichester, UK.
3. DLG, *Risk Management*, [Online, Cited 08/05/2013], URL: http://www.dlg.dk/da/riskmanagement/
4. EU (2005), *Risks and Crisis Management in Agriculture*, Report IP/B/AGRI/ST/2005-30 prepared for the European Parliament, University of Napoli Federico II, Napoli.
5. Hardaker, J. B. (2006), "Farm Risk Management: Past, Present and Prospects", *Journal of Farm Management*, Vol. 12, pp. 593-612.
6. Hardaker, J. B., Huirne, R. B. M., Anderson, J. R. and Lien, G. (2004), *Coping with Risk in Agriculture*, 2nd ed. CABI Publishing, Wallingford, UK.
7. Hubbard, D. W. (2009), *The Failure of Risk Management: Why it's broken and how to fix it*, John Wiley and Sons Inc., Hoboken, New Jersey, USA.
8. ISO (2009), *ISO 31000 / 2009 Risk Management: Principles and Guidelines on Implementation*, International Organization for Standardization, Geneva.
9. Jorion, P. (1997), *Value at Risk: The New Benchmark for Controlling Market Risk*, Irwin-McGraw Hill: Illinois.
10. Kjærulff, U. B. and Madsen, A. L. (2013), *Bayesian Networks and Influence Diagrams – A Guide to Construction and Analysis*, 2nd edition, New York: Springer Publishing Company.
11. Lauritzen, S. L. (1995), "The EM algorithm for graphical association models with missing data", *Computational Statistics & Data Analysis*, Vol.19, pp.191-201.
12. Madsen, A. L., Jensen, F., Kjærulff, U. B. and Lang, M. (2005), "HUGIN - The Tool for Bayesian Networks and Influence Diagrams", *International Journal of Artificial Intelligence Tools*, Vol.14, pp. 507-543.
13. Madsen, A. L., Lang, M., Kjærulff, U., and Jensen, F. (2003), "The Hugin Tool for Learning Bayesian Networks", *Proceedings of The Seventh European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pp. 549-605.
14. McGrayne, Sharon Bertsch (2011), *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted DownRussian Submarines, and Emerged Triumphant from Two Centuries of Controversy,* Yale University Press.
15. Nartea, G. and Webster, P. (2008), "Should farmers invest in financial assets as a risk management strategy? Some evidence from New Zealand", *The Australian Journal of Agricultural and Resource Economics*, Vol. 52, pp.183 – 202.
16. OECD (2009), *Managing Risk in Agriculture: A Holistic Approach*, OECD, Paris.
17. Schaffnit-Chatterjee, Claire (2010), *Risk Management in Agriculture*, Deutsche Bank Research, Frankfurt.