Simulation Study of Two Variable Measurement Error on Linear Regression

Kevin Yan
Department of Statistics
kevin.yan@mail.utoronto.ca

1 Introduction

In this report, we will be looking at the effect of measurement error on a 2 variable, linear regression model. Linear regression was chosen because it is probably the most well-known and used model in statistics; Hence it is important and useful to see what effect measurement error will have on the outcomes of linear regression models. In addition, we have decided to study the case with 2 linear predictors that have varying degrees of measurement error. This allows us to study some effects that may intuitively be interesting, and may occur in real studies, where measurements are required..

2 Problem Description

There will be two points of interest that we will study in this simulation study. The first will include a 3 variable model, which will be used to observe how changes in correlation between measurement errors of 2 predictor variables affect the bias, power, and Type I Error associated with the parameter estimates and hypothesis tests of linear regression. The second will include the 2-variable model, which will be used to analyze how well we can predict ICC, and recover the original betas given estimates of these betas(under the model involving measurement error).

2.1 Data Generation

2.1.1 3-variable Model

For the first part of the study, data will be generated from the following model:

$$Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \beta_3 \times X_3 + \epsilon$$
$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim Normal(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} \\ \sigma_{2,1} & \sigma_2^2 & \sigma_{2,3} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_3^2 \end{bmatrix}), \epsilon \sim Normal(0,1)$$

The true values of our predictor variables are given by X_1 and X_2, however it is often the case that we are no able to properly measure these in practice; This may be due to issues with the measurement devices, or other outside variabilities that may affect the values we observe. Thus we often observe variables of the form:

$$X_{obs1} = X_1 + u_1$$
$$X_{obs2} = X_2 + u_2$$

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \sim Normal(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, U = \begin{bmatrix} \sigma_{u1}^2 & \sigma_{u12} \\ \sigma_{u21} & \sigma_{u2}^2 \end{bmatrix})$$

2.1.2 2-Variable Model

For the second part of the study, the data will be generated from will be from the following model:

$$Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2 + \epsilon$$
$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim Normal(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, C)$$
$$C = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_2^2 \end{bmatrix}$$
$$\epsilon \sim Normal(0, 1)$$

Like the scenario above, measurement error will be introduced in the following way:

$$X_{obs1} = X_1 + u_1$$
$$X_{obs_2} = X_2 + u_2$$
$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \sim Normal(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, U = \begin{bmatrix} \sigma_{u1}^2 & \sigma_{u12} \\ \sigma_{u21} & \sigma_{u2}^2 \end{bmatrix})$$

Note: Correlation between measurement errors will be denoted as u_{12} and thus
$\sigma_{u12} = \sigma_{u1} * \sigma_{u2} * u_{12}$

2.2 Process

The two values of measurement error, we have assumed come from a Bivariate- Normal Distribution, with 0 mean, and a covariance matrix U. We assume that on average, the measurement error incurred will be 0. In both parts of the study described above, we will be adjusting measurement error correlation (ie: $u_{1,2}$), the difference in betas, as well as difference in the size of the means of X_1 and X_2, and measuring the effect they these changes have on power, bias and type I error. Throughout the entire study, once we have chosen the parameters and the model we will be generating data from, we will do the following:

1. Generate a size 100 sample from the model with the true predictor variables
2. Add measurement error to these observations
3. Fit a linear regression model to the data with measurement error, and to the data without
4. Collect the values of interest to us, which include the estimates, power, bias, and type I error.
5. Repeat the Process 10 000 times, and average the results.

Below we will describe, and go over the parameters that we will be adjusting throughout the study and the reasoning behind them.

2.2.1 Measurement Error Correlation

First of all, we will look at the case where the covariance between u_1 and u_2 is zero, which in this case, will mean that the measurement errors are independent. This is a very likely scenario, because it's possible that we can be measuring these variables in isolation of each other, and with different tools, thus removing any relationship between the measurement errors.
Secondly, it's possible that we will be looking at the case where there is large positive correlation (ie: $u_{1,2}$=0.9) between the measurement errors. This may be the case where we measure both variables on the same subject, and with the same instrument.
Thirdly, to relax the scenario above, we can have low/medium correlation between the measurement errors (ie: $u_{1,2} = 0.5$), when perhaps we use different instruments, but some environmental aspects are the same, and hence will cause some slight correlation in the measurement errors.

2.2.2 Adjusting Betas

In order to produce a more complete analysis, we will also be varying some of the parameters in the distribution of the predictor variables $[X_1, X_2]$, as well as some of the β's in our model. By varying the betas, we are inherently changing the relationship structure between the predictor variables and the response variable, holding all else constant(ie:units of all the variables). It's also an interesting idea to vary betas, because we would like to see if creating greater size discrepancies in betas will result in poorer powers for the smaller beta. In other words, would it be more likely to mistaken a small beta for 0, if we know that it's noticeably smaller than another beta that we believe to be non-zero. We can also observe whether or not having a greater value of beta will increase the power of the test (for that particular beta), which we intuitively believe to be true. This will be looked at in the initial part of the study. In the second part of the study, we will instead observe if these changes in betas will affect our ability to recover the true betas, provided estimates from the model under measurement error.

In order to do vary the beta-discrepancy, we will solely be adjusting the value of β_2. The reason being, β_3 must be held at 0 to test type I error(for part 1 only), and we also want to set the value of β_1 close to 0, such that there's a greater possibility for the hypothesis test to be incorrect. Having this setting of β_1 should make more clear the changes in power due to changes in these other parameters. Thus we will set β_1=2, and β_3=0(for part 1 only), and only vary the value of β_2 through the values of 5, 10, 20.

2.2.3 Adjusting Means

In a similar fashion, we will hold μ_1 and μ_3 at 10. We will solely adjust μ_2 to adjust the discrepancy in the means. Thus we will vary μ_1 through the values 20 and 100. A practical purpose of adjusting the means will be to analyze the effects of changing the units of certain predictor variables. We will also see if having larger differences between the values of different predictor variables will affect any of the characteristics of interest.

3 Analysis on a 3-variable model

In this section, we will be generating data from a 3-variable model as described in section 2. Using this model we will try to understand the effect of changing the size of the betas, size of the means of the predictor variables, as well as the magnitude of correlation between the measurement errors on the characteristics of interest described before.

3.1 Process

3.1.1 Settings of global parameters

Throughout this study, we will hold both ICC_1 and ICC_2 at 0.7. Meaning that X_i explains 70% of the variation in X_{obsi} under a simple linear model. This level is set in order to limit the range and intensity of measurement error. Moreover, we will fix the R-squared in the overall model to a level of 0.7, such that we are dealing with a model which has the ability to somewhat describe the response variable given the predictor variables. I believe that setting R-squared at a somewhat "high" level makes sense as we are more concerned with the case of measurement error interfering with the results from a valid model. Also we will keep all the standard deviations of the predictor variables equal to 5, and set the correlation between X_1 X_2 and X_3 at 0.2.

3.2 Additional Information before the results and analysis

Note that we in this simulation study, we are interested in the model involving measurement error (ie: the model with X_{obs1} and X_{obs2}). Thus, we will always be referring to the values in the table involving As and not Bs (because the As refer to estimates from the model with measurement error). Thus any reference to estimates of betas (and their biases), the power of a test, type I error of a test, will refer to estimates and tests based on the linear model fitted on X_{obs1} and X_{obs2}.

Thus these will be the values displayed in the tables that will be of interest to us:

- TAGA1 = Power of the test of $\beta_1 = 0$ with the estimate of β_1 fitted under the model with measurement error.
- TAGA2 = Power of the test of $\beta_2 = 0$ with the estimate of β_2 fitted under the model with measurement error.
- TAGA3 = Type 1 Error of the test of $\beta_1 = 0$ with the estimate of β_1 fitted under the model with measurement error.
- BIASA1,BIASA2,BIASA3, which will be the bias of the estimates of $\beta_1, \beta_2, \beta_3$ under the model with measurement error, respectively.

We will continue to make it a point to make this as clear as possible in the analysis below.

