APPROVAL SHEET

Title of Dissertation: Bayesian Analysis of Synthetic Data under Multiple Linear Regression, Multivariate Normal and Multivariate Regression Models

Name of Candidate: Abhishek Guin

Doctor of Philosophy, Statistics, 2021

Dissertation and Abstract Approved: _____

Bimal K. Sinha

Professor of Statistics

Department of Mathematics and Statistics

Date Approved: _____

CURRICULUM VITAE

Name: Abhishek Guin

Education: University of Maryland, Baltimore County (UMBC)
Doctor of Philosophy, Statistics, Fall 2016 - Summer 2021

National Institute of Science Education and Research (NISER),
Bhubaneswar, India
Integrated MSc, Mathematics, Fall 2011 - Spring 2016

Work Experience: University of Maryland, Baltimore County (UMBC)
Graduate Teaching Assistant
Department of Mathematics and Statistics, UMBC
Fall 2016 - Present

ABSTRACT

Title of Dissertation: Bayesian Analysis of Synthetic Data under Multiple Linear Regression, Multivariate Normal and Multivariate Regression Models

Abhishek Guin

Doctor of Philosophy, 2021

Dissertation directed by: Bimal K. Sinha

Professor of Statistics

Department of Mathematics and Statistics

University of Maryland, Baltimore County

Anindya Roy

Professor of Statistics

Department of Mathematics and Statistics

University of Maryland, Baltimore County

Statistical Disclosure Control (SDC) methods are used to preserve confidentiality of publicly released microdata, without compromising on its fundamental structure, so as to ensure adequate and accurate statistical analysis of the data. The synthetic data approach is a popular form of SDC methodology where (all or part of) the real data are not released, but are instead used to create synthetic data which are released.

In this dissertation we develop Bayesian inference based on singly or multiply imputed synthetic data, when the original data are derived from the following models: multiple linear regression, multivariate normal and multivariate regression. We assume that the synthetic data are generated by using two methods: plug-in sampling, where unknown parameters in the data model are set equal to observed values of their point estimators based on the original data, and synthetic data are drawn from this estimated version of the model; posterior predictive sampling, where an imputed posterior distribution of the unknown parameters is used to generate a posterior draw, which in turn is plugged in the original model to produce synthetic data. In the single imputation case, the procedures developed here fill the gap in the existing literature where inferential methods are only available for multiple imputation and by being based on exact distributions, it may even be applied to cases where the sample size is small. Simulation results are presented to demonstrate how the proposed methodology performs compared to the theoretical predictions. We also outline some ways to extend the proposed methodology for certain scenarios where the required set of conditions do not hold.

BAYESIAN ANALYSIS OF SYNTHETIC DATA UNDER MULTIPLE LINEAR REGRESSION, MULTIVARATE NORMAL AND MULTIVARIATE REGRESSION MODELS

By

Abhishek Guin

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore County, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Dr. Bimal K. Sinha, Chair/Advisor
Dr. Anindya Roy, Co-advisor
Dr. Thomas Mathew
Dr. Yaakov Malinovsky
Dr. Emanuel Ben-David

Dedication

To all my teachers who have shone light along the way

Acknowledgements

I would first like to thank my advisors, Professor Bimal K. Sinha and Prof Anindya Roy, for all of their guidance, wisdom, expertise throughout the project, from inception to completion. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. I would also like to mention their incredible and unwavering patience in working with me over the past three years.

I would like to acknowledge the many assiduous hours Dr Shyamal Krishna De spent with me during my final year of MSc teaching me the basics of statistics. This work is the tree that sprouted off the seeds you planted. I owe my interest in Mathematics to careful and appreciative nurturing Rahman sir and Subrata sir during high school years, I am immensely grateful to you. The work ethic that I observed from all of them is what I hope to emulate.

I would like to acknowledge all of my course professors (especially Prof Thomas Mathew) and my colleagues (especially Eswar, Bala, Nivya, Chip, Neha, Gaurab, Ahmad, Maria) from the department for creating a fertile atmosphere of research and support.

No amount of appreciation is enough for our department staff Janet Burgee, Maggie Kennedy, Marshall Turner who have helped me out at every turn.

A special thanks to my fellow student Michael Lucagbo for being an awesome source of inspiration and support. I am truly blessed to have had your company throughout this journey.

I have learned a great deal from all the students I have taught and all the professors I have assisted in teaching. I would like to single out Prof Michael Muscedere for being particularly kind, genial and opening my eyes to possibilities, at a time when I really needed that.

In addition, I would like to thank my parents Subhan and Krishna Guin and my

beloved sister Aditi for their wise counsel and sympathetic ear. You are always there for me. I could not have completed this dissertation without the support of all my friends (Tuhin, Avradip, Both, Sanjeev, Kiran, Krunal, Varun, Poorni, Anup, Ajinkya, Nitika, Ayushi, Rishabh), who graciously listen.

This dissertation could not have been completed without the overwhelming support of Anisha Kar in the last few laps. I am grateful to have your presence in my life.

Finally, a big thanks to all the friends of Bill Wilson.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Statistical agencies are often faced with two incongruous objectives, namely: collect and publish useful datasets for designing public policies and building scientific theories; as well as protect confidentiality of survey respondents which is essential to uphold public trust, leading to better response rates and data accuracy. When releasing microdata to the public, methods of statistical disclosure control (SDC) are used to protect confidential data. SDC methods include data swapping, additive and multiplicative noise, top and bottom coding, and also the creation of synthetic data. The synthetic data approach is a popular method aims to satisfy the two objectives, and some statistical agencies now release synthetic data products.

Generally, there are two types of synthetic data discussed in the literature: *fully synthetic data* and *partially synthetic data*, and methodology for drawing inference based on synthetic data has been developed using concepts of multiple imputation (Rubin, 1987). In fully synthetic data methodology, all units in the population not selected in the sample are treated as missing, and are multiply imputed based on the information from sampled units, to create multiple synthetic populations. A sample is then drawn

from each synthetic population, and these samples are released to the public. This approach was suggested by Rubin (1993), and methods for drawing inference based on the synthetic data generated using this approach were developed by Raghunathan et al. (2003). In the partially synthetic data approach, the released data comprise only the originally sampled units, but any responses deemed to be confidential are replaced by multiple imputations. For any particular variable, the responses could be deemed as confidential for some or all respondents. This approach was suggested by Little (1993), and methods for drawing inference based on synthetic data under this approach were developed by Reiter (2003). We refer to the monograph by Drechsler (2011) for a detailed and general discussion on synthetic data methodology.

In comparison with the standard SDC methods, multiple imputation techniques presents many advantages dealing with many real data problems that other methods cannot. It preserves the joint distribution of the original data offering a better quality analysis; is applicable to both categorical and continuous variables; released fully synthetic datasets gives a very small disclosure risk; with partially synthetic datasets generation one may only synthesize the records at risk, maintaining intact the records that have no need to be protected; it allows the possibility to impute missing values before generating synthetic datasets having no need to give up on some records; preserves linear constraints; allows the analyst to decide if valid results will be given from the synthetic data based on the meta-data information. Some drawbacks exist as well. Since it is a perturbation method there is a question on the utility limit of the data and only the statistical properties gathered by the model are preserved (An and Little, 2007; Drechsler, 2010).

In this work, we will be concerned solely with partially synthetic data and their utility and necessity is described below. There are several examples where partially synthetic data products have been produced based on major data sources. Some examples in

the United States include the Survey of Income and Program Participation (Abowd et al., 2006; Benedetto et al., 2013), the American Community Survey Group Quarters data (Hawala, 2008), OnTheMap data on where workers live and where they work (Machanavajjhala et al., 2008), and the Longitudinal Business Database (Kinney et al., 2011; Kinney et al., 2014). To obtain valid inference on population quantities using synthetic data, the current practice requires multiple synthetic datasets to be released, but there are cases where it is indeed desirable to release only a single partially synthetic dataset. For example, the Synthetic Longitudinal Business Database, accessible through the VirtualRDC at Cornell University, is a partially synthetic version of the U.S. Census Bureau's Longitudinal Business Database (LBD). As discussed in Kinney et al. (2011) and Kinney et al. (2014), the decision was made to release only a single version of the LBD in the synthetic file, instead of multiple copies, to avoid the perception of high disclosure risk. Similarly, in the application of partially synthetic data to American Community Survey Group Quarters data presented by Hawala (2008), only a single synthetic dataset is released, because of the concern that releasing multiple synthetic copies may increase disclosure risk.

The motivations for this current research are thus twofold. First, although synthetic data methodology calls for releasing multiple synthetic versions of the original data, there are situations where it might not be possible. Secondly, since synthetic data generation is indeed model-based, it becomes imperative to develop rigorous model-based finite sample inference.

The primary purpose of this work is to develop Bayesian analytic tools for drawing inference based on a singly or multiply imputed partially synthetic dataset(s) arising from the subsequent models: *multiple linear regression* (MLR), *multivariate normal* (MVN) and *multivariate regression* (MVR). This synthetic data problem fits into the framework of partially synthetic data, and hence the methodology of Reiter (2003)

can be used to obtain approximately valid inference if the sample size is sufficiently large and the number of multiply imputed synthetic datasets available is $m > 1$. However, given the specific structure of each problem, we shall instead exploit the model structure to derive Bayesian inference for the parameters. While the methodology we derive is specific to the problem at hand, it yields exact inference for both large and small samples using the $m \geq 1$ multiply imputed synthetic datasets that are available. We essentially extend the work done in a series of recent papers Klein et al. (2014), Klein and Sinha (2015a), Klein and Sinha (2015b), Moura (2016), Moura et al. (2017a), Moura et al. (2017b), Klein et al. (2019), Moura et al. (2021) that developed exact parametric inferential methods based on singly or multiply imputed synthetic data for several probability models, to the Bayesian domain.

1.1 Generating Synthetic Data

We consider two ways of generating the $m \geq 1$ synthetic copies of the original data namely, *plug-in sampling* and *posterior predictive sampling*. In the former method, parameter estimates are plugged in the model to generate synthetic data. In the latter one, posterior draws of the parameter are generated using an imputed prior, which are then fed into the original model to beget synthetic data.

Plug-in Sampling. The basic mechanism for generating synthetic data via *plug-in sampling* (PIS) is described as follows: let $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ be the original confidential data, which are jointly distributed according to the probability density function (pdf) $f_{\boldsymbol{\theta}}(\boldsymbol{Y})$, where $\boldsymbol{\theta}$ is the unknown (scalar or vector) parameter. To generate partially synthetic data, let $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\boldsymbol{Y})$ be the observed value of a point estimator of $\boldsymbol{\theta}$, and we plug it into the joint pdf of $\boldsymbol{Y}$. The resulting pdf, with the unknown $\boldsymbol{\theta}$ replaced by the observed value $\hat{\boldsymbol{\theta}}(\boldsymbol{Y})$ of the point estimator, is denoted by $f_{\hat{\boldsymbol{\theta}}}$. The singly

imputed synthetic data, denoted by $\boldsymbol{Z}$, are then generated by drawing $\boldsymbol{Z}$ from the joint pdf $f_{\hat{\boldsymbol{\theta}}}$. For the multiple imputation case, we draw $m > 1$ samples $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m$ independently from $f_{\hat{\boldsymbol{\theta}}}$.

Posterior Predictive Sampling. An alternative method to generate partially synthetic data is to use *posterior predictive sampling* (PPS) which proceeds as follows: suppose that $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ are the original data which are jointly distributed according to the pdf $f_{\boldsymbol{\theta}}(\boldsymbol{Y})$, where $\boldsymbol{\theta}$ is the unknown (scalar or vector) parameter. Assume a prior $\pi(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$, then the posterior distribution of $\boldsymbol{\theta}$ given $\boldsymbol{Y}$ is obtained as $\pi(\boldsymbol{\theta} \,|\, \boldsymbol{Y}) \propto \pi(\boldsymbol{\theta}) f_{\boldsymbol{\theta}}(\boldsymbol{Y})$, and used to draw $m \geq 1$ replications $\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_m^*$ (known as posterior draws). Next, for each posterior draw of $\boldsymbol{\theta}$, a corresponding replicate of $\boldsymbol{Y}$ is generated, namely $\boldsymbol{Z}_j = (\boldsymbol{z}_{j1}, \ldots, \boldsymbol{z}_{jn})'$ drawn from the pdf $f_{\boldsymbol{\theta}_j^*}(\boldsymbol{X})$ independently for $j = 1, \ldots, m$.

The organization of the rest of the thesis is as follows. The first two Chapters 1 and 2 set in motion the development of the theory under the MLR model, the structure of which is paralleled in the latter models as well. In Section 2.1, we carry out Bayesian inference based on singly imputed synthetic data under the MLR model generated using the plug-in sampling method. In Section 2.2, we derive Bayesian inference based on singly imputed synthetic data generated using posterior predictive sampling. Here we use a general form of the prior $\pi(\boldsymbol{\beta}, \sigma^2)$, involving a hyperparameter α. We present simulation results for both setups in Section 2.3 to demonstrate how the proposed methodology performs compared to the theoretical predictions. In Section 2.4 we extend the previous methodology to a more general scenario of the original data being only partially sensitive. In Section 3.1, we carry out Bayesian inference based on multiply imputed synthetic data generated using the plug-in sampling method and also include simulation results. In Section 3.2, we mention our progress in deriving Bayesian inference based on multiply imputed synthetic data generated using pos-

terior predictive sampling with the aforementioned prior. In Section 3.3 we discuss inference for partially sensitive data in context of the multiply imputed MLR setup. In Chapters 4 and 5 we develop Bayesian inference in a similar manner as above for singly and multiply imputed data respectively, when the original data arrives from the MVN model. Similarly the succeeding two Chapters 7 and 8 deal with the MVR model. Finally, in Chapter 8, we discuss the direction we would be taking with future work.

1.2 An Important Lemma

We conclude this chapter with an observation regarding the existence of *sufficient statistics* in the context of synthetic data.

Lemma 1.2.1. Suppose that when the original data $\boldsymbol{Y}$ are observed, $T(\boldsymbol{Y})$ is a sufficient statistic for $\boldsymbol{\theta}$. Then when the synthetic data $\boldsymbol{\mathcal{V}} = (\boldsymbol{V}_1, \ldots, \boldsymbol{V}_M)$ are observed, $(T(\boldsymbol{V}_1), \ldots, T(\boldsymbol{V}_M))$ is jointly sufficient for $\boldsymbol{\theta}$. Furthermore, if $M = 1$, the sufficient statistic is simply $T(\boldsymbol{V}_1)$, and if $M > 1$, then $\sum_{i=1}^{M} T(\boldsymbol{V}_i)$ is sufficient if $f_{\boldsymbol{\theta}}(\boldsymbol{Y}) = h(\boldsymbol{Y})\psi(\boldsymbol{\theta}) \exp\{\gamma(\boldsymbol{\theta})' T(\boldsymbol{Y})\}$, i.e., if $f_{\boldsymbol{\theta}}(\boldsymbol{Y})$ belongs to the exponential family.

Proof. Suppose based on the original data $\boldsymbol{Y}$, $\boldsymbol{T}(\boldsymbol{Y})$ is a sufficient statistic for the unknown parameter $\boldsymbol{\theta}$ in the original model $f_{\boldsymbol{\theta}}(\boldsymbol{Y})$. Then we can write $f_{\boldsymbol{\theta}}(\boldsymbol{Y}) = h(\boldsymbol{Y}) g_{\boldsymbol{\theta}}[\boldsymbol{T}(\boldsymbol{Y})]$, and the pdf of the synthetic data $\boldsymbol{\mathcal{V}} = (\boldsymbol{V}_1, \ldots, \boldsymbol{V}_M)$ is

$$\begin{aligned}\int \left\{ \prod_{i=1}^{M} f_{\hat{\boldsymbol{\theta}}(\boldsymbol{Y})}(\boldsymbol{V}_i) \right\} f_{\boldsymbol{\theta}}(\boldsymbol{Y}) d\boldsymbol{Y} &= \int \left\{ \prod_{i=1}^{M} g_{\hat{\boldsymbol{\theta}}(\boldsymbol{Y})}(T(\boldsymbol{V}_i)) h(\boldsymbol{V}_i) \right\} f_{\boldsymbol{\theta}}(\boldsymbol{Y}) d\boldsymbol{Y} \\ &= \left\{ \prod_{i=1}^{M} h(\boldsymbol{V}_i) \right\} \int \left\{ \prod_{i=1}^{M} g_{\hat{\boldsymbol{\theta}}(\boldsymbol{Y})}(T(\boldsymbol{V}_i)) \right\} f_{\boldsymbol{\theta}}(\boldsymbol{Y}) d\boldsymbol{Y}\end{aligned}$$

□

Chapter 2

Bayesian Analysis of Singly Imputed Synthetic Data under the Multiple Linear Regression Model

Throughout, we would be dealing with the case of a standard MLR model involving a *sensitive response* variable y and a $p \times 1$ dimensional vector of *non-sensitive predictors* $\boldsymbol{x}$. We assume that

$$y_1, \ldots, y_n \text{ are independent such that } y_i \sim N(\boldsymbol{x}_i'\boldsymbol{\beta},\, \sigma^2) \tag{2.1}$$

where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are fixed $p \times 1$ vectors, and $\boldsymbol{\beta}$ and σ^2 are both unknown. Thus the original data consist of $\{(y_i, \boldsymbol{x}_i) : i = 1, \ldots, n\}$. We define $\boldsymbol{y} = (y_1, \ldots, y_n)'$ as the $n \times 1$ dimensional vector of response variables, and $\boldsymbol{X} = [\boldsymbol{x}_1 \cdots \boldsymbol{x}_n]'$ as the $n \times p$ dimensional matrix of predictor variables, and we assume that $\text{rank}(\boldsymbol{X}) = p < n$. Based on the original data, $\hat{\boldsymbol{\beta}} = \boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ is the maximum likelihood estimator (MLE) and uniformly minimum variance unbiased estimator (UMVUE) of

$\boldsymbol{\beta}$, and $\hat{\sigma}^2 = \text{RSS}/(n-p)$ is the UMVUE of σ^2 where $\text{RSS} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) = \boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{y}$ with $\boldsymbol{I}_k$ as the k-dimensional identity matrix and $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is the orthogonal projection matrix to the column space of $\boldsymbol{X}$. Furthermore, $\boldsymbol{b}$ and RSS are independently distributed such that

$$\begin{aligned} \boldsymbol{b} &\sim N_p(\boldsymbol{\beta},\, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}) \\ \text{RSS} &\sim \sigma^2\chi^2_{n-p} \end{aligned} \tag{2.2}$$

When the original data are observed, $\boldsymbol{b}$ and RSS are jointly sufficient for $\boldsymbol{\beta}$ and σ^2.

Since $\boldsymbol{y}$ is sensitive and hence cannot be released, instead it is replaced with a singly imputed synthetic copy which is released. The synthetic data is generated by two methods as described earlier: *plug-in sampling* and *posterior predictive sampling*. In the former method, parameter estimates are plugged in the MLR model to generate synthetic data. In the latter one, posterior draws of the parameter are generated using an imputed prior, which are then fed into the MLR model to generate synthetic data. The development builds on the exact likelihood based procedures developed in Klein and Sinha (2015a) and Klein and Sinha (2015b).

2.1 Plug-In Sampling method

The singly imputed synthetic data in this case consist of a single synthetic version of $\boldsymbol{y} = (y_1, \ldots, y_n)'$, which is denoted as $\boldsymbol{z} = (z_1, \ldots, z_n)'$, and obtained by drawing

$$z_1, \ldots, z_n \text{ independently such that } z_i \sim N\left(\boldsymbol{x}_i'\boldsymbol{b},\, \frac{\text{RSS}}{n-p}\right) \tag{2.3}$$

Thus the released data will be of the form $\{(z_i, \boldsymbol{x}_i) : i = 1, \ldots, n\}$, and our goal is to discuss Bayesian inference on $\boldsymbol{\beta}$ and σ^2 based on this released data.

It is convenient to identify the latent structure of the pseudo randomization involved in the released data. For what follows we would write identities that are sometimes algebraic but also sometimes distributional. The exact case should be clear from the context. Specifically, we could write

$$\boldsymbol{z} \stackrel{d}{=} \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\sigma}\boldsymbol{W}$$

where $\boldsymbol{W} = (w_1, \ldots, w_n)' \sim N_n(\boldsymbol{0}, \boldsymbol{I}_n)$ with $w_i \overset{\text{iid}}{\sim} N(0,1)$. Then by Lemma 1.2.1 the sufficient statistics based on the released data are

$$\begin{aligned} \boldsymbol{b}^* &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{z} \stackrel{d}{=} \hat{\boldsymbol{\beta}} + \hat{\sigma}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W} \stackrel{d}{=} \hat{\boldsymbol{\beta}} + \hat{\sigma}\boldsymbol{C}\boldsymbol{U}_1 \\ \text{RSS}^* &= \boldsymbol{z}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{z} \stackrel{d}{=} \hat{\sigma}^2\boldsymbol{W}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{W} \stackrel{d}{=} \hat{\sigma}^2 V \end{aligned} \tag{2.4}$$

where $\boldsymbol{U}_1 \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$ and $V \sim \chi^2_{n-p}$, and $\boldsymbol{C}$ is a full rank square root of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ such that $\boldsymbol{C}\boldsymbol{C}' = (\boldsymbol{X}'\boldsymbol{X})^{-1}$. It is easy to check that $\boldsymbol{b}^*$ is independent of RSS* by using the following result: If $\boldsymbol{y} \sim N_p(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{B}_{k\times p}$ and $\boldsymbol{A}_{p\times p}$ are constant matrices, then $\boldsymbol{B}\boldsymbol{y}$ and $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}$ are independent if and only if $\boldsymbol{B}\Sigma\boldsymbol{A} = \boldsymbol{O}$. Next, we can write

$$\begin{aligned} \boldsymbol{b}^* &\stackrel{d}{=} \hat{\boldsymbol{\beta}} + \sigma(\hat{\sigma}/\sigma)\boldsymbol{C}\boldsymbol{U}_1 \stackrel{d}{=} \hat{\boldsymbol{\beta}} + \sigma\sqrt{\psi}\boldsymbol{C}\boldsymbol{U}_1 \\ \text{RSS}^* &\stackrel{d}{=} \sigma^2(\hat{\sigma}/\sigma)^2 V_1 \stackrel{d}{=} \sigma^2\psi V \end{aligned} \tag{2.5}$$

where $\psi = (\hat{\sigma}/\sigma)^2$ is a latent quantity. From (2.2), we have $\hat{\boldsymbol{\beta}} \stackrel{d}{=} \boldsymbol{\beta} + \sigma\boldsymbol{C}\boldsymbol{U}_2$ where $\boldsymbol{U}_2 \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$ independent of $\boldsymbol{U}_1$ and hence from (2.5) conditional on the parameters, we could write

$$\boldsymbol{b}^* \stackrel{d}{=} \boldsymbol{\beta} + \sigma\sqrt{1+\psi}\boldsymbol{C}\boldsymbol{U}_3$$

where $\boldsymbol{U}_3 \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$. Thus the likelihood based on the released data for the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \psi)$ is given by

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2, \psi \,|\, \boldsymbol{b}^*, \text{RSS}^*) = \phi_p(\boldsymbol{b}^*; \boldsymbol{\beta}, \sigma^2(1+\psi)(\boldsymbol{X}'\boldsymbol{X})^{-1})\, h(\text{RSS}^*; n-p, \sigma^2\psi) \tag{2.6}$$

where $\phi_k(\boldsymbol{w}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density of $\boldsymbol{w} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $h(v; d, s)$ is the density of $v \sim s\,\chi^2_d$. For full Bayesian specification, we need priors on the unknown quantities $(\boldsymbol{\beta}, \sigma^2, \psi)$. The prior on ψ is naturally imposed by the original MLR model and the single imputation mechanism. Thus, a priori

$$\psi \sim \pi(\psi) = h(\psi; n-p, (n-p)^{-1})$$

For Bayesian inference on the other unknown parameters we assume non-informative improper priors and assume that all unknown quantities are a priori independent. Specifically, we assume

$$\pi(\boldsymbol{\beta}, \sigma^2) = \pi(\boldsymbol{\beta})\pi(\sigma^2)$$

where $\pi(\boldsymbol{\beta}) \propto 1$ and $\pi(\sigma) \propto \sigma^{-\delta}$ and hence the induced prior on σ^2 is $\pi(\sigma^2) \propto (\sigma^2)^{-\frac{\delta+1}{2}}$ for $\delta > 0$. The posterior distribution can then be computed in the following manner:

$$\pi(\boldsymbol{\beta}, \sigma^2, \psi \,|\, \boldsymbol{b}^*, \mathrm{RSS}^*) \propto \mathcal{L}(\boldsymbol{\beta}, \sigma^2, \psi \,|\, \boldsymbol{b}^*, \mathrm{RSS}^*)\, \pi(\psi)\, \pi(\boldsymbol{\beta}, \sigma^2)$$

$$\pi(\boldsymbol{\beta}, \sigma^2, \psi \,|\, \boldsymbol{b}^*, \mathrm{RSS}^*) = \pi(\boldsymbol{\beta} \,|\, \boldsymbol{b}^*, \mathrm{RSS}^*, \sigma^2, \psi)\, \pi(\sigma^2 \,|\, \boldsymbol{b}^*, \mathrm{RSS}^*, \psi)\, \pi(\psi \,|\, \boldsymbol{b}^*, \mathrm{RSS}^*)$$

The conditional posteriors follow from observing that from the above two equations the product of the likelihood of the parameters and their priors break up into three conditional posterior distributions as follows

$$\boldsymbol{\beta} \,|\, \boldsymbol{b}^*, \mathrm{RSS}^*, \sigma^2, \psi \;\sim\; N_p\left(\boldsymbol{b}^*, \sigma^2(1+\psi)(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) \tag{2.7}$$

$$\sigma^2 \,|\, \boldsymbol{b}^*, \mathrm{RSS}^*, \psi \;\sim\; \text{Scale-inv-}\chi^2\left(n-p+\delta-1, \frac{\mathrm{RSS}^*}{\psi(n-p+\delta-1)}\right) \tag{2.8}$$

$$\psi \;\sim\; (n-p)^{-1}\chi^2_{n-p+\delta-1} \tag{2.9}$$

The posterior distributions are proper as long as $n > \max\{p, p-\delta+1\}$.

We observe that $\frac{\sigma^2\psi}{\text{RSS}^*} \,|\, \text{RSS}^*, \psi \sim \text{inv-}\chi^2_{n-p+\delta-1}$ so that

$$\frac{\sigma^2\psi}{\text{RSS}^*} \sim \text{inv-}\chi^2_{n-p+\delta-1} \tag{2.10}$$

unconditionally and $\frac{\sigma^2\psi}{\text{RSS}^*}$ is independent of the data and ψ. Here we use the fact that if $X \sim \text{Scale-inv-}\chi^2(\nu, \tau^2)$ then $\frac{X}{\tau^2\nu} \sim \text{inv-}\chi^2_\nu$.

Marginal Posterior of parameters

$$\boldsymbol{\beta} \,|\, \boldsymbol{b}^*, \text{RSS}^*, \psi \sim \boldsymbol{t}_{n-p+\delta-1}\left(\boldsymbol{b}^*, \frac{\text{RSS}^*(1+\psi)}{\psi(n-p+\delta-1)}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right)$$

$$\pi(\sigma^2 \,|\, \text{RSS}^*) \propto (\sigma^2)^{-\frac{n-p+\delta+1}{2}} \,\text{K}_0\left(\sqrt{\frac{(n-p)\text{RSS}^*}{\sigma^2}}\right)$$

where $\text{K}_\nu(z)$ is the modified Bessel function of the second kind as defined in Tweedie (1957).

Marginal Likelihood of data

$$\pi(\boldsymbol{b}^*, \text{RSS}^*) = \int \pi(\boldsymbol{b}^*, \text{RSS}^*, \psi \,|\, \boldsymbol{\beta}, \sigma^2)\, \pi(\boldsymbol{\beta}, \sigma^2)\, d\boldsymbol{\beta}\, d\sigma^2 d\psi \propto (\text{RSS}^*)^{-\frac{\delta+1}{2}}$$

Posterior Predictive Density

Let D be the original dataset and D_{new} be the new dataset with $(\tilde{\boldsymbol{b}}^*, \widetilde{\text{RSS}}^*)$ as the sufficient statistic.

$$\pi(D_{\text{new}} \,|\, D) = \int \pi(D_{\text{new}} \,|\, \boldsymbol{\beta}, \sigma^2, \psi)\, \pi(\boldsymbol{\beta}, \sigma^2, \psi \,|\, D)\, d\boldsymbol{\beta}\, d\sigma^2\, d\psi$$

$$\propto (\widetilde{\text{RSS}}^*)^{\frac{n-p}{2}-1} \int \left[\frac{(\tilde{\boldsymbol{b}}^* - \boldsymbol{b}^*)'(\boldsymbol{X}'\boldsymbol{X})(\tilde{\boldsymbol{b}}^* - \boldsymbol{b}^*)}{2\,(1+\psi)} + \frac{\widetilde{\text{RSS}}^* + \text{RSS}^*}{\psi}\right]^{-\frac{2n-p+\delta-1}{2}} \frac{e^{-(n-p)\psi}}{\psi^2(1+\psi)^{\frac{p}{2}}} d\psi$$

Bayes Estimators of $\boldsymbol{\beta}$ and σ^2

The Bayes estimators for the parameters are calculated as follows:

$$\hat{\boldsymbol{\beta}}_{\text{BAYES}} = \text{E}(\boldsymbol{\beta} \,|\, \boldsymbol{b}^*, \text{RSS}^*) = \text{E}_\psi \, \text{E}_{\sigma^2} \, \text{E}(\boldsymbol{\beta} \,|\, \boldsymbol{b}^*, \text{RSS}^*, \sigma^2, \psi) = \text{E}_\psi \, \text{E}_{\sigma^2}(\boldsymbol{b}^*) = \boldsymbol{b}^*$$

$$\begin{aligned}\hat{\sigma}^2_{\text{BAYES}} &= \text{E}(\sigma^2 \,|\, \boldsymbol{b}^*, \text{RSS}^*) = \text{E}_\psi \, \text{E}(\sigma^2 \,|\, \boldsymbol{b}^*, \text{RSS}^*, \psi) = \text{E}_\psi(\tfrac{\text{RSS}^*}{\psi(n-p+\delta-3)}) = \tfrac{\text{RSS}^*}{(n-p+\delta-3)} \, \text{E}_\psi(\tfrac{1}{\psi}) \\ &= \frac{(n-p)\text{RSS}^*}{(n-p+\delta-3)^2}\end{aligned}$$

as long as $n > \max\{p, p-\delta+3\}$. Here we use the result that if $X \sim$ Scale-inv-$\chi^2(\nu, \tau^2)$ then $\text{E}(X) = \dfrac{\tau^2 \nu}{\nu - 2}$ for $\nu > 2$.

Credible Sets for $\boldsymbol{\beta}$ and σ^2

We will compute pivots (we are misusing the definition a bit, we merely mean a function of data and parameters whose posterior distribution does not depend on parameters) for σ^2 and $\boldsymbol{\beta}$. Given that we have closed form posterior expressions in the above equations, we can write down exact posterior intervals in terms of credibility and coverage.

A pivot for σ^2 can be defined as

$$V \coloneqq \frac{\text{RSS}^*}{\sigma^2}$$

whose distribution is calculated as $V \sim V_1 \times V_2$ where $(n-p)V_1$, V_2 are independent $\chi^2_{n-p+\delta-1}$ random variables (r.v.'s) due to (2.10). A $(1-\gamma)$ level credible set for σ^2 based on $V = \text{RSS}^*/\sigma^2$ is

$$\left[\frac{\text{RSS}^*}{b_{n,p,\delta;\gamma}}, \frac{\text{RSS}^*}{a_{n,p,\delta;\gamma}}\right]$$

where $a_{n,p,\delta;\gamma}$ and $b_{n,p,\delta;\gamma}$ are any two constants that satisfy $1 - \gamma = P(a_{n,p,\delta;\gamma} \leq V \leq$

$b_{n,p,\delta;\gamma}$). The length of the credible interval is $\text{RSS}^* \left(\frac{1}{a_{n,p,\delta;\gamma}} - \frac{1}{b_{n,p,\delta;\gamma}} \right)$.

Next we define a pivot for $\boldsymbol{\beta}$. From (2.7)

$$\frac{\boldsymbol{C}^{-1/2}(\boldsymbol{\beta} - \boldsymbol{b}^*)}{\sqrt{\text{RSS}^*}} \stackrel{d}{=} \boldsymbol{Y}_1$$

where $\boldsymbol{Y}_1 \stackrel{d}{=} \sqrt{\frac{1}{V_2}\left(\frac{1}{V_1} + 1\right)}\, \boldsymbol{U}$ such that V_1, V_2 are defined as before and are independent of $\boldsymbol{U} \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$. Finally we define the pivot for $\boldsymbol{\beta}$ as

$$T^2 := \frac{(\boldsymbol{\beta} - \boldsymbol{b}^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{\beta} - \boldsymbol{b}^*)}{\text{RSS}^*}$$

whose distribution is given by

$$T^2 \sim \frac{p}{n - p + \delta - 1} \left(\frac{n - p}{\chi^2_{n-p+\delta-1}} + 1 \right) F_{p,\, n-p+\delta-1}$$

where the χ^2 and F-distributions above are independent. A $(1 - \gamma)$ level credible ellipsoid for $\boldsymbol{\beta}$ based on T^2 is given by

$$\{\boldsymbol{\beta} \ : \ T^2 \leq c_{n,p,\delta;\gamma}\}$$

where $c_{n,p,\delta;\gamma}$ satisfies $1 - \gamma = P(T^2 \leq c_{n,p,\delta;\gamma})$. The volume of the credible ellipsoid is

$$V_{\boldsymbol{\beta}}(\boldsymbol{z}, \boldsymbol{X}) = \frac{\pi^{p/2}}{\Gamma\left(\frac{p}{2} + 1\right)} \left(c_{n,p,\delta;\gamma} \text{RSS}^*\right)^{p/2} |\boldsymbol{X}'\boldsymbol{X}|^{-1/2}$$

The above expression follows from the fact that if $\mathcal{A}$ is a $p \times p$ dimensional positive definite (PD) matrix, $\boldsymbol{a} \in \mathbb{R}^p$, and $C > 0$, then the volume of the ellipsoid $\{\boldsymbol{b} \in \mathbb{R}^p \ : (\boldsymbol{b} - \boldsymbol{a})'\mathcal{A}(\boldsymbol{b} - \boldsymbol{a}) \leq C\}$ is $[\pi^{p/2}/\Gamma(\frac{p}{2} + 1)] C^{p/2} |\mathcal{A}|^{-1/2}$. It is worth noting here that it is easy to show that none of the credible intervals are confidence intervals.

Remark 2.1.1. *If one is interested in the credible set of a single regression coefficient or more generally in the credible set of a vector of linear combination of $\boldsymbol{\beta}$, namely, $\boldsymbol{A\beta} = \boldsymbol{\eta}$ where $\boldsymbol{A}$ is a $k \times p$ dimensional matrix with rank$(\boldsymbol{A}) = k \leq p$, we define $T^2_{\boldsymbol{\eta}} = (\boldsymbol{\eta} - \boldsymbol{Ab}^*)'\{\boldsymbol{A}(\boldsymbol{X'X})^{-1}\boldsymbol{A'}\}^{-1}(\boldsymbol{\eta} - \boldsymbol{Ab}^*)/RSS^*$, and proceed by noting that*

$$T^2_{\boldsymbol{\eta}} \sim \frac{k(n-p)}{n-p+\delta-1}\left(\frac{1}{\chi^2_{n-p+\delta-1}} + 1\right) F_{k,\, n-p+\delta-1}$$

where the χ^2 and F-distributions above are independent.

2.2 Posterior Predictive Sampling method

We now proceed as follows to generate the singly imputed synthetic data $\boldsymbol{z} = (z_1, \ldots, z_n)$ under *posterior predictive sampling*. We start from a joint prior distribution $\pi(\boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{-\frac{\alpha+1}{2}}$ for $\boldsymbol{\beta} \in \mathbb{R}^p$, $\sigma^2 > 0$ and $\alpha > 0$ resulting in the posterior

$$\sigma^2 \,|\, \boldsymbol{b}, \mathrm{RSS} \;\sim\; \text{Scale-inv-}\chi^2\left(n-p+\alpha-1, \frac{\mathrm{RSS}}{n-p+\alpha-1}\right) \tag{2.11}$$

$$\boldsymbol{\beta} \,|\, \boldsymbol{b}, \mathrm{RSS}, \sigma^2 \;\sim\; N_p\left(\boldsymbol{b}, \sigma^2(\boldsymbol{X'X})^{-1}\right) \tag{2.12}$$

We assume throughout that $n + \alpha > p + 1$. We first draw $(\boldsymbol{\beta}^*, \sigma^*)$ from the above posterior, and then independently $z_i \sim N(\boldsymbol{x}_i'\boldsymbol{\beta}^*, (\sigma^*)^2)$, $i = 1, \ldots, n$. As before, $\boldsymbol{b}^* = (\boldsymbol{X'X})^{-1}\boldsymbol{X'z}$ and $\mathrm{RSS}^* = (\boldsymbol{z} - \boldsymbol{Xb}^*)'(\boldsymbol{z} - \boldsymbol{Xb}^*)$, which are jointly sufficient for $(\boldsymbol{\beta}, \sigma^2)$ by Lemma 1.2.1.

Similarly as in the last section we can write

$$\boldsymbol{z} \overset{d}{=} \boldsymbol{X\beta}^* + \sigma^*\boldsymbol{W}$$

where $\boldsymbol{W} \sim N_n(\boldsymbol{0}, \boldsymbol{I}_n)$. Then the sufficient statistics based on the released data can be written as

$$\begin{aligned} \boldsymbol{b}^* = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{z} &\stackrel{d}{=} \boldsymbol{\beta}^* + \sigma^*(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W} \stackrel{d}{=} \boldsymbol{\beta}^* + \sigma^*\boldsymbol{C}\boldsymbol{U}_1 \\ \text{RSS}^* = \boldsymbol{z}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{z} &\stackrel{d}{=} \sigma^{*2}\boldsymbol{W}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{W} \stackrel{d}{=} \sigma^{*2}V \end{aligned} \tag{2.13}$$

where $\boldsymbol{U}_1 \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$, $V \sim \chi^2_{n-p}$, $\boldsymbol{C}$ is such that $\boldsymbol{C}\boldsymbol{C}' = (\boldsymbol{X}'\boldsymbol{X})^{-1}$, $\boldsymbol{b}^*$ and RSS^* are independent. Thus, we get

$$\begin{aligned} \boldsymbol{b}^* &\stackrel{d}{=} \boldsymbol{\beta}^* + \sigma(\sigma^*/\sigma)\boldsymbol{C}\boldsymbol{U}_1 \stackrel{d}{=} \boldsymbol{\beta}^* + \sigma\sqrt{\psi}\boldsymbol{C}\boldsymbol{U}_1 \\ \text{RSS}^* &\stackrel{d}{=} \sigma^2(\sigma^*/\sigma)^2 V \stackrel{d}{=} \sigma^2\psi V \end{aligned} \tag{2.14}$$

where $\psi = (\sigma^*/\sigma)^2$ is a latent quantity. From (2.12) and (2.2), we have

$$\boldsymbol{\beta}^* \stackrel{d}{=} \boldsymbol{b} + \sigma^*\boldsymbol{C}\boldsymbol{U}_0 \stackrel{d}{=} \boldsymbol{\beta} + \sigma\boldsymbol{C}\boldsymbol{U}^0 + \sigma^*\boldsymbol{C}\boldsymbol{U}_0 \stackrel{d}{=} \boldsymbol{\beta} + \sigma\sqrt{1+\psi}\boldsymbol{C}\boldsymbol{U}_2$$

where $\boldsymbol{U}_0, \boldsymbol{U}^0, \boldsymbol{U}_2 \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$ are all independent of each other and of $\boldsymbol{U}_1$ and hence from (2.14) conditional on the parameters, we could write

$$\boldsymbol{b}^* \stackrel{d}{=} \boldsymbol{\beta} + \sigma\sqrt{1+2\psi}\boldsymbol{C}\boldsymbol{U}_3$$

where $\boldsymbol{U}_3 \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$. Thus the likelihood based on the released data for the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \psi)$ is given by

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2, \psi \,|\, \boldsymbol{b}^*, \text{RSS}^*) = \phi_p(\boldsymbol{b}^*; \boldsymbol{\beta}, \sigma^2(1+2\psi)(\boldsymbol{X}'\boldsymbol{X})^{-1})\, h(\text{RSS}^*; n-p, \sigma^2\psi) \tag{2.15}$$

The prior on ψ is naturally imposed by the original MLR model and the single imputation method. From (2.11), $\text{RSS}/\sigma^{*2} \,|\, \text{RSS} \sim \chi^2_{n-p+\alpha-1}$ and thus unconditionally $\text{RSS}/\sigma^{*2} \sim \chi^2_{n-p+\alpha-1}$ which also implies RSS/σ^{*2} is independent of RSS. Hence

$$\psi = \frac{\sigma^{*2}}{\sigma^2} = \frac{\text{RSS}/\sigma^2}{\text{RSS}/\sigma^{*2}} \stackrel{d}{=} \frac{n-p}{n-p+\alpha-1} F_{n-p,\, n-p+\alpha-1} \stackrel{d}{=} \beta'\left(\frac{n-p}{2}, \frac{n-p+\alpha-1}{2}\right)$$

For Bayesian inference on the other unknown parameters we consider the same independent non-informative improper priors as before. Thus for $\delta > 0$ we assume $\pi(\boldsymbol{\beta}, \sigma^2) = \pi(\boldsymbol{\beta})\pi(\sigma^2) \propto (\sigma^2)^{-\frac{\delta+1}{2}}$. The conditional posteriors can be determined similarly as in the last section as follows:

$$\begin{aligned}
\boldsymbol{\beta} \,|\, \boldsymbol{b}^*, \mathrm{RSS}^*, \sigma^2, \psi &\sim N_p\left(\boldsymbol{b}^*, \sigma^2(1+2\psi)(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) && (2.16)\\
\sigma^2 \,|\, \boldsymbol{b}^*, \mathrm{RSS}^*, \psi &\sim \text{Scale-inv-}\chi^2\left(n-p+\delta-1, \frac{\mathrm{RSS}^*}{\psi(n-p+\delta-1)}\right) && (2.17)\\
\psi &\sim \beta'\left(\frac{n-p+\delta-1}{2}, \frac{n-p+\alpha-\delta}{2}\right) && (2.18)
\end{aligned}$$

The posterior distributions are proper as long as $n > \max\{p, p-\delta+1, p-\alpha+1, p-\alpha+\delta\}$ and as before $\dfrac{\mathrm{RSS}^*}{\sigma^2\psi} \sim \chi^2_{n-p+\delta-1}$ and thus $\dfrac{\mathrm{RSS}^*}{\sigma^2\psi}$ is independent of the data and ψ.

Marginal Posterior of parameters

$$\boldsymbol{\beta} \,|\, \boldsymbol{b}^*, \mathrm{RSS}^*, \psi \sim \boldsymbol{t}_{n-p+\delta-1}\left(\boldsymbol{b}^*, \frac{\mathrm{RSS}^*(1+2\psi)}{\psi(n-p+\delta-1)}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right)$$

$$\pi(\sigma^2 \,|\, \mathrm{RSS}^*) \propto (\sigma^2)^{-\frac{n-p+\delta+1}{2}}\, \mathrm{U}\left(\frac{2n-2p+\alpha-1}{2}, 1, \frac{\mathrm{RSS}^*}{2\sigma^2}\right)$$

where $\mathrm{U}(a, b, x)$ is the confluent hypergeometric function of the second kind.

Marginal Likelihood of data

$$\pi(\boldsymbol{b}^*, \mathrm{RSS}^*) = \int \pi(\boldsymbol{b}^*, \mathrm{RSS}^*, \psi \,|\, \boldsymbol{\beta}, \sigma^2)\, \pi(\boldsymbol{\beta}, \sigma^2)\, d\boldsymbol{\beta}\, d\sigma^2 d\psi \propto (\mathrm{RSS}^*)^{-\frac{\delta+1}{2}}$$

Posterior Predictive Density

$$\begin{aligned}
&\pi(D_{\text{new}} \,|\, D) = \int \pi(D_{\text{new}} \,|\, \boldsymbol{\beta}, \sigma^2, \psi)\, \pi(\boldsymbol{\beta}, \sigma^2, \psi \,|\, D)\, d\boldsymbol{\beta}\, d\sigma^2\, d\psi\\
&\propto (\widetilde{\mathrm{RSS}}^*)^{\frac{n-p}{2}-1} \int \left[\frac{(\tilde{\boldsymbol{b}}^* - \boldsymbol{b}^*)'(\boldsymbol{X}'\boldsymbol{X})(\tilde{\boldsymbol{b}}^* - \boldsymbol{b}^*)}{2\,(1+2\psi)} + \frac{\widetilde{\mathrm{RSS}}^* + \mathrm{RSS}^*}{\psi}\right]^{-\frac{2n-p+\delta-1}{2}} \frac{(1+\psi)^{-2n-2p+\alpha-1}}{\psi^2(1+2\psi)^{\frac{p}{2}}} d\psi
\end{aligned}$$

Bayes Estimators of $\boldsymbol{\beta}$ and σ^2

$$\hat{\boldsymbol{\beta}}_{\text{BAYES}} = \text{E}(\boldsymbol{\beta} \,|\, \boldsymbol{b}^*, \text{RSS}^*) = \text{E}_\psi \, \text{E}_{\sigma^2} \, \text{E}(\boldsymbol{\beta} \,|\, \boldsymbol{b}^*, \text{RSS}^*, \sigma^2, \psi) = \text{E}_\psi \, \text{E}_{\sigma^2}(\boldsymbol{b}^*) = \boldsymbol{b}^*$$

$$\begin{aligned}\hat{\sigma}^2_{\text{BAYES}} &= \text{E}(\sigma^2 \,|\, \boldsymbol{b}^*, \text{RSS}^*) = \text{E}_\psi \, \text{E}(\sigma^2 \,|\, \boldsymbol{b}^*, \text{RSS}^*, \psi) = \text{E}_\psi(\tfrac{\text{RSS}^*}{\psi(n-p+\delta-3)}) = \tfrac{\text{RSS}^*}{(n-p+\delta-3)} \, \text{E}_\psi(\tfrac{1}{\psi}) \\ &= \frac{(n-p+\alpha-\delta)\text{RSS}^*}{(n-p+\delta-3)^2}\end{aligned}$$

as long as $n > \max\{p, p-\delta+3, p-\alpha+1, p-\alpha+\delta\}$. Here we make use of the facts that if $X \sim \beta'(\alpha, \beta)$ then $X^{-1} \sim \beta'(\beta, \alpha)$ and $\text{E}(X) = \frac{\alpha}{\beta-1}$ for $\beta > 1$.

Credible Sets for $\boldsymbol{\beta}$ and σ^2

As $\dfrac{\text{RSS}^*}{\sigma^2\psi}$ is independent of ψ so a pivot for σ^2 can be defined as

$$N := \frac{\text{RSS}^*}{\sigma^2} = \left(\frac{\text{RSS}^*}{\sigma^2\psi}\right)\psi = N_1 \times N_2$$

where $N_1 \sim \chi^2_{2\zeta}$ is independent of $N_2 \sim \beta'(\zeta, \eta)$ where $\eta = \frac{n-p+\alpha-\delta}{2}$, $\zeta = \frac{n-p+\delta-1}{2}$.
A $(1-\gamma)$ level credible set for σ^2 based on $N = \text{RSS}^*/\sigma^2$ is

$$\left[\frac{\text{RSS}^*}{b_{n,p,\alpha,\delta;\gamma}}, \frac{\text{RSS}^*}{a_{n,p,\alpha,\delta;\gamma}}\right]$$

where $a_{n,p,\alpha,\delta;\gamma}$ and $b_{n,p,\alpha,\delta;\gamma}$ are any two constants that satisfy $1-\gamma = P(a_{n,p,\alpha,\delta;\gamma} \leq N \leq b_{n,p,\alpha,\delta;\gamma})$. The length of the credible interval is $\text{RSS}^* \left(\dfrac{1}{a_{n,p,\alpha,\delta;\gamma}} - \dfrac{1}{b_{n,p,\alpha,\delta;\gamma}}\right)$.

Let us now consider

$$T^2 := \frac{(\boldsymbol{\beta}-\boldsymbol{b}^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{\beta}-\boldsymbol{b}^*)}{\text{RSS}^*}$$

We will compute the posterior distribution of $T^2 \,|\, \boldsymbol{b}^*, \text{RSS}^*$. Observe that we can write

$$T^2 = \left[\frac{(\boldsymbol{\beta}-\boldsymbol{b}^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{\beta}-\boldsymbol{b}^*)}{\sigma^2(1+2\psi)}\right]\left[\frac{\sigma^2\psi}{\text{RSS}^*}\right]\left[\frac{1+2\psi}{\psi}\right] = T_1 \times T_2 \times T_3$$

Now

(a) $T_1 \,|\, \boldsymbol{b}^*, \mathrm{RSS}^*, \sigma^2, \psi \sim \chi^2_p$ and hence $T_1 \sim \chi^2_p$ unconditionally. This also shows that T_1 is independent of $(\boldsymbol{b}^*, \mathrm{RSS}^*, \sigma^2, \psi)$ and thus T_1 is independent of T_2 and T_3.

(b) $T_2 \sim \text{inv-}\chi^2_{2\zeta}$ and is independent of T_3.

(c) $T_3 - 2 \sim \beta'(\eta, \zeta)$ or alternatively $T_3 \stackrel{d}{=} 1 + \frac{1}{M}$ where $M \sim \beta(\zeta, \eta)$. This is because if $X \sim \beta'(\alpha, \beta)$ then $\frac{1}{X} \sim \beta'(\beta, \alpha)$ and $\frac{1}{1+X} \sim \beta(\beta, \alpha)$.

Hence finally we see that T^2 is a pivot for β and

$$T^2 \sim \frac{p}{n-p+\delta-1} F_{p,\, n-p+\delta-1} \left(1 + \frac{1}{M}\right) \qquad \text{where } M \sim \beta(\zeta, \eta)$$

A $(1-\gamma)$ level credible ellipsoid for $\boldsymbol{\beta}$ based on T^2 is given by

$$\{\boldsymbol{\beta} \;:\; T^2 \leq c_{n,p,\alpha,\delta;\gamma}\}$$

where $c_{n,p,\alpha,\delta;\gamma}$ satisfies $1-\gamma = P(T^2 \leq c_{n,p,\alpha,\delta;\gamma})$. The volume of the credible ellipsoid is

$$V_{\boldsymbol{\beta}}(\boldsymbol{z}, \boldsymbol{X}) = \frac{\pi^{p/2}}{\Gamma\left(\frac{p}{2}+1\right)} \left(c_{n,p,\alpha,\delta;\gamma} \mathrm{RSS}^*\right)^{p/2} |\boldsymbol{X}'\boldsymbol{X}|^{-1/2}$$

Remark 2.2.1. *If one is interested in the credible set of a single regression coefficient or more generally in the credible set of a vector of linear combination of $\boldsymbol{\beta}$, namely, $\boldsymbol{A\beta} = \boldsymbol{\eta}$ where $\boldsymbol{A}$ is a $k \times p$ dimensional matrix with $rank(\boldsymbol{A}) = k \leq p$, we define $T^2_{\boldsymbol{\eta}} = (\boldsymbol{\eta} - \boldsymbol{Ab}^*)'\{\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\}^{-1}(\boldsymbol{\eta} - \boldsymbol{Ab}^*)/RSS^*$, and proceed by noting that*

$$T^2 \sim \frac{k}{n-p+\delta-1} F_{k,\, n-p+\delta-1} \left(1 + \frac{1}{M}\right) \qquad \textit{where } M \sim \beta(\zeta, \eta)$$

2.3 Simulation Studies

In order to conduct the simulation, the population distribution is taken to be the linear regression model (2.1) with

$$p = 10, \qquad \boldsymbol{x}_i = \begin{pmatrix} 1 \\ x_{1i} \\ x_{2i} \\ x_{3i} \\ x_{4i} \\ I(x_{5i} = 2) \\ I(x_{5i} = 3) \\ I(x_{5i} = 4) \\ I(x_{5i} = 5) \\ I(x_{5i} = 6) \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \\ \beta_9 \\ \beta_{10} \end{pmatrix} = \begin{pmatrix} 10 \\ 2 \\ 2 \\ -3 \\ -1 \\ -2 \\ 1 \\ 2 \\ 2 \\ 4 \end{pmatrix}, \qquad \sigma^2 = 1. \tag{2.19}$$

The regressor variables in $\boldsymbol{x}_i$ are generated one time at the beginning of the simulation, and then held fixed from one iteration to the next. We generate the regressor variables (all independently) as follows:

$$x_{1i} \sim N(1,1), \qquad \log x_{2i} \sim N(0,1), \qquad x_{3i} \sim \text{Exponential(mean} = 1),$$

$$x_{4i} \sim \text{Poisson}(1), \qquad x_{5i} = \begin{cases} 1 \text{ with probability } 0.2 \\ 2 \text{ with probability } 0.1 \\ 3 \text{ with probability } 0.2 \\ 4 \text{ with probability } 0.2 \\ 5 \text{ with probability } 0.2 \\ 6 \text{ with probability } 0.1 \end{cases}$$

Based on Monte Carlo simulation with 10^4 iterations, we compute an estimate of the coverage probability, the volume or length (as appropriate) of the respective credible sets and the Bayes estimators of the parameters, where in all cases, the level of credibility is set at 0.95.

