



HAL
open science

Product versus additive threshold models for analysis of reproduction outcomes in animal genetics

Ingrid David, Loys L. Bodin, Daniel Gianola, Andres Legarra, Eduardo Manfredi, Christèle Robert-Granié

► To cite this version:

Ingrid David, Loys L. Bodin, Daniel Gianola, Andres Legarra, Eduardo Manfredi, et al.. Product versus additive threshold models for analysis of reproduction outcomes in animal genetics. *Journal of Animal Science*, 2009, 87 (8), pp.2510-2518. 10.2527/jas.2008-1586 . hal-02656727

HAL Id: hal-02656727

<https://hal.inrae.fr/hal-02656727>

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

JOURNAL OF ANIMAL SCIENCE

The Premier Journal and Leading Source of New Knowledge and Perspective in Animal Science

Product versus additive threshold models for analysis of reproduction outcomes in animal genetics

I. David, L. Bodin, D. Gianola, A. Legarra, E. Manfredi and C. Robert-Granié

J ANIM SCI 2009, 87:2510-2518.

doi: 10.2527/jas.2008-1586 originally published online April 24, 2009

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://www.journalofanimalscience.org/content/87/8/2510>



American Society of Animal Science

www.asas.org

Product versus additive threshold models for analysis of reproduction outcomes in animal genetics¹

I. David,*² L. Bodin,* D. Gianola,† A. Legarra,* E. Manfredi,* and C. Robert-Granié*

*Institut National de la Recherche Agronomique UR 631, Station d'Amélioration Génétique des Animaux, 31320 Castanet-Tolosan, France; and †Department of Animal Sciences, University of Wisconsin, Madison 53706

ABSTRACT: The phenotypic observation of some reproduction traits (e.g., insemination success, interval from lambing to insemination) is the result of environmental and genetic factors acting on 2 individuals: the male and female involved in a mating couple. In animal genetics, the main approach (called additive model) proposed for studying such traits assumes that the phenotype is linked to a purely additive combination, either on the observed scale for continuous traits or on some underlying scale for discrete traits, of environmental and genetic effects affecting the 2 individuals. Statistical models proposed for studying human fecundability generally consider reproduction outcomes as the product of hypothetical unobservable variables. Taking inspiration from these works, we propose a

model (product threshold model) for studying a binary reproduction trait that supposes that the observed phenotype is the product of 2 unobserved phenotypes, 1 for each individual. We developed a Gibbs sampling algorithm for fitting a Bayesian product threshold model including additive genetic effects and showed by simulation that it is feasible and that it provides good estimates of the parameters. We showed that fitting an additive threshold model to data that are simulated under a product threshold model provides biased estimates, especially for individuals with high breeding values. A main advantage of the product threshold model is that, in contrast to the additive model, it provides distinct estimates of fixed effects affecting each of the 2 unobserved phenotypes.

Key words: additive model, Gibbs sampling, product model, reproduction traits

©2009 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 2009. 87:2510–2518
doi:10.2527/jas.2008-1586

INTRODUCTION

The phenotypic observation of some reproduction traits is due to effects contributed by the male and the female. For such traits depending on several individuals, a main approach that has been proposed in animal genetics models the observed phenotype (for continuous traits), or an underlying variable (for binary traits analyzed under the threshold model assumption), as the sum of environmental and genetic effects contributed by the different individuals involved (Wilham, 1963; Varona and Noguera, 2001; Piles et al., 2005; Bijma et al., 2007). In some cases, these models give perplexing results (e.g., high negative estimates of the cor-

relation between maternal and direct genetic effects; Robinson, 1996; Koerhuis and Thompson, 1997). Some authors have pointed out that the combination of the different sources of variation may not be purely additive (on either the observed or underlying scales). Actually, statistical models proposed for studying human fecundability generally consider the reproduction result as the product of hypothetical unobservable variables. For instance, Speirs et al. (1983) proposed a model in which the number of observed embryo implantations is the product of the number of viable embryos and a measure of uterus receptivity. For binary traits, Zhou and Weinberg (1996) assumed that the probability of conception in a menstrual cycle is the product of the probability that the cycle is viable and the probability that conception would occur. This product combination of factors for studying reproduction results has the advantage of being biologically more natural and easier to construe than the purely additive combination. Nevertheless, such models have not been used for studying reproduction traits in animal genetics. In line with human studies, we propose a product threshold model for studying the outcome of an insemination taking into account the action of the male and female and the ge-

¹The authors thank the Ministère de l'Agriculture for supporting this study in the framework of a BELIA action, directed by the Association Nationale Insémination Ovine and Institut National de la Recherche Agronomique. The authors thank Wendy Brand-Williams for editing of the English language. The Midi-Pyrénées region is acknowledged for supporting D. Gianola as Chaire D'Excellence Pierre de Fermat.

²Corresponding author: Ingrid.David@toulouse.inra.fr

Received October 27, 2008.

Accepted April 22, 2009.

netic relationships between individuals. The objectives of this study are 1) to present the theory of a product threshold model in a genetic context, and 2) to examine, with illustrations, the feasibility of this model.

MATERIALS AND METHODS

Animal Care and Use Committee approval was not obtained for this study because no animals were used.

