

Modelos de ecuaciones estructurales como alternativa a los modelos
multicarácter en mejoramiento genético: su impacto sobre la respuesta a la
selección

*Tesis presentada para optar al título de Magister de la Universidad de Buenos Aires,
Área Biometría y Mejoramiento*

Martin Bonamy

Médico Veterinario - Universidad Nacional de La Plata - 2012

Lugar de trabajo: Cátedra de Producción Bovina. Facultad de Ciencias Veterinarias.
Universidad Nacional de La Plata.



Escuela para Graduados Ing. Agr. Alberto Soriano
Facultad de Agronomía – Universidad de Buenos Aires



COMITÉ CONSEJERO

Director de Tesis

Sebastián Munilla Leguizamón

Ingeniero Agrónomo (Universidad de Buenos Aires)

Doctor en Ciencias Agropecuarias (Universidad de Buenos Aires)

Co-director de Tesis

Guillermo Giovambattista

Licenciado en Biología (Universidad Nacional de La Plata)

Doctor en Ciencias Naturales (Universidad Nacional de La Plata)

JURADO DE TESIS

JURADO

Olga Ravagnolo Gumila

Ingeniera Agrónoma (Universidad de la República)

Master of Science (University of Georgia)

Doctor of Philosophy (University of Georgia)

JURADO

Pedro Alejandro Vozzi

Licenciado en Genética (Universidad Nacional de Misiones)

Master en Ciencias Biológicas (Universidad de São Paulo)

Doctor en Ciencias Biológicas (Universidad de São Paulo)

JURADO

Matias Bermann

Ingeniero Agrónomo (Universidad de Buenos Aires)

Doctor of Philosophy (University of Georgia)

DEDICATORIA

A Natalí.

AGRADECIMIENTOS

Son muchas las personas que han hecho posible mi conclusión de esta maestría, y a quienes debo un agradecimiento.

En primer lugar, al Dr. Sebastián Munilla quien no solo acepto ser mi director sino que me estimuló a trabajar sobre esta temática y realizó importantes aportes con su mirada crítica sobre los problemas, y en ese proceso además se volvió un amigo.

Al Dr. Guillermo Giovambattista quien además de apoyar en este proceso de formación permanentemente está en la búsqueda de extender el campo de trabajo de nuestro grupo, vincularse con nuevos grupos de trabajo y afrontar nuevos desafíos.

A los jurados de esta tesis la Dra. Olga Ravagnolo y los Dres. Alejandro Vozzi y Matías Bermann, por tomarse el tiempo de leer mi trabajo y realizar sus valiosas sugerencias para mejorarlo y luego intercambiar sus opiniones durante la defensa.

A todos los docentes del programa de maestría, quienes me abrieron la cabeza a un nuevo espacio poco explorado por mí.

A mis compañeros de cursada, en especial a Gastón y Esteban, sin duda los desafíos se enfrentan mejor en grupo.

A mis compañeros de trabajo en la FCV UNLP Andrés, Alberto, Roberto, Fede, Enrique, Nico, Mari, Juli, Enzo, Joaco, Andrés y María, quienes apoyaron esta iniciativa y cubrieron ausencias durante estos los años de cursadas e hicieron aportes desde su visión particular a este trabajo.

Al grupo de Mejoramiento Genético de la FAUBA, por hacerme un lugar en su mesa y los intercambios durante mis estancias en la facultad.

Por último, son los afectos quienes sostienen todos mis proyectos en especial Natalí, el mayor apoyo mis proyectos y mi contención ante las situaciones complicadas. Mi familia y mis amigos quienes son mi cable a tierra.

DECLARACIÓN

Declaro que el material incluido en esta tesis es, a mi mejor saber y entender, original producto de mi propio trabajo (salvo en la medida en que se identifique explícitamente las contribuciones de otros), y que este material no lo he presentado, en forma parcial o total, como una tesis en esta u otra institución.

Martin Bonamy

PUBLICACIONES DERIVADAS DE LA TESIS

Bonamy, M., Fernández, M. E., Giovambattista, G., & Munilla, S. 2023. Conditioning on the causal network prevents indirect response to selection. *Journal of Animal Breeding and Genetics*. <https://doi.org/10.1111/jbg.12824>

ÍNDICE GENERAL

DEDICATORIA.....	iii
AGRADECIMIENTOS	iv
DECLARACIÓN	v
PUBLICACIONES DERIVADAS DE LA TESIS	vi
ÍNDICE GENERAL	vii
ÍNDICE DE TABLAS.....	ix
ÍNDICE DE FIGURAS.....	x
ABREVIATURAS.....	xiv
RESUMEN.....	xv
ABSTRACT	xvii
Capítulo 1. <i>Introducción</i>	1
Capítulo 2. <i>Modelos de ecuaciones estructurales</i>	6
Capítulo 3. <i>Impacto del uso de modelos de ecuaciones estructurales en la respuesta a la selección</i>	12
3.1 MATERIALES Y MÉTODOS	13
3.2 RESULTADOS	18
3.2.2 Respuesta a la selección cuando efectos causales y correlaciones genéticas actúan en forma simultánea	23
3.3 DISCUSIÓN	27
3.3.1 Respuesta a la selección cuando los caracteres están causal pero no genéticamente relacionados entre sí.....	28
3.3.2 Respuesta a la selección cuando los efectos causales y las correlaciones genéticas actúan de forma simultánea	32
3.3.3 Implicancias de estos resultados en los programas de mejoramiento genético	34
Capítulo 4. <i>Estimación de los parámetros del modelo</i>	37
4.1 MATERIALES Y MÉTODOS	38
4.2 RESULTADOS	40
4.3 DISCUSIÓN	45
Capítulo 5. <i>Conclusiones</i>	47

BIBLIOGRAFÍA.....	49
ANEXOS	52

ÍNDICE DE TABLAS

Tabla 3.1. Varianzas y covarianzas del modelo multicarácter (MTM) equivalente derivadas de los parámetros del modelo de ecuaciones estructurales (SEM) simulado.

Tabla 3.2. Respuesta acumulada luego de seis generaciones de selección al utilizar como criterio de selección estimaciones de los valores de cría obtenidas mediante un modelo animal multicarácter (MTM) o mediante un modelo de ecuaciones estructurales (SEM) en el escenario sin correlaciones genéticas entre caracteres.

Tabla 3.3. Respuesta acumulada luego de seis generaciones de selección al utilizar como criterio de selección estimaciones de los valores de cría obtenidas mediante un modelo animal multicarácter (MTM) o mediante un modelo de ecuaciones estructurales (SEM) en el escenario con correlaciones genéticas entre caracteres.

Tabla 4.1. Errores de estimación de los coeficientes causales en el escenario sin correlaciones genéticas entre los caracteres.

Tabla 4.2. Errores de estimación de las varianzas y covarianzas genéticas del modelo de ecuaciones estructurales en el escenario sin correlaciones genéticas entre los caracteres.

Tabla 4.3. Errores de estimación de los coeficientes causales en el escenario con correlaciones genéticas entre los caracteres. Los valores de la tabla corresponden al error de estimación medio, su desvío estándar (DE) y su intervalo de alta densidad del 95%, todos ellos obtenidos sobre las 100 réplicas de la simulación.

Tabla 4.4. Errores de estimación de las varianzas y covarianzas genéticas del modelo de ecuaciones estructurales en el escenario con correlaciones genéticas entre los caracteres. Los valores de la tabla corresponden al error de estimación medio, su desvío estándar (DE) y su intervalo de alta densidad del 95%, todos ellos obtenidos sobre las 100 réplicas de la simulación.

ÍNDICE DE FIGURAS

Figura 2.1. Grafo acíclico directo (DAG) para una red causal compuesta por tres caracteres.

Figura 3.1. Estructura causal simulada.

Figura 3.2. Representación esquemática del diseño del experimento de simulación.

Figura 3.3. Respuesta a la selección a lo largo de las generaciones al utilizar como criterio de selección estimaciones de los valores de cría para el carácter 1 obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) en el escenario sin correlaciones genéticas entre caracteres.

Figura 3.4. Respuesta a la selección a lo largo de las generaciones al utilizar como criterio de selección estimaciones de los valores de cría para el carácter 4 obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) en el escenario sin correlaciones genéticas entre caracteres.

Figura 3.5. Respuesta a la selección a lo largo de las generaciones al utilizar como criterio de selección estimaciones de los valores de cría para el carácter 5 obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) en el escenario sin correlaciones genéticas entre caracteres.

Figura 3.6. Respuesta a la selección a lo largo de las generaciones al utilizar como criterio de selección estimaciones de los valores de cría para el carácter 1 obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) en el escenario con correlaciones genéticas entre caracteres distintas de cero.

Figura 3.7. Respuesta a la selección a lo largo de las generaciones al utilizar como criterio de selección estimaciones de los valores de cría para el carácter 4 obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) en el escenario con correlaciones genéticas entre caracteres distintas de cero.

Figura 3.8. Respuesta a la selección a lo largo de las generaciones al utilizar como criterio de selección estimaciones de los valores de cría para el carácter 5 obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) en el escenario con correlaciones genéticas entre caracteres distintas de cero.

Figura 3.9. **A.** Modelo causal simulado en ausencia de correlaciones genéticas y estimado utilizando modelos de ecuaciones estructurales (SEM). **B.** Representación del empleo de modelo animal multicarácter (MTM) como modelo de estimación de la red descripta por la Figura A.

Figura A1.1. Evolución de los valores de cría reales (simulados) al utilizar como criterio de selección estimaciones de los valores de cría del carácter 1, obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) con caracteres genéticamente independientes. Valores promedio de 100 réplicas de simulación.

Figura A1.2. Evolución de los valores de cría estimados al utilizar como criterio de selección estimaciones de los valores de cría del carácter 1, obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) con caracteres genéticamente independientes. Valores promedio de 100 réplicas de simulación.

Figura A1.3. Evolución de los valores de cría reales (simulados) al utilizar como criterio de selección estimaciones de los valores de cría del carácter 4, obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) con caracteres genéticamente independientes. Valores promedio de 100 réplicas de simulación.

Figura A1.4. Evolución de los valores de cría estimados al utilizar como criterio de selección estimaciones de los valores de cría del carácter 4, obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) con caracteres genéticamente independientes. Valores promedio de 100 réplicas de simulación.

Figura A1.5. Evolución de los valores de cría reales (simulados) al utilizar como criterio de selección estimaciones de los valores de cría del carácter 5, obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) con caracteres genéticamente independientes. Valores promedio de 100 réplicas de simulación.

Figura A1.6. Evolución de los valores de cría estimados al utilizar como criterio de selección estimaciones de los valores de cría del carácter 5, obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) con caracteres genéticamente independientes. Valores promedio de 100 réplicas de simulación.

Figura A2.1. Evolución de los valores de cría reales (simulados) al utilizar como criterio de selección estimaciones de los valores de cría del carácter 1, obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) con caracteres genéticamente correlacionados. Valores promedio de 100 réplicas de simulación.

Figura A2.2. Evolución de los valores de cría estimados al utilizar como criterio de selección estimaciones de los valores de cría del carácter 1, obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) con caracteres genéticamente correlacionados. Valores promedio de 100 réplicas de simulación.

Figura A2.3. Evolución de los valores de cría reales (simulados) al utilizar como criterio de selección estimaciones de los valores de cría del carácter 4, obtenidas mediante un modelo

animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente correlacionados. Valores promedio de 100 réplicas de simulación.

Figura A2.4. Evolución de los valores de cría estimados al utilizar como criterio de selección estimaciones de los valores de cría del carácter 4, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente correlacionados. Valores promedio de 100 réplicas de simulación.

Figura A2.5. Evolución de los valores de cría reales (simulados) al utilizar como criterio de selección estimaciones de los valores de cría del carácter 5, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente correlacionados. Valores promedio de 100 réplicas de simulación.

Figura A2.6. Evolución de los valores de cría estimados al utilizar como criterio de selección estimaciones de los valores de cría del carácter 5, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente correlacionados. Valores promedio de 100 réplicas de simulación.

Figura A3.1. Distribución empírica de los parámetros de causalidad obtenidos mediante análisis de confirmación de factores (CFA) en ausencia de correlaciones genéticas

Figura A3.2. Distribución empírica de las varianzas y covarianzas genéticas bajo un modelo multicarácter equivalente estimadas mediante un algoritmo AI-REML en ausencia de correlaciones genéticas.

Figura A3.3. Distribución empírica de las varianzas y covarianzas del error bajo un modelo multicarácter equivalente estimadas mediante un algoritmo AI-REML en ausencia de correlaciones genéticas.

Figura A3.4. Distribución empírica de las varianzas y covarianzas genéticas bajo un modelo de ecuaciones estructurales obtenidas operacionalmente en ausencia de correlaciones genéticas.

Figura A3.5. Distribución empírica de las varianzas y covarianzas del error bajo un modelo de ecuaciones estructurales obtenidas operacionalmente en ausencia de correlaciones genéticas.

Figura A4.1. Distribución empírica de los parámetros de causalidad obtenidos estimaciones del valor de covariables utilizando muestreo de Gibbs sobre un modelo animal multicarácter en ausencia de correlaciones genéticas.

Figura A4.2. Distribución empírica de las varianzas y covarianzas genéticas obtenidas utilizando muestreo de Gibbs sobre un modelo animal multicarácter declarando las relaciones causales como covariables en ausencia de correlaciones genéticas.

Figura A4.3. Distribución empírica de las varianzas y covarianzas del error obtenidas utilizando muestreo de Gibbs sobre un modelo animal multicarácter declarando las relaciones causales como covariables en ausencia de correlaciones genéticas.

Figura A5.1. Distribución empírica de los parámetros de causalidad obtenidos mediante análisis de confirmación de factores (CFA) caracteres correlacionados genéticamente.

Figura A5.2. Distribución empírica de las varianzas y covarianzas genéticas bajo un modelo multicarácter equivalente estimadas mediante un algoritmo AI-REML caracteres correlacionados genéticamente.

Figura A5.3. Distribución empírica de las varianzas y covarianzas del error bajo un modelo multicarácter equivalente estimadas mediante un algoritmo AI-REML caracteres correlacionados genéticamente.

Figura A5.4. Distribución empírica de las varianzas y covarianzas genéticas obtenidas operacionalmente en caracteres correlacionados genéticamente.

Figura A5.5. Distribución empírica de las varianzas y covarianzas del error obtenidas operacionalmente en caracteres correlacionados genéticamente.

Figura A6.1. Distribución empírica de los parámetros de causalidad obtenidos estimaciones del valor de covariables utilizando muestreo de Gibbs sobre un modelo animal multicarácter con caracteres correlacionados genéticamente.

Figura A6.2. Distribución empírica de las varianzas y covarianzas genéticas obtenidas utilizando muestreo de Gibbs sobre un modelo animal multicarácter declarando las relaciones causales como covariables con caracteres correlacionados genéticamente.

Figura A6.3. Distribución empírica de las varianzas y covarianzas del error obtenidas utilizando muestreo de Gibbs sobre un modelo animal multicarácter declarando las relaciones causales como covariables con caracteres correlacionados genéticamente.

ABREVIATURAS

AI-REML	Información promedio – máxima verosimilitud restringida (del inglés <i>average information – restricted maximum likelihood</i>)
BLUP	Predicción lineal insesgada de mínima varianza (del inglés <i>best lineal unbiased prediction</i>)
CFA	Análisis de confirmación de factores (del inglés <i>confirmatory factor analysis</i>)
DAG	Grafo acíclico directo (del inglés <i>directed acyclic graph</i>)
DE	Desvío estándar
EE	Error estándar
HDI95%	Intervalo de alta densidad 95% (del inglés <i>95% highest density Interval</i>)
MTM	Modelo multicarácter (del inglés <i>multiple trait model</i>)
SEM	Modelo de ecuaciones estructurales (del inglés <i>structural equation model</i>)

RESUMEN

Los modelos animales multicarácter permiten la estimación de los valores de cría para varios caracteres en forma simultánea. Estos modelos asumen que las relaciones entre los caracteres son recíprocas y las especifican a través de parámetros asociados a correlaciones genéticas y ambientales entre ellos. Sin embargo, las relaciones entre los caracteres pueden no ser recíprocas, sino que un carácter actúa sobre otro en forma unidireccional. En estos casos, y dada una red causal subyacente, los modelos de ecuaciones estructurales se consideran una metodología más apropiada. Aunque se ha demostrado que ambos modelos son paramétricamente equivalentes, los valores genéticos obtenidos bajo uno u otro deben interpretarse de manera diferente. En esta tesis, se evaluó a través de un experimento de simulación estocástica el impacto de utilizar estas estimaciones en la respuesta a la selección dentro de una red causal que comprende cinco caracteres diferentes. Se probaron tres criterios de selección diferentes, definidos por la posición que ocupaba el carácter bajo selección en la red causal. En primer lugar, se consideró el caso en el que los caracteres estaban relacionados causalmente, pero eran genéticamente independientes. Bajo esta situación los modelos multicarácter al absorber las relaciones causales como correlaciones genéticas, modifican la respuesta realizada a la selección en comparación con los modelos de ecuaciones estructurales. Los resultados obtenidos respaldaron esta hipótesis. Cuando la selección fue ejercida sobre caracteres ubicados aguas arriba de la red causal se obtuvo la misma respuesta a la selección independientemente del modelo utilizado. En cambio, al ejercer presión de selección en caracteres aguas abajo, tanto la respuesta directa sobre ese carácter como la indirecta sobre otros caracteres de la red se modificó al utilizar uno u otro modelo de estimación de los valores de cría. También se consideró un escenario en el que los efectos causales y las correlaciones genéticas actúan simultáneamente, hallando los mismos

resultados: la selección basada en valores de cría estimados bajo un modelo de ecuaciones estructurales disminuyó la respuesta indirecta en caracteres que se encuentran aguas arriba en la red causal. En este último escenario resulta desafiante identificar por separado los coeficientes causales y los parámetros de varianza y estimarlos con precisión. Como un objetivo colateral de esta tesis, entonces, y capitalizando los datos generados por el experimento de simulación, se evaluaron dos enfoques diferentes para abordar este desafío. Entre ambos, la estrategia basada en ajustar un modelo multicarácter declarando los fenotipos de los ancestros causales como covariables fue la que produjo las estimaciones más precisas de estos parámetros.

Palabras clave: modelos de ecuaciones estructurales; modelo animal multicarácter; estimación de valores de cría; relaciones causales.

ABSTRACT

Multiple trait animal models enable the simultaneous estimation of breeding values for multiple traits. These models assume that the relationships between traits are reciprocal and specify them through parameters associated with genetic and environmental correlations. However, trait relationships may not be reciprocal; instead, one trait may act unidirectionally on another. In such cases, considering an underlying causal network, structural equation models are deemed a more appropriate methodology. Although both models have been demonstrated to be parametrically equivalent, the genetic values obtained under each model must be interpreted differently. In this thesis, the impact of utilizing these estimates on selection response within a causal network comprising five different traits was evaluated through a stochastic simulation experiment. Three different selection criteria were tested, defined by the position of the selected trait in the causal network. Firstly, the case where traits were causally related but genetically independent was considered. In this situation, multiple trait models, by absorbing causal relationships as genetic correlations, altered the selection response compared to structural equation models. The results supported this hypothesis. When selection was applied to traits located upstream in the causal network, the same selection response was obtained regardless of the model used. Conversely, when selection pressure was applied to downstream traits, both the direct response on that trait and the indirect response on other network traits were modified depending on the estimation model used. A scenario was also considered where causal effects and genetic correlations acted simultaneously, yielding similar results: selection based on breeding values estimated under a structural equation model decreased the indirect response in traits located upstream in the causal network. In this latter scenario, it is challenging to separately identify causal coefficients and variance parameters and estimate them accurately. As a collateral objective

of this thesis, leveraging the data generated by the simulation experiment, two different approaches to address this challenge were evaluated. Between the two, the strategy based on fitting a multiple trait model by declaring phenotypes of causal ancestors as covariates produced the most precise estimates of these parameters.

Keywords: structural equation models; multiple trait models; breeding value estimation; causal relationships.

.

Capítulo 1. *Introducción*

Capítulo 1

Introducción

El principal objetivo del mejoramiento genético es inducir cambios en las medias de los fenotipos de las poblaciones para determinados caracteres de interés económico (Bourdon, 2014, Cap. 2). Desde la perspectiva de la producción animal, el interés en estos cambios radica en que son de naturaleza permanente, acumulativa e independiente de acciones de manejo específicas. Para perseguir este objetivo la disciplina se vale fundamentalmente de dos herramientas: la selección y la asignación de servicios. La selección de los animales superiores se sustenta en objetivos definidos previamente y en criterios de superioridad de un reproductor frente a otro. En este sentido, el valor de cría es la medida por excelencia del valor de un individuo como reproductor.

El valor de cría se define como la desviación media de la performance de los hijos de un determinado individuo respecto de la media poblacional para un carácter de interés. Si bien este valor no es directamente observable, puede estimarse a partir de registros fenotípicos, información genealógica y, más recientemente, información genómica. La estimación se basa en el empleo de modelos estadísticos de la familia de los modelos mixtos: los denominados ‘modelos animales’ (cf. Henderson, 1984). En el contexto de un modelo animal, el valor de cría se estima a partir de su predictor lineal insesgado de mínima varianza (BLUP, por su acrónimo en inglés).

En la práctica, el objetivo de selección conjuga la valoración de diferentes caracteres que, en algunos casos, pueden estar genéticamente correlacionados entre sí. Estas correlaciones genéticas son atribuidas fundamentalmente al efecto pleiotrópico de genes que

controlan la expresión de los varios caracteres. En estos casos, es posible emplear modelos estadísticos que estiman los valores de cría de los diferentes caracteres en simultáneo capturando la estructura de correlación genética entre ellos, lo que permite una estimación más precisa de los valores de cría. Estos modelos reciben el nombre de modelos multicarácter (MTM, Henderson y Quaas, 1976).

Ahora bien, la correlación entre dos caracteres medidos sobre un mismo individuo puede no ser exclusivamente genética, es decir, producto de la acción de genes con efectos pleiotrópicos. Por ejemplo, si consideramos el largo de gestación y la dificultad de partos veremos que gestaciones largas producen fetos a término más pesados y estos son naturalmente más difíciles de parir por una hembra. Sin embargo, no es biológicamente posible que la mayor o menor dificultad al parto afecte el largo de gestación (López de Maturana y col., 2009, 2010). La relación que existe entre estos dos caracteres es de tipo causal y no es de naturaleza recíproca.

Los modelos de ecuaciones estructurales (SEM) son una variante de los modelos multivariados (Wright, 1921; Haavelmo, 1943) que permiten incluir relaciones causales entre los caracteres evaluados. La inferencia causal se aplica en diversos campos, como la filosofía (Spirtes y col., 1993), la computación (Pearl 2009) y la biología (Shipley, 2016). En el campo de la mejora genética su empleo ha surgido luego del trabajo seminal de Gianola y Sorensen (2004), quienes derivaron las equivalencias entre modelos MTM y SEM, y luego fue aplicado por varios autores con el fin de identificar redes causales entre caracteres en diferentes especies de interés ganadero, como vacas lecheras (Wu y col., 2007, König y col., 2008; Heringstad y col., 2009), cabras lecheras (de los Campos y col., 2006), cerdos (Varona y col., 2007) y vacas de carne (Inoue y col., 2016). Más recientemente se han aplicado también

modelos de ecuaciones estructurales en el campo de la genética molecular (Peñagaricano y col., 2015, Pegolo y col., 2021). Parte de estos resultados y sus implicancias fueron recientemente revisadas por Varona y González-Recio (2023).

Sin embargo, poco se ha estudiado sobre el impacto que podrían tener el empleo de las estimaciones de los valores de cría obtenidas bajo SEM como criterio de selección en poblaciones en producción. Valente y col. (2013, pág. 563) señalan que, en el contexto de un SEM, el valor de cría de un carácter es el resultado del efecto directo de los genes que lo determinan, libre de los efectos de los genes que gobiernan los caracteres que tienen una influencia causal sobre él. Aun así, consideran que el uso de los SEM no produciría una ventaja adicional sobre los MTM ya que, si bien producen más información o lo hacen de una forma más detallada, el criterio apropiado para la toma de decisiones de selección es el efecto genético total, el cual es correctamente calculado por los MTM. Esta afirmación no ha sido puesta a prueba.