3.3 $\beta_1 = 2$, $\beta_2 = 5$, $\beta_3 = 0$

3.3.1 $\mu_1 = 10$, $\mu_2 = 20$

BETA0 = 10 BETA1 = 2 BETA2=5 BETA3=0

MU1 = 10, MU2 = 20, MU3 = 10

U12 = 0

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.218	0.171
TAGB1	0.999	0.001
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.029	73.625
BIASB1	0.003	0.157
BIASB2	0.000	0.164
BIASB3	0.001	0.160
TAGA0	0.984	0.016
TAGA1	0.954	0.044
TAGA2	1.000	0.000
TAGA3	0.115	0.102
BIASA0	30.470	93.035
BIASA1	-0.445	0.175
BIASA2	-1.492	0.179
BIASA3	0.385	0.254

U12 = 0.5

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.218	0.171
TAGB1	0.999	0.001
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.029	73.625
BIASB1	0.003	0.157
BIASB2	0.000	0.164
BIASB3	0.001	0.160
TAGA0	0.996	0.004
TAGA1	0.629	0.233
TAGA2	1.000	0.000
TAGA3	0.161	0.135
BIASA0	36.411	96.386
BIASA1	-0.959	0.201
BIASA2	-1.593	0.207
BIASA3	0.507	0.273

U12 = 0.9

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.218	0.171
TAGB1	0.999	0.001
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.029	73.625
BIASB1	0.003	0.157
BIASB2	0.000	0.164
BIASB3	0.001	0.160
TAGA0	0.998	0.002
TAGA1	0.240	0.182
TAGA2	1.000	0.000
TAGA3	0.193	0.156
BIASA0	39.091	97.848
BIASA1	-1.396	0.229
BIASA2	-1.548	0.234
BIASA3	0.585	0.283

Figure 1: Table 1 Corr = 0, Table 2 Corr = 0.5, Table 3 Corr = 0.9

In this case we are considering the case where there's a low discrepancy in the betas, as well as a low discrepancy in the means of the predictor variables. We notice that the power of the test on β_1 decreases as we increase the correlation between the measurement errors, dropping from 95.35% to 23.98%. The power of the test on β_2 is 100% for all cases, which can be attributed to the fact that β_2 is noticeably different from both 0 and β_1. Moreover, as we see that Type I Error increases as we increase correlation of the measurement error, from a low of 11.5% when there was no correlation, to a high of 19.3% when it was 0.9. Lastly, we also notice that the sum of the absolute value of biases also increases along with correlation. Therefore, as we increase correlation between measurement error, it thus introduces multi-collinearity, into our model, and causes havoc in our estimates as well as our tests. Lastly, we also observe that the variances of our estimates of power, bias and error are all higher for alphas, compared to the betas. Thus with the introduction of measurement error, the tests, and estimates of the model parameters exhibit less confidence compared to the case when measurement error is non-existent.

3.3.2 $\mu_1 = 10$, $\mu_2 = 100$

In this scenario, we increase the discrepancy in the means by increasing μ_2 from 20 to 100. Here we notice that aside from the bias and estimates of the intercepts, that all other values are virtually identical to the case before. Thus we imply that the differences in the properties we're interested in due to a change in means of the second predictor variable is negligible, as can be seen below. Therefore we observe, and come to the same conclusions as the case above.

BETA0 = 10 BETA1 = 2 BETA2=5 BETA3=0

MU1 = 10, MU2 = 100, MU3 = 10

U12 = 0

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.061	0.058
TAGB1	0.999	0.001
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.039	1562.593
BIASB1	0.003	0.157
BIASB2	0.000	0.164
BIASB3	0.001	0.160
TAGA0	0.965	0.033
TAGA1	0.954	0.044
TAGA2	1.000	0.000
TAGA3	0.115	0.102
BIASA0	149.813	1722.189
BIASA1	-0.445	0.175
BIASA2	-1.492	0.179
BIASA3	0.385	0.254

U12 = 0.5

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.061	0.058
TAGB1	0.999	0.001
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.039	1562.593
BIASB1	0.003	0.157
BIASB2	0.000	0.164
BIASB3	0.001	0.160
TAGA0	0.973	0.027
TAGA1	0.629	0.233
TAGA2	1.000	0.000
TAGA3	0.161	0.135
BIASA0	163.844	1943.114
BIASA1	-0.959	0.201
BIASA2	-1.593	0.207
BIASA3	0.507	0.273

U12 = 0.9

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.061	0.058
TAGB1	0.999	0.001
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.039	1562.593
BIASB1	0.003	0.157
BIASB2	0.000	0.164
BIASB3	0.001	0.160
TAGA0	0.958	0.040
TAGA1	0.240	0.182
TAGA2	1.000	0.000
TAGA3	0.193	0.156
BIASA0	162.895	2150.530
BIASA1	-1.396	0.229
BIASA2	-1.548	0.234
BIASA3	0.585	0.283

Figure 2: Table 1 Corr = 0, Table 2 Corr = 0.5, Table 3 Corr = 0.9

3.4 $\beta_1 = 2, \beta_2 = 10, \beta_3 = 0$

3.4.1 $\mu_1 = 10, \mu_2 = 20$

BETA0 = 10 BETA1 = 2 BETA2=10 BETA3=0

MU1 = 10, MU2 = 20, MU3 = 10

U12 = 0

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.101	0.091
TAGB1	0.778	0.173
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.054	249.878
BIASB1	0.006	0.532
BIASB2	0.000	0.558
BIASB3	0.002	0.542
TAGA0	0.957	0.042
TAGA1	0.602	0.240
TAGA2	1.000	0.000
TAGA3	0.107	0.095
BIASA0	57.221	320.952
BIASA1	-0.265	0.605
BIASA2	-3.056	0.614
BIASA3	0.660	0.874

U12 = 0.5

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.101	0.091
TAGB1	0.778	0.173
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.054	249.878
BIASB1	0.006	0.532
BIASB2	0.000	0.558
BIASB3	0.002	0.542
TAGA0	0.985	0.015
TAGA1	0.123	0.108
TAGA2	1.000	0.000
TAGA3	0.147	0.125
BIASA0	65.450	320.704
BIASA1	-1.342	0.670
BIASA2	-3.033	0.687
BIASA3	0.869	0.906

U12 = 0.9

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.101	0.091
TAGB1	0.778	0.173
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.054	249.878
BIASB1	0.006	0.532
BIASB2	0.000	0.558
BIASB3	0.002	0.542
TAGA0	0.990	0.010
TAGA1	0.067	0.062
TAGA2	1.000	0.000
TAGA3	0.182	0.149
BIASA0	67.747	314.885
BIASA1	-2.321	0.735
BIASA2	-2.726	0.754
BIASA3	1.003	0.909

Figure 3: Table 1 Corr = 0, Table 2 Corr = 0.5, Table 3 Corr = 0.9

In this scenario we observe the case with moderate difference in betas. We can see that the power of the test for β_1 under the model with measurement error decreases as we increase the correlation between the measurement errors. It starts from a value of 60.16% when there's no correlation, and drops to a 6.67% when correlation is at 0.9. Thus when the predictor variables develop a stronger linear relationship, the ability for it to reject β_1 when it's non-zero, begins to falter, partially because of multi-collinearity. We also see a steady increase of Type I Error, represented by TAGA3, as we increase correlation. . Thus we are more likely to reject the null hypothesis, even when it's true. Also, the intercept aside, we see that the sum of the absolute values of the biases is greater as we increase the correlation. In addition we notice that the variances of TAGA1,TAGA3, and the non-intercept biases increase with correlation, which indicates that we're less confident of when to reject the null hypothesis, as well as in our estimates of the parameters/slopes respectively.

3.4.2 $\mu_1 = 10$, $\mu_2 = 100$

Surprisingly, we see that the change of the mean of X_2 from 20 to 100 results in no differences in the power, Type I Error, or biases aside from the intercept. Thus we see that the performance of the t-tests, and estimates of linear regression are unaffected by a change in the mean of the actual predictor variable.

BETA0 = 10 BETA1 = 2 BETA2=10 BETA3=0

MU1 = 10, MU2 = 100, MU3 = 10

U12 = 0

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.056	0.053
TAGB1	0.778	0.173
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.072	5303.347
BIASB1	0.006	0.532
BIASB2	0.000	0.558
BIASB3	0.002	0.542
TAGA0	0.976	0.023
TAGA1	0.602	0.240
TAGA2	1.000	0.000
TAGA3	0.107	0.095
BIASA0	301.700	5928.686
BIASA1	-0.265	0.605
BIASA2	-3.056	0.614
BIASA3	0.660	0.874

U12 = 0.5

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.056	0.053
TAGB1	0.778	0.173
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.072	5303.347
BIASB1	0.006	0.532
BIASB2	0.000	0.558
BIASB3	0.002	0.542
TAGA0	0.973	0.026
TAGA1	0.123	0.108
TAGA2	1.000	0.000
TAGA3	0.147	0.125
BIASA0	308.092	6461.599
BIASA1	-1.342	0.670
BIASA2	-3.033	0.687
BIASA3	0.869	0.906

U12 = 0.9

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.056	0.053
TAGB1	0.778	0.173
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.072	5303.347
BIASB1	0.006	0.532
BIASB2	0.000	0.558
BIASB3	0.002	0.542
TAGA0	0.942	0.055
TAGA1	0.067	0.062
TAGA2	1.000	0.000
TAGA3	0.182	0.149
BIASA0	285.806	6926.542
BIASA1	-2.321	0.735
BIASA2	-2.726	0.754
BIASA3	1.003	0.909

