R^2s for Correlated Data: Phylogenetic Models, LMMs, and GLMMs

Anthony R. Ives, Department of Integrative Biology, UW-Madison, Madison, WI 53706

arives@wisc.edu

Appendix 2: More comparisons among the R^2s.

To give a comprehensive assessment of the R^2s, this supplement discusses and plots a complete set of simulations, from which a subset was presented in the main text. The different models (LMM, PGLS, GLMM, and PLOG) are presented in turn.

For the comparisons, R^2_{glmm} and R^2_{ols} are included where applicable. Nakagawa and Schielzeth (2013) present two R^2s. The conditional $R^2_{glmm(c)}$ is the total R^2 that includes the "explained" variance by both fixed and random effects. The marginal $R^2_{glmm(m)}$ give that component of the total variance that is attributed to the fixed effects. Note that $R^2_{glmm(m)}$ is not a partial R^2, but instead is given as

$$R^2_{glmm(m)} = 1 - \frac{\sigma_d^2 + \hat{\sigma}_l^2}{\hat{\sigma}_f^2 + \hat{\sigma}_l^2 + \sigma_d^2}$$

To give a comparable measure for the contribution of the random effects, I define

$$R^2_{glmm(v)} = 1 - \frac{\sigma_d^2 + \hat{\sigma}_f^2}{\hat{\sigma}_f^2 + \hat{\sigma}_l^2 + \sigma_d^2}$$

In the comparisons for LMMs and GLMMs, I compare the partial R^2s of R^2_{ls}, R^2_{lr} and R^2_{ce} for the fixed effect to $R^2_{glmm(m)}$ and the partial R^2s for the random effect to $R^2_{glmm(v)}$. This illustrates the differences between using partial R^2s and marginal R^2s.

Figure Captions

Figure A2.1: Simulation results for a Linear Mixed Model (LMM) giving R^2_{ls}, R^2_{lr}, R^2_{ce}, R^2_{glmm}, and R^2_{ols} versus the log likelihood ratio (LLR) between full and reduced models. The simulation model (equation (7)) contained both a fixed effect β for a continuous variable x and a random effect b for a categorical variable u. For (a), (b) and (c), data were simulated without the random effect ($\beta = 1$, $\sigma = 0$), and for (d), (e) and (f), data were simulated without the fixed effect ($\beta = 0$, $\sigma = 1.5$). Simulations for (g), (h), and (i) contained both fixed and random effects. Columns give different partial R^2s for each method. Specifically, (a), (d), and (g) give the partial R^2s in which the reduced model removes the fixed effect for x: therefore, these give partial R^2s for the fixed effect. Panels (b), (e), and (h) give the partial R^2s in which the reduced model removes the random effect for u: therefore, these give partial R^2s for the random effect. In panels (c), (f) and (i), the reduced model removes both fixed and random effects, giving the total R^2s. Each data set consisted of 100 simulated points, x was simulated as a normal (0, 1) random variable, and u had 10 levels with b is simulated as a normal (0, σ). All analyses were performed with the function lmer().

Figure A2.2: Simulation results for a Linear Mixed Model (LMM) giving associations between R^2_{ls}, R^2_{lr}, R^2_{ce}, R^2_{glmm}, and R^2_{ols}. Data are the same as presented in figure A2.1.

Figure A2.3: Simulation results for a Linear Mixed Model (LMM) showing means and standard deviations of R^2_{ls}, R^2_{lr}, R^2_{ce}, R^2_{glmm}, and R^2_{ols} versus sample size. The simulation model (equation (7)) contained both a fixed effect β for a continuous variable x and a random effect b for a categorical variable u. For each level of u, from 4 to 16 replicates were simulated. (a), (b), and (c) give means of each R^2 were calculated for 500 simulations at each sample, and (d), (e), and (f) give standard deviations. Columns give different partial R^2s, with (a) and (d) giving the partial R^2s for the fixed effect, (b) and (e) giving the partial R^2s for the random effect, and (c) and (f) giving the total R^2s. In the simulations, x is simulated as a normal (0, 1) random variable with $\beta = 1$; u has 10 levels and b is simulated as a normal (0, $\sigma = 1.5$) random variable; and residuals e are independent (0, 1) random variables. All analyses were performed with the function lmer().

Figure A2.4: Simulation results for the phylogenetic model with a continuous predictor variable x giving R^2_{ls}, R^2_{lr}, and R^2_{ce} versus the log likelihood ratio (LLR) between full and reduced models. For each simulation, a phylogenetic tree was first simulated, and the values of x were simulated up the phylogeny assuming Brownian Motion evolution. Data were simulated using equation (7) with $b = 0$, and residuals e_i were simulated from a multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Sigma}(\lambda) = (1 - \lambda)\mathbf{I} + \lambda\mathbf{\Sigma}_{\mathrm{BM}}$. For (a), (b) and (c), data were simulated without phylogenetic signal ($\lambda = 0$, $\beta = 1$), and for (d), (e) and (f), data were simulated without the fixed effect ($\lambda = 0.5$, $\beta = 0$). Simulations for (g), (h), and (i) contained both fixed and phylogenetic effects ($\beta = 1$, $\lambda = 0.5$). (a), (d), and (g) give the partial R^2s for the fixed effect. Panels (b), (e), and (h) give the partial R^2s for the phylogenetic effect. In panels (c), (f) and (i), the reduced model removes both fixed and phylogenetic effects, giving the total R^2s. All analyses were performed with the function phylolm().

Figure A2.5: Simulation results for a PGLS model giving associations between R^2_{ls}, R^2_{lr}, and R^2_{ce}. Data are the same as presented in figure A2.3.

Figure A2.6: Simulation results for the phylogenetic model with a continuous response variable showing means and standard deviations of R^2_{ls}, R^2_{lr}, and R^2_{ce} versus sample size. For each simulation, a phylogenetic tree was first simulated, and the values of the predictor variable x were simulated up the phylogeny assuming Brownian Motion evolution. In equation 9, residuals e_i were simulated from a multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Sigma}(\lambda) = (1 - \lambda)\mathbf{I} + \lambda\mathbf{\Sigma}_{\mathrm{BM}}$, and the parameter values were $\lambda = 0.5$, $\beta = 1$, and $b = 0$. (a), (b), and (c) give means of each R^2 were calculated for 500 simulations at each sample, and (d), (e), and (f) give standard deviations. Columns give different partial R^2s, with (a) and (d) giving the partial R^2s for x, (b) and (e) giving the partial R^2s for phylogenetic signal λ, and (c) and (f) giving the total R^2s. All analyses were performed with the function phylolm().

Figure A2.7: Simulation results for a binary Generalized Linear Mixed Model (GLMM) giving R^2_{ls}, R^2_{lr}, R^2_{ce}, R^2_{glmm}, and R^2_{ols} versus the log likelihood ratio (LLR) between full and reduced models. The simulation model (equation (7)) contained both a fixed effect β for a continuous variable x and a random effect b for a categorical variable u. For (a), (b) and (c), data were simulated without the random effect ($\beta = 1.8$, $\sigma = 0$), and for (d), (e) and (f), data were simulated without the fixed effect ($\beta = 0$, $\sigma = 1.8$). Simulations for (g), (h), and (i) contained both fixed and random effects. (a), (d), and (g) give the partial R^2s for the fixed effect, and panels (b), (e), and (h) give the partial R^2s for the random effect. In panels (c), (f) and (i) give total R^2s. In the simulations, x is simulated as a normal (0, 1) random variable and u has 10 levels and b is simulated as a normal (0, σ). All analyses were performed with the function glmer().

