Figure and Table Placeholders

Methods

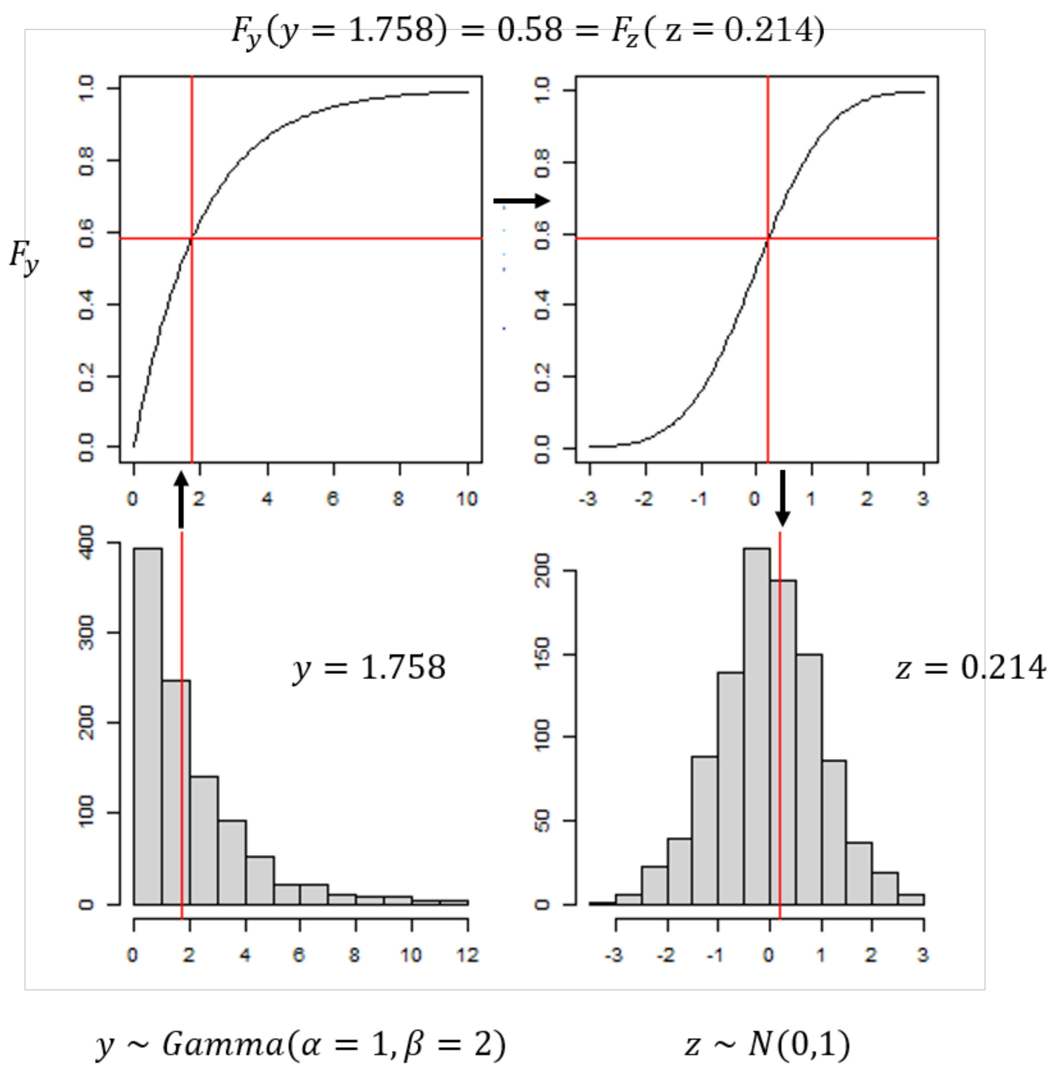


Figure 1: Bottom left: An observation, y, plotted against its distribution, Gamma(alpha = 1, beta = 2). Top left: The cdf value of the observation y given the distribution parameters. Top right: The same cdf value plotted on a standard normal cdf curve. Bottom right: The inverse cdf of the value is plotted on a standard normal distribution.

```
## Warning in ks.test(pear.res, "pnorm"): ties should not be present for the
## Kolmogorov-Smirnov test
```

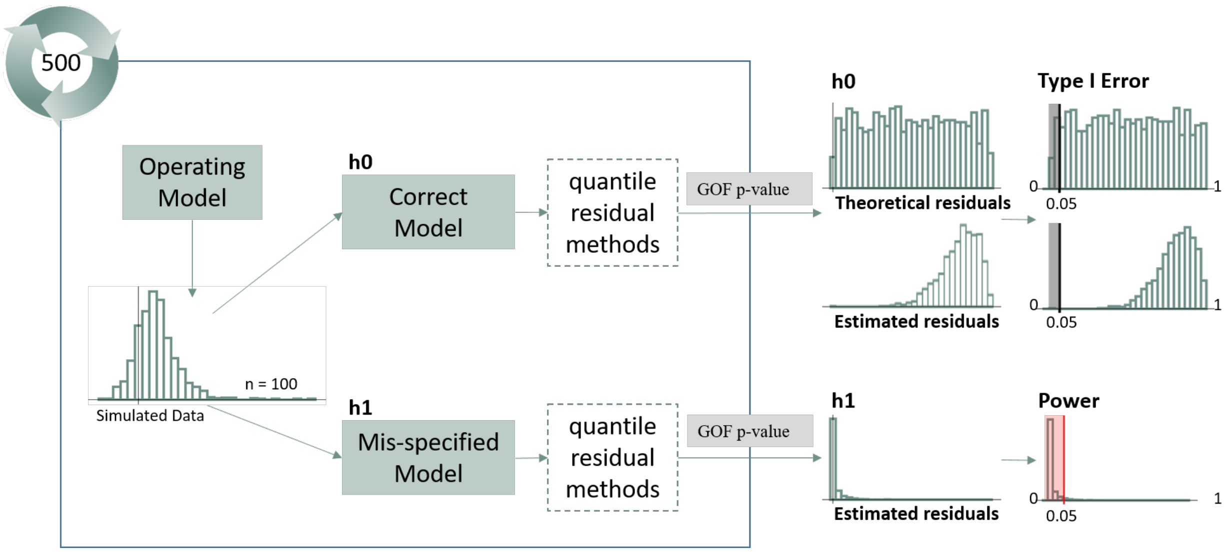
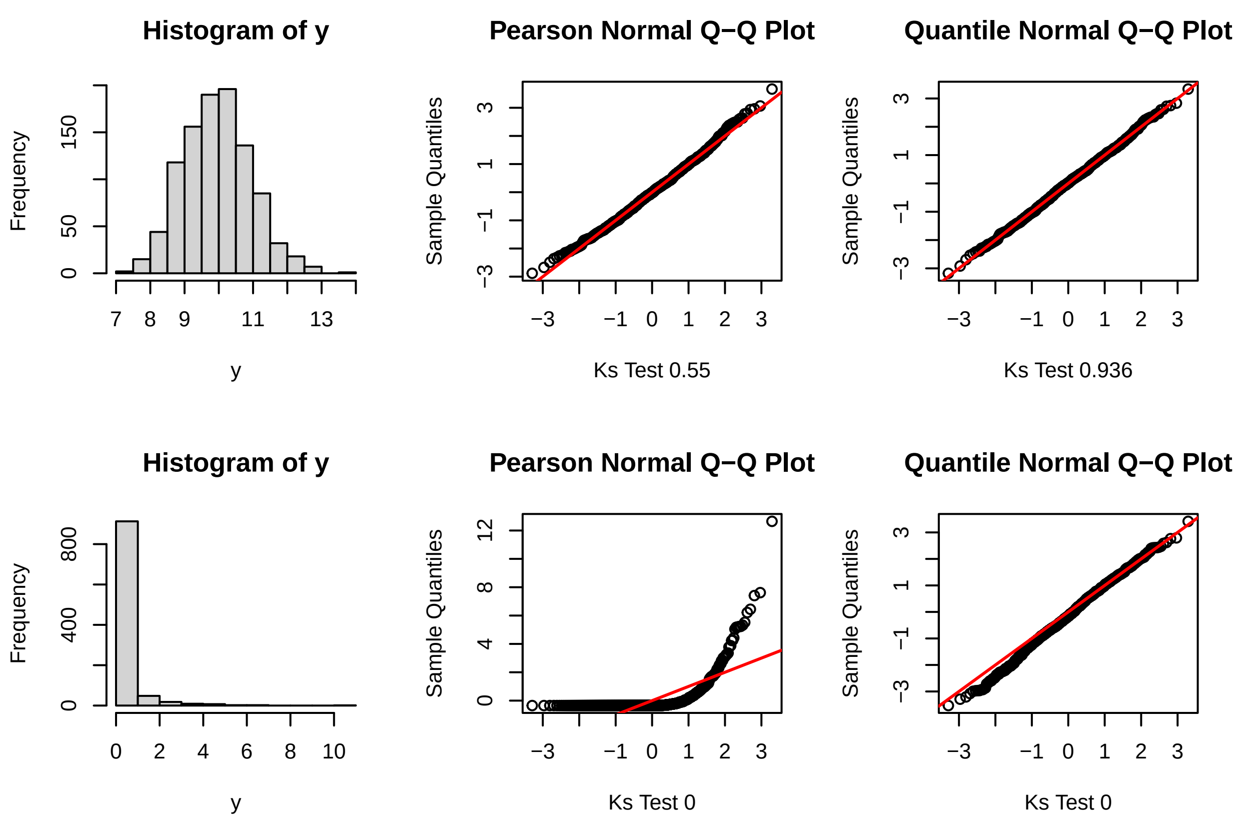
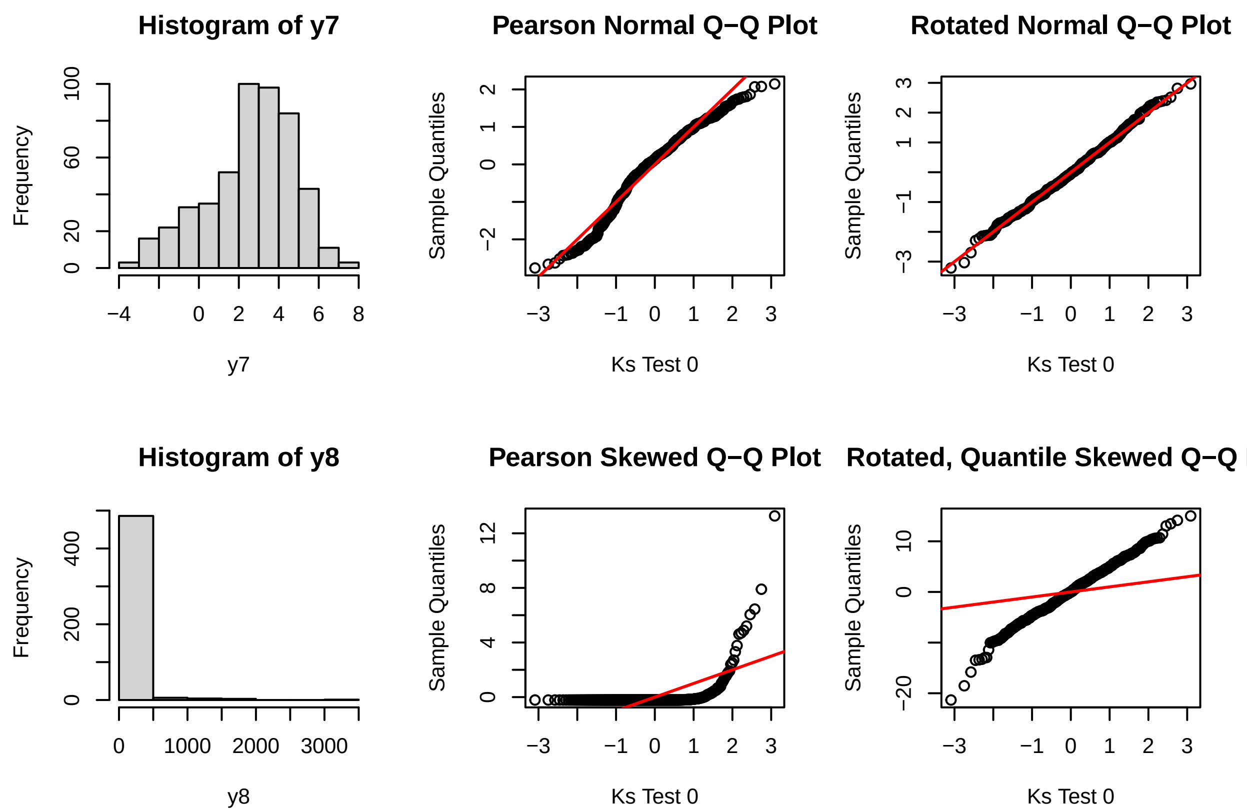


