

Application of Bayesian causal inference and structural equation model to animal breeding

Keiichi Inoue 

National Livestock Breeding Center, Nishigo,
Fukushima, Japan

Correspondence

Keiichi Inoue, National Livestock Breeding
Center, Nishigo, Fukushima 961-8511, Japan.
Email: k1inoue@nlbc.go.jp

Abstract

Optimized breeding goals and management practices for the improvement of target traits requires knowledge regarding any potential functional relationships between them. Fitting a structural equation model (SEM) allows for inferences about the magnitude of causal effects between traits to be made. In recent years, an adaptation of SEM was proposed in the context of quantitative genetics and mixed models. Several studies have since applied the SEM in the context of animal breeding. However, fitting the SEM requires choosing a causal structure with prior biological or temporal knowledge. The inductive causation (IC) algorithm can be used to recover an underlying causal structure from observed associations between traits. The results of the papers, which are introduced in this review, showed that using the IC algorithm to infer a causal structure is a helpful tool for detecting a causal structure without proper prior knowledge or with uncertain relationships between traits. The reports also presented that fitting the SEM could infer the effects of interventions, which are not given by correlations. Hence, information from the SEM provides more insights into and suggestions on breeding strategy than that from a multiple-trait model, which is the conventional model used for multitrait analysis.

KEYWORDS

animal breeding system, Bayesian causal inference, external intervention, inductive causation algorithm, structural equation model

1 | INTRODUCTION

In animal breeding strategies, there are two types of breeding goal to ensure the ability of animal to fit the demands of producers, consumers, and markets. One is the fundamental or sustainable breeding goal, which is not dependent on market trends and focuses on properties such as reproduction, feed efficiency, and longevity. The other is the flexible breeding goal that reflects temporal market or economical demands and focuses on properties such as the qualities of milk and meat products. When we are confronted with the need

to improve a new trait because of changing market demands, such as fatty acid composition, which is a meat quality trait and is related to the sweet aroma (Sakuma et al., 2012) and flavor of beef (Suzuki, Yokota, Shioura, Shimazu, & Iida, 2013) in Japanese Black cattle, we add the new trait to conventional selection traits and improve multiple traits simultaneously. In such a case, the most important thing is to confirm the genetic correlations between the traits by estimating the genetic parameters fitted by the multiple-trait model (MTM; Henderson & Quass, 1976) to prevent negative effects on the conventional economical traits. However, there are not only

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Animal Science Journal* published by John Wiley & Sons Australia, Ltd on behalf of Japanese Society of Animal Science

'correlations' but there could also be 'causal relationships' between the traits. For example, in the case of the relationship between gestation length (GL) and calving difficulty (CD), GL affects CD, because a longer GL causes a bigger infant or harmonization. In contrast, the opposite relationship is not possible; CD never affects GL under biological conditions.

When causal relationships exist between traits, fitting a structural equation model (SEM) allows for inferences of the magnitude of the causal effects between traits (Havelmo, 1943; Wright, 1921). SEMs have been applied in various fields, such as economics, psychology, and sociology. In recent years, Gianola and Solensen (2004) proposed adaptations of the SEM in the context of quantitative genetics and mixed-effect models. Ever since their proposal, several studies have applied the SEM in the context of animal breeding in dairy goats (de los Campos, Gianola, Boettcher, & Moroni, 2006), dairy cattle (de los Campos, Gianola, & Heringstad, 2006; Heringstad, Wu, & Gianola, 2009; Konig, Wu, Gianola, Heringstad, & Simianer, 2008; Wu, Heringstad, Chang, de los Campos, & Gianola, 2007; Wu, Heringstad, & Gianola, 2008; etc.), pigs (Ibanez-Escriche, Lopez de Maturana, Noguera, & Varona, 2010; Varona, Sorensen, & Thompson, 2007), and beef cattle (Inoue, Hosono, & Tanimoto, 2017; Inoue et al., 2016).

Fitting the SEM requires choosing a causal structure with prior biological or temporal knowledge. However, the inductive causation (IC) algorithm (Pearl, 2000; Verma & Pearl, 1990), which is one of the Bayesian causal inference, can be used to recover an underlying causal structure from observed associations between traits. The IC algorithm can output the causal relationships with a directed acyclic graph (DAG), which is a set of variables connected by directed edges, or a partially oriented graph. Searching for the causal structure in the algorithm is based on a conditional independence (Dawid, 1980) between variables.

