

BAYESIAN MEASUREMENT ERROR MODELING WITH APPLICATION TO
THE AREA UNDER THE CURVE SUMMARY MEASURE

by

Jennifer Lee Weeding

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Statistics

MONTANA STATE UNIVERSITY
Bozeman, Montana

April, 2016

©COPYRIGHT

by

Jennifer Lee Weeding

2016

All Rights Reserved

DEDICATION

I dedicate this dissertation to my family, who have provided limitless support and motivation throughout my graduate career.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Mark Greenwood, for the support, encouragement, and patience he has provided throughout my graduate student career. His guidance has helped me grow as a professional and I will be forever grateful for all he has done for me. I cannot over-emphasize how lucky I am to have been able to work with such a great mentor.

I would also like to thank my other committee members, Dr. Steve Cherry, Dr. John Borkowski, Dr. Jim Robison-Cox, Dr. Shannon Moreaux, and Dr. Megan Higgs. I am most thankful for the guidance that you all have provided throughout my graduate studies. Thank you for the invaluable comments you provided on this dissertation. I cannot thank you enough for helping to mold me into the statistician I am today.

I would like to thank the entire Mathematical Sciences department for their kindness and support throughout my studies. The office staff (specifically, Becky, Josie, and Katie) have all been great to work with and extremely helpful in making sure I did not miss any important deadlines. I would also like to thank my undergraduate advisor, Dr. John Lund. He has been a mentor and a friend throughout this journey and his advice and sense of humor are both things I will always be thankful for.

Finally, I would like to thank Dr. Mark Owkes for providing and maintaining the LaTeX template that was used for this dissertation. Numerous questions relating to formatting matters were quickly answered and example code was often provided to fix any problems. Graduate students at Montana State University are extremely fortunate to have the help of Dr. Owkes.

TABLE OF CONTENTS

| | |
|--|----|
| 1. INTRODUCTION | 1 |
| References | 10 |
| 2. AN INTRODUCTION TO THE AREA UNDER THE CURVE SUMMARY MEASURE..... | 11 |
| Contribution of Authors and Co-Authors..... | 11 |
| Manuscript Information Page..... | 12 |
| Abstract | 13 |
| 2.1 Introduction | 13 |
| 2.2 Modifications..... | 15 |
| 2.2.1 Two Modified Versions of the AUC..... | 16 |
| 2.2.2 Confusion Regarding the Three AUCs..... | 17 |
| 2.3 Other Considerations | 20 |
| 2.3.1 Sampling Design..... | 20 |
| 2.3.2 Reproducible Research..... | 22 |
| 2.3.3 AUCs in Analyses..... | 23 |
| 2.4 Discussion | 24 |
| References | 27 |
| 3. BAYESIAN ESTIMATION OF VARIANCE FOR AUC STUDIES | 30 |
| Contribution of Authors and Co-Authors..... | 30 |
| Manuscript Information Page..... | 31 |
| Abstract | 32 |
| 3.1 Introduction | 32 |
| 3.2 Methods..... | 35 |
| 3.2.1 Formula-Based Estimators..... | 36 |
| 3.2.2 Bayesian ANOVA Model..... | 37 |
| 3.2.3 Bayesian AR(1) Model..... | 38 |
| 3.3 Applications | 40 |
| 3.3.1 Simulated Data Study..... | 42 |
| 3.3.2 Equine Data | 48 |
| 3.4 Discussion | 51 |
| References | 57 |

TABLE OF CONTENTS - CONTINUED

| | |
|--|-----|
| 4. A COMPARISON OF MEASUREMENT ERROR CORRECTION METHODS WITH AND WITHOUT KNOWN MEASUREMENT ERRORS..... | 59 |
| Contribution of Authors and Co-Authors..... | 59 |
| Manuscript Information Page..... | 60 |
| Abstract..... | 61 |
| 4.1 Introduction | 61 |
| 4.2 Measurement Error in Simple Linear Regression | 63 |
| 4.3 Review of Three Common Correction Methods | 65 |
| 4.3.1 Method of Moments..... | 65 |
| 4.3.2 SIMEX | 66 |
| 4.3.3 Bayesian Measurement Error Models | 68 |
| 4.4 Simulation Studies | 70 |
| 4.4.1 Known Measurement Error | 71 |
| 4.4.2 Unknown Measurement Error | 73 |
| 4.5 Discussion | 75 |
| 4.6 Appendix | 81 |
| 4.6.1 Bayesian Measurement Error Model Formulation..... | 81 |
| 4.6.2 Simulation Study 1 Code | 81 |
| 4.6.3 Simulation Study 2 Code | 83 |
| References | 87 |
| 5. IMPACTS OF MEASUREMENT ERROR IN AN AREA UNDER THE CURVE STUDY..... | 89 |
| Contribution of Authors and Co-Authors..... | 89 |
| Manuscript Information Page..... | 90 |
| Abstract..... | 91 |
| 5.1 Introduction | 91 |
| 5.2 Methods | 93 |
| 5.2.1 Equine Data..... | 93 |
| 5.2.2 Statistical Analyses..... | 94 |
| 5.3 Results | 103 |
| 5.3.1 Glucose Parameter Estimation..... | 103 |
| 5.3.2 Simple Linear Regression Analyses..... | 105 |
| 5.4 Discussion | 107 |
| 5.5 Appendix | 109 |
| 5.5.1 Bayesian Model Specifications | 109 |

TABLE OF CONTENTS - CONTINUED

| | |
|--|-----|
| 5.5.2 Analysis Code | 112 |
| References | 116 |
| 6. A BAYESIAN MEASUREMENT ERROR CORRECTION MODEL FOR ERRORS IN BOTH EXPLANATORY AND RESPONSE VARIABLES | 119 |
| Contribution of Authors and Co-Authors..... | 119 |
| Manuscript Information Page..... | 120 |
| Abstract | 121 |
| 6.1 Introduction | 121 |
| 6.2 Methods | 124 |
| 6.3 Simulation Studies | 126 |
| 6.3.1 Simulation Study: Case 3..... | 128 |
| 6.3.2 Simulation Study: Case 4..... | 129 |
| 6.4 Discussion | 133 |
| 6.5 Appendix | 139 |
| References | 143 |
| 7. CONCLUSIONS AND FUTURE WORK | 144 |
| References | 149 |
| REFERENCES CITED..... | 151 |

LIST OF TABLES

| Table | Page |
|---|------|
| 3.1 $SD(\widehat{AUC})$ for each method corresponding to Figure 3.2..... | 45 |
| 3.2 $SD(\widehat{AUC})$ for each method corresponding to Figure 3.3..... | 47 |
| 3.3 Variances used in the non-constant variance simulation..... | 48 |
| 3.4 $SD(\widehat{AUC})$ for each method corresponding to Figure 3.4..... | 49 |
| 3.5 Estimated AUCs for the equine data..... | 51 |
| 3.6 $SD(\widehat{AUC})$ for the equine data..... | 51 |
| 4.1 Simulation Study 1 parameter values. | 72 |
| 4.2 Simulation 1 results for $\hat{\beta}_1$ for each method..... | 73 |
| 4.3 Simulation Study 2 parameter values. | 75 |
| 4.4 Simulation 2 results for $\hat{\beta}_1$ for each method..... | 75 |
| 5.1 Insulin and glucose data fro each horse. | 106 |
| 5.2 Parameter estimates and 90% credible intervals from the naive model and the measurement error model..... | 107 |
| 6.1 Simulation study parameter values for each Case considered. | 128 |
| 6.2 Case 3 simulation results for $\hat{\beta}_1$ for each method. | 131 |
| 6.3 Case 4.1 simulation results for $\hat{\beta}_1$ for each method..... | 132 |
| 6.4 Case 4.2 simulation results for $\hat{\beta}_1$ and $\hat{\rho}_{u_i, q_i}$ for the two Bayesian models..... | 133 |
| 6.5 Posterior means and 90% credible intervals for $\hat{\beta}_1$ and $\hat{\rho}_{u_i, q_i}$ from one simulation for each scenario in Case 4.2..... | 133 |

LIST OF FIGURES

| Figure | Page |
|--|------|
| 2.1 Graphical displays of AUC_G , AUC_I , and AUC_{AB} | 17 |
| 2.2 AUC_G and AUC_I computed for the same individual..... | 19 |
| 2.3 AUC_G and AUC_I computed for two individuals on the same day..... | 20 |
| 2.4 AUC computed three ways from the same data. | 23 |
| 3.1 An example of two patterns commonly seen in equine-based glucose AUC data..... | 41 |
| 3.2 Estimated curves from each model using the mean values. | 44 |
| 3.3 Estimated curves for each model using $m_i = 2, 4$ and $\sigma_\epsilon = 4, 100$ | 46 |
| 3.4 Estimated curves for each model using data simulated with non-constant variance. | 55 |
| 3.5 Estimated curves for each model using the equine data..... | 56 |
| 4.1 Example of the bias present in β_1 for three values of k | 65 |
| 4.2 Example of the three SIMEX extrapolation functions. | 68 |
| 4.3 Simulation 1 results for $\beta_1 = 1$ for each method..... | 78 |
| 4.4 Simulation 1 results for $\beta_1 = 5$ for each method..... | 79 |
| 4.5 Simulation 2 results for $\beta_1 = 5$ for each method..... | 80 |
| 5.1 Observed glucose curves for each horse. | 94 |
| 5.2 Scatterplot of the equine data. | 102 |
| 5.3 The estimated curve resulting from the AR(1) model for each horse. | 105 |
| 6.1 Case 3 simulation results for $\beta_1 = 5$ for each method. | 130 |
| 6.2 Case 4.1 simulation results for $\beta_1 = 5$ for each method..... | 136 |
| 6.3 Case 4.2 simulation results for $\beta_1 = 5$ for the two Bayesian models. | 137 |
| 6.4 Posterior distributions from one simulation for β_1 | 138 |

ABSTRACT

Measurement errors arise in a variety of circumstances and occur when a variable cannot be observed exactly, but instead is observed with error. For example, summary measures contain measurement error, as the true value of the variable is estimated from observed data that contain sampling variability. Measurement errors should be accounted for when they are present, as the impacts of ignoring measurement errors include bias in parameter estimates and a loss of power to detect effects. Measurement error models are used to account for measurement errors and correct parameter estimates for the bias induced from variables measured with error. To account for measurement errors when present, most correction methods require that the measurement error variance be known (or estimated). Common correction methods include the method of moments correction, the SIMEX correction, and Bayesian correction methods.

The area under the curve (AUC) summary measure is commonly used in pharmaceutical studies to estimate the total concentration of a substance present in the blood over a given time interval. Other areas, such as Ecology, use the AUC to estimate the total count of a species present over a specified time interval. In situations where the AUC is estimated, a measure of the uncertainty associated with it is often desired. Due to the longitudinal nature of AUC data, the estimation of its variance is often not straightforward.

In this research we develop a Bayesian method to estimate the variance of the AUC, where our focus is on accounting for the possible correlation structure between repeated observations on the same subject. This estimate can then be used in measurement error models to account for the measurement error induced from estimating the AUC. We study the performance of three measurement error correction methods in the simple linear regression setting, where measurement errors are present in the explanatory variable, the response variable, or both. We extend the Bayesian correction methods to account for uncorrelated and correlated measurement errors between variables. The methods were validated using both simulated and real data collected from an equine study of blood glucose measurements.

CHAPTER ONE

INTRODUCTION

The purpose of this research is to develop and study statistical methods that account for measurement error when present in continuous random variables and to develop an estimator for the variance associated with the area under the curve (AUC) summary measure for use in measurement error models. The motivation originates from the use of summary measures in subsequent analyses, in particular the AUC summary measure. An introduction to the AUC summary measure is provided in Chapter 2, attempting to clarify a variety of aspects of the AUC that are important in using it. In Chapter 3 we propose a Bayesian method to estimate the variance associated with an estimated AUC when the linear trapezoidal method is used to estimate the AUC. When estimated AUCs are used in statistical models, measurement error is present and the results in Chapter 3 for estimating the variance in the estimated AUCs provide the information required to account for measurement error in these models. In Chapter 4, comparisons of three common measurement error correction methods are presented for the simple linear regression setting where measurement error is present in the explanatory variable only. An application involving estimated AUCs that are used in subsequent analyses is presented in Chapter 5, where a Bayesian measurement error model is used to correct for the measurement error, combining ideas presented in Chapters 2, 3, and 4. A novel Bayesian measurement error model that accounts for correlated measurement errors between the response and explanatory variables in the simple linear regression setting is proposed in Chapter 6, where simulation studies are used to study the performance of the method.

Measurement error occurs when a variable of interest cannot be observed exactly, but is instead observed with error (Buonaccorsi, 2010). For example, when there is miscalibration of a measuring instrument, sampling error, misclassification, or when responses are estimates, measurement errors arise. Variables measured with error are often used in statistical models where the measurement errors are ignored. The consequences of ignoring measurement errors in analyses include bias in parameter estimates and a loss of power (Carroll et al., 2006).

Measurement error in a continuous random variable can be classified as either Berkson or classical, where the fundamental difference between the two types of error can be seen in the distribution assumed for the errors. Berkson measurement error (Berkson, 1950) occurs when a researcher is trying to achieve a target value (W_i) but instead achieves a true value (X_i) that is unknown. Classical measurement error occurs when a researcher is trying to measure the true value of a variable (X_i) but instead measures the variable with error resulting in an observed value (W_i). The Berkson measurement error model assumes a distribution for the true values given the observed values ($X_i|W_i$), while the classical measurement error model assumes a distribution for the observed values given the true values ($W_i|X_i$).

The additive error structure is the most common measurement error structure used and the focus here, however the linear and multiplicative error structures can also be used. The classical additive measurement error model is

$$W_i = X_i + U_i, \tag{1.1}$$

where U_i represents the measurement errors, $E(U_i|x_i) = 0$, and $Var(U_i|x_i) = \sigma_u^2$. A direct result of the additive error structure is that the mis-measured variable (W_i) is an unbiased estimator of the true variable (X_i).

The simple linear regression (SLR) model is implemented in situations where a linear relationship between two quantitative variables is of interest. Measurement error can arise in the explanatory variable (X), the response variable (Y), or both, where a relatively simple statistical model changes into something much more complicated. If classical additive measurement error is assumed in both the response and explanatory variables, and (D_i, W_i) represent the mis-measured response and explanatory variables, the SLR model becomes

$$\begin{aligned} Y_i + Q_i &= \beta_0 + \beta_1(X_i + U_i) + \epsilon_i \\ \Rightarrow D_i &= \beta_0 + \beta_1 W_i + \epsilon_i, \end{aligned} \tag{1.2}$$

where $Q_i|y_i \sim N(0, \sigma_q^2)$, $U_i|x_i \sim N(0, \sigma_u^2)$, $Cov(Q_i, U_i) = \sigma_{uq}$ and $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. The slope coefficient (β_1) is of primary interest in this setting and Buonaccorsi (2010) provides the following equation for the ordinary least squares estimator of β_1 when measurement error is present in both the response and explanatory variables as a function of the true slope, β_1 :

$$\beta_{1,naive} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_u^2} \beta_1 + \frac{\sigma_{uq}}{\sigma_X^2 + \sigma_u^2}. \tag{1.3}$$

The estimator of β_1 using least squares is biased toward 0, called attenuation, when measurement error is present in the explanatory variable only or when $\sigma_{uq} = 0$. In this scenario, the ordinary least squares estimator of β_1 becomes

$$\beta_{1,naive} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_u^2} \beta_1 = k\beta_1, \tag{1.4}$$

where $k = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_u^2}$ and is referred to as the reliability ratio (Fuller, 2009). Most measurement error research is focused on correcting for the attenuation caused by

the measurement error in the explanatory variable, and commonly ignore the effects of measurement error in the response variable. Carroll et al. (2006) conclude that the only effects of additive error in the response variable are an increase in the variability of the fitted lines and a decrease in power to detect effects. However, we believe they are referring to the situation where measurement error is present in the response variable only, or uncorrelated with the measurement error in the explanatory variable, as Equation 1.3 shows that correlated measurement errors in the response and explanatory variable have an effect on the estimation of β_1 .

Measurement error models are statistical models that account for measurement errors and adjust parameter estimates for the bias induced from variables measured with error. The type of measurement error present (Berkson or classical), the form of the error structure (additive, linear, multiplicative, etc.), and knowledge of the measurement error variances must all be known to account for measurement errors when present. Common measurement error correction methods for continuous variables measured with error include the method of moments correction (Buonaccorsi, 2010), the SIMEX correction (Cook and Stefanski, 1994), regression calibration (Carroll et al., 2006, Chapter 4), and Bayesian methods (Carroll et al., 2006, Chapter 9).

Summary measures reduce multiple responses on an individual into a single number through a function of the observed data. The true underlying value of a variable is often estimated from observed data using a summary measure; for example, the mean, the median, or the peak value. If a summary measure is an unbiased estimator of the true value, classical additive measurement error is present where the summary measure represents the mis-measured variable (W_i) and the true value of the variable is represented using X_i . Summary measures contain sampling variability, as different realizations of the information used to calculate the statistic produce different estimates. The measurement error variance associated with the summary measure,

σ_u^2 , can be estimated using the estimate of the variance associated with the summary measure (if one is known).

The AUC summary measure is used to estimate the total magnitude of a response over time, and is defined as the definite integral of the true underlying function of the response over the given time interval (Stewart, 2001). In most studies, the true underlying function of the response is unknown and the AUC must be estimated from observed data. Research by Purves (1992) and Yeh and Kwan (1978) look at the performance of multiple different integration techniques, including the linear trapezoidal method, and conclude that all methods studied contain bias and that the direction and magnitude of the bias depends on the true underlying function of the response.

The linear trapezoidal method is often preferred over more complex AUC estimation methods due to its simplicity (Yuan, 1993), along with the fact that it performs the same or better in situations with a small number of data points (Allison et al., 1995). The formula for estimating the AUC using the linear trapezoidal method is:

$$\begin{aligned} \widehat{AUC} &= \frac{(y_1 + y_2) * (t_2 - t_1)}{2} + \frac{(y_2 + y_3) * (t_3 - t_2)}{2} + \dots + \frac{(y_{n-1} + y_n) * (t_n - t_{n-1})}{2} \\ &= \sum_{i=1}^{n-1} \frac{(y_i + y_{i+1}) * (t_{i+1} - t_i)}{2} \end{aligned} \quad (1.5)$$

where (y_i, t_i) represent the responses and the corresponding time points for an individual, and $i = 1, 2, \dots, n$ index the responses collected. The simplicity of this method can be seen when examining Equation 1.5, as this estimate of the AUC is

found by summing the areas of all trapezoids formed using consecutive time points and their responses.

One major limitation of the linear trapezoidal method of estimating the AUC is that in most situations there is not a straight-forward method to calculate the variance associated with it (Hilborn et al., 1999). Equation 1.5 shows that the AUC is estimated using a linear combination of the responses (y_i 's), however due to the longitudinal nature of the responses in AUC data, the correlation between repeated observations should be accounted for. The correlation between repeated observations is unknown in most situations, which makes the estimation of the variance of the \widehat{AUC} that incorporates this dependency difficult.

Modifications of the AUC summary measure have been proposed that measure different characteristics of the response profiles. Two commonly used alternatives include the area under the curve with respect to increase, denoted AUC_I , and the area under the curve above the baseline (or initial) value, denoted AUC_{AB} . Both modified measures are measuring an index of change, however they are measuring a different underlying characteristic of the data. The exact characteristic of the data that each version of the AUC is measuring is often misunderstood, leaving researchers unsure of which version of the AUC they should use in their research. In Chapter 2 we examine these three versions of AUCs and explore the characteristic that each AUC is estimating. We provide recommendations for choosing the appropriate AUC based on the study design, and present specific details that need to be considered and reported when using AUCs.

AUCs that are estimated from observed data contain classical, additive measurement error, which should be accounted for in subsequent analyses in order to obtain accurate parameter estimates. In order to account for the measurement error induced from estimating the AUC, an estimate of the variance associated with the

\widehat{AUC} is needed. We present a Bayesian approach to modeling AUC data in Chapter 3, using two different statistical models, where the variance-covariance matrix of the repeated observations can be estimated using values computed from the resulting posterior distributions. An ANOVA model and a first-order autoregressive (AR(1)) model are considered, where a new estimate of the variance associated with the \widehat{AUC} is obtained using the corresponding estimated variance-covariance matrix provided from each model. We study the performance of the method using simulated data, where lower and upper bounds for the variance of the \widehat{AUC} are computed assuming the correlation between repeated observations is either 0 (independent observations) or 1 (perfectly positively correlated observations). The methods are applied to an equine-based glucose data set, where some difficulties that may arise when applying these methods to real data are explored.

The method of moments correction, the SIMEX correction, and Bayesian measurement error models are all common methods used to correct for measurement errors when present. In Chapter 4 we compare the performance of the three methods in the simple linear regression setting where classical additive measurement errors are present in the explanatory variable. We explore situations where W_i is based on a single observation and other situations where it is based on the mean or a function of a set of means of observations. In these situations, researchers have the potential to design a study to impact their measurement error variances by increasing the amount of information used to calculate the mis-measured variable. In other situations, measurement errors could be impacted by improving the instrument used to obtain the measurement(s) which could have smaller variances. Two simulation studies are conducted, where the measurement errors were either assumed known or estimated from replicate data, to study the performance of the methods in relation to

the magnitude of the measurement errors present and changing sample sizes. When the measurement error variance was assumed known, the simulation results showed similar performance of the three methods when the measurement errors were small, however differences in the methods emerged as the magnitude of the measurement errors increased. Overall, the Bayesian method displayed better performance when the measurement errors were assumed known. The second simulation study was used to explore the performance of the three methods when the measurement error variance is estimated from replicate data. The three correction methods all produce approximately unbiased estimates of the slope when the measurement errors were small, however as the magnitude of the measurement errors increase, the method of moments correction and the SIMEX correction are highly variable and depend on how well the measurement error variance was estimated. Results from the Bayesian method contained less bias in all cases considered, and were less variable than the other methods when the measurement errors were unknown.

A data analysis from an equine study involving glucose AUCs is presented in Chapter 5, where the SLR model is used. We first obtain an estimate of the measurement error variance associated with the \widehat{AUC} using the methods presented in Chapter 3, then perform an analysis that accounts for the measurement error in the \widehat{AUC} explanatory variable using a Bayesian measurement error model (given in Chapter 4). We compare the results from the measurement error model to results obtained from a model where the measurement errors are ignored to highlight the importance of accounting for measurement errors when they are present.

The simple linear regression model has four scenarios in which measurement errors can be present: measurement error in the explanatory variable only, measurement error in the response variable only, uncorrelated measurement error in both the response and explanatory variable, and correlated measurement error in the response

and explanatory variable. Chapter 4 examines the performance of three common correction methods when measurement error was present in the explanatory variable only. In Chapter 6, we explore the two other scenarios where measurement error induces bias in the estimate of β_1 in the SLR setting. A new Bayesian method is presented to account for correlated measurement errors, and simulation studies are used to compare the proposed Bayesian method to the method of moments correction, the SIMEX correction (which does not account for correlated measurement errors), and the GSIMEX correction (Ronning and Rosemann, 2008), which is an extension of the SIMEX correction that corrects for correlated measurement errors. Future research and other considerations are presented in Chapter 7.

References

- Allison, D. B., Paultre, F., Maggio, C., Mezzitis, N., Pi-Sunyer, F. X., 1995. The use of areas under curves in diabetes research. *Diabetes care* 18 (2), 245–250.
- Berkson, J., 1950. Are there two regressions? *Journal of the american statistical association* 45 (250), 164–180.
- Buonaccorsi, J. P., 2010. *Measurement error: models, methods, and applications*. CRC Press.
- Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M., 2006. *Measurement error in nonlinear models: a modern perspective*. CRC press.
- Cook, J. R., Stefanski, L. A., 1994. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical association* 89 (428), 1314–1328.
- Fuller, W. A., 2009. *Measurement error models*. Vol. 305. John Wiley & Sons.
- Hilborn, R., Bue, B. G., Sharr, S., 1999. Estimating spawning escapements from periodic counts: a comparison of methods. *Canadian Journal of Fisheries and Aquatic Sciences* 56 (5), 888–896.
- Purves, R. D., 1992. Optimum numerical integration methods for estimation of area-under-the-curve (auc) and area-under-the-moment-curve (aumc). *Journal of pharmacokinetics and biopharmaceutics* 20 (3), 211–226.
- Ronning, G., Rosemann, M., 2008. Simex estimation in case of correlated measurement errors. *AStA Advances in Statistical Analysis* 92 (4), 391–404.
- Stewart, J., 2001. *Calculus concepts and contexts*, 2nd edn (pacific grove, ca: Brooks/cole).
- Yeh, K., Kwan, K., 1978. A comparison of numerical integrating algorithms by trapezoidal, lagrange, and spline approximation. *Journal of pharmacokinetics and biopharmaceutics* 6 (1), 79–98.
- Yuan, J., 1993. Estimation of variance for auc in animal studies. *Journal of pharmaceutical sciences* 82 (7), 761–763.

CHAPTER TWO

AN INTRODUCTION TO THE AREA UNDER THE CURVE SUMMARY
MEASURE

Contribution of Authors and Co-Authors

Author: Jennifer L. Weeding

Contributions: Responsible for the majority of the writing.

Co-Author: Dr. Mark C. Greenwood

Contributions: Provided feedback on statistical analysis and drafts of the manuscript.

Manuscript Information Page

Jennifer L. Weeding and Mark C. Greenwood

Journal of Dairy Science

Status of Manuscript:

Prepared for submission to a peer-reviewed journal

Officially submitted to a peer-review journal

Accepted by a peer-reviewed journal

Published in a peer-reviewed journal

Abstract

The area under the curve (AUC) summary measure reduces repeated measures data on an individual into a single number through a function of the observed data that captures cumulative information on the individual over some period of time. AUCs are often estimated from observed data and the linear trapezoidal method is the most commonly used method to estimate them. Variants of the AUC, such as area under the curve increase (AUC_I) and area under the curve above baseline (AUC_{AB}), have led to a lack of clarity regarding what aspect of the data each version of the AUC is capturing. Here, the differences among the modifications of the AUC summary measures are explained, scenarios are discussed where one version of AUC is favored over others, important characteristics of the study design in regards to computing accurate and precise AUCs are detailed, and we provide recommendations for information that needs to be given for research that involves AUC responses to be reproducible.

2.1 Introduction

The total concentration of a substance present in the blood over time is often of interest to researchers studying pharmacodynamic and pharmacokinetic properties in either humans (for example, Kashyap et al. (2013)) or animals (for example, Moreaux et al. (2011)); the same techniques have been used in Ecology where the total count of a species present over a time interval (for example, Parken et al. (2003)) may be of interest. The area under the curve (AUC) is a summary measure that captures the total magnitude of a response over a specified time interval and has a long history in pharmacology (Armitage et al., 2008). In general, a summary measure reduces longitudinal data (or other types of multiple responses) on an individual into a single number through a function of the observed data for use in some secondary analysis of a data set of n of these summary measures. The most common summary measure is the mean, but other measures such as the median or the rate of change (slope coefficient) are also used; any function of the observed data that captures an important feature

of the data could be a useful summary measure and analyzed, discarding the raw observations.

The effect of a treatment on the total concentration of a substance in the blood is the focus of many designed pharmacodynamic studies. The AUC is an estimate of the total concentration and often serves as the response variable in these studies. The AUC is often used to compress a period of hours over a single day into a single “magnitude” measure, and these responses are compared across multiple groups on a single day. Other studies track changes in the AUCs over time, where repeated AUCs on the same subjects are computed on multiple days, requiring a model that accounts for these repeated measures. The AUC is most often used as a response variable in a study where different treatment groups are compared. However there are situations where it can be used as an explanatory variable as well. For example, another important aspect in pharmacodynamic studies is the peak value of a substance in the blood. The total concentration (AUC) of one substance may help predict the size of the peak value of another substance.

The AUC is defined as the definite integral of the true underlying function of the response, $f(t)$, over the time interval (a, b) . Oftentimes, the true underlying function is unknown, such as in experimental studies that are focused on the effect of a treatment on the total concentration of a substance present in the blood. When this occurs, the AUC must be estimated. The performance of multiple different numerical integration techniques was the focus of research carried out by Purves (1992) and Yeh and Kwan (1978). The results from both studies illustrate that all methods considered contain bias, however the direction and magnitude of the bias depends on the true underlying function of the substance. The linear trapezoidal method is one of the many numerical integration techniques they studied and is found by simply summing the areas of all trapezoids formed using consecutive time points and their

responses. For example, if (y_i, t_i) represent the responses and the corresponding time points for an individual, and $i = 1, 2, \dots, n$ index the responses collected, the formula for estimating the AUC using the linear trapezoidal method is:

$$\begin{aligned} \widehat{AUC} &= \frac{(y_1 + y_2) * (t_2 - t_1)}{2} + \frac{(y_2 + y_3) * (t_3 - t_2)}{2} + \dots + \frac{(y_{n-1} + y_n) * (t_n - t_{n-1})}{2} \\ &= \sum_{i=1}^{n-1} \frac{(y_i + y_{i+1}) * (t_{i+1} - t_i)}{2}. \end{aligned} \quad (2.1)$$

This method is often preferred over more complex AUC estimation methods due to its simplicity (Yuan, 1993), along with the fact that it has good performance in situations involving a small number of data points (Allison et al., 1995).

The AUC as defined above has been modified to create other similar summary measures that focus on different aspects of change over time. These alternative AUC measures are presented in Section 2.2 along with a discussion of the confusion they have created. Section 2.3 examines different sampling schemes, reproducible research, and the effects of using different versions of AUCs in analyses, while Section 2.4 contains concluding remarks about the AUC summary measure.

2.2 Modifications

Various modifications of the AUC summary measure have been proposed that measure different characteristics of the response profiles. We will denote the AUC discussed previously as AUC_G , with the G subscript indicating that it is the AUC calculated with respect to ground, or a response value of 0, which corresponds to an absence of the substance. Some researchers use a T subscript to represent this same AUC, with the subscript indicating that it is the total AUC.

2.2.1 Two Modified Versions of the AUC

The most commonly used alternative to AUC_G is the area under the curve with respect to increase, denoted AUC_I , which is defined as the area under the curve above the baseline value (the baseline value is often, but not always, the initial measurement) minus the area above the curve that is below the baseline value (Le Floch et al., 1990). AUC_I can be negative, therefore it should not be thought of as a ‘true’ area, but instead as an index of change in the response over the sampling period. To calculate the AUC_I , you first find the AUC_G , then subtract off the AUC that would result if the responses had remained constant over the entire time interval. Given that there are n total measurements and y_1 is the initial measurement, the mathematical formula is:

$$\widehat{AUC}_I = \widehat{AUC}_G - [y_1 * (t_n - t_1)]. \quad (2.2)$$

Another commonly used variation of AUC is the area under the curve above the baseline value, denoted AUC_{AB} . This version of AUC calculates the area under the curve above the baseline value, ignoring any area that falls below the baseline value (Wolever and Jenkins, 1986). This modification of the AUC_G is also an index of change, and is strictly related to the total positive change from the baseline value over the time interval. The formula for calculating the AUC_{AB} is more complicated, as you must find the point(s) of intersection where the estimated function of the response crosses the baseline value, and then compute only the area above the baseline value. Fekedulegn et al. (2007) provide a detailed example of calculating the AUC_{AB} . A graphical representation of each AUC version can be seen in Figure 2.1.

