Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

SCIENZE STATISTICHE

Ciclo 35

Settore Concorsuale: 13/D1 - STATISTICA

Settore Scientifico Disciplinare: SECS-S/01 - STATISTICA

ON THE EFFECT OF CONFOUNDING IN LINEAR REGRESSION MODELS: AN APPROACH BASED ON THE THEORY OF QUADRATIC FORMS

Presentata da: Martina Narcisi

Coordinatore Dottorato

Monica Chiogna

Supervisore

Carlo Trivisano

Esame finale anno 2023

Abstract

The main topic of this thesis is confounding in linear regression models. It arises when a relationship between an observed process, the covariate, and an outcome process, the response, is influenced by an unmeasured process, the confounder, associated with both. Consequently, the estimators for the regression coefficients of the measured covariates might be severely biased, less efficient and characterized by misleading interpretations. In fact, confounding is an issue when the primary target of the work is estimation of the regression parameters. The central point of the dissertation is the evaluation of the sampling properties of parameter estimators. This work aims to extend the spatial confounding framework, widely addressed in the literature, to general structured settings and to understand the behaviour of confounding as a function of the data generating process structure parameters in several scenarios focusing on the joint covariate-confounder structure. In line with the spatial statistics literature, our purpose is to quantify the sampling properties of the regression coefficient estimators and, in turn, to identify the most prominent quantities depending on the generative mechanism impacting confounding. Once the bias, variance and mean square error of the estimator conditionally on the covariate process are derived as ratios of dependent quadratic forms in Gaussian random variables, we provide an analytic expression of the marginal sampling properties of the estimator by means of Carlson's R function. This allows the computation of the target quantities without simulation studies. In addition, we propose a representative scalar quantity for the extent of confounding as a proxy of the bias, its first order Laplace approximation. To conclude, we work under several frameworks that consider spatial and temporal data with specific assumptions regarding the covariance and cross-covariance functions used to generate the processes involved. This study allows us to claim that the variability of the confounder-covariate interaction and of the covariate plays the most relevant role in determining the principal marker of the magnitude of confounding, the estimator bias, and the other estimator sampling properties.

Contents

Chapter 1

Introduction

1.1 Overview

Usually, in linear regression models, there are unobserved variables that contribute to explain some of the variability of the response. The lack of knowledge of the latent processes, that influence the predictor of such models, leads to biased estimates of the parameters of interest. To overcome this problem in the attempt to account for the unmeasured variables, it is common practice to use a linear regression model in which a specific type of error structure matrix is assumed to be known. Otherwise, random effects are included in the model to account for the variation of the correlation among observations according to some unknown pattern. Often, neither being aware of residual covariance matrix nor modeling the random effect variations is sufficient to obtain reliable estimates.

Confounding arises when a relationship between an observed process, the covariate, and an outcome process, the response, is affected by an unmeasured process, the confounder, associated with both (see Figure 1.1). As a result, the estimates for the regression coefficients of the measured covariates can be severely biased, less efficient and characterized by a misleading interpretation. Taking into account confounding effects in the parameter estimation is a hard task. When the primary target is estimating the relationship between the response and the covariates through regression coefficients, it is critical to be aware of the consequences of confounding. The central point of this thesis is the evaluation of the sampling properties of parameter estimators under several scenarios and in a simple setup with one covariate.

Confounding may occur in a wide variety of research areas, such as epidemiology, environmental sciences, public health and physics. Frequently spatial models are employed in these fields because they are regression models that take into account the similarity of spatially near observations. The existence of an unmeasured spatial variable introduces a spatial structure into the residuals, which, in turn, requires inserting correlated random effects into the model. This introduction acts as a local adjustment to the regression term due to unmeasured covariates to avoid unobserved variable bias in the estimation of other covariate effects. Unfortunately, as previously mentioned, this approach may not be able to solve the bias issue. It does not fix the problem when the posited statistical model tries to model data from a generative mechanism characterized by spatial confounding.

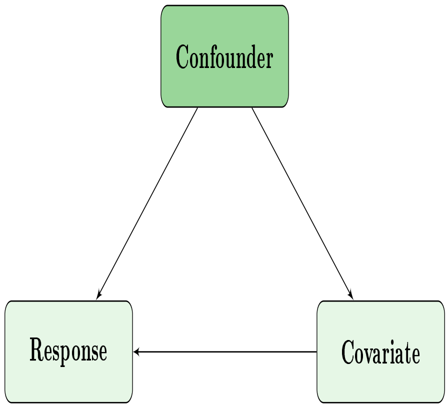

Figure 1.1: Illustration of a confounded statistical relationship in a simple setup with one covariate. The confounder influences, at the same time, the covariate and the response.

Over the past few decades, the statistical literature has dealt extensively with spatial confounding. Clayton et al. (1993) is the first reference to the spatial confounding. They point out that when "the pattern of variation of the covariate is similar to the disease risk, the location may act as a confounder". Consequently, when a spatial effect is included in the model, changes in the regression coefficient estimates are not surprising. In addition, the authors highlight the critical issue calling "confounding by location" the situation where the estimates of a regression coefficient associated with spatially structured covariate are affected by the presence of a spatial random effect in the model. Bernardinelli et al. (1995) give rise to the belief that adding a spatially correlated random effect adjusts fixed effect estimates for spatially structured missing covariate. This idea is debunked in presence of spatial confounding (Hodges and Reich, 2010). Reich et al. (2006) observe that by introducing a spatially correlated random effect the change in the regression coefficient estimates can be owed to collinearity between fixed effects and spatial random effects. Believing this changeover in the estimation was erroneous, Reich et al. (2006) and Hodges and Reich (2010) propose, as a possible redress, a method called restricted spatial regression (RSR). They conjecture that the spatial random effects mask the association between response and covariate when there is spatial confounding. For this reason they retain in the model only the random effects lying in the orthogonal space of fixed effects. In the same period, Paciorek (2010) notes that controlling for spatial confounders in classical spatial models, for example through kriging, cannot reduce bias if the correlation is strong, because the insertion of random effects may confound the effect of the covariates. In other words, including spatial random effects in the regression model as a proxy of unobserved spatial confounders does not succeed addressing the confounding problem. In fact, it will be shown along the work that even assuming a suitable correlation structure for the residuals, the bias still exists.

In light of the discussion above, it is well-reported that spatial confounding arises in the presence of multicollinearity between the covariates and the spatial random effect (Hanks et al., 2015; Hefley et al., 2017; Thaden and Kneib, 2018; Prates et al., 2019; Guan et al., 2022). Actually, in these models, spatial confounding takes place when, in addition to the response, spatially varying covariates modeling the mean of the response are correlated with spatial latent variables involved in the generative model. This notion is consistent with the standard causal definition of confounding (Papadogeorgou et al., 2018; Schnell and Papadogeorgou, 2020; Gilbert et al., 2021). In this regard, Khan and Berrett (2023) identify that there are at least two distinct phenomena currently conflated with the term spatial confounding: one about the multicollinearity characterizing the posited model and one inherent the data from a generative mechanism featured by correlated observed covariates and unmeasured spatial variables.
From the econometric point of view, spatial confounding might be seen as a type of endogeneity, with observed covariate the endogenous variable and the unmeasured component, or some proxy of it, an exogenous variable. Thaden and Kneib (2018) and Khan and Calder (2022) try to give a formal and strict definition of confounding and no confounding respectively, but still there is no unifying one in more general settings. At first glance, it can be briefly defined as the impossibility of disengaging marginal covariate effect from spatial random one when they are dependent. In this case estimation methods can lead to misleading results.

Spatial confounding is a contentious and active spatial statistics' area of research. The relevant literature can be split into two strands. In the first one, the researchers try to quantify, evaluate and understand the impact that spatial confounding has on regression coefficients (Paciorek, 2010; Page et al., 2017; Nobre et al., 2021). In the second one, methods capable of accounting for spatial confounding are developed in order to obtain accurate estimates of the target parameters (Reich et al., 2006; Hodges and Reich, 2010; Hughes and Haran, 2013; Hanks et al., 2015; Hefley et al., 2017; Thaden and Kneib, 2018; Papadogeorgou et al., 2018; Guan et al., 2022; Dupont et al., 2021; Yang, 2021; Reich et al., 2021; Marques et al., 2022; Hui and Bondell, 2022). Our work contributes to the vein of the literature focusing on understanding spatial confounding, whereas future researches may consider the development of new methodologies.

