CHAPTER VI
SUMMARY AND CONCLUSION

The rate of change, time immemorial has fascinated the history of humanity, in particular the fields of Physics, Mathematics and Statistics. Vital characteristics of a process are captured through rate of change of the components that constitutes the process. In Statistics, it plays critical role, as in the case of Cramer-Rao lower bound, in deciding the lower bound of variance of estimators.

In longitudinal studies related to Clinical and Epidemiological nature, a major objective is to estimate the change in the response variable, measured over a fixed or varying time interval. The change in values over time, measured as a number describes both direction and quantum or steepness of that particular measure. This measure of rate of change, also referred to as slope, helps in identifying the patterns of interest and of providing proper intervention in the field of Medical sciences and Industry.

When the data under consideration is complete, in the sense that it has no missing values, the estimation of slope is a straightforward affair. This is not the case with longitudinal studies, in which more drop outs, also known as censored objects, are expected, due to various reasons. When the data has missing structures, the
estimation of slope becomes complicated due to the nature of treatment adopted for the missing values that are, in most of the cases assumed to be independent of the variable of observation. The Problem of slope estimation culminates to its peak, if the drop out pattern becomes dependent on the slope of the response variable under consideration. In addition, most of the bio-medical studies warrant the estimation of the slope for each individual subject in the sample, in order to closely monitor their improvement pattern and to intervene at crucial time points, indicated by the rate of change at that particular time point. This increases the number of parameters to be estimated and turns the problem to a multivariate optimization process, dealing with complex algorithms that are more time consuming in nature.

In this thesis, initially, some of the basic methods of measuring the slope are discussed. These methods include Unweighted slope estimator and Fearn’s Weighted slope estimator. The Unweighted slope estimator is nothing but the ordinary least square estimator of the slope in the linear model, with response value as the dependent variable and time as independent variable. The Weighted estimators in general and the Fearn’s Weighted slope estimator in particular, attach differentiable weights to individual subject’s measurement, in order to arrive at an estimator for the parameters under consideration. Invariably, the weights are chosen as a function of overall variation present among the subjects and the subject–specific variation over
time. This phenomenon, in some literature, also is referred to as “between subject” variation and “within subject variation” respectively.

The problem posed by the increased number of unknown individual slope parameters is circumvented using empirical Bayes method that can be viewed as an approximation to the standard Bayesian method of a hierarchical model, in which the parameters at the highest level of hierarchy are set to their most likely values, instead of being integrated out. In this thesis, empirical Bayes procedure is extensively used in the process of estimating the population and individual slope parameters.

Marginal likelihood estimation, considered in this thesis, derives estimators of slopes for univariate and bivariate outcomes with informative missing structures. It uses the independent sufficient statistics, to capture all the information about the intended individual slope parameters.

In this thesis, longitudinal data structure with informative and non-informative drop out patterns are generated through simulation process governed by real time parameters and using R programming Language. Also the estimation process with known and unknown variance structure are discussed.

In chapter II, Random slopes for 100 individuals are simulated from $N(\beta, \sigma_{\beta}^2)$ and using them with a truncated Geometric distribution, the distribution for number of observations of each subject is generated over the set $M=\{2, 3, 4, 5, 6, 7\}$. The parameters involved in this process are $\beta$, $\sigma_{\beta}^2$, $\gamma_0$ and $\gamma_1$. These parameters are varied.
to generate different drop out patterns. This chapter generates 500 sets of 100 subjects with drop out patterns depending on the random slope parameters and the censoring parameters $\gamma_0$ and $\gamma_1$. For arriving at unique results the random seed of the simulation process is kept at value ‘7’ in most of the cases. In particular, with $\beta = -0.1, \sigma^2_\epsilon = 0.1, \sigma^2_\beta = 0.05, \gamma_0 = -2.06$ and $\gamma_1 = 2.3$ the data are simulated and the Unweighted and Weighted slope estimators are found. This chapter considers the scenario of informative right censoring and non-informative right censoring under both known and unknown variance structures. The efficiency of the estimators is measured using MSE.

