Production, Manufacturing and Logistics

Causal modeling alternatives in operations research: Overview and application

Ronald D. Anderson [a,1], Gyula Vastag [b,*]

[a] *Kelley School of Business, Indiana University, 801 West Michigan Street, BS4053, Indianapolis, IN 46202-5151, USA*
[b] *Kelley School of Business, Indiana University, 801 West Michigan Street, BS4027, Indianapolis, IN 46202-5151, USA*

Abstract

This paper uses the relationships between three basic, fundamental and proven concepts in manufacturing (resource commitment to improvement programs, flexibility to changes in operations, and customer delivery performance) as the empirical context for reviewing and comparing two casual modeling approaches (structural equation modeling and Bayesian networks). Specifically, investments in total quality management (TQM), process analysis, and employee participation programs are considered as resource commitments. The paper begins with the central issue of the requirements for a model of associations to be considered causal. This philosophical issue is addressed in reference to probabilistic causation theory. Then, each method is reviewed in the context of a unified causal modeling framework consistent with probabilistic causation theory and applied to a common dataset. The comparisons include concept representation, distribution and functional assumptions, sample size and model complexity considerations, measurement issues, specification search, model adequacy, theory testing and inference capabilities. The paper concludes with a summary of relative advantages and disadvantages of the methods and highlights the findings relevant to the literature on TQM and on-time deliveries.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Manufacturing; TQM; Delivery performance; Causal modeling; Structural equation modeling; Bayesian networks

1. Introduction

Interest in causal modeling methodologies in the social sciences stems from the desire to establish patterns of regularities or laws analogous to those in the physical sciences. A fundamental appeal of causal modeling is the ability to combine cause–effect information, based on theoretical construction, with statistical data to provide a quantitative assessment of relationships among the studied variables. The purposes for employing causal modeling in the study of operations are to develop an explanation of relationships and to provide a basis for inference. The portrayal, evaluation and summarization of assumed causal relationships are the components of explanation. These relationships are then used to develop

* Corresponding author. Tel.: +1-317-278-3681; fax: +1-317-274-3312/278-3312.
E-mail addresses: randers@iupui.edu (R.D. Anderson), gvastag@iupui.edu (G. Vastag).
[1] Tel.: +1-317-274-2446; fax: +1-317-274-3312.

inferences for diagnostic reasoning from effects to causes and for the prediction of outcomes that would follow from a policy or procedure intervention. Available modeling methods offer differing functional advantages and limitations. However, any method should have potential managerial usefulness by providing outputs with clear interpretation and the capability to assess the impact of potential changes in the modeled process.

Ideally, a causal study would take the form of a randomized controlled experiment conducted over an appropriate time period. Such a research design would minimize construct, internal, external, and statistical threats to validity (Cook and Campbell, 1979), and allow the possibility of causal conclusions to be reached. Unfortunately, randomized controlled experiments can seldom, if ever, be utilized to provide causal knowledge for strategy and policy issues. Thus, causal modeling methods for non-experimental data are of interest.

Bayesian networks and structural equation models (SEM) are the causal modeling methods for non-experimental data reviewed and compared in this paper. The paper begins with the central issue of the requirements for a model of associations to be considered causal. This philosophical issue is addressed in reference to probabilistic causation theory. Then, each method is reviewed in the context of a unified causal modeling framework consistent with probabilistic causation theory, and applied to a common dataset. The comparisons include concept representation, distribution and functional assumptions, sample size and model complexity, measurement, specification search, model adequacy, theory testing and inference capabilities. The paper concludes with a summary of the relative advantages and disadvantages of the methods.

2. Probabilistic causation theory

The area of causation has been extremely active over the past twenty years with numerous interactions between the fields of philosophy, statistics and computer science. This activity has spawn spirited controversy on a wide variety of concep-

tual and methodological issues (McKim and Turner, 1997). Causality, as a theoretical postulate, has been the subject of highly contested discussions since the reductive account offered by Hume (1969). Hume characterized causation by the regularity of constantly conjoined pairs of events (Effect $= f$(Cause)), under conditions of temporal priority (a cause must precede an effect), and contiguity (a cause is temporally adjacent to an effect). However, Hume's account does not provide for imperfect regularities nor does it have the ability to distinguish between a genuine causal relation and a spurious association. These weaknesses motivated development of theories of causation that cast causal relationships between general events in terms of stochastic descriptions (Suppes, 1970).

The key feature of probabilistic causation is a paradigm switch from the absolute determination of an effect due to the occurrence of a cause to the occurrence of a cause increasing the probability of an effect. An assumption underlying this perspective is that incomplete knowledge of causes results in uncertain cause–effect relationships. This conceptualization, labeled as pseudo-indeterminism (Spirtes et al., 1993), assumes that specified causes do not alone determine an effect, but do so in conjunction with unspecified unobserved causes. Thus, pseudo-indeterminism assumes that sets of independent specified causes and unspecified causes are the direct causes ($\rightarrow$) of an effect: specified causes $\rightarrow$ effect $\leftarrow$ unspecified causes.

Cause–effect relationships, under the assumption of pseudo-indeterminism, may be encoded into a graphical structure known as a directed acyclic graph or simply a DAG. Each arrow in a DAG depicts causal dependence and the absence of a connecting arrow indicates causal independence. The encoded structure is characterized as directed, since two-headed arrows depicting noncausal association are not allowed and as acyclic, since feedback loops (e.g., $X \rightarrow Y \rightarrow X$) are not allowed.

The common cause principle states if two variables in a population are associated and neither is a cause of the other, they must share a common cause (Reichenbach, 1956). The term association is used, in reference to probabilistic dependence

$p(X, Y) > p(X)p(Y)$, rather than the narrower term correlation that implies a measure of linearity. When an event C is a common cause of X and Y, $X \leftarrow C \rightarrow Y$, or is intermediate to X and Y, $X \rightarrow C \rightarrow Y$, C is said to screen off values of X from values of Y resulting in probabilistic independence $p(X, Y|C) = p(X|C)p(Y|C)$. Thus, a causal system with asymmetric relations is assumed to always exist in reference to population associations, but usually is not known with certainty and must be inferred from sample data.

The common cause principle is the justification for the causal Markov condition that states every effect variable, conditional on its direct causes, is independent of all variables that are not its effects (Spirtes et al., 1993). The causal Markov condition can also be expressed in probability statements: if X does not cause Y, then $p(X|Y \& \text{ direct causes of } X) = p(X|\text{direct causes of } X)$. The notion of an underlying causal mechanism generating observations implies the converse of the causal Markov condition. That is, empirical regularities of conditional independence relations observed from a population are due to a causal structure not coincidence (Scheines, 1997). This assumption, labeled as the stability condition (Pearl, 2000) or the faithfulness condition (Spirtes et al., 1993), states probabilistic independencies are a stable result of causal structure and not due to happenstance or specific parameter values. Therefore, the joint population probability distribution over a defined variable set is assumed stable or faithful to the underlying causal structure as specified in the DAG. Lastly, if the variable set includes all relevant common causes, it is said to be causally sufficient. Assertions portrayed by a DAG are assumed to be causal when combined with causal sufficiency, the causal Markov and faithfulness conditions, and independence of specified and unspecified causes.

SEM and Bayesian networks are viewed as causal models when the above conditions are satisfied. Druzdzel and Simon (1993) demonstrate a Bayesian network can be represented by a simultaneous equation model with hidden variables and address causal interpretation in the context of the underlying principles of SEM. Pearl (1995) provides a detailed exposition of causal diagrams in empirical research and emphasizes inferences rest on causal assumptions.

The above causation account is not universally accepted in the philosophy of science, even by those that adopt a probabilistic viewpoint (Cartwright, 1997). There is an often-cited exception to the common cause principle and the causal Markov condition in the microscopic world of quantum mechanics (Hausman and Woodward, 1999). At issue in this environment is the behavior of complementary pairs of particles that are not probabilistically dependent on one another or the effects of a common cause. Yet under manipulation of one particle, the other instantly behaves in an identical manner. Various elements of this paradox have been debated in physics and philosophy since publication of the Einstein–Podolsky–Rosen thought experiment (Einstein et al., 1935). The implications of these debates are that the probabilistic behavior of some phenomena do not conform to the pseudo-indeterminism perspective, but is due to inherently stochastic properties in Nature.

