Table Placeholders

Methods

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

	Data Generating Model	Parameters	Data Fitting Model
Correct	$X_i \sim N(0,1)$ $\mu_{i,j} = X_i\beta$ $y_{i,j} \sim N(\mu_{i,j}, \sigma_y)$	$\beta = (4, -5)$ $\sigma_y = 1$	$X_i \sim N(0,1)$ $\mu_{i,j} = X_i\beta$ $y_{i,j} \sim N(\mu_{i,j}, \sigma_y)$
Mis-specified	$X_i \sim N(0,1)$ $\mu_{i,j} = X_i\beta$ $y'_{i,j} \sim N(\mu_{i,j}, exp(\sigma_y))$		$X_i \sim N(0,1)$ $\mu_{i,j} = X_i\beta$ $y'_{i,j} \sim N(\mu_{i,j}, \sigma_y)$

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

Data Generating Model	Parameters	Data Fitting Model	Data Generating Model	Parameters	Data Fitting Model
Linear Mixed Model			Generalized Linear Mixed Model		
Correct			Correct		
$X_i \sim Unif(-0.5, 0.5)$ $u_j \sim N(0, \sigma_u)$ $\mu_{i,j} = X_i\beta + u_j$ $y_{i,j} \sim N(\mu_{i,j}, \sigma_y)$	$\beta = (4, -8)$ $\sigma_u = 2$ $\sigma_y = 0.5$	$X_i \sim Unif(-0.5, 0.5)$ $u_j \sim N(0, \sigma_u)$ $\mu_{i,j} = X_i\beta + u_j$ $y_{i,j} \sim N(\mu_{i,j}, \sigma_y)$	$u_j \sim N(0, \sigma_u)$ $\mu_{i,j} = exp(\beta + u_j)$ $y_{i,j} \sim NBinom(\mu_{i,j}, size)$	$\beta = log(2)$ $size = 1$ $\sigma_u = 1$	$u_j \sim N(0, \sigma_u)$ $\mu_{i,j} = exp(\beta + u_j)$ $y_{i,j} \sim NBinom(\mu_{i,j}, size)$
Mis-specified			Mis-specified		
Correct $X_i \sim Unif(-0.5, 0.5)$ $u_j \sim N(0, \sigma_u)$ $\mu_{i,j} = X_i\beta + u_j$ $y_{i,j} \sim N(\mu_{i,j}, \sigma_y)$		Missing Random Effect $X_i \sim Unif(-0.5, 0.5)$ $\mu_{i,j} = X_i\beta$ $y_{i,j} \sim N(\mu_{i,j}, \sigma_y)$ Missing Covariate $u_j \sim N(0, \sigma_u)$ $\mu_{i,j} = \beta + u_j$ $y_{i,j} \sim N(\mu_{i,j}, \sigma_y)$	Correct $u_j \sim N(0, \sigma_u)$ $\mu_{i,j} = exp(\beta + u_j)$ $y_{i,j} \sim NBinom(\mu_{i,j}, size)$		Missing Random Effect $\mu_{i,j} = exp(\beta)$ $y_{i,j} \sim NBinom(\mu_{i,j}, size)$ Mis-specified Distribution (Poisson) $u_j \sim N(0, \sigma_u)$ $\mu_{i,j} = exp(\beta + u_j)$ $y_{i,j} \sim Poisson(\mu_{i,j})$
Lognormal Random Effect $X_i \sim Unif(-0.5, 0.5)$ $u_j \sim N(0, \sigma_u)$ $\mu'_{i,j} = X_i\beta + e^{u_j}$ $y'_{i,j} \sim N(\mu'_{i,j}, \sigma_y)$		Correct $X_i \sim Unif(-0.5, 0.5)$ $\mu_{i,j} = X_i\beta + u_j$ $y'_{i,j} \sim N(\mu_{i,j}, \sigma_y)$	Gamma Random Effect $u_j \sim Gamma(1, 1)$ $\mu'_{i,j} = exp(\beta + u_j)$ $y'_{i,j} \sim NBinom(\mu_{i,j}, size)$		Correct $u_j \sim N(0, \sigma_u)$ $\mu_{i,j} = exp(\beta + u_j)$ $y_{i,j} \sim NBinom(\mu_{i,j}, size)$

Table 3: Linear Mixed Model Simulation: data generating models, parameter values, and mis-specifications.