Figure 4: Table 1 Corr = 0, Table 2 Corr = 0.5, Table 3 Corr = 0.9

3.5 $\beta_1 = 2$, $\beta_2 = 20$, $\beta_3 = 0$

3.5.1 $\mu_1 = 10$, $\mu_2 = 20$

In this scenario we observe the case with a large discrepancy in betas. There does not appear to be any difference in power or Type I Error for the betas as we vary the correlation between the measurement errors. However, as we begin to analyze the power of the test for β_1, we see that there exhibits a parabola-like pattern as we increase the measurement error correlation, going from 28.7% to 4.73% back up to 26.15%. However for Type I Error, we see that as we increase the correlation between errors, we see the increase in Type I Error from 10.13% to a high of 17.33%. In terms of biases, if we accumulate the absolute values of all the biases excluding the intercepts, we see that as we increase the correlation, the more total "absolute" bias we get. Thus we can clearly state, that as we increase the correlation that the Type I Error and Bias increase. However, unlike the cases we examined previously, the power of the tests for the slopes of the first and second linear predictors do not exactly decrease as we decrease measurement error correlation. The variance of TAGA1 across

BETA0 = 10 BETA1 = 2 BETA2=20 BETA3=0
MU1 = 10, MU2 = 20, MU3 = 10

U12 = 0

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.065	0.061
TAGB1	0.291	0.206
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.104	937.044
BIASB1	0.011	1.994
BIASB2	0.000	2.091
BIASB3	0.005	2.031
TAGA0	0.925	0.069
TAGA1	0.287	0.205
TAGA2	1.000	0.000
TAGA3	0.101	0.091
BIASA0	110.721	1216.604
BIASA1	0.095	2.296
BIASA2	-6.184	2.326
BIASA3	1.212	3.313

U12 = 0.5

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.065	0.061
TAGB1	0.291	0.206
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.104	937.044
BIASB1	0.011	1.994
BIASB2	0.000	2.091
BIASB3	0.005	2.031
TAGA0	0.967	0.032
TAGA1	0.047	0.045
TAGA2	1.000	0.000
TAGA3	0.138	0.119
BIASA0	123.528	1186.385
BIASA1	-2.109	2.479
BIASA2	-5.913	2.542
BIASA3	1.594	3.350

U12 = 0.9

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.065	0.061
TAGB1	0.291	0.206
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.104	937.044
BIASB1	0.011	1.994
BIASB2	0.000	2.091
BIASB3	0.005	2.031
TAGA0	0.976	0.024
TAGA1	0.262	0.193
TAGA2	1.000	0.000
TAGA3	0.173	0.143
BIASA0	125.061	1137.509
BIASA1	-4.170	2.655
BIASA2	-5.082	2.725
BIASA3	1.840	3.281

Figure 5: Table 1 Corr = 0, Table 2 Corr = 0.5, Table 3 Corr = 0.9

correlations also exhibits a similar parabolic pattern. We are alarmingly more sure of the rejections when correlation = 0.5, compared to the other correlation values, when in fact it is wrong the most, with a meager 4.73% accuracy.

3.5.2 $\mu_1 = 10$, $\mu_2 = 100$

Notice that aside from the bias of the intercepts, there does not appear to any difference between the powers, Type I Errors and biases after we changed the mean of the second predictor variable X2 from 20 to 50. Thus we exhibit a negligible effect of changes in mean for these beta values.

BETA0 = 10 BETA1 = 2 BETA2=20 BETA3=0
MU1 = 10. MU2 = 100. MU3 = 10

U12 = 0

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.054	0.051
TAGB1	0.291	0.206
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.139	19887.550
BIASB1	0.011	1.994
BIASB2	0.000	2.091
BIASB3	0.005	2.031
TAGA0	0.980	0.020
TAGA1	0.287	0.205
TAGA2	1.000	0.000
TAGA3	0.101	0.091
BIASA0	605.466	22448.883
BIASA1	0.095	2.296
BIASA2	-6.184	2.326
BIASA3	1.212	3.313

U12 = 0.5

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.054	0.051
TAGB1	0.291	0.206
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.139	19887.550
BIASB1	0.011	1.994
BIASB2	0.000	2.091
BIASB3	0.005	2.031
TAGA0	0.971	0.028
TAGA1	0.047	0.045
TAGA2	1.000	0.000
TAGA3	0.138	0.119
BIASA0	596.585	23897.135
BIASA1	-2.109	2.479
BIASA2	-5.913	2.542
BIASA3	1.594	3.350

U12 = 0.9

The MEANS Procedure

Variable	Mean	Variance
TAGB0	0.054	0.051
TAGB1	0.291	0.206
TAGB2	1.000	0.000
TAGB3	0.048	0.046
BIASB0	-0.139	19887.550
BIASB1	0.011	1.994
BIASB2	0.000	2.091
BIASB3	0.005	2.031
TAGA0	0.924	0.070
TAGA1	0.262	0.193
TAGA2	1.000	0.000
TAGA3	0.173	0.143
BIASA0	531.632	25039.830
BIASA1	-4.170	2.655
BIASA2	-5.082	2.725
BIASA3	1.840	3.281

Figure 6: Table 1 Corr = 0, Table 2 Corr = 0.5, Table 3 Corr = 0.9

3.6 Further Analysis

In summary of what we've just observed, I will list what we've learned from the 6 cases presented above:

- An increase in the discrepancy of the means (at least not an increase of 40 units) doesn't appear to have any significant/noticeable impact on the properties we are interested in.
- In general, as we increase the correlation between the measurement errors, we observe a decrease in the power of the tests for the slope of the first predictor variable, as well as an increase in Type I Error associated with the test for the slope of third predictor variable. In addition, the biases get larger with the change in correlation as well.

Because the change in the mean did not have a significant impact, we can instead focus strictly on the change in the values of beta, as well as the change in measurement error correlation. Also, because the accuracy of the power prediction for the slope of X2 is always 100% we need only consider the power of the test for the slope of X1, as well as the slope of X3. Just as a reminder, we are only considering TAGA1 and TAGA3, which represent the power and Type I Error for our model fitted with X_{obs1} and X_{obs2}.

3.6.1 Power for slope of X1

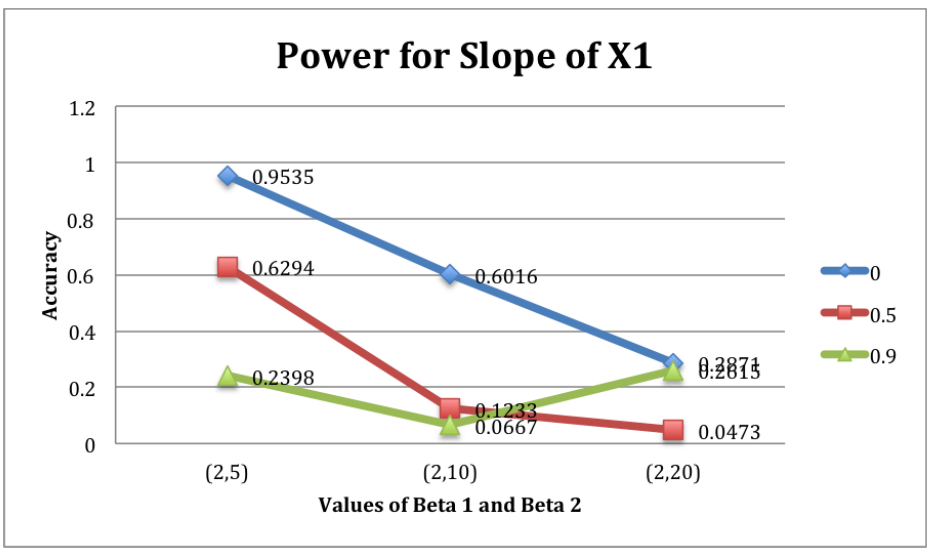

Figure 7: Power of the test for slope of X1

In the plot above we're able to see more clearly what was presented in the previous section(for the values of TAGA1). Again we see higher power for lower levels of measurement error correlation(in general). The values on the X axis were chosen in order to represent an increase in the discrepancy between values for β_1 and β_2. Thus we have (2,5) on the left, (2,10) in the middle and (2,20) on the right. By plotting the figure like this, we are able to see that there is a tendency for power to decrease as we increase the value of $u_{1,2}$ (with the exception of $u_{1,2}$ = 0.9, which drops but bounces back up).

3.6.2 Type I Error for slope of X3

In this plot, we can clearly see that as correlation increases, the Type I Error does as well. Moreover, we're able to see that Type I Error actually tends to drop as we increase the the discrepancy in betas, which is a good thing.