In the recent literature, studies concerning spatial confounding focus on the strength of spatial association characterizing the covariate and confounder, in order to evaluate its impact on the sampling properties of the regression coefficient estimators. To date, what is evident from the previous studies is that the parameters influencing spatial autocorrelation of the covariate and confounder are of major relevance. Moreover, the most widespread idea is that a confounder smoother than the covariates leads to a lower bias, e.g. less confounding (Paciorek, 2010). However, Page et al. (2017) note that this idea cannot be pooled regardless the generative mechanism that drives the data.

The major aim of this dissertation is to extend the current literature on spatial confounding to general structured settings, e.g. to any generating process characterized by autocorrelation structures. Besides, another goal of the work is understanding the behavior of confounding as a function of the data generating process (DGP) structure parameters in several scenarios depending on the possible different ways in which the joint covariate-confounder structure can be built. After discussing the difference between smoothness and variability characterizing a structured random process, here we point out that the bias mostly depends on the ratio of interaction's variability between confounder and covariate, i.e. their expected sampling covariance, and the expected sampling variance of the covariate. Assuming a smoother confounder, in particular cases, leads to smaller global variability and, thus, to lower bias.

1.2 Main contributions

We provide a clear statistical framework to understand confounding. Following the line adopted by Paciorek (2010), Page et al. (2017) and Nobre et al. (2021), a stochastic generative model is assumed as data generating process along the thesis. This is useful because allows us to obtain plausible values for the response, the covariate and the confounder that may arise in real applications. In order to investigate in a more clear way the effects of confounding in the estimation of regression coefficients, all data generating process parameters are posited to be known throughout the thesis except for the regression coefficients. In line with the spatial statistics literature, we aim to quantify the sampling properties of the regression coefficient estimators, and in turn, identify the most relevant quantities depending on DGP impacting confounding.

Assuming that the data generating process characterized by confounding has a specific behaviour, we study the sampling properties of the generalized least square estimator in the linear regression model with omitted variables and on the maximum likelihood estimator in the linear mixed regression model. The former model simply omits the confounding variable, while the second accounts for it by considering a random effect that attempts to represent its correlation structure. First of all, the estimators' sampling properties, such as bias, variance and mean square error, conditionally on the covariate process are derived. They are random variables giving rise to ratios of dependent quadratic forms in Gaussian random variables (Provost and Mathai, 1992; Paolella, 2018). Using the law of total expectation and variance and, following Sawa (1978) and Cressie (1993), it is possible to obtain their expected value, providing an analytic expression of the marginal sampling properties of the parameter estimator by means of the Carlson's R function (Carlson, 1963; Lauricella, 1893).

Starting from the straightforward case in which spherical Gaussian processes generate the covariate and confounder (unstructured DGP), the conditional bias of the estimator is deterministic. It is function of the confounder-covariate covariance and covariate variance parameters. This result suggests that the bias depends on the intensity of the relationship between the confounder and covariate, and on the variability of the covariate process. However, moving on the assumption of structured DGP, complications rise because the variability of a process does not match the variance parameter. To overcome this problem, we propose some quantities that synthesize such variability. Indeed, to summarize that of a structured process, we specify the expected sampling variance and to consider the variability explained by the interaction of the two processes, we define the expected sampling covariance. As they are defined, such quantities allow us to note that the eigenvalues of a covariance matrix are crucial to take into account the variability of a process in the evaluation of the problem of confounding. Furthermore, given the literature's interest in the smoothness of a process considered, we show how both the variability and smoothness are captured in the specification of a process. They can be quantified through the process covariance and correlation structure, respectively.

Since the primary goal of this dissertation is to understand the key issues of confounding and derive informative analytic results for a wider setting, we work under several frameworks with specific assumptions regarding the covariance and cross-covariance functions used to generate the processes involved in the generative mechanism. In particular, we consider different scenarios in order to better understand how they affect the sampling properties of the estimators. In a geostatistical framework, we consider an exponential correlation function that depends on spatial range parameters to produce the covariate and confounder processes. In an areal framework, we assume that the processes follow a conditional autoregressive model depending on structure parameter and adjacency matrix. In order to explore the temporal field, we assume that the covariate and confounder follow an autoregressive process of order one. Then, in light of evidences supplied and regarding the estimator bias as the principal marker for the magnitude of confounding, we point out that the confounder smoothness is not the most relevant measure in determining the bias. Rather, the covariate-confounder interaction and the variability of covariate, play the most prominent role in determining the bias. Based on this fact, we propose a representative scalar quantity for the confounding extent as a proxy of the marginal bias of the estimator of the target parameter, its first order Laplace approximation.

The research reported in this dissertation contributes to the literature with an extensive comprehension of confounding in linear regression models. Regarding future developments and work in progress we would examine the sources of confounding in more depth: considering other kind of covariance and cross-covariance function for the design of the data generating process, studying regression model with more than one covariate and evaluating the predictions' uncertainty in presence of confounding. In addiction, aware about the important role of the matrices' eigenvalues characterizing ratios of quadratic forms that define the estimator sampling properties, we deem to supply boundaries for them. Finally, it would be desirable finding a way to adjust for confounding in a Bayesian framework.

The thesis is organized as follows: in Chapter 2 we define the data generating process and the posited model for parameter estimation putting forward our comprehensive notation used along the work. In Chapter 3 we present the main features of confounding in terms of quadratic forms: the marginal sampling properties of the estimator and the quantities used to evaluate the issue. In Chapter 4 we provide a brief review of the main results of the statistic literature regarding the assessment of spatial confounding adding our own consideration on the matter and we implement an application study aimed at creating scenarios similar to what we get in practice looking at geostatistical, areal and temporal data.

Chapter 2

Analytic framework

One of the objectives of statistical modeling is to figure out the influence of variables (called covariates, regressors or independent variables) on a measure of interest (called response or dependent variable). A general framework to perform this kind of analyses is supplied by regression models. In particular, we focus on linear regression models.

The main topic of this thesis concerns the evaluation of the impact of unobserved relevant information, related to the observed one, on the estimation of regression parameters, an issue known in the literature as confounding. To provide a formal study of the sampling properties of regression parameter estimators in presence of confounding, we start by defining the Data Generating Process (Section 2.1) and a separate posited model for parameter estimation (Section 2.2). This is intended to mimic the workflow of statistical analysis where data are interpreted as realizations of a random mechanism that the researcher tries to infer toward statistical modeling. In this spirit, confounding is about the bias affecting the regression coefficient estimates when the postulated model misses some relevant feature of the DGP. Hypothesis underlying the posited model lead to different estimators of the regression coefficients such as Ordinary Least Squares (OLS, in regression with spherical disturbances), Generalized Least Squares (GLS, in regression with structured disturbances) and more generally Maximum Likelihood (ML) estimators in mixed linear regression models. As we point out in Section 2.2, these estimators can be cast as the same linear function of the data with appropriate weighting matrices (in particular, this is achievable in mixed models after marginalization of the random effects): this allows a unified treatment of the sampling distribution of all these estimators, which is provided in Chapter 3.

2.1 The Data Generating Process

To introduce the problem of confounding a stochastic generative model, i.e. the DGP, is considered. Its specification starts by explicitly stating the linear dependence of the response variable on the covariates:

$$\boldsymbol{Y} = \mathcal{B}_{y\cdot 0(xz)}\mathbf{1}_n + \mathcal{B}_{y\cdot x(z)}\boldsymbol{X} + \mathcal{B}_{y\cdot z(x)}\boldsymbol{Z} + \boldsymbol{\varepsilon}_{y|x,z}, \qquad \boldsymbol{\varepsilon}_{y|x,z} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_{y|x,z}), \tag{2.1}$$

where $\mathbf{1}_n$ is the n-dimensional vector of ones and $\boldsymbol{\Sigma}_{y|x,z}$ is the covariance matrix of the error term that expresses the variability of the dependent variable $\boldsymbol{Y}$ not explained by the regression on $\boldsymbol{X}$ and $\boldsymbol{Z}$. Moreover, $\mathcal{B}_{y\cdot 0(xz)}$, $\mathcal{B}_{y\cdot x(z)}$ and $\mathcal{B}_{y\cdot z(x)}$ are the *partial regression coefficients* that determine the strength and direction of the corresponding covariate's influence. The notation on subscript of the partial regression coefficients is aimed at pointing out that they quantify the relationship between the response (before the dot) and the covariate which is referred to (after the dot), in the presence of another variable (within the brackets) (for a schematic representation see Figure 2.2).