Plug-In Sampling Tables 2.1, 2.2, 2.3 includes the simulation results for a plug-in sampling data where the sample size n equals 500, 1000 and 10000 respectively for different values of the tuning parameter δ. Some interesting observations are in order. The coverage for σ^2 gets slightly better initially as we increase δ, starts worsening beyond $\delta \geq 10$, and at large values of δ it is significantly worse. This effect is more prominent when n is small, in which case the coverage is not the best anyway as is to be expected. The same effect is observed for the coverage of $\boldsymbol{\beta}$ though not as severe. The coverage of $\boldsymbol{\beta}$ decreases at a much slower rate compared to that of σ^2 with increasing δ. The size of the credible sets shrink for both the parameters as n or δ increases. With decreasing n or increasing δ there seems to be no effect on the Bayes estimator of $\boldsymbol{\beta}$, while the Bayes estimator of σ^2 becomes slightly worse, which is what we expect since $\hat{\boldsymbol{\beta}}_{\text{BAYES}}$ does not involve δ while σ^2_{BAYES} has δ in the denominator. All of this suggests that there is a sweet spot for the choice of δ to ensure maximum coverage along with the smallest possible size of the credible sets of the parameters. For both σ^2 and $\boldsymbol{\beta}$, from Table 2.3 asymptotically the results imply that the Bernstein-von Mises theorem holds, with the caveat that inference worsens with increasing δ, quicker for σ^2 than for $\boldsymbol{\beta}$. In the asymptotic case, the credible sets are tighter and the Bayes estimators perform admirably for both the parameters, as expected. The behavior of the coverage of σ^2 and $\boldsymbol{\beta}$ with respect to different values of δ in the case $n = 500$ (depicted by alternating dashes and dots), $n = 1000$ (depicted by solid lines), asymptotic case $n = 10000$ (depicted by dashed lines) are represented in Figure 2.1(a) and Figure 2.1(b) respectively.

Posterior Predictive Sampling The general trend of Bayesian inference for model parameters observed under PIS is also mirrored when data is generated by posterior predictive sampling, as illustrated in Tables 2.4, 2.5, 2.6, 2.7, 2.8 and 2.9. Overall for σ^2, compared to PIS, the coverage is lower, the credible interval is wider, but the Bayes estimator performs similarly well. For $\boldsymbol{\beta}$, compared to PIS, the coverage is similar, the Bayes estimator performs similarly well, but the volume of the credible ellipsoid is one order of magnitude bigger. The interaction of the hyperparameter α and tuning parameter δ is also pretty interesting to observe. Increasing α seems to have no effect on the coverage of the parameters but the size of the credible sets narrow down marginally, although asymptotically there seems to be no significant difference (as seen by comparing Tables 2.6 and 2.9). We should be able to find a combination of the two that yields the best inference. The inference for $\boldsymbol{\beta}$ seems to be unaffected by the increase in α, except again for the fact that the credible set for $\boldsymbol{\beta}$ contracts a bit. The behavior of the coverage of σ^2 and $\boldsymbol{\beta}$ with respect to different values of δ in the case $n = 500$ (depicted by alternating dashes and dots), $n = 1000$ (depicted by solid lines), asymptotic case $n = 10000$ (depicted by dashed lines) are represented in Figures 2.1(c), 2.1(e) and Figures 2.1(d), 2.1(f) respectively.

After assessing the results, the recommendation would be to use $2 \leq \delta \leq 4$.

Table 2.1: Inference for $\boldsymbol{\beta}$ and σ^2 for PIS data with $n = 500$

	σ^2			β		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.953	0.360	1.011	0.953	1.064e-03	(10.002, 2.000, 2.000, -2.999, -1.000, -2.000, 0.997, 1.998, 1.998, 3.998)'
0.5	0.948	0.359	1.010	0.950	8.814e-04	(10.001, 1.999, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 4.000)'
0.8	0.952	0.359	1.009	0.950	8.248e-04	(9.998, 2.000, 2.000, -2.999, -1.001, -2.000, 1.006, 2.001, 2.002, 4.005)'
1	0.951	0.359	1.010	0.949	9.708e-04	(10.002, 2.001, 2.000, -3.000, -0.999, -2.004, 0.997, 1.995, 1.998, 3.998)'
2	0.951	0.357	1.005	0.949	8.054e-04	(10.000, 2.001, 2.000, -2.999, -0.999, -2.003, 0.998, 1.996, 1.998, 3.997)'
3	0.948	0.355	1.000	0.947	3.988e-04	(9.998, 2.000, 2.000, -3.000, -1.000, -2.001, 1.002, 1.998, 2.000, 3.997)'
4	0.945	0.353	0.996	0.946	6.732e-04	(9.997, 2.001, 2.001, -3.000, -0.999, -2.000, 1.002, 1.999, 2.001, 4.000)'
10	0.933	0.342	0.972	0.944	4.310e-04	(10.000, 2.000, 2.000, -3.001, -1.000, -2.002, 1.000, 2.002, 2.001, 4.001)'
20	0.863	0.326	0.934	0.931	4.779e-04	(10.001, 2.000, 2.001, -3.001, -1.001, -2.002, 1.000, 2.002, 2.003, 4.000)'
30	0.745	0.310	0.899	0.919	4.963e-04	(9.997, 2.000, 2.000, -2.999, -1.000, -2.000, 1.002, 2.001, 2.003, 4.000)'
50	0.426	0.282	0.834	0.898	3.861e-04	(10.000, 2.000, 2.001, -2.999, -1.001, -2.003, 1.002, 1.998, 1.998, 3.998)'
100	0.010	0.226	0.697	0.825	1.940e-04	(9.998, 2.000, 2.000, -3.000, -1.000, -2.000, 1.001, 2.000, 2.000, 4.000)'

Table 2.2: Inference for $\boldsymbol{\beta}$ and σ^2 for PIS data with $n = 1000$

	σ^2			β		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.951	0.251	1.006	0.949	2.475e-05	(10.002, 1.999, 2.000, -3.000, -1.000, -2.001, 1.000, 1.999, 1.999, 4.001)'
0.5	0.953	0.251	1.004	0.951	2.130e-05	(10.002, 1.999, 2.000, -3.000, -1.000, -2.000, 0.999, 1.997, 1.999, 4.000)'
0.8	0.951	0.250	1.003	0.953	2.190e-05	(10.000, 2.000, 2.000, -3.000, -1.000, -1.998, 1.000, 2.000, 2.000, 4.000)'
1	0.950	0.251	1.004	0.951	2.215e-05	(10.002, 1.999, 2.000, -3.000, -1.000, -2.001, 0.998, 1.998, 1.997, 3.997)'
2	0.949	0.250	1.004	0.946	2.362e-05	(10.000, 2.000, 2.000, -3.001, -1.001, -2.000, 1.002, 2.000, 2.000, 3.999)'
3	0.949	0.250	1.001	0.947	2.472e-05	(10.000, 2.000, 2.000, -3.000, -1.000, -1.998, 1.001, 2.001, 1.990, 3.997)'
4	0.948	0.248	0.997	0.949	2.174e-05	(10.000, 2.000, 2.000, -3.000, -1.000, -1.998, 0.999, 2.000, 2.000, 4.000)'
10	0.939	0.245	0.986	0.943	1.562e-05	(10.000, 2.000, 2.000, -3.000, -1.000, -2.001, 1.000, 2.002, 2.001, 3.999)'
20	0.906	0.239	0.965	0.943	2.197e-05	(10.001, 2.000, 2.000, -3.000, -1.000, -1.999, 1.001, 1.998, 1.998, 3.996)'
30	0.843	0.233	0.947	0.935	2.156e-05	(10.001, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 1.999, 2.000, 4.000)'
50	0.661	0.222	0.912	0.926	1.534e-05	(9.996, 2.000, 2.001, -3.001, -1.000, -1.997, 1.003, 2.005, 2.004, 4.004)'
100	0.133	0.198	0.830	0.899	1.121e-05	(10.000, 1.999, 2.000, -3.001, -1.000, -1.998, 1.001, 2.001, 1.999, 4.002)'

Table 2.3: Inference for $\boldsymbol{\beta}$ and σ^2 for PIS data with $n = 10000$

	σ^2			β		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.947	0.078	1.000	0.945	1.811e-10	(10.000, 2.000, 2.000, -3.000, -1.000, -2.001, 1.000, 1.999, 2.000, 4.000)'
0.5	0.949	0.078	1.000	0.950	1.938e-10	(10.000, 2.000, 2.000, -3.000, -1.000, -2.001, 1.000, 2.000, 2.000, 4.001)'
0.8	0.950	0.078	1.001	0.951	2.032e-10	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.001, 2.001, 2.000, 4.001)'
1	0.948	0.078	1.000	0.951	2.100e-10	(10.000, 2.000, 2.000, -3.000, -2.000, -2.000, 1.000, 2.000, 2.000, 3.999)'
2	0.950	0.078	1.000	0.950	2.001e-10	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.001, 4.000)'
3	0.949	0.078	1.000	0.947	2.087e-10	(9.999, 2.000, 2.000, -3.000, -1.000, -1.999, 1.001, 2.001, 2.000, 4.001)'
4	0.951	0.078	1.000	0.949	2.054e-10	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.001, 2.000, 4.000)'
10	0.947	0.078	0.999	0.952	1.925e-10	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.001, 2.000, 2.000, 4.001)'
20	0.946	0.078	0.997	0.951	1.898e-10	(10.000, 2.000, 3.000, -3.000, -1.000, -1.999, 1.001, 2.001, 2.000, 4.001)'
30	0.944	0.078	0.995	0.948	1.912e-10	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 4.000)'
50	0.927	0.078	0.991	0.944	1.828e-10	(10.000, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.000, 2.001, 4.000)'
100	0.823	0.077	0.981	0.946	1.612e-10	(10.000, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.000, 2.001, 4.001)'

Table 2.4: Inference for $\boldsymbol{\beta}$ and σ^2 for PPS data with $\alpha = 2$, $n = 500$

	σ^2			β		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.955	0.443	1.017	0.951	7.190e-04	(10.003, 2.001, 2.000, -3.001, -1.000, -2.004, 0.995, 1.999, 1.997, 3.997)'
0.5	0.955	0.441	1.015	0.949	5.615e-03	(9.996, 2.001, 2.000, -3.000, -1.000, -1.997, 1.005, 2.004, 2.004, 4.005)'
0.8	0.949	0.441	1.013	0.946	5.364e-03	(10.000, 1.999, 2.000, -2.998, -1.001, -1.996, 0.999, 2.001, 1.999, 3.999)'
1	0.947	0.441	1.013	0.949	5.072e-03	(9.999, 2.000, 2.000, -3.000, -0.999, -1.998, 1.000, 2.001, 2.000, 4.000)'
2	0.944	0.438	1.006	0.950	4.864e-03	(10.001, 2.000, 2.000, -3.000, -1.000, -1.996, 0.998, 1.999, 2.002, 3.998)'
3	0.951	0.435	1.000	0.946	6.296e-03	(9.998, 2.001, 2.000, -3.000, -0.999, -2.000, 1.000, 1.999, 1.999, 4.002)'
4	0.948	0.432	0.994	0.952	8.090e-03	(10.001, 1.999, 2.000, -2.999, -1.001, -1.998, 1.000, 2.002, 2.002, 3.999)'
10	0.927	0.415	0.958	0.940	5.257e-03	(10.002, 2.000, 2.000, -3.001, -0.998, -2.005, 0.998, 1.996, 1.995, 3.997)'
20	0.818	0.389	0.901	0.929	5.398e-03	(9.998, 2.000, 2.000, -2.999, -1.000, -2.002, 1.000, 2.002, 2.003, 4.003)'
30	0.638	0.366	0.848	0.919	3.985e-03	(10.001, 2.000, 2.000, -3.000, -1.000, -1.998, 0.999, 1.998, 1.997, 3.996)'
50	0.232	0.323	0.752	0.891	2.997e-03	(10.001, 2.001, 2.001, -3.000, -1.000, -2.002, 0.999, 1.996, 1.998, 3.998)'
100	4e-04	0.239	0.559	0.805	1.247e-03	(9.998, 2.001, 2.000, -3.000, -1.000, -1.997, 1.003, 2.002, 2.003, 4.003)'

Table 2.5: Inference for $\boldsymbol{\beta}$ and σ^2 for PPS data with $\alpha = 2$, $n = 1000$

	σ^2			β		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.950	0.308	1.009	0.947	2.381e-04	(10.000, 2.001, 2.000, -3.000, -1.000, -1.999, 1.003, 1.999, 2.002, 4.004)'
0.5	0.947	0.308	1.008	0.951	1.560e-04	(10.000, 2.000, 2.000, -3.000, -1.000, -2.003, 1.001, 2.001, 2.000, 4.000)'
0.8	0.949	0.308	1.007	0.953	1.782e-04	(10.005, 1.999, 1.999, -3.000, -1.001, -2.002, 0.995, 1.997, 1.997, 3.996)'
1	0.945	0.308	1.007	0.949	1.602e-04	(10.000, 2.000, 2.000, -3.000, -1.000, -2.002, 0.999, 2.001, 2.002, 4.000)'
2	0.950	0.307	1.004	0.949	1.479e-04	(9.999, 2.000, 2.000, -3.000, -1.000, -2.000, 0.999, 1.999, 1.999, 4.003)'
3	0.951	0.306	1.000	0.950	2.086e-04	(9.999, 2.001, 2.000, -3.000, -1.000, -2.003, 0.998, 2.001, 2.001, 3.998)'
4	0.949	0.304	0.996	0.952	1.674e-04	(9.999, 2.000, 2.000, -3.000, -1.000, -1.999, 1.001, 2.001, 2.003, 4.006)'
10	0.937	0.299	0.979	0.944	1.424e-04	(9.999, 2.000, 2.000, -2.999, -1.000, -2.001, 1.000, 2.001, 2.001, 3.999)'
20	0.885	0.289	0.949	0.935	1.091e-04	(10.000, 2.000, 2.000, -3.000, -1.000, -2.002, 0.999, 2.000, 1.998, 3.998)'
30	0.796	0.280	0.921	0.937	1.361e-04	(10.002, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.001, 2.001, 4.001)'
50	0.521	0.263	0.867	0.922	1.195e-04	(10.000, 2.000, 2.000, -3.000, -1.001, -2.000, 1.000, 2.000, 2.000, 4.000)'
100	0.030	0.226	0.749	0.893	8.032e-05	(10.001, 2.000, 2.000, -2.999, -1.001, -2.004, 1.000, 1.998, 2.000, 3.999)'

Table 2.6: Inference for $\boldsymbol{\beta}$ and σ^2 for PPS data with $\alpha = 2$, $n = 10000$

	σ^2			β		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.951	0.096	1.001	0.948	1.536e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.000, 2.000, 4.001)'
0.5	0.954	0.096	1.001	0.953	1.519e-09	(9.999, 2.000, 2.000, -3.000, -1.000, -1.999, 1.001, 2.000, 2.000, 4.000)'
0.8	0.952	0.096	1.001	0.950	1.522e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.002, 0.999, 2.000, 1.999, 4.000)'
1	0.949	0.096	1.000	0.952	1.541e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 4.001)'
2	0.951	0.096	1.001	0.945	1.546e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 4.000)'
3	0.952	0.096	1.000	0.951	1.524e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.001, 4.000)'
4	0.949	0.096	1.000	0.950	1.589e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.001, 2.001, 4.001)'
10	0.948	0.096	0.998	0.950	1.592e-09	(10.001, 2.000, 2.000, -3.000, -1.000, -2.001, 1.000, 2.000, 2.000, 3.999)'
20	0.945	0.096	0.995	0.946	1.500e-09	(9.999, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.000, 2.000, 4.000)'
30	0.935	0.095	0.992	0.946	1.557e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.001, 1.000, 2.000, 2.000, 3.999)'
50	0.910	0.095	0.986	0.946	1.466e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.000, 2.000, 4.001)'
100	0.773	0.093	0.972	0.942	1.512e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.001, 0.999, 2.000, 2.000, 4.000)'

Table 2.7: Inference for $\boldsymbol{\beta}$ and σ^2 for PPS data with $\alpha = 50$, $n = 500$

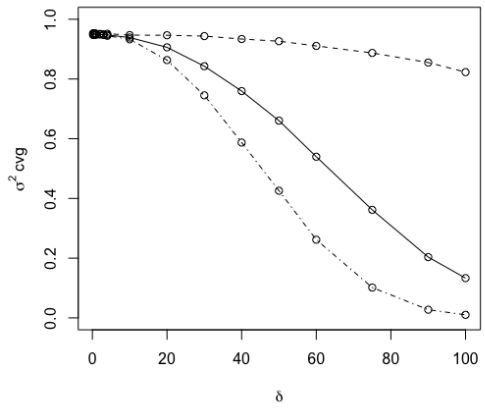
	σ^2			β		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.951	0.436	1.017	0.949	4.017e-03	(10.002, 2.000, 2.000, -3.000, -1.000, -1.998, 1.000, 1.999, 1.998, 3.999)'
0.5	0.952	0.435	1.014	0.953	4.325e-03	(10.000, 2.001, 2.000, -3.000, -0.999, -2.004, 1.000, 1.997, 1.996, 3.994)'
0.8	0.947	0.433	1.011	0.945	5.193e-03	(10.002, 1.999, 1.999, -3.000, -0.999, -2.003, 1.000, 1.999, 1.998, 4.001)'
1	0.953	0.433	1.009	0.953	4.016e-03	(9.998, 2.000, 2.001, -2.999, -1.001, -2.002, 1.001, 1.997, 2.000, 4.004)'
2	0.949	0.431	1.006	0.947	6.528e-03	(10.002, 2.000, 2.000, -3.000, -1.001, -2.001, 1.000, 1.996, 2.000, 4.000)'
3	0.949	0.428	0.999	0.949	5.483e-03	(10.001, 1.999, 2.000, -2.999, -0.999, -2.003, 0.998, 1.999, 1.998, 4.001)'
4	0.944	0.425	0.994	0.949	4.120e-03	(9.998, 2.001, 2.000, -3.000, -1.000, -1.999, 0.999, 1.997, 2.000, 4.003)'
10	0.920	0.409	0.959	0.938	4.192e-03	(10.002, 2.000, 2.000, -3.000, -1.002, -2.001, 1.000, 2.001, 1.999, 3.999)'
20	0.822	0.385	0.904	0.931	4.663e-03	(9.998, 2.000, 2.000, -3.000, -1.000, -1.997, 1.004, 2.001, 2.002, 4.001)'
30	0.640	0.361	0.852	0.919	3.832e-03	(10.001, 1.999, 2.000, -3.001, -1.000, -2.000, 0.999, 1.998, 2.000, 4.007)'
50	0.254	0.321	0.761	0.891	2.084e-03	(9.996, 2.001, 2.000, -3.000, -1.000, -1.995, 1.006, 2.005, 2.002, 4.003)'
100	4e-04	0.239	0.571	0.809	9.473e-04	(10.001, 2.000, 2.000, -3.001, -1.000, -1.999, 0.998, 1.999, 2.001, 3.996)'

Table 2.8: Inference for $\boldsymbol{\beta}$ and σ^2 for PPS data with $\alpha = 50$, $n = 1000$

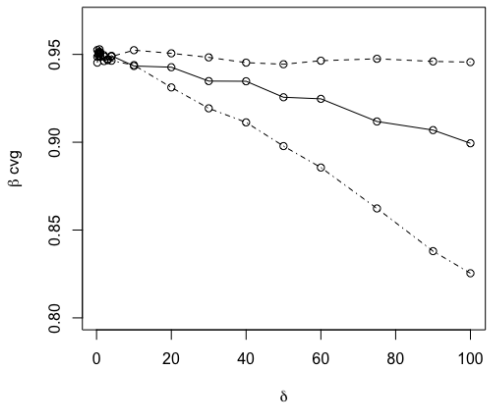
	σ^2			β		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.947	0.306	1.010	0.951	1.405e-04	(10.000, 2.000, 2.000, -3.000, -1.001, -2.004, 1.001, 2.000, 1.998, 3.999)'
0.5	0.948	0.305	1.007	0.950	1.726e-04	(10.002, 2.000, 2.000, -3.001, -1.000, -2.002, 0.997, 1.999, 1.999, 3.999)'
0.8	0.949	0.305	1.007	0.948	1.417e-04	(9.998, 2.000, 2.000, -2.999, -1.000, -1.999, 1.001, 1.999, 1.999, 3.999)'
1	0.953	0.305	1.005	0.945	1.535e-04	(10.000, 1.999, 2.000, -3.000, -1.000, -1.999, 1.001, 2.003, 2.001, 4.000)'
2	0.949	0.304	1.003	0.951	1.675e-04	(9.998, 2.000, 2.000, -3.000, -1.000, -2.001, 1.001, 2.002, 2.002, 4.003)'
3	0.947	0.303	1.000	0.948	1.135e-04	(10.002, 2.000, 2.000, -3.001, -1.001, -1.999, 1.001, 2.000, 2.003, 3.990)'
4	0.949	0.302	0.998	0.949	1.703e-04	(10.000, 2.000, 2.000, -3.000, -1.000, -1.997, 1.002, 2.001, 2.001, 4.000)'
10	0.936	0.297	0.980	0.947	1.364e-04	(10.002, 1.999, 2.000, -3.000, -1.001, -1.999, 0.999, 1.999, 2.000, 4.001)'
20	0.882	0.287	0.951	0.937	1.186e-04	(9.998, 2.001, 2.000, -3.000, -1.001, -1.996, 1.005, 2.002, 2.005, 4.002)'
30	0.791	0.278	0.922	0.933	1.424e-04	(10.000, 2.000, 2.000, -2.999, -1.000, -2.000, 1.000, 1.997, 2.000, 3.998)'
50	0.535	0.262	0.871	0.924	8.791e-05	(10.001, 2.000, 2.000, -3.000, -1.000, -2.003, 1.000, 1.999, 1.999, 4.003)'
100	0.034	0.225	0.752	0.888	4.545e-05	(9.999, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.000, 2.002, 3.997)'

Table 2.9: Inference for $\boldsymbol{\beta}$ and σ^2 for PPS data with $\alpha = 50$, $n = 10000$

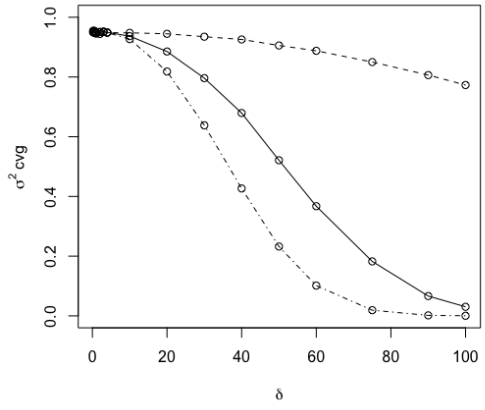
	σ^2			β		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.953	0.096	1.001	0.950	1.484e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.001, 1.000, 1.999, 1.999, 3.999)'
0.5	0.947	0.096	1.001	0.950	1.622e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.001, 2.000, 4.000)'
0.8	0.950	0.096	1.001	0.950	1.464e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.001, 2.000, 2.000, 4.001)'
1	0.947	0.096	1.001	0.952	1.604e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 0.999, 2.000, 2.000, 4.000)'
2	0.948	0.096	1.000	0.950	1.395e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 1.998, 1.999, 3.999)'
3	0.952	0.096	1.000	0.951	1.419e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -1.999, 1.001, 2.000, 2.001, 4.000)'
4	0.946	0.096	1.000	0.949	1.382e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.001, 3.000, 4.000)'
10	0.949	0.096	0.998	0.951	1.318e-09	(10.001, 2.000, 2.000, -3.000, -1.000, -2.000, 0.999, 2.000, 1.999, 3.999)'
20	0.940	0.095	0.994	0.948	1.449e-09	(9.999, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.001, 2.000, 4.000)'
30	0.935	0.095	0.992	0.946	1.519e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.000, 2.001, 4.000)'
50	0.909	0.095	0.986	0.950	1.437e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.001, 4.000)'
100	0.773	0.093	0.971	0.944	1.376e-09	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 4.001)'



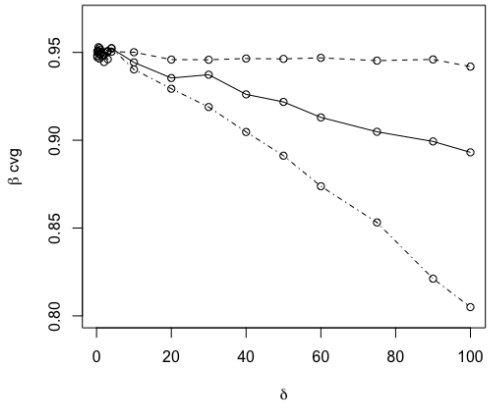
(a) σ^2 (PIS)



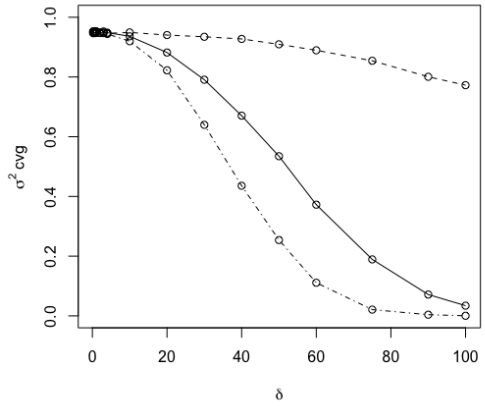
(b) $\boldsymbol{\beta}$ (PIS)



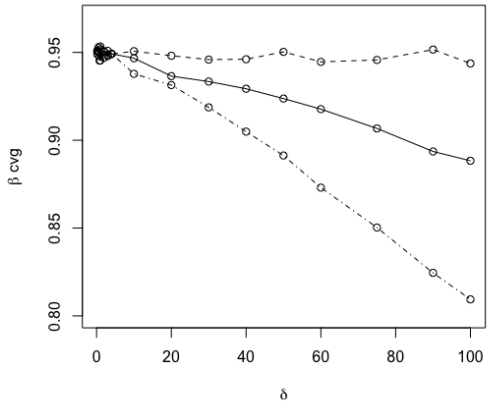
(c) σ^2, $\alpha = 2$ (PPS)



(d) $\boldsymbol{\beta}$, $\alpha = 2$ (PPS)



(e) σ^2, $\alpha = 50$ (PPS)



(f) $\boldsymbol{\beta}$, $\alpha = 50$ (PPS)

Figure 2.1: Variation in coverage of $\boldsymbol{\beta}$ and σ^2 with respect to δ for SI MLR data ($-\cdot-\cdot-$ $n = 500$, — $n = 1000$, $---$ $n = 10000$)

The PIS method offers smaller radius of the confidence sets than the PPS method and also gives estimates of the parameters closer to the ones obtained from the original data, despite giving slightly higher levels of disclosure risk (Moura (2016)). So we have a trade off between data utility and data privacy.

In general, the Bayesian posterior intervals, credible intervals and HPD intervals need not have valid frequentist coverage. This is because the Bayesian intervals are not derived using a repeated sampling paradigm; their objective is to characterize reasonable parameter values that conform with the specific model and prior for a given situation. However, some researchers have advocated a more principled approach to the practice where the Bayes intervals are calibrated to frequentist calculations so that Bayesian statements can be rejected based on empirical tests. Such calibrated Bayes approach (Rubin, 1984; Little, 2006) looks for reconciliation between the two paradigms. Another approach for reconciliation (asymptotically) is to choose priors that provided credible intervals with accurate frequentist coverage. Such priors are called *Probability Matching Priors* (Datta and Ghosh, 1995).

Usually, Bayesian credible intervals have good frequentist properties provided the problem admits some type of Bernstein-von Mises theorem. In the present case however, in the presence of latent structure, such Bernstein-von Mises results may not be readily available. From the limited empirical investigation it seems that the coverage of the credible intervals depends on the δ in the prior even asymptotically. It will be interesting to determine the limits of coverage as $\delta > 0$ varies. We will pursue such investigation in the future.

2.4 Partially Sensitive Data

We have assumed so far that all the n observations $\boldsymbol{y} = (y_1, \ldots, y_n)'$ in the multiple linear regression model are sensitive. Of course, this need not be the case, and quite generally we can partition $\boldsymbol{y}$ into two parts: $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ of dimensions r and $(n-r)$, respectively, and assume that the first r observations $\boldsymbol{y}_1$ are sensitive, thus requiring privacy protection, and the remaining $(n-r)$ observations $\boldsymbol{y}_2$ are non-sensitive, and can remain unprotected. Let $\boldsymbol{X} = [\boldsymbol{X}_1' \, \boldsymbol{X}_2']'$ be the corresponding partitioning of the matrix $\boldsymbol{X}$, so that $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are of dimensions $r \times p$ and $(n-r) \times p$, respectively. The reasons for some of the y-values being sensitive can vary depending on the context. For example, for income data, large incomes (extreme values) may be sensitive. The sensitive nature of y may also depend on the (extreme) values of the corresponding covariates $\boldsymbol{x}$. We outline below two data analysis procedures when the latter situation holds, namely, the sensitivity of the first r values of $\boldsymbol{y}$ is due to the nature of the covariates, which makes r a *non-random* integer.

Method I: Using only estimates of sensitive part to impute synthetic data

Plug-In Sampling

We propose to synthesize the r sensitive y-values $\boldsymbol{y}_1$ by applying the plug-in sampling method based on these r y-values, as discussed in Section 2.1. The reason for using only the sensitive part of the data for imputing synthetic data is to ensure that in the released data the synthetic part and the non-sensitive part are independent. The synthetic version of $\boldsymbol{y}_1$ is $\boldsymbol{y}_1^* = (y_1^*, \ldots, y_r^*)'$ such that $y_i^* \sim N(\boldsymbol{x}_i'\boldsymbol{b}_1, \hat{\sigma}_1^2)$ generated independently for $i = 1, \ldots, r$, where $\boldsymbol{b}_1 = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y}_1$ and $\text{RSS}_1 = \boldsymbol{y}_1'(\boldsymbol{I}_r -$

$\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y}_1$ are the sufficient statistics of $\boldsymbol{y}_1$, and $\hat{\sigma}_1^2 = \text{RSS}_1/(r-p)$. We assume that $r > p$ and $n - r > p$ so that we can draw valid inference about the p regression coefficients $\boldsymbol{\beta}$ separately for each data set. Thus similarly, $\boldsymbol{b}_2 = (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{y}_2$ and $\text{RSS}_2 = \boldsymbol{y}_2'(\boldsymbol{I}_{n-r} - \boldsymbol{P}_{\boldsymbol{X}_2})\boldsymbol{y}_2$ are the sufficient statistics of $\boldsymbol{y}_2$. The released data is $\boldsymbol{y}^* = (\boldsymbol{y}_1^{*\prime}, \boldsymbol{y}_2{}')'$. Then by Lemma 1.2.1 the sufficient statistics for the imputed data are

$$\boldsymbol{b}_1^* = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y}_1^* \stackrel{d}{=} \boldsymbol{b}_1 + \hat{\sigma}_1\boldsymbol{C}_1\boldsymbol{U}_0 \stackrel{d}{=} \boldsymbol{\beta} + \sigma\sqrt{1+\psi}\boldsymbol{C}_1\boldsymbol{U}_1$$

$$\text{RSS}_1^* = \boldsymbol{y}_1^{*\prime}(\boldsymbol{I}_r - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y}_1^* \stackrel{d}{=} \hat{\sigma}_1^2\boldsymbol{W}_1'(\boldsymbol{I}_r - \boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{W}_1 \stackrel{d}{=} \sigma^2\psi V_1$$

where $\boldsymbol{U}_0, \boldsymbol{U}_1 \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$ independently, $\boldsymbol{C}_1\boldsymbol{C}_1' = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}$, $\psi = (\hat{\sigma}_1/\sigma)^2$ is a latent quantity, $\boldsymbol{W}_1 \sim N_r(\boldsymbol{0}, \boldsymbol{I}_r)$ and $V_1 \sim \chi^2_{r-p}$. Now suppose we represent $\boldsymbol{b}_1^* = \boldsymbol{B}\boldsymbol{y}_1^*$, $\text{RSS}_1^* = \boldsymbol{y}_1^{*\prime}\boldsymbol{A}\boldsymbol{y}_1^*$ and $\boldsymbol{y}_1^* \sim N_r(\boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1, \boldsymbol{\Sigma})$, then $\boldsymbol{b}_1^*$ is independent of RSS_1^* since $\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{A} = \hat{\sigma}_1^2(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'(\boldsymbol{I}_r - \boldsymbol{P}_{\boldsymbol{X}_1}) = \boldsymbol{O}$. Thus the likelihood based on the released data for the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \psi)$ is given by

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2, \psi \,|\, \boldsymbol{b}_1^*, \text{RSS}_1^*, \boldsymbol{y}_2) = \phi_p(\boldsymbol{b}_1^*; \boldsymbol{\beta}, \sigma^2(1+\psi)(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1})\, h(\text{RSS}_1^*; r-p, \sigma^2\psi)$$
$$\phi_{n-r}(\boldsymbol{y}_2; \boldsymbol{X}_2\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_{n-r})$$

The prior distribution on the parameters is given by for $\delta > 0$

$$\pi(\boldsymbol{\beta}, \sigma^2, \psi) = \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\psi) \propto (\sigma^2)^{-\frac{\delta+1}{2}}\psi^{\frac{r-p}{2}-1}e^{\frac{-(r-p)\psi}{2}}$$

The posterior distribution can be computed in the following manner:

$$\pi(\boldsymbol{\beta}, \sigma^2, \psi \,|\, \boldsymbol{b}_1^*, \text{RSS}_1^*, \boldsymbol{y}_2) \propto \mathcal{L}(\boldsymbol{\beta}, \sigma^2, \psi \,|\, \boldsymbol{b}_1^*, \text{RSS}_1^*, \boldsymbol{y}_2)\, \pi(\psi)\, \pi(\boldsymbol{\beta}, \sigma^2)$$

$$\pi(\boldsymbol{\beta}, \sigma^2, \psi \,|\, \boldsymbol{b}_1^*, \text{RSS}_1^*, \boldsymbol{y}_2) = \pi(\boldsymbol{\beta} \,|\, \boldsymbol{b}_1^*, \text{RSS}_1^*, \boldsymbol{y}_2, \sigma^2, \psi)\, \pi(\sigma^2 \,|\, \boldsymbol{b}_1^*, \text{RSS}_1^*, \boldsymbol{y}_2, \psi)\, \pi(\psi \,|\, \boldsymbol{b}_1^*, \text{RSS}_1^*, \boldsymbol{y}_2)$$

so that we observe, as before, multiplying likelihood of the released data with our prior results in the product splitting up into three distinct posterior distributions.

The conditional posteriors are as follows

$$\boldsymbol{\beta} \,|\, \sigma^2, \psi, \boldsymbol{b}_1^*, \boldsymbol{b}_2 \sim N_p \left[\left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi} + \boldsymbol{X}_2'\boldsymbol{X}_2 \right)^{-1} \left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi} \boldsymbol{b}_1^* + \boldsymbol{X}_2'\boldsymbol{X}_2 \boldsymbol{b}_2 \right), \sigma^2 \left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi} + \boldsymbol{X}_2'\boldsymbol{X}_2 \right)^{-1} \right]$$

$$\sigma^2 \,|\, \psi, \boldsymbol{b}_1^*, \mathrm{RSS}_1^*, \boldsymbol{b}_2, \mathrm{RSS}_2 \sim \text{Scale-inv-}\chi^2 \left[n-p+\delta-1, \frac{1}{n-p+\delta-1} \left(\frac{\mathrm{RSS}_1^*}{\psi} + \mathrm{RSS}_2 + (\boldsymbol{b}_1^* - \boldsymbol{b}_2)' \left((1+\psi) \left(\boldsymbol{X}_1'\boldsymbol{X}_1 \right)^{-1} + \left(\boldsymbol{X}_2'\boldsymbol{X}_2 \right)^{-1} \right)^{-1} (\boldsymbol{b}_1^* - \boldsymbol{b}_2) \right) \right]$$

$$\pi(\psi \,|\, \boldsymbol{b}_1^*, \mathrm{RSS}_1^*, \boldsymbol{b}_2, \mathrm{RSS}_2) \propto \left| \frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi} + \boldsymbol{X}_2'\boldsymbol{X}_2 \right|^{-\frac{1}{2}} (1+\psi)^{-\frac{p}{2}} \psi^{-1} e^{-\frac{(r-p)\psi}{2}}$$
$$\times \left\{ (\boldsymbol{b}_1^* - \boldsymbol{b}_2)' \left((1+\psi) \left(\boldsymbol{X}_1'\boldsymbol{X}_1 \right)^{-1} + \left(\boldsymbol{X}_2'\boldsymbol{X}_2 \right)^{-1} \right)^{-1} (\boldsymbol{b}_1^* - \boldsymbol{b}_2) + \frac{\mathrm{RSS}_1^*}{\psi} + \mathrm{RSS}_2 \right\}^{-\frac{n-p+\delta-1}{2}}$$

We see that the expressions match the case when all of $\boldsymbol{y}$ is sensitive as in Section 2.1 by deleting all quantities involving $\boldsymbol{y}_2$, $\boldsymbol{X}_2$; replacing $\boldsymbol{X}_1$ by $\boldsymbol{X}$, $\boldsymbol{b}_1^*$ by $\boldsymbol{b}^*$ and r by n. The posterior distributions are proper as long as $r > p$, $n > \max\{r+p, p-\delta+1\}$.

Now as $\pi(\psi)$ (we use this shorthand from here on) is a non-standard pdf, we devise a sampling scheme below using the Accept-Reject method. Let us denote

$$Q(\psi) = (\boldsymbol{b}_1^* - \boldsymbol{b}_2)' \left((1+\psi) \left(\boldsymbol{X}_1'\boldsymbol{X}_1 \right)^{-1} + \left(\boldsymbol{X}_2'\boldsymbol{X}_2 \right)^{-1} \right)^{-1} (\boldsymbol{b}_1^* - \boldsymbol{b}_2) + \frac{\mathrm{RSS}_1^*}{\psi} + \mathrm{RSS}_2$$

$$\boldsymbol{X}'\boldsymbol{X}_\psi = \frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi} + \boldsymbol{X}_2'\boldsymbol{X}_2$$

We notice that, if we had started with only the sole assumption $r > p$, since $\boldsymbol{X}'\boldsymbol{X}_\psi > 0$ $\forall\, \psi > 0$ (as it is a covariance matrix), then letting $\psi \to \infty$ would yield $\boldsymbol{X}_2'\boldsymbol{X}_2 > 0$ and thus $n - r > p$, necessitating both of those assumptions in the first place. Now turning our attention to $Q(\psi)$, we see that $Q(\psi) > 0 \,\forall\, \psi > 0$ by definition and also by design. Since the r.v.'s $(\boldsymbol{b}_1^*, \boldsymbol{b}_2, \mathrm{RSS}_1^*, \mathrm{RSS}_2)$ embroiled in the expression of $Q(\psi)$ are mutually independent, $Q(\psi) > 0$ even when RSS_1^* is arbitrarily small, hence $Q(\psi) \geq \frac{\mathrm{RSS}_1^*}{\psi}$.

This coupled with the fact that $\boldsymbol{X}'\boldsymbol{X}_\psi > \dfrac{\boldsymbol{X}'\boldsymbol{X}}{1+\psi} \implies |\boldsymbol{X}'\boldsymbol{X}_\psi| > (1+\psi)^{-\frac{p}{2}} |\boldsymbol{X}'\boldsymbol{X}|$ (as $\boldsymbol{A} > \boldsymbol{B} \implies \lambda_i(\boldsymbol{A}) > \lambda_i(\boldsymbol{B}) \; \forall\, i = 1, \ldots, n$ where $\{\lambda_i(\boldsymbol{A}) \; : \; i = 1, \ldots, n\}$ and $\{\lambda_i(\boldsymbol{B}) \; : \; i = 1, \ldots, n\}$ are the ordered eigenvalues of $n \times n$ PD matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively) produces $\pi(\psi) \leq Lg(\psi)$ where

$$L = \frac{|\boldsymbol{X}'\boldsymbol{X}|^{-\frac{1}{2}} \, 2^{n-p+\delta-1} \Gamma\left(\frac{n-p+\delta-1}{2}\right)}{((r-p)\mathrm{RSS}_1^*)^{\frac{n-p+\delta-1}{2}}}$$

and $g(\psi)$ is the pdf of a $\frac{\chi^2_{n-p+\delta-1}}{r-p} \equiv \Gamma\left(\frac{n-p+\delta-1}{2}, \frac{r-p}{2}\right)$ r.v.

Algorithm for sampling from $\pi(\psi)$:

1. We have the i-th sample $\psi^{(i)}$.
2. Draw a sample $\psi' \sim g(\psi)$ where $g(\psi) \sim \frac{\chi^2_{n-p+\delta-1}}{r-p}$ and also draw $u \sim U[0,1]$.
3. If $u \leq \dfrac{\pi(\psi)}{Lg(\psi)}$ then $\psi^{(i+1)} = \psi'$, else discard ψ' and go back to Step 2.

Theorem 2.4.1. The joint pdf of $(\boldsymbol{b}_1^*, \mathrm{RSS}_1^*, \boldsymbol{b}_2, \mathrm{RSS}_2)$ is given by

$$\begin{aligned}
& f_{\boldsymbol{\beta},\sigma^2}(\boldsymbol{b}_1^*, \mathrm{RSS}_1^*, \boldsymbol{b}_2, \mathrm{RSS}_2) \\
\propto & \int_0^\infty \phi_p \left[\boldsymbol{\beta}; \left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi} + \boldsymbol{X}_2'\boldsymbol{X}_2 \right)^{-1} \left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi} \boldsymbol{b}_1^* + \boldsymbol{X}_2'\boldsymbol{X}_2 \boldsymbol{b}_2 \right), \sigma^2 \left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi} + \boldsymbol{X}_2'\boldsymbol{X}_2 \right)^{-1} \right] \\
& \times \frac{(\mathrm{RSS}_1^*)^{\frac{r-p}{2}-1} (\mathrm{RSS}_2)^{\frac{n-r-p}{2}-1}}{(\sigma^2)^{\frac{n-p}{2}}} \, e^{-\frac{1}{2\sigma^2}\left[\left(\boldsymbol{b}_1^* - \boldsymbol{b}_2\right)' \left((1+\psi)\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} + \left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1} \right)^{-1} \left(\boldsymbol{b}_1^* - \boldsymbol{b}_2\right) + \frac{\mathrm{RSS}_1^*}{\psi} + \mathrm{RSS}_2 \right]} \\
& \times \left| \frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi} + \boldsymbol{X}_2'\boldsymbol{X}_2 \right|^{-\frac{1}{2}} (1+\psi)^{-\frac{p}{2}} \psi^{-1} e^{-\frac{(r-p)\psi}{2}} \, d\psi
\end{aligned}$$

Posterior Predictive Sampling

We similarly synthesize r sensitive y-values $\boldsymbol{y}_1$ by applying the posterior predictive sampling method based on these r y-values, as discussed in Section 2.2. The synthetic version of $\boldsymbol{y}_1$ is $\boldsymbol{y}_1^* = (y_1^*, \ldots, y_r^*)'$ such that $y_i^* \sim N(\boldsymbol{x}_i'\boldsymbol{\beta}_1^*, \sigma_1^{*2})$ generated indepen-

dently for $i = 1, \ldots, r$, where following from equations (2.11) and (2.12), $(\boldsymbol{\beta}_1^*, \sigma_1^{*2})$ are drawn from

$$\begin{aligned} \sigma_1^2 \,|\, \boldsymbol{b}_1, \mathrm{RSS}_1 &\sim \text{Scale-inv-}\chi^2 \left(r - p + \alpha - 1, \frac{\mathrm{RSS}_1}{r - p + \alpha - 1} \right) \\ \boldsymbol{\beta_1} \,|\, \boldsymbol{b}_1, \mathrm{RSS}_1, \sigma_1^2 &\sim N_p \left(\boldsymbol{b}_1, \sigma_1^2 (\boldsymbol{X_1}' \boldsymbol{X_1})^{-1} \right) \end{aligned}$$

where we assume throughout that $r + \alpha > p + 1$.

Then the sufficient statistics for the imputed data are

$$\begin{aligned} \boldsymbol{b}_1^* &= (\boldsymbol{X}_1' \boldsymbol{X}_1)^{-1} \boldsymbol{X}_1' \boldsymbol{y}_1^* \overset{d}{=} \boldsymbol{\beta}_1^* + \sigma_1^* \boldsymbol{C}_1 \boldsymbol{U}_0 \overset{d}{=} \boldsymbol{\beta} + \sigma \sqrt{1 + 2\psi} \boldsymbol{C}_1 \boldsymbol{U}_1 \\ \mathrm{RSS}_1^* &= \boldsymbol{y}_1^{*'} (\boldsymbol{I}_r - \boldsymbol{P}_{\boldsymbol{X}_1}) \boldsymbol{y}_1^* \overset{d}{=} \sigma_1^{*2} \boldsymbol{W}_1' (\boldsymbol{I}_r - \boldsymbol{P}_{\boldsymbol{X}_1}) \boldsymbol{W}_1 \overset{d}{=} \sigma^2 \psi V_1 \end{aligned}$$

where $\boldsymbol{U}_0, \boldsymbol{U}_1 \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$ independently, $\boldsymbol{C}_1 \boldsymbol{C}_1' = (\boldsymbol{X}_1' \boldsymbol{X}_1)^{-1}$, $\psi = (\sigma_1^*/\sigma)^2$ is a latent quantity, $\boldsymbol{W}_1 \sim N_r(\boldsymbol{0}, \boldsymbol{I}_r)$ and $V_1 \sim \chi^2_{r-p}$. Next we can basically adapt the same procedure as before to obtain the conditional posteriors as follows

$$\boldsymbol{\beta} \,|\, \sigma^2, \psi, \boldsymbol{b}_1^*, \boldsymbol{b}_2 \sim N_p \left[\left(\frac{\boldsymbol{X}_1' \boldsymbol{X}_1}{1 + 2\psi} + \boldsymbol{X}_2' \boldsymbol{X}_2 \right)^{-1} \left(\frac{\boldsymbol{X}_1' \boldsymbol{X}_1}{1 + 2\psi} \boldsymbol{b}_1^* + \boldsymbol{X}_2' \boldsymbol{X}_2 \boldsymbol{b}_2 \right), \sigma^2 \left(\frac{\boldsymbol{X}_1' \boldsymbol{X}_1}{1 + 2\psi} + \boldsymbol{X}_2' \boldsymbol{X}_2 \right)^{-1} \right]$$

$$\begin{aligned} \sigma^2 \,|\, \psi, \boldsymbol{b}_1^*, \mathrm{RSS}_1^*, \boldsymbol{b}_2, \mathrm{RSS}_2 \sim \text{Scale-inv-}\chi^2 \Bigg[n - p + \delta - 1, \frac{1}{n - p + \delta - 1} \Bigg(\frac{\mathrm{RSS}_1^*}{\psi} + \mathrm{RSS}_2 + \\ (\boldsymbol{b}_1^* - \boldsymbol{b}_2)' \left((1 + 2\psi) (\boldsymbol{X}_1' \boldsymbol{X}_1)^{-1} + (\boldsymbol{X}_2' \boldsymbol{X}_2)^{-1} \right)^{-1} (\boldsymbol{b}_1^* - \boldsymbol{b}_2) \Bigg) \Bigg] \end{aligned}$$

$$\begin{aligned} \pi(\psi \,|\, \boldsymbol{b}_1^*, \mathrm{RSS}_1^*, \boldsymbol{b}_2, \mathrm{RSS}_2) \propto \left| \frac{\boldsymbol{X}_1' \boldsymbol{X}_1}{1 + 2\psi} + \boldsymbol{X}_2' \boldsymbol{X}_2 \right|^{-\frac{1}{2}} (1 + 2\psi)^{-\frac{p}{2}} (1 + \psi)^{-\frac{2r - 2p + \alpha - 1}{2}} \psi^{-1} \\ \times \left\{ (\boldsymbol{b}_1^* - \boldsymbol{b}_2)' \left((1 + 2\psi) (\boldsymbol{X}_1' \boldsymbol{X}_1)^{-1} + (\boldsymbol{X}_2' \boldsymbol{X}_2)^{-1} \right)^{-1} (\boldsymbol{b}_1^* - \boldsymbol{b}_2) + \frac{\mathrm{RSS}_1^*}{\psi} + \mathrm{RSS}_2 \right\}^{-\frac{n - p + \delta - 1}{2}} \end{aligned}$$

The posterior distributions are proper as long as $r > \max\left\{p, p - \alpha + 1, \frac{n + p - \alpha + \delta}{2}\right\}$, $n > \max\{r + p, p - \delta + 1\}$ and they match our results in Section 2.2 when $r = n$.

Algorithm for sampling from $\pi(\psi)$:

1. We have the i-th sample $\psi^{(i)}$.

2. Draw a sample $\psi' \sim g(\psi)$ where $g(\psi) \sim \beta'\left(\frac{n-p+\delta-1}{2}, \frac{2r-n-p+\alpha-\delta}{2}\right)$ and also draw $u \sim U[0,1]$. This necessitates the assumption $2r+\alpha > n+p+\delta$.

3. If $u \leq \dfrac{\pi(\psi)}{Lg(\psi)}$ then $\psi^{(i+1)} = \psi'$, else discard ψ' and go back to Step 2. Here

$$L = \frac{|\boldsymbol{X}'\boldsymbol{X}|^{-\frac{1}{2}}\,\mathrm{B}\left(\frac{n-p+\delta-1}{2}, \frac{2r-n-p+\alpha-\delta}{2}\right)}{(\mathrm{RSS}_1^*)^{\frac{n-p+\delta-1}{2}}}$$

where $\mathrm{B}(a,b)$ is the Beta function.