El objetivo de esta tesis fue evaluar la respuesta a la selección que puede esperarse al utilizar como criterio de selección valores de cría estimados mediante un modelo de ecuaciones estructurales y un modelo multicarácter, cuando los caracteres involucrados están relacionados causalmente. La hipótesis de este trabajo es que los modelos multicarácter, al absorber parte de las relaciones causales en correlaciones genéticas, inducen una respuesta indirecta a la selección en caracteres no seleccionados. Esta hipótesis fue puesta a prueba mediante un experimento de simulación estocástica.

Por otro lado, la implementación de estos modelos requiere estimar sus parámetros como paso previo a la obtención de los valores de cría. En el caso de los MTM, es necesario

estimar las varianzas y covarianzas genéticas y ambientales entre los caracteres (Mrode, 2014). Por su parte, para implementar los SEM se necesita estimar $\frac{q^2}{2} - \frac{q}{2}$ parámetros adicionales, donde q el número de caracteres evaluados, correspondientes a los coeficientes de causalidad. La estimación de los parámetros asociados a los modelos de ecuaciones estructurales es un desafío en sí mismo. Como objetivo colateral, entonces, se propuso evaluar dos metodologías para llevar a cabo esta tarea, con el fin de determinar la factibilidad de la aplicación de los SEM en el mejoramiento genético.

Capítulo 2. *Modelos de ecuaciones estructurales*

Capítulo 2

Modelos de ecuaciones estructurales

Los modelos de ecuaciones estructurales (SEM) son modelos estadísticos pertenecientes a la familia de los modelos multivariados, derivados del método de análisis de senderos (*path analysis*) descrito originalmente por Wright (1921). En el contexto de la mejora genética animal, permiten estimar y evaluar relaciones causales entre caracteres que no pueden ser evaluadas mediante los modelos lineales estándares empleado en la disciplina (Rosa y col., 2011).

El uso de los SEM requiere de una definición a priori de una red causal, donde cada variable, que constituye un nodo, se relaciona con otras a través de relaciones causales, representadas por flechas. Esta estructura causal puede formalizarse matemáticamente de forma gráfica mediante un grafo acíclico directo (DAG, por su acrónimo en inglés; Bello y col., 2018). A modo de ejemplo, la Figura 2.1 presenta una red causal de tres caracteres donde el valor fenotípico que toma cada uno de ellos ($y_i, i = 1, 2, 3$) depende de un valor genético individual o valor de cría (a_i) y un residuo o error (e_i). Adicionalmente, y_2 e y_3 dependen del valor que hayan tomado otros caracteres en la red (y_1 o y_1 e y_2 , respectivamente). Esta red causal formaliza por ejemplo las relaciones causales entre el largo de gestación (y_1), la dificultad de parto (y_2) y la probabilidad de natimorto (y_3) de acuerdo con López de Maturana y col. (2010).

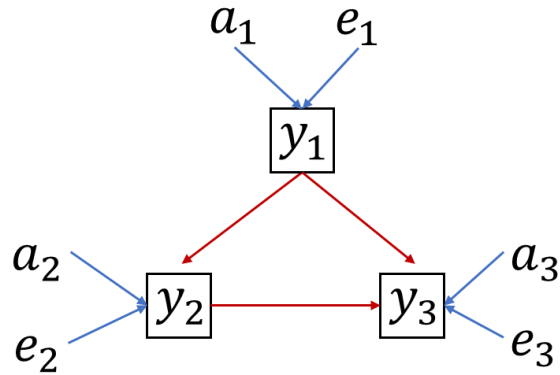


Figura 2.1. Grafo acíclico directo (DAG) para una red causal compuesta por tres caracteres.

El conjunto de ecuaciones que representan la red causal en la Figura 2.1 será

$$\begin{cases} y_1 = a_1 + e_1 \\ y_2 = \lambda_{21}y_1 + a_2 + e_2 \\ y_3 = \lambda_{31}y_1 + \lambda_{32}y_2 + a_3 + e_3 \end{cases} \quad [1]$$

donde puede notarse la recursión en los caracteres que actúan como *ancestros causales* (ver Pearl, 2020) ubicados aguas arriba, como causa de aquellos caracteres ubicados aguas abajo. Los coeficientes λ_{ij} , representados por las flechas rojas en la Figura 2.1, se interpretan como el cambio esperado en el valor fenotípico del carácter j por un cambio unitario en el valor fenotípico del carácter i . En el trabajo de López de Maturana y col. (2010), por ejemplo, estos coeficientes fueron estimados en 0,04 y 0,025, respectivamente, para el efecto del largo de gestación sobre la dificultad de parto y la probabilidad de natimorto. Por su parte, el efecto causal de la dificultad de parto sobre la probabilidad de natimorto fue estimado en 0,318. En este caso, las estimaciones están definidas para gestaciones de entre 280 y 291 días y los valores son diferentes para otros largos de gestación (López de Maturana y col., 2010).

En general, sean $y_1, \dots, y_q$ un conjunto de caracteres relacionados causalmente, entonces el modelo estadístico SEM se define en términos de las siguientes ecuaciones (en notación matricial)

$$\mathbf{y}_i = \mathbf{\Lambda} \mathbf{y}_i + \mathbf{a}_i + \mathbf{e}_i \quad [2]$$

donde $\mathbf{y}_i^t = [y_1 \ \dots \ y_q]$ representa el vector de valores fenotípicos de los caracteres 1 a q medidos en el i -ésimo animal. Por su parte, $\mathbf{\Lambda}$ representa la matriz de orden q de coeficientes causales, cuya estructura es triangular inferior con ceros en la diagonal principal:

$$\mathbf{\Lambda} = \begin{bmatrix} 0 & & \dots & 0 \\ \lambda_{21} & 0 & & \vdots \\ \vdots & & 0 & \\ \lambda_{q1} & \dots & \lambda_{q(q-1)} & 0 \end{bmatrix} \quad [3]$$

Finalmente, $\mathbf{a}_i$ y $\mathbf{e}_i$ representan los vectores de valores de cría y errores, respectivamente, con distribución conjunta:

$$\begin{bmatrix} \mathbf{a}_i \\ \mathbf{e}_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_0 & 0 \\ 0 & \mathbf{R}_0 \end{bmatrix} \right) \quad [4]$$

donde $\mathbf{\Sigma}_0$ y $\mathbf{R}_0$ son las correspondientes matrices de varianza y covarianzas. Para n animales, estas matrices se extienden a $\mathbf{\Sigma} = \mathbf{\Sigma}_0 \otimes \mathbf{A}$ y $\mathbf{R} = \mathbf{R}_0 \otimes \mathbf{I}$, donde $\mathbf{A}$ representa la matriz de relaciones aditivas (o cualquier otra matriz de parentescos), $\mathbf{I}$ representa una matriz identidad y $\otimes$ representa un producto Kronecker.

Como señalan Gianola y Sorensen (2004), operando con las ecuaciones del modelo SEM para el animal i se puede obtener un sistema de ecuaciones equivalente:

$$\begin{aligned}
\mathbf{y}_i &= \Lambda \mathbf{y}_i + \mathbf{a}_i + \mathbf{e}_i \\
&\Leftrightarrow \\
\mathbf{y}_i - \Lambda \mathbf{y}_i &= \mathbf{a}_i + \mathbf{e}_i \\
&\Leftrightarrow \\
(\mathbf{I} - \Lambda) \mathbf{y}_i &= \mathbf{a}_i + \mathbf{e}_i \\
&\Leftrightarrow \\
\mathbf{y}_i &= (\mathbf{I} - \Lambda)^{-1} \mathbf{a}_i + (\mathbf{I} - \Lambda)^{-1} \mathbf{e}_i
\end{aligned} \tag{5}$$

Este sistema representa, de hecho, las ecuaciones del siguiente modelo animal multicarácter (MTM, Henderson y Quaas, 1976):

$$\mathbf{y}_i = \mathbf{a}_i^* + \mathbf{e}_i^* \tag{6}$$

con

$$\mathbf{a}_i^* = (\mathbf{I} - \Lambda)^{-1} \mathbf{a}_i \tag{7}$$

y

$$\mathbf{e}_i^* = (\mathbf{I} - \Lambda)^{-1} \mathbf{e}_i. \tag{8}$$

Ambos modelos, SEM y MTM, son equivalentes dado

$$\Sigma_0^* = (\mathbf{I} - \Lambda)^{-1} \Sigma_0 \left[(\mathbf{I} - \Lambda)^{-1} \right]^t \tag{9}$$

y

$$\mathbf{R}_0^* = (\mathbf{I} - \Lambda)^{-1} \mathbf{R}_0 \left[(\mathbf{I} - \Lambda)^{-1} \right]^t \tag{10}$$

donde Σ_0^* y R_0^* representan las matrices de varianza y covarianzas de los valores de cría y de los errores bajo este MTM.

Esta equivalencia permite al usuario pasar de un modelo a otro. En particular, los valores de cría predichos bajo un modelo SEM se pueden obtener fácilmente ajustando un MTM con un programa estándar de resolución de ecuaciones de modelos mixtos y luego resolviendo

$$\mathbf{a}_i = (\mathbf{I} - \mathbf{\Lambda}) \mathbf{a}_i^* \quad [11]$$

Capítulo 3. *Impacto del uso de modelos de ecuaciones estructurales en la respuesta a la selección*

CAPÍTULO 3

IMPACTO DEL USO DE MODELOS DE ECUACIONES ESTRUCTURALES EN LA RESPUESTA A LA SELECCIÓN

La hipótesis principal planteada para esta tesis expresa que la absorción de las relaciones causales en correlaciones genéticas induce respuesta a la selección en caracteres no seleccionados. Para poner a prueba esta hipótesis se llevó a cabo un experimento de simulación estocástica. El abordaje metodológico permitió controlar todas las fuentes de variación durante el proceso de simulación, construir así escenarios definidos y conocer durante todo el proceso el valor real que quiere estimarse.

3.1 MATERIALES Y MÉTODOS

El experimento consistió en crear una población estructurada en grupos de hembras (rodeos) y un único pool de machos de modo que todos los machos produzcan descendencia en todos los grupos de hembras. Cada individuo fue evaluado para cinco caracteres relacionados causalmente de acuerdo con la red propuesta en la Figura 3.1. La elección de esta estructura en lugar de otras tuvo que ver con la posibilidad de evaluar el impacto en un amplio rango de relaciones, donde ocurren efectos directos e indirectos, positivos o negativos.

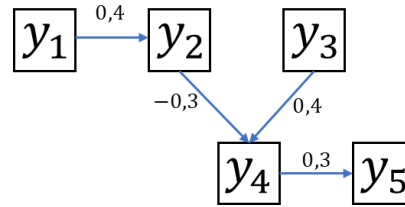


Figura 3.1. Estructura causal simulada. Las cajas representan caracteres evaluados y las flechas describen relaciones causales entre caracteres, los números sobre estas representan los coeficientes de causalidad.

Más específicamente, la población simulada estaba compuesta por cuatro rodeos de 250 hembras cada uno y un único pool de 50 machos. En cada generación, cada macho se apareaba con cinco hembras de cada rodeo generando así 250 crías, 125 machos y 125 hembras. De esta progenie, 75 hembras fueron seleccionadas para reemplazar a un número equivalente de hembras del rodeo que eran eliminadas al azar (lo cual representaba una tasa de reposición del 30%). En el caso de los machos, del total de animales nacidos se seleccionaban 25 sin importar el rodeo de origen para reponer la misma cantidad en el pool, también descartados al azar (50% de reposición).

Durante las primeras cinco generaciones la selección de la reposición se realizó al azar. A partir de la sexta generación, los reproductores selectos para reposición se obtuvieron por truncación de la distribución de los valores de cría estimados para algún carácter determinado por el escenario en evaluación. Las estimaciones fueron obtenidas mediante MTM y SEM en ejecuciones independientes de la simulación, pero iniciadas con el mismo generador de números aleatorios (semilla). Este proceso se repitió a lo largo de seis generaciones.

Los fenotipos de la población simulada se generaron empleando la ecuación [2], donde ahora y_i es un vector de dimensión 5×1 que contiene los valores fenotípicos para los

cinco caracteres evaluados en el individuo i , Λ es la matriz de coeficientes estructurales, definida de acuerdo con lo expresado en la Figura 3.1 como

$$\Lambda = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -0.3 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0.3 & 0 \end{bmatrix} \quad [14]$$

y $\mathbf{a}_i$ y $\mathbf{e}_i$ son los vectores que contienen los valores de cría y los errores.

En la primera generación, los valores de cría fueron muestreados de la siguiente distribución Normal multivariada:

$$\mathbf{a}_i \sim N(\mathbf{0}, \Sigma_0) \quad [15]$$

En una primera instancia, la matriz de co-varianzas genéticas fue definida como

$$\Sigma_0 = 0,5 \mathbf{I}_5$$

lo cual implica una ausencia de correlaciones genéticas entre los caracteres de la red causal. Si bien este escenario representa una situación poco probable en sentido práctico, garantizó que no haya superposición entre efectos causales y correlaciones genéticas, facilitando así la evaluación del impacto del uso de ambos modelos en la respuesta a la selección.

A partir de la segunda generación, los valores de cría se generaron mediante la ecuación

$$\mathbf{a}_i = \frac{1}{2}(\mathbf{a}_s + \mathbf{a}_d) + \phi_i \quad [16]$$

donde a_s y a_d son los valores de cría del padre y la madre respectivamente y ϕ_i es el residuo de segregación mendeliana (Foulley y Chevallet, 1981), obtenido por muestreo de la distribución

$$\phi_i \sim N\left(\mathbf{0}, \frac{1}{2} \left[1 - \frac{F_s + F_d}{2}\right] \Sigma_0\right) \quad [17]$$

Por su parte, los errores fueron muestreados de

$$e_i \sim N(\mathbf{0}, \mathbf{R}_0) \quad [18]$$

donde

$$\mathbf{R}_0 = \mathbf{I}_5$$

de forma tal de garantizar la identificabilidad del modelo causal (Varona y col., 2007; Wu y col., 2010; Varona y González-Recio, 2023).

En cada generación donde se realizó selección, los valores de cría estimados para cada animal fueron obtenidos ajustando el MTM equivalente mediante el software BLUPF90 (Misztal y col., 2018). Los parámetros de covarianza correspondientes a este modelo se especificaron según las ecuaciones [9] y [10]. Cabe destacar aquí que Σ_0 y $\mathbf{R}_0$ son las matrices de varianza y covarianzas de los valores de cría y de los errores usadas en el proceso de simulación, de modo evitar errores en las estimaciones de los valores de cría inducidos por el uso de parámetros de co-varianza incorrectos. Una vez obtenidas las soluciones del MTM, los valores de cría estimados bajo SEM se obtuvieron resolviendo la ecuación

$$\hat{\mathbf{a}} = (\mathbf{I} - \mathbf{\Lambda}) \hat{\mathbf{a}}^* \quad [21]$$

El proceso descrito previamente fue replicado en tres escenarios diferentes de acuerdo con el carácter que se utilizó como criterio de selección: 1, 4 o 5, seleccionando en todos los casos por truncación a los animales con valores de cría estimados superiores.

Adicionalmente, se replicaron también todos estos escenarios considerando una estructura de covarianzas genéticas no diagonal, definida por:

$$\Sigma_0 = \begin{bmatrix} 0.5 & 0.3 & 0 & 0 & 0 \\ & 0.5 & 0 & -0.2 & 0 \\ & & 0.5 & 0.25 & 0 \\ & Sym. & & 0.5 & 0.3 \\ & & & & 0.5 \end{bmatrix} \quad [22]$$

Finalmente, y con el objetivo de descartar errores debidos a la aleatorización, el proceso completo de simulación fue replicado 100 veces, modificando los generadores de números aleatorios (semillas). En la Figura 3.2 se muestra una representación esquemática del diseño del experimento de simulación. El programa de simulación se desarrolló en R (R Core Team, 2022). El código se encuentra disponible en el siguiente repositorio de GitHub: https://github.com/mcbonamy/SEM_popsim.

Nivel 1	Semilla	$\left\{ \begin{array}{c} 1 \\ 2 \\ \vdots \\ 100 \end{array} \right.$
Nivel 2	Matriz de covarianzas genéticas	$\left\{ \begin{array}{c} \text{Diagonal} \\ \text{No diagonal} \end{array} \right.$
Nivel 3	Carácter	$\left\{ \begin{array}{c} 1 \\ 4 \\ 5 \end{array} \right.$
Nivel 4	Modelo de estimación	$\left\{ \begin{array}{c} \text{MTM} \\ \text{SEM} \end{array} \right.$

Figura 3.2. Representación esquemática del diseño del experimento de simulación.

Una vez finalizada la ejecución del programa de simulación, la respuesta a la selección fue evaluada en forma gráfica a través del cambio en las medias fenotípicas de los animales nacidos en cada generación. En cada escenario se contrastó estadísticamente, además, la respuesta acumulada luego de las seis generaciones de selección al utilizar como criterio de selección estimaciones de los valores de cría obtenidas mediante un MTM vs un SEM. En particular, se realizó una prueba de diferencia de medias normales basadas en un estadístico t (Student, 1908) a partir de las 100 réplicas de la simulación.

3.2 RESULTADOS

Tal como se describió en el Capítulo 2, para cualquier SEM existe un MTM equivalente. En particular, las equivalencias explicitadas en las ecuaciones [9] y [10] implican que las relaciones causales son absorbidas por los parámetros de varianzas y covarianzas al ajustar un MTM. Esto se ve claramente en la matriz de covarianzas genéticas del modelo MTM

equivalente al modelo SEM simulado, donde se aprecian valores mayores en términos absolutos para las varianzas de los caracteres 2, 4 y 5 y valores no nulos para las covarianzas genéticas que fueron especificadas en cero en la simulación (Tabla 3.1).

Tabla 3.1. Varianzas y covarianzas del modelo multicarácter (MTM) equivalente derivadas de los parámetros del modelo de ecuaciones estructurales (SEM) simulado. Los valores superiores en la diagonal y en el triángulo superior corresponden al escenario sin covarianzas genéticas, mientras que los valores inferiores en la diagonal y triángulo inferior al escenario con algunas covarianzas genéticas diferentes de cero

	σ_1^*	σ_2^*	σ_3^*	σ_4^*	σ_5^*
σ_1^*	0,500 0,500	0,200	0	-0,060	-0,018
σ_2^*	0,500	0,580 0,820	0	-0,174	-0,052
σ_3^*	0	0	0,500 0,500	0,200	0,060
σ_4^*	-0,150	-0,446	0,450	0,632 0,974	0,190
σ_5^*	-0,045	-0,134	0,135	0,592	0,557 0,768

3.2.1 Respuesta a la selección en ausencia de correlaciones genéticas

Al ejercer presión de selección sobre el carácter 1, aquel ubicado más arriba en la red causal, se obtuvo exactamente la misma respuesta a la selección en los cinco caracteres independientemente de si criterio de selección fueron los valores de cría estimados mediante un SEM o los estimados mediante un MTM. De hecho, los verdaderos valores de cría de los progenitores seleccionados como padres a lo largo de las generaciones fueron los mismos en ambas poblaciones simuladas (recuérdese que se utilizaba la misma semilla para generar ambas poblaciones). El carácter 1 mostró una respuesta superior al resto de los caracteres,

seguida por la respuesta en el carácter 2 y el carácter 4. Los caracteres 3 y 5, por su parte, no evidenciaron respuesta a la selección (Figura 3.3). En la Tabla 3.2 se presentan las medias fenotípicas y sus correspondientes errores estándares calculados sobre las 100 réplicas simuladas luego de seis generaciones de selección.

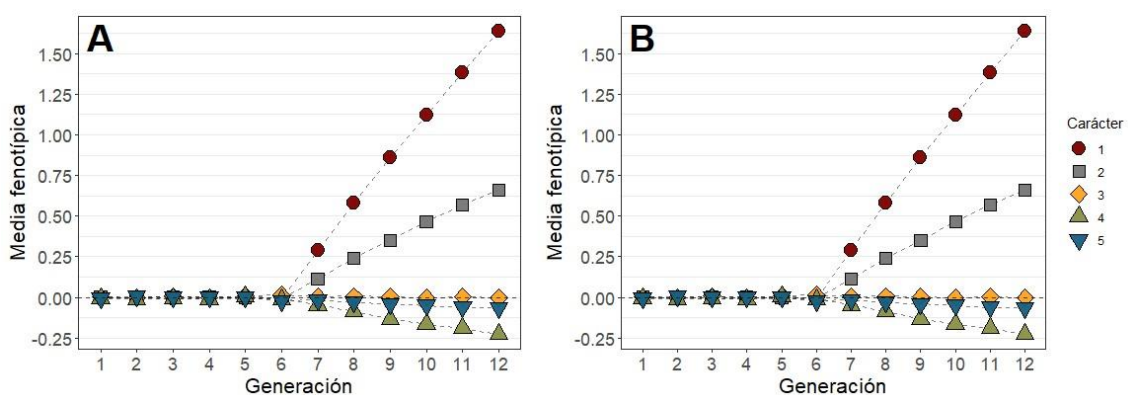


Figura 3.3. Respuesta a la selección a lo largo de las generaciones al utilizar como criterio de selección estimaciones de los valores de cría para el carácter 1 obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) en el escenario sin correlaciones genéticas entre caracteres. Cada punto del gráfico representa la media fenotípica por generación promediada sobre las 100 réplicas de la simulación.

Cuando, en cambio, la presión de selección fue ejercida sobre el carácter 4, si bien se obtuvo respuesta tanto al utilizar estimaciones obtenidas mediante el MTM como mediante el SEM como criterio de selección, la magnitud de esta fue mayor en el primer caso. Por otro lado, en ambos casos se observó también una respuesta indirecta a la selección en el carácter 5. En contraste, sólo se obtuvo respuesta a la selección indirecta en los caracteres 1, 2 y 3 cuando se aplicaron como criterio de selección los valores de cría estimados bajo el MTM (Tabla 3.2 y Figura 3.4).

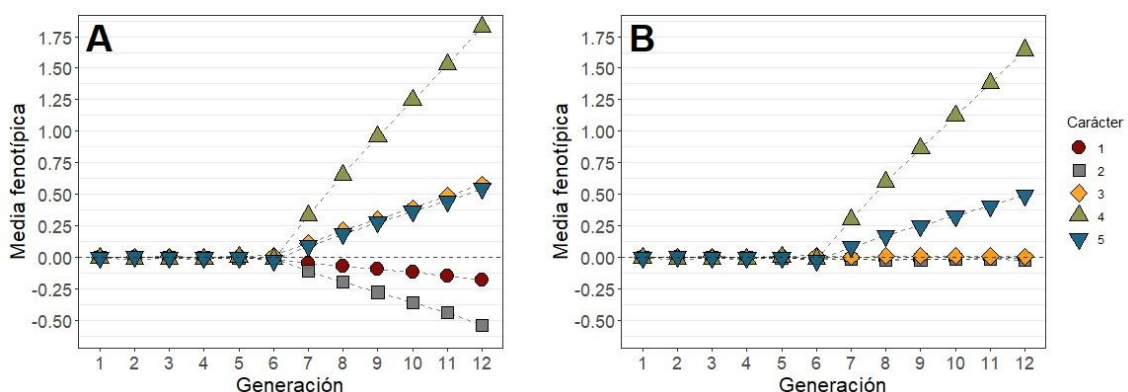


Figura 3.4. Respuesta a la selección a lo largo de las generaciones al utilizar como criterio de selección estimaciones de los valores de cría para el carácter 4 obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) en el escenario sin correlaciones genéticas entre caracteres. Cada punto del gráfico representa la media fenotípica por generación promediada sobre las 100 réplicas de la simulación.

