Corrections to *Faster Fitting for Joint Models of Survival and Multivariate Longitudinal Data*

In the corrected thesis the following changes have occurred:

Overall

Corrected grammatical, spelling, and small errors throughout, as prescribed in the corrections document.

The declaration has had the word count updated, and a signature added above the author's name.

A Glossary (Page 159) has been created, which seeks to give a comprehensive overview of notation used in more than one part of the Thesis.

Abstract

The abstract has been extended (new paragraph, Para 3), separating the '*extensive simulation studies*' and '*thorough application*' clauses, detailing now the estimation capabilities from these simulations, as well as the high levels of agreement with established methodology, compared to which the approximate EM demonstrated faster computation times.

Chapter 1

A section has been added to Section 1.2.1 (Page 3, Para 2) discussing the shared random effects and current value parameterisations, with some references given and the interested reader pointed towards a curated table in existing literature comparing parameterisation approaches. Additionally in this section interpretations of both parameterisations are given.

Chapter 2

The sections 2.6.1–2.6.3 (and corresponding Figures 2.1–2.3) on Pages 30–33 have been reworked with parameters resulting in survival times which exhibit the the skewness usually associated with survival data. Some discussion has been rewritten as a result.

The Section 2.6.4 *A note on statistical power for joint models* has been moved to Appendix A.3 (Page 165). All further references to this previously-existing section have been updated throughout the thesis accordingly. A new section, Section 8.1.2 (Page 155), has been added to Chapter 8 referencing this appendix and discusses power in joint models as a potentially interesting future avenue of research.

Chapter 3

On Page 35, Last Para, Line 3, the passage '*Adaptive quadrature methods sought to decrease the number of points required integrals were evaluated at*' has been rephrased as '*Adaptive quadrature methods sought to reduce the number of weights and abscissae used in numerical integration*'.

When the 'standard' simulation scenarios are discussed in Section 3.4.1, in light of the new figures in Sections 2.6.1-2.6.3, '*The resultant distribution of failure times then resembling the corresponding flatter distributions...*' are referenced, in place of another figure showing this on Page 36, sixth line from bottom.

The simulation studies for t-distributed random effects which begin on Page 57 has been re-done with the degrees of freedom capped at $\nu = 30$, with note that this approximates the normal. Underlying problems with the code generating Table 3.1 were also fixed, leading to nominal coverage amongst all

parameters. Discussion surrounding estimation for these variance-covariance parameters has changed slightly, now mentioning that they achieve nominal coverage as $\nu \to 30$ on Page 59, first paragraph.

Discussion surrounding potential 'knock-on' effects of misspecification in the the t-distribution simulation study has been added on Page 59, last para; the section now concluding by citing appropriate literature detailing how misspecification of the random effects can lead to problems and inefficiencies with other aspects of the fitted joint model, such as individualised predictions.

Chapter 4

Section 4.2.1 (Page 66) has been shortened considerably, with the restatement of advantages of GLMMs removed in favour of a single cross-reference to why one may be interested in doing this.

Table 4.1 on Page 70 which outlines all considered exponential families has been tidied such that the link functions are given as e.g. h: $\log(\mathbb{E}[\boldsymbol{Y}_i])$ instead of $h(\mathbb{E}[\boldsymbol{Y}_i])$: log. The Gaussian case has also been added to this Table and the reader is referred back to Chapter 2 for details surrounding its specification.

Section 4.5.7 has been greatly expanded to discuss the poorer estimation capability of the binomial and GP responses on Page 89, first para, '*We noted broadly good performance in the estimation of* $\hat{\boldsymbol{\Omega}}$; *however, this is not without its caveats...*'. This is posited as being due to the nature of the response, sparsely populated simulated responses, or the underlying normal approximation being less appropriate, with references to the justification for the normal approximation given as a segue. Further recommendations for those interested in using the approach is given in the final paragraph of the section again on Page 89, '*We perhaps revisit then the potential use of the approximate EM algorithm...*'.

Chapter 5

Two new figures, Figures 5.1 and 5.2, present 'cropped' versions of the Gaussian and Binomial scatterplots in the visualisation section starting on Page 97. The full figures were not implemented due to some LaTeX issues which didn't have an obvious fix, but are still present in the Appendix C.2.1 starting on Page 193.

