CHAPTER 4

SIMULATION STUDY RESULTS AND DISCUSSION

4.1 Fitting the Models to the Simulated Data

Parameter estimates were obtained for models under 29 different covariance structures, 25 of which had valid parameter estimates and four of which had invalid parameter estimates (Table 3.1). These parameter estimates were used to fit models to the simulated data sets using SAS PROC MIXED (ver. 9.1) (Appendix C2). Under each covariance specification, including a model with invalid covariance parameter estimates, 250 data sets were simulated, resulting in 26 different covariance structure specifications. The three other models with invalid covariance parameter estimates could not be considered as their random effects covariance matrix was not positive definite. All of the models under consideration were fitted to each of the simulated data sets. Parameter estimates and their confidence intervals were extracted for the fixed effects of the model, as well as the AIC, BIC and AICc each time a model was fitted to a simulated data set.

Convergence did not occur in all cases; therefore some models had fewer fitted results than others. For data sets modelled with UN error covariance structure with no random effects, I investigated the fitted model under $\boldsymbol{\omega}_i$ = CS and Σ = CSH, which was only successfully fitted to 139 out of the 250 data sets. In order to determine if nonconvergence was due to the number of iterations or if it was true nonconvergence, I increased the number of iterations from 50 to 50000 for one of the simulated data sets. The model still did not converge for this data. The convergence criterion gradually reached a constant value, greater than the required level, remaining at this value after each iteration, and therefore would not meet the convergence requirement. The cases where this model was successfully fitted were also investigated, and I found that in these cases non-valid values for the variance components were obtained. This fitted model performed very poorly under all the simulated models, only being successfully fitted to 3063 out of a total of 6500 data sets (Table 4.1).

Table 4.1: Summary of models fitted to the simulated data, where a (-) indicates a model which could not be used to simulate data as the random effects covariance was not positive-definite.

Assumed Covariance Structures		**Number Converged**	**Percentage Converged**	**Percentage Converged when model correct**	**Percentage Converged when model not correct**
ω_i	Σ				
VC	None	6500	100.00%	100.00%	100.00%
	Intercept only	6500	100.00%	100.00%	100.00%
	VC	6500	100.00%	100.00%	100.00%
	CS	6500	100.00%	100.00%	100.00%
	CSH	6329	97.40%	100.00%	97.26%
	ARH(1)	6391	98.32%	100.00%	98.26%
	UN	6500	100.00%	100.00%	100.00%
CS	None	6500	100.00%	100.00%	100.00%
	CSH	3063	47.12%	54.00%	46.85%
	TOEP	5470	84.15%	-	84.15%
CSH	None	6500	100.00%	100.00%	100.00%
	CSH	3336	51.32%	49.60%	51.39%
	ARH(1)	3339	51.37%	53.20%	51.30%
	UN	3705	57.00%	63.60%	56.74%
AR(1)	None	6500	100.00%	100.00%	100.00%
	Int. only	6500	100.00%	100.00%	100.00%
	VC	6500	100.00%	100.00%	100.00%
	CSH	6038	92.89%	98.40%	92.67%
	ARH(1)	6041	92.94%	99.20%	92.69%
	UN	6500	100.00%	100.00%	100.00%
ARH(1)	None	6500	100.00%	100.00%	100.00%
	VC	6365	97.92%	94.80%	98.05%
	CSH	5540	85.23%	79.20%	85.47%
	ARH(1)	5554	85.45%	80.00%	85.66%
	UN	5771	88.78%	76.80%	89.26%
TOEP	None	6500	100.00%	100.00%	100.00%
	CS	6091	93.71%	-	93.71%
	TOEP	6284	96.68%	-	96.68%
UN	None	6500	100.00%	100.00%	100.00%

For a model to be considered as a robust model, it needed to be successfully fitted to all data sets under each simulated covariance structures. Non-convergence or invalid parameter estimates is generally an indicator of problems with the parameterisation of the model (Verbeke & Molenberghs, 2000), or failure of the optimisation procedure, therefore not all models are appropriate under all conditions. Those combinations of covariance structures which did not fit all data sets were therefore not considered as robust models. In some cases (see Appendix A1) the model from which a data set was generated was not successfully fitted to the data, notably the model under $\boldsymbol{\omega}_{\boldsymbol{i}}$ = CSH and $\boldsymbol{\Sigma}$ = CSH, which was only fitted successfully to 124 out of 250 data sets generated from the same model. Models fitted to the data which had non-positive definite random effects covariance matrices estimated for the original data set, namely models with $\boldsymbol{\omega}_{\boldsymbol{i}}$ = CS and $\boldsymbol{\Sigma}$ = TOEP, with $\boldsymbol{\omega}_{\boldsymbol{i}}$ = TOEP and $\boldsymbol{\Sigma}$ = CS, and with $\boldsymbol{\omega}_{\boldsymbol{i}}$ = TOEP and $\boldsymbol{\Sigma}$ = TOEP, were investigated. It was found that although these models showed a relatively high percentage of convergence, the random effects covariance structures estimated were non-positive definite. The model which obtained a zero estimate for the diagonal CS parameter of the random effects covariance matrix when fitted to the original data set (model with $\boldsymbol{\omega}_{\boldsymbol{i}}$ = VC and $\boldsymbol{\Sigma}$ = CS), obtained estimates for all data sets. On investigation of the covariance parameters estimated, it was found that the diagonal CS parameter of the random effects covariance matrix was still estimated as zero, resulting in equal estimates for all variances and covariances of the random effects covariance matrix.