En la Figura 3.5 se presenta el caso cuando la presión de selección fue aplicada sobre el carácter 5, al final de la red causal. Al igual que en caso anterior, se observó una respuesta directa en este carácter independientemente del criterio de selección empleado, pero de mayor magnitud en el caso de los valores de cría estimados mediante un MTM (Tabla 3.2). Por otro lado, sólo se observó una respuesta indirecta a la selección en este último caso. En este punto es interesante notar que esta respuesta tuvo el mismo sentido que las relaciones causales entre los caracteres involucrados. Por el contrario, cuando se utilizó como criterio de selección los valores de cría estimados mediante un SEM no se generó respuesta indirecta a la selección aguas arriba de la red causal.

Tabla 3.2. Respuesta acumulada luego de seis generaciones de selección al utilizar como criterio de selección estimaciones de los valores de cría obtenidas mediante un modelo animal multicarácter (MTM) o mediante un modelo de ecuaciones estructurales (SEM) en el escenario sin correlaciones genéticas entre caracteres.

Carácter seleccionado	Carácter respuesta	Respuesta a la selección ¹		<i>p</i> -valor contraste ²
		MTM	SEM	
		Promedio ± EE	Promedio ± EE	
1	1	1,638 ± 0,014	1,638 ± 0,014	1
	2	0,662 ± 0,017	0,662 ± 0,017	1
	3	-0,003 ± 0,018	-0,003 ± 0,018	1
	4	-0,227 ± 0,019	-0,227 ± 0,019	1
	5	-0,066 ± 0,018	-0,066 ± 0,018	1
4	1	-0,176 ± 0,018	-0,006 ± 0,018	< 0,001
	2	-0,536 ± 0,020	-0,025 ± 0,017	< 0,001
	3	0,577 ± 0,017	0,008 ± 0,018	< 0,001
	4	1,825 ± 0,015	1,643 ± 0,015	0,031
	5	0,543 ± 0,018	0,491 ± 0,016	< 0,001
5	1	-0,052 ± 0,016	0,004 ± 0,017	0,019
	2	-0,150 ± 0,018	-0,003 ± 0,018	< 0,001
	3	0,196 ± 0,017	0,002 ± 0,019	< 0,001
	4	0,585 ± 0,018	-0,019 ± 0,020	< 0,001
	5	1,716 ± 0,015	1,608 ± 0,015	< 0,001

¹ Los valores corresponden a las medias fenotípicas estimadas y sus errores estándares (EE) sobre las 100 réplicas de la simulación; ²Contrastes entre los promedios de fila.

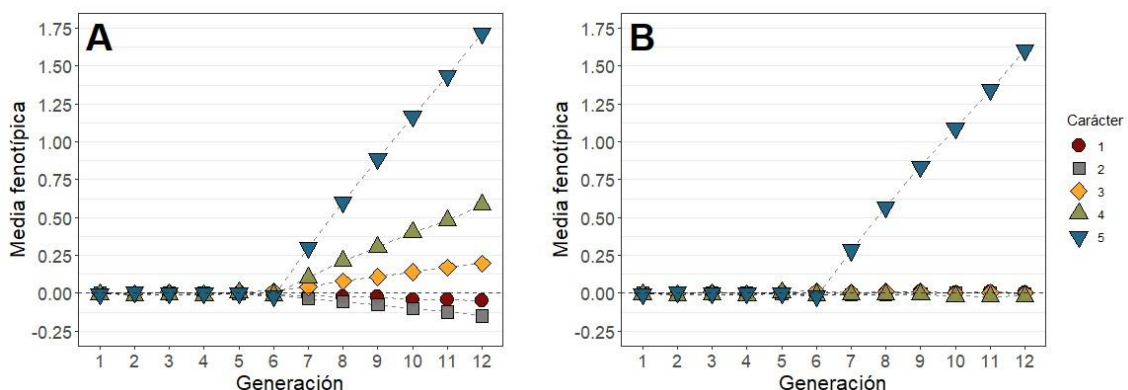


Figura 3.5. Respuesta a la selección a lo largo de las generaciones al utilizar como criterio de selección estimaciones de los valores de cría para el carácter 5 obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) en el escenario sin correlaciones genéticas entre caracteres. Cada punto del gráfico representa la media fenotípica por generación promediada sobre las 100 réplicas de la simulación.

3.2.2 Respuesta a la selección cuando efectos causales y correlaciones genéticas actúan en forma simultánea

Como se explicara en la sección de Materiales y Métodos, también se evaluaron escenarios de simulación en que los que efectos causales y correlaciones genéticas actuaban en simultáneo en la determinación del fenotipo. El patrón de resultados en términos de la respuesta a la selección bajo estos escenarios fue muy similar al caso donde las correlaciones genéticas estaban ausentes.

En primer lugar, la respuesta lograda al ejercer presión de selección sobre el carácter 1 fue la misma sin importar el modelo utilizado para estimar los valores de cría (MTM vs SEM; Figura 3.6). Notoriamente, la mayor respuesta observada en esta situación se dio en el carácter 2 (respuesta indirecta) y, al igual que en el escenario donde las correlaciones

genéticas estaban ausentes, no hubo respuesta en el carácter 3. Estos resultados se reportan numéricamente en la Tabla 3.3.

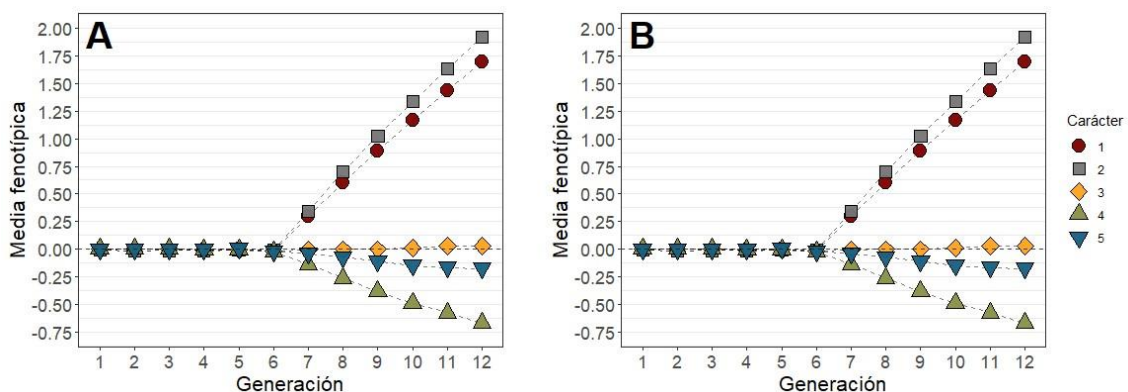


Figura 3.6. Respuesta a la selección a lo largo de las generaciones al utilizar como criterio de selección estimaciones de los valores de cría para el carácter 1 obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) en el escenario con correlaciones genéticas entre caracteres distintos de cero. Cada punto del gráfico representa la media fenotípica por generación promediada sobre las 100 réplicas de la simulación.

Cuando la presión de selección fue trasladada al carácter 4 o al carácter 5, la respuesta directa a la selección lograda difirió dependiendo del criterio de selección y, como era esperable, de la magnitud de las correlaciones genéticas. Puede verse, por ejemplo, que la respuesta a la selección por el carácter 4 indujo una respuesta directa levemente mayor cuando el criterio de selección se basó en estimaciones de los valores de cría obtenidos mediante un MTM. Por su parte, la respuesta indirecta a la selección también difirió según el criterio de selección, donde se destacan una respuesta indirecta menor en y_1 y una mayor en y_5 al utilizar estimaciones del valor de cría obtenidas mediante SEM (Figura 3.7, Tabla 3.3).

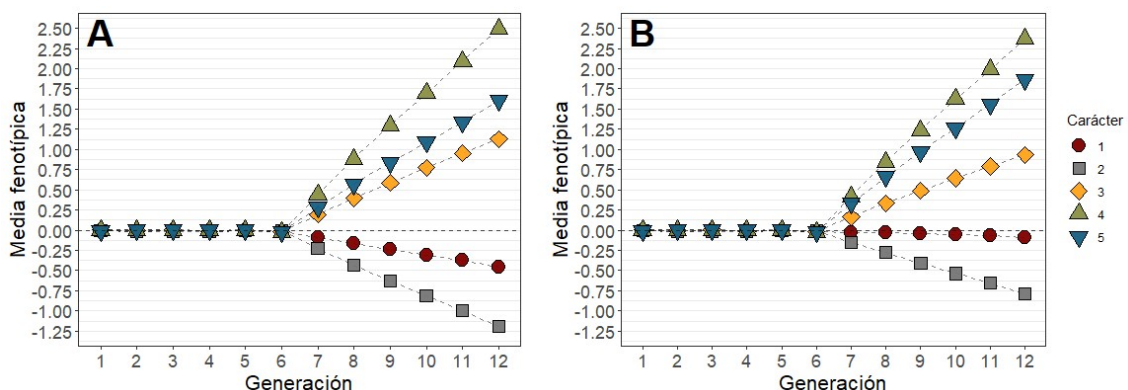


Figura 3.7. Respuesta a la selección a lo largo de las generaciones al utilizar como criterio de selección estimaciones de los valores de cría para el carácter 4 obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) en el escenario con correlaciones genéticas entre caracteres distintas de cero. Cada punto del gráfico representa la media fenotípica por generación promediada sobre las 100 réplicas de la simulación.

La Figura 3.8 muestra los cambios en las medias fenotípicas en respuesta a la selección por el carácter 5, al final de la red causal. Se observó un comportamiento similar al del caso anterior, con una mayor respuesta relativa cuando el criterio de selección fueron los valores de cría estimados mediante un MTM. Esta diferencia fue particularmente notoria en los caracteres 1, 2 y 3, los cuales prácticamente no mostraron respuesta cuando el criterio de selección fue el alternativo. Los resultados numéricos se presentan en la Tabla 3.3.

Tabla 3.3. Respuesta acumulada luego de seis generaciones de selección al utilizar como criterio de selección estimaciones de los valores de cría obtenidas mediante un modelo animal multicarácter (MTM) o mediante un modelo de ecuaciones estructurales (SEM) en el escenario con correlaciones genéticas entre caracteres.

Carácter seleccionado	Carácter respuesta	Respuesta a la selección ¹		<i>p</i> -valor contraste ²
		MTM	SEM	
		Promedio ± EE	Promedio ± EE	
1	1	1,696 ± 0,013	1,696 ± 0,013	1
	2	1,921 ± 0,018	1,921 ± 0,018	1
	3	0,028 ± 0,016	0,028 ± 0,016	1
	4	-0,667 ± 0,021	-0,667 ± 0,021	1
	5	-0,175 ± 0,020	-0,175 ± 0,020	1
4	1	-0,453 ± 0,016	-0,085 ± 0,015	< 0,001
	2	-1,195 ± 0,020	-0,787 ± 0,021	< 0,001
	3	1,127 ± 0,014	0,933 ± 0,017	< 0,001
	4	2,490 ± 0,018	2,373 ± 0,020	< 0,001
	5	1,599 ± 0,020	1,864 ± 0,017	< 0,001
5	1	-0,130 ± 0,015	0,026 ± 0,016	< 0,001
	2	-0,461 ± 0,021	-0,088 ± 0,022	< 0,001
	3	0,450 ± 0,017	0,095 ± 0,016	< 0,001
	4	1,827 ± 0,022	1,269 ± 0,025	< 0,001
	5	2,170 ± 0,019	2,077 ± 0,018	< 0,001

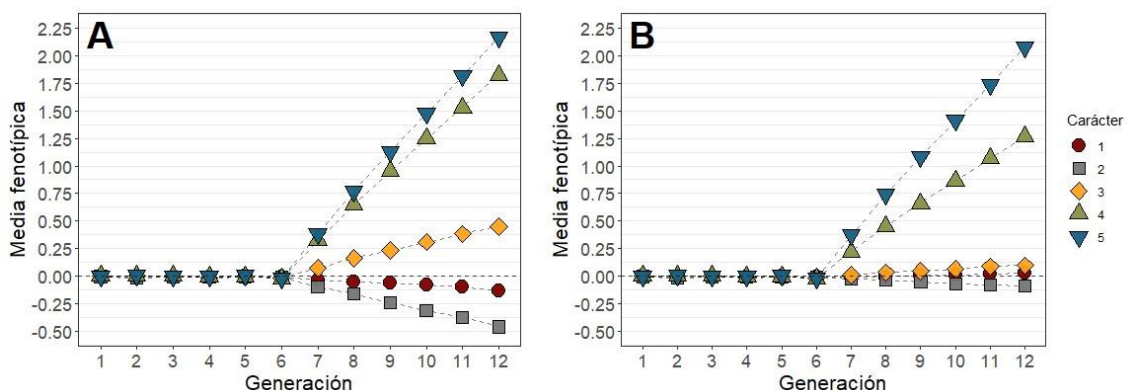


Figura 3.8. Respuesta a la selección a lo largo de las generaciones al utilizar como criterio de selección estimaciones de los valores de cría para el carácter 5 obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) en el escenario con correlaciones genéticas entre caracteres distintas de cero. Cada punto del gráfico representa la media fenotípica por generación promediada sobre las 100 réplicas de la simulación.

3.3 DISCUSIÓN

En la Discusión de su trabajo sobre las posibles ventajas de los modelos de ecuaciones estructurales (SEM) para el mejoramiento genético, Valente y col. (2013) afirman que, aun cuando existan relaciones complejas entre los fenotipos, para seleccionar con base en valores de cría estimados y para estimar respuestas correlacionadas no es necesario conocer la estructura causal subyacente. En ese mismo trabajo, en la sección dedicada a redes causales con efectos recursivos, esto es, cuando relaciones causales y correlaciones genéticas se manifiestan en simultáneo, afirman que los modelos multicarácter (MTM) tienen la misma capacidad predictiva que los SEM, independientemente de la red causal, que, de hecho, no es relevante para un programa de mejora (Valente y col., 2013, p. 565). Estos argumentos dieron lugar a algunos interrogantes que motivaron la presente investigación. En primer lugar, suponiendo que la relación observada en un par de caracteres obedece a una relación

de naturaleza causal pero no a una correlación genética, ¿no podría este conocimiento contribuir a aplicar presión de selección sobre el carácter situado aguas debajo de la red sin afectar la respuesta en su ancestro causal? Por otro lado, asumiendo una estructura causal como la que proponen los autores en la Figura 2 (Valente y col., 2013, P.564), si fuese posible separar el efecto causal de la correlación genética, identificando por separado y estimando de forma precisa los coeficientes de causalidad y los parámetros de covarianza genética, ¿no podría esta nueva fuente de información contribuir a minimizar la respuesta indirecta a la selección en caracteres que no forman parte del objetivo de selección?

Para responder a estas preguntas, se diseñó y ejecutó el experimento de simulación descrito en este capítulo. Primero, se simuló una estructura causal donde los caracteres no están correlacionados genéticamente, con la hipótesis de que, en contraste con los SEM, los MTM absorberían las relaciones causales entre los caracteres en los parámetros que cuantifican las correlaciones genéticas y, como consecuencia, el uso de valores de cría estimados mediante este modelo como criterio de selección resultaría en cambios en la respuesta a la selección lograda en caracteres localizados aguas arriba respecto del carácter objetivo de selección. Los resultados obtenidos confirmaron esta hipótesis. Resulta evidente, al menos bajo este escenario, que conocer el modelo causal puede ayudar a los mejoradores proporcionando herramientas que ajusten mejor sus objetivos a sus criterios de selección. A continuación, se desarrollará esta misma idea con más detalle.

3.3.1 Respuesta a la selección cuando los caracteres están causal pero no genéticamente relacionados entre sí

Cuando la presión de selección fue aplicada sobre un carácter que actúa como “raíz” causal (*causal root* en inglés; no hay otras variables que influyan a esta causalmente; Shipley,

2016) el uso de valores de cría estimados mediante SEM o MTM como criterio de selección indujo la misma respuesta a la selección, tanto directa como indirecta. Esto puede interpretarse como una consecuencia directa de la equivalencia entre modelos derivada por Gianola y Sorensen (2004). Si consideramos que el carácter en cuestión no tiene ancestros causales, es decir que no hay flechas que apunten hacia este, no hay entonces recursión en otros caracteres, y los valores de cría bajo uno u otro modelo se definen de la misma manera.

Sin embargo, al trasladar la presión de selección aguas abajo en la red causal, las respuestas a la selección tanto directa como indirecta difieren entre los criterios de selección. La diferencia más notoria es la ausencia de respuesta correlacionada al seleccionar mediante SEM en caracteres ubicados aguas arriba cuando la correlación genética es igual a cero. Este resultado es debido a la independencia condicional entre caracteres (Pearl, 2000; Shipley, 2016): al condicionar en la red causal los valores de cría del carácter objetivo de selección se vuelven independientes de los valores de cría de los caracteres ubicados aguas arriba.

Este último argumento puede ser intuitivamente interpretado apelando a la teoría de grafos (Pearl, 2000). Para ilustrar este punto resulta útil revisar la Figura 3.9.A, mostrada a continuación. En esta figura los recuadros representan variables (o vértices) y las flechas denotan las relaciones causales entre estas. Dado que no existen correlaciones entre efectos genéticos y entre efectos ambientales, ya que no hay flechas dobles entre estos, la información causal solo puede viajar siguiendo las flechas. Cuando una variable se encuentra en un estado no condicionado (en este caso implica que el carácter no está siendo seleccionado) permite el flujo de información a través de la red. Esto explica por ejemplo porqué dada la vía

$$y_1 \rightarrow y_2 \rightarrow y_4 \rightarrow y_5$$

la selección por el carácter 1 tiene consecuencias en la respuesta a la selección de los demás caracteres. Por contrario, condicionar en una variable ubicada más abajo en la red no permite que la información se transmita aguas arriba hacia sus ancestros causales. Entonces, cuando el objetivo de selección se ubica en el carácter 4, no se registra respuesta en los caracteres 2 y 3 debido a la naturaleza direccional de los efectos causales.

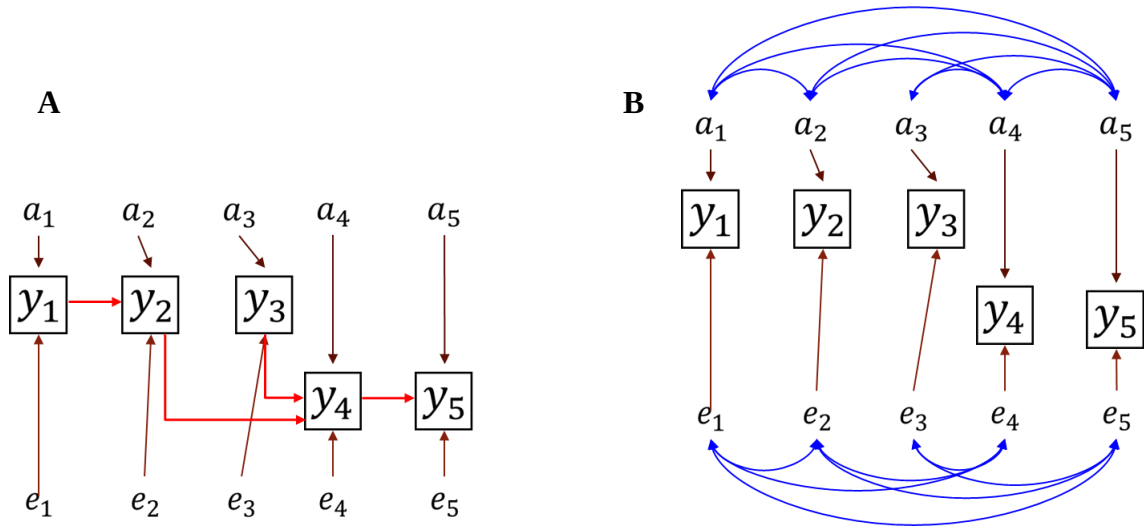


Figura 3.9. A. Modelo causal simulado en ausencia de correlaciones genéticas y estimado utilizando modelos de ecuaciones estructurales (SEM). **B.** Representación del empleo de modelo animal multicarácter (MTM) como modelo de estimación de la red descrita por la figura A. Las flechas rojas indican relaciones causales entre los caracteres mientras que las flechas azules indican covarianzas entre efectos genéticos y ambientales producto de la absorción de las relaciones causales en estos parámetros.

Cuando el criterio de selección establecido se basa en el uso de valores de cría estimados mediante un MTM, la selección por caracteres ubicados a lo largo de la red causal se traslada a los demás caracteres a través de covarianzas entre los valores genéticos y del error, dado que estos efectos son recíprocos entre las variables involucradas. En este caso particular, donde no existen correlaciones genéticas verdaderas entre las variables, las

correlaciones asumidas por el modelo de estimación inducen un sesgo en las estimaciones de los valores de cría, y traccionan la respuesta indirecta a la selección (Figura 3.9.B).

Esta situación se evidenció en las diferencias entre los valores de cría estimados por uno u otro modelo, y en su sesgo respecto al verdadero valor de cría simulado (Anexo 1). Al seleccionar por el carácter 1 ambos modelos arrojaron las mismas estimaciones de los valores de cría. Sin embargo, al mover el objetivo de selección aguas abajo, los valores de cría estimados mediante el MTM fueron sesgados para los caracteres situados aguas arriba, con la magnitud y signo de ese sesgo atado a los verdaderos valores de los coeficientes de causalidad.

Más allá de las diferencias cualitativas descritas, al comparar la respuesta a la selección lograda al utilizar MTM o SEM como criterio de selección se observaron también diferencias cuantitativas. Para describirlas, es útil centrarse nuevamente en lo que ocurre al definir al carácter 4 como objetivo de selección. En este caso, la respuesta directa a la selección lograda al utilizar los valores de cría estimados mediante un MTM fue significativamente mayor que aquella obtenida mediante el uso de valores de cría estimados mediante un SEM. Nótese que los valores de cría estimados mediante un MTM son lineales en los valores de cría estimados de los caracteres aguas arriba de la red causal. En este caso particular, los candidatos seleccionados tendieron, en virtud de las correlaciones genéticas, a ser aquellos con valores de cría estimados más bajos para el carácter 2 ($\sigma_{42} = -0,17$) y más altos para el carácter 3 ($\sigma_{43} = 0,20$). Luego, la selección indirecta sobre estos últimos caracteres produjo respuesta a nivel fenotípico y, por lo tanto, considerando el efecto causal que estos caracteres ejercen sobre el carácter 4 (objetivo de selección), la respuesta a la selección alcanzada en la generación siguiente fue mayor en comparación con lo que se obtiene al utilizar estimaciones bajo un SEM.

Esta mayor respuesta a la selección fue a expensas de cambios en las medias fenotípicas de caracteres que podrían no estar involucrados en los objetivos de selección. Si esta respuesta involuntaria no implica un problema en el plan de mejora, y se plantea como objetivo maximizar la respuesta en el carácter objetivo en el marco del empleo de estimaciones de los valores de cría mediante un SEM, entonces pueden aplicarse índices de selección (Hazel, 1943) combinando los valores de crías estimados de los caracteres aguas arriba con el fin de alcanzar una respuesta similar a la obtenida al utilizar estimaciones obtenidas mediante MTM.

En este caso, por ejemplo, la respuesta a la selección lograda al seleccionar por los caracteres 4 y 5 utilizando estimaciones de los valores de cría obtenidas por SEM igualaría a aquella lograda utilizando estimaciones obtenidas por MTM si se aplicara un índice que pondera las estimaciones de los caracteres ubicados aguas arriba con los apropiados coeficientes de causalidad (ver la ecuación [7]). Explícitamente:

$$\begin{aligned}\hat{a}_4^* &= \lambda_{21}\lambda_{42}\hat{a}_1 + \lambda_{42}\hat{a}_2 + \lambda_{43}\hat{a}_3 + \hat{a}_4 \\ \hat{a}_5^* &= \lambda_{21}\lambda_{42}\lambda_{54}\hat{a}_1 + \lambda_{42}\lambda_{54}\hat{a}_2 + \lambda_{43}\lambda_{54}\hat{a}_3 + \lambda_{54}\hat{a}_4 + \hat{a}_5\end{aligned}$$

3.3.2 Respuesta a la selección cuando los efectos causales y las correlaciones genéticas actúan de forma simultánea

Si bien el escenario planteado hasta aquí puede parecer un tanto ingenuo o irreal, es útil para interpretar las consecuencias de seleccionar con base en uno u otro criterio de selección. A continuación, se discuten los resultados del escenario más realista en el cual los efectos causales y las correlaciones genéticas actuaban en conjunto.