Valente, Rosa, de los Campos, Gianola, and Silva (2010) adapted the IC algorithm to a mixed-model context and showed that applying this method to the posterior distribution of the residual (co) variance matrix of a standard MTM recovered the expected network in simulated data. They also presented an application of the methodology to real field data of European quail. After their work, several groups have tried to explore the causal structure from the real field data.

This review will first describe an overview of the IC algorithm and SEM, and then introduce some papers that apply these techniques to the context of animal breeding.

2 | THE IC ALGORITHM

The IC algorithm, which was proposed by Pearl (2000) and Verma and Pearl (1990), allows to search for causal networks. The IC algorithm performs a series of statistical decisions based on partial correlations between traits and consists of the following three steps (an example is shown in Figure 1):

2.1 | Step 1

If all partial correlations of two traits conditional on each possible set of the other traits are different from 0, the two traits are connected by an undirected edge (e.g., $y_1 - y_2$). For example, when there are four target traits for inferring the causal structure (y_1, y_2, y_3 , and y_4), as shown on the left graph of Step 1 in Figure 1, if all partial correlations of two arbitrary traits (e.g., y_1 and y_2) conditional on every set of the other traits (i.e., $\rho_{y_1y_2|y_3}$, $\rho_{y_1y_2|y_4}$, and $\rho_{y_1y_2|y_3y_4}$) and correlation between y_1 and y_2 ($\rho_{y_1y_2}$) are different from 0 (significant), the two traits are connected by an undirected edge ($y_1 - y_2$). The graph with undirected edges is finally generated by iterating the same process for the remaining possible set of traits (the right graph of Step 1 in Figure 1).

2.2 | Step 2

In the undirected graph obtained by step 1, if partial correlations of two nonadjacent traits (e.g., y_1 and y_3) with a common adjacent trait (e.g., y_2 in $y_1 - y_2 - y_3$) are dependent conditional on any possible set including the adjacent trait (y_2), that is, all partial correlations ($\rho_{y_1y_3}$, $\rho_{y_1y_3|y_2}$, $\rho_{y_1y_3|y_4}$, and $\rho_{y_1y_3|y_2y_4}$) are not significantly different from 0, arrowheads pointing to the common adjacent trait (y_2) are added ($y_1 \rightarrow y_2 \leftarrow y_3$ in the graph of Step 2 in Figure 1). This kind

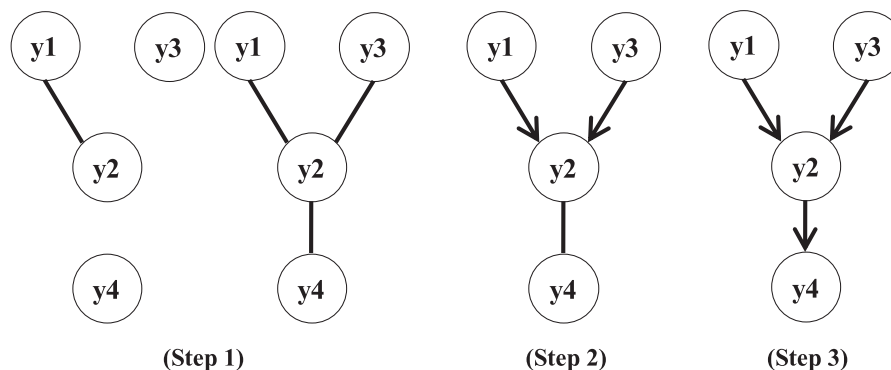


FIGURE 1 An example of the steps for the inductive causation algorithm. y_1 to y_4 represent traits, '-' and ' \rightarrow ' (\leftarrow) represent undirected and directed edges, respectively

of 'V' structure is called an 'unshielded collider.' The same process is repeated for the remaining possible set of traits.

2.3 | Step 3

In the partially oriented graph obtained from step 2, as many undirected edges as possible are oriented without creating a new unshielded collider or cycle. In the case of the example in Figure 1, y_2 and y_4 are connected by an undirected edge. If an arrowhead pointing to y_2 from y_4 is added to the undirected edge ($y_2 \leftarrow y_4$), new unshielded colliders are created ($y_1 \rightarrow y_2 \leftarrow y_4$ and $y_3 \rightarrow y_2 \leftarrow y_4$); therefore, a new directed edge from y_4 to y_2 cannot be added. In contrast, arrowheads pointing to y_4 from y_2 can be added to the undirected edge ($y_2 \rightarrow y_4$) because no new unshielded colliders are created. Finally, the graph of the causal structure is completed, shown as the graph in Step 3 of Figure 1. Additionally, an example of a cyclic graph is as follows: $y_1 \rightarrow y_2 \rightarrow y_3 \rightarrow y_1$.