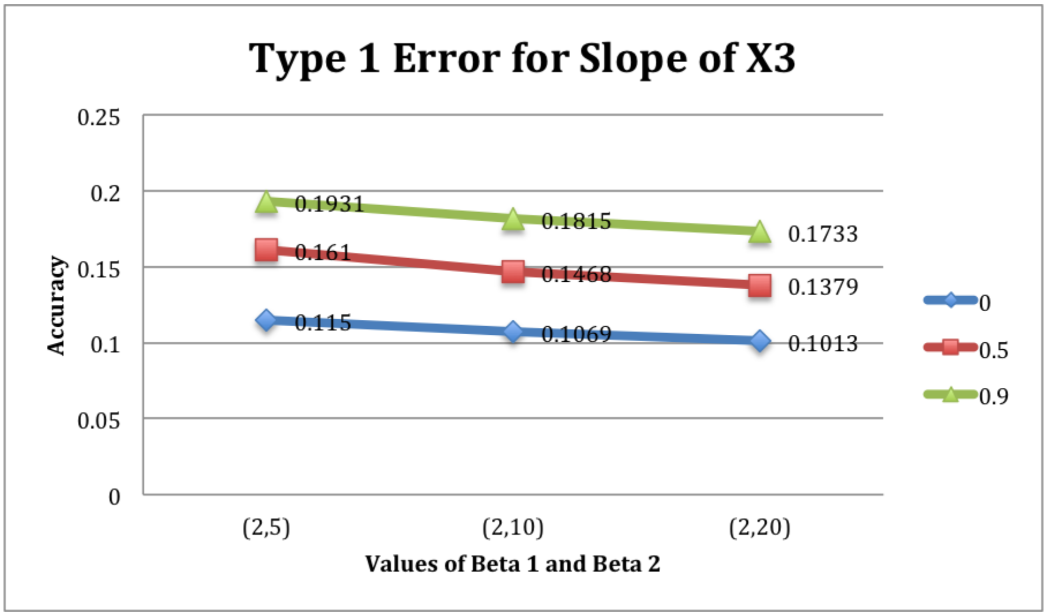

Figure 8: Type I Error of the test for slope of X3

3.6.3 Biases

In the plot below, we can clearly see that as we increase $u_{1,2}$, the total bias(sum of absolute values of BIAS1,BIAS2,BIAS3) also increase. Additionally, it appears that if we increase the discrepancy of the betas, that the sum of the absolute bias of our estimates actually increases. In other words, as we increase the value of β_2, the amount of bias exhibited by our estimates increases. Moreover, the increase seems more significant as we move from (2,10) to (2,20), compared to the increase from (2,5) to (2,20). Thus it appears as though larger increases in the value of β_2 will result in greater increases in the total bias.

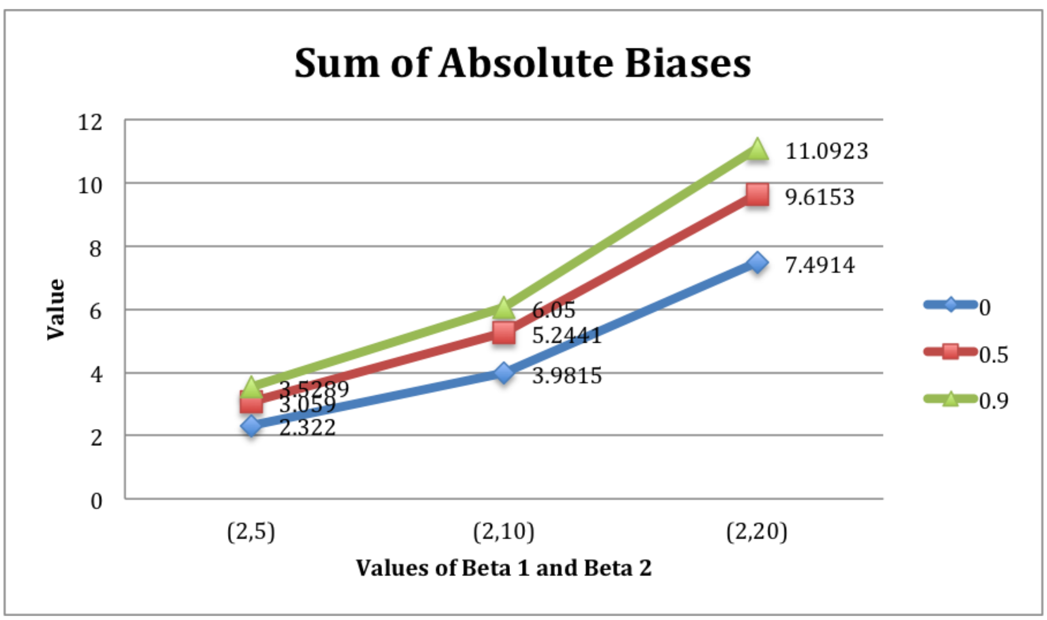

Figure 9: Biases

3.6.4 Section Summary

Overall, it appears as though increasing the correlation between the measurement errors u_1 and u_2 tend to introduce/increase multi-colinearity in the model; hence in general we observe poorer results in our characteristics of interest.

However the picture of how increasing discrepancy in betas is not as clear. In terms of Type 1 Error, we actually do see that it tends to decrease as the beta discrepancy enlarges. On the other hand, the power of the test of the slope of X_{obs1} tends to decrease as well, when we increase the value of β_2. Moreover, the sum of absolute biases is shown to increase as a result of increasing β_2. Thus we conclude that overall, increase the beta-discrepancy (ie: increasing β_2 results in poorer performance in power and biases, but better performance in the level of type I error.

4 Analysis of parameter adjustments on a 2-variable model

Now we will begin the section observing the effects of measurement error correlation, beta-discrepancy, and mean-discrepancy on the ability for us to recover true betas under the fitted model using predictor variables infused with measurement error.

4.1 Derivations

To begin, we will derive the equations that will be used in order to recover the true betas using estimates made under the regression model fitted under X_{obs}(we will refer to these as alphas).
In order to do this We must first assume that there exists a linear relationship between the predictor variables w/ measurement error, and the response variable Y (ie: $Y = \alpha_0 + \alpha_1 X_{obs1} + \alpha_2 X_{obs2} + \delta$, where δ is standard normal).

$$\begin{aligned}
cov(Y, X_{obs1}) &= cov(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon, X_1 + u_1) \\
&= \beta_1 \sigma_1^2 + \beta_2 \sigma_{1,2} \\
cov(Y, X_{obs1}) &= cov(\alpha_0 + \alpha_1 X_{obs1} + \alpha_2 X_{obs2} + \delta, X_{obs1}) \\
&= \alpha_1 \sigma_{obs1}^2 + \alpha_2 \sigma_{obs1,obs2} \\
&\implies \beta_1 \sigma_1^2 + \beta_2 \sigma_{1,2} = \alpha_1 \sigma_{obs1}^2 + \alpha_2 \sigma_{obs1,obs2}
\end{aligned}$$

by symmetry we also have the following:

$$\implies \beta_2 \sigma_2^2 + \beta_1 \sigma_{1,2} = \alpha_2 \sigma_{obs2}^2 + \alpha_1 \sigma_{obs1,obs2}$$

Thus we recover the following:

$$\begin{aligned}
\implies \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma 1,2 & \sigma_2^2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} &= \begin{bmatrix} \sigma_{obs1}^2 & \sigma_{obs1,obs2} \\ \sigma_{obs1,obs2} & \sigma_{obs2}^2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \\
\implies \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} &= \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma 1,2 & \sigma_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{obs1}^2 & \sigma_{obs1,obs2} \\ \sigma_{obs1,obs2} & \sigma_{obs2}^2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \\
\implies \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} &= \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma 1,2 & \sigma_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_1^2 + \sigma_{u1}^2 & \sigma_{1,2} + \sigma_{u1,u2} \\ \sigma_{1,2} + \sigma_{u1,u2} & \sigma_2^2 + \sigma_{u2}^2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}
\end{aligned}$$

Note, that if we interpret the ICC_1 and ICC_2 as the total variance in X_{obs1}, and X_{obs2} explained by X_1, and X_2 respectively, then we have:

$$ICC_1 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_{u1}^2} \implies \sigma_1^2 = ICC_1(\sigma_1^2 + \sigma_{u1}^2)$$

Equivalently:

$$ICC_2 = \frac{\sigma_2^2}{\sigma_2^2 + \sigma_{u2}^2} \implies \sigma_2^2 = ICC_2(\sigma_2^2 + \sigma_{u2}^2)$$

Aside: We can guarantee these ICC values by fixing the variances of measurement errors:

$$\begin{aligned}
\sigma_{u1}^2 &= \frac{1 - ICC_1}{ICC_1} \sigma_1^2 \\
\sigma_{u2}^2 &= \frac{1 - ICC_2}{ICC_2} \sigma_2^2
\end{aligned}$$

By using these substitutions, we can solve for betas in the following way:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} ICC_1(\sigma_1^2 + \sigma_{u1}^2) & \sigma_{1,2} \\ \sigma 1,2 & ICC_2(\sigma_2^2 + \sigma_{u2}^2) \end{bmatrix}^{-1} \begin{bmatrix} \sigma_1^2 + \sigma_{u1}^2 & \sigma_{1,2} + \sigma_{u1,u2} \\ \sigma_{1,2} + \sigma_{u1,u2} & \sigma_2^2 + \sigma_{u2}^2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$$

4.2 Results and Analysis

We will now take a look at the performance of this adjustment process in recovering the true betas. Again, we discovered that altering the mean of the second predictor (ie: adjusting the mean discrepancy) had no noticeable effect on the results, so we will only be using $\mu_1 = 10$ and $\mu_2 = 20$ for the rest of this section. We use the same settings used in section 3, and also iterate over the 3 settings of betas, and 3 levels of $u_{1,2}$. Also, we will always be referring to the values in the table involving A's and not B's (ie: when we refer to estimates of betas, we will be referring to A1,A2). This is because we are interested in the effects of these changes(in beta and $u_{1,2}$) on the model involving X_{obs} as the predictor variables, and how these changes affect the adjusted estimates of the true betas using $\hat{\alpha}$'s (e.g ADJA1,ADJA2) via the equations above.