Methods

Background: Linear Additive Model for a Continuous Trait Depending on 2 Individuals. Statistical methodology for studying a quantitative genetic trait dependent on several individuals has already been developed under the assumption of the additive model (Wilham, 1963; Bijma et al., 2007). To define the notation that will be used in this paper, we recall the corresponding theory as in Bijma et al. (2007). The model for a phenotype dependent on 2 individuals $\{i, j\}$, consists in expressing the observed phenotypic value, P_{ij} as the sum of 2 unobserved phenotypic effects: a direct effect ($P_{D,i}$) of individual i and an associative effect ($P_{S,j}$) of the associated individual j :

$$P_{ij} = P_{D,i} + P_{S,j}. \quad [1.1]$$

Phenotypic direct and associative effects can both be decomposed into a heritable component, referred to as the breeding value (\mathbf{A}), and a nonheritable environmental component (\mathbf{E}). Thus, $P_{ij} = \mathbf{A}_{D,i} + \mathbf{E}_{D,i} + \mathbf{A}_{S,j} + \mathbf{E}_{S,j}$, where $\mathbf{A}_{D,i}$ is the direct breeding value (DBV) and $\mathbf{A}_{S,j}$ is the associative breeding value (SBV), which pertain to genetically distinct traits. In matrix notation, the corresponding animal model is as follows:

$$l = \mathbf{X}\mathbf{b} + \mathbf{Z}_D\mathbf{a}_D + \mathbf{Z}_S\mathbf{a}_S + \varepsilon, \quad [1.2]$$

where $l = \{l_{ij}\}$ is the vector of observed phenotypes; β is a vector of fixed effects, with incidence matrix \mathbf{X} ; \mathbf{a}_D and \mathbf{a}_S are vectors of DBV and SBV, respectively (each having an order equal to the number of animals in the pedigree), with corresponding incidence matrices \mathbf{Z}_D and \mathbf{Z}_S . The distributional assumption is $\begin{bmatrix} a_D \\ a_S \end{bmatrix} \sim N \left(0, \begin{bmatrix} \sigma_{A_D}^2 & \sigma_{A_{DS}} \\ \sigma_{A_{DS}} & \sigma_{A_S}^2 \end{bmatrix} \otimes \mathbf{A} \right)$, where \mathbf{A} is the numerator relationship matrix and \otimes represents the Kronecker product. The residuals ε are independent and identically distributed with variance σ_ε^2 .

Threshold Models for a Binary Trait Depending on 2 Individuals. For dichotomous traits, the observed phenotype P_{ij}^* (1 or 0) is modeled using generalized linear mixed model assumptions. The most

widely used approach adopts the same linear predictor as for a linear mixed model with a probit link function:

$$\Pr(P_{ij}^* = 1) = \Phi(A_{D,i} + E_{D,i} + A_{S,j} + E_{S,j}),$$

where $\Pr(P_{ij}^* = 1)$ is the conditional (given the A and E variables) probability of success for the couple $\{i, j\}$ and Φ is the standard cumulative distribution function of the normal distribution. This model is called the additive threshold model. This model is equivalent, under the threshold model theory (Wright, 1934; Gianola, 1982), to

$$P_{ij}^* = \begin{cases} 1 & \text{if } l_{ij} > T \\ 0 & \text{if } l_{ij} \leq T \end{cases}$$

where T is a threshold value and l_{ij} is an underlying variable linked to the environmental and genetic effect as in model [1.2].

Suppose now that the 2 unobserved phenotypes ($P_{D,i}^*, P_{S,j}^*$) are dichotomous traits as well, such that $\Pr(P_{D,i}^* = 1) = \Phi(A_{D,i} + E_{D,i})$ and $\Pr(P_{S,j}^* = 1) = \Phi(A_{S,j} + E_{S,j})$. In this case, the relationship between the probability of success for the unobserved phenotypes and the observed phenotype is not obvious with the additive threshold model. A more intuitive way to link observed and unobserved phenotypes in the case of dichotomous traits is to assume that a success is observed only if there is a success for each of the 2 unobserved phenotypes: $P_{ij}^* = P_{D,i}^* \times P_{S,j}^*$. For example, conception in the couple is the result of success of each member. If the probabilities of success for the 2 unobserved phenotypes are conditionally independent, then

$$\begin{aligned} \Pr(P_{ij}^* = 1) &= \Pr(P_{D,i}^* = 1) \times \Pr(P_{S,j}^* = 1) \\ &= \Phi(A_{D,i} + E_{D,i}) \times \Phi(A_{S,j} + E_{S,j}). \end{aligned} \quad [1.3]$$

According to the classical threshold model theory, this product threshold model is equivalent to

$$P_{ij}^* = \begin{cases} 1 & \text{if } P_{D,i}^* = 1 \text{ and } P_{S,j}^* = 1 \\ 0 & \text{elsewhere} \end{cases},$$

where $P_{D,i}^*$ and $P_{S,j}^*$ are related to 2 underlying continuous variables, d and s , respectively.

$$P_{D,i}^* = \begin{cases} 1 & \text{if } d > T_D \\ 0 & \text{if } d \leq T_D \end{cases} \quad \text{and} \quad P_{S,j}^* = \begin{cases} 1 & \text{if } s > T_S \\ 0 & \text{if } s \leq T_S \end{cases},$$

where T_D and T_S are the threshold values for d and s , respectively. Further, let