En este caso, los resultados fueron muy similares a los obtenidos en el escenario donde no existían las correlaciones genéticas, con las diferencias numéricas propias del caso.

En primer lugar, la presión de selección aplicada sobre el carácter 1 produjo la misma respuesta en todos los caracteres localizados aguas abajo en la red causal, independientemente de que los valores de cría fueran los estimados por SEM o MTM. En segundo lugar, cuando la presión de selección fue trasladada a un carácter localizado aguas abajo en la red causal, se observó una respuesta correlacionada en los caracteres ubicados aguas arriba cuando los valores de cría fueron estimados mediante un MTM. Y aunque también hubo una respuesta al seleccionar utilizando estimaciones basadas en un SEM, la magnitud fue marcadamente menor. Estos resultados ponen de manifiesto, una vez más, el impacto de la absorción de las relaciones causales en las covarianzas genéticas.

Un punto que merece ser tenido en cuenta es que los valores de cría obtenidos ajustar un SEM no son los mismos que aquellos obtenidos mediante un MTM, y deben ser interpretados de forma diferente. Tal como expresan Gianola y Sorensen (2004), los primeros representan una estimación del componente genético aditivo asociado a

$$y_i - (I - \Lambda) y_i$$

Así, por ejemplo, el valor de cría estimado mediante SEM para el carácter 4 estima la contribución genética sobre

$$y_4 - \lambda_{42} y_2 - \lambda_{43} y_3.$$

En contraste, el valor de cría estimado mediante MTM estima el efecto genético total ejercido sobre el carácter 4, incluyendo el efecto de genes que controlan la expresión de este carácter en particular y de genes que ejercen su efecto sobre otros caracteres, relacionados causalmente con este (Anexo 2).

En el desarrollo de este trabajo se asumió una estructura de co-varianzas diagonal para los errores. De esta manera, se garantizó que el modelo fuera identificable (Varona y

González-Recio, 2023). Si, en cambio, se hubiera especificado una matriz de covarianzas del error no diagonal entre los caracteres evaluados, la respuesta a la selección hubiese resultado diferente a la obtenida en esta simulación (Deng y Kibota, 1995). Sin embargo, las diferencia en los resultados en términos absolutos se interpretaría de la misma manera: la absorción de los coeficientes de causalidad en los valores estimados de las varianzas y covarianzas (genéticas y del error) bajo MTM, generaría una respuesta a la selección en caracteres situados aguas arriba en la estructura causal cuya magnitud hubiera sido diferente a la obtenida bajo SEM.

3.3.3 Implicancias de estos resultados en los programas de mejoramiento genético

Desde la publicación del trabajo de Gianola y Sorensen (2004), los SEM han sido utilizados para evaluar redes causales para diferentes caracteres asociados a especies de interés ganadero. Los resultados expuestos en esta tesis proponen que los SEM pueden ser modelos útiles en los programas de mejoramiento.

Entre los ejemplos de aplicación reportados puede citarse a Jafaroghli y col. (2021), quienes estudiaron en ovinos la relación entre la edad al primer cordero, el número de corderos nacidos, el peso al nacimiento de la camada de corderos, el número de corderos destetados y el peso de al destete de la camada. Luego de estimar los coeficientes causales, concluyeron que la edad al primer parto es independiente del resto de los caracteres y que el número de corderos nacidos ejerce un efecto causal indirecto sobre el peso de la camada al destete. En el trabajo citado, se determinó que existe una relación causal positiva entre el peso al nacimiento y el peso al destete, por lo que mediante el empleo de estimaciones de los valores de cría obtenidas mediante SEM, podría aumentarse el peso al destete evitando una respuesta indirecta e indeseada en el peso al nacimiento; al menos hasta el límite impuesto

por las correlaciones genéticas. Algo similar puede plantearse para el peso al nacimiento y el número de corderos nacidos.

Otras oportunidades para la aplicación de SEM en el mejoramiento animal se presentan en función de reportes de relaciones causales entre caracteres al parto en vacas primíparas (Lopez de Maturana y col., 2009). En este caso, se observó una relación causal heterogénea entre el largo de gestación, la dificultad de parto y la probabilidad de natimorto. En gestaciones más cortas que los 280 días (duración fisiológica de la gestación 280 – 285 días), el largo de gestación tiene un efecto negativo sobre la probabilidad de natimortos. En otras palabras, cuanto más corta es la gestación mayor es la probabilidad de que la cría nazca muerta. En cambio, para gestaciones que superan los 280 días la probabilidad de tener un natimorto aumenta a medida que la gestación se alarga; es decir, el signo del coeficiente causal se invierte. A su vez en gestaciones de duración menor que 267, el largo de gestación no tiene impacto en la dificultad de parto, mientras que en gestaciones mas largas, la dificultad de parto aumenta a medida que lo hace el largo de gestación.

En cerdas en lactancia, Mora y col. (2022) determinaron las relaciones causales entre el consumo de alimento de la cerda, la ganancia de peso de las crías, los cambios de peso de la cerda, la acumulación de grasa de cobertura y el peso metabólico, y resaltan la importancia de la medición del consumo individual en planteos de mejora. Además, encuentran poco probable reducir en consumo individual de la cerda mientras se busca mejorar la productividad sin impactar en el engrasamiento.

En todos los casos mencionados se validaron las redes causales y se estimaron los parámetros de causalidad, varianza y covarianza. Esto es un prerrequisito importante para la aplicación de modelos de ecuaciones estructurales. Afortunadamente, la metodología para la determinación de la estructura subyacente a la relación entre las variables de una red causal

propuesta por Pearl (2000; Verma y Pearl, 1990), el algoritmo de causalidad inductiva (del inglés *inductive causation algorithm*), fue adaptada al mejoramiento animal por Valente y col (2010).

Capítulo 4. *Estimación de los parámetros del modelo*

CAPÍTULO 4

ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

La estimación de parámetros de varianza y covarianza constituye una tarea desafiante. A la hora de ajustar un modelo mixto, estos parámetros se asumen conocidos. En particular, los dos modelos de estimación utilizados en esta tesis (SEM y MTM) asumen que las varianzas y covarianzas genéticas son conocidas. Los SEM, además, asumen que los coeficientes que determinan las relaciones causales también se conocen. En los resultados derivados del capítulo anterior se utilizaron los verdaderos valores de estos parámetros a la hora de obtener los valores de cría predichos. Sin embargo, en una situación práctica con datos de campo sería necesario realizar una estimación previa de estos parámetros antes de ajustar los modelos para obtener los valores de cría estimados.

Aprovechando los datos generados por el experimento de simulación, entonces, se evaluaron dos metodologías para la estimación de los coeficientes de causalidad y los parámetros de varianza y covarianza de los SEM. Ambas metodologías se implementaron sobre la base de registros fenotípicos de las seis primeras generaciones, donde aún no se había ejercido presión de selección y la selección de la reposición era realizada al azar.

4.1 MATERIALES Y MÉTODOS

La primera metodología empleada en esta tesis involucró un procedimiento en dos etapas. En la primera de ellas se ajustó un análisis de confirmación de factores (CFA, por el acrónimo en inglés de *confirmatory factor análisis*) para obtener los coeficientes de causalidad. Luego, en la segunda etapa, se estimaron los componentes de varianza y covarianza bajo un MTM

utilizando el algoritmo de *average information restricted maximum likelihood* (AI-REML).

Los correspondientes parámetros bajo SEM se obtuvieron luego resolviendo

$$\Sigma_0 = (I - \Lambda) \Sigma_0^* (I - \Lambda)^t \quad [23]$$

$$R_0 = (I - \Lambda) R_0^* (I - \Lambda)^t \quad [24].$$

Este enfoque no había sido evaluado previamente y se evaluó dada su practicidad y la escasa demanda de recursos informáticos. Los análisis fueron ejecutados utilizando la función *cfa* del paquete *lavaan* (Rosseel, 2012) en R (R core team, 2022) y AIREMLF90 perteneciente a la familia de programas de BLUPF90 (Misztal y col., 2018).

La segunda metodología empleada para estimar los parámetros fue la estrategia llevada adelante en el reciente trabajo de Mora y col. (2022; Mora *comunicación personal*), previamente reportada por Rosa y col. (2011). La estrategia consiste en implementar un análisis bayesiano de un MTM declarando a los ancestros causales directos como covariables en las ecuaciones de sus descendientes (tal como se simboliza en la ecuación [2]) y considerando que las varianzas y covarianzas de los errores tienen una estructura diagonal. Bajo esta metodología, los coeficientes de regresión asociados a las covariables son interpretados como los coeficientes de causalidad, mientras que los componentes de varianza estimados son aquellos correspondientes al SEM. Estos análisis fueron llevados a cabo utilizando el software GibbsF90+ (Misztal y col., 2018), un software diseñado para estimar parámetros bajo MTM a través del empleo de un algoritmo de muestreo de Gibbs. Para la implementación del mismo se obtuvieron 80.000 muestras, eliminando las primeras 30.000 como período de calentamiento (*burn-in*) y almacenando, luego, una muestra por cada 100 ciclos de muestreo. La convergencia fue evaluada considerando el tamaño efectivo de la

muestra. Las medias de las distribuciones posteriores de los coeficientes de causalidad y las varianzas y covarianzas genéticas y del error fueron interpretadas como estimaciones puntuales de estos parámetros.

Para evaluar la calidad de las estimaciones, tanto en el escenario sin correlaciones genéticas como en el escenario con correlaciones genéticas, se calculó el error de estimación de cada uno de los parámetros, definido como la diferencia entre el valor estimado y el verdadero. Luego, a partir de los valores obtenidos sobre las 100 réplicas de la población simulada se obtuvo la distribución empírica de este error de estimación. Las estimaciones fueron consideradas apropiadas cuando el valor cero (0) estaba incluido dentro del intervalo definido por los cuantiles 0,025 y 0,975 de esta distribución, una medida denominada intervalo de alta densidad del 95% (HDI95% por su acrónimo en inglés).

4.2 RESULTADOS

En el escenario de simulación donde las covarianzas genéticas entre caracteres eran nulas, los errores de estimación de los coeficientes causales estuvieron centrados en el valor cero independientemente del método de estimación ensayado (Tabla 4.1). En todos los casos, los intervalos alta densidad del 95% incluyeron al verdadero valor del parámetro (ver Anexos 3 y 4, donde se presentan las distribuciones empíricas de los parámetros).

Con respecto a las varianzas y covarianzas genéticas del MTM, las estimaciones obtenidas mediante el algoritmo AI-REML estuvieron, en todos los casos, dentro de los límites de los correspondientes intervalos de alta densidad del 95% (Anexo 3). El empleo de estas estimaciones en conjunto con aquellas obtenidas para los coeficientes causales devolvió estimaciones para las varianzas y covarianzas genéticas bajo el modelo SEM con errores

centrados en cero (Tabla 4.2). Vale la pena destacar que las distribuciones empíricas de todas las covarianzas genéticas, obtenidas con cualquiera de los dos métodos de estimación ensayados, se concentraron alrededor del cero, su verdadero valor (Anexo 3).

Tabla 4.1. Errores de estimación de los coeficientes causales en el escenario sin correlaciones genéticas entre los caracteres. Los valores de la tabla corresponden al error de estimación medio, su desvío estándar (DE) y su intervalo de alta densidad del 95%, todos ellos obtenidos sobre las 100 réplicas de la simulación.

Estimador del parámetro	Valor verdadero	Método de estimación			
		CFA		Gibbs	
		Error de estimación $\pm$ DE	HDI95%	Error de estimación $\pm$ DE	HDI95%
$\widehat{\lambda}_{21}$	0,4	0,0010 $\pm$ 0,015	-0,032; 0,028	0,0025 $\pm$ 0,021	-0,037; 0,043
$\widehat{\lambda}_{42}$	-0,3	-0,0001 $\pm$ 0,013	-0,023; -0,024	-0,0011 $\pm$ 0,022	-0,041; -0,033
$\widehat{\lambda}_{43}$	0,4	0,0012 $\pm$ 0,016	-0,027; 0,033	-0,0044 $\pm$ 0,025	-0,045; 0,051
$\widehat{\lambda}_{54}$	0,3	0,0005 $\pm$ 0,015	-0,028; 0,034	-0,0044 $\pm$ 0,021	-0,051; 0,031

Referencias:

$\widehat{\lambda}_{ij}$ estimador del efecto causal ejercido por el carácter j en el carácter i .

CFA: método de estimación basado en un análisis de confirmación de factores.

Gibbs: método de estimación basado en definir las relaciones causales como covariables.

En cuanto a las varianzas y covarianzas de los términos de error, las estimaciones obtenidas también estuvieron dentro de los correspondientes intervalos de alta densidad en todos los casos. Para más detalle puede consultarse los Anexos 3 y 4.

Tabla 4.2. Errores de estimación de las varianzas y covarianzas genéticas del modelo de ecuaciones estructurales en el escenario sin correlaciones genéticas entre los caracteres. Los valores de la tabla corresponden al error de estimación medio, su desvío estándar (DE) y su intervalo de alta densidad del 95%, todos ellos obtenidos sobre las 100 réplicas de la simulación.

Estimador del parámetro	Valor verdadero	Método de estimación			
		Dos etapas		Gibbs	
		Error de estimación $\pm$ DE	HDI95%	Error de estimación $\pm$ DE	HDI95%
$\hat{\sigma}_1^2$	0,5	0,0033 $\pm$ 0,039	-0,064; 0,083	0,0084 $\pm$ 0,040	-0,061; 0,080
$\hat{\sigma}_2^2$	0,5	0,0048 $\pm$ 0,044	-0,073; 0,095	0,0082 $\pm$ 0,043	-0,067; 0,083
$\hat{\sigma}_3^2$	0,5	0,0053 $\pm$ 0,041	-0,071; 0,080	0,0079 $\pm$ 0,043	-0,089; 0,083
$\hat{\sigma}_4^2$	0,5	-0,0008 $\pm$ 0,044	-0,085; 0,080	0,0174 $\pm$ 0,044	-0,079; 0,094
$\hat{\sigma}_5^2$	0,5	0,0079 $\pm$ 0,045	-0,067; 0,102	0,0169 $\pm$ 0,044	-0,069; 0,106
$\hat{\sigma}_{12}$	0,0	0,0038 $\pm$ 0,024	-0,042; 0,046	-0,0080 $\pm$ 0,037	-0,078; 0,056
$\hat{\sigma}_{13}$	0,0	-0,0065 $\pm$ 0,026	-0,052; 0,047	0,0019 $\pm$ 0,019	-0,032; 0,037
$\hat{\sigma}_{14}$	0,0	-0,0073 $\pm$ 0,027	-0,059; 0,052	-0,0029 $\pm$ 0,024	-0,047; 0,051
$\hat{\sigma}_{15}$	0,0	0,0021 $\pm$ 0,030	-0,051; 0,065	0,0023 $\pm$ 0,021	-0,035; 0,041
$\hat{\sigma}_{23}$	0,0	-0,0005 $\pm$ 0,032	-0,063; 0,061	-0,0012 $\pm$ 0,023	-0,035; 0,046
$\hat{\sigma}_{24}$	0,0	0,000 $\pm$ 0,027	-0,053; 0,046	0,0011 $\pm$ 0,041	-0,068; 0,094
$\hat{\sigma}_{25}$	0,0	0,0002 $\pm$ 0,030	-0,063; 0,057	-0,0034 $\pm$ 0,025	-0,055; 0,036
$\hat{\sigma}_{34}$	0,0	0,0015 $\pm$ 0,026	-0,048; 0,050	0,0073 $\pm$ 0,043	-0,087; 0,084
$\hat{\sigma}_{35}$	0,0	-0,0005 $\pm$ 0,026	-0,049; 0,057	0,0019 $\pm$ 0,024	-0,041; 0,043
$\hat{\sigma}_{45}$	0,0	0,0061 $\pm$ 0,030	-0,052; 0,065	0,0091 $\pm$ 0,037	-0,060; 0,078

Referencias:

Dos etapas: método de estimación basado en combinar las estimaciones de los coeficientes de causalidad con las estimaciones de máxima verosimilitud restringida (REML) de las varianzas y covarianzas del modelo animal multicarácter equivalente.

Gibbs: método de estimación basado en definir un modelo con las relaciones causales como covariables y obtener las estimaciones de sus parámetros mediante un algoritmo del muestreo de Gibbs.

La inclusión de correlaciones genéticas entre los caracteres en el proceso de simulación introduce una nueva fuente de confusión que puede afectar la identificabilidad de

los parámetros. En este escenario, las estimaciones de todos los coeficientes de causalidad obtenidas mediante el análisis de confirmación de factores se desviaron positivamente de los verdaderos valores simulados (Tabla 4.3). En cambio, el método basado en el muestreo de Gibbs de los coeficientes de regresión de los caracteres causales arrojó valores muy próximos a los verdaderos, con errores de estimación centrados en cero.

Tabla 4.3. Errores de estimación de los coeficientes causales en el escenario con correlaciones genéticas entre los caracteres. Los valores de la tabla corresponden al error de estimación medio, su desvío estándar (DE) y su intervalo de alta densidad del 95%, todos ellos obtenidos sobre las 100 réplicas de la simulación.

Estimador del parámetro	Valor verdadero	Método de estimación			
		CFA		Gibbs	
		Error de estimación $\pm$ DE	HDI95%	Error de estimación $\pm$ DE	HDI95%
$\widehat{\lambda}_{21}$	0,4	0,2002 $\pm$ 0,015	0,174; 0,299	-0,0024 $\pm$ 0,025	-0,057; 0,047
$\widehat{\lambda}_{42}$	-0,3	-0,0997 $\pm$ 0,013	-0,127; -0,076	0,0013 $\pm$ 0,025	-0,046; 0,048
$\widehat{\lambda}_{43}$	0,4	0,1643 $\pm$ 0,015	0,138; 0,194	-0,0068 $\pm$ 0,022	-0,047; 0,030
$\widehat{\lambda}_{54}$	0,3	0,1343 $\pm$ 0,013	0,109; 0,158	-0,0035 $\pm$ 0,023	-0,057; 0,041

Referencias:

$\widehat{\lambda}_{ij}$ estimador del efecto causal ejercido por el carácter j en el carácter i .

CFA: método de estimación basado en un análisis de confirmación de factores.

Gibbs: método de estimación basado en definir las relaciones causales como covariables.

Con respecto a los parámetros de varianza y covarianzas del MTM bajo el escenario con correlaciones genéticas, las estimaciones obtenidas mediante el algoritmo AI-REML estuvieron también dentro de los límites de los correspondientes intervalos de alta densidad del 95% (Anexo 5). Sin embargo, los errores de estimación de los coeficientes causales indujeron un sesgo en las estimaciones de los parámetros correspondientes del modelo SEM.

En contraste, este sesgo no se manifestó en las estimaciones obtenidas mediante el método alternativo (Tabla 4.4).

Tabla 4.4. Errores de estimación de las varianzas y covarianzas genéticas del modelo de ecuaciones estructurales en el escenario con correlaciones genéticas entre los caracteres. Los valores de la tabla corresponden al error de estimación medio, su desvío estándar (DE) y su intervalo de alta densidad del 95%, todos ellos obtenidos sobre las 100 réplicas de la simulación.

Estimador del parámetro	Valor verdadero	Método de estimación			
		Dos etapas		Gibbs	
		Error de estimación $\pm$ DE	HDI95%	Error de estimación $\pm$ DE	HDI95%
$\hat{\sigma}_1^2$	0,5	0,0021 $\pm$ 0,041	-0,068; 0,084	-0,0061 $\pm$ 0,042	-0,072; 0,078
$\hat{\sigma}_2^2$	0,5	-0,0995 $\pm$ 0,037	-0,168; -0,028	0,0037 $\pm$ 0,049	-0,085; 0,094
$\hat{\sigma}_3^2$	0,5	0,0015 $\pm$ 0,044	-0,071; 0,086	0,0133 $\pm$ 0,042	-0,062; 0,099
$\hat{\sigma}_4^2$	0,5	-0,0991 $\pm$ 0,033	-0,164; -0,031	0,0150 $\pm$ 0,047	-0,082; 0,107
$\hat{\sigma}_5^2$	0,5	-0,0646 $\pm$ 0,041	-0,143; 0,016	0,0165 $\pm$ 0,045	-0,067; 0,120
$\hat{\sigma}_{12}$	0,3	-0,1018 $\pm$ 0,024	-0,153; -0,060	0,0067 $\pm$ 0,039	-0,062; 0,083
$\hat{\sigma}_{13}$	0,0	-0,0008 $\pm$ 0,029	-0,061; 0,058	-0,0061 $\pm$ 0,023	-0,052; 0,043
$\hat{\sigma}_{14}$	0,0	0,0468 $\pm$ 0,024	0,007; 0,090	-0,0013 $\pm$ 0,032	-0,061; 0,056
$\hat{\sigma}_{15}$	0,0	0,0186 $\pm$ 0,030	-0,034; 0,074	-0,0030 $\pm$ 0,023	-0,048; 0,038
$\hat{\sigma}_{23}$	0,0	0,0036 $\pm$ 0,026	-0,044; 0,053	0,0001 $\pm$ 0,022	-0,046; 0,041
$\hat{\sigma}_{24}$	-0,2	0,0504 $\pm$ 0,024	0,002; 0,096	-0,0102 $\pm$ 0,046	-0,108; -0,064
$\hat{\sigma}_{25}$	0,0	0,0506 $\pm$ 0,025	0,000; 0,090	-0,0036 $\pm$ 0,031	-0,061; 0,046
$\hat{\sigma}_{34}$	0,25	-0,0868 $\pm$ 0,023	-0,138; -0,046	0,0050 $\pm$ 0,037	-0,063; 0,076
$\hat{\sigma}_{35}$	0,0	-0,0596 $\pm$ 0,027	-0,106; -0,005	0,0049 $\pm$ 0,030	-0,055; 0,055
$\hat{\sigma}_{45}$	0,3	-0,0709 $\pm$ 0,025	-0,124; -0,027	0,0079 $\pm$ 0,049	-0,078; 0,109

Referencias:

Dos etapas: método de estimación basado en combinar las estimaciones de los coeficientes de causalidad con las estimaciones de máxima verosimilitud restringida (REML) de las varianzas y covarianzas del modelo animal multicarácter equivalente.

Gibbs: método de estimación basado en definir un modelo con las relaciones causales como covariables y obtener las estimaciones de sus parámetros mediante un algoritmo del muestreo de Gibbs.

4.3 DISCUSIÓN

La implementación práctica de un SEM requiere de la estimación de los componentes de varianza y covarianza (tal como en un MTM), y además requiere de la estimación de los coeficientes de causalidad (dada una red causal conocida). Aunque para evaluar la respuesta a la selección se ajustaron modelos donde todos estos parámetros se asumieron conocidos, el desarrollo de la simulación permitió también evaluar este problema en forma colateral. En particular, se evaluaron dos métodos alternativos de estimación y la comparación se basó en el error de estimación a través de réplicas independientes de las poblaciones.

Para la estructura de datos obtenida de la simulación diseñada para este trabajo, la mejor estrategia en términos de precisión fue aquella basada en la estimación de los parámetros de varianzas y covarianzas genéticas de un MTM que incluía en su formulación las relaciones causales entre los fenotipos como covariables. El método resultó eficiente en términos de requerimientos computacionales y de tiempos de cómputo. La principal limitación de esta metodología de estimación es que asume independencia entre los términos del error de los caracteres involucrados en la red causal (matriz de covarianzas de los errores con estructura diagonal). Esta situación es comúnmente asumida en muchos de los trabajos de investigación que aplican SEM en la búsqueda de relaciones causales entre caracteres (Texeira Silva et al., 2021; Mora et al., 2022) y se basa en el principio de suficiencia causal, el cual implica la presunción de que una vez que la red causal es conocida, no hay otras causas de asociación ocultas entre dos o más variables (Rosa y col., 2011).