Statistical decisions regarding declaring partial correlations as null or not null are based on highest posterior density (HPD) intervals. If the interval contained the value 0, the correlation is declared as null. Outputs may differ according to the probability content used for the decisions.

In the context of the mixed-model analysis, after fitting the MTM to the traits of interest, the IC algorithm can be applied to samples from the posterior distribution of the residual (co)variance matrix obtained from the multiple-trait analysis because the residual (co)variances are considered to be information from the joint distributions of the traits conditional on genetic effects, which corrects the search for confounding because of such effects when they are correlated (Valente et al., 2010). The program ICPS written in R (R Development Core Team, 2009), which can carry out the IC algorithm analysis, is provided by Valente and Rosa (2013).

3 | STRUCTURAL EQUATION MODEL

In the case of existing causal relationships between traits, as mentioned in the previous section, fitting a SEM (Haavelmo, 1943; Wright, 1921) can infer the causal coefficients between the traits. A SEM with a causal structure and random additive genetic effects, which was first proposed by Gianola and Solensen (2004), has been reviewed (Rosa et al., 2011; Wu, Heringstad, & Gianola, 2010) and the model with t traits can be written as follows:

$$\mathbf{y}_i = \Lambda \mathbf{y}_i + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i + \mathbf{e}_i$$

where \mathbf{y}_i is a $t \times 1$ vector of phenotypic records on animal i ; Λ is a $t \times t$ matrix of structural coefficients, which is filled by 0 except for the off-diagonal elements that correspond to the causal structure; $\boldsymbol{\beta}$ is a vector of fixed effects; \mathbf{u}_i is a vector of random additive genetic effects; \mathbf{e}_i is a vector of random residual effects; and \mathbf{X}_i is a known incidence

matrix. The joint distribution of vectors \mathbf{u}_i and \mathbf{e}_i in the equation is as follows:

$$\begin{bmatrix} \mathbf{u}_i \\ \mathbf{e}_i \end{bmatrix} \sim N \left\{ \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G}_0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_0 \end{bmatrix} \right\}$$

where \mathbf{G}_0 is the additive genetic (co)variance matrix and $\boldsymbol{\Psi}_0$ is the diagonal residual variance matrix.

By reducing the above SEM for \mathbf{y} , the model is transformed, as indicated below (Gianola & Solensen, 2004; Varona et al., 2007):

$$\begin{aligned} (\mathbf{I} - \Lambda) \mathbf{y}_i &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i + \mathbf{e}_i \\ \mathbf{y}_i &= (\mathbf{I} - \Lambda)^{-1} \mathbf{X}_i \boldsymbol{\beta} + (\mathbf{I} - \Lambda)^{-1} \mathbf{u}_i + (\mathbf{I} - \Lambda)^{-1} \mathbf{e}_i \\ &= \boldsymbol{\mu}_i^* + \mathbf{u}_i^* + \mathbf{e}_i^* \end{aligned}$$

where $\boldsymbol{\mu}_i^* = (\mathbf{I} - \Lambda)^{-1} \mathbf{X}_i \boldsymbol{\beta}$, $\mathbf{u}_i^* = (\mathbf{I} - \Lambda)^{-1} \mathbf{u}_i$, and $\mathbf{e}_i^* = (\mathbf{I} - \Lambda)^{-1} \mathbf{e}_i$. In addition, the joint distribution of \mathbf{u}_i^* and \mathbf{e}_i^* is as follows:

$$\begin{bmatrix} \mathbf{u}_i^* \\ \mathbf{e}_i^* \end{bmatrix} \sim N \left\{ \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G}_0^* & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_0^* \end{bmatrix} \right\}$$

with $\mathbf{G}_0^* = (\mathbf{I} - \Lambda)^{-1} \mathbf{G}_0 (\mathbf{I} - \Lambda)^{-1}$, $\mathbf{R}_0^* = (\mathbf{I} - \Lambda)^{-1} \boldsymbol{\Psi}_0 (\mathbf{I} - \Lambda)^{-1}$. Here, \mathbf{u}_i^* , \mathbf{e}_i^* , \mathbf{G}_0^* , and \mathbf{R}_0^* are the similar vectors as described above (\mathbf{G}_0^* is the additive genetic (co)variance matrix and \mathbf{R}_0^* is the residual (co)variance matrix). Like this, the SEM can be transformed into a MTM, which ignores causal relationships between traits. Therefore, the likelihood between MTM and SEM is equivalent (Varona et al., 2007). However, genetic effects estimated from the two models have different meaning. The MTM only infers 'overall' genetic effects, which include all 'direct' and 'indirect' genetic effects on each trait, as illustrated later. Here, the indirect genetic effects are those that are mediated by other phenotype traits and the direct genetic effects are those that are not. In contrast, the SEM can infer the direct genetic effects, which are separated from the causal effects (corresponding to the indirect genetic effects; Valente, Rosa, Gianola, Wu, & Weigel, 2013). For instance, assuming the causal structure between two traits shown in Figure 2a, the SEM can be represented as follows:

