

# Heritable Environmental Variance Causes Nonlinear Relationships Between Traits: Application to Birth Weight and Stillbirth of Pigs

Herman A. Mulder,<sup>\*,\*1</sup> William G. Hill,<sup>†</sup> and Egbert F. Knol<sup>‡</sup>

<sup>\*</sup>Animal Breeding and Genomics Centre, Wageningen University, 6700 AH Wageningen, The Netherlands, <sup>†</sup>Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom, and <sup>‡</sup>TOPIGS Research Center IPG BV, 6640 AA Beuningen, The Netherlands

**ABSTRACT** There is recent evidence from laboratory experiments and analysis of livestock populations that not only the phenotype itself, but also its environmental variance, is under genetic control. Little is known about the relationships between the environmental variance of one trait and mean levels of other traits, however. A genetic covariance between these is expected to lead to nonlinearity between them, for example between birth weight and survival of piglets, where animals of extreme weights have lower survival. The objectives were to derive this nonlinear relationship analytically using multiple regression and apply it to data on piglet birth weight and survival. This study provides a framework to study such nonlinear relationships caused by genetic covariance of environmental variance of one trait and the mean of the other. It is shown that positions of phenotypic and genetic optima may differ and that genetic relationships are likely to be more curvilinear than phenotypic relationships, dependent mainly on the environmental correlation between these traits. Genetic correlations may change if the population means change relative to the optimal phenotypes. Data of piglet birth weight and survival show that the presence of nonlinearity can be partly explained by the genetic covariance between environmental variance of birth weight and survival. The framework developed can be used to assess effects of artificial and natural selection on means and variances of traits and the statistical method presented can be used to estimate trade-offs between environmental variance of one trait and mean levels of others.

**KEYWORDS** genetics of environmental variance; optimum trait; fitness; complex traits; mass selection

**I**N the classic model of quantitative genetics, the phenotype is assumed to be the sum of an effectively infinite number of genes and environmental effects; *i.e.*,  $P = G + E$  (Falconer and Mackay 1996; Lynch and Walsh 1998). The phenotypic variance is the sum of the genetic ( $V_G$ ) and environmental ( $V_E$ ) variance. It is assumed that the environmental variance is homogeneous across genotypes, but recently there is evidence from laboratory experiments and analysis of livestock populations that not only the phenotype itself, but also its  $V_E$ , is under genetic control (Hill and Mulder 2010). This might be expressed as genetic heterogeneity in the within-

individual variance of a trait with repeat measurements, such as weight of individual piglets in a litter or total litter weight across parities, or as a difference in within-family variance for traits such as juvenile body weight expressed once.

The median genetic coefficient of variation for  $V_E$  ( $GCV_{V_E} = \sigma_{A_v}/V_E$ , where  $\sigma_{A_v}$  is additive genetic variance in  $V_E$ ) is  $\sim 0.3$  based on a review of 14 studies with estimates of genetic variance in  $V_E$  primarily on livestock populations (Hill and Mulder 2010). This indicates that  $V_E$  could be increased or decreased by 30% if changed by one genetic standard deviation. These analyses were mainly based on genetic analysis of pedigreed livestock populations; but with 300 isofemale lines of *Drosophila melanogaster*, thereby avoiding possible confounding by within-line genetic heterogeneity, Mackay and Lyman (2005) showed a significant genetic variance in CV and  $V_E$ . In a selection experiment in rabbits, selection on estimated breeding value for  $V_E$  of birth weight successfully changed the within-litter variability of

Copyright © 2015 by the Genetics Society of America  
doi: 10.1534/genetics.114.173070

Manuscript received November 26, 2014; accepted for publication January 23, 2015; published Early Online January 27, 2015.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.173070/-/DC1>.

<sup>1</sup>Corresponding author: Animal Breeding and Genomics Centre, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands.  
E-mail: han.mulder@wur.nl

birth weight (Garreau *et al.* 2008). There are more recent data on genetic control of  $V_E$  at the level of individual loci from genome-wide association studies (see review by Geiler-Samerotte *et al.* 2013). Those in experimental species are very clear cut: for example, Jimenez-Gomez *et al.* (2011) and Nelson *et al.* (2013) located QTL for phenotypic variability in gene expression in *Arabidopsis* and yeast, respectively, and Shen *et al.* (2012) identified regions controlling the phenotypic variability of traits in *Arabidopsis*. In vertebrates, the FTO genotype was shown to be associated with phenotypic variability in body mass index in humans, with the two homozygotes differing 7% in standard deviation (SD) (Yang *et al.* 2012). In laying hens, a QTL region on chromosome 4 was identified that explained 30% and 16% of the genetic variance in mean and within-individual SD, respectively, of egg weight (Wolc *et al.* 2012); and in dairy cattle the phenotypic variability of somatic cell score, an indicator for mastitis, was shown to be determined by many loci, while that with the largest effect explained ~3% of the genetic variance in  $V_E$  (Mulder *et al.* 2013a). It has been argued that findings such as those of Yang *et al.* (2012) can also be explained in terms of a scale effect (Sun *et al.* 2013), but the extreme transformations required are not biologically plausible (Shen and Rönnegård 2013).

In agriculture, there is interest in reducing phenotypic variance to improve uniformity of animals and animal products. For instance, penalties are imposed by slaughterhouses for meat animals delivered with excessive variability because uniform products are demanded by retailers and packing efficiency of products is higher (Hennessy 2005). It has been shown both in theory (SanCristobal-Gaudy *et al.* 1998; Sorensen and Waagepetersen 2003; Mulder *et al.* 2007, 2008) and from experiment (Garreau *et al.* 2008) that uniformity can be increased by selection.

Before implementing selection to reduce  $V_E$  in livestock breeding programs it is necessary to know what correlated changes in means of the target trait and other traits of interest would be expected. A particular concern is that animals with small  $V_E$  would not have the capacity to react to unpredictable environments, *i.e.*, have reduced plasticity. There is some evidence that plasticity and  $V_E$  are positively correlated, *i.e.*, that genotypes with higher  $V_E$  have higher plasticity (Mulder *et al.* 2013b; Tonsor *et al.* 2013). Whether plastic genotypes with high  $V_E$  or stable genotypes with small  $V_E$  are better capable of handling unpredictable environments is under debate. In livestock little is known about trade-offs between  $V_E$  and means of other traits, except that uniformity of birth weight in litter-bearing animals does increase the survival of the offspring, *e.g.*, in pigs (Kapell *et al.* 2011) and rabbits (Garreau *et al.* 2008).

Genetic variation in  $V_E$  of a trait has been shown analytically to lead to a nonlinear relationship between the breeding value of  $V_E$  and the phenotypic deviation from the mean (Mulder *et al.* 2007). Therefore, if there is a genetic correlation between the mean of one trait and  $V_E$  of another,

which we term an M-V genetic correlation between traits, the phenotypic relationship between them is expected to be nonlinear. Nonlinear parent-offspring regression and nonlinear regression of one trait on another within half-sib groups have been employed to explore nonlinearity between traits by Sölkner and Fuerst-Waltl (1996) and Fuerst-Waltl *et al.* (1997, 1998), who assumed this was caused by nonlinearity between additive genetic effects of different traits. They also showed that phenotypic relationships were often more linear than the genetic relationships due to dilution with environmental covariance. Nonlinear relationships might also be expected when one of the traits, for example fitness, has an optimum value (Lande and Arnold 1983; Falconer and Mackay 1996; Kingsolver *et al.* 2001, 2012). Such nonlinear relationships are often observed or hypothesized in natural populations, typically as “fitness profiles” between trait values and fitness (*e.g.*, Falconer and Mackay 1996) that show whether phenotypes on one side of the distribution (directional), in the middle (stabilizing), or at extremes (disruptive) have highest fitness. If there is an optimum phenotype, a maximum would imply a negative covariance between  $V_E$  and fitness; *i.e.*, families with small variance would have selective advantage if there is stabilizing selection, and correspondingly families with a large variance would have an advantage if there is disruptive or strong directional selection (Hill and Zhang 2004; Mulder *et al.* 2007). Examples of nonlinear trait relationships include that between birth weight and survival of infants in humans in historic data when medical care was less optimized (Karn and Penrose 1951 in Schluter 1988), survival and size in house sparrows (Schluter and Smith 1986; Schluter 1988), and between litter size and birth weight in piglets (Varona and Sorensen 2014). These nonlinear relationships might be due in part to a nonnormality of at least one of the traits such as survival (binomial). To our knowledge no studies have yet been undertaken of the degree to which nonlinearity between traits is caused by an M-V genetic correlation between them. It is possible that low  $V_E$  may yield better fitness in some environments (a conservative bet-hedging strategy) or high  $V_E$  in others (a coin-flipping strategy) (Olofsson *et al.* 2009). Therefore, while insight into the nonlinear relationships between traits should increase understanding of the existence and if so of the position of phenotypic and genetic optima, no framework is available to study the role of M-V genetic correlations between traits and the consequent effects of selection.

The objectives this article are to (1) derive mathematically nonlinear relationships between traits when one exhibits genetic variance in  $V_E$ , (2) predict responses to selection in mean and variance of both traits, (3) predict genetic and phenotypic optima, (4) predict changes in genetic parameters consequent on selection shifting the population mean toward the fitness optimum, and (5) investigate the relationship between piglet birth weight and survival in pigs as a practical application of the theory. We assumed

multivariate normality for deriving the expressions, whereas survival is binomially distributed. Therefore, we extended the predictions toward assuming that survival has an underlying normal distribution and predictions can be transformed to the observed binomial scale.

## Quantitative Genetic Model

Consider two traits for which the phenotypic deviations from the mean are the sum of additive genetic and environmental effects, with no dominance or epistasis (Falconer and Mackay 1996). The first trait, with phenotypic value  $P_1$ , is also subject to additive genetic differences in its environmental variance, so-called genetic heterogeneity of environmental variance (SanCristobal-Gaudy *et al.* 1998), whereas the second trait (phenotypic value  $P_2$ ) has homogeneous environmental variance. Here we assume the exponential model for environmental variance (SanCristobal-Gaudy *et al.* 1998; Hill and Mulder 2010), because of its statistical tractability, as the environmental variance cannot be negative and the model appears linear on the log scale. Different models to model genetic differences in environmental variance appear similar when genetic variance in  $V_E$  is small. Further discussion and comparison of models for genetic analysis of the environmental variance are given elsewhere (Hill and Mulder 2010). The genetic model is

$$P_1 = \mu_1 + A_{m,1} + E_1 = \mu_1 + A_{m,1} + \chi \exp\left[\frac{1}{2}\left(\ln(\sigma_{e_1}^2) + A_{v,1}\right)\right] \quad (1)$$

$$P_2 = \mu_2 + A_{m,2} + E_2, \quad (2)$$

where  $\mu_i$  is the population mean,  $A_{m,i}$  is the breeding value, and  $E_i$  is the environmental deviation for trait  $i$ . The environmental variance for trait 2 is assumed constant,  $\sigma_{e_2}^2$ , whereas  $\sigma_{e_1}^2$  is the expected environmental variance for trait 1 when  $A_{v,1} = 0$ , but varies among individuals according to their breeding value  $A_{v,1}$ . The standard normal deviate  $\chi$  is the standardized environmental deviation for trait 1. The breeding values  $A_{m,1}$ ,  $A_{v,1}$ , and  $A_{m,2}$  are assumed to be the sum of effects of an infinite number of additive loci and are multivariate normally distributed,

$$\begin{bmatrix} A_{m,1} \\ A_{v,1} \\ A_{m,2} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_{A_{m,1}}^2 & \text{cov}_{A_{m,1}, A_{v,1}} & \text{cov}_{A_{m,1}, A_{m,2}} \\ 0 & \sigma_{A_{v,1}}^2 & \text{cov}_{A_{v,1}, A_{m,2}} \\ \text{symmetric} & & \sigma_{A_{m,2}}^2 \end{bmatrix} \otimes \mathbf{A}\right),$$

where  $\mathbf{A}$  is the additive genetic or numerator relationship matrix;  $\sigma_{A_{m,1}}^2$ ,  $\sigma_{A_{v,1}}^2$ , and  $\sigma_{A_{m,2}}^2$  are the additive genetic variances in  $A_{m,1}$ ,  $A_{v,1}$ , and  $A_{m,2}$ ; and  $\text{cov}_{A_{m,1}, A_{v,1}}$ ,  $\text{cov}_{A_{m,1}, A_{m,2}}$ , and  $\text{cov}_{A_{v,1}, A_{m,2}}$  are the corresponding additive genetic covariances. The environmental effects  $E_1 (= \chi \exp[(1/2)(\ln(\sigma_{e_1}^2) + A_{v,1}])$ ) and  $E_2$  are bivariate normally distributed,

$$\begin{bmatrix} E_1 \\ E_2 \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_{e_1}^2 \exp(A_{v,1}) & r_e \sigma_{e_1} \exp\left(\frac{1}{2}A_{v,1}\right) \sigma_{e_2} \\ \text{symmetric} & \sigma_{e_2}^2 \end{bmatrix}\right),$$

where  $\sigma_{e_2}^2$  is the environmental variance of trait 2 and  $r_e$  is the environmental correlation.

