Invalid Performance Evaluation of Imputation Methods: A Warning Against Using the (Root) Mean Squared Error

E P Swens

July 1, 2021

Abstract

This abstract will be added 29 Jun

1 Introduction

Almost all studies involve missing data. For instance, non-response in surveys. Even in well-designed and controlled studies, patients can be lost to follow-up before the end of the study. In a study by Leurent et al. (2018), only 14 out of 47 randomized controlled trials reported a proportion of complete cases of more than 75%.

Missing data presents various problems. Firstly, missing observations reduce the sample size, which limits statistical power. Secondly, samples can lose representativeness if there is a disparity between responders and non-responders. Thirdly, missing data may complicate the analysis of the study. For example, it can bias the estimation of parameters. Any of these problems may challenge the validity of studies (Kang, 2013).

To deal with missing values, practitioners rely on the following methods: case deletion, weighting, model-based, and imputation-based procedures (Schafer, 1997; Little and Rubin, 2019). While the first method involves omitting incomplete records, the other procedures fill in missing values. In this paper, we focus on the fourth procedure: imputation.

Imputation is a universal term for filling in missing data with substituted values; it is often a data preprocessing step before the data analysis. It is a favored method because it provides completed data. Therefore, inferences can be obtained from all sample cases. However, improper imputation could lead to systematic errors in the data analysis, introducing biased and confidence invalid estimates (Murray et al., 2018; Rubin, 1987). Common causes of improper imputation include: omitting important variables from the imputation procedure, dealing with non-normally distributed variables, and violation of the missing data mechanism assumption (Sterne et al., 2009).

We believe another common pitfall is disregarded: invalid performance evaluation of imputation methods. To elaborate, many authors have adopted a method that aims to best recover the original values (Xu et al., 2020; Jadhav et al., 2019; Friedjungová et al., 2019). This method can mathematically be defined as the (root) mean squared error of the imputed values

$$RMSE = \sqrt{\frac{1}{n_{mis}} \sum_{n=1}^{n_{mis}} (y_i - \dot{y}_i)^2} \tag{1}$$

where $y_i - \dot{y}_i$ represents the error between the true value and imputed value of the i-th record. The errors are averaged over the number of missing records (n_{mis}). For the general case, the minimum RMSE is achieved by predicting the missing y_i by the linear model with the regression weights set to their least squares estimates, otherwise known as regression imputation (Van Buuren, 2018).

Unfortunately, regression imputation does not consider an error term in the imputation model, as the imputed values fit perfectly along the regression line without any residual variance. This method artificially strengthens the relations in the data, suggesting greater precision than is warranted (Little and Rubin, 2019). Furthermore, this method could lead to biased parameter estimates (Schafer and Graham, 2002). Whereas, the goal of imputation should be to produce estimates that are unbiased and confidence valid (Rubin, 1996). This evidence suggests that the RMSE is an invalid performance measure of imputation methods, as it appears to favor regression imputation.

In this study, we investigate the validity of the RMSE as a performance measure of imputation methods. We will assess the validity by comparing the quality of imputation methods. Following Rubin's (1996) goal of imputation, we will determine the imputation quality based on the bias and coverage rate of estimates. Intuitively, the RMSE favors the method with the best prediction accuracy. However, we will empirically demonstrate that this does not necessarily select an unbiased and confidence valid imputation procedure. First, we will briefly discuss imputation methods and missing data mechanisms. Then, we will outline our simulation study and present our findings in the second and third sections. In the fourth section, we discuss the implications of our findings.

1.1 Imputation Methods

Over the last decades, there have been considerable developments in general statistical methods for missing data (Kim and Shao, 2013; Little and Rubin, 2019). For instance, the `MICE` (Buuren and Groothuis-Oudshoorn, 2010) package in `R` provides various imputation methods for inferences in a wide range of missing-data problems. In this section, we will briefly discuss the imputation models applied in our simulation study.

Table 1 provides an overview imputation models. For more detailed information, see Flexible Imputation of Missing Data (Van Buuren, 2018).

Table 1: Description of imputation methods applied in the simulation study

Imputation Method	**Description**	**№ Imputations**
Mean	Missing values are replaced with the observed mean for that variable	$m = 1$
Regression	Missing values are replaced by a linear regression using observed data as explanatory variables	$m = 1$
Stochastic	Similar to regression imputation with additional noise based on the residual error of the regression model	$m \geq 1$
Bayesian	Missing values are replaced by Bayesian linear regression. Bayesian methods draw the parameters directly from their posterior distributions, introducing parameter uncertainty	$m \geq 1$
Predictive mean matching (pmm)	Missing values are replaced by a donor that has predicted values closest to the predicted value for the missing entry	$m \geq 1$

1.1.1 Single & Multiple Imputation

Stochastic imputation, Bayesian imputation, and predictive mean matching (pmm) incorporate uncertainty in the imputation prediction. Therefore, these methods can create $m > 1$ varying complete datasets, known as multiple imputation. To obtain a single estimate plus standard error the m datasets are pooled according to Rubin's (1987) rules. This approach introduces between-imputation variance and generally leads to better estimates of the standard error (Van Buuren, 2018).

In contrast, mean and regression imputation are not stochastic processes. Therefore, a single imputation $m = 1$ yields the same result regardless of the number of imputations.

1.2 Missing Data Mechanisms

The reasons for missing data can differ between studies. Consider an $n \times p$ matrix containing the dataset Y with p variables and n cases. Then, we introduce an $n \times p$ binary matrix R; its values, either a 0 or 1, represent whether an observation is missing or not. Here r_{ij} and y_{ij} denote the elements of R and Y, respectively, where $i = 1, \ldots, n$ and $j = 1, \ldots, p$. The set of all missing values, Y_{mis}, contains all elements y_{ij} where $r_{ij} = 0$. Conversely, the set of observed records, Y_{obs}, contains all elements y_{ij} where $r_{ij} = 1$. Combined $Y = (Y_{obs}, Y_{mis})$ contains the complete data, however in practice Y_{mis} if often unknown (Van Buuren, 2018; Rubin, 1987).

In theory, every data point has a probability of being an element of Y_{obs} or Y_{mis}. Rubin (1976) introduced three missing data mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). When observations are MCAR, the probability of missingness

$$P(R = 0 \,|\, \psi) \tag{2}$$

does not depend on any observed or missing variables. Thus, the probability of being missing only depends on ψ, the overall probability of being missing. Whereas, MAR occurs when

$$P(R = 0 \,|\, Y_{obs}, \psi) \tag{3}$$

so the missingness is conditional on other observed variables. Finally, MNAR implies that

$$P(R = 0 \,|\, Y_{obs}, Y_{mis}, \psi) \tag{4}$$

so the probability of a value being missing depends on unknown information.

The information from Y_{obs} can give some evidence if Y_{mis} is MCAR. However, from Y_{obs} alone, it is impossible to evaluate whether Y_{mis} is MAR or MNAR. Generally, statistical software implements imputation methods under the assumption of MCAR or MAR (Donders et al., 2006).

2 Methods

To investigate the validity of the RMSE as performance evaluation of imputation methods, we performed a model-based simulation in `R` (R Core Team, 2020). To this end, samples are drawn from a known probability distribution. To reflect reality, the Pima Indians Diabetes Database formed the motivating example for generating the simulated datasets; it contains diagnostic criteria for diabetes from a group of Native Americans (Smith et al., 1988). For this simulation study, we focus on the relations between the continuous values: Blood Pressure (X_1) Insulin (X_2), and BMI (Y). Figure 1 summarizes the design of the experiment.