The simpler models tended to fit all of the simulated data sets (Table 4.1). The models that did fit all datasets successfully were the models fitted without random effects; the two random intercept models with fitted error covariance structures $\boldsymbol{\omega}_{\boldsymbol{i}}$ = VC and $\boldsymbol{\omega}_{\boldsymbol{i}}$ =

AR(1); all random intercept and slope models where $\boldsymbol{\omega}_i$ = VC, except when heterogeneous random effects covariance structures were fitted; and random intercept and slope models where $\boldsymbol{\omega}_i$ = AR(1) and Σ = UN and where $\boldsymbol{\omega}_i$ = AR(1) and Σ = VC. These models which were successfully fitted either had five or less parameters, or were models without random effects.

4.2 Criteria for Selecting the Best Model

The AIC, BIC and AICc were extracted for each model fitted. These are the criteria recommended for choosing between models with different covariance structures (Wolfinger, 1993; Davis, 2002; Demidenko, 2004). The mean, standard error, minimum and maximum of the information criteria were calculated for each fitted model for each set of simulated data. For each simulated data set, it was noted which models obtained AIC, BIC, and AICc values within two units of the minimum (the threshold value for differences between AIC values of the best fitting models specified by Duong (1984), Burnham and Anderson (2002) and Jones (1993)) for the models fitted to the data set. This information was then summarised across all of the data sets that were simulated from the same model. For the data sets simulated under each model, for each of the different models fitted, the percentage of occurrences where the model was within two units of the minimum AIC, BIC or AICc was calculated. These results are summarised and discussed in Section 4.3, with full results appearing in Appendix A1 (summary measures), and A2 (percentage within two units of the minimum values).

To determine the robustness of the linear mixed effect model against misspecification of the covariance structure, the method described in Verbeke and Lesaffre (1997) and Jacqmin-Gadda *et al.* (2007) was used. In these simulation studies the authors investigated the robustness of the maximum likelihood estimator of fixed effects from a linear mixed model when the error distribution was misspecified. In order to determine the robustness of the estimates, the authors used the coverage rates for the 95% confidence interval of the fixed parameter estimates. The proportion, $\hat{\pi}$, of 95% confidence intervals containing the true population parameter could then be calculated along with a confidence interval for π, the true proportion of confidence intervals that contain the true population parameter. I used the Wilson score confidence interval for π, as recommended by Brown, Cai and DasGupta (2001), because the standard interval for percentages using the normal approximation behaves very poorly when percentages are close to 0% or 100%, leading to confidence intervals that exceed 100% or that are less than 0% (Wilson, 1927). The Wilson confidence interval is defined as

$$\mathrm{CL_W} = \tilde{\pi} \pm \frac{zn^{1/2}}{n+z^2}\left(\hat{\pi}(1-\hat{\pi}) + \frac{z^2}{4n}\right)^{1/2}$$

where $\pi = \varsigma / n$, $\tilde{\pi} = \tilde{\varsigma} / \tilde{n}$, $\tilde{\varsigma} = \varsigma + z^2 / 2$, and $\tilde{n} = n + z^2$. ς is the number of data sets where the true parameter value was contained within the confidence interval, *n* is the total number of data sets, and *z* is the $100(1\text{-}\alpha/2)^{\text{th}}$ percentile of the standard normal distribution. When calculating the confidence interval for a set of simulated data, where there are 250 data sets, the Wilson confidence intervals and those obtained using the normal approximation to the binomial are quite different, where the upper limit of the normal confidence interval often exceeds 100%. The Wilson confidence limits when calculated for all the simulated data (i.e. 6500 data sets) are very similar to those obtained by the normal approximation. When estimated values are obtained for the full set of simulated data, if an estimated coverage probability falls outside the limits of between 94% and 96%, then the confidence interval of the estimated proportion will not contain 95%, and so the estimated value for the coverage probability will be significantly different from 95%. These results are summarised and discussed in Section 4.4 and the full results for each simulated model are presented in Appendix A3, which includes the Wilson confidence interval for each estimated coverage probability.

4.3 Analysis of Information Criteria

The information criteria, AIC, BIC and AICc, were used in two ways to determine which models fit the simulated data best. Firstly, for each simulated model, the mean AIC, BIC and AICc values were calculated for each fitted model (Appendix A1). The standard errors for these means were also calculated, and in general, were close to one for most models. To summarise these values into one value for each fitted model, the overall mean AIC, BIC and AICc values across all simulated models were calculated for each fitted model. Models with overall AIC, BIC or AICc values within two units of the minimum mean values, and which were successfully fitted to all simulated data sets, were considered as generally good fitting models. A summary of the AIC, BIC and AICc values for the models successfully fitted to all the simulated data sets appears in Table 4.2.