## Methodological Framework for Nonlinear Relationships

### Multiple-regression framework

We used a multiple-regression framework to predict the nonlinear relationship between traits caused by M-V genetic correlations between them and to predict responses in mean and variance due to selection, following the methodology of Mulder *et al.* (2007). In Mulder *et al.* (2007), breeding values and selection responses for mean and variance were predicted using linear, quadratic, and cubic regression on the phenotypic deviation of an individual phenotype. This methodology formally requires multivariate normality, which holds only approximately because genetic variance in  $V_E$  for  $P_1$  causes small deviations from normality (Mulder *et al.* 2007). We also used selection index theory (Hazel 1943), which is essentially an application of multiple regression assuming that fixed effects are known without error.

### Monte Carlo simulation

To check the accuracy of derived relationships and predicted responses to selection, we simulated 1000 replicates each with 1 million animals according to the genetic models in Equations 1 and 2 and calculated the expected value of trait 2 given trait 1 by calculating the average  $P_2$  given  $P_1$  within 601 successive intervals of  $0.01\sigma_{P,1}$  between  $-3$  and  $3\sigma_{P,1}$ . Selection responses to directional mass selection on  $P_1$  and  $P_2$  were calculated as the mean  $A_{m,1}$ ,  $A_{v,1}$ , and  $A_{m,2}$  of all selected animals with  $P \geq x$  in each replicate and averaged over replicates, where  $x$  is the truncation point.

### Deriving the phenotypic relationship between $P_2$ and $P_1$

The phenotypic relationship between  $P_2$  and  $P_1$  was derived as the multiple regression of  $P_2$  on  $P_1$  and  $P_1^2$ , because genetic variance in  $V_E$  causes genetic variance in  $P_1^2$  (Mulder *et al.* 2007). The regression can be written as

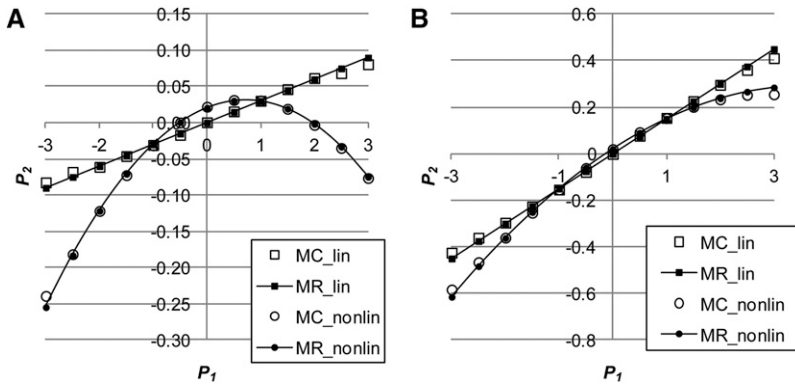
$$E(P_2 | P_1, P_1^2) = b_1 P_1 + b_2 P_1^2, \quad (3)$$

where

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \mathbf{P}^{-1} \mathbf{c}, \quad (4)$$

$$\mathbf{P} = \begin{bmatrix} \text{var}(P_1) & \text{cov}(P_1, P_1^2) \\ \text{cov}(P_1, P_1^2) & \text{var}(P_1^2) \end{bmatrix},$$

and



**Figure 1** (A and B) The relationship between  $P_2$  and  $P_1$  either assuming no genetic correlation ( $r_{A_{v,1}, A_{m,2}} = 0$ ) between environmental variance in  $P_1$  and the mean of  $P_2$  using Monte Carlo simulation (MC\_lin) and multiple regression (MR\_lin) or assuming a negative genetic correlation ( $r_{A_{v,1}, A_{m,2}} = -0.5$ ) between them (MC\_nonlin, MR\_nonlin) when  $r_{A_{m,1}, A_{m,2}} = 0.1$  (A) or when  $r_{A_{m,1}, A_{m,2}} = 0.5$  (B) ( $\sigma_{A_{m,1}}^2 = \sigma_{A_{m,2}}^2 = 0.3$ ,  $\sigma_{A_{v,1}}^2 = 0.05$ ,  $\sigma_{e_1}^2 \exp((1/2)\sigma_{A_{v,1}}^2) = 0.7$ ,  $\sigma_{e_2}^2 = 0.7$ ,  $r_{A_{m,1}, A_{m,2}} = 0.1$ ,  $r_{A_{m,1}, A_{v,1}} = 0$ ,  $r_e = 0$ ).

$$\mathbf{c} = \begin{bmatrix} \text{cov}(P_2, P_1) \\ \text{cov}(P_2, P_1^2) \end{bmatrix}.$$

The elements of  $\mathbf{P}$  and  $\mathbf{c}$  were derived using the higher-order moments of the normal distribution (Stuart and Ord 1994) and using the laws of total (co)variance and standard variance–covariance rules (see Appendix A for full derivations),

$$\text{var}(P_1) = \sigma_{A_{m,1}}^2 + \sigma_{e_1}^2 \exp\left(\frac{1}{2}\sigma_{A_{v,1}}^2\right) = \sigma_{P_1}^2 \quad (5)$$

$$\text{cov}(P_1, P_1^2) = 3 \text{cov}_{A_{m,1}, A_{v,1}} \sigma_{e_1}^2 \exp\left(\frac{1}{2}\sigma_{A_{v,1}}^2\right) \quad (6)$$

$$\begin{aligned} \text{var}(P_1^2) &= 2\sigma_{P_1}^4 + 6 \text{cov}_{A_{m,1}, A_{v,1}}^2 \sigma_{e_1}^2 \exp\left(\frac{1}{2}\sigma_{A_{v,1}}^2\right) \\ &\quad + 3\sigma_{e_1}^4 \exp\left(\sigma_{A_{v,1}}^2\right) \left(\exp\left(\sigma_{A_{v,1}}^2\right) - 1\right) \end{aligned} \quad (7)$$

$$\text{cov}(P_2, P_1) = \text{cov}_{A_{m,1}, A_{m,2}} + \text{cov}_{e_1 e_2} \quad (8)$$

$$\begin{aligned} \text{cov}(P_2, P_1^2) &= \text{cov}_{A_{v,1}, A_{m,2}} \sigma_{e_1}^2 \exp\left(\frac{1}{2}\sigma_{A_{v,1}}^2\right) \\ &\quad + \text{cov}_{e_1 e_2} \text{cov}_{A_{m,1}, A_{v,1}}, \end{aligned} \quad (9)$$

where  $\text{cov}_{e_1 e_2} = r_e \sigma_{e_1} \exp((1/4)\sigma_{A_{v,1}}^2) \sigma_{e_2}$ . Note that Equation 3 is about the relationship between single observations of  $P_2$  and  $P_1$ . In Equation 7, the two rightmost terms are related to the skewness (Equation 6) and kurtosis of the distribution, respectively. All elements of  $\mathbf{P}$  and  $\mathbf{c}$  (equal to Equations 5–9) were checked using Monte Carlo simulation. The multiple-regression prediction of  $P_2$  from  $P_1$  shows very good agreement with the Monte Carlo simulation (Figure 1 and Figure 2), with  $R^2 > 0.995$  for all cases evaluated. Appendix C shows a natural extension to multiple random effects, such as permanent and common environmental effects.

### Deriving response to mass selection

**Deterministic predictions:** Response to selection ( $R$ ) is predicted as the mean deviation of offspring from the population mean regressed on the selection differential of the parents  $R = bS$ , which is equal to the regression of the

breeding value ( $A$ ) on the selection differential ( $S$ )  $\Delta A = bS$  (Falconer and Mackay 1996). We showed previously that response to selection in  $A_{m,1}$  and  $A_{v,1}$  could be predicted for different kinds of mass selection, e.g., directional, stabilizing, or disruptive (Mulder *et al.* 2007). Here we extend this framework to the correlated responses in (1)  $P_2$  from mass selection on  $P_1$  and (2)  $A_{m,1}$  and  $A_{v,1}$  from mass selection on  $P_2$ . In both cases, the general equation to predict selection responses in

$$\Delta \mathbf{a} = \begin{bmatrix} \Delta A_{m,1} \\ \Delta A_{v,1} \\ \Delta A_{m,2} \end{bmatrix}$$

in terms of selection differentials ( $\Delta \mathbf{x}$ ) is

$$\Delta \mathbf{a} = \mathbf{P}^{-1} \mathbf{G} \Delta \mathbf{x} = \mathbf{B}' \Delta \mathbf{x}, \quad (10)$$

where  $\mathbf{P}$  is the variance–covariance matrix among selection differentials in  $\Delta \mathbf{x}$ , because selection differentials are random variables under the quantitative genetic model assumed;  $\mathbf{G}$  is the covariance matrix between  $\Delta \mathbf{a}$  and  $\Delta \mathbf{x}$ ; and  $\mathbf{B}$  is a matrix of regression coefficients. First, if mass selection on  $P_1$  is performed, then the selection differentials on  $P_1$  and  $P_1^2$  are

$$\Delta \mathbf{x} = \begin{bmatrix} \overline{P_{1,s}} - \mu_1 \\ \overline{P_{1,s}^2} - \sigma_{P_1}^2 \end{bmatrix} = \begin{bmatrix} \overline{P_{1,s}} - E(P_1) \\ (\overline{P_{1,s}} - E(P_1))^2 - E\{(P_1 - E(P_1))^2\} \end{bmatrix}. \quad (11)$$

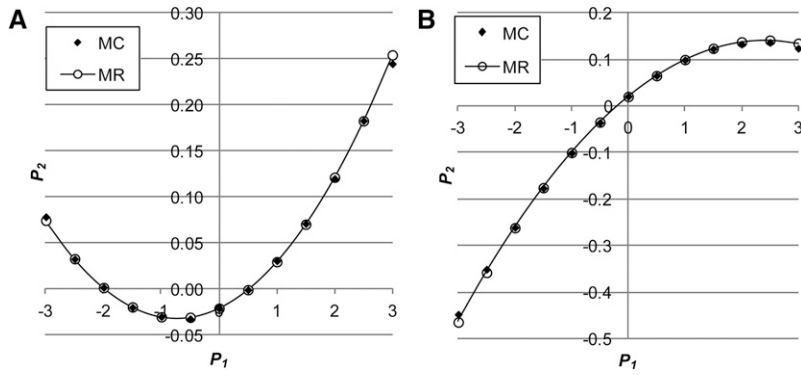
If it is assumed that  $\mu_1 = 0$ ,

$$\begin{bmatrix} \Delta A_{m,1} \\ \Delta A_{v,1} \\ \Delta A_{m,2} \end{bmatrix} = \mathbf{B}'_{P_1} \begin{bmatrix} \overline{P_{1,s}} \\ \overline{P_{1,s}^2} - \sigma_{P_1}^2 \end{bmatrix} = \mathbf{B}'_{P_1} \begin{bmatrix} i\sigma_{P_1} \\ ix\sigma_{P_1}^2 \end{bmatrix}, \quad (12)$$

where  $\mathbf{B}'_{P_1}$  is the transpose of  $\mathbf{B}_{P_1} = \mathbf{P}^{-1} \mathbf{G}_{P_1}$ ,  $\mathbf{P}$  is the same as in Equation 3, and

$$\mathbf{G}_{P_1} = \begin{bmatrix} \sigma_{A_{m,1}}^2 & & \\ & \text{cov}_{A_{m,1}, A_{v,1}} & \\ & & \text{cov}_{A_{m,1}, A_{m,2}} \end{bmatrix},$$

i.e., the covariance matrix between



**Figure 2** (A and B) The relationship between  $P_2$  and  $P_1$ , using Monte Carlo simulation (MC) and multiple regression (MR) either with positive genetic correlation (A:  $r_{A_{v,1}, A_{m,2}} = 0.5, r_e = 0$ ) between environmental variance in  $P_1$  and the mean of  $P_2$  or with negative genetic correlation (B: when  $r_{A_{v,1}, A_{m,2}} = -0.5, r_e = 0.1$ ) ( $\sigma_{A_{m,1}}^2 = \sigma_{A_{m,2}}^2 = 0.3, \sigma_{A_{v,1}}^2 = 0.05, \sigma_{e_1}^2 \exp((1/2)\sigma_{A_{v,1}}^2) = 0.7, \sigma_{e_2}^2 = 0.7, r_{A_{m,1}, A_{m,2}} = 0.1, r_{A_{m,1}, A_{v,1}} = 0$ ).