Figure A2.8: Simulation results for a Generalized Linear Mixed Model (GLMM) giving associations between R^2_{ls}, R^2_{lr}, R^2_{ce}, R^2_{glmm}, and R^2_{ols}. Data are the same as presented in figure A2.7.

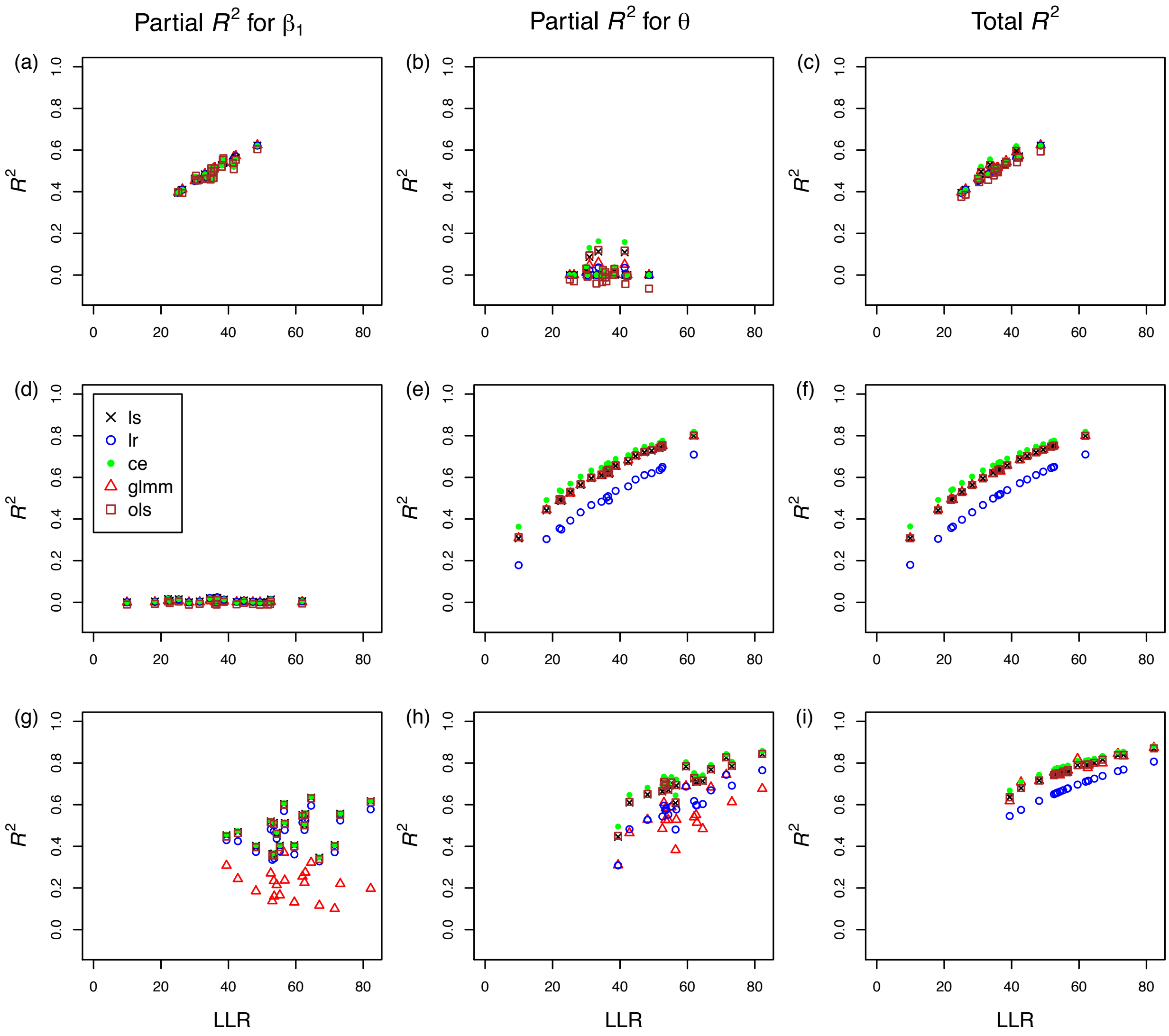
Figure A2.9: Simulation results for a binary Generalized Linear Mixed Model (GLMM) showing means and standard deviations of R^2_{ls}, R^2_{lr}, R^2_{ce}, R^2_{glmm}, and R^2_{ols} versus sample size. The simulation model (equation (7)) contained both a fixed effect β for a continuous variable x and a random effect b for a categorical variable u. For each level of u, from 4 to 16 replicates were simulated. (a), (b), and (c) give means of each R^2 were calculated for 1000 simulations at each sample, and (d), (e), and (f) give standard deviations. Columns give different partial R^2s, with (a) and (d) giving the partial R^2s for the fixed effect, (b) and (e) giving the partial R^2s for the random effect, and (c) and (f) giving the total R^2s. In the simulations, x is simulated as a normal (0, 1) random variable with $\beta = 1.8$ and u has 10 levels and b is simulated as a normal ($0, \sigma = 1.8$) random variable. All analyses were performed with the function glmer().

Figure A2.10: Simulation results for the phylogenetic model with a continuous predictor variable x giving R^2_{ls}, R^2_{lr}, and R^2_{ce} versus the log likelihood ratio (LLR) between full and reduced models. For each simulation, a phylogenetic tree was first simulated, and the values of x were simulated up the phylogeny assuming Brownian Motion evolution. Data were simulated using equation (7) with $b = 0$, and residuals e_i were simulated from a multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Sigma}(\lambda) = (1 - \lambda)\mathbf{I} + \lambda\mathbf{\Sigma}_{\mathrm{BM}}$. For (a), (b) and (c), data were simulated without phylogenetic signal ($\lambda = 0$, $\beta = 1.5$), and for (d), (e) and (f), data were simulated without the fixed effect ($\lambda = 2$, $\beta = 0$). Simulations for (g), (h), and (i) contained both fixed and phylogenetic effects ($\lambda = 2$, $\beta = 1.5$). (a), (d), and (g) give the partial R^2s for the fixed effect. Panels (b), (e), and (h) give the partial R^2s for the phylogenetic effect. In panels (c), (f) and (i), the reduced model removes both fixed and phylogenetic effects, giving the total R^2s. All analyses were performed with the function phylolm().

Figure A2.11: Simulation results for a phylogenetic logistic regression model giving associations between R^2_{ls}, R^2_{lr}, and R^2_{ce}. Data are the same as presented in figure A2.10.

Figure A2.12: Simulation results for the phylogenetic model with a binary response variable showing means and standard deviations of R^2_{ls}, R^2_{lr}, and R^2_{ce} versus sample size. For each simulation, a phylogenetic tree was first simulated, and residuals e_i (equation (7)) were simulated from a multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Sigma}(\lambda) = \lambda\mathbf{\Sigma}_{\mathrm{BM}}$. Values of the predictor variable x were assumed to be independently distributed by a (0,1) normal distribution, and the parameter values were $\lambda = 2$, $\beta = 1.5$, and $b = 0$. (a), (b), and (c) give means of each R^2 were calculated for 500 simulations at each sample, and (d), (e), and (f) give standard deviations. Columns give different partial R^2s, with (a) and (d) giving the partial R^2s for x, (b) and (e) giving the partial R^2s for phylogenetic signal λ, and (c) and (f) giving the total R^2s. Calculations of R^2_{lr} were performed with a modified version of the function phyloglm() and the function glm(). Calculations of R^2_{ls} and R^2_{ce} were performed with the function binaryPGLMM().