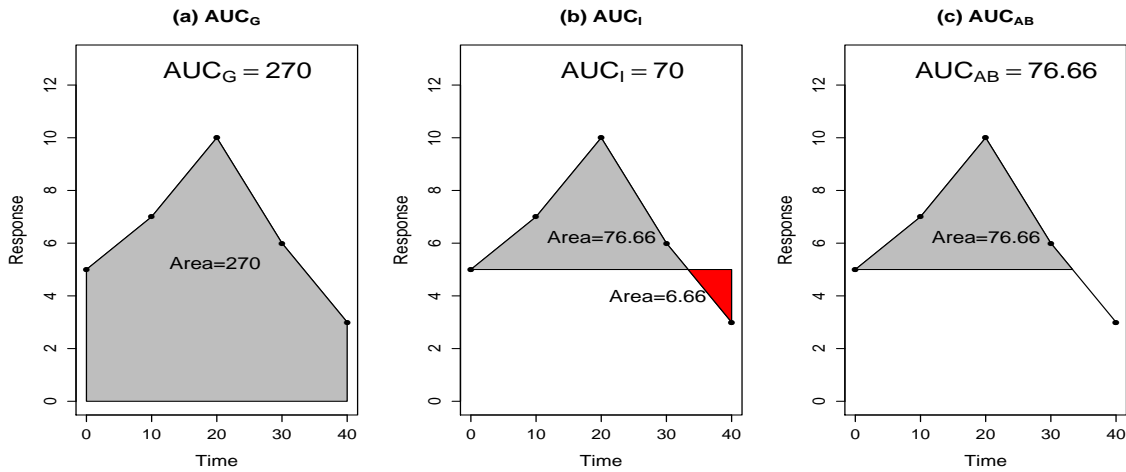


Figure 2.1: Graphical displays showing the areas used in the calculation of (a) AUC_G , (b) AUC_I , and (c) AUC_{AB} .

2.2.2 Confusion Regarding the Three AUCs

Modifications of a known measure can sometimes lead to confusion regarding the measures themselves. The three versions of AUC have often been misunderstood and assumed to be measuring the same characteristic of the data. For example, Cardoso et al. (2011), Jacobsen et al. (2004), and Potteiger et al. (2002) all conduct studies where separate analyses are conducted for each version of the AUC, and conclude that the results of their study depend on which measure of AUC was used. Because the three versions of AUC are measuring different characteristics of the data, this should not be a shocking result. AUC_G is measuring the *total* concentration of the response over the study period, so it is going to be much larger in magnitude when compared to AUC_I and AUC_{AB} , which are both measuring a *change* in concentration from the baseline value of the response over the time period. Although AUC_I and AUC_{AB} are both measuring a change in concentration from the baseline value, they are each measuring different types of change. AUC_I is measuring the total change

in concentration from the baseline value, either positive or negative, while AUC_{AB} is only capturing the positive change in concentration from the baseline value. AUC_I and AUC_{AB} will be the same if the response doesn't fall below the baseline value, however the two measures become more dissimilar as the response falls below the baseline value more often.

2.2.2.1 Recommendations The study design, along with the goals of the study, should determine which version of the AUC summary measure to use. If the study is interested in determining the long-term effects of a treatment on the concentrations of a substance, and responses are collected over a given time interval across multiple days, then the AUC_G is recommended. The AUC_G then would measure the total concentration on each day and could be used to monitor whether the total concentration has changed over time, whereas AUC_I and AUC_{AB} are both measuring a change in the response from the baseline value on each day and might fail to capture the characteristics of the data that are of interest. For example, consider Figure 2.2 where the AUC_G is decreasing across days, however the values of AUC_I remain nearly constant across days. In situations like this, the AUC_I would miss the overall decreasing trend in the substance across the days and fail to detect the long-term effect of the treatment.

On the other hand, the goal of the study might be to determine whether there is a treatment effect on a particular day. In this scenario, the AUC_I is recommended, as it captures the change in the substance from the baseline value over the time interval. The AUC_G could also be used in this setting, however then it would be recommended that the model include the baseline value as a covariate to control for subject to subject variability in baseline values, resulting in a more complicated ANCOVA-type model (Box, 1950). Figure 2.3 shows an example of two individual

response curves where the AUC_G values are similar for two individuals, however the corresponding AUC_I values are very different. Even controlling for differences across subjects' starting values may not provide an analysis of the AUC_G that detects the treatment effect.

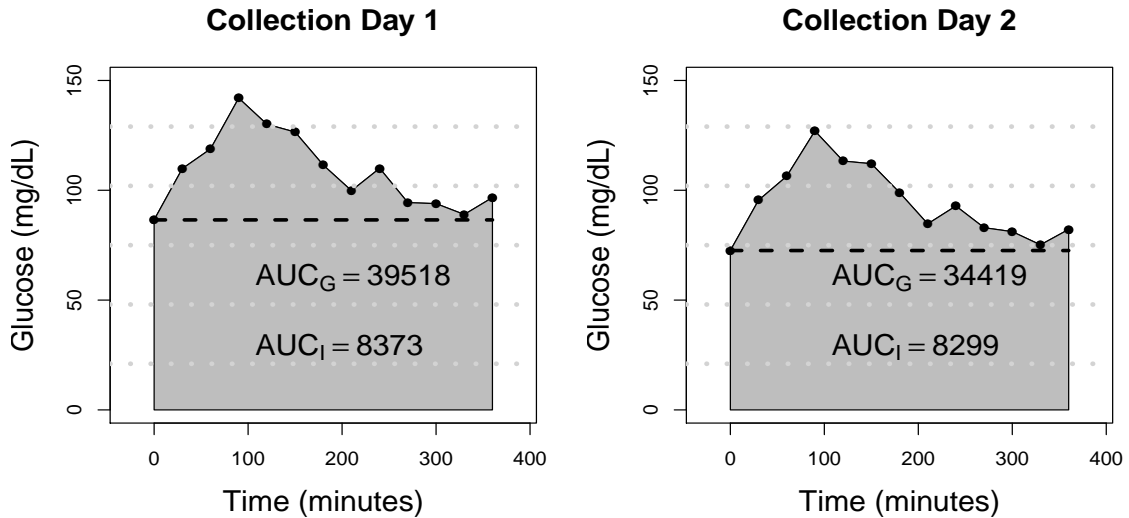


Figure 2.2: The AUC_G and AUC_I computed for the same individual on two collection days. A 360 minute sampling period was used with samples taken every 30 minutes. The bold dashed horizontal line represents the baseline value and the area below is what is subtracted from AUC_G when computing AUC_I .

AUC_{AB} rarely seems to be preferred over AUC_I , as many researchers feel that ignoring any area that falls below the baseline value is throwing away information. The AUC_{AB} calculation is potentially a linear combination of fewer observations so may be less precisely estimated than the other AUC measures and its precision could also vary across observations in the same data set. However, if the goal of the study is to determine whether a treatment increases the concentration of a substance over the time interval, then AUC_{AB} is the preferred AUC since it captures only the positive change in concentration over the time interval.

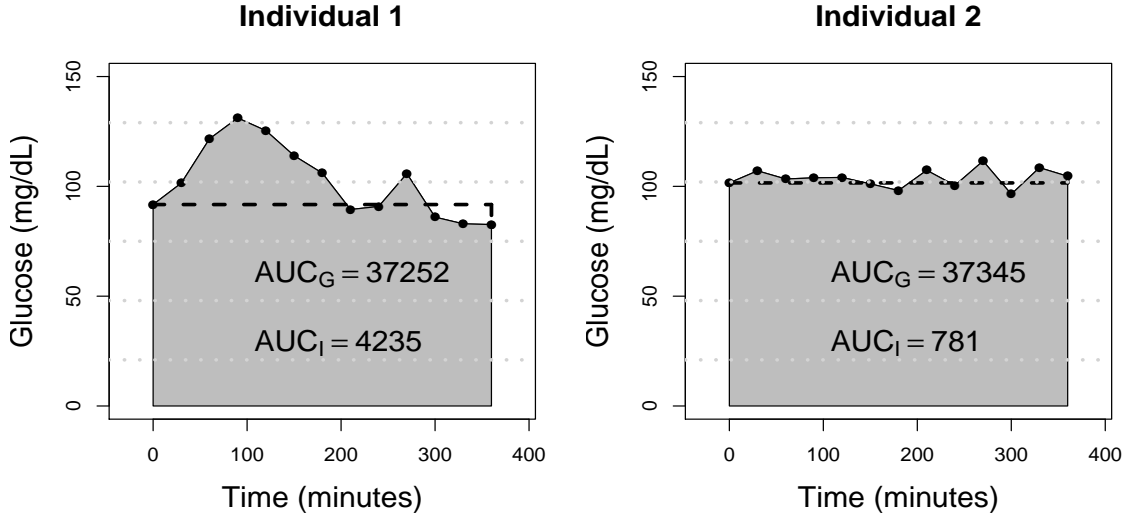


Figure 2.3: The AUC_G and AUC_I computed for two individuals, over the same 360 minute sampling period with samples taken every 30 minutes. The bold dashed horizontal line represents the baseline value and the area below it is what is subtracted from AUC_G when computing AUC_I .

2.3 Other Considerations

2.3.1 Sampling Design

The sampling interval and timing of data collection are important aspects to consider when using the AUC summary measure, as it is based on measurements taken over time. Available resources often dictate the total number of samples that can be obtained, however the process itself may also constrain how many observations or how closely in time measurements can be obtained. It is important to have some knowledge about the expected behavior of the responses in order to sample efficiently. Sampling times need to be chosen, as best as possible, to prevent missed peaks or troughs in the data (Grice and Jackson, 2004) which could severely bias the estimate of AUC. Also, more frequent sampling times result in less bias associated with the

linear trapezoidal method estimator of AUC (Stewart, 2001). There will always be a precision-cost trade-off, however knowledge about the expected behavior of the response can be helpful in choosing sampling times to obtain the most accurate and precise estimator of the AUC.

The width of the sampling interval may vary over the time period, and the expected behavior of the response should drive when sampling occurs. The sampling frequency should be increased during time intervals where the response is expected to be changing rapidly, whereas less sampling effort can be applied during time intervals where the response is expected to change at a constant rate. In the absence of prior knowledge of the timing of extremes or high variability in that timing, equally spaced sampling is recommended as it provides equal chances of sampling near the peak (or trough).

Variability in the data near peaks (or troughs) tends to be higher as compared to other places in the data (Yuan, 1993). Therefore it is recommended that multiple independent measurements (replicates) be obtained at each time point in intervals where peaks (or troughs) are anticipated, if possible. The replicated measurements can be averaged to obtain a more precise estimate of the true response at that time. Also, the baseline value (or initial measurement) plays an extremely important role in both modified versions of the AUC, as it is subtracted from all subsequent measurements. In order to reduce the variability associated with the baseline value, obtaining multiple measurements of the baseline value is recommended when using either the AUC_I or the AUC_{AB} , if possible. In general, whenever the variability in the data is expected to be high, using the mean of multiple measurements at each time point in the AUC calculation will reduce the variability, as compared to using only one measurement at those time points.

2.3.2 Reproducible Research

Recent efforts to increase reproducibility of research (for example, Whitlock et al. (2010)) by publishing data sets can alleviate confusion in methods and models used. Barring that complete openness, certain aspects of the study need to be provided in order for research to be reproducible by other researchers. When using the AUC summary measure, the exact version of the AUC (AUC_G , AUC_I , or AUC_{AB}) that was used in the study should be specified, along with how it was calculated (linear trapezoidal method, spline-based methods, etc.). The unit of time associated with the sampling times used in calculating the AUC should also be stated, as it determines the interpretation of the AUCs. For example, Pruessner et al. (2003) note that if the sampling intervals are constant across the entire sampling period, the time interval can be set to ‘1’ in order to simplify the AUC calculation. The statistical analysis is unaffected by this change, as the simplified version of AUC is linearly related to the AUC that reflects the real time units, however the magnitude of the simplified version of AUC could be much smaller or larger than the AUC computed using the original time scale. Figure 2.4 displays the results of using three different units of time in the calculation of the AUC_G for data collected every 30 minutes over a 6 hour sampling interval, resulting in AUCs that are different in terms of magnitude even though they are all perfectly correlated.

Including the units associated with the AUC is another important detail that will help other researchers fully understand analyses of AUC data. For example, in panel (a) of Figure 2.4 the AUC_G value is 39,607.76 mg/dL per 360 minutes, while in panel (b) the value is 660.13 mg/dL per 6 hours, and the value in panel (c) is 1320.26 mg/dL per 12 equally spaced time intervals. The total sampling period also affects the magnitude of the AUC, since the total concentration of a response is likely to be

larger for longer total sampling intervals. Therefore, when comparing the magnitude of AUC results between studies, one needs to know the total sampling period and units of time used in the calculation as well as the units of the concentration being observed. Researchers rarely fail to report the concentration units but do often fail to provide enough information about time units in AUCs.

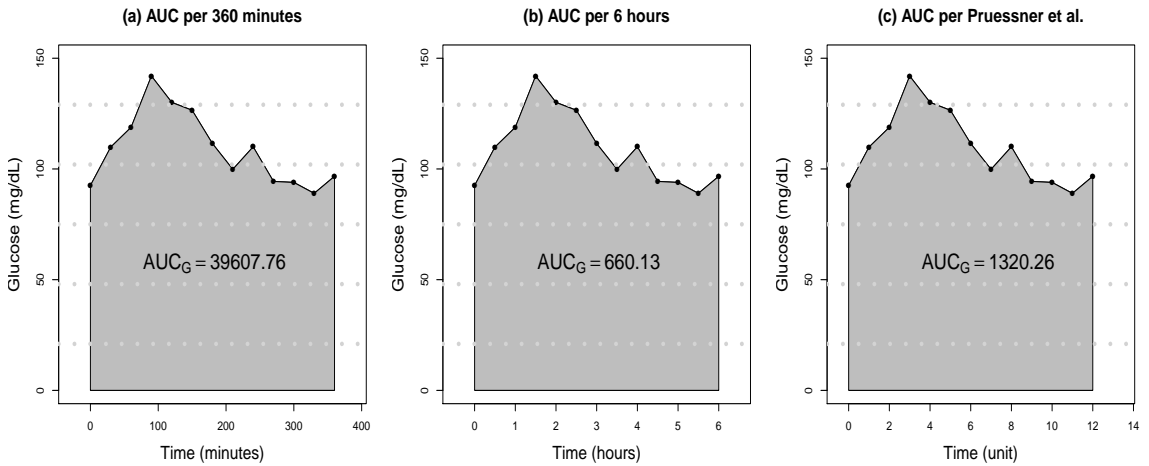


Figure 2.4: Three versions of the AUC_G calculated using the same responses and varying the “Time” scale. Panel (a) represents the AUC resulting from using a 30 minute time interval in the calculation, (b) represents the AUC resulting from using a 0.5 hour time interval in the calculation, and (c) represents the AUC resulting from using Pruessner et al.’s simplified version of the calculation.

2.3.3 AUCs in Analyses

The use of summary measures in analyses dates as far back as 1938 when John Wishart analyzed regression coefficients (Wishart, 1938). However, researchers should be careful when using summary measures as variables in subsequent analyses. In general, the potential explanatory variables or covariates should not vary with time or across replicate measurements (Fitzmaurice et al., 2008). If a potential covariate varies over the sampling interval used in the AUC calculation, choosing

the appropriate value of the covariate to use in the model may be arbitrary or even impossible. If this occurs, the original observations should be used in a more complex model that allows for time-varying covariate values in a model that explains the variations in repeated measurements over time for each subject.

One common assumption of most statistical analyses is that the response variable has constant variance across individuals. If a summary measure is used as a response variable, the constant variance assumption should be checked carefully. Either a transformation or a more complex model, such as an extended regression model (Galecki and Burzykowski, 2013), should be used if this assumption has not been met. Another issue to be aware of when using summary measures in analyses is that since the true value of the variable cannot be observed, but is instead estimated from the data, measurement error is induced. Ignoring measurement errors in analyses can lead to bias in parameter estimates and a loss of power (Carroll et al., 2006).

Although the constant variance assumption can be checked using diagnostic plots, an estimate of the variance must be known in order to account for measurement errors. An estimate of the precision or variability associated with the AUCs is important in order to understand the magnitude of error associated with each AUC. A major limitation of the AUC summary measure is that in most situations there is not a straight-forward method available to calculate an estimate of the variability associated with the AUC summary measure (Hilborn et al., 1999). However there is currently work being done in this area (see Chapter 3).

2.4 Discussion

The AUC summary measure is frequently used as a measure of the total concentration, or total count, of a variable of interest measured over some interval of

time. However, modifications of the original AUC summary measure have generated confusion. Two commonly used modifications of the AUC summary measure were discussed, detailing the characteristics of the data that each are capturing. The goals of the study, along with the study design, dictate which version of the AUC summary measure is most appropriate to use.

The sampling design is extremely important when considering the bias and variance of the AUC estimator. Nedelman et al. (1995) explain that having more sampling times with fewer observations at each sampling time will decrease the bias of the estimate of the AUC but also decrease the precision of the estimate (and vice versa). Therefore, certain trade-offs must be accepted when a fixed number of observations are to be allocated over the entire sampling interval. The expected behavior of the response should be taken into account in order to obtain the best estimate of AUC. Samples should be obtained more frequently in periods where the response is expected to change rapidly; and obtaining replicate measurements at individual time points where the variability is expected to be high is also recommended. In the versions of AUCs that depend heavily on baseline values, additional effort can be usefully placed on better estimating the baseline.

Any scientific research should contain enough details for the research to be reproducible. At a minimum, certain aspects of the study need to be specified when using AUCs, such as which version of AUC was used, the method used to calculate it, and the unit of time used in the calculation. Researchers should also select a single AUC measure for analysis, as Type I error rates will be inflated when multiple AUC measures (computed from the same observed data) are analyzed without adjustment for multiple testing. The consequences of using summary measures in analyses are also important to consider and include the possibility of non-constant variance across individuals and measurement error being present, both of which require more

complicated statistical models or transformations (which complicate interpretations) to be properly accounted for.

References

- Allison, D. B., Paultre, F., Maggio, C., Mezzitis, N., Pi-Sunyer, F. X., 1995. The use of areas under curves in diabetes research. *Diabetes care* 18 (2), 245–250.
- Armitage, P., Berry, G., Matthews, J. N. S., 2008. *Statistical methods in medical research*. John Wiley & Sons.
- Box, G. E., 1950. Problems in the analysis of growth and wear curves. *Biometrics* 6 (4), 362–389.
- Cardoso, F., Sears, W., LeBlanc, S., Drackley, J., 2011. Technical note: Comparison of 3 methods for analyzing areas under the curve for glucose and nonesterified fatty acids concentrations following epinephrine challenge in dairy cows. *Journal of dairy science* 94 (12), 6111–6115.
- Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M., 2006. *Measurement error in nonlinear models: a modern perspective*. CRC press.
- Fekedulegn, D. B., Andrew, M. E., Burchfiel, C. M., Violanti, J. M., Hartley, T. A., Charles, L. E., Miller, D. B., 2007. Area under the curve and other summary indicators of repeated waking cortisol measurements. *Psychosomatic medicine* 69 (7), 651–659.
- Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G., 2008. *Longitudinal data analysis*. CRC Press.
- Gałecki, A., Burzykowski, T., 2013. *Linear mixed-effects models using R: A step-by-step approach*. Springer Science & Business Media.
- Grice, J. E., Jackson, R. V., 2004. Letter to the editor: Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change: A letter to the editor in response to jc preussner et al.(2003) *psychoneuroendocrinology* 28, 916-931. *Psychoneuroendocrinology* 29 (4), 563–564.
- Hilborn, R., Bue, B. G., Sharr, S., 1999. Estimating spawning escapements from periodic counts: a comparison of methods. *Canadian Journal of Fisheries and Aquatic Sciences* 56 (5), 888–896.
- Jacobsen, D., Bailey, B., LeCheminant, J., Hill, J., Mayo, M., Donnelly, J., 2004. A comparison of three methods of analyzing post-exercise oxygen consumption. *International journal of sports medicine* 26 (1), 34–38.
- Kashyap, S. R., Bhatt, D. L., Wolski, K., Watanabe, R. M., Abdul-Ghani, M., Abood, B., Pothier, C. E., Brethauer, S., Nissen, S., Gupta, M., et al., 2013. *Metabolic*

- effects of bariatric surgery in patients with moderate obesity and type 2 diabetes analysis of a randomized control trial comparing surgery with intensive medical treatment. *Diabetes care* 36 (8), 2175–2182.
- Le Floch, J.-P., Escuyer, P., Baudin, E., Baudon, D., Perlemuter, L., 1990. Blood glucose area under the curve: methodological aspects. *Diabetes care* 13 (2), 172–175.
- Moreaux, S. J. J., Nichols, J. L., Bowman, J. G., Hatfield, P. G., 2011. Psyllium lowers blood glucose and insulin concentrations in horses. *Journal of Equine Veterinary Science* 31 (4), 160–165.
- Nedelman, J. R., Gibiansky, E., Lau, D. T., 1995. Applying bailer’s method for auc confidence intervals to sparse sampling. *Pharmaceutical research* 12 (1), 124–128.
- Parken, C. K., Bailey, R. E., Irvine, J. R., 2003. Incorporating uncertainty into area-under-the-curve and peak count salmon escapement estimation. *North American Journal of Fisheries Management* 23 (1), 78–90.
- Potteiger, J., Jacobsen, D., Donnelly, J., 2002. A comparison of methods for analyzing glucose and insulin areas under the curve following nine months of exercise in overweight adults. *International Journal of Obesity & Related Metabolic Disorders* 26 (1).
- Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., Hellhammer, D. H., 2003. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology* 28 (7), 916–931.
- Purves, R. D., 1992. Optimum numerical integration methods for estimation of area-under-the-curve (auc) and area-under-the-moment-curve (aumc). *Journal of pharmacokinetics and biopharmaceutics* 20 (3), 211–226.
- Stewart, J., 2001. *Calculus concepts and contexts*, 2nd edn (pacific grove, ca: Brooks/cole).
- Whitlock, M. C., McPeck, M. A., Rausher, M. D., Rieseberg, L., Moore, A. J., 2010. Data archiving. *The American Naturalist* 175 (2), 145–146.
- Wishart, J., 1938. Growth-rate determinations in nutrition studies with the bacon pig, and their analysis. *Biometrika* 30 (1/2), 16–28.
- Wolever, T. M., Jenkins, D., 1986. The use of the glycemic index in predicting the blood glucose response to mixed meals. *The American journal of clinical nutrition* 43 (1), 167–172.

Yeh, K., Kwan, K., 1978. A comparison of numerical integrating algorithms by trapezoidal, lagrange, and spline approximation. *Journal of pharmacokinetics and biopharmaceutics* 6 (1), 79–98.

Yuan, J., 1993. Estimation of variance for auc in animal studies. *Journal of pharmaceutical sciences* 82 (7), 761–763.

CHAPTER THREE

BAYESIAN ESTIMATION OF VARIANCE FOR AUC STUDIES

Contribution of Authors and Co-Authors

Author: Jennifer L. Weeding

Contributions: Responsible for the majority of the writing.

Co-Author: Dr. Mark C. Greenwood

Contributions: Provided feedback on statistical analysis and drafts of the manuscript.

Co-Author: Dr. Shannon J.J. Moreaux

Contributions: Provided application expertise.

Co-Author: Jyme L. Nichols

Contributions: Data Collection.

Manuscript Information Page

Jennifer L. Weeding, Mark C. Greenwood, Shannon J.J. Moreaux, and Jyme L. Nichols

PLoS ONE

Status of Manuscript:

Prepared for submission to a peer-reviewed journal

Officially submitted to a peer-review journal

Accepted by a peer-reviewed journal

Published in a peer-reviewed journal

Abstract

The area under the curve (AUC) summary measure reduces repeated measures data on an individual into a single number through a function of the observed data, and is often estimated using the linear trapezoidal method. It provides an estimate of total concentration or amount over a period of time and is popular in pharmacokinetic applications. The uncertainty associated with the estimated AUC generally goes unaccounted for in subsequent analyses, as the estimation of the variance associated with it is not straightforward in most scenarios. We present a Bayesian approach to modeling AUC data, using two different statistical models, where the variance-covariance matrix of the repeated observations can be estimated using values computed from the resulting posterior distributions. The estimated variance-covariance matrix is then used to compute an estimate of the variance of the \widehat{AUC} . The performances of the two methods are studied using simulated data. An application to an equine-based glucose data set displays some difficulties that can arise when working with real data.

3.1 Introduction

The total concentration of a substance present in the blood over time is often of interest to researchers studying pharmacodynamic and pharmacokinetic properties in either humans or animals; the same techniques have been used in Ecology where the total count of a species present over a time interval is of interest. The area under the curve (AUC) is a summary measure that captures the total magnitude of a response over a specified time interval and has a long history in pharmacology (Armitage et al., 2008). However, recently it is becoming popular in other areas such as fisheries and animal science based research (for example, Parken et al. (2003) or Moreaux et al. (2011)).

The AUC is defined as the definite integral of the true underlying function of the response, $f(t)$, over the time interval (a, b) . Usually, the true underlying function of the response is unknown and the AUC must be estimated from the observed data. The

linear trapezoidal method is one of many numerical integration techniques that can be used to estimate the AUC, and is often preferred over more complex estimation methods due to its simplicity (Yuan, 1993), along with the fact that it has good performance in situations involving a small number of data points (Allison et al., 1995). The formula for estimating the AUC using the linear trapezoidal method is:

$$\begin{aligned}\widehat{AUC} &= \frac{(y_1 + y_2) * (t_2 - t_1)}{2} + \frac{(y_2 + y_3) * (t_3 - t_2)}{2} + \dots + \frac{(y_{n-1} + y_n) * (t_n - t_{n-1})}{2} \\ &= \sum_{i=1}^{n-1} \frac{(y_i + y_{i+1}) * (t_{i+1} - t_i)}{2},\end{aligned}\tag{3.1}$$

where (y_i, t_i) represent the responses and the corresponding time points for an individual, and $i = 1, 2, \dots, n$ index the responses collected. The simplicity of this method can be seen when examining Equation 3.1, as this estimate of the AUC is found by summing the areas of all trapezoids formed using consecutive time points and their responses.

A major limitation of the linear trapezoidal method of calculating the AUC is that in most scenarios there is not a straight-forward method available to estimate the variance associated with it (Hilborn et al., 1999). Equation 3.1 can be rewritten as:

$$\widehat{AUC} = \frac{t_2 - t_1}{2} * y_1 + \frac{t_3 - t_1}{2} * y_2 + \frac{t_4 - t_2}{2} * y_3 + \dots + \frac{t_n - t_{n-2}}{2} * y_{n-1} + \frac{t_n - t_{n-1}}{2} * y_n,\tag{3.2}$$

which shows more clearly that the \widehat{AUC} is a linear combination of the responses (y_i 's). If we denote the multiplier in front of each y_i as c_i , the formula for the variance of

the estimated AUC is:

$$Var(\widehat{AUC}) = Var\left(\sum_{i=1}^n c_i y_i\right) = \sum_{i=1}^n c_i^2 Var(y_i) + 2 \sum_{1 \leq i < i' \leq n} c_i c_{i'} Cov(y_i, y_{i'}). \quad (3.3)$$

In most situations, the $Var(y_i)$ and $Cov(y_i, y_{i'})$ are unknown making estimation of $Var(\widehat{AUC})$ using Equation 3.3 challenging. Frequently, multiple observations are obtained at each time point, resulting in observing y_{ij} , where $j = 1, 2, \dots, m_i$ for each i . These repeated observations may come from new independent observations (replicates) of the process (for example, a new blood draw) or, more commonly, from measuring another portion of the initial blood draw (a pseudo-replicate which contains less independent information than from replicated measurements). In either scenario, the y_i 's in Equations 3.1, 3.2, and 3.3 can be replaced with \bar{y}_i 's, where \bar{y}_i represents the mean of the m_i observations at time i . It is important to note that it would not be typical to observe a new trajectory, say y_{i2} , that provides an independent measurement of the $i = 1, 2, \dots, n$ observations over time for the same subject, but that the replicates are occurring as repeated observations for each time point. If replicate time trajectories were observed for each subject, the data would have enough information to use standard mixed models (Galecki and Burzykowski, 2013) with autocorrelated error structures to provide the information required to estimate the variance in the \widehat{AUC} s. These methods do not work with replicate or pseudo-replicate measurements for each time point as are commonly encountered in AUC data obtained from blood samples.

The primary focus of this paper is to present a Bayesian approach to modeling AUC data where measurements are taken over time, potentially with multiple independent replicates measured at each time point. The variance-covariance matrix of the true underlying means across time points can be estimated using values

computed from the appropriate posterior distributions and then used to estimate the variance of the \widehat{AUC} . Different statistical models are considered for the time-trajectories, possibly with replicate measurements, then the linear trapezoidal method is used to compute an estimate of the AUC using posterior means obtained at each time point. Two Bayesian models are presented in Section 3.2. The performance of the methods are studied using simulated data in Section 3.3, along with an application to real data. Extensions of this method to other AUC measures, along with recommendations and concluding remarks, are given in Section 3.4.

3.2 Methods

To model the time-trajectory of the observations used in the AUC calculation, analysis of variance (ANOVA) and first-order autoregressive (AR(1)) models are considered here. In order to be able to incorporate replicate or pseudo-replicate information, the statistical models are developed using a Bayesian framework, resulting in posterior distributions for the true underlying means at each time point from which an estimate of the mean and variance can be computed. The AUC is estimated using the linear trapezoidal method, where the estimated means at each time point are used in Equation 3.2. The estimated variance-covariance matrix of the means over time can be obtained using the model-based estimated variances and covariances (when applicable), and an estimate of the $Var(\widehat{AUC})$ is then computed using Equation 3.3. Two formula-based estimators are also computed for comparison, one where the repeated observations are assumed to be independent across time and one where the repeated observations are assumed to be perfectly correlated, representing the extremes of potential impacts of covariances on the estimation of $Var(\widehat{AUC})$.

3.2.1 Formula-Based Estimators

Equation 3.3 can be used to compute (an estimate of) the variance of the \widehat{AUC} when $Var(y_i)$ and $Cov(y_i, y_{i'})$ are known, assumed known, or can be estimated from the data. If only one observation is available at each time point ($m_i = 1$), $Var(y_i)$ and $Cov(y_i, y_{i'})$ must be known, or assumed known, in order to compute an estimate of $Var(\widehat{AUC})$. If $m_i > 1$, the $Var(\bar{y}_i)$ can be estimated from the data, while $Cov(\bar{y}_i, \bar{y}_{i'})$ is often more difficult to estimate using standard techniques that also estimate $Var(\bar{y}_i)$. In this scenario, assumptions must be made regarding the correlation structure of the repeated observations across time.

For comparison to the Bayesian methods presented in the next sections, we assume the measurements over time are either independent or perfectly (positively) correlated. If they are positively auto-correlated, the two estimators presented will serve as a lower bound (independence) and an upper bound (auto-correlation of 1) for the estimated $Var(\widehat{AUC})$. If the observations at different sampling times are assumed to be independent, Equation 3.3 becomes

$$Var(\widehat{AUC}) = \sum_{i=1}^n c_i^2 Var(y_i), \quad (3.4)$$

since $Cov(y_i, y_{i'}) = 0$ for independent observations. We denote this estimator as FB0, to represent the formula-based estimator assuming a correlation of 0. If the repeated observations are assumed to be perfectly (positively) correlated, Equation 3.3 becomes

$$Var(\widehat{AUC}) = \sum_{i=1}^n c_i^2 Var(y_i) + 2 \sum_{1 \leq i < i' \leq n} c_i c_{i'} * \sigma_{y_i} * \sigma_{y_{i'}}, \quad (3.5)$$

where $\rho = 1$, since

$$Cov(y_i, y_{i'}) = \rho * \sigma_{y_i} * \sigma_{y_{i'}} \quad (3.6)$$

for correlated observations. We denote this estimator as FB1, to represent the formula-based estimator assuming a correlation of 1. Note that if the estimated autocorrelation is negative, the variance estimator can be lower than the result FB0 provides (Cryer and Chan, 2008).

3.2.2 Bayesian ANOVA Model

Observations on the same subject over time are commonly assumed to be correlated, however this correlation may be negligible in some cases. A potential model at the replicate (y_{ij}) level is an ANOVA type model where the observations at each time point are considered a group and the group means are computed for the n time points. The *means model* specification of this model is:

$$y_{ij} = \alpha_i + \epsilon_{ij}, \quad (3.7)$$

where $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m_i$, $m_i \geq 1$, y_{ij} represents the j^{th} observation taken at time i , and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. The ANOVA model is suggested for cases where observations across time have negligible correlation present, as this model assumes the observations are independent over time.