En este trabajo, se ha optado por utilizar una estructura de covarianza diagonal para los términos del error en el proceso de simulación de datos. Esta elección facilita la identificabilidad del modelo de manera sencilla. Sin embargo, si esta estructura no puede

asumirse en los datos debido a la presencia de correlaciones no esperadas entre los errores, la identificabilidad del modelo puede garantizarse mediante la incorporación de variables instrumentales. Estas se definen como covariables que afectan directamente a caracteres ubicados aguas arriba y que no tienen un efecto directo sobre los caracteres situados aguas abajo (Varona y González-Recio, 2023). Un ejemplo podría ser la relación entre la probabilidad de distocia y la probabilidad de nacimiento de una cría muerta, donde el estado corporal de la madre tiene un efecto directo sobre la dificultad al parto dado que se engrasa el canal de parto, pero no afecta de forma directa a la probabilidad de nacimiento de una cría muerta. En estos casos, los parámetros deben obtenerse utilizando algoritmos de muestreo MCMC para el modelo definido (Gianola y Sorensen, 2004; Wu et al., 2007). El muestreo MCMC es preferible porque permite obtener estimaciones más precisas de los parámetros en presencia de estructuras de covarianza complejas, y facilita la convergencia de las estimaciones en modelos con alta dimensionalidad, en lugar de un modelo animal multicarácter que defina la estructura causal en términos de covariables como efectos fijos.

Capítulo 5. Conclusiones.

CAPÍTULO 5

CONCLUSIONES

En un escenario donde un conjunto de caracteres está causalmente pero no genéticamente relacionado, la estimación de valores de cría utilizando modelos de ecuaciones estructurales permite aplicar presión de selección sobre un carácter objetivo sin inducir respuesta indirecta a la selección en caracteres ubicados aguas arriba en la red causal. Este modelo puede ser útil también cuando existen correlaciones genéticas que actúan en forma simultánea a los efectos causales entre fenotipos, en la medida que se conozca la estructura causal y se disponga de suficiente información para estimar con precisión los parámetros necesarios.

En la medida en la que pueda asumirse una estructura de covarianzas diagonal en los errores, el empleo de modelos animales para estimar los parámetros de un modelo SEM declarando la estructura causal en términos de covariables, produce estimaciones con errores pequeños y de poca varianza, sin demandas importantes en capacidad de cómputo.

BIBLIOGRAFÍA

- Bello, N. M., Ferreira, D., Gianola, D. y Rosa, G. J. M. 2018. Conceptual framework for investigating causal effects from observational data in livestock. *Journal of Animal Science*, 96, 4045–4062.
- Bourdon, R. M. 2014. *Understanding animal breeding*. 2da edición. Ed. Pearson.
- De los Campos, G., Gianola, D., Boettcher, P. y Moroni, P. 2006. A structural equation model for describing relationships between somatic cell score and milk yield in dairy goats. *Journal of Animal Science*, 84, 2934–2941.
- Deng, H. W., y Kibota, T. 1995. The importance of the environmental variance-covariance structure in predicting evolutionary responses. *Evolution*, 49(3), 572 – 574.
- Foulley, J. L., y Chevalet, C. 1981. Méthode de prise en compte de la consanguinité dans un modèle simple de simulation de performances. *Genetics Selection Evolution*, 13, 189.
- Gianola, D. y Sorensen, D. 2004. Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. *Genetics*, 167, 1407–424.
- Haavelmo T. 1943. The statistical implications of a system of simultaneous equations. *Econometrica*, 11:1-12
- Hazel, L. N. 1943. The genetic basis for constructing selection indexes. *Genetics*, 28, 476–490.
- Henderson, C. R. 1984. *Applications of linear models in animal breeding*. University of Guelph.
- Henderson, C. R. y Quass, R. L. 1976. Multiple trait evaluation using relatives records. *Journal of Animal Science*, 43, 118–1197.
- Heringstad, B., Wu, X. L., & Gianola, D. 2009. Inferring relationships between health and fertility in Norwegian Red cows using recursive models. *Journal of dairy science*, 92(4), 1778-1784.
- Inoue, K., Valente, B. D., Shoji, N., Honda, T., Oyama, K. y Rosa, G. J. M. 2016. Inferring phenotypic causal structures among meat quality traits and the application of a structural equation model in Japanese Black cattle. *Journal of Animal Science*, 94, 4133–4142.
- Jafaroghli, M., Ghafouri-Kesbi, F., Khorami, S. J., Barazandeh, A. y Mokhtari, M. 2022. Application of structural equation models for genetic evaluation of lifetime reproductive traits and age at first lambing in Moghani sheep. *Small Ruminant Research*, 214, 106761.
- Konig, S., Wu, X. L., Gianola, D., Heringstad, B. y Simianer, H. 2008. Exploration of Relationships between Claw Disorders and Milk Yield in Holstein Cows via Recursive Linear and Threshold Models. *Journal of Dairy Science*, 91(1), 395-406.

- Lopez de Maturana, E., Wu, X. L., Gianola, D., Weigel, K. A. y Rosa, G. J. M. 2009. Exploring biological relationships between calving traits in primiparous cattle with a Bayesian recursive model. *Genetics*, 181, 277–287.
- Lopez de Maturana, E., de los Campos, G., Wu, X. L., Gianola, D., Weigel, K. A. y Rosa, G. J. 2010. Modeling relationships between calving traits: a comparison between standard and recursive mixed models. *Genetics Selection Evolution*, 42, 1-9.
- Misztal, I., Tsuruta, D., Lourenco, D. A. L., Masuda, Y., Aguilar, I., Legarra, A. Vitezica, Z. 2018. Manual for BLUPF90 family programs. University of Georgia.
- Mora, M., David, I., Gilbert, H., Rosa, G. J. M., Sanchez, J. P. y Piles, M. 2022. Analysis of the causal structure of traits involved in sow lactation feed efficiency. *Genetics Selection Evolution*, 54(1), 16–53.
- Mora, M., Velasco-Galilea, M., Sanchez, J. P., Ramayo-Caldas, Y. y Piles, M. 2022. Disectangling the causal relationships between rabbit growth and cecal microbiota through structural equation models. *Genetics, Selection, Evolution*, 54, 81.
- Mrode, R. 2014. Linear models for the prediction of animal breeding values. 3ra edición. CAB International.
- Pearl, J. 2000. Causality: Models, reasoning and inference 1er edición. Cambridge University Press.
- Pearl, J y Mackenzie, D. 2020 The book of why: the new science of cause and effect. Basic Books.
- Pegolo, S., Yu, H., Morota, G., Bisutti, V., Rosa, G. J., Bittante, G. y Cecchinato, A. 2021. Structural equation modeling for unraveling the multivariate genomic architecture of milk proteins in dairy cattle. *Journal of Dairy Science*, 104(5), 5705-5718.
- Peñagaricano, F., Valente, B. D., Steibel, J. P., Bates, R. O., Ernst, C. W., Khatib, H. y Rosa, G. J. M. 2015. Searching for causal networks involving latent variables in complex traits: Application to growth, carcass, and meat quality traits in pigs. *Journal of Animal Science*, 93, 4617–4623.
- R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- Rosa G., Valente B. D., de los Campos, G., Wu, X. L., Gianola, D. Y Silva M. A. 2011. Inferring causal phenotype networks using structural equation models. *Genetics Selection Evolution* 43:6
- Rosseel Y. 2012. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36.
- Shipley, B. 2016. Cause and correlation in biology. 2da edición. Cambridge University Press.

- Spirtes P, Glymour C y Scheines R. 1993. Causation, Prediction and Search. 2da edición. Cambridge, MA: MIT Press.
- Student. 1908. The probable error of a mean. *Biometrika* 6 (1): 1-25.
- Texeira, S. H., Paiva, J. T., Bothelo, M. E., Carrara, E. R., Lopes, P. S., Fonseca, S. F., Veroneze, R., Ferraz, J. B., Pereira Eler, J., Chicaroni, M. E. y Gaya, L. G. 2021. Searching for causal relationships among latent variables concerning performance, carcass, and meat quality in broilers. *Journal of Animal Breeding and Genetics*, 139(2), 181–192.
- Valente, B. D., Rosa, G. J. M., de los Campos, G., Gianola, D. y Silva, M. A. 2010. Searching for recursive causal structures in multivariate quantitative genetics mixed models. *Genetics*, 185, 633–644.
- Valente, B. D., Rosa, G. J. M., Gianola, D., Wu, X. L. y Weigel, K. 2013. Is structural equation modeling advantageous for the genetic improvement of multiple traits. *Genetics*, 194, 561–572.
- Varona, L. y González-Recio, O. 2023. Invited review: Recursive models in animal breeding: Interpretation, limitations, and extensions. *Journal of Dairy Science*, 106, 2198–2212.
- Varona, L., Sorensen, R. y Thompson, R. 2007. Analysis of litter size and average litter weight in pigs using a recursive model. *Genetics*, 177, 1791–1799.
- Verma, T. y Pearl, J. 1990. Equivalence and synthesis of causal models. In *Proceedings of the 6th conference on uncertainty in artificial intelligence* (pp. 220–227). Elsevier.
- Wright, S. 1921. Correlation and causation. *Journal of Agricultural Research*, 20, 557–585.
- Wu, X. L., Heringstad, B., Chang, Y. M., de los Campos, G. y Gianola, D. 2007. Inferring relationships between somatic cell score and Milk yield using simultaneous and recursive models. *Journal of Dairy Science*, 90, 3508–3521.
- Wu, X. L., Heringstad, B., & Gianola, D. 2010. Models for inferring relationships between phenotypes: A review of methodology, identifiability, and applications. *Journal of Animal Breeding and Genetics*, 127, 3–15.
- Wu, X. L., Parker Gaddis, K. L., Buchard, J., Norman, H. D., Nicolazzi, E., Connor, E. E., Cole, J. B., Durr, J. 2021. An alternative interpretation of residual feed intake by phenotypic recursive relationships in dairy cattle. *JDS Communications*, 2, 371-375.

ANEXOS

ANEXO 1. Evolución de los valores de cría reales (simulados) y estimados en las poblaciones simuladas al utilizar como criterio de selección estimaciones de los valores de cría de distintos caracteres genéticamente independientes, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B). Valores promedio de 100 réplicas de simulación.

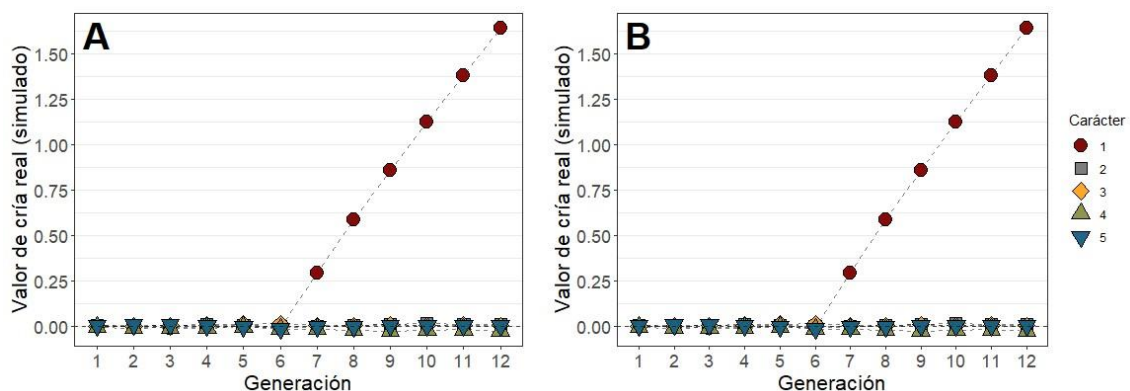


Figura A1.1. Evolución de los valores de cría reales (simulados) al utilizar como criterio de selección estimaciones de los valores de cría del carácter 1, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente independientes. Valores promedio de 100 réplicas de simulación.

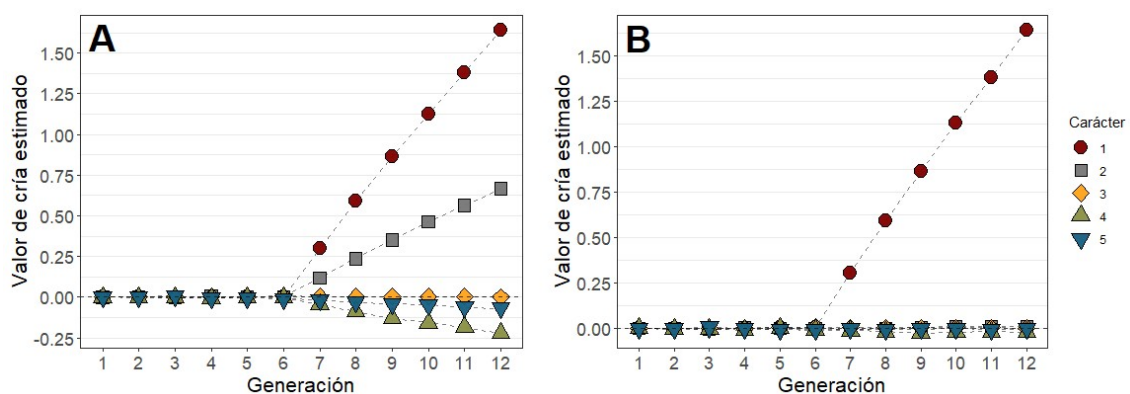


Figura A1.2. Evolución de los valores de cría estimados al utilizar como criterio de selección estimaciones de los valores de cría del carácter 1, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente independientes. Valores promedio de 100 réplicas de simulación.

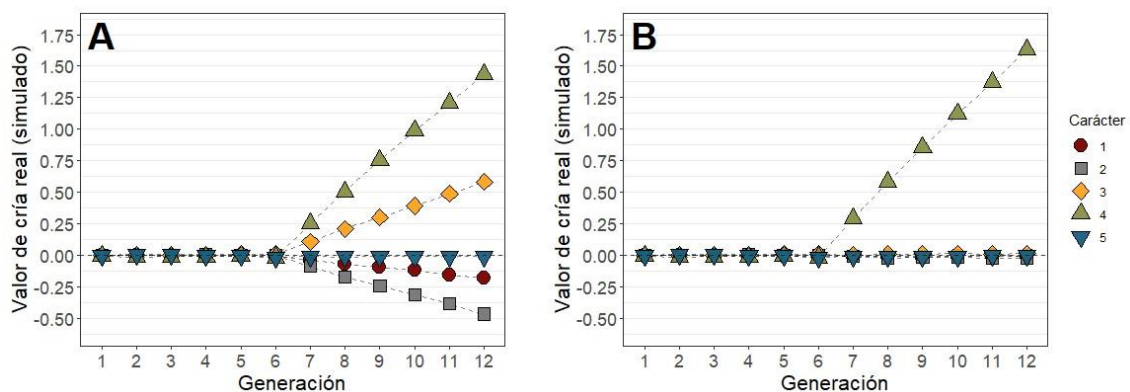


Figura A1.3. Evolución de los valores de cría reales (simulados) al utilizar como criterio de selección estimaciones de los valores de cría del carácter 4, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente independientes. Valores promedio de 100 réplicas de simulación.

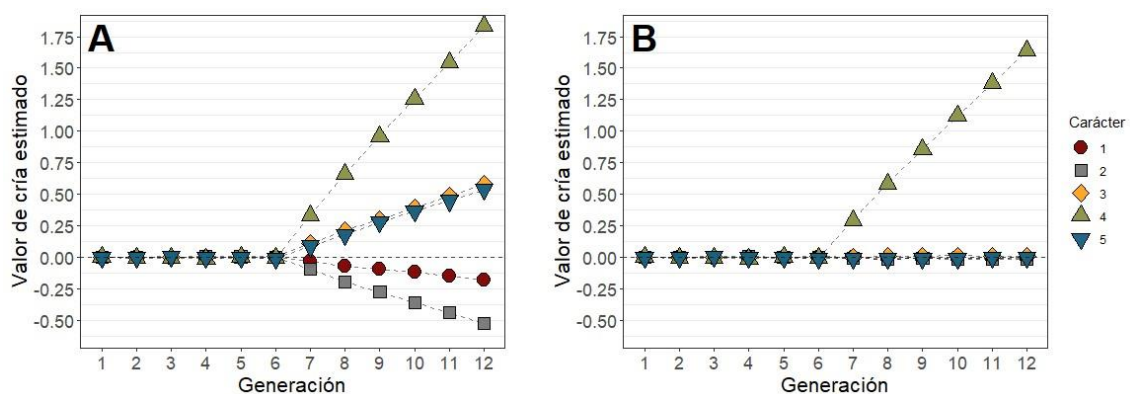


Figura A1.4. Evolución de los valores de cría estimados al utilizar como criterio de selección estimaciones de los valores de cría del carácter 4, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente independientes. Valores promedio de 100 réplicas de simulación.

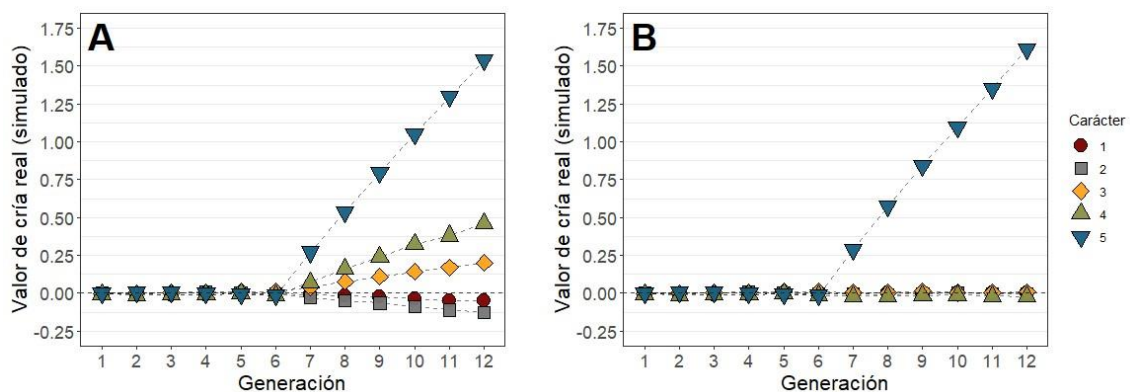


Figura A1.5. Evolución de los valores de cría reales (simulados) al utilizar como criterio de selección estimaciones de los valores de cría del carácter 5, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente independientes. Valores promedio de 100 réplicas de simulación.

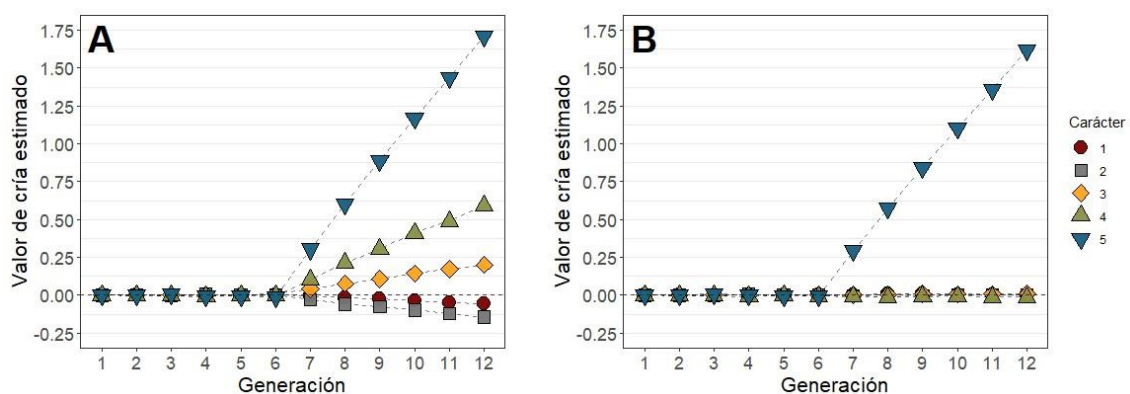


Figura A1.6. Evolución de los valores de cría estimados al utilizar como criterio de selección estimaciones de los valores de cría del carácter 5, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente independientes. Valores promedio de 100 réplicas de simulación.

ANEXO 2. Evolución de los valores de cría reales (simulados) y estimados en las poblaciones simuladas al utilizar como criterio de selección estimaciones de los valores de cría de distintos caracteres genéticamente correlacionados, obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**). Valores promedio de 100 réplicas de simulación.

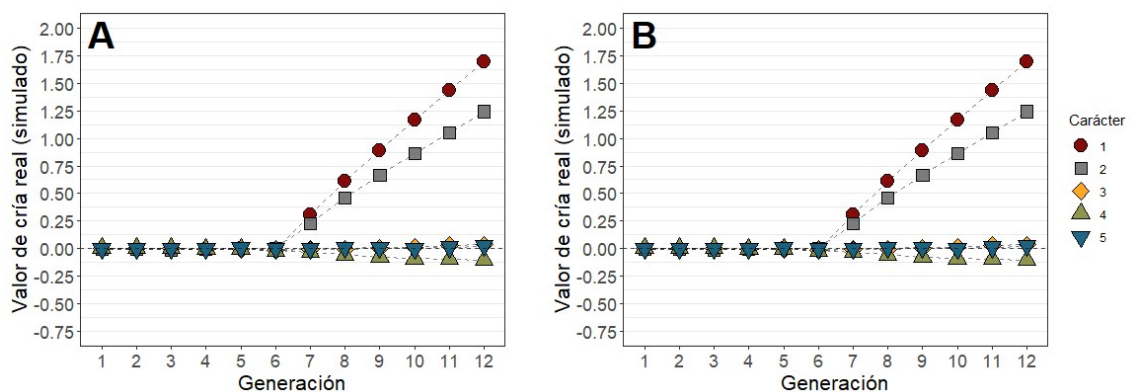


Figura A2.1. Evolución de los valores de cría reales (simulados) al utilizar como criterio de selección estimaciones de los valores de cría del carácter 1, obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) con caracteres genéticamente correlacionados. Valores promedio de 100 réplicas de simulación.

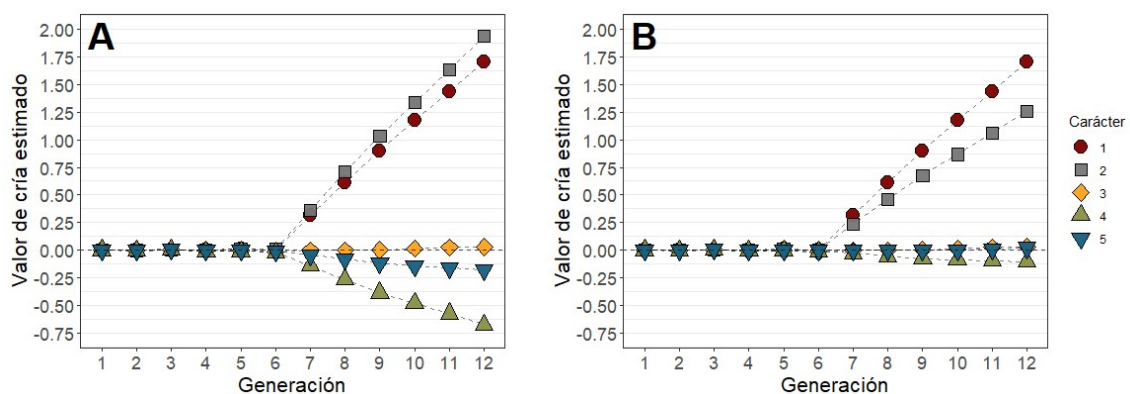


Figura A2.2. Evolución de los valores de cría estimados al utilizar como criterio de selección estimaciones de los valores de cría del carácter 1, obtenidas mediante un modelo animal multicarácter (**A**) o mediante un modelo de ecuaciones estructurales (**B**) con caracteres genéticamente correlacionados. Valores promedio de 100 réplicas de simulación.