The treatment of $\boldsymbol{X}$ and $\boldsymbol{Z}$ as random processes has also been adopted by Paciorek (2010) and Page et al. (2017). The stochastic approach allows for some analytic results and is further justified in that the variation that an unmeasured $\boldsymbol{Z}$ induces in $\boldsymbol{Y}$ is necessarily treated stochastically as a part of the residual.

In our setting, the response, the covariate, and the confounder are multivariate Gaussian random variables characterized by the following joint distribution:

$$\begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{X} \\ \boldsymbol{Z} \end{pmatrix} \sim \mathcal{N}_{3n}\left(\begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_z \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_y & \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yz} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}_{zy} & \boldsymbol{\Sigma}_{zx} & \boldsymbol{\Sigma}_z \end{bmatrix} \right), \tag{2.2}$$

where $\boldsymbol{\Sigma}_y, \boldsymbol{\Sigma}_x, \boldsymbol{\Sigma}_z \in \mathcal{S}^n_{++}$ [1] are the positive definite marginal covariance matrices defined as

$$\boldsymbol{\Sigma}_x = \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}) = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^\top\right],$$

and $\boldsymbol{\Sigma}_{xz}$, such that $\boldsymbol{\Sigma}_{xz} = \boldsymbol{\Sigma}_{zx}^\top$, is the cross-covariance matrix defines as

$$\boldsymbol{\Sigma}_{xz} = \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Z}) = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}])^\top\right].$$

Equation (2.1) can be expressed through the conditional distribution of the response

[1] Let $\mathcal{S}^n$ denote the vector space of $n \times n$ real symmetric matrices. Recall that by the spectral theorem any matrix $\mathbf{A} \in \mathcal{S}^n$ is diagonalizable in an orthonormal basis and has real eigenvalues. Let $\mathcal{S}^n_{++}$ ($\mathcal{S}^n_{+}$) denote the set of positive (semi)definite matrices, i.e. the set of real symmetric matrices having strictly positive (non-negative) eigenvalues. For a matrix $\mathbf{A} \in \mathcal{S}^n_{++}$ ($\mathbf{A} \in \mathcal{S}^n_{+}$) we will use the notation $\mathbf{A} \succ 0$ ($\mathbf{A} \succeq 0$).

process as follows:

$$\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z} \sim \mathcal{N}_n\left(\mathcal{B}_{y\cdot 0(xz)}\mathbf{1}_n + \mathcal{B}_{y\cdot x(z)}\boldsymbol{X} + \mathcal{B}_{y\cdot z(x)}\boldsymbol{Z}, \boldsymbol{\Sigma}_{y|x,z}\right). \tag{2.3}$$

Recalling the results in Appendix B.2, valid joint distribution of $(\boldsymbol{Y}^\top, \boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top$ may be obtained assuming any combination of two conditional covariance matrices and one marginal positive definite covariance matrix, e.g. $\boldsymbol{\Sigma}_{y|x,z} \succ 0$, $\boldsymbol{\Sigma}_{x|z} \succ 0$ and $\boldsymbol{\Sigma}_z \succ 0$. Otherwise, selecting all positive definite marginal covariance matrices and choosing valid cross-covariance function leads to a valid joint distribution.

We are interested in the ensuing linear predictor, i.e. the mean of random vector $\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z} \in \mathbb{R}^n$,

$$\mathbb{E}_Y[\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z}] = \mathcal{B}_{y\cdot 0(xz)}\mathbf{1}_n + \mathcal{B}_{y\cdot x(z)}\boldsymbol{X} + \mathcal{B}_{y\cdot z(x)}\boldsymbol{Z}, \tag{2.4}$$

obtained by integration on the domain of $\boldsymbol{Y}$. A significant aspect of Equation (2.4), as emphasized by the notation, is that the expectation has to be taken with respect to the distribution of $\boldsymbol{Y}$. The following proposition is aimed at clarifying the notation adopted for the expected value when the random vector $\boldsymbol{Y}$ is expressed as a function of $\boldsymbol{X}$ and $\boldsymbol{Z}$.

Proposition 2.1.1. *Let $\boldsymbol{X}$ and $\boldsymbol{Z}$ be continuous random vectors with joint probability density function (pdf) $f_{X,Z}(\boldsymbol{x},\boldsymbol{z})$ and let $\boldsymbol{Y}$ be a continuous random vector with pdf $f_Y(\boldsymbol{y})$. Considering the transformation $h(\cdot)$ such that*

$$\boldsymbol{Y} = h(\boldsymbol{X},\boldsymbol{Z}),$$

by a change of variables argument, it can be established that

$$\begin{aligned}\mathbb{E}_Y[\boldsymbol{Y}] &= \int_{\mathcal{Y}} \boldsymbol{y} f_Y(\boldsymbol{y})\, d\boldsymbol{y} \\ &= \int_{\mathcal{X}}\int_{\mathcal{Z}} h(\boldsymbol{x},\boldsymbol{z}) f_{X,Z}(\boldsymbol{x},\boldsymbol{z})\, d\boldsymbol{x}d\boldsymbol{z} \\ &= \mathbb{E}_{X,Z}[h(\boldsymbol{X},\boldsymbol{Z})],\end{aligned}$$

under the condition

$$\int_{\mathcal{X}}\int_{\mathcal{Z}} |h(\boldsymbol{x},\boldsymbol{z})| f_{X,Z}(\boldsymbol{x},\boldsymbol{z})\, d\boldsymbol{x}d\boldsymbol{z} < \infty,$$

is the expected value of the random vector $\boldsymbol{Y}$ with respect to the random vectors $\boldsymbol{X}$ and $\boldsymbol{Z}$ on the product space $\mathcal{X} \times \mathcal{Z}$.

To understand the consequences of lack of information concerning the variable $\boldsymbol{Z}$ on the estimation of the regression coefficient of $\boldsymbol{Y}$ on the measured variable $\boldsymbol{X}$, we will consider the conditional distribution $\boldsymbol{Y}|\boldsymbol{X}$ marginalized over $\boldsymbol{Z}$:

$$\boldsymbol{Y}|\boldsymbol{X} \sim \mathcal{N}_n(\mathcal{B}_{y\cdot 0(x)}\mathbf{1}_n + \mathbf{A}_{y\cdot x}\boldsymbol{X},\, \mathcal{B}^2_{y\cdot z(x)}\boldsymbol{\Sigma}_{z|x} + \boldsymbol{\Sigma}_{y|x,z}), \tag{2.5}$$

where

$$\begin{aligned}\mathbf{A}_{y\cdot x} &= \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1} \\ &= \mathcal{B}_{y\cdot x(z)}\mathbf{I}_n + \mathcal{B}_{y\cdot z(x)}\mathbf{A}_{z\cdot x} \end{aligned} \tag{2.6}$$

is the *regression matrix* of $\boldsymbol{Y}$ on $\boldsymbol{X}$.

When $\mathbf{A}_{x\cdot z}$ is not a scalar matrix[2], that is the most stimulating case, also the coefficients' interpretation problem arises because they are not understandable as partial derivatives anymore (LeSage, 2008; Golgher and Voss, 2015). This is an interesting topic that could be the target of a future insights. More broadly, whether there were some relationships between the multivariate components of the random vectors in a model, they certainly would not want to be ignored. In that case, the regression matrices would no longer be scalar but will have some sparse or full structure characterized by dependence they share.