A hierarchical Bayesian ANOVA model is used here, where the likelihood is specified by:

$$\begin{aligned} y_{ij} &\sim N(\alpha_i, \sigma_\epsilon^2), \\ \alpha_i &\sim N(\mu, \sigma_\alpha^2), \end{aligned} \quad (3.8)$$

and non-informative prior distributions are used for the parameters μ , σ_ϵ , and σ_α for the given application ($\mu \sim N(0, 10^5)$, $\sigma_\epsilon \sim Uniform(0, 50)$, and $\sigma_\alpha \sim Uniform(0, 50)$). Non-constant variance across time points can be accounted for

by replacing σ_ϵ^2 in Equation 3.8 with $\sigma_{\epsilon,i}^2$ when $m_i > 1$. If $m_i = 1$, an empirical or an informative prior for σ_ϵ should be used, as the model has difficulties distinguishing σ_ϵ and σ_α . Although the hierarchical Bayesian ANOVA model specification used here is similar to the specification used by Jackman (2009, Chapter 7), we allow for the possibility of non-constant variance if needed.

This model assumes the observations taken across time are independent, therefore an estimate of the $Var(\widehat{AUC})$ can be computed using Equation 3.4, where the variance associated with the mean at each time point is computed from the posterior distribution of each α_i . This model is used to provide the information required to calculate what we call the ANOVA variance estimate.

3.2.3 Bayesian AR(1) Model

A strong pattern, even a curve, is often evident in the time profiles used to calculate AUCs. In this situation, the observations across time points are expected to be positively correlated and the statistical model used should account for this correlation. The AR(1) model is a statistical model that accounts for correlated observations by assuming the current value is a linear combination of the most recent value plus a random error term, where the sampling interval between consecutive observations is assumed equal. The correlation structure implied by this model is that observations obtained closer in time are more correlated than observations that are more distant in time. The AR(1) model is specified by:

$$y_i = c + \phi y_{i-1} + \epsilon_i, \tag{3.9}$$

where $\phi < |1|$, c is the intercept, $\epsilon_i \sim N(0, \sigma_\epsilon^2)$, and $i = 1, 2, \dots, n$. AR(1) models are typically used when only one observation is collected across multiple time points, but

they can be extended for cases where multiple observations are obtained at each time point. In this scenario, Equation 3.9 extends to:

$$y_{ij} = c + \phi\mu_{i-1} + \epsilon_{ij}, \quad (3.10)$$

where $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m_i$, $m_i \geq 1$, y_{ij} represents the j^{th} observation taken at time i , μ_i represent the true underlying mean at time i , $\phi < |1|$, c is the intercept, and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$.

AR(1) models often lose the initial observation due to the structure of the model, since there is no observation obtained at time 0. We specify a Bayesian AR(1) model, similar to the one in Lunn et al. (2012, Chapter 11), however our specification retains the initial observation by placing a prior distribution on an assumed $t = 0$ measurement, μ_0 . The AR(1) model given in Equation 3.10 can be specified in a Bayesian framework using a likelihood defined by

$$\begin{aligned} y_{ij} &\sim N(\mu_i, \sigma_\epsilon^2), \\ \mu_i &\sim N(c + \phi\mu_{i-1}, \sigma_\mu^2), \end{aligned} \quad (3.11)$$

where non-informative priors are used for c , ϕ , σ_ϵ , and μ_0 ($c \sim N(0, 10^5)$, $\phi \sim \text{Uniform}(-1, 1)$, $\sigma_\epsilon \sim \text{Uniform}(0, 50)$, $\sigma_\mu \sim \text{Uniform}(0, 50)$, and $\mu_0 \sim N(\mu, 10^4)$ where μ is the overall mean calculated from the data). Non-constant variance across time points can be accounted for by replacing σ_ϵ^2 in Equation 3.11 with $\sigma_{\epsilon,i}^2$ when $m_i > 1$. If $m_i = 1$, an empirical or informative prior for σ_ϵ should be used, as the model has difficulties distinguishing σ_ϵ and σ_μ . The prior we chose for ϕ aligns with the constraint that $\phi < |1|$; if an explosive AR(1) model ($\phi > 1$) is desired, a non-informative Normal prior can be applied (for example, $\phi \sim N(0, 100)$).

The AR(1) model assumes that the sampling intervals are equal and that observations obtained one time point apart have a correlation of ϕ , two time steps apart have a correlation of ϕ^2 , and k apart have a correlation of ϕ^k . The resulting covariance of observations at time i and i' is

$$\text{Cov}(y_i, y_{i'}) = \phi^k * \sigma_{y_i} * \sigma_{y_{i'}}, \quad (3.12)$$

where y_i and $y_{i'}$ are obtained k intervals apart. This result is equivalent to Equation 3.6 if you let $\rho = \phi^k$. The AR(1) model can be used to estimate $\text{Cov}(y_i, y_{i'})$ when repeated observations are assumed to have this correlation structure, where an estimate of ϕ is obtained by computing the mean of the posterior distribution of ϕ and the standard deviation associated with the mean at each time point is computed from the posterior distribution of each μ_i . An estimate of $\text{Var}(\widehat{AUC})$ can then be computed using Equation 3.3 based on these results.

3.3 Applications

Figure 3.1 displays examples of two patterns that often arise in equine-based studies where glucose data are collected and AUCs calculated. The glucose response is often measured for an extended period of time after a treatment is applied and the resulting AUCs may have a single peak where the glucose response strictly increases until the peak value is reached, then decreases over the rest of the sampling interval (Pattern 1). The pattern may also be less structured, with the glucose response having multiple increasing and decreasing intervals (Pattern 2). The performance of the two statistical models discussed in Section 3.2 were evaluated using data generated from both scenarios, where a 6 hour sampling period was used with 30 minute sampling intervals. Three different m_i values were considered ($m_i = 1, 2, 4$) to see the effects

of increasing the number of replicate values obtained at each time point. The m_i replicate data were generated such that the mean structure of the original data was retained and three variance structures were considered (small constant variance, large constant variance, non-constant variance that increased near the peak) for the replicate values. Constant variance was assumed when applying the Bayesian models in order to assess model performance when the error structure was mis-specified. The performance of each model when using the mean of the replicate values was also compared to using the replicate values themselves in the analyses. Finally, the models were applied to an equine-based glucose data set, to see how they perform in situations involving real data.

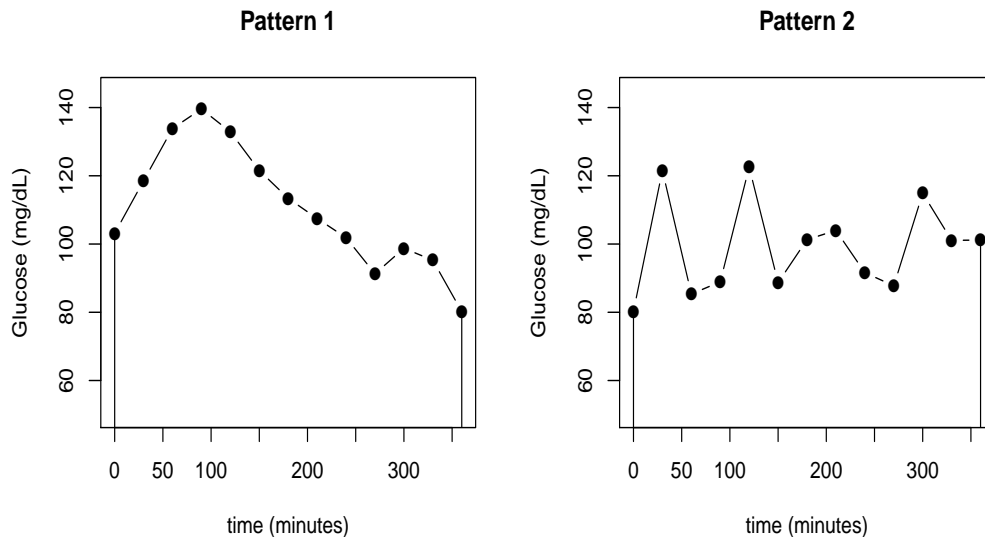


Figure 3.1: Two generated curves with characteristics commonly seen in equine-based glucose AUC data.

The study was carried out in R (R Core Team, 2016) and used the R-package R2jags (Su and Yajima, 2016). The models were fit using a Bayesian framework,

where JAGS (Plummer, 2003) was used to implement the Gibbs sampling algorithm (Geman and Geman, 1984). Three independent chains were used and the first 1000 samples were discarded as *burn-in*. Each chain was run for 50,000 iterations and a thinning rate of 5 was applied, resulting in a final chain length of 10,000 iterations. Convergence was monitored using both sample path plots and potential scale reduction factors (Gelman and Rubin, 1992), where a threshold of $\hat{R} < 0.05$ was used to determine convergence. In scenarios where $m_i = 1$, σ_ϵ in Equations 3.8 and 3.11 is fixed at the data generating values to reflect the situation where these parameters must be known to incorporate their variability into the model.

3.3.1 Simulated Data Study

3.3.1.1 $m_i = 1$ versus $m_i > 1$ AUC data often consist of replicate observations ($m_i > 1$) at individual time points from which the mean at each time point is estimated. The resulting means are used in Equation 3.1 and, in the conventional analysis, the replicate data are generally ignored. Replicate data were simulated at each time point ($m_i = 1, 2, 4$) to assess the effects of using the mean values versus using all of the measurements in each of the three models proposed, where the variance of the replicate values at each time point is 100 ($\sigma_\epsilon^2 = 100$). Figure 3.2 displays the fitted curves for each model using only the mean values. The models both track the observed data pattern better for Pattern 1 as compared to Pattern 2 when $m_i = 1$ and follow the observed data pattern better when the means are based on more replicate values for both patterns. In general, increasing m_i results in smaller estimated variance in the \widehat{AUC} s, as can be seen in Table 3.1. We feel that if the models do not track the observed data well, any estimated variances are not trustworthy and therefore the

results in Table 3.1 corresponding to panel (d) in Figure 3.2 are not to be trusted, as the models did not fit the observed data well in this scenario.

The fitted curves for each model where the replicate data were used are displayed in Figure 3.3, where panels (b), (d), (f), and (h) correspond to the curves in Figure 3.2. The models using the replicate-level data perform similarly to those using the means, however they do not require σ_ϵ to be known. Tables 3.1 and 3.2 show the estimated standard deviations of the \widehat{AUC} are slightly smaller for each model when the mean values are used (only rows in Table 3.2 where $\sigma_\epsilon^2 = 100$ can be compared), however the models that use the replicate data estimate σ_ϵ , resulting in more variability than the models that use only the means (where the data generating value is plugged in for σ_ϵ) since these estimates reflect the uncertainty associated with σ_ϵ that is missed when the values are assumed known.

Increasing the number of replicate values obtained at each time point results in overall better performance of the models and less variability associated with the \widehat{AUC} , as can be seen in Figure 3.3 and Table 3.2. When the observed data pattern is less structured and large variance is present, the models require more replicate values at each time point for the results to track the observed pattern, as can be seen in panels (f) and (h) of Figure 3.3.

3.3.1.2 Different Variance Structures In order to assess how well the models perform when non-constant variance is present, replicate data ($m_i = 2, 4$) for the two patterns were generated with non-constant variance using the values in Table 3.3 based on the results from the original data set. The models were fit using two assumptions, one where the variance structure was mis-specified (the models assumed constant variance) and one where the non-constant variance structure was accounted for. An empirical Bayes approach was used when $m_i = 2$, where the observed data

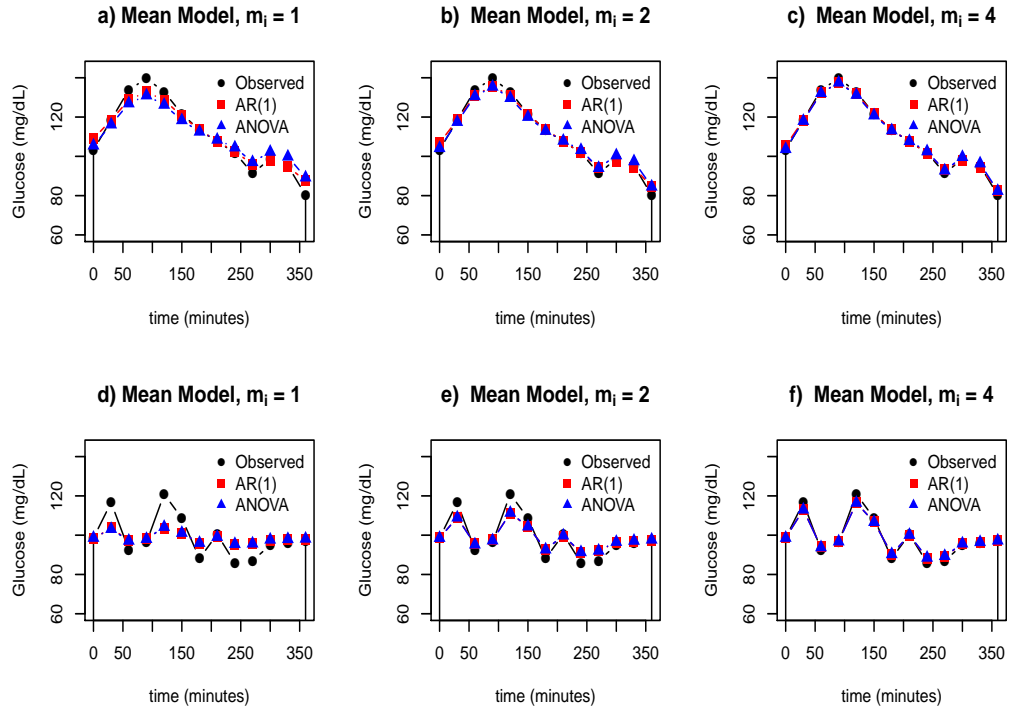


Figure 3.2: Estimated curves from each model where the mean values are used (instead of the replicate values) are displayed for both patterns. The observed curve, along with the fitted curve for each method are displayed, where posterior means at each time point are plotted. Results for Pattern 1 are displayed in panels (a)-(c), and results for Pattern 2 are displayed in panels (d)-(f).

were used to estimate the priors for each $\sigma_{\epsilon,i}$ in the non-constant variance model. Prior distributions were placed on each $\sigma_{\epsilon,i}$ when $m_i = 4$ ($\sigma_{\epsilon,i} \sim Uniform(0, 20)$).

The fitted curves for the models that assume constant variance are shown in panels (a), (c), (e), and (g) of Figure 3.4. The models both have good performance when the variance structure is mis-specified, and have even better performance when more replicate values are available at each time point. The fitted curves for the models that account for non-constant variance are shown in panels (b), (d), (f), and (h) of Figure 3.4. The models all track the observed data pattern well when the non-constant

Table 3.1: Standard deviations of the \widehat{AUC} s for each method computed using the mean value at each time point, corresponding to the estimated curves in Figure 3.2. $\hat{\phi}$ is estimated from the AR(1) model and is an estimate of the lag 1 autocorrelation. FB0 and FB1 correspond to the formula-based estimators assuming a correlation between repeated observations of 0 or 1, respectively. * indicates that the models tracked the observed data pattern, but didn't fit the observed means exactly and ** indicates that the models fit the observed data poorly.

| Pattern | m_i | FB0 | ANOVA | AR(1) ($\hat{\phi}$) | FB1 |
|---------|-------|---------|----------|------------------------|---------|
| 1 | 1 | 1017.35 | 892.21* | 1616.89* (0.67) | 3600.00 |
| | 2 | 719.37 | 678.22 | 1341.86 (0.72) | 2545.58 |
| | 4 | 508.67 | 493.12 | 1049.60 (0.75) | 1800.00 |
| 2 | 1 | 1017.35 | 610.34** | 581.48** (-0.07) | 3600.00 |
| | 2 | 719.37 | 597.50* | 621.05* (0.03) | 2545.58 |
| | 4 | 508.67 | 468.81 | 500.31 (0.06) | 1800.00 |

variance structure is taken into account. Table 3.4 contains the estimated standard deviations of the \widehat{AUC} s for each method. The uncertainty in the $\sigma_{\epsilon,i}$'s is incorporated in the non-constant variance models that use prior distributions (panels (d) and (h) of Figure 3.4), and for comparison, we have included results in Table 3.4 for $m_i = 4$ where an empirical Bayes approach is used (where the variance parameters are set at the data-based values). The resulting \widehat{AUC} s from the models that estimate $\sigma_{\epsilon,i}$ are more variable than those from the models that use an empirical Bayes approach, however the estimates reflect the uncertainty associated with each $\sigma_{\epsilon,i}$ that is missed when the values are assumed known.

3.3.1.3 Effects of Different Patterns Equation 3.3 shows that repeated observations which are positively correlated across time should have a larger variance associated with the AUC than when repeated observations are uncorrelated, assuming the variances at the individual time points are held constant (since $Cov(y_i, y_{i'}) > 0$ for positively correlated observations). The patterns displayed in Figure 3.1 are

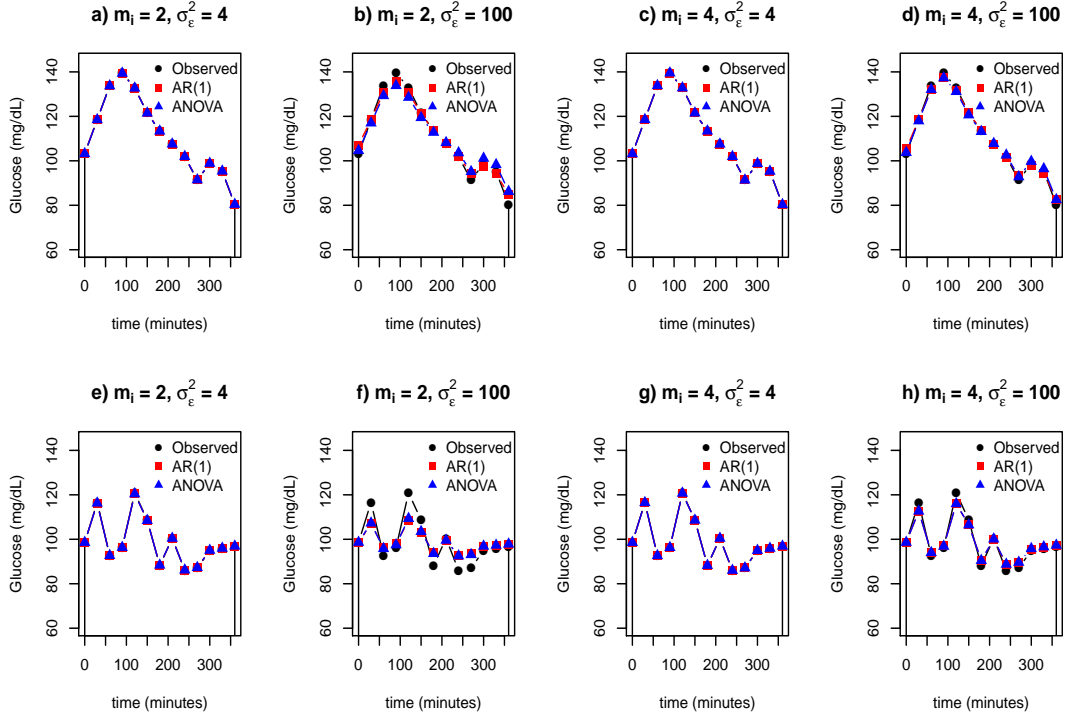


Figure 3.3: Estimated curves from each model where different values of replication are considered ($m_i = 2, 4$) and different variances associated with the replicate values ($\sigma_\epsilon^2 = 4, 100$). The observed curve, along with the fitted curve for each method are displayed, where posterior means at each time point are plotted. Results for Pattern 1 are displayed in panels (a)-(d), and results for Pattern 2 are displayed in panels (e)-(h).

expected to have different underlying correlation structures, where it appears that the repeated observations in Pattern 1 are more correlated than the repeated observations in Pattern 2. Estimated ϕ values obtained from the AR(1) model are included in Tables 3.1, 3.2, and 3.4, and indicate that overall Pattern 1 has a strong positive autocorrelation present (≈ 0.75) and Pattern 2 has negligible autocorrelation present (≈ 0.05). The data generated for the two patterns across the different scenarios shown in Figures 3.2 and 3.3 contain exactly the same random errors, therefore the resulting standard deviations for the two patterns can be compared (Tables 3.1 and 3.2).

Table 3.2: Estimated standard deviations of the \widehat{AUC} 's for each method for different magnitudes of variance and values of replication (m_i), corresponding to the estimated curves in Figure 3.3. $\hat{\phi}$ is estimated from the AR(1) model and is an estimate of the lag 1 autocorrelation. FB0 and FB1 correspond to the formula-based estimators assuming a correlation between repeated observations of 0 or 1, respectively. * indicates that the models tracked the observed data pattern, but didn't fit the observed means exactly.

| Pattern | m_i | σ_ϵ^2 | FB0 | ANOVA | AR(1) ($\hat{\phi}$) | FB1 |
|---------|-------|---------------------|--------|---------|------------------------|---------|
| 1 | 2 | 4 | 143.87 | 164.51 | 380.47 (0.78) | 509.12 |
| 1 | 2 | 100 | 719.37 | 767.91 | 1366.21 (0.72) | 2545.58 |
| 1 | 4 | 4 | 101.73 | 105.53 | 251.92 (0.78) | 360.00 |
| 1 | 4 | 100 | 508.67 | 514.77 | 1066.21 (0.75) | 1800.00 |
| 2 | 2 | 4 | 143.87 | 162.81 | 174.97 (0.07) | 509.12 |
| 2 | 2 | 100 | 719.37 | 626.52* | 626.30* (-0.01) | 2545.58 |
| 2 | 4 | 4 | 101.73 | 105.50 | 112.97 (0.07) | 360.00 |
| 2 | 4 | 100 | 508.67 | 488.02 | 522.01 (0.06) | 1800.00 |

The results for the estimated standard deviation of the \widehat{AUC} obtained from the ANOVA model are very similar for the two patterns across the different scenarios studied, aligning with its assumption of independent observations across time. The results from the formula-based estimator that assumes independent observations (FB0) should be similar to those obtained from the ANOVA model, since both models assume the true underlying means are independent. The formula-based estimator (FB0) produces estimated standard deviations that are smaller than those produced by the ANOVA model in most scenarios, because the ANOVA model incorporates the extra uncertainty in the σ_ϵ 's that the formula-based estimator misses.

The AR(1) model assumes that observations across time are positively correlated based on distances apart, therefore results for the standard deviation of the \widehat{AUC} obtained from Pattern 1 are expected to be larger than those obtained from Pattern 2 since autocorrelation is strong in Pattern 1. For all scenarios studied, the estimated

Table 3.3: Variances used in the non-constant variance simulation for each pattern.

| i | Pattern 1 $\sigma_{\epsilon,i}^2$ | Pattern 2 $\sigma_{\epsilon,i}^2$ |
|-----|--------------------------------------|--------------------------------------|
| 1 | 18.75 | 40.83 |
| 2 | 25.52 | 4.28 |
| 3 | 37.63 | 8.37 |
| 4 | 47.22 | 365.75 |
| 5 | 99.11 | 67.49 |
| 6 | 98.08 | 63.14 |
| 7 | 107.41 | 4.50 |
| 8 | 163.58 | 192.82 |
| 9 | 3.15 | 2.08 |
| 10 | 50.78 | 60.55 |
| 11 | 17.12 | 44.81 |
| 12 | 57.76 | 0.83 |
| 13 | 62.59 | 1.40 |

standard deviations produced from the AR(1) model are larger for Pattern 1 than for Pattern 2, as can be seen in Tables 3.1 and 3.2.

The two formula-based estimators provide a lower and upper bound for where the Bayesian model estimated standard deviations are expected to fall for scenarios where the Bayesian models fit the observed data well. The estimated standard deviations produced from the ANOVA model were similar to FB0, as is expected. The estimated standard deviations produced from the AR(1) model fall between the two formula-based estimators in all scenarios for Pattern 1, and for Pattern 2 were similar to the formula-based estimator that assumes independent observations.

3.3.2 Equine Data

An equine-based data set was used to explore the performance of the three models when applied to actual data and also to point out important details to consider when applying these methods. All procedures were approved by Montana

Table 3.4: Estimated standard deviations of the \widehat{AUC} s for each method, corresponding to the estimated curves in Figure 3.4. The replicate data at each time point have non-constant variance, and the models were fit 2 ways: assuming constant variance (denoted C below) and assuming non-constant variance (denoted NC below). E-NC indicates an empirical non-constant variance model was used and P-NC indicates prior distributions were used in a non-constant variance model. $\hat{\phi}$ is estimated from the AR(1) model and is an estimate of the lag 1 autocorrelation. FB0 and FB1 correspond to the formula-based estimators assuming a correlation between repeated observations of 0 or 1, respectively.

| Pattern | Model | m_i | FB0 | ANOVA | AR(1) ($\hat{\phi}$) | FB1 |
|---------|-------|-------|--------|--------|------------------------|---------|
| 1 | C | 2 | 519.89 | 614.98 | 1168.63 (0.74) | 1839.68 |
| 1 | E-NC | 2 | 572.25 | 542.03 | 1047.17 (0.73) | 1863.15 |
| 1 | C | 4 | 367.62 | 404.96 | 879.07 (0.76) | 1300.85 |
| 1 | E-NC | 4 | 404.64 | 393.28 | 809.86 (0.75) | 1317.44 |
| 1 | P-NC | 4 | 404.64 | 530.34 | 1030.98 (0.74) | 1317.44 |
| 2 | C | 2 | 446.35 | 582.92 | 611.58 (0.03) | 1579.46 |
| 2 | E-NC | 2 | 609.37 | 478.96 | 497.71 (0.03) | 1630.76 |
| 2 | C | 4 | 315.62 | 407.73 | 435.50 (0.06) | 1116.84 |
| 2 | E-NC | 4 | 430.89 | 372.82 | 392.64 (0.06) | 1153.12 |
| 2 | P-NC | 4 | 430.89 | 448.25 | 469.56 (0.03) | 1153.12 |

State University’s Animal Care and Use Committee. Eight 11- to 16 yr-old Quarter Horses were used in an exploratory study to examine glucose responses over a 6-hour sampling period, with blood samples obtained every 30 minutes ($n = 13$). Glucose analysis was determined in duplicate ($m_i = 2$) from the same blood sample (pseudo-replicates) using a spectrophotometric method based on glucose hexokinase (Glucose hexokinase kit: Sigma Diagnostics, St. Louis, MO). The data used in this application were initial measurements (Day 0) that were part of a larger study, where the effect of a treatment over time was of interest. More details concerning the data collection process can be found in Peterson et al. (2009).

The $m_i = 2$ pseudo-replicate values at each time point were used in the analysis, as the previous results suggest better performance using the raw values as compared

to the mean values at each time point. A non-constant variance model was used where the $\sigma_{\epsilon,i}$'s are estimated from the observed data (an empirical Bayes approach). The observed data and resulting estimated curves from each method can be seen in Figure 3.5, where posterior means at each time point are plotted.

The posterior means at each time point obtained from both models match the observed data closely for most horses. Both models have similar performance for Horses 4 and 7, where the models have difficulties estimating the observed means. For the other six horses, the AR(1) model and the ANOVA model both track the observed data pattern, but do not always hit the observed means exactly. Obtaining more replicate data at each time point would improve the accuracy of the models, as was observed in Section 3.3.1.

The estimated AUCs and standard deviations for the eight horses can be seen in Tables 3.5 and 3.6, along with the estimated ϕ values for each horse corresponding to the AR(1) model. All models produced \widehat{AUC} s that were similar to the estimate obtained using the observed mean in Equation 3.2. The estimated standard deviations of \widehat{AUC} from the ANOVA model are similar to those produced by the formula-based estimator assuming independent observations. The standard deviations estimated from the AR(1) model fell within the bounds created by the two formula-based estimators for all horses where the correlation (ϕ) was estimated to be positive except for Horse 2. The posterior distributions for each mean for Horse 2 were less variable than the variances of two observations used in FB0, resulting in estimated $SD(\widehat{AUC})$ s that are smaller than the formula-based estimator.

The results from the Bayesian models should only be used if the estimated data pattern closely matches the observed data pattern. An estimate of the AUC can be obtained using the original data in Equation 3.2 when the Bayesian models fail to estimate the original data pattern well. In this scenario, a conservative estimate

of the standard deviation can be obtained using Equation 3.3 if replicate data are available at each time point, where the variance at each time point is estimated from the observed data and the underlying group means are assumed independent.

Table 3.5: \widehat{AUC} s computed for each method corresponding to the data in Figure 3.5. The Formula column was calculated using the observed means in Equation 3.2.

| Horse | Formula | ANOVA | AR(1) |
|-------|---------|-------|-------|
| 1 | 37252 | 37055 | 37079 |
| 2 | 41030 | 40762 | 40650 |
| 3 | 37874 | 37794 | 37812 |
| 4 | 37396 | 37347 | 37460 |
| 5 | 35945 | 36041 | 36051 |
| 6 | 39608 | 39504 | 39696 |
| 7 | 37461 | 37558 | 37576 |
| 8 | 40163 | 39660 | 39729 |

Table 3.6: Estimated standard deviations of the \widehat{AUC} s given in Table 3.5 for each method. FB0 and FB1 indicate the formula-based method (Equation 3.3) assuming a correlation of 0 and 1, respectively, between observations across all 13 time points. * indicates that the models did not fit the observed data well.

| Horse | FB0 | ANOVA | AR(1) ($\hat{\phi}$) | FB1 |
|-------|-------|--------|------------------------|--------|
| 1 | 435.6 | 412.4 | 645.3 (0.6) | 1284.4 |
| 2 | 674.9 | 569.7 | 594.7 (0.1) | 2212.4 |
| 3 | 443.8 | 482.4 | 618.8 (0.3) | 1417.2 |
| 4 | 655.3 | 516.4* | 676.5 (0.4)* | 2043.0 |
| 5 | 609.4 | 406.2 | 338.1 (-0.3) | 1630.8 |
| 6 | 572.2 | 537.0 | 899.9 (0.6) | 1863.2 |
| 7 | 550.8 | 326.1* | 358.7 (0.1)* | 1730.8 |
| 8 | 639.9 | 474.9 | 409.9 (-0.2) | 2019.2 |

3.4 Discussion

The AUC summary measure is used as a measure of the total concentration of a substance measured over some interval of time, however a straight-forward estimate

of the variance associated with it has been lacking. An ANOVA model and an AR(1) model are both models that can be applied to time trajectories to estimate the means at each time point, where each model has different underlying assumptions. The models can be fit using a Bayesian framework, where a posterior distribution for the mean at each time point can be obtained. The corresponding mean and variance of each posterior distribution can be computed, which then can be used to estimate both the AUC and the $Var(\widehat{AUC})$. The ANOVA model assumes uncorrelated observations, therefore $Cov(y_i, y_{i'}) = 0$ in Equation 3.3. The AR(1) model estimates the correlation between observations over time, therefore the mean of the posterior distribution of ϕ can be obtained and used to estimate $Cov(y_i, y_{i'}) = \phi^k * \sigma_{y_i} * \sigma_{y_{i'}}$. This method of estimating the variance associated with the \widehat{AUC} can be used in scenarios where only one observation is available at each time point – but only when the models fit the observed data well and assumptions about the variance at each time point can be made.

The performance of the statistical models was studied using data generated from two patterns that often arise in equine-based AUC studies. The models all estimate the observed means well when the observed pattern is approximately smooth. However, they may require more replicate values at each time point in scenarios where multiple intervals of increasing and decreasing values are observed. Estimates of the standard deviation associated with the \widehat{AUC} obtained from both models matched the underlying assumptions of the model when compared to the formula-based upper and lower bounds.

The models were applied to an equine-based glucose data set, where two measurements were obtained at each time point over a 6-hour sampling interval. A non-constant variance assumption was used, where the $\sigma_{\epsilon,i}$ were estimated from the observed replicate data. If more replicate values were obtained at each time point,

prior distributions could be placed on each $\sigma_{\epsilon,i}$. Although the models performed well for nearly all of the observed glucose-curves, they did have some problems when the observed data pattern was less structured and had large variability in the replicate values.

A top priority when applying these methods to real data is the models' ability to reconstruct the original data pattern. If the models do not estimate the observed means accurately, they should not be used since the resulting estimated standard deviations of the \widehat{AUC} would correspond to a different trajectory than what was observed. When the original data pattern can not be reconstructed by either of the models, the AUC can be estimated using the observed data and an estimate of the standard deviation can be obtained if replicate data are available to estimate the standard deviation at each time point and assumptions regarding the underlying correlation structure are made.