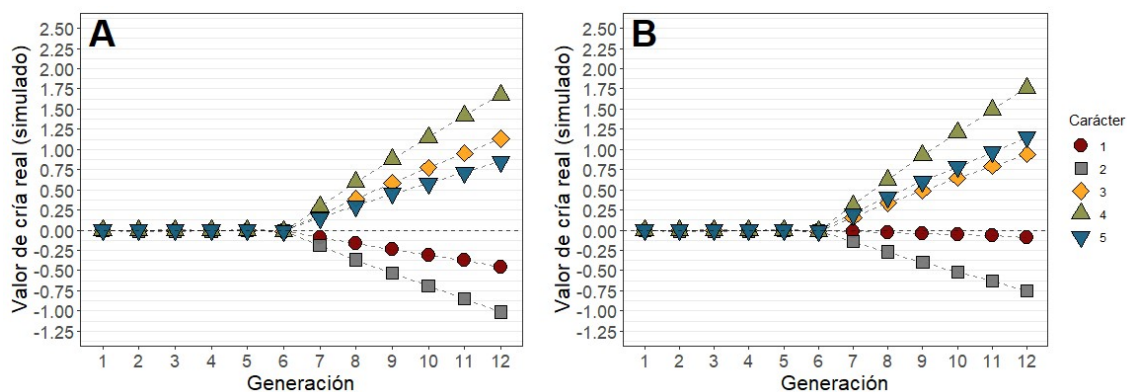


Figura A2.3. Evolución de los valores de cría reales (simulados) al utilizar como criterio de selección estimaciones de los valores de cría del carácter 4, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente correlacionados. Valores promedio de 100 réplicas de simulación.

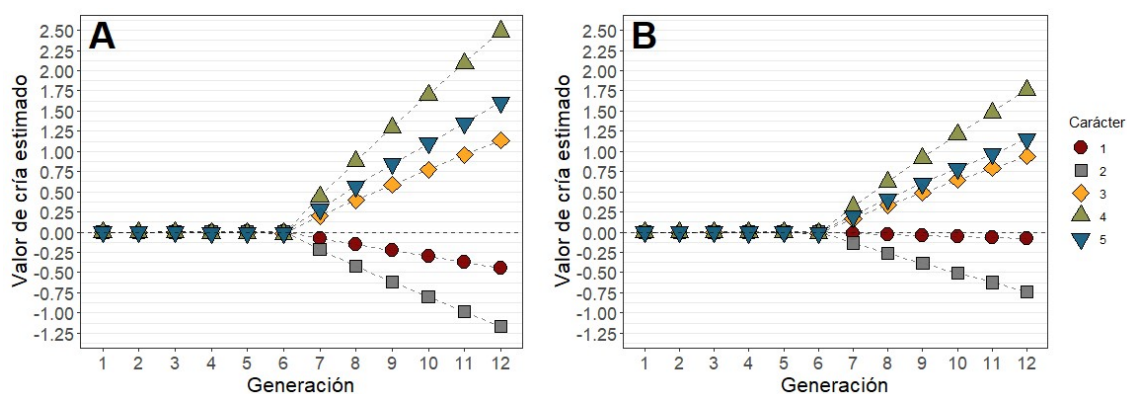


Figura A2.4. Evolución de los valores de cría estimados al utilizar como criterio de selección estimaciones de los valores de cría del carácter 4, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente correlacionados. Valores promedio de 100 réplicas de simulación.

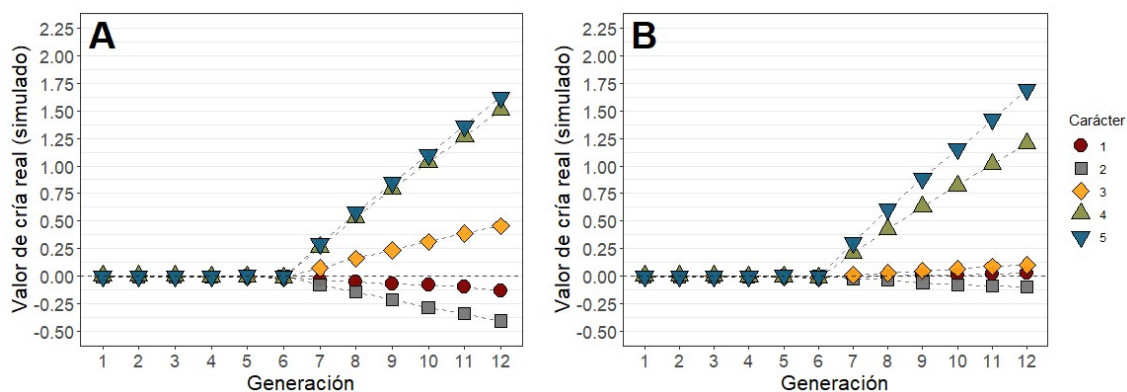


Figura A2.5. Evolución de los valores de cría reales (simulados) al utilizar como criterio de selección estimaciones de los valores de cría del carácter 5, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente correlacionados. Valores promedio de 100 réplicas de simulación.

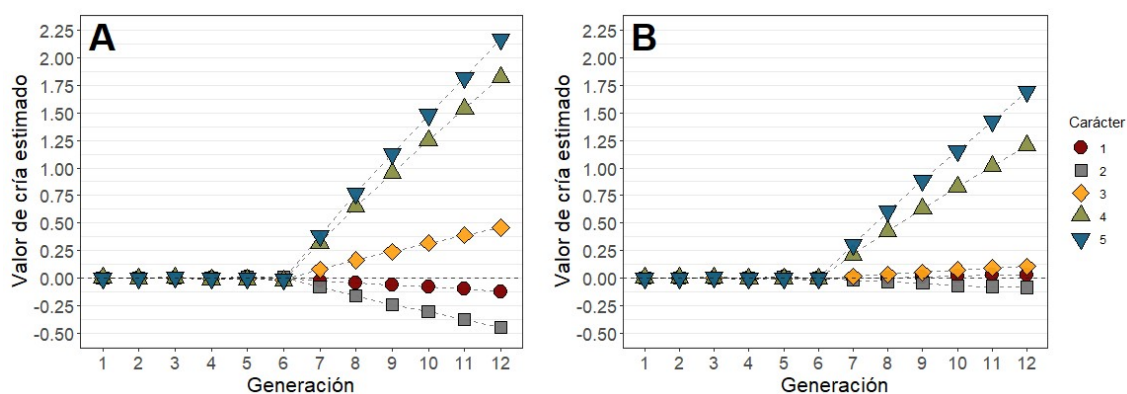


Figura A2.6. Evolución de los valores de cría estimados al utilizar como criterio de selección estimaciones de los valores de cría del carácter 5, obtenidas mediante un modelo animal multicarácter (A) o mediante un modelo de ecuaciones estructurales (B) con caracteres genéticamente correlacionados. Valores promedio de 100 réplicas de simulación.

ANEXO 3. Distribución de las estimaciones de los parámetros del modelo de ecuaciones estructurales obtenidas mediante el método en dos etapas e implementado en poblaciones simuladas en ausencia de correlaciones genéticas entre caracteres de la red causal.

Brevemente, el método se basa en estimar los coeficientes de causalidad mediante un análisis de confirmación de factores y, en paralelo, estimar los componentes de co-varianzas genéticas y del error mediante un algoritmo AIREML bajo el modelo animal multicarácter equivalente. Luego, con los valores obtenidos, se computan los correspondientes parámetros del SEM resolviendo:

$$\Sigma_0 = (I - \Lambda) \Sigma_0^* (I - \Lambda)^t \text{ y } R_0 = (I - \Lambda) R_0^* (I - \Lambda)^t$$

En los gráficos, las curvas representan la función de densidad de probabilidad del estimador del parámetro correspondiente, ajustadas a partir de los valores de las 100 réplicas de la simulación. Por su parte, las líneas punteadas indican el valor verdadero del parámetro objeto de estimación. El espacio sin sombreado acumula el 95%, entre los cuantiles 0,025 y 0,975, de la probabilidad del valor estimado del parámetro.

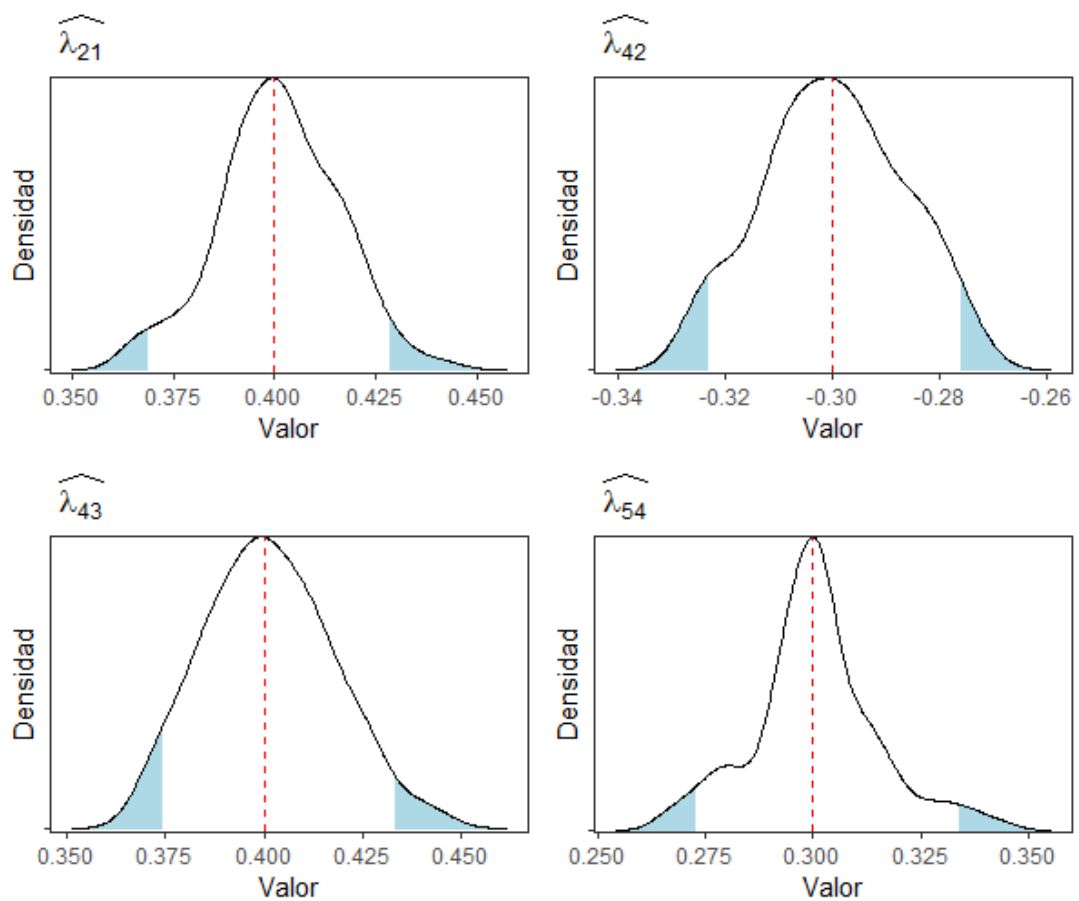


Figura A3.1. Distribución empírica de los parámetros de causalidad obtenidos mediante análisis de confirmación de factores (CFA) en ausencia de correlaciones genéticas

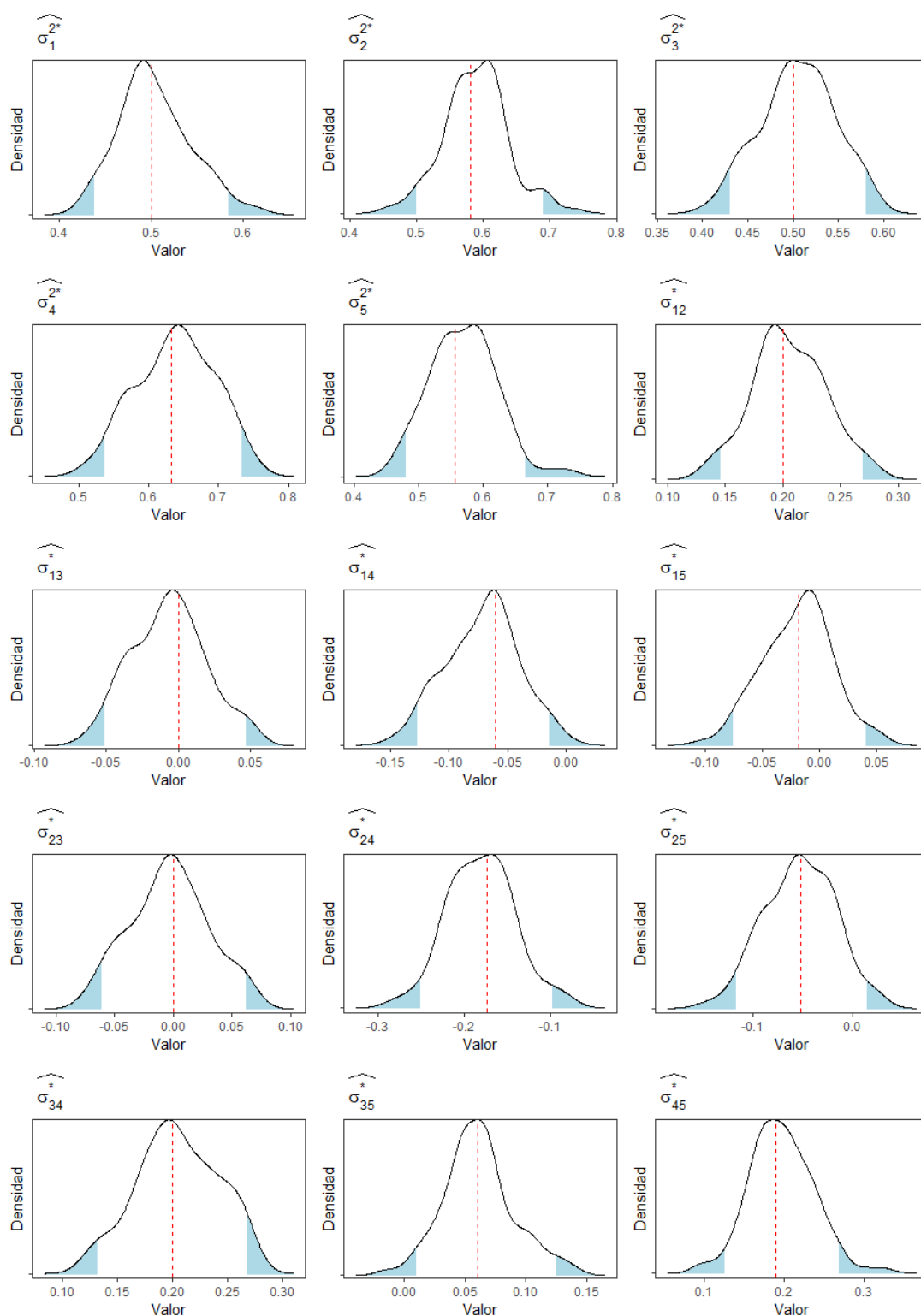


Figura A3.2. Distribución empírica de las varianzas y covarianzas genéticas bajo un modelo multicarácter equivalente estimadas mediante un algoritmo AI-REML en ausencia de correlaciones genéticas.

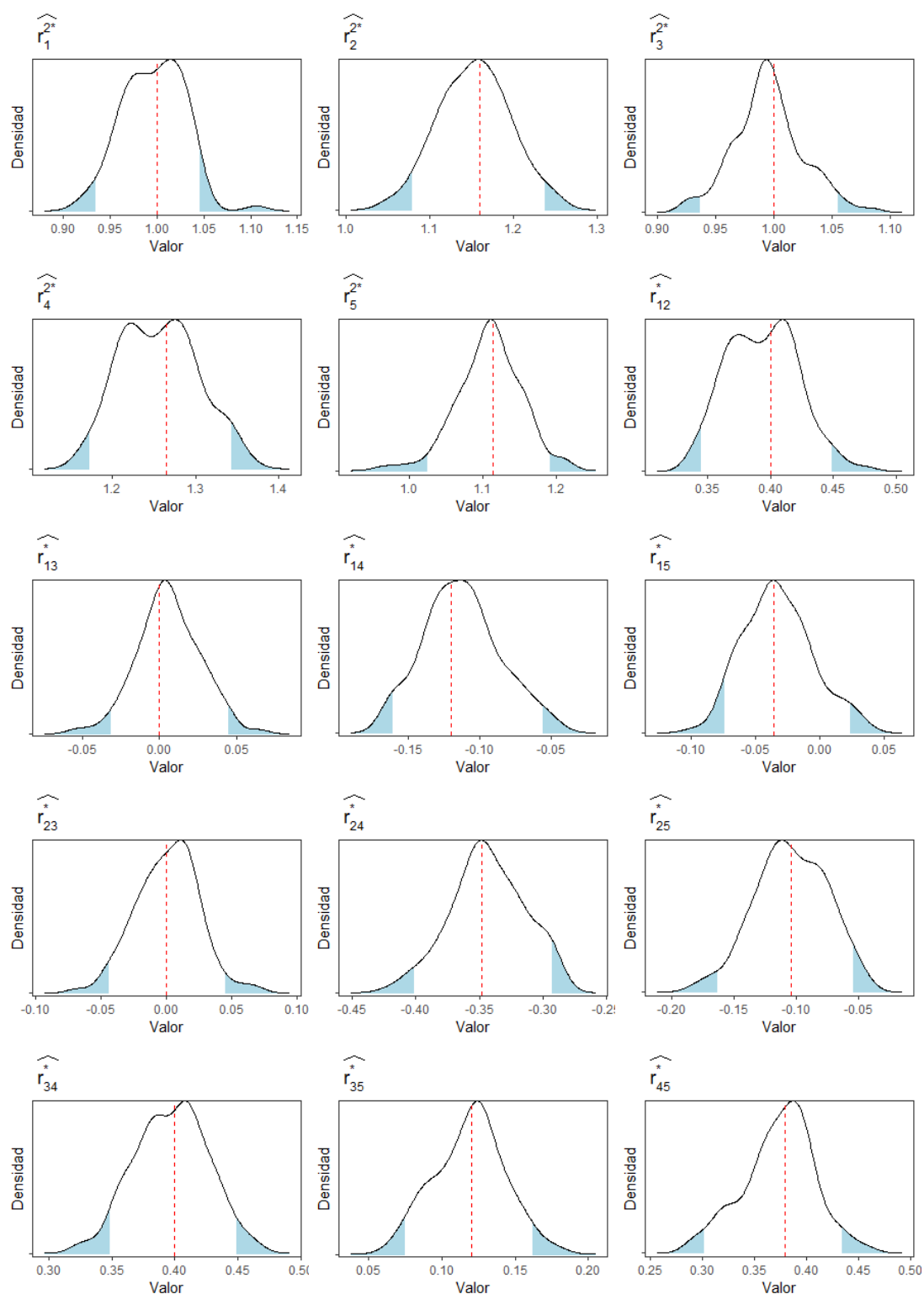


Figura A3.3. Distribución empírica de las varianzas y covarianzas del error bajo un modelo multicarácter equivalente estimadas mediante un algoritmo AI-REML en ausencia de correlaciones genéticas.

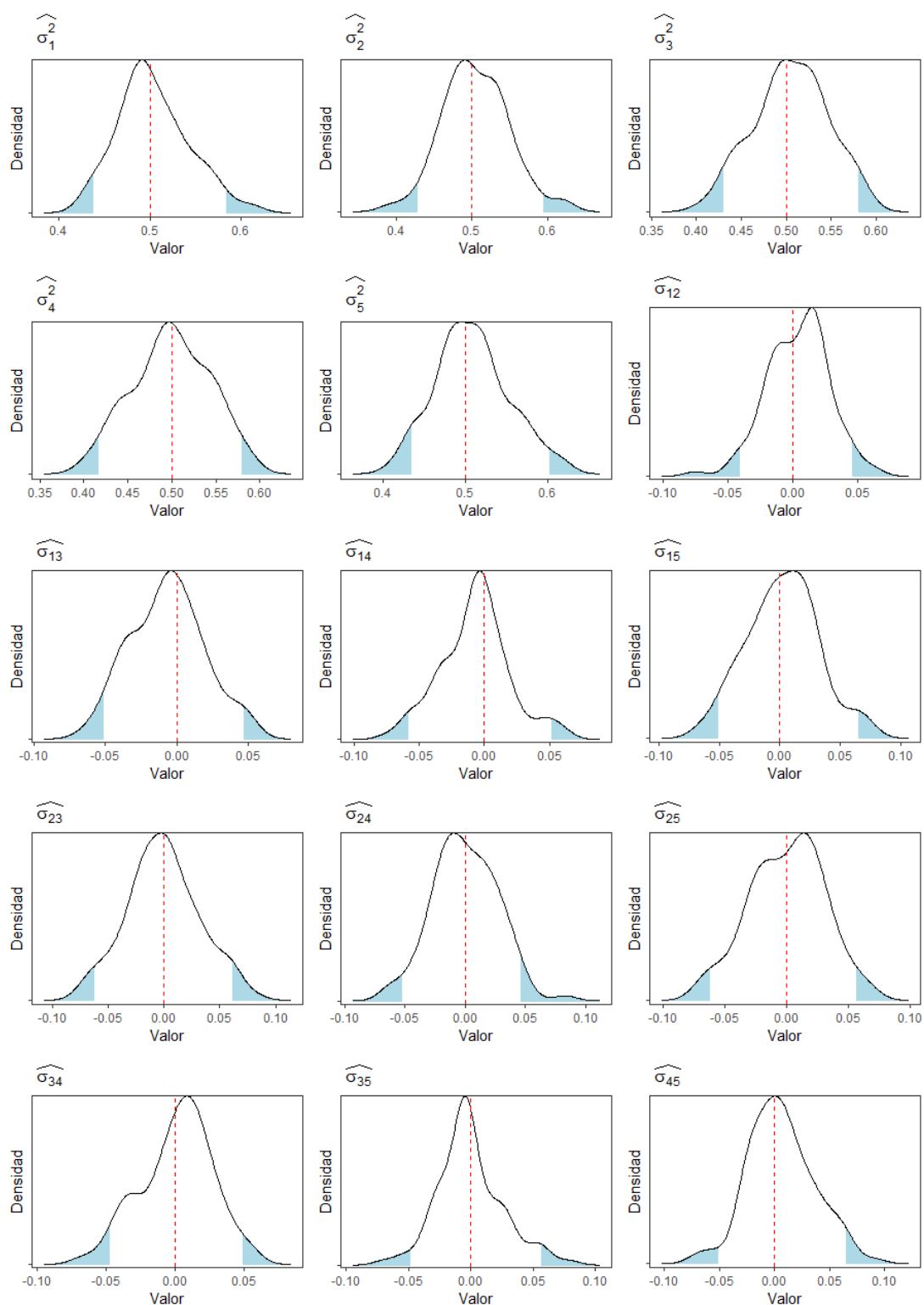


Figura A3.4. Distribución empírica de las varianzas y covarianzas genéticas bajo un modelo de ecuaciones estructurales obtenidas operacionalmente en ausencia de correlaciones genéticas.