2.1. A causal model

A causal model may be expressed as $M = \{S, \boldsymbol{\Theta}_S\}$, where S is the structure of the causal assertions of the variable set V portrayed by a DAG and $\boldsymbol{\Theta}_S$ is a set of parameters compatible with S. A DAG can always be translated into a set of recursive structural equations with independent errors that satisfies the causal Markov condition (Kiiveri and Speed, 1982). A set of recursive structural equations that describe the data-generating process of a DAG is specified by

$$V_i = f_i(\text{Parents}(V_i), U_i) \qquad (1)$$

where V_i is a consequence variable linked by a function f_i to a configuration set of direct causes, $\text{Parents}(V_i)$, and U_i, an error term (Pearl, 2000). Each V_i is represented by an individual equation that corresponds to a distinct causal mechanism where the function f_i is invariant over a range of values of $\text{Parents}(V_i)$ and U_i. The error terms are assumed to represent mutually independent unobserved variables, each with a probability distribution function $p(U_i)$. Eq. (1) reflects the view that

Nature possesses stable causal mechanisms described by f_i that are deterministic functional relationships between variables, while $p(U_i)$ reflects incomplete causal knowledge associated with pseudo-indeterminism: $\mathrm{Parents}(V_i) \rightarrow V_i \leftarrow U_i$. The specification of Eq. (1) can accommodate non-linear relationships, expressed in terms of the estimated probability of an effect, as well as the typical linear specification.

The parameters in $\mathbf{\Theta}_S$ assign a function f_i to each V_i and a probability measure $p(U_i)$ to each U_i. A structure S is taken to portray a set of conditional independence assertions among the m variables under study, which permits the factorization of the associated joint probability distribution as

$$p(V_1, V_2, \ldots, V_m) = p(V_1|\mathrm{Parents}(V_1))p$$
$$\times (V_2|\mathrm{Parents}(V_2))\ldots p$$
$$\times (V_m|\mathrm{Parents}(V_m)) \qquad (2)$$

The factorability condition follows from the causal Markov condition (Hausman and Woodward, 1999).

3. Structural equation models (SEM)

The true score model (Spearman, 1904), common factor analysis (Thurstone, 1935), and statistical factor analysis (Lawley and Maxwell, 1963) provide the foundation to represent the within-concept measurement model in SEM. This representation assumes a measured variable MV_i is caused by two unrelated latent variables; a specified common cause representing concept j, LV_j, and an error term, e_i, for unspecified causes: $\mathrm{LV}_j \rightarrow \mathrm{MV}_i \leftarrow e_i$. The general measurement equation of an indicator, a specialization of Eq. (1), is given by

$$\mathrm{MV}_i = \lambda_{ij}\mathrm{LV}_j + e_i \qquad (3)$$

where λ_{ij} is a path coefficient linking common cause j to measure variable i and e_i is the measurement error.

Path analysis (Wright, 1934) is generally acknowledged as the common parent of SEM (Jöreskog, 1973) and Bayesian networks (Pearl, 1988). Wright's method of path coefficients consists of developing a graphical representation assumed to describe a causal process, then decomposing the correlation coefficient for each pair of variables (V_i, V_j) into a sum of products of path coefficients and residual correlations. The concept equations are an application of path analysis that specializes Eq. (1) into the standard linear equations:

$$\mathrm{LV}_j = \Sigma b_{jk}\mathrm{LV}_k + u_j \qquad (4)$$

where b_{jk} is a path coefficient, u_j is the disturbance, and the index k ranges over all parents of LV_j. Thus, the variable set for SEM contains a set of measured variables and two sets of latent variables, $V = \{\mathrm{MV}, \mathrm{LV}, U\}$, where $U = \{e_i, u_j\}$.

The distribution assumption for SEM is a normal probability density under maximum likelihood estimation, which results in normal conditional probability densities for each LV_j and a multivariate normal joint distribution for the measured variables. Estimation by generalized least-squares or asymptotic distribution free methods are not constrained by the normality assumption, but implementation problems has resulted in sparse usage. The set of matrix parameters consistent with Eqs. (3) and (4) is $\mathbf{\Theta}_S = \{\mathbf{\Lambda}, \mathbf{B}, \mathbf{\Psi_u}, \mathbf{\Theta}_e\}$, where $\mathbf{\Lambda}$ is a matrix of measurement coefficients, $\mathbf{B}$ is a matrix of concept path coefficients, $\mathbf{\Psi_u}$ is a covariance matrix of concept disturbances, and $\mathbf{\Theta}_e$ is a covariance matrix of measurement errors. The measured variable matrix equation is established by substituting Eq. (4) into Eq. (3), $\mathbf{MV} = \mathbf{\Lambda}(\mathbf{I} - \mathbf{B})\mathbf{u} + \mathbf{e}$. The parameter matrices of $\mathbf{\Theta}_S$ can be combined to form the covariance of the multivariate normal joint distribution of the measured variables, $\mathbf{\Sigma} = E[\mathbf{MVMV}'] = \mathbf{\Lambda}(\mathbf{I} - \mathbf{B})\mathbf{\Psi_u}(\mathbf{I} - \mathbf{B}')\mathbf{\Lambda}' + \mathbf{\Theta}_e$.

The objective of the parameter estimation, usually based on the maximum likelihood criterion, is to reproduce the sample covariance as closely as possible with a covariance matrix implied by the structure of measurement and structural assertions. The overall fit of the model is assessed by significance testing of the discrepancy function formed from the differences between the implied covariance matrix and the sample covariance matrix, and by descriptive indexes. Also, unconstrained matrix elements of $\mathbf{\Theta}_S$ can be tested

for significance. Complete details are provided in Bollen (1989).

4. Bayesian networks

An influence diagram provides a graphical scheme for representing conditional dependencies in a decision-making framework (Shachter, 1986). Such diagrams model uncertain knowledge, decisions and utilities to assess the actions that will yield the highest expected utility. Bayesian networks are a subset of influence diagrams that focus on the uncertain knowledge component. A discrete Bayesian network (Pearl, 1988) is a specialization of $M = \{S, \Theta_S\}$, where the structure S implies a set of conditional probability distributions

$$
\begin{aligned}
\Theta_S = \{ &p(V_1|\mathrm{Parents}(V_1), \theta_1), p \\
&\times (V_2|\mathrm{Parents}(V_2), \theta_2), \ldots, p \\
&\times (V_m|\mathrm{Parents}(V_m), \theta_m) \}.
\end{aligned}
$$

Each variable has c_i discrete values or states, and each θ_i is assumed to be a collection of multinomial distributions, one for each parent configuration. The probability assignment may be subjective or based on frequency ratios from a database or a combination of both. The associated joint probability distribution of the network variables follows Eq. (2) as the product of the conditional probabilities in Θ_S, $p(v_1, \ldots, v_m) = \Pi_m p(V_i|\mathrm{Parents}(V_i), \theta_i)$. Conceptually, the framework for Bayesian networks is applicable to continuous or discrete variables or both. However, applications have been concerned with discrete Bayesian networks where the vast majority of computational work has been focused.

The Bayesian network methodology has been closely associated with causation in philosophy (McKim and Turner, 1997) artificial intelligence (Pearl, 1988) and knowledge discovery (Spirtes et al., 1993). In addition, applications have been made in agriculture, computer imaging, computer software, education, information retrieval, medicine, space research and weather forecasting (Jensen, 1996). Cooper (1999) and Heckerman (1999) provide technical reviews of Bayesian networks.

Two approaches have been employed to evaluate the independencies and dependencies in the graphical structure S: an independence-testing method (Spirtes et al., 1993) and a Bayesian scoring method (Cooper and Herskovits, 1992). The independence-testing approach is consistent with traditional statistical methods utilizing likelihood ratio significance tests and maximum likelihood estimation constrained to satisfy to conditional independencies implied by the data. Whereas, the Bayesian approach employs a scoring metric based on the probability of S in the context of conjugate analysis (Bernardo and Smith, 2000). Both approaches allow incorporation of prior knowledge with the search algorithms to specify S.