Fig. A2.1



Partial R^2 for β_1
Partial R^2 for θ
Total R^2
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
R^2
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
× ls
○ lr
● ce
△ glmm
□ ols
LLR

Fig. A2.2

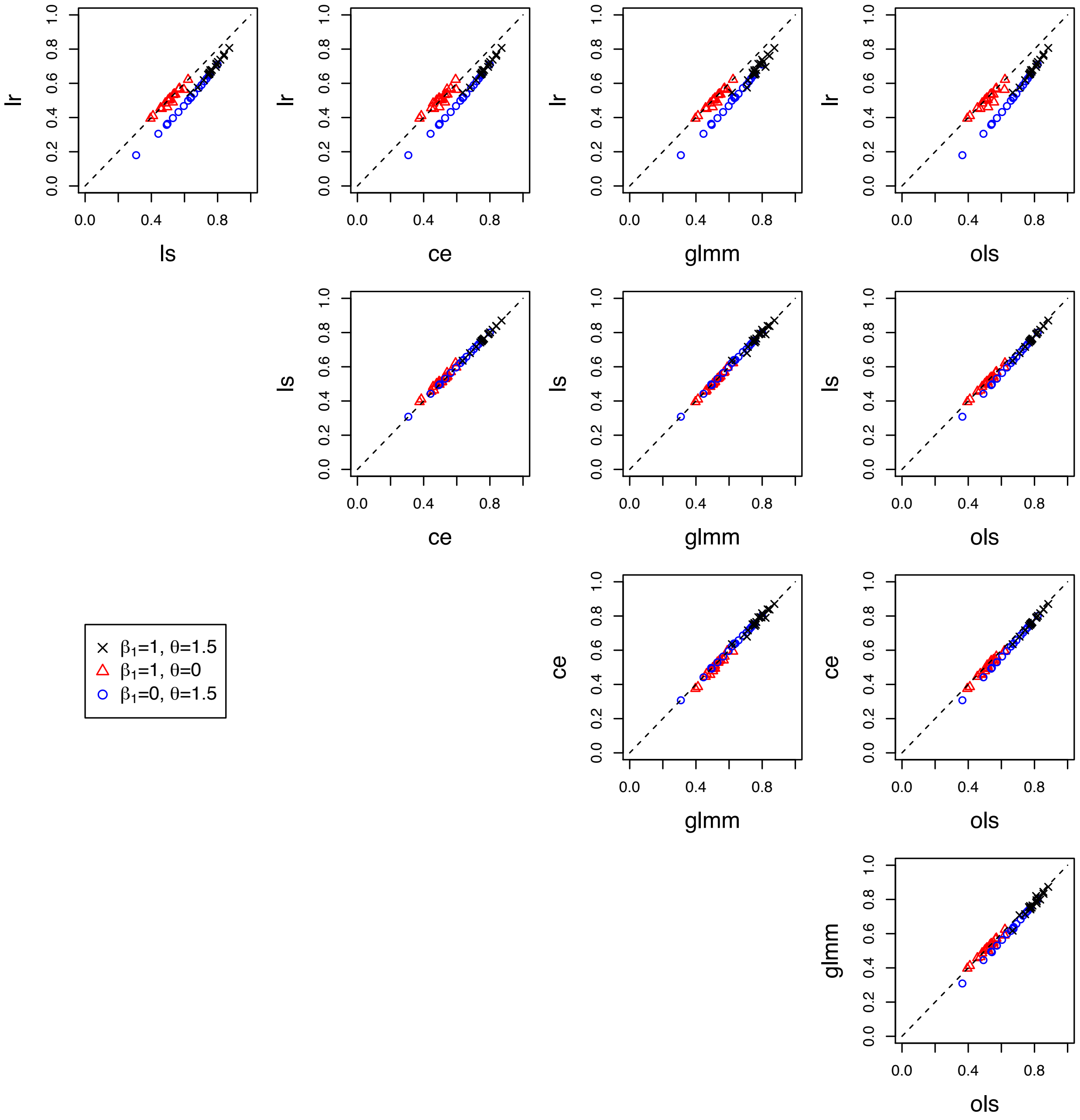


Fig. A2.3

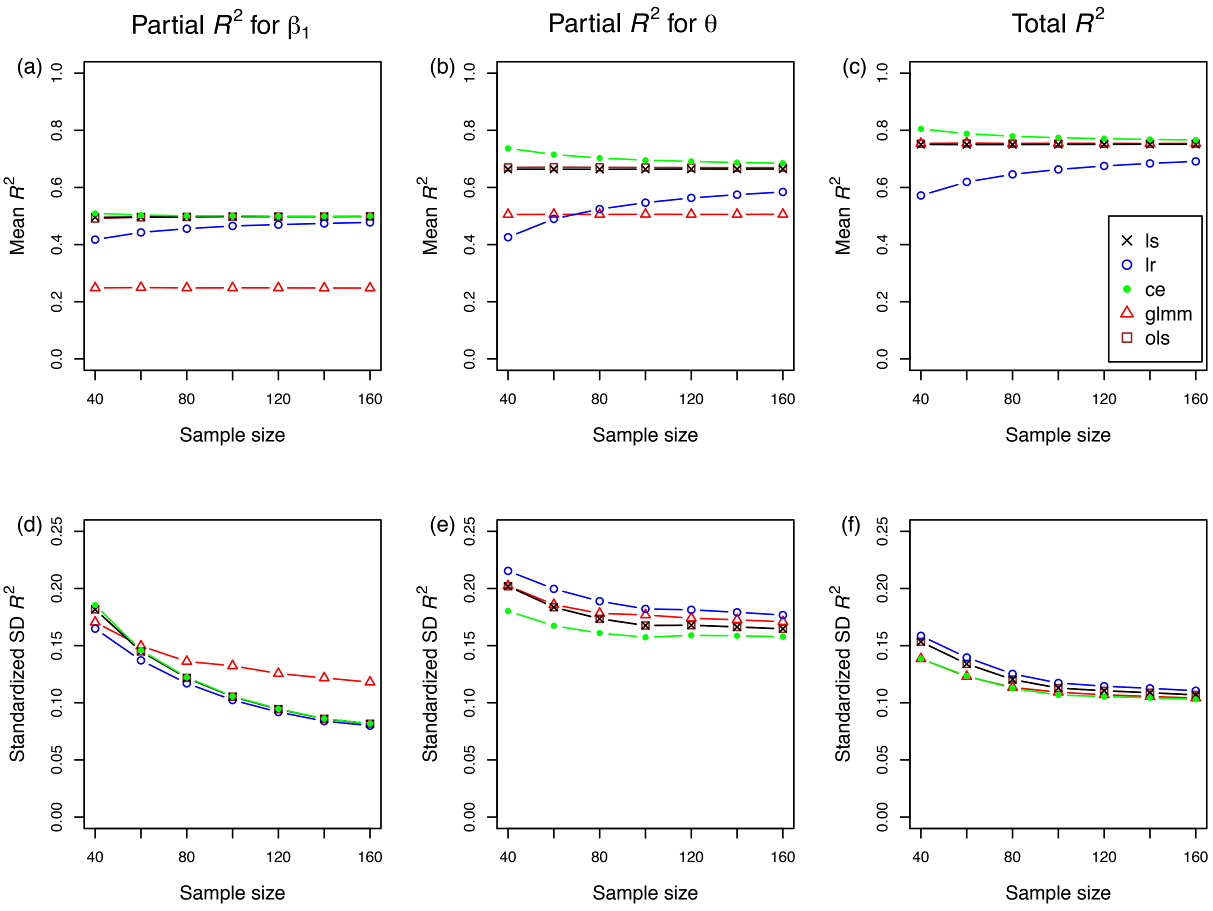

Partial R^2 for β_1
Partial R^2 for θ
Total R^2
(a)
(b)
(c)
(d)
(e)
(f)
Mean R^2
Standardized SD R^2
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.05
0.10
0.15
0.20
0.25
40
80
120
160
Sample size
ls
lr
ce
glmm
ols