$$\begin{bmatrix} \overline{P_{1,s}} \\ \overline{P_{1,s}^2} - \sigma_{P_1}^2 \end{bmatrix} \text{ and } \begin{bmatrix} \Delta A_{m,1} \\ \Delta A_{v,1} \\ \Delta A_{m,2} \end{bmatrix},$$

$i$  is the selection intensity when selecting the best  $p$  animals, and  $x$  is the truncation point corresponding to  $p$  based on a normal distribution (Mulder *et al.* 2007).

Second, if mass selection is on  $P_2$ , Equation 10 becomes

$$\begin{bmatrix} \Delta A_{m,1} \\ \Delta A_{v,1} \\ \Delta A_{m,2} \end{bmatrix} = \mathbf{b}'_{P_2} [\overline{P_{2,s}}] = \mathbf{b}'_{P_2} i \sigma_{P_2} \quad (13)$$

where  $\mathbf{b}_{P_2} = \text{var}(P_2)^{-1} \mathbf{g}_{P_2}$  and  $\mathbf{g}_{P_2} = \begin{bmatrix} \text{cov}_{A_{m,1}A_{m,2}} & \text{cov}_{A_{v,1}A_{m,2}} \\ \sigma_{e_1}^2 \exp((1/2)\sigma_{A_{v,1}}^2) & \sigma_{A_{m,2}}^2 \end{bmatrix}$ , the covariance vector between  $\overline{P_{2,s}}$  and

$$\begin{bmatrix} \Delta A_{m,1} \\ \Delta A_{v,1} \\ \Delta A_{m,2} \end{bmatrix}.$$

**Evaluation of predictions:** Genetic selection differentials in  $A_{m,1}$ ,  $A_{v,1}$ , and  $A_{m,2}$  with directional mass selection on either  $P_1$  (Equation 12) or  $P_2$  (Equation 13) were shown by Monte Carlo simulation to be very accurately predicted, with differences only at the third decimal (Table 1). Correlated responses in both  $A_{v,1}$  and  $A_{m,2}$  are expected following mass selection on  $P_1$ , even if genetic correlations between  $A_{m,1}$  and  $A_{v,1}$  or  $A_{m,2}$  are zero, provided that the genetic correlation between  $A_{v,1}$  and  $A_{m,2}$  is nonzero. This is because, with strong directional selection (e.g.,  $P \geq 20\%$ ), animals with higher  $A_{v,1}$  (i.e., environmental variance in  $P_1$ ) have a higher probability of selection (Hill and Zhang 2004; Mulder *et al.* 2007). In other words, with directional selection genotypes with a higher variance are more likely to be parents of the next generation than genotypes with a low variance.

#### Response to selection on estimated breeding value

In animal breeding, selection is often on predicted breeding values rather than on raw phenotypes. For instance, if the breeding goal is to increase trait 2, then the predicted breeding value  $\hat{A}_{m,2}$  using information of  $P_1$  and  $P_1^2$  is

$$\hat{A}_{m,2} = \mathbf{b}'_{a_2} \begin{bmatrix} P_1 - \mu_1 \\ P_1^2 - \sigma_{P_1}^2 \end{bmatrix}, \quad (14)$$

where  $\mathbf{b}_{a_2} = \mathbf{P}^{-1} \mathbf{g}_{a_2}$ ,  $\mathbf{P}$  is given in Equation 3, and

$$\mathbf{g}_{a_2} = \begin{bmatrix} \text{cov}_{A_{m,1}, A_{m,2}} \\ \text{cov}_{A_{v,1}, A_{m,2}} \sigma_{e_1}^2 \exp\left(\frac{1}{2} \sigma_{A_{v,1}}^2\right) \end{bmatrix}.$$

The expected response to selection would be equal to that using the generalized form of the breeders equation (Hazel 1943):

$$\Delta A_{m,2} = \frac{i \mathbf{b}'_{a_2} \mathbf{g}_{a_2}}{\sqrt{\mathbf{b}'_{a_2} \mathbf{P}^{-1} \mathbf{b}_{a_2}}}. \quad (15)$$

An alternative to using squared phenotypic deviations is to use squared residuals of  $P_1$  corrected for its leverage ( $l$ ), which is the diagonal element of the “hat matrix” for each observation (Hoaglin and Welsh 1978) as used in the double-hierarchical generalized linear model (DHGLM) (Rönnegård *et al.* 2010):

$$\frac{e_1^2}{(1-l)} = \frac{[(1-b_{A_{m,1}})P_1]^2}{[1-l]}. \quad (16)$$

In the absence of any fixed effects,  $1-l = 1-b_{A_{m,1}}$ , Equation 16 reduces to  $e_1^2/(1-l) = (1-b_{A_{m,1}})P_1^2$  and

$$\hat{A}_{m,2} = \mathbf{b}'_{a_2e} \begin{bmatrix} P_1 \\ e_1^2/(1-l) \end{bmatrix}, \quad (17)$$

where  $\mathbf{b}_{a_2e} = \mathbf{P}_e^{-1} \mathbf{g}_{a_2e}$ ,

$$\mathbf{P}_e = \begin{bmatrix} \text{var}(P_1) & \text{cov}(P_1, e_1^2/(1-l)) \\ \text{cov}(P_1, e_1^2/(1-l)) & \text{var}(e_1^2/(1-l)) \end{bmatrix} = \mathbf{W} \mathbf{P} \mathbf{W},$$

and  $\mathbf{g}_{a_2e} = \mathbf{W} \mathbf{g}_{a_2}$ , where

$$\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & (1-b_{A_{m,1}}) \end{bmatrix}$$

(see derivation in Appendix B). Use of either  $P_1^2$  or  $e_1^2/(1-l)$  would yield identical accuracy of prediction  $\hat{A}_{m,2}$  and of

**Table 1** Genetic selection differentials in  $A_{m,1}$ ,  $A_{v,1}$ , and  $A_{m,2}$  when selecting on  $P_1$  or on  $P_2$  for different selected proportions ( $p$ ) and when  $r_{A_{m,1},A_{v,1}}$  and  $r_{A_{m,1},A_{m,2}}$  are the same (0 or 0.5)

Selection	$r_{A_{m,1},A_{v,1}} = r_{A_{m,1},A_{m,2}}$	$p$	$A_{m,1}$		$A_{v,1}$		$A_{m,2}$	
			MR	MC	MR	MC	MR	MC
$P_1$	0	0.80	0.105	0.105	-0.005	-0.005	0.006	0.006
		0.50	0.239	0.242	0.000	0.000	0.000	0.000
		0.20	0.420	0.421	0.020	0.020	-0.024	-0.025
	0.5	0.80	0.104	0.103	0.017	0.017	0.063	0.061
		0.50	0.239	0.241	0.048	0.049	0.123	0.121
		0.20	0.422	0.430	0.099	0.102	0.180	0.179
$P_2$	0	0.80	0.000	0.000	-0.021	-0.021	0.105	0.105
		0.50	0.000	0.000	-0.049	-0.049	0.239	0.239
		0.20	0.000	0.000	-0.086	-0.086	0.420	0.420
	0.5	0.80	0.052	0.052	-0.021	-0.021	0.105	0.105
		0.50	0.120	0.120	-0.049	-0.049	0.239	0.239
		0.20	0.210	0.210	-0.086	-0.086	0.420	0.420

Values were obtained analytically [multiple regression (MR): Equations 12 and 13] and with Monte Carlo simulation (MC) ( $\sigma_{A_{m,1}}^2 = \sigma_{A_{m,2}}^2 = 0.3$ ,  $\sigma_{A_{v,1}}^2 = 0.05$ ,  $\sigma_{e_1}^2 \exp((1/2)\sigma_{A_{v,1}}^2) = 0.7$ ,  $\sigma_{e_2}^2 = 0.7$ ,  $r_{A_{v,1},A_{m,2}} = -0.5$ ,  $r_e = 0$ ).

response to selection  $\Delta A_{m,2}$  in this case where only their own phenotypes provide information and there are no fixed effects; otherwise squared residuals corrected for leverages are expected to give higher accuracy and response than using squared phenotypic deviations.

#### Determining whether the parabola have a maximum or a minimum

The multiple-regression Equation 3 is quadratic and therefore has a parabolic shape. Such quadratic relationships may indicate a biological optimum, for instance birth weight and survival in humans prior to modern medical intervention (Karn and Penrose 1951 in Schluter 1988). Expressions for  $b_1$  and  $b_2$  (from Equation 3) are given by

$$b_1 = \frac{[\text{var}(P_1^2)c_1 - \text{cov}(P_1, P_1^2)c_2]}{\det(\mathbf{P})} \quad (18)$$

$$b_2 = \frac{[-\text{cov}(P_1, P_1^2)c_1 + \text{var}(P_1)c_2]}{\det(\mathbf{P})}, \quad (19)$$

where  $c_1 = \text{cov}(P_2, P_1)$  (Equation 8) and  $c_2 = \text{cov}(P_2, P_1^2)$  (Equation 9) and  $\mathbf{P}$  is shown below Equation 4. Whether there is a maximum or a minimum depends on the sign of the quadratic regression coefficient ( $b_2$ ), *i.e.*,  $P_1^2$ . If  $b_2 > 0$ , it has a minimum and if  $b_2 < 0$ , it has a maximum. The sign of the regression coefficient depends on multiple parameters (Equations 18 and 19). As Figure 3 shows,  $b_2$  is linear in  $r_{A_{v,1},A_{m,2}}$  because it is a linear function of  $c_2$  (Equation 19) and  $c_2$  is the genetic covariance between  $A_{v,1}$  and  $A_{m,2}$ . When  $r_{A_{v,1},A_{m,2}} = 0$  and other correlations are zero,  $b_2 = 0$ , *i.e.*, when the straight-line relationship between  $P_2$  and  $P_1$  passes through (0, 0). If other correlations are nonzero,  $b_2 = 0$  shifts to a point dependent mainly on  $r_{A_{v,1},A_{m,2}}$ , with other parameters having less influence on  $b_2$ . In summary, a moderate to strong negative M-V genetic correlation between traits ( $r_{A_{v,1},A_{m,2}}$ ) yields a parabola with a maximum, *e.g.*, animals with phenotypes close to the optimum have

maximum fitness, a fitness optimum; while a moderate to strong positive genetic correlation yields one with a minimum; *e.g.*, animals with phenotypes largely deviating from the minimum have the highest fitness.