4.2.1 β_1=2, β_2=5

BETA0 = 10, BETA1 = 2, BETA2=5, BETA3=0
MU1 = 10, MU2 = 20

U12=0

The MEANS Procedure

Variable	Mean
MB1	1.997
BIASB1	-0.003
MB2	5.003
BIASB2	0.003
MA1	1.591
BIASA1	-0.409
MA1ADJ	1.854
BIASA1ADJ	-0.146
MA1ADJT	1.983
BIASA1ADJT	-0.017
MA2	3.559
BIASA2	-1.441
MA2ADJ	5.141
BIASA2ADJ	0.141
MA2ADJT	5.006
BIASA2ADJT	0.006

U12=0.5

The MEANS Procedure

Variable	Mean
MB1	1.997
BIASB1	-0.003
MB2	5.003
BIASB2	0.003
MA1	1.082
BIASA1	-0.918
MA1ADJ	1.854
BIASA1ADJ	-0.146
MA1ADJT	1.981
BIASA1ADJT	-0.019
MA2	3.465
BIASA2	-1.535
MA2ADJ	5.141
BIASA2ADJ	0.141
MA2ADJT	5.003
BIASA2ADJT	0.003

U12=0.9

The MEANS Procedure

Variable	Mean
MB1	1.997
BIASB1	-0.003
MB2	5.003
BIASB2	0.003
MA1	0.647
BIASA1	-1.353
MA1ADJ	1.853
BIASA1ADJ	-0.147
MA1ADJT	1.981
BIASA1ADJT	-0.019
MA2	3.513
BIASA2	-1.487
MA2ADJ	5.142
BIASA2ADJ	0.142
MA2ADJT	5.001
BIASA2ADJT	0.001

Figure 10: $\beta_1 = 2$, $\beta_2 = 5$

Here we notice that when we introduce measurement error, that we constantly underestimate for betas (ie: MA1 < MB1, MA2 < MB2). In terms of the estimates for the slope (MA1), we notice that as we increase the correlation, that the estimates tend to become lower. However, for MA2, it exhibits a V-like pattern instead.

4.2.2 β_1=2, β_2=10

In this section we notice that we consistently underestimate the value of β_1 and β_2 (under the X_{obs} model). Moreover, the estimates MA1 tend to decrease as we increase the correlation, while the estimates MA2 tend to increase with correlation. Something we also notice, is that for once, when the correlation is 0, we see that the bias for A1 is actually slightly lower than the bias for ADJA1. This may be caused by poor estimates in variances/ICC_1.

BETA0 = 10, BETA1 = 2, BETA2=10, BETA3=0
MU1 = 10, MU2 = 20

	U12=0	U12=0.5	U12=0.9
	The MEANS Procedure	The MEANS Procedure	The MEANS Procedure
Variable	Mean	Mean	Mean
MB1	1.995	1.995	1.995
BIASB1	-0.005	-0.005	-0.005
MB2	10.006	10.006	10.006
BIASB2	0.006	0.006	0.006
MA1	1.796	0.728	-0.248
BIASA1	-0.204	-1.272	-2.248
MA1ADJ	1.780	1.779	1.777
BIASA1ADJ	-0.220	-0.221	-0.223
MA1ADJT	1.971	1.968	1.967
BIASA1ADJT	-0.029	-0.032	-0.033
MA2	7.032	7.067	7.378
BIASA2	-2.968	-2.933	-2.622
MA2ADJ	10.280	10.280	10.282
BIASA2ADJ	0.280	0.280	0.282
MA2ADJT	10.010	10.004	10.001
BIASA2ADJT	0.010	0.004	0.001

Figure 11: $\beta_1 = 2$, $\beta_2 = 10$

4.2.3 β_1=2, β_2=20

BETA0 = 10, BETA1 = 2, BETA2=20, BETA3=0
MU1 = 10, MU2 = 20

	U12=0	U12=0.5	U12=0.9
	The MEANS Procedure	The MEANS Procedure	The MEANS Procedure
Variable	Mean	Mean	Mean
MB1	1.989	1.989	1.989
BIASB1	-0.011	-0.011	-0.011
MB2	20.011	20.011	20.011
BIASB2	0.011	0.011	0.011
MA1	2.208	0.020	-2.038
BIASA1	0.208	-1.980	-4.038
MA1ADJ	1.634	1.630	1.623
BIASA1ADJ	-0.366	-0.370	-0.377
MA1ADJT	1.946	1.939	1.937
BIASA1ADJT	-0.054	-0.061	-0.063
MA2	13.977	14.271	15.107
BIASA2	-6.023	-5.729	-4.893
MA2ADJ	20.557	20.558	20.564
BIASA2ADJ	0.557	0.558	0.564
MA2ADJT	20.019	20.007	20.000
BIASA2ADJT	0.019	0.007	0.000

Figure 12: $\beta_1 = 2$, $\beta_2 = 20$

In this section we notice that the estimates for β_2 (under the X_{obs} model) are always lower than the true value of β_2, and are in fact increasing with correlation. On the other hand, MA1 values tend to decrease with correlation, initally over-estimating for β_1, but under-estimating for $u_{1,2}$ = 0.5, 0.9. We also observe that again, for correlation 0, that the bias for A1 is lower than the bias of the adjusted estimate. In fact, this time the bias is even greater(now 0.366 vs 0.220 for the previous case). Thus we see that as the value of β_2 increased, our ability to estimate the variances and ICC_1 got a little bit worse.

4.3 Further Analysis

Below we will perform a quick re-cap of what was observed above:

- In general, there was an under-estimation of the betas under the model fitted with measuremenet error.
- In general we noticed that there was a tendency for the estimates of β_2 under the X_{obs} model) to increase as we increased the correlation (w/ the exception of (2,5)).
- The estimates for β_1 always decrease as we increased $u_{1,2}$.
- The adjusted estimates of β_2 are always much better than the original estimates. The same can be said for β_1 with the exception.
- There are 2 cases, $\beta_2 = 10, 20$ and $u_{1,2}$ where the bias of the original estimate was indeed slightly lower than the adjusted estimate. We can possibly attribute this to poor estimates of σ_{u1}^2 and σ_1^2, which would lead to poor estimates of ICC_1, and thus adversely affecting ADJA1.

Below we have plots of the absolute values of biases under these different settings of betas and measurement error correlation:

4.3.1 Analyzing bias of ADJA1

First off we will analyze the absolute bias of ADJA1. ADJA1 represents the estimated value of β_1 using the method presented in section 4.1; thus $\hat{\beta_1}$ = ADJA1. In general we see that as we increase correlation, the bias tends to increase along with it (with the slight exception of (2,5)). Therefore as we increase the amount of correlation in the measurement errors, it becomes harder for us to actually estimate the true value of β_1 using the adjustment mechanism in section 4.1. Moreover, we see that as we increase the beta-discrepancy, the level of bias will also increase. Thus as the difference between the betas enlarges, recovering the actual value of β_1 using our estimates $\hat{\alpha}$ from the model under X_{obs} becomes more difficult.