Fig. A2.4

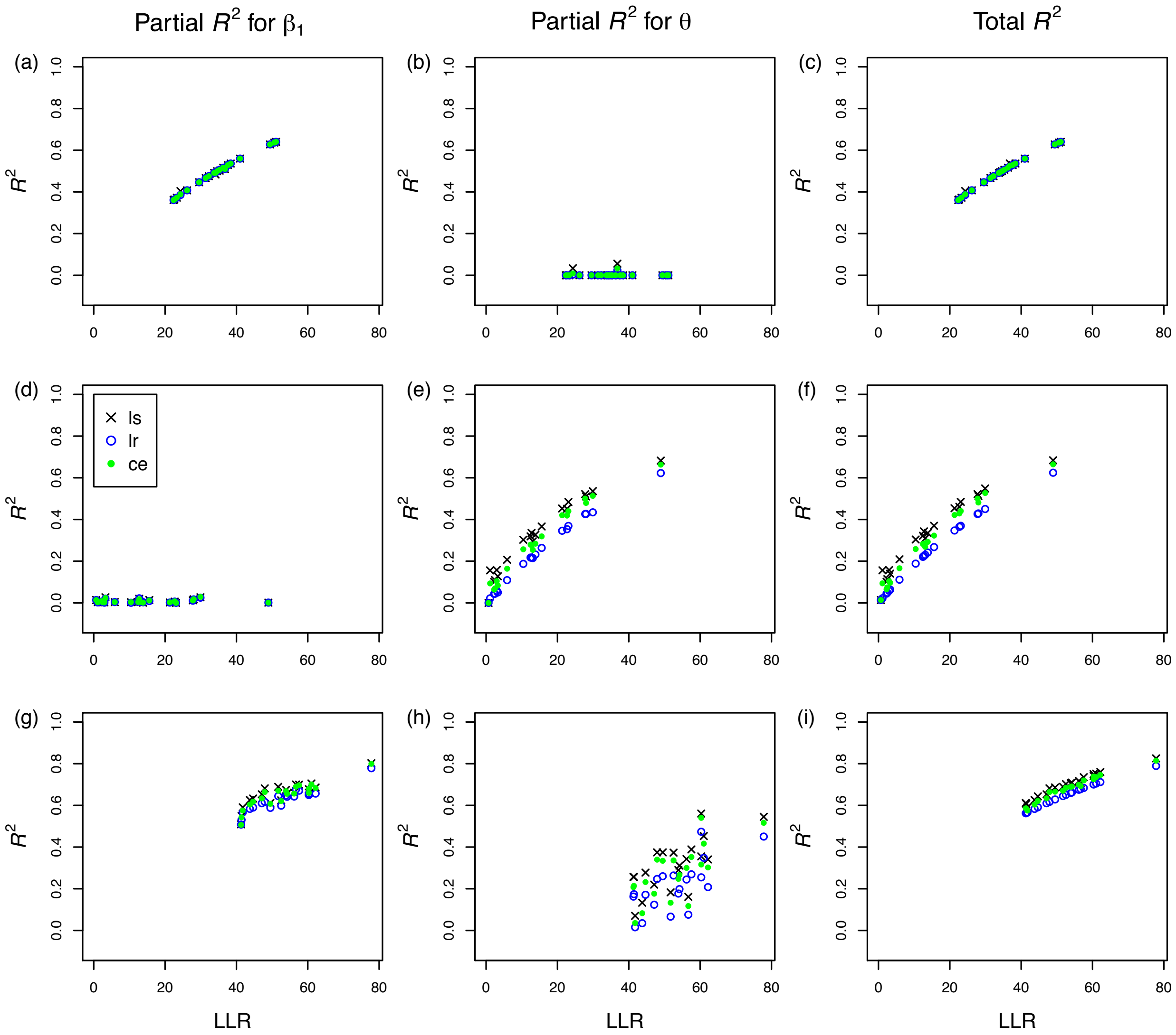
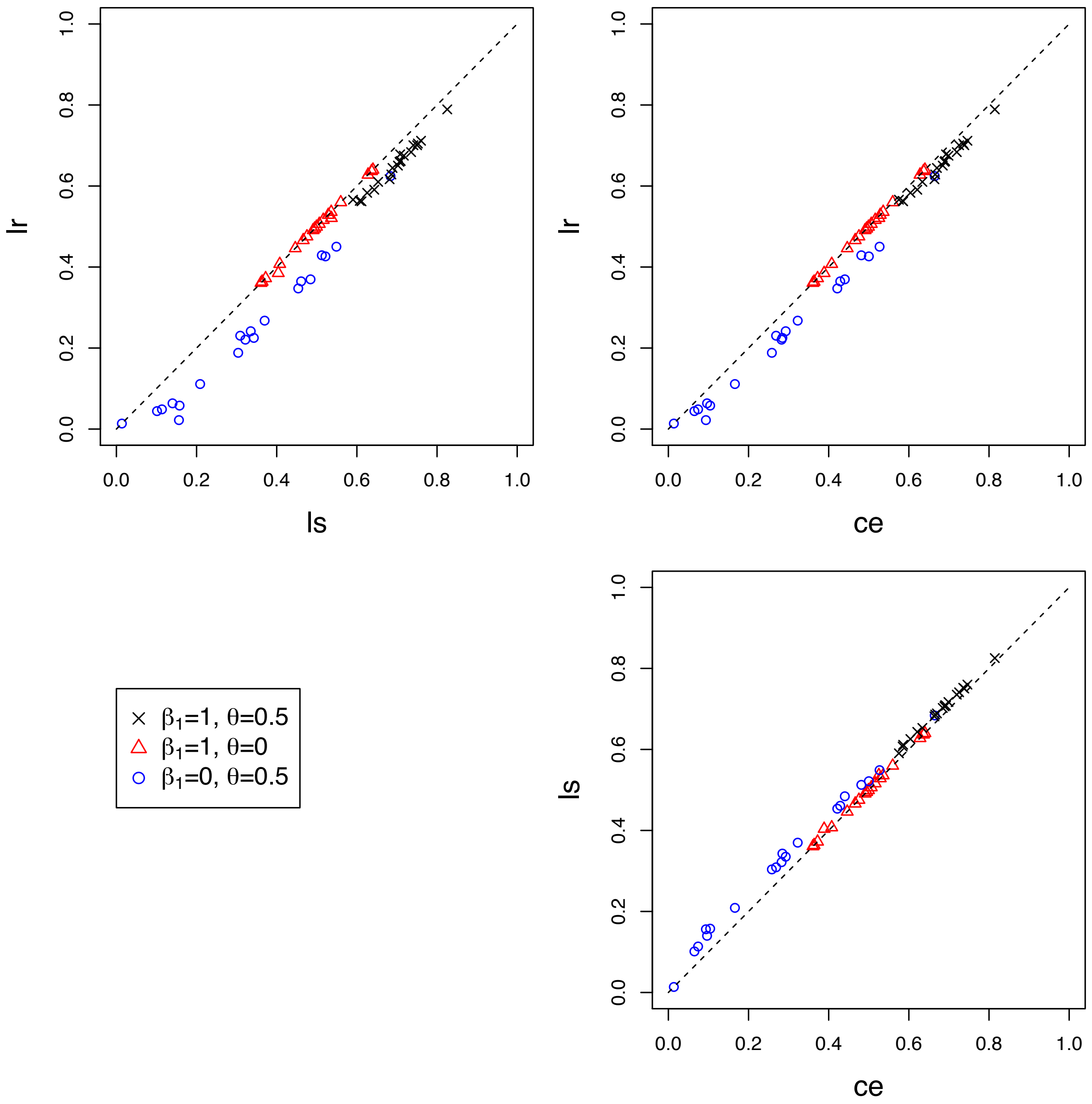