The AR(1) model assumes the length of the time between consecutive observations is held constant over the entire sampling period, which is not always true in AUC data sets. However, the Bayesian AR(1) model can handle missing data, therefore in these scenarios the unobserved data that could have been obtained at time points associated with equal sampling intervals can be treated as missing data before the models are applied. Individual time points are treated independently in the ANOVA model, therefore this model can be applied to AUC data sets without modification to accurately model data collected with unequally spaced sampling times.

Two commonly used modifications of the AUC summary measure are the area under the curve with respect to increase (Le Floch et al., 1990) and the area under the curve above the baseline value (Wolever and Jenkins, 1986). These modifications, often denoted AUC_I and AUC_{AB} , are both calculated using a linear combination of the mean values at each time point. The estimated mean and variance of the

modified AUC measures can be computed using the methods presented here with minimal modification.

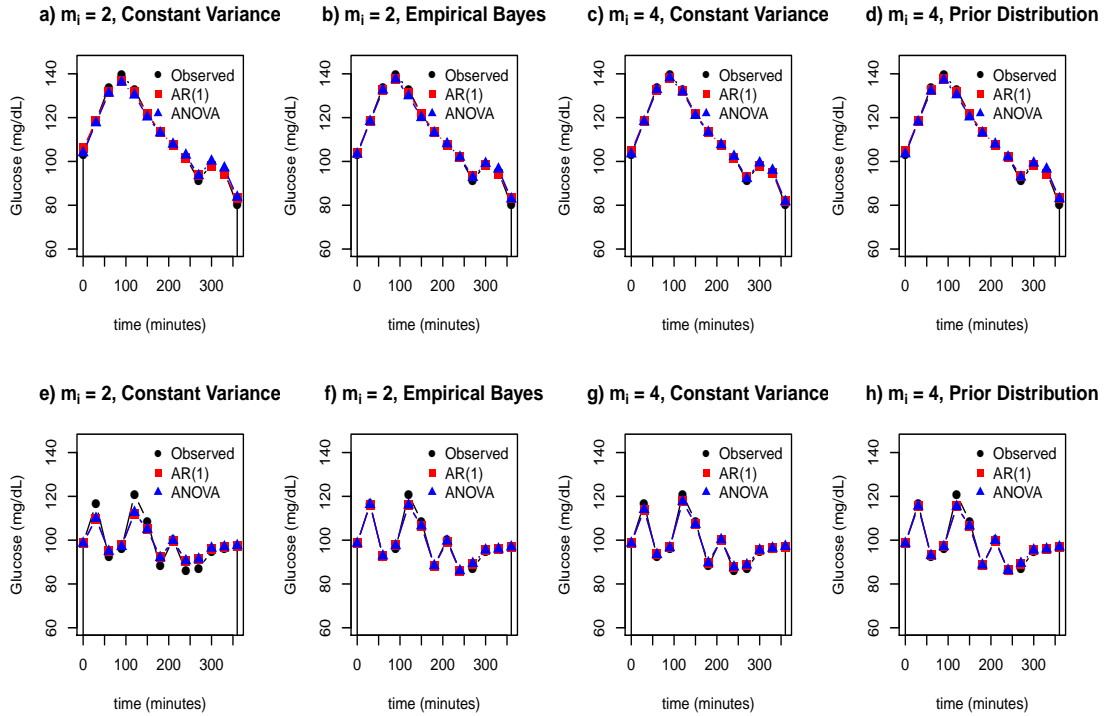


Figure 3.4: Estimated curves from each model where the replicate data have non-constant variance across time points. The observed curve, along with the fitted curve for each method are displayed, where posterior means at each time point are plotted. Results for Pattern 1 are displayed in panels (a)-(d), and results for Pattern 2 are displayed in panels (e)-(h). The panels labeled ‘Constant Variance’ use the models assuming the variance of the replicate values at each time point is constant, the models in panels labeled ‘Empirical Bayes’ use the estimated standard deviation of the data at each time point in the model, and the models in panels labeled ‘Prior Distribution’ use a prior distribution to estimate the standard deviation of the data at each time point.

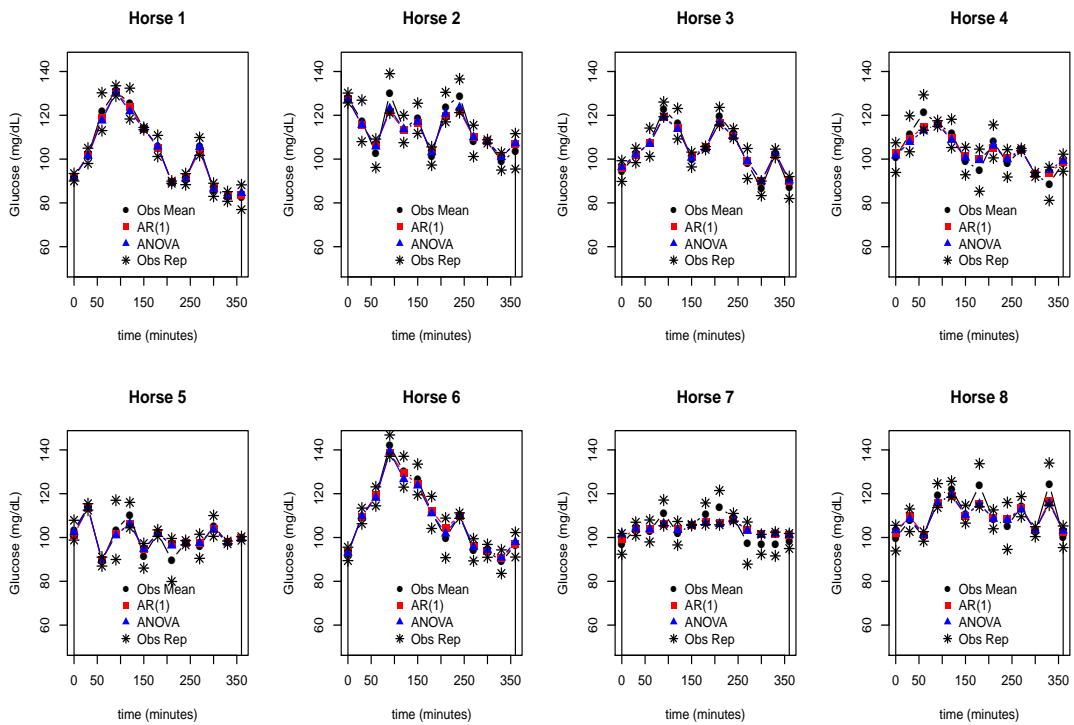


Figure 3.5: The observed curve for each horse, along with the estimated curve from each method. The replicate values at each time point are displayed to show the variability present in the replicate values across the different time points.

References

- Allison, D. B., Paultre, F., Maggio, C., Mezzitis, N., Pi-Sunyer, F. X., 1995. The use of areas under curves in diabetes research. *Diabetes care* 18 (2), 245–250.
- Armitage, P., Berry, G., Matthews, J. N. S., 2008. *Statistical methods in medical research*. John Wiley & Sons.
- Cryer, J., Chan, K., 2008. *Time Series Analysis with Applications to R*. Vol. 2. Springer.
- Gałecki, A., Burzykowski, T., 2013. *Linear mixed-effects models using R: A step-by-step approach*. Springer Science & Business Media.
- Gelman, A., Rubin, D. B., 1992. Inference from iterative simulation using multiple sequences. *Statistical science*, 457–472.
- Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (6), 721–741.
- Hilborn, R., Bue, B. G., Sharr, S., 1999. Estimating spawning escapements from periodic counts: a comparison of methods. *Canadian Journal of Fisheries and Aquatic Sciences* 56 (5), 888–896.
- Jackman, S., 2009. *Bayesian analysis for the social sciences*. Vol. 846. John Wiley & Sons.
- Le Floch, J.-P., Escuyer, P., Baudin, E., Baudon, D., Perlemuter, L., 1990. Blood glucose area under the curve: methodological aspects. *Diabetes care* 13 (2), 172–175.
- Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2012. *The BUGS book: A practical introduction to Bayesian analysis*. CRC press.
- Moreaux, S. J. J., Nichols, J. L., Bowman, J. G., Hatfield, P. G., 2011. Psyllium lowers blood glucose and insulin concentrations in horses. *Journal of Equine Veterinary Science* 31 (4), 160–165.
- Parken, C. K., Bailey, R. E., Irvine, J. R., 2003. Incorporating uncertainty into area-under-the-curve and peak count salmon escapement estimation. *North American Journal of Fisheries Management* 23 (1), 78–90.
- Peterson, J., Moreaux, S., Bowman, J., Hatfield, P., Berardinelli, J., Olsen, J., 2009. Metabolic and physical effects of psyllium supplementation on quarter horses. In: *American Society of Animal Science*. p. 176.

- Plummer, M., 2003. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In: Proceedings of the 3rd international workshop on distributed statistical computing. Vol. 124. Technische Universit at Wien Wien, Austria, p. 125.
- R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL <http://www.R-project.org/>
- Su, Y.-S., Yajima, M., 2016. R2jags: A Package for Running jags from R. R package version 0.04-03.
URL <http://CRAN.R-project.org/package=R2jags>
- Wolever, T. M., Jenkins, D., 1986. The use of the glycemic index in predicting the blood glucose response to mixed meals. The American journal of clinical nutrition 43 (1), 167–172.
- Yuan, J., 1993. Estimation of variance for auc in animal studies. Journal of pharmaceutical sciences 82 (7), 761–763.

CHAPTER FOUR

A COMPARISON OF MEASUREMENT ERROR CORRECTION METHODS
WITH AND WITHOUT KNOWN MEASUREMENT ERRORS

Contribution of Authors and Co-Authors

Author: Jennifer L. Weeding

Contributions: Responsible for the majority of the writing.

Co-Author: Dr. Mark C. Greenwood

Contributions: Provided feedback on statistical analysis and drafts of the manuscript.

Manuscript Information Page

Jennifer L. Weeding and Mark C. Greenwood

Journal of Statistical Computation and Simulation

Status of Manuscript:

Prepared for submission to a peer-reviewed journal

Officially submitted to a peer-review journal

Accepted by a peer-reviewed journal

Published in a peer-reviewed journal

Abstract

The impacts of ignoring explanatory variable measurement error in regression models have been well documented and include bias in parameter estimates and a loss of power. There are many methods available to correct for measurement error. However, each method may not be appropriate in every situation. In this research, we compare the performance of three measurement error correction methods, the method of moments correction, the Simulation-Extrapolation (SIMEX) correction, and a Bayesian correction, in a simple linear regression setting where additive measurement errors are present in the explanatory variable. Simulation studies show the three methods have similar performance in regards to estimating the slope coefficient when the measurement error variance is small. However, as the measurement error variance increases the Bayesian measurement error model has overall better performance compared to the other two methods. Therefore, we recommend using the Bayesian measurement error model when classical additive measurement error is present, especially when sample sizes are not large.

4.1 Introduction

Measurement error can arise in many research situations and occurs when a variable of interest cannot be observed exactly, but instead is observed with error. Measurement error, also referred to as *errors-in-variables* (Stefanski, 2000), can occur in a variety of circumstances, including, but not limited to miscalibration of a measuring instrument, sampling error, misclassification, modeling of abundance estimates, and whenever the analyzed responses are estimated. The consequences of ignoring measurement error in analyses have been well documented and include bias in parameter estimation and a loss of power (Carroll et al., 2006).

Measurement error in a continuous random variable can be classified as either Berkson or classical. Berkson measurement error (Berkson, 1950) occurs when a researcher attempts to achieve a target value (W_i), but instead achieves a true value (X_i). Classical measurement error, which occurs when the true value of the continuous

random variable (X_i) is measured with error resulting in an observed value (W_i), will be the focus of this work. The classical additive measurement error model for the mismeasured random variable W_i is

$$W_i = X_i + U_i, \tag{4.1}$$

where U_i represents the measurement error, $E(U_i|x_i) = 0$, and $Var(U_i|x_i) = \sigma_u^2$. A result of the mean 0, additive error structure is that W is an unbiased estimator of X since $E(W_i|x_i) = x_i$. Although the additive error structure is the focus here, there are other ways for measurement errors to occur (for example, linear: $W_i = \theta_0 + \theta_1 X_i + U_i$ or multiplicative: $W_i = X_i * U_i$).

To account for measurement errors when present, the measurement error variance, σ_u^2 , must either be assumed known or estimated from replicate data (repeated independent observations on the same subject at the same time). Many methods exist to correct for measurement error, however each method is not appropriate in every situation. The purpose of this study is to compare the performance of popular correction methods in the simple linear regression (SLR) setting when classical additive measurement error is present in the explanatory variable. Regression calibration is a commonly used measurement error correction method; Bartlett and Keogh (2016) compare a Bayesian approach to regression calibration, concluding that the Bayesian method has a number of statistical advantages over regression calibration techniques. In our simulation studies, we compare three measurement error correction methods: the Method of Moments correction (see Buonaccorsi (2010)), the SIMEX correction (Cook and Stefanski, 1994), and a Bayesian method similar to what is presented in Section 9.8 of Carroll et al. (2006). We believe the results from this study will be helpful for researchers when deciding which method to use.

This paper is organized as follows. In Section 2, a brief overview of the effects of measurement error in a SLR model are given. In Section 3, we review the three correction methods. In Section 4, we conduct a simulation study to compare the performance of the three correction methods in different scenarios. Section 5 concludes with a summary of our findings. The R code (R Core Team, 2016) used for the simulation study can be found in the Appendix.

4.2 Measurement Error in Simple Linear Regression

Measurement error in regression models has a long history, yet it still seems to be rare to find subject-area usage of these techniques even though it is present in many data sets. Reasons why researchers may fail to account for measurement error is that extra information (or assumptions) about measurement error parameters are needed to account for measurement error and historically there has been a lack of statistical software readily available to deal with measurement error problems. This work will examine the effects of measurement error when using a SLR model, however the methods presented easily extend to multiple regression models.

The SLR model is implemented in situations where a linear relationship between a quantitative response variable and a single quantitative explanatory variable is of interest. If we let Y and X denote the response and explanatory variables respectively, the SLR model is

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad (4.2)$$

where $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. Often it is not possible to measure the true value of the response variable, explanatory variable, or sometimes both, resulting in measurement error in one or both variables in the SLR model. Here we focus on measurement error in the explanatory variable where (Y_i, W_i) represent the observed response and mis-

measured explanatory variable containing classical additive measurement error. The SLR model becomes

$$Y_i = \beta_0 + \beta_1(X_i + U_i) + \epsilon_i$$

$$\Rightarrow Y_i = \beta_0 + \beta_1 W_i + \epsilon_i, \quad (4.3)$$

where $U_i|x_i \sim N(0, \sigma_u^2)$ and $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. Here, the measurement error variances (σ_u^2) are assumed to be constant across observations. However, the model and notation above can be altered to allow them to vary across observations (see Buonaccorsi (2010)).

In the SLR setting, the naive (ordinary least squares regression of Y_i on W_i) estimator of the slope is biased towards 0 (called attenuation) because

$$\beta_{1,naive} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_u^2} \beta_1 = k\beta_1, \quad (4.4)$$

where $k = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_u^2}$ (referred to as the reliability ratio (Fuller, 2009)), β_1 is the true slope, σ_X^2 is the variance of the random variable X , and σ_u^2 is the measurement error variance associated with the mis-measured explanatory variable. Larger measurement error variance associated with the mis-measured explanatory variable (σ_u^2) results in $\beta_{1,naive}$ being a more biased estimator, as can be seen in Equation 4.4 and Figure 4.1. The estimator of $\beta_{1,naive}$ is unbiased when measurement error is present in the response variable only, as Equation 4.4 only contains σ_X^2 and σ_u^2 even when the response is measured with error. However, $\sigma_{\epsilon,naive}^2$ is inflated resulting in less precise inferences for β_1 and a loss of power for detecting effects. We do not focus on this scenario here.

The measurement error variance must be known or estimated to account for measurement errors when present. In some situations, σ_u^2 is assumed known,

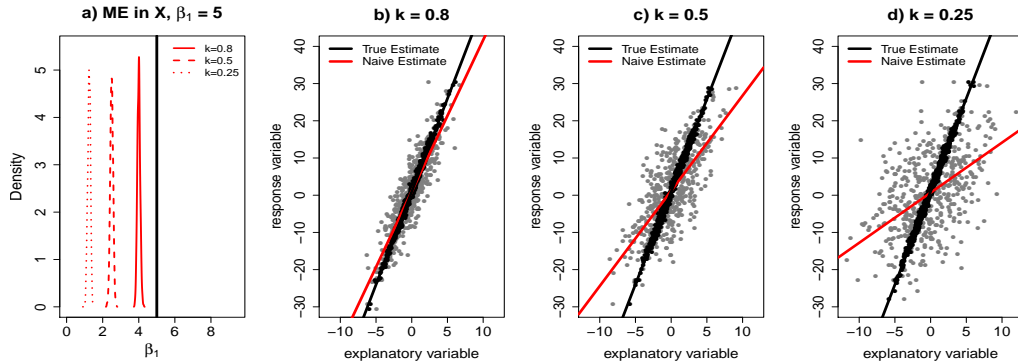


Figure 4.1: Data containing measurement error in the explanatory variable were generated with a true slope of $\beta_1 = 5$. Three values of the reliability ratio, $k = (0.8, 0.5, 0.25)$, were considered to show the effect of increasing measurement error on the estimate of β_1 . 500 simulations were used for each value of k , with each data set consisting of $n = 500$ observations. (a) Distributions of $\hat{\beta}_{1,naive}$ for each value of k . Vertical bold line represents the true value of β_1 . (b)-(d) The estimated regression line for the uncontaminated data (black data points) and contaminated data (grey data points) for one realization for each value of k .

and is based on calibration parameters or results from previous studies. In most situations, it must be estimated. For example, measurement error is present when summary measures (mean, median, etc.) are used and σ_u^2 can be estimated using the standard estimate of the variability associated with the measure. In more complex situations, the errors could be derived from models for each individual response such as abundance estimates from mark-recapture studies or analysis of regression coefficients.

4.3 Review of Three Common Correction Methods

4.3.1 Method of Moments

The method of moments measurement error correction is a formula-based correction method derived by equating moments specific to the model parameters of interest (here, SLR parameters). Buonaccorsi (2010) contains moment-based formulas to correct for measurement error for many different modeling situations including

SLR. The moment-based formula for the measurement error corrected estimator of the slope in the SLR setting is

$$\hat{\beta}_{1,MOM} = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_W^2 - \hat{\sigma}_u^2} = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X^2}. \quad (4.5)$$

where $\hat{\sigma}_{XY}$ is the estimated covariance between the observed response and unobserved explanatory variable, $\hat{\sigma}_W^2$ is the estimated variance of the mis-measured explanatory variable (W), and $\hat{\sigma}_u^2$ is the estimated measurement error variance, with $\hat{\sigma}_X^2 = \hat{\sigma}_W^2 - \hat{\sigma}_u^2$.

If there are $j = 1, 2, \dots, m_i$ replicates of the mis-measured explanatory variable for the i^{th} subject, then $W_{ij} = X_i + U_{ij}$. In this scenario, the overall measurement error variance, $\hat{\sigma}_u^2$, can be estimated using the average of the estimated variances across the n observations, $\hat{\sigma}_u^2 = \frac{\sum \hat{\sigma}_{u,i}^2}{n}$. Otherwise, the researcher must make assumptions about the value of the measurement error parameter.

4.3.2 SIMEX

Cook and Stefanski (1994) introduced the SIMEX approach for accounting for measurement errors when the measurement error variance is known or can be reasonably well estimated. This method is a general framework applicable to many situations and is available in the R package ‘simex’ (Lederer and Kchenhoff, 2013). The SIMEX method defines the mis-measured explanatory variable as

$$W_i = X_i + \sigma_u Z_i, \quad (4.6)$$

where Z_i is a standard normal random variable independent of Y_i and X_i , and σ_u^2 is the assumed known measurement error variance.

The first step of the SIMEX procedure is the simulation of pseudo data that contains additional measurement error than what was actually observed. This occurs

through the use of a λ_j vector ($\lambda_j > 0$) and

$$W_{b,i}(\lambda_j) = W_i + \sqrt{\lambda_j} \sigma_u Z_{b,i}, \quad (4.7)$$

where $(Z_{b,i})_{i=1}^n$ are mutually independent and identically distributed standard normal random variables, $\lambda_j = (0.5, 1, 1.5, 2)$, and $b = 1, 2, \dots, B$. For each value of λ_j , B simulated data sets containing additional measurement error are generated and the original model is fit and parameter estimates are obtained (estimates for both the intercept and slope are obtained). Once this process is complete for all values of λ_j , a linear, quadratic, or nonlinear model is fit using the mean of the B estimates for each λ_j as the response variable and the λ_j vector as the explanatory variable. Finally, the SIMEX corrected β_0 and β_1 estimates are obtained by extrapolating this model back to $\lambda_j = -1$.

One important aspect of using the SIMEX method is specifying which extrapolation model to use when modeling the estimated coefficients from the simulation process. To be explicit, the extrapolation models are as follows: the linear model ($\mu(\lambda_j) = b_0 + b_1 \lambda_j$), the quadratic model ($\mu(\lambda_j) = b_0 + b_1 \lambda_j + b_2^2 \lambda_j^2$) and the nonlinear model ($\mu(\lambda_j) = a + \frac{b}{c + \lambda_j}$), with $\mu(\lambda_j)$ representing the mean of the B estimates for each value of λ_j . Figure 4.2 shows the results from using each of the three extrapolation models on three simulated data sets, with each data set containing a different amount of measurement error. As the measurement error increases in these three data sets, the linear and quadratic extrapolation functions result in more biased estimates of the slope parameter, while the nonlinear extrapolation function seems to do better when estimating the slope parameter across all three values of the reliability ratio. The simulation study below explores the performance of the quadratic and nonlinear extrapolation functions in more detail.

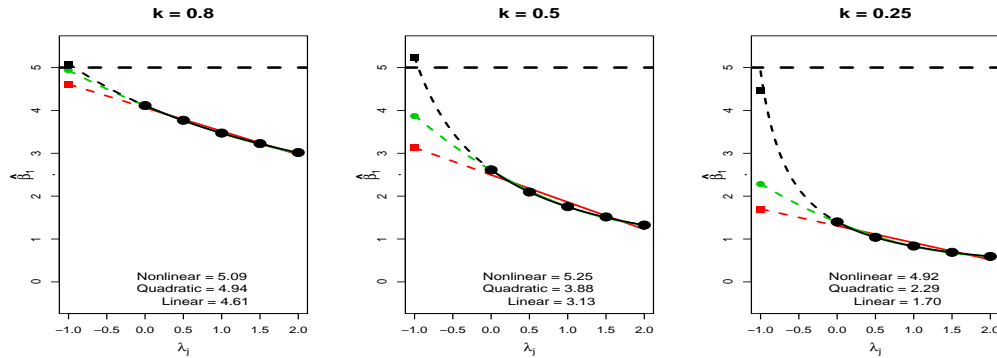


Figure 4.2: 500 observations were generated with a true slope of $\beta_1 = 5$ containing measurement error in the explanatory variable. Three values of the reliability ratio, $k = (0.8, 0.5, 0.25)$, were considered to show the effect of increasing measurement error on the estimate of β_1 for each extrapolation function.

4.3.3 Bayesian Measurement Error Models

Although a variety of Bayesian methods exist to correct for measurement error in the explanatory variable, they all have the same underlying steps. Chapter 9 in Carroll et al. (2006) is devoted to Bayesian measurement error models and describes five general steps in developing a Bayesian measurement error model as: (i) specify the likelihood model as if X were observed, (ii) select the appropriate measurement error model (additive, linear, multiplicative, etc.), (iii) form the likelihood function as if X were observed, (iv) select appropriate prior distributions, and (v) compute the complete conditional distributions. Here, we present a Bayesian measurement error model that corrects for classical additive measurement error when present in the explanatory variable in the SLR setting.

In a Bayesian analysis where measurement error is present in the explanatory variable, the true *unobserved* value (X_i) is treated like a latent variable and given a prior distribution. Carroll et al. (2006), along with Gilks et al. (1996), both describe the Bayesian SLR measurement error model using three model pieces: an outcome

model, a measurement model, and an exposure model. The outcome model is a model for the outcomes that you would obtain if measurement error was not present and is

$$Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma_\epsilon^2). \quad (4.8)$$

The measurement model is a model for the mis-measured variable given the true variable and is

$$W_i | X_i \sim N(X_i, \sigma_u^2). \quad (4.9)$$

This is a direct result of Equation 4.1, which implies that when classical additive measurement error is present in the explanatory variable, $E(W_i | x_i) = x_i$ and $Var(W_i | x_i) = \sigma_u^2$. The exposure model is the model for the true unobserved variable (X_i) and, by centering the observed variable, is

$$X_i \sim N(0, \sigma_x^2). \quad (4.10)$$

In this model, σ_u^2 is assumed known and all other parameters (β_0 , β_1 , $\tau_X = \frac{1}{\sigma_x^2}$, $\tau_\epsilon = \frac{1}{\sigma_\epsilon^2}$) are given prior distributions, which depend on the particular situation and are often selected to be non-informative.

One major challenge in accounting for measurement errors is the estimation of the measurement error variance (Muff et al., 2015). If replicate data are available (for example, observing W_{ij} replicates for each X_i , where $j = 1, 2, \dots, m_i$), that information can be used to generate a prior distribution for the measurement error variance which can then be estimated from the model. This feature is unique to the Bayesian measurement error methods, as other correction methods require the measurement error variances be known, or estimated from replicate data and then assumed known, in order to use the method. The impacts of using estimated measurement errors

are considered below. The Bayesian method is able to capture the uncertainty in σ_u^2 that the other methods are missing, since they are assuming σ_u^2 is known with complete certainty. Also, situations may arise where non-constant measurement error is present. The model presented here can extend to a non-constant measurement error setting by replacing σ_u^2 in Equation 4.9 with $\sigma_{u,i}^2$.

4.4 Simulation Studies

Carroll et al. (2006) state that the quadratic extrapolation function results in a conservative correction for β_1 . However, the documentation related to the `simex` package in R (Lederer and Kchenhoff, 2013) notes that the nonlinear extrapolation function is numerically unstable and advise against using this option. In the simulation studies, both the quadratic and nonlinear extrapolation functions were considered, while the linear extrapolation function was omitted since Cook and Stefanski (1994) report the bias associated with this extrapolation function is an order of magnitude larger than the bias associated with either the quadratic or nonlinear extrapolation functions.

For the Bayesian method, JAGS (Plummer, 2003) was used to implement the Gibbs sampling algorithm (Geman and Geman, 1984), resulting in samples from the joint posterior distribution of the parameters. Three independent chains, each with random starting values, were used and the first 1000 samples were discarded as *burn-in*. Each chain was run for 60,000 iterations and a thinning rate of 12 was applied, resulting in a final chain length of 5,000 iterations. Convergence was assessed using sample path plots and potential scale reduction factors, where a threshold of $\hat{R} < 0.05$ was used to determine convergence. The prior distributions used were $\beta_0 \sim N(0, 10^5)$, $\beta_1 \sim N(0, 10^5)$, $\tau_X \sim \text{gamma}(0.5, 2)$, and $\tau_\epsilon \sim \text{gamma}(0.5, 2)$. The

$\text{gamma}(a, b)$ notation in JAGS specifies a Gamma distribution with mean a/b and variance a/b^2 . The mean of the posterior distribution from each simulation is used as the estimate of β_1 .

4.4.1 Known Measurement Error

4.4.1.1 Simulation Study 1 Setup We performed simulation studies to look at the performance of the three correction methods in different scenarios and, for comparison, the naive SLR model (the ordinary least squares model with no measurement error correction) results are included. To see the effects of changing sample size and measurement error on the different correction methods, two sample sizes ($n = 25, 500$) and three values of the measurement error variance ($\sigma_u^2 = 1, 4, 12$) were considered. Overall, twelve different cases were considered with 500 simulations for each case. The parameter values used in the simulation study are given in Table 4.1. The simulation studies were conducted in R (R Core Team, 2016) and used the R-packages `simex` (Lederer and Kchenhoff, 2013) and `R2jags` (Su and Yajima, 2016). Results are summarized numerically and using beanplots (Kampstra, 2008) which display each result (small lines), the mean of the simulations (large line), and nonparametric density estimates.

All three correction methods require the measurement error variance be known when replicate data are not available. Simulation study 1 used the true parameter value of the measurement error variance ($\sigma_u^2 = 1, 4, \text{ or } 12$) as the known/estimated measurement error variance for each of the methods.

4.4.1.2 Simulation Study 1 Results The distributions of $\hat{\beta}_1$ resulting from the simulations are shown in Figures 4.3 and 4.4, and numerical summaries are provided in Table 4.2. The attenuation of the naive estimator can be seen in all cases, with

Table 4.1: Simulation Study 1 parameter values.

| n | β_0 | β_1 | σ_ϵ^2 | σ_X^2 | σ_u^2 |
|-----------|-----------|-----------|---------------------|--------------|--------------|
| (25, 500) | 1 | (1, 5) | 4 | 4 | (1, 4, 12) |

greater attenuation occurring for increasing measurement errors and larger β_1 . In general, increasing the sample size results in less variable estimators that contain less bias than smaller sample sizes.

The measurement error correction methods studied here have comparable performance when the measurement error variance is small and the sample size is large, however distinctions between the methods become more clear when the measurement error variance becomes larger ($\sigma_u^2 = 4, 12$). The method of moments estimator is approximately unbiased in all cases when the sample size is large, but becomes highly variable and depends on how well the measurement error correction parameters are estimated when the measurement error variance is large and/or the sample size is small as can be seen in panels (b) and (c) of Figures 4.3 and 4.4. The quadratic SIMEX method performs well when measurement errors are small, however this method becomes more biased (conservative) as the measurement error increases or as the size of the effect (β_1) increases. For the larger sample size, many of the nonlinear SIMEX corrections look quite reasonable but the method produces occasional massive over-corrections that dramatically bias its average and increase its variance. The instability of the method is especially pronounced when the sample size is small, as can be seen in the variability of the simulated estimates in panels (a)-(c) of Figures 4.3 and 4.4. The Bayesian measurement error correction method has small bias when the sample size is small, as can be seen in in Table 4.2 and panels

(a)-(c) of Figures 4.3 and 4.4, however it is still the least biased option of the methods considered. When the sample size is large, the Bayesian measurement error correction method and the method of moments correction have similar performance in regards to bias, as can be seen in panels (d)-(f) of Figures 4.3 and 4.4. However, the method of moments correction becomes more variable as the measurement error increases. Overall, the Bayesian measurement error correction method appears to be the best correction method of those considered, producing estimates that are routinely among the best in terms of bias and are as variable or less variable than those produced from the other methods studied here.