When summarised over all simulated data sets, the standard error of the mean information criteria was 0.22 to 0.23, except for the OLS model which had a standard error of 0.27. The overall mean AIC, BIC, and AICc values for all fitted models were 444.29, 444.66, and 448.96 respectively, the overall minimum and maximum AIC values were 432.12 and 492.05 respectively, the overall minimum and maximum BIC values were 437.30 and 505.00 respectively, and the overall minimum and maximum AICc values were 432.12 and 494.41 respectively.

The models that obtained low mean AIC, BIC and AICc values tended to be those models with less complex error covariance structures, such as VC, CS, AR(1) or TOEP, and consequently fewer covariance parameters (six or less). To demonstrate this, the mean information criteria were plotted against the number of covariance parameters in the fitted models (Fig. 4.1). Kruskal-Wallis non-parametric tests were calculated to accompany this plot, and show that models with more complex error covariance structures (i.e. UN, CSH or ARH(1) structures) obtained significantly higher mean information criteria at the 5% level of significance, but not at the 1% level. The models with heterogeneous covariance parameters, as well as for unstructured covariance matrices, obtained comparatively higher mean AIC, BIC and AICc values. Both the AIC and AICc, which in general obtained very similar mean values, show that the no random effects model with $\boldsymbol{\omega}_i$ = TOEP obtained the lowest mean value compared to all other fitted models. The means for the BIC, which penalises more heavily for extra parameters compared to the AIC and AICc, indicate that the random intercept model with $\boldsymbol{\omega}_i$ = VC performed the best, with the no random effects model with $\boldsymbol{\omega}_i$ = CS obtaining the second lowest mean value. The mean AIC and AICc values were generally quite close, as shown in Fig. 4.1.

Table 4.2: Summary measures of AIC, BIC and AICc values for models successfully fitted to all simulated data sets.

Fitted Model			AIC				BIC			
ω_i	**Σ**	**No. Par**	mean	se	min	max	mean	se	min	max
VC	None	1	483.33	0.27	480.96	486.26	484.62	0.27	482.26	487.55
	Int. only	2	439.74	0.23	435.00	484.19	442.30	0.23	437.59	486.11
	VC	3	439.55	0.23	434.95	484.28	442.95	0.23	438.91	486.30
	CS	3	441.96	0.23	436.56	485.18	444.71	0.23	439.22	488.43
	UN	4	439.60	0.22	435.47	486.17	444.35	0.22	439.80	490.61
CS	None	2	439.75	0.23	435.00	484.69	442.35	0.23	437.59	487.28
CSH	None	10	441.88	0.23	437.10	487.49	448.36	0.23	443.58	493.97
AR(1)	None	2	446.14	0.22	439.97	484.67	448.73	0.22	442.56	487.26
	Int. only	3	439.50	0.22	435.25	485.10	443.30	0.22	439.12	488.29
	VC	4	439.32	0.22	435.26	485.18	443.91	0.22	439.97	488.48
	UN	6	439.11	0.22	432.42	486.94	445.14	0.22	438.88	492.66
ARH(1)	None	10	448.45	0.22	441.00	487.49	454.93	0.22	447.48	493.96
TOEP	None	4	439.05	0.22	432.12	486.48	444.24	0.22	437.30	491.66
UN	None	10	442.63	0.23	433.94	492.05	455.59	0.23	446.90	505.00

Fitted Model			AICc			
ω_i	**Σ**	**No. Par**	mean	se	min	max
VC	None	1	483.36	0.27	481.00	486.30
	Int. only	2	439.85	0.23	435.12	484.27
	VC	3	439.55	0.23	434.95	484.28
	CS	3	442.09	0.23	436.68	485.36
	UN	4	439.95	0.22	435.84	486.48
CS	None	2	439.87	0.23	435.12	484.80
CSH	None	10	442.49	0.23	437.71	488.10
AR(1)	None	2	446.26	0.22	440.09	484.79
	Int. only	3	439.73	0.22	435.49	485.28
	VC	4	439.32	0.22	435.26	485.18
	UN	6	439.65	0.22	433.03	487.43
ARH(1)	None	10	449.07	0.22	441.61	488.10
TOEP	None	4	439.05	0.22	432.12	486.48
UN	None	10	445.00	0.23	436.31	494.41

Note:
Int. only = Intercept only
No. Par = Number of parameters
se = standard error
min = minimum
max = maximum