4.1. Independence-testing approach

The frequentist viewpoint, where a parameter $\theta_i = p(U_i)$ is assumed as an unknown fixed quantity and the estimator of θ_i is a random variable, underlies the independence-testing approach. In general, classical test statistics can be utilized under the assumption the data results from multinomial sampling. The validity of the claim that X causes Y, with no screening off a common cause or intermediate variable, implies the population probability statement $p(Y|X) \neq p(Y)$. This causal assertion can be evaluated by sample tests of independence employing the maximum likelihood chi-square test statistic, G^2. Rejection of the null hypothesis of independence of X and Y,

$$H_0 : p(Y|X) = p(Y),$$

provides evidence that $p(Y|X) \neq p(Y)$ in support of the causal claim that $X \to Y$. When an common cause or intermediate screening variable is postulated as a claim that X and Y are independent conditional on C, i.e., C screens off X from Y, a sample test of conditional independence of X and Y given C will provide evidence for support or rejection of the claim.

The approach is a two-step procedure implemented by the PC algorithm of TETRAD II (Scheines et al., 1994), which first tests conditional independence relationships in a saturated structure or a structure constrained by prior knowledge. A

sequential testing procedure is employed to eliminate dependency linkages at a specified significance level. Thus, a reject or retain decision of conditional independence is obtained for each tested linkage. The rejected independence relationships are taken as dependency linkages and assembled to provide the graphical structure. The output of this assessment is a pattern that represents a set of equivalent DAGs, rather than a unique set of directional dependencies, that entail the same independence relations consistent with user-provided background knowledge. Search algorithms other than the PC algorithm are available for evaluating causal assertions (Spirtes et al., 1993).

In the second step, maximum likelihood estimates of the conditional probabilities $\mathbf{\Theta}_S$ are constrained to conform to the structure of the first step. When a database is used for estimation, the conditional probabilities are relative frequency ratios that are calculated as the joint frequency of V_i and its parents divided by the frequency of the parents.

4.2. Bayesian scoring approach

In the alternative Bayesian approach, the estimator of θ_i is assumed fixed, conditional on the data D, and the parameter θ_i is viewed as a random variable. The prior probability distribution of θ_i, $p(\theta_i)$, describes what is known about θ_i without knowledge of the data. The posterior probability distribution of θ_i, $p(\theta_i|D)$, describes what is known about θ_i with knowledge of the data. The relationship between the posterior distribution and the prior distribution is provided by Bayes' theorem (Bayes, 1763), $p(\theta_i|D) = p(D|\theta_i)p(\theta_i)/p(D)$, where $p(D|\theta_i)$ is the likelihood of the data given θ_i.

The Bayesian approach assumes that the structure S and the conditional probabilities in $\mathbf{\Theta}_S$ are stochastic variables with prior probability distributions provided from prior knowledge. Bayes' theorem provides the mechanism to revise the prior distribution of a model M, given the data D, to the posterior distribution, $p(M|D) = p(M, D)/p(D) = p(M)p(D|M)/p(D)$. The $p(D)$ does not change over the values of M, and can be viewed as a normalizing constant needed to scale $p(M)p(D|M)$ to sum to unity over all outcomes of M. Thus, the posterior distribution of M is proportional to the product of the prior distribution of M and the likelihood of the data D given M, i.e., Posterior Distribution $\propto$ Prior Distribution $\times$ Likelihood of the Data or $p(M|D) \propto p(M)p(D|M)$.

Prior knowledge of $\mathbf{\Theta}_S$ is usually very minimal or non-existent and the estimation of the conditional probability distributions relies primarily on available data. The Bayesian approach uses the proportionality relationship, Posterior Distribution $\propto$ Prior Distribution $\times$ Likelihood of the Data, to estimate the maximum posterior probability for variable i in state k for parent configuration j. The estimation task is greatly simplified when the prior and posterior distributions have the same functional form, which is termed as being conjugate. The family of Dirichlet distributions is conjugate for multinomial sampling and provides the basis for prior-to-posterior analysis in many computational programs (Cowell et al., 1999, appendix A). A Dirichlet prior distribution of θ_i when combined with the data frequency counts yields a Dirichlet posterior distribution (Ramoni and Sebastiani, 1999).

When the situation of initial ignorance exists, the prior probability of θ_i is generally taken as a uniform distribution (Geiger and Herkerman, 1997). Parameter specification of the prior distribution provides a summary of prior experience, and is specified by a quantity called the prior distribution precision. The value of the precision necessary for uniform priors is judgmental, and conceptually considered the number of cases representing prior experience, often termed as the equivalent sample size (Winkler, 1967). An equivalent sample size of 1 indicates the lowest level of confidence in prior estimation.

4.3. Combining independence-testing and Bayesian scoring methods

Although the philosophical positions underlying the independence-testing and Bayesian scoring methods are very different, there has been an effort to combine the two approaches in developing Bayesian networks. The PC algorithm, with large samples, will recover all causal relationship from

observational data assuming the causal Markov and faithfulness conditions, causal sufficiency, independent sources of variation, and valid statistical testing (Cooper, 1999). Scheines (1999) adopted a hybrid approach of first using independence-testing for construction of causal model structure S. Then he employed the Bayesian scoring method, using the multinomial-Dirichlet distribution to estimate Θ_S. This hybrid approach is adopted in the following application.

5. A manufacturing application: Concepts, objectives and data

A study of the relationships between three concepts (resource commitment to improvement programs, flexibility to changes in operations and customer delivery performance) provides the empirical context for reviewing and comparing the Bayesian network and SEM methodologies. The comparison and tests of the two approaches used data from the second round survey of the Global Manufacturing Research Group (GMRG). Whybark and Vastag (1993) have the questionnaire used in the survey and it provides details of this global data gathering effort primarily focusing on two industries: small machine tools and non-fashion textile manufacturing. A brief overview of this project is provided in Appendix A.

The variables associated with resource commitment are programs in which the responding company had invested money, time and human resources in the past two years. Three specific programs were measured on five-point scales ($1 =$ none to $5 =$ a large amount): Total quality management (labeled TQM), process analysis (labeled Analysis) and employee participation (labeled Involve). The concept flexibility to change was measured by flexibility to change product (labeled Product) and flexibility to change output volume (labeled Volume). Customer delivery performance was measured by delivery speed (labeled Speed) and delivery as promised (labeled On-Time). The flexibility and performance variables were measured on five-point scales in reference to competitors ($1 =$ far worse than competitors to $3 =$ about the same as competitors to $5 =$ far

better than competitors). However, scale responses 1 and 2 were combined due to extreme sparseness in the lowest scale category.

The general goals of causal modeling, explanation and inference, are translated into three objectives specific to the application. The first objective is to provide an adequate explanation of how resource allocation for performance improvement influences flexibility and delivery performances. The concepts are assumed to have the following temporal ordering: resource commitment precedes flexibility to change and delivery performance, and flexibility to change precedes delivery performance. Potential explanations, at the concept level, include a common effect model where resource commitment is a direct cause of flexibility to change and delivery performance, and flexibility to change is a direct cause of delivery performance. An alternative is a mediation model where flexibility to change is a mediating variable between resource commitment and delivery performance.

The second objective is to utilize the selected model to develop predictions of observable performance outcomes given assumed resource allocation levels in the TQM program. The final objective is to provide diagnostics from the selected model for assessment of relative changes in variables when an intervention manipulates the state of the On-Time variable to certainty. The last two objectives require a modeling method to have the capability to translate the explanatory content at the conceptual level to inferential content at the empirical level. The methods of SEM and Bayesian networks are applied below to data.