A figure for the binomial case (Figure 5.6, Page 103) has been added in Section 5.2.3 (starting on Page 100). Subsequent discussion has been rearranged and references to the remaining figures in the appendices altered.

In addition to the above, a brief investigation into the proportion of the complete data log-likelihood accounted for by the longitudinal/survival part is given on Page 102, first paragraph, second line, '*Taking this idea further, we undertake a brief simulation study and consider the Gaussian, Poisson and binomial families...*'. This elucidates slightly the problems with the binomial fit; the proportion of this likelihood accounted for by the *longitudinal* part is quite variable hence the approximation may be suffering slightly.

Chapter 6

The text immediately following Equation (6.17) on Page 124 has been expanded upon, describing what this equation is calculating exactly i.e. the components of prediction error calculated by each term.

Note that the equation for $\widehat{\mathrm{PE}}(w)$ on Page 125 has remained unchanged since the optimism seeks to correct for overly favourable performance on the original data. Since lower values of $\widehat{\mathrm{PE}}$ are better, this optimism is then *added* to the original $\widehat{\mathrm{PE}}$. Conversely the optimism in AUC is subtracted, since higher values of this is preferable.

Chapter 7

Figure 7.1 on Page 127 has been updated with correct axes, and the caption updated to explain that the histogram shows failure times occurring during follow-up.

Limitations of the package `gmvjoint` are mentioned in the fifth line of the first paragraph in Section 7.3.1 on Page 129 when discussing the lack of interactions in survival sub-models. The interested reader is now referred to a new section, Appendix D.4 on Page 221, dedicated to highlighting some limitations of the package.

Figure 7.4 on Page 132 has had the colours changed to hopefully be more readable.

An interpretation for the γ term attached to serum bilirubin has been added in Section 7.4.1 on Page 133, third Para, sixth line, '*Focussing specifically on serum bilirubin ...*'.

Figure 7.5 (univariate) on Page 134 and Figure 7.6 (multivariate) on Page 135 have had the association parameters attached to prothrombin scaled an amount representative of the 80th percentile of the estimated random effects at $t = 1$. This has been mentioned in the caption in both cases, as well as in-text on Page 133, fourth Para '*A very large value for $\hat{\gamma}$ is observed...*' for the univariate case; and for the multivariate on Page 135, third Para '*Here we note that for 'Blood clotting and flow'...*'.

Tables 7.4 (Page 138), 7.5 (Page 140) and 7.6 (Page 141) have had parameter estimates which are significant at the 5% level put in **bold**face font to aid visual comparison.

Further model diagnostics from the final bivariate fitted joint model are provided in new Appendices C.3.5, C.3.6, C.3.7, C.3.8, and C.3.9 (starting on Page 215) and discussed in main-text on Page 141 '*Several forms of model diagnostics (for the joint model fit by approximate EM only) are carried out...*'.

Chapter 8

The opening paragraphs on Page 151 which resembled the Abstract too strongly has been reworded and restructured to make this a less obvious cut-and-paste job.

Added '*However, the performance of the algorithm was notably weakest when a binary response was considered, and over-dispersed generalised Poisson, with both shortcomings discussed*' to the 'thesis recap/conclusions' part on Page 153, second Para, third line from bottom to draw specific attention to these shortcomings.

The conflation in naming a competing risks model as a Cox PH on Page 155, third line, has been removed.

A short sub-section detailing potential use of separate γ terms on the intercepts and slopes '*Finally, within the specification of the survival sub-model in the joint modelling framework...*' has been added on Page 155, third Para.

A new Section 8.1.2 '*Power for joint models*' (Page 155) has been added, which references the new Appendix A.3 and draws attention to this being a vital line for future research in multivariate joint models.

A few sentences have been added when discussing potential avenues of further modelling counts '*Perhaps the most obvious avenues here are modelling a structural excess of zeroes...*' on Page 156, fourth Para, fourth line, therein giving some idea as to what a hurdle model essentially does.

When discussing the linear scan algorithm, a couple of sentences have been added on Page 157, second Para, sixth line, to attempt to better explain where this supposed uptick in computational efficiency comes from '*A linear scan algorithm sequentially...*'.

All changes to the thesis can be found at `https://github.com/jamesmurray7/thesis`