Figure 2: Overview of Simulation Study. Data were first simulated under the Operating, or True Model. Data were then fit to two separate models: the same operating model and the mis-specified model. For each model fit, quantile residuals and subsequent GOF p-values were calculated for each method. This simulation was repeated 500 times and resulted in a distribution of p-values for each method under the correct and mis-specified model.




```
## Warning in ks.test(r8.pears, "pnorm"): ties should not be present for the
```

Kolmogorov-Smirnov test



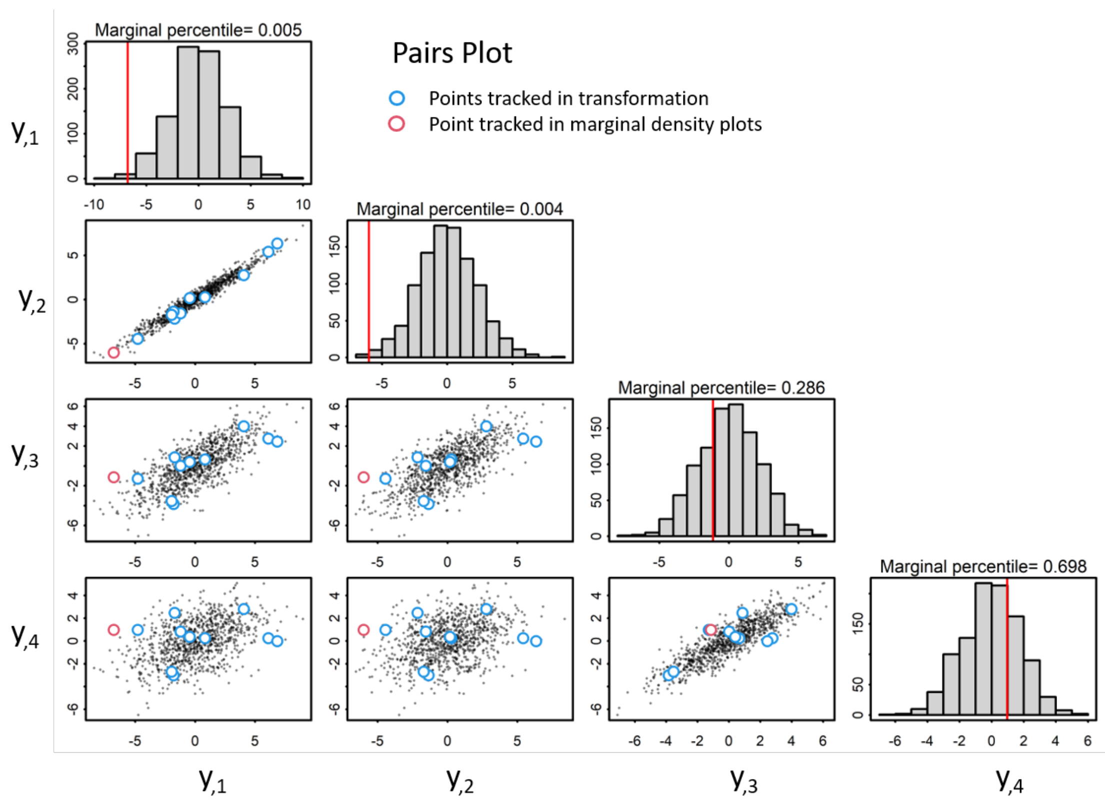


Figure 3: Given zero-centered multivariate data with a covariance matrix, Sigma. Pairs plo=ts visualize the correlation structure of the data. Blue and red indicate points tracked in transformation. The red points correspond with the marginal percentile in the histogram.