Similarly, Equation 14 shows whether the genetic parabola has a maximum or a minimum. The regression coefficients now are

$$b_{a_21} = \frac{[\text{var}(P_1^2)g_{a_21} - \text{cov}(P_1, P_1^2)g_{a_22}]}{\det(\mathbf{P})} \quad (20)$$

$$b_{a_22} = \frac{[-\text{cov}(P_1, P_1^2)g_{a_21} + \sigma_{P_1}^2 g_{a_22}]}{\det(\mathbf{P})}, \quad (21)$$

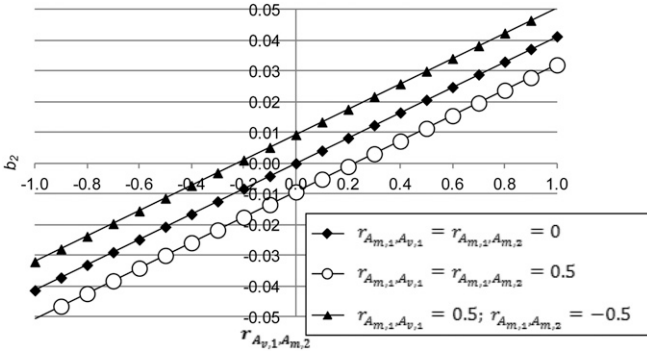
where  $g_{a_21}$  and  $g_{a_22}$  are the elements of  $\mathbf{g}_{a_2}$  (just below Equation 14). Equations 20 and 21 differ from Equations 18 and 19 only in  $g_{a_21}$  and  $g_{a_22}$ . If  $r_e = 0$ , then  $g_{a_21} = c_1$  and  $g_{a_22} = c_2$  and consequently  $b_{a_21} = b_1$  and  $b_{a_22} = b_2$ .

#### Can phenotypic and genetic parabolas have different shapes?

The interest here is in whether there are conditions when the phenotypic parabola has a maximum and the genetic parabola a minimum or vice versa. Because  $r_e$  affects only the phenotypic parabola, it determines the relative shapes of the parabolas, and so it is helpful to calculate the magnitude of  $r_e$  when  $b_2 = 0$  and switches to a different sign than that of  $b_{a_22}$ . The most obvious situation is when the environmental correlation  $r_e$  and the genetic correlation  $r_{A_{m,1},A_{m,2}}$  have opposite signs, but  $b_2 = 0$  only when  $\text{cov}_{A_{m,1},A_{v,1}} \neq 0$ . Calculations show that, for the parabolas to be of different shape,  $r_e$  has to be strong, in which case the regression is almost linear. So in practice, parabolas of opposite sign are likely to occur only rarely, if at all.

#### Determining the phenotypic and genetic stationary points

The maximum (minimum) phenotype  $P_2$  can be found at the stationary point of  $P_1$  ( $P_{1\text{stat}}$ ). The stationary point is



**Figure 3** The quadratic regression coefficient  $b_2$  as a function of the genetic correlation  $r_{A_{v,1}, A_{m,2}}$  for different values of  $r_{A_{m,1}, A_{v,1}}$  and  $r_{A_{m,1}, A_{m,2}}$  ( $\sigma_{A_{m,1}}^2 = \sigma_{A_{m,2}}^2 = 0.3$ ,  $\sigma_{A_{v,1}}^2 = 0.05$ ,  $\sigma_{e_1}^2 \exp((1/2)\sigma_{A_{v,1}}^2) = 0.7$ ,  $\sigma_{e_2}^2 = 0.7$ ,  $r_e = 0$ ).

mathematically the point where the first derivative is zero, e.g., a minimum or maximum. The stationary point of the quadratic relationship of  $P_2$  as a function of  $P_1$  can be calculated as

$$P_{1\text{stat}} = \frac{-b_1}{2b_2} = \frac{-\text{var}(P_1^2)c_1 + \text{cov}(P_1, P_1^2)c_2}{-2 \text{cov}(P_1, P_1^2)c_1 + 2\sigma_{P_1}^2 c_2}. \quad (22)$$

If we assume for simplicity that  $\text{cov}(P_1, P_1^2) = 0$  when  $\text{cov}_{A_{m,1}, A_{v,1}} = 0$ , then  $P_{1\text{stat}} = -\text{var}(P_1^2)c_1/2\sigma_{P_1}^2 c_2$ . The population mean is at a stationary point of zero when  $c_1 = r_{A_{m,1}, A_{m,2}} = 0$ . A positive (negative) stationary point implies that the current population mean is lower (higher) than it is. In general,  $P_{1\text{stat}} = 0$  if  $\text{var}(P_1^2)c_1 = \text{cov}(P_1, P_1^2)c_2$ .

The genetic stationary point can be determined similarly:

$$P_{1g\text{stat}} = \frac{-b_{a_2 1}}{2b_{a_2 2}} = \frac{-\text{var}(P_1^2)g_{a_2 1} + \text{cov}(P_1, P_1^2)g_{a_2 2}}{-2 \text{cov}(P_1, P_1^2)g_{a_2 1} + 2\sigma_{P_1}^2 g_{a_2 2}}. \quad (23)$$

The genetic and phenotypic stationary points differ unless  $r_e = 0$ . They are shown in Figure 4 as functions of  $r_{A_{v,1}, A_{m,2}}$  (Figure 4A) and  $r_{A_{m,1}, A_{m,2}}$  (Figure 4B), and although they differ, their shapes do not. Both are near zero when  $r_{A_{v,1}, A_{m,2}}$  is close to  $-1$  or  $1$ , but large in absolute terms when  $r_{A_{v,1}, A_{m,2}} \sim 0$  (Figure 4A), i.e., a hyperbolic form, because  $r_{A_{v,1}, A_{m,2}}$  largely determines the curvature of the parabolas; and both increase linearly with  $r_{A_{m,1}, A_{m,2}}$  (Figure 4B). These phenotypic and genetic stationary points can be interpreted as phenotypic and genetic optima in the case of stabilizing selection.

### Effects of selection on genetic parameters

Both natural and artificial selection may bring the population mean to a fitness optimum. As shown in Figure 4, there is a relationship between the distance between the stationary point and the current population mean and the genetic correlations, such that the genetic correlations may change as selection shifts the population toward the stationary point. The formulas underlying Figure 4 can also be expressed to show how genetic correlations depend on the

distance between the current population mean and the stationary point, e.g., a fitness optimum ( $\Omega$ ), assuming that the parabola has fixed shape. The regression equation can be written as

$$P_2 = b_2(P_1 - \Omega)^2 + d, \quad (24)$$

where  $d$  is a constant. If we assume  $d = 0$  and  $\text{cov}(P_1, P_1^2) = 0$ , then  $b_2 = c_2/\text{var}(P_1^2)$ . From Equation 22,  $b_1 = -2b_2\Omega$ , substituting  $b_2 = c_2/\text{var}(P_1^2)$  and equating to the expression for  $b_1$  in Equation 18 assuming  $\text{cov}(P_1, P_1^2) = 0$ ,

$$b_1 = \frac{-2c_2\Omega}{\text{var}(P_1^2)} = \frac{c_1}{\sigma_{P_1}^2}. \quad (25)$$

The change in genetic correlation  $r_{A_{m,1}, A_{m,2}}$  can be predicted as the population mean moves to the stationary point. For instance, if all other parameters stay equal,

$$c_1 = \frac{-2\sigma_{P_1}^2 c_2 \Omega}{\text{var}(P_1^2)} \quad (26)$$

$$\begin{aligned} r_{A_{m,1}, A_{m,2}} &= \frac{-2\sigma_{P_1}^2 c_2 \Omega}{\text{var}(P_1^2)\sigma_{A_{m,1}}\sigma_{A_{m,2}}} \\ &= \frac{-2\sigma_{P_1}^2 \Omega (\text{cov}_{A_{v,1}, A_{m,2}}\sigma_{e_1}^2 \exp((1/2)\sigma_{A_{v,1}}^2) + \text{cov}_{e_1 e_2} \text{cov}_{A_{m,1}, A_{v,1}})}{\text{var}(P_1^2)\sigma_{A_{m,1}}\sigma_{A_{m,2}}}. \end{aligned} \quad (27)$$

Equation 27 shows that  $r_{A_{m,1}, A_{m,2}}$  and  $\Omega$  are linearly related. Similarly, for the genetic correlation between  $r_{A_{v,1}, A_{m,2}}$ ,

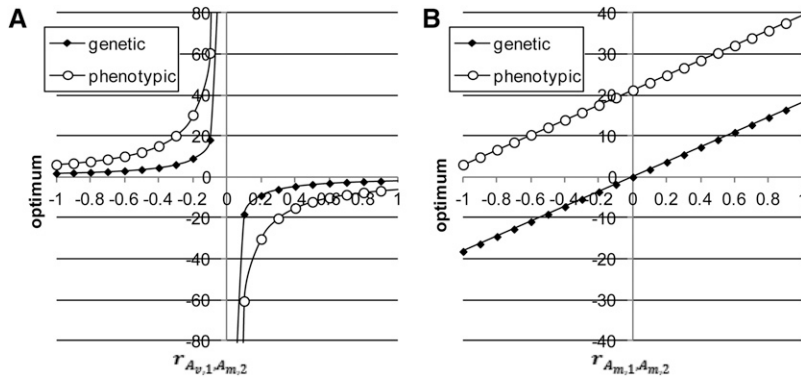
$$c_2 = \frac{c_1 \text{var}(P_1^2)}{-2\sigma_{P_1}^2 \Omega} \quad (28)$$

$$r_{A_{v,1}, A_{m,2}} = \frac{c_1 \text{var}(P_1^2)}{-2\sigma_{P_1}^2 \Omega \sigma_{A_{v,1}} \sigma_{A_{m,2}}} - \frac{\text{cov}_{e_1 e_2} \text{cov}_{A_{m,1}, A_{v,1}}}{\sigma_{A_{v,1}} \sigma_{A_{m,2}}}, \quad (29)$$

showing that  $r_{A_{v,1}, A_{m,2}}$  has a hyperbolic relationship with  $\Omega$  (Figure 4A).

### Example: index selection with changing genetic correlation $r_{A_{m,1}, A_{m,2}}$

If index selection is applied to improve trait 2 using the phenotypic deviation of trait 1 and its square, i.e., Equations 14 and 15, the mean of trait 1 will tend to move to the optimum value (Figure 5). If only  $r_{A_{m,1}, A_{m,2}}$  changes according to Equation 27, the genetic correlation will become smaller as the population mean approaches the optimum. The response in trait 1 is nearly linear up to generation seven and becomes almost zero after generation eight, whereas the response in trait 2 is curvilinear. The response in trait 2 in the first seven generations is due mainly to bringing the population mean for trait 1 to the optimum and subsequently is due to reducing the phenotypic variance in trait 1.



**Figure 4** (A and B) Phenotypic and genetic optima as functions of  $r_{A_{v,1}, A_{m,2}}$  (A) and  $r_{A_{m,1}, A_{m,2}}$  (B) when  $r_e = 0.5$  ( $\sigma_{A_{m,1}}^2 = \sigma_{A_{m,2}}^2 = 0.3$ ,  $\sigma_{A_{v,1}}^2 = 0.05$ ,  $\sigma_{e_1}^2 \exp((1/2)\sigma_{A_{v,1}}^2) = 0.7$ ,  $\sigma_{e_2}^2 = 0.7$ ,  $r_{A_{m,1}, A_{v,1}} = 0$ ; A,  $r_{A_{m,1}, A_{m,2}} = 0.5$ ; B,  $r_{A_{v,1}, A_{m,2}} = -0.2$ ).

### Deriving the phenotypic relationship between $P_2$ and $P_1$ when $P_2$ is a binary trait

Until now, we assumed multivariate normality but nonlinear relationships between traits occur when one of the traits is not normally distributed, such as survival or presence/absence of disease. Therefore we extend the prediction of  $P_2$  on  $P_1$  (Equation 3) when  $P_2$  is a 0/1 binomial trait. We assume that  $P_2$  is normally distributed on the underlying/liability scale ( $P_{2,liab}$ ) (Gianola 1982; Foulley 1992) and therefore,  $\widehat{P_{2,liab}}$  can be predicted using Equation 3. On the observed scale ( $P_{2,obs}$ ), the proportion of animals that survive (dead = 0; alive = 1) is predicted as

$$P_{2,obs} = P_{survival} = \int_{x - \widehat{P_{2,liab}}}^{\infty} \phi(t) dt, \quad (30)$$

where  $x$  is the truncation point of the normal distribution corresponding to the average survival in the population ( $\overline{P_{survival}}$ ),  $x - \widehat{P_{2,liab}}$  represents the predicted truncation point given  $P_1$ , and  $\phi(t)$  is the standard normal density function. Using Equation 30 together with Equation 3 to predict  $P_{2,obs}$  from  $P_1$  shows very good agreement with results from Monte Carlo simulation (Figure 6), with  $R^2 > 0.995$  for all cases evaluated. (Fortran scripts for all theoretical predictions and Monte Carlo simulations are provided in [Supporting Information, File S3](#)).