Partial R^2 for β_1
Partial R^2 for θ
Total R^2
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
R^2
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
ls
lr
ce
LLR

Fig. A2.5



lr
ls
lr
ce
ls
ce
0.0
0.2
0.4
0.6
0.8
1.0
β_1=1, θ=0.5
β_1=1, θ=0
β_1=0, θ=0.5

Fig. A2.6

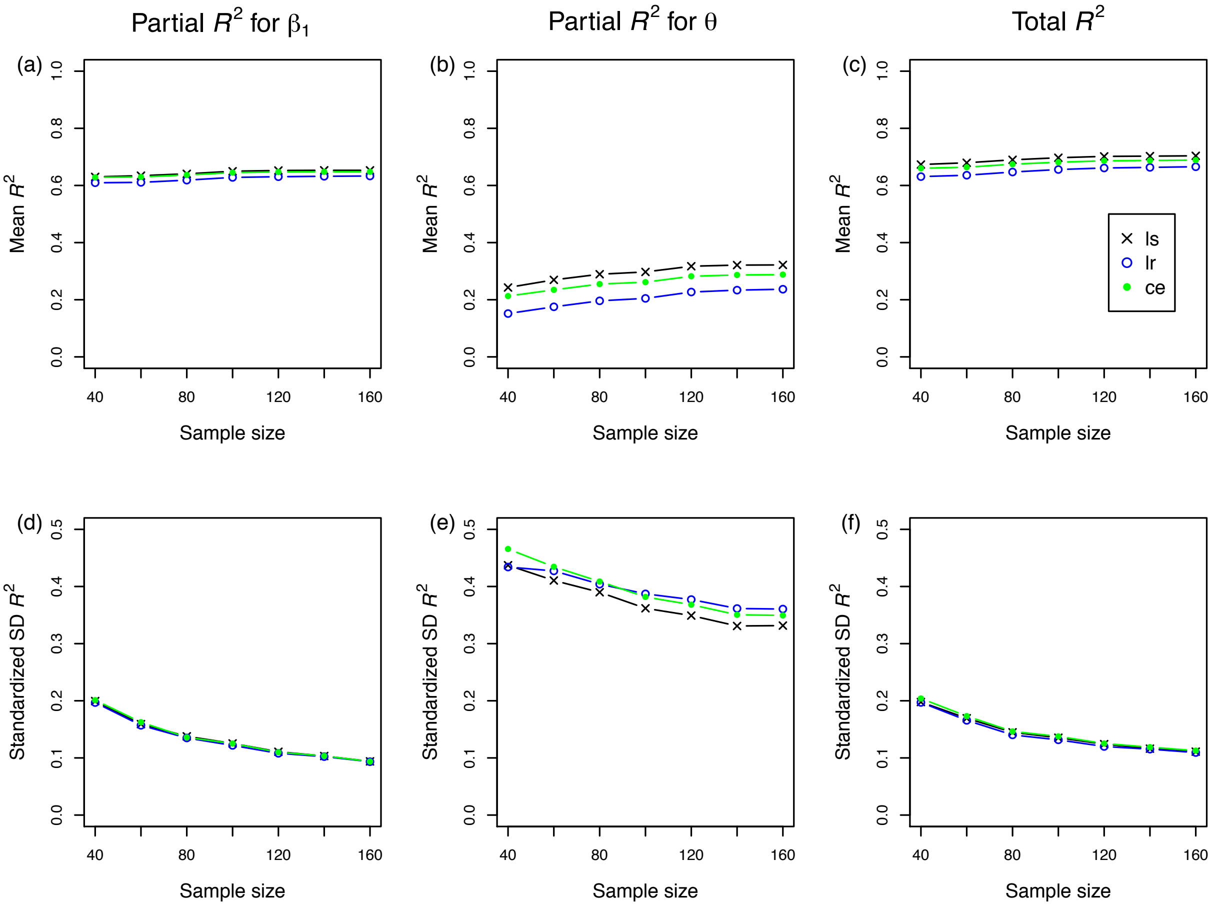
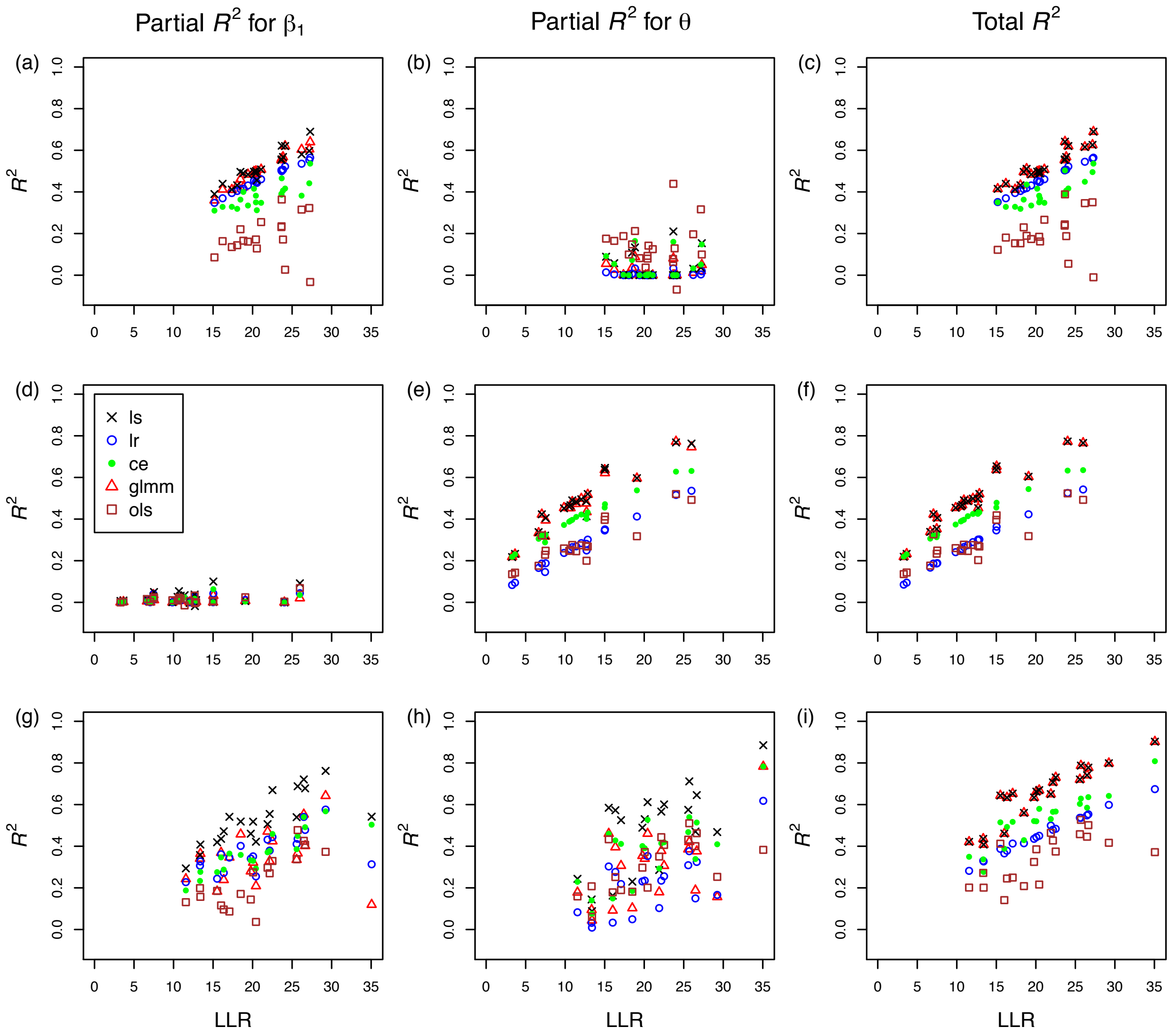