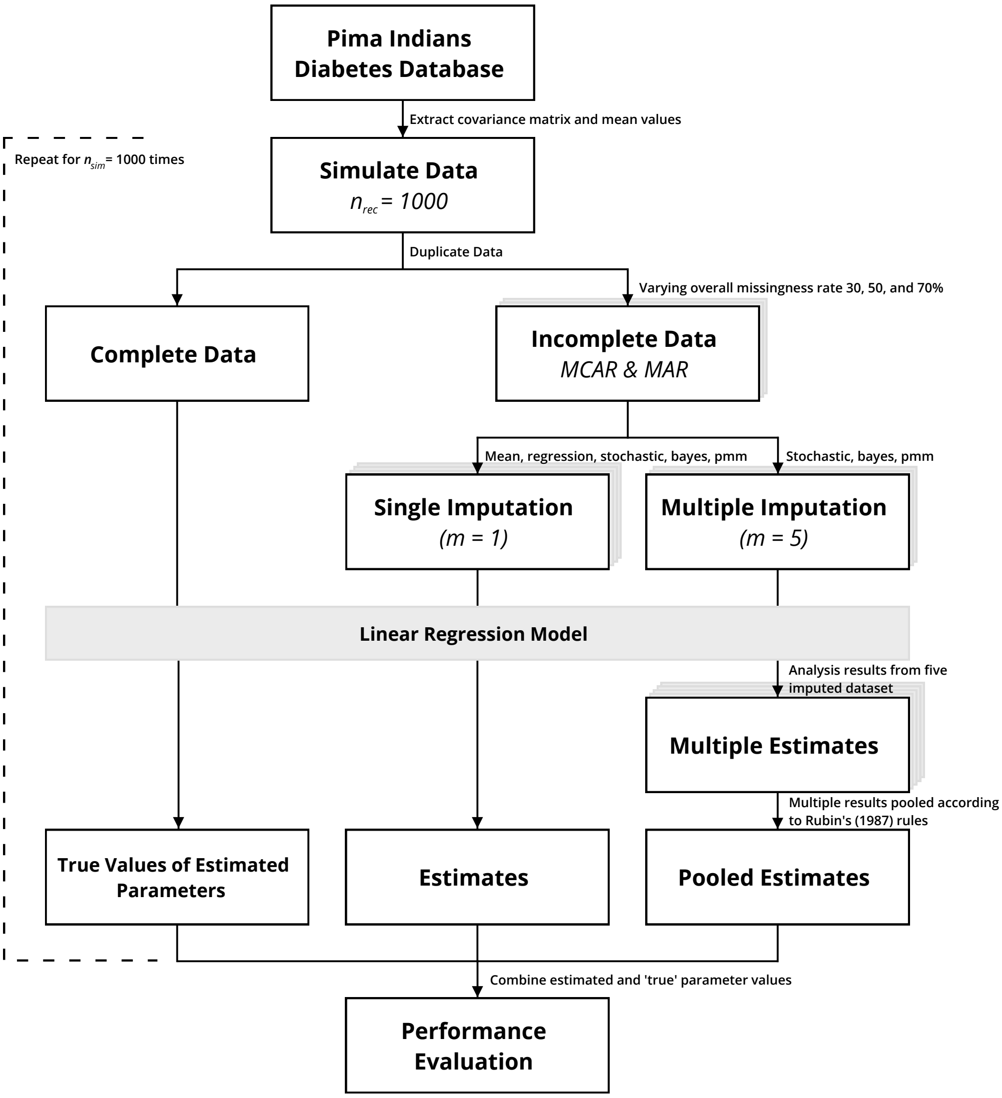

Figure 1: Design of simulation study

2.1 Data Generation

Data ($n_{rec} = 1000$) was drawn from a multivariate normal distribution, $\mathcal{N}(\mu, \Sigma)$, using `mvrnorm` from the `MASS` package (Ripley et al., 2013). Where the mean vector consisted of

$$\mu = \begin{bmatrix} 70 \\ 5 \\ 35 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} 155 & 1 & 1.5 \\ 1 & 0.3 & 25 \\ 1.5 & 25 & 50 \end{bmatrix}$$

represented the co-variance matrix. Subsequently, a log transformation was applied on X_2, introducing a positive skewness of $(e^{0.3} + 2)\sqrt{e^{0.3} - 1}$. Note, non-normality categorizes as one of the common pitfalls of imputations, which could lead to improper imputation.

2.2 Amputation Procedure

The amputation procedure (i.e. introducing missingness) involved all variables under MCAR and MAR mechanisms. Under MAR conditions, the type of missingness was kept constant at right-tailed missingness, i.e. data deletion predominantly occurred in the higher values. This is considered to be one of the more difficult missingness types, as the distribution of the observed data shifts to the left, introducing skewness in the probability distributions (Vink et al., 2014). For each mechanism, an overall missingness rate (λ) of 30, 50, and 70% was imposed to resemble intermediate, severe, and extreme missingness scenarios. Furthermore, three missing data patterns were drawn from all possible combinations of patterns, excluding trivial ones (complete missingness and complete cases). Missing values were generated with the multivariate amputation function: `ampute` (Schouten et al., 2018).

2.3 Imputation Procedure

We performed the imputation procedure with `MICE` (Buuren and Groothuis-Oudshoorn, 2010), applying the aforementioned imputation techniques (section 1.1). For multiple imputation techniques, five imputed datasets ($m = 5$) were combined into a single inference following Rubin's (1987) rules.

2.4 Regression Model

The evaluation of estimates of scientific interest comprised of a linear regression model in the form

$$X_3 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \tag{5}$$

where Insulin (X_2) remains log-normally distributed to investigate imputation methods' ability to deal with positive skewness. The main parameters of interest from the regression model include regression estimates, residual variance, and coefficient of determination. Supportive software packages for the analysis consisted of `tidyverse` (Wickham, 2017).

2.5 Evaluation

The imputation methods were evaluated in terms of RMSE, raw bias, coverage rate, and average confidence interval width. Firstly, as previously discussed, the RMSE evaluates imputation methods based on their performance to recover the original data. For all variables of interest, the RMSE was derived according to formula 1. Secondly, the raw bias for a given estimator was defined as a systematic deviation from the actual value

$$B_\theta(\hat{\theta}) = E(\hat{\theta}\,|X) - \theta \tag{6}$$

here $E(\hat{\theta}\,|X)$ represents the estimator's expected value over all possible samples for X, and the true value of the parameter being estimated is represented by θ. Thirdly, the coverage rate (or coverage probability) refers to the proportion of confidence intervals of an estimation that contain the true value:

$$P_\theta(\theta \in [L(X), U(X)]) \geq 1 - \alpha \tag{7}$$

where $[L(X), U(X)]$ is the 95% confidence interval of a parameter θ. A statistical test with a nominal rejection rate of $\alpha = 0.05$ rejects the null hypothesis at most 5% of the simulations when the null hypothesis is true (Van Buuren, 2018). Lastly, the average confidence interval is the average difference between the upper and lower bounds

$$E_\theta(U(X) - L(X)) \tag{8}$$

where ideally the length is as small as possible, however not so narrow that the coverage rate will decrease below the nominal level.

2.6 Parameters of Interest

Table 2 presents an overview of the parameters of interest and the accompanying evaluation measures. For all parameters, we determined the raw biased. For the means and regression coefficients, we extended our analysis to the coverage rate and average confidence interval width.

Table 2: Description of parameters of interest for the performance measures: †Raw bias, ‡Coverage rate ‡‡Average confidence interval width

Parameter of Interest		**Performance Measure**		
Name	Symbol	RB†	CR‡	CW‡‡
Means	μ	✓	✓	✓
Variances	σ^2	✓	-	-
Correlations	ρ	✓	-	-
Regression coefficients	β	✓	✓	✓
Coefficient of determination	R^2	✓	-	-
Residual variance	σ_e^2	✓	-	-

2.7 Number of simulation

A moderately independent simulation strategy (Burton et al., 2006) was adopted. Between simulations ($n_{sim} = 1000$), variance existed because of sampling and amputation differences. Whereas within simulations the data remained constant for all imputation methods. This approach strengthened comparisons between different methods as it eliminates any sampling and amputation variability.

2.8 Ethical Considerations

The Pima Indians Diabetes Database was provided under a General Public License (GNU, 1989). Data anonymization ensured privacy protection. Furthermore, the data application was limited to ascertain probability distribution parameters for the simulated data.

3 Results

The results provide insight in the validity of the RMSE as a performance evaluation of imputation methods. First, we compare the RMSE between imputation methods. Next, we evaluate the imputation quality in terms of raw bias, coverage rates, and average confidence width of estimates.

3.1 Root Mean Square Error

To assess the method that best recovers the true data, we studied the error between the original and imputed values. On average, over the 1000 simulations, regression imputation resulted in the lowest RMSE (Figure 2). This holds true for all variables: Blood Pressure (X_1), Insulin (X_2), and BMI (Y). With mean imputation, the second-best RMSE was achieved. The remaining imputation methods performed worse, their average RMSE exceeded the maximum RMSE of the prior methods. Simulations with higher missingness rate showed reduced variation around the average RMSE, whereas the average results remained constant (Appendix A). Overall, mean and regression imputation best recovered the original data according to the RMSE.