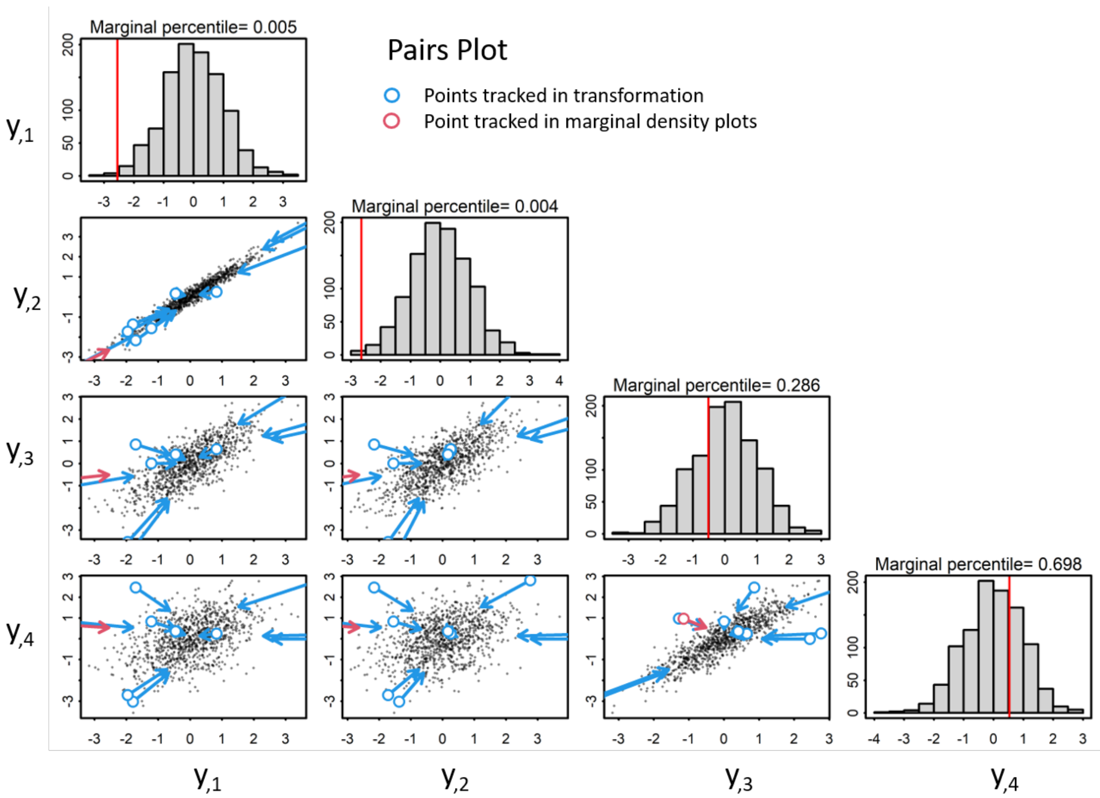


Figure 4: When observations are scaled to a unit variance, data are transformed to standardized normal space, yet correlation structure is retained.

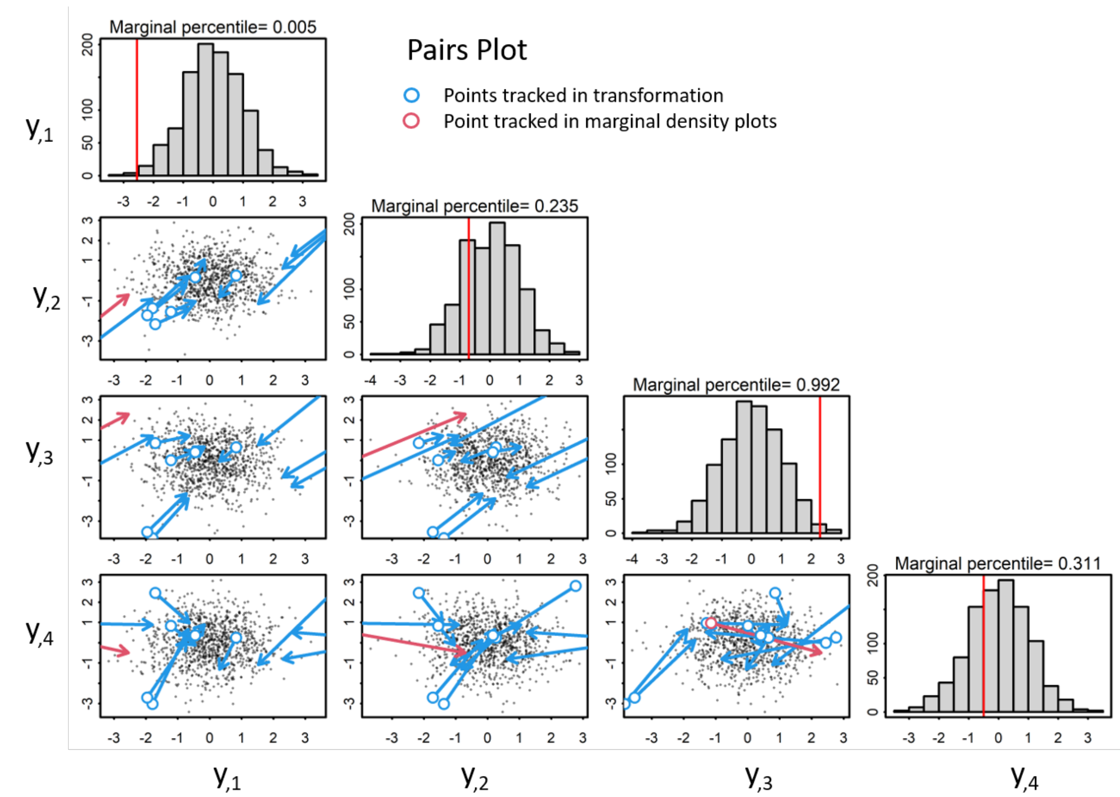


Figure 5: In order to properly decorrelate the data, we need to apply a decoorelation method, such as the cholesky transformation. In this approach, we calculate the cholesky decomposition of the covariance matrix, Sigma, with which we use to transform the data to iid standardized normal space via both a scaling and a rotaion.

Table 1: Linear Model Simulation: data generating models, parameter values, and mis-specifications.

Data Generating Model	Parameters	Mis-specified Model
$X_i \sim N(0,1)$ $\mu_{i,j} = X_i\beta$ $y_{i,j} \sim N(\mu_{i,j}, \sigma_y)$	$\beta = (4, -5)$ $\sigma_y = 1$	Data simulated with lognormal overdispersion: $\mu_{i,j} = X_i\beta + exp(\epsilon)$ $\epsilon \sim N(0,1)$ Data fit to model without drift term

Table 2: Mixed Model Simulation: data generating models, parameter values, and mis-specifications.

Data Generating Model	Parameters	Mis-specified Model
$X_i \sim N(0,1)$ $u_j \sim N(0, \sigma_u)$ $\mu_{i,j} = X_i\beta + u_j$ $y_{i,j} \sim N(\mu_{i,j}, \sigma_y)$	$\beta = (4, -8)$ $\sigma_u = 2$ $\sigma_y = 0.5$	Data simulated with covariate term Data fit to model without covariate term
$X_i \sim Unif(-0.5, 0.5)$ $u_j \sim N(0, \sigma_u)$ $\mu_{i,j} = X_i\beta + u_j$ $y_{i,j} \sim N(\mu_{i,j}, \sigma_y)$	$\beta = (4, -8)$ $\sigma_u = 2$ $\sigma_y = 0.5$	Data simulated with covariate term Data fit to model without covariate term
$u_j \sim N(0, \sigma_u)$ $\mu_{i,j} = exp(\beta_0 + u_j)$ $y_{i,j} \sim Tweedie(\mu_{i,j}, \phi, p)$	$\beta = 1.5$ $\sigma_u = 1.4$ $\phi = 1.4$ $p = 1.2$	Data simulated with random effect term Data fit to model without random effect term

Table 3: Randomwalk Simulation: data generating models, parameter values, and mis-specifications.

Data Generating Model	Parameters	Mis-specified Model
$\mu_i = u_{i-1} + a$ $u_i \sim N(\mu_i, \tau)$ $y_i \sim N(u_i, \sigma)$	$a = 0.75$ $\tau = 1$ $\sigma = 1$	Data simulated with drift term, a Data fit to model without drift term

Table 4: Spatial Simulation: data generating models, parameter values, and mis-specifications.

Data Generating Model	Parameters	Mis-specified Model
$\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = \beta_0 + \omega_i$ $y \sim N(\eta, \sigma_y)$	spatial range = 50 $\kappa = \sqrt{8}/50$ $\sigma_\omega^2 = 1$ $\beta_0 = 1$ $\sigma_y = 1$	Data simulated with $\exp(\omega_i)$ Data fit to model without covariate term
$\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = \beta_0 + \omega_i$ $y \sim Pois(exp(\eta))$	spatial range = 50 $\kappa = \sqrt{8}/50$ $\sigma_\omega^2 = 2$ $\beta_0 = 0.5$	Data simulated with random effect term Data fit to model without random effect term