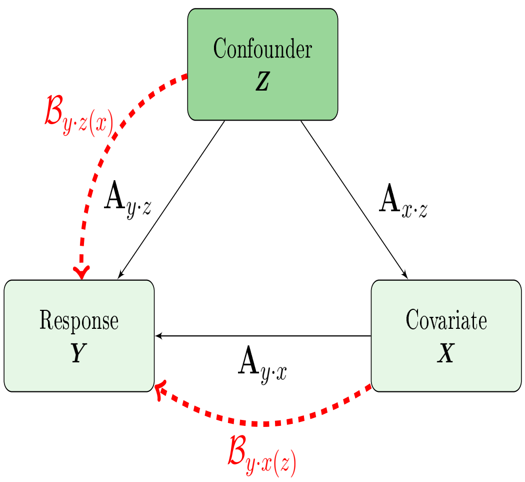

Figure 2.1: Formalization of the confounded simple regression model in (2.1) with scalar partial regression coefficients (red) and regression matrices (black).

Under such setting, if $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are conditionally independent given $\boldsymbol{X}$, i.e.

$$\mathcal{B}_{y\cdot z(x)} = 0$$

and, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are independent, i.e

$$\mathbf{A}_{z\cdot x} = \mathbf{0} \iff \boldsymbol{\Sigma}_{zx} = \mathbf{0},$$

the simple regression matrix of $\boldsymbol{Y}$ on $\boldsymbol{X}$ coincides with the corresponding scalar matrix containing the partial regression coefficient related to $\boldsymbol{X}$, i.e. $\mathbf{A}_{y\cdot x} = \mathcal{B}_{y\cdot x(z)}\mathbf{I}_n$. It means that there is no confounding in the model inasmuch as the presence of a latent process

[2] A scalar matrix is a diagonal matrix with equal diagonal entries.

in DGP does not change the measurement of relationship between the response and the covariate. Consequently, regression parameter estimates are unbiased.

2.1.1 Unstructured data generating process

We start considering an unstructured DGP. Assuming a scalar error covariance matrix (assumed throughout the whole thesis), i.e.

$$\mathbf{\Sigma}_{y|x,z} = \sigma^2_{y|x,z}\mathbf{I}_n,$$

and treating the regressors as spherical[3] random variables, such that

$$\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu}_x, \sigma^2_x\mathbf{I}_n) \qquad \text{and} \qquad \boldsymbol{Z} \sim \mathcal{N}_n(\boldsymbol{\mu}_z, \sigma^2_z\mathbf{I}_n),$$

the partial regression coefficients can be expressed in term of the simple regression coefficients as:

$$\begin{aligned}\mathcal{B}_{y\cdot x(z)} &= \frac{\mathcal{B}_{y\cdot x} - \mathcal{B}_{y\cdot z}\mathcal{B}_{z\cdot x}}{1 - \mathcal{B}_{x\cdot z}\mathcal{B}_{z\cdot x}} \\ \mathcal{B}_{y\cdot z(x)} &= \frac{\mathcal{B}_{y\cdot z} - \mathcal{B}_{y\cdot x}\mathcal{B}_{x\cdot z}}{1 - \mathcal{B}_{x\cdot z}\mathcal{B}_{z\cdot x}}.\end{aligned} \tag{2.7}$$

where $\mathcal{B}_{y\cdot x}$ is the *simple regression coefficient* of $\boldsymbol{Y}$ on $\boldsymbol{X}$.
Thus, $\mathcal{B}_{y\cdot x(z)}$ can be interpreted as the regression coefficient of $\boldsymbol{Y}$ on $\boldsymbol{X}$ modified for the presence of $\boldsymbol{Z}$. Moreover, if the random variables $\boldsymbol{X}$ and $\boldsymbol{Z}$ are not correlated, one has that $\mathcal{B}_{z\cdot x} = \mathcal{B}_{x\cdot z} = 0$, and hence, $\mathcal{B}_{y\cdot x(z)} = \mathcal{B}_{y\cdot x}$. The same is valid for $\mathcal{B}_{y\cdot z(x)}$ (Anderson, 1984; Allen, 1997). In this spherical setup, given the pair of random vectors $(\boldsymbol{Y}, \boldsymbol{X})$, the *i-th* component of the vector $\boldsymbol{Y}$, Y_i, is affected exclusively by X_i, in a proportional way through $\mathcal{B}_{y\cdot x}$, for all $i = 1, \ldots, n$. Thereby, we have $\mathbf{A}_{y\cdot x} = \mathcal{B}_{y\cdot x}\mathbf{I}_n$, and so, in this case, the regression coefficient $\mathcal{B}_{y\cdot x}$ can be expressed in terms of the partial regression coefficients of the full model using the Equations in (2.7) as follows

$$\mathcal{B}_{y\cdot x} = \mathcal{B}_{y\cdot x(z)} + \mathcal{B}_{y\cdot z(x)}\mathcal{B}_{z\cdot x}. \tag{2.8}$$

Equation (2.8) shows that the estimation will be biased when $\mathcal{B}_{y\cdot x} \neq \mathcal{B}_{y\cdot x(z)}$, that is the case in which there is confounding in the model, which is when the covariate and confounder processes are correlated and the response and the confounder are conditionally dependent given the covariate. This intuition will be formalized for a broader context in Section 2.1.3.

[3] Spherical distributions are considered an extension of the standard multivariate normal distribution characterized by no unit marginal variance and mean not necessarily zero.

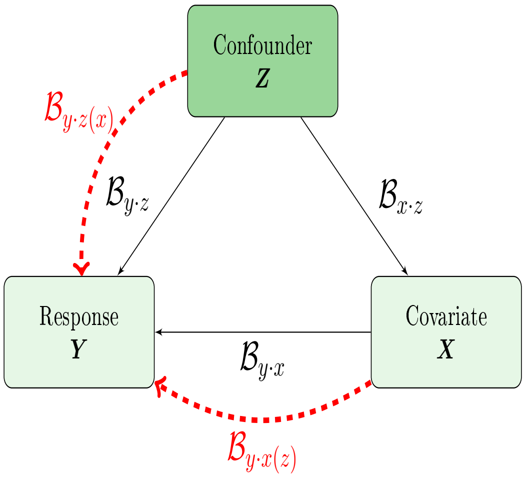

Figure 2.2: Formalization of the confounded simple regression model in (2.1) considering an unstructured DGP characterized the simple (black) and partial (red) regression coefficients quantifying the relations between processes.

2.1.2 Structured data generating process

After tackling the most basic case, in this section we consider structured covariance and cross-covariance matrices for the distributive assumption in (2.2). Subsequently, what has been reported in Section 2.1.1 becomes more complex. In this regard we start by remarking how the joint distribution for (2.2) can be expressed in dependence of regression matrices since expressing the linear predictor in terms of partial regression coefficients, as in (2.4), is no longer possible. Particular attention is devoted to conditions that have to be met in order to obtain a valid joint covariance matrix.

2.1.2.1 Joint distribution of Gaussian vectors through regression matrices

Consider a bivariate random vector $(\boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top \sim \mathcal{N}_{2n}(\boldsymbol{\mu}_{x,z}, \boldsymbol{\Sigma}_{x,z})$, where $\boldsymbol{\mu}_{x,z} = \left(\boldsymbol{\mu}_x^\top, \boldsymbol{\mu}_z^\top\right)^\top$ and

$$\boldsymbol{\Sigma}_{x,z} = \begin{bmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}_{zx} & \boldsymbol{\Sigma}_z \end{bmatrix}. \tag{2.9}$$

From standard multivariate normal theory (see Appendix B), this distribution can be obtained starting from the marginal distribution of $\boldsymbol{Z}$, $\mathcal{N}_n(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z)$, and the conditional distribution $\boldsymbol{X}|\boldsymbol{Z} \sim \mathcal{N}_n(\boldsymbol{\mu}_{x|z}, \boldsymbol{\Sigma}_{x|z})$ which has moments:

$$\begin{aligned} \boldsymbol{\mu}_{x|z} &= \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_z) \\ &= \boldsymbol{\mu}_x + \mathbf{A}_{x\cdot z}(\boldsymbol{z} - \boldsymbol{\mu}_z), \\ \boldsymbol{\Sigma}_{x|z} &= \boldsymbol{\Sigma}_x - \mathbf{A}_{x\cdot z}\boldsymbol{\Sigma}_z\mathbf{A}_{x\cdot z}^\top. \end{aligned} \tag{2.10}$$