The MSE of Unweighted slope estimator is 1.0302 times that of the Weighted slope estimator in case of informative right censoring under known variance. The MSE of Unweighted slope estimator is 1.0218 times that of the Weighted slope estimator in case of non-informative right censoring under known variance. The MSE of Unweighted slope estimator is 1.0542 times that of the Weighted slope estimator in case of informative right censoring under unknown variance. The MSE of Unweighted slope estimator is 1.0422 times that of the Weighted slope estimators in case of non-informative right censoring under unknown variance.

In all the four combinations, it is seen that the Weighted slope estimators have uniformly less MSE compared to their corresponding Unweighted slope estimators. But in all the four combinations the MSE reduction factor is considerably low and
this can be justified, as none of the censoring distribution parameters or their estimates are taken into account in the estimation process of the individual slopes. Also, it is seen that in both the informative and non-informative cases, Weighted slope estimators behave better compared to the Unweighted slope estimators. This, to some extent indicates the supremacy of the Weighted slope estimator in both the informative and non-informative scenarios, as it makes use of the information of the variance in an effective manner. In the case of known and unknown variances, again, the Weighted slope estimator behave better compared to the Unweighted slope estimator. This implies that the Weighted slope estimators better handles the variance factor compared to its Unweighted counterpart.

Thus, from this chapter, it is concluded that the Weighted individual slope estimators behaves better compared to the Unweighted individual slope estimators, in having reduced MSE.

Chapter III mainly deals with empirical Bayes estimation of individual slope parameters with data generated from Normal distribution and having a truncated Geometric drop out structure. This chapter also deals with LMVUB estimation, as a special case with $\sigma_\beta^2 = 0$.

The estimators described include Unweighted, Fearn’s Weighted and empirical Bayes slope estimators under informative and non-informative right censoring with known and unknown variances. In the case of informative right censoring with known variance parameters, Empirical Bayes estimator possesses the
least MSE. In the case of non-informative right censoring with known variance parameters, Fearn’s Weighted slope estimator possesses the least MSE compared to the Unweighted and Empirical Bayes estimators. These two observations put together implies when there is no informative right censoring, the Fearn’s Weighted estimator, which does not make use of the censoring parameter $\gamma_1$, performs better. In this case the Empirical Bayes method, possibly by estimating the practically non-existing parameter $\gamma_1$, goes down in not reducing the MSE. This observation encourages simpler models for simple scenarios and complicated models to be used only when situation warrants. The scenario under unknown variance is similar to that of the known case and the margin of gain achieved by the Empirical Bayes estimator considerably reduces here. In the case of informative right censoring with unknown variance parameters, Empirical Bayes estimator possesses the least MSE. In the case of non-informative right censoring with unknown variance parameters, Fearn’s Weighted slope estimator possesses the least MSE compared to the Unweighted and Empirical Bayes estimator. This again implies when there is no informative right censoring, the Fearn’s Weighted estimator, which does not make use of the censoring parameter $\gamma_1$ performs better. In this case the Empirical Bayes method fails as per the justification provided in the case of known variance. In the case of informative right censoring with unknown variance parameters, LMVUB estimator possesses the least MSE. In the case of non-informative right censoring with unknown variance
parameters, LMVUB slope estimator possesses the least MSE compared to the Unweighted estimator. This implies that, irrespective of informative or non-informative right censoring, the LMVUB estimator performs better.

In short, Empirical Bayes estimator behaves better in case of informative right censoring under both known and unknown variance structures. Under non-informative right censoring, the Fearn’s Weighted slope estimator has the least MSE. But when considering the LMVUB estimator, they always behave effectively in all the scenarios of informative and non-informative with known and unknown variance structures.