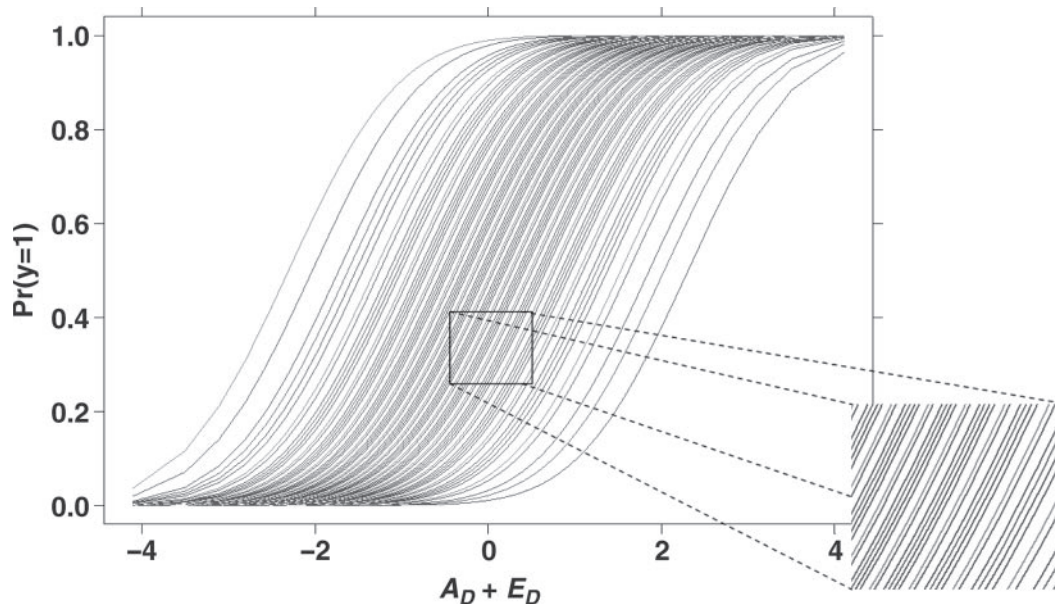


Figure 1. Expected probability of success for a couple (i, j) under the assumption of the threshold additive model $\Pr(P_{ij}^* = 1) = \Phi(A_{D,i} + E_{D,i} + A_{S,j} + E_{S,j})$, as a function of the direct value $\delta_{D,i} = A_{D,i} + E_{D,i}$, with one curve for each associative value $\delta_{S,j} = A_{S,j} + E_{S,j}$, where P_{ij}^* is the observed phenotype, $A_{D,i}, A_{S,j}$ are the direct and associative breeding values, and $E_{D,i}, E_{S,j}$ are the direct and associative environmental component, respectively.

$$d = X_D b_D + Z_D a_D + e_D \quad [1.4]$$

and

$$s = X_S b_S + Z_S a_S + e_S, \quad [1.5]$$

where the notations and distributional assumptions for the genetic effects are the same as in [1.2]. In addition, β_D and β_S are vectors of fixed effects, with incidence matrices X_D and X_S linking unobserved direct and associative phenotypes to fixed effects, respectively, and $\varepsilon_D, \varepsilon_S$ are mutually independent residuals with variances $\sigma_{\varepsilon_D}^2$ and $\sigma_{\varepsilon_S}^2$, respectively.

Contrary to the additive threshold model, the interpretation of unobserved phenotypes is straightforward in the product threshold model. For instance, in the study of AI success, if individual j is the male, $\Pr(P_{S,j}^* = 1)$ is the chance that the male produces a sperm capable of fertilizing the ovum if the latter is viable. Conversely, $\Pr(P_{D,i}^* = 1)$ is the chance that the ovum can be fertilized, provided that the sperm is fecundant.

Product and additive threshold models support different biological assumptions. Figures 1 and 2 depict for the additive and product threshold models, respectively, variations of $\Pr(P_{ij} = 1)$ with respect to $\delta_{D,i} = A_{D,i} + E_{D,i}$, with one curve for each $\delta_{S,j} = A_{S,j} + E_{S,j}$. In the additive threshold model, the probability of success $\Pr(P_{ij} = 1)$ can reach very high values as soon as $\delta_{D,i}$ is sufficiently large in comparison with $\delta_{S,j}$. In contrast, in the product threshold model,

the increase of $\Pr(P_{ij} = 1)$ with $\delta_{D,i}$ is penalized by the value of $\delta_{S,j}$ and reaches a maximum value that is generally less than 1. Assuming a product threshold model may be more realistic in some cases. For instance, a sterile male will never have progeny, even if he is mated to highly fertile females. This case corresponds to the change of the lowest lines in Figures 1 and 2. For intermediate $\delta_{D,i}$ and $\delta_{S,j}$, the change of $\Pr(P_{ij} = 1)$ for different $\delta_{S,j}$ values when $\delta_{D,i}$ increases are similar and independent of $\delta_{S,j}$ values in the additive threshold model. On the other hand, in the product threshold model these changes in probability of success depend on the values of $\delta_{S,j}$, and increase faster for greater $\delta_{S,j}$. However, for reduced $\delta_{D,i}$ and $\delta_{S,j}$, the models behave similarly because, in this case, the odds ratio is close to the relative risk.

Heritabilities. In the additive threshold model, the total variance of the underlying variate l is $\sigma_l^2 = \sigma_{A_D}^2 + \sigma_{A_S}^2 + \sigma_\varepsilon^2 + 2r\sigma_{A_{DS}}$, where r denotes the additive genetic relationship between associates in a couple. The heritability of direct and associative effects on the underlying scale is then $h_{A_D(x-scale)}^2 = \frac{\sigma_{A_D}^2}{\sigma_l^2}$ and

$h_{A_S(x-scale)}^2 = \frac{\sigma_{A_S}^2}{\sigma_l^2}$, respectively. An approximate transfor-

mation of these heritabilities to the observed scale (p-scale) is obtained by multiplying each heritability by $z^2 / \{\Pr(P^* = 1)[1 - \Pr(P^* = 1)]\}$, where z is the ordinate of a standard normal distribution function correspond-

ing to a threshold = $\Phi^{-1}[\Pr(P^* = 1)]$ (Dempster and Lerner, 1950).