$$\mathbf{y}_1 = \mathbf{u}_1 + \mathbf{e}_1$$

$$\mathbf{y}_2 = \lambda_{21} \mathbf{y}_1 + \mathbf{u}_2 + \mathbf{e}_2$$

where \mathbf{y}_1 and \mathbf{y}_2 are the phenotypes; \mathbf{u}_1 and \mathbf{u}_2 are the additive genetic effects which affect directly on \mathbf{y}_1 and \mathbf{y}_2 , respectively (direct genetic effects), but \mathbf{u}_1 is also the indirect genetic effect on \mathbf{y}_2 ; \mathbf{e}_1 and \mathbf{e}_2 are the residual effects; and λ_{21} is the causal effect from \mathbf{y}_1 to \mathbf{y}_2 . In addition, the equation described above can be transformed as follows:

$$\mathbf{y}_1 = \mathbf{u}_1 + \mathbf{e}_1 = \mathbf{u}_1^* + \mathbf{e}_1^*$$

$$\mathbf{y}_2 = \lambda_{21} (\mathbf{u}_1 + \mathbf{e}_1) + \mathbf{u}_2 + \mathbf{e}_2 = (\lambda_{21} \mathbf{u}_1 + \mathbf{u}_2) + (\lambda_{21} \mathbf{e}_1 + \mathbf{e}_2) = \mathbf{u}_2^* + \mathbf{e}_2^*$$

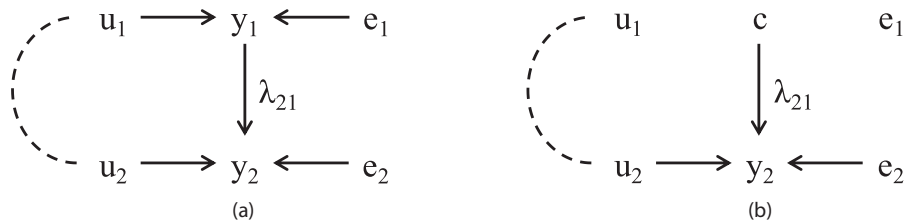


FIGURE 2 Hypothetical recursive causal structure involving two traits under no external intervention (a) and under external intervention (b). y_1 and y_2 are phenotypes; u_1 and u_2 are the additive genetic effects, which affect directly on y_1 and y_2 , respectively, but u_1 is also the indirect genetic effect on y_2 ; e_1 and e_2 are the residual effects; λ_{21} is the causal coefficient; and c is a constant. Dashed arc represents genetic correlation between the additive genetic effects

where u_1^* and u_2^* , e_1^* and e_2^* are the additive genetic and residual effects in the MTM, respectively. This equation shows that the genetic effects in the SEM affect not only the corresponding phenotypes directly but also the other phenotypes indirectly via the causal structure. In addition, it suggests that the genetic effects in the MTM represent the overall effects, which also include the direct and indirect effects via the causal effects ($u_2^* = \lambda_{21}u_1^* + u_2^*$).

4 | INTERVENTIONS IN THE SEM

As shown in the previous section, the SEM can infer direct and indirect genetic effects separately but the MTM cannot. Therefore, a breeding strategy based only on estimated genetic parameters obtained from the MTM would cause a delay in achieving the breeding goal if causal relationships and external ‘interventions,’ which would block indirect genetic effects, exist among the traits (Valente et al., 2013). An example of the intervention for y_1 in Figure 2a is shown in Figure 2b. The value of y_1 is fixed. The equation, which represents the intervention shown in Figure 2b, is as follows:

$$y_1 = c$$

$$y_2 = \lambda_{21}c + u_2 + e_2$$

where c is a constant and the indirect genetic effect of u_1 to y_2 via the causal structure no longer exists. Examples of the external intervention in management practices in livestock production systems are cross-fostering for litter size in sows and cesarean sections for CD.