5.1. Structural equation model

The concepts of resource commitment, flexibility to change and delivery performance are represented by the latent variables labeled as PROGRAMS, FLEXIBLE and DELIVERY. The set of model variables for the SEM application is defined by $V = \{$PROGRAMS, FLEXIBLE, DELIVERY, TQM, Analysis, Involve, Product, Volume, Speed, On-time$\}$ and $U = \{u_1, u_2, u_3, e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$. The hypothesized within-concept

structures are the basis for the following measurement equations:

$$TQM = \lambda_1 PROGRAMS + e_1,$$
$$Analysis = \lambda_2 PROGRAMS + e_2,$$
$$Involve = \lambda_3 PROGRAMS + e_3,$$
$$Product = \lambda_4 FLEXIBLE + e_4,$$
$$Volume = \lambda_5 FLEXIBLE + e_5,$$
$$Speed = \lambda_6 DELIVERY + e_6,$$
$$On\text{-}time = \lambda_7 DELIVERY + e_7.$$

5.1.1. Distribution assumptions

The analysis began by testing the assumption that the joint probability distribution of the measured variables is multivariate normal. Excessive kurtosis is known to yield a maximum likelihood test statistic that is a poor approximation of a chi-square variate (Babakus et al., 1987; Johnson and Creech, 1983). If multivariate kurtosis is extreme, the Satorra–Bentler scaled chi-square can be used to adjust the inflated chi-square goodness-of-fit statistic, with robust standard error estimates for model parameters (Bentler, 1995). However, the Mardia standardized coefficient of multivariate kurtosis (Mardia, 1970) indicated that deviation from multivariate normality was minor ($Z = 3.29$).

5.1.2. Unidimensionality, convergent validity and discriminant validity

Confirmatory factor analysis was applied to assess the measurement properties of unidimensionality, convergent validity and discriminant validity. Unidimensionality, which assumes each set of indicators reflect a single concept, was evaluated by the goodness-of-fit of the congeneric factor model (Jöreskog, 1971) and was supported ($\chi^2 = 13.45$, df = 11, $p = 0.27$). The congeneric factor model was also employed to evaluate convergent validity, the degree of agreement among indicators used to measure the same concept. The minimum requirement for convergent validity is evidence of significant link coefficients between each indicator and the common factor (Pedhazur and Schmelkin, 1991). The ratio of the coefficient estimate to the estimated standard error, which is approximately distributed as a normal variate, provides the test statistic for the null hypothesis the coefficient is equal to zero. Each λ-coefficient in the congeneric model was at least fifteen standard errors from the hypothesized value of zero, providing evidence of convergent validity.

Discriminant validity, the degree to which a latent variable and its indicators differ from other latent variables and their indicators, was assessed by fixing the correlation between each pair of concepts to 1.0, then estimating the constrained model, and comparing the resulting value of chi-square with the unconstrained model chi-square (Pedhazur and Schmelkin, 1991). The null hypothesis of a unit (perfect) correlation between two concepts was rejected each of the pairs of latent variables, at the 0.05 significance level, thus providing evidence of discriminant validity.

5.1.3. Tau-equivalent measures and reliability

Since the congeneric model was acceptable, tau-equivalent confirmatory factor analyses was applied to determine if the within-concept λ-coefficients are equal (Jöreskog and Sörbom, 1993). The tau-equivalent model was found to have an acceptable fit ($\chi^2 = 20.82$, df = 15, $p = 0.14$). Further, in comparison with the congeneric model, the composite hypothesis that ($\lambda_1 = \lambda_2 = \lambda_3$, $\lambda_4 = \lambda_5$, $\lambda_6 = \lambda_7$) was supported ($\Delta\chi^2 = 7.37$, $\Delta df = 4$, $p = 0.12$). Thus, each LV_j was assumed to contribute equally to each measured variable within each concept and the three models of concept relationships were constrained to tau-equivalent measurement.

Tau-equivalent or parallel measures are required for appropriate assessment of internal consistency by Cronbach's alpha coefficient of reliability (Bollen, 1989, p. 217). The construct reliabilities for PROGRAMS, FLEXIBLE and DELIVERY are 0.75, 0.71, and 0.81, respectively, and are in the acceptable range.

5.1.4. Structural models

The common effect model and the mediation model are potential explanations of the relationships between the assumed latent variables. These structures are specified as

Common Effect Model:

$$\text{PROGRAMS} = u_1$$

$$\text{FLEXIBLE} = b_1\text{PROGRAMS} + u_2$$

$$\text{DELIVERY} = b_2\text{PROGRAMS}$$
$$+ b_3\text{FLEXIBLE} + u_3;$$

Mediation Model:

$$\text{PROGRAMS} = u_1$$

$$\text{FLEXIBLE} = b_1\text{PROGRAMS} + u_2$$

$$\text{DELIVERY} = b_3\text{FLEXIBLE} + u_3.$$

The common effect model assumes that FLEXIBLE and DELIVERY are the common effects of PROGRAMS. The model had an adequate statistical fit ($\chi^2 = 20.82, \text{df} = 15, p = 0.14$) and excellent descriptive fit indexes (CFI $= 0.99$, AGFI $= 0.98$, RMSEA $= 0.03$). The mediation model also showed an acceptable statistical fit ($\chi^2 = 21.36$, df $= 16$, $p = 0.16$) and descriptive fit (CFI $= 0.99$, AGFI $= 0.98$, RMSEA $= 0.02$). The difference chi-square comparing the common effect model to the mediated model and evaluating the hypothesis $b_2 = 0$ was not significant ($\chi^2 = 0.54$, df $= 1, p = 0.46$). Each concept equation coefficient for the mediation model was significant ($b_1 = 0.31, \text{s.e.} = 0.04; b_3 = 0.63, \text{s.e.} = 0.05$). Thus, the more complex common effect model was rejected in favor of the parsimonious mediation model, PROGRAMS → FLEXIBLE → DELIVERY.

The standardized parameters for the mediation model are displayed in Fig. 1.
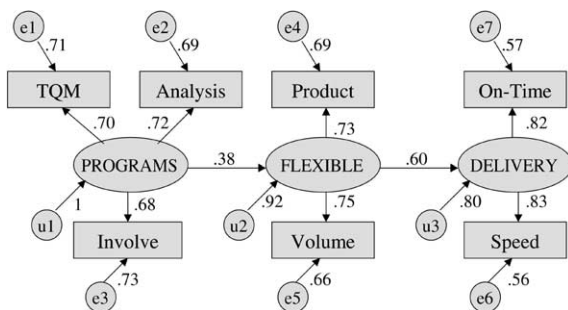

Fig. 1. SEM mediation model: standardized parameter estimates.

Each set of standardized measurement coefficients show the relative influence of a concept variable and an error variable on a measured variable. The square of a standardized coefficient shows the proportion of observed variance attributed to specified causes, the LV variable, and unspecified causes, the error term. For example, PROGRAMS contributes 49% (0.70^2) of the unit variance of TQM and the unspecified causes attribute 51% (0.71^2). The standardized concept coefficients similarly show the relative influence of a specified concept variable cause and unspecified causes, the u-variables. Thus, PROGRAMS variance is entirely due to unspecified causes, 15% (0.38^2) of the FLEXIBLE variance is attributed to PROGRAMS and 85% (0.92^2) to unspecified causes, and 36% (0.60^2) of the DELIVERY variables is accounted for by FLEXIBLE, with 64% (0.80^2) assigned to unspecified causes. The path model interpretation of Fig. 1 shows a one standard deviation change in FLEXIBLE is expected to result in a 0.60 standard deviation change in DELIVERY. A one standard deviation change in PROGRAMS is expected to result in a 0.38 standard deviation change in FLEXIBLE, and a 0.23 ($= 0.38 \times 0.60$) standard deviation change in DELIVERY.

The prediction of the impacts of changes in TQM or On-Time on other observed variables is not possible unless the latent variables can be measured. However, measurement of latent variables is indeterminate (Acito and Anderson, 1986). Thus, prediction and diagnostics of measured variables is not a capability of SEM.

5.2. A Bayesian network

The set of model variables for the Bayesian network application is defined by $V = \{\text{TQM},$ Analysis, Involve, Product, Volume, Speed, On-Time$\}$ and $U = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$. A hybrid combination of the independence testing and Bayesian approaches, as discussed above, was adopted for developing the network structure and estimating the conditional probabilities. Independence testing was employed to build the structure and a Dirichlet distribution conjugate analysis was used for probability estimation.