By analogy, it is possible to do the same with the variables $\boldsymbol{X}$ and $\boldsymbol{Z}|\boldsymbol{X}$. Thus, the joint distribution of $\boldsymbol{X}$ and $\boldsymbol{Z}$ can be expressed as follows:

$$\begin{pmatrix}\boldsymbol{X}\\ \boldsymbol{Z}\end{pmatrix} \sim \mathcal{N}_{2n}\left(\begin{pmatrix}\boldsymbol{\mu}_{x|z} - \mathbf{A}_{x\cdot z}(\boldsymbol{z}-\boldsymbol{\mu}_z)\\ \boldsymbol{\mu}_z\end{pmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{x|z} + \mathbf{A}_{x\cdot z}\boldsymbol{\Sigma}_z\mathbf{A}_{x\cdot z}^\top & \mathbf{A}_{x\cdot z}\boldsymbol{\Sigma}_z\\ (\mathbf{A}_{x\cdot z}\boldsymbol{\Sigma}_z)^\top & \boldsymbol{\Sigma}_z\end{bmatrix}\right). \tag{2.11}$$

According to Harville (1997) and Zhang (2005), conditions ensuring the property of the covariance matrix in (2.11) are $\boldsymbol{\Sigma}_z \succ 0$ and $\boldsymbol{\Sigma}_{x|z} \succ 0$. Hence, one needs to specify the matrices $\boldsymbol{\Sigma}_z \succ 0$ and $\boldsymbol{\Sigma}_{x|z} \succ 0$ to get a valid covariance matrix, while no conditions need to be met by the regression matrix $\mathbf{A}_{x\cdot z}$.

Considering the trivariate random vector $(\boldsymbol{Y}^\top, \boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top$, in a similar manner as previously done in Equation (2.6), we define the regression matrix $\mathbf{A}_{y\cdot xz} \in \mathbb{R}^{n\times 2n}$ that expresses the structure of the regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$ and $\boldsymbol{Z}$, as

$$\mathbf{A}_{y\cdot xz} = \boldsymbol{\Sigma}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}^{-1}. \tag{2.12}$$

The conditional distribution $\boldsymbol{Y}|(\boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top \sim \mathcal{N}_n(\boldsymbol{\mu}_{y|x,z}, \boldsymbol{\Sigma}_{y|x,z})$ is characterized by the following moments:

$$\begin{aligned}\boldsymbol{\mu}_{y|x,z} &= \boldsymbol{\mu}_y + \mathbf{A}_{y\cdot xz}\begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix},\\ \boldsymbol{\Sigma}_{y|x,z} &= \boldsymbol{\Sigma}_y - \mathbf{A}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}\mathbf{A}_{y\cdot xz}^\top.\end{aligned} \tag{2.13}$$

As for the bivariate case, we express the joint covariance matrix $\boldsymbol{\Sigma}_{y,x,z}$ as a function of the regression matrices introduced above:

$$\begin{aligned}\boldsymbol{\mu}_{y,x,z} &= \begin{pmatrix}\boldsymbol{\mu}_{y|x,z} - \mathbf{A}_{y\cdot xz}\begin{pmatrix}\boldsymbol{x} - [\boldsymbol{\mu}_{x|z} - \mathbf{A}_{x\cdot z}(\boldsymbol{z}-\boldsymbol{\mu}_z)]\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix}\\ \boldsymbol{\mu}_{x|z} - \mathbf{A}_{x\cdot z}(\boldsymbol{z}-\boldsymbol{\mu}_z)\\ \boldsymbol{\mu}_z\end{pmatrix}\\ \boldsymbol{\Sigma}_{y,x,z} &= \begin{bmatrix}\boldsymbol{\Sigma}_{y|x,z} + \mathbf{A}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}\mathbf{A}_{y\cdot xz}^\top & \mathbf{A}_{y\cdot x}\boldsymbol{\Sigma}_x & \mathbf{A}_{y\cdot z}\boldsymbol{\Sigma}_z\\ \mathbf{A}_{x\cdot y}\boldsymbol{\Sigma}_y & \boldsymbol{\Sigma}_{x|z} + \mathbf{A}_{x\cdot z}\boldsymbol{\Sigma}_z\mathbf{A}_{x\cdot z}^\top & \mathbf{A}_{x\cdot z}\boldsymbol{\Sigma}_z\\ \mathbf{A}_{z\cdot y}\boldsymbol{\Sigma}_y & \mathbf{A}_{z\cdot x}\boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_z\end{bmatrix}\end{aligned} \tag{2.14}$$

According to Theorem B.2.1, the sufficient conditions that ensure the property of $\boldsymbol{\Sigma}_{y,x,z}$ are $\boldsymbol{\Sigma}_{x,z} \succ 0$ and $\boldsymbol{\Sigma}_{y|x,z} \succ 0$. Hence, regression matrices $\mathbf{A}_{y\cdot x}$, $\mathbf{A}_{y\cdot z}$, $\mathbf{A}_{x\cdot z}$, $\mathbf{A}_{y\cdot xz}$ do not impact positive definiteness of $\boldsymbol{\Sigma}_{y,x,z}$.

Furthermore, from Equation (2.12) it is possible to identify the sub-matrices $\mathbf{A}_{y\cdot x(z)} \in$

$\mathbb{R}^{n\times n}$ and $\mathbf{A}_{y\cdot z(x)} \in \mathbb{R}^{n\times n}$, called *partial regression matrices*, such that:

$$\mathbf{A}_{y\cdot xz} = \left[\mathbf{A}_{y\cdot x(z)} \quad : \quad \mathbf{A}_{y\cdot z(x)}\right], \tag{2.15}$$

where $\mathbf{A}_{y\cdot x(z)}$ is the matrix that describes the relationship between the variable $\boldsymbol{Y}$ and $\boldsymbol{X}$, holding $\boldsymbol{Z}$ constant. Matrices $\mathbf{A}_{x\cdot z}$, $\mathbf{A}_{y\cdot x}$ and $\mathbf{A}_{y\cdot z}$ are referred to as *simple regression matrices* in what follows when we need to distinguish them from the partial ones.

2.1.2.2 Relationship between partial and simple regression matrices

In the following we focus on how simple and partial regression matrices are related to each other (see Figure 2.1).

Proposition 2.1.2. *The simple regression matrices that link* $\boldsymbol{Y}$ *to* $\boldsymbol{X}$ *and* $\boldsymbol{Z}$ *can be expressed as a function of the partial regression matrices* $\boldsymbol{A}_{y\cdot x(z)}$*,* $\boldsymbol{A}_{y\cdot z(x)}$ *and the simple regression matrices between* $\boldsymbol{X}$ *and* $\boldsymbol{Z}$ *as follows:*

(i)

$$\boldsymbol{A}_{y\cdot x} = \boldsymbol{A}_{y\cdot z(x)}\boldsymbol{A}_{z\cdot x} + \boldsymbol{A}_{y\cdot x(z)}, \tag{2.16}$$

(ii)

$$\boldsymbol{A}_{y\cdot z} = \boldsymbol{A}_{y\cdot x(z)}\boldsymbol{A}_{x\cdot z} + \boldsymbol{A}_{y\cdot z(x)}. \tag{2.17}$$

Proof. From (2.11), (2.12) and (2.15), one gets:

$$\begin{aligned}
\boldsymbol{\Sigma}_{y\cdot xz} &= \mathbf{A}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z} \\
&= \left[\mathbf{A}_{y\cdot x(z)} \quad : \quad \mathbf{A}_{y\cdot z(x)}\right] \begin{bmatrix} \boldsymbol{\Sigma}_x & \mathbf{A}_{x\cdot z}\boldsymbol{\Sigma}_z \\ (\mathbf{A}_{x\cdot z}\boldsymbol{\Sigma}_z)^\top & \boldsymbol{\Sigma}_z \end{bmatrix} \\
&= \left[(\mathbf{A}_{y\cdot x(z)} + \mathbf{A}_{y\cdot z(x)}\mathbf{A}_{z\cdot x})\boldsymbol{\Sigma}_x \quad : \quad (\mathbf{A}_{y\cdot x(z)}\mathbf{A}_{x\cdot z} + \mathbf{A}_{y\cdot z(x)})\boldsymbol{\Sigma}_z\right].
\end{aligned}$$

Then, recalling how $\boldsymbol{\Sigma}_{y\cdot xz}$ is built from (B.2) brings to (i) and (ii). □