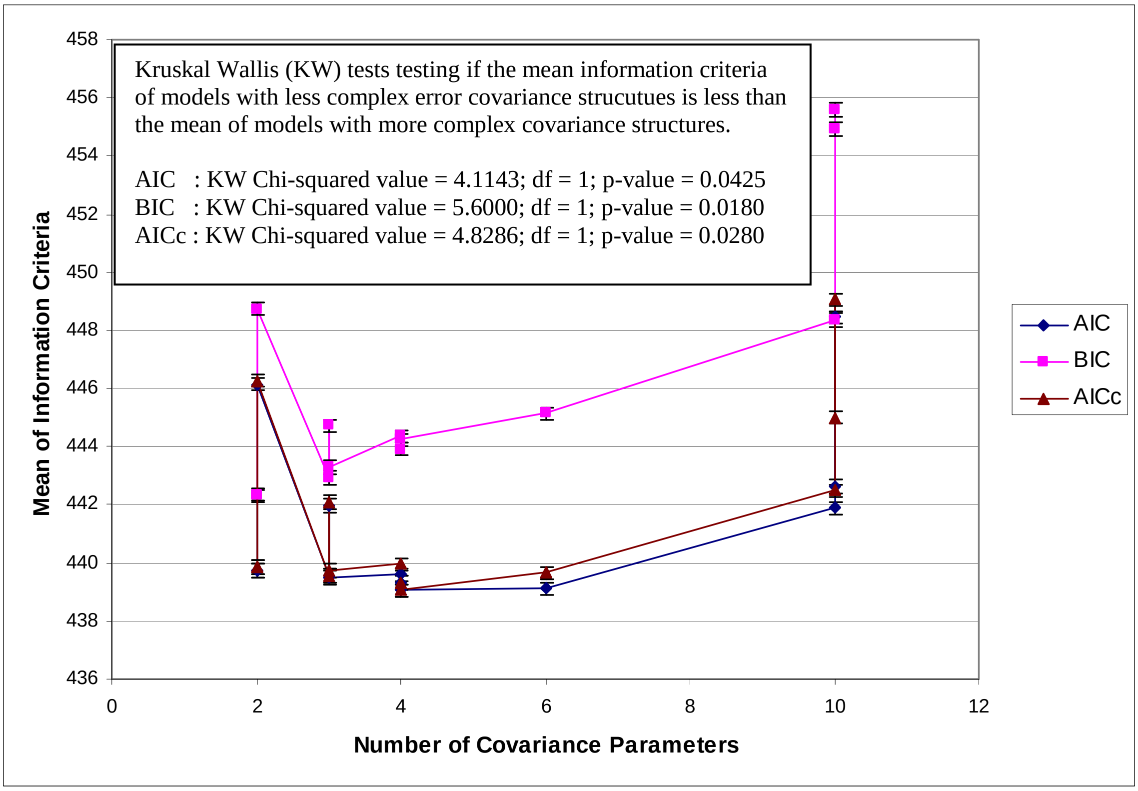

Fig. 4.1: Mean of the AIC, BIC, and AICc values as a function of the number of covariance parameters. The OLS model has been excluded. Kruskal Wallis tests testing if the mean information criteria are significantly higher for models with either UN, CSH or ARH(1) error covariance structures are included.

The mean AIC, BIC, and AICc values for the OLS model were about 10% larger than the values obtained for the best fitting models. Therefore the OLS model, which had the smallest number of parameters (one covariance parameter), did not fit the data sufficiently well, as evidenced by the large AIC, BIC and AICc values.

The overall mean AIC, BIC and AICc values can only give an indication of which models are performing better than others when the variability of these criteria between different simulated data sets is small. As the standard errors of the mean AIC, BIC and AICc values were small (Table 4.2), ranging between 0.22 and 0.23 for all models except the OLS model, which had a standard error of 0.27, this method of comparing fitted models can be justified in this case. A clearer indication that a model was performing well would be if the fitted model consistently obtained information criteria within two units of the minimum across all simulated data sets. Therefore a second means of comparing between models was developed based on the AIC, BIC and AICc values. For each simulated model, the percentage of times a model produced an AIC, BIC or AICc within two units of the respective minimum was calculated for each fitted model (Appendix A2). In order to have one percentage value with which to compare different fitted models, each fitted model's percentage values were averaged over all simulated models. These results appear in Table 4.3.

The mean percentage values showed a similar trend to the mean AIC, BIC and AICc values. The simpler models tended to have information criteria within two units of the respective minimum more often than the more complicated models. Models which performed particularly well under this analysis were the random intercept model with $\boldsymbol{\omega}_i$ = VC, and the no random effects model with $\boldsymbol{\omega}_i$ = CS, both models having only two covariance parameters. Both of these models were within two units of the minimum BIC more than 55% of the time on average, and within two units of the minimum AIC and AICc values more than 43% of the time. The lowest percentages for both of these models were recorded for data simulated under AR(1) or ARH(1) models with no random effects. The CS model with no random effects and the random intercept model with independent errors should have similar results as these two models result in the same parameterisation of the variance, $\mathbf{V}_i$ (Davis 2002, p. 137). The top performing model according the AIC and AICc values was the model with $\boldsymbol{\omega}_i$ = VC and $\boldsymbol{\Sigma}$ = VC, which obtained values for these criteria within two units of the minimum more than 45% of the time. The model with $\boldsymbol{\omega}_i$ = VC and $\boldsymbol{\Sigma}$ = UN, the two random intercept models with $\boldsymbol{\omega}_i$ = VC and with $\boldsymbol{\omega}_i$ = AR(1), and the no random effects model with $\boldsymbol{\omega_i}$ = TOEP, obtained AIC and AICc values with 2 units of the respective minimum more than 30% of the time.