In contrast to the additive threshold model, it is possible to assess the heritabilities of the unobservable individual phenotypes P_D^* and P_S^* in the product threshold model. The parametric definitions are

$$h_{P_{D(x-scale)}}^2 = \frac{\sigma_{A_D}^2}{\sigma_{A_D}^2 + \sigma_{\varepsilon_D}^2} \text{ and } h_{P_{S(x-scale)}}^2 = \frac{\sigma_{A_S}^2}{\sigma_{A_S}^2 + \sigma_{\varepsilon_S}^2} \text{ on the}$$

underlying scale. The genetic variances on the observed scale (p-scale) are given by $\sigma_{A_{D(p-scale)}}^2 = z_{P_D}^2 h_{P_{D(x-scale)}}^2$

and $\sigma_{A_{S(p-scale)}}^2 = z_{P_S}^2 h_{P_{S(x-scale)}}^2$ (Dempster and Lerner, 1950) for the direct and associate genetic effects, respectively, where z_{P_D} and z_{P_S} are the ordinates of a standard normal distribution function corresponding to thresholds equal to $\Phi^{-1}[\Pr(P_D^* = 1)]$ and $\Phi^{-1}[\Pr(P_S^* = 1)]$, respectively. Therefore, the heritabilities of direct and associative effects on the observed scale are

$$h_{A_{D(p-scale)}}^2 = h_{P_{D(x-scale)}}^2 \times z_{P_D}^2 / \left\{ \Pr(P^* = 1) [1 - \Pr(P^* = 1)] \right\}$$

and

$$h_{A_{S(p-scale)}}^2 = h_{P_{S(x-scale)}}^2 \times z_{P_S}^2 / \left\{ \Pr(P^* = 1) [1 - \Pr(P^* = 1)] \right\},$$

respectively. Because P_D^* and P_S^* are unobservable, one has to estimate $\Pr(P_D^* = 1)$ and $\Pr(P_S^* = 1)$.

Parameter Estimation. Parameter estimation methods in the additive threshold model are well known (Gianola and Foulley, 1983; Harville and Mee, 1984; Gilmour et al., 1985; Hoeschele and Tier, 1995; Moreno et al., 1997) and will not be presented here. Bayesian inference in the product threshold model can be carried out using the Gibbs sampling algorithm with a modification of the data augmentation algorithm described in Sorensen et al. (1995). In the classical additive threshold model, the idea is to sample the underlying variable from its conditional posterior distribution (a truncated normal distribution, on the left if $P = 1$ and on the right otherwise). The next step is to sample each parameter from its conditional posterior distribution, with the underlying variable replacing the observed data, so that the Gibbs sampler is as for the linear model (Sorensen et al., 1995). Implementation of the product threshold model can be done using the same approach and a supplementary step, which is the sampling of $P_{D,i}$ and $P_{S,j}$. Consider the following notation,

$$\Pr(P_{D,i} = 1) = \Phi(w'_i \theta_D) \text{ and } \Pr(P_{S,j} = 1) = \Phi(w'_j \theta_S),$$

where $P_{D,i}$ and $P_{S,j}$ are the unobserved phenotypes, θ_D and θ_S are the vectors of parameters (i.e., fixed and random effects) for direct and associative phenotypes, respectively; w'_i and w'_j are row incidence matrices linking θ_D and θ_S to the unobserved phenotypes. $P_{D,i}$ and

$P_{S,j}$ are sampled in the n th step of the Gibbs sampling from their joint conditional distribution.

$$\begin{aligned} & \text{If } P_{ij} = 1, \text{ then } P_{D,i} = 1 \text{ and } P_{S,j} = 1; \\ & \text{if } P_{ij} = 0, \text{ then } \Pr(P_{D,i}^n = 1, P_{S,j}^n = 1 | P_{ij} = 0, \theta_D^{n-1}, \theta_S^{n-1}) = 0. \\ & \Pr(P_{D,i}^n = 0, P_{S,j}^n = 1 | P_{ij} = 0, \theta_D^{n-1}, \theta_S^{n-1}) \\ & = \left[1 - \Phi(w'_i \theta_D^{n-1}) \right] \times \Phi(w'_j \theta_S^{n-1}) / P_t; \\ & \Pr(P_{D,i}^n = 1, P_{S,j}^n = 0 | P_{ij} = 0, \theta_D^{n-1}, \theta_S^{n-1}) \\ & = \Phi(w'_i \theta_D^{n-1}) \times \left[1 - \Phi(w'_j \theta_S^{n-1}) \right] / P_t; \\ & \Pr(P_{D,i}^n = 0, P_{S,j}^n = 0 | P_{ij} = 0, \theta_D^{n-1}, \theta_S^{n-1}) \\ & = \left[1 - \Phi(w'_i \theta_D^{n-1}) \right] \times \left[1 - \Phi(w'_j \theta_S^{n-1}) \right] / P_t, \end{aligned}$$

where $P_t = 1 - \Phi(w'_i \theta_D^{n-1}) \times \Phi(w'_j \theta_S^{n-1})$. Given $P_{D,i}$ and $P_{S,j}$, parameter sampling is identical to the additive model by regarding $P_{D,i}$ and $P_{S,j}$ as the data. We added this supplementary step to the TM program developed by A. Legarra (Institut National de la Recherche Agronomique-Station d'Amélioration Génétique des Animaux, Castanet-Tolosan, France, personal communication).

Illustrations

A study with 4 different simulated designs (Table 1) was performed to illustrate the feasibility of the product threshold model. The base population consisted of 60 unrelated sires and 200 or 2,000 unrelated dams without records. Sires were randomly mated to dams to generate 600 or 4,200 progeny. Dam (DBV) and sire (SBV) breeding values were generated using bivariate normal distributions with parameters corresponding to the designs shown in Table 1. Genetic values of the progeny were sampled as the mean of sire and dam genetic values plus the Mendelian sampling. Part of the progeny were randomly assigned to be females and have an unobserved direct phenotype (400 or 4000 individuals i), whereas the other part were considered to be males and have an associative unobserved phenotype (200 individuals j). The observed phenotypes, resulting from the association of individuals i and j , were generated according to the product threshold model. The underlying d and s values were generated for each couple using Eq. [1.4] and [1.5]. Two fixed effects were included in the model. The first effect f_1 was a cross-classified effect with 3 levels that affected direct and associative unobserved phenotypes. The second effect f_2 was a cross-classified effect with 2 levels affecting associative unobserved phenotypes only. Here, $\beta'_D = [0, 1.5, 1]$ and $\beta'_S = [0, -1.5, 0.5, 0, 2]$, where the 3 first elements in β_S corresponded to the effects of f_1 and the fourth and