Lopez de Maturana, Wu, Gianola, Weigel, and Rosa (2009) inferred the causal effects of GL, CD, and stillbirth (SB) by fitting a mixed-effects threshold SEM to the causal structure, assuming that GL affects CD and SB, and CD affects SB. Their results showed that the causal effects of GL to CD, GL to SB, and CD to SB were 0.47, 0.23, and 0.60, respectively. Using this result as an example, Valente et al. (2013) illustrated numerically that the artificial interventions in CD, such as cesarean sections in cows, would invert the ranking of genetic effects on SB because the indirect genetic effects of GL via CD and direct genetic effects of CD would be blocked in the causal structure. Therefore, in any breeding and production system based

on multiple traits, it is important to study the potential causal relationships among the traits (Rosa et al., 2011).

If the causal effects can be inferred by fitting the SEM to the causal structure, giving an artificial intervention, such as management practice, to the variation in causing traits can phenotypically improve the following traits. In this way, the application of causal inference and fitting the SEM are effective not only for breeding programs but also for planning for management practices in livestock production systems (Rosa & Valente, 2013). Therefore, when causal relationships exist between target traits, application of the SEM to the traits should be considered.

5 | APPLYING THE IC ALGORITHM AND SEM FOR ANIMAL BREEDING

Ever since Gianola and Solensen (2004) proposed adaptations of the SEM in the context of quantitative genetics and mixed-effect models, several studies have applied the SEM to different species and traits in the context of animal breeding. de los Campos, Gianola, Boettcher, et al. (2006) and de los Campos, Gianola, and Heringstad (2006) studied the relationship between somatic cell score (SCS) and milk yield (MY) in dairy cows and dairy goats. This work is the first application of the SEM to field data. Similarly, in dairy cattle, there are many reports exploring the relationships between SCS or clinical mastitis and MY in Norwegian Red cows (Wu et al., 2007, 2008); between claw disorders and MY in Holstein cows (Konig et al., 2008); between health and fertility traits in Norwegian Red cows (Heringstad et al., 2009); among GL, CD, and SB in US Holstein cows (Lopez de Maturana et al., 2009, 2010); and between SCS and MY in Canadian Holstein cows (Jamrozik, Bohmanova, & Schaeffer, 2010; Jamrozik & Schaeffer, 2010). Some of the authors proposed extensions for the SEM, such as accounting for population heterogeneity (Wu et al., 2007), the Gaussian-threshold Bayesian hierarchical model (Wu et al., 2008), the heterogeneous causal model (Wu et al., 2010), and the random regression model (Jamrozik et al., 2010; Jamrozik & Schaeffer, 2010). In swine, Varona et al. (2007) used a recursive model to analyze the relationship between litter size and average litter weight in Landrace and Yorkshire pigs and found likelihood equivalence between MTM and the SEM. Ibanez-Escriche et al. (2010) developed a change-point recursive model

using change-point detection techniques (Carlin, Gelfand, & Smith, 1992; Chib, 1998) and fitted the model to the litter size and number of SBs in Large White pigs. However, these studies assumed that causal structures were known a priori.

Recently, some studies fitted SEMs based on a data-driven causal structure search using the IC algorithm. The IC algorithm fitted to the posterior distribution of residual (co)variance matrices that are considered phenotypic correlations conditional on additive genetic correlations between traits; this approach was proposed by Valente et al. (2010). They used simulated data and showed that this approach recovered the correct causal network. Then, Valente, Rosa, Silva, Teixeira, and Torres (2011) presented the first application of such a methodology to a real dataset by exploring the space of causal structures among five productive and reproductive traits in European quail. From the application of the IC algorithm with different HPD intervals, three undirected graphs were obtained, but all had no directed edge or unshielded collider. To construct the structures among the five traits and fit SEMs to the structures, the authors oriented the edges of the graphs according to temporal information as a prior knowledge.

Bouwman, Valente, Janss, Bovenhuis, and Rosa (2014) explored causal networks between 14 bovine milk fatty acids, which are highly correlated traits, by applying the IC algorithm. They retrieved undirected graphs using the IC algorithm with different HPD intervals, and then oriented the undirected graph obtained with the 95% HPD interval according to prior biological knowledge of *de novo* synthesis to fit the SEM: C4:0 → C6:0 → C8:0 → C10:0 → C12:0. The authors showed that inferred structural coefficients ranged from 0.85 to 1.05 and the SEM could be more plausible than the MTM by the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002). In the genetic parameters obtained from the SEM, the genetic variances for downstream traits in the causal structure showed a gradual decrease (i.e., 0.460 for C4:0, 0.114 for C6:0, 0.073 for C8:0, 0.066 for C10:0, and 0.004 for C12:0), suggesting that indirect genetic effects from upstream traits were gradually explaining a larger portion of genetic variability.