5.2.1. Building the network structure

The process of constructing the graphical structure started with a within-concept temporal ordering. The TQM is the most general improvement program, and is assumed to be a common cause of process analysis and employee involvement programs (Analysis ← TQM → Involve). Further, a process analysis program is assumed to be a cause of employee involvement programs (Analysis → Involve). In terms of flexibility, product flexibility is assumed to be a direct cause of volume flexibility (Product → Volume). For delivery, speed of delivery is viewed as a direct cause of on-time delivery (Speed → On-Time).

The PC algorithm, which permits the incorporation of variable order specification and restriction of the parent relationships, was used for conformation of the within-concept structure and to search for between-concept relationships. The within-concept structure was confirmed by rejection of the hypothesis of independence for each relationship ($p < 0.001$). Additionally, four between-concept relationships were deemed to be significant. Independence between TQM and product flexibility was rejected ($p < 0.001$), and independence between employee involvement programs and product flexibility was rejected ($p < 0.005$). These results imply dependencies of TQM → Product and Involve → Product. Further, the independence testing results implied the dependencies of Product → On-Time and Volume → Speed ($p < 0.001$).
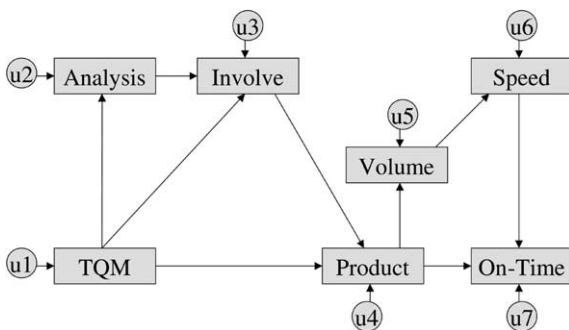
The complete structure of the Bayesian network is shown in Fig. 2.

The structure implies the following equations:

$$\text{TQM} = f_1(u_1),$$

$$\text{Analysis} = f_2(\text{TQM}, u_2),$$

$$\text{Involve} = f_3(\text{TQM, Analysis}, u_3),$$

$$\text{Product} = f_4(\text{TQM, Involve}, u_4),$$

$$\text{Volume} = f_5(\text{Product}, u_5),$$

$$\text{Speed} = f_6(\text{Volume}, u_6),$$

$$\text{On-Time} = f_7(\text{Product, Speed}, u_7).$$

5.2.2. Conditional probability estimation

The Bayesian approach was used to estimate the conditional probabilities for the network. Each $p(U_i) = \theta_i$ was assumed to be a collection of multinomial distributions, one for each parent configuration. The precision of the prior estimation was set at an equivalent sample size of 1. The Knowledge Discoverer program (Sebastiani and Ramoni, 2000) was used to estimate the posterior probabilities for the sets of parents of each consequence variable.

5.2.3. Probabilistic inference

Probabilistic inference is concerned with revising probabilities for a variable or set of variables, called the query, when an intervention fixes the values of another variable or set of variables, called the evidence. Any of the variables in the Bayesian network can serve as a query or as evidence, thus allowing forward inference from causes to effects (prediction) or backward inference from effects to causes (diagnostics). The simplest type of intervention is one where a single variable is forced to take on a fixed value. The intervention replaces the functional mechanism for the evidence variable, $E_i = f_i(pa_{ij}, u_i)$ with a fixed value, $E_i = e_i$, assumed known with certainty, in all equations. This creates a new model by removal of the network arrows from the parent set of the instantiated variable, which represents the system's behavior under the intervention. The relative magnitude of an influence from an intervention may be measured as the percent change from pre-intervention probabilities to post-intervention probabilities of the query for evidence in different states.
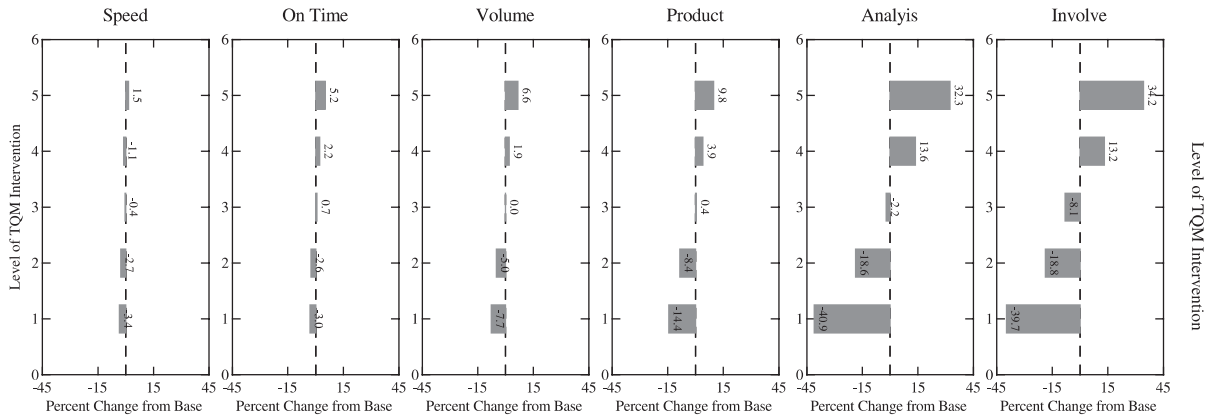

Fig. 2. Bayesian network structure.

Fig. 3. Summary of changes due to TQM interventions.

5.2.4. Prediction

The objective of predicting the impact of potential changes in TQM allocation levels on flexibility and delivery variables was accomplished by TQM interventions as the evidence. The magnitude of the impacts of the interventions was greatest on the direct effects of Involve, Analysis, and Product, as would be expected from the Bayesian network. The indirect influences on Volume, Speed and On-Time were relatively small.

The distributions resulting from the TQM interventions can be summarized by expected values and compared to the expected values of pre-intervention distributions. Fig. 3 provides a summary of expected changes given TQM interventions.

Fig. 3 follows the basic principles of Trellis displays (Cleveland, 1993). The variables from the left to the right show an increasing impact of TQM intervention. The level of TQM intervention ranges from 1 = None (at the bottom of the graphs) to 5 = Large Amount (at the top). Over this range of interventions, the customer delivery performance measures (Speed and On-Time) show the least variation. Investments in employee participation and process analysis (Involve and Analysis) show the greatest variation. The pre-intervention expected value of Involve was 2.72 (where 1 = no allocation and 5 = large allocation). The expected value of Involve with the TQM intervention set to no allocation was 1.64 or an expected decrease of 39.7%, as displayed in Fig. 3. In a similar fashion, the TQM intervention at the largest allocation

level showed an expected increase in Involve of 34.2%. Each expected percent change in Fig. 3 has a similar interpretation.

These results are consistent with our current understanding of TQM programs and systems. There is a saying attributed to Kaoru Ishikawa that summarizes the people aspect of this program succinctly: "TQM starts with education and ends with education." Furthermore, process improvements and continuous improvement programs have always been essential parts of quality management philosophies.

The expected changes in the delivery variables were small. The range of expected changes in Speed of delivery was from −3.4% for no TQM allocations to 1.5% for the highest TQM allocation level. The On-Time delivery variable also displayed a narrow range of expected percent changes over the TQM interventions. Similar small impacts would be obtained from interventions on Analysis and Involve since the influences are mediated by the flexibility variables.