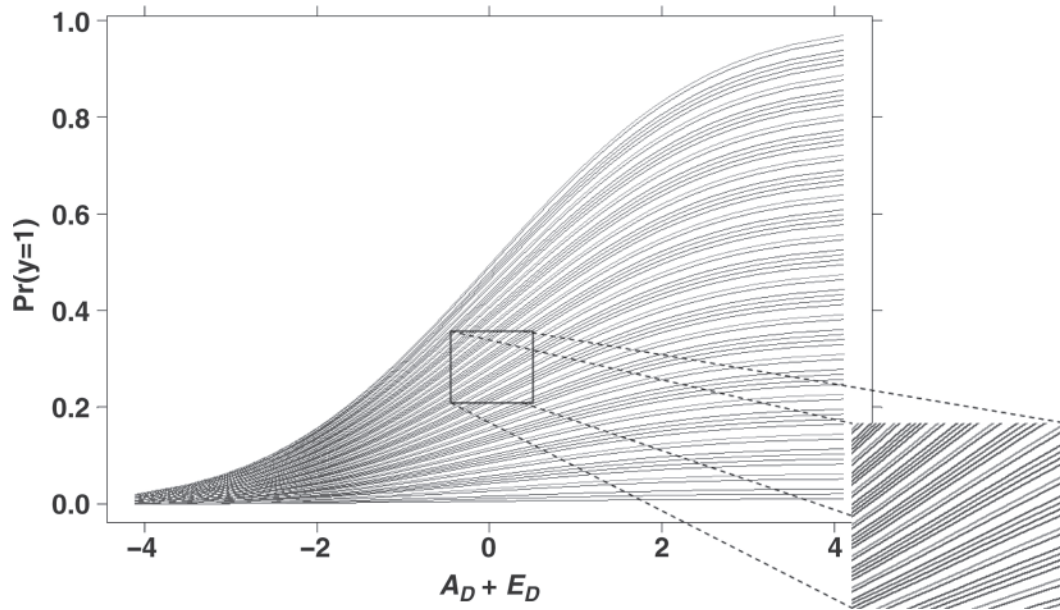


Figure 2. Expected probability of success for a couple (i, j) under the assumption of the threshold product model $\Pr(P_{ij}^* = 1) = \Phi(A_{D,i} + E_{D,i}) \times \Phi(A_{S,j} + E_{S,j})$, as a function of the direct value $\delta_{D,i} = A_{D,i} + E_{D,i}$, with one curve for each associative value $\delta_{S,j} = A_{S,j} + E_{S,j}$, where P_{ij}^* is the observed phenotype, $A_{D,i}, A_{S,j}$ are the direct and associative breeding values, and $E_{D,i}, E_{S,j}$ are the direct and associative environmental component, respectively.

fifth to the effects of f_2 . Levels of the fixed effects were randomly assigned to each couple with equal probability for each level. We fixed $\sigma_{\varepsilon_D}^2 = \sigma_{\varepsilon_S}^2 = 1$ and randomly assigned ε_D and ε_S to each couple. Discrete phenotypes P_D^*, P_S^* , and P^* were generated as follows: $P_D^* = 1$ if $d > T_D$, $P_D^* = 0$ otherwise; $P_S^* = 1$ if $s > T_S$, $P_S^* = 0$ otherwise, and $P^* = P_D^* \times P_S^*$. Values of T_D and T_S were fixed to 0 in all designs.

The first 2 designs had perfectly balanced data: each of the 400 “female” individuals i was associated to each of the 200 “male” individuals j . The first design assumed a large ($\sigma_{A_D}^2 = 1$) variance for the direct genetic effect and a moderate ($\sigma_{A_S}^2 = 0.25$) variance for associative genetic effect; the second employed a very small (0.05) variance for the direct genetic effects. In the third design, there were unbalanced data (males were

randomly mated with females) with a small variance for the direct genetic effect. In the fourth design, few observations per individual i were considered with moderate variances for direct and associative effects.

Each design was analyzed using 3 models: the product threshold model, the additive threshold model, and a multiple-trait threshold model, which considers P_D^* and P_S^* as observed phenotypes. Results obtained with the multiple-trait threshold model were a reference considered as the best parameter estimates (called best) that can be obtained. In practice, this third model cannot be applied to real data, because P_D^* and P_S^* are unobserved phenotypes.

For all models, the Gibbs sampler analysis was carried out using 1 chain consisting of 300,000 iterations. After discarding the first 30,000 iterations, samples of the parameters of interest were saved every 100 itera-

Table 1. Description of simulated designs

Item	Design 1	Design 2	Design 3	Design 4
No. of dams	200	200	200	2,000
No. of individuals i	400	400	400	4,000
No. of individuals j	200	200	200	200
No. of records	80,000	80,000	80,000	12,000
$\sigma_{A_D}^2$	1.00	0.05	0.05	0.36
$\sigma_{A_S}^2$	0.25	0.25	0.25	0.25
$r(u_D, u_S)$	0.30	0	0.30	0.30
Type of design	Balanced	Balanced	Unbalanced	Unbalanced
$\Pr(P_D = 1)^1$	0.73	0.72	0.73	0.73
$\Pr(P_S = 1)^1$	0.47	0.46	0.47	0.45
$\Pr(P = 1)^1$	0.32	0.30	0.32	0.30

¹Probability computed using the simulated data.