In Chapter IV, the censoring and the response variables are generated using truncated Geometric and Normal distributions respectively, similar to the one done in Chapter II. This chapter compares MSEs of individual slope estimators among Unweighted, Fearn’s Weighted, Empirical Bayes, IRC1, IRC2 and IRC3 slope estimators under informative and non-informative right censoring with known and unknown variance structures.

In the case of informative right censoring with known variance parameters, IRC3 estimator possesses the least MSE. This is followed by Fearn’s Weighted, Empirical Bayes, Unweighted, IRC1 and IRC2 estimators. In the case of non-informative right censoring with known variance parameters, IRC3 slope estimator possesses the least MSE, followed by Fearn’s Weighted, Empirical Bayes, IRC1, Unweighted and IRC2 estimators. These two observations put together implies that
under both cases of informative and non-informative right censoring, the IRC3 estimators predicts better the individual slope parameters, when the variance structure is known.

The scenario under unknown variance is similar to that of the known variance case. In the case of informative right censoring with unknown variance parameters, replaced by Henderson’s variance estimate, IRC3 estimator possesses the least MSE. This is followed by IRC1, Fearn’s Weighted, Empirical Bayes, IRC2 and Unweighted estimators. In the case of non-informative right censoring with unknown variance parameters, replaced by Henderson’s variance estimate, IRC3 estimator possesses the least MSE. This is followed by IRC1, Fearn’s Weighted, Empirical Bayes, IRC2 and Unweighted estimators. In case of informative right censoring with unknown variance parameters, replaced by Mori’s variance estimate, IRC3 estimator possesses the least MSE. This is followed by IRC1, Fearn’s Weighted, Empirical Bayes, IRC2 and Unweighted estimators. In the case of non-informative right censoring with unknown variance parameters, replaced by Mori’s variance estimate, IRC3 estimator possesses the least MSE. This is followed by IRC1, Fearn’s Weighted, Empirical Bayes, IRC2 and Unweighted estimators. On consolidation, it seems that the IRC3 estimator behaves better in terms of reduced MSE in case of informative and non-informative right censoring, under both known and unknown variance structures, in the selected parametric values. In the case of unknown variance structure, a comparison between the Henderson’s estimator and Mori’s
estimator reveals that Henderson estimator has uniformly less MSE under all scenarios.

In Chapter V, the censoring and the response variables are generated using truncated Geometric and Normal distributions respectively, maintaining the structure in chapter II, to enable better comparison. Both univariate and bivariate slope models are considered. Also under bivariate model different correlation structures are discussed. This chapter compares MSEs of individual slope estimators among Unweighted, Fearn’s Weighted, Empirical Bayes, IRC1, IRC2, IRC3, MIRC1, MIRC3, MIRC4, MIRC5, MIRC6, MIRC7 and MRLWS estimators, under informative and non-informative right censoring, with unknown variance structures. As this chapter focuses more on the marginal likelihood estimation, the case of unknown variance parameters is considered and the case of known variances is not discussed.

In the case of univariate model with informative right censoring with unknown variance parameters, IRC3 estimator possesses the least MSE. Going by the least MSE values, it is seen that IRC3 behaves better in both the informative and non-informative censoring scenarios. The structure of IRC3 estimates the censoring parameters using the Weighted Least square technique. In this case, the variance parameters are estimated using Henderson method. This gives the edge for the IRC3 estimator over other estimators, which involve partial or full utilization of marginal
likelihood estimates. This highlights the comparatively poor performance of the marginal likelihood estimation procedure. This is even more evident by the fact that IRC3 serves better in terms of reduced MSE under both informative and non-informative right censoring, a pattern reflected in Chapter IV. Thus, it is recommended that for the univariate model with informative or non-informative right censoring, IRC3 can be used for effective estimation of the individual slope parameters.