Theorem 2.4.2. The joint pdf of $(\boldsymbol{b}_1^*, \mathrm{RSS}_1^*, \boldsymbol{b}_2, \mathrm{RSS}_2)$ is given by

$$\begin{aligned}
&f_{\boldsymbol{\beta},\sigma^2}(\boldsymbol{b}_1^*, \mathrm{RSS}_1^*, \boldsymbol{b}_2, \mathrm{RSS}_2) \\
\propto &\int_0^\infty \phi_p\left[\boldsymbol{\beta}; \left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+2\psi} + \boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+2\psi}\boldsymbol{b}_1^* + \boldsymbol{X}_2'\boldsymbol{X}_2\boldsymbol{b}_2\right), \sigma^2\left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+2\psi} + \boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\right] \\
&\times \frac{(\mathrm{RSS}_1^*)^{\frac{r-p}{2}-1}(\mathrm{RSS}_2)^{\frac{n-r-p}{2}-1}}{(\sigma^2)^{\frac{n-p}{2}}}\, e^{-\frac{1}{2\sigma^2}\left[\left(\boldsymbol{b}_1^*-\boldsymbol{b}_2\right)'\left((1+2\psi)\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}+\left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}\right)^{-1}\left(\boldsymbol{b}_1^*-\boldsymbol{b}_2\right)+\frac{\mathrm{RSS}_1^*}{\psi}+\mathrm{RSS}_2\right]} \\
&\times \left|\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+2\psi} + \boldsymbol{X}_2'\boldsymbol{X}_2\right|^{-\frac{1}{2}} (1+2\psi)^{-\frac{p}{2}}(1+\psi)^{-\frac{2r-2p+\alpha-1}{2}}\psi^{-1}\,d\psi
\end{aligned}$$

Method II: Using whole data estimates to impute synthetic data

Plug-In Sampling

We can relax the assumption $n-r > p$ needed before if we use estimates of the entire data to impute the r sensitive y-values $\boldsymbol{y}_1$. In that case, the synthetic version of $\boldsymbol{y}_1$ is $\boldsymbol{y}_1^* = (y_1^*, \ldots, y_r^*)'$ such that $y_i^* \sim N(\boldsymbol{x}_i'\boldsymbol{b}, \mathrm{RSS})$ generated independently for $i = 1, \ldots, r$ and $\boldsymbol{y}_2$ is defined as before. The likelihood of the released data is proportional

to what follows below, since we only retain quantities containing parameters $(\boldsymbol{\beta}, \sigma^2)$ necessary for posterior distribution calculation, also using the fact that $\boldsymbol{y}_2 \,|\, \boldsymbol{b}, \text{RSS}$ is independent of $(\boldsymbol{\beta}, \sigma^2)$ by the definition of sufficient statistic

$$
\begin{aligned}
&\pi(\boldsymbol{y}_1^*, \boldsymbol{y}_2 \,|\, \boldsymbol{\beta}, \sigma^2) \\
&= \int \pi(\boldsymbol{y}_1^*, \boldsymbol{y}_2 \,|\, \boldsymbol{b}, \text{RSS})\, \pi(\boldsymbol{b}, \text{RSS} \,|\, \boldsymbol{\beta}, \sigma^2)\, d\boldsymbol{b}\, d\text{RSS} \\
&= \int \pi(\boldsymbol{y}_1^* \,|\, \boldsymbol{y}_2, \boldsymbol{b}, \text{RSS})\, \pi(\boldsymbol{y}_2 \,|\, \boldsymbol{b}, \text{RSS})\, \pi(\boldsymbol{b}, \text{RSS} \,|\, \boldsymbol{\beta}, \sigma^2)\, d\boldsymbol{b}\, d\text{RSS} \\
&\propto \int \pi(\boldsymbol{y}_1^* \,|\, \boldsymbol{b}, \text{RSS})\, \pi(\boldsymbol{b} \,|\, \boldsymbol{\beta}, \sigma^2)\, \pi(\text{RSS} \,|\, \sigma^2)\, d\boldsymbol{b}\, d\text{RSS} \\
&\propto \int \frac{1}{(\sigma^2\psi)^{\frac{r}{2}}} \exp\left[-\frac{1}{2\sigma^2\psi}(\boldsymbol{y}_1^* - \boldsymbol{X}_1\boldsymbol{b})'(\boldsymbol{y}_1^* - \boldsymbol{X}_1\boldsymbol{b})\right] \\
&\quad \times \frac{1}{(\sigma^2)^{\frac{p}{2}}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{b} - \boldsymbol{\beta})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{b} - \boldsymbol{\beta})\right] \times \psi^{\frac{n-p}{2}-1} e^{-\frac{(n-p)\psi}{2}}\, d\boldsymbol{b}\, d\psi
\end{aligned} \tag{2.20}
$$

The last line is due to a change in variable $\text{RSS}/(n-p)\sigma^2 = \psi$. Next we collect terms for $\boldsymbol{b}$ as

$$
\begin{aligned}
&\frac{1}{\psi}(\boldsymbol{y}_1^* - \boldsymbol{X}_1\boldsymbol{b})'(\boldsymbol{y}_1^* - \boldsymbol{X}_1\boldsymbol{b}) + (\boldsymbol{b} - \boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{b} - \boldsymbol{\beta}) \\
&= \boldsymbol{b}'\left(\boldsymbol{X}'\boldsymbol{X} + \frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right)\boldsymbol{b} - 2\boldsymbol{b}\left(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \frac{\boldsymbol{X}_1'\boldsymbol{y}_1^*}{\psi}\right) + \frac{\boldsymbol{y}_1^{*\prime}\boldsymbol{y}_1^*}{\psi} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\
&= \left(\boldsymbol{b} - \left(\boldsymbol{X}'\boldsymbol{X} + \frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \frac{\boldsymbol{X}_1'\boldsymbol{y}_1^*}{\psi}\right)\right)'\left(\boldsymbol{X}'\boldsymbol{X} + \frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right) \\
&\quad \left(\boldsymbol{b} - \left(\boldsymbol{X}'\boldsymbol{X} + \frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \frac{\boldsymbol{X}_1'\boldsymbol{y}_1^*}{\psi}\right)\right) + \frac{\boldsymbol{y}_1^{*\prime}\boldsymbol{y}_1^*}{\psi} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\
&\quad - \left(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \frac{\boldsymbol{X}_1'\boldsymbol{y}_1^*}{\psi}\right)'\left(\boldsymbol{X}'\boldsymbol{X} + \frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \frac{\boldsymbol{X}_1'\boldsymbol{y}_1^*}{\psi}\right)
\end{aligned} \tag{2.21}
$$

where we know $\boldsymbol{X}'\boldsymbol{X} + \dfrac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}$ is invertible because $\boldsymbol{X}'\boldsymbol{X} + \dfrac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi} > 0$ due to $\boldsymbol{X}'\boldsymbol{X} > 0$, $\boldsymbol{X}_1'\boldsymbol{X}_1 > 0$, $\psi > 0$. The last three quantities above simplify to

$$\boldsymbol{\beta}'\left[\boldsymbol{X}'\boldsymbol{X}-\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}+\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\right]-2\boldsymbol{\beta}'\left[\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}+\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right)^{-1}\frac{\boldsymbol{X}_1'\boldsymbol{y}_1^*}{\psi}\right]$$
$$+\boldsymbol{y}_1^{*\prime}\left[\frac{\boldsymbol{I}_r}{\psi}-\frac{\boldsymbol{X}_1}{\psi}\left(\boldsymbol{X}'\boldsymbol{X}+\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right)^{-1}\frac{\boldsymbol{X}_1'}{\psi}\right]\boldsymbol{y}_1^* \tag{2.22}$$

from which it is clear that the (conditional) posterior variance of $\boldsymbol{\beta}$ is

$$\begin{aligned}
&\boldsymbol{X}'\boldsymbol{X}-\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}+\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}>0\\
\iff&(\boldsymbol{X}'\boldsymbol{X})^{1/2}\left[\boldsymbol{I}_p-(\boldsymbol{X}'\boldsymbol{X})^{1/2}\left(\boldsymbol{X}'\boldsymbol{X}+\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right)^{-1}(\boldsymbol{X}'\boldsymbol{X})^{1/2}\right](\boldsymbol{X}'\boldsymbol{X})^{1/2}>0\\
\iff&\boldsymbol{I}_p>(\boldsymbol{X}'\boldsymbol{X})^{1/2}\left(\boldsymbol{X}'\boldsymbol{X}+\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right)^{-1}(\boldsymbol{X}'\boldsymbol{X})^{1/2}\\
\iff&(\boldsymbol{X}'\boldsymbol{X})^{-1}>\left(\boldsymbol{X}'\boldsymbol{X}+\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right)^{-1}\iff\boldsymbol{X}'\boldsymbol{X}+\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}>\boldsymbol{X}'\boldsymbol{X}\\
\iff&\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}>0\quad\forall\,\psi>0\iff\boldsymbol{X}_1'\boldsymbol{X}_1>0\iff r>p
\end{aligned}$$

so that we still have to respect the condition $r > p$ while employing this method. Thus (2.22) further simplifies to

$$\begin{aligned}
&\boldsymbol{\beta}'\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}+\psi(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right]^{-1}\boldsymbol{\beta}-2\boldsymbol{\beta}'\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}+\psi(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right]^{-1}\boldsymbol{b}_1^*\\
&+\boldsymbol{b}_1^{*\prime}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}+\psi(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right]^{-1}\boldsymbol{b}_1^*+\frac{\boldsymbol{y}_1^{*\prime}\boldsymbol{y}_1^*}{\psi}-\boldsymbol{b}_1^{*\prime}\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\boldsymbol{b}_1^*\\
&=(\boldsymbol{\beta}-\boldsymbol{b}_1^*)'\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}+\psi(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right]^{-1}(\boldsymbol{\beta}-\boldsymbol{b}_1^*)+\frac{\boldsymbol{y}_1^{*\prime}\boldsymbol{y}_1^*}{\psi}-\boldsymbol{y}_1^{*\prime}\frac{\boldsymbol{P}_{\boldsymbol{X}_1}\boldsymbol{P}_{\boldsymbol{X}_1}}{\psi}\boldsymbol{y}_1^*\\
&=(\boldsymbol{\beta}-\boldsymbol{b}_1^*)'\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}+\psi(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right]^{-1}(\boldsymbol{\beta}-\boldsymbol{b}_1^*)+\boldsymbol{y}_1^{*\prime}\frac{(\boldsymbol{I}_r-\boldsymbol{P}_{\boldsymbol{X}_1})}{\psi}\boldsymbol{y}_1^*\\
&=(\boldsymbol{\beta}-\boldsymbol{b}_1^*)'\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}+\psi(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right]^{-1}(\boldsymbol{\beta}-\boldsymbol{b}_1^*)+\frac{\mathrm{RSS}_1^*}{\psi}\boldsymbol{y}_1^*
\end{aligned}\tag{2.23}$$

where $\boldsymbol{b}_1^*=(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y}_1^*$, $\mathrm{RSS}_1^*=\boldsymbol{y}_1'(\boldsymbol{I}_r-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y}_1$ are sufficient statistics for $\boldsymbol{y}_1^*$.

So integrating out $\boldsymbol{b}$ from (2.20) using (2.21) and (2.23) and multiplying by our usual prior $\pi(\boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{-\frac{\delta+1}{2}}$ we get the joint posterior distribution to be

$$\begin{aligned}
&\pi(\boldsymbol{\beta}, \sigma^2, \psi \,|\, \boldsymbol{y}_1^*, \boldsymbol{y}_2) \\
&\propto \phi_p\left(\boldsymbol{\beta}; \boldsymbol{b}_1^*, \sigma^2\left[(\boldsymbol{X}'\boldsymbol{X})^{-1} + \psi(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right]\right) \\
&\times \frac{(\mathrm{RSS}_1^*/\psi)^{\frac{r-p+\delta-1}{2}}}{(\sigma^2)^{\frac{r-p+\delta-1}{2}+1}} \exp\left[-\frac{\mathrm{RSS}_1^*}{2\sigma^2\psi}\right] \\
&\times \left|(\boldsymbol{X}'\boldsymbol{X})^{-1} + \psi(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right|^{\frac{1}{2}} \left|\boldsymbol{X}'\boldsymbol{X} + \frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right|^{-\frac{1}{2}} \psi^{\frac{n-r-p}{2}-1} e^{-\frac{(n-p)\psi}{2}} \psi^{\frac{r-p+\delta-1}{2}}
\end{aligned}$$

where after observing

$$\left|(\boldsymbol{X}'\boldsymbol{X})^{-1} + \psi(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right|^{\frac{1}{2}} \left|\boldsymbol{X}'\boldsymbol{X} + \frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\psi}\right|^{-\frac{1}{2}} = \psi^{\frac{p}{2}}$$

it leads us to the hierarchical (conditional) posterior distributions as follows

$$\begin{aligned}
\boldsymbol{\beta} \,|\, \sigma^2, \psi, \boldsymbol{b}_1^* &\sim N_p\left(\boldsymbol{b}_1^*, \sigma^2\left((\boldsymbol{X}'\boldsymbol{X})^{-1} + \psi(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right)\right) && (2.24) \\
\sigma^2 \,|\, \psi, \mathrm{RSS}_1^* &\sim \text{Scale-inv-}\chi^2\left(r-p+\delta-1, \frac{\mathrm{RSS}_1^*}{\psi(r-p+\delta-1)}\right) && (2.25) \\
\psi &\sim \frac{\chi^2_{n-p+\delta-1}}{n-p} \equiv \Gamma\left(\frac{n-p+\delta-1}{2}, \frac{n-p}{2}\right) && (2.26)
\end{aligned}$$

The posterior distributions are proper as long as $r > \max\{p, p-\delta+1\}$ and the results match our expressions from Section 2.1 when $r = n$. An advantage of this method is that the sampling of ψ is straightforward.

Posterior Predictive Sampling

We synthesize $\boldsymbol{y}_1^* = (y_1^*, \ldots, y_r^*)'$ such that $y_i^* \sim N(\boldsymbol{x}_i'\boldsymbol{\beta}^*, \sigma^{*2})$ generated independently for $i = 1, \ldots, r$, where $(\boldsymbol{\beta}^*, \sigma^{*2})$ are drawn from (2.11) and (2.12). The likelihood of

the released data is given by

$$
\begin{aligned}
&\pi(\boldsymbol{y}_1^*, \boldsymbol{y}_2 \,|\, \boldsymbol{\beta}, \sigma^2) \\
&= \int \pi(\boldsymbol{y}_1^*, \boldsymbol{y}_2 \,|\, \boldsymbol{\beta}^*, \sigma^{*2}, \boldsymbol{b}, \mathrm{RSS})\, \pi(\boldsymbol{\beta}^*, \sigma^{*2} \,|\, \boldsymbol{b}, \mathrm{RSS}))\, \pi(\boldsymbol{b}, \mathrm{RSS} \,|\, \boldsymbol{\beta}, \sigma^2)\, d\boldsymbol{\beta}^*\, d\sigma^{*2}\, d\boldsymbol{b}\, d\mathrm{RSS} \\
&\propto \int \pi(\boldsymbol{y}_1^* \,|\, \boldsymbol{\beta}^*, \sigma^{*2})\, \pi(\boldsymbol{\beta}^* \,|\, \sigma^{*2}, \boldsymbol{b})\, \pi(\sigma^{*2} \,|\, \mathrm{RSS})\, \pi(\boldsymbol{b} \,|\, \boldsymbol{\beta}, \sigma^2)\, \pi(\mathrm{RSS} \,|\, \sigma^2)\, d\boldsymbol{\beta}^*\, d\sigma^{*2}\, d\boldsymbol{b}\, d\mathrm{RSS} \\
&\propto \int \frac{1}{(\sigma^{*2})^{\frac{r}{2}}} \exp\left[-\frac{1}{2\sigma^{*2}}(\boldsymbol{y}_1^* - \boldsymbol{X}_1\boldsymbol{\beta}^*)'(\boldsymbol{y}_1^* - \boldsymbol{X}_1\boldsymbol{\beta}^*)\right] \\
&\quad \times \frac{1}{(\sigma^{*2})^{\frac{p}{2}}} \exp\left[-\frac{1}{2\sigma^{*2}}(\boldsymbol{\beta}^* - \boldsymbol{b})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{\beta}^* - \boldsymbol{b})\right] \times \frac{\exp\left[-\frac{\mathrm{RSS}}{2\sigma^{*2}}\right](\mathrm{RSS})^{\frac{n-p+\alpha-1}{2}}}{(\sigma^{*2})^{\frac{n-p+\alpha+1}{2}}} \\
&\quad \times \frac{1}{(\sigma^{2})^{\frac{p}{2}}} \exp\left[-\frac{1}{2\sigma^{2}}(\boldsymbol{b} - \boldsymbol{\beta})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{b} - \boldsymbol{\beta})\right] \times \frac{\exp\left[-\frac{\mathrm{RSS}}{2\sigma^{2}}\right](\mathrm{RSS})^{\frac{n-p}{2}-1}}{(\sigma^{2})^{\frac{n-p}{2}}}\, d\boldsymbol{\beta}^*\, d\sigma^{*2}\, d\boldsymbol{b}\, d\mathrm{RSS}
\end{aligned}
$$

We begin by collecting terms for $\boldsymbol{\beta}^*$ as

$$
\begin{aligned}
&(\boldsymbol{y}_1^* - \boldsymbol{X}_1\boldsymbol{\beta}^*)'(\boldsymbol{y}_1^* - \boldsymbol{X}_1\boldsymbol{\beta}^*) + (\boldsymbol{\beta}^* - \boldsymbol{b})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{\beta}^* - \boldsymbol{b}) \\
&= \boldsymbol{\beta}^{*\prime}\left(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_1'\boldsymbol{X}_1\right)\boldsymbol{\beta}^* - 2\boldsymbol{\beta}^*\left(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} + \boldsymbol{X}_1'\boldsymbol{y}_1^*\right) + \boldsymbol{y}_1^{*\prime}\boldsymbol{y}_1^* + \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} \\
&= \left(\boldsymbol{\beta}^* - \left(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} + \boldsymbol{X}_1'\boldsymbol{y}^*\right)\right)'\left(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_1'\boldsymbol{X}_1\right) \\
&\quad \left(\boldsymbol{\beta}^* - \left(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} + \boldsymbol{X}_1'\boldsymbol{y}_1^*\right)\right) + \boldsymbol{y}_1^{*\prime}\boldsymbol{y}_1^* + \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} \\
&- \left(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} + \boldsymbol{X}_1'\boldsymbol{y}_1^*\right)'\left(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} + \boldsymbol{X}_1'\boldsymbol{y}_1^*\right)
\end{aligned}
$$

After integrating out $\boldsymbol{\beta}^*$ the likelihood stands at

$$
\begin{aligned}
&\int \frac{1}{(\sigma^{*2})^{\frac{r}{2}}} \exp\left[-\frac{1}{2\sigma^{*2}}\left(\boldsymbol{y}_1^{*\prime}\boldsymbol{y}_1^* + \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} - \left(\boldsymbol{X}_1{}'\boldsymbol{y}_1^* + \boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}\right)'\left(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\left(\boldsymbol{X}_1{}'\boldsymbol{y}_1^* + \boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}\right)\right)\right] \\
&\times \frac{\exp\left[-\frac{\mathrm{RSS}}{2\sigma^{*2}}\right]}{(\sigma^{*2})^{\frac{n-p+\alpha+1}{2}}} \times \frac{\exp\left[-\frac{\mathrm{RSS}}{2\sigma^{2}}\right]}{(\sigma^{2})^{\frac{n-p}{2}}} \times (\mathrm{RSS})^{\frac{2n-2p+\alpha-1}{2}-1} \\
&\times \frac{1}{(\sigma^{2})^{\frac{p}{2}}} \exp\left[-\frac{1}{2\sigma^{2}}(\boldsymbol{b} - \boldsymbol{\beta})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{b} - \boldsymbol{\beta})\right] d\sigma^{*2}\, d\boldsymbol{b}\, d\mathrm{RSS}
\end{aligned}
$$

Next we collect terms for $\boldsymbol{b}$ as follows

$$
\frac{1}{\sigma^{*2}}\left(\boldsymbol{y}_1^{*\prime}\boldsymbol{y}_1^* + \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} - \left(\boldsymbol{X}_1{}'\boldsymbol{y}_1^* + \boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}\right)'\left(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\left(\boldsymbol{X}_1{}'\boldsymbol{y}_1^* + \boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}\right)\right)
$$

$$
\begin{aligned}
&+\frac{1}{\sigma^2}(\boldsymbol{b}-\boldsymbol{\beta})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{b}-\boldsymbol{\beta}) \\
&=\boldsymbol{b}'\left[(\boldsymbol{X}'\boldsymbol{X})\left(\frac{1}{\sigma^2}+\frac{1}{\sigma^{*2}}\right)-\frac{\boldsymbol{X}'\boldsymbol{X}}{\sigma^{*2}}\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\right]\boldsymbol{b} \\
&-2\boldsymbol{b}'\left[\frac{\boldsymbol{X}'\boldsymbol{X}}{\sigma^{*2}}\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1{}'\boldsymbol{y}_1^*+\frac{\boldsymbol{X}'\boldsymbol{X}}{\sigma^2}\boldsymbol{\beta}\right]+\frac{\boldsymbol{y}_1^{*\prime}\boldsymbol{y}_1^*}{\sigma^{*2}} \\
&-\frac{\boldsymbol{y}_1^{*\prime}}{\sigma^{*2}}\boldsymbol{X}_1\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1{}'\boldsymbol{y}_1^*+\boldsymbol{\beta}'\frac{\boldsymbol{X}'\boldsymbol{X}}{\sigma^2}\boldsymbol{\beta} \qquad (2.27)
\end{aligned}
$$

We can figure out what the variance-covariance matrix will be when we would integrate out $\boldsymbol{b}$, and thus by definition after a change of variable $\sigma^{*2}/\sigma^2=\psi$ we have

$$
\begin{aligned}
&(1+\psi)\,\boldsymbol{X}'\boldsymbol{X}-\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}'\boldsymbol{X}>0 \\
\iff &(1+\psi)\,\boldsymbol{I}_p>(\boldsymbol{X}'\boldsymbol{X})^{1/2}\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}(\boldsymbol{X}'\boldsymbol{X})^{1/2} \\
\iff &(1+\psi)\,(\boldsymbol{X}'\boldsymbol{X})^{-1}>\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} \\
\iff &\frac{\boldsymbol{X}'\boldsymbol{X}}{1+\psi}<\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1 \iff \boldsymbol{X}_1'\boldsymbol{X}_1+\frac{\psi}{1+\psi}\boldsymbol{X}'\boldsymbol{X}>0
\end{aligned}
$$

which is true for all values of $\psi>0$. We let $\psi\to 0$ to get $\boldsymbol{X}_1'\boldsymbol{X}_1>0$, so $r>p$ and

$$
(1+\psi)\,\boldsymbol{X}'\boldsymbol{X}-\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}'\boldsymbol{X}=\psi\boldsymbol{X}'\boldsymbol{X}+\left((\boldsymbol{X}'\boldsymbol{X})^{-1}+(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right)^{-1}
$$

Here we use the following fact: for any two PD matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,

$$
\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}\boldsymbol{A}^{-1}=\boldsymbol{A}^{-1}(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}\boldsymbol{B}^{-1}=(\boldsymbol{A}+\boldsymbol{B})^{-1} \qquad (2.28)
$$

Next we follow up from (2.27) to get, after taking out the common factor $\dfrac{1}{\sigma^2\psi}$ from all quantities involved

$$
\begin{aligned}
&\left(\boldsymbol{b}-\left[\psi\boldsymbol{X}'\boldsymbol{X}+\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}\right]^{-1}\left[\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1{}'\boldsymbol{y}_1^*+\psi\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}\right]\right)'\\
&\left[\psi\boldsymbol{X}'\boldsymbol{X}+\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}\right]\\
&\left(\boldsymbol{b}-\left[\psi\boldsymbol{X}'\boldsymbol{X}+\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}\right]^{-1}\left[\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1{}'\boldsymbol{y}_1^*+\psi\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}\right]\right)\\
&+\boldsymbol{y}_1^{*\prime}\boldsymbol{y}_1^*-\boldsymbol{y}_1^{*\prime}\boldsymbol{X}_1\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1{}'\boldsymbol{y}_1^*+\psi\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}-\\
&\left[\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1{}'\boldsymbol{y}_1^*+\psi\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}\right]'\left[\psi\boldsymbol{X}'\boldsymbol{X}+\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}\right]^{-1}\\
&\left[\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1{}'\boldsymbol{y}_1^*+\psi\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}\right]
\end{aligned}
$$

The last three lines give us by repeated application of (2.28)

$$
\begin{aligned}
&\boldsymbol{\beta}'\left(\psi\boldsymbol{X}'\boldsymbol{X}-\psi\boldsymbol{X}'\boldsymbol{X}\left[\psi\boldsymbol{X}'\boldsymbol{X}+\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}\right]^{-1}\psi\boldsymbol{X}'\boldsymbol{X}\right)\boldsymbol{\beta}\\
&-2\boldsymbol{\beta}'\left(\psi\boldsymbol{X}'\boldsymbol{X}\left[\psi\boldsymbol{X}'\boldsymbol{X}+\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}\right]^{-1}\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1{}'\boldsymbol{X}_1\right)\boldsymbol{b}_1^*\\
&+\boldsymbol{y}_1^{*\prime}(\boldsymbol{I}_r-\boldsymbol{P}_{\boldsymbol{X}_1})\boldsymbol{y}_1^*+\boldsymbol{b}_1^{*\prime}\left(\boldsymbol{X}_1'\boldsymbol{X}_1-\boldsymbol{X}_1'\boldsymbol{X}_1\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1{}'\boldsymbol{X}_1\right)\boldsymbol{b}_1^{*\prime}\\
&-\boldsymbol{b}_1^{*\prime}\boldsymbol{X}_1'\boldsymbol{X}_1\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\left[\psi\boldsymbol{X}'\boldsymbol{X}+\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}\right]^{-1}\\
&\quad\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1{}'\boldsymbol{X}_1\boldsymbol{b}_1^*\\
&=\boldsymbol{\beta}'\left(\frac{1}{\psi}(\boldsymbol{X}'\boldsymbol{X})^{-1}+\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}\boldsymbol{\beta}\\
&-2\boldsymbol{\beta}'\left(\frac{1}{\psi}(\boldsymbol{X}'\boldsymbol{X})^{-1}+\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}\boldsymbol{b}_1^*\\
&+\mathrm{RSS}_1^*+\boldsymbol{b}_1^{*\prime}\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}\boldsymbol{b}_1^*\\
&-\boldsymbol{b}_1^{*\prime}\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}\left[\psi\boldsymbol{X}'\boldsymbol{X}+\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}\right]^{-1}\\
&\quad\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}\boldsymbol{b}_1^*\\
&=(\boldsymbol{\beta}-\boldsymbol{b}_1^*)'\left(\frac{1}{\psi}(\boldsymbol{X}'\boldsymbol{X})^{-1}+\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}+\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\right)^{-1}(\boldsymbol{\beta}-\boldsymbol{b}_1^*)+\mathrm{RSS}_1^* \qquad (2.29)
\end{aligned}
$$

We integrate out $\boldsymbol{b}$ to find the likelihood to be

$$\int \frac{1}{(\sigma^2)^{\frac{p}{2}}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{b}_1^*)'\left((1+\psi)(\boldsymbol{X}'\boldsymbol{X})^{-1}+\psi(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right)^{-1}(\boldsymbol{\beta}-\boldsymbol{b}_1^*)\right]$$
$$\times\frac{\exp\left[-\frac{\mathrm{RSS}_1^*}{2\sigma^2\psi}\right]}{\psi^{\frac{n+r-2p+\alpha+1}{2}}}\times\frac{\exp\left[-\frac{\mathrm{RSS}}{2\sigma^2}\left(1+\frac{1}{\psi}\right)\right]}{(\sigma^2)^{\frac{2n+r-2p+\alpha+1}{2}-1}}\times(\mathrm{RSS})^{\frac{2n-2p+\alpha-1}{2}-1}$$
$$\times(\sigma^2)^{\frac{p}{2}}\left|\psi\boldsymbol{X}'\boldsymbol{X}+\left((\boldsymbol{X}'\boldsymbol{X})^{-1}+(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right)^{-1}\right|^{-\frac{1}{2}}d\psi\,d\mathrm{RSS}$$

Next integrating out RSS we have

$$\int \frac{1}{(\sigma^2)^{\frac{p}{2}}} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{b}_1^*)'\left((1+\psi)(\boldsymbol{X}'\boldsymbol{X})^{-1}+\psi(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right)^{-1}(\boldsymbol{\beta}-\boldsymbol{b}_1^*)\right]$$
$$\times\frac{\exp\left[-\frac{\mathrm{RSS}_1^*}{2\sigma^2\psi}\right]}{\psi^{\frac{n+r-2p+\alpha+1}{2}}}\times\frac{\left(\frac{\sigma^2\psi}{1+\psi}\right)^{\frac{2n-2p+\alpha-1}{2}}}{(\sigma^2)^{\frac{2n+r-2p+\alpha+1}{2}-1}}$$
$$\times(\sigma^2)^{\frac{p}{2}}\left|\psi\boldsymbol{X}'\boldsymbol{X}+\left((\boldsymbol{X}'\boldsymbol{X})^{-1}+(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right)^{-1}\right|^{-\frac{1}{2}}d\psi$$

Finally multiplying the integrand by our regular prior $\pi(\boldsymbol{\beta},\sigma^2)\propto(\sigma^2)^{-\frac{\delta+1}{2}}$, we find that the product breaks up into exactly three parts corresponding to the following posterior distributions

$$\boldsymbol{\beta}\,|\,\sigma^2,\psi,\boldsymbol{b}_1^* \sim N_p\left(\boldsymbol{b}_1^*,\sigma^2\left((1+\psi)(\boldsymbol{X}'\boldsymbol{X})^{-1}+\psi(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right)\right) \tag{2.30}$$
$$\sigma^2\,|\,\psi,\mathrm{RSS}_1^* \sim \text{Scale-inv-}\chi^2\left(r-p+\delta-1,\frac{\mathrm{RSS}_1^*}{\psi(r-p+\delta-1)}\right) \tag{2.31}$$
$$\psi \sim \beta'\left(\frac{n-p+\delta-1}{2},\frac{n-p+\alpha-\delta}{2}\right) \tag{2.32}$$

The posterior distributions are proper as long as $r>\max\{p,p-\delta+1,p-\alpha+1\}$, $n>p-\alpha+\delta$ and they match our results in Section 2.2 when $r=n$.

Remark 2.4.1. *We can think of an r based decision rule to analyze synthetic MLR data as follows. If $r<p$ we ignore the part of the data that is sensitive and base*

our analysis only on the non-sensitive part. This makes sense in the light of our simulation data where $n \gg p$. If $r > p$, then we use Method II (use whole data estimates to impute synthetic data). If $r > p$, $n - r > p$ then we use Method I (use sensitive part estimates to impute synthetic data). If $r = n$ then we use our regular methods of analyses outlined in Sections 2.1 and 2.2.

Chapter 3

Bayesian Analysis of Multiply Imputed Synthetic Data under the Multiple Linear Regression Model

3.1 Plug In Sampling method

Procedure I

We return to the case of a standard MLR model involving a *sensitive response* variable y and a $p \times 1$ dimensional vector of *non-sensitive predictors* $\boldsymbol{x}$ as in Section 2.1, but the analysis will take a slightly different route. In what follows, we emulate the development in Klein et al. (2019). To generate synthetic data $\boldsymbol{z}_1 = (z_{11}, \ldots, z_{1n})', \ldots, \boldsymbol{z}_m = (z_{m1}, \ldots, z_{mn})'$ for $m > 1$ under *plug-in sampling*, we start from the point estimates $\boldsymbol{b}$ and $\text{RSS}/(n-p)$, of $\boldsymbol{\beta}$ and σ^2, respectively. The synthetic data are obtained by drawing $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m \overset{\text{iid}}{\sim} N_n\left(\boldsymbol{X}\boldsymbol{b}, \frac{\text{RSS}}{n-p}\boldsymbol{I}_n\right)$. Equivalently, the synthetic data are obtained by drawing $z_{ji} \sim N(\boldsymbol{x}_i'\boldsymbol{b}, \frac{\text{RSS}}{n-p})$, independently for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

Let $\bar{z}_i = \frac{1}{m}\sum_{j=1}^{m} z_{ji}$, $S_{zi}^2 = \sum_{j=1}^{m}(z_{ji} - \bar{z}_i)^2$, and $S_z^2 = \sum_{i=1}^{n} S_{zi}^2$. If $m > 1$, then it follows that, conditional on $\boldsymbol{b}$ and RSS,

$$S_z^2 \sim \frac{\text{RSS}}{(n-p)}\chi^2_{n(m-1)}, \quad \bar{z}_i \sim N\left(\boldsymbol{x}_i'\boldsymbol{b}, \frac{\text{RSS}}{m(n-p)}\right),\ i = 1, \ldots, n$$

with these terms being (conditionally) independent. If $m = 1$, then the situation reduces to $\bar{z}_i = z_{1i}$ and $S_{zi}^2 = 0$ for $i = 1, \ldots, n$, and hence $S_z^2 = 0$.

Let $\bar{\boldsymbol{z}} = (\bar{z}_1, \ldots, \bar{z}_n)'$ and $\boldsymbol{b}_j^* = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{z}_j$. We define $\overline{\boldsymbol{b}^*} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\bar{\boldsymbol{z}} = \frac{1}{m}\sum_{j=1}^{m}\boldsymbol{b}_j^*$ and $S_{\text{comb}}^2 = S_z^2 + m(\bar{\boldsymbol{z}} - \boldsymbol{X}\overline{\boldsymbol{b}^*})'(\bar{\boldsymbol{z}} - \boldsymbol{X}\overline{\boldsymbol{b}^*})$, and note that, conditionally given $\boldsymbol{b}$ and RSS,

$$\overline{\boldsymbol{b}^*} \sim N_p\left(\boldsymbol{b}, \frac{\text{RSS}}{m(n-p)}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right), \quad S_{\text{comb}}^2 \sim \frac{\text{RSS}}{(n-p)}\chi^2_{n(m-1)+n-p}$$

which are (conditionally) independent and can be shown to be jointly sufficient for $(\boldsymbol{\beta}, \sigma^2)$. From Klein et al. (2019), we have the following result.

Theorem 3.1.1. The joint pdf of $(\overline{\boldsymbol{b}^*}, S_{\text{comb}}^2)$ is given by

$$f_{\boldsymbol{\beta},\sigma^2}(\overline{\boldsymbol{b}^*}, S_{\text{comb}}^2) \propto \int_0^\infty e^{-\frac{1}{2}\left[\frac{(\overline{\boldsymbol{b}^*}-\boldsymbol{\beta})'(\boldsymbol{X}'\boldsymbol{X})(\overline{\boldsymbol{b}^*}-\boldsymbol{\beta})}{\sigma^2(1+\frac{\psi}{m(n-p)})} + \frac{(n-p)S_{\text{comb}}^2}{\sigma^2\psi} + \psi\right]} \frac{(S_{\text{comb}}^2)^{\frac{nm-p}{2}-1}}{\sigma^{nm}\psi^{\frac{n(m-1)+p+2}{2}}}\left[1 + \frac{m(n-p)}{\psi}\right]^{-p/2} d\psi$$

Posterior distributions of $\boldsymbol{\beta}$ and σ^2

We choose the same prior on the parameters and multiply it with the above pdf so that product splits up into the following (conditional) posterior distributions:

$$\boldsymbol{\beta} \,|\, \overline{\boldsymbol{b}^*}, S_{\text{comb}}^2, \sigma^2, \psi \sim N_p\left(\overline{\boldsymbol{b}^*}, \sigma^2(1 + \frac{\psi}{m(n-p)})(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) \tag{3.1}$$

$$\sigma^2 \,|\, \overline{\boldsymbol{b}^*}, S_{\text{comb}}^2, \psi \sim \text{Scale-inv-}\chi^2\left(nm - p + \delta - 1, \frac{(n-p)S_{\text{comb}}^2}{\psi(nm - p + \delta - 1)}\right) \tag{3.2}$$

$$\psi \sim \chi^2_{n-p+\delta-1} \tag{3.3}$$

The posterior distributions are proper as long as $n > \max\{p, p-\delta+1\}$ (this also ensures that $nm - p + \delta - 1 > 0$ since $m > 1$, which is necessary for the posterior distribution of σ^2 to be proper). It is interesting to observe that plugging in $m = 1$ in the above formulas yields the same results we obtained for singly imputed plug-in sampling data as in Section 2.1.

Bayes Estimators of $\boldsymbol{\beta}$ and σ^2

$$\hat{\boldsymbol{\beta}}_{\text{BAYES}} = \mathrm{E}(\boldsymbol{\beta} \mid \overline{\boldsymbol{b}^*}, S^2_{\text{comb}}) = \mathrm{E}_\psi \, \mathrm{E}_{\sigma^2} \, \mathrm{E}(\boldsymbol{\beta} \mid \overline{\boldsymbol{b}^*}, S^2_{\text{comb}}, \sigma^2, \psi) = \mathrm{E}_\psi \, \mathrm{E}_{\sigma^2}(\overline{\boldsymbol{b}^*}) = \overline{\boldsymbol{b}^*}$$

$$\begin{aligned}\hat{\sigma}^2_{\text{BAYES}} &= \mathrm{E}(\sigma^2 \mid \overline{\boldsymbol{b}^*}, S^2_{\text{comb}}) = \mathrm{E}_\psi \, \mathrm{E}(\sigma^2 \mid \overline{\boldsymbol{b}^*}, S^2_{\text{comb}}, \psi) = \mathrm{E}_\psi(\tfrac{(n-p)S^2_{\text{comb}})}{\psi(nm-p+\delta-3)}) = \tfrac{(n-p)S^2_{\text{comb}}}{(nm-p+\delta-3)} \, \mathrm{E}_\psi(\tfrac{1}{\psi}) \\ &= \frac{(n-p)S^2_{\text{comb}}}{(nm-p+\delta-3)(n-p+\delta-3)}\end{aligned}$$

Credible Sets for $\boldsymbol{\beta}$ and σ^2

We will compute pivots for σ^2 and $\boldsymbol{\beta}$. For that we will first need to define and find the distribution of a few variables.

Consider $U \coloneqq \dfrac{1}{\sigma^2}$. It is easy to see that

$$U \mid S^2_{\text{comb}}, \psi \sim \Gamma\left(\frac{nm-p+\delta-1}{2}, \frac{(n-p)S^2_{\text{comb}}}{2\psi}\right) \tag{3.4}$$

Then a pivot for σ^2 can be defined as

$$\begin{aligned} K &\coloneqq \frac{S^2_{\text{comb}}}{\sigma^2} = U \times S^2_{\text{comb}} \\ \implies K \mid S^2_{\text{comb}}, \psi &\sim \Gamma\left(\frac{nm-p+\delta-1}{2}, \frac{(n-p)}{2\psi}\right) \\ \implies K \mid \psi &\sim \Gamma\left(\frac{nm-p+\delta-1}{2}, \frac{(n-p)}{2\psi}\right)\end{aligned}$$

where the second line above follows from (3.4) and the fact that if $X \sim \Gamma(\alpha, \beta)$ then $cX \sim \Gamma(\alpha, \beta/c)$.

Hence the pivot for σ^2 is computed as

$$\begin{aligned} \psi &\sim \chi^2_{n-p+\delta-1} \\ K \,|\, \psi &\sim \Gamma\left(\frac{nm-p+\delta-1}{2}, \frac{(n-p)}{2\psi}\right) \end{aligned}$$

A $(1-\gamma)$ level credible set for σ^2 based on $K = \mathrm{RSS}^*/\sigma^2$ is

$$\left[\frac{S^2_{\text{comb}}}{b_{n,p,\delta;\gamma}}, \frac{S^2_{\text{comb}}}{a_{n,p,\delta;\gamma}}\right]$$

where $a_{n,p,\delta;\gamma}$ and $b_{n,p,\delta;\gamma}$ are any two constants that satisfy $1-\gamma = P(a_{n,p,\delta;\gamma} \leq K \leq b_{n,p,\delta;\gamma})$. The length of the credible interval is $S^2_{\text{comb}}\left(\frac{1}{a_{n,p,\delta;\gamma}} - \frac{1}{b_{n,p,\delta;\gamma}}\right)$.

Now consider

$$V \coloneqq \frac{(\boldsymbol{\beta} - \overline{\boldsymbol{b}^*})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{\beta} - \overline{\boldsymbol{b}^*})}{\sigma^2(1 + \frac{\psi}{m(n-p)})}$$

Then $V \,|\, \overline{\boldsymbol{b}^*}, S^2_{\text{comb}}, \sigma^2, \psi \sim \chi^2_p$ and thus unconditionally $V \sim \chi^2_p$. Also V is independent of $(\overline{\boldsymbol{b}^*}, S^2_{\text{comb}}, \sigma^2, \psi)$ and thus V is independent of U. If

$$U^* \coloneqq \frac{U(n-p)S^2_{\text{comb}}}{\psi}$$

then $U^* \,|\, S^2_{\text{comb}}, \psi \sim \chi^2_{nm-p+\delta-1}$, and unconditionally $U^* \sim \chi^2_{nm-p+\delta-1}$. Now as V is independent of U, it is independent of U^*. Finally we define the pivot for $\boldsymbol{\beta}$ as

$$T^2_m \coloneqq \frac{(\boldsymbol{\beta} - \overline{\boldsymbol{b}^*})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{\beta} - \overline{\boldsymbol{b}^*})}{S^2_{\text{comb}}}$$

Then since T^2_m is not dependent on $(\overline{\boldsymbol{b}^*}, S^2_{\text{comb}})$ as V, U^* are both independent of $(\overline{\boldsymbol{b}^*}, S^2_{\text{comb}})$, we have

$$
\begin{aligned}
T_m^2 \,|\, \psi &= \frac{\sigma^2 V \left(1 + \dfrac{\psi}{m(n-p)}\right)}{\dfrac{\psi U^*}{U(n-p)}} \\
&= \frac{V}{U^*}\left[\frac{1+\frac{\psi}{m(n-p)}}{\psi}\right](n-p) \qquad\qquad [\because \sigma^2 = \frac{1}{U}] \\
&\sim \frac{\chi_p^2}{\chi_{nm-p+\delta-1}^2}\left(\frac{1}{m} + \frac{n-p}{\psi}\right) \\
&\sim \frac{\chi_p^2}{\chi_{nm-p+\delta-1}^2}\left(\frac{1}{m} + \frac{n-p}{\psi}\right) = \left[\frac{p}{nm-p+\delta-1}\right] F_{p,\, n-p+\delta-1}\left(\frac{1}{m} + \frac{n-p}{\psi}\right)
\end{aligned}
$$

Hence the pivot for $\boldsymbol{\beta}$ is computed as

$$
\begin{aligned}
\psi &\sim \chi_{n-p+\delta-1}^2 \\
T_m^2 \,|\, \psi &\sim \left[\frac{p}{nm-p+\delta-1}\right]\left[\frac{1}{m} + \frac{n-p}{\psi}\right] F_{p,\, nm-p+\delta-1}
\end{aligned}
$$

A $(1-\gamma)$ level credible ellipsoid for $\boldsymbol{\beta}$ based on T_m^2 is given by

$$
\{\boldsymbol{\beta} \;:\; T_m^2 \le d_{n,p,\delta,m;\gamma}\}
$$

where $d_{n,p,\delta,m;\gamma}$ satisfies $1-\gamma = P(T_m^2 \le d_{n,p,\delta,m;\gamma})$. The volume of the credible ellipsoid is

$$
V_{\boldsymbol{\beta}}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m, \boldsymbol{X}) = \frac{\pi^{p/2}}{\Gamma\left(\frac{p}{2}+1\right)}\left(d_{n,p,\delta,m;\gamma} S_{\text{comb}}^2\right)^{p/2} |\boldsymbol{X}'\boldsymbol{X}|^{-1/2}
$$

Remark 3.1.1. *If one is interested in the credible set of a single regression coefficient or more generally in the credible set of a linear combination of $\boldsymbol{\beta}$, namely, $\boldsymbol{A\beta} = \boldsymbol{\eta}$ where $\boldsymbol{A}$ is a $k \times p$ dimensional matrix with $rank(\boldsymbol{A}) = k < p$, we define $T_{m,\boldsymbol{\eta}}^2 = (\boldsymbol{\eta} - \boldsymbol{Ab}^*)'\{\boldsymbol{A}(\boldsymbol{XX}')^{-1}\boldsymbol{A}'\}^{-1}(\boldsymbol{\eta} - \boldsymbol{Ab}^*)/RSS^*$, and proceed by noting that*

$$
T_{m,\boldsymbol{\eta}}^2 \,|\, \psi \sim \left[\frac{k}{nm-p+\delta-1}\right]\left[\frac{1}{m} + \frac{n-p}{\psi}\right] F_{k,\, nm-p+\delta-1} \quad \textit{and} \quad \psi \sim \chi_{n-p+\delta-1}^2
$$

Procedure II

We can adapt the procedure in Section 2.1 so that the sufficient statistics for the released data is $(\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_m^*, \mathrm{RSS}_1^*, \ldots, \mathrm{RSS}_m^*)$ to obtain

$$\boldsymbol{\beta} \,|\, \sigma^2, \psi, \overline{\boldsymbol{b}^*} \sim N_p\left(\overline{\boldsymbol{b}^*}, \frac{\sigma^2}{m}(1+\psi)(\boldsymbol{X}\boldsymbol{X}')^{-1}\right) \tag{3.5}$$

$$\sigma^2 \,|\, \psi, \boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_m^*, \mathrm{RSS}_1^*, \ldots, \mathrm{RSS}_m^* \sim \text{Scale-inv-}\chi^2(\nu, \tau^2) \tag{3.6}$$

$$\pi(\psi \,|\, \boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_m^*, \mathrm{RSS}_1^*, \ldots, \mathrm{RSS}_m^*) \propto \psi^{-\frac{(m-1)(n-p)}{2}-1}(1+\psi)^{-\frac{(m-1)p}{2}} e^{-\frac{(n-p)\psi}{2}} \left\{\nu\tau^2\right\}^{-\frac{\nu}{2}} \tag{3.7}$$

where $\overline{\boldsymbol{b}^*} = \frac{1}{m}\sum_{j=1}^m$, $\overline{\mathrm{RSS}^*} = \frac{1}{m}\sum_{j=1}^m \mathrm{RSS}_j^*$ and

$$\nu = nm - p + \delta - 1$$

$$\nu\tau^2 = \sum_{j=1}^m \left(\boldsymbol{b}_j^* - \overline{\boldsymbol{b}^*}\right)' \frac{\boldsymbol{X}'\boldsymbol{X}}{1+\psi}\left(\boldsymbol{b}_j^* - \overline{\boldsymbol{b}^*}\right) + \frac{m}{\psi}\,\overline{\mathrm{RSS}^*}$$

We can sample ψ from (3.7) using Accept-Reject algorithm as follows:

$$\pi(\psi \,|\, \text{data}) \leq \frac{2^{\frac{n+(m-2)p+\delta-1}{2}}\, \Gamma\left(\frac{n+(m-2)p+\delta-1}{2}\right)}{(n-p)^{\frac{n+(m-2)p+\delta-1}{2}} \left(\sum_{j=1}^m \mathrm{RSS}_j^*\right)^{\frac{nm-p+\delta-1}{2}}} f_{\text{ScaledChi}}(\psi)$$

where $f_{\text{ScaledChi}}(\psi)$ is the pdf of a Gamma$\left(\frac{n+(m-2)p+\delta-1}{2}, \frac{n-p}{2}\right)$ distribution. This would require the additional assumption $n + (m-2)p + \delta - 1 > 0$.

Simulation studies

The setup is same as in Section 2.3. As before in the last chapter the general trend observed when $m = 1$ is still followed here, which entails lower coverage for σ^2 compared to $\boldsymbol{\beta}$ that improves asymptotically but not quite (the effect is more prominent

for large values of δ); the coverage for σ^2 increases with increasing δ for a while and then it dips. Both of the effects are observed for $\boldsymbol{\beta}$ as well but it is not as dramatic as the coverage for $\boldsymbol{\beta}$ is close to 0.95 to begin with and asymptotically the Bernstein-von Mises theorem holds. Increasing δ decreases the size of credible sets and slightly worsens the Bayes estimator of σ^2 but not that of $\boldsymbol{\beta}$. Lastly, credible sets shrink and Bayes estimators for both parameters work reasonably well in the asymptotic case. Now let us get to the peculiarities of the situation at hand, i.e. the case when $m > 1$. The maximum coverage for σ^2 is actually higher when $m = 1$, with $m > 5$ being slightly better off than $m > 10$ overall for coverage of σ^2. For large δ, the coverage for σ^2 is better in the case of $m > 1$ in the regular case, and $m = 1$ is better in the asymptotic case. The coverage for $\boldsymbol{\beta}$ seems to be close to 0.95 for both $m = 1$ and $m > 1$, with $m > 1$ being very slighly better off than $m = 1$ for large δ. The coverage for $\boldsymbol{\beta}$ is also slightly better overall for $m = 5$ than $m = 10$. The size of credible sets decrease with increasing m, and they become even more tighter asymptotically. We observe this quirk for both the parameters, suggesting that the gain in efficiency of estimation due to an increase in m is paid for by the decrease in coverage. The Bayes estimator for σ^2 is better in the case $m > 1$ for large values of δ in the regular, although asymptotically there seems to be no difference. The Bayes estimator for $\boldsymbol{\beta}$ behaves similarly for all values of m. Finally, for $\boldsymbol{\beta}$ overall, the asymptotic case seems identical in $m = 1$ and $m > 1$. The behavior of the coverage of σ^2 and $\boldsymbol{\beta}$ with respect to different values of δ in the case $n = 500$ (depicted by alternating dashes and dots), $n = 1000$ (depicted by solid lines), asymptotic case $n = 10000$ (depicted by dashed lines) are represented in Figures 3.1(a), 3.1(c) and Figures 3.1(b), 3.1(d) respectively.