Table 4.3: Percentage occurrences where fitted models produced AIC, BIC or AICc values within two units of the minimum.

Fitted Model		Percentage fits where model was within 2 units of the AIC	Percentage fits where model was within 2 units of the BIC	Percentage fits where model was within 2 units of the AICc
$\boldsymbol{\omega_i}$	$\boldsymbol{\Sigma}$			
VC	None	3.02	3.42	3.05
	Int. only	43.81	57.14	46.07
	VC	47.11	32.80	43.93
	CS	25.08	29.91	25.70
	UN	30.39	10.50	24.59
CS	None	43.70	55.85	45.82
CSH	None	12.65	4.15	11.45
AR(1)	None	14.33	17.50	14.86
	Int. only	44.01	19.90	36.98
	VC	33.56	13.60	27.48
	UN	26.16	9.14	21.45
ARH(1)	None	4.16	1.75	3.99
TOEP	None	32.61	19.41	30.70
UN	None	7.90	0.68	4.14

The percentage of cases where the AIC, BIC or AICc values of the OLS model were within two units of the minimum was much lower (close to 3%) compared to the best fitting models (close to 50%) (Table 4.3). On closer inspection, the cases where the OLS model did have AIC or BIC values close to the minimum were all for data simulated under the OLS model (Table 4.4).

Table 4.4: Percentage occurrences where fitted models produced AIC, BIC or AICc values within two units of the minimum when fitted to data simulated from the same model and the percentage when fitted to data simulated from other models.

Fitted Model		Percentage fits where correct model was within 2 units of the AIC	Percentage fits where incorrect model was within 2 units of the AIC	Percentage fits where correct model was within 2 units of the BIC	Percentage fits where incorrect model was within 2 units of the BIC	Percentage fits where correct model was within 2 units of the AICc	Percentage fits where incorrect model was within 2 units of the AICc
$\boldsymbol{\omega}_i$	$\boldsymbol{\Sigma}$						
VC	None	81.60	0.00	92.40	0.00	82.40	0.00
	Int. only	75.60	43.23	89.20	56.35	78.40	45.50
	VC	69.20	46.78	37.20	32.80	58.80	43.71
	CS	71.20	23.31	56.00	28.74	64.80	24.19
	UN	42.00	30.16	7.20	10.59	34.80	24.34
CS	None	75.60	43.12	83.20	55.20	77.60	45.26
CSH	None	40.00	11.74	14.00	3.84	36.80	2.82
AR(1)	None	77.60	12.30	90.80	15.12	80.40	12.77
	Int. only	79.60	43.81	88.00	20.30	51.20	37.10
	VC	38.00	33.92	11.20	13.86	28.40	27.90
	UN	52.40	24.08	24.00	7.86	48.40	19.26
ARH(1)	None						
TOEP	None	68.00	30.34	52.00	17.34	67.20	28.29
UN	None	23.6	7.36	3.60	0.59	13.2	3.82

Table 4.4 compares how well a fitted model performs when fitted to data simulated from the same model versus when it is fitted to data simulated under a different model (i.e. when it is incorrectly fitted to a data set). This table shows that, with respect to the AIC and AICc values, the percentage of model fits with information criteria within two units of the minimum when a model is fitted to data simulated from the same model is always higher compared to the percentage when fitted to data simulated under a different model. The percentage values for the BIC indicate that the model with $\boldsymbol{\omega}_i$ = VC and $\boldsymbol{\Sigma}$ = UN and the model with $\boldsymbol{\omega}_i$ = AR(1) and $\boldsymbol{\Sigma}$ = VC have

lower percentage of model fits with BIC within two units of the minimum where data was fit the correct model versus when fit to the incorrect model. Both the percentages for the AIC and for the AICc showed the smallest difference between correct and incorrect model fits for the model with $\boldsymbol{\omega_i}$ = AR(1) and $\boldsymbol{\Sigma}$ = VC, where the difference for AICc was less than 1% and for AIC was close to 4%. The smallest difference in the BIC percentage values between correct and incorrect model fits was for the model with $\boldsymbol{\omega_i}$ = VC and $\boldsymbol{\Sigma}$ = UN. The largest difference between correct and incorrect models for all three information criteria was for the no random effects model with $\boldsymbol{\omega_i}$ = AR(1). The differences in percentages for the AIC and AICc values were both over 65%, and the difference for the BIC was over 75%. There were large differences in percentage values for the AIC and AICc value for the model with $\boldsymbol{\omega_i}$ = VC and $\boldsymbol{\Sigma}$ = CS. This was the model fitted with equal elements in the covariance structure for the random effects.