### Application to Birth Weight and Stillbirth in Pigs

The theoretical framework developed above shows that an M-V genetic correlation between traits leads to a nonlinear relationship between their phenotypes. The subsequent analyses demonstrated that the genetic and phenotypic optima could differ in position, but that selection could bring the population mean to the genetic optimum. As a practical example, we investigate such a nonlinear relationship between birth weight and stillbirth in pigs. We used 32,450 birth weight records from 2129 litters from 813 sows. Piglets were crossbred whereas the sows were purebred Landrace pigs from TOPIGS. Piglet survival at birth was recorded for each piglet as one (alive) or zero (dead). More details on the data are given elsewhere

(Sell-Kubiak *et al.* 2015). Data are provided in [Supporting Information, File S2](#).

### Statistical analysis

We analyzed piglet birth weight, its environmental variance ( $V_E$ ), and piglet survival at birth as three traits with a DHGLM in ASReml4 (Rönnegård *et al.* 2010; Felleki *et al.* 2012; Gilmour *et al.* 2014). For piglet survival we assumed a linear model with homogeneous residual variance because combining a logistic model for piglet survival with DHGLM for birth weight is not yet feasible and indeed outside the scope of this article. Fixed effects fitted were herd-year-season, parity of the sow, and sex of the piglet for all three traits. The random effects included were maternal genetic effects, litter effects, and residuals for all three traits, with each random effect trivariate normally distributed. The model is

$$\begin{aligned} \begin{bmatrix} \mathbf{y}_{m,1} \\ \mathbf{y}_{v,1} \\ \mathbf{y}_{m,2} \end{bmatrix} &= \begin{bmatrix} \mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X} \end{bmatrix} \begin{bmatrix} \mathbf{b}_{m,1} \\ \mathbf{b}_{v,1} \\ \mathbf{b}_{m,2} \end{bmatrix} + \begin{bmatrix} \mathbf{Z} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{m,1} \\ \mathbf{a}_{v,1} \\ \mathbf{a}_{m,2} \end{bmatrix} \\ &+ \begin{bmatrix} \mathbf{W} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{W} \end{bmatrix} \begin{bmatrix} \mathbf{c}_{m,1} \\ \mathbf{c}_{v,1} \\ \mathbf{c}_{m,2} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_{m,1} \\ \mathbf{e}_{v,1} \\ \mathbf{e}_{m,2} \end{bmatrix}, \end{aligned} \quad (31)$$

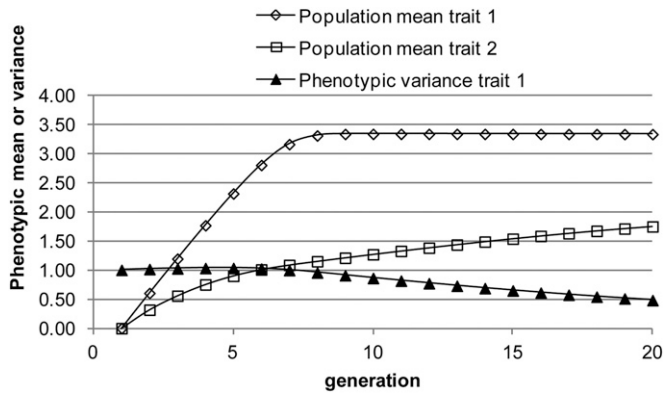
where  $\mathbf{y}_{m,1}$ ,  $\mathbf{y}_{v,1}$ , and  $\mathbf{y}_{m,2}$  are the vectors of observations for piglet birth weight, the response variable for  $V_E$  of birth weight, and piglet survival; and  $\mathbf{X}$ ,  $\mathbf{Z}$ , and  $\mathbf{W}$  are the incidence matrices relating fixed effects ( $\mathbf{b}$ ), random breeding values, and random litter effects to observations. The vectors  $\mathbf{a}_{m,1}$ ,  $\mathbf{a}_{v,1}$ , and  $\mathbf{a}_{m,2}$  are maternal breeding values for birth weight, its  $V_E$ , and survival

$$\begin{bmatrix} \mathbf{a}_{m,1} \\ \mathbf{a}_{v,1} \\ \mathbf{a}_{m,2} \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_{A_{m,1}}^2 & cov_{A_{m,1}, A_{v,1}} & cov_{A_{m,1}, A_{m,2}} \\ & \sigma_{A_{v,1}}^2 & cov_{A_{v,1}, A_{m,2}} \\ & & \text{symmetric} & \sigma_{A_{m,2}}^2 \end{bmatrix} \otimes \mathbf{I} \right);$$

$\mathbf{c}_{m,1}$ ,  $\mathbf{c}_{v,1}$ , and  $\mathbf{c}_{m,2}$  are the corresponding vectors for random litter effects

$$\begin{bmatrix} \mathbf{c}_{m,1} \\ \mathbf{c}_{v,1} \\ \mathbf{c}_{m,2} \end{bmatrix} \sim N \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_{C_{m,1}}^2 & cov_{C_{m,1}, C_{v,1}} & cov_{C_{m,1}, C_{m,2}} \\ & \sigma_{C_{v,1}}^2 & cov_{C_{v,1}, C_{m,2}} \\ & & \text{symmetric} & \sigma_{C_{m,2}}^2 \end{bmatrix} \otimes \mathbf{I} \right),$$



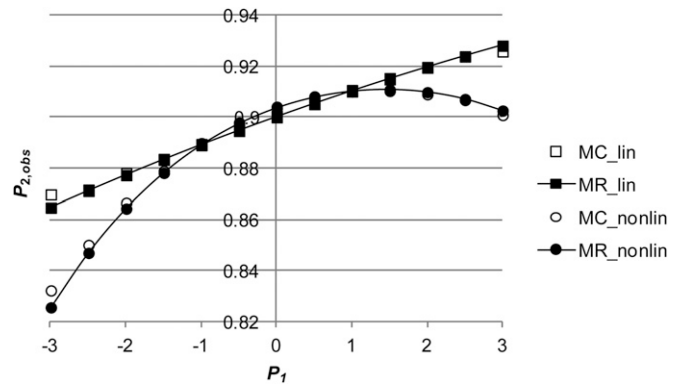


**Figure 5** Population mean for traits 1 and 2 and the phenotypic variance of trait 1 as a function of generation when applying index selection for trait 2 using the phenotype deviation of trait 1 and the squared phenotypic deviation of trait 1 in the index (starting values are as follows: selected proportion is 5%,  $\sigma_{A_{m,1}}^2 = \sigma_{A_{m,2}}^2 = 0.3$ ,  $\sigma_{A_{v,1}}^2 = 0.05$ ,  $\sigma_{e_1}^2 \exp((1/2)\sigma_{A_{v,1}}^2) = 0.7$ ,  $\sigma_{e_2}^2 = 0.7$ ,  $r_{A_{m,1},A_{m,2}} = 0.5$ ,  $r_{A_{m,1},A_{v,1}} = 0.1$ ,  $r_e = 0$ ,  $r_{A_{v,1},A_{m,2}} = -0.5$ ).

where  $\mathbf{I}$  is the identity matrix;  $\sigma_{C_{m,1}}^2$ ,  $\sigma_{C_{v,1}}^2$ , and  $\sigma_{C_{m,2}}^2$  are the litter variances; and  $cov_{C_{m,1},C_{v,1}}$ ,  $cov_{C_{m,1},C_{m,2}}$ , and  $cov_{C_{v,1},C_{m,2}}$  are the corresponding covariances of  $\mathbf{c}_{m,1}$ ,  $\mathbf{c}_{v,1}$ , and  $\mathbf{c}_{m,2}$ . Although sows had repeated observations, at the risk of biasing estimates of genetic covariance estimates, permanent sow effects were excluded from the model because their inclusion yielded estimates at the lower boundaries. Subsequently, these estimates of parameters were used to predict the regression of piglet survival on birth weight and were compared to the averages of survival rate for each of 20 ranked bins with equal numbers of piglet birth weight records. Because the statistical model included litter effects, the extension in *Appendix C* was used to predict the regression equation when the residual covariances between  $V_E$  and birth weight and survival are ignored.

## Results

The parameter estimates (Table 2 and Table 3) indicate that the genetic correlation between mean piglet birth weight and survival is negative, as is that between  $V_E$  of birth weight and survival, the M-V genetic correlation between traits. This implies that lower birth weight or lower  $V_E$  is genetically associated with a higher survival rate and, importantly, that uniform litters have higher mean survival rates. The multiple-regression equation based on the estimated variance components describes the nonlinear relationship between individual survival rate and individual birth weight quite well (Figure 7). Because a linear model for survival is used, the regression equation goes beyond a survival rate of 1.0. Accounting for the binomial nature of survival, using Equation 30, solves the problem of predicting a survival rate beyond 1.0, but the goodness of fit was slightly worse ( $R^2 = 0.78$ ) than for the regression assuming multivariate normality ( $R^2 = 0.79$ ). The relationship between individual birth weight and individual survival rate is,



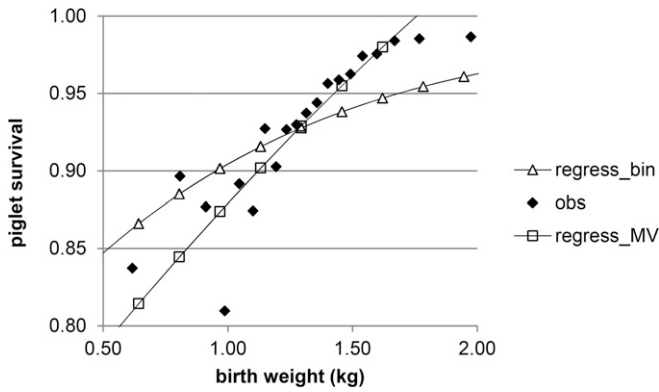
**Figure 6** The relationship between the proportion of surviving animals  $P_{2,obs}$  and  $P_1$  using Monte Carlo simulation (MC\_lin) and multiple regression (MR\_lin) either with no genetic correlation ( $r_{A_{v,1},A_{m,2}} = 0$ ) or with a negative genetic correlation ( $r_{A_{v,1},A_{m,2}} = -0.5$ ) (MC\_nonlin, MR\_nonlin) between environmental variance in  $P_1$  and the mean of  $P_{2,liab}$  on the underlying scale ( $\sigma_{A_{m,1}}^2 = \sigma_{A_{m,2}}^2 = 0.3$ ,  $\sigma_{A_{v,1}}^2 = 0.05$ ,  $\sigma_{e_1}^2 \exp((1/2)\sigma_{A_{v,1}}^2) = 0.7$ ,  $\sigma_{e_2}^2 = 0.7$ ,  $r_{A_{m,1},A_{m,2}} = 0.2$ ,  $r_{A_{m,1},A_{v,1}} = 0$ ,  $r_e = 0$ , average proportion for  $P_{2,obs} = 0.9$ ).

however, almost linear because the residual covariance between birth weight and survival substantially dilutes the nonlinear genetic relationship. Similarly Sölkner and Fuerst-Waltl (1996) found that environmental or residual correlations can greatly reduce the nonlinearity of the genetic relationship such that the phenotypic relationship is almost linear. From a biological point of view, a nonlinear relationship between piglet birth weight and survival rate at litter level can arise because extremely small piglets decrease survival rate, as also do very large piglets but to a lesser extent, because the very large piglets induce early farrowing.