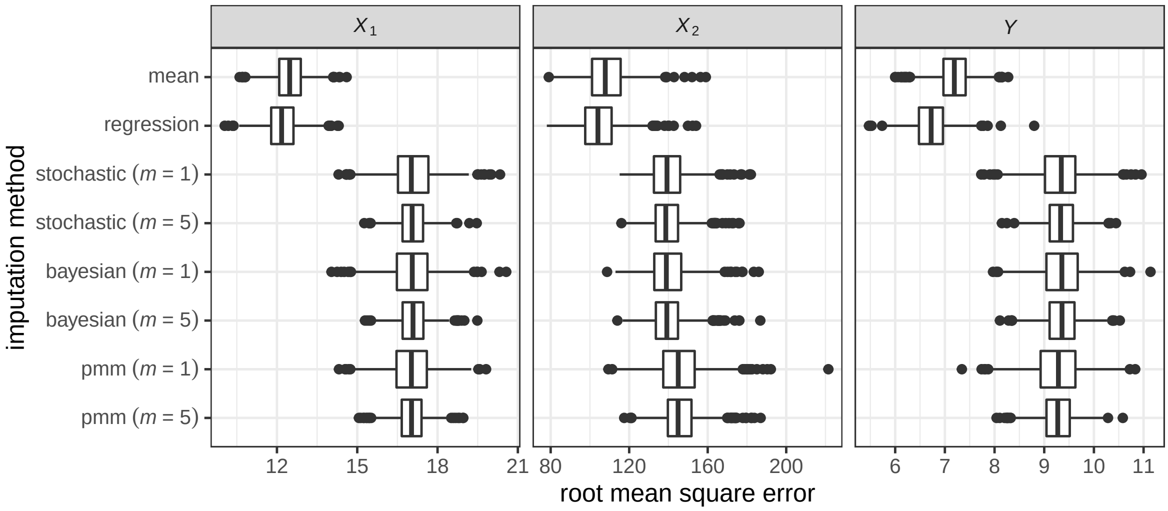

Figure 2: Root mean square error for imputation methods under MAR mechanism and 50% missingness rate

3.2 Raw Bias

We studied the raw bias to estimate the systematic deviation between an estimate and actual parameter value. Figure 3 presents the raw bias of the mean, variance, and correlation estimates for X_1, X_2 and Y under MAR and 50% missingness rate. Mean imputation underestimated all aforementioned parameters. Right-tailed missingness caused the mean to shift left. Furthermore, imputing the mean undervalued the variance and correlations. Regression imputation minimized the variance and distorted correlations between variables. Where mean imputation reduced the correlation between variables, regression imputation attuned correlations upwards. Stochastic and Bayesian imputations slightly undervalued and overvalued the variance of X_2 and Y, respectively. This also corresponded to an upwards bias for $\rho_{x2,x3}$. In contrast, predictive mean matching (pmm) provided unbiased estimates. Generally, the raw bias of estimates appeared not to be influenced by the number of imputations. As the missingness rate increased biases amplified in their respective direction (Appendix B). In general, mean and regression imputations severely yielded biased estimates.

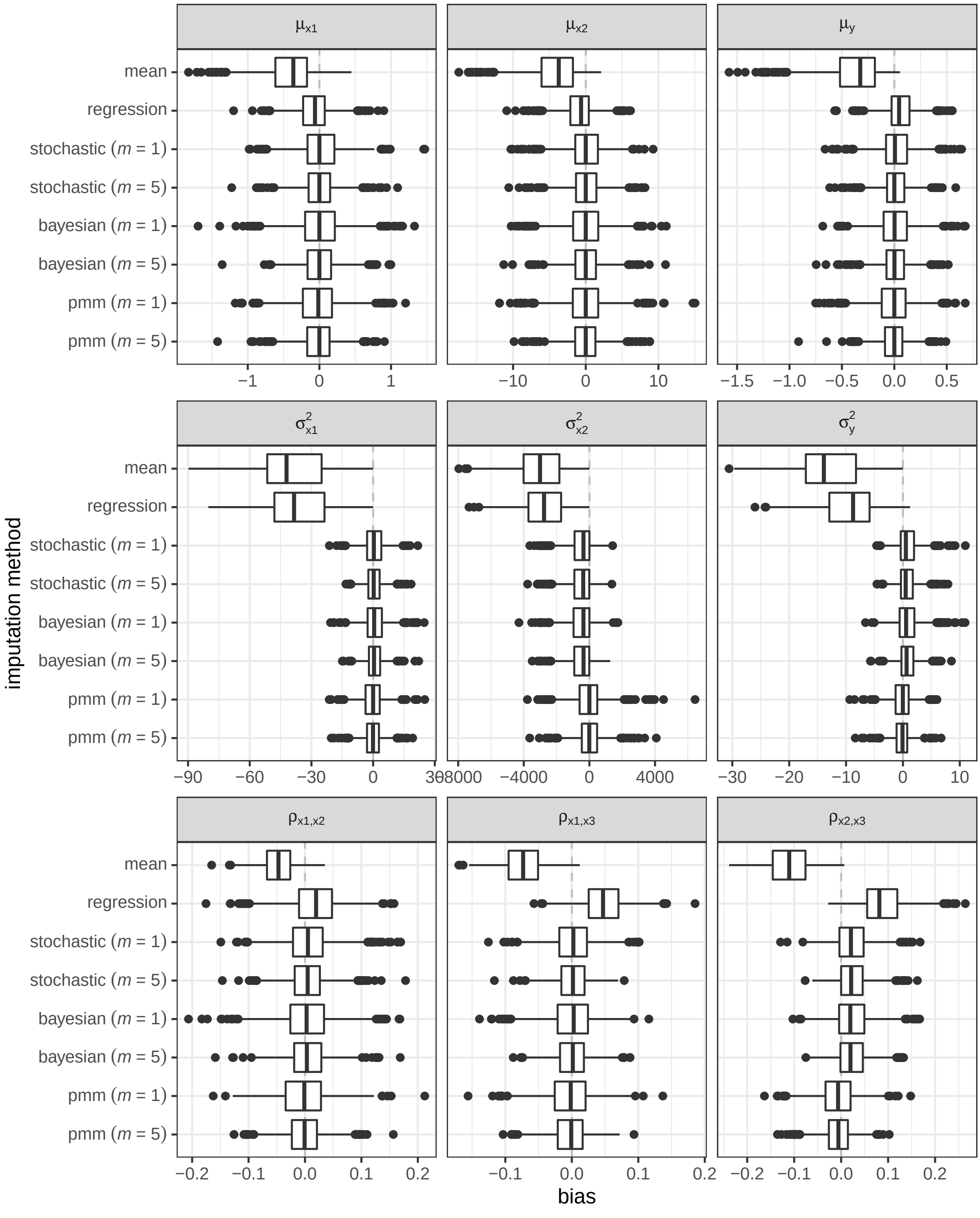

Figure 3: Mean, variance, and correlation bias for imputation methods under MAR mechanism and 50% missingness rate

Figure 4 displays the biases for the regression estimates under MAR for 50% missingness rate. Regression coefficients were severely biased under mean and regression imputation. Mean imputation undervalued the effect of covariates, while regression imputation overvalued the effect of covariates. These results are in accordance with the bias in the correlation estimates, as mean imputation weakened correlations and regression imputation strengthened correlations. Furthermore, these methods disturbed the estimates for the residual variance and R^2. Estimations for the residual variance from both mean and regression imputation suffered from upward bias. Regression imputation inflated the coefficient of determination. This means an overestimation in the proportion of the variance in Y that is predictable from X_1 and X_2. For mean imputation, we observed the opposite effect. Bayesian and stochastic regression biased the effect of X_2 upwards to the 25th percentile. Similarly, the estimations for R^2 were also attuned upwards. Only predictive mean matching resulted in unbiased regression estimates. A higher missingness rate increased the presented biases in their respective directions (Appendix B). Again, mean and regression imputation resulted in severely biased estimates.