Table 3.1: Inference for $\boldsymbol{\beta}$ and σ^2 for MI PIS data with $m = 5$, $n = 500$

	σ^2			$\boldsymbol{\beta}$		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.949	0.277	1.007	0.95	4.966e-05	(10.000, 2.000, 1.999, -3.000, -0.999, -2.001, 1.000, 1.999, 2.000, 3.999)'
0.5	0.947	0.277	1.007	0.95	5.845e-05	(10.000, 2.000, 2.000, -3.000, -1.000, -2.002, 0.998, 2.001, 2.001, 4.000)'
0.8	0.952	0.277	1.006	0.951	4.979e-05	(10.001, 2.000, 2.000, -3.001, -0.999, -2.002, 0.999, 2.000, 2.000, 4.001)'
1	0.946	0.276	1.004	0.953	4.031e-05	(10.001, 2.000, 2.000, -3.001, -0.999, -2.002, 1.001, 2.000, 1.999, 4.002)'
2	0.950	0.275	1.002	0.95	5.778e-05	(10.000, 2.000, 2.000, -3.000, -1.000, -2.001, 1.002, 1.999, 1.999, 3.999)'
3	0.951	0.275	1.001	0.948	6.429e-05	(10.001, 2.000, 2.000, -2.999, -1.000, -2.003, 0.999, 2.000, 1.998, 3.999)'
4	0.946	0.273	0.998	0.948	7.037e-05	(9.999, 2.001, 2.000, -3.000, -0.999, -2.000, 0.999, 2.001, 2.000, 4.000)'
10	0.939	0.268	0.984	0.942	4.706e-05	(10.000, 2.000, 2.000, -3.000, -1.001, -2.003, 0.999, 2.000, 2.001, 4.002)'
20	0.898	0.259	0.960	0.938	6.336e-05	(10.000, 2.000, 2.000, -2.999, -1.000, -2.001, 0.998, 1.999, 2.000, 3.999)'
30	0.828	0.251	0.938	0.929	4.428e-05	(9.999, 2.001, 2.000, -3.000, -1.000, -2.003, 1.001, 1.999, 2.002, 3.999)'
50	0.594	0.236	0.894	0.918	3.348e-05	(9.999, 2.000, 2.000, -2.999, -1.000, -1.999, 0.999, 2.000, 2.000, 4.001)'
100	0.083	0.204	0.804	0.873	2.036e-05	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.001, 2.000, 2.002, 3.999)'

Table 3.2: Inference for $\boldsymbol{\beta}$ and σ^2 for MI PIS data with $m = 5$, $n = 1000$

	σ^2			$\boldsymbol{\beta}$		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.947	0.194	1.003	0.952	1.316e-06	(10.001, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 1.999, 1.999, 3.997)'
0.5	0.949	0.194	1.004	0.951	1.912e-06	(10.001, 2.000, 2.000, -3.00, -0.999, -2.000, 1.000, 1.998, 1.997, 3.997)'
0.8	0.956	0.194	1.003	0.950	1.681e-06	(10.001, 2.000, 2.000, -3.000, -1.000, -2.002, 0.999, 2.000, 1.999, 3.999)'
1	0.951	0.194	1.003	0.955	1.925e-06	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 3.999)'
2	0.947	0.193	1.002	0.952	1.145e-06	(9.999, 2.000, 2.000, -3.000, -1.000, -1.999, 1.001, 2.000, 2.000, 4.001)'
3	0.945	0.193	1.000	0.948	1.665e-06	(9.999, 2.000, 2.000, -3.000, -1.000, -1.998, 1.002, 2.001, 2.001, 4.004)'
4	0.950	0.193	0.999	0.947	1.511e-06	(10.001, 2.000, 2.000, -3.001, -1.000, -2.001, 1.001, 2.000, 2.000, 3.999)'
10	0.946	0.191	0.991	0.949	1.538e-06	(10.000, 2.000, 2.000, -3.000, -1.001, -2.002, 1.001, 2.000, 2.002, 3.999)'
20	0.925	0.188	0.980	0.943	1.510e-06	(10.003, 2.000, 2.000, -3.000, -1.001, -2.001, 0.998, 1.998, 1.998, 3.998)'
30	0.894	0.185	0.969	0.941	1.690e-06	(10.001, 2.000, 2.000, -3.000, -1.001, -1.999, 1.000, 2.000, 2.000, 4.001)'
50	0.775	0.179	0.946	0.931	1.673e-06	(10.000, 2.000, 2.000, -3.000, -1.000, -2.003, 1.000, 2.001, 2.000, 3.999)'
100	0.335	0.166	0.894	0.917	8.961e-07	(9.999, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.001, 2.000, 4.001)'

Table 3.3: Inference for $\boldsymbol{\beta}$ and σ^2 for MI PIS data with $m = 5$, $n = 10000$

	σ^2			$\boldsymbol{\beta}$		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.953	0.061	1.000	0.952	1.551e-11	(10.000, 2.000, 2.000, -3.000, -1.000, -1.999, 1.001, 2.001, 2.000, 4.001)'
0.5	0.948	0.061	1.000	0.95	1.539e-11	(10.000, 2.000, 2.000, -3.000, -1.000, -2.001, 1.000, 2.000, 2.000, 4.000)'
0.8	0.948	0.061	1.000	0.946	1.425e-11	(10.000, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.001, 2.000, 4.000)'
1	0.946	0.061	1.000	0.950	1.540e-11	(9.999, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.000, 2.001, 4.000)'
2	0.950	0.061	1.000	0.951	1.556e-11	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 4.001)'
3	0.954	0.061	1.000	0.952	1.556e-11	(10.000, 2.000, 2.000, -3.000, -1.000, -1.999, 1.001, 2.000, 2.001, 4.000)'
4	0.952	0.061	1.000	0.950	1.542e-11	(10.000, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.000, 2.000, 4.000)'
5	0.697	0.061	0.999	0.949	1.034e-32	(10.000, 2.000, 1.999, -3.000, -1.000, -2.000, 1.000, 1.999, 2.000, 3.999)'
10	0.948	0.061	0.999	0.952	1.480e-11	(10.001, 2.000, 2.000, -3.000, -1.000, -2.001, 1.000, 1.999, 2.000, 4.000)'
20	0.947	0.061	0.998	0.951	1.474e-11	(10.000, 2.000, 2.000, -3.000, -1.000, -2.001, 1.000, 2.000, 2.000, 4.000)'
30	0.944	0.061	0.997	0.952	1.659e-11	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 4.000)'
50	0.933	0.060	0.994	0.947	1.648e-11	(10.000, 2.000, 2.000, -3.000, -1.000, -2.001, 1.000, 2.000, 2.000, 4.000)'
100	0.881	0.060	0.988	0.948	1.456e-11	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 3.999)'

Table 3.4: Inference for $\boldsymbol{\beta}$ and σ^2 for MI PIS data with $m = 10$, $n = 500$

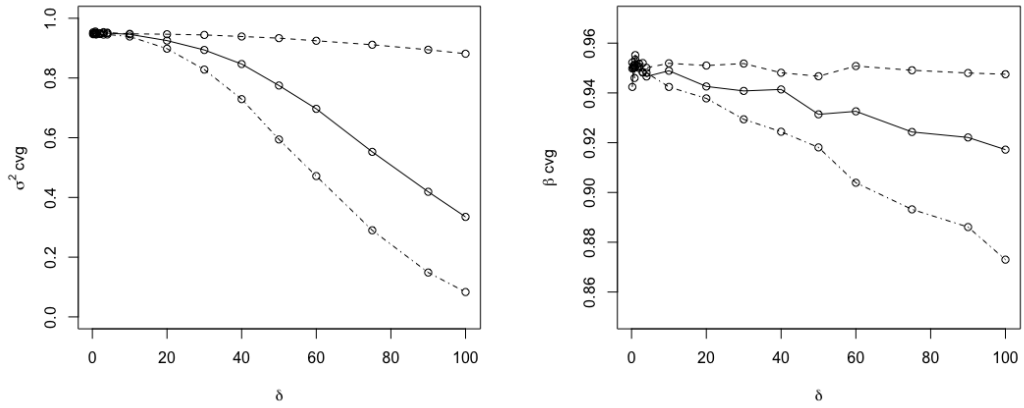
	σ^2			$\boldsymbol{\beta}$		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.952	0.265	1.007	0.954	3.167e-05	(10.002, 2.000, 2.000, -2.999, -1.001, -2.002, 1.000, 1.999, 2.000, 4.000)'
0.5	0.948	0.265	1.006	0.952	4.888e-05	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 0.999, 2.000, 1.998, 4.000)'
0.8	0.949	0.264	1.005	0.947	2.500e-05	(10.001, 2.000, 2.000, -3.000, -1.000, -2.001, 0.999, 2.000, 1.998, 3.999)'
1	0.949	0.265	1.005	0.954	2.831e-05	(9.999, 2.000, 2.000, -3.000, -1.000, -2.001, 1.001, 2.001, 2.002, 3.999)'
2	0.944	0.264	1.002	0.953	4.065e-05	(9.999, 2.000, 2.000, -3.000, -0.999, -2.000, 1.000, 1.998, 1.999, 4.000)'
3	0.949	0.263	1.000	0.950	3.059e-05	(10.000, 2.000, 2.000, -3.001, -1.000, -1.998, 1.000, 2.000, 1.998, 4.000)'
4	0.946	0.262	0.997	0.952	3.543e-05	(10.001, 1.999, 2.000, -3.000, -0.999, -1.999, 1.000, 2.000, 2.000, 4.000)'
10	0.941	0.257	0.985	0.948	3.629e-05	(10.001, 1.999, 2.000, -3.000, -1.000, -1.998, 0.998, 1.999, 1.999, 3.999)'
20	0.901	0.249	0.963	0.933	2.978e-05	(10.000, 2.000, 2.000, -3.001, -0.999, -2.001, 1.001, 2.000, 2.001, 4.000)'
30	0.831	0.242	0.942	0.933	2.293e-05	(10.000, 2.000, 2.000, -3.000, -0.999, -1.997, 1.001, 2.004, 2.002, 4.000)'
50	0.624	0.227	0.903	0.915	2.201e-05	(10.002, 2.000, 2.000, -3.000, -1.001, -2.000, 0.999, 2.001, 1.999, 4.001)'
100	0.105	0.198	0.819	0.880	1.463e-05	(10.001, 2.000, 2.000, -3.000, -1.000, -2.002, 1.001, 2.000, 1.998, 4.000)'

Table 3.5: Inference for $\boldsymbol{\beta}$ and σ^2 for MI PIS data with $m = 10$, $n = 1000$

	σ^2			$\boldsymbol{\beta}$		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.95	0.185	1.002	0.947	1.271e-06	(10.002, 1.999, 2.000, -2.999, -1.000, -2.000, 0.998, 1.999, 2.000, 3.999)'
0.5	0.949	0.186	1.003	0.952	7.804e-07	(10.000, 1.999, 2.000, -3.000, -1.000, -2.000, 0.999, 2.002, 2.000, 3.999)'
0.8	0.951	0.185	1.002	0.95	1.159e-06	(10.000, 2.000, 2.000, -3.000, -1.000, -2.002, 1.000, 1.999, 2.000, 3.999)'
1	0.95	0.185	1.001	0.95	8.669e-07	(9.999, 2.001, 2.000, -3.000, -0.999, -2.001, 0.998, 2.000, 1.998, 3.997)'
2	0.949	0.185	1.001	0.947	1.070e-06	(10.000, 2.000, 2.000, -3.000, -1.000, -2.001, 1.000, 2.000, 2.001, 4.000)'
3	0.949	0.185	0.999	0.951	1.210e-06	(10.000, 2.000, 2.000, -3.000, -1.000, -1.999, 1.001, 2.000, 2.000, 4.001)'
4	0.956	0.184	0.999	0.95	8.738e-07	(10.001, 2.000, 2.000, -3.000, -1.001, -2.001, 1.001, 2.000, 1.999, 3.998)'
10	0.944	0.183	0.992	0.946	9.520e-07	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.001, 2.000, 2.000, 4.000)'
20	0.929	0.180	0.982	0.944	9.566e-07	(9.999, 2.000, 2.000, -3.000, -1.000, -2.000, 1.001, 2.001, 2.000, 3.998)'
30	0.895	0.177	0.971	0.939	1.043e-06	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 3.997)'
50	0.790	0.172	0.951	0.938	7.767e-07	(10.000, 2.000, 2.000, -3.000, -1.000, -1.999, 1.000, 2.000, 2.000, 4.000)'
100	0.374	0.160	0.902	0.920	6.855e-07	(10.001, 2.001, 2.000, -3.000, -1.000, -2.000, 0.999, 2.000, 2.000, 4.000)'

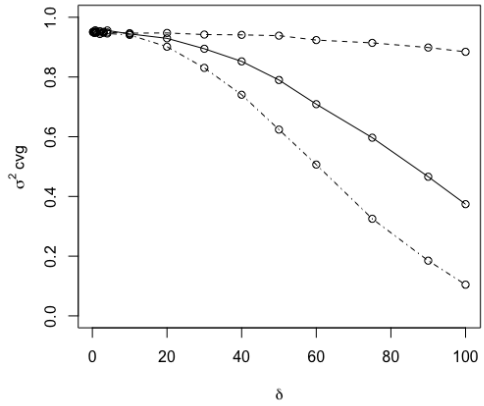
Table 3.6: Inference for $\boldsymbol{\beta}$ and σ^2 for MI PIS data with $m = 10$, $n = 10000$

	σ^2			$\boldsymbol{\beta}$		
	avg	*est*	*Bayes*	*avg*	*est*	*Bayes*
δ	*cvg*	*len*	*est*	*cvg*	*vol*	*est*
0.2	0.95	0.058	1.001	0.949	9.535e-12	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 0.999, 2.000, 2.000, 4.000)'
0.5	0.95	0.058	1.000	0.954	9.863e-12	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 4.000)'
0.8	0.946	0.058	1.000	0.952	9.799e-12	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 4.000)'
1	0.953	0.058	1.000	0.953	9.854e-12	(9.999, 2.000, 2.000, -3.000, -1.000, -2.000, 1.001, 2.001, 2.001, 4.001)'
2	0.953	0.058	1.000	0.951	1.011e-11	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.001, 2.000, 2.000, 4.000)'
3	0.951	0.058	1.000	0.95	1.006e-11	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 4.000)'
4	0.949	0.058	1.000	0.950	9.727e-12	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 4.000)'
10	0.947	0.058	0.999	0.947	9.419e-12	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.001, 2.000, 2.000, 4.000)'
20	0.948	0.058	0.998	0.948	9.377e-12	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.000, 4.000)'
30	0.942	0.058	0.997	0.951	9.503e-12	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 1.000, 2.000, 2.001, 4.000)'
50	0.938	0.058	0.995	0.949	9.948e-12	(10.001, 2.000, 2.000, -3.000, -1.000, -2.001, 1.000, 2.000, 2.000, 4.000)'
100	0.884	0.057	0.989	0.946	9.302e-12	(10.000, 2.000, 2.000, -3.000, -1.000, -2.000, 0.999, 2.000, 2.000, 4.000)'

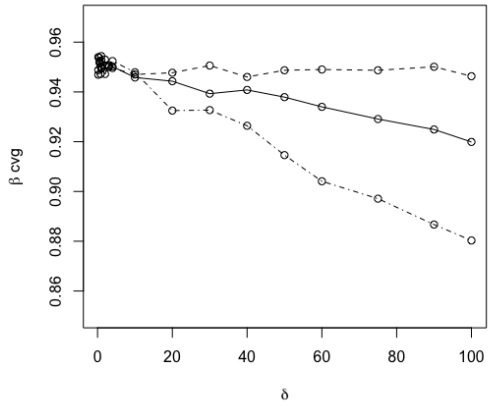


(a) σ^2, $m = 5$

(b) $\boldsymbol{\beta}$, $m = 5$



(c) σ^2, $m = 10$



(d) $\boldsymbol{\beta}$, $m = 10$

Figure 3.1: Variation in coverage of $\boldsymbol{\beta}$ and σ^2 with respect to δ for MI PIS MLR data (–·–·– $n = 500$, — $n = 1000$, – – – $n = 10000$)

3.2 Posterior Predictive Sampling method

Procedure I

We consider the setup described in Section 2.2. The synthetic data are generated by repeating the following steps below independently for each $j = 1, \ldots, m$.

(a) Draw $(\boldsymbol{\beta}_j^*, \sigma_j^{*2})$ from the posterior distribution (2.11) and (2.12).

(b) Draw $\boldsymbol{z}_j = (z_{j1}, \ldots, z_{jn})' \sim N_n(\boldsymbol{X}\boldsymbol{\beta}_j^*, \sigma_j^{*2}\boldsymbol{I}_n)$.

The released synthetic data are $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m$ along with the matrix of predictor variables $\boldsymbol{X}$. Similar as before the sufficient statistics for the synthetic data are: $\boldsymbol{b}_j^* = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{z}_j$ and $\text{RSS}_j^* = (\boldsymbol{z}_j - \boldsymbol{X}\boldsymbol{b}_j^*)'(\boldsymbol{z}_j - \boldsymbol{X}\boldsymbol{b}_j^*)$, for $j = 1, \ldots, m$. It can be shown that $(\boldsymbol{b}_1^*, \text{RSS}_1^*), \ldots, (\boldsymbol{b}_m^*, \text{RSS}_m^*)$ are jointly sufficient for $(\boldsymbol{\beta}, \sigma^2)$. In view of the sampling mechanism above, it follows that from the frequentist perspective, the joint distribution of $\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_m^*$, $\text{RSS}_1^*, \ldots, \text{RSS}_m^*$, $\boldsymbol{\beta}_1^*, \ldots, \boldsymbol{\beta}_m^*$, $\sigma_1^{*2}, \ldots, \sigma_m^{*2}$, $\boldsymbol{b}$ and RSS has the following hierarchical structure:

$$
\begin{aligned}
\boldsymbol{b}_j^* \,|\, \text{RSS}_1^*, \ldots, \text{RSS}_m^*, \boldsymbol{\beta}_1^*, \ldots, \boldsymbol{\beta}_m^*, \sigma_1^{*2}, \ldots, \sigma_m^{*2}, \boldsymbol{b}, \text{RSS} &\sim N_p(\boldsymbol{\beta}_j^*,\, \sigma_j^{*2}(\boldsymbol{X}'\boldsymbol{X})^{-1}) \\
\text{RSS}_j^* \,|\, \boldsymbol{\beta}_1^*, \ldots, \boldsymbol{\beta}_m^*, \sigma_1^{*2}, \ldots, \sigma_m^{*2}, \boldsymbol{b}, \text{RSS} &\sim \sigma_j^{*2}\chi^2_{n-p} \\
\boldsymbol{\beta}_j^* \,|\, \sigma_1^{*2}, \ldots, \sigma_m^{*2}, \boldsymbol{b}, \text{RSS} &\sim N_p(\boldsymbol{b},\, \sigma_j^{*2}(\boldsymbol{X}'\boldsymbol{X})^{-1}) \\
\sigma_j^{*2} \,|\, \boldsymbol{b}, \text{RSS} &\sim \frac{\text{RSS}}{\chi^2_{n-p+\alpha-1}} \\
\boldsymbol{b} &\sim N_p(\boldsymbol{\beta},\, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}) \\
\text{RSS} &\sim \sigma^2\chi^2_{n-p}
\end{aligned}
$$

which are generated independently for $j = 1, \ldots, m$, whenever applicable.

Hence we have $f(\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_m^*, \text{RSS}_1^*, \ldots, \text{RSS}_m^*, \boldsymbol{\beta}_1^*, \ldots, \boldsymbol{\beta}_m^*, \sigma_1^{*2}, \ldots, \sigma_m^{*2}, \boldsymbol{b}, \text{RSS})$

$$
\begin{aligned}
&= \prod_{j=1}^{m} (2\pi\sigma_j^{*2})^{-p/2} \left|\boldsymbol{X}'\boldsymbol{X}\right|^{1/2} \exp\left[-\frac{1}{2\sigma_j^{*2}}(\boldsymbol{b}_j^* - \boldsymbol{\beta}_j^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{b}_j^* - \boldsymbol{\beta}_j^*)\right] \\
&\times \prod_{j=1}^{m} \frac{(\text{RSS}_j^*)^{\frac{n-p}{2}-1}}{2^{\frac{n-p}{2}}\,\Gamma(\frac{n-p}{2})} (\sigma_j^{*2})^{-(n-p)/2} \exp\left[-\frac{\text{RSS}_j^*}{2\sigma_j^{*2}}\right] \\
&\times \prod_{j=1}^{m} (2\pi\sigma_j^{*2})^{-p/2} \left|\boldsymbol{X}'\boldsymbol{X}\right|^{1/2} \exp\left[-\frac{1}{2\sigma_j^{*2}}(\boldsymbol{\beta}_j^* - \boldsymbol{b})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{\beta}_j^* - \boldsymbol{b})\right] \\
&\times \prod_{j=1}^{m} \frac{(\text{RSS})^{(n-p+\alpha-1)/2}}{2^{(n-p+\alpha-1)/2}\,\Gamma\left(\frac{n-p+\alpha-1}{2}\right)} (\sigma_j^{*2})^{-(n-p+\alpha-1)/2-1} \exp\left[-\frac{\text{RSS}}{2\sigma_j^{*2}}\right] \\
&\times (2\pi\sigma^2)^{-p/2} \left|\boldsymbol{X}'\boldsymbol{X}\right|^{1/2} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{b} - \boldsymbol{\beta})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{b} - \boldsymbol{\beta})\right] \\
&\times \frac{(\text{RSS})^{\frac{n-p}{2}-1}}{2^{\frac{n-p}{2}}\,\Gamma(\frac{n-p}{2})} (\sigma^2)^{-(n-p)/2} \exp\left[-\frac{\text{RSS}}{2\sigma^2}\right]
\end{aligned}
$$

After integrating out $\boldsymbol{\beta}_j^*$'s, $\boldsymbol{b}$, RSS, simplifying we get,

$$
\begin{aligned}
&(2\pi)^{-p/2} \frac{\left(\frac{1}{\sigma^2}\right)^{p/2} \left(\sum_{j=1}^m \frac{1}{2\sigma_j^{*2}}\right)^{p/2}}{\left(\frac{1}{\sigma^2} + \sum_{j=1}^m \frac{1}{2\sigma_j^{*2}}\right)^{p/2}} \left|\boldsymbol{X}'\boldsymbol{X}\right|^{1/2} \\
&\times \exp\left[-\frac{1}{2}\frac{\left(\frac{1}{\sigma^2}\right)\left(\sum_{j=1}^m \frac{1}{2\sigma_j^{*2}}\right)}{\left(\frac{1}{\sigma^2} + \sum_{j=1}^m \frac{1}{2\sigma_j^{*2}}\right)} \left(\boldsymbol{\beta} - \frac{\sum_{j=1}^m \frac{\boldsymbol{b}_j^*}{2\sigma_j^{*2}}}{\sum_{j=1}^m \frac{1}{2\sigma_j^{*2}}}\right)' (\boldsymbol{X}'\boldsymbol{X}) \left(\boldsymbol{\beta} - \frac{\sum_{j=1}^m \frac{\boldsymbol{b}_j^*}{2\sigma_j^{*2}}}{\sum_{j=1}^m \frac{1}{2\sigma_j^{*2}}}\right)\right] \\
&\times \prod_{j=1}^{m} (2\pi)^{-p/2} \left(2\sigma_j^{*2}\right)^{-p/2} \left|\boldsymbol{X}'\boldsymbol{X}\right|^{1/2} \exp\left[-\frac{1}{2}\cdot\frac{1}{2\sigma_j^{*2}} \left(\boldsymbol{b}_j^* - \frac{\sum_{j=1}^m \frac{\boldsymbol{b}_j^*}{2\sigma_j^{*2}}}{\sum_{j=1}^m \frac{1}{2\sigma_j^{*2}}}\right)' (\boldsymbol{X}'\boldsymbol{X}) \left(\boldsymbol{b}_j^* - \frac{\sum_{j=1}^m \frac{\boldsymbol{b}_j^*}{2\sigma_j^{*2}}}{\sum_{j=1}^m \frac{1}{2\sigma_j^{*2}}}\right)\right] \\
&\times \frac{\left(\sum_{j=1}^m \frac{1}{2\sigma_j^{*2}}\right)^{-\frac{p}{2}} \left(\frac{1}{\sigma^2} + \sum_{j=1}^m \frac{1}{\sigma_j^{*2}}\right)^{-\frac{(m+1)(n-p)+m(\alpha-1)}{2}}}{(\sigma^2)^{\frac{n-p}{2}} \left(\prod_{j=1}^m (\sigma_j^{*2})^{\frac{n-p+\alpha-1}{2}} + \frac{n-p}{2} + 1\right)} \exp\left[-\sum_{j=1}^m \frac{\text{RSS}_j^*}{2\sigma_j^{*2}}\right] \\
&\times \frac{\Gamma\left(\frac{m(n-p+\alpha-1)+(n-p)}{2}\right)}{\left(\Gamma\left(\frac{n-p+\alpha-1}{2}\right)\right)^m \left(\Gamma\left(\frac{n-p}{2}\right)\right)^{m+1}} (2\pi)^{p/2} \left|\boldsymbol{X}'\boldsymbol{X}\right|^{-1/2} \left(\prod_{j=1}^m \frac{(\text{RSS}_j^*)^{\frac{n-p}{2}-1}}{2^{\frac{n-p}{2}}}\right)
\end{aligned}
\tag{3.8}
$$

It is clear from the expression above that the part involving $\boldsymbol{\beta}$ separates out in the first line and it's posterior distribution is obvious. We also observe that the quantity inside the exponential in the second line vanishes for $m = 1$. We multiply the joint distribution of $(\boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_m^*, \mathrm{RSS}_1^*, \ldots, \mathrm{RSS}_m^*, \sigma_1^{*2}, \ldots, \sigma_m^{*2})$ in (3.8) by our usual prior, and the parameters separate out as follows:

$$\boldsymbol{\beta} \,|\, \sigma^2, \sum_{j=1}^m \frac{\boldsymbol{b}_j^*}{\sigma_j^{*2}}, \sum_{j=1}^m \frac{1}{\sigma_j^{*2}} \sim N_p \left(\frac{\sum_{j=1}^m \frac{\boldsymbol{b}_j^*}{\sigma_j^{*2}}}{\sum_{j=1}^m \frac{1}{\sigma_j^{*2}}}, \left(1 + \sigma^2 \left(\sum_{j=1}^m \frac{1}{\sigma_j^{*2}} \right) \right) (\boldsymbol{X}'\boldsymbol{X})^{-1} \right) \quad (3.9)$$

$$\sigma^2 \left(\sum_{j=1}^m \frac{1}{\sigma_j^{*2}} \right) \sim \beta' \left(\frac{m(n-p+\alpha-1) - \delta + 1}{2}, \frac{n-p+\delta-1}{2} \right) \quad (3.10)$$

so that $\sigma^2 \left(\sum_{j=1}^m 1/\sigma_j^{*2} \right)$ is independent of both latent variables and data. The posterior distributions are proper as long as $n > \max\left\{p, p-\delta+1, p-\alpha+1, p-\alpha+1+\frac{\delta-1}{m}\right\}$. The latent variables have the following distribution

$$\begin{aligned}
&g(\sigma_1^{*2}, \ldots, \sigma_m^{*2} \,|\, \boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_m^*, \mathrm{RSS}_1^*, \ldots, \mathrm{RSS}_m^*) \\
&\propto \frac{\left(\prod_{j=1}^m (\sigma_j^{*2}) \right)^{-\frac{2n-p+\alpha+1}{2}}}{\left(\sum_{j=1}^m \frac{1}{\sigma_j^{*2}} \right)^{\frac{m(n-p+\alpha-1)+p-\delta}{2}}} \\
&\exp\left[-\frac{1}{4} \sum_{j=1}^m \frac{1}{\sigma_j^{*2}} \left\{ \left(\boldsymbol{b}_j^* - \frac{\sum_{j=1}^m \frac{\boldsymbol{b}_j^*}{\sigma_j^{*2}}}{\sum_{j=1}^m \frac{1}{\sigma_j^{*2}}} \right)' (\boldsymbol{X}\boldsymbol{X}') \left(\boldsymbol{b}_j^* - \frac{\sum_{j=1}^m \frac{\boldsymbol{b}_j^*}{\sigma_j^{*2}}}{\sum_{j=1}^m \frac{1}{\sigma_j^{*2}}} \right) + 2\mathrm{RSS}_j^* \right\} \right]
\end{aligned}$$

If we make the transformation where $\sigma_j^2 = 1/\sigma_j^{*2}$, then the distribution looks like

$$\begin{aligned}
&h(\sigma_1^2, \ldots, \sigma_m^2 \,|\, \boldsymbol{b}_1^*, \ldots, \boldsymbol{b}_m^*, \mathrm{RSS}_1^*, \ldots, \mathrm{RSS}_m^*) \\
&\propto \frac{\left(\prod_{j=1}^m (\sigma_j^2) \right)^{\frac{2n-p+\alpha-1}{2}}}{\left(\sum_{j=1}^m \sigma_j^2 \right)^{\frac{m(n-p+\alpha-1)+p-\delta}{2}}} \\
&\exp\left[-\frac{1}{4} \sum_{j=1}^m \sigma_j^2 \left\{ \left(\boldsymbol{b}_j^* - \frac{\sum_{j=1}^m \sigma_j^2 \boldsymbol{b}_j^*}{\sum_{j=1}^m \sigma_j^2} \right)' (\boldsymbol{X}\boldsymbol{X}') \left(\boldsymbol{b}_j^* - \frac{\sum_{j=1}^m \sigma_j^2 \boldsymbol{b}_j^*}{\sum_{j=1}^m \sigma_j^2} \right) + 2\mathrm{RSS}_j^* \right\} \right]
\end{aligned} \quad (3.11)$$

Let us denote the quantity inside the exponential of (3.11) as Q. Our goal here is to sample from (3.11). If we make the following transformation

$$u_1 = \sum_{j=1}^{m} \sigma_j^2, u_2 = \frac{\sigma_j^2}{\sum_{j=1}^{m} \sigma_j^2}, \ldots, u_m = \frac{\sigma_m^2}{\sum_{j=1}^{m} \sigma_j^2} \tag{3.12}$$

then with some abuse of notation for Q, we can sample $(y_1, y_2, \ldots, y_m)$ using conditional sampling and Accept-Reject algorithm in the following manner:

$$u_1 \,|\, u_2, \ldots, u_m, \text{data} \sim \frac{2\,\chi^2_{nm-p+\delta-1}}{Q} \tag{3.13}$$

$$\pi(u_2, \ldots, u_m \,|\, \text{data}) \leq \frac{\mathbf{B}\left(\frac{2n-p+\alpha-1}{2}, \ldots, \frac{2n-p+\alpha-1}{2}\right)}{(\text{RSS}^*_{\min})^{\frac{nm-p+\delta-1}{2}}}\, g_{\text{Dir}}(u_2, \ldots, u_m) \tag{3.14}$$

where $\mathbf{B}(\alpha_1, \ldots, \alpha_m)$ is the multivariate beta function, and $g_{\text{Dir}}(u_2, \ldots, u_m)$ is the pdf of an m^{th} order Dirichlet $\left(\frac{2n-p+\alpha-1}{2}, \ldots, \frac{2n-p+\alpha-1}{2}\right)$ distribution.

Other transformations can be made along similar lines. Let us denote

$$w_1 = \sum_{j=1}^{m} \sigma_j^2, w_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}, w_3 = \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}, \ldots, w_m = \frac{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_{m-1}^2}{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_{m-1}^2 + \sigma_m^2} \tag{3.15}$$

Then $(w_1, \ldots, w_m)$ can be sampled similarly as before

$$w_1 \,|\, w_2, \ldots, w_m, \text{data} \sim \frac{2\,\chi^2_{nm-p+\delta-1}}{Q} \tag{3.16}$$

$$\pi(w_2, \ldots, w_m \,|\, \text{data}) \leq \frac{\prod_{j=2}^{m} \text{B}\left(\frac{(2n-p+\alpha-1)(j-1)}{2}, \frac{2n-p+\alpha-1}{2}\right)}{(\text{RSS}^*_{\min})^{\frac{nm-p+\delta-1}{2}}} \left(\prod_{j=2}^{m} g_j(w_j)\right) \tag{3.17}$$

where $\text{B}(a, b)$ is the regular beta function, and $g_j(w_j)$ is the pdf of a $\text{Beta}\left(\frac{(2n-p+\alpha-1)(j-1)}{2}, \frac{2n-p+\alpha-1}{2}\right)$ distribution, independently for $j = 2, \ldots, m$. One advantage of this transformation is that the proposal distribution uses all independent random variables. Immediately we can observe that, due to the fact if $X \sim \text{Beta}(a, b)$

then $1 - X \sim \text{Beta}(b, a)$, the following transformation would also work

$$w_1^* = \sum_{j=1}^{m} \sigma_j^2, w_2^* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}, w_3^* = \frac{\sigma_3^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}, \ldots, w_m^* = \frac{\sigma_m^2}{\sum_{j=1}^{m} \sigma_j^2} \tag{3.18}$$

with the only change being the parameters would be switched in the Beta function and the parameters of the independent Beta distributions in (3.17).

The most intuitive transformation given the problem structure seems to be

$$v_1 = \sigma_1^2, v_2 = \frac{\sigma_2^2}{\sigma_1^2}, \ldots, v_m = \frac{\sigma_m^2}{\sigma_1^2} \tag{3.19}$$

so that $(v_1, v_2, \ldots, v_m)$ can be sampled as

$$v_1 \,|\, v_2, \ldots, v_m, \text{data} \sim \frac{2\,\chi^2_{nm-p+\delta-1}}{Q} \tag{3.20}$$

$$\pi(v_2, \ldots, v_m \,|\, \text{data}) \leq \frac{\mathbf{B}\left(\frac{2n-p+\alpha-1}{2}, \ldots, \frac{2n-p+\alpha-1}{2}\right)}{(\text{RSS}^*_{\text{min}})^{\frac{nm-p+\delta-1}{2}}}\, g_{\text{InvDir}}(v_2, \ldots, v_m) \tag{3.21}$$

where $g_{\text{InvDir}}(v_2, \ldots, v_m)$ is the pdf of an m^{th} order Inverse-Dirichlet $\left(\frac{2n-p+\alpha-1}{2}, \ldots, \frac{2n-p+\alpha-1}{2}\right)$ distribution.

Lastly, the following transformation can also be made

$$y_1 = \sigma_1^2, y_2 = \frac{\sigma_1^2}{\sigma_2^2}, y_3 = \frac{\sigma_1^2 + \sigma_2^2}{\sigma_3^2}, \ldots, y_m = \frac{\sigma_1^2 + \cdots + \sigma_{m-1}^2}{\sigma_m^2} \tag{3.22}$$

so that $(y_1, v_2, \ldots, y_m)$ can be sampled as

$$y_1 \,|\, y_2, \ldots, y_m, \text{data} \sim \frac{2\,\chi^2_{nm-p+\delta-1}}{Q} \tag{3.23}$$

$$\pi(y_2, \ldots, y_m \,|\, \text{data}) \leq \frac{\prod_{j=2}^{m} \text{B}\left(\frac{(2n-p+\alpha-1)(j-1)}{2}, \frac{2n-p+\alpha-1}{2}\right)}{(\text{RSS}^*_{\text{min}})^{\frac{nm-p+\delta-1}{2}}} \left(\prod_{j=2}^{m} g_j'(y_j)\right) \tag{3.24}$$

where $g_j'(y_j)$ is the pdf of a Beta-Prime $\left(\frac{(2n-p+\alpha-1)(j-1)}{2}, \frac{2n-p+\alpha-1}{2}\right)$ distribution, independently for $j = 2, \ldots, m$. Since if $X \sim \beta'(a, b)$ then $X^{-1} \sim \beta'(b, a)$, the reciprocal transformation would also work in (3.24).

Procedure II

We can adapt the approach in Section 2.2 to get

$$\boldsymbol{\beta} \,|\, \sigma^2, \psi_1, \ldots, \psi_m, \text{data} \sim N_p \left[\left(\sum_{j=1}^{m} \frac{\boldsymbol{X'X}}{1+2\psi_j} \right)^{-1} \left(\sum_{j=1}^{m} \frac{\boldsymbol{X'X}}{1+2\psi_j} \boldsymbol{b}_j^* \right), \sigma^2 \left(\sum_{j=1}^{m} \frac{\boldsymbol{X'X}}{1+2\psi_j} \right)^{-1} \right] \tag{3.25}$$

$$\begin{aligned} \sigma^2 \,|\, \psi_1, \ldots, \psi_m, \text{data} \sim \text{Scale-inv-}\chi^2 \Bigg[& nm - p + \delta - 1, \frac{1}{nm - p + \delta - 1} \Bigg(\sum_{j=1}^{m} \frac{\text{RSS}_j^*}{\psi_j} + \\ & \sum_{j=1}^{m} \left(\boldsymbol{b}_j^* - \left(\sum_{j=1}^{m} (1+2\psi_j)^{-1} \right)^{-1} \left(\sum_{j=1}^{m} \frac{\boldsymbol{b}_j^*}{1+2\psi_j} \right) \right)' \frac{\boldsymbol{X'X}}{1+2\psi_j} \\ & \left(\boldsymbol{b}_j^* - \left(\sum_{j=1}^{m} (1+2\psi_j)^{-1} \right)^{-1} \left(\sum_{j=1}^{m} \frac{\boldsymbol{b}_j^*}{1+2\psi_j} \right) \right) \Bigg) \Bigg] \end{aligned} \tag{3.26}$$

$$\begin{aligned} \pi(\psi_1, \ldots, \psi_m \,|\, \text{data}) \propto & |\boldsymbol{X'X}|^{-\frac{1}{2}} \left(\sum_{j=1}^{m} \frac{1}{1+2\psi_j} \right)^{-\frac{p}{2}} \left(\prod_{j=1}^{m} \psi_j^{-1} (1+\psi)^{-\frac{2n-2p+\alpha-1}{2}} (1+2\psi_j)^{-\frac{p}{2}} \right) \\ \times & \left\{ \sum_{j=1}^{m} \left(\boldsymbol{b}_j^* - \left(\sum_{j=1}^{m} (1+2\psi_j)^{-1} \right)^{-1} \sum_{j=1}^{m} \frac{\boldsymbol{b}_j^*}{1+2\psi_j} \right)' \frac{\boldsymbol{X'X}}{1+2\psi_j} \right. \\ & \left. \left(\boldsymbol{b}_j^* - \left(\sum_{j=1}^{m} (1+2\psi_j)^{-1} \right)^{-1} \sum_{j=1}^{m} \frac{\boldsymbol{b}_j^*}{1+2\psi_j} \right) + \sum_{j=1}^{m} \frac{\text{RSS}_j^*}{\psi_j} \right\}^{-\frac{nm-p+\delta-1}{2}} \end{aligned} \tag{3.27}$$

Now since

$$\sum_{j=1}^{m} \frac{1}{1+2\psi_j} \geq \prod_{j=1}^{m} \frac{1}{1+2\psi_j} \quad ; \quad \{\cdot\} \geq \sum_{j=1}^{m} \frac{\text{RSS}_j^*}{\psi_j} \geq m \left(\prod_{j=1}^{m} \frac{\text{RSS}_j^*}{\psi_j} \right)^{\frac{1}{m}} \tag{3.28}$$

We could have also done $\sum_{j=1}^{m} \frac{\text{RSS}_j^*}{\psi_j} \geq \sum_{j=1}^{m} \frac{\text{RSS}_{\min}^*}{\psi_j} \geq m\text{RSS}_{\min}^* \prod_{j=1}^{m} \left(\frac{1}{\psi_j}\right)^{\frac{1}{m}}$ instead. Now we can use Accept-Reject algorithm to sample the latent variables as follows

$$\pi(\psi_1, \ldots, \psi_m \,|\, \text{data}) \leq \frac{\left(\prod_{j=1}^{m} \text{B}\left(\frac{nm-p+\delta-1}{2m}, \frac{nm-(2m-1)p+\alpha m-\delta-m+1}{2}\right)\right)}{|\boldsymbol{X}'\boldsymbol{X}|^{\frac{1}{2}} \left(m \prod_{j=1}^{m} \text{RSS}_j^{*\frac{1}{m}}\right)^{\frac{nm-p+\delta-1}{2}}} \left(\prod_{j=1}^{m} \tilde{g_j}'(\psi_j)\right)$$

where $\tilde{g_j}'(\psi_j)$ is the pdf of a Beta-Prime $\left(\frac{nm-p+\delta-1}{2m}, \frac{nm-(2m-1)p+\alpha m-\delta-m+1}{2}\right)$ distribution, independently for $j = 1, \ldots, m$. An advantage of this method is that the latent variables need not be transformed before applying Accept-Reject algorithm.

Procedure III

In an attempt to reduce the complexity of the posterior distributions, Moura et al. (2017a) propose drawing only a single posterior draw of the parameter, using which we can then generate m replicates of the original data. This method is known as Fixed-Posterior Predictive Sampling (FPPS). So we draw $(\boldsymbol{\beta}^*, \sigma^*)$ from (2.11) and (2.12), so that our released data is $\boldsymbol{z}_j = (z_{j1}, \ldots, z_{jn})' \overset{\text{iid}}{\sim} N_n(\boldsymbol{X}\boldsymbol{\beta}^*, \sigma^{*2}\boldsymbol{I}_n)$ for each $j = 1, \ldots, m$. Using the same notation as in Section 3.1, we can derive the posterior distribution of the parameters as follows with exactly the same conditions for existence as in Section 2.2

$$\boldsymbol{\beta} \,|\, \overline{\boldsymbol{b}^*}, \sigma^2, t \;\sim\; N_p\left(\overline{\boldsymbol{b}^*}, \left(\sigma^2 + \frac{1+\frac{1}{m}}{t}\right)(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) \tag{3.29}$$

$$t\sigma^2 \;\sim\; \beta'\left(\frac{n-p+\alpha-\delta}{2}, \frac{n-p+\delta-1}{2}\right) \tag{3.30}$$

$$tS_{\text{comb}}^2 \;\sim\; \chi^2_{nm-p+\delta-1} \tag{3.31}$$

As we can see, we avoid the mess of complicated distribution of the latent variables, and the independence of $t\sigma^2$ and tS_{comb}^2 makes it easy to construct an pivot for σ^2.

The posterior distributions can be parametrized in a slightly different manner as follows, again with the same conditions for existence as in Section 2.2

$$\begin{aligned}
\boldsymbol{\beta} \,|\, \overline{\boldsymbol{b}^*}, \sigma^2, \psi &\sim N_p\left(\overline{\boldsymbol{b}^*}, \sigma^2\left(1+\left(1+\frac{1}{m}\right)\psi\right)(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) && (3.32)\\
\sigma^2 \,|\, {}^2_{\text{comb}}, \psi &\sim \text{Scale-inv-}\chi^2\left(nm-p+\delta-1, \frac{S^2_{\text{comb}}}{\psi(nm-p+\delta-1)}\right) && (3.33)\\
\psi &\sim \beta'\left(\frac{n-p+\delta-1}{2}, \frac{n-p+\alpha-\delta}{2}\right) && (3.34)
\end{aligned}$$

3.3 Partially Sensitive Data

Method I: Using only estimates of sensitive part to impute synthetic data

Plug-In Sampling

The original data now has the same setup as in Section 2.4 with both assumptions $r > p$, $n - r > p$ in effect. We synthesize m copies of the original data $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$ given by $\{\boldsymbol{y}^{*j} = (\boldsymbol{y}_1^{*j}, \boldsymbol{y}_2) : j = 1, \ldots, m\}$ whose sufficient statistics are given by $(\boldsymbol{b}_1^{*1}, \ldots, \boldsymbol{b}_1^{*m}, \text{RSS}_1^{*1}, \ldots, \text{RSS}_1^{*m}, \boldsymbol{b}_2, \text{RSS}_2)$. We denote $\overline{\boldsymbol{b}_1^*} = \frac{1}{m}\sum_{j=1}^m \boldsymbol{b}_1^{*j}$. Then we can derive the following posterior distributions in a similar manner as before

$$\boldsymbol{\beta} \,|\, \sigma^2, \psi, \overline{\boldsymbol{b}_1^*}, \boldsymbol{b}_2 \sim N_p\left[\left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi}+\frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{m}\right)^{-1}\left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi}\overline{\boldsymbol{b}_1^*}+\frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{m}\boldsymbol{b}_2\right), \frac{\sigma^2}{m}\left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi}+\frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{m}\right)^{-1}\right]$$

$$\sigma^2 \,|\, \psi, \boldsymbol{b}_1^{*1}, \ldots, \boldsymbol{b}_1^{*m}, \text{RSS}_1^{*1}, \ldots, \text{RSS}_1^{*m}, \boldsymbol{b}_2, \text{RSS}_2 \sim \text{Scale-inv-}\chi^2\left(\nu, \tau_1^2\right)$$

$$\pi(\psi \,|\, \boldsymbol{b}_1^{*1}, \ldots, \boldsymbol{b}_1^{*m}, \text{RSS}_1^{*1}, \ldots, \text{RSS}_1^{*m}, \boldsymbol{b}_2, \text{RSS}_2)$$

$$\propto \left| \frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi} + \frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{m} \right|^{-\frac{1}{2}} \psi^{-\frac{(m-1)(r-p)}{2}-1} (1+\psi)^{-\frac{mp}{2}} e^{\frac{(r-p)\psi}{2}} \left\{\nu\tau_1^2\right\}^{-\frac{\nu}{2}}$$

where $\nu = n + (m-1)r - p + \delta - 1$ and

$$\nu\tau_1^2 = \sum_{j=1}^{m} \left(\boldsymbol{b}_1^{*j} - \overline{\boldsymbol{b}_1^*}\right)' \frac{\boldsymbol{X}_1{}'\boldsymbol{X}_1}{1+\psi} \left(\boldsymbol{b}_1^{*j} - \overline{\boldsymbol{b}_1^*}\right)$$
$$+ m\left(\overline{\boldsymbol{b}_1^*} - \boldsymbol{b}_2\right)' \left((1+\psi)\left(\boldsymbol{X}_1{}'\boldsymbol{X}_1\right)^{-1} + \left(\boldsymbol{X}_2{}'\boldsymbol{X}_2\right)^{-1}\right)^{-1} \left(\boldsymbol{b}_1^{*j} - \overline{\boldsymbol{b}_1^*}\right) + \sum_{j=1}^{m} \frac{\mathrm{RSS}_1^{*j}}{\psi} + \mathrm{RSS}_2$$

The posterior distributions are proper as long as $r > p$, $n - r > \max\{p, p - rm - \delta + 1\}$.

Let us denote $\boldsymbol{X}'\boldsymbol{X}_{\psi,m} = \frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{1+\psi} + \frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{m}$. Now since

$$\boldsymbol{X}'\boldsymbol{X}_{\psi,m} \geq \frac{\boldsymbol{X}'\boldsymbol{X}}{m(1+\psi)} \implies \left|\boldsymbol{X}'\boldsymbol{X}_{\psi,m}\right|^{-\frac{1}{2}} \leq m^{\frac{p}{2}}(1+\psi)^{\frac{p}{2}} \left|\boldsymbol{X}'\boldsymbol{X}\right|^{-\frac{1}{2}}$$
$$\nu\tau_1^2 \geq \sum_{j=1}^{m} \frac{\mathrm{RSS}_1^{*j}}{\psi}$$

and using the fact that $(1+\psi)^{-\frac{(m-1)p}{2}} \leq 1$ (as $\psi > 0$), we can sample from the distribution of latent variables using the Accept-Reject algorithm as follows:

$$\pi(\psi \,|\, \text{data}) \leq \frac{m^{\frac{p}{2}} \left|\boldsymbol{X}'\boldsymbol{X}\right|^{-\frac{1}{2}} 2^{\frac{n+(m-2)p+\delta-1}{2}} \Gamma\left(\frac{n+(m-2)p+\delta-1}{2}\right)}{(r-p)^{\frac{n+(m-2)p+\delta-1}{2}} \left(\sum_{j=1}^{m} \mathrm{RSS}_1^{*j}\right)^{\frac{n+(m-1)r-p+\delta-1}{2}}} \tilde{f}_{\text{ScaledChi}}(\psi)$$

where $\tilde{f}_{\text{ScaledChi}}(\psi)$ is the pdf of a Gamma$\left(\frac{n+(m-2)p+\delta-1}{2}, \frac{r-p}{2}\right)$ distribution, which interestingly is the same proposal distribution we used in Section 3.1 when $r = n$. We would need another assumption $n + (m-2)p + \delta - 1 > 0$. All the expressions coincide with our results earlier when all of $\boldsymbol{y}$ is sensitive.

Posterior Predictive Sampling

We follow the same process as in Section 2.4 for the PPS case to derive the following posterior distributions

$$\begin{aligned}
&\boldsymbol{\beta} \,|\, \sigma^2, \psi_1, \ldots, \psi_m, \boldsymbol{b}_1^{*1}, \ldots, \boldsymbol{b}_1^{*m}, \boldsymbol{b}_2 \\
&\sim N_p \left[\left(\sum_{j=1}^{m} \frac{\boldsymbol{X}_1' \boldsymbol{X}_1}{1+2\psi_j} + \boldsymbol{X}_2' \boldsymbol{X}_2 \right)^{-1} \left(\sum_{j=1}^{m} \frac{\boldsymbol{X}_1' \boldsymbol{X}_1}{1+2\psi_j} \boldsymbol{b}_1^{*j} + \boldsymbol{X}_2' \boldsymbol{X}_2 \boldsymbol{b}_2 \right), \sigma^2 \left(\sum_{j=1}^{m} \frac{\boldsymbol{X}_1' \boldsymbol{X}_1}{1+2\psi_j} + \boldsymbol{X}_2' \boldsymbol{X}_2 \right)^{-1} \right]
\end{aligned}$$

$$\sigma^2 \,|\, \psi_1, \ldots, \psi_m, \boldsymbol{b}_1^{*1}, \ldots, \boldsymbol{b}_1^{*m}, \mathrm{RSS}_1^{*1}, \ldots, \mathrm{RSS}_1^{*m}, \boldsymbol{b}_2, \mathrm{RSS}_2 \sim \text{Scale-inv-}\chi^2 \left(\nu, \tau_1^2 \right)$$

$$\begin{aligned}
&\pi(\psi_1, \ldots, \psi_m \,|\, \boldsymbol{b}_1^{*1}, \ldots, \boldsymbol{b}_1^{*m}, \mathrm{RSS}_1^{*1}, \ldots, \mathrm{RSS}_1^{*m}, \boldsymbol{b}_2, \mathrm{RSS}_2) \\
&\propto \left| \sum_{j=1}^{m} \frac{\boldsymbol{X}_1' \boldsymbol{X}_1}{1+2\psi_j} + \boldsymbol{X}_2' \boldsymbol{X}_2 \right|^{-\frac{1}{2}} \left(\prod_{j=1}^{m} \psi_j^{-1} (1+2\psi_j)^{-\frac{p}{2}} (1+\psi_j)^{-\frac{2r-2p+\alpha-1}{2}} \right) \left\{ \nu \tau_2^2 \right\}^{-\frac{\nu}{2}}
\end{aligned}$$

where $\nu = n + (m-1)r - p + \delta - 1$ as before and

$$\begin{aligned}
&\nu \tau_2^2 \\
&= \sum_{j=1}^{m} \left(\boldsymbol{b}_1^{*j} - \left(\sum_{j=1}^{m} \frac{1}{1+2\psi_j} \right)^{-1} \left(\sum_{j=1}^{m} \frac{\boldsymbol{b}_1^{*j}}{1+2\psi_j} \right) \right)' \frac{\boldsymbol{X}_1{}' \boldsymbol{X}_1}{1+2\psi_j} \left(\boldsymbol{b}_1^{*j} - \left(\sum_{j=1}^{m} \frac{1}{1+2\psi_j} \right)^{-1} \left(\sum_{j=1}^{m} \frac{\boldsymbol{b}_1^{*j}}{1+2\psi_j} \right) \right) \\
&+ \left(\left(\sum_{j=1}^{m} \frac{1}{1+2\psi_j} \right)^{-1} \left(\sum_{j=1}^{m} \frac{\boldsymbol{b}_1^{*j}}{1+2\psi_j} \right) - \boldsymbol{b}_2 \right)' \left(\left(\sum_{j=1}^{m} \frac{1}{1+2\psi_j} \right)^{-1} \left(\boldsymbol{X}_1{}' \boldsymbol{X}_1 \right)^{-1} + \left(\boldsymbol{X}_2{}' \boldsymbol{X}_2 \right)^{-1} \right)^{-1} \\
&\left(\left(\sum_{j=1}^{m} \frac{1}{1+2\psi_j} \right)^{-1} \left(\sum_{j=1}^{m} \frac{\boldsymbol{b}_1^{*j}}{1+2\psi_j} \right) - \boldsymbol{b}_2 \right) + \sum_{j=1}^{m} \frac{\mathrm{RSS}_1^{*j}}{\psi_j} + \mathrm{RSS}_2
\end{aligned}$$

The posterior distributions are proper as long as $r > \max\left\{ p, p - \alpha + 1, \frac{n+(2m-1)p-\alpha m+\delta+m-1}{m+1} \right\}$, $n - r > \max\{p, p - rm - \delta + 1\}$.

Now using (3.28) we can write

$$\left(\sum_{j=1}^{m}\frac{1}{1+2\psi_j}\right)\boldsymbol{X}_1'\boldsymbol{X}_1+\boldsymbol{X}_2'\boldsymbol{X}_2 \geq \left(\prod_{j=1}^{m}\frac{1}{1+2\psi_j}\right)\boldsymbol{X}_1'\boldsymbol{X}_1+\boldsymbol{X}_2'\boldsymbol{X}_2 \geq \left(\prod_{j=1}^{m}\frac{1}{1+2\psi_j}\right)\boldsymbol{X}'\boldsymbol{X} \tag{3.35}$$

Finally using (3.28) and (3.35) we can use Accept-Reject algorithm to sample the latent variables as follows

$$\pi(\psi_1,\ldots,\psi_m\,|\,\text{data}) \leq \frac{\left(\prod_{j=1}^{m}\mathrm{B}\left(\frac{n+(m-1)r-p+\delta-1}{2m},\frac{(m+1)r-n-(2m-1)p+m\alpha-\delta-m+1}{2m}\right)\right)}{|\boldsymbol{X}'\boldsymbol{X}|^{\frac{1}{2}}\left(m\prod_{j=1}^{m}\mathrm{RSS}_j^{*\frac{1}{m}}\right)^{\frac{n+(m-1)r-p+\delta-1}{2}}}\left(\prod_{j=1}^{m}\overline{g_j}'(\psi_j)\right)$$

where $\overline{g_j}'(\psi_j)$ is the pdf of a Beta-Prime $\left(\frac{n+(m-1)r-p+\delta-1}{2m},\frac{(m+1)r-n-(2m-1)p+m\alpha-\delta-m+1}{2m}\right)$ distribution, independently for $j=1,\ldots,m$.

Method II: Using whole data estimates to impute synthetic data

Plug-In Sampling

As in Section 2.4, our analysis in this case will be based solely on the synthetic part. We require only $r>p$. The posterior distributions are given by

$$\begin{aligned}
\boldsymbol{\beta}\,|\,\sigma^2,\psi,\overline{\boldsymbol{b}_1^*} &\sim N_p\left(\overline{\boldsymbol{b}_1^*},\sigma^2\left((\boldsymbol{X}'\boldsymbol{X})^{-1}+\frac{\psi}{m}(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\right)\right)\\
\sigma^2\,|\,\psi,\boldsymbol{b}_1^{*1},\ldots,\boldsymbol{b}_1^{*m},\mathrm{RSS}_1^{*1},\ldots,\mathrm{RSS}_1^{*m} &\sim \text{Scale-inv-}\chi^2\left(\tilde{\nu},\tilde{\tau}_1^2\right)\\
\psi &\sim \frac{\chi^2_{n-p+\delta-1}}{n-p}\equiv\Gamma\left(\frac{n-p+\delta-1}{2},\frac{n-p}{2}\right)
\end{aligned}$$

where $\tilde{\nu}=rm-p+\delta-1$, $\tilde{\nu}\tilde{\tau_1}^2=\frac{1}{\psi}\left(\sum_{j=1}^{m}\mathrm{RSS}_1^{*j}+\sum_{j=1}^{m}\left(\boldsymbol{b}_1^{*j}-\overline{\boldsymbol{b}_1^*}\right)'\left(\boldsymbol{X}_1{}'\boldsymbol{X}_1\right)\left(\boldsymbol{b}_1^{*j}-\overline{\boldsymbol{b}_1^*}\right)\right)$. Interestingly, the last quantity is a sum of variation within samples and variation between samples. The posterior distributions are proper as long as $r>\max\left\{p,\frac{p-\delta+1}{m}\right\}$, $n>p-\delta+1$ and the results match our expressions from Section 3.1 when $r=n$.

Posterior Predictive Sampling

$$\boldsymbol{\beta} \,|\, \sigma^2, \psi_1, \ldots, \psi_m, \boldsymbol{b}_1^{*1}, \ldots, \boldsymbol{b}_1^{*m}$$
$$\sim N_p \left[\left(\sum_{j=1}^{m} \psi_j^{-1} \right)^{-1} \left(\sum_{j=1}^{m} \frac{\boldsymbol{b}_1^{*j}}{\psi_j} \right), \sigma^2 \left(\left(\sum_{j=1}^{m} \psi_j^{-1} \right)^{-1} \left((\boldsymbol{X}_1' \boldsymbol{X}_1)^{-1} + (\boldsymbol{X}' \boldsymbol{X})^{-1} \right) + (\boldsymbol{X}' \boldsymbol{X})^{-1} \right) \right]$$

$$\sigma^2 \,|\, \psi_1, \ldots, \psi_m, \boldsymbol{b}_1^{*1}, \ldots, \boldsymbol{b}_1^{*m}, \mathrm{RSS}_1^{*1}, \ldots, \mathrm{RSS}_1^{*m} \sim \text{Scale-inv-}\chi^2 \left(\nu, \tilde{\tau_2}^2 \right)$$

$$\pi(\psi_1, \ldots, \psi_m \,|\, \boldsymbol{b}_1^{*1}, \ldots, \boldsymbol{b}_1^{*m}, \mathrm{RSS}_1^{*1}, \ldots, \mathrm{RSS}_1^{*m})$$
$$\propto \left(\prod_{j=1}^{m} \psi_j \right)^{-\frac{n-p+r+\alpha+1}{2}} \left(\sum_{j=1}^{m} \psi_j^{-1} \right)^{-\frac{p}{2}} \left(1 + \left(\sum_{j=1}^{m} \psi_j^{-1} \right) \right)^{-\frac{2n-2p+\alpha-1}{2}} \left\{ \nu \tilde{\tau_2}^2 \right\}^{-\frac{\nu}{2}}$$

where $\nu = m(r-2) - p + \delta + 1$ as before and

$$\nu \tilde{\tau_2}^2 = \sum_{j=1}^{m} \left(\boldsymbol{b}_1^{*j} - \left(\sum_{j=1}^{m} \psi_j^{-1} \right)^{-1} \left(\sum_{j=1}^{m} \frac{\boldsymbol{b}_1^{*j}}{\psi_j} \right)^{-1} \right)' \frac{\left((\boldsymbol{X}_1' \boldsymbol{X}_1)^{-1} + (\boldsymbol{X}' \boldsymbol{X})^{-1} \right)^{-1}}{\psi_j}$$
$$\left(\boldsymbol{b}_1^{*j} - \left(\sum_{j=1}^{m} \psi_j^{-1} \right)^{-1} \left(\sum_{j=1}^{m} \frac{\boldsymbol{b}_1^{*j}}{\psi_j} \right)^{-1} \right) + \sum_{j=1}^{m} \frac{\mathrm{RSS}_1^{*j}}{\psi_j}$$

The latent variables can be sampled using Accept-Reject algorithm similarly as before, using an Inverse-Dirichlet distribution as proposal distribution.

Chapter 4

Bayesian Analysis of Singly Imputed Synthetic Data under the Multivariate Normal Model

In this chapter we present the Bayesian approach for analysis of singly imputed synthetic data generated from a MVN population with both mean vector and covariance matrix unknown. Assume the original confidential data are

$$\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \overset{\text{iid}}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{4.1}$$

where $n > p$, and define $\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i$ (sample mean) and $\boldsymbol{W} = \boldsymbol{\mathcal{S}}_x/(n-1)$ (sample variance) where $\boldsymbol{\mathcal{S}}_x = \sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$ is the sample Wishart matrix, to be the unbiased estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ respectively. We know that $\bar{\boldsymbol{x}} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/n)$, $\boldsymbol{\mathcal{S}}_x \sim \mathcal{W}_p(\boldsymbol{\Sigma}, n-1)$, $\bar{\boldsymbol{x}}$ is independent of $\boldsymbol{\mathcal{S}}_x$ and $(\bar{\boldsymbol{x}}, \boldsymbol{W})$ are jointly sufficient for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ when the original data are observed.