In Japanese Black cattle, Inoue et al. (2016) inferred phenotypic causal structures and causal effects among meat quality traits. They applied the IC algorithm with different HPD intervals to meat quality traits, including beef marbling score (BMS), beef color score (BCS), firmness of beef (FIR), texture of beef (TEX), beef fat color score (BFS), and the ratio of monounsaturated fatty acids to saturated fatty acids (MUS). They obtained a partially oriented graph except for BFS by using the IC algorithm with 80% of HPD intervals. The obtained graph had an undirected edge between FIR and TEX; therefore, two competing causal structures (i.e., DAGs), with either the arrow $FIR \rightarrow TEX$ or $FIR \leftarrow TEX$, were fitted using a SEM to infer structural coefficients between the selected traits. The authors finally suggested that the DAG with the directed edge of $FIR \leftarrow TEX$ was more feasible for the structures based on the results of DICs of the SEMs. The structural coefficient for the path from MUS and BCS to BMS showed that a 1-unit improvement in MUS or BCS resulted in an increase of 0.85 or a decrease of 0.54 in BMS in the DAG. They

also reported that the genetic variances in the downstream traits (BMS, FIR, and TEX) in the causal structure from the SEM were smaller than those obtained from the MTM, whereas the variances in the upstream traits (BCS and MUS), which were not conditional on any of the other traits in the causal structure, had no significant differences between the SEM and MTM. This indicated that a breeding strategy based only on inferred genetic parameters obtained from the MTM would cause a delay in achieving the breeding goal if external interventions occurred in the upstream traits.

Inoue et al. (2017) also inferred causal structures and the causal effects among reproductive traits (CD and GL) and calf size (CS: birth weight, withers height, and chest girth of calves) in Japanese Black cattle. The IC algorithm with different HPD intervals was applied to the posterior distribution of the residual (co)variance matrix obtained from a threshold multiple-trait sire-maternal grand sire (MGS) model. The results of the IC algorithm with 95% and 90% of HPD intervals only retrieved an undirected graph, which had edges between GL and CS and between CS and CD (i.e., $GL - CS - CD$). Then, they fitted the SEM to the causal structure that was assigned directions on the edges detected by the IC algorithm according to prior biological knowledge and graph theory (i.e., $GL \rightarrow CS \rightarrow CD$). The structural coefficients of GL on CD via CS on the observable scale showed that an extra day of GL increased the risk of dystocia by 0.2% (via chest girth) and 0.3% (via birth weight and withers height), in the causal structure. The authors also suggested that the application of the IC algorithm to the residual variance components from the sire-MGS model could lead to the detection of an incorrect structure because maternal genetic effects might not be removed completely from the residual variance components in the model.

More recently, techniques for inferring causal relationships by using the IC algorithm and SEM have applied to genome-wide association studies (GWAS). Momen et al. (2018) applied SEM to GWAS on chickens, including causal relationships among breast meat (BM), body weight (BW), and hen-house production (HHP), to compare the results obtained from both an SEM-based GWAS (SEM-GWAS) and a traditional multitrait association analysis (MTM-GWAS). They set a causal structure, $BM \rightarrow BW$ and HHP and $BW \rightarrow HHP$, found by the IC algorithm and fitted the SEM to the structure. As a result, the application of SEM-GWAS could separate SNP effects into direct, indirect, and total effects. The authors concluded that the SEM-GWAS delivered a more comprehensive understanding of SNP effects than did the MTM-GWAS.

6 | CONCLUSIONS

The results of the papers that are introduced in this review showed that the application of the IC algorithm to infer a causal structure is one of the most helpful tools for detecting an underlying causal structure without proper prior knowledge or with uncertain relationships between traits. The reports also indicated that visualization of the causal structure among traits on a graph and fitting the SEM to the causal structure could infer the effects of interventions, which

are not given by correlations. Hence, information from the SEM provides more insights into and suggestions on breeding strategies than does that from a MTM, which is a conventional model usually used for multiple-trait analysis. These new procedures could be utilized in the field of animal breeding.