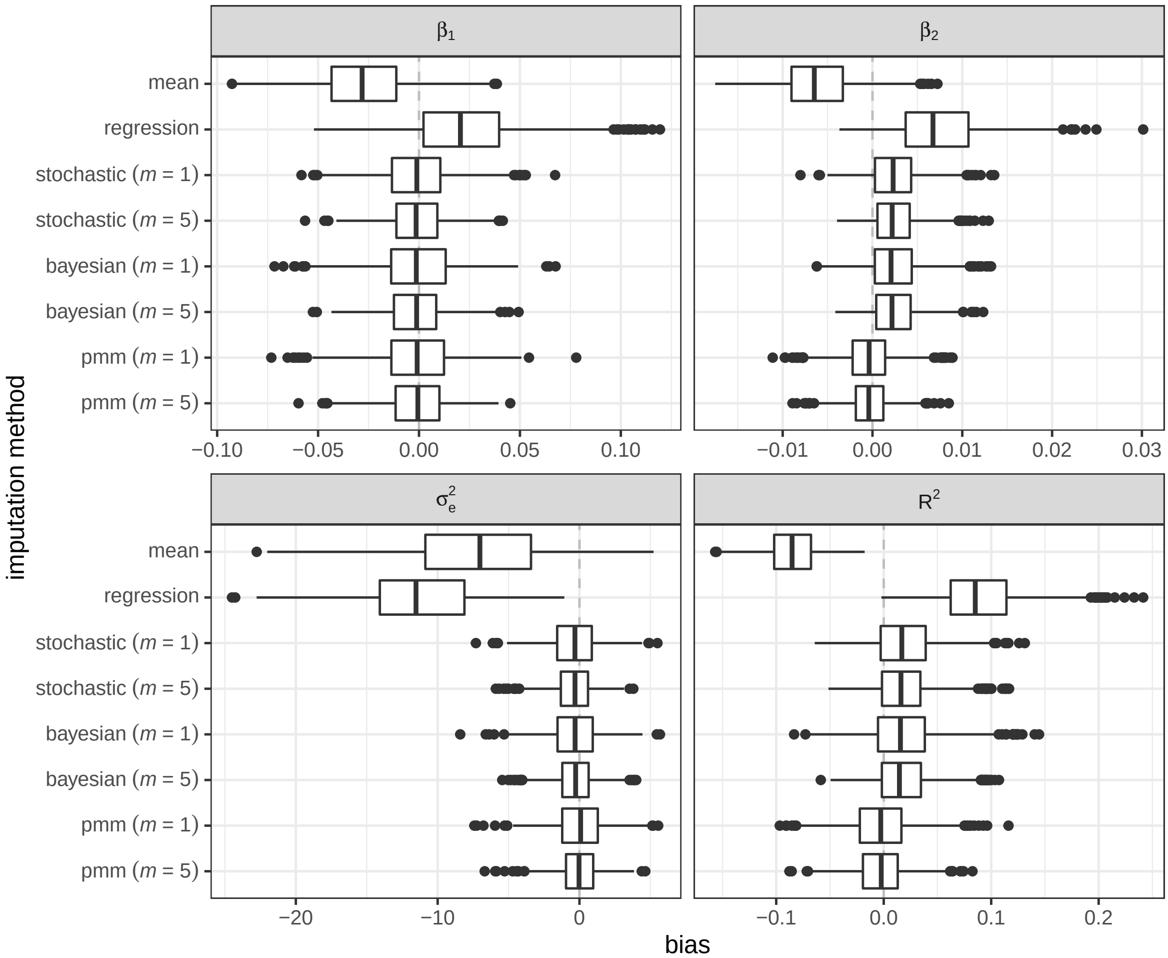

Figure 4: Linear regression estimates bias for imputation methods under MAR mechanism and 50% missingness rate

3.3 Confidence Rate

To investigate the validity of the standard error of estimates, we studied the proportion of confidence intervals of an estimation that contain the true value. Figure 6 presents the coverage rates (CR) under MAR for the missingness rates: 30, 50, and 70%. In general, a higher missingness rate negatively impacts the coverage rate. Between imputation methods, mean imputation resulted in coverage rates $\leq$ 95% for all missingness rates. In the most extreme missingness case, λ = 70%, we observed undercoverage for all single imputation methods. While multiple imputation methods consistently produced confidence valid results.

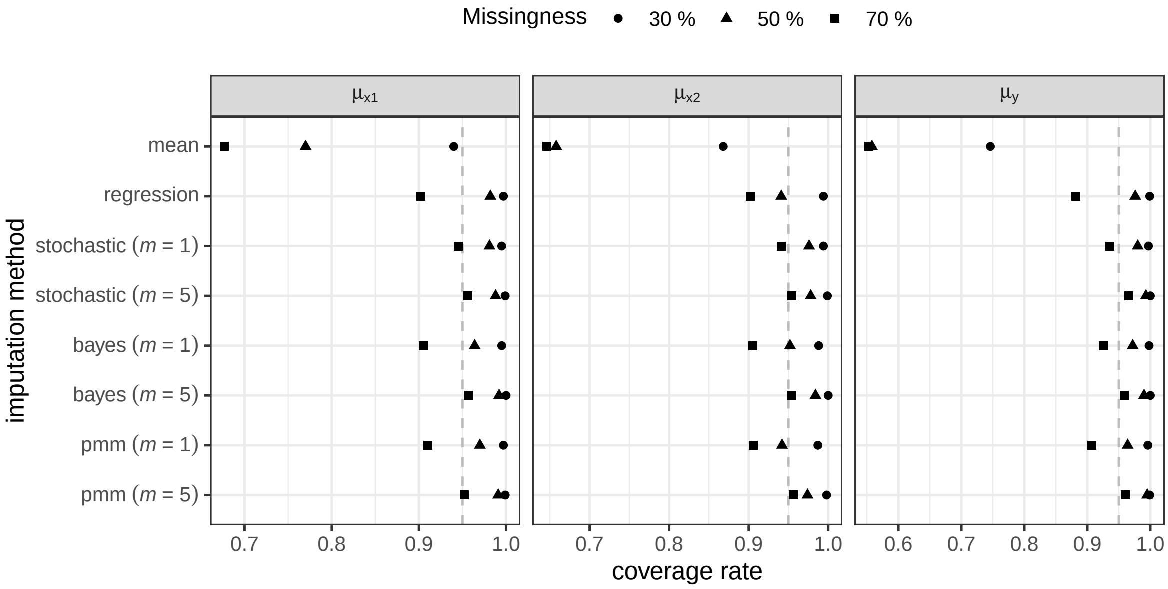

Figure 5: Coverage rate of the means for 30%, 50%, 70% missingness under MAR mechanism

For the regression estimates of Blood Pressure (β_1) and Insulin (β_2), mean and regression imputation are prone to undercoverage (Figure 6). The remaining single imputation methods only achieved confidence valid results for β_1 under $\lambda = 30\%$. The multiple imputation procedures obtained confidence valid results for β_1. Only predictive mean matching resulted in proper coverage for both regression estimates, except for a slight undercoverage of β_2 under severe missingness.

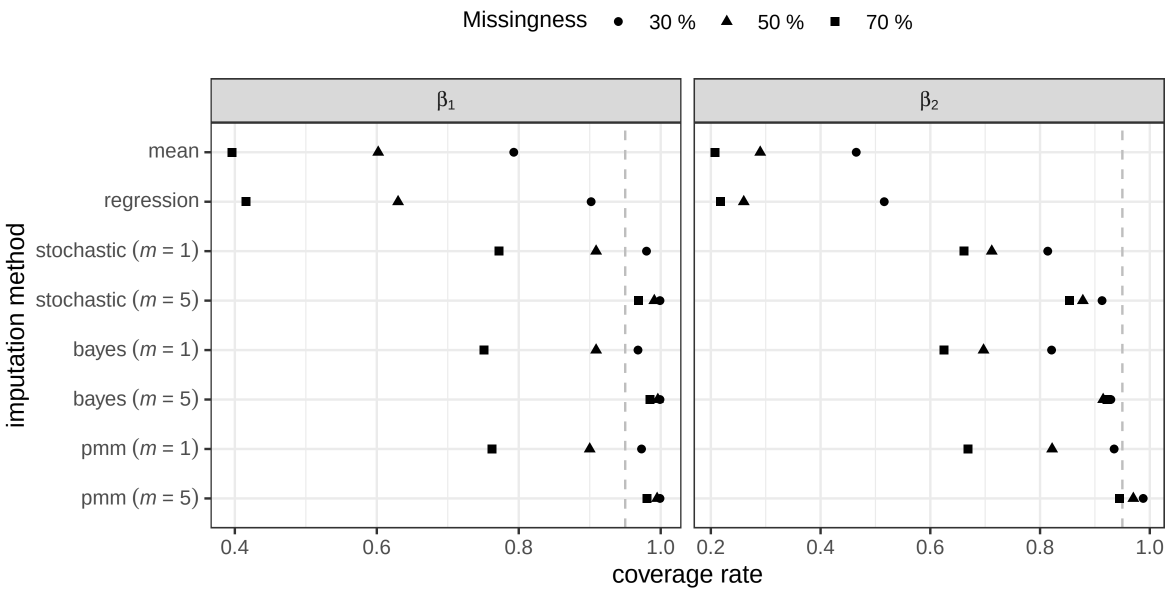

Figure 6: Regression estimates coverage rate for 30%, 50%, 70% missingness under MAR mechanism

To summarize, multiple imputation methods resulted in better coverage rates compared to single imputation methods. For more skewed data, predictive mean matching obtained the best results.

4 Discussion

The results question the validity of the RMSE as a performance evaluation of imputation methods. The simulation study empirically demonstrates that the RMSE favors biased and confidence invalid imputation procedures.