Partial R^2 for β_1
Partial R^2 for θ
Total R^2
(a)
(b)
(c)
(d)
(e)
(f)
Mean R^2
Standardized SD R^2
Sample size
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.3
0.5
40
80
120
160
× ls
○ lr
● ce

Fig. A2.7



Partial R^2 for β_1
Partial R^2 for θ
Total R^2
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
R^2
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
35
ls
lr
ce
glmm
ols
LLR

Fig. A2.8

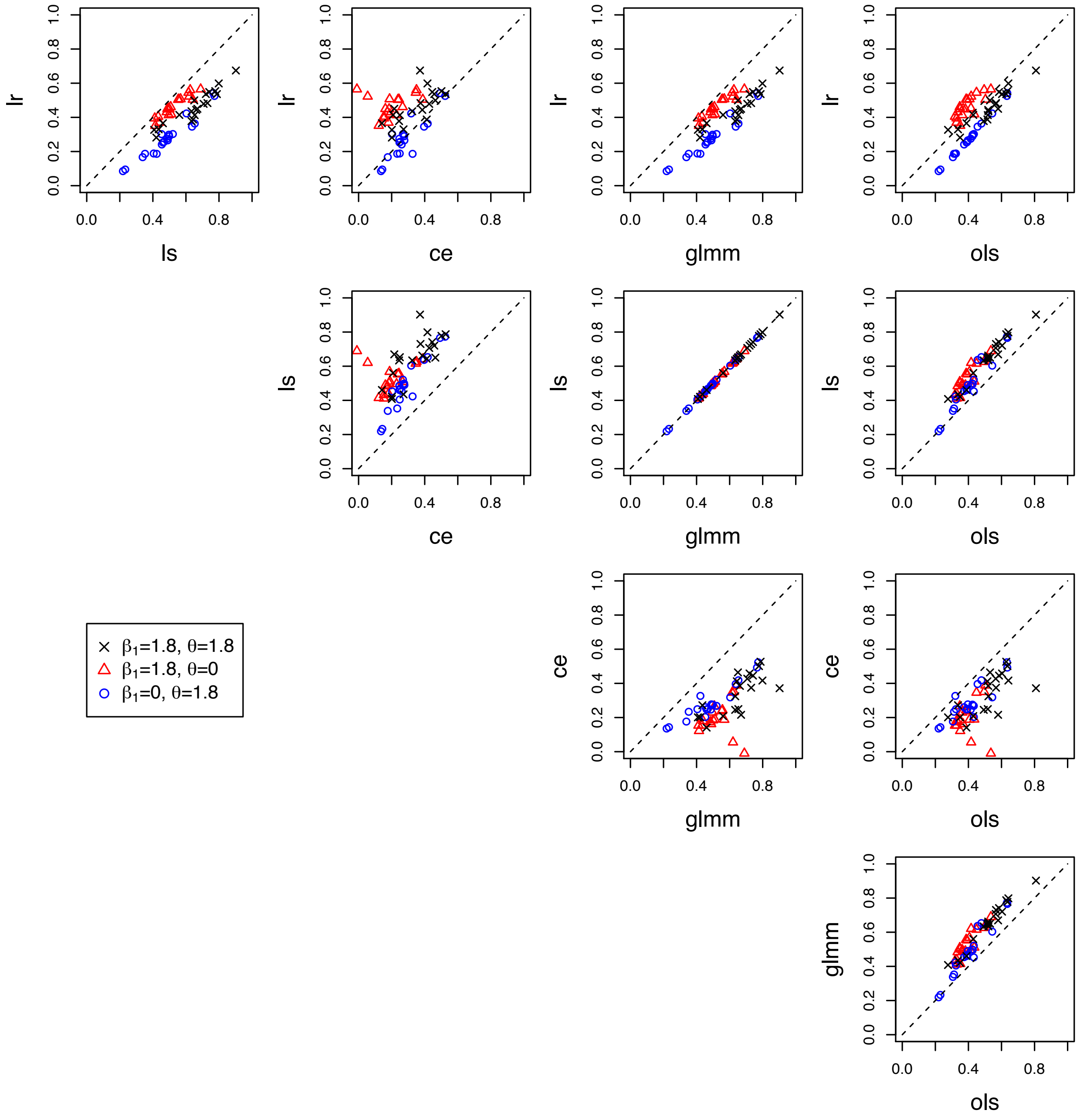


Fig. A2.9

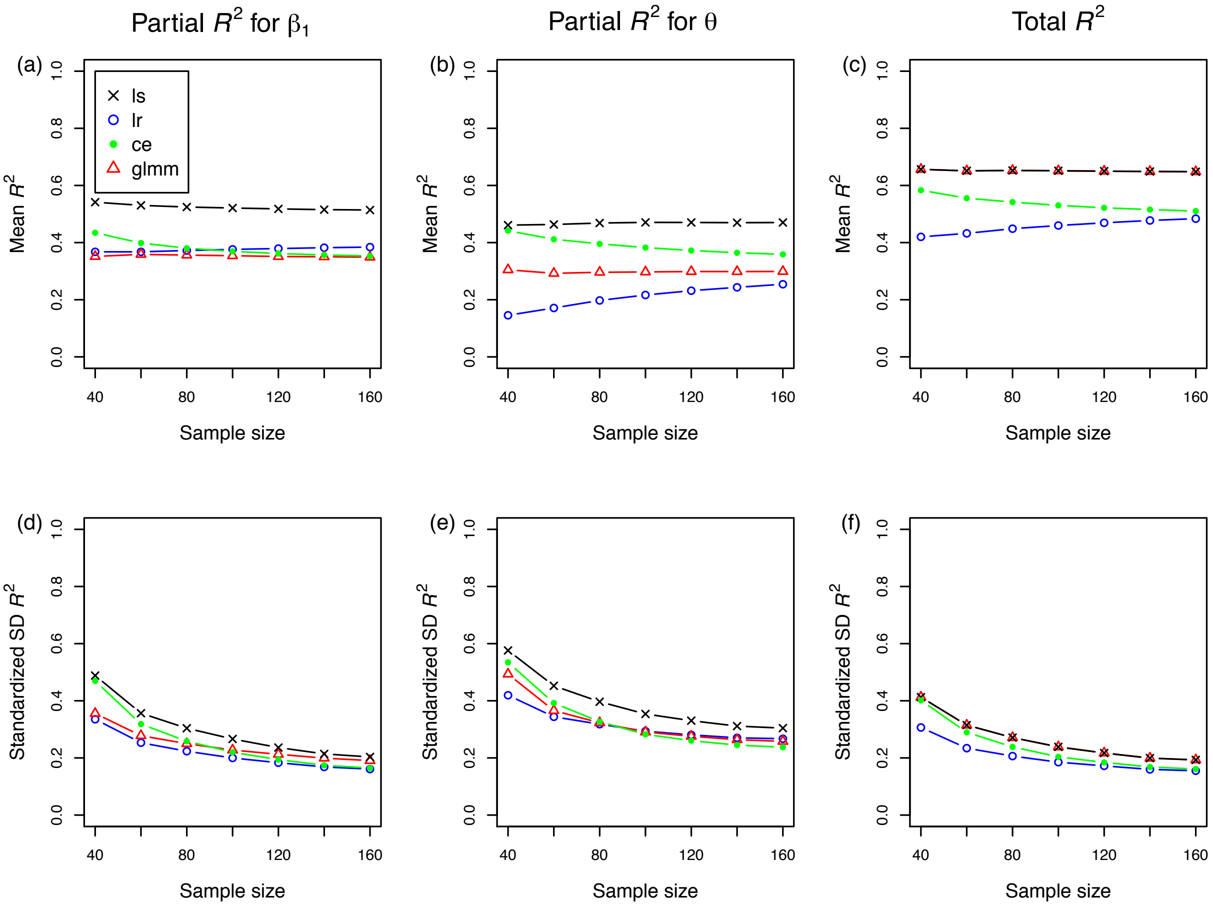

Partial R^2 for β_1
Partial R^2 for θ
Total R^2
(a)
(b)
(c)
(d)
(e)
(f)
Mean R^2
Standardized SD R^2
ls
lr
ce
glmm
0.0
0.2
0.4
0.6
0.8
1.0
40
80
120
160
Sample size