Data Generating Model	Parameters	Data Fitting Model	Data Generating Model	Parameters	Data Fitting Model
Linear Mixed Model			Generalized Linear Mixed Model		
Correct			Correct		
$\mu_i = u_{i-1} + a$ $u_i \sim N(\mu_i, \sigma_u)$ $y_i \sim N(u_i, \sigma_y)$	$a = 2$ $u[1] = 0$ $\sigma_u = 1$ $\sigma_y = 1$	$\mu_i = u_{i-1} + a$ $u_i \sim N(\mu_i, \sigma_u)$ $y_i \sim N(u_i, \sigma_y)$	$\mu_i = u_{i-1} + a$ $u_i \sim N(\mu_i, \sigma_u)$ $y_i \sim Gamma(\frac{1}{CV}^2, e^{u_i}CV^2)$	$a = .01$ $u[1] = 0$ $\sigma_u = 0.05$ $CV = 0.5$	$\mu_i = u_{i-1} + a$ $u_i \sim N(\mu_i, \sigma_u)$ $y_i \sim Gamma(\frac{1}{CV}^2, e^{u_i}CV^2)$
Mis-specified			Mis-specified		
Correct $\mu_i = u_{i-1} + a$ $u_i \sim N(\mu_i, \sigma_u)$ $y_i \sim N(u_i, \sigma_y)$		Missing Random Effect $y_i \sim N(a(1:n), \sigma_y)$ Missing Drift Term $\mu_i = u_{i-1}$ $u_i \sim N(\mu_i, \sigma_u)$ $y_i \sim N(u_i, \sigma_y)$	Correct $\mu_i = u_{i-1} + a$ $u_i \sim N(\mu_i, \sigma_u)$ $y_i \sim Gamma(\frac{1}{CV}^2, e^{u_i}CV^2)$		Mising Random Effect $y_i \sim Gamma(\frac{1}{CV}^2, e^{a}CV^2)$ Missing Drift Term $\mu_i = u_{i-1}$ $u_i \sim N(\mu_i, \sigma_u)$ $y_i \sim Gamma(\frac{1}{CV}^2, e^{u_i}CV^2)$
Heteroscedasticity $\mu_i = u_{i-1} + a$ $u_i \sim N(\mu_i, \sigma_u)$ $\sigma_y = \sqrt{(1:\frac{n}{2}}^{1.3}$ $y_i^{'} \sim N(u_i, \sigma_y)$		Correct $\mu_i = u_{i-1} + a$ $u_i \sim N(\mu_i, \sigma_u)$ $y_i^{'} \sim N(u_i, \sigma_y)$			Mis-specified Distribution Normal $\mu_i = u_{i-1} + a$ $u_i \sim N(\mu_i, \sigma_u)$ $y_i \sim N(u_i, \sigma_y)$
Lognormal Random Effect $\mu_i = u_{i-1} + a$ $u_i \sim N(\mu_i, \sigma_u)$ $y_i^{'} \sim N(e^{u_i}, \sigma_y)$		Correct $\mu_i = u_{i-1} + a$ $u_i \sim N(\mu_i, \sigma_u)$ $y_i^{'} \sim N(u_i, \sigma_y)$	Gamma Random Effect $u_i = u_{i-1} + Gamma(0.5, 20)$ $y_i^{'} \sim Gamma(\frac{1}{CV}^2, e^{u_i}CV^2)$		Correct $\mu_i = u_{i-1} + a$ $u_i \sim N(\mu_i, \sigma_u)$ $y_i \sim Gamma(\frac{1}{CV}^2, e^{u_i}CV^2)$