Regression imputation yielded the lowest RMSE. The results might suggest that regression imputation best recovered the original data. Still, it did not result in unbiased and confidence valid estimates. For instance, estimates of the correlations and study effects of Insulin and BMI were severely biased upwards. Yielding greater precision than was warranted. Results indicate a tendency towards a type I error in the analysis (a false-positive conclusion). This corresponds with the $\leq 95\%$ coverage rates for all missingness rates: $0.3, 0.5, 0.7$. Theoretical work from Little (1992) and Little and Rubin (2019) support our findings.

Mean imputation resulted in the second-best RMSE. In contrast to regression imputation, mean imputation led to biased downwards estimations of the correlations and study effects. This method appears prone to a type II error in the analysis (a false-negative conclusion). Similarly, coverage rates were $\leq 95\%$ for all missingness rates. The outcomes are in line with the findings of Donders et al. (2006) and Van Buuren (2018).

The average RMSE of the remaining imputation methods exceeded the maximum RMSE of the prior methods, suggesting that stochastic imputation, Bayesian imputation, and predictive mean matching recovered the original data poorly. In contrast, the overall estimators were unbiased and confidence valid, apart from estimates involving Insuline (X_2). The distribution of X_2 was positively skewed. Stochastic and Bayesian regression resulted in biased upward estimators for the correlation and study effects of X_2. This outcome exhibits a limitation of stochastic and Bayesian regression as both methods assume the data follows a normal distribution. In contrast, predictive mean matching does not consider an explicit model for the distribution of the missing values (Van Buuren, 2018). For this reason, predictive mean matching yielded unbiased results. As for the confidence intervals, multiple imputation ($m = 5$) resulted in higher coverage rates compared to single imputation ($m = 1$). Rubin's (1996) rules introduce additional between-imputation variance. This extra variance resulted in wider confidence intervals, which in turn resulted in better coverage rates.

4.1 Limitations

Potential limitations of this study include data generation conditions, missingness generation assumptions, and performance evaluation. Firstly, in this study, we confined ourselves to continuous data. Therefore, restricting the generalizability of our results across data types. Nevertheless, the data generation process aimed to resemble real-life data by taking the Pima Indians Diabetes Database as an example. Secondly, the missingness mechanism was limited to MCAR and MAR. It can be argued that MNAR is the more likely mechanism for real-life missingness scenarios. Although, a performance evaluation of imputation methods should still favor proper imputation methods under MCAR and MAR. Thirdly, we restricted the performance evaluated to the bias, coverage rate, and average confidence interval width. Vink (2016) proposed three additional evaluation metrics: distributional characteristics, the plausibility of the imputed values, and convergence of the algorithm. These metrics could further support our claims. However, based on the findings of the biases and coverage rates, we can still maintain our claims about the RMSE.

4.2 Implications & Recommendations

Empirically we demonstrated that the RMSE is not a valid evaluation technique of imputation methods. This result also casts doubt on the validity of the conclusion from preceding studies adopting the RMSE.

If mean,

If regression,

Besides, it raises the question: "how to properly evaluate imputation procedures?". To reiterate the goal of imputation from Rubin (YEAR), producing estimates that are unbiased and confidence valid. Therefore, the quality of the imputation method should be evaluated with respect to this goal.

References

Leurent, B., Gomes, M., & Carpenter, J. R. (2018). Missing data in trial-based cost-effectiveness analysis: An incomplete journey. *Health economics*, *27*(6), 1024–1040.

Kang, H. (2013). The prevention and handling of the missing data. *Korean journal of anesthesiology*, *64*(5), 402.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. CRC press.

Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793). John Wiley & Sons.

Murray, J. S. et al. (2018). Multiple imputation: A review of practical and theoretical findings. *Statistical Science*, *33*(2), 142–159.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *Bmj*, *338*.

Xu, X., Xia, L., Zhang, Q., Wu, S., Wu, M., & Liu, H. (2020). The ability of different imputation methods for missing values in mental measurement questionnaires. *BMC medical research methodology*, *20*(1), 1–9.

Jadhav, A., Pramod, D., & Ramanathan, K. (2019). Comparison of performance of data imputation methods for numeric dataset. *Applied Artificial Intelligence*, *33*(10), 913–933.

Friedjungová, M., Jiřina, M., & Vašata, D. (2019). Missing features reconstruction and its impact on classification accuracy. *International Conference on Computational Science*, 207–220.

Van Buuren, S. (2018). *Flexible imputation of missing data*. CRC press.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological methods*, *7*(2), 147.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American statistical Association*, *91*(434), 473–489.

Kim, J. K., & Shao, J. (2013). *Statistical methods for handling incomplete data*. CRC press.

Buuren, S. v., & Groothuis-Oudshoorn, K. (2010). Mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, 1–68.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.

Donders, A. R. T., Van Der Heijden, G. J., Stijnen, T., & Moons, K. G. (2006). A gentle introduction to imputation of missing values. *Journal of clinical epidemiology*, *59*(10), 1087–1091.

R Core Team, B. (2020). R: A language and environment for statistical computing.

Smith, J. W., Everhart, J. E., Dickson, W., Knowler, W. C., & Johannes, R. S. (1988). Using the adap learning algorithm to forecast the onset of diabetes mellitus. *Proceedings of the annual symposium on computer application in medical care*, 261.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass'. *Cran r*, *538*, 113–120.

Vink, G., Frank, L. E., Pannekoek, J., & Van Buuren, S. (2014). Predictive mean matching imputation of semicontinuous variables. *Statistica Neerlandica*, *68*(1), 61–90.

Schouten, R. M., Lugtig, P., & Vink, G. (2018). Generating missing values for simulation purposes: A multivariate amputation procedure. *Journal of Statistical Computation and Simulation*, *88*(15), 2909–2930.

Wickham, H. (2017). Package tidyverse. *Easily Install and Load the 'Tidyverse*.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in medicine*, *25*(24), 4279–4292.

GNU, L. (1989). Version 2, june 1991. *Copyright (C)*, *1991*, 02111–1307.

Little, R. J. (1992). Regression with missing x's: A review. *Journal of the American statistical association*, *87*(420), 1227–1237.

Vink, G. (2016). Towards a standardized evaluation of multiple imputation routines.

Appendices

A Appendix RMSE

	λ		**Mean**	**Regression**	**Stoch (m=1)**	**Stoch (m=1)**	**Bayes (m=1)**	**Bayes (m=1)**	**pmm (m=1)**	**pmm (m=5)**
	0.3	X_1	12.45 (0.77)	12.25 (0.83)	17.06 (1.05)	17.07 (0.73)	17.05 (1.14)	17.1 (0.74)	17.03 (1.07)	17.06 (0.73)
	0.3	X_2	101.41 (12.64)	98.77 (12.74)	137.28 (12.24)	137.47 (10.34)	137.33 (12.3)	137.75 (10.39)	137.1 (14.2)	137.62 (11.3)
	0.3	Y	7.05 (0.43)	6.62 (0.44)	9.31 (0.63)	9.33 (0.44)	9.32 (0.62)	9.32 (0.44)	9.33 (0.61)	9.29 (0.42)
	0.5	X_1	12.46 (0.58)	12.24 (0.62)	17.07 (0.87)	17.08 (0.58)	17.05 (0.87)	17.12 (0.6)	16.97 (0.83)	17.05 (0.59)
MCAR	0.5	X_2	101.59 (9.72)	98.59 (9.89)	136.98 (9.86)	137.24 (8.53)	137.33 (10.02)	137.45 (8.57)	136.81 (11.66)	137.16 (9.24)
	0.5	Y	7.06 (0.33)	6.65 (0.34)	9.32 (0.5)	9.35 (0.34)	9.35 (0.51)	9.35 (0.34)	9.3 (0.5)	9.3 (0.34)
	0.7	X_1	12.43 (0.51)	12.22 (0.55)	17.08 (0.79)	17.07 (0.56)	17.1 (0.78)	17.11 (0.57)	17.05 (0.78)	17.06 (0.57)
	0.7	X_2	101.43 (8.6)	98.52 (8.79)	136.92 (9.28)	137.23 (8.03)	137.35 (9.72)	137.29 (8.21)	136.53 (10.41)	136.99 (8.77)
	0.7	Y	7.08 (0.29)	6.69 (0.33)	9.33 (0.46)	9.33 (0.34)	9.34 (0.47)	9.33 (0.35)	9.28 (0.45)	9.29 (0.34)
	0.3	X_1	12.59 (0.77)	12.25 (0.8)	17.1 (1.08)	17.1 (0.72)	17.14 (1.11)	17.11 (0.75)	17.05 (1.12)	17.05 (0.72)
	0.3	X_2	111.52 (15.12)	107.12 (13.73)	142.26 (12.99)	142.05 (11.25)	142.21 (12.74)	142.27 (11.37)	148.75 (16.19)	149.34 (12.57)
	0.3	Y	7.26 (0.46)	6.69 (0.43)	9.37 (0.6)	9.37 (0.41)	9.37 (0.59)	9.39 (0.42)	9.23 (0.6)	9.25 (0.42)
	0.5	X_1	12.48 (0.61)	12.19 (0.64)	17.08 (0.88)	17.07 (0.59)	17.06 (0.9)	17.08 (0.6)	17.04 (0.85)	17.04 (0.59)
MAR	0.5	X_2	108.54 (10.93)	104.81 (10.33)	139.7 (10.25)	139.58 (9.12)	139.73 (10.72)	139.82 (9.23)	145.78 (12.78)	145.98 (10.21)
	0.5	Y	7.18 (0.36)	6.73 (0.37)	9.33 (0.47)	9.34 (0.36)	9.36 (0.49)	9.36 (0.38)	9.28 (0.5)	9.27 (0.35)
	0.7	X_1	12.46 (0.5)	12.22 (0.54)	17.06 (0.76)	17.07 (0.54)	17.11 (0.8)	17.09 (0.55)	17.08 (0.75)	17.07 (0.54)
	0.7	X_2	105.26 (8.95)	102.12 (8.68)	137.4 (8.95)	137.49 (8)	137.84 (9.36)	137.66 (8.08)	141.88 (10.92)	142.39 (8.95)
	0.7	Y	7.12 (0.27)	6.8 (0.4)	9.31 (0.44)	9.34 (0.31)	9.35 (0.46)	9.35 (0.32)	9.28 (0.43)	9.27 (0.31)