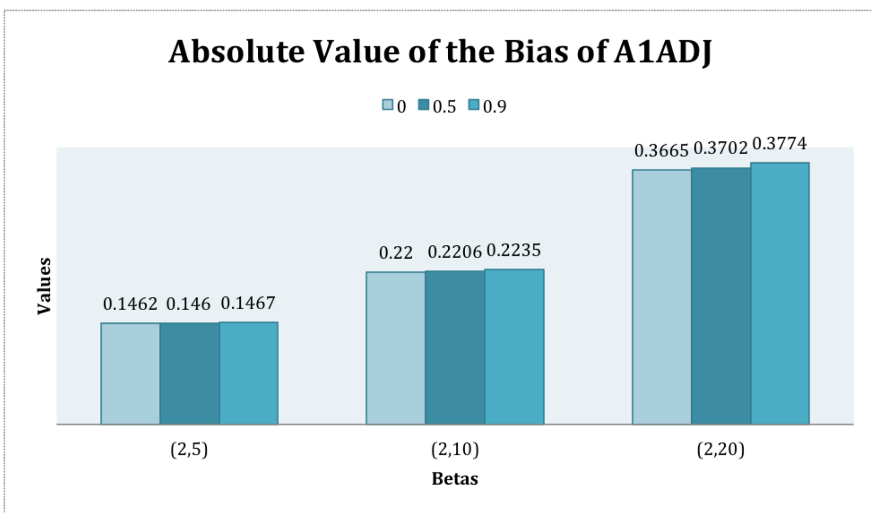

Figure 13: ADJA1 BIAS

4.3.2 Analyzing bias of ADJA2

Next we will take a look at the absolute bias of ADJA2. ADJA2 represents the estimated value of β_2 using the method presented in section 4.1; thus $\hat{\beta_2}$ = ADJA2. Here we notice that in all settings of the betas, the bias increases as we increase the measurement error correlation. Initially in section 3, we noticed that as we increased correlation, the bias of the estimates increased as well; thus it makes sense that the worse the estimates get, the adjustments we make on those estimates will also follow a similar trend. Next, we notice that within different settings of betas, the average value of the biases increase as well. We see that for the low setting(of beta discrepancy), the values hover around 0.14, while as we increase the discrepancy we get values around 0.28 and finally around 0.55 for the highest setting. Because we see a similar pattern above for ADJA1, it appears as though increasing the discrepancy in betas will make it more difficult for us to properly estimate and recover the true value of betas.

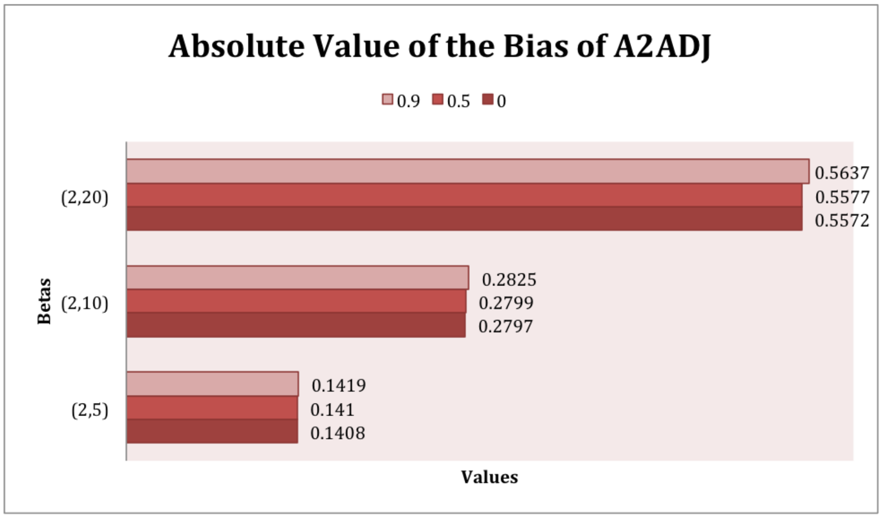

Figure 14: ADJA2 BIAS

4.3.3 Issues

Overall, we see that the adjustment process results in very good estimates of the true betas. In fact, the BIAS of A1 can get as bad as 4.04, and the BIAS of A2 can be as bad 6.02; When compared with the max biases of ADJA1 (0.377), and ADJA2 (0.56), we see that the adjustment process actually does a really good job in recovering the betas. However a key issue is the fact that we need to estimate:

$$\sigma_1, \sigma_2, \sigma_{u1}, \sigma_{u2}$$

Estimating these values will require repeated measurements of each predictor variable for each individual in the sample. In the analysis above, we simulated the scenario where to took 2 measurements of both X_1 and X_2 for each individual/rep. Even so, we observe that with just the one extra measurement, we're able to estimate the standard deviations with enough precision to make extremely good estimates of the true betas despite measurement error. In order to estimate these values, we will use the following information:

$$E(MSB) = m\sigma_x^2 + \sigma_u^2$$
$$E(MSW) = \sigma_u^2$$
$$\implies \sigma_x^2 = \frac{E(MSB) - E(MSW)}{m}$$

where m is the number of repeated measurements, n is the sample size and:

$$MSB = \frac{m\sum_{i=1}^{n}(\overline{X_i} - \overline{X})^2}{n-1}$$
$$MSW = \frac{\sum_{i=1}^{n}\left(\frac{\sum_{j=1}^{m}(X_{ij}-\overline{X_i})^2}{m-1}\right)}{n}$$

Thus assuming consistency of these estimators, we know that we can better get more accurate estimates of the sigmas when we increase our sample size. Currently we are dealing with a sample size of 100, however in the following section, we will look at the case when we halve this amount to n=50.

4.4 Reducing sample size to 50 from 100

In the previous section, we had 2 repeated measurements on 100 subjects in total. This was a valid number considering the fact that there were 100 people per group used as input to the regression. However, now we will consider the case when we try to estimate the ICC with fewer than 100 people, and see how the estimates react to such a change. We will try to estimate the ICC and recover true betas using n=50 samples, with 2 repeated measurements per predictor variable.

4.4.1 $\beta_1 = 2, \beta_2 = 5$

BETA0 = 10, BETA1 = 2, BETA2=5, BETA3=0
MU1 = 10, MU2 = 20

U12=0

The MEANS Procedure

Variable	Mean
MB1	1.997
BIASB1	-0.003
MB2	5.003
BIASB2	0.003
MA1	1.591
BIASA1	-0.409
MA1ADJ	1.852
BIASA1ADJ	-0.148
MA1ADJT	1.983
BIASA1ADJT	-0.017
MA2	3.559
BIASA2	-1.441
MA2ADJ	5.204
BIASA2ADJ	0.204
MA2ADJT	5.006
BIASA2ADJT	0.006

U12=0.5

The MEANS Procedure

Variable	Mean
MB1	1.997
BIASB1	-0.003
MB2	5.003
BIASB2	0.003
MA1	1.082
BIASA1	-0.918
MA1ADJ	1.852
BIASA1ADJ	-0.148
MA1ADJT	1.981
BIASA1ADJT	-0.019
MA2	3.465
BIASA2	-1.535
MA2ADJ	5.204
BIASA2ADJ	0.204
MA2ADJT	5.003
BIASA2ADJT	0.003

U12=0.9

The MEANS Procedure

Variable	Mean
MB1	1.997
BIASB1	-0.003
MB2	5.003
BIASB2	0.003
MA1	0.647
BIASA1	-1.353
MA1ADJ	1.851
BIASA1ADJ	-0.149
MA1ADJT	1.981
BIASA1ADJT	-0.019
MA2	3.513
BIASA2	-1.487
MA2ADJ	5.204
BIASA2ADJ	0.204
MA2ADJT	5.001
BIASA2ADJT	0.001

Figure 15: $\beta_1 = 2$, $\beta_2 = 5$

4.4.2 $\beta_1 = 2, \beta_2 = 10$

BETA0 = 10, BETA1 = 2, BETA2=10, BETA3=0
MU1 = 10, MU2 = 20

U12=0

The MEANS Procedure

Variable	Mean
MB1	1.995
BIASB1	-0.005
MB2	10.006
BIASB2	0.006
MA1	1.796
BIASA1	-0.204
MA1ADJ	1.761
BIASA1ADJ	-0.239
MA1ADJT	1.971
BIASA1ADJT	-0.029
MA2	7.032
BIASA2	-2.968
MA2ADJ	10.411
BIASA2ADJ	0.411
MA2ADJT	10.010
BIASA2ADJT	0.010

U12=0.5

The MEANS Procedure

Variable	Mean
MB1	1.995
BIASB1	-0.005
MB2	10.006
BIASB2	0.006
MA1	0.728
BIASA1	-1.272
MA1ADJ	1.759
BIASA1ADJ	-0.241
MA1ADJT	1.968
BIASA1ADJT	-0.032
MA2	7.067
BIASA2	-2.933
MA2ADJ	10.412
BIASA2ADJ	0.412
MA2ADJT	10.004
BIASA2ADJT	0.004

U12=0.9

The MEANS Procedure

Variable	Mean
MB1	1.995
BIASB1	-0.005
MB2	10.006
BIASB2	0.006
MA1	-0.248
BIASA1	-2.248
MA1ADJ	1.756
BIASA1ADJ	-0.244
MA1ADJT	1.967
BIASA1ADJT	-0.033
MA2	7.378
BIASA2	-2.622
MA2ADJ	10.414
BIASA2ADJ	0.414
MA2ADJT	10.001
BIASA2ADJT	0.001