## Discussion

### Methodological framework and quantitative genetic aspects of nonlinear relationships

This study shows that a genetic covariance between the environmental variance of one trait and the mean of another trait, *i.e.*, an M-V genetic correlation between them, can lead to nonlinear relationships between these traits and presented a framework based on multiple regression to investigate the consequences of selection in the presence of such an M-V genetic correlation. Although all relationships between breeding values are assumed to be linear, the phenotypic relationship between traits becomes curvilinear because the breeding value for environmental variance for the first trait  $A_{v,1}$  is associated with the squared phenotype  $P_1^2$  and its covariance to the mean of trait 2. Most quantitative genetic machinery is based on multivariate normality and thus linear relationships (Bulmer 1980), but we show that it can be extended to incorporate nonlinear relationships using methods of Mulder *et al.* (2007) for predicting breeding values and selection responses when there is genetic variation in  $V_E$ . Mulder *et al.* (2007) used the additive model for genetic heterogeneity of  $V_E$ , but here

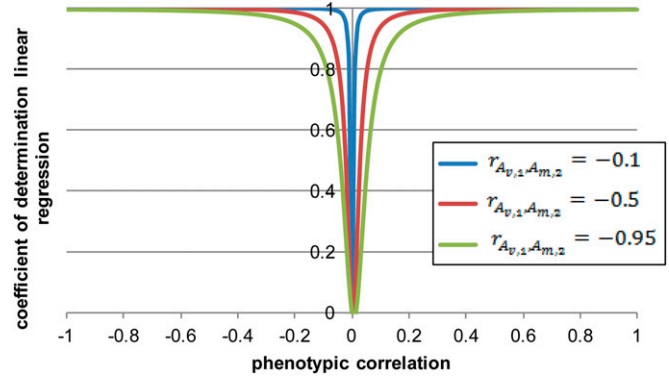


**Figure 7** The phenotypic relationship between survival and individual piglet birth weight (kg) when averages were calculated for 20 bins after sorting on birth weight (obs), regression based on Equation 3 assuming multivariate normality (regress\_MV), and regression based on Equations 3 and 30 to account for the binomial nature of survival (regress\_bin).

the exponential model was used because of its statistical tractability and implementation in the DHGLM methodology to estimate genetic variance in  $V_E$ . The additive and exponential models are similar, assuming that the genetic variance in  $V_E$  is small and parameters can be converted from one model to the other (Mulder *et al.* 2007; Hill and Mulder 2010). Here the focus is on the principles rather than the details and therefore the choice of model does not have a major impact on the non-linear relationships caused by M-V genetic correlations between traits.

We extended the DHGLM methodology to estimate genetic and environmental variances and the correlations between piglet birth weight, its  $V_E$ , and survival. The current algorithm can be run in ASREML4 (code provided in Supporting Information, File S1). It could, for example, also be used to estimate the trade-offs between  $V_E$  and other traits and thereby help further unravel the biological mechanisms and causes of genetic differences in  $V_E$ . Genetic correlations may change due to selection, as predicted in this study, even under infinitesimal model assumptions, and therefore reestimation of genetic parameters seems important if there are curvilinear relationships between traits. Large family sizes are required to get accurate estimates for genetic variance in  $V_E$  (Hill 2004; Hill and Mulder 2010), and consequently very large data sets are required for estimating M-V genetic correlations between traits with sufficient precision (inferred from Robertson 1959).

The regression in Figure 1 shows substantial nonlinearity, but in our examples the coefficient of determination ( $R^2$ ) would not be greatly reduced if a linear regression was applied even though the true relationship is quadratic because of an M-V genetic correlation between traits. The linear regression would generally explain  $\geq 90\%$  of the variation, except when the phenotypic correlation between the traits is close to zero (*i.e.*,  $-0.15 < \text{phenotypic correlation} < 0.15$ ) (Figure 8). Therefore, a linear regression between traits is sufficient when the population is far from the



**Figure 8** The coefficient of determination ( $R^2$ ) of a linear regression when the true relationship is quadratic as a function of the phenotypic correlation ( $= r_{A_{m,1}, A_{m,2}} = r_e$ ) between trait 1 and trait 2 for different values of  $r_{A_{v,1}, A_{m,2}}$  ( $\sigma_{A_{m,1}}^2 = \sigma_{A_{m,2}}^2 = 0.3$ ,  $\sigma_{A_{v,1}}^2 = 0.05$ ,  $\sigma_{e_1}^2 \exp((1/2))\sigma_{A_{v,1}}^2 = 0.7$ ,  $\sigma_{e_2}^2 = 0.7$ ,  $r_{A_{m,1}, A_{v,1}} = 0$ ).

optimum, but if there is a nonzero M-V genetic correlation between them and the population mean is close to the optimum, quadratic regression should be used.

A related question is how much of the nonlinearity observed is due to an M-V genetic correlation between the traits. In the pig example, the quadratic relationship improved  $R^2$  only from 0.787 to 0.793, whereas the best-fitting quadratic regression  $R^2$  would explain 0.794. Assuming that the true relationship is equal to the best-fitting quadratic regression, it can be concluded that the genetic covariance between  $V_E$  of birth weight and survival is the main driver for the quadratic relationship, and no other factors need to be invoked. The environmental correlations seem to have a dominant influence on diluting the nonlinearity, which accords with findings of Sölkner and Fuerst-Waltl (1996) who investigated the effect of nonlinear relationships between traits due only to nonlinearity of their breeding values for trait means. If the environmental relationship between traits is approximately linear, the phenotypic relationship will also be approximately linear, especially when the heritabilities are small.

In this study, we investigated the contribution of M-V genetic correlations between traits to nonlinearity between them. Nonnormality of at least one of the traits can cause a nonlinear relationship between them (Varona and Sorensen 2014). Survival is binomially distributed, although we assumed multivariate normality of birth weight and survival, but estimates of additive genetic correlations on the linear scale are expected to be equal to those on the underlying scale (Vinson *et al.* 1976; Gianola 1982) and therefore not biased by the current procedure. Furthermore, we showed that the prediction assuming multivariate normality fitted the observed survival rates better than when assuming that survival is binomially distributed. Nevertheless, Equation 30 is a useful extension of the framework to predict phenotypes of traits that are not normally distributed, but have an underlying liability scale that is normally distributed. More

**Table 2** Estimates of genetic, litter, and residual variance for birth weight (kg) (BW), its environmental variance  $V_E$  (BWv), and survival rate (SU) in pigs (standard error in parentheses)

Variance component	BW ( $\times 10^{-3}$ )	BWv ( $\times 10^{-3}$ )	SU ( $\times 10^{-3}$ )
Genetic variance	16.4 (1.51)	46.5 (7.55)	0.968 (0.22)
Litter variance	15.0 (0.80)	75.0 (7.53)	2.74 (0.26)
Residual variance <sup>a</sup>	74.9 (0.61)	1803 (14.7)	63.4 (0.50)

<sup>a</sup>Residual variance for BW is based on a model with homogeneous residual variance.

research is needed to explore the effects of nonnormality on estimation of such M-V correlations.

The focus in the present study was on understanding the nonlinear relationship between traits caused by M-V correlations between traits and its consequences for artificial and natural selection. We focused on mass selection, but the same framework could be used to predict responses to selection when using information of relatives, *e.g.*, sibs or offspring as shown by Mulder *et al.* (2007, 2008). With progeny testing, accuracy will be greatly improved, especially for the breeding value for  $V_E$ . In Mulder *et al.* (2008) we showed that when profit has an optimum value, artificial selection will first bring the population mean to the optimum and subsequently selection will be to reduce the variance, similar to that in Figure 5. It should be noted that current calculations as well as those in Mulder *et al.* (2007, 2008) do not take into account reduced genetic variance due to gametic phase disequilibrium (Bulmer 1971).

The framework provides estimates of positions and heights of phenotypic and genetic optima, which are functions of genetic and nongenetic (co)variances. As standard errors, particularly on genetic (co)variances, are moderate to high, those on optima are also likely to be large, especially when the M-V genetic correlation between traits is close to zero (Figure 5). The calculated location of optima gives insight into where the current population mean is compared to the optimum. The piglet example shows that phenotypic and genetic optima may be very different: the current population mean is far from the phenotypic optimum ( $\sim 13$  phenotypic SD), because higher birth weight still yields higher survival on average based on the regression, although observed survival rate flattens at  $\sim 1.8$ – $2.0$  kg. The high prediction for the phenotypic optimum is likely due to neglecting the binomial nature of survival. In contrast, the genetic optimum is  $\sim 6$  phenotypic SD below the current population mean, with the discrepancy between the phenotypic and the genetic optimum due to the opposite signs of the genetic *vs.* litter and environmental correlations between birth weight and survival. Yoshimura and Shields (1995) indicate that environmental uncertainty is the driver between differences in optima. If genetic/nongenetic (co)variances differ among environments, the optima, *i.e.*, fitness profiles, would also differ among environments (*e.g.*, Dewitt and Yoshimura 1998).

**Table 3** Estimates for traits as in Table 2 of genetic correlations (above diagonal) and litter correlations (below diagonal) in pigs (standard errors in parentheses)

Trait	BW	BWv	SU
BW		0.51 (0.08)	-0.24 (0.11)
BWv	0.20 (0.05)		-0.19 (0.14)
SU	0.35 (0.05)	0.06 (0.07)	

Note that residual correlation between BW and SU was 0.18.

### Phenotypic variability and fitness in natural populations

Nonlinear relationships are often observed or hypothesized in natural populations, typically as “fitness profiles” between trait values and fitness (Lande and Arnold 1983; Falconer and Mackay 1996; Kingsolver *et al.* 2001, 2012). If the optimum phenotype is intermediate, there would be a negative covariance between  $V_E$  and fitness, such that families with small variance would contribute more survivors, and correspondingly families with a large variance would have an advantage with disruptive or strong directional selection (Hill and Zhang 2004; Mulder *et al.* 2007). The strength of natural selection (*i.e.*, the selection gradient) can be determined by regressing relative fitness on phenotypic value if it is directional or on the squared deviation of phenotypic value from the mean if it is stabilizing or disruptive (Lande and Arnold 1983). Thus if fitness is taken as trait 2, the regression coefficients in Equation 3 indicate the strength of selection. For instance, the strength of stabilizing selection is obtained as  $V_S = -1/2b_2$  (Falconer and Mackay 1996), where  $b_2$  is the regression of fitness on the squared deviation of the trait under selection, *i.e.*, trait 1 in this study. Applying this to our study and assuming  $r_{A_{m,1}A_{v,1}} = 0$ ,  $r_e = 0$ ,  $r_{A_{m,1}A_{m,2}} = 0$ , *i.e.*, the population mean is at the optimum, the strength of stabilizing selection (expressed as the “variance” of the fitness profile standardized by  $V_E$  on trait 1, so small values imply strong selection) would be  $V_S = -1/2b_2 = [2\sigma_{P_1}^4 + 3\sigma_{e_1}^4 \exp(\sigma_{A_{v,1}}^2)(\exp(\sigma_{A_{v,1}}^2) - 1)] / [-2cov_{A_{v,1}A_{m,2}} \sigma_{e_1}^4 \exp(\sigma_{A_{v,1}}^2)]$  (Falconer and Mackay 1996). The estimate of  $V_S$  ( $V_S = -1/2b_2$ ) for the example of piglet survival and piglet birth weight is  $435V_E$ .