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

Data Generating Model	Parameters	Data Fitting Model	Data Generating Model	Parameters	Data Fitting Model
Linear Mixed Model			Generalized Linear Mixed Model		
Correct			Correct		
$\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = \beta_0 + \omega_i$ $y \sim N(\eta, \sigma_y)$	$\beta_0 = 4$ $\sigma_y^2 = 1$ $\phi = 50$ $\kappa = \sqrt{8}/\phi$ $\sigma_\omega^2 = 1$	$\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = \beta_0 + \omega_i$ $y \sim N(\eta, \sigma_y)$	$\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = exp(\beta_0 + \omega_i)$ $y \sim Pois(\eta, \sigma_y)$	$\beta = 0.5$ $\phi = 50$ $\kappa = \sqrt{8}/\phi$ $\sigma_\omega^2 = 0.25$	$\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = exp(\beta_0 + \omega_i)$ $y \sim Pois(\eta, \sigma_y)$
Mis-specified			Mis-specified		
Correct $\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = \beta_0 + \omega_i$ $y \sim N(\eta, \sigma_y)$		Missing Random Effect $\eta_i = \beta_0$ $y \sim N(\eta, \sigma_y)$ Lognormal Error $\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = \beta_0 + \omega_i$ $\sigma_y = exp(N(0,1))$ $y \sim N(\eta, \sigma_y)$	Correct $\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = exp(\beta_0 + \omega_i)$ $y \sim Pois(\eta)$		Missing Random Effect $\eta_i = exp(\beta_0)$ $y \sim Pois(\eta)$
Lognormal Random Effect $\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = \beta_0 + exp(\omega_i)$ $y^{'} \sim N(\eta, \sigma_y)$		Correct $\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = \beta_0 + \omega_i$ $y^{'} \sim N(\eta, \sigma_y)$	Zero-Inflated Poisson $\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = exp(\beta_0 + \omega_i)$ $y^{'} \sim B(1, 0.7) * Pois(\eta)$		Correct $\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = exp(\beta_0 + \omega_i)$ $y \sim Pois(\eta)$
			Lognormal Random Effect $\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = exp(\beta_0 + exp(\omega_i))$ $y^{'} \sim Pois(\eta)$		Correct $\omega \sim GMRF(Q[\kappa, \sigma_\omega^2])$ $\eta_i = exp(\beta_0 + \omega_i)$ $y^{'} \sim Pois(\eta)$

Results

Linear Model

Table 5: Linear Model. Type I error rates and Power evaluated for each analytical and simulation method for theoretical residuals using the KS normality test. Results are partitioned out by residual type (top to bottom).

	Overdispersion	
method	Type I Error	Power
Pearson	0.045	1
one-step Generic	0.045	1
one-step Gaussian	0.045	1
full Gaussian	0.045	1
cdf	0.045	1
Unconditional ecdf, Not Rotated	0.034	1
Conditional ecdf, Not Rotated	0.034	1

Table 6: Linear Model. Type I error rates and Power evaluated for each analytical and simulation method for estimated residuals using the KS normality test. Results are partitioned out by residual type (top to bottom).

	Overdispersion	
method	Type I Error	Power
Pearson	0	0.959
one-step Generic	0	0.959
one-step Gaussian	0	0.959
full Gaussian	0	0.959
cdf	0	0.959
Unconditional ecdf, Not Rotated	0	0.961
Conditional ecdf, Not Rotated	0	0.961

Mixed Model

Table 7: Mixed Model. Type I error rates and Power evaluated for each analytical and simulation method for theoretical residuals using the KS normality test. Results are partitioned out by model mis-specification (from left to right) and residual type (top to bottom).