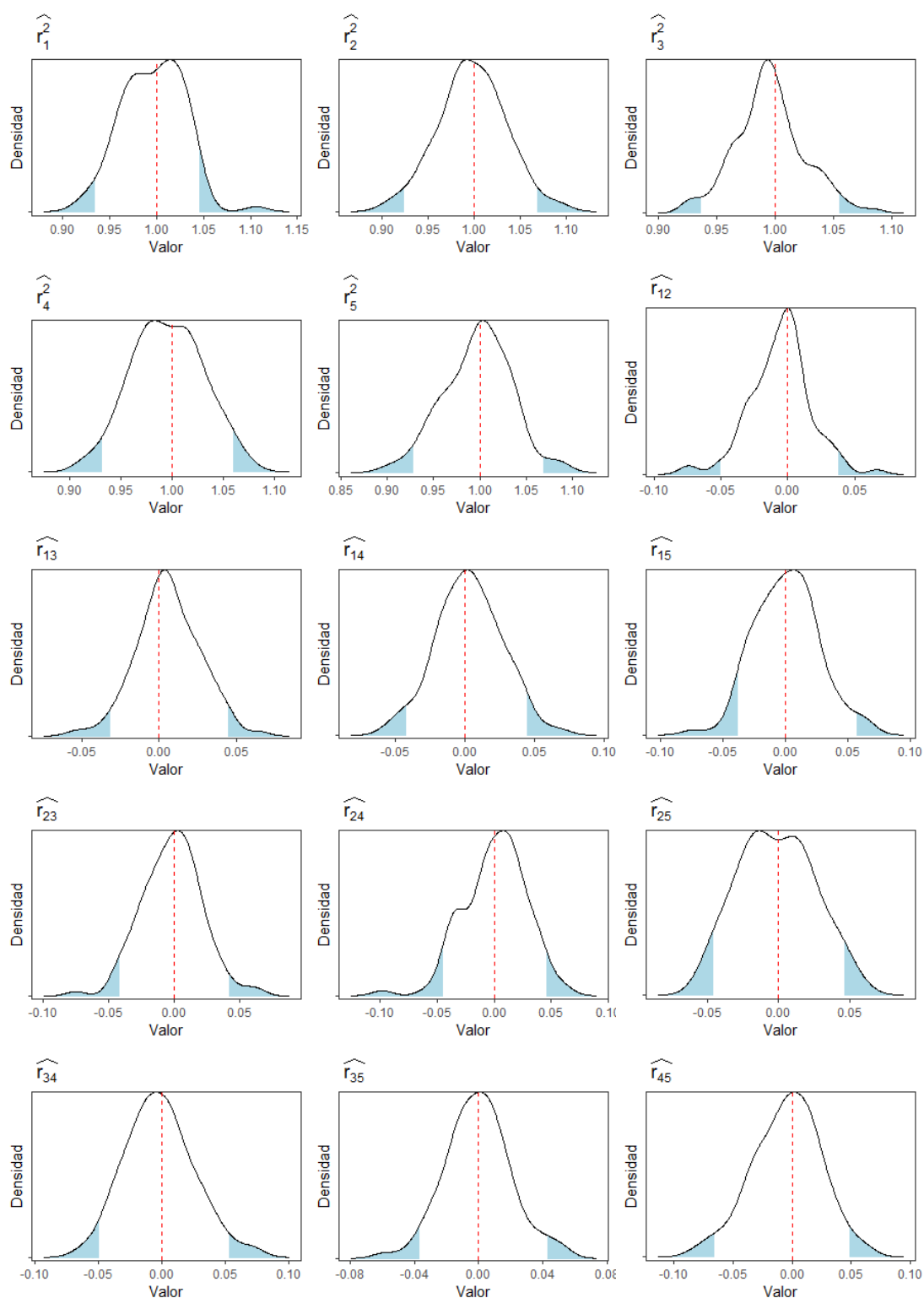


Figura A3.5. Distribución empírica de las varianzas y covarianzas del error bajo un modelo de ecuaciones estructurales obtenidas operacionalmente en ausencia de correlaciones genéticas.

ANEXO 4. Distribución de las estimaciones de los parámetros del modelo de ecuaciones estructurales obtenidas mediante muestreo de Gibbs sobre un modelo animal ajustando la estructura causal como covariables implementado en poblaciones simuladas en ausencia de correlaciones genéticas entre caracteres de la red causal.

Brevemente el modelo consiste en ajustar un modelo animal donde las observaciones están definidas por:

$$\mathbf{y}_i = \mathbf{\Lambda} \mathbf{y}_i + \mathbf{a}_i + \mathbf{e}_i$$

donde $\mathbf{\Lambda} \mathbf{y}_i$ define como covariables a los ancestros causales de cada variable, las varianzas y covarianzas de los efectos genéticos y del error se definen como:

$$\mathbf{\Sigma}_0 = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ & & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ & Sym. & & \sigma_4^2 & \sigma_{45} \\ & & & & \sigma_5^2 \end{bmatrix} \text{ y } \mathbf{R}_0 = \begin{bmatrix} r_1^2 & 0 & \dots & 0 & 0 \\ & r_2^2 & & & 0 \\ & & r_3^2 & & \vdots \\ & Sym. & & r_4^2 & 0 \\ & & & & r_5^2 \end{bmatrix}$$

respectivamente.

En los gráficos, las curvas representan la función de densidad de probabilidad del estimador del parámetro correspondiente, ajustadas a partir de los valores de las 100 réplicas de la simulación. Por su parte, las líneas punteadas indican el valor verdadero del parámetro objeto de estimación. El espacio sin sombreado acumula el 95%, entre los cuantiles 0,025 y 0,975, de la probabilidad del valor estimado del parámetro.

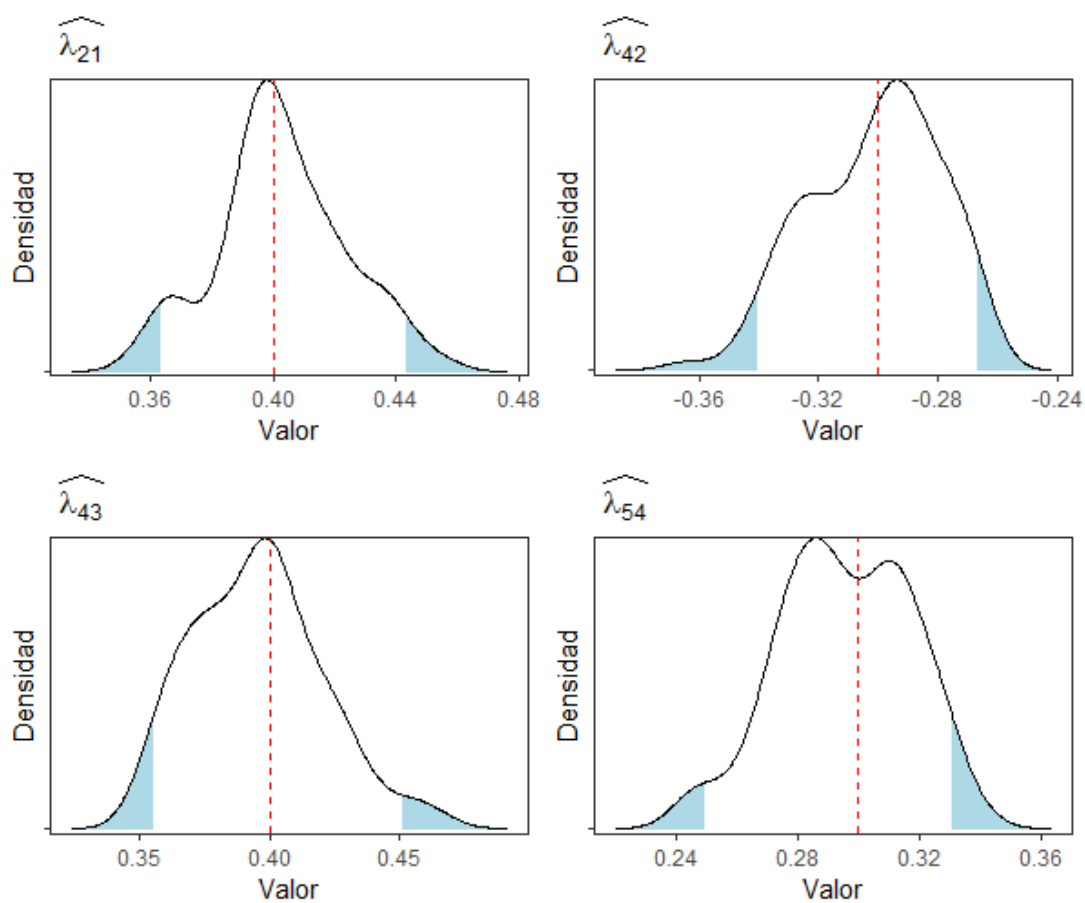


Figura A4.1. Distribución empírica de los parámetros de causalidad obtenidos estimaciones del valor de covariables utilizando muestreo de Gibbs sobre un modelo animal multicarácter en ausencia de correlaciones genéticas.

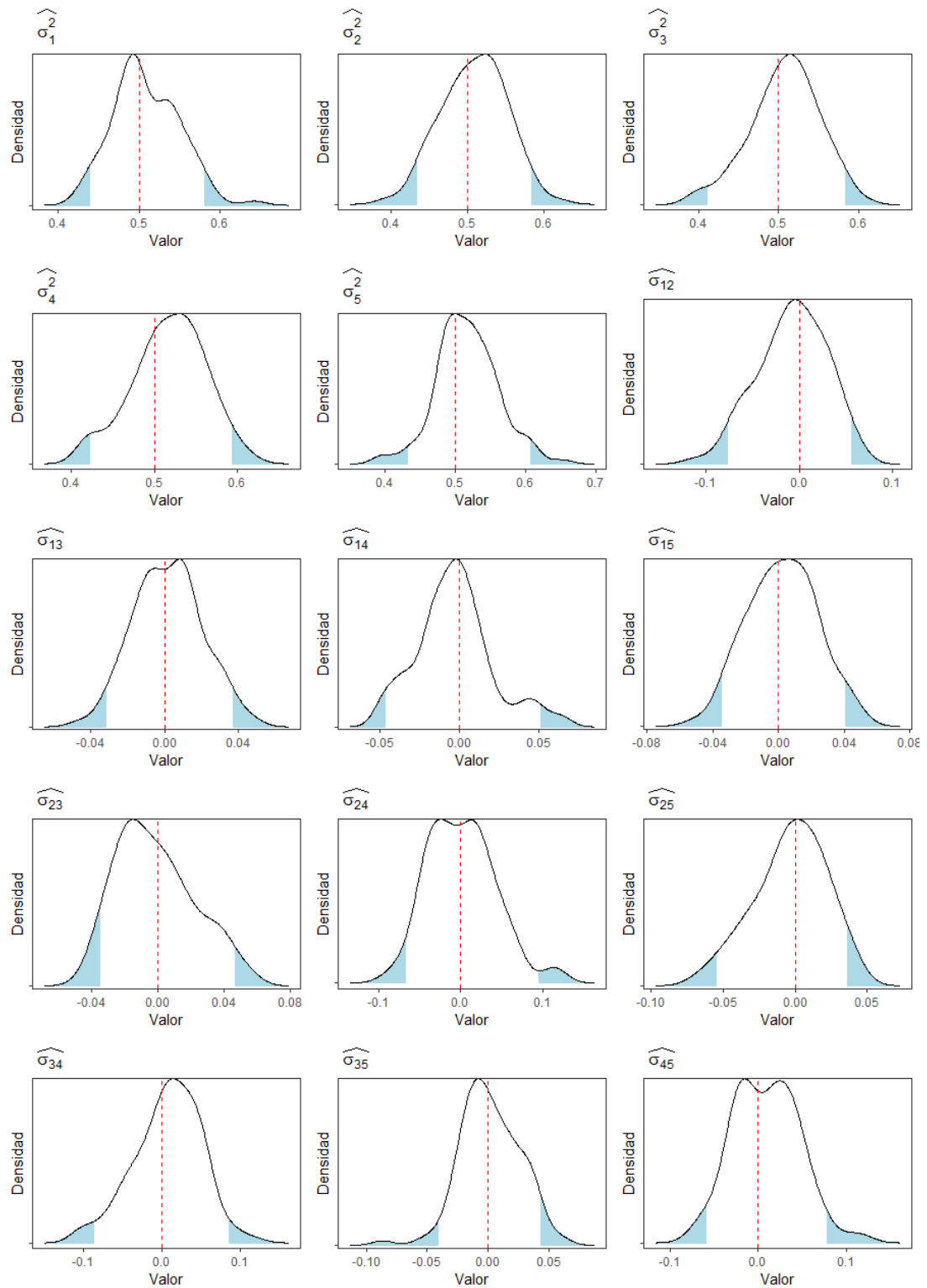


Figura A4.2. Distribución empírica de las varianzas y covarianzas genéticas obtenidas utilizando muestreo de Gibbs sobre un modelo animal multicarácter declarando las relaciones causales como covariables en ausencia de correlaciones genéticas.

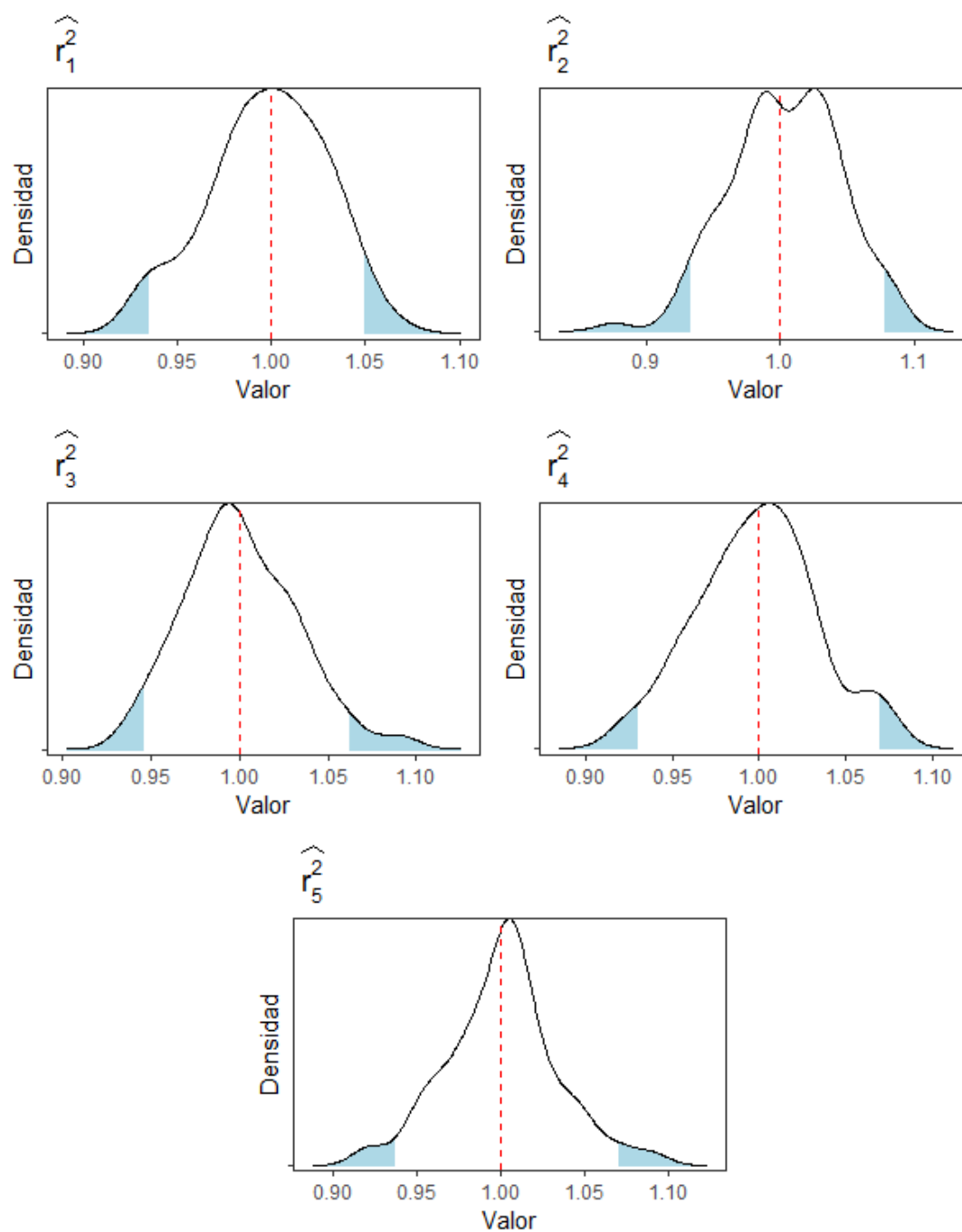


Figura A4.3. Distribución empírica de las varianzas y covarianzas del error obtenidas utilizando muestreo de Gibbs sobre un modelo animal multicarácter declarando las relaciones causales como covariables en ausencia de correlaciones genéticas.

ANEXO 5. Distribución de las estimaciones de los parámetros del modelo de ecuaciones estructurales obtenidas mediante el método en dos etapas e implementado en poblaciones simuladas con caracteres genéticamente correlacionados en la red causal.

Brevemente, el método se basa en estimar los coeficientes de causalidad mediante un análisis de confirmación de factores y, en paralelo, estimar los componentes de co-varianzas genéticas y del error mediante un algoritmo AIREML bajo el modelo animal multicarácter equivalente. Luego, con los valores obtenidos, se computan los correspondientes parámetros del SEM resolviendo:

$$\Sigma_0 = (I - \Lambda) \Sigma_0^* (I - \Lambda)^t \text{ y } R_0 = (I - \Lambda) R_0^* (I - \Lambda)^t$$

En los gráficos, las curvas representan la función de densidad de probabilidad del estimador del parámetro correspondiente, ajustadas a partir de los valores de las 100 réplicas de la simulación. Por su parte, las líneas punteadas indican el valor verdadero del parámetro objeto de estimación. El espacio sin sombreado acumula el 95%, entre los cuantiles 0,025 y 0,975, de la probabilidad del valor estimado del parámetro.

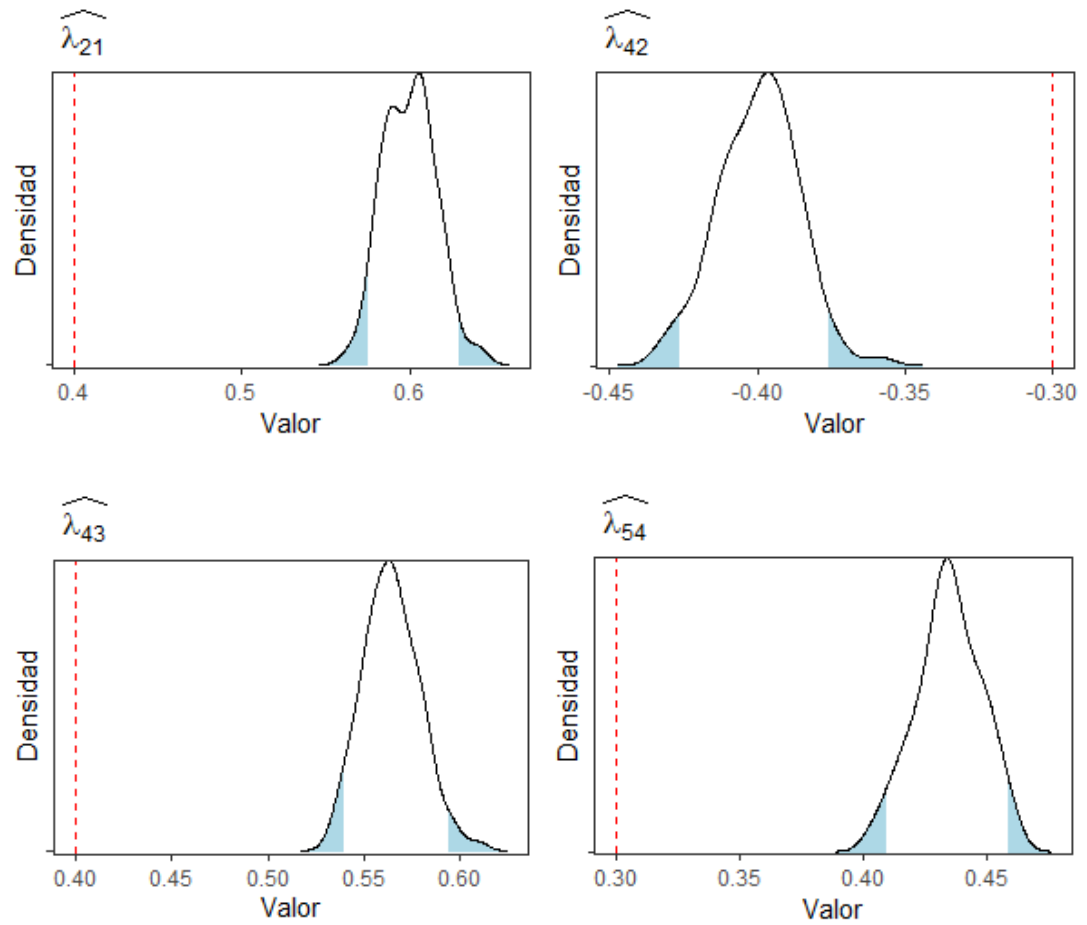


Figura A5.1. Distribución empírica de los parámetros de causalidad obtenidos mediante análisis de confirmación de factores (CFA) caracteres correlacionados genéticamente.

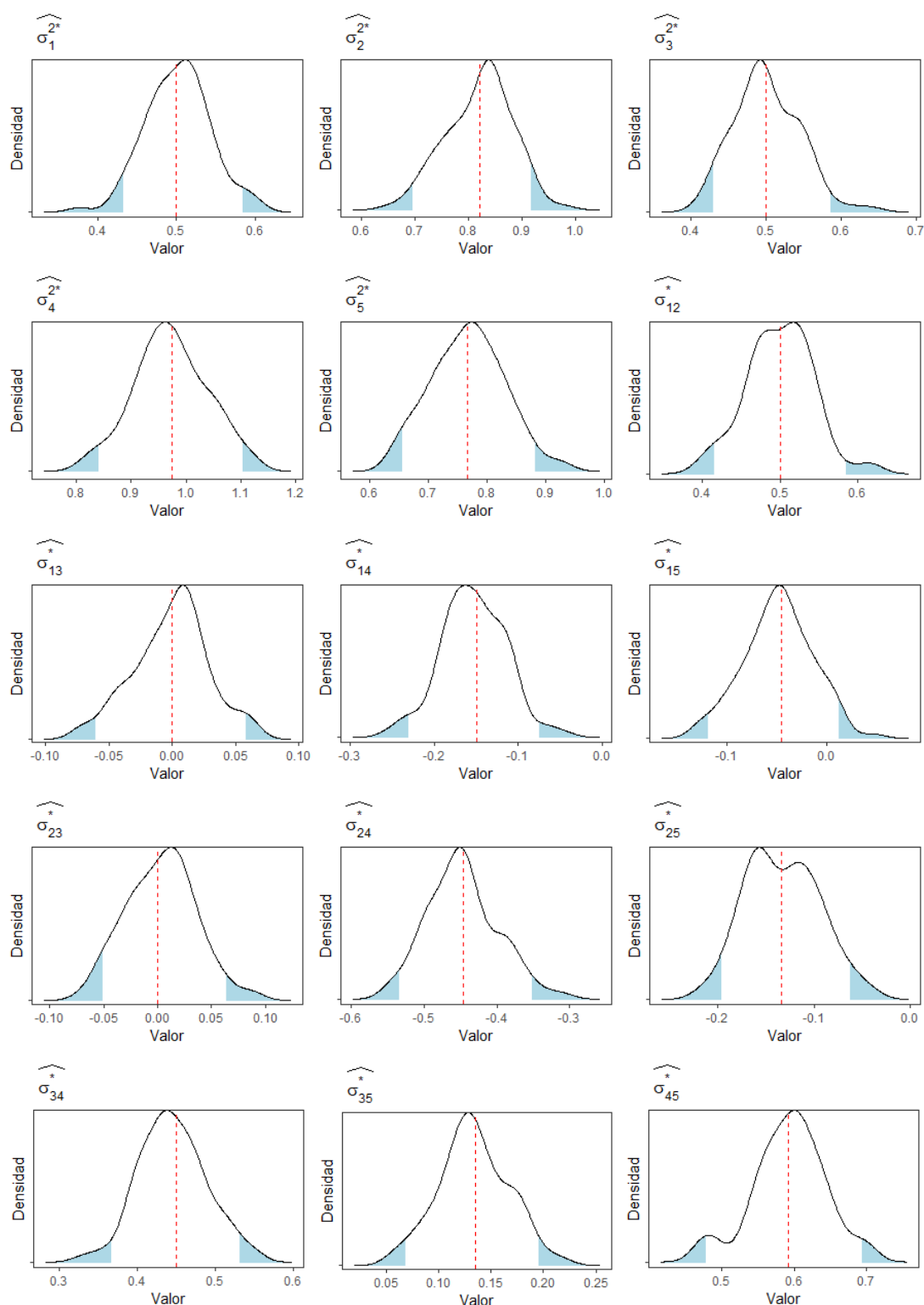


Figura A5.2. Distribución empírica de las varianzas y covarianzas genéticas bajo un modelo multicarácter equivalente estimadas mediante un algoritmo AI-REML caracteres correlacionados genéticamente.