Proposition 2.1.3. *The partial regression matrices can be expressed as a function of the simple regression matrices of every possible pair of random vectors* $\boldsymbol{Y}$*,* $\boldsymbol{X}$ *and* $\boldsymbol{Z}$*, as follows.*

(i)

$$\boldsymbol{A}_{y\cdot z(x)} = \left(\boldsymbol{A}_{y\cdot z} - \boldsymbol{A}_{y\cdot x}\boldsymbol{A}_{x\cdot z}\right)\left(\boldsymbol{I}_n - \boldsymbol{A}_{z\cdot x}\boldsymbol{A}_{x\cdot z}\right)^{-1}; \tag{2.18}$$

(ii)

$$\boldsymbol{A}_{y\cdot x(z)} = \left(\boldsymbol{A}_{y\cdot x} - \boldsymbol{A}_{y\cdot z}\boldsymbol{A}_{z\cdot x}\right)\left(\boldsymbol{I}_n - \boldsymbol{A}_{x\cdot z}\boldsymbol{A}_{z\cdot x}\right)^{-1}. \tag{2.19}$$

Proof. Concerning statement (i), starting from (2.17) and then using (2.16), one gets:

$$\begin{aligned}
\mathbf{A}_{y\cdot z(x)} &= \mathbf{A}_{y\cdot z} - \mathbf{A}_{y\cdot x(z)}\mathbf{A}_{x\cdot z} \\
&= \mathbf{A}_{y\cdot z} - (\mathbf{A}_{y\cdot x} - \mathbf{A}_{y\cdot z(x)}\mathbf{A}_{z\cdot x})\mathbf{A}_{x\cdot z} \\
\mathbf{A}_{y\cdot z(x)} - \mathbf{A}_{y\cdot z(x)}\mathbf{A}_{z\cdot x}\mathbf{A}_{x\cdot z} &= \mathbf{A}_{y\cdot z} - \mathbf{A}_{y\cdot x}\mathbf{A}_{x\cdot z} \\
\mathbf{A}_{y\cdot z(x)}(\mathbf{I}_n - \mathbf{A}_{z\cdot x}\mathbf{A}_{x\cdot z}) &= \mathbf{A}_{y\cdot z} - \mathbf{A}_{y\cdot x}\mathbf{A}_{x\cdot z} \\
\mathbf{A}_{y\cdot z(x)} &= (\mathbf{A}_{y\cdot z} - \mathbf{A}_{y\cdot x}\mathbf{A}_{x\cdot z})\,(\mathbf{I}_n - \mathbf{A}_{z\cdot x}\mathbf{A}_{x\cdot z})^{-1}.
\end{aligned}$$

Alternatively, we can provide the evidence of the statement (i) making use of the Schur-Barachiewicz inverse formula (A.4) through the Schur complement $\mathbf{\Sigma}_{x,z}/\mathbf{\Sigma}_x$:

$$\begin{aligned}
\mathbf{A}_{y\cdot xz} &= \begin{bmatrix}\mathbf{A}_{y\cdot x(z)} & \vdots & \mathbf{A}_{y\cdot z(x)}\end{bmatrix} \\
&= \mathbf{\Sigma}_{y\cdot xz}\mathbf{\Sigma}_{x,z}^{-1} \\
&= \begin{bmatrix}\mathbf{\Sigma}_{yx} & \vdots & \mathbf{\Sigma}_{yz}\end{bmatrix}\begin{bmatrix}\mathbf{\Sigma}_{x} & \mathbf{\Sigma}_{xz} \\ \mathbf{\Sigma}_{zx} & \mathbf{\Sigma}_{z}\end{bmatrix}^{-1} \\
&= \begin{bmatrix}\mathbf{\Sigma}_{yx} & \vdots & \mathbf{\Sigma}_{yz}\end{bmatrix}\begin{bmatrix}\mathbf{\Sigma}_x^{-1} + \mathbf{\Sigma}_x^{-1}\mathbf{\Sigma}_{xz}(\mathbf{\Sigma}_{x,z}/\mathbf{\Sigma}_x)^{-1}\mathbf{\Sigma}_{zx}\mathbf{\Sigma}_x^{-1} & -\mathbf{\Sigma}_x^{-1}\mathbf{\Sigma}_{xz}(\mathbf{\Sigma}_{x,z}/\mathbf{\Sigma}_x)^{-1} \\ -(\mathbf{\Sigma}_{x,z}/\mathbf{\Sigma}_x)^{-1}\mathbf{\Sigma}_{zx}\mathbf{\Sigma}_x^{-1} & (\mathbf{\Sigma}_{x,z}/\mathbf{\Sigma}_x)^{-1}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{\Sigma}_{yx}\mathbf{\Sigma}_x^{-1}\mathbf{\Sigma}_{xz}(\mathbf{\Sigma}_{x,z}/\mathbf{\Sigma}_x)^{-1}\mathbf{\Sigma}_{zx}\mathbf{\Sigma}_x^{-1} & \vdots & -\mathbf{\Sigma}_{yx}\mathbf{\Sigma}_x^{-1}\mathbf{\Sigma}_{xz}(\mathbf{\Sigma}_{x,z}/\mathbf{\Sigma}_x)^{-1}+ \\ +\mathbf{\Sigma}_{yx}\mathbf{\Sigma}_x^{-1} - \mathbf{\Sigma}_{yz}(\mathbf{\Sigma}_{x,z}/\mathbf{\Sigma}_x)^{-1}\mathbf{\Sigma}_{zx}\mathbf{\Sigma}_x^{-1} & \vdots & +\mathbf{\Sigma}_{yz}(\mathbf{\Sigma}_{x,z}/\mathbf{\Sigma}_x)^{-1}\end{bmatrix}.
\end{aligned}$$

According to the definitions of regression matrices and considering that

$$(\mathbf{\Sigma}_{x,z}/\mathbf{\Sigma}_x)^{-1} = \mathbf{\Sigma}_z^{-1}(\mathbf{I}_n - \mathbf{A}_{z\cdot x}\mathbf{A}_{x\cdot z})^{-1},$$

we get

$$\mathbf{A}_{y\cdot xz} = \begin{bmatrix}\mathbf{A}_{y\cdot x}\mathbf{A}_{x\cdot z}(\mathbf{I}_n - \mathbf{A}_{z\cdot x}\mathbf{A}_{x\cdot z})^{-1}\mathbf{A}_{z\cdot x}+ & \vdots & -\mathbf{A}_{y\cdot x}\mathbf{A}_{x\cdot z}(\mathbf{I}_n - \mathbf{A}_{z\cdot x}\mathbf{A}_{x\cdot z})^{-1}+ \\ +\mathbf{A}_{y\cdot x} - \mathbf{A}_{y\cdot z}(\mathbf{I}_n - \mathbf{A}_{z\cdot x}\mathbf{A}_{x\cdot z})^{-1}\mathbf{A}_{z\cdot x} & \vdots & +\mathbf{A}_{y\cdot z}(\mathbf{I}_n - \mathbf{A}_{z\cdot x}\mathbf{A}_{x\cdot z})^{-1}\end{bmatrix}$$

and then, the expected result

$$\mathbf{A}_{y\cdot z(x)} = (\mathbf{A}_{y\cdot z} - \mathbf{A}_{y\cdot x}\mathbf{A}_{x\cdot z})(\mathbf{I}_n - \mathbf{A}_{z\cdot x}\mathbf{A}_{x\cdot z})^{-1}.$$

In order to proof (ii) one can simply repeat the process above using (A.7) instead. □

It is immediate to see that these last propositions are multivariate generalizations of Formulas (2.7) and (2.8) expressing the relations between the simple and partial regression coefficients under unstructured DGP. In fact, when in the conditional mean the partial regression matrices $\mathbf{A}_{y\cdot x(z)}$ and $\mathbf{A}_{y\cdot z(x)}$ are scalar matrices, such that

$$\mathbf{A}_{y\cdot x(z)} = \mathcal{B}_{y\cdot x(z)}\mathbf{I}_n \quad \text{and} \quad \mathbf{A}_{y\cdot z(x)} = \mathcal{B}_{y\cdot z(x)}\mathbf{I}_n.$$

we get the base-line conditional mean in (2.4).