Table 4.2: Simulation 1 results for $\hat{\beta}_1$ for each method based on 500 simulations in each scenario. The method of moments correction is labeled MOM.

| n | β_1 | σ_u^2 | Naive mean (SD) | MOM mean (SD) | SIMEX.Q mean (SD) | SIMEX.NL mean (SD) | Bayes mean (SD) |
|-----|-----------|--------------|--------------------|------------------|----------------------|-----------------------|--------------------|
| 25 | 1 | 1 | 0.792 (0.197) | 1.021 (0.282) | 0.957 (0.246) | 1.062 (0.758) | 1.059 (0.289) |
| | | 4 | 0.494 (0.179) | 1.429 (4.086) | 0.736 (0.279) | 1.025 (4.603) | 1.217 (0.397) |
| | | 12 | 0.247 (0.139) | 2.865 (52.330) | 0.414 (0.243) | 0.738 (5.518) | 1.062 (0.489) |
| 25 | 5 | 1 | 4.008 (0.441) | 5.163 (0.786) | 4.848 (0.588) | 4.963 (2.593) | 5.135 (0.583) |
| | | 4 | 2.524 (0.556) | 7.057 (19.996) | 3.767 (0.891) | 6.969 (24.819) | 5.446 (1.056) |
| | | 12 | 1.275 (0.480) | 13.232 (201.926) | 2.146 (0.849) | 5.917 (169.964) | 5.558 (1.424) |
| 500 | 1 | 1 | 0.801 (0.046) | 1.002 (0.060) | 0.971 (0.057) | 1.001 (0.060) | 1.004 (0.060) |
| | | 4 | 0.500 (0.041) | 1.004 (0.106) | 0.746 (0.063) | 1.009 (0.118) | 1.026 (0.112) |
| | | 12 | 0.250 (0.031) | 1.038 (0.272) | 0.418 (0.053) | 1.442 (0.998) | 1.161 (0.235) |
| 500 | 5 | 1 | 3.996 (0.096) | 4.999 (0.146) | 4.846 (0.131) | 5.018 (0.443) | 5.015 (0.137) |
| | | 4 | 2.494 (0.113) | 5.008 (0.395) | 3.723 (0.182) | 5.043 (0.451) | 4.979 (0.251) |
| | | 12 | 1.246 (0.098) | 5.164 (1.217) | 2.084 (0.171) | 5.036 (0.426) | 4.924 (0.424) |

4.4.2 Unknown Measurement Error

4.4.2.1 Simulation Study 2 Setup A second simulation was run to look at the performance of the three correction methods when replicate data are available to estimate the measurement error variance in contrast to the previous results that assumed the measurement error variance was known. Two replicate sample sizes were

considered to assess model performance based on how well the measurement error variance is known. Data were simulated such that the true underlying measurement error variance was the same for both scenarios. The mean of the replicate values ($\overline{W}_i = \frac{W_{ij}}{n_{rep}}$ for $j = 1, 2, \dots, n_{rep}$) was used as the mis-measured variable (W_i) and the variance of the means are the estimates of the measurement error variances ($\sigma_{u,i}^2 = \frac{Var(w_{ij})}{n_{rep}}$ for $j = 1, 2, \dots, n_{rep}$). The data were generated using a constant measurement error variance model so the mean of the individual measurement error variances was used as the overall estimate of σ_u^2 for all methods. Note that the raw replicate observations can be used in the Bayesian measurement error model, but to allow for direct comparison to the other methods, we used the mean of the replicate values and the estimated measurement error variance in the Bayesian model here. The parameter values used for this simulation are given in Table 4.3.

4.4.2.2 Simulation Study 2 Results The distributions of $\hat{\beta}_1$ resulting from the simulations are shown in Figure 4.5, and numerical results are shown in Table 4.4. To see the effect of how well the measurement error variance is known, panel (a) can be compared to panel (d), panel (b) to panel (e), and panel (c) to panel (f) in Figure 4.5. The correction methods studied here have comparable performance when estimating β_1 when the measurement error variance is small ($\sigma_u^2 = 1$), where the estimates were less variable when more information (replicated values) was available to estimate measurement error variance. The method of moments estimator is dependent on how well the measurement error variances are estimated, and exhibits high variability when measurement errors are large for both values of replication, as can be seen in panels (c) and (f) of Figure 4.5. The quadratic extrapolation function in the SIMEX method has similar performance across the two values of replication, where the bias of the method increases as σ_u^2 increases. The instability of the nonlinear extrapolation

Table 4.3: Simulation Study 2 parameter values.

| N_{rep} | n | β_0 | β_1 | σ_ϵ^2 | σ_X^2 | σ_u^2 |
|-----------|-----|-----------|-----------|---------------------|--------------|--------------|
| (2, 20) | 25 | 1 | 5 | 4 | 4 | (1, 4, 12) |

function in the SIMEX method is clearly visible in Table 4.4, however we also saw noticeable performance issues when the sample size was small in Section 4.4.1. All panels in Figure 4.5 show the Bayesian correction method is the least biased method considered, and the estimates are less variable when more information is available to estimate the measurement error variance as seen in Table 4.4.

Table 4.4: Simulation 2 results for $\hat{\beta}_1$ for each method based on 500 simulations in each scenario. The method of moments correction is labeled MOM.

| N.rep | σ_u^2 | Naive mean (SD) | MOM mean (SD) | SIMEX.Q mean (SD) | SIMEX.NL mean (SD) | Bayes mean (SD) |
|-------|--------------|--------------------|------------------|----------------------|-----------------------|--------------------|
| 2 | 1 | 4.000 (0.446) | 5.209 (1.092) | 4.840 (0.658) | 5.143 (1.042) | 5.081 (0.636) |
| | 4 | 2.502 (0.549) | 7.285 (93.873) | 3.703 (0.916) | 5.717 (11.769) | 5.260 (1.179) |
| | 12 | 1.264 (0.474) | -4.870 (127.874) | 2.092 (0.831) | 5.051 (41.807) | 5.192 (1.499) |
| 20 | 1 | 4.023 (0.467) | 5.183 (0.824) | 4.875 (0.614) | 5.097 (0.772) | 5.171 (0.607) |
| | 4 | 2.511 (0.556) | 5.863 (13.975) | 3.758 (0.930) | 7.718 (49.305) | 5.538 (1.143) |
| | 12 | 1.232 (0.454) | 0.981 (47.924) | 2.071 (0.823) | 13.025 (138.008) | 5.658 (1.454) |

4.5 Discussion

Measurement error in the explanatory variable is often the focus of measurement error correction methods, as the naive estimator of the slope coefficient is biased towards 0. Simulation studies were used to compare three measurement error correction methods in this scenario. The magnitude of the measurement error variance (σ_u^2) relative to the variance of the unobserved explanatory variable (σ_X^2) can have

a big impact on the performance of the methods, as seen in Section 4.4. Although all methods performed relatively well when the measurement error variances were small, the SIMEX method and the method of moments correction had noticeable performance issues when measurement error variances increased. The estimator of β_1 from the quadratic SIMEX method is conservative, as previously documented, but when measurement error variances are small this method performs quite well. The nonlinear SIMEX method appears to have acceptable performance in some cases when the sample size is large, however the instability of the method emerges when the sample size is small. Both SIMEX and the method of moments estimators are highly variable when measurement error variances are large. The Bayesian measurement error model provided an approximately unbiased estimator of β_1 in all cases considered in the simulation study and results were as variable or less variable than the method of moments estimator and the nonlinear SIMEX method.

This study focused on scenarios where the measurement error variance is constant, however there are many situations where non-constant measurement errors arise such as when summary measures are used to estimate the true value of the unknown variable or when the sampling effort changes across different observations in the study. The three measurement error correction methods studied here all have adjustments that allow for non-constant measurement errors. The method of moments correction uses the average of the individual measurement error variances in the correction formula, while both the SIMEX method and the Bayesian measurement error method can directly use the individual measurement error variances for each observation and should provide better corrections than the method of moments correction when varying measurement error is present.

Measurement errors should be taken into account when they are present in order to obtain the most accurate and precise results as the naive slope estimates

and those from the best corrections are quite distinct. We recommend using the Bayesian measurement error model when measurement errors are present. This method provides approximately unbiased results and can extend to cases where non-constant measurement error is present and to other models (multiple linear regression, nonlinear regression, etc.), suggesting a variety of different areas of application for these methods. We realize that each individual researcher has a specific skill set, and that some are not equipped or willing to perform a Bayesian analyses. In these cases, the method of moments correction or the SIMEX method may be preferred as they can be more easily implemented and may provide useful corrections when the reliability ratio is large and/or slope coefficients are smaller in relative magnitude if the sample size is large. It may also be possible to avoid some of the extreme over-corrections that the nonlinear extrapolation function generated in SIMEX in particular applications using subject area knowledge.

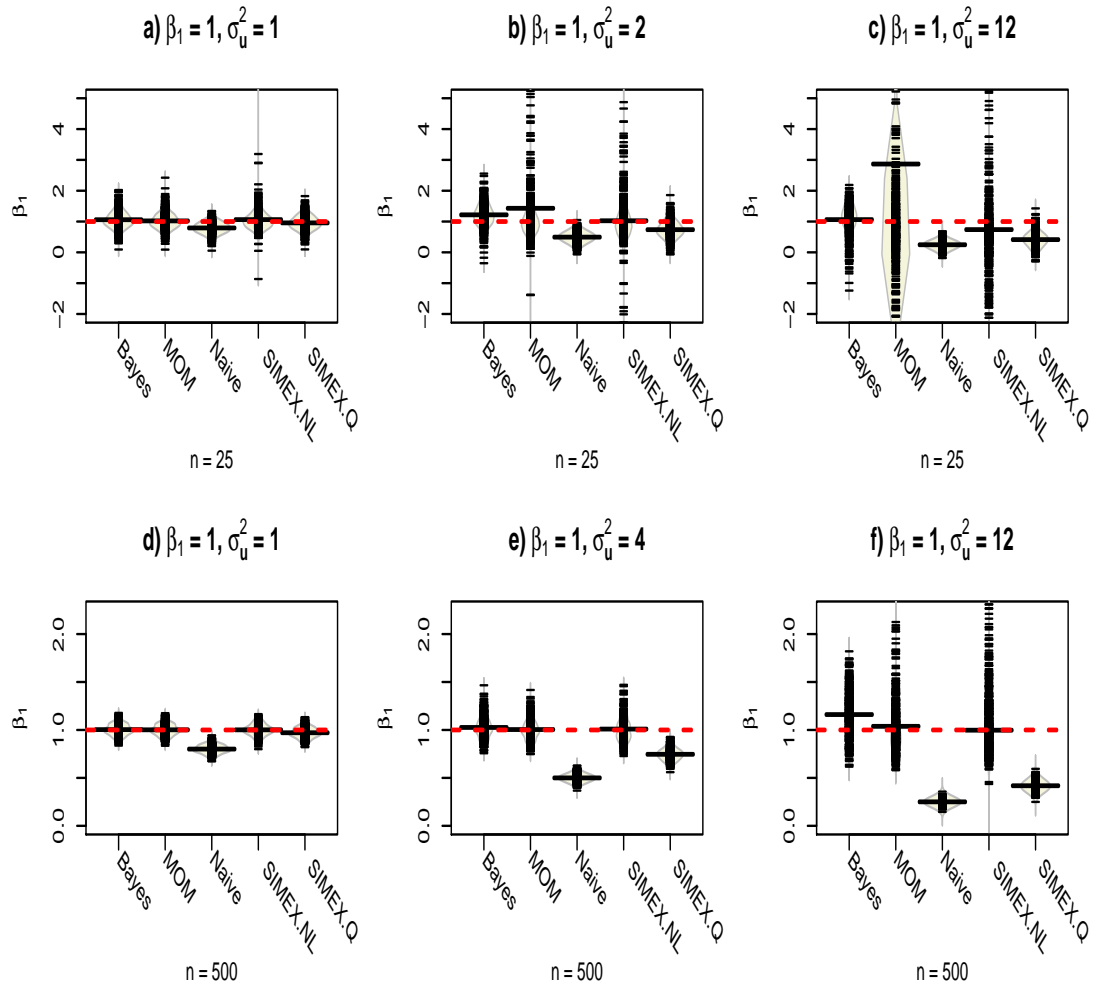


Figure 4.3: Simulation 1 results for $\beta_1 = 1$ for each method based on 500 simulations in each scenario. Horizontal dashed lines appear at the target value ($\beta_1 = 1$). Panels (a)-(c) have a sample size of 25 and panels (d)-(f) have a sample size of 500, both with increasing measurement error as you go from left to right. Estimates outside of $(-2, 5)$ for simulations corresponding to $n = 25$ and outside of $(0, 2.5)$ for simulations corresponding to $n = 500$ were not displayed to allow comparison of the majority of results.

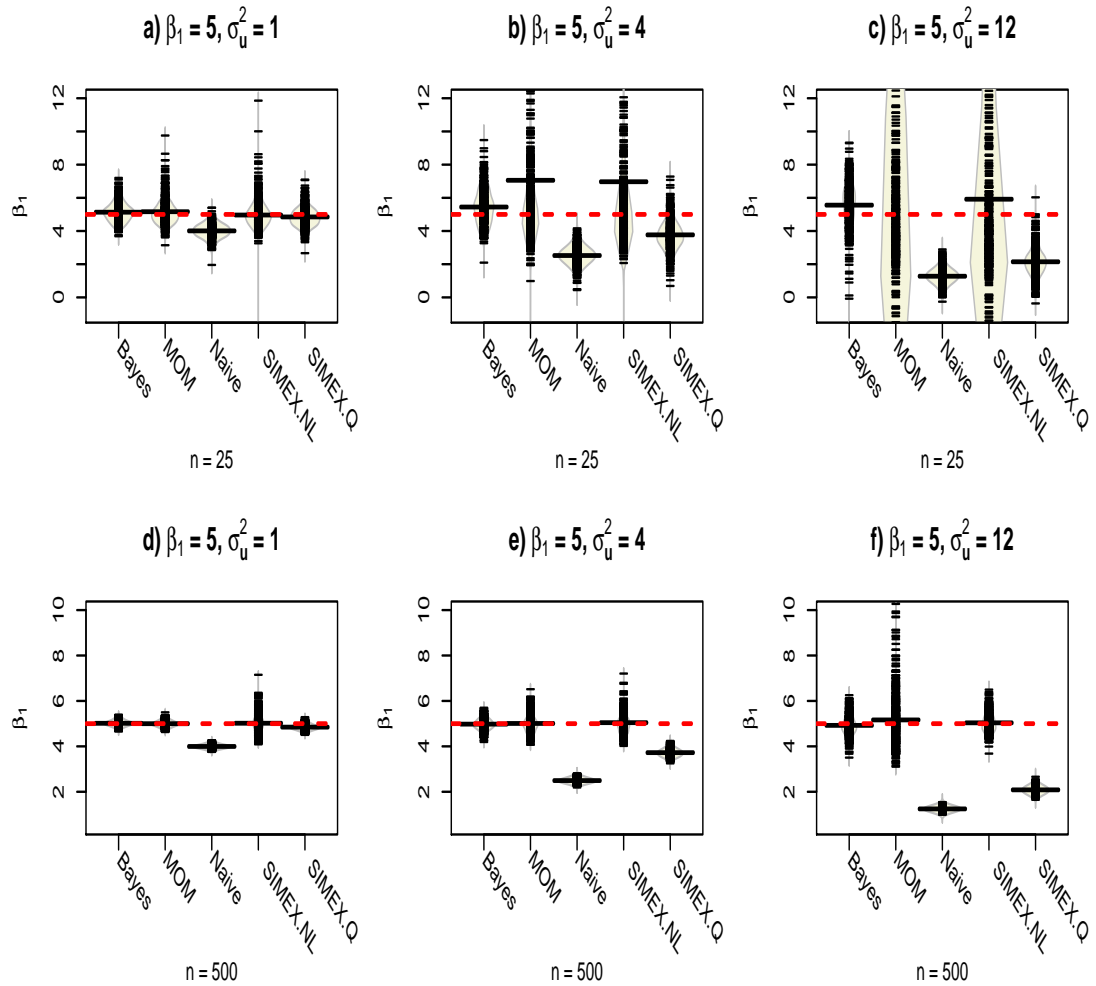


Figure 4.4: Simulation 1 results for $\beta_1 = 5$ for each method based on 500 simulations in each scenario. Horizontal dashed lines appear at the target value ($\beta_1 = 5$). Panels (a)-(c) have a sample size of 25 and panels (d)-(f) have a sample size of 500, with increasing measurement error as you go from left to right. Estimates outside of $(-1, 12)$ for simulations corresponding to $n = 25$ and above 10 for simulations corresponding to $n = 500$ were not displayed to allow comparison of the majority of results.

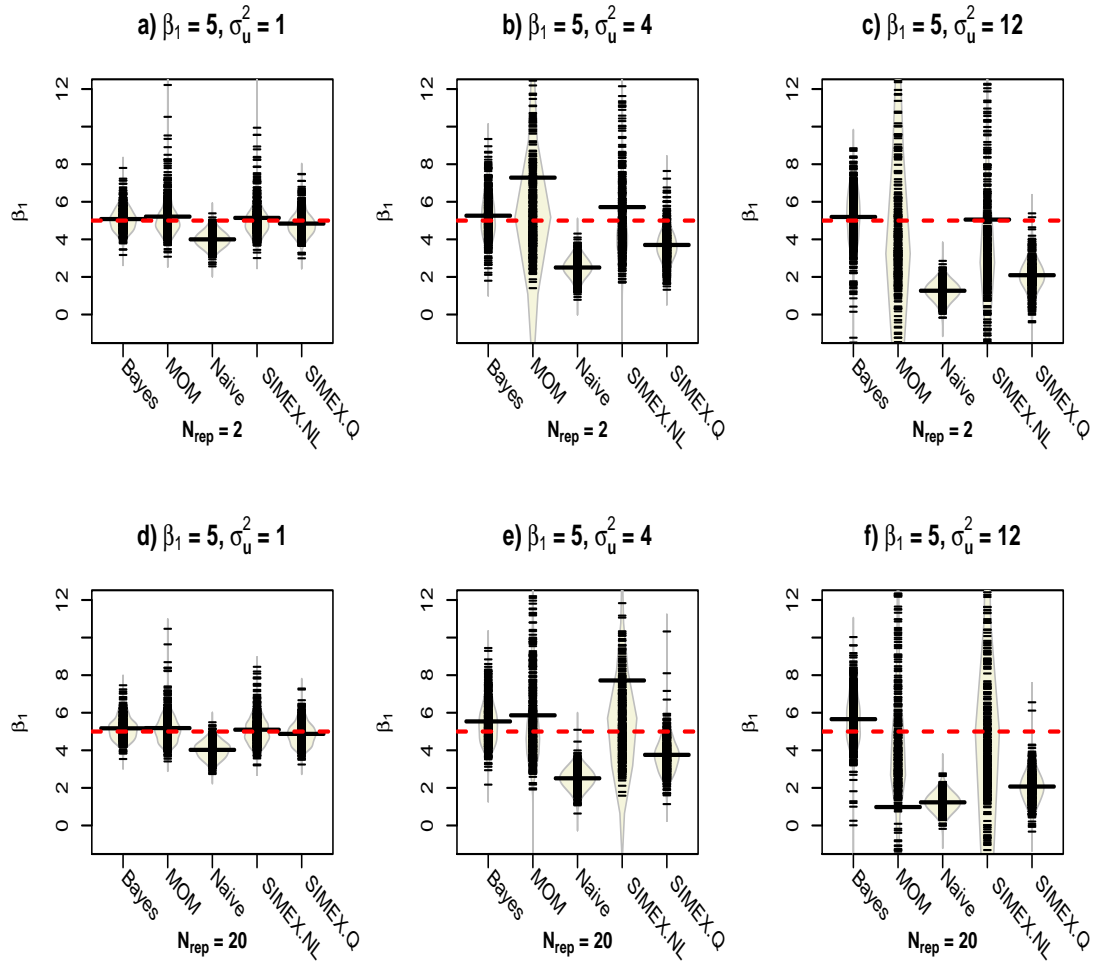


Figure 4.5: Simulation 2 results for $\beta_1 = 5$ for each method based on 500 simulations in each scenario, using a sample size of 25. Horizontal dashed lines appear at the target value ($\beta_1 = 5$). Panels (a)-(c) represent the cases where 2 replicate values were available to estimate σ_u^2 and panels (d)-(f) represent the cases where 20 replicate values were available to estimate σ_u^2 . The measurement error increases from left to right. Estimates outside of $(-1, 12)$ were not displayed to allow comparison of the majority of results.

4.6 Appendix

This section contains the R code for the Bayesian measurement error model and for the simulations performed in this study.

4.6.1 Bayesian Measurement Error Model Formulation

The Bayesian measurement error model is given below. This model is stored in a file named ‘BayesianMEmodel.txt’, which gets called during the simulation.

```
for(i in 1:N){
  Y[i]~dnorm(meanY[i],taueps)
  meanY[i]<-alpha+beta*X[i]
  for(j in 1:Nrep){
    W[i,j]~dnorm(X[i],tauu)
  }
  X[i]~dnorm(0,taux)
}
taueps~dgamma(0.5,2)
taux~dgamma(0.5,2)
alpha~dnorm(0, .000001)
beta~dnorm(0, .000001)
sigmaeps <- 1/sqrt(taueps)
sigmax <- 1/sqrt(taux)
```

4.6.2 Simulation Study 1 Code

The R code used to execute the simulations in Simulation Study 1 is given below. For each simulation, the data are generated and each corrected estimate is obtained.

```

require(simex)
require(R2jags)
N <- 25          ## number of observations (25,500)
N.sim <- 500     ## number of simulations
N.rep <- 1       ## number of reps of the mis-measured
                  explanatory variable
observed.sig.ux <- 1 ## ME SD (1,2,sqrt(12))
theta <- c(1, 1)  ## linear regression parameters (1,1) or (1,5)
sig_x <- 2.0
sig_eps <- 2

mom.slope <- simex.resultsQ <- simex.resultsNL <- post.dist <-
  naive.results <- numeric(0)
for(i in 1:N.sim){
  xstar <- rnorm(N, 0, sig_x)          ## true value of x
  x_true <- cbind(rep(1,N), xstar)    ## design matrix of model
  eps <- rnorm(N, 0, sig_eps)
  ystar <- x_true%*%theta + eps        ## true value of y
  y.obs <- c(ystar)                   ## observed y values
  ux <- rnorm(N, 0, observed.sig.ux)  ## ME
  x.obs <- xstar + ux                 ## observed x values

  mom.slope[i] <- (cov(y.obs,x.obs) - cov.uxuy)/(var(x.obs) -
    observed.sig.ux^2)
  fit.naive <- lm(y.obs ~ x.obs, x=TRUE)
  naive.results[i] <- fit.naive$coef[2]
}

```

```

fit.simexQ <- simex(fit.naive,SIMEXvariable=c("x.obs"),
observed.sig.ux, fitting.method="quadratic")
simex.resultsQ[i] <- fit.simexQ$coef[2]
fit.simexNL <- refit(fit.simexQ, "nonl")
simex.resultsNL[i] <- fit.simexNL$coef[2]

w.obs <- as.matrix(x.obs)
data.jags <- list(N=N, Nrep=1, Y=y.obs, W=w.obs,
tauu=(1/(observed.sig.ux^2)))
params <- c("alpha", "beta", "sigmaeps", "sigmax")
jags.pre <- jags.model("BayesianMEModel.txt",data.jags,
n.adapt=1000,n.chains=3)
jags.out <- coda.samples(jags.pre, params, n.iter=60000,thin=12)
post.dist[i] <- mean(as.matrix(jags.out[,2]))
}

```

4.6.3 Simulation Study 2 Code

The R code used to execute the simulations in Simulation Study 2 is given below. For each simulation, the data are generated and each corrected estimate is obtained.

```

require(simex)
require(R2jags)
N <- 25          ## number of observations (25,500)
N.sim <- 500     ## number of simulations
N.rep.sim <- c(2,20)  ## (2,20)
theta <- c(1, 5)  ## Linear Regression Parameters
sig_x <- 2.0

```

```

sig_eps <- 2
cov.uxuy <- 0

mom.slope <- simex.resultsQ <- simex.resultsQ2 <- post.dist <-
  post.dist2 <- naive.results <-
  true.results <- matrix(0, nrow=N.sim, ncol=2)

for(i in 1:N.sim){

for(h in 1:2){
N.rep <- N.rep.sim[h]
sim.sig.ux<-sqrt(12)*sqrt(N.rep) ## ME SD(1,2,sqrt(12))*sqrt(N.rep)

xstar <- rnorm(N, 0, sig_x)          ## true value of x
x_true <- cbind(rep(1,N), xstar)     ## design matrix of model
eps <- rnorm(N, 0, sig_eps)
ystar <- x_true%*%theta + eps        ## true value of y
y.obs <- c(ystar)                   ## observed y values

W.matrix <- matrix(0, nrow=N, ncol=N.rep)
mn.u <- var.u <- var.mn <- numeric(0)
for(j in 1:N){
ux <- rnorm(N.rep, 0, sim.sig.ux)   ## ME
var.u[j] <- var(ux)
mn.u[j] <- mean(ux)
var.mn[j] <- var(ux)/N.rep

```

```

W.matrix[j,] <- xstar[j] + ux
}

W.obs <- xstar + mn.u
observed.sig2.ux <- mean(var.mn)
observed.sig.ux <- sqrt(observed.sig2.ux)

mom.slope[i,h] <- (cov(y.obs,W.obs) - cov.uxuy)/(var(W.obs) -
                  observed.sig2.ux) ## mom slope estimate
fit.true <- lm(ystar ~ xstar)
true.results[i,h] <- fit.true$coef[2]
fit.naive <- lm(y.obs ~ W.obs, x=TRUE)
naive.results[i,h] <- fit.naive$coef[2]

fit.simexQ <- simex(fit.naive,SIMEXvariable=c("W.obs"),
                  observed.sig.ux, fitting.method="quadratic")
simex.resultsQ[i,h] <- fit.simexQ$coef[2]
fit.simexNL <- refit(fit.simexQ, "nonl")
simex.resultsNL[i,h] <- fit.simexNL$coef[2]

## Use plug-in estimate of sigma_u ##
W.obs3 <- as.matrix(W.obs)
data.jags2 <- list(N=N, Nrep=1, Y=y.obs, W=W.obs3,
                  tauu=(1/observed.sig2.ux))
params2 <- c("alpha", "beta", "sigmaeps", "sigmax")
jags.pre2 <- jags.model("BayesianMEModel.txt",

```

```
      data.jags2,n.adapt=1000,n.chains=3)
jags.out2 <- coda.samples(jags.pre2, params2, n.iter=60000, thin=12)
post.dist2[i,h] <- mean(as.matrix(jags.out2[,2]))
} }
```

References

- Bartlett, J. W., Keogh, R. H., 2016. Bayesian correction for covariate measurement error: a frequentist evaluation and comparison with regression calibration. arXiv preprint arXiv:1603.06284.
- Berkson, J., 1950. Are there two regressions? *Journal of the American Statistical Association* 45 (250), 164–180.
- Buonaccorsi, J. P., 2010. *Measurement error: models, methods, and applications*. CRC Press.
- Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M., 2006. *Measurement error in nonlinear models: a modern perspective*. CRC press.
- Cook, J. R., Stefanski, L. A., 1994. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* 89 (428), 1314–1328.
- Fuller, W. A., 2009. *Measurement error models*. Vol. 305. John Wiley & Sons.
- Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (6), 721–741.
- Gilks, W. R., Richardson, S., Spiegelhalter, D. J., 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- Kampstra, P., 2008. Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of Statistical Software, Code Snippets* 28 (1), 1–9.
URL <http://www.jstatsoft.org/v28/c01/>
- Lederer, W., Kchenhoff, H., 2013. simex: SIMEX- and MCSIMEX-Algorithm for measurement error models. R package version 1.5.
URL <http://CRAN.R-project.org/package=simex>
- Muff, S., Riebler, A., Held, L., Rue, H., Saner, P., 2015. Bayesian analysis of measurement error models using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 64 (2), 231–252.
- Plummer, M., 2003. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In: *Proceedings of the 3rd international workshop on distributed statistical computing*. Vol. 124. Technische Universit at Wien Wien, Austria, p. 125.

R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

URL <http://www.R-project.org/>

Stefanski, L. A., 2000. Measurement error models. *Journal of the American Statistical Association* 95 (452), 1353–1358.

Su, Y.-S., Yajima, M., 2016. R2jags: A Package for Running jags from R. R package version 0.04-03.

URL <http://CRAN.R-project.org/package=R2jags>

CHAPTER FIVE

IMPACTS OF MEASUREMENT ERROR IN AN AREA UNDER THE CURVE
STUDY

Contribution of Authors and Co-Authors

Author: Jennifer L. Weeding

Contributions: Responsible for the majority of the writing.

Co-Author: Dr. Mark C. Greenwood

Contributions: Provided feedback on statistical analysis and drafts of the manuscript.

Co-Author: Dr. Shannon J.J. Moreaux

Contributions: Provided application expertise.

Co-Author: Jyme L. Nichols

Contributions: Data Collection.

Manuscript Information Page

Jennifer L. Weeding, Mark C. Greenwood, Shannon J.J. Moreaux, and Jyme L. Nichols

Journal of Animal Science

Status of Manuscript:

Prepared for submission to a peer-reviewed journal

Officially submitted to a peer-review journal

Accepted by a peer-reviewed journal

Published in a peer-reviewed journal

Abstract

When researchers analyze variables that are summaries of other information (such as means or, here, areas under curves (AUCs)), those variables contain measurement error that should be accounted for. This analysis demonstrates what is needed to account for measurement errors, one method to account for measurement errors, and illustrates the effects of ignoring measurement errors when they are present. Serum insulin and plasma glucose levels were collected every 30 min for 6 h beginning at 0800 on the day of blood collection from eight 11- to 16-yr-old Quarter Horses. The linear relationship between the peak insulin value and the glucose AUC was explored using a statistical model that corrects for measurement error. Two separate Bayesian analyses are presented, one where the measurement errors are accounted for and one where the measurement errors are ignored, resulting in a measurement error corrected estimate of the slope coefficient (β_1) that was 17% larger than the estimate obtained when the measurement errors are ignored. This modest correction occurred when the measurement errors in the AUCs were quite small.

5.1 Introduction

It is common practice for researchers to collect measurements on the same subject over time, resulting in longitudinal data and multiple responses. Important aspects of these data can often be captured through the use of a summary measure, such as the mean, minimum, maximum, rate of linear change over time (slope coefficient), or the area under the curve (AUC). Combining observations into a single value serves three purposes, it simplifies the statistical models required, it can produce more precise “data” to analyze (one observation versus a mean), and it can allow the analysis to focus on the characteristic of interest.

The use of summary measures as the objects of statistical analyses dates at least as far back as 1938 when John Wishart analyzed regression coefficients of weight gain in pigs (Wishart, 1938). Summary measures are estimated from observed data and contain sampling variability since different samples tend to produce

different estimates. *Measurement error*, sometimes referred to as *errors-in-variables* (Stefanski, 2000), occurs when a variable of interest cannot be observed exactly, but is instead observed with error, and arises in a variety of circumstances including anytime summary measures are used. The consequences of ignoring measurement error in analyses have been well documented and include bias in parameter estimation and a loss of power (Carroll et al., 2006). It is important to note that by measurement error, we are not referring to the unexplainable random variation represented by the error term in most statistical models. We will attempt to clarify these differences and their relevance to Animal Science applications below.

The total concentration of a substance present in the blood over time is often of interest to researchers studying both animals and humans. The AUC is a summary measure that captures the total magnitude of a substance over a specified time interval and has a long history in pharmacology (Armitage et al., 2008). However, it has become increasingly popular in other areas such as fisheries (Parken et al., 2003) and animal sciences (Moreaux et al., 2011). AUCs are estimated from measurements taken over time, resulting in measurement error which should be accounted for when using the results in subsequent analyses in order to obtain more accurate parameter estimates.

An equine-based AUC data set is used to illustrate the impacts of measurement error in a simple linear regression (SLR) framework, where the explanatory variable in the model is an AUC. A method to estimate the variability associated with the AUC summary measure, presented in Chapter 3, is implemented here. The estimated variance of the AUC serves as an estimate of the measurement error variance, which is needed to correct for measurement errors. A brief overview of Bayesian methods is presented in Section 5.2, along with a Bayesian measurement error model that corrects for measurement errors in the explanatory variable in SLR.

The importance of accounting for measurement errors when present can be seen in Section 5.3, where results from a naive analysis (which ignores the measurement error) are compared to results obtained from a Bayesian measurement error model analysis. The Appendix contains all the R code used for the analysis, allowing other researchers access to these methods and making the methods used reproducible. The data set used in this analysis (Weeding et al., 2016) is available through MSU Scholarworks (<http://doi.org/10.15788/M2QP4R>).

5.2 Methods

5.2.1 Equine Data

All procedures used to generate the data set were approved by Montana State University's Animal Care and Use Committee. Data from a study of eight 11- to 16-yr-old Quarter Horses (Peterson et al., 2009) are used here to examine the linear relationship between peak insulin value and glucose AUC (calculated via the linear trapezoidal method). Blood samples were collected every 30 minutes over a 6 hour sampling period beginning at 0800 on June 13, 2008. Two glucose measurements were obtained from each blood sample (pseudo-replicates) at each sampling time, and the mean of the two values is used in subsequent analyses. Glucose analysis was determined in duplicate using a spectrophotometric method based on glucose hexokinase (Glucose hexokinase kit; Sigma Diagnostics, St. Louis, MO). Serum insulin was determined in duplicate by use of a commercially available radioimmunoassay (Insulin kit; Coat-a-Count Diagnostics, Los Angeles, CA) validated for horse blood (Freestone et al., 1992). The observed glucose values for each of the eight horses are displayed in Figure 5.1.