Results

Simple Linear Model

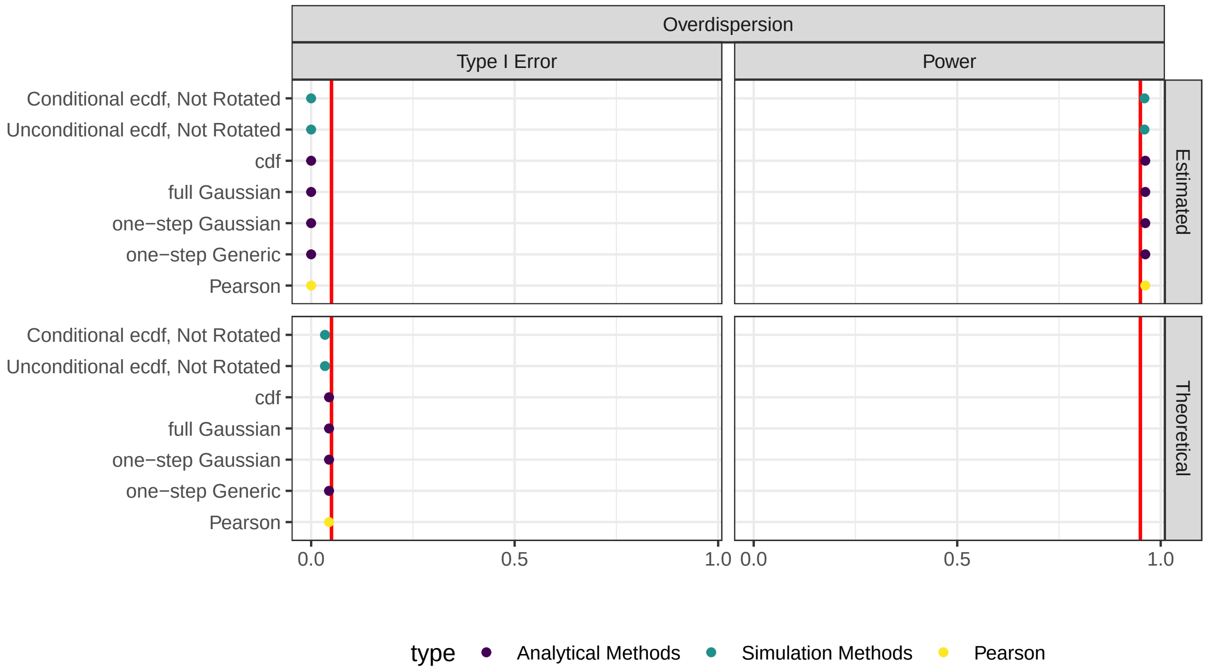


Figure 6: Simple Linear Model. Type I error rates and Power evaluated for each analytical and simulation method. Results are partitioned out by residual type (top to bottom).

Simple Mixed Model

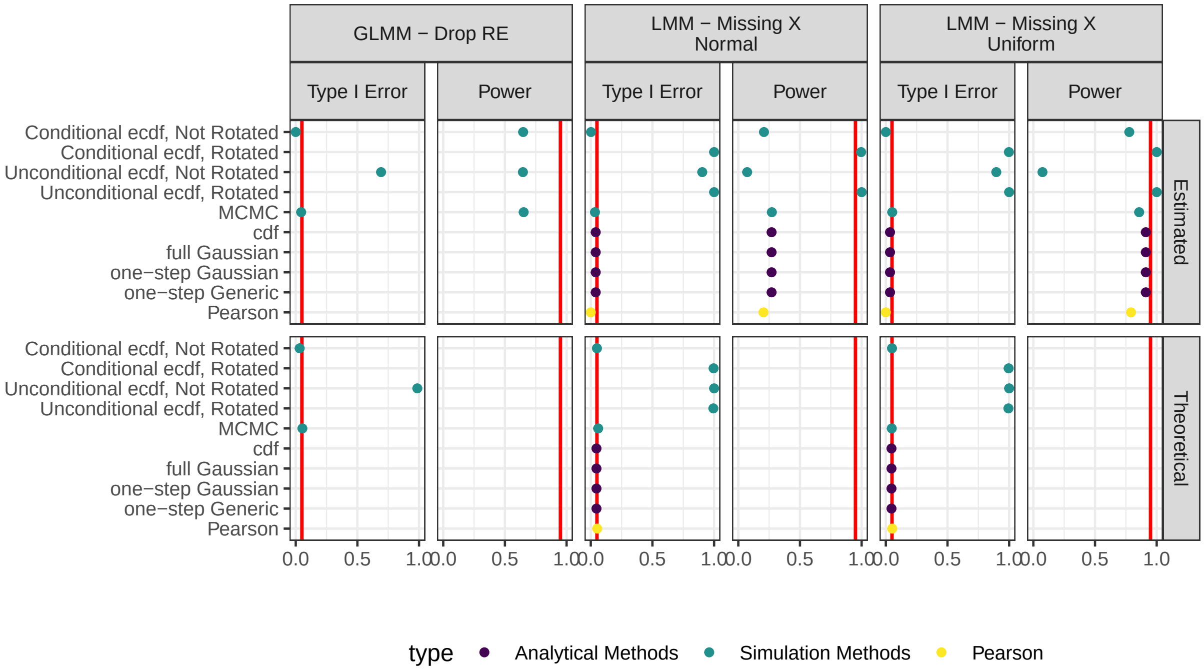


Figure 7: Simple Mixed Model. Type I error rates and Power evaluated for each analytical and simulation method. Results are partitioned out by model mis-specification (from left to right) and residual type (top to bottom).

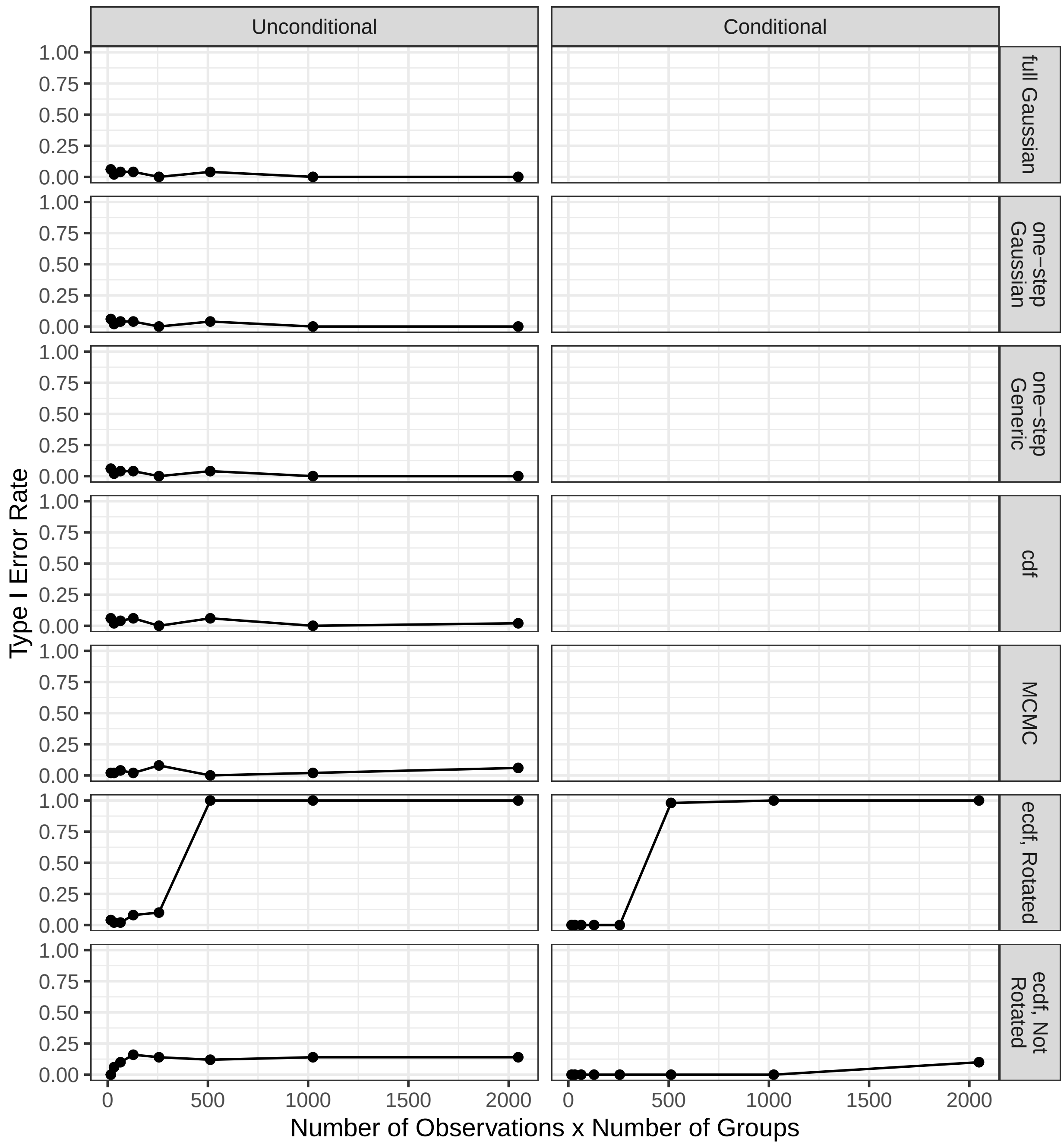


Figure 8: Simple Linear Mixed Model, mispecification = missing normally distributed coviariate. Type I Error rates for different quantile residual methods as within group sample size increases, number of groups fixed at 4. **Take home: Rotated ecdf method has high Type I Error rate when within group sample sizes are high while unroated residuals maintain a low Type I Error rate.**