2.1.3 Defining confounding

The mean of the Gaussian random vector $(\boldsymbol{Y}^\top, \boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top$ can be expressed as a function of the regression matrices $\mathbf{A}_{y\cdot x(z)}$, $\mathbf{A}_{y\cdot z(x)}$ and $\mathbf{A}_{x\cdot z}$ as follows

$$\boldsymbol{\mu}_{y,x,z} = \begin{pmatrix} \boldsymbol{\mu}_{y|x,z} + \mathbf{A}_{y\cdot x(z)}\boldsymbol{\mu}_x + \mathbf{A}_{y\cdot z(x)}\boldsymbol{\mu}_z - \mathbf{A}_{y\cdot x(z)}\boldsymbol{x} - \mathbf{A}_{y\cdot z(x)}\boldsymbol{z} \\ \boldsymbol{\mu}_{x|z} + \mathbf{A}_{x\cdot z}\boldsymbol{\mu}_z - \mathbf{A}_{x\cdot z}\boldsymbol{z} \\ \boldsymbol{\mu}_z \end{pmatrix}.$$

The conditional mean of the random vector $\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{Z}$ con be expressed as:

$$\begin{aligned}\mathbb{E}_Y[\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{Z}] &= \boldsymbol{\mu}_y - \mathbf{A}_{y\cdot x(z)}\boldsymbol{\mu}_x - \mathbf{A}_{y\cdot z(x)}\boldsymbol{\mu}_z + \mathbf{A}_{y\cdot xz}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{z}\end{pmatrix} \\ &= \mathbf{a}_{y\cdot 0(xz)} + \mathbf{A}_{y\cdot x(z)}\boldsymbol{x} + \mathbf{A}_{y\cdot z(x)}\boldsymbol{z} \end{aligned} \tag{2.20}$$

where $\mathbf{a}_{y\cdot 0(xz)}$ is the intercept vector. In line with Paciorek (2010) and Page et al. (2017), from now on we suppose $\boldsymbol{\mu}_x = \boldsymbol{\mu}_z = \mathbf{0}$ because $\boldsymbol{\mu}_z = \mathbf{0}$ is a common assumption in random effects models and $\boldsymbol{\mu}_x = \mathbf{0}$ is analogous to centering the covariate, without loss of generality. Applying Propositions 2.1.2 and 2.1.3, and recalling our goal to understand the consequences of an unmeasured confounder, we show the conditional mean in (2.20) marginalized with respect to $\boldsymbol{Z}$:

$$\mathbb{E}_Y[\boldsymbol{Y}|\boldsymbol{X}] = \mathbf{a}_{y\cdot 0(x)} + \left(\mathbf{A}_{y\cdot x(z)} + \mathbf{A}_{y\cdot z(x)}\mathbf{A}_{z\cdot x}\right)\boldsymbol{x}. \tag{2.21}$$

Analogously to Section 2.1.1, it appears that confounding occurs when $\mathbf{A}_{y\cdot x(z)} \neq \mathbf{A}_{y\cdot x}$. Indeed, the conditional mean in Equation (2.21) shows how the product matrix $\mathbf{A}_{y\cdot z(x)}\mathbf{A}_{z\cdot x}$ contributes to influence the magnitude of confounding.

After the complete presentation of our overarching notation, hereunder the definition of confounding is formalized. The representation of the regression matrix $\mathbf{A}_{y\cdot x}$ in Equation (2.21) and Proposition 2.1.2 suggest our proposal. The definition of confounding is coherent with the one proposed by Thaden and Kneib (2018), with adapted notation.

Definition 2.1.1. Let $\boldsymbol{Y}$ be the response, $\boldsymbol{X}$ be the covariate and $\boldsymbol{Z}$ be the confounder processes of a linear regression model characterized by the joint distribution expressed in (2.2). Then, the regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$ is defined *confounded* by $\boldsymbol{Z}$ if both the following conditions are verified:

(i) $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are conditionally dependent given $\boldsymbol{X}$ ($\boldsymbol{Y} \not\perp \boldsymbol{Z}|\boldsymbol{X}$), i.e.

$$\mathbb{E}_Y[\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Z}=\boldsymbol{z}] \neq \mathbb{E}_Y[\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}] \Rightarrow \mathbf{A}_{y\cdot z(x)} \neq \mathbf{0};$$

(ii) $\boldsymbol{X}$ and $\boldsymbol{Z}$ are dependent ($\boldsymbol{X} \not\perp \boldsymbol{Z}$), i.e.

$$\mathbb{E}_X[\boldsymbol{X}|\boldsymbol{Z}=\boldsymbol{z}] \neq \mathbb{E}_X[\boldsymbol{X}] \Rightarrow \mathbf{A}_{x\cdot z} \neq \mathbf{0}.$$

An alternative and equivalent way to define confounding, explicitly related to the joint covariance matrix of the distribution of $\boldsymbol{Y}, \boldsymbol{X}$ and $\boldsymbol{Z}$ follows.

Definition 2.1.2. Let $\boldsymbol{Y}$ be the response, $\boldsymbol{X}$ be the covariate and $\boldsymbol{Z}$ be the confounder processes of a linear regression model characterized by the joint distribution expressed in (2.2). The regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$ is confounded by $\boldsymbol{Z}$ if both $\boldsymbol{\Sigma}_{yz} \neq \mathbf{0}$ and $\boldsymbol{\Sigma}_{xz} \neq \mathbf{0}$.

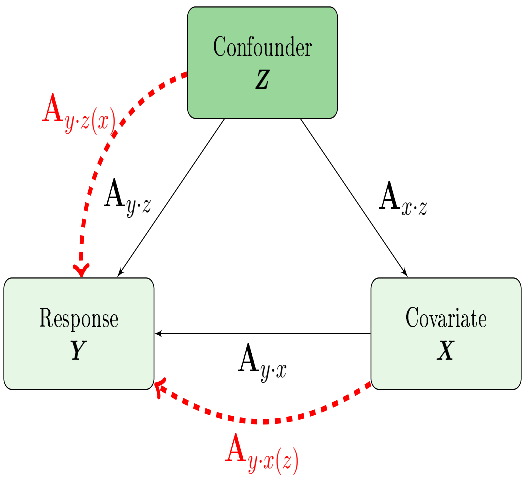

Figure 2.3: Formalization of the confounded simple regression model in (2.20) considering a structured DGP characterized the simple (black) and partial (red) regression matrices.

In other words, confounding occurs, if the unobserved variable is related with both the response and the covariate. As shown in Figure 2.3, the confounder $\boldsymbol{Z}$ influences the response and covariate simultaneously via $\mathbf{A}_{y\cdot z(x)}$ and $\mathbf{A}_{z\cdot x}$, respectively. In an attempt to define confounding in a spatial setting, Thaden and Kneib (2018) describe the separation of direct and indirect covariate effects via path analysis in case of discrete spatial information (Weiber and Mühlhaus, 2014, Chapter 3). The overall spatial information is written as a composition of the direct and indirect spatial effect. In the same spirit, we highlight that the overall effect of the confounder can be expressed trough the regression matrix $\mathbf{A}_{y\cdot z}$ stated in Equation (2.17). The so called indirect effects is linked to the simple regression matrix $\mathbf{A}_{x\cdot z}$, while the direct one is linked to the partial regression matrix $\mathbf{A}_{y\cdot z(x)}$.

In addition, the covariance matrix of $\boldsymbol{Y}|\boldsymbol{X}$ in function of the other conditional ones and the partial regression matrices is reported for future use along the thesis. It also justifies the form of the error covariance matrix of the marginalized DGP in (2.5).