4.1 Plug In Sampling method

The singly imputed synthetic data, denoted by $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$, are obtained by drawing

$$\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) \,|\, \boldsymbol{X} \overset{\text{iid}}{\sim} N_p(\bar{\boldsymbol{x}}, \boldsymbol{W}) \tag{4.2}$$

Define $\bar{\boldsymbol{y}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_i$ (sample mean based on $\boldsymbol{Y}$) and $\boldsymbol{\mathcal{S}}_y = \sum_{i=1}^{n}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})'$ (sample Wishart matrix based on $\boldsymbol{Y}$). Clearly $\bar{\boldsymbol{y}} \sim N_p(\bar{\boldsymbol{x}}, n^{-1}\boldsymbol{W})$, $\boldsymbol{\mathcal{S}}_y \sim \mathcal{W}_p(\boldsymbol{W}, n-1)$. It follows from Lemma 1.2.1 that $(\bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y)$ are jointly sufficient for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Also $\bar{\boldsymbol{y}}$ is independent $\boldsymbol{\mathcal{S}}_y$, because $\bar{\boldsymbol{y}}$ is independent of $\boldsymbol{y}_i - \bar{\boldsymbol{y}}$ as $\text{Cov}(\bar{\boldsymbol{y}}, \boldsymbol{y}_i - \bar{\boldsymbol{y}}) = 0 \; \forall \, i = 1, \ldots, n$. The following discussion follows from the elucidation in Klein and Sinha (2015b).

Likelihood of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

The conditional joint pdf of $(\bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y)$, given $(\bar{\boldsymbol{x}}, \boldsymbol{W})$, is given by

$$\begin{aligned}
& f(\bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y \,|\, \bar{\boldsymbol{x}}, \boldsymbol{W}) \\
=& f(\bar{\boldsymbol{y}} \,|\, \bar{\boldsymbol{x}}, \boldsymbol{W}) f(\boldsymbol{\mathcal{S}}_y \,|\, \boldsymbol{W}) \\
\propto & \, |\boldsymbol{W}|^{-1/2} \exp\left[-\frac{n}{2}(\bar{\boldsymbol{y}} - \bar{\boldsymbol{x}})'\boldsymbol{W}^{-1}(\bar{\boldsymbol{y}} - \bar{\boldsymbol{x}})\right] \times \frac{|\boldsymbol{\mathcal{S}}_y|^{(n-p-2)/2}}{|\boldsymbol{W}|^{n-1/2}} \exp\left[-\frac{1}{2}\,\text{tr}(\boldsymbol{\mathcal{S}}_y \boldsymbol{W}^{-1})\right] \\
=& \frac{|\boldsymbol{\mathcal{S}}_y|^{(n-p-2)/2}}{|\boldsymbol{W}|^{n/2}} \exp\left[-\frac{n}{2}(\bar{\boldsymbol{y}} - \bar{\boldsymbol{x}})'\boldsymbol{W}^{-1}(\bar{\boldsymbol{y}} - \bar{\boldsymbol{x}}) - \frac{1}{2}\,\text{tr}(\boldsymbol{\mathcal{S}}_y \boldsymbol{W}^{-1})\right]
\end{aligned} \tag{4.3}$$

A similar calculation yields the joint pdf of $(\bar{\boldsymbol{x}}, \boldsymbol{W})$ as

$$f(\bar{\boldsymbol{x}}, \boldsymbol{W} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \frac{|\boldsymbol{W}|^{(n-p-2)/2}}{|\boldsymbol{\Sigma}|^{n/2}} \exp\left[-\frac{n}{2}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) - \frac{n-1}{2}\,\text{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^{-1})\right] \tag{4.4}$$

We now combine the terms involving $\bar{\boldsymbol{x}}$ from the two exponents as

$$
\begin{aligned}
&(\bar{\boldsymbol{y}}-\bar{\boldsymbol{x}})'\boldsymbol{W}^{-1}(\bar{\boldsymbol{y}}-\bar{\boldsymbol{x}})+(\bar{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}) \\
=&\left\{\bar{\boldsymbol{x}}-[\boldsymbol{W}^{-1}+\boldsymbol{\Sigma}^{-1}]^{-1}[\boldsymbol{W}^{-1}\bar{\boldsymbol{y}}+\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}]\right\}'\left\{\boldsymbol{W}^{-1}+\boldsymbol{\Sigma}^{-1}\right\} \\
&\left\{\bar{\boldsymbol{x}}-[\boldsymbol{W}^{-1}+\boldsymbol{\Sigma}^{-1}]^{-1}[\boldsymbol{W}^{-1}\bar{\boldsymbol{y}}+\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}]\right\} \\
-&\left\{\boldsymbol{W}^{-1}\bar{\boldsymbol{y}}+\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right\}'\left\{\boldsymbol{W}^{-1}+\boldsymbol{\Sigma}^{-1}\right\}^{-1}\left\{\boldsymbol{W}^{-1}\bar{\boldsymbol{y}}+\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right\}+\bar{\boldsymbol{y}}'\boldsymbol{W}^{-1}\bar{\boldsymbol{y}}+\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \\
=&\left\{\bar{\boldsymbol{x}}-[\boldsymbol{W}^{-1}+\boldsymbol{\Sigma}^{-1}]^{-1}[\boldsymbol{W}^{-1}\bar{\boldsymbol{y}}+\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}]\right\}'\left\{\boldsymbol{W}^{-1}+\boldsymbol{\Sigma}^{-1}\right\} \\
&\left\{\bar{\boldsymbol{x}}-[\boldsymbol{W}^{-1}+\boldsymbol{\Sigma}^{-1}]^{-1}[\boldsymbol{W}^{-1}\bar{\boldsymbol{y}}+\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}]\right\}+(\bar{\boldsymbol{y}}-\boldsymbol{\mu})'\left(\boldsymbol{\Sigma}+\boldsymbol{W}\right)^{-1}(\bar{\boldsymbol{y}}-\boldsymbol{\mu})
\end{aligned}
$$

where the simplification is due to (2.28). Now integrating out $\bar{\boldsymbol{x}}$ from the product of the above two pdfs, we arrive at the following result.

Theorem 4.1.1. *The joint pdf of* $(\bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y)$ *is given by*

$$
f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}\left(\bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y\right) \propto \int_{S_n^{++}} \frac{\left|\boldsymbol{\mathcal{S}}_y\right|^{\frac{n-p-2}{2}}\left|\boldsymbol{\Sigma}+\boldsymbol{W}\right|^{-\frac{1}{2}}}{\left|\boldsymbol{\Sigma}\right|^{\frac{n-1}{2}}\left|\boldsymbol{W}\right|^{\frac{p+1}{2}}} e^{-\frac{1}{2}\left[n(\bar{\boldsymbol{y}}-\boldsymbol{\mu})'(\boldsymbol{\Sigma}+\boldsymbol{W})^{-1}(\bar{\boldsymbol{y}}-\boldsymbol{\mu})+\operatorname{tr}(\boldsymbol{\mathcal{S}}_y\boldsymbol{W}^{-1})+(n-1)\operatorname{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^{-1})\right]} d\boldsymbol{W}
$$

Posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

We choose the non-informative joint prior: $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{\delta}{2}}$ on the parameters. The posterior distribution can be computed by multiplying the expression inside the above integral with the prior and the product splits up into exactly three parts corresponding to the three conditional posterior distributions.

$$
\begin{aligned}
&\pi(\bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y, \boldsymbol{W} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
\propto&\ \left(\left|\boldsymbol{\Sigma}+\boldsymbol{W}\right|^{-\frac{1}{2}} e^{-\frac{1}{2}\left[n(\bar{\boldsymbol{y}}-\boldsymbol{\mu})'(\boldsymbol{\Sigma}+\boldsymbol{W})^{-1}(\bar{\boldsymbol{y}}-\boldsymbol{\mu})\right]}\right)\left(\frac{\left|\boldsymbol{W}\right|^{\frac{n-p+\delta-2}{2}}}{\left|\boldsymbol{\Sigma}\right|^{\frac{(n-p+\delta-2)+p+1}{2}}} e^{-\frac{1}{2}\left[\operatorname{tr}((n-1)\boldsymbol{W}\boldsymbol{\Sigma}^{-1})\right]}\right)
\end{aligned}
$$

$$\left(\frac{|\boldsymbol{\mathcal{S}}_y|^{\frac{n-p+\delta-2}{2}}}{|\boldsymbol{W}|^{\frac{(n-p+\delta-2)+p+1}{2}}} e^{-\frac{1}{2}\left[\mathrm{tr}(\boldsymbol{\mathcal{S}}_y \boldsymbol{W}^{-1})\right]} \right)$$

which concedes that the posterior sampling will be done sequentially in the following manner:

$$\begin{aligned} \boldsymbol{W} \,|\, \boldsymbol{\mathcal{S}}_y &\sim \mathcal{W}_p^{-1}\left(\boldsymbol{\mathcal{S}}_y, n-p+\delta-2\right) && (4.5) \\ \boldsymbol{\Sigma} \,|\, \boldsymbol{W} &\sim \mathcal{W}_p^{-1}\left((n-1)\boldsymbol{W}, n-p+\delta-2\right) && (4.6) \\ \boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \boldsymbol{W}, \bar{\boldsymbol{y}} &\sim N_p\left(\bar{\boldsymbol{y}}, \frac{1}{n}\left(\boldsymbol{\Sigma}+\boldsymbol{W}\right)\right) && (4.7) \end{aligned}$$

We can reformulate the above posterior distributions as:

$$\begin{aligned} {\boldsymbol{\mathcal{S}}_y}^{-1/2} \boldsymbol{W} {\boldsymbol{\mathcal{S}}_y}^{-1/2} &\sim \mathcal{W}_p^{-1}\left(\boldsymbol{I}_p, n-p+\delta-2\right) && (4.8) \\ \boldsymbol{W}^{-1/2} \boldsymbol{\Sigma} \boldsymbol{W}^{-1/2} &\sim \mathcal{W}_p^{-1}\left((n-1)\boldsymbol{I}_p, n-p+\delta-2\right) && (4.9) \\ \boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \boldsymbol{W}, \bar{\boldsymbol{y}} &\sim N_p\left(\bar{\boldsymbol{y}}, \frac{1}{n}\left(\boldsymbol{\Sigma}+\boldsymbol{W}\right)\right) && (4.10) \end{aligned}$$

which has the benefit that ${\boldsymbol{\mathcal{S}}_y}^{-1/2} \boldsymbol{W} {\boldsymbol{\mathcal{S}}_y}^{-1/2}$ is independent of $\boldsymbol{W}^{-1/2} \boldsymbol{\Sigma} \boldsymbol{W}^{-1/2}$ and their posterior distributions are unconditional.

The posterior distributions are proper as long as $n > \max\{p, 2p-\delta+1\}$.

Bayes Estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

$$\hat{\boldsymbol{\mu}}_{\mathrm{BAYES}} = \mathrm{E}(\boldsymbol{\mu} \,|\, \bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y) = \mathrm{E}_{\boldsymbol{W}}\, \mathrm{E}_{\boldsymbol{\Sigma}}\, \mathrm{E}(\boldsymbol{\mu} \,|\, \bar{\boldsymbol{y}}, \boldsymbol{\Sigma}, \boldsymbol{W}) = \mathrm{E}_{\boldsymbol{W}}\, \mathrm{E}_{\boldsymbol{\Sigma}}\, \mathrm{E}(\bar{\boldsymbol{y}}) = \bar{\boldsymbol{y}}$$

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{BAYES}} = \mathrm{E}(\boldsymbol{\Sigma} \,|\, \bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y) = \mathrm{E}_{\boldsymbol{W}}\, \mathrm{E}(\boldsymbol{\Sigma} \,|\, \boldsymbol{\mathcal{S}}_y, \boldsymbol{W}) = \mathrm{E}_{\boldsymbol{W}}\left(\tfrac{(n-1)\boldsymbol{W}}{(n-2p+\delta-3)} \,|\, \boldsymbol{\mathcal{S}}_y\right) = \frac{(n-1)\boldsymbol{\mathcal{S}}_y}{(n-2p+\delta-3)^2}$$

$$\widehat{|\boldsymbol{\Sigma}|}_{\mathrm{BAYES}} = \mathrm{E}\left(|\boldsymbol{\Sigma}| \,|\, \bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y\right) = \mathrm{E}_{\boldsymbol{W}}\, \mathrm{E}\left(|\boldsymbol{\Sigma}| \,|\, \boldsymbol{\mathcal{S}}_y, \boldsymbol{W}\right) = \mathrm{E}_{\boldsymbol{W}}\left(|\boldsymbol{W}|\, \mathrm{E}\left(\left|\boldsymbol{W}^{-1/2} \boldsymbol{\Sigma} \boldsymbol{W}^{-1/2}\right|\right) \,|\, \boldsymbol{\mathcal{S}}_y\right)$$

$$= \left(\prod_{j=1}^{p} \frac{n-1}{n-p+\delta-j-3}\right) \mathrm{E}\left(|\boldsymbol{W}| \,|\, \boldsymbol{\mathcal{S}}_y\right) = \left(\prod_{j=1}^{p} \frac{n-1}{(n-p+\delta-j-3)^2}\right) |\boldsymbol{\mathcal{S}}_y|$$

provided $n > \max\{p, 2p - \delta + 5\}$, and we use the results: If $\boldsymbol{S} \sim \mathcal{W}_p^{-1}(\boldsymbol{\Sigma}, \nu)$ then

$$\mathrm{E}\,(\boldsymbol{S}) = (\nu - p - 1)^{-1}\boldsymbol{\Sigma} \quad \text{if } \nu > p + 1 \tag{4.11}$$

$$\mathrm{E}\,(|\boldsymbol{S}|) = |\boldsymbol{\Sigma}| \prod_{j=1}^{p} (\nu - j - 1)^{-1} \quad \text{if } \nu > p + 3 \tag{4.12}$$

Credible Sets for $|\boldsymbol{\Sigma}|$ and $\boldsymbol{\mu}$

We see that $\boldsymbol{\Sigma}^{-1} \,|\, \boldsymbol{W} \sim \mathcal{W}_p\,(\boldsymbol{W}^{-1}/(n-1), n - p + \delta - 2)$, so

$$\frac{|\boldsymbol{\Sigma}^{-1}|}{|\boldsymbol{W}^{-1}/(n-1)|} \sim \prod_{i=1}^{p} u_i, \quad \text{where } u_i \sim \chi^2_{n-p+\delta-i-1} \text{ independently for } i = 1, \ldots, p$$

which also shows that the quantity on the left hand side of the above relation is independent of $\boldsymbol{W}$. Here we use the following result: If $\boldsymbol{S} \sim \mathcal{W}_p(\boldsymbol{\Sigma}, n)$ then

$$\frac{|\boldsymbol{S}|}{|\boldsymbol{\Sigma}|} \sim \prod_{i=1}^{p} t_i, \quad \text{where } t_i \sim \chi^2_{n-i+1} \text{ independently for } i = 1, \ldots, p \tag{4.13}$$

Thus using the above result similarly as before, we can get

$$\frac{|\boldsymbol{W}^{-1}|}{|\boldsymbol{\mathcal{S}}_y^{-1}|} \sim \prod_{j=1}^{p} v_j, \quad \text{where } v_j \sim \chi^2_{n-p+\delta-j-1} \text{ independently for } j = 1, \ldots, p$$

So we can define a pivot for the *generalized variance* $|\boldsymbol{\Sigma}|$ as $N \coloneqq \left|\boldsymbol{\Sigma}\boldsymbol{\mathcal{S}}_y^{-1}\right|$ where

$$N^{-1} \sim \frac{\prod_{i=1}^{p} u_i}{(n-1)^p} \prod_{j=1}^{p} v_j$$

where u_i's and v_j's are as above and they are all pairwise independent. A $(1 - \gamma)$ level credible set for $|\boldsymbol{\Sigma}|$ based on N is

$$\left[a_{n,p,\delta;\gamma}\,|\boldsymbol{\mathcal{S}}_y|\,, b_{n,p,\delta;\gamma}\,|\boldsymbol{\mathcal{S}}_y|\right]$$

where $a_{n,p,\delta;\gamma}$ and $b_{n,p,\delta;\gamma}$ are any two constants that satisfy $1-\gamma = P(a_{n,p,\delta;\gamma} \leq N \leq b_{n,p,\delta;\gamma})$. The length of the credible interval is $|\boldsymbol{\mathcal{S}}_y|\,(b_{n,p,\delta;\gamma} - a_{n,p,\delta;\gamma})$.

Next we define the pivot for $\boldsymbol{\mu}$ as

$$T^2 \coloneqq n(\boldsymbol{\mu} - \bar{\boldsymbol{y}})'\boldsymbol{\mathcal{S}}_y^{-1}(\boldsymbol{\mu} - \bar{\boldsymbol{y}})$$

We will prove that T^2 is a pivot and derive a sampling scheme in what follows. We notice that

$$\sqrt{n}\boldsymbol{\mathcal{S}}_y^{-1/2}(\boldsymbol{\mu} - \bar{\boldsymbol{y}}) \,|\, \boldsymbol{\Sigma}, \boldsymbol{W} \sim N_p(\boldsymbol{0}, \boldsymbol{A})$$

where $\boldsymbol{A} = \boldsymbol{\mathcal{S}}_y^{-1/2}\boldsymbol{W}^{1/2}(\boldsymbol{W}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{W}^{-1/2} + \boldsymbol{I}_p)\boldsymbol{W}^{1/2}\boldsymbol{\mathcal{S}}_y^{-1/2}$, which is obviously defined through the parameters $(\boldsymbol{\Sigma}, \boldsymbol{W})$. If we can prove that the distribution of $\boldsymbol{A}$ is free of $(\boldsymbol{\Sigma}, \boldsymbol{W})$, then by using the fact that if $\boldsymbol{Z} \sim N_p(\boldsymbol{0}, \boldsymbol{A})$ then $\boldsymbol{Z}'\boldsymbol{Z} \sim \sum_{i=1}^p \lambda_i \chi_{1i}^2$ where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of $\boldsymbol{A}$ and χ_{1i}^2 are independent χ_1^2 variables, we can conclude that T^2 is a pivot. Taking $\boldsymbol{Z} = \sqrt{n}\boldsymbol{\mathcal{S}}_y^{-1/2}(\boldsymbol{\mu} - \bar{\boldsymbol{y}})$, $\boldsymbol{B} = \boldsymbol{\mathcal{S}}_y^{-1/2}\boldsymbol{W}\boldsymbol{\mathcal{S}}_y^{-1/2}$ it finally follows that:

(a) the conditional distribution of $T^2 \,|\, \boldsymbol{A}$ is $\sum_{i=1}^p \lambda_i \chi_{1i}^2$ where $\lambda_1, \ldots, \lambda_p$ are the roots of $|\boldsymbol{A} - \lambda \boldsymbol{I}_p| = 0$ such that $\boldsymbol{A} \,|\, \boldsymbol{B} \stackrel{d}{=} \mathcal{W}_p^{-1}\left((n-1)\boldsymbol{B}, n-p+\delta-2\right) + \boldsymbol{B}$ by (4.9) and $\boldsymbol{B} \sim \mathcal{W}_p^{-1}\left(\boldsymbol{I}_p, n-p+\delta-2\right)$ by (4.8); and

(b) the unconditional distribution of T^2 is obtained by averaging over the joint distribution of the roots $\lambda_1, \ldots, \lambda_p$.

We have shown that T^2 is a pivotal quantity, and therefore a $(1-\gamma)$ credible ellipsoid for $\boldsymbol{\mu}$ based on T^2 is given by

$$\left\{\boldsymbol{\mu} \;:\; T^2 \leq c_{n,p,\delta;\gamma}\right\}$$

where $c_{n,p,\delta;\gamma}$ satisfies $1-\gamma = P(T^2 \leq c_{n,p,\delta;\gamma})$. From the above discussion, it follows that the cut-off point $c_{n,p,\delta;\gamma}$ can be obtained by simulating the distribution of T^2 as follows.

1. Generate $\boldsymbol{B} \sim \mathcal{W}_p^{-1}(\boldsymbol{I}_p, n-p+\delta-2)$.
2. Generate $\boldsymbol{A} \,|\, \boldsymbol{B} \sim \mathcal{W}_p^{-1}((n-1)\boldsymbol{B}, n-p+\delta-2) + \boldsymbol{B}$.
3. Generate $\lambda_1, \ldots, \lambda_p$, the roots of $|\boldsymbol{A} - \lambda \boldsymbol{I}_p| = 0$.
4. Generate $T^2 = \sum_{i=1}^{p} \lambda_i \chi_{1i}^2$ where χ_{1i}^2 are independent χ_1^2 variables.

The volume of the credible ellipsoid is given by

$$V_{\boldsymbol{\mu}}(\boldsymbol{Y}) = \frac{\pi^{p/2}}{\Gamma\left(\frac{p}{2}+1\right)} \left(c_{n,p,\delta;\gamma}/n\right)^{p/2} |\boldsymbol{\mathcal{S}}_y|^{1/2}$$

4.2 Posterior Predictive Sampling method

We return to the setup of the last section. Under the *posterior predictive sampling* method, starting with a vague prior $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\alpha/2}$, the joint (imputed) posterior distribution of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, given $\boldsymbol{X}$, can be represented as

$$\begin{aligned} \boldsymbol{\Sigma} \,|\, \boldsymbol{X} &\sim \mathcal{W}_p^{-1}((n-1)\boldsymbol{W}, n-p+\alpha-2) \\ \boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \boldsymbol{X} &\sim N_p\left(\bar{\boldsymbol{x}}, n^{-1}\boldsymbol{\Sigma}\right) \end{aligned} \tag{4.14}$$

We assume throughout that $n+\alpha > 2p+1$. We now draw $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ from the above posterior, resulting in $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, and then draw a random sample $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ as iid from $N_p(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$, which form the singly imputed synthetic data that are released. Define $\bar{\boldsymbol{y}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_i$ (sample mean based on $\boldsymbol{Y}$) and $\boldsymbol{\mathcal{S}}_y = \sum_{i=1}^{n} (\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})'$

(sample Wishart matrix based on $\boldsymbol{Y}$) which are jointly sufficient for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ by Lemma 1.2.1. From Klein and Sinha (2015a), we have the following result.

Theorem 4.2.1. *The joint pdf of* $\bar{\boldsymbol{y}}$ *and* $\boldsymbol{\mathcal{S}}_y$ *is obtained by integrating out* $\boldsymbol{\Sigma}^*$ *from the joint pdf of* $(\bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y, \boldsymbol{\Sigma}^*)$ *given by*

$$f(\bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y, \boldsymbol{\Sigma}^*) \propto e^{-\frac{1}{2}\left[n(\bar{\boldsymbol{y}}-\boldsymbol{\mu})'(\boldsymbol{\Sigma}+2\boldsymbol{\Sigma}^*)^{-1}(\bar{\boldsymbol{y}}-\boldsymbol{\mu})+\mathrm{tr}(\boldsymbol{\mathcal{S}}_y\boldsymbol{\Sigma}^{*-1})\right]} \left|\boldsymbol{\Sigma}+2\boldsymbol{\Sigma}^*\right|^{-\frac{1}{2}} \left|\boldsymbol{\Sigma}\right|^{\frac{n-p+\alpha-2}{2}}$$
$$\left|\boldsymbol{\Sigma}+\boldsymbol{\Sigma}^*\right|^{-\frac{2n-p+\alpha-3}{2}} \left|\boldsymbol{\Sigma}^*\right|^{-\frac{p+1}{2}} \left|\boldsymbol{\mathcal{S}}_y\right|^{\frac{n-p-2}{2}}$$

Posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

We choose the same prior $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{\delta}{2}}$ as before and attempt to compute the posterior distribution as before by multiplying the expression inside the above integral with the prior and the product should split up into exactly three parts corresponding to the three conditional posterior distributions.

$$\pi(\bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y, \boldsymbol{W} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \left(\left|\boldsymbol{\Sigma}+2\boldsymbol{\Sigma}^*\right|^{-\frac{1}{2}} e^{-\frac{1}{2}\left[n(\bar{\boldsymbol{y}}-\boldsymbol{\mu})'(\boldsymbol{\Sigma}+2\boldsymbol{\Sigma}^*)^{-1}(\bar{\boldsymbol{y}}-\boldsymbol{\mu})\right]}\right)$$
$$\times \left(\left|\boldsymbol{\Sigma}\right|^{\frac{n-p+\alpha-\delta-2}{2}} \left|\boldsymbol{\Sigma}+\boldsymbol{\Sigma}^*\right|^{-\frac{2n-p+\alpha-3}{2}} \left|\boldsymbol{\Sigma}^*\right|^{-\frac{p+1}{2}} e^{-\frac{1}{2}\left[\mathrm{tr}(\boldsymbol{\mathcal{S}}_y\boldsymbol{\Sigma}^{*-1})\right]}\right) \tag{4.15}$$

We see that the part involving $\boldsymbol{\mu}$ separates out nicely in front and thus it's posterior distribution is obvious. We will now work with the part inside the second parenthesis involving just the determinants below.

$$\begin{aligned} &\left|\boldsymbol{\Sigma}\right|^{\frac{n-p+\alpha-\delta-2}{2}} \left|\boldsymbol{\Sigma}+\boldsymbol{\Sigma}^*\right|^{-\frac{2n-p+\alpha-3}{2}} \left|\boldsymbol{\Sigma}^*\right|^{-\frac{p+1}{2}} \\ &= \left|\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1/2}\right|^{\frac{n-p+\alpha-\delta-2}{2}} \left|I_p+\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1/2}\right|^{-\frac{2n-p+\alpha-3}{2}} \left|\boldsymbol{\Sigma}^*\right|^{-\frac{n+p+\delta}{2}} \end{aligned} \tag{4.16}$$

Next we combine equations 4.15 and 4.16 and multiply by the Jacobian of the transformation $\boldsymbol{\Sigma} \mapsto \boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1/2}$ which is $|\boldsymbol{\Sigma}^*|^p$ to get

$$
\begin{aligned}
&\pi(\bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y, \boldsymbol{W} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
\propto\ &\left(\left|\boldsymbol{\Sigma} + 2\boldsymbol{\Sigma}^*\right|^{-\frac{1}{2}} e^{-\frac{1}{2}\left[n(\bar{\boldsymbol{y}}-\boldsymbol{\mu})'(\boldsymbol{\Sigma}+2\boldsymbol{\Sigma}^*)^{-1}(\bar{\boldsymbol{y}}-\boldsymbol{\mu})\right]} \right) \\
&\left(\left|\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1/2}\right|^{\frac{n-p+\alpha-\delta-2}{2}} \left|I_p + \boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1/2}\right|^{-\frac{2n-p+\alpha-3}{2}} \right) \\
&\left(\left|\boldsymbol{\Sigma}^*\right|^{-\frac{n-p+\delta}{2}} e^{-\frac{1}{2}\left[\mathrm{tr}(\boldsymbol{\mathcal{S}}_y\boldsymbol{\Sigma}^{*-1})\right]} \right)
\end{aligned}
$$

which indicates that the posterior sampling will be done sequentially in the following manner:

$$
\begin{aligned}
\boldsymbol{\Sigma}^* \mid \boldsymbol{\mathcal{S}}_y &\sim \mathcal{W}_p^{-1}\left(\boldsymbol{\mathcal{S}}_y, n-2p+\delta-1\right) && (4.17)\\
\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1/2} &\sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}\left(\frac{n+\alpha-\delta-1}{2}, \frac{n-p+\delta-2}{2}\right) && (4.18)\\
\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \boldsymbol{\Sigma}^*, \bar{\boldsymbol{y}} &\sim N_p\left(\bar{\boldsymbol{y}}, \frac{1}{n}\left(\boldsymbol{\Sigma}+2\boldsymbol{\Sigma}^*\right)\right) && (4.19)
\end{aligned}
$$

where $\mathrm{B}_{\mathrm{p}}^{\mathrm{II}}(a,b)$ denotes the matrix variate beta type II distribution as described in Gupta and Nagar (2000). We can reformulate the above posterior distributions as:

$$
\begin{aligned}
\boldsymbol{\mathcal{S}}_y^{-1/2}\boldsymbol{\Sigma}^*\boldsymbol{\mathcal{S}}_y^{-1/2} &\sim \mathcal{W}_p^{-1}\left(\boldsymbol{I}_p, n-2p+\delta-1\right) && (4.20)\\
\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1/2} &\sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}\left(\frac{n+\alpha-\delta-1}{2}, \frac{n-p+\delta-2}{2}\right) && (4.21)\\
\boldsymbol{\mu} \mid \boldsymbol{\Sigma}, \boldsymbol{\Sigma}^*, \bar{\boldsymbol{y}} &\sim N_p\left(\bar{\boldsymbol{y}}, \frac{1}{n}\left(\boldsymbol{\Sigma}+2\boldsymbol{\Sigma}^*\right)\right) && (4.22)
\end{aligned}
$$

which has the benefit that $\boldsymbol{\mathcal{S}}_y^{-1/2}\boldsymbol{\Sigma}^*\boldsymbol{\mathcal{S}}_y^{-1/2}$ is independent of $\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1/2}$ and it's posterior distribution is unconditional.

The posterior distributions are proper as long as $n > \max\{p, 2p-\alpha+1, 3p-\delta, p-\alpha+\delta, 2p-\delta+1\}$.

Bayes Estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

$$\hat{\boldsymbol{\mu}}_{\text{BAYES}} = \mathrm{E}(\boldsymbol{\mu} \,|\, \bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y) = \mathrm{E}_{\boldsymbol{\Sigma}^*}\, \mathrm{E}_{\boldsymbol{\Sigma}}\, \mathrm{E}(\boldsymbol{\mu} \,|\, \bar{\boldsymbol{y}}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}^*) = \mathrm{E}_{\boldsymbol{\Sigma}^*}\, \mathrm{E}_{\boldsymbol{\Sigma}}\, \mathrm{E}(\bar{\boldsymbol{y}}) = \bar{\boldsymbol{y}}$$

Finding $\hat{\boldsymbol{\Sigma}}_{\text{BAYES}}$ seems very difficult.

$$\begin{aligned}
&\widehat{|\boldsymbol{\Sigma}|}_{\text{BAYES}} = \mathrm{E}\left(|\boldsymbol{\Sigma}| \,|\, \bar{\boldsymbol{y}}, \boldsymbol{\mathcal{S}}_y\right) = \mathrm{E}_{\boldsymbol{\Sigma}^*}\, \mathrm{E}\left(|\boldsymbol{\Sigma}| \,|\, \boldsymbol{\mathcal{S}}_y, \boldsymbol{\Sigma}^*\right) \\
&= \mathrm{E}_{\boldsymbol{\Sigma}^*}\left(|\boldsymbol{\Sigma}^*|\, \mathrm{E}\left(\left|\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1/2}\right|\right) \,|\, \boldsymbol{\mathcal{S}}_y\right) = \left(\prod_{j=1}^{p} \frac{n+\alpha-\delta-j}{n-p+\delta-j-3}\right) \mathrm{E}\left(|\boldsymbol{\Sigma}^*| \,|\, \boldsymbol{\mathcal{S}}_y\right) \\
&= \left(\prod_{j=1}^{p} \frac{n+\alpha-\delta-j}{(n-p+\delta-j-3)(n-2p+\delta-j-2)}\right) |\boldsymbol{\mathcal{S}}_y|
\end{aligned}$$

provided that $n > \max\{p, 2p-\alpha+1, p-\alpha+\delta, 3p-\delta+4\}$. We use (4.12) and the following result for the above derivation: If $\boldsymbol{V} \sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}(a, b)$ then

$$\mathrm{E}\left(|\boldsymbol{V}|\right) = \prod_{j=1}^{p} \frac{a - \frac{1}{2}(j-1)}{b - \frac{1}{2}(j+1)} \quad \text{if } a > \frac{p-1}{2},\, b > \frac{p+1}{2} \tag{4.23}$$

Credible Sets for $|\boldsymbol{\Sigma}|$ and $\boldsymbol{\mu}$

Let $\boldsymbol{C} = \boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1/2}$. Then by (4.21), we have $\boldsymbol{C}^{-1} \sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}\left(\frac{n-p+\delta-3}{2}, \frac{n+\alpha-\delta}{2}\right)$. Also by (4.13) we can get,

$$\frac{\left|\boldsymbol{\Sigma}^{*-1}\right|}{\left|\boldsymbol{\mathcal{S}}_y^{-1}\right|} \sim \prod_{i=1}^{p} v_j, \quad \text{where } v_j \sim \chi^2_{n-2p+\delta-j} \text{ independently for } j = 1, \ldots, p$$

We can define a pivot for $|\boldsymbol{\Sigma}|$ in the same manner as in the last section to be $N \coloneqq \left|\boldsymbol{\Sigma}\boldsymbol{\mathcal{S}}_y^{-1}\right|$ where

$$N^{-1} \sim |\boldsymbol{M}| \prod_{j=1}^{p} v_j$$

where v_j's are defined as above, $\boldsymbol{M} \sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}\left(\frac{n-p+\delta-3}{2}, \frac{n+\alpha-\delta}{2}\right)$ and $\boldsymbol{M}$ is independent of $v_j\ \forall j$. Since the distribution of N is free of $(\boldsymbol{\Sigma}, \boldsymbol{\Sigma}^*)$ we conclude that it is a pivot. A $(1-\gamma)$ level credible set for $|\boldsymbol{\Sigma}|$ is

$$\left[a_{n,p,\alpha,\delta;\gamma}\left|\boldsymbol{\mathcal{S}}_y\right|, b_{n,p,\alpha,\delta;\gamma}\left|\boldsymbol{\mathcal{S}}_y\right|\right]$$

where $a_{n,p,\alpha,\delta;\gamma}$ and $b_{n,p,\alpha,\delta;\gamma}$ are any two constants that satisfy $1-\gamma = P(a_{n,p,\alpha,\delta;\gamma} \leq N \leq b_{n,p,\alpha,\delta;\gamma})$. The length of the credible interval is $\left|\boldsymbol{\mathcal{S}}_y\right|(b_{n,p,\alpha,\delta;\gamma} - a_{n,p,\alpha,\delta;\gamma})$.

Next we define the pivot similarly as in the last section for $\boldsymbol{\mu}$ as

$$T^2 \coloneqq n(\boldsymbol{\mu}-\bar{\boldsymbol{y}})'\boldsymbol{\mathcal{S}}_y^{-1}(\boldsymbol{\mu}-\bar{\boldsymbol{y}})$$

We will prove that T^2 is a pivot and derive a sampling scheme in what follows. We notice that

$$\sqrt{n}\boldsymbol{\mathcal{S}}_y^{-1/2}(\boldsymbol{\mu}-\bar{\boldsymbol{y}}) \mid \boldsymbol{\Sigma}, \boldsymbol{\Sigma}^* \sim N_p(\boldsymbol{0}, \boldsymbol{A})$$

where $\boldsymbol{A} = \boldsymbol{\mathcal{S}}_y^{-1/2}\boldsymbol{\Sigma}^{*1/2}(\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1/2} + 2\boldsymbol{I}_p)\boldsymbol{\Sigma}^{*1/2}\boldsymbol{\mathcal{S}}_y^{-1/2}$, which is obviously defined through the parameters $(\boldsymbol{\Sigma}, \boldsymbol{\Sigma}^*)$. If we can prove that the distribution of $\boldsymbol{A}$ is free of $(\boldsymbol{\Sigma}, \boldsymbol{\Sigma}^*)$, then by using the fact that if $\boldsymbol{Z} \sim N_p(\boldsymbol{0}, \boldsymbol{A})$ then $\boldsymbol{Z}'\boldsymbol{Z} \sim \sum_{i=1}^p \lambda_i \chi_{1i}^2$ where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of $\boldsymbol{A}$ and χ_{1i}^2 are independent χ_1^2 variables, we can conclude that T^2 is a pivot. Taking $\boldsymbol{Z} = \sqrt{n}\boldsymbol{\mathcal{S}}_y^{-1/2}(\boldsymbol{\mu}-\bar{\boldsymbol{y}})$, $\boldsymbol{B} = \boldsymbol{\mathcal{S}}_y^{-1/2}\boldsymbol{\Sigma}^*\boldsymbol{\mathcal{S}}_y^{-1/2}$ it finally follows that:

(a) the conditional distribution of $T^2 \mid \boldsymbol{A}$ is $\sum_{i=1}^p \lambda_i \chi_{1i}^2$ where $\lambda_1, \ldots, \lambda_p$ are the roots of $|\boldsymbol{A} - \lambda \boldsymbol{I}_p| = 0$ such that $\boldsymbol{B} \sim \mathcal{W}_p^{-1}(\boldsymbol{I}_p, n-2p+\delta-1)$ by (4.20) and

$$\boldsymbol{A} \mid \boldsymbol{B} \overset{d}{=} \mathrm{GB}_{\mathrm{p}}^{\mathrm{II}}\left(\frac{n+\alpha-\delta}{2}, \frac{n-p+\delta-3}{2}; \boldsymbol{B}, \boldsymbol{O}\right) + 2\boldsymbol{B}$$

where $\mathrm{GB}_{\mathrm{p}}^{\mathrm{II}}(a, b; \boldsymbol{\Omega}, \boldsymbol{\Psi})$ denotes the generalized matrix variate beta type II distribution as described in Gupta and Nagar (2000). The above derivation follows from (4.21) and the result: If $\boldsymbol{V} \sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}(a, b)$, $\boldsymbol{A}_{p\times p}$ is a constant, non-singular matrix then $\boldsymbol{AVA'} \sim \mathrm{GB}_{\mathrm{p}}^{\mathrm{II}}(a, b; \boldsymbol{AA'}, \boldsymbol{O})$.

(b) the unconditional distribution of T^2 is obtained by averaging over the joint distribution of the roots $\lambda_1, \ldots, \lambda_p$.

We have shown that T^2 is a pivotal quantity, and therefore a $(1-\gamma)$ credible ellipsoid for $\boldsymbol{\mu}$ based on T^2 is given by

$$\left\{\boldsymbol{\mu} \ : \ T^2 \leq c_{n,p,\alpha,\delta;\gamma}\right\}$$

where $c_{n,p,\alpha,\delta;\gamma}$ satisfies $1-\gamma = P(T^2 \leq c_{n,p,\alpha,\delta;\gamma})$. From the above discussion, it follows that the cut-off point $c_{n,p,\alpha,\delta;\gamma}$ can be obtained by simulating the distribution of T^2 as follows.

1. Generate $\boldsymbol{B} \sim \mathcal{W}_p^{-1}(\boldsymbol{I}_p, n-2p+\delta-1)$.
2. Generate $\boldsymbol{A} \,|\, \boldsymbol{B} \sim \mathrm{GB}_{\mathrm{p}}^{\mathrm{II}}\left(\frac{n+\alpha-\delta}{2}, \frac{n-p+\delta-3}{2}; \boldsymbol{B}, \boldsymbol{O}\right) + 2\boldsymbol{B}$.
3. Generate $\lambda_1, \ldots, \lambda_p$, the roots of $|\boldsymbol{A} - \lambda \boldsymbol{I}_p| = 0$.
4. Generate $T^2 = \sum_{i=1}^{p} \lambda_i \chi_{1i}^2$ where χ_{1i}^2 are independent χ_1^2 variables.

The volume of the credible ellipsoid is given by

$$V_{\boldsymbol{\mu}}(\boldsymbol{Y}) = \frac{\pi^{p/2}}{\Gamma\left(\frac{p}{2}+1\right)} \left(c_{n,p,\alpha,\delta;\gamma}/n\right)^{p/2} |\boldsymbol{S}_y|^{1/2}$$

4.3 Simulation Studies

To conduct the simulation, the population distribution is taken to be the multivariate normal model (4.1) with

$$p = 10, \quad \boldsymbol{\mu} = 0.1 \times \begin{pmatrix} 1 & 2 & \dots & 10 \end{pmatrix}', \quad \boldsymbol{\Sigma} = 0.25\boldsymbol{I}_p + 0.75\boldsymbol{J}_p, \tag{4.24}$$

where $\boldsymbol{I}_p$ is the $p \times p$ dimensional identity matrix and $\boldsymbol{J}_p$ is the $p \times p$ matrix of 1's. Based on Monte Carlo simulation with 10^4 iterations, we compute an estimate of the coverage probability, the volume or length (as appropriate) of the respective credible sets and the Bayes estimators of $\boldsymbol{\mu}$ and $|\boldsymbol{\Sigma}|$, where in all cases, the level of credibility is set at 0.95.

In both PIS and PPS cases, increasing δ increases the coverage of $|\boldsymbol{\Sigma}|$ before it drops off, the effect hastened for small values of n. We thus see the reverse-sigmoid shape of the curve in all situations, it is wider in the PIS case, and the curve seems to shift to the right with increasing α in the PPS case. So the best choice of δ to ensure maximum coverage of $|\boldsymbol{\Sigma}|$ would increase with increasing α, in the PPS case.

For $\boldsymbol{\mu}$, we see that increasing δ slightly increases the coverage before decreasing steadily, albeit at a much slower pace than that of $|\boldsymbol{\Sigma}|$.

The size of the credible sets of both quantities shrink with either increasing n or δ. Asymptotically the results conform to our expectations, with the inference worsening for higher δ, quicker for $|\boldsymbol{\Sigma}|$ than $\boldsymbol{\mu}$. The better inference we get off the PIS method than the PPS method attests to the trade-off between data utility and data privacy.

The recommendation is to use $\delta = 10$ in the PIS case, PPS case with $\alpha = 2$, and $\delta = 20$ in the PPS case with $\alpha = 50$.

Table 4.1: Inference for $\boldsymbol{\mu}$ and $|\boldsymbol{\Sigma}|$ for SI PIS MVN data with $n = 1000$

	$\lvert\boldsymbol{\Sigma}\rvert$		$\boldsymbol{\mu}$	
	avg	*est*	*avg*	*est*
δ	*cvg*	*len*	*cvg*	*vol*
0.2	0.8093	3.0625e-05	0.9531	1.0384e-09
0.5	0.8125	3.0607e-05	0.9552	1.0399e-09
0.8	0.8319	3.0254e-05	0.9595	1.1277e-09
1	0.8294	3.0172e-05	0.9581	1.0437e-09
2	0.8661	2.9942e-05	0.9572	1.0776e-09
3	0.8755	2.8962e-05	0.9501	1.0287e-09
4	0.8874	2.7952e-05	0.9533	1.0288e-09
5	0.9108	2.7679e-05	0.9542	1.0313e-09
10	0.9517	2.4692e-05	0.9494	9.6189e-10
20	0.8491	2.0096e-05	0.95	9.0293e-10
30	0.5135	1.6400e-05	0.9396	8.0198e-10
50	0.0303	1.1083e-05	0.9297	6.9183e-10
100	0	4.1404e-06	0.8979	4.9770e-10

Table 4.2: Inference for $\boldsymbol{\mu}$ and $|\boldsymbol{\Sigma}|$ for SI PIS MVN data with $n = 10000$

	$\lvert\boldsymbol{\Sigma}\rvert$		$\boldsymbol{\mu}$	
	avg	*est*	*avg*	*est*
δ	*cvg*	*len*	*cvg*	*vol*
0.2	0.9353	7.5066e-06	0.9475	9.0069e-15
0.5	0.9347	7.4725e-06	0.9494	9.2580e-15
0.8	0.9379	7.4664e-06	0.947	9.1447e-15
1	0.9386	7.4609e-06	0.9491	8.9631e-15
2	0.9408	7.5066e-06	0.9516	9.4023e-15
3	0.9464	7.5633e-06	0.9505	9.4476e-15
4	0.9401	7.3873e-06	0.9492	9.0631e-15
5	0.9483	7.4442e-06	0.9462	8.7338e-15
10	0.9477	7.3689e-06	0.9478	9.3546e-15
20	0.9422	7.3092e-06	0.9442	8.9443e-15
30	0.9136	7.1060e-06	0.9532	9.6100e-15
50	0.7604	6.7308e-06	0.9519	9.2865e-15
100	0.1918	6.1571e-06	0.9465	8.3714e-15

Table 4.3: Inference for $\boldsymbol{\mu}$ and $|\boldsymbol{\Sigma}|$ for SI PPS MVN data with $\alpha = 2$, $n = 1000$

	$\lvert\boldsymbol{\Sigma}\rvert$		$\boldsymbol{\mu}$	
	avg	*est*	*avg*	*est*
δ	*cvg*	*len*	*cvg*	*vol*
0.2	0.8259	4.0296e-05	0.9552	1.0947e-09
0.5	0.8561	4.0295e-05	0.9523	1.0695e-09
0.8	0.8521	4.0643e-05	0.9571	1.0774e-09
1	0.8745	4.0285e-05	0.9588	1.0814e-09
2	0.8862	3.8019e-05	0.9586	1.0862e-09
3	0.9157	3.7622e-05	0.9591	1.1022e-09
4	0.9444	3.6284e-05	0.9542	1.0461e-09
5	0.9556	3.5436e-05	0.9502	1.0046e-09
10	0.9864	3.0662e-05	0.9536	1.0498e-09
20	0.7985	2.2145e-05	0.9472	9.0976e-10
30	0.2629	1.6346e-05	0.9381	7.8479e-10
50	0.0001	8.8689e-06	0.9298	6.6995e-10
100	0	1.9857e-06	0.8784	3.9465e-10

Table 4.4: Inference for $\boldsymbol{\mu}$ and $|\boldsymbol{\Sigma}|$ for SI PPS MVN data with $\alpha = 2$, $n = 10000$

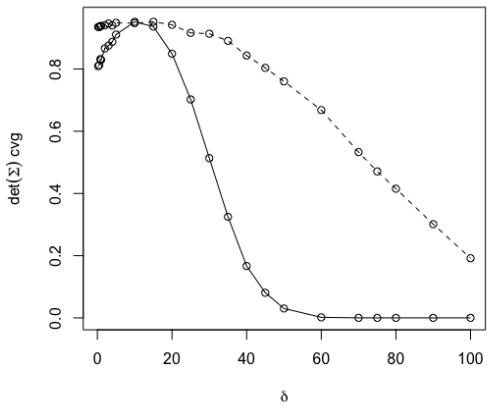
	$\lvert\boldsymbol{\Sigma}\rvert$		$\boldsymbol{\mu}$	
	avg	*est*	*avg*	*est*
δ	*cvg*	*len*	*cvg*	*vol*
0.2	0.974	9.4585e-06	0.9481	9.0549e-15
0.5	0.9722	9.2176e-06	0.955	9.4496e-15
0.8	0.975	9.2989e-06	0.949	9.3798e-15
1	0.9729	9.3212e-06	0.9552	9.8925e-15
2	0.9728	9.0660e-06	0.9522	9.6263e-15
3	0.9799	9.1529e-06	0.9477	8.8715e-15
4	0.9789	9.2158e-06	0.9569	9.8428e-15
5	0.9813	9.2459e-06	0.9541	1.0064e-14
10	0.9835	8.9542e-06	0.9494	8.8645e-15
20	0.9672	8.6689e-06	0.9489	8.7975e-15
30	0.924	8.4805e-06	0.9502	9.0701e-15
50	0.7123	8.1242e-06	0.9449	8.5905e-15
100	0.0291	6.8814e-06	0.9431	8.3882e-15

Table 4.5: Inference for $\boldsymbol{\mu}$ and $|\boldsymbol{\Sigma}|$ for SI PPS MVN data with $\alpha = 50$, $n = 1000$

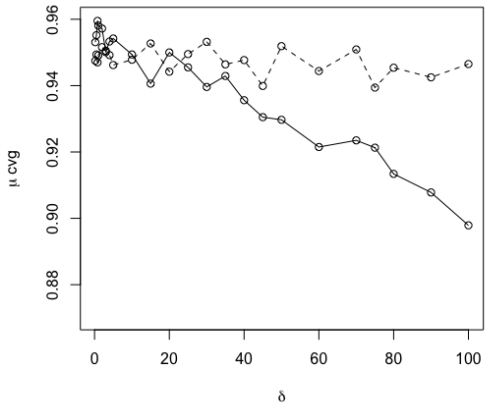
	$\lvert\boldsymbol{\Sigma}\rvert$		$\boldsymbol{\mu}$	
	avg	*est*	*avg*	*est*
δ	*cvg*	*len*	*cvg*	*vol*
0.2	0.0742	6.4298e-05	0.961	1.1587e-09
0.5	0.0841	6.3811e-05	0.9641	1.3207e-09
0.8	0.0826	6.4185e-05	0.9616	1.2371e-09
1	0.1073	6.3435e-05	0.9596	1.2032e-09
2	0.1248	6.1302e-05	0.9654	1.2527e-09
3	0.1657	6.0413e-05	0.9601	1.2380e-09
4	0.1925	5.8477e-05	0.9541	1.1280e-09
5	0.2418	5.7401e-05	0.9602	1.1985e-09
10	0.5246	4.7836e-05	0.9598	1.1788e-09
20	0.9323	3.5425e-05	0.953	1.0537e-09
30	0.963	2.6544e-05	0.948	9.4866e-10
50	0.1188	1.4721e-05	0.9335	7.2119e-10
100	0	3.2798e-06	0.9007	4.6731e-10

Table 4.6: Inference for $\boldsymbol{\mu}$ and $|\boldsymbol{\Sigma}|$ for SI PPS MVN data with $\alpha = 50$, $n = 10000$

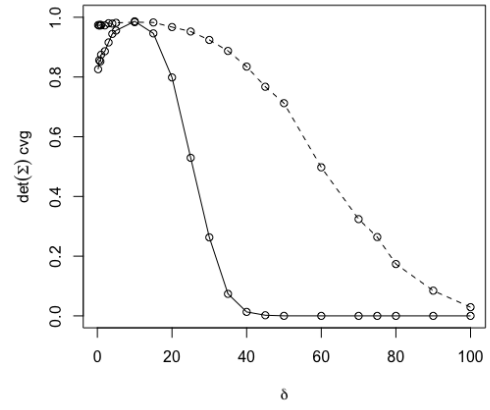
	$\lvert\boldsymbol{\Sigma}\rvert$		$\boldsymbol{\mu}$	
	avg	*est*	*avg*	*est*
δ	*cvg*	*len*	*cvg*	*vol*
0.2	0.8845	9.7714e-06	0.9542	9.4786e-15
0.5	0.8776	9.7090e-06	0.9532	9.3734e-15
0.8	0.8921	9.7755e-06	0.9502	9.3442e-15
1	0.8949	9.6187e-06	0.9463	9.0667e-15
2	0.9002	9.7131e-06	0.9549	1.0231e-14
3	0.9110	9.7131e-06	0.9534	9.7385e-15
4	0.9151	9.6263e-06	0.9501	9.2920e-15
5	0.9147	9.5988e-06	0.9495	9.5697e-15
10	0.9489	9.4674e-06	0.9532	9.2661e-15
20	0.981	9.3402e-06	0.9496	9.1140e-15
30	0.9822	8.9777e-06	0.951	9.4796e-15
50	0.897	8.4004e-06	0.9466	8.8161e-15
100	0.1373	7.2256e-06	0.9427	8.1619e-15



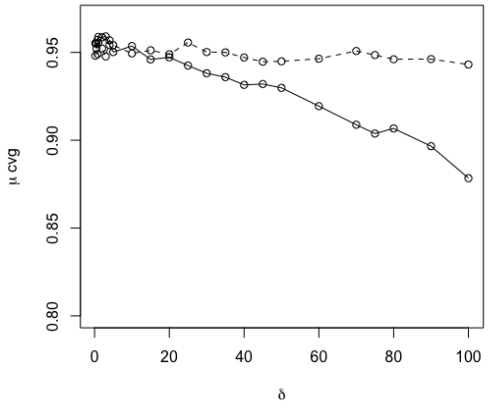
(a) $|\boldsymbol{\Sigma}|$ (PIS)



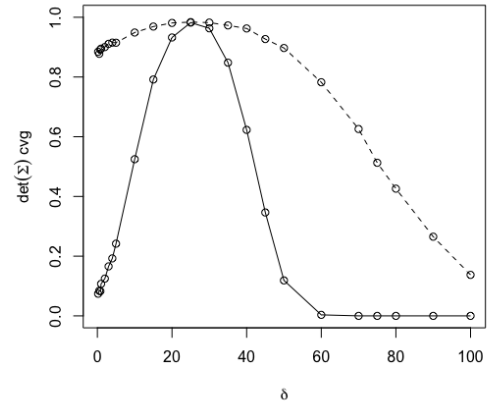
(b) $\boldsymbol{\mu}$ (PIS)



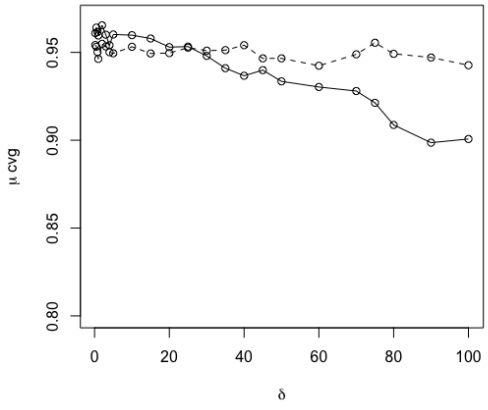
(c) $|\boldsymbol{\Sigma}|$, $\alpha = 2$ (PPS)



(d) $\boldsymbol{\mu}$, $\alpha = 2$ (PPS)



(e) $|\boldsymbol{\Sigma}|$, $\alpha = 50$ (PPS)



(f) $\boldsymbol{\mu}$, $\alpha = 50$ (PPS)

Figure 4.1: Variation in coverage of $\boldsymbol{\mu}$ and $|\boldsymbol{\Sigma}|$ with respect to δ for SI MVN data (— $n = 1000$, - - - $n = 10000$)

4.4 Partially Sensitive Data

Method I: Using only estimates of sensitive part to impute synthetic data

Plug-In Sampling

Let us now assume, following from (4.1) that $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$ is sensitive and the rest $(\boldsymbol{x}_{r+1}, \ldots, \boldsymbol{x}_n)$ is not. The sufficient statistics for the sensitive part, assuming $r > p$, is given by

$$\bar{\boldsymbol{x}}_r = \frac{1}{r}\sum_{i=1}^{r} \boldsymbol{x}_i \sim N_p\left(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{r}\right)$$
$$\boldsymbol{W}^{(r)} = \frac{1}{r-1}\sum_{i=1}^{r}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_r)(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_r)' \sim \frac{\mathcal{W}_p(\boldsymbol{\Sigma}, r-1)}{r-1}$$

and the sufficient statistics for the non-sensitive part, assuming $n - r > p$, is given by

$$\bar{\boldsymbol{x}}_{n-r} = \frac{1}{n-r}\sum_{i=r+1}^{n} \boldsymbol{x}_i \sim N_p\left(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{n-r}\right)$$
$$\boldsymbol{W}^{(n-r)} = \frac{1}{n-r-1}\sum_{i=r+1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{n-r})(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{n-r})' \sim \frac{\mathcal{W}_p(\boldsymbol{\Sigma}, n-r-1)}{n-r-1}$$

We will impute the synthetic counterparts to the sensitive data using only the sufficient statistics for the sensitive part so as to ensure the imputed data is independent of the non-sensitive data. Thus we generate

$$\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r \overset{\text{iid}}{\sim} N_p\left(\bar{\boldsymbol{x}}_r, \boldsymbol{W}^{(r)}\right)$$

so that the released data is $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r, \boldsymbol{x}_{r+1}, \ldots, \boldsymbol{x}_n)$. Since $(\bar{\boldsymbol{x}}_r, \boldsymbol{W}^{(r)}, \bar{\boldsymbol{x}}_{n-r}, \boldsymbol{W}^{(n-r)})$ is the sufficient statistics of the original data, so by Lemma 1.2.1 $(\bar{\boldsymbol{y}}_r, \boldsymbol{W}_y^{(r)}, \bar{\boldsymbol{x}}_{n-r}, \boldsymbol{W}^{(n-r)})$

is the sufficient statistics for the released data where

$$\bar{\boldsymbol{y}}_r = \frac{1}{r}\sum_{i=1}^{r}\boldsymbol{y}_i, \quad \bar{\boldsymbol{y}}_r \,|\, \bar{\boldsymbol{x}}_r, \boldsymbol{W}^{(r)} \sim N_p\left(\bar{\boldsymbol{x}}_r, \frac{\boldsymbol{W}^{(r)}}{r}\right)$$
$$\boldsymbol{W}_y^{(r)} = \frac{1}{r-1}\sum_{i=1}^{r}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_r)(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_r)', \quad \boldsymbol{W}_y^{(r)} \,|\, \boldsymbol{W}^{(r)} \sim \frac{\mathcal{W}_p(\boldsymbol{W}^{(r)}, r-1)}{r-1}$$

We can then compute the likelihood of the released data and multiply it with our regular prior $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{\frac{\delta}{2}}$ to find the following posterior distributions

$$\begin{aligned}
&\boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \bar{\boldsymbol{y}}_r, \boldsymbol{W}^{(r)}, \bar{\boldsymbol{x}}_{n-r} \\
&\sim N_p\left[\left(r\left(\boldsymbol{\Sigma} + \boldsymbol{W}^{(r)}\right)^{-1} + (n-r)\boldsymbol{\Sigma}^{-1}\right)^{-1}\left(r(\boldsymbol{\Sigma} + \boldsymbol{W}^{(r)})^{-1}\bar{\boldsymbol{y}}_r + (n-r)\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{x}}_{n-r}\right),\right. \\
&\qquad \left.\left(r\left(\boldsymbol{\Sigma} + \boldsymbol{W}^{(r)}\right)^{-1} + (n-r)\boldsymbol{\Sigma}^{-1}\right)^{-1}\right]
\end{aligned} \tag{4.25}$$

$$\begin{aligned}
&\pi(\boldsymbol{\Sigma}, \boldsymbol{W}^{(r)} \,|\, \bar{\boldsymbol{y}}_r, \boldsymbol{W}_y^{(r)}, \bar{\boldsymbol{x}}_{n-r}, \boldsymbol{W}^{(n-r)}) \\
&\propto \exp\left[-\tfrac{r(n-r)}{2}(\bar{\boldsymbol{y}}_r - \bar{\boldsymbol{x}}_{n-r})'(n\boldsymbol{\Sigma} + (n-r)\boldsymbol{W}^{(r)})^{-1}(\bar{\boldsymbol{y}}_r - \bar{\boldsymbol{x}}_{n-r})\right] \\
&\quad \left|n\boldsymbol{\Sigma} + (n-r)\boldsymbol{W}^{(r)}\right|^{-\frac{1}{2}}\left|\boldsymbol{W}^{(r)}\right|^{-\frac{p+1}{2}}|\boldsymbol{\Sigma}|^{-\frac{n+\delta-2}{2}} \\
&\quad \exp\left[-\tfrac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}((n-r-1)\boldsymbol{W}^{(n-r)} + (r-1)\boldsymbol{W}^{(r)}))\right] \\
&\quad \exp\left[-\tfrac{r-1}{2}\operatorname{tr}({\boldsymbol{W}^{(r)}}^{-1}\boldsymbol{W}_y^{(r)})\right]
\end{aligned} \tag{4.26}$$

The distributions of $\boldsymbol{\Sigma}$ and the latent matrix $\boldsymbol{W}^{(r)}$ are inextricably entangled, we would have to find a way to sample them computationally.