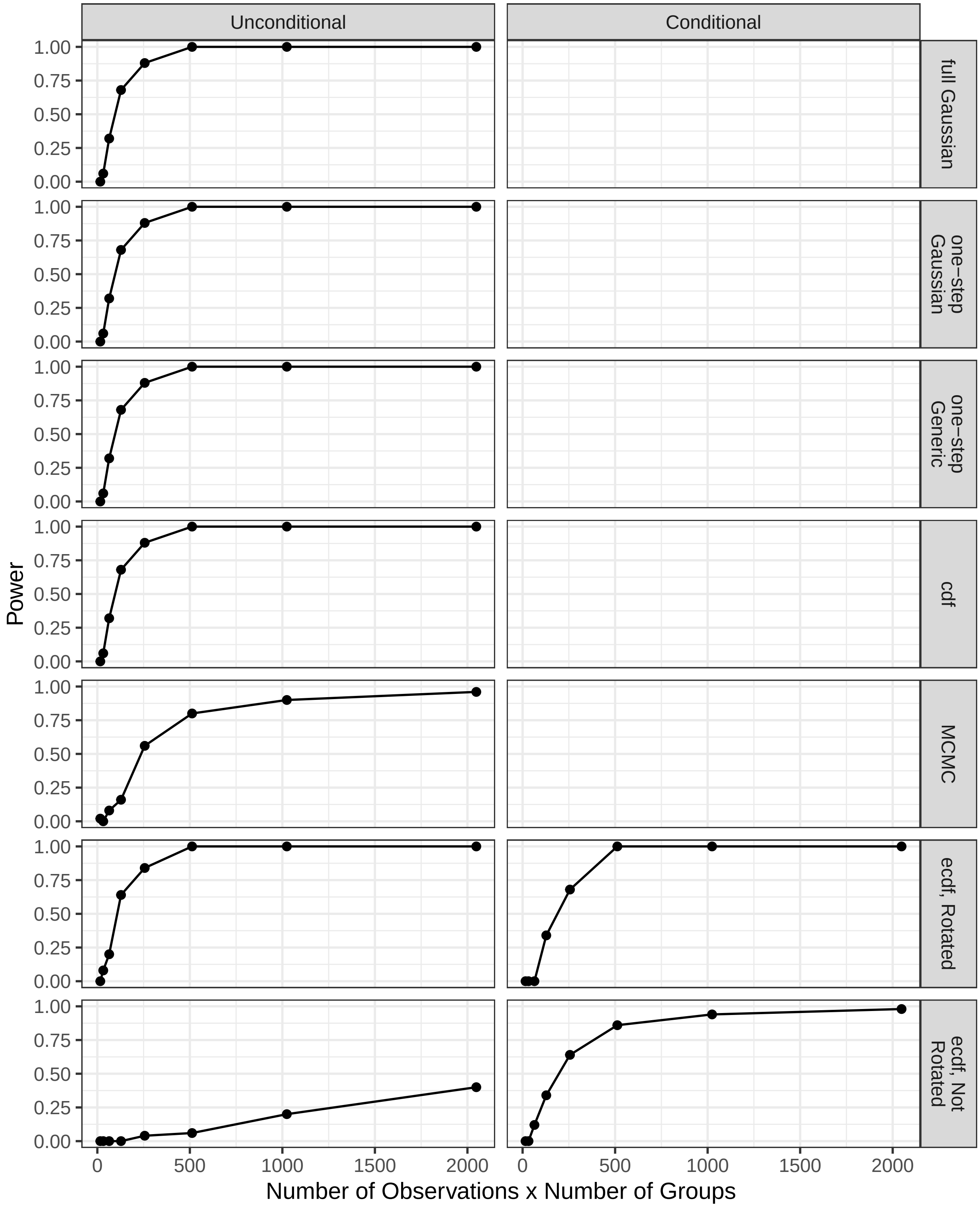


Figure 9: Simple Linear Mixed Model, mispecification = missing normally distributed coviariate. Power to detect mis-specification for different quantile residual methods as within group sample size increases, number of groups fixed at 4. **Take home: The unrotated ecdf method applied to unconditional residuals has low power to detect mis-specification.**

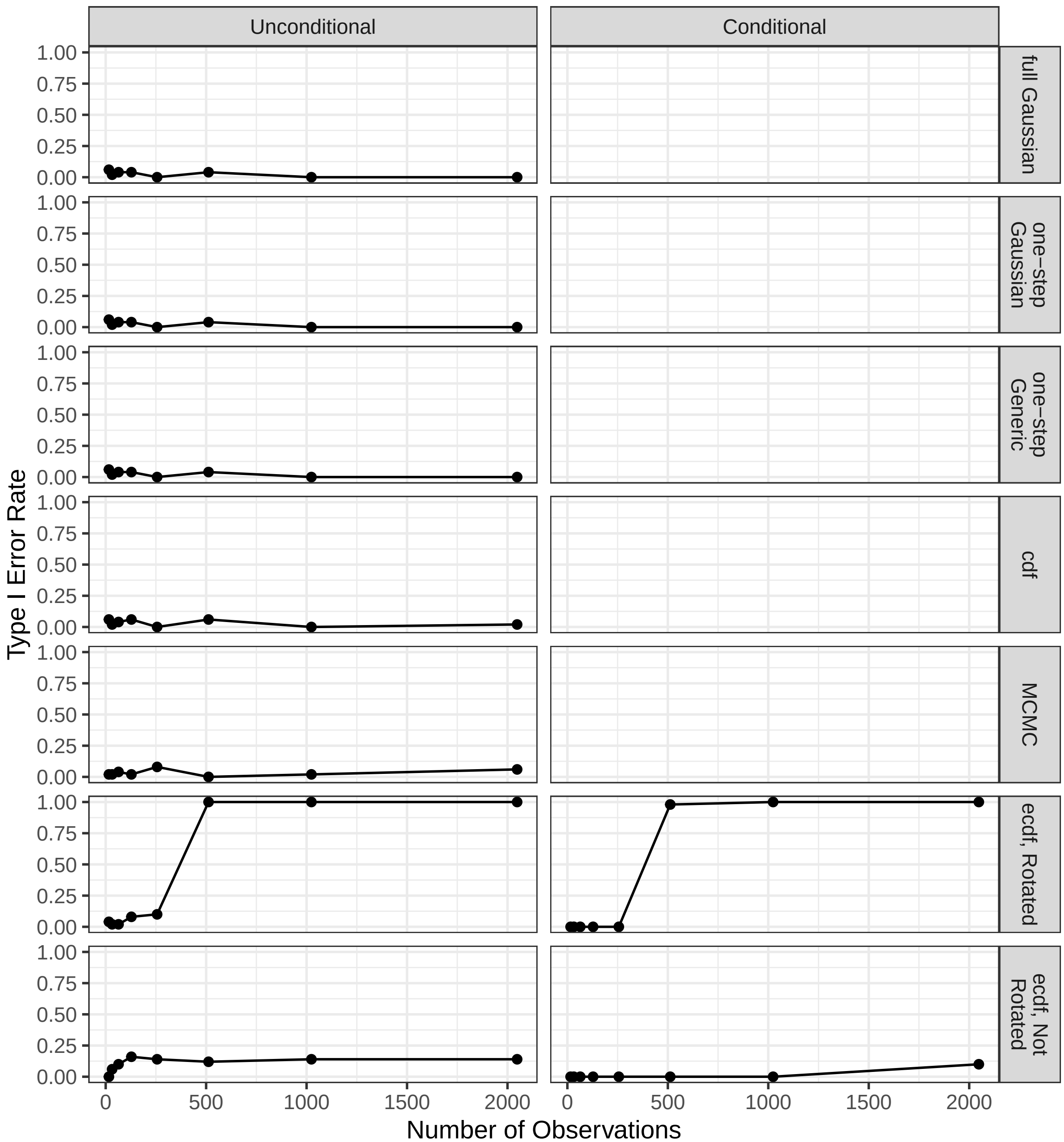


Figure 10: Simple Linear Mixed Model, mispecification = missing normally distributed coviariate. Type I Error rates for different quantile residual methods as betweem group sample size increases, number of observations within groups fixed at 8. **Take home: Results same as above with varying within group sample size.**