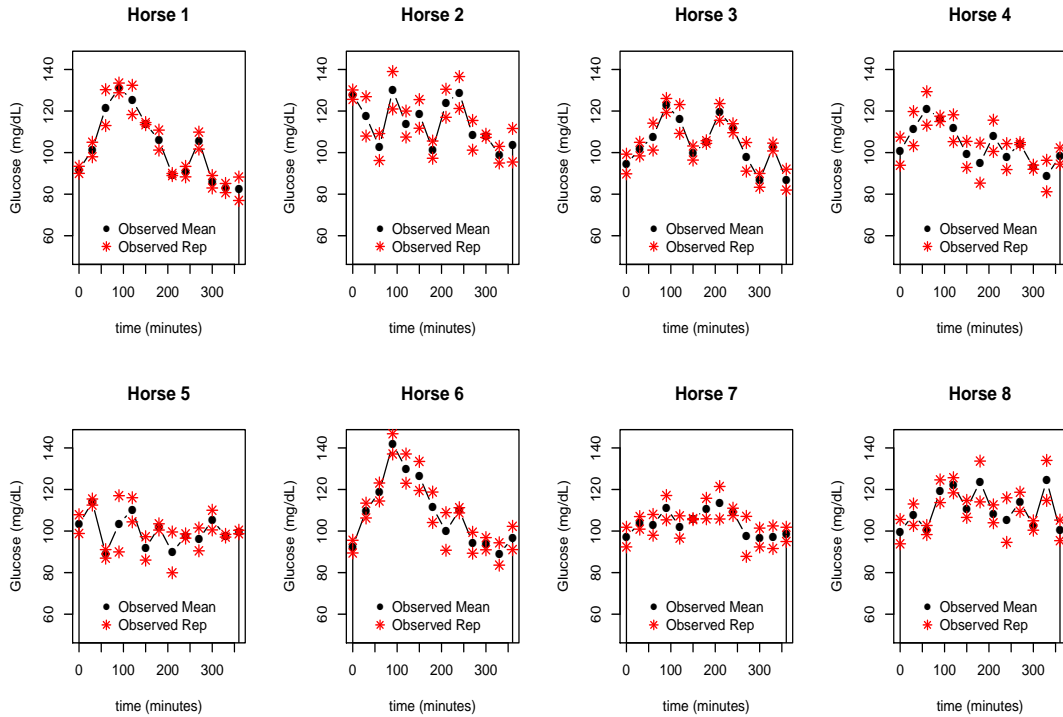


Figure 5.1: Observed glucose curves for the eight horses used in the exploratory study. The mean value, along with two replicate values, were obtained at each time point. Glucose was measured every 30 minutes over a 360-minute sampling interval.

5.2.2 Statistical Analyses

5.2.2.1 Area Under the Curve The AUC is theoretically defined as the definite integral of the true underlying function of the response, $f(t)$, over the time interval (a, b) , where points a and b are the time points associated with the initial measurement and final measurement, respectively. The true underlying function is often unknown, therefore the AUC must be estimated from observations taken between times a and b . The linear trapezoidal method is one of many AUC estimation methods available and performs the same or better in situations with a small number of data points (Allison et al., 1995). This method estimates the AUC by simply summing the area of all

trapezoids formed using responses over consecutive time points, and is often used due to its simplicity (Yuan, 1993). For example, if (z_i, t_i) represent the responses and corresponding time points for an individual and $i = 1, 2, \dots, n$ indexes the responses collected, the formula for estimating the AUC using the linear trapezoidal method is:

$$\widehat{AUC} = \frac{(z_1 + z_2) * (t_2 - t_1)}{2} + \frac{(z_2 + z_3) * (t_3 - t_2)}{2} + \dots + \frac{(z_{n-1} + z_n) * (t_n - t_{n-1})}{2} \quad (5.1)$$

Frequently, replicate observations are obtained at each time point, resulting in observing z_{ij} data, where $j = 1, 2, \dots, m_i$ for each i . These replicates may come from new independent observations of the process (for example, a new blood draw) or, more commonly, from measuring another portion of the initial blood draw (a pseudo-replicate). In either scenario, the z_i 's in Equation 5.1 are replaced with \bar{z}_i 's, where \bar{z}_i represents the mean of the m_i observations at time i .

AUCs that are estimated contain measurement error, since the true value of the AUC is unknown and if we could take the same measurements again we would likely get different results. The variability we will estimate in the AUC quantifies this measurement error variability. An estimate of the measurement error variance is needed in order to account for measurement errors. Although estimating the AUC using the linear trapezoidal method results in a simple formula, one major limitation of it is that in most situations there is not a straight-forward method to calculate the variance associated with it (Hilborn et al., 1999).

To develop an estimate of $Var(\widehat{AUC})$ note that \widehat{AUC} is a linear combination of the responses (z_i 's), as can be seen by rewriting Equation 5.1 as

$$\widehat{AUC} = \frac{t_2 - t_1}{2} * z_1 + \frac{t_3 - t_1}{2} * z_2 + \frac{t_4 - t_2}{2} * z_3 + \dots + \frac{t_n - t_{n-2}}{2} * z_{n-1} + \frac{t_n - t_{n-1}}{2} * z_n. \quad (5.2)$$

If we denote the multiplier in front of each z_i as c_i , the formula for the variance of the estimated AUC from rules for variances of linear combinations of random variables is:

$$Var(\widehat{AUC}) = Var\left(\sum_{i=1}^n c_i z_i\right) = \sum_{i=1}^n c_i^2 Var(z_i) + 2 \sum_{1 \leq i < i' \leq n} c_i c_{i'} Cov(z_i, z_{i'}). \quad (5.3)$$

In the analyses here, we will use the Bayesian models presented in Chapter 3 to model the means and variances-covariances of observations over time to obtain estimates of the true underlying mean values at each time point (z_i), which are then used in conjunction with the linear trapezoidal method to compute an estimate of the AUC. Additionally, $Var(z_i)$ and $Cov(z_i, z_{i'})$ are estimated from the same Bayesian model (discussed in Section 5.3.1) and used in Equation 5.3 to obtain an estimate of the measurement error variance.

5.2.2.2 Measurement Error Models A class of statistical models that take into account measurement error and correct parameter estimates for the bias induced from variables measured with error can be described as measurement error models. The type of measurement error (Berkson or classical), the form of the error structure, and knowledge of the size of the measurement error (its variance) must all be known in order to use a measurement error version of more conventional statistical models.

Measurement error in a continuous random variable can be classified as either Berkson or classical. Berkson measurement error (Berkson, 1950) occurs when the researcher is trying to achieve a target (observed) value (W_i), but instead achieves a true value (X_i) that is unknown. Classical measurement error occurs when the true value of the variable (X_i) is measured with error, resulting in an observed value (W_i). The fundamental difference between the two types of measurement error can be seen

in the distribution assumed for the errors where the Berkson measurement error model assumes a distribution for the true values given the observed values ($X_i|W_i$), whereas the classical measurement error model assumes a distribution for the observed values given the true values ($W_i|X_i$).

Although measurement errors are generally assumed to be additive (Buonaccorsi, 2010), other error structures do exist (linear, multiplicative, etc.) and could be required in some applications. The classical additive measurement error model for the mis-measured random variable, W_i is

$$W_i = X_i + U_i, \quad (5.4)$$

where U_i represents the measurement errors, which are assumed to have a mean of 0 ($E(U_i|x_i) = 0$) and variance σ_u^2 ($Var(U_i|x_i) = \sigma_u^2$). The additive error structure implies the mis-measured variable W is an unbiased estimator of X , since the measurement errors are centered at 0 resulting in $E(W_i|x_i) = x_i$. The additive Berkson measurement error model is obtained by switching W_i and X_i in the above model.

The measurement error variance (σ_u^2) must be assumed known or estimated from replicate (repeated observations on the same subject) data in order to fit a measurement error model and can be allowed to vary across observations if non-constant measurement error variance is present. When summary measures are used, such as the AUC, the measurement error variance is estimated using the variance associated with the summary measure.

5.2.2.3 Measurement Error in Simple Linear Regression The simple linear regression (SLR) model is implemented in situations where a linear relationship

between two quantitative variables is of interest. If we let Y and X denote the response and explanatory variables respectively, the simple linear regression model is

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad (5.5)$$

where the random, unexplained residual errors, ϵ_i , are assumed to be independent and normally distributed, centered at 0, with variance σ_ϵ^2 (denoted $\epsilon_i \sim N(0, \sigma_\epsilon^2)$, where ‘ \sim ’ means ‘distributed as’ and $N(a, b)$ generically denotes a Normal distribution with mean a and variance b). Measurement error may be present in the response variable, explanatory variable, or both, but here we focus on measurement error in the explanatory variable where (Y_i, W_i) represent the observed (un-contaminated) response and mis-measured explanatory variable containing classical additive measurement error. Combining the classical additive error model in Equation 5.4 with the SLR model in Equation 5.5 results in

$$Y_i = \beta_0 + \beta_1(X_i + U_i) + \epsilon_i, \quad (5.6)$$

where the U_i are independent of the ϵ_i 's.

Biased parameter estimates are a known result of ignoring measurement errors in analyses. In the simple linear regression setting, the naive (ordinary least squares regression of Y_i on W_i) estimator of the slope ($\beta_{1,naive}$) is biased towards 0 (called attenuation) because

$$\beta_{1,naive} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_u^2} \beta_1 = k \beta_1, \quad (5.7)$$

where $k = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_u^2}$ (referred to as the reliability ratio (Fuller, 2009), β_1 is the true slope, σ_X^2 is the variance of the true random variable X , and σ_u^2 is the measurement error variance associated with the mis-measured explanatory variable. Equation 5.7 shows

that larger measurement error variance associated with the mis-measured explanatory variable (σ_u^2) results in $\beta_{1,naive}$ being a more biased estimator of the true value, β_1 .

Common methods available to correct for measurement errors include the method of moments correction, SIMEX (Cook and Stefanski, 1994), regression calibration, and Bayesian methods; more details can be found in standard measurement error texts (Buonaccorsi (2010), Carroll et al. (2006), and Gilks et al. (1996)). The focus here will be on Bayesian methods, as they provide better performance (Chapter 4 and Weeding and Greenwood (2014)), especially when sample sizes are not large.

5.2.2.4 Overview of Bayesian Methods Bayesian modeling is becoming a common tool in many research areas, including Animal Sciences. Cernicchiaro et al. (2013) recently used a hierarchical Bayesian model when studying average daily gain (ADG) in cohorts of commercial feedlot cattle where non-constant variance was present. Many animal-science researchers, including Cai et al. (2012), are utilizing Bayesian models for studies involving genomic data. A brief overview of Bayesian inference is presented here, where more details can be found in standard Bayesian texts such as Gelman et al. (2014), Gelman and Hill (2006), Jackman (2009), and Lunn et al. (2012).

Bayesian models are defined by specifying a likelihood model for the data and prior distributions for each unknown parameter. The likelihood model is a probability distribution for the observed data (y) given the parameters (θ) and is denoted $p(y|\theta)$. Prior distributions for each parameter must be specified, which express prior beliefs about the parameters, and the set of prior distributions for all parameters is denoted $p(\theta)$. The posterior distribution of the parameters given the data, denoted $p(\theta|y)$, is obtained through Bayes' Rule by multiplying the likelihood by the prior distributions, and is the distribution through which Bayesian inferences are made. The posterior

distribution is computed as

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)}. \quad (5.8)$$

$p(y)$ represents the probability distribution of the data and does not depend on θ so is often omitted, resulting in an equivalent form of the posterior distribution:

$$p(\theta|y) \propto p(y|\theta)p(\theta). \quad (5.9)$$

Inferences for the model parameters (θ) are available directly from the posterior distribution.

Prior distributions for each parameter can be informative or non-informative. Informative prior distributions can incorporate prior information available about the parameter, whereas non-informative prior distributions attempt to express a lack of prior information about the parameter. Choosing an appropriate informative prior distribution can require extra thought to ensure the prior information is portrayed correctly. However, non-informative prior distributions may also require careful consideration, as some parameters have restrictions on the values they can take on (for example, variance parameters must be greater than zero).

Gibbs sampling (Geman and Geman, 1984) is a Markov chain Monte Carlo (MCMC) sampling algorithm which is often used to obtain samples from the joint posterior distribution of the parameters. The samples obtained from a Gibbs sampler are ordered and referred to as a *chain*. Early draws may be unrepresentative of the posterior distribution, but, eventually, the draws will converge to the posterior distribution (Gelman et al., 2014). The samples obtained before the chain has converged are referred to as the *burn-in period* and are discarded, leaving only the

samples from the converged chain for posterior inference. There are many available methods to assess convergence (Cowles and Carlin, 1996), two of which are examining sample path plots from multiple independent chains (each with different random starting values) and using potential scale reduction factors (Gelman and Rubin, 1992). To ensure convergence, multiple convergence assessment tools should be used.

5.2.2.5 Bayesian Measurement Error Model: SLR Setting The SLR model can be fit using a Bayesian framework by specifying a likelihood model for the data (Y_i) and prior distributions for the unknown parameters $\theta = (\beta_0, \beta_1, \sigma_\epsilon^2)$. The likelihood model, corresponding to Equation 5.5, is specified as

$$Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma_\epsilon^2) \quad (5.10)$$

where the corresponding prior distributions for the unknown parameters are generally non-informative.

The SLR model that can correct for measurement error in the explanatory variable can be fit using a Bayesian framework by specifying a likelihood model for the data (Y_i) and prior distributions for the unknown parameters $\theta = (\beta_0, \beta_1, \sigma_\epsilon^2, \sigma_X^2)$. The likelihood model, corresponding to Equation 5.6, is specified as

$$\begin{aligned} Y_i &\sim N(\beta_0 + \beta_1 X_i, \sigma_\epsilon^2), \\ W_i | X_i &\sim N(X_i, \sigma_u^2), \text{ and} \\ X_i &\sim N(0, \sigma_X^2), \end{aligned} \quad (5.11)$$

where σ_u^2 is assumed known and the corresponding prior distributions for the unknown parameters are generally non-informative. A measurement error model that assumes non-constant measurement error variance across observations can be used by replacing

σ_u^2 with $\sigma_{u,i}^2$. If replicate data are available (for example, observing W_{ij} replicates for each X_i , where $j = 1, 2, \dots, m_i$), that information can be used to generate a prior distribution for σ_u^2 .

5.2.2.6 Application of Methods A SLR model with peak insulin as the response variable and glucose AUC as the explanatory variable was considered for the data displayed in Figure 5.2 and provided in Table 5.1. The Bayesian SLR measurement error model was applied and, for comparison, the naive Bayesian SLR model (ignoring measurement error) was also considered. Non-informative prior distributions were used for the unknown parameters in both models ($\beta_0 \sim N(0, 10^6)$, $\beta_1 \sim N(0, 10^6)$, $\sigma_\epsilon \sim Uniform(0, 5000)$, and $\sigma_X \sim Uniform(0, 5000)$).

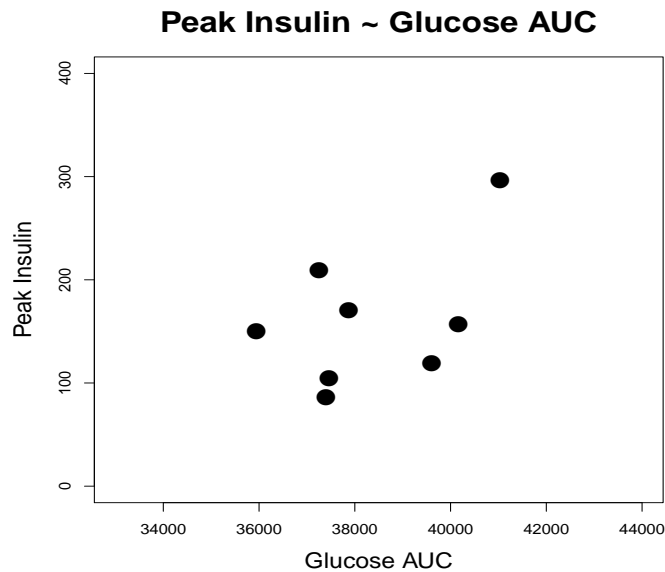


Figure 5.2: Scatterplot showing the relationship between peak insulin and glucose AUC for the eight horses used in the exploratory study.

Statistical analyses were conducted in R (R Core Team, 2016) using the Gibbs sampler JAGS (Plummer, 2003), and the R-package R2jags (Su and Yajima, 2016).

Three independent chains were used, each with random starting points. Each chain was run for 20,000 iterations and for computational efficiency every 4th iteration was saved. Each final chain consisted of 5,000 iterations of which the first 1,000 iterations were discarded as *burn-in*. Convergence was assessed using sample path plots and potential scale reduction factors (a threshold of $\hat{R} < 0.05$ was used to determine convergence). All R code is provided in the Appendix.

5.3 Results

5.3.1 Glucose Parameter Estimation

Glucose AUC for each horse was computed using a Bayesian first-order autoregressive (AR(1)) model (Lunn et al., 2012), where the two replicate values at each time point were used in a non-constant variance, Bayesian AR(1) model. The non-constant variance AR(1) model is a statistical model that accounts for correlated observations by assuming the current value is a linear combination of the most recent value plus a random error term, and is specified by

$$\begin{aligned} z_{ij} &= \mu_i + \epsilon_{ij}, \\ \mu_i &= c + \phi\mu_{i-1} + \epsilon_\mu, \end{aligned} \tag{5.12}$$

where $\phi < |1|$, $\epsilon_{ij} \sim N(0, \sigma_{\epsilon,i}^2)$, μ_i represent the true underlying mean at time i , $\epsilon_\mu \sim N(0, \sigma_\mu^2)$, c is an intercept, and $i = 1, 2, \dots, n$. The AR(1) model assumes that the responses are taken at equal intervals across time and that observations one time point apart have a correlation of ϕ , two apart have a correlation of ϕ^2 , and k apart have a correlation of ϕ^k . This model can be used to estimate the covariance in Equation 5.3 using the estimated ϕ and $\sigma_{\epsilon,i}$'s, where two observations obtained k

sampling times apart have an estimated covariance of

$$\widehat{Cov}(z_i, z_{i'}) = \hat{\phi}^k * \hat{\sigma}_{\epsilon,i} * \hat{\sigma}_{\epsilon,i'}. \quad (5.13)$$

An empirical Bayes approach (Gilks et al., 1996) was used since only two replicate glucose values were available at each time point, where the observed data were used to estimate the variances at each time point, which then are used as the priors for each $\sigma_{\epsilon,i}^2$ in the non-constant variance model. If more replicate values were available at each time point, the AR(1) model could be modified to estimate the variance at each time point. However, with only two replicates available at each time point there is not enough information available to accurately estimate these values. Estimates of the means at each time point were computed from the corresponding posterior distributions and used in Equation 5.2. The variance of the AUC was computed similarly, where $Var(\bar{z}_i)$ was estimated from the posterior distribution of the mean at each time point and the estimate of the correlation used in $Cov(\bar{z}_i, \bar{z}_{i'})$ was estimated from its posterior distribution. See Chapter 3 for complete details concerning estimation of $Var(\widehat{AUC})$.

The estimated glucose curves for each horse resulting from the AR(1) model can be seen in Figure 5.3. The AR(1) model tracked the observed means closely for all horses except Horse 4 and Horse 7. The \widehat{AUC} and $\widehat{SD}(\widehat{AUC})$ corresponding to these two horses was computed only using the observed means and variances, whereas for all other horses they were computed using the estimates of the mean, variance, and correlation obtained from the AR(1) model. The estimated standard deviation of individual glucose \widehat{AUC} values are an estimate of the measurement error standard deviation, and are small compared to the \widehat{AUC} values themselves (Table 5.1). The

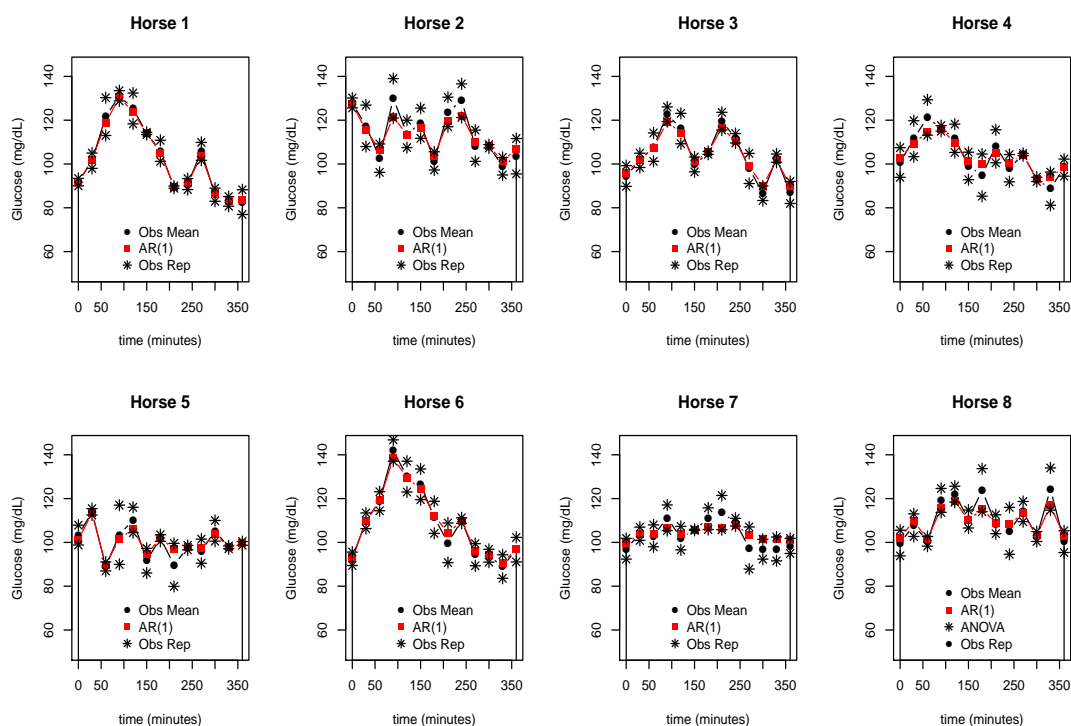


Figure 5.3: The observed curve for each horse, along with the estimated curve resulting from the AR(1) model. The replicate values at each time point are displayed to show the variability present in the replicate values across the different time points.

average estimated reliability ratio, \hat{k} , is 0.85, (Equation 5.7) indicating relatively modest measurement error variance for these data.

5.3.2 Simple Linear Regression Analyses

A Bayesian measurement error model that allows for non-constant measurement error variances across horses was implemented, where the standard deviations presented in Table 5.1 are used as estimates of $\sigma_{u,i}$ in the likelihood model (Equation 5.11). Estimates and 90% credible intervals from both the naive SLR model and the SLR measurement error model are presented in Table 5.2. The estimate of β_1 from the measurement error model is 0.0217, indicating that peak

| Horse | Glucose \widehat{AUC} | $SD(\widehat{AUC})$ | CV (%) | Peak Insulin |
|-------|-------------------------|---------------------|--------|--------------|
| 1 | 37079 | 645.3 | 1.74 | 209.1 |
| 2 | 40650 | 594.7 | 1.46 | 296.2 |
| 3 | 37812 | 618.8 | 1.64 | 170.2 |
| 4* | 37396 | 655.3 | 1.75 | 86.2 |
| 5 | 36051 | 338.1 | 0.94 | 149.8 |
| 6 | 39696 | 899.9 | 2.27 | 118.9 |
| 7* | 37461 | 550.8 | 1.47 | 104.2 |
| 8 | 39729 | 409.9 | 1.03 | 156.8 |

Table 5.1: Peak insulin (measured as $\mu IU/mL$) and estimated glucose AUC (measured as $mg/dL/360$ min) values for each horse calculated over the 360 minute sampling interval where samples were obtained every 30 minutes. Standard deviations and coefficients of variation (CV) corresponding to the glucose \widehat{AUC} for each horse are also given. * indicates the glucose \widehat{AUC} and $SD(\widehat{AUC})$ were computed using the observed means and variances in Equations 5.2 and 5.3, where the correlation between repeated observations is assumed to be 0.

insulin, on average, increases by $0.0217 \mu IU/mL$ for a one mg/dL per 360 minutes increase in glucose AUC. The estimate of β_1 from the naive model is smaller because of attenuation due to the measurement error present and is 0.0186, indicating that peak insulin, on average, increases by $0.0186 \mu IU/mL$ for a one mg/dL per 360 minutes increase in glucose AUC. The 90% credible interval for β_1 obtained from the measurement error model is wider than the interval obtained from the naive model, however the measurement error model is accounting for the extra uncertainty present in the observed explanatory variable that the naive model ignores.

In this application the measurement errors are relatively small. However, the estimate of β_1 from the measurement error model is 17% larger than the corresponding estimate from the naive model. The results presented here reaffirm the importance of accounting for measurement errors when they are present in explanatory variables, regardless of how small they may be.

| Parameter | Naive Model | | Measurement Error Model | |
|-----------|-------------|---------------------|-------------------------|--------------------|
| | Estimate | 90% CI | Estimate | 90% CI |
| β_0 | -552.244 | (-602.37, -501.531) | -666.349 | (-716.22, -617.17) |
| β_1 | 0.0186 | (-0.0145, 0.0523) | 0.0217 | (-0.0176, 0.0599) |

Table 5.2: Parameter estimates and 90% credible intervals from the naive model and the measurement error model.

5.4 Discussion

Ignoring measurement errors in analyses when they are present will cause bias in parameter estimates and may result in an important effect going undetected. Researchers often do not account for measurement errors, even when they know they are present (Buonaccorsi, 2010), as they are unaware of the potential impacts they can have on their analyses. Other reasons why researchers do not account for measurement errors when they know they are present include that there is a lack of easily available software that can handle measurement error models and the replicate data needed to estimate the measurement error variance are not available, or are sometimes discarded unknowingly.

The AUC summary measure is generally estimated from observed data, and therefore contains classical additive measurement error. An estimate of the variance associated with the \widehat{AUC} is needed to account for the measurement error, and was obtained through the use of a Bayesian AR(1) model and the linear trapezoidal method. The estimated variance of the two replicate values at each time point was used in the Bayesian model, however the uncertainty of the estimate based on such a small number of observations can be high. The Bayesian model can be adjusted to estimate the variance at each time point by using appropriate prior distributions when more replicate values are available at each time point.

Measurement errors should be taken into account when they are present in order to obtain the least biased results. Bayesian measurement error correction methods have shown better performance than other common correction methods when measurement errors are present in the explanatory variable in the SLR setting (Chapter 4 and Weeding and Greenwood (2014)). Bayesian models can be computationally challenging to set up and require careful thought prior to the analysis in order to choose the appropriate likelihood and prior distributions, however the results are easily interpretable, as inferences are made directly from the posterior distributions.

The equine data provided a setting in which to estimate the measurement error variance associated with the AUC summary measure and then explore the impacts that measurement errors have on subsequent analyses. The measurement errors associated with the glucose \widehat{AUC} values were small relative to the magnitude of the glucose \widehat{AUC} values, however the measurement error corrected estimate of the slope was 17% larger than the estimate obtained ignoring the measurement errors. In applications with larger relative magnitudes of measurement error, the corrections using these methods would be more dramatic. For example, if the observed measurement errors in this application had been larger and resulted in an estimated reliability ratio of, say, $\hat{k} = 0.5$, the corresponding estimate of β_1 would be 0.0486, which is more than 2.6 times larger than the naive estimate. All R code used for the application presented here can be found in the Appendix and can serve as a template for other simple linear regression measurement error applications.

5.5 Appendix

All R code used for the analyses in this study is presented here, along with the Bayesian model specification for the models presented in Section 5.2.2.5.

5.5.1 Bayesian Model Specifications

The JAGS specification of the models presented in Sections 5.2.2.1 and 5.2.2.5 are given below. In JAGS, a Normal distribution is specified using the precision ($\frac{1}{\sigma^2}$) instead of the variance. In the sections below, $\sim dnorm(a, b)$ corresponds to a variable being distributed Normally with mean a and variance $\frac{1}{b}$. Also, $\sim dunif(c, d)$ corresponds to a variable being Uniformly distributed with minimum value c and maximum value d . All models can be run in R, using the *R2jags* package. The # symbol in R denotes a comment. The models are saved in separate text documents (named ‘anova.txt’, ‘SLR.txt’ and ‘MExOnly.txt’) and are called from R when needed.

5.5.1.1 Bayesian AR(1) Model A Bayesian AR(1) model was discussed in Section 5.2.2.1 as a method to obtain estimates of the AUC and $SD(\widehat{AUC})$ and is given here. The j index refers to the number of time points (13 for this analysis) and the i index refers to the number of replicate values obtained at each time point (N.rep = 2 for this analysis). This model assumes the variances at each time point (σ_j^2) are known, since the line specifying the prior distribution for σ_j is commented out. If more replicate data are available at each time point, the model can estimate σ_j by using an appropriate prior distribution (remove the # symbol from the sig[j] line below). This model is saved in a text file named ‘AR1.txt’.

```
model{
  for(j in 1:13){
```

```

for(i in 1:N.rep){
  Y[i,j] ~ dnorm(alpha[j], tau[j])
} }

mu[1] <- b0+ phi*mu0
alpha[1] ~ dnorm(mu[1], tau.a)

for(j in 2:13){
  mu[j] <- b0 + phi*mu[j-1]
  alpha[j] ~ dnorm(mu[j], tau.a)
}

for(j in 1:13){
  tau[j] <- pow(sig[j], -2)
# sig[j] ~ dunif(0,1000)
}

mu0 ~ dnorm(0, 0.001)
b0 ~ dnorm(0.0, 1.0E-6)
phi ~ dunif(-1,1)
sigma.a ~ dunif(0,50)
tau.a <- pow(sigma.a, -2)
}

```

5.5.1.2 Bayesian SLR Naive Model The Bayesian SLR model corresponding to Equation 5.10 is given here. N refers to the total number of subjects in the analysis.