Table 4.5 shows, for each set of simulated data, pairs of fitted models which both obtained AICc values within two units of the minimum more than 40% of the time. This table was created to show which models are "interchangeable", in that they obtain similar fits to the data, making it difficult to distinguish which model is best using the information criteria. The AICc was chosen as it is corrected for bias compared to the regular AIC, and does not penalise for additional parameters as severely as the BIC, as discussed in Chapter 2 (McQuarrie & Tsai, 1998). Models which together obtained low AICc values close to the minimum the most often are the random effects model with $\boldsymbol{\omega_i}$ = VC and $\boldsymbol{\Sigma}$ = VC, the random intercept model with $\boldsymbol{\omega_i}$ = VC, and the no random effects model with $\boldsymbol{\omega_i}$ = CS. Together with these three models, the random intercept model with $\boldsymbol{\omega_i}$ = AR(1) also obtained close to the minimum AICc for many of the simulated models. Only the OLS model included itself and the AR(1) no random effects model, along with the three previously mentioned models, as most frequent best fitting models. The OLS model had the highest number of model pairs obtaining AICc values close to the minimum for more than 40% of the simulated data sets. Data simulated under the random intercept models with $\boldsymbol{\omega}_{\boldsymbol{i}}$ = AR(1) and $\boldsymbol{\omega}_{\boldsymbol{i}}$ = ARH(1) were two of only twelve models with best fitted models differing from those mentioned above. In these two cases the no random effects model, random intercept model, and the random effects model with $\boldsymbol{\Sigma}$ = VC, each with AR(1) error structure, most often obtained the lowest AICc values. The data simulated under $\boldsymbol{\omega}_{\boldsymbol{i}}$ = AR(1) random effects models with either $\boldsymbol{\Sigma}$ = CSH, ARH(1) and UN had minimum AIC values when either of these three model were fit to the data. The models with the most number of parameters (i.e. the heterogeneous error models with heterogeneous random effects or UN random effects structure, as well as the no random effects model with $\boldsymbol{\Sigma}$ = UN) did not have any pairs of models frequently obtain minimum AICc values.

Table 4.5: For each simulated model set the fitted model pairs that both obtained within 2 units of the minimum AICc value for more than 40% of the simulated data sets.

True Model		Fitted model pairs which both obtained AICc values within 2 units of the minimum for more than 40% of the simulated data sets
ω_i	Σ	
VC	None	1-2; 1-3; 1-8; 1-15; 2-3; 2-8; 2-15; 3-8; 3-15; 8-15;
	Int. only	2-3; 2-8; 2-16; 3-8; 8-16
	VC	2-3; 2-8; 2-16; 3-8; 8-16
	CS	-
	CSH	2-3; 2-8; 3-8
	ARH(1)	2-3; 2-8; 2-16; 3-8; 8-16
	UN	2-3; 2-8; 2-16; 3-8; 8-16
CS	None	2-3; 2-8; 2-16; 3-8; 8-16
	CSH	2-3; 2-8; 2-16; 3-8; 8-16
CSH	None	2-3; 2-8; 2-16; 3-8; 8-16
	CSH	-
	ARH(1)	-
	UN	-
AR(1)	None	15-16; 15-17; 16-17
	Int. only	2-3; 2-8; 2-16; 3-8; 8-16
	VC	2-3; 2-8; 3-8
	CSH	18-19; 18-20; 19-20
	ARH(1)	18-19; 18-20; 19-20
	UN	18-19; 18-20; 19-20
ARH(1)	None	15-16; 15-17; 16-17
	VC	2-8
	CSH	-
	ARH(1)	-
	UN	-
TOEP	None	2-8
UN	None	-

Model numbers for Assumed Models

1 : ω_i = VC, Σ = None;
2 : ω_i = VC, Σ = intercept only;
3 : ω_i = VC, Σ = VC;
4 : ω_i = VC, Σ = CS;
5 : ω_i = VC, Σ = CSH;
6 : ω_i = VC, Σ = ARH(1);
7 : ω_i = VC, Σ = UN;
8 : ω_i = CS, Σ = None;
9 : ω_i = CS, Σ = CSH;
10 : ω_i = CS, Σ = TOEP;
11 : ω_i = CSH, Σ = None;
12 : ω_i = CSH, Σ = CSH;
13 : ω_i = CSH, Σ = ARH(1);
14 : ω_i = CSH, Σ = UN;
15 : ω_i = AR(1), Σ = None;
16 : ω_i = AR(1), Σ = intercept only;
17 : ω_i = AR(1), Σ = VC;
18 : ω_i = AR(1), Σ = CSH;
19 : ω_i = AR(1), Σ = ARH(1);
20 : ω_i = AR(1), Σ = UN;
21 : ω_i = ARH(1), Σ = None;
22 : ω_i = ARH(1), Σ = VC;
23 : ω_i = ARH(1), Σ = UN;
24 : ω_i = ARH(1), Σ = ARH(1);
25 : ω_i = ARH(1), Σ = UN;
26 : ω_i = TOEP, Σ = None;
27 : ω_i = TOEP, Σ = CS;
28 : ω_i = TOEP, Σ = TOEP;
29 : ω_i = UN, Σ = None

4.4 Analysis of Coverage Probabilities

To determine the robustness of the models, the coverage probabilities were calculated for the 95% confidence intervals of the fixed effects for each fitted model under each simulated model. As there were a large number of comparisons to make, these coverage probabilities appear in Appendix A3. The average coverage probabilities were calculated for each fitted model over all the simulated data sets. These results are presented in Table 4.6 for those models which were successfully fitted to all of the simulated data sets.