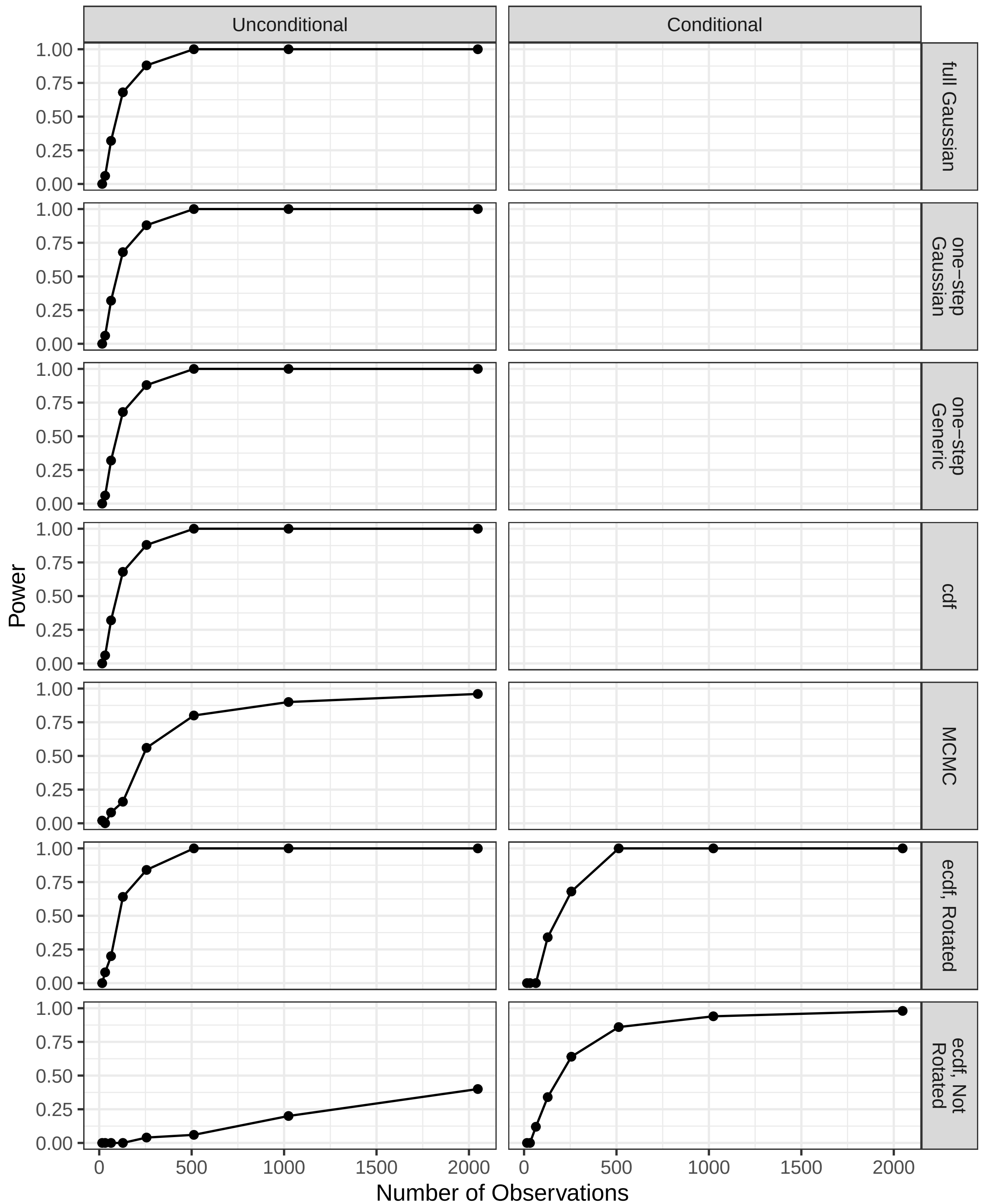


Figure 11: Simple Linear Mixed Model, mispecification = missing normally distributed coviariate. Power to detect mis-specification for different quantile residual methods as betweem group sample size increases, number of observations within groups fixed at 8. **Take home: Results same as above with varying within group sample size.**

Randomwalk

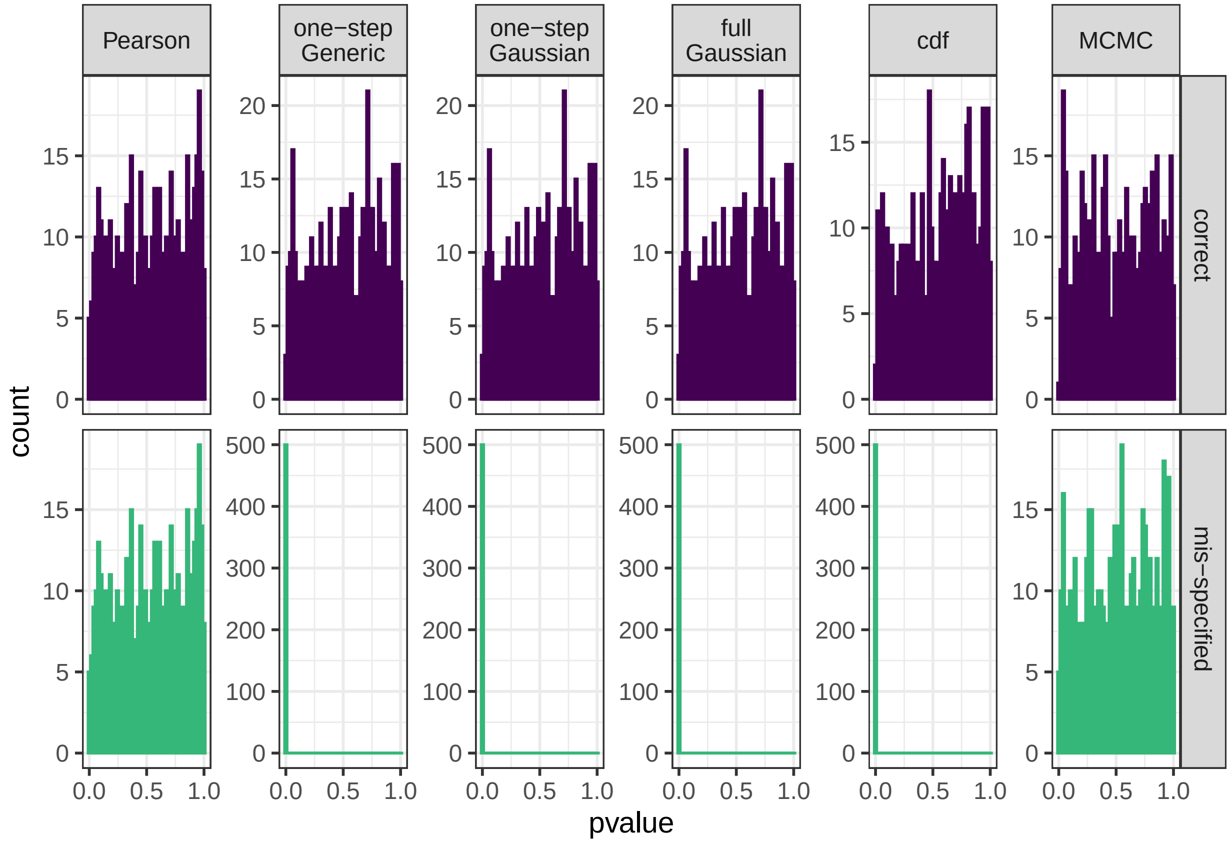


Figure 12: Randomwalk. Distribution of theoretical p-values under the correct model evaluated for each analytical method when true parameters are known. **Take home: all return approx uniform p-value distributions, including Pearson.**

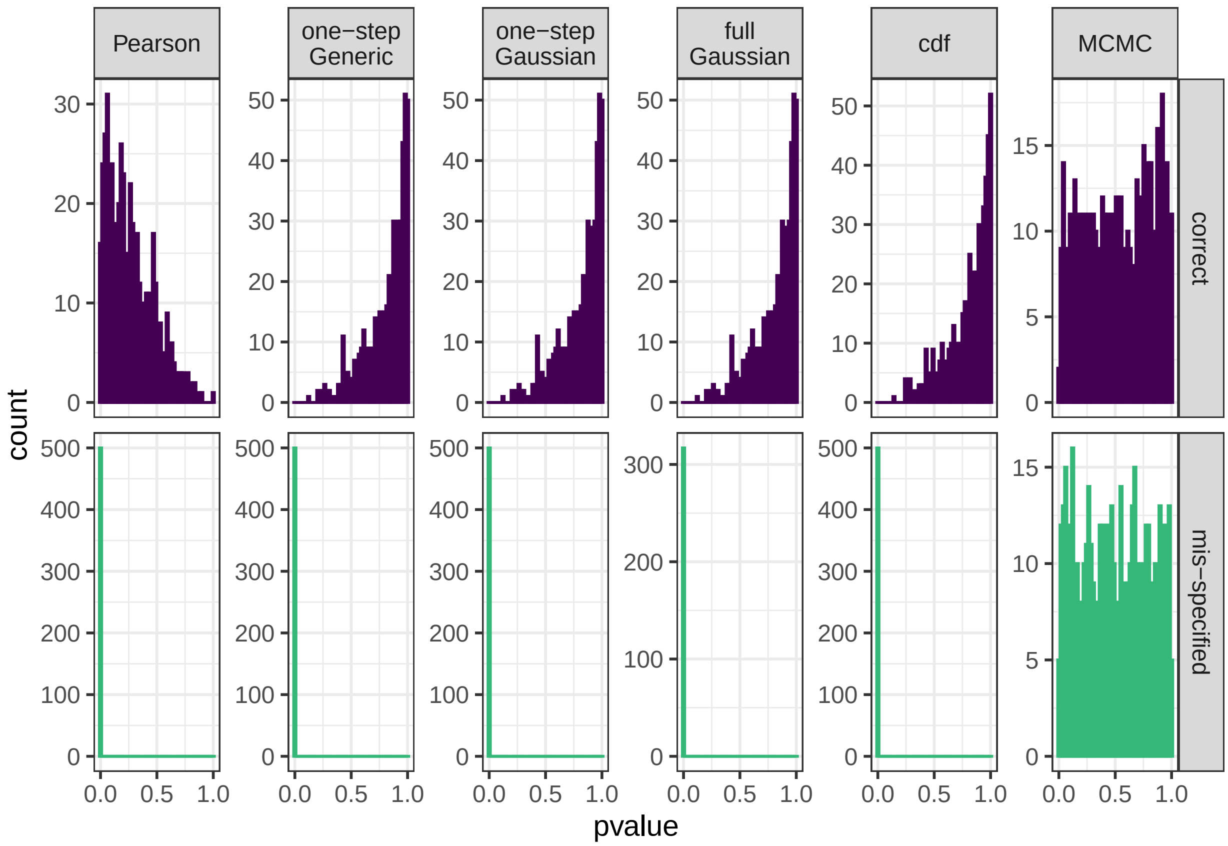


Figure 13: Randomwalk. Distribution of estimated p-values evaluated for each analytical method when parameters are estimated under the correct (top) and mis-specified (bottom) models. **Take home: Pearson rejects the correct model more than expected, MCMC fails to detect mis-specification.**