B Appendix Raw Bias

	λ		**Mean**	**Regression**	**Stoch (m=1)**	**Stoch (m=5)**	**Bayes (m=1)**	**Bayes (m=5)**	**pmm (m=1)**	**pmm (m=5)**
	0.3	μ_{x1}	0 (0.17)	0 (0.16)	0.01 (0.22)	0 (0.18)	0 (0.23)	0 (0.18)	0 (0.23)	0 (0.18)
	0.3	μ_{x2}	-0.03 (1.37)	-0.03 (1.36)	-0.06 (1.77)	0 (1.4)	-0.01 (1.75)	-0.04 (1.46)	-0.01 (1.9)	0.01 (1.47)
	0.3	μ_y	0 (0.09)	0 (0.09)	0 (0.12)	0 (0.09)	0.01 (0.12)	0 (0.1)	0 (0.13)	0 (0.1)
	0.3	σ_{x1}	-23.32 (10.58)	-21.74 (9.82)	-0.16 (4.24)	-0.15 (3.24)	0 (4.17)	0 (3.29)	-0.12 (3.99)	-0.23 (3.21)
	0.3	σ_{x2}	-1534.06 (793.46)	-1368.7 (719.38)	0.8 (448.59)	-9.46 (419.22)	0.75 (463.53)	2.43 (424.11)	-4.37 (622.07)	-3.88 (450.02)
	0.3	σ_y	-7.52 (3.47)	-6.4 (3.07)	0.02 (1.31)	0.02 (1.02)	0.04 (1.32)	0.03 (1.05)	0.07 (1.3)	-0.03 (1)
	0.3	$\rho_{x1,x2}$	-0.02 (0.02)	0.01 (0.02)	0 (0.02)	0 (0.02)	0 (0.03)	0 (0.02)	0 (0.02)	0 (0.02)
	0.3	$\rho_{x1,x3}$	-0.03 (0.02)	0.03 (0.02)	0 (0.02)	0 (0.02)	0 (0.02)	0 (0.02)	0.01 (0.02)	0 (0.02)
	0.3	$\rho_{x2,x3}$	-0.04 (0.02)	0.03 (0.02)	0 (0.02)	0 (0.02)	0 (0.02)	0 (0.02)	0 (0.02)	0 (0.02)
	0.5	μ_{x1}	0 (0.24)	0 (0.23)	0 (0.3)	0 (0.24)	0 (0.32)	0 (0.26)	0 (0.34)	-0.01 (0.25)
	0.5	μ_{x2}	-0.03 (1.99)	-0.02 (1.9)	-0.09 (2.46)	0.01 (2.03)	-0.04 (2.65)	-0.02 (2.04)	0 (2.58)	-0.01 (2.11)
	0.5	μ_y	0 (0.14)	0 (0.12)	-0.01 (0.15)	0 (0.13)	0 (0.18)	0 (0.13)	0 (0.18)	0 (0.14)
	0.5	σ_{x1}	-39.1 (18.44)	-36.81 (17.19)	-0.2 (5.31)	-0.27 (4.54)	-0.13 (5.88)	0.22 (4.71)	-0.49 (5.63)	-0.45 (4.62)
MCAR	0.5	σ_{x2}	-2636.09 (1338.93)	-2392.96 (1209.25)	2.97 (692.04)	-8.07 (653.64)	12.73 (691.03)	6.58 (658.72)	-1.54 (814.31)	-11.26 (689.32)
	0.5	σ_y	-12.38 (5.66)	-10.25 (4.95)	0.09 (1.73)	0.09 (1.4)	0.15 (1.76)	0.14 (1.45)	0.04 (1.79)	-0.01 (1.41)
	0.5	$\rho_{x1,x2}$	-0.03 (0.02)	0.02 (0.04)	0 (0.03)	0 (0.03)	0 (0.04)	0 (0.03)	0 (0.04)	0 (0.03)
	0.5	$\rho_{x1,x3}$	-0.06 (0.03)	0.05 (0.03)	0 (0.03)	0 (0.03)	0 (0.03)	0 (0.03)	0.01 (0.03)	0.01 (0.03)
	0.5	$\rho_{x2,x3}$	-0.08 (0.04)	0.07 (0.03)	0 (0.03)	0 (0.03)	0 (0.03)	0 (0.03)	-0.01 (0.03)	-0.01 (0.03)
	0.7	μ_{x1}	0.01 (0.31)	0 (0.3)	0 (0.37)	0.01 (0.31)	0.02 (0.43)	0.01 (0.32)	0.01 (0.43)	0.01 (0.32)
	0.7	μ_{x2}	0.14 (2.62)	0.14 (2.56)	0.04 (3.09)	0.14 (2.7)	0.06 (3.52)	0.13 (2.77)	0.18 (3.62)	0.11 (2.74)
	0.7	μ_y	0 (0.17)	0 (0.17)	0 (0.21)	-0.01 (0.17)	0 (0.23)	0 (0.18)	0 (0.25)	-0.01 (0.18)
	0.7	σ_{x1}	-53.26 (24.5)	-50.69 (22.88)	0.42 (6.8)	0.25 (5.93)	0.58 (7.76)	0.7 (6.06)	0.14 (6.97)	0.12 (5.87)
	0.7	σ_{x2}	-3576.93 (1738.14)	-3298.46 (1570.26)	38.54 (834.27)	29.97 (809.42)	72.51 (904.27)	54.63 (819.54)	-12.69 (1047.9)	2.27 (855.53)
	0.7	σ_y	-17.74 (7.57)	-14.12 (6.82)	0.04 (2.25)	-0.01 (1.92)	0.08 (2.49)	0.06 (2)	-0.19 (2.32)	-0.16 (1.88)
	0.7	$\rho_{x1,x2}$	-0.05 (0.03)	0.03 (0.06)	0 (0.05)	0 (0.04)	0 (0.05)	0 (0.04)	0 (0.05)	0 (0.04)
	0.7	$\rho_{x1,x3}$	-0.1 (0.05)	0.08 (0.06)	0 (0.04)	0 (0.04)	0 (0.05)	0 (0.04)	0.01 (0.05)	0.01 (0.04)
	0.7	$\rho_{x2,x3}$	-0.13 (0.05)	0.1 (0.06)	0 (0.04)	0 (0.04)	0 (0.05)	0 (0.04)	-0.01 (0.05)	-0.01 (0.04)
	0.3	μ_{x1}	-0.3 (0.23)	-0.05 (0.17)	0.01 (0.22)	0.01 (0.18)	0 (0.25)	0.01 (0.19)	-0.01 (0.24)	-0.01 (0.19)
	0.3	μ_{x2}	-2.96 (2.21)	-0.64 (1.6)	-0.02 (1.95)	-0.03 (1.67)	-0.1 (2.1)	-0.04 (1.66)	-0.1 (2.26)	-0.01 (1.72)
	0.3	μ_y	-0.28 (0.17)	0.03 (0.1)	0.02 (0.13)	0.02 (0.1)	0.02 (0.14)	0.02 (0.1)	0 (0.13)	0 (0.1)
	0.3	σ_{x1}	-22.49 (10.97)	-20.69 (9.97)	0.29 (3.95)	0.04 (3.23)	0.43 (4.28)	0.19 (3.24)	-0.21 (4.19)	-0.26 (3.3)
	0.3	σ_{x2}	-1880.37 (943.91)	-1718.27 (872.9)	-349.84 (576.31)	-368.09 (564.98)	-354.55 (590.16)	-362.53 (564.6)	-29.26 (783.43)	-24.25 (615.88)
	0.3	σ_y	-7.9 (3.66)	-5.64 (2.9)	0.76 (1.59)	0.71 (1.34)	0.68 (1.64)	0.75 (1.35)	-0.05 (1.41)	-0.08 (1.2)
	0.3	$\rho_{x1,x2}$	-0.04 (0.02)	0.01 (0.03)	0.01 (0.03)	0.01 (0.03)	0.01 (0.04)	0.01 (0.03)	0 (0.03)	0 (0.03)
	0.3	$\rho_{x1,x3}$	-0.05 (0.02)	0.03 (0.02)	0 (0.02)	0 (0.02)	0 (0.03)	0 (0.02)	0 (0.03)	0 (0.02)
	0.3	$\rho_{x2,x3}$	-0.08 (0.03)	0.05 (0.03)	0.02 (0.03)	0.02 (0.03)	0.02 (0.04)	0.02 (0.03)	-0.01 (0.03)	0 (0.03)
	0.5	μ_{x1}	-0.42 (0.34)	-0.07 (0.25)	0.02 (0.31)	0 (0.26)	0.01 (0.35)	0 (0.27)	-0.03 (0.34)	-0.02 (0.27)
	0.5	μ_{x2}	-4.21 (3.38)	-0.87 (2.29)	0.1 (2.65)	0.02 (2.32)	-0.04 (2.94)	-0.03 (2.38)	-0.08 (3.15)	-0.03 (2.49)
	0.5	μ_y	-0.37 (0.26)	0.06 (0.15)	0.02 (0.17)	0.01 (0.14)	0.01 (0.18)	0.01 (0.15)	0 (0.19)	-0.01 (0.14)
	0.5	σ_{x1}	-38.65 (17.34)	-36.1 (15.86)	0.57 (5.35)	0.43 (4.45)	0.85 (6)	0.61 (4.55)	-0.11 (5.84)	-0.1 (4.8)
MAR	0.5	σ_{x2}	-2980.95 (1485.74)	-2759.32 (1358.03)	-495.72 (743.63)	-506.18 (715.93)	-491.9 (760.03)	-492.45 (727.99)	-25.43 (984.93)	2.55 (817.19)
	0.5	σ_y	-12.64 (5.87)	-9.26 (4.83)	0.8 (1.99)	0.71 (1.67)	0.8 (2.2)	0.8 (1.73)	-0.12 (1.91)	-0.19 (1.59)
	0.5	$\rho_{x1,x2}$	-0.05 (0.03)	0.02 (0.04)	0.01 (0.04)	0.01 (0.04)	0 (0.05)	0.01 (0.04)	0 (0.05)	0 (0.04)
	0.5	$\rho_{x1,x3}$	-0.07 (0.03)	0.05 (0.03)	0 (0.03)	0 (0.03)	0 (0.03)	0 (0.03)	0 (0.04)	0 (0.03)
	0.5	$\rho_{x2,x3}$	-0.11 (0.05)	0.09 (0.05)	0.02 (0.04)	0.02 (0.04)	0.02 (0.04)	0.02 (0.04)	-0.01 (0.04)	-0.01 (0.03)
	0.7	μ_{x1}	-0.47 (0.48)	-0.11 (0.35)	-0.02 (0.39)	-0.03 (0.34)	-0.03 (0.46)	-0.03 (0.36)	-0.05 (0.46)	-0.04 (0.36)
	0.7	μ_{x2}	-4.16 (3.92)	-0.71 (2.73)	0.19 (3.23)	0.21 (2.75)	0.29 (3.7)	0.22 (2.84)	-0.16 (3.86)	-0.06 (2.98)
	0.7	μ_y	-0.37 (0.33)	0.1 (0.21)	0 (0.22)	0 (0.19)	0 (0.25)	0 (0.21)	-0.02 (0.25)	-0.01 (0.2)
	0.7	σ_{x1}	-54.38 (24.9)	-51.35 (22.93)	0.5 (7.19)	0.32 (6.15)	0.97 (8.12)	0.79 (6.41)	0.13 (7.96)	-0.02 (6.74)
	0.7	σ_{x2}	-3835.66 (1824.17)	-3588.44 (1692.89)	-463.79 (912.77)	-470.02 (880.31)	-431.28 (933.36)	-458.54 (887.77)	15.73 (1305.19)	27.22 (1028.08)
	0.7	σ_y	-17.41 (8.14)	-12.4 (7)	0.68 (2.36)	0.72 (2.06)	0.84 (2.8)	0.84 (2.18)	-0.25 (2.42)	-0.25 (1.98)
	0.7	$\rho_{x1,x2}$	-0.05 (0.04)	0.04 (0.07)	0.01 (0.06)	0.01 (0.05)	0.01 (0.06)	0.01 (0.05)	0 (0.06)	0 (0.05)
	0.7	$\rho_{x1,x3}$	-0.1 (0.04)	0.08 (0.06)	0 (0.05)	0 (0.04)	0 (0.05)	0 (0.04)	0 (0.05)	0 (0.04)
	0.7	$\rho_{x2,x3}$	-0.14 (0.06)	0.13 (0.08)	0.02 (0.05)	0.02 (0.05)	0.02 (0.06)	0.02 (0.05)	-0.02 (0.06)	-0.01 (0.05)