	A: Missing RE		B: Missing X		C: Misp RE		
method	Type I Error	Power	Type I Error	Power	Type I Error	Power	NA
LMM	Pearson	0.045	1	0.045	1.000	0.045	0.995
LMM	one-step Generic	0.041	1	0.041	1.000	0.041	0.045
LMM	one-step Gaussian	0.041	1	0.041	1.000	0.041	0.045
LMM	full Gaussian	0.041	1	0.041	1.000	0.041	0.045
LMM	cdf	0.043	1	0.043	1.000	0.043	0.048
LMM	MCMC	0.032	1	0.032	1.000	0.032	0.054
LMM	Unconditional ecdf, Rotated	0.052	1	0.052	1.000	0.052	0.057
LMM	Unconditional ecdf, Not Rotated	0.905	1	0.905	0.821	0.905	1.000
LMM	Conditional ecdf, Rotated	0.041	1	0.041	1.000	0.041	0.998
LMM	Conditional ecdf, Not Rotated	0.029	1	0.029	1.000	0.029	0.995

Table 8: Mixed Model. Type I error rates and Power evaluated for each analytical and simulation method for estimated residuals using the KS normality test. Results are partitioned out by model mis-specification (from left to right) and residual type (top to bottom).

	A: Missing RE		B: Missing X		C: Misp RE		
method	Type I Error	Power	Type I Error	Power	Type I Error	Power	NA
LMM	Pearson	0.000	0.449	0.000	0.159	0.000	0.000
LMM	one-step Generic	0.032	0.449	0.032	0.324	0.032	0.016
LMM	one-step Gaussian	0.032	0.449	0.032	0.324	0.032	0.016
LMM	full Gaussian	0.032	0.449	0.032	0.324	0.032	0.016
LMM	cdf	0.032	0.449	0.032	0.324	0.032	0.016
LMM	MCMC	0.039	0.449	0.039	0.186	0.039	0.034
LMM	Unconditional ecdf, Rotated	0.034	0.372	0.034	0.315	0.034	0.016
LMM	Unconditional ecdf, Not Rotated	0.456	0.460	0.456	0.002	0.456	0.163
LMM	Conditional ecdf, Rotated	0.002	0.381	0.002	0.093	0.002	0.002
LMM	Conditional ecdf, Not Rotated	0.000	0.442	0.000	0.159	0.000	0.000

Randomwalk

Table 9: Randomwalk Model. Type I error rates and Power evaluated for each analytical and simulation method for theoretical residuals using the KS normality test. Results are partitioned out by model misspecification (from left to right) and residual type (top to bottom).

	A: Missing RE		B: Heterosced.		D: Misp RE		
method	Type I Error	Power	Type I Error	Power	Type I Error	Power	NA
LMM	Pearson	0.059	1	0.059	1.000	0.059	1
LMM	one-step Generic	0.036	1	0.036	1.000	0.036	1
LMM	one-step Gaussian	0.036	1	0.036	1.000	0.036	1
LMM	full Gaussian	0.036	1	0.036	1.000	0.036	1
LMM	cdf	0.034	1	0.034	1.000	0.034	1
LMM	MCMC	0.039	1	0.039	1.000	0.039	1
LMM	Unconditional ecdf, Rotated	0.045	1	0.045	1.000	0.045	1
LMM	Unconditional ecdf, Not Rotated	0.989	1	0.989	0.764	0.989	1
LMM	Conditional ecdf, Rotated	0.061	1	0.061	1.000	0.061	1
LMM	Conditional ecdf, Not Rotated	0.057	1	0.057	1.000	0.057	1

Table 10: Randomwalk Model. Type I error rates and Power evaluated for each analytical and simulation method for estimated residuals using the KS normality test. Results are partitioned out by model misspecification (from left to right) and residual type (top to bottom).

	A: Missing RE		B: Heterosced.		D: Misp RE		
method	Type I Error	Power	Type I Error	Power	Type I Error	Power	NA
LMM	Pearson	0.135	1	0.135	0.000	0.135	1.000
LMM	one-step Generic	0.000	1	0.000	0.016	0.000	0.979
LMM	one-step Gaussian	0.000	1	0.000	0.016	0.000	0.979
LMM	full Gaussian	0.000	1	0.000	0.017	0.000	0.984
LMM	cdf	0.000	1	0.000	0.016	0.000	1.000
LMM	MCMC	0.060	1	0.060	0.472	0.060	0.037
LMM	Unconditional ecdf, Rotated	0.000	1	0.000	0.283	0.000	0.882
LMM	Unconditional ecdf, Not Rotated	0.998	1	0.998	0.773	0.998	1.000
LMM	Conditional ecdf, Rotated	0.081	1	0.081	0.000	0.081	1.000
LMM	Conditional ecdf, Not Rotated	0.128	1	0.128	0.000	0.128	1.000

Spatial

Table 11: Spatial Model. Type I error rates and Power evaluated for each analytical and simulation method for theoretical residuals using the KS normality test. Results are partitioned out by model mis-specification (from left to right) and residual type (top to bottom).