This model does **not** account for measurement errors and was referred to as the *naive* model. This model is saved in a text file named ‘SLR.txt’.

```

model{
  for(i in 1:N){
    Y[i]~dnorm(meanY[i],taueps)
    meanY[i]<-alpha+beta*X[i]
  }
  taueps <- pow(sigmaeps, -2)
  sigmaeps ~ dunif(0, 5000)

  alpha~dnorm(0, .000001)
  beta~dnorm(0, .000001)
}

```

5.5.1.3 Bayesian SLR Measurement Error Model The Bayesian SLR *measurement error* model corresponding to Equation 5.11 is given here, where we account for non-constant measurement error variances. N refers to the total number of subjects in the analysis and N_{rep} refers to the total number of replicate values available for subject i ($N_{\text{rep}} = 1$ for this analysis). This model currently assumes the measurement error variances are known, since the line specifying the prior distribution for σ_u is commented out. If replicate values of the mis-measured variable are available, the model can estimate σ_u by using an appropriate prior distribution (remove the $\#$ symbol from the ‘sig.u’ line below). This model is saved in a text file named ‘MeasurementErrorSLR.txt’.

```

model{

```

```

for(i in 1:N){
  Y[i]~dnorm(meanY[i],taueps)
  meanY[i]<-alpha+beta*X[i]
for(j in 1:Nrep){
  W[i,j]~dnorm(X[i],tauu[i])
  }
  X[i]~dnorm(0,taux)
  tauu[i] <- pow(sig.u[i], -2) # must provide estimates of sig.u!
# sig.u[i] ~ dunif(0, 5000)
  }
taueps <- pow(sigmaeps, -2)
sigmaeps ~ dunif(0, 5000)

taux <- pow(sigmax, -2)
sigmax ~ dunif(0, 5000)

alpha~dnorm(0, .000001)
beta~dnorm(0, .000001)
}

```

5.5.2 Analysis Code

All of the code used for the analysis in Section 5.3 is presented here. You will need to save the JAGS models in Section 5.5.1 in the directory you are working in before running the code below. The equine glucose data set (Weeding et al., 2016) has 3 columns: column 1 represents the identification number for each horse, column 2 represent the time of data collection (in minutes), and column 3 represents the

observed glucose values. The data set is available upon request or through MSU Scholarworks (<http://doi.org/10.15788/M2QP4R>).

```
install.packages("R2jags")
require(R2jags)
equine.data <- read.csv("EquineGlucoseData.csv", header=T)
```

5.5.2.1 Obtaining AUCs and $sd(AUC)$ The code below is used to obtain the estimated AUC for each individual horse. The data to use in the model below should be organized where each column represents an individual time point and each row represents a replicate value at at given time point.

```
Horse1 <- equine.data[equine.data$HorseID==1,]
Nrep <- 2
Y.data <- matrix(0, nrow=Nrep, ncol=13)
for(i in 1:13){
  Y.data[,i] <- Horse1[Horse1$Time==i,3]
}
sd.data <- apply(Y.data,2,sd) # obtaining SDs to use in the model
data.AR1 <- list(Y=Y.data, N.rep=2, sig=sd.data)
params.AR1 <- c("alpha","b0","mu0","phi","sigma.a")
jags.AR1 <- jags.model("AR1.txt",data.AR1,n.adapt=1000,n.chains=3)
jags.outAR1 <- coda.samples(jags.AR1,params.AR1,n.iter=50000,thin=5)

alpha1 <- as.matrix(jags.outAR1)[,1]
alpha2 <- as.matrix(jags.outAR1)[,2]
...
```

```

alpha13 <- as.matrix(jags.outAR1)[,13]

posterior.means <- data.frame(alpha1,alpha2,alpha3,alpha4,alpha5,
    alpha6,alpha7,alpha8,alpha9,alpha10,alpha11,alpha12,alpha13)
apply(posterior.means, 2, mean)  ## these are the estimates of
    ## the mean at each time point for Horse 1
apply(posterior.means, 2, sd)   ## these are the estimates of
    ## the standard deviation at each time point for Horse 1
phi <- as.matrix(jags.outAR1)[,16]
mean(phi)  ## this is the estimate of the correlation between
    ## consecutive observations

```

The mean of ϕ is the estimate of the correlation (ϕ^k) used in $Cov(y_i, y_{i'})$ for each horse. The code above is repeated for each horse, each time obtaining an estimate of the AUC and $SD(\widehat{AUC})$. At the end, all estimated AUCs and $SD(\widehat{AUC})$ are put in a vector.

```

glucose.AUC <- c(37079,40650,37812,37396,36051,39696,37461,39729)
glucose.AUC.sd <- c(645.3,594.7,618.8,655.3,338.1,899.9,550.8,409.9)

```

5.5.2.2 Naive SLR Analysis The code below is for the naive SLR model (ignoring the measurement errors). The explanatory variable should be centered before using either SLR model below.

```

glucose.AUC.centered <- glucose.AUC - mean(glucose.AUC)
pk.insulin <- c(209.1,296.2,170.2,86.2,149.8,118.9,104.2,156.8)
N <- length(pk.insulin)

```


References

- Allison, D. B., Paultre, F., Maggio, C., Mezzitis, N., Pi-Sunyer, F. X., 1995. The use of areas under curves in diabetes research. *Diabetes care* 18 (2), 245–250.
- Armitage, P., Berry, G., Matthews, J. N. S., 2008. *Statistical methods in medical research*. John Wiley & Sons.
- Berkson, J., 1950. Are there two regressions? *Journal of the American Statistical Association* 45 (250), 164–180.
- Buonaccorsi, J. P., 2010. *Measurement error: models, methods, and applications*. CRC Press.
- Cai, W., Kaiser, M., Dekkers, J., 2012. Bayesian analysis of the effect of selection for residual feed intake on growth and feed intake curves in yorkshire swine. *Journal of animal science* 90 (1), 127–141.
- Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M., 2006. *Measurement error in nonlinear models: a modern perspective*. CRC press.
- Cernicchiaro, N., Renter, D., Xiang, S., White, B., Bello, N., 2013. Hierarchical bayesian modeling of heterogeneous variances in average daily weight gain of commercial feedlot cattle. *Journal of animal science* 91 (6), 2910–2919.
- Cook, J. R., Stefanski, L. A., 1994. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* 89 (428), 1314–1328.
- Cowles, M. K., Carlin, B. P., 1996. Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* 91 (434), 883–904.
- Freestone, J., Shoemaker, K., Bessin, R., Wolfsheimer, K., 1992. Insulin and glucose response following oral glucose administration in well-conditioned ponies. *Equine Veterinary Journal* 24 (S11), 13–16.
- Fuller, W. A., 2009. *Measurement error models*. Vol. 305. John Wiley & Sons.
- Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2014. *Bayesian data analysis*. Vol. 2. Taylor & Francis.
- Gelman, A., Hill, J., 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Gelman, A., Rubin, D. B., 1992. Inference from iterative simulation using multiple sequences. *Statistical science*, 457–472.

- Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (6), 721–741.
- Gilks, W. R., Richardson, S., Spiegelhalter, D. J., 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- Hilborn, R., Bue, B. G., Sharr, S., 1999. Estimating spawning escapements from periodic counts: a comparison of methods. *Canadian Journal of Fisheries and Aquatic Sciences* 56 (5), 888–896.
- Jackman, S., 2009. *Bayesian analysis for the social sciences*. Vol. 846. John Wiley & Sons.
- Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2012. *The BUGS book: A practical introduction to Bayesian analysis*. CRC press.
- Moreaux, S. J. J., Nichols, J. L., Bowman, J. G., Hatfield, P. G., 2011. Psyllium lowers blood glucose and insulin concentrations in horses. *Journal of Equine Veterinary Science* 31 (4), 160–165.
- Parken, C. K., Bailey, R. E., Irvine, J. R., 2003. Incorporating uncertainty into area-under-the-curve and peak count salmon escapement estimation. *North American Journal of Fisheries Management* 23 (1), 78–90.
- Peterson, J., Moreaux, S., Bowman, J., Hatfield, P., Berardinelli, J., Olsen, J., 2009. Metabolic and physical effects of psyllium supplementation on quarter horses. In: *American Society of Animal Science*. p. 176.
- Plummer, M., 2003. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In: *Proceedings of the 3rd international workshop on distributed statistical computing*. Vol. 124. Technische Universit at Wien Wien, Austria, p. 125.
- R Core Team, 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL <http://www.R-project.org/>
- Stefanski, L. A., 2000. Measurement error models. *Journal of the American Statistical Association* 95 (452), 1353–1358.
- Su, Y.-S., Yajima, M., 2016. R2jags: A Package for Running jags from R. R package version 0.04-03.
URL <http://CRAN.R-project.org/package=R2jags>

- Weeding, J., Greenwood, M., Moreaux, S., Nichols, J., 2016. Equine Glucose Data [dataset]. Montana State University Scholarworks, Vienna, Austria.
URL <http://doi.org/10.15788/M2QP4R>
- Weeding, J., Greenwood, M. D., 2014. A bayesian approach to modeling measurement errors. In: JSM Proceedings, Section on Bayesian Statistical Science. American Statistical Association, Alexandria, VA, pp. 1443–1452.
- Wishart, J., 1938. Growth-rate determinations in nutrition studies with the bacon pig, and their analysis. *Biometrika* 30 (1/2), 16–28.
- Yuan, J., 1993. Estimation of variance for auc in animal studies. *Journal of pharmaceutical sciences* 82 (7), 761–763.

CHAPTER SIX

A BAYESIAN MEASUREMENT ERROR CORRECTION MODEL FOR
ERRORS IN BOTH EXPLANATORY AND RESPONSE VARIABLES

Contribution of Authors and Co-Authors

Author: Jennifer L. Weeding

Contributions: Responsible for the majority of the writing.

Co-Author: Dr. Mark C. Greenwood

Contributions: Provided feedback on statistical analysis and drafts of the manuscript.

Manuscript Information Page

Jennifer L. Weeding and Mark C. Greenwood

Journal: To Be Determined

Status of Manuscript:

Prepared for submission to a peer-reviewed journal

Officially submitted to a peer-review journal

Accepted by a peer-reviewed journal

Published in a peer-reviewed journal

Abstract

Measurement error research is commonly focused on correcting parameter estimates for bias induced from measurement errors in the explanatory variable, however the response variable may also contain measurement error. In this research, we focus on the simple linear regression setting where classical, additive measurement errors are present in both the response and explanatory variables. We compare the performance of three existing measurement error correction methods, the method of moments correction, the Simulation-Extrapolation (SIMEX) correction, and the GSIMEX correction, to a proposed Bayesian correction method, in scenarios where the measurement errors are either uncorrelated or correlated. Simulation studies show the four correction methods all estimate the slope coefficient well when the measurement error variances are small. However, the Bayesian model has better overall performance compared to the other methods when the measurement error variances increase. The correlation between the measurement errors can be estimated using the proposed Bayesian method, whereas the other methods require the value of the correlation be provided. Simulation studies compared the proposed Bayesian method using a ‘known’ value of the correlation in the model to the proposed Bayesian method where the correlation was estimated by the model, and similar results were obtained, suggesting that it is not necessary to know the correlation between the measurement errors associated with the mis-measured response and explanatory variables to account for it.

6.1 Introduction

The effects of ignoring measurement errors in analyses have been well documented and include bias in parameter estimates and a loss of power to detect effects (Carroll et al., 2006). A statistical model that accounts for measurement errors when present is called a measurement error model. The measurement error variance must be assumed known, or estimated from replicate data, in order to account for measurement errors in analyses. Common measurement error correction methods include the Method of Moments correction (see Buonaccorsi (2010)), the simulation-extrapolation (SIMEX) correction (Cook and Stefanski, 1994), the generalized SIMEX (GSIMEX) correction (Ronning and Rosemann, 2008), and Bayesian methods (see

Carroll et al. (2006) and Gilks et al. (1996)). However, each method is not appropriate in every situation and performance of the methods vary depending on the situation.

Correcting parameter estimates for the bias induced from variables measured with error is the focus of many measurement error correction methods. The simple linear regression (SLR) model was considered in Chapter 4, where classical, additive measurement error was present in the explanatory variable only. However, in many applications measurement error may be present in the response variable only, or measurement error may be present in both the response and explanatory variables. If we let D_i represent the mis-measured response variable and W_i represent the mis-measured explanatory variable, both containing classical additive measurement error, then

$$D_i = Y_i + Q_i \quad \text{and} \quad W_i = X_i + U_i, \quad (6.1)$$

where Q_i and U_i represent the measurement errors, with $Q_i|y_i \sim N(0, \sigma_q^2)$ and $U_i|x_i \sim N(0, \sigma_u^2)$. The notation here assumes the measurement error variances are constant, but can be altered to allow non-constant measurement error variance when needed by replacing σ_q^2 with $\sigma_{q,i}^2$ and σ_u^2 with $\sigma_{u,i}^2$. If measurement error is present in both variables, the measurement errors may be correlated, where $Cov(Q_i, U_i) = \sigma_{uq}$ represents the covariance between the measurement errors. In either situation, the SLR model for variables containing classical additive measurement errors is defined as

$$\begin{aligned} Y_i + Q_i &= \beta_0 + \beta_1(X_i + U_i) + \epsilon_i \\ \Rightarrow D_i &= \beta_0 + \beta_1 W_i + \epsilon_i, \end{aligned} \quad (6.2)$$

where $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. Buonaccorsi (2010) provides the following equation for the ordinary least squares estimator of β_1 when measurement error is present in both the response and explanatory variables:

$$\beta_{1,naive} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_u^2} \beta_1 + \frac{\sigma_{uq}}{\sigma_X^2 + \sigma_u^2}, \quad (6.3)$$

which shows that the the naive estimator of the slope is biased if the measurement errors are ignored.

This research focuses on the performance of common measurement error methods that correct the estimate of the slope when classical additive measurement errors are present in the SLR setting. There are four distinct cases where measurement errors can arise in the SLR setting, which include: Case 1 - measurement error is present in the explanatory variable only; Case 2 - measurement error is present in the response variable only; Case 3 - uncorrelated measurement errors present in the response and explanatory variables; and Case 4 - correlated measurement errors present in the response and explanatory variables. Case 1 was considered in Chapter 4. Equation 6.3 does not depend on σ_q^2 , therefore the naive estimator of the slope is unbiased in Case 2. Here we focus on Cases 3 and 4, where measurement error in both variables leads to bias in the naive estimator of the slope and the bias in Case 4 is different from Case 1. A Bayesian measurement error model is proposed in Section 6.2, where uncorrelated and correlated measurement errors are accounted for. Simulation studies are performed for each case in Section 6.3, where the performance of common measurement error corrections are compared. A discussion of the results and a direction for future research are provided in Section 6.4.

6.2 Methods

The method of moments measurement error correction is a formula-based correction method that is derived by equating moments specific to the model of interest (here, SLR under the assumptions of Case 4). Buonaccorsi (2010) contains moment-based formulas to correct for measurement error for many different modeling situations, including SLR. The method of moments measurement error corrected estimator of the slope for Case 4 in the SLR setting is

$$\hat{\beta}_{1,MOM} = \frac{\hat{\sigma}_{WD} - \hat{\sigma}_{uq}}{\hat{\sigma}_W^2 - \hat{\sigma}_u^2} = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X}, \quad (6.4)$$

where $\hat{\sigma}_{WD}$ is the estimated covariance between the mis-measured response and explanatory variables, $\hat{\sigma}_{uq}$ is the estimated covariance between the measurement errors (u_i, q_i), $\hat{\sigma}_W^2$ is the estimated variance of the mis-measured explanatory variable (W), and $\hat{\sigma}_u^2$ is the estimated measurement error variance associated with the mis-measured explanatory variable, where $\hat{\sigma}_{XY} = \hat{\sigma}_{WD} - \hat{\sigma}_{uq}$ and $\hat{\sigma}_X^2 = \hat{\sigma}_W^2 - \hat{\sigma}_u^2$. $\hat{\beta}_{1,MOM}$ simplifies in the obvious way for Cases 2 ($\hat{\sigma}_u^2 = \hat{\sigma}_{uq} = 0$) and 3 ($\hat{\sigma}_{uq} = 0$).

The SIMEX correction developed by Cook & Stefanski (1994) is a general framework that is applicable to many situations, and is purported to account for uncorrelated measurement error in both the explanatory and response variable. This method obtains corrected parameter estimates through a simulation-extrapolation process, where additional (uncorrelated) errors are simulated for both mis-measured variables in the simulation step, before using the extrapolation model to obtain corrected parameter estimates, as discussed in detail in Chapter 4. Correlated measurement error is not discussed in the primary references on SIMEX. The GSIMEX correction (Ronning and Rosemann, 2008) is a generalization of the SIMEX

method that was developed to attempt to account for correlated measurement errors. This method obtains corrected parameter estimates in the same fashion as the SIMEX method, but the additional errors for the two mis-measured variables are simulated with correlation in the simulation step from a multivariate normal distribution.

In a Bayesian analysis where measurement error is present, the true unobserved value of a variable measured with error is treated like a latent variable and given a prior distribution. Bayesian measurement error models are described using three pieces: an outcome model, a measurement model, and an exposure model (Carroll et al. (2006); Gilks et al. (1996)). The outcome model is a model for the outcomes that would be obtained if measurement error was not present and in the SLR setting is

$$Y_i \sim N(\beta_0 + \beta_1 X_i, \sigma_\epsilon^2). \quad (6.5)$$

The measurement model is a model for the mis-measured variables given the true variables and is different for each of the cases considered in the SLR setting. Bayesian measurement error models are often developed for specific applications and so there may be previous approaches to this problem, but we have not come across a model that accounts for correlated measurement errors. For Cases 3 and 4, we propose the following measurement model:

$$\begin{pmatrix} D_i | Y_i \\ W_i | X_i \end{pmatrix} \sim MVN \left(\begin{pmatrix} Y_i \\ X_i \end{pmatrix}, \begin{pmatrix} \sigma_q^2 & \sigma_{uq} \\ \sigma_{uq} & \sigma_u^2 \end{pmatrix} \right), \quad (6.6)$$

where $\sigma_{uq} = 0$ in Case 3. The exposure model is a model for the true unobserved explanatory variable (X_i) and is present in Cases 1, 3, and 4. Explanatory variables are often centered in Bayesian models for convergence purposes, and by centering the

observed explanatory variable, the exposure model is

$$X_i \sim N(0, \sigma_X^2). \quad (6.7)$$

The parameters associated with the measurement errors (σ_q^2 , σ_u^2 , and σ_{uq}) are assumed known in Equation 6.6. However, if replicate data are available (for example, observing D_{ij} replicates for each Y_i and W_{ij} replicates for each X_i , where $j = 1, 2, \dots, m_i$) they can be estimated. Although it is often difficult to disentangle $\hat{\sigma}_q^2$ from $\hat{\sigma}_\epsilon^2$, the two variances can be separated if $m_i > 1$ replicates of Y_i are available. All other parameters (β_0 , β_1 , $\tau_\epsilon = \frac{1}{\sigma_\epsilon^2}$, and $\tau_X = \frac{1}{\sigma_X^2}$) are given prior distributions, which depend on the particular situation and are often chosen to be non-informative.

It is also possible to estimate the correlation between the measurement errors in the Bayesian model. By placing a prior distribution for the correlation parameter of $\rho_{u_i, q_i} \sim Uniform(-1, 1)$, which non-informatively constrains the parameter to be between -1 and 1, we are able to obtain estimates of the correlation parameter. This is especially helpful in potentially applying these methods as there is little information about estimating measurement error variances in the literature and no discussion of estimating the correlation in measurement errors. Ronning and Rosemann (2008) require this correlation in their GSIMEX approach, but do not discuss how to obtain it; they just try different values to assess impacts of varying amounts of correlation.

6.3 Simulation Studies

Performances of the correction methods discussed in Section 6.2 were compared using simulation studies for each case, and, for comparison, the naive SLR model (the ordinary least squares model with no measurement error correction) results are included. The results for the SIMEX method using the nonlinear extrapolation

function are omitted here, as the instability of the method was shown in Chapter 4 and similar results were found here. Additionally, Ronning and Rosemann (2008) only considered the linear and quadratic extrapolation functions. Because of this, the results presented here for both the GSIMEX and SIMEX methods use the quadratic extrapolation function.

The sample sizes and parameter values used in each simulation study are given in Table 6.1. The true values of the explanatory variable (X_i) were simulated using a normal distribution ($X_i \sim N(0, \sigma_X^2)$) and the measurement errors (U_i, Q_i) were simulated using a multivariate normal distribution, centered at 0, with the covariance matrix corresponding to parameter values given in Table 6.1 for each case.

The simulation studies were conducted in R (R Core Team, 2016) and used the R-packages `simex` (Lederer and Kchenhoff, 2013) and `R2jags` (Su and Yajima, 2016). An R implementation of GSIMEX that was used is provided in the Appendix. Results are summarized numerically and using beanplots (Kampstra, 2008) which display each result (small lines), the mean of the simulations (large line), and nonparametric density estimates.

For the Bayesian method, JAGS (Plummer, 2003) was used to implement the Gibbs sampling algorithm (Geman and Geman, 1984), resulting in samples from the joint posterior distribution of the parameters. Three independent chains, each with random starting values, were used and the first 1,000 samples were discarded as *burn-in*. A final chain length of 5,000 was implemented in each simulation. Convergence was monitored using sample path plots and potential scale reduction factors, where all chains in each simulation converged (a threshold of $\hat{R} < 0.05$ was used to determine convergence). The prior distributions used were $\beta_0 \sim N(0, 10^6)$, $\beta_1 \sim N(0, 10^6)$, $\tau_X \sim \text{gamma}(0.5, 2)$, and $\tau_\epsilon \sim \text{gamma}(0.5, 2)$. The *gamma*(a, b) notation in JAGS

Table 6.1: Simulation study parameter values for each Case considered.

| Case | n | (β_0, β_1) | $(\sigma_\epsilon^2, \sigma_X^2)$ | σ_q^2 | σ_u^2 | ρ_{u_i, q_i} |
|------|-----|----------------------|-----------------------------------|--------------|--------------|-------------------|
| 3 | 100 | (1, 5) | (4, 4) | (1, 4, 12) | (1, 4, 12) | 0 |
| 4.1 | 100 | (1, 5) | (4, 4) | (1, 4, 12) | (1, 4, 12) | (0.5, 0.7) |
| 4.2 | 100 | (1, 5) | (4, 4) | (1, 4) | (1, 4) | (0.5, 0.7) |

specifies a Gamma distribution with mean a/b and variance a/b^2 . The mean of the posterior distribution from each simulation is used as the estimate of β_1 .

All of the correction methods require the measurement error variances be known when replicate data are not available to estimate them. The simulation studies presented in Sections 6.3.1 and 6.3.2.1 use the data generating measurement error parameters as the assumed known values for each of the methods. The Bayesian measurement error model has the potential to estimate the correlation present between the measurement errors (assuming the measurement error variances are known); the simulations in Section 6.3.2.2 compare the performance of the Bayesian measurement error model where the correlation is estimated to the Bayesian measurement error model that assumes the correlation is known.

6.3.1 Simulation Study: Case 3

The GSIMEX correction method is a generalization of the SIMEX correction method that accounts for correlated measurement errors when present. When measurement errors are uncorrelated, the GSIMEX method produces the same results as the SIMEX method, therefore results from the GSIMEX method are not included here. To see the effects of changing measurement error variance on the different correction methods, three combinations of the measurement error variances were

considered $((\sigma_q^2, \sigma_u^2) = (1,1), (4,4), \text{ and } (12, 12))$. A sample size of 100 was used and 500 simulations were run.

Equation 6.3 shows that the addition of uncorrelated measurement error in the response variable results in the same bias associated with the naive estimator of the slope when measurement error is present in the explanatory variable only, therefore the results from this simulation are very similar to the results obtained in Chapter 4, where measurement error was present in the explanatory variable only. The distributions of $\hat{\beta}_1$ resulting from the three simulations are shown in Figure 6.1 and numerical summaries are provided in Table 6.2. The bias present in the naive estimator is visible in all scenarios considered and becomes more prominent as the measurement error variances increase. The three correction methods have similar performance when the measurement errors are small ($\sigma_u^2 = \sigma_q^2 = 1$), but distinctions between the methods become more clear when the measurement error variances become larger. The method of moments correction and the Bayesian correction both have similar average estimates in each scenario, where a small positive bias appears as the measurement error variances become large. Table 6.2 shows the variability of the method of moments correction is larger than that of the Bayesian correction, especially when the measurement error variances are large. The variability of the SIMEX method using the quadratic extrapolation function is smaller than the other corrections, but it is a conservative correction, as seen here and in Chapter 4.

6.3.2 Simulation Study: Case 4

When correlated measurement errors are present, the correlation (ρ_{u_i, q_i}) can be estimated in the Bayesian model if an appropriate prior distribution is specified. The method of moments correction and the GSIMEX correction require the correlation to be known or estimated in order to account for correlated measurement errors.

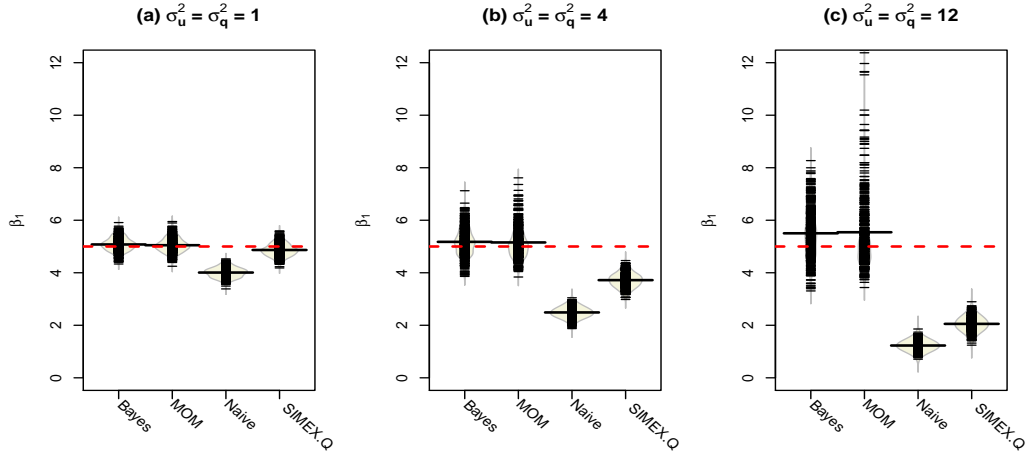


Figure 6.1: Case 3 simulation results for $\beta_1 = 5$ for each method based on 500 simulations in each scenario. Horizontal dashed lines appear at the target value ($\beta_1 = 5$). A sample size of 100 was used, with increasing measurement error from left to right. Estimates outside of $(0, 12)$ were not displayed to allow comparison of the majority of results.

In order to fairly compare the Bayesian method to the other correction techniques, we assume the correlation is known and provide the true, data generating value to all correction methods for the simulation in Section 6.3.2.1. We conduct another simulation in Section 6.3.2.2 where we compare results from two Bayesian models, one where the correlation is assumed known and fixed at the data generating value and one where we let the model estimate the correlation.

6.3.2.1 Case 4.1: Known Correlation Although the SIMEX correction method does not account for correlated measurement errors, we include the results from this method here to see the effects of accounting for the measurement error variances but ignoring the correlation present between the two measurement errors. To see the effects of changing measurement error variance and correlation on the different correction methods, three combinations of the measurement error variances $((\sigma_q^2, \sigma_u^2))$

Table 6.2: Case 3 simulation results for $\hat{\beta}_1$ for each method based on 500 simulations in each scenario.

| (σ_u^2, σ_q^2) | Naive | MOM | SIMEX | Bayes |
|----------------------------|---------------|---------------|---------------|---------------|
| | mean (SD) | mean (SD) | mean (SD) | mean (SD) |
| (1, 1) | 4.008 (0.195) | 5.049 (0.276) | 4.868 (0.250) | 5.078 (0.264) |
| (4, 4) | 2.490 (0.204) | 5.155 (0.588) | 3.720 (0.269) | 5.177 (0.518) |
| (12, 12) | 1.229 (0.183) | 5.543 (1.782) | 2.053 (0.265) | 5.503 (0.907) |

= (1,1), (4,4), and (12,12)) and two values of correlation ($\rho = 0.5, 0.7$) were considered. A sample size of 100 was used and 500 simulations were run.

The distributions of $\hat{\beta}_1$ resulting from the simulations are shown in Figure 6.2 and numerical summaries are provided in Table 6.3. The attenuation of the naive estimator is seen in all cases, where greater attenuation occurs with larger measurement error variances. The measurement error correction methods studied here have comparable performance when the measurement error variances are small ($(\sigma_u^2, \sigma_q^2) = (1, 1)$) for both values of correlation. The SIMEX method ignores the correlation present between the measurement errors and results from this method contain bias when the measurement error variances increase. Results from the GSIMEX method appear biased when the measurement errors are larger, and contain even more bias than the results from the SIMEX method, which ignores the correlation present in the measurement errors. The method of moments correction and the Bayesian correction have similar means, resulting in a small positive bias as the measurement error variances increase. The results from the method of moments correction are more variable than those produced from the Bayesian method, as can be seen in Table 6.3. All methods studied have similar performance with respect to bias and variability for the two values of ρ_{u_i, q_i} considered.

Table 6.3: Case 4.1 simulation results for $\hat{\beta}_1$ for each method based on 500 simulations in each scenario.

| ρ_{u_i, q_i} | (σ_u^2, σ_q^2) | Naive mean (SD) | MOM mean (SD) | GSIMEX mean (SD) | SIMEX.Q mean (SD) | Bayes mean (SD) |
|-------------------|----------------------------|--------------------|------------------|---------------------|----------------------|--------------------|
| 0.5 | (1, 1) | 4.098 (0.178) | 5.025 (0.262) | 4.870 (0.232) | 4.969 (0.240) | 5.060 (0.255) |
| | (4, 4) | 2.740 (0.185) | 5.102 (0.556) | 3.848 (0.240) | 4.082 (0.238) | 5.135 (0.493) |
| | (12, 12) | 1.615 (0.173) | 5.281 (3.274) | 2.364 (0.250) | 2.691 (0.234) | 5.385 (0.844) |
| 0.7 | (1, 1) | 4.136 (0.172) | 5.021 (0.254) | 4.873 (0.222) | 5.017 (0.232) | 5.057 (0.249) |
| | (4, 4) | 2.840 (0.176) | 5.092 (0.537) | 3.896 (0.228) | 4.230 (0.227) | 5.127 (0.478) |
| | (12, 12) | 1.766 (0.165) | 5.292 (2.378) | 2.483 (0.240) | 2.944 (0.216) | 5.347 (0.810) |

6.3.2.2 Case 4.2: Unknown Correlation A second simulation study was performed for Case 4 to compare the performance of the Bayesian model when ρ_{u_i, q_i} is assumed known to when we allow the model to estimate it. 100 observations were generated using the parameter values given in Table 6.1. The distributions of $\hat{\beta}_1$ resulting from 250 simulations are shown in Figure 6.3 and numerical summaries are provided in Table 6.4. The two Bayesian models have nearly identical performance in terms of bias and variability, but the estimated correlation model does not require specification (or knowledge) of ρ_{u_i, q_i} . The correlation parameter may be underestimated with this sample size but seems to be reasonably well-estimated.

The resulting posterior distributions of $\hat{\beta}_1$ from one simulation for each Bayesian model are shown in Figure 6.4, where it can be seen that the posterior distributions from both Bayesian models are centered at nearly the same place. Table 6.5 shows more variability associated with the estimate obtained from the Bayesian model that estimates the correlation, as compared to the model that assumes the correlation is known. Estimated ρ_{u_i, q_i} values and either standard deviations or 90% credible intervals are included in Tables 6.4 and 6.5, where results show the model provides

better estimates of the true value of the correlation when larger measurement errors are present. The reason for this improved performance is not immediately obvious.

Table 6.4: Case 4.2 simulation results for $\hat{\beta}_1$ and $\hat{\rho}_{u_i, q_i}$ for the two Bayesian models based on 250 simulations in each scenario.

| ρ_{u_i, q_i} | (σ_u^2, σ_q^2) | Known ρ_{u_i, q_i} | Estimated ρ_{u_i, q_i} | |
|-------------------|----------------------------|------------------------------|------------------------------|--------------------------------------|
| | | $\hat{\beta}_1$ mean (SD) | $\hat{\beta}_1$ mean (SD) | $\hat{\rho}_{u_i, q_i}$ mean (SD) |
| 0.5 | (1, 1) | 5.063 (0.244) | 5.043 (0.240) | 0.334 (0.136) |
| 0.5 | (4, 4) | 5.135 (0.477) | 5.125 (0.476) | 0.490 (0.073) |
| 0.7 | (1, 1) | 5.058 (0.235) | 5.039 (0.231) | 0.466 (0.124) |
| 0.7 | (4, 4) | 5.123 (0.458) | 5.068 (0.436) | 0.623 (0.062) |

Table 6.5: Posterior means and 90% credible intervals for $\hat{\beta}_1$ and $\hat{\rho}_{u_i, q_i}$ from **one** simulation for each scenario in Case 4.2.

| ρ_{u_i, q_i} | (σ_u^2, σ_q^2) | Known ρ_{u_i, q_i} | Estimated ρ_{u_i, q_i} | |
|-------------------|----------------------------|----------------------------------|----------------------------------|--|
| | | $\hat{\beta}_1$ mean (90% CI) | $\hat{\beta}_1$ mean (90% CI) | $\hat{\rho}_{u_i, q_i}$ mean (90% CI) |
| 0.5 | (1, 1) | 4.842 (4.433, 5.283) | 4.827 (4.409, 5.261) | 0.472 (-0.150, 0.948) |
| 0.5 | (4, 4) | 4.621 (3.900, 5.427) | 4.677 (3.857, 5.637) | 0.574 (0.080, 0.943) |
| 0.7 | (1, 1) | 4.837 (4.450, 5.243) | 4.827 (4.425, 5.250) | 0.621 (0.098, 0.965) |
| 0.7 | (4, 4) | 4.620 (3.951, 5.394) | 4.615 (3.880, 5.452) | 0.708 (0.309, 0.972) |

6.4 Discussion

Measurement error research is generally focused on correcting for the bias caused by measurement error in the explanatory variable, and the effects of measurement error in the response variable are commonly ignored as the naive estimator is unbiased if additive measurement error is present in the response variable only. Measurement error in the response variable and in combinations with measurement error in the explanatory variable does matter in inferences generated from the models. Simulation

studies were used to compare four measurement error correction methods when measurement error is present in both the response and the explanatory variable in the SLR setting.