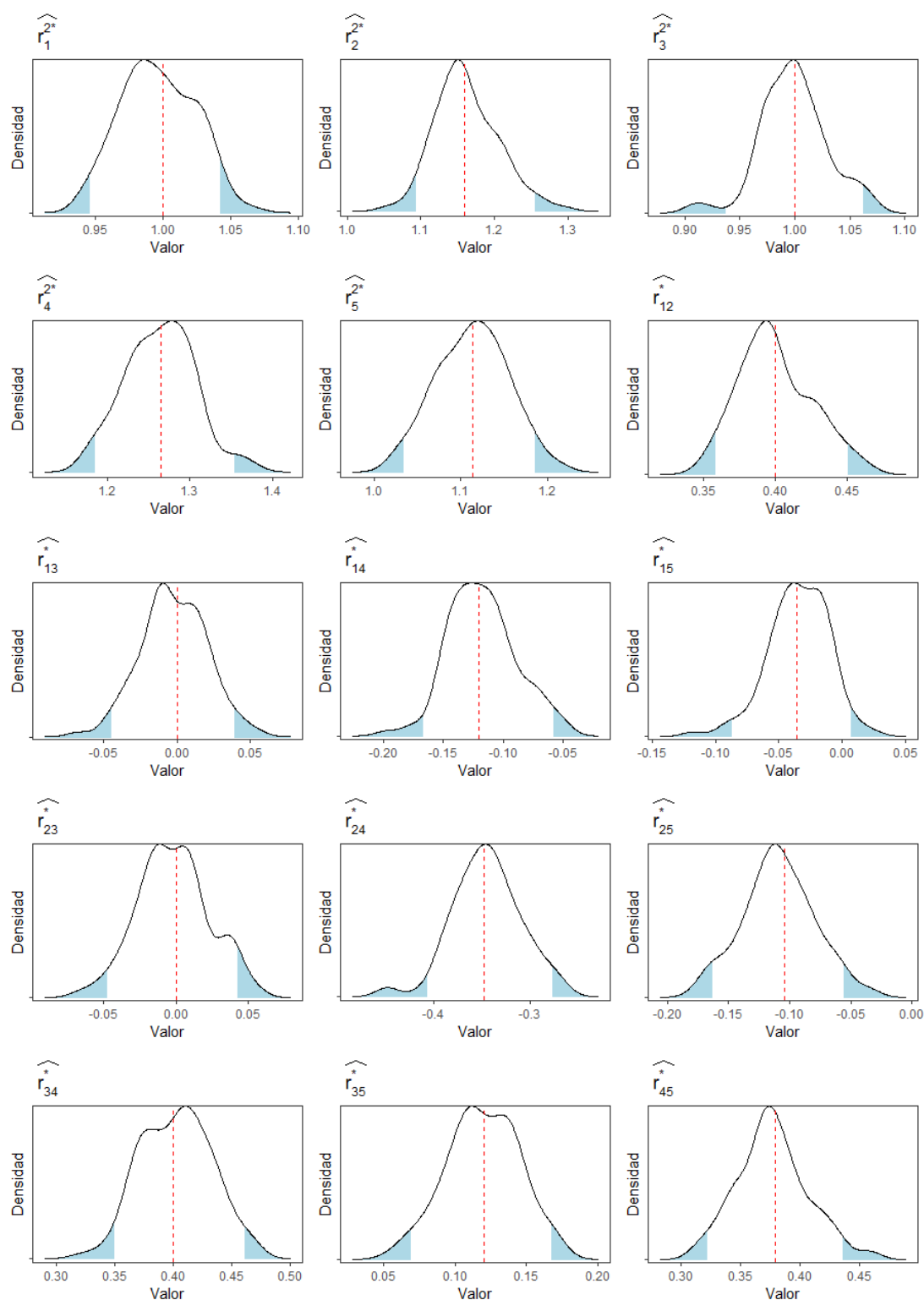


Figura A5.3. Distribución empírica de las varianzas y covarianzas del error bajo un modelo multicarácter equivalente estimadas mediante un algoritmo AI-REML caracteres correlacionados genéticamente.

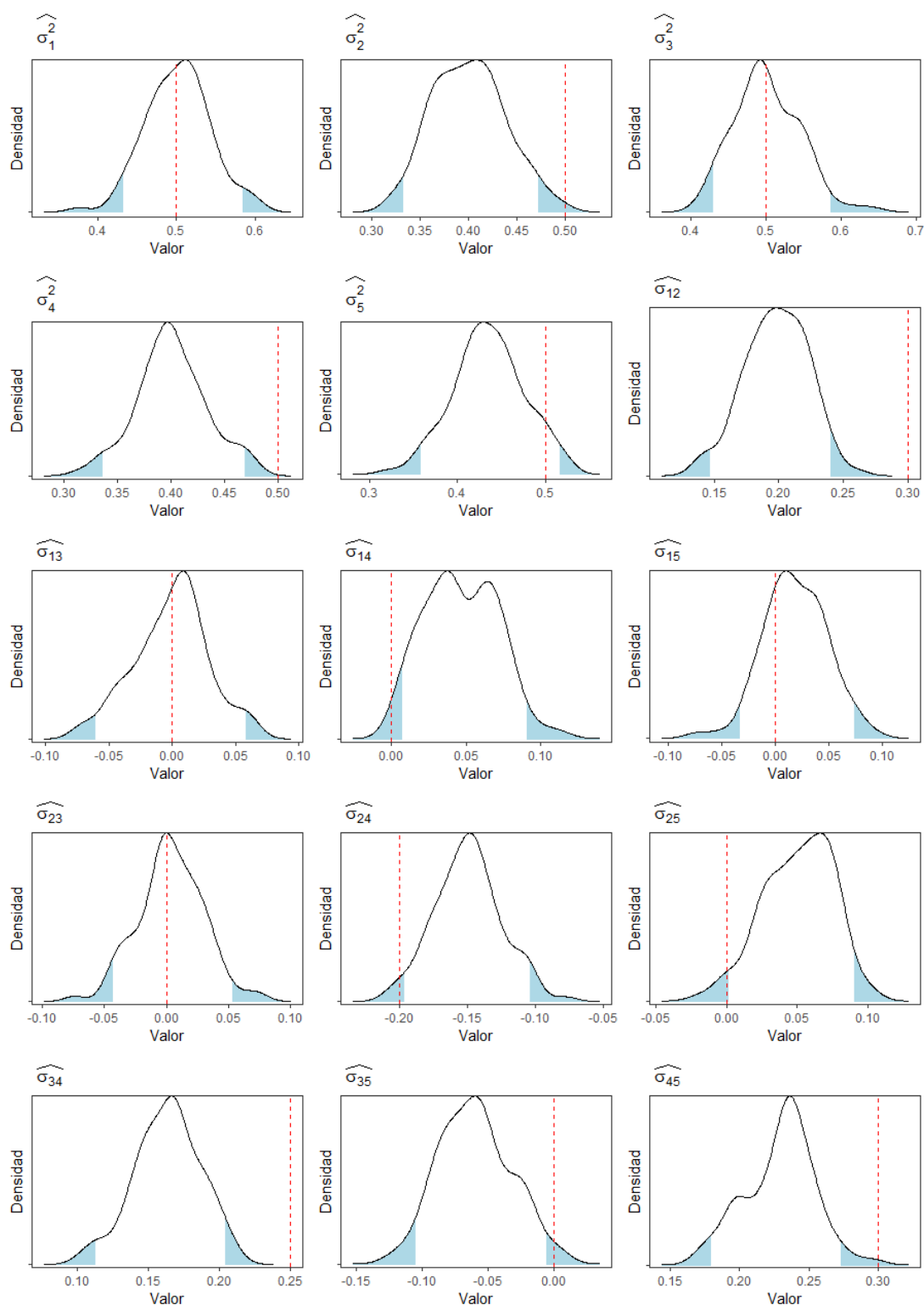


Figura A5.4. Distribución empírica de las varianzas y covarianzas genéticas obtenidas operacionalmente en caracteres correlacionados genéticamente.

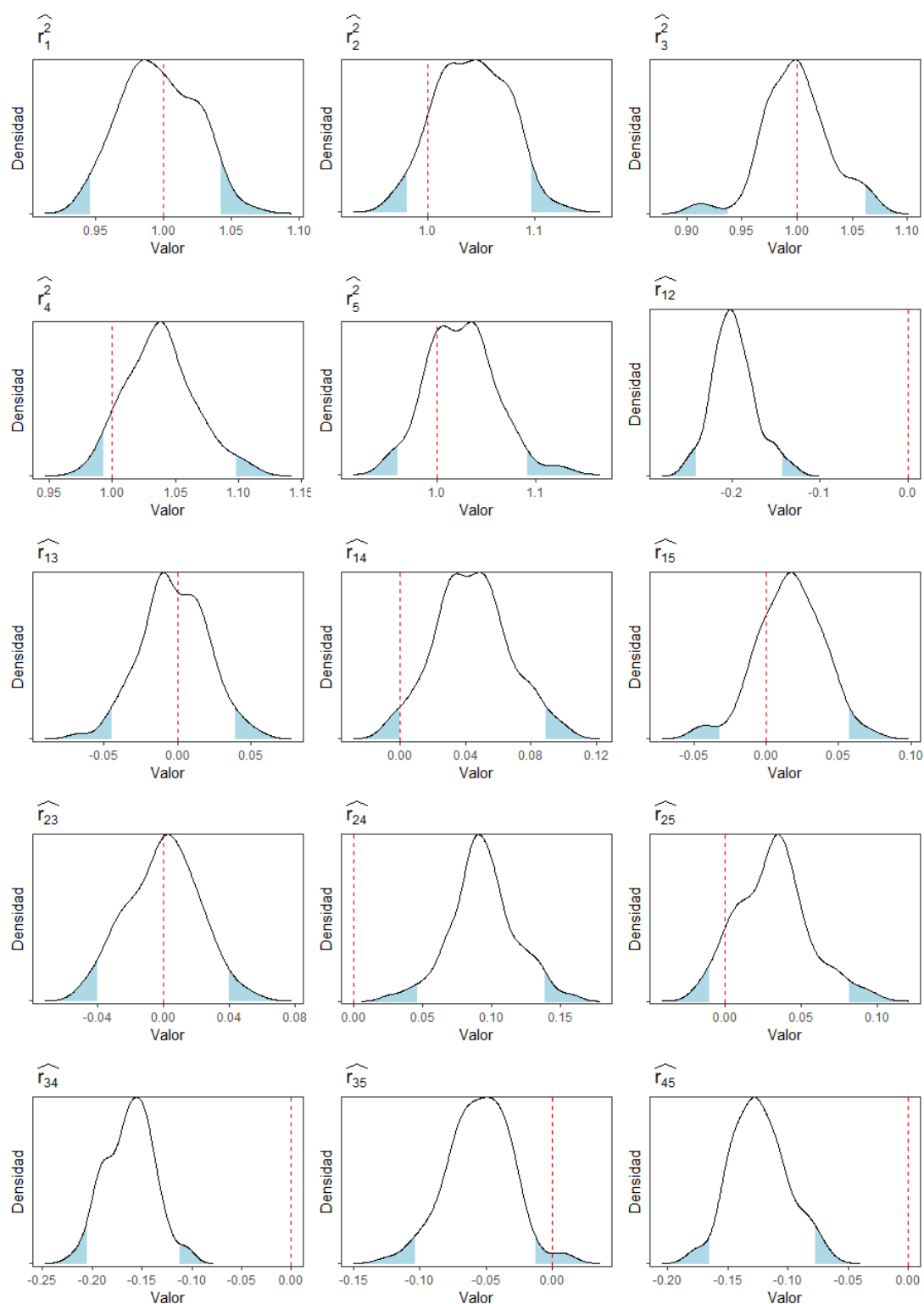


Figura A5.5. Distribución empírica de las varianzas y covarianzas del error obtenidas operacionalmente en caracteres correlacionados genéticamente.

ANEXO 6. Distribución de las estimaciones de los parámetros del modelo de ecuaciones estructurales obtenidas mediante muestreo de Gibbs sobre un modelo animal ajustando la estructura causal como covariables implementado en poblaciones simuladas con caracteres genéticamente correlacionados.

Brevemente el modelo consiste en ajustar un modelo animal donde las observaciones están definidas por:

$$\mathbf{y}_i = \mathbf{\Lambda} \mathbf{y}_i + \mathbf{a}_i + \mathbf{e}_i$$

donde $\mathbf{\Lambda} \mathbf{y}_i$ define como covariables a los ancestros causales de cada variable, las varianzas y covarianzas de los efectos genéticos y del error se definen como:

$$\mathbf{\Sigma}_0 = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ & & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ & & & \sigma_4^2 & \sigma_{45} \\ & & & & \sigma_5^2 \end{bmatrix} \text{ y } \mathbf{R}_0 = \begin{bmatrix} r_1^2 & 0 & \cdots & 0 & 0 \\ & r_2^2 & & & 0 \\ & & r_3^2 & & \vdots \\ & & & r_4^2 & 0 \\ & & & & r_5^2 \end{bmatrix}$$

respectivamente.

En los gráficos, las curvas representan la función de densidad de probabilidad del estimador del parámetro correspondiente, ajustadas a partir de los valores de las 100 réplicas de la simulación. Por su parte, las líneas punteadas indican el valor verdadero del parámetro objeto de estimación. El espacio sin sombreado acumula el 95%, entre los cuantiles 0,025 y 0,975, de la probabilidad del valor estimado del parámetro.

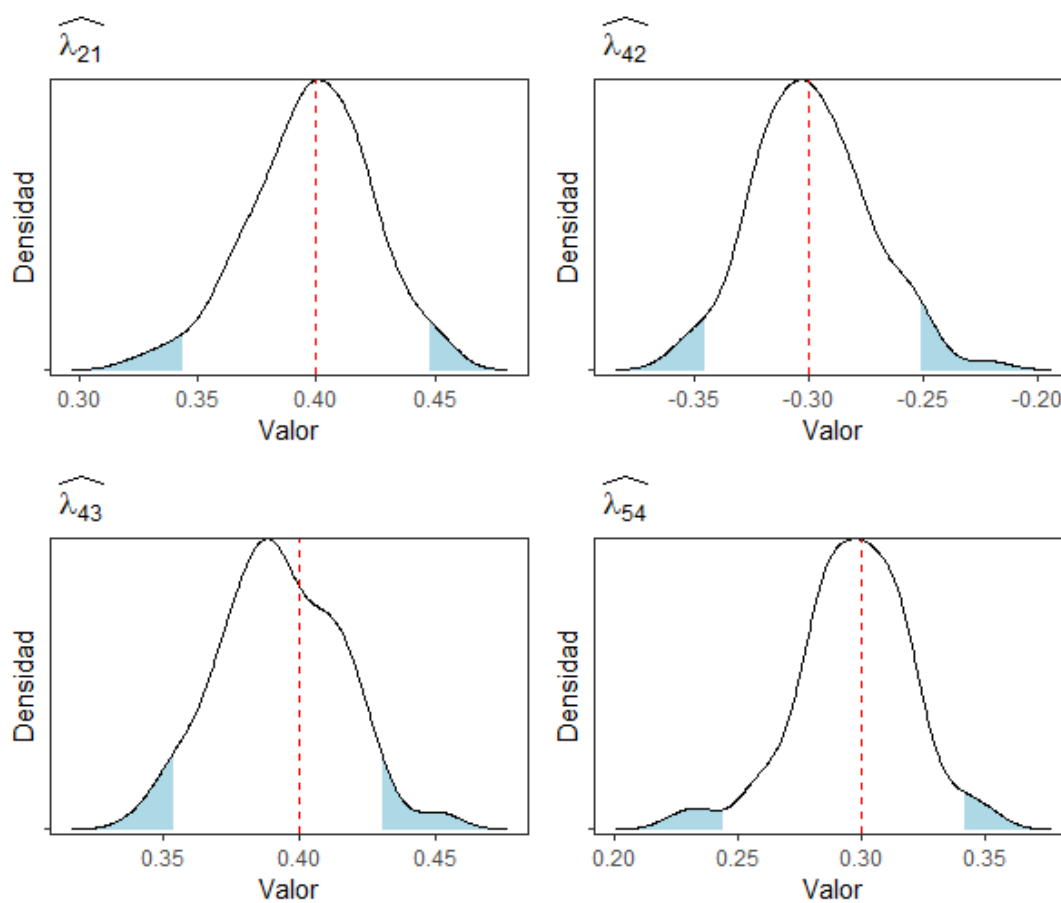


Figura A6.1. Distribución empírica de los parámetros de causalidad obtenidos estimaciones del valor de covariables utilizando muestreo de Gibbs sobre un modelo animal multicarácter con caracteres correlacionados genéticamente.

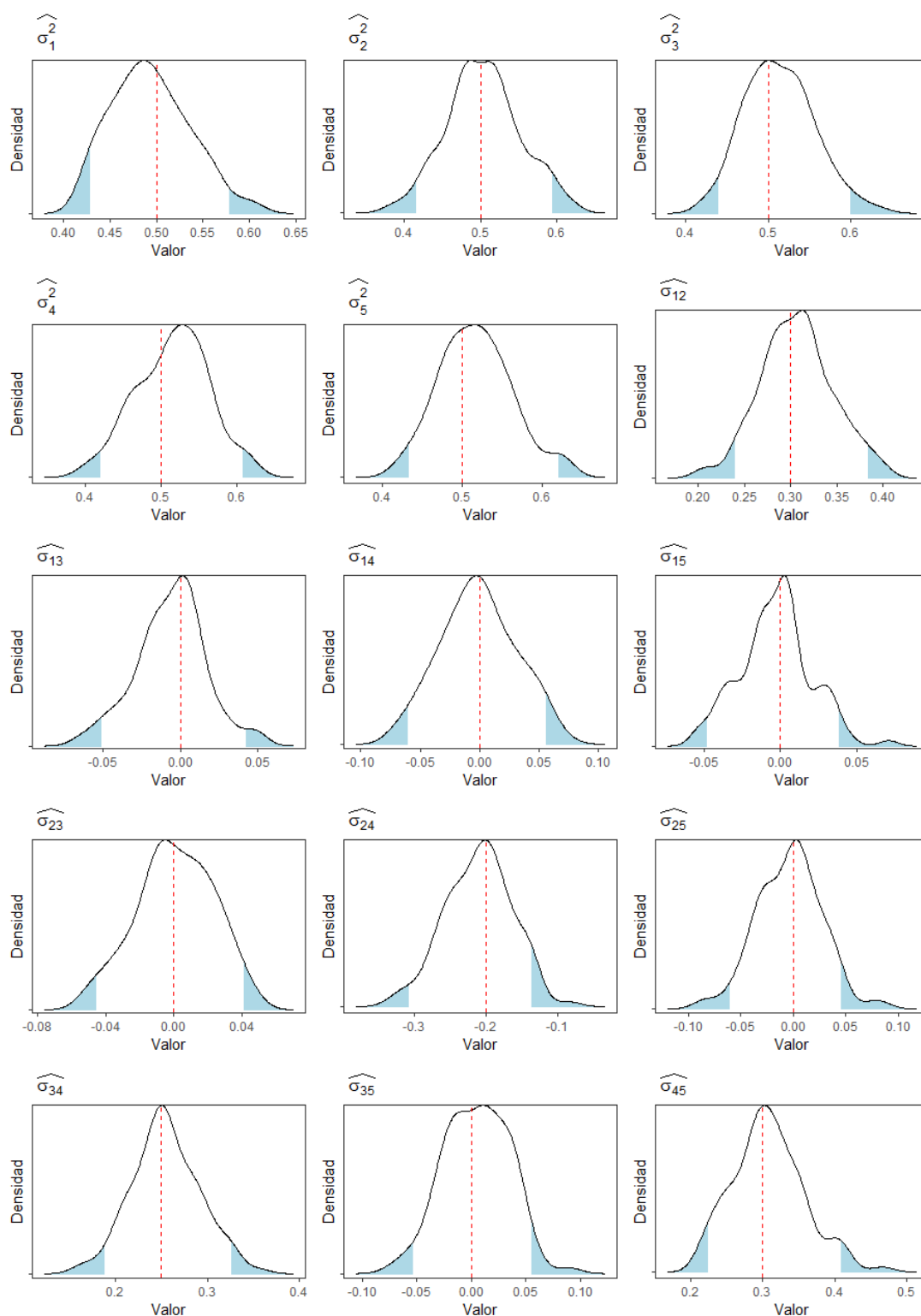


Figura A6.2. Distribución empírica de las varianzas y covarianzas genéticas obtenidas utilizando muestreo de Gibbs sobre un modelo animal multicarácter declarando las relaciones causales como covariables con caracteres correlacionados genéticamente.

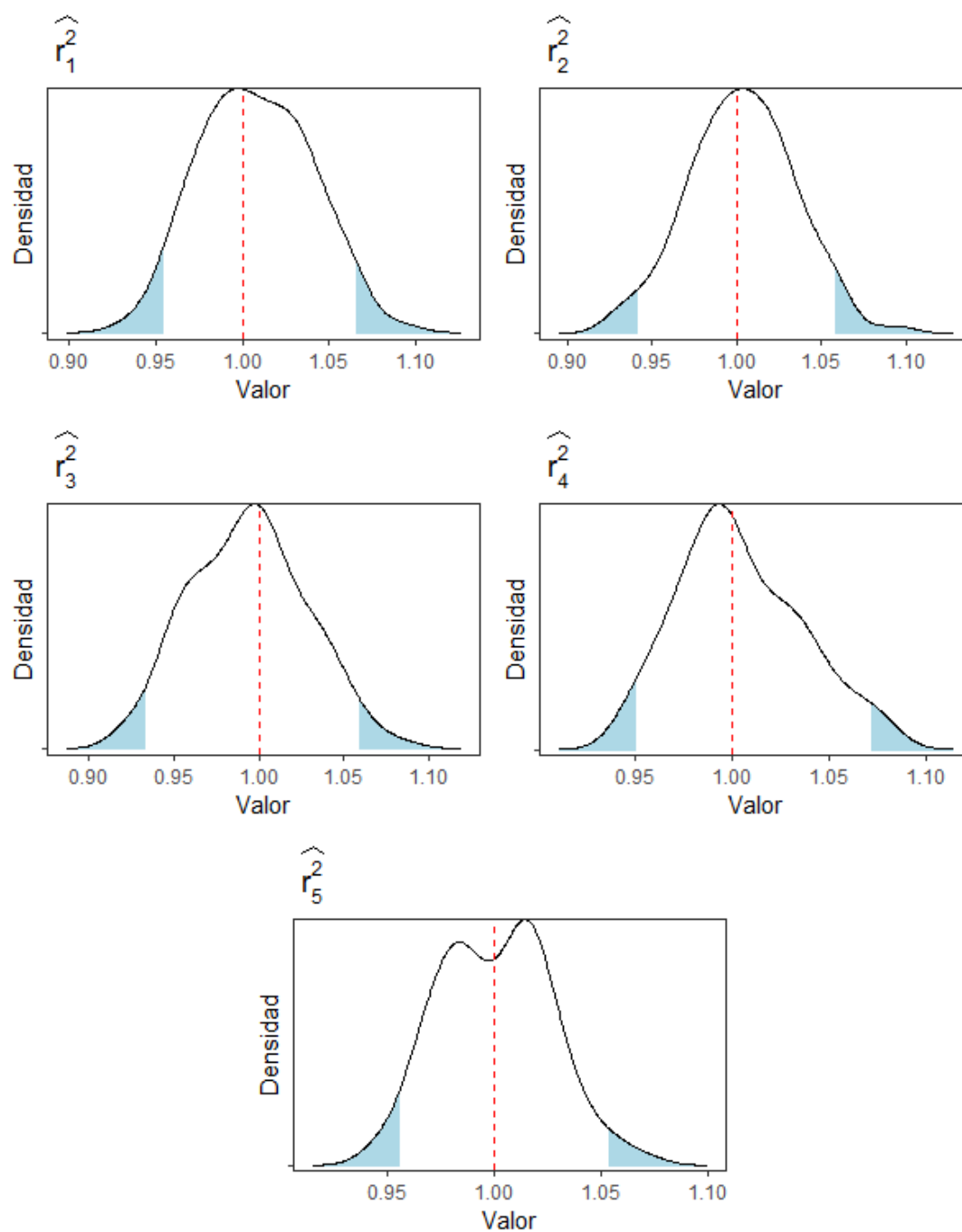


Figura A6.3. Distribución empírica de las varianzas y covarianzas del error obtenidas utilizando muestreo de Gibbs sobre un modelo animal multicarácter declarando las relaciones causales como covariables con caracteres correlacionados genéticamente.