	λ		**Mean**	**Regression**	**Stoch (m=1)**	**Stoch (m=5)**	**Bayes (m=1)**	**Bayes (m=5)**	**pmm (m=1)**	**pmm (m=5)**
MCAR	0.3	β_{x1}	-0.01 (0.01)	0.01 (0.01)	0 (0.01)	0 (0.01)	0 (0.01)	0 (0.01)	0 (0.01)	0 (0.01)
	0.3	β_{x2}	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	0.3	σ_e^2	-4.56 (2.98)	-6.85 (2.54)	-0.07 (1.25)	-0.05 (1)	-0.03 (1.28)	-0.07 (0.99)	-0.02 (1.22)	-0.06 (0.97)
	0.3	R^2	-0.04 (0.01)	0.04 (0.02)	0 (0.02)	0 (0.01)	0 (0.02)	0 (0.01)	0 (0.02)	0 (0.01)
	0.5	β_{x1}	-0.02 (0.02)	0.02 (0.03)	0 (0.02)	0 (0.02)	0 (0.02)	0 (0.02)	0 (0.02)	0 (0.02)
	0.5	β_{x2}	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	0.5	σ_e^2	-7.56 (4.89)	-11.2 (3.94)	-0.07 (1.71)	-0.04 (1.39)	0.04 (1.8)	0 (1.43)	-0.02 (1.72)	-0.04 (1.37)
	0.5	R^2	-0.07 (0.02)	0.07 (0.03)	0 (0.03)	0 (0.02)	0 (0.03)	0 (0.02)	0 (0.03)	0 (0.02)
	0.7	β_{x1}	-0.04 (0.03)	0.04 (0.05)	0 (0.03)	0 (0.02)	0 (0.03)	0 (0.02)	0.01 (0.03)	0.01 (0.02)
	0.7	β_{x2}	-0.01 (0.01)	0.01 (0.01)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	0.7	σ_e^2	-11.21 (6.73)	-15.66 (5.21)	-0.1 (2.22)	-0.15 (1.88)	-0.03 (2.55)	-0.12 (1.92)	-0.13 (2.42)	-0.1 (1.89)
	0.7	R^2	-0.1 (0.03)	0.12 (0.05)	0 (0.03)	0 (0.03)	0 (0.04)	0 (0.03)	0 (0.04)	0 (0.03)
MAR	0.3	β_{x1}	-0.02 (0.02)	0.01 (0.02)	0 (0.01)	0 (0.01)	0 (0.01)	0 (0.01)	0 (0.01)	0 (0.01)
	0.3	β_{x2}	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	0.3	σ_e^2	-3.81 (2.94)	-6.8 (2.55)	-0.16 (1.23)	-0.16 (1.06)	-0.14 (1.34)	-0.12 (1.05)	0.12 (1.37)	0.05 (1.07)
	0.3	R^2	-0.06 (0.02)	0.05 (0.03)	0.01 (0.03)	0.01 (0.02)	0.01 (0.03)	0.01 (0.02)	0 (0.02)	0 (0.02)
	0.5	β_{x1}	-0.03 (0.02)	0.02 (0.03)	0 (0.02)	0 (0.02)	0 (0.02)	0 (0.02)	0 (0.02)	0 (0.02)
	0.5	β_{x2}	-0.01 (0)	0.01 (0.01)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	0.5	σ_e^2	-7.18 (5.01)	-11.23 (4.03)	-0.32 (1.76)	-0.37 (1.44)	-0.31 (1.84)	-0.28 (1.46)	0.02 (1.86)	-0.05 (1.54)
	0.5	R^2	-0.08 (0.02)	0.09 (0.04)	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)	0.02 (0.03)	0 (0.03)	0 (0.03)
	0.7	β_{x1}	-0.04 (0.03)	0.04 (0.05)	0 (0.03)	0 (0.02)	0 (0.03)	0 (0.02)	0 (0.03)	0 (0.02)
	0.7	β_{x2}	-0.01 (0.01)	0.01 (0.01)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	0.7	σ_e^2	-10.85 (7.09)	-15.59 (5.43)	-0.49 (2.26)	-0.4 (1.95)	-0.3 (2.67)	-0.36 (2)	0.06 (2.5)	-0.03 (2.03)
	0.7	R^2	-0.11 (0.03)	0.15 (0.07)	0.02 (0.04)	0.02 (0.04)	0.02 (0.05)	0.02 (0.04)	-0.01 (0.04)	0 (0.03)