Table 2. Variance components estimated (posterior mean) with the additive and product threshold models for the different designs (credibility interval in brackets)¹

Item	$\sigma_{A_D}^2$	$\sigma_{A_S}^2$	$r(a_D, a_S)$	$h_{D(p-scale)}^2$	$h_{S(p-scale)}^2$
Design 1 ²					
Simulated data	1.00	0.25	0.30	0.25	0.14
Product threshold model	1.07	0.22	0.12	0.26	0.13
	[0.94,1.21]	[0.18,0.27]	[-0.12,0.36]	[0.24,0.28]	[0.11,0.15]
Additive threshold model	0.58	0.03	0.06	0.21	0.01
	[0.51,0.66]	[0.02,0.03]	[-0.20,0.32]	[0.19,0.23]	[0.01,0.01]
Design 2 ³					
Simulated data	0.05	0.25	0.00	0.03	0.15
Product threshold model	0.05	0.19	0.21	0.03	0.12
	[0.04,0.06]	[0.15,0.23]	[-0.12,0.53]	[0.02,0.03]	[0.10,0.14]
Additive threshold model	0.03	0.02	0.24	0.02	0.01
	[0.05,0.04]	[0.02,0.03]	[-0.11,0.57]	[0.01,0.02]	[0.01,0.02]
Design 3 ³					
Simulated data	0.05	0.25	0.30	0.02	0.15
Product threshold model	0.06	0.27	0.62	0.03	0.15
	[0.05,0.07]	[0.22,0.33]	[0.33,0.63]	[0.02,0.03]	[0.13,0.18]
Additive threshold model	0.03	0.02	0.53	0.02	0.01
	[0.03,0.04]	[0.02,0.03]	[0.23,0.80]	[0.01,0.02]	[0.01,0.02]
Design 4 ⁴					
Simulated data	0.36	0.25	0.30	0.14	0.15
Product threshold model	0.39	0.19	0.59	0.13	0.11
	[0.30,0.49]	[0.11,0.28]	[0.25,0.88]	[0.11,0.16]	[0.07,0.16]
Additive threshold model	0.21	0.01	0.65	0.10	<0.01
	[0.17,0.26]	[0.00,0.02]	[0.2,0.99]	[0.08,0.12]	

¹ a_D, a_S : direct and associative breeding values.

²400 individuals i , 200 individuals j , balanced data.

³400 individuals i , 200 individuals j , unbalanced data.

⁴4,000 individuals i , 200 individuals j , unbalanced data.

tions. Flat priors were used for all parameters, and beginning values were randomly sampled.

RESULTS

Except for $\sigma_{A_D}^2$ in design 2, genetic parameter estimates obtained with the product threshold model were in accordance with the simulated data (Table 2). Similar posterior means and credibility intervals were obtained with the multiple trait model (results not shown). In nearly all designs, heritability estimates obtained with the additive threshold model understated true values, whereas genetic correlation estimates aligned with those obtained with the product threshold model. Fixed effects were well estimated with the product threshold model but not with the additive one (Table 3). However, fixed effects in product and additive threshold models do not have the same interpretation and cannot be properly compared. Even if posterior means were similar, credibility intervals obtained for fixed effects with the product threshold model were 3 to 4 times larger than those obtained with the multiple trait model (results not shown).

Correlations between “best estimated” and predicted direct or associative breeding values were greater for the product threshold model than for the additive threshold model in all designs (Table 4). A plot between the “best” and estimated breeding values from the product threshold model did not show any trend,

whereas the additive model underestimated the largest breeding values (Figure 3). Nevertheless, correlations between breeding values estimated with product and additive threshold models were high.

DISCUSSION

For studying reproduction results, we proposed the product threshold model, which supposes that a success for P^* is observed when there is a success for P_D^* and P_S^* . A similar approach has been proposed for modeling embryo implantation after in vitro fertilization in humans: the EU model (Speirs et al., 1983), where E is the number of viable embryos and U is a binary indicator of uterine receptivity. Dukic and Hogan (2002) showed that in the extreme case where the number of embryos transferred is equal to one for all subjects, none of the model parameters is identifiable from observed data. Our illustrations imply a similar situation where each female event is associated with a unique male event. Nevertheless, our results (with input parameters essentially recovered within high credibility regions) suggest that parameters are identifiable, at least under the conditions considered. In general, identification is very difficult to study in complex hierarchical models such as the one considered here. In any case, if proper priors are adapted for all parameters, this provides unambiguous Bayesian learning (Carlin and Louis, 2000; Sorensen and Gianola, 2002). However, it may be the case

Table 3. Fixed effects (f_1 or f_2) estimated with the additive threshold and product threshold models for the different designs (credibility interval in brackets)

Item	f_1				f_2
	Level 2-Level 1		Level 3-Level 2		Level 2-Level 1
	Direct	Associative	Direct	Associative	Associative
Simulated data	1.50	-1.50	1.00	0.50	2.00
Design 1 ¹					
Product threshold model	1.53 [1.49,1.58]	-1.53 [-1.60,-1.46]	0.99 [0.96,1.02]	0.52 [0.44,0.61]	2.00 [1.96,2.07]
Additive threshold model		0.46 [0.43,0.48]		0.93 [0.90,0.95]	0.75 [0.73,0.80]
Design 2 ²					
Product threshold model	1.56 [1.49,1.63]	-1.48 [-1.58,-1.38]	1.00 [0.97,1.03]	0.44 [0.32,0.55]	1.88 [1.79,1.98]
Additive threshold model		0.70 [0.68,0.72]		1.03 [1.01,1.05]	0.67 [0.65,0.69]
Design 3 ³					
Product threshold model	1.47 [1.42,1.52]	-1.57 [-1.69,-1.46]	1.03 [1.00,1.06]	0.38 [0.25,0.50]	2.08 [1.99,2.17]
Additive threshold model		0.71 [0.69,0.73]		1.05 [1.03,1.07]	0.66 [0.64,0.68]
Design 4 ⁴					
Product threshold model	1.38 [1.21,1.58]	-1.52 [-1.85,-1.28]	1.03 [0.95,1.12]	0.58 [0.22,0.96]	2.16 [1.88,2.64]
Additive threshold model		0.60 [0.54,0.66]		0.97 [0.90,1.03]	0.72 [0.67,0.77]

¹400 individuals i , 200 individuals j , balanced data, high direct genetic variance.