Posterior Predictive Sampling

Assuming $r > \max\{p, 2p - \alpha + 1\}$, we generate a posterior draw $(\boldsymbol{\mu}_r^*, \boldsymbol{\Sigma}_r^*)$ from

$$\begin{aligned}
\boldsymbol{\Sigma} \,|\, \boldsymbol{W}^{(r)} &\sim \mathcal{W}_p^{-1}\left((r-1)\boldsymbol{W}^{(r)}, r - p + \alpha - 2\right) \\
\boldsymbol{\mu} \,|\, \bar{\boldsymbol{x}}_r, \boldsymbol{\Sigma} &\sim N_p\left(\bar{\boldsymbol{x}}_r, \frac{\boldsymbol{\Sigma}}{r}\right)
\end{aligned}$$

so that the released data is $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r, \boldsymbol{x}_{r+1}, \ldots, \boldsymbol{x}_n)$ where $\boldsymbol{y}_i \overset{\text{iid}}{\sim} N_p(\boldsymbol{\mu}_r^*, \boldsymbol{\Sigma}_r^*)$ for $i = 1, \ldots, r$. Thus the sufficient statistics for released data is $(\bar{\boldsymbol{y}}_r, \boldsymbol{W}_y^{(r)}, \bar{\boldsymbol{x}}_{n-r}, \boldsymbol{W}^{(n-r)})$, the quantities are defined as in the preceding page, whose distributions are as follows

$$\begin{aligned}
\bar{\boldsymbol{y}}_r \,|\, \boldsymbol{\mu}_r^*, \boldsymbol{\Sigma}_r^* &\sim N_p\left(\boldsymbol{\mu}_r^*, \frac{\boldsymbol{\Sigma}_r^*}{r}\right); \quad \boldsymbol{W}_y^{(r)} \,|\, \boldsymbol{\Sigma}_r^* \sim \frac{\mathcal{W}_p(\boldsymbol{\Sigma}_r^*, r-1)}{r-1} \\
\bar{\boldsymbol{x}}_{n-r} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma} &\sim N_p\left(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{r}\right); \quad \boldsymbol{W}^{(n-r)} \,|\, \boldsymbol{\Sigma} \sim \frac{\mathcal{W}_p(\boldsymbol{\Sigma}, n-r-1)}{n-r-1}
\end{aligned}$$

where the last two quantities need the assumption $n - r > p$ for their distributions to be defined. Then the likelihood of the released data is computed as

$$\begin{aligned}
&\int \pi(\bar{\boldsymbol{y}}_r \,|\, \boldsymbol{\mu}_r^*, \boldsymbol{\Sigma}_r^*)\, \pi(\boldsymbol{W}_y^{(r)} \,|\, \boldsymbol{\Sigma}_r^*)\, \pi(\boldsymbol{\mu}_r^* \,|\, \bar{\boldsymbol{x}}_r, \boldsymbol{\Sigma})\, \pi(\boldsymbol{\Sigma}_r^* \,|\, \boldsymbol{W}^{(r)})\, \pi(\bar{\boldsymbol{x}}_r \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{W}^{(r)} \,|\, \boldsymbol{\Sigma}) \\
&\quad \pi(\bar{\boldsymbol{x}}_{n-r} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{W}^{(n-r)} \,|\, \boldsymbol{\Sigma})\; d\boldsymbol{\mu}_r^*\, d\boldsymbol{\Sigma}_r^*\, d\bar{\boldsymbol{x}}_r\, d\boldsymbol{W}^{(r)}
\end{aligned}$$

We integrate out $\boldsymbol{\mu}_r^*$, $\bar{\boldsymbol{x}}_r$, $\boldsymbol{W}^{(r)}$ one by one from the above likelihood and then multiply with our usual prior $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{\frac{\delta}{2}}$ to obtain the following posterior distributions

$$\begin{aligned}
&\boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_r^*, \bar{\boldsymbol{y}}_r, \bar{\boldsymbol{x}}_{n-r} \\
&\sim N_p\Big[\left(r\left(\boldsymbol{\Sigma} + 2\boldsymbol{\Sigma}_r^*\right)^{-1} + (n-r)\boldsymbol{\Sigma}^{-1}\right)^{-1}\left(r\left(\boldsymbol{\Sigma} + 2\boldsymbol{\Sigma}_r^*\right)^{-1}\bar{\boldsymbol{y}}_r + (n-r)\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{x}}_{n-r}\right), \\
&\qquad \left(r\left(\boldsymbol{\Sigma} + 2\boldsymbol{\Sigma}_r^*\right)^{-1} + (n-r)\boldsymbol{\Sigma}^{-1}\right)^{-1}\Big] \qquad (4.27)
\end{aligned}$$

$$
\begin{aligned}
&\pi(\boldsymbol{\Sigma}, \boldsymbol{\Sigma}_r^* \,|\, \bar{\boldsymbol{y}}_r, \boldsymbol{W}_y^{(r)}, \bar{\boldsymbol{x}}_{n-r}, \boldsymbol{W}^{(n-r)}) \\
&\propto \exp\left[-\tfrac{r(n-r)}{2}(\bar{\boldsymbol{y}}_r - \bar{\boldsymbol{x}}_{n-r})'(n\boldsymbol{\Sigma} + 2(n-r)\boldsymbol{\Sigma}_r^*)^{-1}(\bar{\boldsymbol{y}}_r - \bar{\boldsymbol{x}}_{n-r})\right] \\
&\quad |n\boldsymbol{\Sigma} + 2(n-r)\boldsymbol{\Sigma}_r^*|^{-\frac{1}{2}} \, |\boldsymbol{\Sigma}_r^*|^{\frac{2r-2p-4}{2}} \, |\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_r^*|^{-\frac{2r-p+\alpha-3}{2}} \, |\boldsymbol{\Sigma}|^{-\frac{n-2r+p-\alpha+\delta+1}{2}} \\
&\quad \exp\left[-\tfrac{n-r-1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^{(n-r)})\right] \exp\left[-\tfrac{r-1}{2}\operatorname{tr}(\boldsymbol{\Sigma}_r^{*-1}\boldsymbol{W}_y^{(r)})\right] \qquad (4.28)
\end{aligned}
$$

We can check that the results of this section match the case when all responses are sensitive, those obtained in Sections 4.1 and 4.2, by suppressing the quantities $\bar{\boldsymbol{x}}_{n-r}, \boldsymbol{W}^{(n-r)}$; replacing $\boldsymbol{W}_y^{(r)} = \dfrac{\boldsymbol{\mathcal{S}}_y}{n-1}$, $\bar{\boldsymbol{y}}_r = \bar{\boldsymbol{y}}$, $r = n$ and $\boldsymbol{W}_y^{(r)} = \boldsymbol{W}$ (in the PIS case), $\boldsymbol{\Sigma}_r^* = \boldsymbol{\Sigma}^*$ (in the PPS case).

Method II: Using whole data estimates to impute synthetic data

Plug-In Sampling

In this method, the released data is $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r, \boldsymbol{x}_{r+1}, \ldots, \boldsymbol{x}_n)$ where $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r \overset{\text{iid}}{\sim} N_p(\bar{\boldsymbol{x}}, \boldsymbol{W})$. Then the portion of the likelihood of the released data required to calculate the posterior distributions is computed as

$$
\int \left(\prod_{i=1}^{r} \pi(\boldsymbol{y}_i \,|\, \bar{\boldsymbol{x}}, \boldsymbol{W})\right) \pi(\bar{\boldsymbol{x}} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, \pi(\boldsymbol{W} \,|\, \boldsymbol{\Sigma}) \, d\bar{\boldsymbol{x}} \, d\boldsymbol{W}
$$

We integrate out $\bar{\boldsymbol{x}}$ from the above likelihood and then multiply with our usual prior $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{\frac{\delta}{2}}$ to obtain the following posterior distributions

$$
\begin{aligned}
\boldsymbol{W}_y^{(r)^{-1/2}} \boldsymbol{W} \boldsymbol{W}_y^{(r)^{-1/2}} &\sim \mathcal{W}_p^{-1}((r-1)\boldsymbol{I}_p, r-p+\delta-2) && (4.29) \\
\boldsymbol{W}^{-1/2} \boldsymbol{\Sigma} \boldsymbol{W}^{-1/2} &\sim \mathcal{W}_p^{-1}((n-1)\boldsymbol{I}_p, n-p+\delta-2) && (4.30) \\
\boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \boldsymbol{W}, \bar{\boldsymbol{y}}_r &\sim N_p\left(\bar{\boldsymbol{y}}_r, \frac{\boldsymbol{\Sigma}}{n} + \frac{\boldsymbol{W}}{r}\right) && (4.31)
\end{aligned}
$$

where $\boldsymbol{W}_y^{(r)} = \sum_{i=1}^{r}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_r)(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_r)'$, which is equivalent to $\boldsymbol{\mathcal{S}}_y$ when $r = n$.

The distributions are proper as long as $n > p$, $r > 2p - \delta + 1$.

Posterior Predictive Sampling

We draw $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ from (4.14) so that the released data is $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_r, \boldsymbol{x}_{r+1}, \ldots, \boldsymbol{x}_n)$ where $\boldsymbol{y}_i \overset{\text{iid}}{\sim} N_p(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ for $i = 1, \ldots, r$. Then the portion of the likelihood of the released data required to calculate the posterior distributions is computed as

$$\int \left(\prod_{i=1}^{r} \pi(\boldsymbol{y}_i \,|\, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) \right) \pi(\boldsymbol{\mu}^* \,|\, \bar{\boldsymbol{x}}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{\Sigma}^* \,|\, \boldsymbol{W})\, \pi(\bar{\boldsymbol{x}} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{W} \,|\, \boldsymbol{\Sigma})\, d\boldsymbol{\mu}^*\, d\boldsymbol{\Sigma}^*\, d\bar{\boldsymbol{x}}\, d\boldsymbol{W}$$

We integrate out $\boldsymbol{\mu}^*$, $\bar{\boldsymbol{x}}$, $\boldsymbol{W}$ one by one from the above likelihood and then multiply with our usual prior $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{\frac{\delta}{2}}$ to obtain the following posterior distributions

$$\begin{aligned}
\boldsymbol{\mathcal{S}}_{y,r}^{-1/2} \boldsymbol{\Sigma}^* \boldsymbol{\mathcal{S}}_{y,r}^{-1/2} &\sim \mathcal{W}_p^{-1}\left(\boldsymbol{I}_p, r - 2p + \delta - 1\right) && (4.32)\\
\boldsymbol{\Sigma}^{*-1/2} \boldsymbol{\Sigma}\, \boldsymbol{\Sigma}^{*-1/2} &\sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}\left(\frac{n + \alpha - \delta - 1}{2}, \frac{n - p + \delta - 2}{2}\right) && (4.33)\\
\boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}^*, \bar{\boldsymbol{y}}_r &\sim N_p\left(\bar{\boldsymbol{y}}_r, \frac{1}{n}\left(\boldsymbol{\Sigma} + \left(1 + \frac{n}{r}\right)\boldsymbol{\Sigma}^*\right)\right) && (4.34)
\end{aligned}$$

where $\boldsymbol{\mathcal{S}}_{y,r} = \sum_{i=1}^{r}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_r)(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_r)'$, which is equivalent to $\boldsymbol{\mathcal{S}}_y$ when $r = n$.

The conditions for existence are $r > 3p - \delta$, $n > \max\{p, 2p - \alpha + 1, p - \alpha + \delta, 2p - \delta + 1\}$.

All the conditions for existence throughout this work can also be expressed as inequalities for δ, since once we have the data at hand, that would enable us to choose a proper value of δ to get the best inference.

Chapter 5

Bayesian Analysis of Multiply Imputed Synthetic Data under the Multivariate Normal Model

5.1 Plug In Sampling method

We return to the case of a standard MVN model where the original data has the same structure as in Section 4.1. The multiply imputed synthetic data, denoted by $(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m)$, are obtained by drawing

$$\boldsymbol{Y}_j = (\boldsymbol{y}_{1j}, \ldots, \boldsymbol{y}_{nj}) \,|\, \boldsymbol{X} \overset{\text{iid}}{\sim} N_p\left(\bar{\boldsymbol{x}},\ \boldsymbol{W}\right) \quad \text{independently for } j = 1, \ldots, m \tag{5.1}$$

We will now proceed to calculate the pdf of the multiply imputed synthetic data, for which we will need to define a few quantities. Let $\bar{\boldsymbol{y}}_j = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_{ij}$, $\overline{\boldsymbol{y}^*} = \frac{1}{m}\sum_{j=1}^{m} \bar{\boldsymbol{y}}_j$ and $\boldsymbol{S}_{y,j} = \sum_{i=1}^{n}(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_j)(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_j)'$.

Likelihood of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Using equation (4.3) and the fact that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$

$$
\begin{aligned}
&f(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m \,|\, \bar{\boldsymbol{x}}, \boldsymbol{W}) = \prod_{j=1}^{m} f(\boldsymbol{Y}_j \,|\, \bar{\boldsymbol{x}}, \boldsymbol{W}) = \prod_{j=1}^{m} f(\boldsymbol{y}_{1j}, \ldots, \boldsymbol{y}_{nj} \,|\, \bar{\boldsymbol{x}}, \boldsymbol{W}) \\
&\propto \prod_{j=1}^{m} \frac{|\boldsymbol{\mathcal{S}}_{y,j}|^{(n-p-2)/2}}{|\boldsymbol{W}|^{n/2}} \exp\left\{ -\frac{1}{2} \left[n(\bar{\boldsymbol{y}}_j - \bar{\boldsymbol{x}})' \boldsymbol{W}^{-1} (\bar{\boldsymbol{y}}_j - \bar{\boldsymbol{x}}) + \mathrm{tr}(\boldsymbol{W}^{-1} \boldsymbol{\mathcal{S}}_{y,j}) \right] \right\} \\
&= |\boldsymbol{W}|^{-nm/2} \left(\prod_{j=1}^{m} |\boldsymbol{\mathcal{S}}_{y,j}|^{(n-p-2)/2} \right) \\
&\times \exp\left\{ -\frac{1}{2} \left[n \sum_{j=1}^{m} (\bar{\boldsymbol{y}}_j - \bar{\boldsymbol{x}})' \boldsymbol{W}^{-1} (\bar{\boldsymbol{y}}_j - \bar{\boldsymbol{x}}) + \mathrm{tr}(\boldsymbol{W}^{-1} \sum_{j=1}^{m} \boldsymbol{\mathcal{S}}_{y,j}) \right] \right\} \\
&= |\boldsymbol{W}|^{-nm/2} \left(\prod_{j=1}^{m} |\boldsymbol{\mathcal{S}}_{y,j}|^{(n-p-2)/2} \right) \exp\left\{ -\frac{1}{2} \left[nm(\overline{\boldsymbol{y}^*} - \bar{\boldsymbol{x}})' \boldsymbol{W}^{-1} (\overline{\boldsymbol{y}^*} - \bar{\boldsymbol{x}}) + \mathrm{tr}(\boldsymbol{W}^{-1} \boldsymbol{\mathcal{S}}_y^*) \right] \right\}
\end{aligned}
\tag{5.2}
$$

where $\boldsymbol{\mathcal{S}}_y^* = \sum_{j=1}^{m} \boldsymbol{\mathcal{S}}_{y,j} + n \sum_{j=1}^{m} (\bar{\boldsymbol{y}}_j - \overline{\boldsymbol{y}^*})(\bar{\boldsymbol{y}}_j - \overline{\boldsymbol{y}^*})'$, whose distribution can be computed as follows. $\boldsymbol{\mathcal{S}}_{y,j}$ is the sample variance of iid r.v.'s $\boldsymbol{y}_{1j}, \ldots, \boldsymbol{y}_{nj}$ and thus $\boldsymbol{\mathcal{S}}_{y,j} \overset{\text{iid}}{\sim} \mathcal{W}_p(\boldsymbol{W}, n-1)$ for $j = 1, \ldots, m$, hence $\sum_{j=1}^{m} \boldsymbol{\mathcal{S}}_{y,j} \sim \mathcal{W}_p(\boldsymbol{W}, m(n-1))$. Similarly, it can be shown that $n \sum_{j=1}^{m} (\bar{\boldsymbol{y}}_j - \overline{\boldsymbol{y}^*})(\bar{\boldsymbol{y}}_j - \overline{\boldsymbol{y}^*})' \sim \mathcal{W}_p(\boldsymbol{W}, m-1)$. Now since $\mathrm{Cov}(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_j, \bar{\boldsymbol{y}}_j - \overline{\boldsymbol{y}^*}) = 0$, all of the terms in the expression of the sum that is $\boldsymbol{\mathcal{S}}_y^*$ are independent and $\boldsymbol{\mathcal{S}}_y^* \sim \mathcal{W}_p(\boldsymbol{W}, mn-1)$. The final expression in equation 5.2 also proves that given $\bar{\boldsymbol{x}}$ and $\boldsymbol{W}$, $\overline{\boldsymbol{y}^*}$ is independent of $\boldsymbol{\mathcal{S}}_y^*$, and $(\overline{\boldsymbol{y}^*}, \boldsymbol{\mathcal{S}}_y^*)$ is jointly sufficient for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The simplification in the last step comes from the following calculation:

$$
\begin{aligned}
&n \sum_{j=1}^{m} (\bar{\boldsymbol{y}}_j - \bar{\boldsymbol{x}})' \boldsymbol{W}^{-1} (\bar{\boldsymbol{y}}_j - \bar{\boldsymbol{x}}) - nm(\overline{\boldsymbol{y}^*} - \bar{\boldsymbol{x}})' \boldsymbol{W}^{-1} (\overline{\boldsymbol{y}^*} - \bar{\boldsymbol{x}}) \\
&= n \sum_{j=1}^{m} \bar{\boldsymbol{y}}_j' \boldsymbol{W}^{-1} \bar{\boldsymbol{y}}_j - 2n \left(\sum_{j=1}^{m} \bar{\boldsymbol{y}}_j' \right) \boldsymbol{W}^{-1} \bar{\boldsymbol{x}} + nm \bar{\boldsymbol{x}}' \boldsymbol{W}^{-1} \bar{\boldsymbol{x}} \\
&- nm \overline{\boldsymbol{y}^*}' \boldsymbol{W}^{-1} \overline{\boldsymbol{y}^*} + 2nm \overline{\boldsymbol{y}^*}' \boldsymbol{W}^{-1} \bar{\boldsymbol{x}} - nm \bar{\boldsymbol{x}}' \boldsymbol{W}^{-1} \bar{\boldsymbol{x}}
\end{aligned}
$$

$$
\begin{aligned}
&= n\left[\sum_{j=1}^{m} \bar{\boldsymbol{y}}_j' \boldsymbol{W}^{-1} \bar{\boldsymbol{y}}_j - m\overline{\boldsymbol{y}^*}' \boldsymbol{W}^{-1} \overline{\boldsymbol{y}^*}\right] \\
&= n\left[\sum_{j=1}^{m} (\bar{\boldsymbol{y}}_j - \overline{\boldsymbol{y}^*})' \boldsymbol{W}^{-1} (\bar{\boldsymbol{y}}_j - \overline{\boldsymbol{y}^*})\right] \\
&= n\left[\sum_{j=1}^{m} \operatorname{tr}((\bar{\boldsymbol{y}}_j - \overline{\boldsymbol{y}^*})' \boldsymbol{W}^{-1} (\bar{\boldsymbol{y}}_j - \overline{\boldsymbol{y}^*}))\right] = n \operatorname{tr}(\boldsymbol{W}^{-1} \sum_{j=1}^{m} (\bar{\boldsymbol{y}}_j - \overline{\boldsymbol{y}^*})(\bar{\boldsymbol{y}}_j - \overline{\boldsymbol{y}^*})')
\end{aligned}
$$

We now combine the terms involving $\bar{\boldsymbol{x}}$ from the two exponents of equations (5.2) and (4.4) as

$$
\begin{aligned}
& m(\overline{\boldsymbol{y}^*} - \bar{\boldsymbol{x}})' \boldsymbol{W}^{-1} (\overline{\boldsymbol{y}^*} - \bar{\boldsymbol{x}}) + (\bar{\boldsymbol{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu}) \\
&= \left\{\bar{\boldsymbol{x}} - [m\boldsymbol{W}^{-1} + \boldsymbol{\Sigma}^{-1}]^{-1} [m\boldsymbol{W}^{-1} \overline{\boldsymbol{y}^*} + \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}]\right\}' \left\{m\boldsymbol{W}^{-1} + \boldsymbol{\Sigma}^{-1}\right\} \\
&\quad \left\{\bar{\boldsymbol{x}} - [m\boldsymbol{W}^{-1} + \boldsymbol{\Sigma}^{-1}]^{-1} [m\boldsymbol{W}^{-1} \overline{\boldsymbol{y}^*} + \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}]\right\} \\
&- \left\{m\boldsymbol{W}^{-1} \overline{\boldsymbol{y}^*} + \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}\right\}' \left\{m\boldsymbol{W}^{-1} + \boldsymbol{\Sigma}^{-1}\right\}^{-1} \left\{m\boldsymbol{W}^{-1} \overline{\boldsymbol{y}^*} + \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}\right\} + m\overline{\boldsymbol{y}^*}' \boldsymbol{W}^{-1} \overline{\boldsymbol{y}^*} + \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\
&= \left\{\bar{\boldsymbol{x}} - [m\boldsymbol{W}^{-1} + \boldsymbol{\Sigma}^{-1}]^{-1} [m\boldsymbol{W}^{-1} \overline{\boldsymbol{y}^*} + \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}]\right\}' \left\{m\boldsymbol{W}^{-1} + \boldsymbol{\Sigma}^{-1}\right\} \\
&\quad \left\{\bar{\boldsymbol{x}} - [m\boldsymbol{W}^{-1} + \boldsymbol{\Sigma}^{-1}]^{-1} [m\boldsymbol{W}^{-1} \overline{\boldsymbol{y}^*} + \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}]\right\} + (\overline{\boldsymbol{y}^*} - \boldsymbol{\mu})' \left(\boldsymbol{\Sigma} + \frac{\boldsymbol{W}}{m}\right)^{-1} (\overline{\boldsymbol{y}^*} - \boldsymbol{\mu})
\end{aligned}
$$

where again the last step uses the result (2.28). Now integrating out $\bar{\boldsymbol{x}}$ from the product of the two pdfs of equations (5.2) and (4.4), we arrive at the following result.

Theorem 5.1.1. *The joint pdf of* $(\overline{\boldsymbol{y}^*}, \boldsymbol{\mathcal{S}}_y^*)$ *is given by*

$$
f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\overline{\boldsymbol{y}^*}, \boldsymbol{\mathcal{S}}_y^*) \propto \int_{S_n^{++}} \frac{\left|\boldsymbol{\Sigma} + \frac{\boldsymbol{W}}{m}\right|^{-\frac{1}{2}}}{|\boldsymbol{\Sigma}|^{\frac{n-1}{2}} |\boldsymbol{W}|^{\frac{nm-n+p+1}{2}}} \left(\prod_{j=1}^{m} |\boldsymbol{\mathcal{S}}_{y,j}|^{(n-p-2)/2}\right) \\
e^{-\frac{1}{2}\left[n\left(\overline{\boldsymbol{y}^*} - \boldsymbol{\mu}\right)' \left(\boldsymbol{\Sigma} + \frac{\boldsymbol{W}}{m}\right)^{-1} \left(\overline{\boldsymbol{y}^*} - \boldsymbol{\mu}\right) + \operatorname{tr}(\boldsymbol{\mathcal{S}}_y^* \boldsymbol{W}^{-1}) + (n-1)\operatorname{tr}(\boldsymbol{W}\boldsymbol{\Sigma}^{-1})\right]} d\boldsymbol{W}
$$

Posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

We choose the same prior on the parameters and follow the same procedure as in Section 4.1 to get the following (conditional) posterior distributions:

$$\begin{aligned} \boldsymbol{W} \,|\, \boldsymbol{\mathcal{S}}_y^* &\sim \mathcal{W}_p^{-1}\left(\boldsymbol{\mathcal{S}}_y^*, nm - p + \delta - 2\right) && (5.3)\\ \boldsymbol{\Sigma} \,|\, \boldsymbol{W} &\sim \mathcal{W}_p^{-1}\left((n-1)\boldsymbol{W}, n - p + \delta - 2\right) && (5.4)\\ \boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \boldsymbol{W}, \overline{\boldsymbol{y}^*} &\sim N_p\left(\overline{\boldsymbol{y}^*}, \frac{1}{n}\left(\boldsymbol{\Sigma} + \frac{\boldsymbol{W}}{m}\right)\right) && (5.5) \end{aligned}$$

We can reformulate the above posterior distributions as:

$$\begin{aligned} \boldsymbol{\mathcal{S}}_y^{*-1/2}\boldsymbol{W}\boldsymbol{\mathcal{S}}_y^{*-1/2} &\sim \mathcal{W}_p^{-1}\left(\boldsymbol{I}_p, nm - p + \delta - 2\right) && (5.6)\\ \boldsymbol{W}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{W}^{-1/2} &\sim \mathcal{W}_p^{-1}\left((n-1)\boldsymbol{I}_p, n - p + \delta - 2\right) && (5.7)\\ \boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \boldsymbol{W}, \overline{\boldsymbol{y}^*} &\sim N_p\left(\overline{\boldsymbol{y}^*}, \frac{1}{n}\left(\boldsymbol{\Sigma} + \frac{\boldsymbol{W}}{m}\right)\right) && (5.8) \end{aligned}$$

which has again the same benefit that $\boldsymbol{\mathcal{S}}_y^{*-1/2}\boldsymbol{W}\boldsymbol{\mathcal{S}}_y^{*-1/2}$ is independent of $\boldsymbol{W}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{W}^{-1/2}$ and their posterior distributions are unconditional. It is interesting to observe that plugging in $m = 1$ in the above formulas yields the same results we obtained for singly imputed plug-in sampling data as in Section 4.1.

The posterior distributions are proper as long as $n > \max\{p, 2p - \delta + 1\}$.

Bayes Estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

$$\begin{aligned} \hat{\boldsymbol{\mu}}_{\text{BAYES}} &= \mathrm{E}(\boldsymbol{\mu} \,|\, \overline{\boldsymbol{y}^*}, \boldsymbol{\mathcal{S}}_y^*) = \mathrm{E}_{\boldsymbol{W}}\,\mathrm{E}_{\boldsymbol{\Sigma}}\,\mathrm{E}(\boldsymbol{\mu} \,|\, \overline{\boldsymbol{y}^*}, \boldsymbol{\Sigma}, \boldsymbol{W}) = \mathrm{E}_{\boldsymbol{W}}\,\mathrm{E}_{\boldsymbol{\Sigma}}\,\mathrm{E}(\overline{\boldsymbol{y}^*}) = \overline{\boldsymbol{y}^*}\\ \hat{\boldsymbol{\Sigma}}_{\text{BAYES}} &= \mathrm{E}(\boldsymbol{\Sigma} \,|\, \overline{\boldsymbol{y}^*}, \boldsymbol{\mathcal{S}}_y^*) = \mathrm{E}_{\boldsymbol{W}}\,\mathrm{E}(\boldsymbol{\Sigma} \,|\, \boldsymbol{\mathcal{S}}_y^*, \boldsymbol{W}) = \mathrm{E}_{\boldsymbol{W}}\left(\tfrac{(n-1)\boldsymbol{W}}{(n-2p+\delta-3)} \,|\, \boldsymbol{\mathcal{S}}_y^*\right)\\ &= \frac{(n-1)\boldsymbol{\mathcal{S}}_y^*}{(nm - 2p + \delta - 3)(n - 2p + \delta - 3)} \end{aligned}$$

$$\widehat{|\boldsymbol{\Sigma}|}_{\text{BAYES}} = \mathrm{E}\left(|\boldsymbol{\Sigma}| \mid \overline{\boldsymbol{y}^*}, \boldsymbol{\mathcal{S}}_y^*\right) = \mathrm{E}_{\boldsymbol{W}}\, \mathrm{E}\left(|\boldsymbol{\Sigma}| \mid \boldsymbol{\mathcal{S}}_y^*, \boldsymbol{W}\right) = \mathrm{E}_{\boldsymbol{W}}\left(|\boldsymbol{W}|\, \mathrm{E}\left(\left|\boldsymbol{W}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{W}^{-1/2}\right|\right) \mid \boldsymbol{\mathcal{S}}_y^*\right)$$
$$= \left(\prod_{j=1}^{p} \frac{n-1}{n-p+\delta-j-3}\right) \mathrm{E}\left(|\boldsymbol{W}| \mid \boldsymbol{\mathcal{S}}_y^*\right) = \left(\prod_{j=1}^{p} \frac{n-1}{(nm-2p+\delta-3)(n-2p+\delta-3)}\right) \left|\boldsymbol{\mathcal{S}}_y^*\right|$$

provided $n > \max\{p, 2p-\delta+5\}$, and we use the results (4.11) and (4.12).

Credible Sets for $|\boldsymbol{\Sigma}|$ and $\boldsymbol{\mu}$

In the same way as in Section 4.1 we can define a pivot for $|\boldsymbol{\Sigma}|$ as $N_m \coloneqq \left|\boldsymbol{\Sigma}\boldsymbol{\mathcal{S}}_y^{-1}\right|$ where

$$N_m^{-1} \sim \frac{\prod_{i=1}^{p} u_i}{(n-1)^p} \prod_{j=1}^{p} v_j$$

where $u_i \sim \chi^2_{n-p+\delta-i-1}$ independently for $i = 1, \ldots, p$; $v_j \sim \chi^2_{nm-p+\delta-j-1}$ independently for $j = 1, \ldots, p$; and they are all pairwise independent. A $(1-\gamma)$ level credible set for $|\boldsymbol{\Sigma}|$ based on N_m is

$$\left[a_{n,p,m,\delta;\gamma} \left|\boldsymbol{\mathcal{S}}_y\right|, b_{n,p,m,\delta;\gamma} \left|\boldsymbol{\mathcal{S}}_y\right|\right]$$

where $a_{n,p,m,\delta;\gamma}$ and $b_{n,p,m,\delta;\gamma}$ are any two constants that satisfy $1-\gamma = P(a_{n,p,m,\delta;\gamma} \leq N_m \leq b_{n,p,m,\delta;\gamma})$. The length of the credible interval is $\left|\boldsymbol{\mathcal{S}}_y\right| (b_{n,p,m,\delta;\gamma} - a_{n,p,m,\delta;\gamma})$.

Next we define the pivot for $\boldsymbol{\mu}$ as

$$T_m^2 \coloneqq n(\boldsymbol{\mu} - \overline{\boldsymbol{y}^*})' \boldsymbol{\mathcal{S}}_y^{*-1} (\boldsymbol{\mu} - \overline{\boldsymbol{y}^*})$$

We will derive the distribution of T_m^2 similarly as in Section 4.1. We notice that

$$\sqrt{n}\boldsymbol{\mathcal{S}}_y^{*-1/2}(\boldsymbol{\mu} - \overline{\boldsymbol{y}^*}) \mid \boldsymbol{\Sigma}, \boldsymbol{W} \sim N_p(\boldsymbol{0}, \boldsymbol{A})$$

where $\boldsymbol{A} = \boldsymbol{S}_y^{*-1/2}\boldsymbol{W}^{1/2}(\boldsymbol{W}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{W}^{-1/2} + m^{-1}\boldsymbol{I}_p)\boldsymbol{W}^{1/2}\boldsymbol{S}_y^{*-1/2}$ which is obviously defined through the parameters $(\boldsymbol{\Sigma}, \boldsymbol{W})$. If we can prove that the distribution of $\boldsymbol{A}$ is free of $(\boldsymbol{\Sigma}, \boldsymbol{W})$, then by using the fact that if $\boldsymbol{Z} \sim N_p(\boldsymbol{0}, \boldsymbol{A})$ then $\boldsymbol{Z}'\boldsymbol{Z} \sim \sum_{i=1}^{p} \lambda_i \chi_{1i}^2$ where $\lambda_1, \ldots, \lambda_p$ are the eigenvalues of $\boldsymbol{A}$ and χ_{1i}^2 are independent χ_1^2 variables, we can conclude that T^2 is a pivot. Taking $\boldsymbol{Z} = \sqrt{n}\boldsymbol{S}_y^{*-1/2}(\boldsymbol{\mu} - \overline{\boldsymbol{y}^*})$, $\boldsymbol{B} = \boldsymbol{S}_y^{*-1/2}\boldsymbol{W}\boldsymbol{S}_y^{*-1/2}$ it finally follows that:

(a) the conditional distribution of $T^2 \,|\, \boldsymbol{A}$ is $\sum_{i=1}^{p} \lambda_i \chi_{1i}^2$ where $\lambda_1, \ldots, \lambda_p$ are the roots of $|\boldsymbol{A} - \lambda \boldsymbol{I}_p| = 0$ such that $\boldsymbol{A} \,|\, \boldsymbol{B} \stackrel{d}{=} \mathcal{W}_p^{-1}\left((n-1)\boldsymbol{B}, n - p + \delta - 2\right) + m^{-1}\boldsymbol{B}$ by (5.7) and $\boldsymbol{B} \sim \mathcal{W}_p^{-1}\left(\boldsymbol{I}_p, nm - p + \delta - 2\right)$ by (5.6); and

(b) the unconditional distribution of T^2 is obtained by averaging over the joint distribution of the roots $\lambda_1, \ldots, \lambda_p$.

We have shown that T^2 is a pivotal quantity, and therefore a $(1-\gamma)$ credible ellipsoid for $\boldsymbol{\mu}$ based on T^2 is given by

$$\left\{\boldsymbol{\mu} \;:\; T^2 \leq c_{n,p,m,\delta;\gamma}\right\}$$

where $c_{n,p,m,\delta;\gamma}$ satisfies $1 - \gamma = P(T^2 \leq c_{n,p,m,\delta;\gamma})$. From the above discussion, it follows that the cut-off point $c_{n,p,m,\delta;\gamma}$ can be obtained by simulating the distribution of T^2 as follows.

1. Generate $\boldsymbol{B} \sim \mathcal{W}_p^{-1}\left(\boldsymbol{I}_p, nm - p + \delta - 2\right)$.
2. Generate $\boldsymbol{A} \,|\, \boldsymbol{B} \sim \mathcal{W}_p^{-1}\left((n-1)\boldsymbol{B}, n - p + \delta - 2\right) + m^{-1}\boldsymbol{B}$.
3. Generate $\lambda_1, \ldots, \lambda_p$, the roots of $|\boldsymbol{A} - \lambda \boldsymbol{I}_p| = 0$.
4. Generate $T^2 = \sum_{i=1}^{p} \lambda_i \chi_{1i}^2$ where χ_{1i}^2 are independent χ_1^2 variables.

The volume of the credible ellipsoid is given by

$$V_{\boldsymbol{\mu}}(\boldsymbol{Y}) = \frac{\pi^{p/2}}{\Gamma\left(\frac{p}{2}+1\right)} \left(c_{n,p,m,\delta;\gamma}/n\right)^{p/2} \left|\boldsymbol{S}_y^*\right|^{1/2}$$

Simulation Studies We have the same setup as in the last chapter, and we notice similar behavior, except as m is big inference is obviously better, but there is an increased perceived disclosure risk, which again points to the trade-off between quality of inference and privacy. We also see that asymptotically the coverage for μ is unaffected by increase in δ. As we increase m, coverage improves, which is apparent for higher δ when we compare to the case of $m = 1$.

Table 5.1: Inference for $\boldsymbol{\mu}$ and $|\boldsymbol{\Sigma}|$ for MI PIS MVN data with $m = 5$, $n = 1000$

	$\lvert\boldsymbol{\Sigma}\rvert$		$\boldsymbol{\mu}$	
	avg	*est*	*avg*	*est*
δ	*cvg*	*len*	*cvg*	*vol*
0.2	0.8608	2.0972e-05	0.9532	7.7485e-11
0.5	0.8654	2.1152e-05	0.9576	8.083e-11
0.8	0.8714	2.1071e-05	0.9536	7.7724e-11
1	0.8769	2.0950e-05	0.9524	7.4056e-11
2	0.8968	2.1075e-05	0.9549	8.3019e-11
3	0.901	2.0453e-05	0.9486	7.3772e-11
4	0.9094	2.0498e-05	0.9528	7.7908e-11
5	0.9271	2.0013e-05	0.9491	7.3588e-11
10	0.9473	1.8518e-05	0.9498	7.6129e-11
20	0.8914	1.6498e-05	0.9533	7.5433e-11
30	0.6842	1.4602e-05	0.9388	6.4412e-11
50	0.1581	1.1732e-05	0.9441	6.3722e-11
100	0	6.2396e-06	0.9218	4.9502e-11

Table 5.2: Inference for $\boldsymbol{\mu}$ and $|\boldsymbol{\Sigma}|$ for MI PIS MVN data with $m = 5$, $n = 10000$

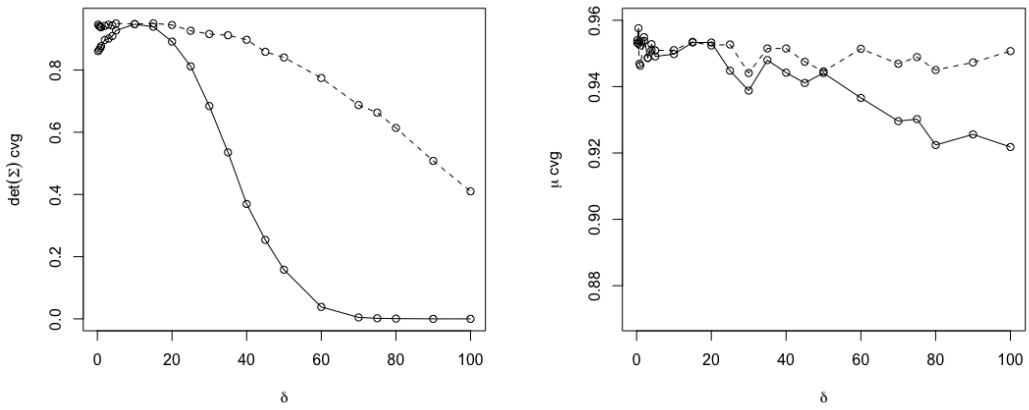
	$\lvert\boldsymbol{\Sigma}\rvert$		$\boldsymbol{\mu}$	
	avg	*est*	*avg*	*est*
δ	*cvg*	*len*	*cvg*	*vol*
0.2	0.946	5.8077e-06	0.954	7.5415e-16
0.5	0.9412	5.7425e-06	0.9529	7.4118e-16
0.8	0.9392	5.6669e-06	0.9469	6.9455e-16
1	0.9372	5.7574e-06	0.9463	7.4841e-16
2	0.9419	5.7320e-06	0.9538	7.5244e-16
3	0.9457	5.7057e-06	0.9487	6.7614e-16
4	0.9432	5.7221e-06	0.951	7.3456e-16
5	0.9496	5.7975e-06	0.9509	7.159e-16
10	0.9478	5.624e-06	0.951	7.437e-16
20	0.9448	5.6056e-06	0.9524	7.3618e-16
30	0.9158	5.4522e-06	0.9441	6.7509e-16
50	0.8404	5.4233e-06	0.9446	6.5821e-16
100	0.4101	5.0887e-06	0.9507	6.6923e-16

Table 5.3: Inference for $\boldsymbol{\mu}$ and $|\boldsymbol{\Sigma}|$ for MI PIS MVN data with $m = 10$, $n = 1000$

	$\lvert\boldsymbol{\Sigma}\rvert$		$\boldsymbol{\mu}$	
	avg	*est*	*avg*	*est*
δ	*cvg*	*len*	*cvg*	*vol*
0.2	0.8796	1.9733e-05	0.9501	4.9617e-11
0.5	0.8708	1.9997e-05	0.9498	5.0604e-11
0.8	0.8827	1.9946e-05	0.9553	5.2025e-11
1	0.8763	1.9716e-05	0.9528	4.9584e-11
2	0.8945	1.9689e-05	0.9581	5.2159e-11
3	0.9162	1.9539e-05	0.9494	4.8426e-11
4	0.9225	1.9659e-05	0.9532	4.9263e-11
5	0.923	1.8594e-05	0.9551	5.2248e-11
10	0.9495	1.8103e-05	0.9469	4.6541e-11
20	0.8951	1.5746e-05	0.952	4.7451e-11
30	0.7074	1.4186e-05	0.9459	4.4338e-11
50	0.1661	1.1251e-05	0.9379	3.9822e-11
100	0	6.7094e-06	0.919	3.0661e-11

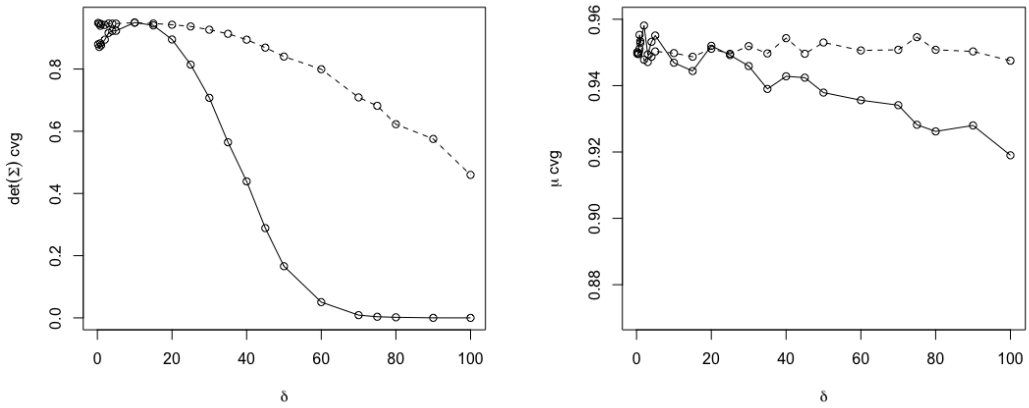
Table 5.4: Inference for $\boldsymbol{\mu}$ and $|\boldsymbol{\Sigma}|$ for MI PIS MVN data with $m = 10$, $n = 10000$

	$\lvert\boldsymbol{\Sigma}\rvert$		$\boldsymbol{\mu}$	
	avg	*est*	*avg*	*est*
δ	*cvg*	*len*	*cvg*	*vol*
0.2	0.9482	5.6025e-06	0.9496	4.6263e-16
0.5	0.9456	5.5416e-06	0.9495	4.4142e-16
0.8	0.939	5.4727e-06	0.9511	4.5672e-16
1	0.9445	5.486e-06	0.9534	4.8759e-16
2	0.9417	5.405e-06	0.9478	4.6308e-16
3	0.9463	5.4666e-06	0.9471	4.3834e-16
4	0.9458	5.4953e-06	0.9487	4.5177e-16
5	0.946	5.4846e-06	0.9503	4.6882e-16
10	0.949	5.4524e-06	0.9498	4.3575e-16
20	0.9425	5.3725e-06	0.9511	4.7028e-16
30	0.9264	5.3382e-06	0.9519	4.8632e-16
50	0.8401	5.1636e-06	0.953	4.6459e-16
100	0.4597	4.9643e-06	0.9475	4.5159e-16



(a) $|\boldsymbol{\Sigma}|$, $m = 5$

(b) $\boldsymbol{\mu}$, $m = 5$



(c) $|\boldsymbol{\Sigma}|$, $m = 10$

(d) $\boldsymbol{\mu}$, $m = 10$

Figure 5.1: Variation in coverage of $\boldsymbol{\mu}$ and $|\boldsymbol{\Sigma}|$ with respect to δ for MI PIS MVN data (— $n = 1000$, - - - $n = 10000$)

Extensions of the Methodology to Other Scenarios

Only Part of $\boldsymbol{y}$ is Sensitive

Method I: Using only estimates of sensitive part to impute synthetic data

With the same notation as in Section 4.4, we impute m-copies of the sensitive part of the data as follows, independently for $j = 1, \ldots, m$

$$\boldsymbol{y}_{1j}, \ldots, \boldsymbol{y}_{rj} \overset{\text{iid}}{\sim} N_p\left(\bar{\boldsymbol{x}}_r, \boldsymbol{W}^{(r)}\right)$$

so that the released data is $\{\boldsymbol{Y}_j = (\boldsymbol{y}_{1j}, \ldots, \boldsymbol{y}_{rj}, \boldsymbol{x}_{r+1}, \ldots, \boldsymbol{x}_n) : j = 1, \ldots, m\}$. Assuming both $r > p$ and $n - r > p$, the sufficient statistics for the released data is given by $(\bar{\boldsymbol{y}}_1^{(r)}, \boldsymbol{W}_{y,1}^{(r)}, \ldots, \bar{\boldsymbol{y}}_m^{(r)}, \boldsymbol{W}_{y,m}^{(r)}, \bar{\boldsymbol{x}}_{n-r}, \boldsymbol{W}^{(n-r)})$ where independently for $j = 1, \ldots, m$

$$\begin{aligned}
\bar{\boldsymbol{y}}_j^{(r)} &= \frac{1}{r}\sum_{i=1}^{r} \boldsymbol{y}_{ij}, \quad \bar{\boldsymbol{y}}_j^{(r)} \,|\, \bar{\boldsymbol{x}}_r, \boldsymbol{W}^{(r)} \sim N_p\left(\bar{\boldsymbol{x}}_r, \frac{\boldsymbol{W}^{(r)}}{r}\right) \\
\boldsymbol{W}_{y,j}^{(r)} &= \frac{1}{r-1}\sum_{i=1}^{r}(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_j^{(r)})(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_j^{(r)})', \quad \boldsymbol{W}_{y,j}^{(r)} \,|\, \boldsymbol{W}^{(r)} \sim \frac{\mathcal{W}_p(\boldsymbol{W}^{(r)}, r-1)}{r-1} \\
\bar{\boldsymbol{x}}_{n-r} &= \frac{1}{n-r}\sum_{i=r+1}^{n} \boldsymbol{x}_i \sim N_p\left(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{n-r}\right) \\
\boldsymbol{W}^{(n-r)} &= \frac{1}{n-r-1}\sum_{i=r+1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{n-r})(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{n-r})' \sim \frac{\mathcal{W}_p(\boldsymbol{\Sigma}, n-r-1)}{n-r-1}
\end{aligned}$$

Let us denote

$$\overline{\boldsymbol{y}_r^*} = \frac{1}{m}\sum_{j=1}^{m} \bar{\boldsymbol{y}}_j^{(r)}$$

Then the likelihood of the released data is computed as

$$\int \left(\prod_{j=1}^{m} \pi(\bar{\boldsymbol{y}}_j^{(r)} \,|\, \bar{\boldsymbol{x}}_r, \boldsymbol{W}^{(r)})\, \pi(\boldsymbol{W}_{y,j}^{(r)} \,|\, \boldsymbol{W}^{(r)})\right) \pi(\bar{\boldsymbol{x}}_r \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{W}^{(r)} \,|\, \boldsymbol{\Sigma})\, \pi(\bar{\boldsymbol{x}}_{n-r} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma})$$
$$\pi(\boldsymbol{W}^{(n-r)} \,|\, \boldsymbol{\Sigma})\, d\bar{\boldsymbol{x}}_r\, d\boldsymbol{W}^{(r)}$$

We integrate out $\bar{\boldsymbol{x}}_r$ from the above likelihood and then multiply with our usual prior $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{\frac{\delta}{2}}$ to obtain the following posterior distributions

$$
\begin{aligned}
&\boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \overline{\boldsymbol{y}_r^*}, \boldsymbol{W}^{(r)}, \bar{\boldsymbol{x}}_{n-r} \\
&\sim N_p \left[\left(r \left(\boldsymbol{\Sigma} + \frac{\boldsymbol{W}^{(r)}}{m} \right)^{-1} + m(n-r)\boldsymbol{\Sigma}^{-1} \right)^{-1} \left(r \left(\boldsymbol{\Sigma} + \frac{\boldsymbol{W}^{(r)}}{m} \right)^{-1} \overline{\boldsymbol{y}_r^*} + m(n-r)\boldsymbol{\Sigma}^{-1} \bar{\boldsymbol{x}}_{n-r} \right), \right. \\
&\qquad \left. \left(r \left(\boldsymbol{\Sigma} + \frac{\boldsymbol{W}^{(r)}}{m} \right)^{-1} + m(n-r)\boldsymbol{\Sigma}^{-1} \right)^{-1} \right] \qquad (5.9)
\end{aligned}
$$

$$
\begin{aligned}
&\pi \left(\boldsymbol{\Sigma}, \boldsymbol{W}^{(r)} \,|\, \overline{\boldsymbol{y}_r^*}, \boldsymbol{W}_{y,1}^{(r)}, \ldots, \boldsymbol{W}_{y,m}^{(r)}, \bar{\boldsymbol{x}}_{n-r}, \boldsymbol{W}^{(n-r)} \right) \\
&\propto \left| (m(n-r)+r)\boldsymbol{\Sigma} + (n-r)\boldsymbol{W}^{(r)} \right|^{-\frac{1}{2}} \left| \boldsymbol{W}^{(r)} \right|^{-\frac{r(m-1)+p+1}{2}} |\boldsymbol{\Sigma}|^{-\frac{m(n-r)+r+\delta-2}{2}} \\
&\quad \exp \left[-\tfrac{mr(n-r)}{2} \left(\overline{\boldsymbol{y}_r^*} - \bar{\boldsymbol{x}}_{n-r} \right)' \left((m(n-r)+r)\,\boldsymbol{\Sigma} + (n-r)\boldsymbol{W}^{(r)} \right)^{-1} \left(\overline{\boldsymbol{y}_r^*} - \bar{\boldsymbol{x}}_{n-r} \right) \right] \\
&\quad \exp \left[-\tfrac{1}{2} \operatorname{tr} \left(\boldsymbol{\Sigma}^{-1} \left(m(n-r-1)\boldsymbol{W}^{(n-r)} + (r-1)\boldsymbol{W}^{(r)} \right) \right) \right] \\
&\quad \exp \left[-\tfrac{r-1}{2} \operatorname{tr} \left(\boldsymbol{W}^{(r)^{-1}} \left(\sum_{j=1}^{m} \boldsymbol{W}_{y,j}^{(r)} \right) \right) \right] \qquad (5.10)
\end{aligned}
$$

Method II: Using whole data estimates to impute synthetic data

For this method, we impute m-copies of the sensitive part of the data as follows, independently for $j = 1, \ldots, m$

$$
\boldsymbol{y}_{1j}, \ldots, \boldsymbol{y}_{rj} \overset{\text{iid}}{\sim} N_p \left(\bar{\boldsymbol{x}}, \boldsymbol{W} \right)
$$

so that the released data is $\{\boldsymbol{Y}_j = (\boldsymbol{y}_{1j}, \ldots, \boldsymbol{y}_{rj}, \boldsymbol{x}_{r+1}, \ldots, \boldsymbol{x}_n) : j = 1, \ldots, m\}$. Then the portion of the likelihood of the released data required to calculate the posterior distributions is computed as

$$\int \left(\prod_{j=1}^{m} \pi(\boldsymbol{y}_{1j}, \ldots, \boldsymbol{y}_{rj} \,|\, \bar{\boldsymbol{x}}, \boldsymbol{W}) \right) \pi(\bar{\boldsymbol{x}} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{W} \,|\, \boldsymbol{\Sigma})\, d\bar{\boldsymbol{x}}\, d\boldsymbol{W}$$

We integrate out $\bar{\boldsymbol{x}}$ from the above likelihood and then multiply with our usual prior $\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{\frac{\delta}{2}}$ to obtain the following posterior distributions

$$\begin{aligned} \boldsymbol{S}_y^{*(r)^{-1/2}} \boldsymbol{W} \boldsymbol{S}_y^{*(r)^{-1/2}} &\sim \mathcal{W}_p^{-1}\left(\boldsymbol{I}_p, rm - p + \delta - 2\right) && (5.11) \\ \boldsymbol{W}^{-1/2} \boldsymbol{\Sigma} \boldsymbol{W}^{-1/2} &\sim \mathcal{W}_p^{-1}\left((n-1)\boldsymbol{I}_p, n - p + \delta - 2\right) && (5.12) \\ \boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \boldsymbol{W}, \overline{\boldsymbol{y}_r^*} &\sim N_p\left(\overline{\boldsymbol{y}_r^*}, \frac{\boldsymbol{\Sigma}}{n} + \frac{\boldsymbol{W}}{rm}\right) && (5.13) \end{aligned}$$

where $\boldsymbol{S}_y^{*(r)} = \sum_{j=1}^{m} \sum_{i=1}^{r} (\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}_r^*})(\boldsymbol{y}_{ij} - \overline{\boldsymbol{y}_r^*})' = \sum_{j=1}^{m} \sum_{i=1}^{r} (\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_j)(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_j)' + r \sum_{j=1}^{m} (\bar{\boldsymbol{y}}_j - \overline{\boldsymbol{y}_r^*})(\bar{\boldsymbol{y}}_j - \overline{\boldsymbol{y}_r^*})'$, so $\boldsymbol{S}_y^{*(r)}$ coincides with $\boldsymbol{S}_y^*$ when $r = n$; it is an analogue of $\boldsymbol{S}_y^*$ in the case of partially sensitive data.