For the bivariate slope model, going by the MSE values, it is seen that MIRC4 and MIRC5 possess relatively less MSE, in most of the cases. For the first response, in case of informative right censoring with negative and zero correlations, MIRC4 possesses the least MSE. In case of informative right censoring with positive correlation, MIRC5 possesses the least MSE. Similarly, for the first response, in case of non-informative right censoring with negative and zero correlations, MIRC7 possesses the least MSE. In case of non-informative right censoring with positive correlation, Unweighted estimator possesses the least MSE and this leaves no idea or reason for its better behaviour compared to the other estimators that make use of more information from the sample.

For the second response, in case of informative right censoring with positive and negative correlations, MIRC5 possesses the least MSE. In case of informative right censoring with zero correlation, MIRC4 possesses the least MSE. Similarly, for the second response, in case of non-informative right censoring with negative and
zero correlations, MIRC7 possesses the least MSE. In case of non-informative right censoring with positive correlation, MRC5 estimator possesses the least MSE.

In the case of bivariate model, the performance of MIRC estimators under different scenarios needs to be justified for possible reasons. In case of informative right censoring, MIRC4 and MIRC5 tend to behave better, in terms of reduced MSE, whereas in the case of non-informative right censoring MIRC7 behaves better. The structure of MIRC4 and MIRC5 estimates the censoring parameters using Geometric component of likelihood and weighted least square technique respectively. In both the cases, the variance parameters are estimated using Henderson method. This gives the edge for the MIRC4 and MIRC5 estimators over MIRC6 and MIRC7 estimators, which involves partial or full utilization of marginal likelihood estimates. This highlights the comparatively poor performance of the marginal likelihood estimation procedure. Thus, it is recommended that for the scenario with informative right censoring MIRC4 or MIRC5 can be used for effective estimation of the individual slope parameters. For the case with non-informative right censoring MIRC6 or MIRC7 can be used for effective estimation of the individual slope parameters.

The research work finally ends up with preferred estimation process, for data under different scenarios. The Weighted slope estimators behave better in comparison to the Unweighted, under informative right censoring with known variance structure, informative right censoring with unknown variance structure, non-informative censoring with known variance structure and non-informative censoring
with unknown variance structure. For the model described for empirical Bayesian structure, empirical Bayes estimator behaves better in comparison to the Unweighted and Fearn’s Weighted estimators, under informative right censoring with known variance structure. In case of non-informative right censoring with known variance parameters, Fearn’s Weighted slope estimator possesses the least MSE compared to Unweighted and empirical Bayes estimators. The case of unknown variance is similar to that of the known variance case. The highlight of the thesis is the development of IRC3 estimator, which out performs all the remaining individual slope estimators considered under most of the scenarios. For bivariate model with the scenario with informative right censoring MIRC4 or MIRC5 can be used for effective estimation of the individual slope parameters. For the case with non-informative right censoring MIRC6 or MIRC7 can be used.

**SUGGESTIONS FOR FURTHER RESEARCH**

The parameter structures considered are varied considerably and it is seen that the results derived does not undergo major changes, though this was not presented and discussed in this thesis. Some sort of sensitivity analysis in this direction is possible, that may bring out some interaction effects of the parameters in higher dimensional spaces. Different algorithms, apart from GA or ‘optim ()’ functions could be used to verify the global optimal nature of the likelihood estimators. Seed variation in the simulation process does not bring major changes in the result
patterns. The sample size and replications, in the case of marginal likelihood estimation process in chapter V could have been increased, of course with considerable increase in the runtime of the algorithms. As shown in the screen shot of the summary of the GA output for the bivariate slope model of individual parameters with informative right censoring with correlation 0.4, the processing time is nearly four hours. For some other combinations the processing time even touched 23 hours. This is the case with Intel-i5 processor with 4GB RAM configuration on Windows platform. Higher sample size with increased replications can be achieved, possibly by some parallel processing or cloud computing systems. This thesis uses the MSE as a measure of accuracy for estimators and for further research the coverage probability can be used to improve the simulation process.