Table 4.6: Mean coverage probabilities of fixed effects for each fitted model which was successfully fitted to all the simulated data.

ω_i	Σ	Intercept	Gender (male = 1)	Age	Gender×Age
VC	None	**98.86**	**98.55**	**99.31**	**99.11**
	Int. only	94.18	*93.75*	*92.42*	*92.09*
	VC	*93.14*	*92.66*	*93.06*	*93.03*
	CS	*91.65*	*91.51*	95.14	94.97
	UN	95.77	95.25	95.69	95.28
CS	None	94.18	*93.75*	*92.43*	*92.09*
CSH	None	94.00	*93.77*	*92.48*	*92.49*
AR(1)	None	**98.05**	**97.65**	**98.40**	**98.18**
	Int. only	94.91	94.46	*93.65*	*92.89*
	VC	*93.72*	*92.97*	*93.71*	*93.18*
	UN	95.65	95.05	95.29	94.92
ARH(1)	None	**97.83**	**97.80**	**98.22**	**98.17**
TOEP	None	95.69	95.08	94.38	94.12
UN	None	*93.35*	*92.26*	*92.95*	*92.28*

Test for H_o: The coverage probability is 95%. This hypothesis is rejected for values less than 94% or greater than 96%.

Values less than 94% (italics) with background.

Values between 94% and 96% (normal font) with background.

Values more than 96% (bold) with background.

Each model was fitted to 26 sets of simulated data, each set containing 250 data sets. In total, each model was fitted to 6500 data sets. To get an approximate non-rejection region for the hypothesis that the true coverage probability of the 95% confidence interval is 95%, the normal approximation was used. To do this, it needs to be assumed that the 6500 data sets represent a random sample of data sets. To reject this hypothesis, the average coverage probability needs to be less than 94% or more than 96%. These results are displayed in Table 4.6. A robust model should result in accurate estimates of the fixed effects, as well as accurate standard errors of these estimates. If a model is predicting the fixed effects well under all of the simulated models then the coverage probability of the 95% should be between the 94% and 96%. If the coverage probabilities are above 96%, then this means that the variances have been overestimated, resulting in an overly conservative model. If the coverage probabilities are below 94%, then estimates are either biased or the variance is under estimated.

The model with $\boldsymbol{\omega}_{\boldsymbol{i}}$ = TOEP and no random effects, as well as the random intercept and slope models with $\boldsymbol{\omega}_{\boldsymbol{i}}$ = VC and Σ = UN and with $\boldsymbol{\omega}_{\boldsymbol{i}}$ = AR(1) and Σ = UN, obtained coverage probabilities that were within the limits. All three of these models were determined by the AIC and BIC analysis to fit the data relatively well. The random intercept model with $\boldsymbol{\omega}_{\boldsymbol{i}}$ = VC, the no random effects model with $\boldsymbol{\omega}_{\boldsymbol{i}}$ = CS, and the random intercept and slope model with $\boldsymbol{\omega}_{\boldsymbol{i}}$ = VC and $\boldsymbol{\Sigma}$ = VC had coverage probabilities that were significantly below 95%, despite being the models that, according to the AIC and BIC measures, fit the data the best. The AR(1) and VC error models with no random effects had coverage probabilities that were significantly more than 95%, indicating that these models were overestimating the variance.

The performance of these models was then investigated for each simulated model (Appendix A3). Each model's performance was based on how many confidence intervals of the coverage probabilities contained 95%. Of the three models that obtained coverage probabilities close to 95%, the random intercept and slope model with $\boldsymbol{\omega}_i$ = VC and $\boldsymbol{\Sigma}$ = UN performed the best, with only two coverage probabilities significantly higher, and only one coverage probability significantly lower than 95%. The random intercept and slope model with $\boldsymbol{\omega}_i$ = AR(1) and $\boldsymbol{\Sigma}$ = UN also performed well, with six coverage probabilities significantly more than 95% and one coverage probability significantly lower than 95%. The random effects with $\boldsymbol{\omega}_i$ = TOEP performed slightly worse, with eight coverage probabilities significantly above and four coverage probabilities significantly below 95%.