5.2 Posterior Predictive Sampling method

We again consider the setup described in Section 4.2. The synthetic data are generated by repeating the following steps below independently for each $j = 1, \ldots, m$.

(a) Draw $(\boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*)$ from the posterior distribution (4.14).

(b) Draw $\boldsymbol{Y}_j = (y_{j1}, \ldots, y_{jn})' \sim N_p(\boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*)$.

We mention the distributions of all the quantities involved below.

The pdf of released data is

$$f(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m \,|\, \boldsymbol{\mu}_1^*, \ldots, \boldsymbol{\mu}_m^*, \boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^*)$$

$$\propto \prod_{j=1}^{m} \left|\boldsymbol{\Sigma}_j^*\right|^{-\frac{n}{2}} \left|\boldsymbol{\mathcal{S}}_{y,j}\right|^{\frac{n-p-2}{2}} \exp\left\{-\frac{1}{2}\left[\operatorname{tr}\left(\boldsymbol{\Sigma}_j^{*-1}\boldsymbol{\mathcal{S}}_{y,j}\right) + n(\bar{\boldsymbol{y}}_j - \boldsymbol{\mu}_j^*)'\boldsymbol{\Sigma}_j^{*-1}(\bar{\boldsymbol{y}}_j - \boldsymbol{\mu}_j^*)\right]\right\} \tag{5.14}$$

where $\bar{\boldsymbol{y}}_j = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_{ij}$, $\boldsymbol{\mathcal{S}}_{y,j} = \sum_{i=1}^{n}(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_j)(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_j)'$ as in the last section.

The pdf of the imputed parameters is

$$\begin{aligned} &f(\boldsymbol{\mu}_1^*, \ldots, \boldsymbol{\mu}_m^*, \boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^* \,|\, \bar{\boldsymbol{x}}, \boldsymbol{W}) \\ &\propto \prod_{j=1}^{m} \frac{|\boldsymbol{W}|^{\frac{n-p+\alpha-2}{2}}}{\left|\boldsymbol{\Sigma}_j^*\right|^{\frac{n+\alpha}{2}}} \exp\left\{-\frac{1}{2}\left[n(\boldsymbol{\mu}_j^* - \bar{\boldsymbol{x}})'\boldsymbol{\Sigma}_j^{*-1}(\boldsymbol{\mu}_j^* - \bar{\boldsymbol{x}}) + (n-1)\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{\Sigma}_j^{*-1}\right)\right]\right\} \end{aligned} \tag{5.15}$$

Finally the pdf of the original data is given by equation (4.4).

Likelihood of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

We want the likelihood of released data given the parameters. To that end, we compute the product of equations (5.14), (5.15) and (4.4) to obtain the joint pdf of all the quantities at play.

$$\begin{aligned} &f(\boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_m^*, \boldsymbol{\mu}_1^*, \ldots, \boldsymbol{\mu}_m^*, \boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^*, \bar{\boldsymbol{x}}, \boldsymbol{W} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\propto \exp\left[-\frac{1}{2}\sum_{j=1}^{m}\left\{\operatorname{tr}\left(\boldsymbol{\Sigma}_j^{*-1}\boldsymbol{\mathcal{S}}_{y,j}\right) + n(\bar{\boldsymbol{y}}_j - \boldsymbol{\mu}_j^*)'\boldsymbol{\Sigma}_j^{*-1}(\bar{\boldsymbol{y}}_j - \boldsymbol{\mu}_j^*) + \right.\right. \\ &\qquad \left.\left. n(\boldsymbol{\mu}_j^* - \bar{\boldsymbol{x}})'\boldsymbol{\Sigma}_j^{*-1}(\boldsymbol{\mu}_j^* - \bar{\boldsymbol{x}}) + (n-1)\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{\Sigma}_j^{*-1}\right)\right\}\right] \\ &\times \left(\prod_{j=1}^{m}\left|\boldsymbol{\Sigma}_j^*\right|^{-\frac{2n+\alpha}{2}}\left|\boldsymbol{\mathcal{S}}_{y,j}\right|^{\frac{n-p-2}{2}}\right)|\boldsymbol{W}|^{\frac{m(n-p+\alpha-2)+n-p-2}{2}}|\boldsymbol{\Sigma}|^{-\frac{n}{2}} \\ &\times \exp\left[-\frac{1}{2}\left\{n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu}) + (n-1)\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{\Sigma}^{-1}\right)\right\}\right] \end{aligned}$$

Our goal is to integrate out everything else from the joint pdf except for the released data. We first attempt to integrate out $\boldsymbol{\mu}_j^*$'s. Combining terms for $\boldsymbol{\mu}_j^*$ from inside first exponent we get

$$\begin{aligned}
&(\bar{\boldsymbol{y}}_j - \boldsymbol{\mu}_j^*)' \boldsymbol{\Sigma}_j^{*-1} (\bar{\boldsymbol{y}}_j - \boldsymbol{\mu}_j^*) + (\boldsymbol{\mu}_j^* - \bar{\boldsymbol{x}})' \boldsymbol{\Sigma}_j^{*-1} (\boldsymbol{\mu}_j^* - \bar{\boldsymbol{x}}) \\
=& 2 \left(\boldsymbol{\mu}_j^* - \frac{\bar{\boldsymbol{y}}_j + \bar{\boldsymbol{x}}}{2} \right)' \boldsymbol{\Sigma}_j^{*-1} \left(\boldsymbol{\mu}_j^* - \frac{\bar{\boldsymbol{y}}_j + \bar{\boldsymbol{x}}}{2} \right) + \frac{1}{2} (\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}_j)' \boldsymbol{\Sigma}_j^{*-1} (\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}_j)
\end{aligned}$$

After integrating out all the $\boldsymbol{\mu}_j^*$'s we are left with the following

$$\begin{aligned}
&f(\boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_m^*, \boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^*, \bar{\boldsymbol{x}}, \boldsymbol{W} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
\propto& \exp \left[-\frac{1}{2} \sum_{j=1}^{m} \left\{ \operatorname{tr} \left(\boldsymbol{\Sigma}_j^{*-1} \boldsymbol{\mathcal{S}}_{y,j} \right) + \frac{n}{2} (\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}_j)' \boldsymbol{\Sigma}_j^{*-1} (\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}_j) + (n-1) \operatorname{tr} \left(\boldsymbol{W} \boldsymbol{\Sigma}_j^{*-1} \right) \right\} \right] \\
\times& \left(\prod_{j=1}^{m} \left| \boldsymbol{\Sigma}_j^* \right|^{-\frac{2n+\alpha-1}{2}} \left| \boldsymbol{\mathcal{S}}_{y,j} \right|^{\frac{n-p-2}{2}} \right) |\boldsymbol{W}|^{\frac{m(n-p+\alpha-2)+n-p-2}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \\
\times& \exp \left[-\frac{1}{2} \left\{ n (\bar{\boldsymbol{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu}) + (n-1) \operatorname{tr} \left(\boldsymbol{W} \boldsymbol{\Sigma}^{-1} \right) \right\} \right]
\end{aligned}$$

Next we combine terms for $\bar{\boldsymbol{x}}$ from the exponents to get

$$\begin{aligned}
&\frac{1}{2} \sum_{j=1}^{n} (\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}_j)' \boldsymbol{\Sigma}_j^{*-1} (\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}_j) + (\bar{\boldsymbol{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu}) \\
=& \left(\bar{\boldsymbol{x}} - \left(\frac{1}{2} \sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} + \boldsymbol{\Sigma}^{-1} \right)^{-1} \left(\frac{1}{2} \sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j + \boldsymbol{\Sigma}^{-1} \mu \right) \right)' \left(\frac{1}{2} \sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} + \boldsymbol{\Sigma}^{-1} \right) \\
& \left(\bar{\boldsymbol{x}} - \left(\frac{1}{2} \sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} + \boldsymbol{\Sigma}^{-1} \right)^{-1} \left(\frac{1}{2} \sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j + \boldsymbol{\Sigma}^{-1} \mu \right) \right) \\
&+ \left(\boldsymbol{\mu} - \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right) \right)' \left(\boldsymbol{\Sigma} + 2 \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \right)^{-1} \\
& \left(\boldsymbol{\mu} - \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right) \right) \\
&+ \frac{1}{2} \left[\sum_{j=1}^{m} \bar{\boldsymbol{y}}_j' \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j - \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right)' \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right) \right]
\end{aligned}$$

After integrating out $\bar{\boldsymbol{x}}$ what remains is

$$
\begin{aligned}
& f(\boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_m^*, \boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^*, \boldsymbol{W} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
& \propto \exp\left[-\frac{n}{2} \left(\boldsymbol{\mu} - \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right) \right)' \left(\boldsymbol{\Sigma} + 2 \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \right)^{-1} \right. \\
& \quad \left. \left(\boldsymbol{\mu} - \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right) \right) \right] \\
& \times \left| \frac{1}{2} \sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} + \boldsymbol{\Sigma}^{-1} \right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \left\{ \sum_{j=1}^{m} \operatorname{tr}\left(\boldsymbol{\Sigma}_j^{*-1} \boldsymbol{\mathcal{S}}_{y,j} \right) + \right. \right. \\
& \quad \left. \left. \frac{n}{2} \left(\sum_{j=1}^{m} \bar{\boldsymbol{y}}_j' \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j - \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right)' \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right) \right) \right\} \right] \\
& \times \left(\prod_{j=1}^{m} \left| \boldsymbol{\Sigma}_j^* \right|^{-\frac{2n+\alpha-1}{2}} \left| \boldsymbol{\mathcal{S}}_{y,j} \right|^{\frac{n-p-2}{2}} \right) |\boldsymbol{W}|^{\frac{m(n-p+\alpha-2)+n-p-2}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \\
& \times \exp\left[-\frac{n-1}{2} \operatorname{tr}\left(\boldsymbol{W} \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} + \boldsymbol{\Sigma}^{-1} \right) \right) \right]
\end{aligned}
$$

The posterior distribution of $\boldsymbol{\mu}$ is already evident, since we are going to use a vague prior involving only $\boldsymbol{\Sigma}$. Finally integrating out $\boldsymbol{W}$, we have

$$
\begin{aligned}
& f(\boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_m^*, \boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^* \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
& \propto \left| \boldsymbol{\Sigma} + 2 \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \right|^{-\frac{1}{2}} \exp\left[-\frac{n}{2} \left(\boldsymbol{\mu} - \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right) \right)' \right. \\
& \quad \left. \left(\boldsymbol{\Sigma} + 2 \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \right)^{-1} \left(\boldsymbol{\mu} - \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right) \right) \right] \\
& \times \exp\left[-\frac{1}{2} \left\{ \sum_{j=1}^{m} \operatorname{tr}\left(\boldsymbol{\Sigma}_j^{*-1} \boldsymbol{\mathcal{S}}_{y,j} \right) + \right. \right. \\
& \quad \left. \left. \frac{n}{2} \left(\sum_{j=1}^{m} \bar{\boldsymbol{y}}_j' \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j - \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right)' \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right) \right) \right\} \right]
\end{aligned}
$$

$$\times \left(\prod_{j=1}^{m} \left| \boldsymbol{\Sigma}_j^* \right|^{-\frac{2n+\alpha-1}{2}} \left| \boldsymbol{\mathcal{S}}_{y,j} \right|^{\frac{n-p-2}{2}} \right) |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \left| \sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} + \boldsymbol{\Sigma}^{-1} \right|^{-\frac{m(n-p+\alpha-2)+n-1}{2}}$$

$$\times \left| \frac{1}{2} \sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} + \boldsymbol{\Sigma}^{-1} \right|^{-\frac{1}{2}} \left| \boldsymbol{\Sigma} + 2 \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right) \right|^{-1} \Bigg|^{\frac{1}{2}}$$

Theorem 5.2.1. *The likelihood of* $(\boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_m^*)$ *can be obtained by integrating out* $\boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^*$ *from*

$$f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_m^*, \boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^*) \propto \left| \boldsymbol{\Sigma} + 2 \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \right|^{-\frac{1}{2}} \exp \left[-\frac{n}{2} \left(\boldsymbol{\mu} - \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \right.\right.$$

$$\left.\left. \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right) \right)' \left(\boldsymbol{\Sigma} + 2 \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \right)^{-1} \left(\boldsymbol{\mu} - \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right) \right) \right]$$

$$\times \exp \left[-\frac{1}{2} \left\{ \sum_{j=1}^{m} \operatorname{tr} \left(\boldsymbol{\Sigma}_j^{*-1} \boldsymbol{\mathcal{S}}_{y,j} \right) + \right.\right.$$

$$\left.\left. \frac{n}{2} \left(\sum_{j=1}^{m} \bar{\boldsymbol{y}}_j' \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j - \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right)' \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right)^{-1} \left(\sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \bar{\boldsymbol{y}}_j \right) \right) \right\} \right]$$

$$\times \left(\prod_{j=1}^{m} \left| \boldsymbol{\Sigma}_j^* \right|^{-\frac{2n+\alpha-1}{2}} \left| \boldsymbol{\mathcal{S}}_{y,j} \right|^{\frac{n-p-2}{2}} \right) \left| \sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} \right|^{-\frac{1}{2}} |\boldsymbol{\Sigma}|^{-\frac{n-1}{2}} \left| \sum_{j=1}^{m} \boldsymbol{\Sigma}_j^{*-1} + \boldsymbol{\Sigma}^{-1} \right|^{-\frac{m(n-p+\alpha-2)+n-1}{2}} \tag{5.16}$$

Posterior distributions of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

We multiply the same prior as before with the likelihood above, which results in the posterior distribution neatly separating out into three parts as follows:

$$\pi(\boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^*, \boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_m^*)\, \pi(\boldsymbol{\Sigma} \,|\, \boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^*, \boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_m^*) \pi(\boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^* \,|\, \boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_m^*)$$

$$= \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^* \,|\, \boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_m^*)$$

$$\propto \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^*, \boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_m^*)$$

$$= \pi(\boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_m^*, \boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^* \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \pi(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$
\begin{aligned}
&\propto \left|\boldsymbol{\Sigma}+2\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{-1}\right|^{-\frac{1}{2}} \exp\left[-\frac{n}{2}\left(\boldsymbol{\mu}-\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{-1}\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\bar{\boldsymbol{y}}_j\right)\right)'\right. \\
&\quad \left.\left(\boldsymbol{\Sigma}+2\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{-1}\right)^{-1}\left(\boldsymbol{\mu}-\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{-1}\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\bar{\boldsymbol{y}}_j\right)\right)\right] \\
&\times \left|\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{1/2}\boldsymbol{\Sigma}\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{1/2}\right|^{\frac{m(n-p+\alpha-2)-\delta}{2}} \\
&\times \left|\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{1/2}\boldsymbol{\Sigma}\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{1/2}+\boldsymbol{I}_p\right|^{-\frac{m(n-p+\alpha-2)+n-1}{2}} \\
&\times \frac{\left(\prod_{j=1}^{m}\left|\boldsymbol{\Sigma}_j^{*}\right|^{-\frac{2n+\alpha-1}{2}}\right)}{\left|\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right|^{\frac{m(n-p+\alpha-2)-\delta+1}{2}}}\exp\left[-\frac{1}{2}\left\{\sum_{j=1}^{m}\mathrm{tr}\left(\boldsymbol{\Sigma}_j^{*-1}\mathfrak{S}_{y,j}\right)+\right.\right. \\
&\quad \left.\left.\frac{n}{2}\left(\sum_{j=1}^{m}\bar{\boldsymbol{y}}_j'\boldsymbol{\Sigma}_j^{*-1}\bar{\boldsymbol{y}}_j-\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\bar{\boldsymbol{y}}_j\right)'\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{-1}\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\bar{\boldsymbol{y}}_j\right)\right)\right\}\right]
\end{aligned}
$$

From the above set of equations, the requisite posterior distributions are as follows (obtained after multiplying by the Jacobian of the transformation $\boldsymbol{\Sigma}\mapsto\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{1/2}\boldsymbol{\Sigma}\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{1/2}$ which is $\left|\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right|^{-p}$)

$$
\boldsymbol{\mu}\,|\,\boldsymbol{\Sigma},\boldsymbol{\Sigma}_1^*,\ldots,\boldsymbol{\Sigma}_m^*,\boldsymbol{Y}_1^*,\ldots,\boldsymbol{Y}_m^*\sim N_p\left(\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{-1}\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\bar{\boldsymbol{y}}_j\right),\frac{1}{n}\left(\boldsymbol{\Sigma}+2\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{-1}\right)\right) \tag{5.17}
$$

$$
\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{1/2}\boldsymbol{\Sigma}\left(\sum_{j=1}^{m}\boldsymbol{\Sigma}_j^{*-1}\right)^{1/2}\sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}\left(\frac{m(n-p+\alpha-2)+p-\delta+1}{2},\frac{n-p+\delta-2}{2}\right) \tag{5.18}
$$

and the latent matrices have the following distribution

$$
\begin{aligned}
& g(\boldsymbol{\Sigma}_1^*, \ldots, \boldsymbol{\Sigma}_m^* \,|\, \boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_m^*) \\
& \propto \frac{\left|\sum_{j=1}^m \boldsymbol{\Sigma}_j^{*-1}\right|^{-\frac{m(n-p+\alpha-2)+2p-\delta+1}{2}}}{\left(\prod_{j=1}^m \left|\boldsymbol{\Sigma}_j^*\right|^{\frac{2n+\alpha-1}{2}}\right)} \exp\left[-\frac{1}{2}\left\{\sum_{j=1}^m \operatorname{tr}\left(\boldsymbol{\Sigma}_j^{*-1}\boldsymbol{\mathcal{S}}_{y,j}\right) + \right.\right. \qquad (5.19) \\
& \left.\left. \frac{n}{2}\left(\sum_{j=1}^m \bar{\boldsymbol{y}}_j'\boldsymbol{\Sigma}_j^{*-1}\bar{\boldsymbol{y}}_j - \left(\sum_{j=1}^m \boldsymbol{\Sigma}_j^{*-1}\bar{\boldsymbol{y}}_j\right)'\left(\sum_{j=1}^m \boldsymbol{\Sigma}_j^{*-1}\right)^{-1}\left(\sum_{j=1}^m \boldsymbol{\Sigma}_j^{*-1}\bar{\boldsymbol{y}}_j\right)\right)\right\}\right] \\
& \propto \frac{\left|\sum_{j=1}^m \boldsymbol{\Sigma}_j^{*-1}\right|^{-\frac{m(n-p+\alpha-2)+2p-\delta+1}{2}}}{\left(\prod_{j=1}^m \left|\boldsymbol{\Sigma}_j^*\right|^{\frac{2n+\alpha-1}{2}}\right)} \exp\left[-\frac{1}{2}\sum_{j=1}^m \operatorname{tr}\left(\boldsymbol{\Sigma}_j^{*-1}\boldsymbol{\mathcal{S}}_{y,j}\right)\right] \\
& \exp\left[-\frac{n}{4}\sum_{j=1}^m \left(\bar{\boldsymbol{y}}_j - \left(\sum_{j=1}^m \boldsymbol{\Sigma}_j^{*-1}\right)^{-1}\left(\sum_{j=1}^m \boldsymbol{\Sigma}_j^{*-1}\bar{\boldsymbol{y}}_j\right)\right)' \boldsymbol{\Sigma}_j^{*-1}\left(\bar{\boldsymbol{y}}_j - \left(\sum_{j=1}^m \boldsymbol{\Sigma}_j^{*-1}\right)^{-1}\left(\sum_{j=1}^m \boldsymbol{\Sigma}_j^{*-1}\bar{\boldsymbol{y}}_j\right)\right)\right] \qquad (5.20)
\end{aligned}
$$

where we can notice that the quantity inside the last exponential vanishes when $m = 1$.

The posterior distributions for the parameters exist provided
$n > \max\left\{p, 2p - \delta + 1, p - \alpha + 2 + \frac{\delta-2}{m}\right\}$.

Following the development before for the MI PPS MLR case, here the following transformation should work to sample from the above distribution

$$
\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1^*, \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_1^{*-\frac{1}{2}}\boldsymbol{\Sigma}_2^*\boldsymbol{\Sigma}_1^{*-\frac{1}{2}}, \ldots, \boldsymbol{\Sigma}_m = \boldsymbol{\Sigma}_1^{*-\frac{1}{2}}\boldsymbol{\Sigma}_m^*\boldsymbol{\Sigma}_1^{*-\frac{1}{2}} \qquad (5.21)
$$

so that the proposal density of $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \boldsymbol{\Sigma}_m)$ when we apply Accept-Reject would be matrix variate Dirichlet Type-II (matrix analogue of Inverted Dirichlet distribution).

FPPS

For the FPPS case, we produce a single copy $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ from the posterior distribution (4.14) and use it to generate m samples $\boldsymbol{Y}_j = (y_{j1}, \ldots, y_{jn})' \overset{\text{iid}}{\sim} N_p(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ for each $j = 1, \ldots, m$. Using the same notation as in Section 5.1, we can derive the posterior distribution of the parameters as follows with exactly the same conditions for existence as in Section 4.2

$$\begin{aligned}
\boldsymbol{\Sigma}^* \,|\, \boldsymbol{\mathcal{S}}_y^* &\sim \mathcal{W}_p^{-1}\left(\boldsymbol{\mathcal{S}}_y^*, nm - 2p + \delta - 1\right) && (5.22)\\
\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1/2} &\sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}\left(\frac{n+\alpha-\delta-1}{2}, \frac{n-p+\delta-2}{2}\right) && (5.23)\\
\boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}^*, \overline{\boldsymbol{y}^*} &\sim N_p\left(\overline{\boldsymbol{y}^*}, \frac{1}{n}\left(\boldsymbol{\Sigma} + \left(1 + \frac{1}{m}\right)\boldsymbol{\Sigma}^*\right)\right) && (5.24)
\end{aligned}$$

We can reformulate the above posterior distributions as:

$$\begin{aligned}
\boldsymbol{\mathcal{S}}_y^{*-1/2}\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\mathcal{S}}_y^* &\sim \mathcal{W}_p^{-1}\left(\boldsymbol{I}_p, nm - 2p + \delta - 1\right) && (5.25)\\
\boldsymbol{\Sigma}^{*-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1/2} &\sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}\left(\frac{n+\alpha-\delta-1}{2}, \frac{n-p+\delta-2}{2}\right) && (5.26)\\
\boldsymbol{\mu} \,|\, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}^*, \overline{\boldsymbol{y}^*} &\sim N_p\left(\overline{\boldsymbol{y}^*}, \frac{1}{n}\left(\boldsymbol{\Sigma} + \left(1 + \frac{1}{m}\right)\boldsymbol{\Sigma}^*\right)\right) && (5.27)
\end{aligned}$$

Chapter 6

Bayesian Analysis of Singly Imputed Synthetic Data under the Multivariate Regression Model

In terms of the MVR model, in our context, we consider several *sensitive* response variables y_j, $j = 1, \ldots, m$, originating the vector of response variables $\boldsymbol{y} = (y_1, \ldots, y_m)'$, and a vector of p *non-sensitive* predictors $\boldsymbol{x} = (x_1, \ldots, x_p)'$. We assume that $\boldsymbol{y} \,|\, \boldsymbol{x} \sim N_m(\boldsymbol{B}'\boldsymbol{x}, \boldsymbol{\Sigma})$, with $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ unknown, and the original data consist of $\{(y_{1i}, \ldots, y_{mi}, x_{1i}, \ldots, x_{pi}) : i = 1, \ldots, n\}$. We write $\boldsymbol{Y} = (\boldsymbol{y}_1 \cdots \boldsymbol{y}_n)'$ with $\boldsymbol{y}_i = (y_{1i}, \ldots, y_{mi})'$ and $\boldsymbol{X} = (\boldsymbol{x}_1 \cdots \boldsymbol{x}_n)'$ with $\boldsymbol{x}_i = (x_{1i}, \ldots, x_{pi})'$. We also assume that $\text{rank}(\boldsymbol{X}) = p < n$ and $n \geq m + p$. We are thus considering the following regression model

$$\boldsymbol{Y}_{n\times m} = \boldsymbol{X}_{n\times p}\boldsymbol{B}_{p\times m} + \mathbb{E}_{n\times m} \tag{6.1}$$

where $\mathbb{E}_{n\times m} \sim N_{n,m}(\boldsymbol{O}, \boldsymbol{\Sigma}\otimes\boldsymbol{I}_n)$. It is well known that based on the original data, $\hat{\boldsymbol{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ is the MLE and the UMVUE of $\boldsymbol{B}$, distributed as $\hat{\boldsymbol{B}} \sim N_{p,m}(\boldsymbol{B}, \boldsymbol{\Sigma} \otimes$

$(\boldsymbol{X}'\boldsymbol{X})^{-1})$, independent of $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{B}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{B}}) = \frac{1}{n}\boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{Y}$ which is the MLE of $\boldsymbol{\Sigma}$, with $n\hat{\boldsymbol{\Sigma}} \sim \mathcal{W}_m(\boldsymbol{\Sigma}, n-p)$. Therefore $\boldsymbol{S} = \frac{n\hat{\boldsymbol{\Sigma}}}{n-p}$ will be the UMVUE of $\boldsymbol{\Sigma}$. We use the following two standard results for deriving of the distribution of the MLEs, that we will keep using continuously, which can be found in Gupta and Nagar (2000) or Kollo and Rosen (2005) or Anderson (2003):

- If $\boldsymbol{X} \sim N_{n,p}(\boldsymbol{M}, \boldsymbol{V} \otimes \boldsymbol{U})$, then for $\boldsymbol{D}_{r\times n}$, $C_{p\times s}$ with $\text{rank}(D) = r \leq n$, $\text{rank}(C) = s \leq p$, we have $\boldsymbol{DXC} \sim N_{r,s}(\boldsymbol{DMC}, \boldsymbol{C}'\boldsymbol{VC} \otimes \boldsymbol{DUD}')$.
- If $\boldsymbol{X} \sim N_{n,p}(\boldsymbol{O}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Psi})$ where $\boldsymbol{\Psi}_{n\times n}$ is symmetric idempotent with $\text{rank}(\boldsymbol{\Psi}) = q \geq p$. Then $\boldsymbol{X}'\boldsymbol{X} \sim \mathcal{W}_p(\boldsymbol{\Sigma}, q)$.

6.1 Plug In Sampling method

The synthetic data consist of a single synthetic version of $\boldsymbol{Y}$ generated based on the Plug- in method as described below. From the original data $\{(y_{1i}, \ldots, y_{mi}, x_{1i}, \ldots, x_{pi}) : i = 1, \ldots, n\}$, after estimating $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$ by $\hat{\boldsymbol{B}}$ and $\boldsymbol{S}$, respectively, we generate the synthetic data, denoted as $\boldsymbol{V} = (\boldsymbol{v}_1 \cdots \boldsymbol{v}_n)'$ where $\boldsymbol{v}_i = (v_{1i}, \ldots, v_{mi})'$, are independently distributed as

$$\boldsymbol{v}_i \,|\, \hat{\boldsymbol{B}}, \boldsymbol{S} \sim N_m(\hat{\boldsymbol{B}}'\boldsymbol{x}_i, \boldsymbol{S}),\ i = 1, \ldots, n \tag{6.2}$$

so that $\boldsymbol{V} \sim N_{n,m}(\boldsymbol{X}\hat{\boldsymbol{B}}, \boldsymbol{S} \otimes \boldsymbol{I}_n)$. Our goal is to draw inference on $\boldsymbol{B}$ based on the synthetic data $\{(v_{1i}, \ldots, v_{mi}, x_{1i}, \ldots, x_{pi}) : i = 1, \ldots, n\}$. Towards this end, let us define

$$\boldsymbol{B}^* = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V} \sim N_{p,m}(\hat{\boldsymbol{B}}, \boldsymbol{S} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}) \tag{6.3}$$

$$(n-p)\boldsymbol{S}^* = (\boldsymbol{V} - \boldsymbol{X}\boldsymbol{B}^*)'(\boldsymbol{V} - \boldsymbol{X}\boldsymbol{B}^*) = \boldsymbol{V}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{V} \sim \mathcal{W}_m(\boldsymbol{S}, n-p) \tag{6.4}$$

From Moura et al. (2021) we have the following result.

Theorem 6.1.1. *The joint pdf of* $(\boldsymbol{B}^*, \boldsymbol{S}^*)$ *is given by*

$$f_{\boldsymbol{B},\boldsymbol{\Sigma}}(\boldsymbol{B}^*, \boldsymbol{S}^*) \propto \int_{S_n^{++}} \frac{|\boldsymbol{S}^*|^{\frac{n-p-m-1}{2}} |\boldsymbol{\Sigma}+\boldsymbol{S}|^{-\frac{p}{2}}}{|\boldsymbol{\Sigma}|^{\frac{n-p}{2}} |\boldsymbol{S}|^{\frac{m+1}{2}}} e^{-\frac{1}{2}\operatorname{tr}\left[(\boldsymbol{\Sigma}+\boldsymbol{S})^{-1}(\boldsymbol{B}^*-\boldsymbol{B})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{B}^*-\boldsymbol{B})+(n-p)\left(\boldsymbol{S}^*\boldsymbol{S}^{-1}+\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right)\right]} d\boldsymbol{S}$$

We see that $\boldsymbol{B}^*$ and $\boldsymbol{S}^*$ are separable above, showing that they are independent, with $\boldsymbol{B}^* \mid \boldsymbol{S} \sim N_{p,m}(\boldsymbol{B}, (\boldsymbol{\Sigma}+\boldsymbol{S}) \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1})$ and $(n-p)\boldsymbol{S}^* \mid \boldsymbol{S} \sim \mathcal{W}_m(\boldsymbol{S}, n-p)$.

Posterior distributions of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$

We choose the standard non-informative prior $\pi(\boldsymbol{B}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{\delta}{2}}$, and multiply it with the likelihood of the released data to obtain the following

$$\begin{aligned}
\boldsymbol{S} \mid \boldsymbol{S^*} &\sim \mathcal{W}_m^{-1}((n-p)\boldsymbol{S^*}, n-p-m+\delta-1) && (6.5)\\
\boldsymbol{\Sigma} \mid \boldsymbol{S} &\sim \mathcal{W}_m^{-1}((n-p)\boldsymbol{S}, n-p-m+\delta-1) && (6.6)\\
\boldsymbol{B} \mid \boldsymbol{\Sigma}, \boldsymbol{S}, \boldsymbol{B}^* &\sim N_{p,m}(\boldsymbol{B}^*, (\boldsymbol{\Sigma}+\boldsymbol{S}) \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}) && (6.7)
\end{aligned}$$

We can reformulate the above posterior distributions as:

$$\begin{aligned}
\boldsymbol{S^*}^{-1/2}\boldsymbol{S}\boldsymbol{S^*}^{-1/2} &\sim \mathcal{W}_m^{-1}((n-p)\boldsymbol{I}_m, n-p-m+\delta-1) && (6.8)\\
\boldsymbol{S}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{S}^{-1/2} &\sim \mathcal{W}_m^{-1}((n-p)\boldsymbol{I}_m, n-p-m+\delta-1) && (6.9)\\
\boldsymbol{B} \mid \boldsymbol{\Sigma}, \boldsymbol{S}, \boldsymbol{B}^* &\sim N_{p,m}(\boldsymbol{B}^*, (\boldsymbol{\Sigma}+\boldsymbol{S}) \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}) && (6.10)
\end{aligned}$$

which has the benefit that $\boldsymbol{S^*}^{-1/2}\boldsymbol{S}\boldsymbol{S^*}^{-1/2} \perp \boldsymbol{S}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{S}^{-1/2}$ and their posterior distributions are unconditional.

The posterior distributions are proper as long as $n > \max\{m+p-1, 2m+p-\delta\}$.

We can get the equivalent results for MVN model by plugging $p = 1$, $m = p$; and that of the MLR model by plugging $m = 1$.

Bayes Estimators of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$

$$\hat{\boldsymbol{B}}_{\text{BAYES}} = \mathrm{E}(\boldsymbol{B} \mid \boldsymbol{B}^*, \boldsymbol{S}^*) = \mathrm{E}_{\boldsymbol{S}}\, \mathrm{E}_{\boldsymbol{\Sigma}}\, \mathrm{E}(\boldsymbol{B} \mid \boldsymbol{B}^*, \boldsymbol{\Sigma}, \boldsymbol{S}) = \mathrm{E}_{\boldsymbol{S}}\, \mathrm{E}_{\boldsymbol{\Sigma}}\, \mathrm{E}(\boldsymbol{B}^*) = \boldsymbol{B}^*$$

$$\hat{\boldsymbol{\Sigma}}_{\text{BAYES}} = \mathrm{E}(\boldsymbol{\Sigma} \mid \boldsymbol{B}^*, \boldsymbol{S}^*) = \mathrm{E}_{\boldsymbol{S}}\, \mathrm{E}(\boldsymbol{\Sigma} \mid \boldsymbol{S}^*, \boldsymbol{S}) = \mathrm{E}_{\boldsymbol{S}} \left(\tfrac{(n-p)\boldsymbol{S}}{(n-p-2m+\delta-2)} \mid \boldsymbol{S}^* \right)$$

$$= \frac{(n-p)^2 \boldsymbol{S}^*}{(n-p-2m+\delta-2)^2}$$

$$\widehat{|\boldsymbol{\Sigma}|}_{\text{BAYES}} = \mathrm{E}\left(|\boldsymbol{\Sigma}| \mid \boldsymbol{B}^*, \boldsymbol{S}^*\right) = \mathrm{E}_{\boldsymbol{S}}\, \mathrm{E}\left(|\boldsymbol{\Sigma}| \mid \boldsymbol{S}, \boldsymbol{S}^*\right) = \mathrm{E}_{\boldsymbol{S}} \left(|\boldsymbol{S}|\, \mathrm{E}\left(\left| \boldsymbol{S}^{-1/2} \boldsymbol{\Sigma} \boldsymbol{S}^{-1/2} \right| \right) \mid \boldsymbol{S}^* \right)$$

$$= \left(\prod_{j=1}^{m} \frac{n-p}{n-p-m+\delta-j-2} \right) \mathrm{E}\left(|\boldsymbol{S}| \mid \boldsymbol{S}^*\right) = \left(\prod_{j=1}^{m} \frac{n-p}{n-p-m+\delta-j-2} \right)^2 |\boldsymbol{S}^*|$$

provided $n > \max\{m+p-1, 2m+p-\delta+4\}$, and we use the results (4.11) and (4.12).

Credible Sets for $|\boldsymbol{\Sigma}|$ and $\boldsymbol{B}$

We see that $\boldsymbol{\Sigma}^{-1} \mid \boldsymbol{S} \sim \mathcal{W}_m\left(\boldsymbol{S}^{-1}/(n-p), n-p-m+\delta-1\right)$, so by using (4.13)

$$\frac{|\boldsymbol{\Sigma}^{-1}|}{|\boldsymbol{S}^{-1}/(n-p)|} \sim \prod_{i=1}^{m} u_i, \quad \text{where } u_i \sim \chi^2_{n-p-m+\delta-i} \text{ independently for } i = 1, \ldots, m$$

which also shows that the quantity on the left hand side of the above relation is independent of $\boldsymbol{S}$. Similarly we can get

$$\frac{|\boldsymbol{S}^{-1}|}{\left|\boldsymbol{S}^{*-1}/(n-p)\right|} \sim \prod_{j=1}^{m} v_j, \quad \text{where } v_j \sim \chi^2_{n-p-m+\delta-j} \text{ independently for } j = 1, \ldots, m$$

So we can define a pivot for the *generalized variance* $|\boldsymbol{\Sigma}|$ as $N \coloneqq \left|\boldsymbol{\Sigma}\boldsymbol{S}^{*-1}\right|$ where

$$(n-p)^{-2m}\, N^{-1} \sim \left(\prod_{i=1}^{m} u_i\right)\left(\prod_{j=1}^{m} v_j\right)$$

where u_i's and v_j's are as above and they are all pairwise independent. A $(1-\gamma)$ level credible set for $|\boldsymbol{\Sigma}|$ based on N is

$$\left[a_{n,p,\delta;\gamma}\,|\boldsymbol{S}^*|\,, b_{n,p,\delta;\gamma}\,|\boldsymbol{S}^*|\right]$$

where $a_{n,p,\delta;\gamma}$ and $b_{n,p,\delta;\gamma}$ are any two constants that satisfy $1-\gamma = P(a_{n,p,\delta;\gamma} \leq N \leq b_{n,p,\delta;\gamma})$. The length of the credible interval is $|\boldsymbol{S}^*|\,(b_{n,p,\delta;\gamma} - a_{n,p,\delta;\gamma})$.

Next we define the pivot for $\boldsymbol{B}$ as

$$T \coloneqq \frac{|(\boldsymbol{B}-\boldsymbol{B}^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{B}-\boldsymbol{B}^*)|}{|(n-p)\boldsymbol{S}^*|}$$

We will prove that T is a pivot and derive a sampling scheme in what follows. We first notice that the Wishart distribution is defined through the matrix normal distribution in the following manner: Let $\boldsymbol{Y} \sim N_{n,p}(\boldsymbol{O}, \boldsymbol{\Psi}\otimes\boldsymbol{I}_n)$ and define $\boldsymbol{V} = \boldsymbol{Y}'\boldsymbol{Y}$, $n \geq p$, then $\boldsymbol{V} \sim \mathcal{W}_p(\boldsymbol{\Psi}, n)$. That fact combined with $(\boldsymbol{X}'\boldsymbol{X})^{1/2}(\boldsymbol{B}-\boldsymbol{B}^*) \sim N_{p,m}(\boldsymbol{O}, (\boldsymbol{\Sigma}+\boldsymbol{S})\otimes\boldsymbol{I}_p)$ due to (6.7) gives us, for $p \geq m$,

$$(\boldsymbol{B}-\boldsymbol{B}^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{B}-\boldsymbol{B}^*)\,|\,\boldsymbol{\Sigma}, \boldsymbol{S} \sim \mathcal{W}_m(\boldsymbol{\Sigma}+\boldsymbol{S}, p)$$

Therefore, we have $T = |\boldsymbol{H}| \times |\boldsymbol{G}/(n-p)|$ where

$$\boldsymbol{H} = (\boldsymbol{\Sigma}+\boldsymbol{S})^{-1/2}(\boldsymbol{B}-\boldsymbol{B}^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{B}-\boldsymbol{B}^*)(\boldsymbol{\Sigma}+\boldsymbol{S})^{-1/2}$$

$$\boldsymbol{G} = \boldsymbol{S}^{*-1/2}\boldsymbol{S}^{1/2}(\boldsymbol{S}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{S}^{-1/2} + \boldsymbol{I}_m)\boldsymbol{S}^{1/2}\boldsymbol{S}^{*-1/2}$$

we have $\boldsymbol{H} \mid \boldsymbol{\Sigma}, \boldsymbol{S} \sim \mathcal{W}_m(\boldsymbol{I}_m, p)$, so $\boldsymbol{H}$ is independent of $(\boldsymbol{\Sigma}, \boldsymbol{S})$ and $\boldsymbol{H} \sim \mathcal{W}_m(\boldsymbol{I}_m, p)$. That implies $\boldsymbol{H}$ is independent of $\boldsymbol{G}$, and we show that the distribution of $\boldsymbol{G}$ is free of $(\boldsymbol{\Sigma}, \boldsymbol{S})$. In a similar manner as in Section 4.1, we notice that conditionally $\boldsymbol{G} \mid \boldsymbol{G}_0 \sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{G}_0, n-p-m+\delta-1\right)+\boldsymbol{G}_0$, and $\boldsymbol{G}_0 \sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{I}_m, n-p-m+\delta-1\right)$ due to (6.9) and (6.8) respectively. This concludes the proof of the fact that T is a pivot for $\boldsymbol{B}$. Using (4.13) the distribution of T is given by

$$\begin{aligned} T \mid \boldsymbol{G} &\sim \left(\prod_{i=1}^{m} \chi^2_{p-i+1}\right) \times |\boldsymbol{G}/(n-p)| && (6.11)\\ \boldsymbol{G} \mid \boldsymbol{G}_0 &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{G}_0, n-p-m+\delta-1\right) + \boldsymbol{G}_0 && (6.12)\\ \boldsymbol{G}_0 &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{I}_m, n-p-m+\delta-1\right) && (6.13) \end{aligned}$$

Remark 6.1.1. *If one wants to test the significance of a set of regression coefficients or more generally of a linear combination of these regression coefficients,* $\boldsymbol{AB} = \boldsymbol{C}$ *where* $\boldsymbol{A}$ *is a* $k \times p$ *matrix with* $rank(\boldsymbol{A}) = k \leq p$ *and* $k \geq m$*, one may define*

$$T_{\boldsymbol{C}} \coloneqq \frac{\left|(\boldsymbol{C}-\boldsymbol{AB}^*)'\left\{\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right\}^{-1}(\boldsymbol{C}-\boldsymbol{AB}^*)\right|}{|(n-p)\boldsymbol{S}^*|} \tag{6.14}$$

and proceed by noting that

$$\begin{aligned} T_{\boldsymbol{C}} \mid \boldsymbol{G} &\sim \left(\prod_{i=1}^{m} \chi^2_{k-i+1}\right) \times |\boldsymbol{G}/(n-p)| && (6.15)\\ \boldsymbol{G} \mid \boldsymbol{G}_0 &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{G}_0, n-p-m+\delta-1\right) + \boldsymbol{G}_0 && (6.16)\\ \boldsymbol{G}_0 &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{I}_m, n-p-m+\delta-1\right) && (6.17) \end{aligned}$$

Remark 6.1.2. *To infer about* $\boldsymbol{ABD} = \boldsymbol{\Delta}$ *where* $\boldsymbol{A}$ *is a* $k \times p$ *matrix as before and* $\boldsymbol{D}$ *is a* $m \times r$ *matrix with* $rank(\boldsymbol{D}) = r \leq k$*, we start with* $\boldsymbol{\Delta}^* = \boldsymbol{AB}^*\boldsymbol{D}$ *and propose*

to use

$$T_{\boldsymbol{\Delta}} \coloneqq \frac{\left|(\boldsymbol{\Delta}-\boldsymbol{\Delta}^*)'\left\{\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right\}^{-1}(\boldsymbol{\Delta}-\boldsymbol{\Delta}^*)\right|}{|(n-p)\boldsymbol{D}'\boldsymbol{S}^*\boldsymbol{D}|} \tag{6.18}$$

whose distribution is obtained as follows

$$T_{\boldsymbol{\Delta}} \,|\, \tilde{\boldsymbol{G}} \;\sim\; \left(\prod_{i=1}^{m}\chi^2_{r-i+1}\right) \times \left|\tilde{\boldsymbol{G}}/(n-p)\right| \tag{6.19}$$

$$\tilde{\boldsymbol{G}} \,|\, \tilde{\boldsymbol{G}}_0 \;\sim\; \mathcal{W}_r^{-1}\left((n-p)\tilde{\boldsymbol{G}}_0, n-p-m+\delta-1\right) + \tilde{\boldsymbol{G}}_0 \tag{6.20}$$

$$\tilde{\boldsymbol{G}}_0 \;\sim\; \mathcal{W}_r^{-1}\left((n-p)\boldsymbol{I}_r, n-p-m+\delta-1\right) \tag{6.21}$$

6.2 Posterior Predictive Sampling method

We return to the setup of the last section. Under the *posterior predictive sampling* method, starting with a vague prior $\pi(\boldsymbol{B},\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\alpha/2}$, the joint (imputed) posterior distribution of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$, given $\boldsymbol{Y}$, can be represented as

$$\begin{aligned} \boldsymbol{\Sigma} \,|\, \boldsymbol{S} &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{S}, n-p+\alpha-m-1\right) \\ \boldsymbol{B} \,|\, \boldsymbol{\Sigma}, \hat{\boldsymbol{B}} &\sim N_{p,m}(\hat{\boldsymbol{B}}, \boldsymbol{\Sigma}\otimes(\boldsymbol{X}'\boldsymbol{X})^{-1}) \end{aligned} \tag{6.22}$$

We assume $n+\alpha > 2p+m$ throughout. We now draw $(\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Sigma}})$ from the above posterior, and then draw a random sample $\boldsymbol{V} \sim N_{nm}(\boldsymbol{X}\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Sigma}}\otimes\boldsymbol{I}_n)$, which form the singly imputed synthetic data. Define

$$\boldsymbol{B}^* = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V} \sim N_{p,m}(\tilde{\boldsymbol{B}}, \tilde{\boldsymbol{\Sigma}}\otimes(\boldsymbol{X}'\boldsymbol{X})^{-1}) \tag{6.23}$$

$$(n-p)\boldsymbol{S}^* = \boldsymbol{V}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{V} \sim \mathcal{W}_m^{-1}(\tilde{\boldsymbol{\Sigma}}, n-p) \tag{6.24}$$

From Moura et al. (2017b) we have the following result.

Theorem 6.2.1. *The joint pdf of $\boldsymbol{B}^*$ and $\boldsymbol{S}^*$ is obtained by integrating out $\tilde{\boldsymbol{\Sigma}}$ from*

the joint pdf of $(\boldsymbol{B}^*, \boldsymbol{S}^*, \tilde{\boldsymbol{\Sigma}})$ *given by*

$$f(\boldsymbol{B}^*, \boldsymbol{S}^*, \tilde{\boldsymbol{\Sigma}}) \propto e^{-\frac{1}{2}\operatorname{tr}\left[\left(\boldsymbol{\Sigma}+2\tilde{\boldsymbol{\Sigma}}\right)^{-1}(\boldsymbol{B}^*-\boldsymbol{B})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{B}^*-\boldsymbol{B})+(n-p)\left(\boldsymbol{S}^*\tilde{\boldsymbol{\Sigma}}^{-1}\right)\right]}$$
$$\times \left|\boldsymbol{\Sigma}+2\tilde{\boldsymbol{\Sigma}}\right|^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{\frac{n-p+\alpha-m-1}{2}} \left|\boldsymbol{\Sigma}+\tilde{\boldsymbol{\Sigma}}\right|^{-\frac{2n-2p+\alpha-m-1}{2}} \left|\tilde{\boldsymbol{\Sigma}}\right|^{-\frac{m+1}{2}} |\boldsymbol{S}^*|^{\frac{n-p-m-1}{2}}$$

Posterior distributions of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$

We multiply the above likelihood with our usual prior $\pi(\boldsymbol{B}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{\delta}{2}}$ and the Jacobian of the transformation $\boldsymbol{\Sigma} \mapsto \tilde{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\tilde{\boldsymbol{\Sigma}}^{-1/2}$ which is $\left|\tilde{\boldsymbol{\Sigma}}\right|^m$ to get

$$\begin{aligned}
\tilde{\boldsymbol{\Sigma}} \,|\, \boldsymbol{S}^* &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{S}^*, n-p-2m+\delta\right) && (6.25)\\
\tilde{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\tilde{\boldsymbol{\Sigma}}^{-1/2} &\sim \mathrm{B}_{\mathrm{m}}^{\mathrm{II}}\left(\frac{n-p+\alpha-\delta}{2}, \frac{n-p-m+\delta-1}{2}\right) && (6.26)\\
\boldsymbol{B} \,|\, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{B}^* &\sim N_{p,m}(\boldsymbol{B}^*, (\boldsymbol{\Sigma}+2\tilde{\boldsymbol{\Sigma}}) \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}) && (6.27)
\end{aligned}$$

We can reformulate the above posterior distributions as:

$$\begin{aligned}
\boldsymbol{S}^{*-1/2}\tilde{\boldsymbol{\Sigma}}\boldsymbol{S}^{*-1/2} &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{I}_m, n-p-2m+\delta\right) && (6.28)\\
\tilde{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\tilde{\boldsymbol{\Sigma}}^{-1/2} &\sim \mathrm{B}_{\mathrm{m}}^{\mathrm{II}}\left(\frac{n-p+\alpha-\delta}{2}, \frac{n-p-m+\delta-1}{2}\right) && (6.29)\\
\boldsymbol{B} \,|\, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{B}^* &\sim N_{p,m}(\boldsymbol{B}^*, (\boldsymbol{\Sigma}+2\tilde{\boldsymbol{\Sigma}}) \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}) && (6.30)
\end{aligned}$$

which has the benefit that $\boldsymbol{S}^{*-1/2}\tilde{\boldsymbol{\Sigma}}\boldsymbol{S}^{*-1/2}$ is independent of $\tilde{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\tilde{\boldsymbol{\Sigma}}^{-1/2}$ and it's posterior distribution is unconditional.

The posterior distributions are proper as long as $n > \max\{m+p-1, m+2p-\alpha, 3m+p-\delta-1, m+p-\alpha+\delta-1, 2m+p-\delta\}$.