Fig. A2.10

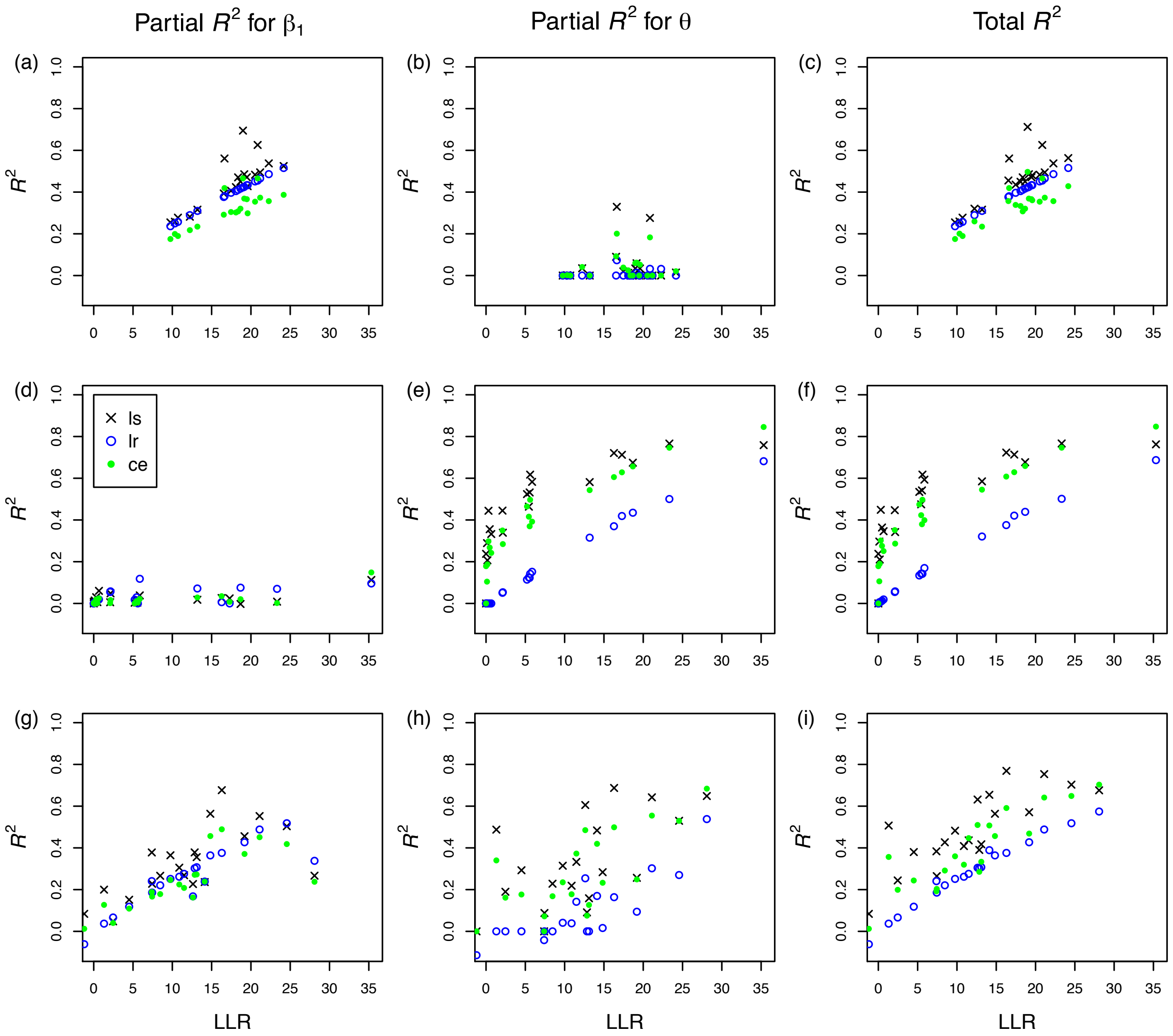
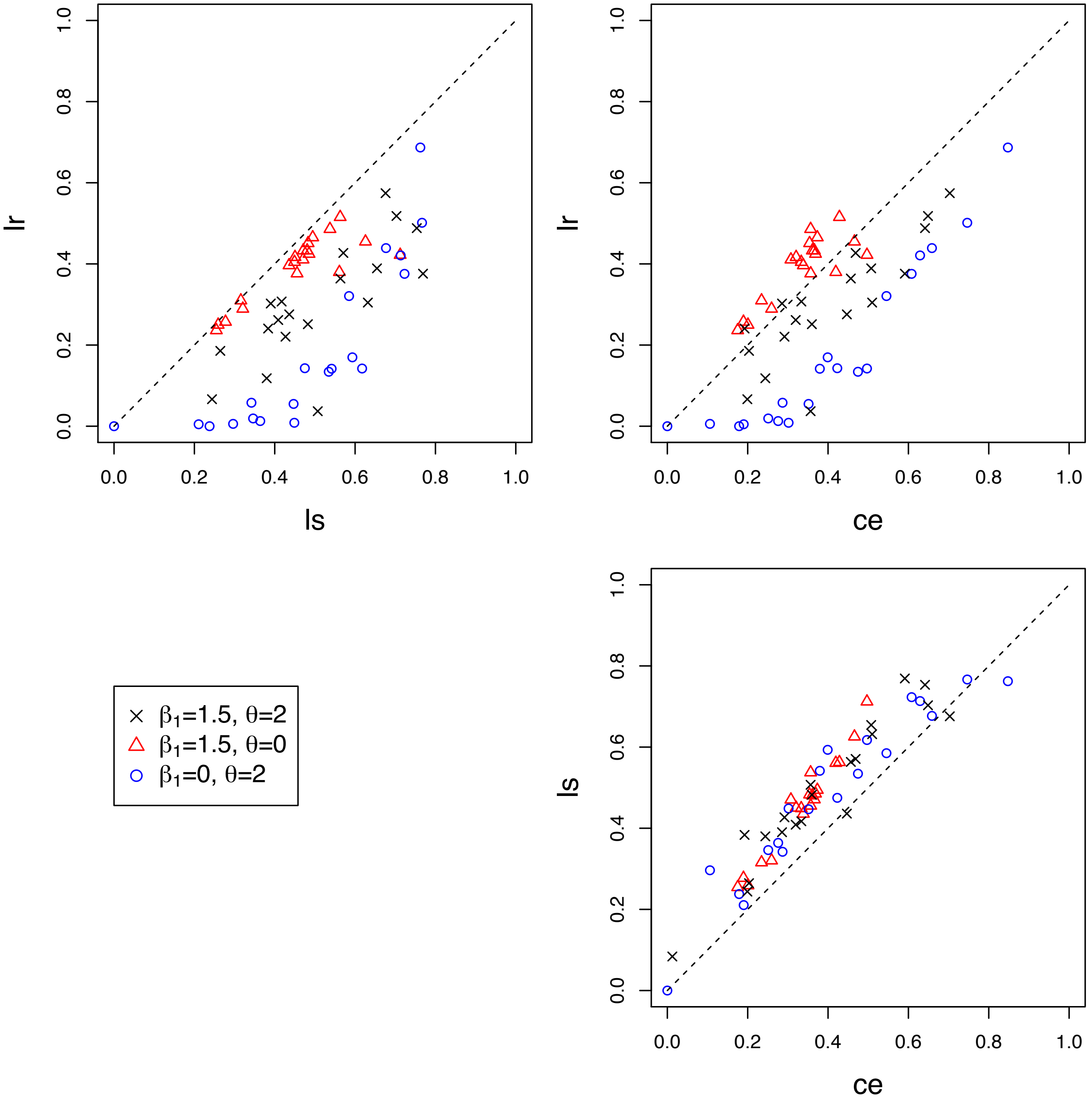
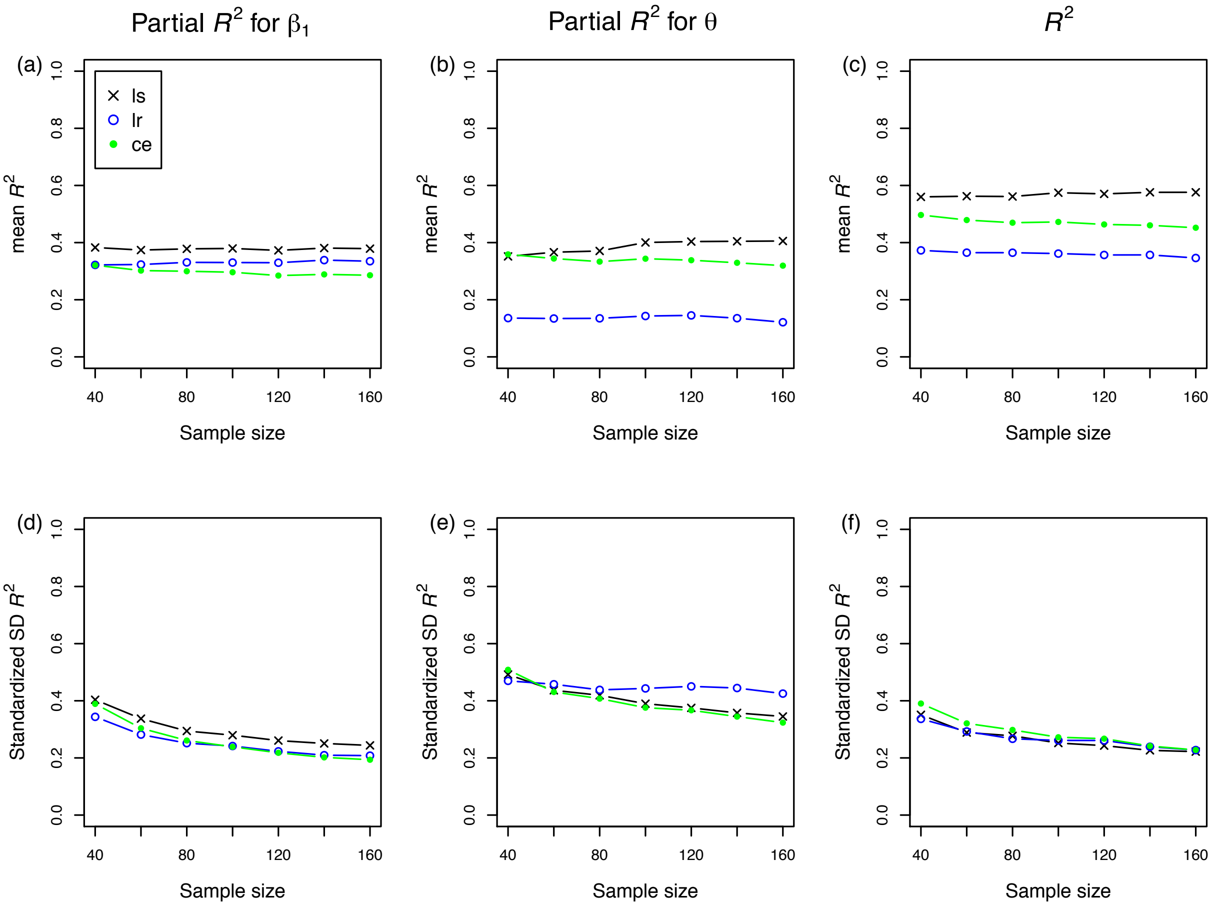

Partial R2 for β1
Partial R2 for θ
Total R2
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
R2
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
35
ls
lr
ce
LLR

Fig. A2.11



lr
ls
lr
ce
ls
ce
0.0
0.2
0.4
0.6
0.8
1.0
β_1=1.5, θ=2
β_1=1.5, θ=0
β_1=0, θ=2

Fig. A2.12



Partial R^2 for β_1
Partial R^2 for θ
R^2
(a)
(b)
(c)
(d)
(e)
(f)
ls
lr
ce
mean R^2
Standardized SD R^2
Sample size
0.0
0.2
0.4
0.6
0.8
1.0
40
80
120
160