	A: Missing RE		B: Lognorm error		C: Misp RE		
method	Type I Error	Power	Type I Error	Power	Type I Error	Power	NA
LMM	Pearson	0.050	0.755	0.050	1	0.050	1.000
LMM	one-step Generic	0.011	0.755	0.011	1	0.011	0.388
LMM	one-step Gaussian	0.011	0.755	0.011	1	0.011	0.388
LMM	full Gaussian	0.011	0.755	0.011	1	0.011	0.388
LMM	cdf	0.011	0.755	0.011	1	0.011	0.390
LMM	MCMC	0.036	0.755	0.036	1	0.036	0.095
LMM	Unconditional ecdf, Rotated	0.014	0.789	0.014	1	0.014	0.438
LMM	Unconditional ecdf, Not Rotated	0.560	0.751	0.560	1	0.560	0.995
LMM	Conditional ecdf, Rotated	0.066	0.800	0.066	1	0.066	1.000
LMM	Conditional ecdf, Not Rotated	0.045	0.762	0.045	1	0.045	1.000

Table 12: Spatial Model. Type I error rates and Power evaluated for each analytical and simulation method for estimated residuals using the KS normality test. Results are partitioned out by model mis-specification (from left to right) and residual type (top to bottom).

	A: Missing RE		B: Lognorm error		C: Misp RE		
method	Type I Error	Power	Type I Error	Power	Type I Error	Power	NA
LMM	Pearson	0.082	0	0.082	1.000	0.082	0.360
LMM	one-step Generic	0.005	0	NA	NA	NA	NA
LMM	one-step Gaussian	0.005	0	0.005	1.000	0.005	0.167
LMM	full Gaussian	0.005	0	0.005	1.000	0.005	0.213
LMM	cdf	0.027	0	0.027	1.000	0.027	0.200
LMM	MCMC	0.039	0	0.039	1.000	0.039	0.155
LMM	Unconditional ecdf, Rotated	0.007	0	0.007	0.991	0.007	0.116
LMM	Unconditional ecdf, Not Rotated	0.139	0	0.139	1.000	0.139	0.360
LMM	Conditional ecdf, Rotated	0.078	0	0.078	0.998	0.078	0.292
LMM	Conditional ecdf, Not Rotated	0.084	0	0.084	1.000	0.084	0.364

Table 13: Overview of issues and recommendations for common classes of models. Correlation and distributions refer to predicted data from a fitted model, against which observed points are compared. A linear rotation refers to a multiplication of the simulated and observed data by a Cholesky decomposition of the estimated covariance matrix of the observed data, z'=Lz, as available in DHARMa.

Model class	Case studies	Issues and causes	Recommendation
Linear model	Linear model	No issues	Pearson residuals
Generalized linear model (GLM)	Skewed Gamma	Non-normality resulting from response variable. Quantile residuals are needed if not approximately normal.	Quantile residual
Linear mixed model (LMM), Multivariate model	Random walk, Spatial LMM, Multinomial	Linear correlations caused by non-independence in observations.	Use a method that linearly decorrelates in order to transform to a unit iid normal. OSA Full Gaussian, OSA one-step Gaussian, or simulation residuals with rotation.
Generalized linear mixed model (GLMM)	Spatial Poisson, Repeated measures Tweedie	Non-normality and non-linear correlations caused by response variable and non-independence in observations.	Needs non-linear decorrelation and quantiles. Needs non-linear decorrelation and quantiles. Best approach depends on case study and sample size.