Bayes Estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

$$\hat{\boldsymbol{B}}_{\text{BAYES}} = \mathrm{E}(\boldsymbol{B} \mid \boldsymbol{B}^*, \boldsymbol{S}^*) = \mathrm{E}_{\tilde{\boldsymbol{\Sigma}}}\, \mathrm{E}_{\boldsymbol{\Sigma}}\, \mathrm{E}(\boldsymbol{B} \mid \boldsymbol{B}^*, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}}) = \mathrm{E}_{\tilde{\boldsymbol{\Sigma}}}\, \mathrm{E}_{\boldsymbol{\Sigma}}\, \mathrm{E}(\boldsymbol{B}^*) = \boldsymbol{B}^*$$

Finding $\hat{\boldsymbol{\Sigma}}_{\text{BAYES}}$ seems very difficult.

$$\widehat{|\boldsymbol{\Sigma}|}_{\text{BAYES}} = \mathrm{E}\left(|\boldsymbol{\Sigma}| \mid \boldsymbol{B}^*, \boldsymbol{S}^*\right) = \mathrm{E}_{\tilde{\boldsymbol{\Sigma}}}\, \mathrm{E}\left(|\boldsymbol{\Sigma}| \mid \tilde{\boldsymbol{\Sigma}}, \boldsymbol{S}^*\right) = \mathrm{E}_{\tilde{\boldsymbol{\Sigma}}}\left(\left|\tilde{\boldsymbol{\Sigma}}\right| \mathrm{E}\left(\left|\tilde{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\tilde{\boldsymbol{\Sigma}}^{-1/2}\right|\right) \mid \boldsymbol{S}^*\right)$$
$$= \left(\prod_{j=1}^{m} \frac{n-p+\alpha-\delta-j+1}{n-p+\delta-m-j-2}\right) \mathrm{E}\left(\left|\tilde{\boldsymbol{\Sigma}}\right| \mid \boldsymbol{S}^*\right)$$
$$= \left(\prod_{j=1}^{m} \frac{n-p+\alpha-\delta-j+1}{(n-p+\delta-m-j-2)(n-p+\delta-2m-j-1)}\right) |\boldsymbol{S}^*|$$

provided that $n > \max\{m+p-1, m+2p-\alpha, m+p-\alpha+\delta-1, 3m+p-\delta+3\}$, by using (4.12) and (4.23).

Credible Sets for $|\boldsymbol{\Sigma}|$ and $\boldsymbol{M}$

Let $\boldsymbol{C} = \tilde{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\tilde{\boldsymbol{\Sigma}}^{-1/2}$. Then by (6.29), we have $\boldsymbol{C}^{-1} \sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}\left(\frac{n-p+\delta-m-1}{2}, \frac{n-p+\alpha-\delta}{2}\right)$. Also by (4.13) we have,

$$\frac{\left|\tilde{\boldsymbol{\Sigma}}^{-1}\right|}{\left|\boldsymbol{S}^{*-1}/(n-p)\right|} \sim \prod_{j=1}^{m} v_j, \quad \text{where } v_j \sim \chi^2_{n-p+\delta-2m-j+1} \text{ independently for } j = 1, \ldots, m$$

We can define a pivot for $|\boldsymbol{\Sigma}|$ in the same manner as in the last section to be $N \coloneqq \left|\boldsymbol{\Sigma}\boldsymbol{S}^{*-1}\right|$ where

$$(n-p)^{-m}\, N^{-1} \sim |\boldsymbol{M}| \prod_{j=1}^{m} v_j$$

where v_j's are defined as above, $\boldsymbol{M} \sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}\left(\frac{n-p+\delta-m-1}{2}, \frac{n-p+\alpha-\delta}{2}\right)$ and $\boldsymbol{M}$ is independent of each v_j. Since the distribution of N is free of $(\boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}})$ we conclude that it is a

pivot. A $(1-\gamma)$ level credible set for $|\boldsymbol{\Sigma}|$ is

$$[a_{n,p,\alpha,\delta;\gamma}\,|\boldsymbol{S}^*|\,, b_{n,p,\alpha,\delta;\gamma}\,|\boldsymbol{S}^*|]$$

where $a_{n,p,\alpha,\delta;\gamma}$ and $b_{n,p,\alpha,\delta;\gamma}$ are any two constants that satisfy $1-\gamma = P(a_{n,p,\alpha,\delta;\gamma} \leq N \leq b_{n,p,\alpha,\delta;\gamma})$. The length of the credible interval is $|\boldsymbol{S}^*|\,(b_{n,p,\alpha,\delta;\gamma} - a_{n,p,\alpha,\delta;\gamma})$.

Next we define the pivot for $\boldsymbol{B}$ as

$$T \coloneqq \frac{|(\boldsymbol{B}-\boldsymbol{B}^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{B}-\boldsymbol{B}^*)|}{|(n-p)\boldsymbol{S}^*|}$$

We will prove that T is a pivot and derive a sampling scheme in what follows. Since $(\boldsymbol{X}'\boldsymbol{X})^{1/2}(\boldsymbol{B}-\boldsymbol{B}^*) \sim N_{p,m}(\boldsymbol{O}, (\boldsymbol{\Sigma}+2\tilde{\boldsymbol{\Sigma}}) \otimes \boldsymbol{I}_p)$ due to (6.30), then for $p \geq m$,

$$(\boldsymbol{B}-\boldsymbol{B}^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{B}-\boldsymbol{B}^*)\,|\,\boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}} \sim \mathcal{W}_m(\boldsymbol{\Sigma}+2\tilde{\boldsymbol{\Sigma}}, p)$$

Therefore, for

$$\boldsymbol{H} = (\boldsymbol{\Sigma}+2\tilde{\boldsymbol{\Sigma}})^{-1/2}(\boldsymbol{B}-\boldsymbol{B}^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{B}-\boldsymbol{B}^*)(\boldsymbol{\Sigma}+2\tilde{\boldsymbol{\Sigma}})^{-1/2}$$
$$\boldsymbol{G} = \boldsymbol{S}^{*-1/2}\tilde{\boldsymbol{\Sigma}}^{1/2}(\tilde{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\tilde{\boldsymbol{\Sigma}}^{-1/2}+2\boldsymbol{I}_m)\tilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{S}^{*-1/2}$$

we have $\boldsymbol{H}\,|\,\boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}} \sim \mathcal{W}_m(\boldsymbol{I}_m, p)$, so $\boldsymbol{H}$ is independent of $(\boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}})$ and $\boldsymbol{H} \sim \mathcal{W}_m(\boldsymbol{I}_m, p)$. That implies $\boldsymbol{H}$ is independent of $\boldsymbol{G}$, and we show that the distribution of $\boldsymbol{G}$ is free of $(\boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}})$. In a similar manner as in Section 4.2, we notice that $\boldsymbol{G}\,|\,\boldsymbol{G}_0 \sim \mathrm{GB}_{\mathrm{p}}^{\mathrm{II}}\left(\frac{n-p+\alpha-\delta}{2}, \frac{n-p-m+\delta-1}{2}; \boldsymbol{G}_0, \boldsymbol{O}\right)+2\boldsymbol{G}_0$, and $\boldsymbol{G}_0 \sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{I}_m, n-p-2m+\delta\right)$ due to (6.29) and (6.28) respectively. This concludes the proof of the fact that T is

a pivot for $\boldsymbol{B}$. Using (4.13) the distribution of T is given by

$$T \,|\, \boldsymbol{G} \sim \left(\prod_{i=1}^{m} \chi^2_{p-i+1} \right) \times |\boldsymbol{G}/(n-p)| \tag{6.31}$$

$$\boldsymbol{G} \,|\, \boldsymbol{G}_0 \sim \mathrm{GB}_{\mathrm{p}}^{\mathrm{II}} \left(\frac{n-p+\alpha-\delta}{2}, \frac{n-p-m+\delta-1}{2}; \boldsymbol{G}_0, \boldsymbol{O} \right) + 2\boldsymbol{G}_0 \tag{6.32}$$

$$\boldsymbol{G}_0 \sim \mathcal{W}_m^{-1} \left((n-p)\boldsymbol{I}_m, n-p-2m+\delta \right) \tag{6.33}$$

Remark 6.2.1. *If one wants to test the significance of a set of regression coefficients or more generally of a linear combination of these regression coefficients, $\boldsymbol{AB} = \boldsymbol{C}$ where $\boldsymbol{A}$ is a $k \times p$ matrix with $rank(\boldsymbol{A}) = k \leq p$ and $k \geq m$, one may define*

$$T_{\boldsymbol{C}} \coloneqq \frac{\left| (\boldsymbol{C} - \boldsymbol{AB}^*)' \left\{ \boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}' \right\}^{-1} (\boldsymbol{C} - \boldsymbol{AB}^*) \right|}{|(n-p)\boldsymbol{S}^*|} \tag{6.34}$$

and proceed by noting that

$$T_{\boldsymbol{C}} \,|\, \boldsymbol{G} \sim \left(\prod_{i=1}^{m} \chi^2_{k-i+1} \right) \times |\boldsymbol{G}/(n-p)| \tag{6.35}$$

$$\boldsymbol{G} \,|\, \boldsymbol{G}_0 \sim \mathrm{GB}_{\mathrm{p}}^{\mathrm{II}} \left(\frac{n-p+\alpha-\delta}{2}, \frac{n-p-m+\delta-1}{2}; \boldsymbol{G}_0, \boldsymbol{O} \right) + 2\boldsymbol{G}_0 \tag{6.36}$$

$$\boldsymbol{G}_0 \sim \mathcal{W}_m^{-1} \left((n-p)\boldsymbol{I}_m, n-p-2m+\delta \right) \tag{6.37}$$

Remark 6.2.2. *To infer about $\boldsymbol{ABD} = \boldsymbol{\Delta}$ where $\boldsymbol{A}$ is a $k \times p$ matrix as before and $\boldsymbol{D}$ is a $m \times r$ matrix with $rank(\boldsymbol{D}) = r \leq k$, we start with $\boldsymbol{\Delta}^* = \boldsymbol{AB}^*\boldsymbol{D}$ and propose to use*

$$T_{\boldsymbol{\Delta}} \coloneqq \frac{\left| (\boldsymbol{\Delta} - \boldsymbol{\Delta}^*)' \left\{ \boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}' \right\}^{-1} (\boldsymbol{\Delta} - \boldsymbol{\Delta}^*) \right|}{|(n-p)\boldsymbol{D}'\boldsymbol{S}^*\boldsymbol{D}|} \tag{6.38}$$

whose distribution is obtained as follows

$$T_{\boldsymbol{\Delta}} \,|\, \tilde{\boldsymbol{G}} \sim \left(\prod_{i=1}^{m} \chi^2_{r-i+1} \right) \times \left| \tilde{\boldsymbol{G}}/(n-p) \right| \tag{6.39}$$

$$\tilde{\boldsymbol{G}} \,|\, \tilde{\boldsymbol{G}}_0 \sim \mathrm{GB}_{\mathrm{r}}^{\mathrm{II}} \left(\frac{n-p+\alpha-\delta}{2}, \frac{n-p-2m+r+\delta-1}{2}; \tilde{\boldsymbol{G}}_0, \boldsymbol{O} \right) + 2\tilde{\boldsymbol{G}}_0 \tag{6.40}$$

$$\tilde{\boldsymbol{G}}_0 \sim \mathcal{W}_r^{-1} \left((n-p)\boldsymbol{I}_r, n-p-2m+\delta \right) \tag{6.41}$$

For (6.40), we use the following result: Let $\boldsymbol{V} \sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}(\mathrm{a}, \mathrm{b})$. Then for a constant matrix $\boldsymbol{A}$ of size $q \times p$ such that $rank(\boldsymbol{A}) = q \leq p$,

$$\boldsymbol{A}\boldsymbol{V}\boldsymbol{A}' \sim \mathrm{GB}_{\mathrm{q}}^{\mathrm{II}} \left(a, b - \frac{1}{2}(p-q); \boldsymbol{A}\boldsymbol{A}', \boldsymbol{O} \right) \tag{6.42}$$

6.3 Random design matrix $\boldsymbol{X}$

When the prediction variables $\boldsymbol{X}$, just as the dependent variables $\boldsymbol{Y}$, are observed, then it is appropriate to consider $\boldsymbol{X}$ as a random matrix. Following Bilodeau and Brenner (1999), the model most commonly encountered assumes

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \mathbb{E}, \quad \mathbb{E} \sim N_{n,m}(\boldsymbol{O}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n), \boldsymbol{X} \sim N_{n,p}(\boldsymbol{O}, \boldsymbol{\Omega} \otimes \boldsymbol{I}_n), \mathbb{E} \perp\!\!\!\perp \boldsymbol{X}$$

where the errors are independently distributed of the prediction variables. The conditional model

$$\boldsymbol{Y} \,|\, \boldsymbol{X} \sim N_{n,m}(\boldsymbol{X}\boldsymbol{B}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$$

is thus identical to the case of a fixed $\boldsymbol{X}$. Then $\hat{\boldsymbol{B}}$ is still the unbiased estimate of $\boldsymbol{B}$, the distribution of $\boldsymbol{S}$ is free of $\boldsymbol{X}$, and $\hat{\boldsymbol{B}}$ is still independent of $\boldsymbol{S}$. We proceed by assuming a Inv-Wishart prior on $\boldsymbol{\Omega}$ that is independent of the prior on $(\boldsymbol{B}, \boldsymbol{\Sigma})$, i.e. $\pi(\boldsymbol{B}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}) = \pi(\boldsymbol{B})\,\pi(\boldsymbol{\Sigma})\,\pi(\boldsymbol{\Omega})$ where a priori $\boldsymbol{\Omega} \sim \mathcal{W}_p^{-1}(\boldsymbol{V}_0, \nu_0)$, provided $\boldsymbol{V}_0 > 0$, $\nu_0 > p-1$. For the sake of simplicity, we consider $\boldsymbol{V}_0 = \boldsymbol{I}_p$. In the PIS case, the

posterior distributions are as follows

$$\begin{aligned}
\boldsymbol{\Omega} &\sim \mathcal{W}_p^{-1}(\boldsymbol{I}_p, \nu_0 + m) && (6.43)\\
\boldsymbol{S} \,|\, \boldsymbol{S}^* &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{S}^*, n-p-m+\delta-1\right) && (6.44)\\
\boldsymbol{\Sigma} \,|\, \boldsymbol{S} &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{S}, n-p-m+\delta-1\right) && (6.45)\\
\boldsymbol{B} \,|\, \boldsymbol{\Sigma}, \boldsymbol{S}, \boldsymbol{B}^*, \boldsymbol{\Omega} &\sim T_{p,m}\left(\frac{n-m-p+1}{2}, \boldsymbol{B}^*, \boldsymbol{\Omega}^{-1}, \boldsymbol{\Sigma}+\boldsymbol{S}\right) && (6.46)
\end{aligned}$$

where $T_{p,m}(n, \boldsymbol{M}, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$ denotes the matrix variate t-distribution as described in Gupta and Nagar (2000). We can reformulate the above posterior distributions as:

$$\begin{aligned}
\boldsymbol{\Omega} &\sim \mathcal{W}_p^{-1}(\boldsymbol{I}_p, \nu_0 + m) && (6.47)\\
\boldsymbol{S}^{*-1/2}\boldsymbol{S}\boldsymbol{S}^{*-1/2} &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{I}_m, n-p-m+\delta-1\right) && (6.48)\\
\boldsymbol{S}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{S}^{-1/2} &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{I}_m, n-p-m+\delta-1\right) && (6.49)\\
\boldsymbol{B} \,|\, \boldsymbol{\Sigma}, \boldsymbol{S}, \boldsymbol{B}^*, \boldsymbol{\Omega} &\sim T_{p,m}\left(\frac{n-m-p+1}{2}, \boldsymbol{B}^*, \boldsymbol{\Omega}^{-1}, \boldsymbol{\Sigma}+\boldsymbol{S}\right) && (6.50)
\end{aligned}$$

The posterior distributions are proper as long as $\nu_0 > p-1$, $n > \max\{m+p-1, 2m+p-\delta\}$.

In the PPS case, the posterior distributions are as follows:

$$\begin{aligned}
\boldsymbol{\Omega} &\sim \mathcal{W}_p^{-1}(\boldsymbol{I}_p, \nu_0 + m) && (6.51)\\
\tilde{\boldsymbol{\Sigma}} \,|\, \boldsymbol{S}^* &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{S}^*, n-p-2m+\delta\right) && (6.52)\\
\tilde{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\tilde{\boldsymbol{\Sigma}}^{-1/2} &\sim \mathrm{B}_{\mathrm{m}}^{\mathrm{II}}\left(\frac{n-p+\alpha-\delta}{2}, \frac{n-p-m+\delta-1}{2}\right) && (6.53)\\
\boldsymbol{B} \,|\, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{B}^*, \boldsymbol{\Omega} &\sim T_{p,m}\left(\frac{n-m-p+1}{2}, \boldsymbol{B}^*, \boldsymbol{\Omega}^{-1}, \boldsymbol{\Sigma}+2\tilde{\boldsymbol{\Sigma}}\right) && (6.54)
\end{aligned}$$

We can reformulate the above posterior distributions as:

$$\begin{aligned}
\boldsymbol{\Omega} &\sim \mathcal{W}_p^{-1}(\boldsymbol{I}_p, \nu_0 + m) && (6.55)\\
\boldsymbol{S}^{*-1/2}\tilde{\boldsymbol{\Sigma}}\boldsymbol{S}^{*-1/2} &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{I}_m, n-p-2m+\delta\right) && (6.56)\\
\tilde{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\tilde{\boldsymbol{\Sigma}}^{-1/2} &\sim \mathrm{B}_{\mathrm{m}}^{\mathrm{II}}\left(\frac{n-p+\alpha-\delta}{2}, \frac{n-p-m+\delta-1}{2}\right) && (6.57)\\
\boldsymbol{B} \mid \boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}}, \boldsymbol{B}^*, \boldsymbol{\Omega} &\sim T_{p,m}\left(\frac{n-m-p+1}{2}, \boldsymbol{B}^*, \boldsymbol{\Omega}^{-1}, \boldsymbol{\Sigma} + 2\tilde{\boldsymbol{\Sigma}}\right) && (6.58)
\end{aligned}$$

The posterior distributions are proper as long as $\nu_0 > p-1$, $n > \max\{m+p-1, m+2p-\alpha, 3m+p-\delta-1, m+p-\alpha+\delta-1, 2m+p-\delta\}$

The Bayes estimators and credible sets in both cases are still exactly the same.

Chapter 7

Bayesian Analysis of Multiply Imputed Synthetic Data under the Multivariate Regression Model

7.1 Plug In Sampling method

With the same setup as in in Section 6.1, we now generate M copies of the original data $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_M$ where $\boldsymbol{V}_i \overset{\text{iid}}{\sim} N_{n,m}(\boldsymbol{X}\hat{\boldsymbol{B}}, \boldsymbol{S} \otimes \boldsymbol{I}_n)$ for $i = 1, \ldots, m$. The sufficient statistics for the released data is given by $(\boldsymbol{B}_1^*, \boldsymbol{S}_1^*), \ldots, (\boldsymbol{B}_M^*, \boldsymbol{S}_M^*)$ where for $i = 1, \ldots, M$

$$\boldsymbol{B}_i^* = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}_i \overset{\text{iid}}{\sim} N_{p,m}(\hat{\boldsymbol{B}}, \boldsymbol{S} \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}) \tag{7.1}$$

$$(n-p)\boldsymbol{S}_i^* = (\boldsymbol{V}_i - \boldsymbol{X}\boldsymbol{B}_i^*)'(\boldsymbol{V}_i - \boldsymbol{X}\boldsymbol{B}_i^*) = \boldsymbol{V}_i'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{V}_i \overset{\text{iid}}{\sim} \mathcal{W}_m(\boldsymbol{S}, n-p) \tag{7.2}$$

It can be shown that, similar to Sections 3.1 and 5.1, $(\overline{\boldsymbol{B}}_M^*, \tilde{\boldsymbol{S}}_M^*)$ is sufficient for $(\boldsymbol{B}, \boldsymbol{\Sigma})$ where $\overline{\boldsymbol{B}}_M^* = \frac{1}{M}\sum_{i=1}^M \boldsymbol{B}_i^*$, $\tilde{\boldsymbol{S}}_M^* = \sum_{i=1}^M (\boldsymbol{B}_i^* - \overline{\boldsymbol{B}}_M^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{B}_i^* - \overline{\boldsymbol{B}}_M^*) + (n-p)\sum_{i=1}^M \boldsymbol{S}_i^*$.

We follow the same procedure as in Section 6.1 to obtain the conditional posterior distributions which we state below without proof

$$\begin{aligned}
\boldsymbol{S} \,|\, \tilde{\boldsymbol{S}}_M^* &\sim \mathcal{W}_m^{-1}\left(\tilde{\boldsymbol{S}}_M^*, Mn-p-m+\delta-1\right) && (7.3)\\
\boldsymbol{\Sigma} \,|\, \boldsymbol{S} &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{S}, n-p-m+\delta-1\right) && (7.4)\\
\boldsymbol{B} \,|\, \boldsymbol{\Sigma}, \boldsymbol{S}, \overline{\boldsymbol{B}}_M^* &\sim N_{p,m}\left(\overline{\boldsymbol{B}}_M^*, (\boldsymbol{\Sigma}+\boldsymbol{S}/M)\otimes(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) && (7.5)
\end{aligned}$$

We can reformulate the above posterior distributions as:

$$\begin{aligned}
{\tilde{\boldsymbol{S}}_M^*}^{-1/2}\boldsymbol{S}{\tilde{\boldsymbol{S}}_M^*}^{-1/2} &\sim \mathcal{W}_m^{-1}\left(\boldsymbol{I}_m, Mn-p-m+\delta-1\right) && (7.6)\\
\boldsymbol{S}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{S}^{-1/2} &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{I}_m, n-p-m+\delta-1\right) && (7.7)\\
\boldsymbol{B} \,|\, \boldsymbol{\Sigma}, \boldsymbol{S}, \overline{\boldsymbol{B}}_M^* &\sim N_{p,m}\left(\overline{\boldsymbol{B}}_M^*, (\boldsymbol{\Sigma}+\boldsymbol{S}/M)\otimes(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) && (7.8)
\end{aligned}$$

which has the benefit that ${\tilde{\boldsymbol{S}}_M^*}^{-1/2}\boldsymbol{S}{\tilde{\boldsymbol{S}}_M^*}^{-1/2}$ is independent of $\boldsymbol{S}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{S}^{-1/2}$ and their posterior distributions are unconditional.

The posterior distributions are proper as long as $n > \max\{m+p-1, 2m+p-\delta\}$.

Bayes Estimators of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$

$$\begin{aligned}
\hat{\boldsymbol{B}}_{\text{BAYES}} &= \mathrm{E}(\boldsymbol{B} \,|\, \overline{\boldsymbol{B}}_M^*, \tilde{\boldsymbol{S}}_M^*) = \mathrm{E}_{\boldsymbol{S}}\,\mathrm{E}_{\boldsymbol{\Sigma}}\,\mathrm{E}(\boldsymbol{B} \,|\, \overline{\boldsymbol{B}}_M^*, \boldsymbol{\Sigma}, \boldsymbol{S}) = \mathrm{E}_{\boldsymbol{S}}\,\mathrm{E}_{\boldsymbol{\Sigma}}\,\mathrm{E}(\overline{\boldsymbol{B}}_M^*) = \overline{\boldsymbol{B}}_M^*\\
\hat{\boldsymbol{\Sigma}}_{\text{BAYES}} &= \mathrm{E}(\boldsymbol{\Sigma} \,|\, \overline{\boldsymbol{B}}_M^*, \tilde{\boldsymbol{S}}_M^*) = \mathrm{E}_{\boldsymbol{S}}\,\mathrm{E}(\boldsymbol{\Sigma} \,|\, \tilde{\boldsymbol{S}}_M^*, \boldsymbol{S}) = \mathrm{E}_{\boldsymbol{S}}\left(\tfrac{(n-p)\boldsymbol{S}}{(n-p-2m+\delta-2)} \,|\, \tilde{\boldsymbol{S}}_M^*\right)\\
&= \frac{(n-p)\tilde{\boldsymbol{S}}_M^*}{(n-p-2m+\delta-2)(Mn-p-2m+\delta-2)}
\end{aligned}$$

$$\begin{aligned}
\widehat{|\boldsymbol{\Sigma}|}_{\text{BAYES}} &= \mathrm{E}\left(|\boldsymbol{\Sigma}| \,|\, \overline{\boldsymbol{B}}_M^*, \tilde{\boldsymbol{S}}_M^*\right) = \mathrm{E}_{\boldsymbol{S}}\,\mathrm{E}\left(|\boldsymbol{\Sigma}| \,|\, \boldsymbol{S}, \tilde{\boldsymbol{S}}_M^*\right) = \mathrm{E}_{\boldsymbol{S}}\left(|\boldsymbol{S}|\,\mathrm{E}\left(\left|\boldsymbol{S}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{S}^{-1/2}\right|\right) \,|\, \tilde{\boldsymbol{S}}_M^*\right)\\
&= \left(\prod_{j=1}^{m}\frac{n-p}{n-p-m+\delta-j-2}\right)\mathrm{E}\left(|\boldsymbol{S}| \,|\, \tilde{\boldsymbol{S}}_M^*\right) = \left(\prod_{j=1}^{m}\tfrac{n-p}{(Mn-p-m+\delta-j-2)(n-p-m+\delta-j-2)}\right)\left|\tilde{\boldsymbol{S}}_M^*\right|
\end{aligned}$$

provided $n > \max\{m+p-1, 2m+p-\delta+4\}$, we use the results (4.11) and (4.12).

Credible Sets for $|\boldsymbol{\Sigma}|$ and $\boldsymbol{B}$

In the same way as in Section 6.1 we can define a pivot for $|\boldsymbol{\Sigma}|$ as $N_M := \left|\boldsymbol{\Sigma}\tilde{\boldsymbol{S}}_M^{*\,-1}\right|$ where

$$(n-p)^{-m}\, N_M^{-1} \sim \left(\prod_{i=1}^{m} u_i\right)\left(\prod_{j=1}^{m} v_j\right)$$

where $u_i \sim \chi^2_{n-p-m+\delta-i}$ independently for $i = 1, \ldots, p$; $v_j \sim \chi^2_{Mn-p-m+\delta-j}$ independently for $j = 1, \ldots, p$; and they are all pairwise independent. A $(1-\gamma)$ level credible set for $|\boldsymbol{\Sigma}|$ based on N_M is

$$\left[a_{n,p,m,\delta;\gamma}\left|\tilde{\boldsymbol{S}}_M^*\right|, b_{n,p,m,\delta;\gamma}\left|\tilde{\boldsymbol{S}}_M^*\right|\right]$$

where $a_{n,p,m,\delta;\gamma}$ and $b_{n,p,m,\delta;\gamma}$ are any two constants that satisfy $1-\gamma = P(a_{n,p,m,\delta;\gamma} \leq N_M \leq b_{n,p,m,\delta;\gamma})$. The length of the credible interval is $\left|\tilde{\boldsymbol{S}}_M^*\right|(b_{n,p,m,\delta;\gamma} - a_{n,p,m,\delta;\gamma})$.

Next we define the pivot for $\boldsymbol{B}$ as

$$T_M := \frac{\left|(\boldsymbol{B}-\overline{\boldsymbol{B}}_M^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{B}-\overline{\boldsymbol{B}}_M^*)\right|}{\left|\tilde{\boldsymbol{S}}_M^*\right|}$$

We will derive the distribution of T_M similarly as in Section 6.1. We first notice that due to (7.5), for $p \geq m$,

$$(\boldsymbol{B}-\overline{\boldsymbol{B}}_M^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{B}-\overline{\boldsymbol{B}}_M^*) \mid \boldsymbol{\Sigma}, \boldsymbol{S} \sim \mathcal{W}_m(\boldsymbol{\Sigma}+\boldsymbol{S}/M, p)$$

Thus for

$$\boldsymbol{H}_M = (\boldsymbol{\Sigma}+\boldsymbol{S}/M)^{-1/2}(\boldsymbol{B}-\overline{\boldsymbol{B}}_M^*)'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{B}-\overline{\boldsymbol{B}}_M^*)(\boldsymbol{\Sigma}+\boldsymbol{S}/M)^{-1/2}$$

$$\boldsymbol{G}_M = \tilde{\boldsymbol{S}}_M^{*\,-1/2}\boldsymbol{S}^{1/2}(\boldsymbol{S}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{S}^{-1/2}+\boldsymbol{I}_m/M)\boldsymbol{S}^{1/2}\tilde{\boldsymbol{S}}_M^{*\,-1/2}$$

we have $\boldsymbol{H}_M \,|\, \boldsymbol{\Sigma}, \boldsymbol{S} \sim \mathcal{W}_m(\boldsymbol{I}_m, p)$, so $\boldsymbol{H}_M$ is independent of $(\boldsymbol{\Sigma}, \boldsymbol{S})$ and $\boldsymbol{H}_M \sim \mathcal{W}_m(\boldsymbol{I}_m, p)$. That implies $\boldsymbol{H}_M$ is independent of $\boldsymbol{G}_M$, and we show that the distribution of $\boldsymbol{G}_M$ is free of $(\boldsymbol{\Sigma}, \boldsymbol{S})$. In a similar manner as in Section 6.1, we notice that $\boldsymbol{G}_M \,|\, \boldsymbol{G}_{0,M} \sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{G}_{0,M}, n-p-m+\delta-1\right) + \boldsymbol{G}_{0,M}/M$, and $\boldsymbol{G}_{0,M} \sim \mathcal{W}_m^{-1}\left(\boldsymbol{I}_m, Mn-p-m+\delta-1\right)$ due to (7.7) and (7.6) respectively. This concludes the proof of the fact that T_M is a pivot for $\boldsymbol{B}$. Using (4.13) the distribution of T is given by

$$
\begin{aligned}
T_M \,|\, \boldsymbol{G}_M &\sim \left(\prod_{i=1}^{m} \chi^2_{p-i+1}\right) \times |\boldsymbol{G}_M/(n-p)| && (7.9)\\
\boldsymbol{G}_M \,|\, \boldsymbol{G}_{0,M} &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{G}_{0,M}, n-p-m+\delta-1\right) + \boldsymbol{G}_{0,M}/M && (7.10)\\
\boldsymbol{G}_{0,M} &\sim \mathcal{W}_m^{-1}\left(\boldsymbol{I}_m, Mn-p-m+\delta-1\right) && (7.11)
\end{aligned}
$$

Remark 7.1.1. *If one wants to test the significance of a set of regression coefficients or more generally of a linear combination of these regression coefficients,* $\boldsymbol{AB} = \boldsymbol{C}$ *where* $\boldsymbol{A}$ *is a* $k \times p$ *matrix with* $rank(\boldsymbol{A}) = k \leq p$ *and* $k \geq m$*, one may define*

$$
T_{\boldsymbol{C},M} \coloneqq \frac{\left|(\boldsymbol{C} - \boldsymbol{A}\overline{\boldsymbol{B}}^*_M)' \left\{\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right\}^{-1} (\boldsymbol{C} - \boldsymbol{A}\overline{\boldsymbol{B}}^*_M)\right|}{\left|(n-p)\tilde{\boldsymbol{S}}^*_M\right|} \tag{7.12}
$$

and proceed by noting that

$$
\begin{aligned}
T_{\boldsymbol{C},M} \,|\, \boldsymbol{G}_M &\sim \left(\prod_{i=1}^{m} \chi^2_{k-i+1}\right) \times |\boldsymbol{G}_M/(n-p)| && (7.13)\\
\boldsymbol{G}_M \,|\, \boldsymbol{G}_{0,M} &\sim \mathcal{W}_m^{-1}\left((n-p)\boldsymbol{G}_{0,M}, n-p-m+\delta-1\right) + \boldsymbol{G}_{0,M}/M && (7.14)\\
\boldsymbol{G}_{0,M} &\sim \mathcal{W}_m^{-1}\left(\boldsymbol{I}_m, Mn-p-m+\delta-1\right) && (7.15)
\end{aligned}
$$

Remark 7.1.2. *To infer about* $\boldsymbol{ABD} = \boldsymbol{\Delta}$ *where* $\boldsymbol{A}$ *is a* $k \times p$ *matrix as before and* $\boldsymbol{D}$ *is a* $m \times r$ *matrix with* $rank(\boldsymbol{D}) = r \leq k$*, we start with* $\boldsymbol{\Delta}^* = \boldsymbol{A}\overline{\boldsymbol{B}}^*_M\boldsymbol{D}$ *and propose*

to use

$$T_{\boldsymbol{\Delta},M} \coloneqq \frac{\left|(\boldsymbol{\Delta}-\boldsymbol{\Delta}^*)'\left\{\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\right\}^{-1}(\boldsymbol{\Delta}-\boldsymbol{\Delta}^*)\right|}{\left|(n-p)\boldsymbol{D}'\tilde{\boldsymbol{S}}_M^*\boldsymbol{D}\right|} \tag{7.16}$$

whose distribution is obtained as follows

$$T_{\boldsymbol{\Delta}} \,|\, \tilde{\boldsymbol{G}}_M \sim \left(\prod_{i=1}^{m}\chi^2_{r-i+1}\right) \times \left|\tilde{\boldsymbol{G}}_M/(n-p)\right| \tag{7.17}$$

$$\tilde{\boldsymbol{G}}_M \,|\, \tilde{\boldsymbol{G}}_{0,M} \sim \mathcal{W}_r^{-1}\Big((n-p)\tilde{\boldsymbol{G}}_{0,M}, n-p-m+\delta-1\Big) + \tilde{\boldsymbol{G}}_{0,M}/M \tag{7.18}$$

$$\tilde{\boldsymbol{G}}_{0,M} \sim \mathcal{W}_r^{-1}(\boldsymbol{I}_r, Mn-p-m+\delta-1) \tag{7.19}$$

7.2 Posterior Predictive Sampling method

We adapt the procedure in Section 6.2 in the multiple imputation scenario in what follows. We begin by producing M posterior draws of the parameters $(\tilde{\boldsymbol{B}}_1, \tilde{\boldsymbol{\Sigma}}_1), \ldots, (\tilde{\boldsymbol{B}}_M, \tilde{\boldsymbol{\Sigma}}_M)$ from (6.22), which we use to spawn M samples of the released data $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_M$ where $\boldsymbol{V}_i \sim N_{n,m}(\boldsymbol{X}\tilde{\boldsymbol{B}}_i, \tilde{\boldsymbol{\Sigma}}_i \otimes \boldsymbol{I}_n)$ drawn independently for $i = 1, \ldots, m$. The sufficient statistics for the released data is given by $(\boldsymbol{B}_1^*, \boldsymbol{S}_1^*), \ldots, (\boldsymbol{B}_M^*, \boldsymbol{S}_M^*)$ where independently for $i = 1, \ldots, M$

$$\boldsymbol{B}_i^* = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}_i \sim N_{p,m}(\tilde{\boldsymbol{B}}_i, \tilde{\boldsymbol{\Sigma}}_i \otimes \boldsymbol{X}'\boldsymbol{X}^{-1}) \tag{7.20}$$

$$(n-p)\boldsymbol{S}_i^* = \boldsymbol{V}_i'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{V}_i \sim \mathcal{W}_m\Big(\tilde{\boldsymbol{\Sigma}}_i, n-p\Big) \tag{7.21}$$

We then follow the same procedure as in Section 6.2 to obtain the conditional posterior distributions which we state below without proof

$$\begin{aligned}&\boldsymbol{B} \,|\, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}}_1, \ldots, \tilde{\boldsymbol{\Sigma}}_M, \boldsymbol{B}_1^*, \ldots, \boldsymbol{B}_M^* \\ &\sim N_{p,m}\left(\left(\sum_{i=1}^{M}\boldsymbol{B}_i^*\tilde{\boldsymbol{\Sigma}}_i^{-1}\right)\left(\sum_{i=1}^{M}\tilde{\boldsymbol{\Sigma}}_i^{-1}\right)^{-1}, \left(\boldsymbol{\Sigma} + 2\left(\sum_{i=1}^{M}\tilde{\boldsymbol{\Sigma}}_i^{-1}\right)^{-1}\right) \otimes \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)\end{aligned} \tag{7.22}$$

$$\left(\sum_{i=1}^{M}\tilde{\boldsymbol{\Sigma}}_i^{-1}\right)^{1/2}\boldsymbol{\Sigma}\left(\sum_{i=1}^{M}\tilde{\boldsymbol{\Sigma}}_i^{-1}\right)^{1/2}\sim \mathrm{B}_{\mathrm{m}}^{\mathrm{II}}\left(\frac{M(n-p+\alpha-m-1)+m-\delta+1}{2},\frac{n-p+\delta-m-1}{2}\right) \tag{7.23}$$

and the latent matrices have the following distribution

$$\begin{aligned}
&g(\tilde{\boldsymbol{\Sigma}}_1,\ldots,\tilde{\boldsymbol{\Sigma}}_M\,|\,\boldsymbol{B}_1^*,\ldots,\boldsymbol{B}_M^*,\boldsymbol{S}_1^*,\ldots,\boldsymbol{S}_M^*)\\
&\propto\frac{\left|\sum_{i=1}^{M}\tilde{\boldsymbol{\Sigma}}_i^{-1}\right|^{-\frac{M(n-p+\alpha-m-1)+2m-\delta+p}{2}}}{\left(\prod_{i=1}^{M}\left|\tilde{\boldsymbol{\Sigma}}_i\right|^{\frac{2n-p+\alpha}{2}}\right)}\operatorname{etr}\left[-\frac{n-p}{2}\sum_{i=1}^{M}\tilde{\boldsymbol{\Sigma}}_i^{-1}\boldsymbol{S}_i^*\right]\\
&\times\operatorname{etr}\left[-\frac{1}{2}\sum_{i=1}^{M}\left(2\tilde{\boldsymbol{\Sigma}}_i\right)^{-1}\left(\boldsymbol{B}_i^*-\left(\sum_{i=1}^{M}\boldsymbol{B}_i^*\tilde{\boldsymbol{\Sigma}}_i^{-1}\right)\left(\sum_{i=1}^{M}\tilde{\boldsymbol{\Sigma}}_i^{-1}\right)^{-1}\right)'\right.\\
&\qquad\left.(\boldsymbol{X}'\boldsymbol{X})\left(\boldsymbol{B}_i^*-\left(\sum_{i=1}^{M}\boldsymbol{B}_i^*\tilde{\boldsymbol{\Sigma}}_i^{-1}\right)\left(\sum_{i=1}^{M}\tilde{\boldsymbol{\Sigma}}_i^{-1}\right)^{-1}\right)\right]
\end{aligned} \tag{7.24}$$

where we can notice that the quantity inside the last exponential vanishes when $M=1$.

The posterior distributions for the parameters exist provided $n>\max\{p,2m-\delta+p,$ $p-\alpha+m+1+\frac{\delta-2}{M}\}$.

We can double check the accuracy of our results in Chapters 6 and 7 by verifying that they match those of Chapters 4 and 5 when we plug in $m=p$, $p=1$.

FPPS

For the FPPS case, we produce a single copy $(\tilde{\boldsymbol{B}},\tilde{\boldsymbol{\Sigma}})$ from the posterior distribution (6.22) and use it to create M samples $\boldsymbol{V}_i\overset{\text{iid}}{\sim}N_{n,m}(\boldsymbol{X}\tilde{\boldsymbol{B}},\tilde{\boldsymbol{\Sigma}}\otimes\boldsymbol{I}_n)$ for $i=1,\ldots,m$. Using the same notation as in Section 7.1, we derive the posterior distribution of the

parameters as follows with exactly the same conditions for existence as in Section 6.2

$$\begin{aligned}
\tilde{\boldsymbol{\Sigma}} \,|\, \tilde{\boldsymbol{S}}_M^* &\sim \mathcal{W}_m^{-1}\left(\tilde{\boldsymbol{S}}_M^*, Mn - p - 2m + \delta\right) && (7.25)\\
\tilde{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\tilde{\boldsymbol{\Sigma}}^{-1/2} &\sim \mathrm{B}_{\mathrm{m}}^{\mathrm{II}}\left(\frac{n-p+\alpha-\delta}{2}, \frac{n-p-m+\delta-1}{2}\right) && (7.26)\\
\boldsymbol{B} \,|\, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}}, \overline{\boldsymbol{B}}_M^* &\sim N_{p,m}\left(\overline{\boldsymbol{B}}_M^*, \left(\boldsymbol{\Sigma} + \left(1 + \frac{1}{M}\right)\tilde{\boldsymbol{\Sigma}}\right) \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\right) && (7.27)
\end{aligned}$$

We can reformulate the above posterior distributions as:

$$\begin{aligned}
{\tilde{\boldsymbol{S}}_M^*}^{-1/2}\tilde{\boldsymbol{\Sigma}}{\tilde{\boldsymbol{S}}_M^*}^{-1/2} &\sim \mathcal{W}_m^{-1}\left(\boldsymbol{I}_m, Mn - p - 2m + \delta\right) && (7.28)\\
\tilde{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\Sigma}\tilde{\boldsymbol{\Sigma}}^{-1/2} &\sim \mathrm{B}_{\mathrm{m}}^{\mathrm{II}}\left(\frac{n-p+\alpha-\delta}{2}, \frac{n-p-m+\delta-1}{2}\right) && (7.29)\\
\boldsymbol{B} \,|\, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\Sigma}}, \overline{\boldsymbol{B}}_M^* &\sim N_{p,m}\left(\overline{\boldsymbol{B}}_M^*, \left(\boldsymbol{\Sigma} + \left(1 + \frac{1}{M}\right)\tilde{\boldsymbol{\Sigma}}\right) \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\right) && (7.30)
\end{aligned}$$

Chapter 8

Future Work

In this concluding chapter, we point out a few directions to which future research can be carried out.

1. We would like to examine how the inference is affected by the choice of other (non-informative) priors, probability matching priors, and also prions which are conjugate in nature with a suitable choice of the hyperparameters so as not to affect data influence.

2. It is clear from the simulation results in the preceding chapters that the coverage is a decreasing function of δ. It is desirable to express the nature of this dependence exactly, or even within bounds.

3. We can also look into construction of highest posterior density (HPD) sets of the parameters discussed in various chapters of this dissertation. This will necessitate a judicial choice of the cut-off points of the proposed credible sets.

4. Another future research topic would be to explore conditions for valid inference for the scenarios we have discussed here and consider extensions of our

methodology to non-ideal situations that frequently mar real life data (i.e., scenarios where the imputer and/or data analyst overfit or underfit the regression model; the imputer's model is the regression of y on x, but the data analyst's model is the regression of x on y; non-normal errors; y's have unequal variances and/or are correlated; only part of y is sensitive; response and covariates are all sensitive; original data contain missing observations; multiple y-variables are synthesized, while multiple x-variables are not; the original data are from a census, not a sample; and so on). It is also imperative to verify our results by application to real-life data, like the application to CPS data as carried out in Klein and Sinha (2015a), Klein and Sinha (2015b), and Klein et al. (2019).

5. In the context of Chapter 3 and Chapter 5, it is desirable to carry out the frequentist analysis of the MI PPS MLR and MI MVN cases as well. We can also look into the issue of sampling of latent matrices in the case of multiply imputed MVN and MVR PPS data which has not been addressed. Another task would be to finish the partially synthetic data case for MI PPS MVN and MVR data.

6. Since one of our prime objectives is to provide valid inference while protecting privacy, we would like to devise methods to quantify privacy in the synthetic data (for e.g., Disclosure Risk Analysis as discussed in Klein and Sinha (2015b)) and observe the trade-off between quality of inference and privacy of survey respondents. It is worth mentioning here that since the data generating methods are still the same as in the frequentist case, the disclosure risk is the same for the cases considered here as in Klein and Sinha (2015a), Klein and Sinha (2015b), and Klein et al. (2019).

7. An excellent new direction of research would be to beyond the models considered here, and to develop both frequentist and Bayesian analysis of singly

and multiply imputed data under a GLM framework; and also based on Noise Multiplied data, building on the work of Klein et al. (2014), Klein and Sinha (2015a) and Klein and Sinha (2015b).

Appendices

Proof of (4.12)

Let us find out $\mathrm{E}(|\boldsymbol{S}|)$ when $\boldsymbol{S} \sim \mathcal{W}_p^{-1}(\nu, \boldsymbol{\Sigma})$. We denote $\Gamma_m(a)$ to be the multivariate gamma function given by $\Gamma_m(a) = \pi^{\frac{m(m-1)}{4}} \prod_{i=1}^{m} \Gamma\left(a - \frac{1}{2}(j-1)\right)$ for $a > \frac{m-1}{2}$.

$$\begin{aligned}
\mathrm{E}(|\boldsymbol{S}|) &= \frac{|\boldsymbol{\Sigma}|^{\frac{\nu}{2}}}{2^{\frac{\nu p}{2}}\,\Gamma_p\left(\frac{\nu}{2}\right)} \int |\boldsymbol{S}|^{-\frac{\nu-2+p+1}{2}} \operatorname{etr}\left(-\tfrac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\right)\right) d\boldsymbol{S} \\
&= \frac{2^{\frac{(\nu-2)p}{2}}\,\Gamma_p\left(\frac{\nu-2}{2}\right)}{2^{\frac{\nu p}{2}}\,\Gamma_p\left(\frac{\nu}{2}\right)} \times |\boldsymbol{\Sigma}| \\
&= \frac{|\boldsymbol{\Sigma}|}{2^p} \times \frac{\Gamma_p\left(\frac{\nu}{2}-1\right)}{\Gamma_p\left(\frac{\nu}{2}\right)} \\
&= \frac{|\boldsymbol{\Sigma}|}{2^p} \times \frac{\prod_{j=1}^{p} \Gamma\left(\frac{\nu}{2}-1-\frac{1}{2}(j-1)\right)}{\prod_{j=1}^{p} \Gamma\left(\frac{\nu}{2}-\frac{1}{2}(j-1)\right)} \\
&= \frac{|\boldsymbol{\Sigma}|}{2^p} \times \frac{\prod_{j=1}^{p} \Gamma\left(\frac{\nu}{2}-1-\frac{1}{2}(j-1)\right)}{\prod_{j=1}^{p} \left(\frac{\nu}{2}-1-\frac{1}{2}(j-1)\right) \Gamma\left(\frac{\nu}{2}-1-\frac{1}{2}(j-1)\right)} \\
&= |\boldsymbol{\Sigma}| \prod_{j=1}^{p} (\nu - j - 1)^{-1}
\end{aligned}$$

The second equality hold true when $\nu > p+3$, to ensure the pdf of the Inverse-Wishart distribution inside the integral is defined.

Proof of (4.23)

Let us find out $\mathrm{E}(|\boldsymbol{V}|)$ when $\boldsymbol{V} \sim \mathrm{B_p^{II}}(a, b)$. We denote $\mathrm{B_m}(a, b)$ to be the multivariate beta function given by $\mathrm{B_m}(a, b) = \Gamma_m(a)\Gamma_m(b)/\Gamma_m(a+b)$.

$$\begin{aligned}
\mathrm{E}(|\boldsymbol{V}|) &= \mathrm{B_p}(a, b)^{-1} \int |\boldsymbol{V}|^{\mathrm{a}+1-\frac{\mathrm{p}+1}{2}} \, |\boldsymbol{I}_\mathrm{p} + \boldsymbol{V}|^{-(\mathrm{a}+1+\mathrm{b}-1)} \, \mathrm{d}\boldsymbol{V} \\
&= \frac{\mathrm{B_p}(a+1, b-1)}{\mathrm{B_p}(a, b)} \\
&= \frac{\Gamma_p(a+1)\Gamma_p(b-1)}{\Gamma_p(a+b)} \times \frac{\Gamma_p(a+b)}{\Gamma_p(a)\Gamma_p(b)} \\
&= \prod_{j=1}^{p} \frac{\Gamma\left(a+1-\frac{1}{2}(j-1)\right)\Gamma\left(b-1-\frac{1}{2}(j-1)\right)}{\Gamma\left(a-\frac{1}{2}(j-1)\right)\Gamma\left(b-\frac{1}{2}(j-1)\right)} = \prod_{j=1}^{p} \frac{a-\frac{1}{2}(j-1)}{b-\frac{1}{2}(j+1)}
\end{aligned}$$

The second equality requires the conditions $a+1 > \frac{p-1}{2}$, $b-1 > \frac{p-1}{2}$, to ensure the pdf of the matrix variate beta type II distribution inside the integral is defined. Now the first condition is already guaranteed, since $\boldsymbol{V} \sim \mathrm{B}_{\mathrm{p}}^{\mathrm{II}}(a,b)$ gives us $a > \frac{p-1}{2}$, and also $b > \frac{p-1}{2}$ which is not enough for the second one. So put together, for this result to hold true, we need $a > \frac{p-1}{2}$, $b > \frac{p+1}{2}$.

Bibliography

Abowd, J., Stinson, M., and Benedetto, G. (2006). Final Report to the Social Security Administration on the SIPP/SSA/IRS Public Use File Project. Technical report.

An, D. and Little, R. J. A. (2007). Multiple imputation: an alternative to top coding for statistical disclosure control. *Journal of the Royal Statistical Society: Series A*, 170(4):923–940.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley, 3 edition.

Benedetto, G., Stinson, M., and Abowd, J. (2013). The Creation and Use of the SIPP Synthetic Beta. Technical report.

Bilodeau, M. and Brenner, D. (1999). *Theory of Multivariate Statistics*. Springer.

Datta, G. and Ghosh, J. (1995). On Priors Providing Frequentist Validity for Bayesian Inference. *Biometrika*, 82(1):37–45.

Drechsler, J. (2010). *Generating Multiply Imputed Synthetic Datasets: Theory and Implementation*. PhD thesis, Otto-Friedrich-University Bamberg.

Drechsler, J. (2011). *Synthetic Datasets for Statistical Disclosure Control*. Springer, New York.

Gupta, A. and Nagar, D. (2000). *Matrix Variate Distributions*. CRC Press.

Hawala, S. (2008). Producing Partially Synthetic Data to Avoid Disclosure. *Proceedings of the Joint Statistical Meetings*. American Statistical Association.

Kinney, S., Reiter, J., and Miranda, J. (2014). SynLBD 2.0: Improving the Synthetic Longitudinal Business Database. *Statistical Journal of the International Association for Official Statistics*, 30(3):129–135.

Kinney, S., Reiter, J., Reznek, A., Miranda, J., Jarmin, R., and Abowd, J. (2011). Towards Unrestricted Public Use Business Microdata: The Synthetic Longitudinal Business Database. *International Statistical Review*, 79(3):362–384.

Klein, M., Mathew, T., and Sinha, B. (2014). Noise Multiplication for Statistical Disclosure Control of Extreme Values in Log-normal Regression Samples. *Journal of Privacy and Confidentiality*, 6(1).

Klein, M. and Sinha, B. (2015a). Inference for Singly Imputed Synthetic Data Based on Posterior Predictive Sampling under Multivariate Normal and Multiple Linear Regression Models. *Sankhya B*, 77(2):293–311.

Klein, M. and Sinha, B. (2015b). Likelihood Based Finite Sample Inference for Singly Imputed Synthetic Data Under the Multivariate Normal and Multiple Linear Regression Models. *Journal of Privacy and Confidentiality*, 7(1).

Klein, M., Zylstra, J., and Sinha, B. (2019). Finite Sample Inference for Multiply Imputed Synthetic Data under a Multiple Linear Regression Model. *Calcutta Statistical Association Bulletin*, 71(2):63–82.

Kollo, T. and Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*. Springer.

Little, R. (1993). Statistical Analysis of Masked Data. *Journal of Official Statistics*, 9(2):407–426.

Little, R. J. (2006). Calibrated Bayes: A Bayes/Frequentist Roadmap. *The American Statistician*, 60(3):213–223.

Machanavajjhala, A., Kifer, D., Abowd, J., Gehrke, J., and Vilhuber, L. (2008). Privacy: Theory Meets Practice on the Map. *IEEE 24th International Conference on Data Engineering*, pages 277–286.

Moura, R. (2016). *Likelihood-based Inference for Multivariate Regression Models using Synthetic Data*. PhD thesis, NOVA University of Lisbon.

Moura, R., Klein, M., Coelho, C., and Sinha, B. (2017a). Inference for multivariate regression model based on synthetic data generated under fixed-posterior predictive sampling: comparison with plug-in sampling. *REVSTAT*, 15(2):155–186.

Moura, R., Klein, M., Zylstra, J., Coelho, C., and Sinha, B. (2021). Inference for Multivariate Regression Model based on Synthetic Data generated using Plug-in Sampling. *Journal of American Statistical Association*, 116(534):720–733.

Moura, R., Sinha, B., and Coelho, C. (2017b). Inference for multivariate regression model based on multiply imputed synthetic data generated via posterior predictive sampling. *AIP Conference Proceedings*, 1836(1):020065.

Raghunathan, T., Reiter, J., and Rubin, D. (2003). Multiple Imputation for Statistical Disclosure Limitation. *Journal of Official Statistics*, 19(1):1–16.

Reiter, J. (2003). Inference for Partially Synthetic, Public Use Microdata Sets. *Survey Methodology*, 29(2):181–188.

Rubin, D. (1984). Bayesianly Justifiable and Relevant Frequency Calculations for the Applies Statistician. *The Annals of Statistics*, 12(4):1151–1172.

Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.

Rubin, D. (1993). Discussion: Statistical Disclosure Limitation. *Journal of Official Statistics*, 9(2):461–468.

Tweedie, M. C. K. (1957). Statistical Properties of Inverse Gaussian Distributions I. *Annals of Mathematical Statistics*, 28(2):362–377.