Figure 16: $\beta_1 = 2$, $\beta_2 = 10$

4.4.3 $\beta_1 = 2$, $\beta_2 = 20$

BETA0 = 10, BETA1 = 2, BETA2=20, BETA3=0
MU1 = 10, MU2 = 20

U12=0

The MEANS Procedure

Variable	Mean
MB1	1.989
BIASB1	-0.011
MB2	20.011
BIASB2	0.011
MA1	2.208
BIASA1	0.208
MA1ADJ	1.578
BIASA1ADJ	-0.422
MA1ADJT	1.946
BIASA1ADJT	-0.054
MA2	13.977
BIASA2	-6.023
MA2ADJ	20.826
BIASA2ADJ	0.826
MA2ADJT	20.019
BIASA2ADJT	0.019

U12=0.5

The MEANS Procedure

Variable	Mean
MB1	1.989
BIASB1	-0.011
MB2	20.011
BIASB2	0.011
MA1	0.020
BIASA1	-1.980
MA1ADJ	1.573
BIASA1ADJ	-0.427
MA1ADJT	1.939
BIASA1ADJT	-0.061
MA2	14.271
BIASA2	-5.729
MA2ADJ	20.828
BIASA2ADJ	0.828
MA2ADJT	20.007
BIASA2ADJT	0.007

U12=0.9

The MEANS Procedure

Variable	Mean
MB1	1.989
BIASB1	-0.011
MB2	20.011
BIASB2	0.011
MA1	-2.038
BIASA1	-4.038
MA1ADJ	1.564
BIASA1ADJ	-0.436
MA1ADJT	1.937
BIASA1ADJT	-0.063
MA2	15.107
BIASA2	-4.893
MA2ADJ	20.833
BIASA2ADJ	0.833
MA2ADJT	20.000
BIASA2ADJT	0.000

Figure 17: $\beta_1 = 2$, $\beta_2 = 20$

4.4.4 Comparison

Below we will provide tables that will help us compare the effect of reducing the number the sample size, and how that affects our ability to recover the true betas.

(2,5)	BIASA1ADJ(50)	BIASA1ADJ(100)	BIASA2ADJ(50)	BIASA2ADJ(100)
0	-0.148	-0.146	0.204	0.141
0.5	-0.148	-0.146	0.204	0.141
0.9	-0.149	-0.147	0.204	0.142

Table 1: Comparison of adjusted estimates for β_1=2, β_2=5

(2,10)	BIASA1ADJ(50)	BIASA1ADJ(100)	BIASA2ADJ(50)	BIASA2ADJ(100)
0	-0.239	-0.220	0.411	0.280
0.5	-0.241	-0.221	0.412	0.280
0.9	-0.244	-0.223	0.414	0.282

Table 2: Comparison of adjusted estimates for β_1=2, β_2=10

(2,20)	BIASA1ADJ(50)	BIASA1ADJ(100)	BIASA2ADJ(50)	BIASA2ADJ(100)
0	-0.422	-0.366	0.826	0.557
0.5	-0.427	-0.370	0.828	0.558
0.9	-0.436	-0.377	0.833	0.564

Table 3: Comparison of adjusted estimates for β_1=2, β_2=20

We notice that in all scenarios, the magnitude of the biases are larger in the case where we use a smaller sample size. This makes sense of course because as we have fewer people, the more variable our estimates will be, and through a law of large numbers argument, the greater the sample size, the closer we are to the true value of ICC, and thus the true value of betas. Moreover, we can also quite easily see that as we increase the value of the correlation between measurement errors, the magnitude of the bias tends to increase. It's also easy to see that as we increase β_2, this results in a general increase in the amount of bias in our estimates.

5 Log-normal measurement errors

In this section, instead of having measurement error follow a joint-normal distribution with 0 mean, we instead let it follow a joint log-normal distribution. This will cover the case when the distribution of the measurement error is skewed(to the right). To implement this we randomly draw both u_1 and u_2 from a standard log-normal distribution, then we standardize both, and implement a correlation structure between the two (similar to what we did to define a multivariate normal distribution). We will then compare the results of section 4.2-4.3 which had measurement errors under a bivariate normal distribution, to what we have observe now, with the joint log-normal measurement errors. Note: We will again be using a sample size of 100 for the ICC estimates.

5.1 Results and observations

5.1.1 β_1=2, β_2=5

BETA0 = 10, BETA1 = 2, BETA2=5, BETA3=0
MU1 = 10, MU2 = 20

U12=0

The MEANS Procedure

Variable	Mean
MB1	1.997
BIASB1	-0.003
MB2	5.003
BIASB2	0.003
MA1	1.638
BIASA1	-0.362
MA1ADJ	1.888
BIASA1ADJ	-0.112
MA1ADJT	2.040
BIASA1ADJT	0.040
MA2	3.695
BIASA2	-1.305
MA2ADJ	5.347
BIASA2ADJ	0.347
MA2ADJT	5.198
BIASA2ADJT	0.198

U12=0.5

The MEANS Procedure

Variable	Mean
MB1	1.997
BIASB1	-0.003
MB2	5.003
BIASB2	0.003
MA1	1.166
BIASA1	-0.834
MA1ADJ	1.892
BIASA1ADJ	-0.108
MA1ADJT	2.115
BIASA1ADJT	0.115
MA2	3.591
BIASA2	-1.409
MA2ADJ	5.287
BIASA2ADJ	0.287
MA2ADJT	5.190
BIASA2ADJT	0.190

U12=0.9

The MEANS Procedure

Variable	Mean
MB1	1.997
BIASB1	-0.003
MB2	5.003
BIASB2	0.003
MA1	0.781
BIASA1	-1.219
MA1ADJ	1.874
BIASA1ADJ	-0.126
MA1ADJT	2.202
BIASA1ADJT	0.202
MA2	3.632
BIASA2	-1.368
MA2ADJ	5.298
BIASA2ADJ	0.298
MA2ADJT	5.206
BIASA2ADJT	0.206

Figure 18

Here we notice that the values of MA1 always decrease as we introduce more correlation, and thus we notice that the bias of our estimate of β_1 (under the model with X_{obs}) increases in magnitude with correlation. Alternatively, the values of MA2 follow a V-shape pattern, and the result is an inverted V-shape pattern for the bias. In addition, in all cases, we underestimate the values of both the slopes (ie:MA1 ¡ 2 and MA2 ¡ 5).

5.1.2 β_1=2, β_2=10

Again, we observe that on average the estimates of β_1 via the model under X_{obs} (ie: MA1), tends to drop as the level of correlation increases; Hence the bias for MA1 tends to increase as this occurs, because MA1 deviates further from the value of β_1 = 2. Next we observe that values of MA2 increase with correlation, and because they are moving towards β_2 = 10, we that the bias for this estimate decreases with correlation. Moreover, it appears to be the case that under the model with measurement error, we underestimate both slopes. The last thing we happen to notice is that our adjusted estimates always seem to underestimate β_1, but over-estimate β_2. In all cases, we underestimate

BETA0 = 10, BETA1 = 2, BETA2=10, BETA3=0
MU1 = 10, MU2 = 20

	U12=0	U12=0.5	U12=0.9
	The MEANS Procedure	The MEANS Procedure	The MEANS Procedure
Variable	**Mean**	**Mean**	**Mean**
MB1	1.995	1.995	1.995
BIASB1	-0.005	-0.005	-0.005
MB2	10.006	10.006	10.006
BIASB2	0.006	0.006	0.006
MA1	1.835	0.841	-0.024
BIASA1	-0.165	-1.159	-2.024
MA1ADJ	1.763	1.768	1.728
BIASA1ADJ	-0.237	-0.232	-0.272
MA1ADJT	2.002	2.158	2.337
BIASA1ADJT	0.002	0.158	0.337
MA2	7.307	7.308	7.583
BIASA2	-2.693	-2.692	-2.417
MA2ADJ	10.713	10.596	10.626
BIASA2ADJ	0.713	0.596	0.626
MA2ADJT	10.405	10.357	10.351
BIASA2ADJT	0.405	0.357	0.351

Figure 19

both β_1 and β_2, except for the case where β_2=20 and there's 0 correlation between measurement errors. Thus the over-estimation of β_2 under the X_{obs} model seems like it's a resultant of estimates of ICC and the correction process.