5.2.5. Diagnostics

The application of Bayes' theorem allows "backward" probabilistic inference with interventions of On-Time delivery. The expected percent changes in the causes of On-Time delivery resulting from On-Time interventions are summarized in Fig. 4. As with Fig. 3, the expected changes are based on comparing pre-intervention and post-intervention expected values.
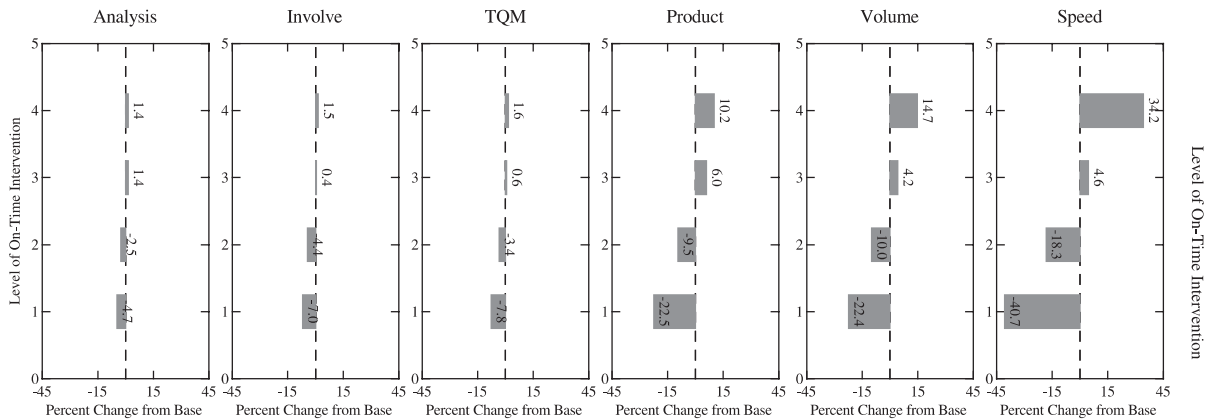
Fig. 4. Summary of changes due to On-Time intervention.

The expected changes in Fig. 4 may be viewed as measures of diagnostic importance for On-Time delivery. The highest level of Speed of delivery is the most important cause of improvement in On-Time delivery. The indirect cause of Volume flexibility is second in importance, followed by the direct cause of Product flexibility. Expected changes in the allocation levels of the improvement programs were minor.

These findings are consistent with the literature on JIT systems, including papers on the linkages between quick set-ups (and thus increased manufacturing flexibility), manufacturing lead-time reduction, and delivery speed improvements (Schmenner, 1988; Vastag and Whybark, 1993; Vastag and Montabon, 2001).

5.3. Specific comparisons

Both the Bayesian network and SEM, based on the results of traditional significance testing, supported the model of flexibility of change mediating resource commitment programs and delivery performance. However, SEM provided the more parsimonious explanation of the mediation model at the concept level. SEM relationships, described in the dimensionless regression metric of standardized values, are appropriate since the latent variables are not measured, and thus have no empirical metric. Although the latent variables are not measured, a great deal of emphasis is placed on the measurement sub-model relating observed

variables to the hypothetical latent variables. The strength of measurement by tau-equivalent evaluations and alpha reliability coefficients of internal consistency play a central role in the SEM analysis. However, the measurement evaluations and the fit of relationships between latent variables provide only an abstract causal description, since prediction and diagnostics are not possible. Thus, SEM could only achieve the explanation objective of the application.

The Bayesian network approach, using only measured variables, allows interventions on the TQM and On-Time network variables that results in predictions and diagnostics. The distribution changes from pre-intervention to post-intervention were summarized to expected values and then described by expected percent changes, as shown in Figs. 3 and 4. Probabilistic Inference could also be conducted with interventions on other network variables.

In summary, SEM provides a parsimonious description of observed and hypothetical variables rich in support by psychometric indexes. Whereas, the Bayesian network provides predictions described in terms of probabilities and percents. If the objective were only a description of theoretical constructs with no interest in current or future inference to observable variables, then SEM is likely be the preferred method. If objectives included prediction and diagnostics of observed variables, then the Bayesian network approach should be selected.

Table 1
Summary of model comparisons

Comparison	SEM	Bayesian network
Concept representation		
Parsimony	High	Low
Measurement assumption	Continuous	Discrete
Relationship assumption	Common cause and causally sufficient	Causally sufficient
Operating characteristics		
Functional form	Linear	Non-linear
Distribution assumption	Normal	Multinomial
Concept network structure	Parsimonious	Complex
Process modeling capability	Limited	Large
Assessment of adequacy		
Goodness-of-fit	Global statistical tests	Tests of independence
	Model and measurement fit indexes	
	Significance tests for measurement and concept coefficients	
Inference	Latent variable prediction	Observed variable prediction

6. General comparisons

Both SEM and Bayesian networks are specified by Eq. (1), and each coveys the causal assertions of a model $M = \{S, \Theta_S\}$ using a DAG to portray the structure of assumed functional relationships. However, fundamental differences exist between methods for the structure S and the set of parameters Θ_S compatible with S. Concept representation, distribution and functional assumptions, sample size, model complexity, measurement, specification search, model adequacy, theory testing and inference capabilities are the characteristics discussed below for each of reviewed methods. Table 1 provides a generalized summary of these comparisons.

6.1. Concept representation

Bayesian networks are often labeled as Bayesian belief networks. However, SEM also portrays a belief structure based on a conceptualization of perceived reality derived from existing theory, empirical evidence and speculation. The primary belief in SEM is an observable variable MV_i is caused by two unrelated unobservable hypothetical variables.

The level of abstraction entailed by the model structure varies across the methods in relation to the form of within-concept representation. Con-cept representation in SEM is highly abstract and parsimonious; with the simple methodologically form of linear regression. SEM portrays covariations between measured variables as the result of a common cause, LV_j, which results in the measurement model of $e_i \rightarrow MV_i \leftarrow LV_j$. Finally, SEM greatly simplifies the DAG since only the LV_j variables require temporal ordering.

6.2. Function, distribution and measurement assumptions

The functional form for discrete Bayesian networks is non-parametric and non-linear. Dirichlet distributions are assumed to result from multinomial sampling, and a subjective estimate of precision is required for the prior distributions. SEM methods are parametric in function and distribution, assuming normality and linearity. Although multivariate normality is usually assumed in SEM, the assumption is often not tested (Breckler, 1990) nor supported in applications (Micceri, 1989).

Continuous measurement is generally assumed for SEM due to the normality assumption. However, SEM offers considerable flexibility in handling non-normal continuous variables (West et al., 1995), ordinal data (Coenders et al., 1997) and categorical data (Muthen and Muthen, 1998). Bayesian networks currently require categorical measurement. Scale compression, such as imple-

mented in the above application, will result in some loss of information in a Bayesian network. The tradeoff is parsimonious intuitive measurement, which is consistent with an emphasis on the analysis of actions, versus full measurement utilization of SEM. If the tenuous assumption of normality for quasi-continuous scales holds in SEM, statistical testing will have greater power than significance testing in a discrete Bayesian network. However, the magnitude of power differences is not known.

6.3. Sample size and model complexity

There is some general agreement that the minimum sample size for SEM should at least 100 (Guadognoli and Velicer, 1988). Minimum sample size requirement for a Bayesian network is an open issue, but we speculate that 100 cases would be a reasonable minimum requirement. However, the issue of minimum sample size for a Bayesian network would seem to be highly related to the number of states in the categorical variables and the maximum number of parents of the variables in the modeled system.

In general, SEM is best suited for modeling rather small processes, as Bentler and Chou (1987) have recommended an analysis be limited to 20 or fewer measured variables. This recommendation is not too restrictive when item parcels employing single indicator SEM are used to represent concepts (Little et al., 2002). Although SEM assumes a linear process, SEM can handle latent variable interactions and quadratic effects (Ping, 1996). However, branching patterns in modeled process that create missing values require separate groups be formed for analysis. Bayesian networks have the capability to be applied to very large processes, with potentially thousands of variables (Spiegelhalter et al., 1993). Further, Bayesian network can accommodate a non-linear process, resulting from branching patterns.

6.4. Specification search

Specification search in causal models is a primary area of algorithm research, as well as a source of philosophical controversy (McKim and Turner, 1997). The controversy centers on acceptable ways of knowledge building and the possibility of inferring causation from association (Glymour and Cooper, 1999, part 3). The magnitude of a search effort for a DAG without temporal order specification is enormous. For example, given the objective to identify the "true" model of relationships between ten unordered variables, 4, 175, 098, 976, 430, 596, 143 DAGs should be searched and evaluated in reference to an appropriate fit criterion. When the background knowledge gives complete variable order for a Bayesian network, casual direction will available in the pattern. However in applications, establishing variable order is often an issue. Genetic algorithms based on permutation assessments have been developed to assist in with this problem (Larranaga et al., 1996).