Based on early estimates of the strength of stabilizing selection, Falconer and Mackay (1996) suggested that strong selection would be  $V_S = 10$  and weak selection  $V_S = 100$ , with “typical” estimates of the order of 20–25 (Falconer and Mackay 1996; Zhang and Hill 2005). Kingsolver *et al.* (2001, 2012), however, reviewed selection gradients estimated in a range of natural populations and found that in only a few studies negative quadratic regression coefficients ( $b_2$ ) were actually obtained, indicating that strong stabilizing selection is rarely detected. The value obtained here for the piglet data implies very weak stabilizing selection. Indeed, as fitness typically has a low heritability ( $< 0.1$ ) and  $\sigma_{A_{v,1}}^2 < 0.1$  ( $GCV_{V_E} \cong \sigma_{A_{v,1}} < 0.32$ ) in most recent estimates using DHGLM or MCMC (Felleki *et al.* 2012; Mulder *et al.* 2013a; Rönnegård *et al.* 2013), the stabilizing selection driven by the M-V

**Table 4** The strength of stabilizing selection ( $V_E$ ) standardized by  $V_E$  for different combinations of the heritability of relative fitness ( $\sigma_{P,fitness}^2 = 0.25$ ), the genetic correlation between  $V_E$  of trait 1 and fitness ( $r_{A_{v,1},A_{m,2}}$ ), and the genetic variance in  $V_E$  ( $\sigma_{A_{v,1}}^2$ )

h <sup>2</sup> fitness	$r_{A_{v,1},A_{m,2}}$	$\sigma_{A_{v,1}}^2$	
		0.05	0.1
0.02	-0.1	1339	983
	-0.5	268	197
	-1.0	134	98
0.1	-0.1	599	440
	-0.5	120	88
	-1.0	60	44

genetic correlation between  $V_E$  and fitness is likely to be weak (examples in Table 4).

### Implications and applications in agriculture

In agriculture, a frequent objective is to increase uniformity of animals and plants for some traits because these may have an optimum with respect to profit, for instance for slaughter traits in livestock because slaughterhouses prefer low variability in carcass weight and meat quality to optimize processing and to provide homogeneous products to retailers (Hennessy 2005). Some conformation traits such as udder depth or front teat placement may have a nonlinear relationship with respect to fitness traits, such as longevity (Larroque and Ducrocq 2001), or with milk production (Fuerst-Waltl *et al.* 1998). In contrast, there is also concern that animals with small  $V_E$  would not have the capacity to react to unpredictable environments, *i.e.*, lack robustness, so it is likely to be worthwhile to explore the relationship between  $V_E$  and fitness traits in different environments. While very little is known about trade-offs between  $V_E$  and means of traits, the current study provides insight in showing that a negative genetic covariance between  $V_E$  and fitness is advantageous for piglet survival and birth weight. Furthermore, selection on  $V_E$  can increase accuracy of selection provided the population is at the optimum and the genetic correlation between trait means is close to zero. Therefore, inclusion of  $V_E$  in index selection is mainly of interest for traits that have a direct economic value for uniformity or where  $V_E$  has a strong genetic correlation with an economically important trait that has a low heritability and is not strongly correlated to other economic traits.

In conclusion, this study provides a framework to study nonlinear relationships between traits caused by M-V genetic correlations between them and to assess the effects of artificial and natural selection on their means and variances. It is shown that phenotypic and genetic optima may differ; that genetic relationships are often more curvilinear than phenotypic relationships, especially for environmentally correlated traits; and that genetic correlations may change when population means change relative to the optimal phenotypes. The analysis of piglet birth weight and survival shows that the nonlinearity can be partly explained by such

an M-V genetic correlation between  $V_E$  of birth weight and survival.

### Acknowledgments

We thank Ian White, University of Edinburgh, for helping with some of the derivations. TOPIGS Research Center IPG is acknowledged for providing the data. This study was supported by Breed4Food, a public-private partnership in the domain of animal breeding and genomics.

### Literature Cited

- Bulmer, M. G., 1971 The effect of selection on genetic variability. *Am. Nat.* 105: 201–211.
- Bulmer, M. G., 1980 *The Mathematical Theory of Quantitative Genetics*. Oxford University Press, Oxford.
- DeWitt, T. J., and J. Yoshimura, 1998 The fitness threshold model: random environmental change alters adaptive landscapes. *Evol. Ecol.* 12: 615–626.
- Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*. Pearson Education, Essex, UK.
- Felleki, M., D. Lee, Y. Lee, A. R. Gilmour, and L. Rönnegård, 2012 Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models. *Genet. Res.* 94: 307–317.
- Foulley, J. L., 1992 Prediction of selection response for threshold dichotomous traits. *Genetics* 132: 1187–1194.
- Fuerst-Waltl, B., A. Essl, and J. Sölkner, 1997 Nonlinear genetic relationships between traits and their implications on the estimation of genetic parameters. *J. Anim. Sci.* 75: 3119–3125.
- Fuerst-Waltl, B., J. Sölkner, A. Essl, I. Hoeschele, and C. Fuerst, 1998 Non-linearity in the genetic relationship between milk yield and type traits in Holstein cattle. *Livest. Prod. Sci.* 57: 41–47.
- Garreau, H., G. Bolet, C. Larzul, C. Robert-Granie, G. Saleil *et al.*, 2008 Results of four generations of a canalising selection for rabbit birth weight. *Livest. Sci.* 119: 55–62.
- Geiler-Samerotte, K. A., C. R. Bauer, S. Li, N. Ziv, D. Gresham *et al.*, 2013 The details in the distributions: why and how to study phenotypic variability. *Curr. Opin. Biotechnol.* 24: 752–759.
- Gianola, D., 1982 Theory and analysis of threshold characters. *J. Anim. Sci.* 54: 1079–1096.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson, 2014 *ASReml User Guide Release 4.0*. VSN International, Hemel Hempstead, UK.
- Hazel, L. N., 1943 The genetic basis for constructing selection indexes. *Genetics* 28: 476–490.
- Hennessy, D. A., 2005 Slaughterhouse rules: animal uniformity and regulating for food safety in meat packing. *Am. J. Agric. Econ.* 87: 600–609.
- Hill, W. G., 2004 Heterogeneity of genetic and environmental variance of quantitative traits. *J. Ind. Soc. Agric. Stat.* 57: 49–63.
- Hill, W. G., and H. A. Mulder, 2010 Genetic analysis of environmental variation. *Genet. Res.* 92: 381–395.
- Hill, W. G., and X. S. Zhang, 2004 Effects on phenotypic variability of directional selection arising through genetic differences in residual variability. *Genet. Res.* 83: 121–132.
- Hoaglin, D. C., and R. E. Welsh, 1978 The hat matrix in regression and ANOVA. *Am. Stat.* 32: 17–22.
- Jimenez-Gomez, J. M., J. A. Corwin, B. Joseph, J. N. Maloof, and D. J. Kliebenstein, 2011 Genomic analysis of QTLs and genes

- altering natural variation in stochastic noise. *PLoS Genet.* 7: 1002295.
- Kapell, D., C. J. Ashworth, P. W. Knap, and R. Roehe, 2011 Genetic parameters for piglet survival, litter size and birth weight or its variation within litter in sire and dam lines using Bayesian analysis. *Livest. Sci.* 135: 215–224.
- Karn, M. N., and L. S. Penrose, 1951 Birth weight and gestation time in relation to maternal age, parity and infant survival. *Ann. Eugen.* 16: 147–164.
- Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri *et al.*, 2001 The strength of phenotypic selection in natural populations. *Am. Nat.* 157: 245–261.
- Kingsolver, J. G., S. E. Diamond, A. M. Siepielski, and S. M. Carlson, 2012 Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evol. Ecol.* 26: 1101–1118.
- Lande, R., and S. J. Arnold, 1983 The measurement of selection on correlated characters. *Evolution* 37: 1210–1226.
- Larroque, H., and V. Ducrocq, 2001 Relationships between type and longevity in the Holstein breed. *Genet. Sel. Evol.* 33: 39–59.
- Lynch, M., and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- Mackay, T. F. C., and R. F. Lyman, 2005 *Drosophila* bristles and the nature of quantitative genetic variation. *Philos. Trans. R. Soc. B Biol. Sci.* 360: 1513–1527.
- Mulder, H. A., P. Bijma, and W. G. Hill, 2007 Prediction of breeding values and selection responses with genetic heterogeneity of environmental variance. *Genetics* 175: 1895–1910.
- Mulder, H. A., P. Bijma, and W. G. Hill, 2008 Selection for uniformity in livestock by exploiting genetic heterogeneity of residual variance. *Genet. Sel. Evol.* 40: 37–59.
- Mulder, H. A., R. E. Crump, M. P. L. Calus, and R. F. Veerkamp, 2013a Unraveling the genetic architecture of environmental variance of somatic cell score using high-density single nucleotide polymorphism and cow data from experimental farms. *J. Dairy Sci.* 96: 7306–7317.
- Mulder, H. A., L. Ronnegard, W. F. Fikse, R. F. Veerkamp, and E. Strandberg, 2013b Estimation of genetic variance for macro- and micro-environmental sensitivity using double hierarchical generalized linear models. *Genet. Sel. Evol.* 45: 23.
- Nelson, R. M., M. E. Pettersson, X. D. Li, and O. Carlborg, 2013 Variance heterogeneity in *Saccharomyces cerevisiae* expression data: trans-regulation and epistasis. *PLoS ONE* 8: e79507.
- Olofsson, H., J. Ripa, and N. Jonzen, 2009 Bet-hedging as an evolutionary game: the trade-off between egg size and number. *Proc. R. Soc. B Biol. Sci.* 276: 2963–2969.
- Robertson, A., 1959 The sampling variance of the genetic correlation coefficient. *Biometrics* 15: 469–485.
- Rönnegård, L., M. Felleki, F. Fikse, H. A. Mulder, and E. Strandberg, 2010 Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models. *Genet. Sel. Evol.* 42: 8.
- Rönnegård, L., M. Felleki, W. F. Fikse, H. A. Mulder, and E. Strandberg, 2013 Variance component and breeding value estimation for genetic heterogeneity of residual variance in Swedish Holstein dairy cattle. *J. Dairy Sci.* 96: 2627–2636.
- SanCristobal-Gaudy, M., J. M. Elsen, L. Bodin, and C. Chevalet, 1998 Prediction of the response to a selection for canalisation of a continuous trait in animal breeding. *Genet. Sel. Evol.* 30: 423–451.
- Schluter, D., 1988 Estimating the form of natural-selection on a quantitative trait. *Evolution* 42: 849–861.
- Schluter, D., and J. N. M. Smith, 1986 Natural-selection on beak and body size in the song sparrow. *Evolution* 40: 221–231.
- Sell-Kubiak, E., P. Bijma, E. F. Knol, and H. A. Mulder, 2015 Comparison of methods to study uniformity of traits: application to birth weight in in pigs. *J. Anim. Sci.* DOI: 10.2527/jas2014-8313.
- Shen, X., and L. Rönnegård, 2013 Issues with data transformation in genome-wide association studies for phenotypic variability. *F1000 Res.* 2: 200.
- Shen, X., M. Pettersson, L. Rönnegård, and O. Carlborg, 2012 Inheritance beyond plain heritability: variance-controlling genes in *Arabidopsis thaliana*. *PLoS Genet.* 8: 8.
- Sölkner, J., and B. Fuerst-Waltl, 1996 Nonlinear heritabilities and curvilinear relationships: genetic factors complicating selection for functional traits. *Interbull Bull.* 12: 226–233.
- Sorensen, D., and R. Waagepetersen, 2003 Normal linear models with genetically structured residual variance heterogeneity: a case study. *Genet. Res.* 82: 207–222.
- Stuart, A., and J. K. Ord, 1994 *Kendall's Advanced Theory of Statistics: Distribution Theory*. Arnold, London.
- Sun, X. Q., R. Elston, N. Morris, and X. F. Zhu, 2013 What is the significance of difference in phenotypic variability across SNP genotypes? *Am. J. Hum. Genet.* 93: 390–397.
- Tonsor, S. J., T. W. Elnaccash, and S. M. Scheiner, 2013 Developmental instability is genetically correlated with phenotypic plasticity, constraining heritability, and fitness. *Evolution* 67: 2923–2935.
- Varona, L., and D. Sorensen, 2014 Joint analysis of binomial and continuous traits with a recursive model: a case study using mortality and litter size of pigs. *Genetics* 196: 643–651.
- Vinson, W. E., J. M. White, and R. H. Kliever, 1976 Overall classification as a selection criterion for improving categorically scored components of type in Holsteins. *J. Dairy Sci.* 59: 2104–2114.
- Wolc, A., J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan *et al.*, 2012 Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. *Anim. Genet.* 43: 87–96.
- Yang, J., R. J. F. Loos, J. E. Powell, S. E. Medland, E. K. Speliotes *et al.*, 2012 FTO genotype is associated with phenotypic variability of body mass index. *Nature* 490: 267–273.
- Yoshimura, J., and W. M. Shields, 1995 Probabilistic optimization of body size - a discrepancy between genetic and phenotypic optima. *Evolution* 49: 375–378.
- Zhang, X.-S., and W. G. Hill, 2005 Evolution of the environmental component of the phenotypic variance: stabilizing selection in changing environments and the cost of homogeneity. *Evolution* 59: 1237–1244.