C Appendix Coverage Rate

	λ		Mean	Regression	Stoch (m=1)	Stoch (m=5)	Bayes (m=1)	Bayes (m=5)	pmm (m=1)	pmm (m=5)
	0.3	μ_{x1}	1,000	1,000	0,998	1,000	0,998	1,000	0,999	1,000
	0.3	μ_{x2}	0,999	0,999	0,998	1,000	1,000	1,000	0,999	1,000
	0.3	μ_y	0,999	0,999	0,999	1,000	0,999	0,999	0,998	0,999
	0.5	μ_{x1}	0,983	0,987	0,985	0,994	0,974	0,994	0,965	0,996
MCAR	0.5	μ_{x2}	0,978	0,985	0,983	0,992	0,969	0,997	0,978	0,990
	0.5	μ_y	0,981	0,993	0,988	0,997	0,979	0,996	0,982	0,996
	0.7	μ_{x1}	0,928	0,937	0,959	0,974	0,920	0,973	0,924	0,975
	0.7	μ_{x2}	0,911	0,923	0,941	0,968	0,929	0,963	0,918	0,963
	0.7	μ_y	0,933	0,953	0,963	0,979	0,940	0,967	0,933	0,978
	0.3	μ_{x1}	0,940	0,997	0,995	0,999	0,995	1,000	0,997	0,999
	0.3	μ_{x2}	0,868	0,994	0,994	0,999	0,988	1,000	0,987	0,998
	0.3	μ_y	0,746	0,999	0,997	1,000	0,998	1,000	0,996	0,999
	0.5	μ_{x1}	0,770	0,982	0,981	0,988	0,964	0,992	0,97	0,991
MAR	0.5	μ_{x2}	0,658	0,941	0,976	0,978	0,952	0,984	0,942	0,974
	0.5	μ_y	0,558	0,976	0,980	0,993	0,972	0,990	0,964	0,995
	0.7	μ_{x1}	0,677	0,902	0,945	0,956	0,905	0,957	0,910	0,952
	0.7	μ_{x2}	0,646	0,902	0,941	0,954	0,905	0,954	0,906	0,956
	0.7	μ_y	0,553	0,882	0,936	0,966	0,925	0,959	0,907	0,960

	λ		Mean	Regression	Stoch (m=1)	Stoch (m=5)	Bayes (m=1)	Bayes (m=5)	pmm (m=1)	pmm (m=5)
	0.3	β_{x1}	0,893	0,895	0,986	0,995	0,984	0,999	0,975	0,999
	0.3	β_{x2}	0,802	0,796	0,984	0,998	0,977	0,997	0,973	0,997
MCAR	0.5	β_{x1}	0,633	0,654	0,907	0,991	0,904	0,989	0,881	0,986
	0.5	β_{x2}	0,407	0,461	0,909	0,98	0,882	0,984	0,9	0,984
	0.7	β_{x1}	0,335	0,46	0,798	0,965	0,763	0,981	0,76	0,971
	0.7	β_{x2}	0,207	0,307	0,775	0,944	0,729	0,978	0,731	0,961
	0.3	β_{x1}	0,793	0,902	0,98	0,999	0,968	0,999	0,973	0,999
	0.3	β_{x2}	0,465	0,516	0,814	0,913	0,821	0,929	0,935	0,988
MAR	0.5	β_{x1}	0,602	0,63	0,909	0,991	0,909	0,996	0,9	0,995
	0.5	β_{x2}	0,29	0,26	0,712	0,878	0,697	0,915	0,822	0,97
	0.7	β_{x1}	0,396	0,416	0,772	0,969	0,751	0,985	0,762	0,981
	0.7	β_{x2}	0,208	0,218	0,661	0,854	0,625	0,922	0,669	0,945

D Appendix Average Confidence Interval Width

	λ		**Mean**	**Regression**	**Stoch (m=1)**	**Stoch (m=5)**	**Bayes (m=1)**	**Bayes (m=5)**	**pmm (m=1)**	**pmm (m=5)**
	0.3	X_1	1,422	1,431	1,544	1,544	1,544	1,544	1,544	1,543
	0.3	X_2	11,667	11,777	12,644	12,637	12,643	12,645	12,639	12,640
	0.3	Y	0,808	0,819	0,878	0,878	0,878	0,878	0,878	0,877
	0.5	X_1	1,331	1,345	1,543	1,543	1,544	1,545	1,542	1,542
MCAR	0.5	X_2	10,889	11,067	12,640	12,634	12,646	12,643	12,635	12,632
	0.5	Y	0,759	0,781	0,878	0,878	0,879	0,879	0,878	0,877
	0.7	X_1	1,242	1,259	1,546	1,546	1,547	1,548	1,545	1,545
	0.7	X_2	10,154	10,385	12,660	12,655	12,679	12,670	12,624	12,637
	0.7	Y	0,700	0,740	0,878	0,877	0,878	0,878	0,876	0,876
	0.3	X_1	1,427	1,437	1,546	1,545	1,547	1,545	1,543	1,543
	0.3	X_2	11,431	11,542	12,428	12,417	12,425	12,420	12,621	12,626
	0.3	Y	0,805	0,826	0,884	0,884	0,883	0,884	0,877	0,877
	0.5	X_1	1,334	1,350	1,547	1,547	1,548	1,547	1,544	1,544
MAR	0.5	X_2	10,634	10,802	12,334	12,328	12,336	12,337	12,617	12,637
	0.5	Y	0,756	0,791	0,884	0,884	0,884	0,884	0,876	0,876
	0.7	X_1	1,234	1,255	1,547	1,546	1,549	1,548	1,545	1,544
	0.7	X_2	9,951	10,159	12,349	12,347	12,369	12,354	12,634	12,648
	0.7	Y	0,703	0,758	0,883	0,884	0,885	0,885	0,875	0,875

	λ		**Mean**	**Regression**	**Stoch (m=1)**	**Stoch (m=5)**	**Bayes (m=1)**	**Bayes (m=5)**	**pmm (m=1)**	**pmm (m=5)**
	0.3	β_{x1}	0,066	0,064	0,064	0,076	0,064	0,078	0,064	0,078
	0.3	β_{x2}	0,008	0,008	0,008	0,009	0,008	0,009	0,008	0,009
MCAR	0.5	β_{x1}	0,068	0,063	0,064	0,086	0,064	0,094	0,064	0,093
	0.5	β_{x2}	0,008	0,008	0,008	0,011	0,008	0,011	0,008	0,011
	0.7	β_{x1}	0,069	0,063	0,064	0,101	0,064	0,122	0,064	0,120
	0.7	β_{x2}	0,008	0,008	0,008	0,012	0,008	0,015	0,008	0,014
	0.3	β_{x1}	0,066	0,063	0,064	0,078	0,064	0,080	0,065	0,080
	0.3	β_{x2}	0,008	0,008	0,008	0,011	0,008	0,011	0,008	0,010
MAR	0.5	β_{x1}	0,068	0,063	0,064	0,085	0,064	0,096	0,064	0,094
	0.5	β_{x2}	0,009	0,008	0,008	0,011	0,008	0,013	0,008	0,012
	0.7	β_{x1}	0,070	0,064	0,064	0,101	0,064	0,119	0,064	0,117
	0.7	β_{x2}	0,009	0,008	0,008	0,013	0,008	0,017	0,008	0,016