SEM methods also generally employ model specification searches in applications. Jöreskog and Sörbom (1993) describe three approaches for SEM construction and development: strictly confirmatory, model generation and model comparison. MacCallum (1995) observes that the strictly confirmatory approach in SEM is very rare in applications, with model modification being very common. Model revisions in SEM are usually based on a local modification index, such as the Lagrange multiplier test, to assess the improvement in fit if a selected subset of fixed parameters were converted into free parameters. Model comparison posits a small number of feasible representations and selects the most acceptable model, based on evidence from nested significance testing or fit indexes. The SEM application in this paper provides an example of specification search by model comparison.

6.5. Model adequacy

Bayesian networks employ significance tests of conditional independence for a statistical assessment of structure adequacy. A maximum scoring approach is also available, where log likelihood scores for various parent sets are computed for each measured variable. SEM uses a global statistical test of goodness-fit, significance tests of measurement and concept coefficients, and

descriptive model and equation indexes to support the adequacy of a given model.

6.6. Theory testing

The qualitative tasks of concept definition, concept representation, temporal ordering and specification of causal relationships provide the major explanations of an investigated theory whether the concepts are portrayed by latent variables or data-dependent composites or direct measured variables. The explanation contributions of a quantitative method, regardless whether a regression or probability metric is employed, are limited to assessment of the model's adequacy reflected by significance testing, descriptive fit indexes, and estimation of the magnitude of causal influences.

Neither a Bayesian network, nor a SEM offers a unique model of reality; rather a class of observationally equivalent models is portrayed that cannot be distinguished by statistical analysis (Verma and Pearl, 1990; MacCallum et al., 1993). Further, Cooper (1999) points out that the interpretation of a Bayesian network must be made relative to the categorical variables represented by the nodes in the DAG. The existence of equivalent models is similar to confounding in experimental design, where different parameterizations cannot be distinguished in reference to fit to the data (Stelzl, 1986).

6.7. Inference

Science typically views theory validation as coming from predictive verification of expected theoretical results based on empirical evidence. A causal model should provide an explanatory description of causal relationships, plus manipulation capabilities for diagnosing the key changes necessary for system improvement and for predicting the impacts of potential change actions. Since each equation, implied by a structure S, represents a distinct causal mechanism, it is conceptually possible to set any represented variable to a specific value, i.e., impose an intervention.

Bayesian networks permit both forward and backward inference, allowing diagnostics and

prediction based on any set of selected measured variables as illustrated in the above application. As noted above, SEM has prediction capabilities only at the hypothetical variable level. Even when all aspects of a SEM representation are completely supported, prediction of observable consequences from potential managerial action is not attainable. Simply, the model assumes that every observable variable is a function of two unobservable variables, a common cause and an unspecified cause that cannot be subject to managerial interventions. The inability to translate knowledge gained from the theoretical model evaluation into a basis for managerial actions is the major weakness SEM.

7. Summary

The focus of this paper is on methods that view causation from a graphical viewpoint. The notion of a DAG is utilized to represent the conditional independence between any two variables, which implies the absence of a direct causal relationship. Further, assuming the nodes of a DAG represent random variables, the joint probability distribution of these variables can be factored in a causal meaningful manner. These general results are summarized in the specification of a causal model as a structure S conforming to a DAG and a set of associated parameters Θ_S consistent with the structure. SEM and Bayesian networks are portrayed as specializations of the general causal model specification, $M = \{S, \Theta_S\}$.

The emphasis of causal modeling applied in operations management has been primarily concentrated on providing increased process understanding by emphasizing psychometric and statistical support for theory-based explanations. SEM have been the most frequently used method for quantifying and evaluating an assumed causal process. The primary objective, under this approach, is to assess whether a postulated theoretical network is a reasonable approximation of the process that generated the study data. Indexes supporting construct validity, measurement reliability, parameter significance, model fit, and causal effects tend to dominate the reported results. Since the analysis is usually based on a high

level of theoretical and knowledge domain support, the confirmatory findings typically support the large majority of hypothesized relationships. Thus, conclusions tend to provide value by incremental extension of existing conceptualizations that cannot be extended to prediction of observed outcomes.

Bayesian networks, in contrast to SEM, assume the main role of causal modeling is to facilitate the analysis of potential and actual actions, rather than focus on theory confirmation. Indeed, Bayesian networks offer the capabilities to explain system relationships and to predict the impacts of potential actions as alternative structures that can be evaluated by traditional tests of significance or by posterior probabilities or both, as demonstrated in the above application. The probability metric provides non-linear detailed relationship information that should be easily consumable by the managers as well as academics. The modeling effort is concerned with observable variables, not hypothetical concepts. Thus, it is possible to introduce a conceptual intervention and evaluate the expected observable changes. More specifically, the posterior probabilities resulting from the intervention can be compared to the pre-intervention probabilities to provide a quantitative measure of expected change.

Appendix A. Description of the global manufacturing research group survey

In the late 1980s and in the mid 1990s, the Global Manufacturing Research Group carried out two worldwide surveys focusing on small machine tools and no-fashion textile manufacturing. Data were collected in about 30 countries representing market, transitional and planned economies.

In each country, directories of trade association members were used to select a random sample of firms. A manufacturing executive from each of the selected companies was contacted by telephone. The executive was presented evidence of the trade association support and was invited to participate in the study. As an incentive, firms were told that if they participated in the survey, the average res-

ponses for firms in their industry would be provided to them. Follow-up telephone calls were made to solicit responses and remind the participants to complete and return the questionnaires. If necessary, a second round of mailing was done to increase the sample size. In this paper, we selected a subset of the GMRG database, countries with large number of responses.

In the subset of data we used in this paper, we had 588 respondents from the second survey. The respondents were from Canada (88), Japan (77), Mexico (93), Russia (92) and the United States (238). Industry representation was classified as machine tool (209), textile (207) and others (172).

References

Acito, F., Anderson, R.D., 1986. Factor score indeterminacy: Correlates and implications for marketing research. Journal of Marketing Research 20 (2), 111–118.

Babakus, E., Ferguson, C.E., Jöreskog, K.G., 1987. The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distribution assumptions. Journal of Marketing Research 37, 72–141.

Bayes, T., 1763. An essay toward solving a problem in the doctrine of chance. Philosophical Transactions of the Royal Society 53, 370–418.

Bentler, P.M., 1995. EQS Structural Equations Program Manual. Multivariate Software, Encino, CA.

Bentler, P.M., Chou, C.P., 1987. Practical issues in structural modeling. Sociological Methods and Research 16, 78–117.

Bernardo, J.M., Smith, A.F.M., 2000. Bayesian Theory. Wiley, New York.

Bollen, K.A., 1989. Structural Equations with Latent Variables. Wiley, New York.

Breckler, S.J., 1990. Application of covariance structure modeling in psychology: Cause for concern? Psychological Bulletin 107, 260–273.

Cartwright, N., 1997. What is a causal structure? In: McKim, V.R., Turner, P.T. (Eds.), Causality in Crisis?. University of Notre Dame Press, Notre Dame, IN, pp. 343–357.

Cleveland, W.S., 1993. Visualizing Data. Hobart Press, Summit, NJ.

Coenders, G., Satorra, A., Saris, W.E., 1997. Alternative approaches to structural modeling of ordinal data. Structural Equation Modeling 4 (4), 261–282.

Cook, T.D., Campbell, D.T., 1979. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Rand McNally, Chicago.

Cooper, G.F., 1999. An overview of the representation and discovery of causal relationships using bayesian networks. In: Glymour, C., Cooper, G.F. (Eds.), Computation,

Causation and Discovery. MIT Press, Menlo Park, CA, pp. 3–62.