ORCID

Keiichi Inoue  <https://orcid.org/0000-0001-5353-8558>

REFERENCES

- Bouwman, A. C., Valente, B. D., Janss, L. L. G., Bovenhuis, H., & Rosa, G. J. M. (2014). Exploring causal networks of bovine milk fatty acids in a multivariate mixed model context. *Genetics Selection Evolution*, *46*, 2. <https://doi.org/10.1186/1297-9686-46-2>
- Carlin, B. P., Gelfand, A. E., & Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics*, *41*, 389–405. <https://doi.org/10.2307/2347570>
- Chib, S. (1998). Estimation and comparison of multiple-change point models. *Journal of Econometrics*, *86*, 221–241. [https://doi.org/10.1016/s0304-4076\(97\)00115-2](https://doi.org/10.1016/s0304-4076(97)00115-2)
- Dawid, A. P. (1980). Conditional independence for statistical operations. *Annals of Statistics*, *8*, 598–617. <https://doi.org/10.1214/aos/1176345011>
- de los Campos, G., Gianola, D., Boettcher, P., & Moroni, P. (2006). A structural equation model for describing relationships between somatic cell score and milk yield in dairy goats. *Journal of Animal Science*, *84*, 2934–2941. <https://doi.org/10.2527/jas.2006-016>
- de los Campos, G., Gianola, D., & Heringstad, B. (2006). A structural equation model for describing relationships between somatic cell score and milk yield in first-lactation dairy cows. *Journal of Dairy Science*, *89*, 4445–4455. [https://doi.org/10.3168/jds.S0022-0302\(06\)72493-6](https://doi.org/10.3168/jds.S0022-0302(06)72493-6)
- Gianola, D., & Sorensen, D. (2004). Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. *Genetics*, *167*, 1407–1424. <https://doi.org/10.1534/genetics.103.025734>
- Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, *11*, 1–12. <https://doi.org/10.2307/1905714>
- Henderson, C. R., & Quaas, R. L. (1976). Multiple trait evaluation using relative's records. *Journal of Animal Science*, *43*, 1188–1197. <https://doi.org/10.2527/jas1976.4361188x>
- Heringstad, B., Wu, X. L., & Gianola, D. (2009). Inferring relationships between health and fertility in Norwegian Red cows using recursive models. *Journal of Dairy Science*, *92*, 1778–1784. <https://doi.org/10.3168/jds.2008-1535>
- Ibanez-Escriche, N., Lopez de Maturana, E., Noguera, J. L., & Varona, L. (2010). An application of change-point recursive models to the relationship between litter size and number of stillborns in pigs. *Journal of Animal Science*, *88*, 3493–3503. <https://doi.org/10.2527/jas.2009-2557>
- Inoue, K., Hosono, M., & Tanimoto, Y. (2017). Inferring causal structures and comparing the causal effects among calving difficulty, gestation length and calf size in Japanese Black cattle. *Animal*, *11*, 2120–2128. <https://doi.org/10.1017/S1751731117000957>
- Inoue, K., Valente, B. D., Shoji, N., Honda, T., Oyama, K., & Rosa, G. J. M. (2016). Inferring phenotypic causal structures among meat quality traits and the application of a structural equation model in Japanese Black cattle. *Journal of Animal Science*, *94*, 4133–4142. <https://doi.org/10.2527/jas.2016-0554>
- Jamrozik, J., Bohmanova, J., & Schaeffer, L. R. (2010). Relationships between milk yield and somatic cell score in Canadian Holsteins from simultaneous and recursive random regression models. *Journal of Dairy Science*, *93*, 1216–1233. <https://doi.org/10.3168/jds.2009-2585>
- Jamrozik, J., & Schaeffer, L. R. (2010). Recursive relationships between milk yield and somatic cell score of Canadian Holsteins from finite mixture random regression models. *Journal of Dairy Science*, *93*, 5474–5486. <https://doi.org/10.3168/jds.2010-3470>
- Konig, S., Wu, X. L., Gianola, D., Heringstad, B., & Simianer, H. (2008). Exploration of relationships between claw disorders and milk yield in Holstein cows via recursive linear and threshold models. *Journal of Dairy Science*, *91*, 395–406. <https://doi.org/10.3168/jds.2007-0170>
- Lopez de Maturana, E., de los Campos, G., Wu, X.-L., Gianola, D., Weigel, K. A., & Rosa, G. J. M. (2010). Modeling relationships between calving traits: A comparison between standard and recursive mixed models. *Genetics Selection Evolution*, *42*, 1. <https://doi.org/10.1186/1297-9686-42-1>
- Lopez de Maturana, E., Wu, X. L., Gianola, D., Weigel, K. A., & Rosa, G. J. M. (2009). Exploring biological relationships between calving traits in primiparous cattle with a Bayesian recursive model. *Genetics*, *181*, 277–287. <https://doi.org/10.1534/genetics.108.094888>
- Momen, M., Ayatollahi Mehrgardi, A., Amiri Roudbar, M., Kranis, A., Mercuri Pinto, R., Valente, B. D., ... Gianola, D. (2018). Including phenotypic causal networks in genome-wide association studies using mixed effects structural equation models. *Frontiers in Genetics*, *9*, 455. <https://doi.org/10.3389/fgene.2018.00455>
- Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, UK: Cambridge University Press.
- R Development Core Team. (2009). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rosa, G. J. M., & Valente, B. D. (2013). Breeding and genetics symposium: Inferring causal effects from observational data in livestock. *Journal of Animal Science*, *91*, 553–564. <https://doi.org/10.2527/jas.2012-5840>
- Rosa, G. J. M., Valente, B. D., de los Campos, G., Wu, X.-L., Gianola, D., & Silva, M. A. (2011). Inferring causal phenotype networks using structural equation models. *Genetics Selection Evolution*, *43*, 6. <https://doi.org/10.1186/1297-9686-43-6>
- Sakuma, H., Saito, K., Sowa, T., Asano, S., Kohira, K., Okumura, T., ... Kawamura, T. (2012). Effect of crude fat content and fatty acid composition on sensory characteristics of *M. longissimus dorsi* of Japanese Black steers. *Nihon Chikusan Gakkaiho*, *83*(3), 291–299. (In Japanese with English summary) <https://doi.org/10.2508/chikusan.83.291>
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, *64*, 583–639. <https://doi.org/10.1111/1467-9868.00353>
- Suzuki, K., Yokota, S., Shioura, H., Shimazu, T., & Iida, F. (2013). Effect of meat grade, gender of the animal, and fatty acid content on the eating quality of Japanese black beef meat determined using testing panel. *Nihon Chikusan Gakkaiho*, *84*(3), 375–382. (In Japanese with English summary) <https://doi.org/10.2508/chikusan.84.375>
- Valente, B. D., & Rosa, G. J. M. (2013). Mixed effects structural equation models and phenotypic causal networks. In C. Gondro, J. van der Werf, & B. Hayes (Eds.), *Genome-wide association studies and genomic prediction* (pp. 449–464). New York, NY: Humana Press.
- Valente, B. D., Rosa, G. J. M., & de los Campos, G., Gianola, D., & Silva, M. A. (2010). Searching for recursive causal structures in multivariate genetics mixed models. *Genetics*, *185*, 633–644. <https://doi.org/10.1534/genetics.109.112979>
- Valente, B. D., Rosa, G. J. M., Gianola, D., Wu, X. L., & Weigel, K. A. (2013). Is structural equation modeling advantageous for the genetic improvement of multiple traits? *Genetics*, *194*, 561–572. <https://doi.org/10.1534/genetics.113.151209>