Approximately 15% of the simulated models had coverage probabilities for the random intercept model with $\boldsymbol{\omega}_i$ = AR(1) that were significantly below 95%. The no random effects model with $\boldsymbol{\omega}_i$ = CS, the random intercept model with $\boldsymbol{\omega}_i$ = VC, and the random intercept and slope models with $\boldsymbol{\omega}_i$ = VC and $\boldsymbol{\Sigma}$ = CS, $\boldsymbol{\omega}_i$ = VC and $\boldsymbol{\Sigma}$ = VC, and with $\boldsymbol{\omega}_i$ = AR(1) and $\boldsymbol{\Sigma}$ = VC did not perform very well in relation to the coverage probabilities as each of these models had more than 25% coverage probabilities below 95%. The no random effects models with $\boldsymbol{\omega}_i$ = AR(1) and $\boldsymbol{\omega}_i$ = VC had coverage probabilities significantly higher than 95% for 76% and 92% of the simulated models respectively. Both of these models produced coverage probabilities that were reasonable for simulated models under no random intercept models with $\boldsymbol{\omega}_i$ = AR(1), $\boldsymbol{\omega}_i$ = ARH(1) and $\boldsymbol{\omega}_i$ = VC.

4.5 Discussion

Verbeke and Lesaffre (1997) showed that the fixed effects estimates of the linear mixed effects model were robust to misspecification of the distribution of the random effects. In their study they assumed that the covariance structure was correctly specified, but that the random effects were not normal. The purpose of this study was to determine the robustness of the fixed effects estimates of the linear mixed effect model when the covariance structure was misspecified, assuming that the random effects were normally distributed. The results show that the linear mixed effects model is not robust to misspecification of the covariance structure, although there are certain covariance structures that perform better when misspecified compared to others.

This study supports the findings of Jacqmin-Gadda *et al*. (2007), who showed that random intercept and slope models were more robust than random intercept models. In this study, the random intercept models fit relatively well according to the information criteria, but had coverage probabilities that were too low. Under certain covariance combinations, the random intercept and slope models obtained coverage probabilities that indicated robustness. This applies to simple covariance structures, and generally when the correlation between observations on the same subject is only described by either the random errors or the random effects covariance structures.

Certain covariance structures were shown to be "interchangeable", as they obtained similar fits consistently, and for data simulated from various covariance structures. In particular, the random effects model with $\boldsymbol{\omega}_i$ = VC and $\boldsymbol{\Sigma}$ = VC, the random intercept model with $\boldsymbol{\omega}_i$ = VC, and the no random effects model with $\boldsymbol{\omega}_i$ = CS obtained similar fits the most often. It would be expected that the random intercept model with $\boldsymbol{\omega}_{\boldsymbol{i}}$ = VC and the no random effects model with $\boldsymbol{\omega}_{\boldsymbol{i}}$ = CS would obtain similar fits as the resulting covariance structure for the responses should be the same under these two models, but where one model includes random effects and the other does not. These three covariance structures did not frequently fit the data well under all generating models. For example, these models did not perform well when fitted to data generated from random intercept models with either AR(1) or ARH(1) errors, random effects models with AR(1) errors and complex random effects structures. Therefore the flexibility of these models is limited.

A comparison was also made between an OLS model, assuming independent errors and no random effects, and other linear mixed effects models. The OLS model fitted the data much worse, as the AIC, BIC and AICc values were much higher compared to the other linear mixed effects models. The coverage probabilities of the OLS model were too high, indicating an overestimation of the standard error. This is what is expected when the correlation between repeated measurements is ignored (Fitzmaurice *et al*. 2004).

As well as fitting the data well, a robust model should obtain parameter estimates of the fixed effects that are close to the true values, as these parameter estimates are largely what a researcher would require from the model, and obtain inferences that accurately describe the expected range of the parameters. Even though quite a number of models were identified by the information criteria analysis as fitting the data similarly, based on the analysis of the coverage probabilities, only three of these models obtained coverage probabilities close to the confidence level. These three models therefore obtained parameter estimates for the fixed effects that were the closest to the true parameter values, as well as standard errors that were reasonable. Good choices for the covariance structure therefore include the random intercept and slope model with $\boldsymbol{\omega}_i$ = VC and $\boldsymbol{\Sigma}$ = UN, the random intercept and slope model with $\boldsymbol{\omega}_i$ = AR(1) and $\boldsymbol{\Sigma}$ = UN, and the no random effects model with $\boldsymbol{\omega}_i$ = TOEP.

This analysis shows that even though the underlying covariance structure of the data may be complicated, it may not be necessary to fit a model to this data with the same parameterisation. A simpler covariance structure can result in a better fit with parameter values closer to the true parameter values. Therefore it appears that the maximisation routine used by default by SAS PROC MIXED (ver. 9.1) is not very successful when fitting complicated models with many parameters. Since certain models were not successfully fitted to simulated data generated from the same model, it indicates that the optimisation procedure is not always successful at fitting models with certain, usually more complex, covariance structures, not because the covariance structure is misspecified, but because the optimum solution is not located by the procedure. This is most likely due to the large number of covariance parameters, but small number of subjects in the PR data set.