In both cases considered (uncorrelated and correlated measurement errors), all correction methods had similar performance when the measurement errors were small. The SIMEX and GSIMEX methods show considerable bias when the measurement errors become larger, and sometimes are not much of an improvement over the naive estimator. The method of moments correction and the Bayesian correction both have similar performance in terms of bias, while the method of moments correction contains more variability.

The naive estimator of the slope given in Equation 6.3 shows that the presence of positive correlation between the measurement errors reduces the attenuation caused by measurement error in the explanatory variable, therefore the naive estimator and the SIMEX correction actually contain less bias when positive correlation is present as compared to when the measurement errors are uncorrelated. The method of moments correction is highly variable when the measurement error variances are large, and contains even more variability when measurement errors are correlated as compared to uncorrelated as can be seen in Tables 6.2 and 6.3.

Although the GSIMEX method attempts to account for correlated measurement errors, estimates obtained in the simulation studies presented here were no better than the SIMEX estimates and sometimes worse. The results in Ronning and Rosemann (2008) conclude that the GSIMEX method performs better than the SIMEX method, but their results were based on only 50 simulations using a limited set of parameter values. More extensive simulations using different parameter values do not agree with these conclusions, as can be seen in Section 6.3.2.1. The GSIMEX method may have better performance than the SIMEX method when measurement errors are negatively

correlated. However, negatively correlated measurement errors seem to be something that would rarely occur, so we did not explore these in our simulation study.

The method of moments correction and the Bayesian correction both show small positive bias when measurement errors are large, however we believe this is related to the sample size used in the simulations. In Chapter 4, a small positive bias was present for both methods in the results obtained using small sample sizes, however both methods were approximately unbiased for larger sample sizes.

The Bayesian measurement error model presented in this research can account for correlated measurement errors in the SLR setting. Corrected estimates of the slope obtained from this model are similar to the method of moments corrected estimates in terms of the bias present, however they are less variable. The Bayesian method is the only method studied here in which an estimate of the correlation is not needed in order to account for correlated measurement errors. This aspect of the model makes it most attractive for potential use in a real application.

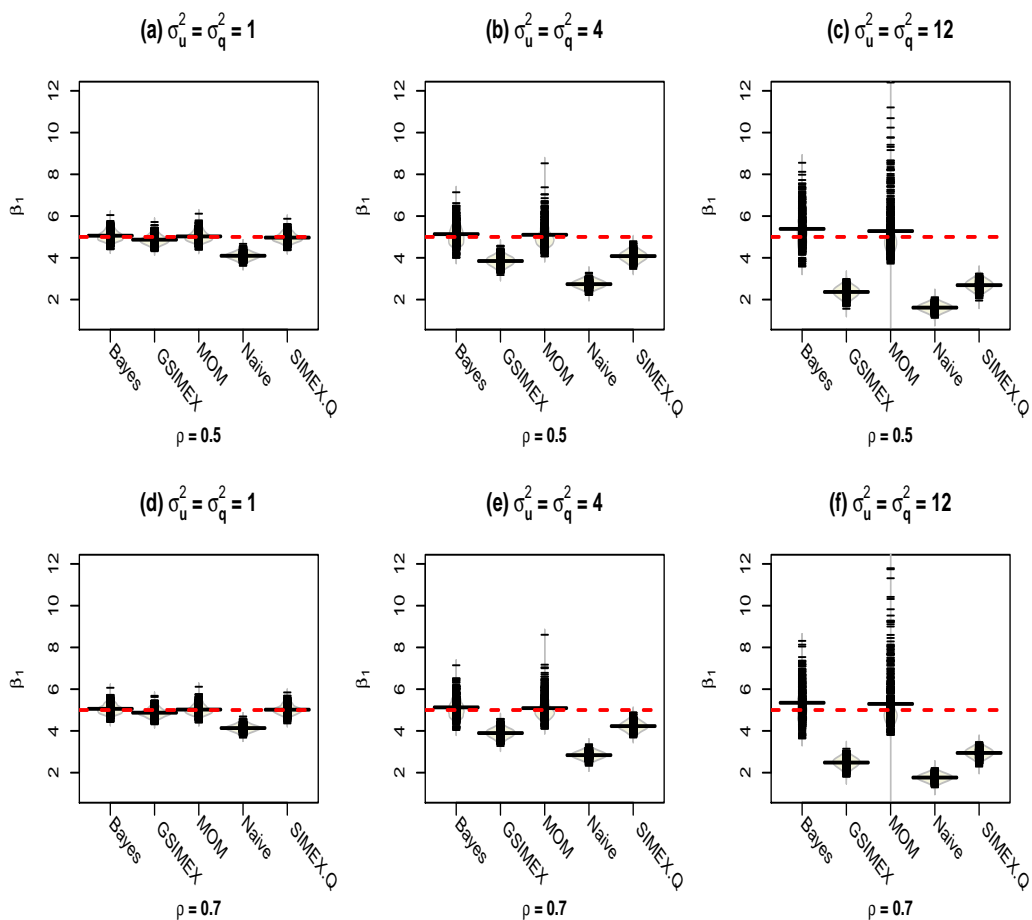


Figure 6.2: Case 4.1 simulation results for $\beta_1 = 5$ for each method based on 500 simulations in each scenario. Horizontal dashed lines appear at the target value ($\beta_1 = 5$). Panels (a)-(c) are for $\rho_{u_i, q_i} = 0.5$ and panels (d)-(f) are for $\rho_{u_i, q_i} = 0.7$. A sample size of 100 was used, with increasing measurement error from left to right in each row. Estimates outside of (1, 12) were not displayed to allow comparison of the majority of results.

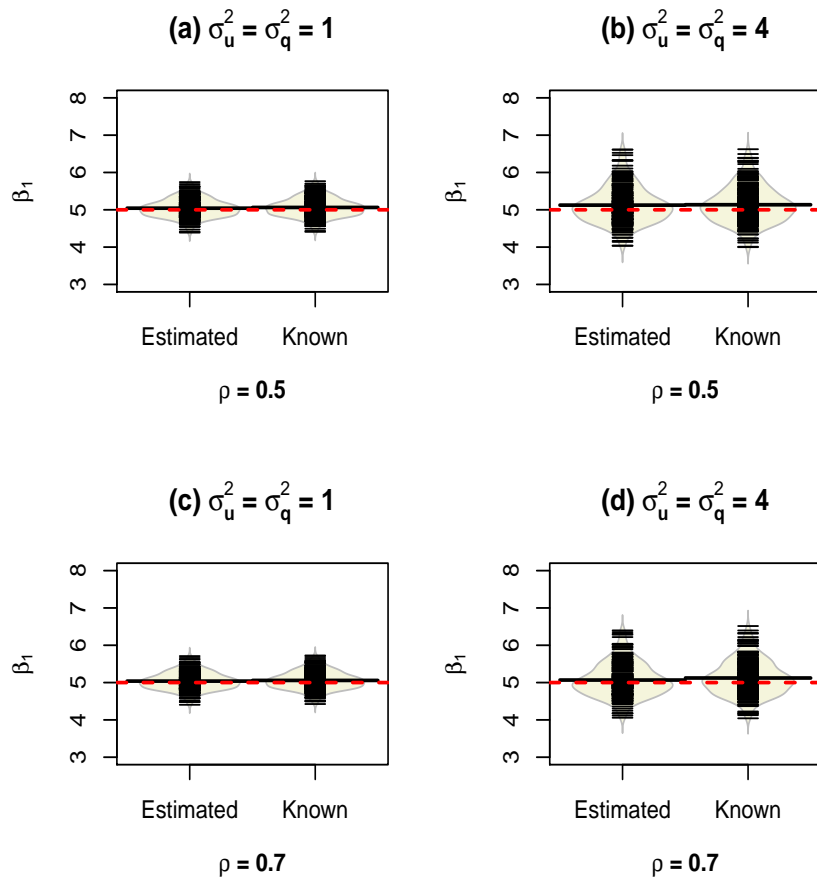


Figure 6.3: Case 4.2 simulation results for $\beta_1 = 5$ for the two Bayesian models based on 250 simulations in each scenario. Horizontal dashed lines appear at the target value ($\beta_1 = 5$). Panels (a) and (b) are for $\rho_{u_i, q_i} = 0.5$ and panels (c) and (d) are for $\rho_{u_i, q_i} = 0.7$.

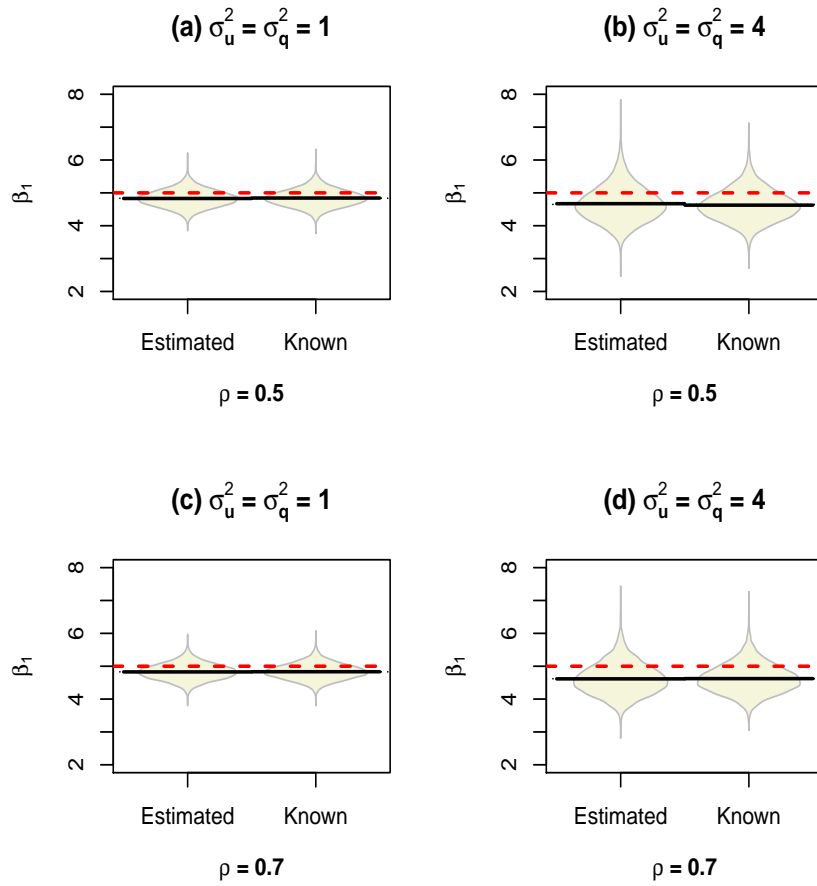


Figure 6.4: Posterior distributions from **one** simulation for β_1 for the two Bayesian models in each scenario for Case 4.2 . Horizontal dashed lines appear at the target value ($\beta_1 = 5$). Panels (a) and (b) are for $\rho_{u_i, q_i} = 0.5$ and panels (c) and (d) are for $\rho_{u_i, q_i} = 0.7$.

6.5 Appendix

The GSIMEX function was originally implemented in GAUSS by Ronning and Rosemann (2008). We implemented the function in R, where our function provides estimates for 5 measurement error correction methods (the naive estimate, the method of moments estimate, the linear and quadratic GSIMEX estimates, and the SIMEX estimate).

The lambda vector we use corresponds directly to what is used as the default in the SIMEX function (Lederer and Kchenhoff, 2013). The SIMEX function uses $\lambda \geq 0$, where $\lambda = 0$ corresponds to the observed mis-measured variable, then obtains corrected estimates by extrapolating back to $\lambda = -1$. We shift λ by 1 unit (therefore, our $\lambda \geq 1$, and $\lambda = 1$ corresponds to the observed mis-measured variable), where in the extrapolation step we extrapolate back to $\lambda = 0$.

To use the GSIMEX function, you need to specify the lambda vector to use to obtain increased measurement error in the simulation step, the number of simulations to perform for each value of lambda (*N.simex*), the mis-measured explanatory variable (W_i), the mis-measured response variable (D_i), the standard deviation of the measurement errors associated with the mis-measured variables (σ_u and σ_q), and the covariance between the measurement errors associated with the mis-measured explanatory and response variables ($Cov(u_i, q_i) = \rho_{u_i, q_i} * \sigma_u * \sigma_q$).

```
gsimex <- function(lambda.vec,x,y,sig.ux,sig.uy,cov.uxuy,N.simex){
require(MASS)

theta.hat.mean <- matrix(0, nrow=length(lambda.vec), ncol=2)

  ## ^ stores the estimates of beta for each lambda in gsimex
```

```

theta.hat.mean2 <- matrix(0, nrow=length(lambda.vec), ncol=2)
  ## ^ stores the estimates of beta for each lambda in simex

N <- length(x)
for(i in 1:length(lambda.vec)){
  factor_x <- sqrt(lambda.vec[i]-1)
  factor_y <- sqrt(lambda.vec[i]-1)

  theta_hat_lambdamue <- matrix(0, nrow=N.simex, ncol=2)
  ## ^ stores the estimates of beta for each simulation in gsimex
  theta_hat_lambdamue2 <- matrix(0, nrow=N.simex, ncol=2)
  ## ^ stores the estimates of beta for each simulation in simex

for(j in 1:N.simex){
  Sigma.GSimex <- matrix(c(sig.ux^2, cov.uxuy, cov.uxuy, sig.uy^2), 2, 2)
  mu <- c(0, 0)
  uxuy <- mvrnorm(N, mu, Sigma.GSimex, empirical=T)
  ux <- uxuy[,1]
  uy <- uxuy[,2]
  x_lambda <- x + factor_x*ux
  y_lambda <- y + factor_y*uy
  dat_x <- cbind(rep(1,N), x_lambda)
  theta_hat_lambdamue[j,] <- t((solve(t(dat_x)%*%dat_x))%*%
    (t(dat_x)%*%y_lambda))

  Sigma.Simex <- matrix(c(sig.ux^2, 0, 0, sig.uy^2), 2, 2)

```

```

uxuy2 <- mvrnorm(N, mu, Sigma.Simex, empirical=T)
ux2 <- uxuy2[,1]
uy2 <- uxuy2[,2]
x_lambda2 <- x + factor_x*ux2
y_lambda2 <- y + factor_y*uy2
dat_x2 <- cbind(rep(1,N), x_lambda2)
theta_hat_lambdamue2[j,] <- t((solve(t(dat_x2)%*%dat_x2))%*%
  (t(dat_x2)%*%y_lambda2))
  }
theta.hat.mean[i,] <- apply(theta_hat_lambdamue, 2, mean)
theta.hat.mean2[i,] <- apply(theta_hat_lambdamue2, 2, mean)
  }
lambda.vec2 <- lambda.vec^2
x.naive <- cbind(rep(1,length(x)), x)
x.mat.lin <- cbind(rep(1,length(lambda.vec)),lambda.vec)
x.mat.quad <- cbind(rep(1,length(lambda.vec)),lambda.vec,lambda.vec2)

naive <- t((solve(t(x.naive)%*%x.naive))%*%(t(x.naive)%*%y)) ##naive
mom.slope <- (cov(y,x) - cov.uxuy)/(var(x) - sig.ux^2) ## mom slope
mom.intercept <- (mean(y) - (mom.slope*mean(x))) ## mom intercept
linear.coeffs <- (solve(t(x.mat.lin)%*%x.mat.lin))%*%
  (t(x.mat.lin)%*%theta.hat.mean)
quadratic.coeffs <- (solve(t(x.mat.quad)%*%x.mat.quad))%*%
  (t(x.mat.quad)%*%theta.hat.mean)
quadratic.coeffs.simex <- (solve(t(x.mat.quad)%*%x.mat.quad))%*%
  (t(x.mat.quad)%*%theta.hat.mean2)

```


References

- Buonaccorsi, J. P., 2010. Measurement error: models, methods, and applications. CRC Press.
- Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M., 2006. Measurement error in nonlinear models: a modern perspective. CRC press.
- Cook, J. R., Stefanski, L. A., 1994. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical association* 89 (428), 1314–1328.
- Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (6), 721–741.
- Gilks, W. R., Richardson, S., Spiegelhalter, D. J., 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- Kampstra, P., 2008. Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of Statistical Software, Code Snippets* 28 (1), 1–9.
URL <http://www.jstatsoft.org/v28/c01/>
- Lederer, W., Kchenhoff, H., 2013. simex: SIMEX- and MCSIMEX-Algorithm for measurement error models. R package version 1.5.
URL <http://CRAN.R-project.org/package=simex>
- Plummer, M., 2003. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In: *Proceedings of the 3rd international workshop on distributed statistical computing*. Vol. 124. Technische Universit at Wien Wien, Austria, p. 125.
- R Core Team, 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL <http://www.R-project.org/>
- Ronning, G., Rosemann, M., 2008. Simex estimation in case of correlated measurement errors. *AStA Advances in Statistical Analysis* 92 (4), 391–404.
- Su, Y.-S., Yajima, M., 2016. R2jags: A Package for Running jags from R. R package version 0.04-03.
URL <http://CRAN.R-project.org/package=R2jags>

CHAPTER SEVEN

CONCLUSIONS AND FUTURE WORK

Measurement errors are common in the variables in many data sets and the impacts of those errors on inferences in regression models built from the mis-measured variables include bias and a loss of power. A suite of different correction methods were reviewed in previous chapters. Across the different scenarios, the performance of different approaches suggested the use of the proposed Bayesian model as opposed to the corrections available in the frequentist approaches. It provided better point estimation performance and a unified framework for inferences that suggests its use over other methods.

Researchers are often unaware that the use of summary measures in analyses induces measurement error, which should be accounted for to obtain less biased parameter estimates. The variability associated with the summary measure is often needed to account for the measurement error induced by the measure, however it is not always obvious how to obtain it. For example, the area under the curve (AUC) summary measure is commonly used by researchers studying pharmacodynamic and pharmacokinetic properties, however in most scenarios there is not a straight-forward method available to estimate the variance associated with it (Hilborn et al., 1999).

Measurement error research is commonly focused on correcting parameter estimates for bias induced from measurement errors in the explanatory variable, while measurement error in both the explanatory and response variable receives much less attention. The overall focus of this dissertation was two-fold:

1. Develop an estimator for the variance associated with the AUC summary measure.

2. Develop a Bayesian measurement error model that accounts for measurement errors in both the response and explanatory variables in the simple linear regression setting.

An application involving measurement error in the simple linear regression setting directed our focus on measurement error modeling. Although our equine application involved only eight observations, it allowed us to explore many aspects of measurement error modeling in the simple linear regression setting, including how to obtain an estimate of the measurement error variance when the AUC summary measure is used in analyses.

A Bayesian AR(1) model was used to obtain estimates of the $SD(\widehat{AUC})$ in Chapter 3. This model assumes the length of the sampling intervals between consecutive observations are held constant over the entire sampling period, which is not always true in AUC data sets. Another statistical tool that can estimate the correlation structure of repeated observations and assumes observations obtained closer in time are more strongly correlated is the semivariogram, developed for use in spatial statistics. Semivariograms can handle unequal distances apart and still provide an estimate of the correlation structure for repeated observations. Future research will explore the use of semivariograms to obtain estimates of the correlation structure when unequal sampling intervals are present, likely using a Bayesian model similar to the AR(1) model.

Comparisons of measurement error correction methods were made for the simple linear regression model where measurement errors were present in the explanatory variable only, uncorrelated measurement errors were present in the response and explanatory variables, and correlated measurement errors were present between the response and explanatory variables. In all scenarios studied, the Bayesian

measurement error model performed as well or better than the other methods considered. Measurement error in the response variable only does not induce bias in the estimate of the slope, however it increases the variability of the fitted lines and decreases the power to detect effects (Carroll et al., 2006). We did not focus on measurement error in the response variable only in this research, however this scenario was studied in Weeding and Greenwood (2014).

The focus of this work was primarily on the bias and variability in measurement error correction methods, however measures of the uncertainty associated with the point estimate are just as important to be able to use a method. The standard error associated with the estimate obtained from the method of moments correction is computed using a formula that is an adjustment of the typical standard error obtained from SLR when there is no measurement error present, whereas the SIMEX method obtains standard errors associated with the regression parameters two ways, either using a jackknife method or an asymptotic approach. Inferences for model parameters are made directly from posterior distributions in Bayesian analyses, therefore the standard errors associated with the regression parameters and any quantile of interest are available directly from the posterior distributions resulting from the Bayesian measurement error model. The Bayesian measurement error model also incorporates the uncertainty associated with other unknown parameters, which is a feature of the Bayesian measurement error model that is missed when using the other measurement error correction methods. Future research will explore the performance of the correction methods in the simple linear regression setting in terms of complete inferences, such as the performance of confidence or credibility intervals or model comparisons. In addition, it would be interesting to examine the performance of the correction methods when extended to other statistical models, such as multiple linear

regression models, generalized linear models (glms), nonlinear models, mixed models, etc.

There are many potential applications of the Bayesian measurement error model proposed here. For example, in climate change research, climatic information such as temperature (Morice et al., 2012) is often summarized from minute-by-minute measurements to daily extrema (min or max) or averages, to monthly or yearly averages of those values, for a single location. Further aggregation before analysis for trends over time often occurs across space, averaging weather station information into regional, hemispheric, or global values. The measurement errors due to aggregation are sometimes presented with results but rarely used in the analyses of the resulting summary values. Often, measurement errors have decreased over time in climate historical records as the number of monitoring stations has increased or the quality of measuring instruments used has improved and this information is rarely used to adjust the trend estimates.

Another area where measurement error is prevalent and rarely considered is in educational testing data. When researchers generate results from an Item Response Theory model (see Rizopoulos (2006), for example), it is basically a prediction from a Bernoulli-response generalized linear mixed model and the predictions are more precisely estimated for subjects that are near the mean of the sample used to build the scoring model. Subjects at extremes of the underlying estimated “ability” spectrum have results that are less well-known, resulting in higher error attached to their predictions. These types of scores are used as responses and explanatory variables, depending on the study goals.

Both of these scenarios lead to commonly used and analyzed variables (average temperature and ability scores) and both have measurement error estimates attached to the values as common practice, however the measurement errors are rarely

incorporated in the subsequent statistical analyses. Future research will target specific applications like these where researchers may be missing out on the potential detection of interesting relationships by ignoring the measurement error information in their responses, explanatory variables, or both.

References

- Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M., 2006. Measurement error in nonlinear models: a modern perspective. CRC press.
- Hilborn, R., Bue, B. G., Sharr, S., 1999. Estimating spawning escapements from periodic counts: a comparison of methods. *Canadian Journal of Fisheries and Aquatic Sciences* 56 (5), 888–896.
- Morice, C. P., Kennedy, J. J., Rayner, N. A., Jones, P. D., 2012. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 data set. *Journal of Geophysical Research: Atmospheres* 117 (D8).
- Rizopoulos, D., 2006. ltm: An r package for latent variable modelling and item response theory analyses. *Journal of Statistical Software* 17 (5), 1–25.
URL <http://www.jstatsoft.org/v17/i05/>
- Weeding, J., Greenwood, M. D., 2014. A bayesian approach to modeling measurement errors. In: *JSM Proceedings, Section on Bayesian Statistical Science*. American Statistical Association, Alexandria, VA, pp. 1443–1452.

REFERENCES CITED

Bibliography

- Allison, D. B., Paultre, F., Maggio, C., Mezzitis, N., Pi-Sunyer, F. X., 1995. The use of areas under curves in diabetes research. *Diabetes care* 18 (2), 245–250.
- Armitage, P., Berry, G., Matthews, J. N. S., 2008. *Statistical methods in medical research*. John Wiley & Sons.
- Bartlett, J. W., Keogh, R. H., 2016. Bayesian correction for covariate measurement error: a frequentist evaluation and comparison with regression calibration. arXiv preprint arXiv:1603.06284.
- Berkson, J., 1950. Are there two regressions? *Journal of the American Statistical Association* 45 (250), 164–180.
- Box, G. E., 1950. Problems in the analysis of growth and wear curves. *Biometrics* 6 (4), 362–389.
- Buonaccorsi, J. P., 2010. *Measurement error: models, methods, and applications*. CRC Press.
- Cai, W., Kaiser, M., Dekkers, J., 2012. Bayesian analysis of the effect of selection for residual feed intake on growth and feed intake curves in yorkshire swine. *Journal of animal science* 90 (1), 127–141.
- Cardoso, F., Sears, W., LeBlanc, S., Drackley, J., 2011. Technical note: Comparison of 3 methods for analyzing areas under the curve for glucose and nonesterified fatty acids concentrations following epinephrine challenge in dairy cows. *Journal of dairy science* 94 (12), 6111–6115.
- Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M., 2006. *Measurement error in nonlinear models: a modern perspective*. CRC press.
- Cernicchiaro, N., Renter, D., Xiang, S., White, B., Bello, N., 2013. Hierarchical bayesian modeling of heterogeneous variances in average daily weight gain of commercial feedlot cattle. *Journal of animal science* 91 (6), 2910–2919.
- Cook, J. R., Stefanski, L. A., 1994. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* 89 (428), 1314–1328.
- Cowles, M. K., Carlin, B. P., 1996. Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* 91 (434), 883–904.
- Cryer, J., Chan, K., 2008. *Time Series Analysis with Applications to R*. Vol. 2. Springer.

- Fekedulegn, D. B., Andrew, M. E., Burchfiel, C. M., Violanti, J. M., Hartley, T. A., Charles, L. E., Miller, D. B., 2007. Area under the curve and other summary indicators of repeated waking cortisol measurements. *Psychosomatic medicine* 69 (7), 651–659.
- Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G., 2008. *Longitudinal data analysis*. CRC Press.
- Freestone, J., Shoemaker, K., Bessin, R., Wolfsheimer, K., 1992. Insulin and glucose response following oral glucose administration in well-conditioned ponies. *Equine Veterinary Journal* 24 (S11), 13–16.
- Fuller, W. A., 2009. *Measurement error models*. Vol. 305. John Wiley & Sons.
- Gałecki, A., Burzykowski, T., 2013. *Linear mixed-effects models using R: A step-by-step approach*. Springer Science & Business Media.
- Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 2014. *Bayesian data analysis*. Vol. 2. Taylor & Francis.
- Gelman, A., Hill, J., 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Gelman, A., Rubin, D. B., 1992. Inference from iterative simulation using multiple sequences. *Statistical science*, 457–472.
- Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (6), 721–741.
- Gilks, W. R., Richardson, S., Spiegelhalter, D. J., 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- Grice, J. E., Jackson, R. V., 2004. Letter to the editor: Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change: A letter to the editor in response to jc preussner et al.(2003) *psychoneuroendocrinology* 28, 916-931. *Psychoneuroendocrinology* 29 (4), 563–564.
- Hilborn, R., Bue, B. G., Sharr, S., 1999. Estimating spawning escapements from periodic counts: a comparison of methods. *Canadian Journal of Fisheries and Aquatic Sciences* 56 (5), 888–896.
- Jackman, S., 2009. *Bayesian analysis for the social sciences*. Vol. 846. John Wiley & Sons.

- Jacobsen, D., Bailey, B., LeCheminant, J., Hill, J., Mayo, M., Donnelly, J., 2004. A comparison of three methods of analyzing post-exercise oxygen consumption. *International journal of sports medicine* 26 (1), 34–38.
- Kampstra, P., 2008. Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of Statistical Software, Code Snippets* 28 (1), 1–9.
URL <http://www.jstatsoft.org/v28/c01/>
- Kashyap, S. R., Bhatt, D. L., Wolski, K., Watanabe, R. M., Abdul-Ghani, M., Abood, B., Pothier, C. E., Brethauer, S., Nissen, S., Gupta, M., et al., 2013. Metabolic effects of bariatric surgery in patients with moderate obesity and type 2 diabetes analysis of a randomized control trial comparing surgery with intensive medical treatment. *Diabetes care* 36 (8), 2175–2182.
- Le Floch, J.-P., Escuyer, P., Baudin, E., Baudon, D., Perlemuter, L., 1990. Blood glucose area under the curve: methodological aspects. *Diabetes care* 13 (2), 172–175.
- Lederer, W., Kchenhoff, H., 2013. simex: SIMEX- and MCSIMEX-Algorithm for measurement error models. R package version 1.5.
URL <http://CRAN.R-project.org/package=simex>
- Lunn, D., Jackson, C., Best, N., Thomas, A., Spiegelhalter, D., 2012. *The BUGS book: A practical introduction to Bayesian analysis*. CRC press.
- Moreaux, S. J. J., Nichols, J. L., Bowman, J. G., Hatfield, P. G., 2011. Psyllium lowers blood glucose and insulin concentrations in horses. *Journal of Equine Veterinary Science* 31 (4), 160–165.
- Morice, C. P., Kennedy, J. J., Rayner, N. A., Jones, P. D., 2012. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 data set. *Journal of Geophysical Research: Atmospheres* 117 (D8).
- Muff, S., Riebler, A., Held, L., Rue, H., Saner, P., 2015. Bayesian analysis of measurement error models using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 64 (2), 231–252.
- Nedelman, J. R., Gibiansky, E., Lau, D. T., 1995. Applying bailer’s method for auc confidence intervals to sparse sampling. *Pharmaceutical research* 12 (1), 124–128.
- Parken, C. K., Bailey, R. E., Irvine, J. R., 2003. Incorporating uncertainty into area-under-the-curve and peak count salmon escapement estimation. *North American Journal of Fisheries Management* 23 (1), 78–90.

- Peterson, J., Moreaux, S., Bowman, J., Hatfield, P., Berardinelli, J., Olsen, J., 2009. Metabolic and physical effects of psyllium supplementation on quarter horses. In: American Society of Animal Science. p. 176.
- Plummer, M., 2003. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In: Proceedings of the 3rd international workshop on distributed statistical computing. Vol. 124. Technische Universit at Wien Wien, Austria, p. 125.
- Potteiger, J., Jacobsen, D., Donnelly, J., 2002. A comparison of methods for analyzing glucose and insulin areas under the curve following nine months of exercise in overweight adults. *International Journal of Obesity & Related Metabolic Disorders* 26 (1).
- Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., Hellhammer, D. H., 2003. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology* 28 (7), 916–931.
- Purves, R. D., 1992. Optimum numerical integration methods for estimation of area-under-the-curve (auc) and area-under-the-moment-curve (aumc). *Journal of pharmacokinetics and biopharmaceutics* 20 (3), 211–226.
- R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL <http://www.R-project.org/>
- Rizopoulos, D., 2006. ltm: An r package for latent variable modelling and item response theory analyses. *Journal of Statistical Software* 17 (5), 1–25.
URL <http://www.jstatsoft.org/v17/i05/>
- Ronning, G., Rosemann, M., 2008. Simex estimation in case of correlated measurement errors. *AStA Advances in Statistical Analysis* 92 (4), 391–404.
- Stefanski, L. A., 2000. Measurement error models. *Journal of the American Statistical Association* 95 (452), 1353–1358.
- Stewart, J., 2001. *Calculus concepts and contexts*, 2nd edn (pacific grove, ca: Brooks/cole).
- Su, Y.-S., Yajima, M., 2016. R2jags: A Package for Running jags from R. R package version 0.04-03.
URL <http://CRAN.R-project.org/package=R2jags>
- Weeding, J., Greenwood, M., Moreaux, S., Nichols, J., 2016. Equine Glucose Data [dataset]. Montana State University Scholarworks, Vienna, Austria.
URL <http://doi.org/10.15788/M2QP4R>

- Weeding, J., Greenwood, M. D., 2014. A bayesian approach to modeling measurement errors. In: JSM Proceedings, Section on Bayesian Statistical Science. American Statistical Association, Alexandria, VA, pp. 1443–1452.
- Whitlock, M. C., McPeck, M. A., Rausher, M. D., Rieseberg, L., Moore, A. J., 2010. Data archiving. *The American Naturalist* 175 (2), 145–146.
- Wishart, J., 1938. Growth-rate determinations in nutrition studies with the bacon pig, and their analysis. *Biometrika* 30 (1/2), 16–28.
- Wolever, T. M., Jenkins, D., 1986. The use of the glycemic index in predicting the blood glucose response to mixed meals. *The American journal of clinical nutrition* 43 (1), 167–172.
- Yeh, K., Kwan, K., 1978. A comparison of numerical integrating algorithms by trapezoidal, lagrange, and spline approximation. *Journal of pharmacokinetics and biopharmaceutics* 6 (1), 79–98.
- Yuan, J., 1993. Estimation of variance for auc in animal studies. *Journal of pharmaceutical sciences* 82 (7), 761–763.