²400 individuals i , 200 individuals j , balanced data, small direct genetic variance.

³400 individuals i , 200 individuals j , unbalanced data, small direct genetic variance.

⁴4,000 individuals i , 200 individuals j , unbalanced data, medium direct genetic variance.

that the prior could always be influential, even for very large samples. An argument suggesting why we obtained good estimations of the parameters using flat priors is as follows. In the standard linear animal model, the genetic additive (a) and permanent environmental (p) effects cannot be disentangled from each other at the level of the conditional likelihood because only their sum is identified (an infinite number of values a

and p produce the same sum). However, when all random effects are integrated out, the corresponding variance components can be identified in the marginal likelihood, provided their covariance matrices are not diagonal (due to the relationship matrix considered for the additive genetic effect). A similar reasoning holds in the product threshold model: at the level of the conditional likelihood, the probabilities $\Pr(P_S = 1)$ and

Table 4. Correlations between breeding values predicted with the different models (between direct breeding values above the diagonal, associative breeding values below the diagonal)

Model	Design	Multiple trait threshold model	Additive threshold model	Product threshold model
Multiple trait threshold model	Design 1 ¹	—	0.98	0.99
	Design 2 ²	—	0.95	0.96
	Design 3 ³	—	0.95	0.97
	Design 4 ⁴	—	0.83	0.88
Additive threshold model	Design 1 ¹	0.94	—	0.98
	Design 2 ²	0.91	—	0.98
	Design 3 ³	0.93	—	0.98
	Design 4 ⁴	0.73	—	0.95
Product threshold model	Design 1 ¹	0.97	0.96	—
	Design 2 ²	0.95	0.96	—
	Design 3 ³	0.97	0.96	—
	Design 4 ⁴	0.83	0.93	—

¹400 individuals i , 200 individuals j , balanced data, high direct genetic variance.

²400 individuals i , 200 individuals j , balanced data, small direct genetic variance.

³400 individuals i , 200 individuals j , unbalanced data, small direct genetic variance.

⁴4,000 individuals i , 200 individuals j , unbalanced data, medium direct genetic variance.

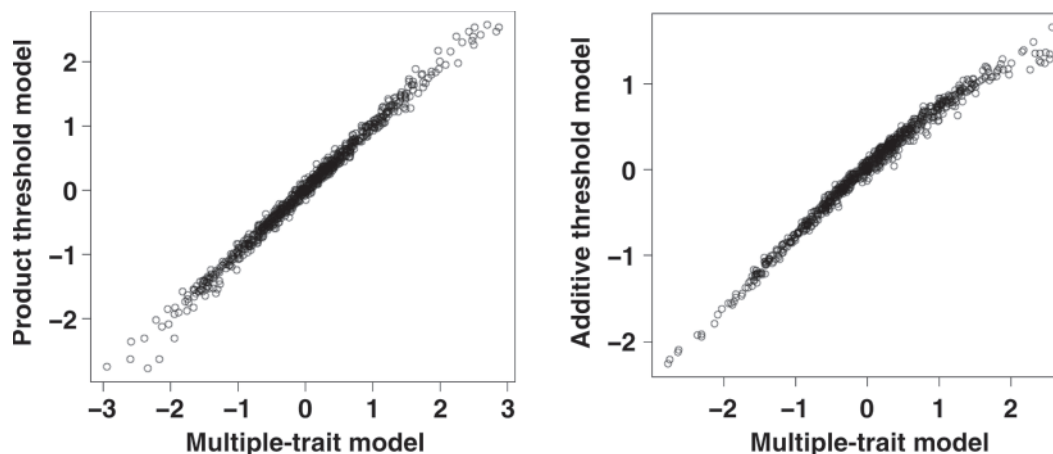


Figure 3. Correlation between direct breeding value obtained with product and additive threshold models in design 1 (200 individuals i , 400 individuals j , balanced data, high direct genetic variance).

$\Pr(P_D = 1)$ are not identified because an infinite number of values yield the same product, $\Pr(P_S = 1) \times \Pr(P_D = 1)$. However, the explanatory structure in the probit link functions are distinct for the 2 components, and further, the “male” and “female” genetic effects have different covariance matrices. This led to parameter identification in the cases examined. In short, our illustration results suggest that identification was attained, even though improper priors were adopted for fixed effects and variance components.

The product threshold model was proposed for studying reproduction results under a double assumption: the observed phenotype is the product of 2 unobserved phenotypes and these unobserved events are conditionally independent. When events are not independent, $\Pr(P_{ij}^* = 1) = \Pr(P_{D,i}^* = 1 | P_{S,j}^* = 1) \times \Pr(P_{S,j}^* = 1)$. In this case, expressing conditional probabilities as a function of genetic and environmental factors is not trivial and needs further investigations. On the other hand, it may be of interest to include a third unobserved phenotype affecting the reproduction result: the embryo survival effect. Even if, theoretically, the product threshold model can consider more than 2 unobserved phenotypes, including a third one could induce convergence problems. Actually, it is well known that the additive threshold model can produce biased estimates of parameters when the amount of data per level of fixed effect is small, known as the extreme category problem (Miształ et al., 1989; Moreno et al., 1997). The risk of an extreme category problem is greater in the product threshold model, where the probabilities of success of the unobserved phenotypes are in essence greater than the probability of success of the observed phenotype. This problem increases as the number of unobserved phenotypes increases. Because P_D^* and P_S^* are unobservable, it does not seem possible to investigate a priori this kind of problem.