Communicating editor: J. B. Wolf

## Appendix A

### Derivation of Elements in the Multiple-Regression Equation of $P_2$ on $P_1$ and $P_1^2$

The elements of the  $\mathbf{P}$  matrix and the  $\mathbf{c}$  vector were derived as follows for the genetic model in Equation 1 ( $P_1 = \mu_1 + A_{m,1} + E_1$ ), in which  $E_1 = \chi \exp[(1/2)(\ln(\sigma_{e_1}^2) + A_{v,1})]$  and  $E(\chi^2 \exp(A_{v,1})) = \exp((1/2)\sigma_{A_{v,1}}^2)$ , where the roman E denotes expectation.

The elements of  $\mathbf{P}$  are

$$E(P_1) = 0$$

$$E(P_1^2) = E(A_{m,1}^2) + 2E(A_{m,1}E_1) + E(E_1^2) = \sigma_{A_{m,1}}^2 + \sigma_{e_1}^2 \exp\left(\frac{1}{2}\sigma_{A_{v,1}}^2\right) = \sigma_P^2$$

$$E(P_1^3) = E(A_{m,1}^3) + 3E(A_{m,1}^2E_1) + 3E(A_{m,1}E_1^2) + E(E_1^3) = 3 \text{cov}_{A_{m,1}, A_{v,1}} \sigma_{e_1}^2 \exp\left(\frac{1}{2}\sigma_{A_{v,1}}^2\right)$$

$$\begin{aligned} E(P_1^4) &= E(A_{m,1}^4) + 4E(A_{m,1}^3E_1) + 6E(A_{m,1}^2E_1^2) + 4E(A_{m,1}E_1^3) + E(E_1^4) \\ &= 3\sigma_{A_{m,1}}^4 + 6 \text{cov}_{A_{m,1}, A_{v,1}}^2 \sigma_{e_1}^2 \exp\left(\frac{1}{2}\sigma_{A_{v,1}}^2\right) + 3\sigma_{e_1}^4 \exp\left(2\sigma_{A_{v,1}}^2\right), \end{aligned}$$

where

$$E(A_{m,1}^2E_1^2) = 2 \text{cov}^2(A_{m,1}E_1) = (2) \left(\frac{1}{2}\right) \text{cov}_{A_{m,1}, A_{v,1}}^2 \sigma_{e_1}^2 \exp\left(\frac{1}{2}\sigma_{A_{v,1}}^2\right),$$

and 1/2 comes from the coefficient of  $A_{v,1}$  in  $E_1 = \chi \exp[(1/2)(\ln(\sigma_{e_1}^2) + A_{v,1})]$ ;

$$\text{var}(P_1) = E(P_1^2) - [E(P_1)]^2 = \sigma_P^2$$

$$\text{cov}(P_1, P_1^2) = E(P_1^3) - E(P_1)E(P_1^2) = 3 \text{cov}_{A_{m,1}, A_{v,1}} \sigma_{e_1}^2 \exp\left(\frac{1}{2}\sigma_{A_{v,1}}^2\right)$$

$$\text{var}(P_1^2) = E(P_1^4) - [E(P_1^2)]^2 = 2\sigma_{P_1}^4 + 6 \text{cov}_{A_{m,1}, A_{v,1}}^2 \sigma_{e_1}^2 \exp\left(\frac{1}{2}\sigma_{A_{v,1}}^2\right) + 3\sigma_{e_1}^4 \exp\left(\sigma_{A_{v,1}}^2\right) \left(\exp\left(\sigma_{A_{v,1}}^2\right) - 1\right).$$

The elements of  $\mathbf{c}$  are

$$\text{cov}(P_2, P_1) = E(A_{m,2}A_{m,1}) + E(E_2E_1) = \text{cov}_{A_{m,1}, A_{m,2}} + \text{cov}_{e_1e_2}$$

$$\begin{aligned} \text{cov}(P_2, P_1^2) &= E(A_{m,2}A_{m,1}^2) + 2E(A_{m,2}A_{m,1}E_1) + E(A_{m,2}E_1^2) + E(E_2A_{m,1}^2) + 2E(E_2A_{m,1}E_1) + E(E_2, E_1^2) \\ &= \text{cov}_{A_{v,1}, A_2} \sigma_{e_1}^2 \exp\left(\frac{1}{2}\sigma_{A_{v,1}}^2\right) + \text{cov}_{A_{m,1}, A_{v,1}} \text{cov}_{e_1e_2}. \end{aligned}$$

## Appendix B

### Derivation of Elements of $\mathbf{P}$ and $\mathbf{g}$ When Using Squared Residuals Corrected for Leverages

Instead of using  $P_1^2$  directly, squared residuals of  $P_1$  corrected for its leverage ( $l$ ) can be used to predict breeding values for the second trait  $A_{m,2}$  (see Equation 17 in main text). The leverage is the diagonal element of the hat matrix for each observation (Hoaglin and Welsh 1978) as used in the DHGLM (Rönnegård *et al.* 2010).

Assuming that only information of the animal itself is used for breeding value estimation and in the absence of fixed effects,  $b_{A_{m,1}} = l$  and hence the elements of  $\mathbf{P}_e$  in Equation 17 are

$$\text{var}\left(\frac{e_1^2}{1-l}\right) = \text{var}\left((1-b_{A_{m,1}})P_1^2\right) = (1-b_{A_{m,1}})^2 \text{var}(P_1^2)$$

$$\text{cov}\left(P_1, \frac{e_1^2}{1-l}\right) = \text{cov}(P_1, (1-b_{A_{m,1}})P_1^2) = (1-b_{A_{m,1}}) \text{cov}(P_1, P_1^2)$$

and similarly the elements of  $\mathbf{g}_{a2e}$  are

$$\text{cov}\left(A_{v1}, \frac{e_1^2}{1-l}\right) = (1-b_{A_{m,1}}) \text{cov}(A_{v1}, P_1^2)$$

$$\text{cov}\left(A_{m,2}, \frac{e_1^2}{1-l}\right) = (1-b_{A_{m,1}}) \text{cov}(A_{m,2}, P_1^2).$$

If  $(1-b_{A_{m,1}})$  is put into a scaling matrix

$$\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & (1-b_{A_{m,1}}) \end{bmatrix},$$

the derived elements in *Appendix A* can be used together with the scaling matrix (see main text below Equation 17). It can be shown that the accuracy of the estimated breeding value  $\widehat{A_{m,2}}$  ( $r_{A_{m,2}, A_{m,2}}$ ) is the same as with using squared phenotypic deviations instead of squared residuals corrected for leverages:

$$r_{A_{m,2}, A_{m,2}} = \frac{\sqrt{(\mathbf{W}^{-1}\mathbf{P}^{-1}\mathbf{W}^{-1}\mathbf{W}\mathbf{g})'\mathbf{W}\mathbf{g}}}{\sigma_{A_{m,2}}} = \frac{\sqrt{(\mathbf{P}^{-1}\mathbf{g})'\mathbf{g}}}{\sigma_{A_{m,2}}}.$$

## Appendix C

### Generalization of Equations Involved in Multiple Regression with Multiple Random Effects

Assume the statistical model for trait 1 ( $= \mathbf{y}$ ), its  $V_E$ , and trait 2 contains fixed effects and a number of random effects ( $n-1$ ) in addition to the residual,

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \sum_{i=1}^{n-1} \mathbf{Z}_i \mathbf{u}_i + \mathbf{e},$$

and  $P_2$  is predicted from  $P_1$  and  $P_1^2$ . The (co)variances in the  $\mathbf{P}$  matrix are  $V_{11} = \sum_{i=1}^{n-1} \nu_{c_{i11}}$ ,  $V_{12} = V_{21} = \sum_{i=1}^{n-1} \nu_{c_{i12}}$ ,  $V_{22} = \sum_{i=1}^{n-1} \nu_{c_{i22}}$ , where  $\nu_{c_{i11}}$  ( $\nu_{c_{i22}}$ ) is the variance for random effect  $i$  for trait 1 ( $V_E$  of trait 1) and  $\nu_{c_{i12}}$  is the covariance for random effect  $i$  between trait 1 and its  $V_E$ :

$$\text{var}(P_1) = V_{11} + \sigma_{e_1}^2 \exp\left(\frac{1}{2}V_{22}\right)$$

$$\text{cov}(P_1, P_1^2) = 3V_{12}\sigma_{e_1}^2 \exp\left(\frac{1}{2}V_{22}\right)$$

$$\text{var}(P_1^2) = 2\left(V_{11} + \sigma_{e_1}^2 \exp\left(\frac{1}{2}V_{22}\right)\right)^2 + 6V_{12}^2\sigma_{e_1}^2 \exp\left(\frac{1}{2}V_{22}\right) + 3\sigma_{e_1}^4 \exp(V_{22})(\exp(V_{22}) - 1).$$

The covariances in  $\mathbf{c}$  are  $V_{13} = \sum_{i=1}^{n-1} \nu_{c_{i13}}$ ,  $V_{23} = \sum_{i=1}^{n-1} \nu_{c_{i23}}$ , where  $\nu_{c_{i13}}$  is the covariance for random effect  $i$  for trait 1 and trait 2 and  $\nu_{c_{i23}}$  is the covariance for random effect  $i$  between trait 1 and its  $V_E$ :

$$\text{cov}(P_2, P_1) = V_{13} + \text{cov}_{e_1 e_2}$$

$$\text{cov}(P_2, P_1^2) = V_{23}\sigma_{e_1}^2 \exp\left(\frac{1}{2}V_{22}\right) + V_{12}\text{cov}_{e_1 e_2}.$$

# GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.173070/-/DC1>

## **Heritable Environmental Variance Causes Nonlinear Relationships Between Traits: Application to Birth Weight and Stillbirth of Pigs**

Herman A. Mulder, William G. Hill, and Egbert F. Knol



**File S1**

**ASReml file for analysis piglet birth weight and survival**

!WORKSPACE 1800 !NOGRAPHICS !DEBUG !LOGFILE !RENAME !ARGS 1 2 // !DOPART \$1

DHGLM model of birth weight and survival

animal 32450 !!

litter 2129 !!

parity 10 !!

sex 2 !!

farm 15 !!

ys 22 !!

sow 7415 !!

bw !M -99

surv !M -99

Gval !=bw !-1.19 !\*V10

Ywt !=1. Gwt !=1. survW !=1.

Ainv.giv

phenotype3.txt !maxit 1000 !skip 1 !DOPART \$1

!PART 1 # normal model

bw ~ mu parity sex farm.ys !r giv(sow,1) litter

residual units

!Part 2

!ASUV !EXTRA 100 !SLOW

# in odd iterations, we use the predicted weights for the primary response

!IF ODD !CALC W1=EXP(R2-Y2) #redefine weights for Y1

!IF EVEN !CALC S1=1./W1; H0=MIN(H1/S1, .9999); Z2=MAX(R1\*R1,.0001)/(1-H0)

!IF EVEN !CALC Y2=LOG(S1)+(Z2-S1)/S1 #redefine Y2

!IF EVEN !CALC W2=(1-H0)/2 #redefine weights for Y2

!ASSIGN gen 0.016 0.005 0.05 0.00 0.00 5.0

!ASSIGN lit 0.015 0.0 0.08 0.0 0.0 5.0

bw Gval surv !Weight Ywt !WT Gwt !WT survW ~ Trait Trait.parity Trait.sex Trait.farm.yr !r us(Trait,\$gen).giv(sow,1)  
us(Trait,\$lit).litter !f mv

**Files S2-S3**

Available for download at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.173070/-/DC1>

**File S2** Phenotypic data and the A-inverse

**File S3** Fortran programs