Cooper, G.F, Herskovits, E., 1992. A Bayesian method for the induction of probabilistic network from data. Machine Learning 9, 309–347.

Cowell, R.G., Dawid, A.P., Lauritzen, S.L., Spiegelhalter, D.J., 1999. Probabilistic Networks and Expert Systems. Springer-Verlag, New York.

Druzdzel, M.J., Simon, H.A., 1993. Causality in Bayesian belief networks. In: Proceedings of Ninth Annual Conference in Artificial Intelligence. Morgan Kaufmann Publishers, San Francisco, pp. 3–11.

Einstein, A., Podolsky, B., Rosen, N., 1935. Can quantum-mechanical description of physical reality be considered complete? Physical Review 41, 777.

Geiger, D., Herkerman, D., 1997. A characterization of Dirichlet distributions through local and global independence. Annals of Statistics 25, 1344–1368.

Glymour, C., Cooper, G.F., 1999. Computation, Causation and Discovery. MIT Press, Cambridge, MA.

Guadognoli, E., Velicer, W.F., 1988. Relationship of sample size to the stability of component patterns. Psychological Bulletin 103, 265–275.

Hausman, D.M., Woodward, J., 1999. Independence, invariance and the causal Markov condition. British Journal for the Philosophy of Science 50, 521–583.

Heckerman, D., 1999. A tutorial on learning with Bayesian networks. In: Jordan, M.I. (Ed.), Learning in Graphical Models. MIT Press, Cambridge, MA, pp. 301–354.

Hume, D., 1969. A Treatise of Human Nature. Penguin Books, London.

Jensen, F., 1996. An Introduction to Bayesian Network. University College London Press, London.

Johnson, D.R., Creech, J.C., 1983. Ordinal measures in multiple indicator models. American Sociological Review 48, 398–407.

Jöreskog, K.G., 1971. Statistical analysis of set of congeneric test. Psychometrika 36, 109–133.

Jöreskog, K.G., 1973. A general method for estimating a linear structural equation system. In: Goldberger, A.S., Duncan, O.D. (Eds.), Structural Equation Models in the Social Sciences. Academic Press, New York, pp. 85–112.

Jöreskog, K.G., Sörbom, D., 1993. LISREL 8, Analysis of Linear Structural Relations by the Method of Maximum Likelihood. International Education Services, Chicago.

Kiiveri, H.T., Speed, T.P., 1982. Structural analysis of multi-variate data: A review. In: Leinhardt, S. (Ed.), Sociological Methodology. Jossy Bass, San Francisco, pp. 209–289.

Larranaga, P., Poza, M., Yurramendi, Y., Murga, R.H., Kuijpers, C.M.H., 1996. Structure learning of Bayesian networks by genetic algorithms: A performance analysis of control parameters. IEEE Journal on Pattern Analysis and Machine Intelligence 18 (9), 912–926.

Lawley, D.N., Maxwell, A.E., 1963. Factor Analysis as a Statistical Method. Butterworth, London.

Little, T.D., Cunningham, W.A., Shahar, G., Widaman, K.F., 2002. To parcel or not to parcel: Exploring the question, weighing the merits. Structural Equation Modeling 9, 151–173.

MacCallum, R.C., 1995. Model specification: Procedures, strategies, and related issues. In: Hoyle, R.H. (Ed.), Structural Equation Modeling: Concepts, Issues and Applications. Sage, Thousand Oaks, CA, pp. 16–36.

MacCallum, R.C., Wegener, D.T., Uchino, B.M., Fabrigar, L.R., 1993. The problem of equivalent models in applications of covariance structure analysis. Psychological Bulletin 114, 185–199.

Mardia, K.V., 1970. Measures of multivariate skewness and kurtosis with applications. Biometrika 57, 519–530.

McKim, V.R., Turner, S.P., 1997. Causality in Crisis? University of Notre Dame Press, Notre Dame, IN.

Micceri, T., 1989. The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin 105, 156–166.

Muthen, B.O., Muthen, L.K., 1998. Mplus: User's Guide. Muthen and Muthen, Los Angeles, CA.

Pearl, J., 1988. Probabilistic Inference in Intelligent Systems. Morgan Kaufmann, San Mateo, CA.

Pearl, J., 1995. Causal diagrams for empirical research. Biometrika 82 (4), 669–710.

Pearl, J., 2000. Causality: Models, Reasoning and Inference. Cambridge University Press, Cambridge, UK.

Pedhazur, E.J., Schmelkin, L., 1991. Measurement, Design, and Analysis. Lawrence Erlbaum Associates, Hillsdale, NJ.

Ping, R., 1996. Latent variable interaction and quadratic effect estimation: A two-step technique using structural equation analysis. Psychological Bulletin 119, 166–175.

Ramoni, M., Sebastiani, P., 1999. Bayesian methods for intelligent data analysis. In: Bethold, M., Hand, D.J. (Eds.), An Introduction to Intelligent Data Analysis. Springer-Verlag, New York, pp. 129–166.

Reichenbach, Hans., 1956. The Direction of Time. University of California Press, Berkeley.

Scheines, R., 1997. An introduction to causal inference. In: McKim, V.R., Turner, P.T. (Eds.), Causality in Crisis?. University of Notre Dame Press, Notre Dame, IN, pp. 185–199.

Scheines, R., 1999. Estimating causal influences: TETRAD II model selection and Bayesian parameter estimation. In: Glymour, C., Cooper, G.F. (Eds.), Computation, Causation and Discovery. MIT Press, Menlo Park, CA, pp. 425–437.

Scheines, R., Spirtes, P., Glymour, C., Meek, C., 1994. TETRAD II: User's Manual. Lawrence Erlbaum, Hillsdale, NJ.

Schmenner, R.W., 1988. The merit of making things fast. Sloan Management Review 30 (1), 11–17.

Sebastiani, P., Ramoni, M., 2000. Bayesian inference with missing data using bound and collapse. Journal of Computational and Graphical Statistics 9 (4), 779–800.

Shachter, R.D., 1986. Evaluating influence diagrams. Operations Research 34 (6), 871–882.

Spearman, C., 1904. General intelligence, objectively determined and measured. American Journal of Psychology 15, 201–293.

Spiegelhalter, D.J., Dawid, A.P., Lauritzen, S.L., Cowell, R.G., 1993. Bayesian analysis with expert systems (with discussion). Statistical Science 8, 219–283.

Spirtes, P., Glymour, C., Schienes, R., 1993. Causation, Prediction, and Search. Springer-Verlag, New York.

Stelzl, I., 1986. Changing causal relationships without changing the fit: Some rules for generating equivalent LISREL-models. Multivariate behavioral Research 21, 309–331.

Suppes, P., 1970. A Probabilistic Theory of Causality. North-Holland Publishing, Amsterdam.

Thurstone, L.L., 1935. Vectors of the Mind. Chicago University Press, Chicago.

Vastag, G., Whybark, D.C., 1993. Global relations between inventory, manufacturing lead time and delivery date promises. International Journal of Production Economics 30–31, 563–569.

Vastag, G., Montabon, F., 2001. Linkages among manufacturing concepts, inventories, delivery service and competitive-ness. International Journal of Production Economics 71, 195–204.

Verma, T., Pearl, J., 1990. Equivalence and synthesis of causal models. In: Proceedings, Sixth Workshop on Uncertainty in Artificial Intelligence. North Holland, Amsterdam, pp. 220–227.

West, S.G., Finch, J.F., Curran, P.J., 1995. Structural equation models with nonnormal variables: Problems and remedies. In: Hoyle, R.H. (Ed.), Structural Equation Modeling: Concepts, Issues and Applications. Sage, Thousand Oaks, CA, pp. 56–75.

Wright, S., 1934. The method of path coefficients. Annals of Mathematical Statistics 5, 161–215.

Whybark, D.C., Vastag, G. (Eds.), 1993. Global Manufacturing Practices: A Worldwide Survey of Practices and Manufacturing Planning and Control. Elsevier, Amsterdam.

Winkler, R.L., 1967. The assessment of prior distributions in Bayesian analysis. Journal of American Statistical Association 62, 776–800.