- Valente, B. D., Rosa, G. J. M., Silva, M. A., Teixeira, R. B., & Torres, R. A. (2011). Searching for phenotypic causal networks involving complex traits: An application to European quail. *Genetics Selection Evolution*, 43, 37. <https://doi.org/10.1186/1297-9686-43-37>
- Varona, L., Sorensen, D., & Thompson, R. (2007). Analysis of litter size and average litter weight in pigs using a recursive model. *Genetics*, 177, 1791–1799. <https://doi.org/10.1534/genetics.107.077818>
- Verma, T., & Pearl, J. (1990). Equivalence and synthesis of causal models. In *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence* (pp. 220–227). Cambridge, MA: Elsevier. <https://doi.org/10.5555/647233.719736>
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557–585.
- Wu, X.-L., Heringstad, B., Chang, Y.-M., de los Campos, G., & Gianola, D. (2007). Inferring relationships between somatic cell score and milk yield using simultaneous and recursive models. *Journal of Dairy Science*, 90, 3508–3521. <https://doi.org/10.3168/jds.2006-762>
- Wu, X. L., Heringstad, B., & Gianola, D. (2008). Exploration of lagged relationships between mastitis and milk yield in dairy cows using a Bayesian structural equation Gaussian-threshold model. *Genetics Selection Evolution*, 40, 333–357. <https://doi.org/10.1051/gse:2008009>
- Wu, X. L., Heringstad, B., & Gianola, D. (2010). Bayesian structural equation models for inferring relationships between phenotypes: A review of methodology, identifiability, and applications. *Journal of Animal Breeding and Genetics*, 127, 3–15. <https://doi.org/10.1111/j.1439-0388.2009.00835.x>

How to cite this article: Inoue K. Application of Bayesian causal inference and structural equation model to animal breeding. *Anim Sci J*. 2020;91:e13359. <https://doi.org/10.1111/asj.13359>