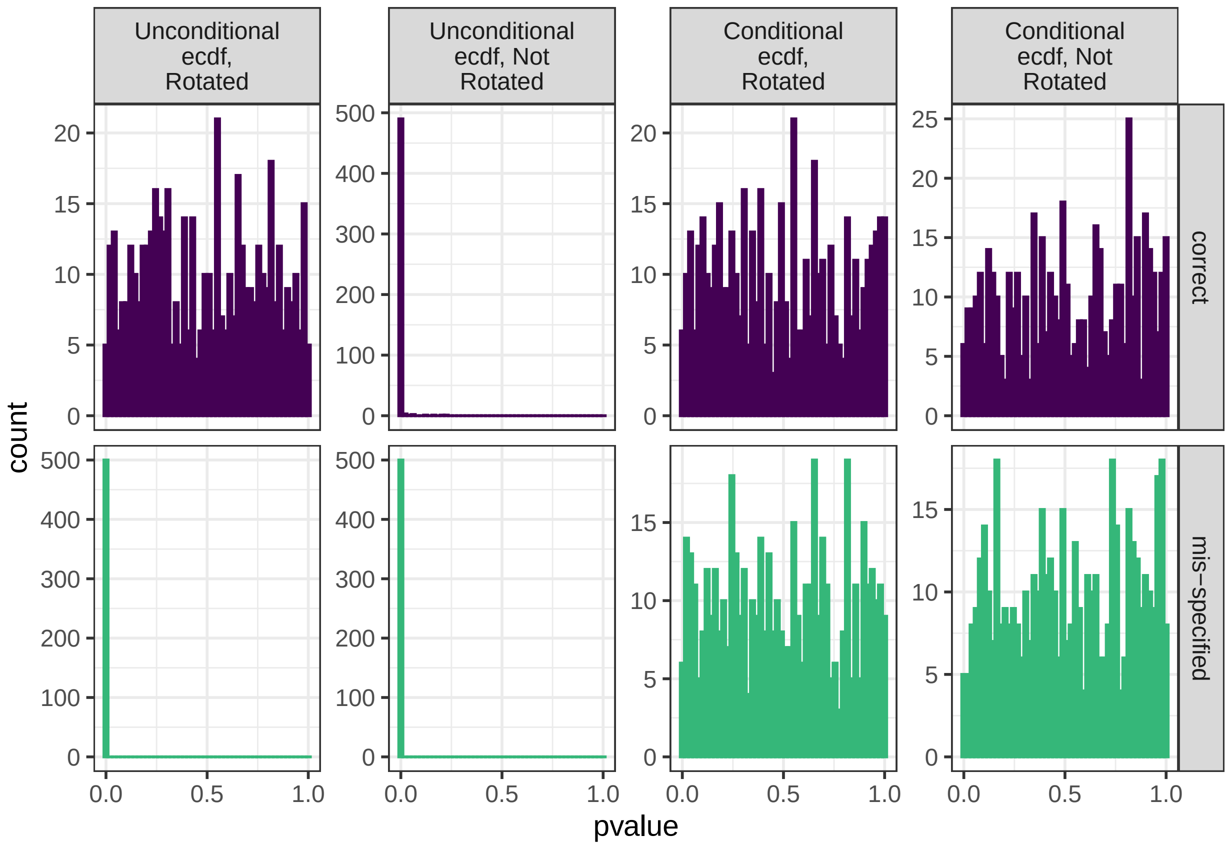


Figure 14: Randomwalk. Distribution of theoretical p-values under the correct model evaluated for each simulation-based method when true parameters are known. **Take home: unconditional ecdf needs to be rotated but rotation does not matter when conditioning on random effects.**

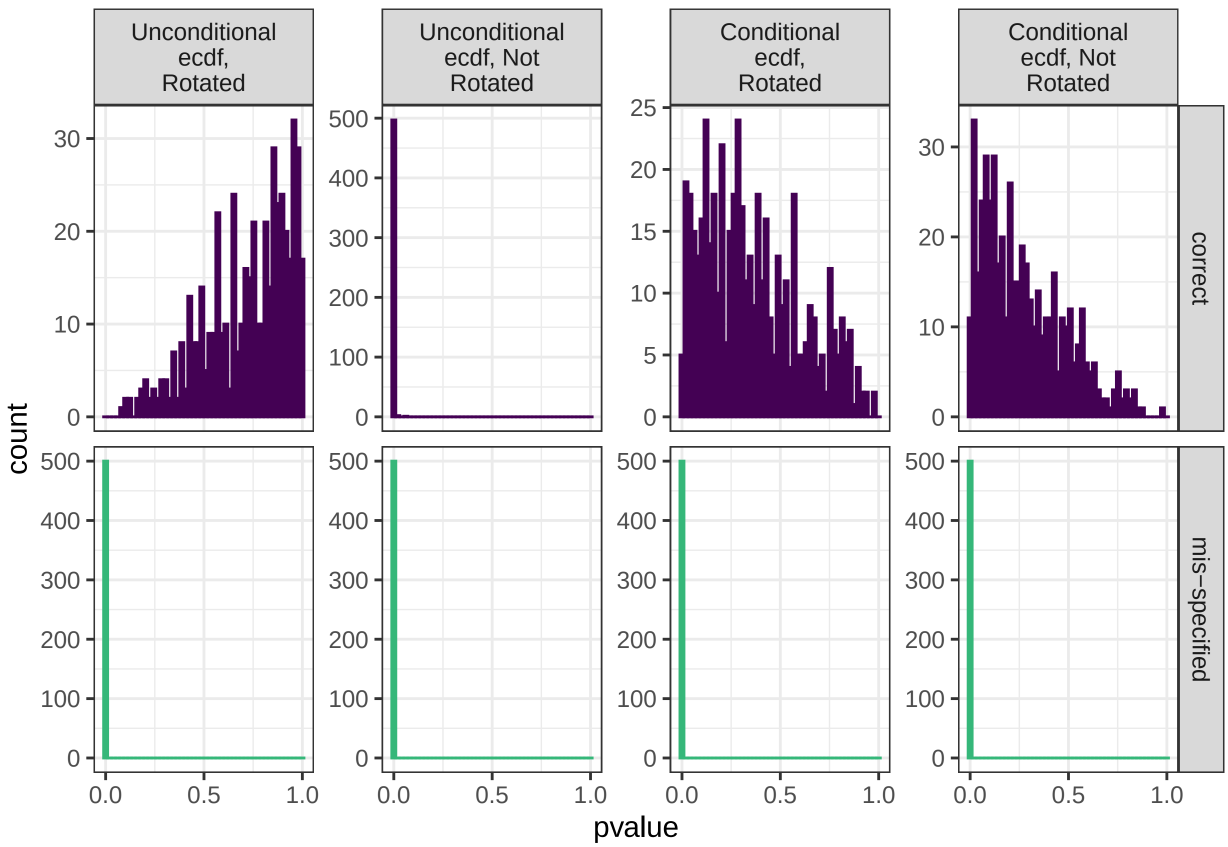


Figure 15: Randomwalk. Distribution of estimated p-values evaluated for each analytical method when parameters are estimated under the correct (top) and mis-specified (bottom) models. **Take home: unconditional ecdf needs to be rotated but rotation does not matter when conditioning on random effects. Conditioning on estimated random effects results in rejecting the model more often than expected, regardless of rotation.**