5.1.3 β_1=2, β_2=20

BETA0 = 10, BETA1 = 2, BETA2=20, BETA3=0
MU1 = 10, MU2 = 20

	U12=0	U12=0.5	U12=0.9
	The MEANS Procedure	The MEANS Procedure	The MEANS Procedure
Variable	**Mean**	**Mean**	**Mean**
MB1	1.989	1.989	1.989
BIASB1	-0.011	-0.011	-0.011
MB2	20.011	20.011	20.011
BIASB2	0.011	0.011	0.011
MA1	2.229	0.192	-1.634
BIASA1	0.229	-1.808	-3.634
MA1ADJ	1.512	1.519	1.437
BIASA1ADJ	-0.488	-0.481	-0.563
MA1ADJT	1.926	2.243	2.606
BIASA1ADJT	-0.074	0.243	0.606
MA2	14.532	14.743	15.484
BIASA2	-5.468	-5.257	-4.516
MA2ADJ	21.445	21.215	21.282
BIASA2ADJ	1.445	1.215	1.282
MA2ADJT	20.820	20.692	20.641
BIASA2ADJT	0.820	0.692	0.641

Figure 20

Once more we observe that on average the estimates of β_1 via the model under X_{obs} (ie: MA1), tends to drop as the level of correlation increases; Hence the magnitude of the bias for MA1 tends to increase as this occurs, because MA1 deviates further from the value of β_1 = 2.Next we observe that values of MA2 increase with correlation, and because they are moving towards β_2 = 10, we that the magnitude of the bias for this estimate decreases with correlation. Moreover, it appears to be the case that under the model with measurement error, we underestimate both slopes.

5.2 Plots and further analysis

5.2.1 ADJA1

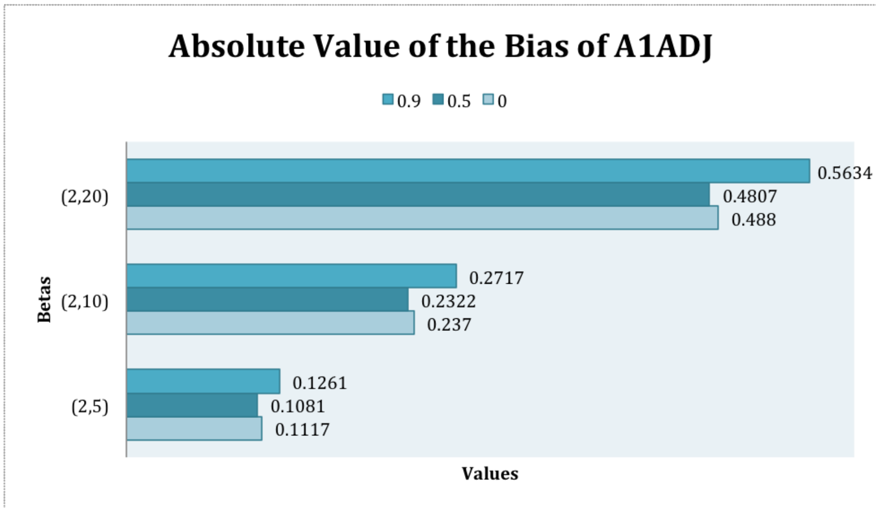

Figure 21

We observe that as we increase the discrepancy in betas, the magnitude of the bias of our adjusted estimate tends to become worse; note that this is the case for all values of $u_{1,2}$. We also notice a V-shaped pattern in how correlation tends to affect the magnitude of the bias of ADJA1. As we increase correlation from 0 to 0.5, we notice a slight decrease in bias, followed by a relatviely larger increase when $u_{1,2}$ becomes 0.9.

5.2.2 ADJA2

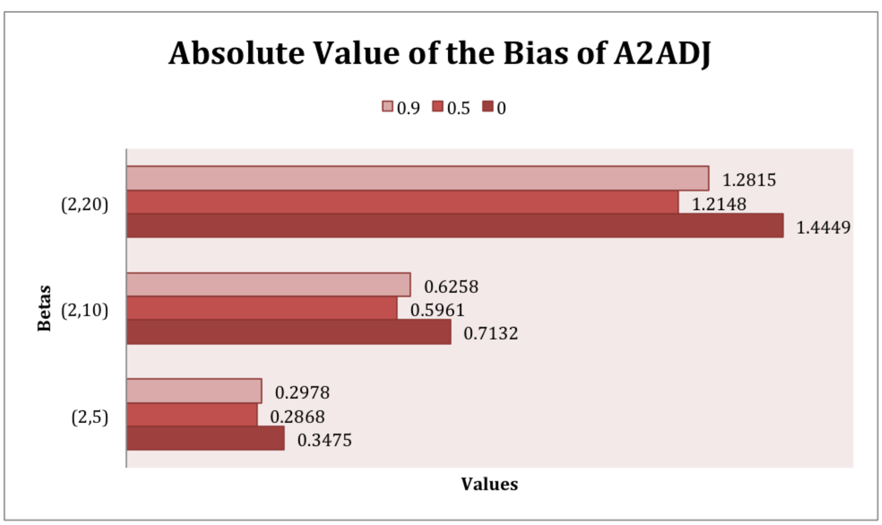

Figure 22

Again, just like ADJA1 we notice that as we increase the beta-discrepancy, that the bias of ADJA2 tends to increase. Thus our adjusted estimates of β_2 tend to become more incorrect as we increase β_2. Moreover, we see that again there is a V-shaped pattern in how the bias is affected by correlation. Instead however, we notice the largest value of bias for 0 correlation, followed by a large(relative) drop in bias for $u_{1,2}$=0.5, and finally a small increase as we move from 0.5 to 0.9.

5.2.3 Table Comparisons

Below we will list some tables comparing the bias of ADJA1, and ADJA2, when the sample size=100(which serves as our benchmark), and the scenario where each measurement error follows a log-normal distribution.

It is only the case for β_1=2, β_2=5 that the bias of ADJA1 (under the model with lognormal measurement error) has lower bias than the original case. In all other scenarios, we see that the bias for both adjusted estimates are much higher for the model with lognormal measurement error compared to the model with just normal measurement error. Again, we notice however, that the amount of bias (absolute value) exhibits a V-like pattern as we increase the correlation levels; Addtionaly, the level of bias tends to increase as we increase β_2 as well.

(2,5)	BIASA1ADJ(log)	BIASA1ADJ(100)	BIASA2ADJ(log)	BIASA2ADJ(100)
0	-0.112	-0.146	0.347	0.141
0.5	-0.108	-0.146	0.287	0.141
0.9	-0.126	-0.147	0.298	0.142

Table 4: Comparison of adjusted estimates for β_1=2, β_2=5

(2,10)	BIASA1ADJ(log)	BIASA1ADJ(100)	BIASA2ADJ(log)	BIASA2ADJ(100)
0	-0.237	-0.220	0.713	0.280
0.5	-0.232	-0.221	0.596	0.280
0.9	-0.272	-0.223	0.626	0.282

Table 5: Comparison of adjusted estimates for β_1=2, β_2=10

(2,20)	BIASA1ADJ(log)	BIASA1ADJ(100)	BIASA2ADJ(log)	BIASA2ADJ(100)
0	-0.488	-0.366	1.445	0.557
0.5	-0.481	-0.370	1.215	0.558
0.9	-0.563	-0.377	1.282	0.564

Table 6: Comparison of adjusted estimates for β_1=2, β_2=20

In addition, we can compare the results from the case where we reduced sample size to 50 with the results we observed from introducing log-normal measurement error. What we notice is that aside from the case for BIASA1ADJ, when $\beta_2 = 5$, log-measurement error causes higher bias values than reducing the sample size by half.

(2,5)	BIASA1ADJ(50)	BIASA1ADJ(log)	BIASA2ADJ(50)	BIASA2ADJ(log)
0	-0.148	-0.112	0.204	0.347
0.5	-0.148	-0.108	0.204	0.287
0.9	-0.149	-0.126	0.204	0.298

Table 7: Comparison between n=50 and log-normal measurement errors

(2,10)	BIASA1ADJ(50)	BIASA1ADJ(log)	BIASA2ADJ(50)	BIASA2ADJ(log)
0	-0.239	-0.237	0.411	0.713
0.5	-0.241	-0.232	0.412	0.596
0.9	-0.244	-0.272	0.414	0.626

Table 8: Comparison between n=50 and log-normal measurement errors

(2,20)	BIASA1ADJ(log)	BIASA1ADJ(50)	BIASA2ADJ(50)	BIASA2ADJ(log)
0	-0.488	-0.422	0.826	1.445
0.5	-0.481	-0.427	0.828	1.215
0.9	-0.563	-0.436	0.833	1.282

Table 9: Comparison between n=50 and log-normal measurement errors

5.3 Summary

Overall we can conclude a few things:

- As we increase the beta-discrepancy/the size of β_2, the the level of bias of our adjusted estimates tends to increase.
- There also appears to be higher levels of bias for more extreme levels of correlation (ie: 0, 0.9 vs 0.5)
- We consistently underestimate the values of the betas, when we use the model with measurement error
- Having skewed measurement errors, makes estimates less accurate, hence increasing the value of bias for all estimates.
- After adjustments, we notice that ADJA1 still under-estimate β_1, while ADJA2 over-estimate β_2
- MA1 tends to decrease as we increase correlations, and thus making the bias of A1 higher
- MA2 tends to increase as increase correlation, thus making the bias lower
- Log-Measurement error seems to make biases worse than reducing the sample size by half.