The illustrations aimed to investigate whether or not the product threshold model is feasible and to evaluate the importance of bias from fitting a (wrong) ad-

ditive threshold model when observations come from a product threshold model. Even though different designs (balanced vs. unbalanced data; low, medium, or high variances; low vs. medium correlation; scant or considerable information per individual) were considered, these represent only a few of the many possible data scenarios. Due to heavy computing time requirement (2 wk per estimation), there were no replications of each design. Therefore, further investigations are needed to evaluate properties of the product threshold model. Nevertheless, we obtained similar results for simulations with some replications and fewer observations (8,000).

The product threshold model can extract more information from the data than the additive threshold model, provided that it holds. Also, it is able to estimate the effect of an environmental factor on either the direct or associative unobserved phenotypes, whereas the additive threshold model only allows estimating a global effect on the observed phenotype. Our results indicated that the additive threshold model tended to temper the effect of a factor by combining the effects on the direct and associative phenotype. The second level of had a slightly positive effect on the observed phenotype in comparison with its first level in the additive threshold model, whereas it had a positive effect on the direct phenotype and a negative one on the associative phenotype. Therefore, strategies for improving P^* may be different in the 2 models. For instance, AI success declines with time in some species (Mackey et al., 2007). In species where fresh semen is used, a product model would estimate year effects on male and female fertility separately and then identify which sex is responsible for the decline. Subsequently, fertility research could be focused on this sex.

Discrepancy between best and predicted breeding values with the additive threshold model was especially marked for individuals with increased breeding values. Consequently, fitting an additive threshold model when data are generated under the product threshold model assumption would lead to a less effective evaluation of

the best animals in the population. This issue requires additional study.

The objective of the study was to propose an alternative model to the additive threshold model for studying reproduction result in an animal genetics context to evaluate if estimation procedures can work well on this model and to evaluate if results obtained with this model differ from those obtained with the classical approach. Because parameters seem estimable, the next step of the analysis would be to compare the predictive ability of the 2 models on real data to provide validation of the product threshold model.

LITERATURE CITED

- Bijma, P., W. M. Muir, and J. A. M. van Arendonk. 2007. Multilevel selection 1: Quantitative genetics of inheritance and response to selection. *Genetics* 175:277–288.
- Carlin, B. P., and T. A. Louis. 2000. *Bayes and Empirical Bayes Methods for Data Analysis*. 2nd ed. Chapman and Hall/CRC Press, Boca Raton, FL.
- Dempster, E. R., and I. M. Lerner. 1950. Heritability of threshold characters. *Genetics* 35:212–236.
- Dukic, V., and J. Hogan. 2002. A hierarchical Bayesian approach to modeling embryo implantation following in vitro fertilization. *Biostatistics* 3:361–377.
- Gianola, D. 1982. Theory and analysis of threshold characters. *J. Anim. Sci.* 54:1079–1096.
- Gianola, D., and J. L. Foulley. 1983. Sire evaluation for ordered categorical data with a threshold model. *Genet. Sel. Evol.* 15:201–224.
- Gilmour, A. R., R. D. Anderson, and A. C. Rae. 1985. The analysis of binomial data by a generalized linear mixed model. *Biometrika* 72:593–599.
- Harville, D., and R. Mee. 1984. A mixed-model procedure for analyzing ordered categorical data. *Biometrics* 40:393–408.
- Hoeschele, I., and B. Tier. 1995. Estimation of variance components of threshold characters by marginal posterior modes and means via Gibbs sampling. *Genet. Sel. Evol.* 27:519–540.
- Koerhuis, A. N. M., and R. Thompson. 1997. Models to estimate maternal effects for juvenile body weight in broiler chickens. *Genet. Sel. Evol.* 29:225–249.
- Mackey, D. R., A. W. Gordon, M. A. McCoy, M. Verner, and C. S. Mayne. 2007. Associations between genetic merit for milk production and animal parameters and the fertility performance of dairy cows. *Animal* 1:29.
- Misztal, I., D. Gianola, and J. L. Foulley. 1989. Computing aspects of a nonlinear method of sire evaluation for categorical data. *J. Dairy Sci.* 72:1557–1568.
- Moreno, C., D. Sorensen, L. A. Garcia-Cortes, L. Varona, and J. Al-tarriba. 1997. On biased inferences about variance components in the binary threshold model. *Genet. Sel. Evol.* 29:145–160.
- Piles, M., O. Rafel, J. Ramon, and L. Varona. 2005. Genetic parameters of fertility in two lines of rabbits with different reproductive potential. *J. Anim. Sci.* 83:340–343.
- Robinson, D. L. 1996. Estimation and interpretation of direct and maternal genetic parameters for weights of Australian Angus cattle. *Livest. Prod. Sci.* 45:1–11.
- Sorensen, D., S. Andersen, D. Gianola, and I. Korsgaard. 1995. Bayesian inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.* 27:229–249.
- Sorensen, D., and D. Gianola. 2002. *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics*. Springer-Verlag, New York, NY.
- Speirs, A. L., A. Lopata, M. J. Grnow, G. N. Kellow, and W. I. H. Johnston. 1983. Analysis of the benefits and risks of multiple embryo transfer. *Fertil. Steril.* 39:468–471.
- Varona, L., and J. L. Noguera. 2001. Variance components of fertility in Spanish Landrace pigs. *Livest. Prod. Sci.* 67:217–221.
- Wilham, R. L. 1963. The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* 19:18–27.
- Wright, S. 1934. An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* 19:506–536.
- Zhou, H., and C. R. Weinberg. 1996. Modeling conception as an aggregated Bernoulli outcome with latent variables via the EM algorithm. *Biometrics* 52:945–954.

References

This article cites 20 articles, 7 of which you can access for free at:
<http://www.journalofanimalscience.org/content/87/8/2510#BIBL>