Spatial

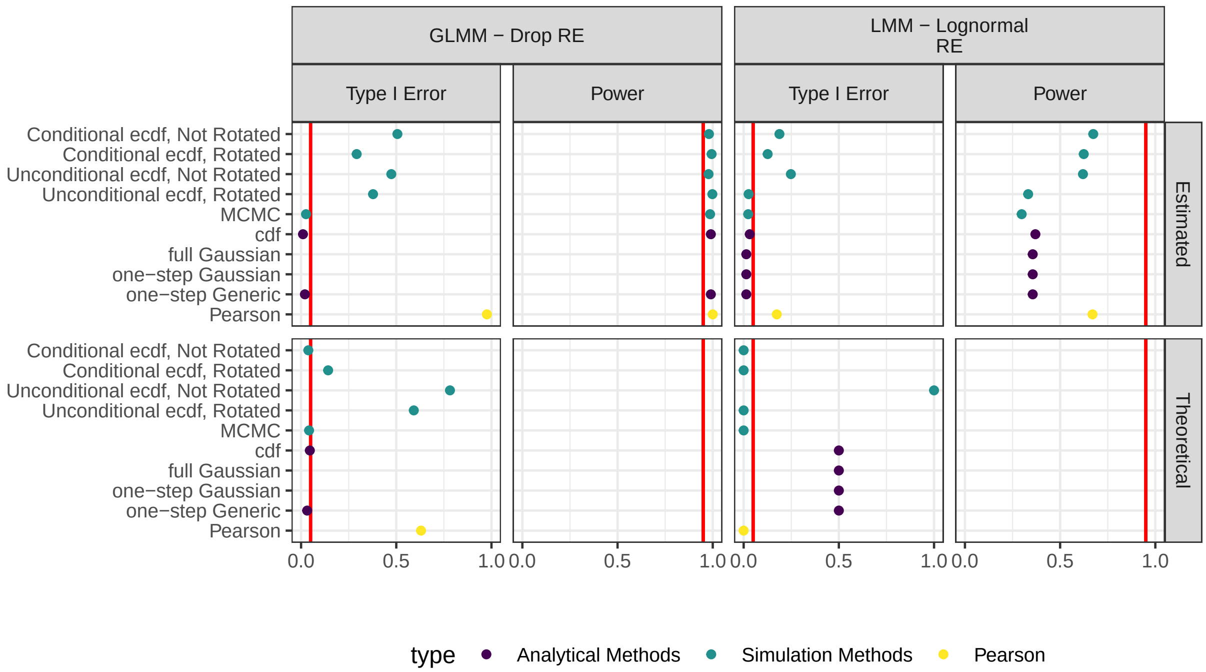
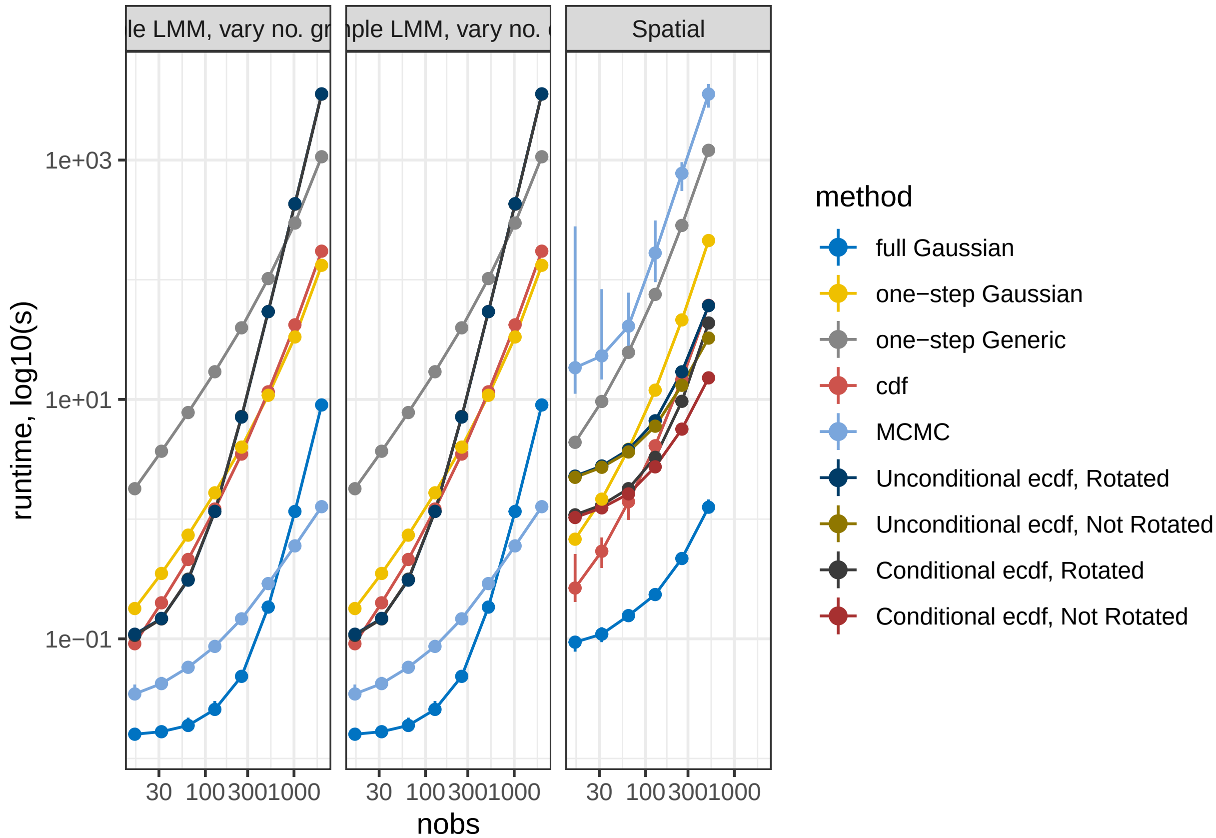
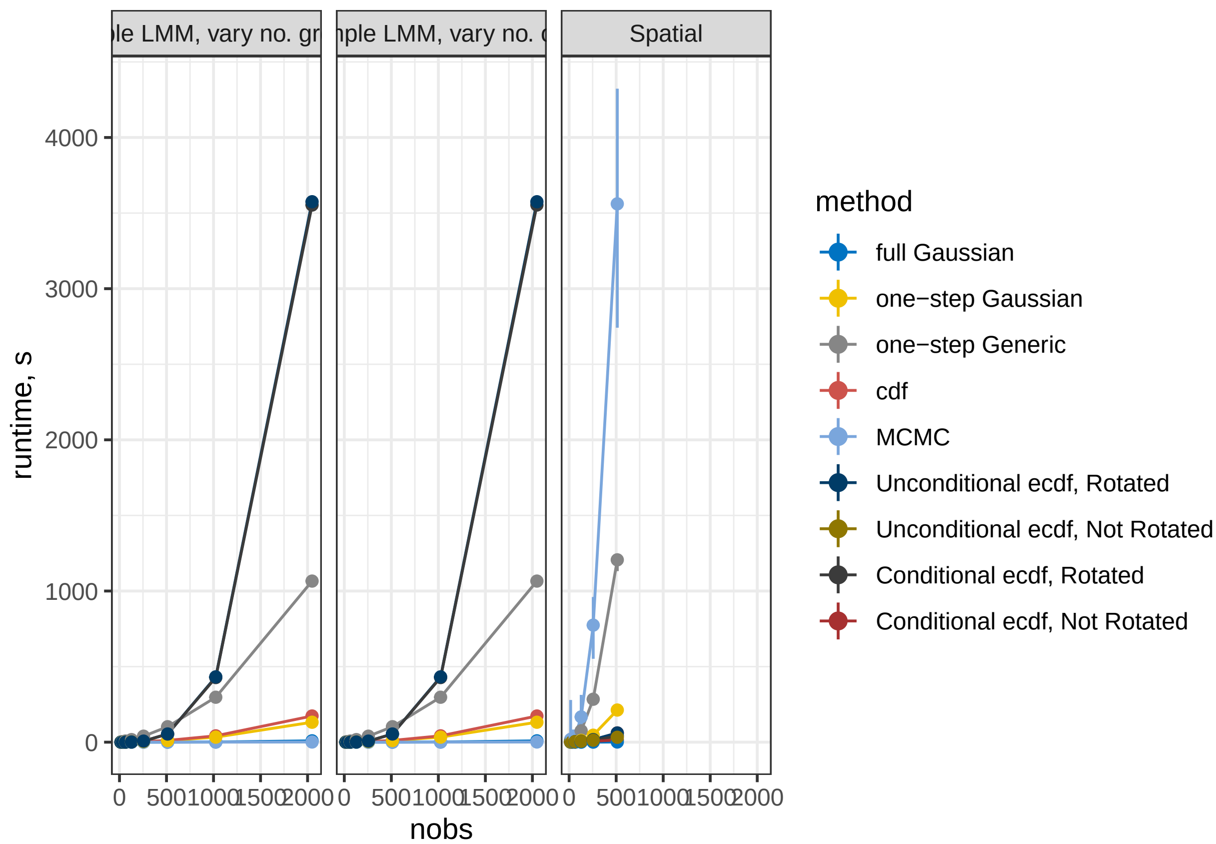


Figure 16: Simple Mixed Model. Type I error rates and Power evaluated for each analytical and simulation method. Results are partitioned out by model mis-specification (from left to right) and residual type (top to bottom).

Model Runtimes





ple LMM, vary no. gr
nple LMM, vary no. c
Spatial
runtime, s
0
1000
2000
3000
4000
0
500
1000
1500
2000
nobs
method
full Gaussian
one−step Gaussian
one−step Generic
cdf
MCMC
Unconditional ecdf, Rotated
Unconditional ecdf, Not Rotated
Conditional ecdf, Rotated
Conditional ecdf, Not Rotated