Proposition 2.1.4. *Considering the random vector* $\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z} \sim \mathcal{N}_n\big(\boldsymbol{a}_{y\cdot 0(xz)} + \boldsymbol{A}_{y\cdot z(x)}\boldsymbol{x} + \boldsymbol{A}_{y\cdot x(z)}\boldsymbol{z},\ \boldsymbol{\Sigma}_{y|x,z}\big)$ *and marginalizing it over* $\boldsymbol{Z}$*, the covariance matrix of* $\boldsymbol{Y}|\boldsymbol{X}$ *is:*

$$\boldsymbol{\Sigma}_{y|x} = \boldsymbol{\Sigma}_{y|x,z} + \boldsymbol{A}_{y\cdot z(x)}\boldsymbol{\Sigma}_{z|x}\boldsymbol{A}^\top_{y\cdot z(x)}. \tag{2.22}$$

Proof. From Equation (2.13) we express $\boldsymbol{\Sigma}_y$ as shown below:

$$\begin{aligned}
\boldsymbol{\Sigma}_y &= \boldsymbol{\Sigma}_{y|x,z} + \begin{bmatrix}\mathbf{A}_{y\cdot x(z)} & : & \mathbf{A}_{y\cdot z(x)}\end{bmatrix} \boldsymbol{\Sigma}_{x,z} \begin{bmatrix}\mathbf{A}^\top_{y\cdot x(z)} \\ \mathbf{A}^\top_{y\cdot z(x)}\end{bmatrix} \\
&= \boldsymbol{\Sigma}_{y|x,z} + \mathbf{A}_{y\cdot x(z)}\boldsymbol{\Sigma}_x\mathbf{A}^\top_{y\cdot x(z)} + 2\mathbf{A}_{y\cdot z(x)}\boldsymbol{\Sigma}_z\mathbf{A}^\top_{x\cdot z}\mathbf{A}^\top_{y\cdot x(z)} + \mathbf{A}_{y\cdot z(x)}\boldsymbol{\Sigma}_z\mathbf{A}^\top_{y\cdot z(x)}.
\end{aligned}$$

Then, using Equation (2.10) and basic algebraic manipulations we get the result as follows

$$\begin{aligned}
\boldsymbol{\Sigma}_{y|x} &= \boldsymbol{\Sigma}_y - \mathbf{A}_{y\cdot x}\boldsymbol{\Sigma}_x\mathbf{A}^\top_{y\cdot x} \\
&= \boldsymbol{\Sigma}_{y|x,z} + \mathbf{A}_{y\cdot x(z)}\boldsymbol{\Sigma}_x\mathbf{A}^\top_{y\cdot x(z)} + 2\mathbf{A}_{y\cdot z(x)}\boldsymbol{\Sigma}_z\mathbf{A}^\top_{x\cdot z}\mathbf{A}^\top_{y\cdot x(z)} + \mathbf{A}_{y\cdot z(x)}\boldsymbol{\Sigma}_z\mathbf{A}^\top_{y\cdot z(x)} + \\
&\quad - (\mathbf{A}_{y\cdot x(z)} + \mathbf{A}_{y\cdot z(x)}\mathbf{A}_{z\cdot x})\boldsymbol{\Sigma}_x(\mathbf{A}_{y\cdot x(z)} + \mathbf{A}_{y\cdot z(x)}\mathbf{A}_{z\cdot x})^\top \\
&= \boldsymbol{\Sigma}_{y|x,z} + 2\mathbf{A}_{y\cdot z(x)}\boldsymbol{\Sigma}_z\mathbf{A}^\top_{x\cdot z}\mathbf{A}^\top_{y\cdot x(z)} + \mathbf{A}_{y\cdot z(x)}(\boldsymbol{\Sigma}_z - \mathbf{A}_{z\cdot x}\boldsymbol{\Sigma}_x\mathbf{A}^\top_{z\cdot x})\mathbf{A}^\top_{y\cdot z(x)} + \\
&\quad - 2\mathbf{A}_{y\cdot z(x)}\mathbf{A}_{z\cdot x}\boldsymbol{\Sigma}_x\mathbf{A}^\top_{y\cdot x(z)} \\
&= \boldsymbol{\Sigma}_{y|x,z} + \mathbf{A}_{y\cdot z(x)}\boldsymbol{\Sigma}_{z|x}\mathbf{A}^\top_{y\cdot z(x)}.
\end{aligned}$$

□

2.2 The statistical models and estimators

In practice, confounding is a problem because only the realizations $\boldsymbol{y}$ and $\boldsymbol{x}$ are observed from the data generating process in (2.3), ignoring the presence of $\boldsymbol{Z}$. Given a phenomenon of interest, different specifications of the model can be proposed. The model specified by researchers, is based on their assumptions: in what follows we revise estimators of the regression coefficient of $\boldsymbol{y}$ on $\boldsymbol{x}$ arising from standard linear regression models and linear mixed effect models noting that these estimators can be seen as the same linear function of the data for an appropriate choice of the error covariance matrix. The marginal sampling properties of such estimators will be studied in Chapter 3.

2.2.1 Linear regression model

The simple linear regression model can be cast as

$$\boldsymbol{Y} = \beta_0 \mathbf{1}_n + \beta_x \boldsymbol{X} + \boldsymbol{u}, \qquad \boldsymbol{u} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{V}),$$

where $\boldsymbol{u}$ is the zero-mean normal error term characterized by the error covariance matrix $\mathbf{V}$. This is equivalent to the following distributional assumption

$$\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x} \sim \mathcal{N}_n\left(\beta_0 \mathbf{1}_n + \beta_x \boldsymbol{x}, \mathbf{V}\right). \tag{2.23}$$

Comparing different structures for $\mathbf{V}$ is useful to understand the dynamics of confounding, given that the data generating process marginalized with respect to $\boldsymbol{Z}$ (2.5) has a covariance matrix, $\boldsymbol{\Sigma}_{y|x}$, whose structure also depends on $\boldsymbol{\Sigma}_{z|x}$ (see Proposition 2.1.4).

Regression parameters are estimated via generalized least squares estimators, with the special case of ordinary least squares when $\mathbf{V} = \sigma^2 \mathbf{I}_n$. A number of assumptions about the regressors, the response variables and their relationships characterize this model. Essentially, the following are the most relevant:

1. linearity (*),(**);
2. no perfect multicollinearity (*),(**), i.e. $\text{rank}[\mathbf{1}_n : \boldsymbol{X}] = 2$,
3. exogeneity (*),(**), i.e. $\text{Cov}(\boldsymbol{X}, \boldsymbol{u}) = \mathbf{0}$;
4. homoschedasticity and no autocorrelation (*), i.e. $\mathbf{V} = \sigma^2 \mathbf{I}_n$.

When the assumptions for OLS (*) or GLS (**) are met, both least square estimators are the best linear unbiased ones (BLUE) (Hayashi, 2000; Wooldridge, 2019). If assumptions are violated, the estimator will no longer be BLUE. In general, the dependent variable is *endogenous* by definition because other variables in the model are assumed to explain its variability. The *least squares* (LS) estimator of regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_x)^\top$ is

$$\hat{\boldsymbol{\beta}}^{LS} = (\tilde{\mathbf{X}}^\top \mathbf{V}^{-1} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^\top \mathbf{V}^{-1} \boldsymbol{Y}, \tag{2.24}$$

where $\tilde{\mathbf{X}} = [\mathbf{1}_n : \boldsymbol{X}]$ is the design matrix. Since the parameter of interest is β_x, our focal point in this thesis will be the studying of the sampling properties of the second component of estimator vector in (2.24).

2.2.2 Linear mixed regression model

Mixed models are applied in many disciplines where multiple correlated measurements are made on each unit of interest. This correlation characterizing the data are modeled through the random effects considered in the models. For detailed treatment of the topic see Jiang and Nguyen (2021). In some cases the random effects themselves are of interest, such as in small area estimation (e.g. Rao and Molina (2015)). In some other cases, such as in the analysis of longitudinal data (e.g. Diggle et al. (2002), Pusponegoro et al. (2017)), the random effects are mostly used to model the correlation among the data in order to get correct measure of uncertainty. Sometimes, such as in genetic studies (e.g. Yang et al. (2010), Dandine-Roulland and Perdry (2019)), the variances of random effects are of major importance. In some cases, random effects are included with the aim of capturing the effect of unobserved covariates. This is our case of interest.

A linear mixed model can be expressed as

$$\boldsymbol{Y} = \tilde{\mathbf{X}}\boldsymbol{\beta} + \mathbf{H}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{2.25}$$

where $\tilde{\mathbf{X}} \in \mathbb{R}^{n\times p}$ is the covariate matrix (in our case $p = 2$), $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown regression coefficients which are called fixed effects, $\mathbf{H} \in \mathbb{R}^{n\times q}$ is the random effects design matrix, $\boldsymbol{\gamma} \in \mathbb{R}^q$ is a vector of random effects, and $\boldsymbol{\varepsilon}$ is the error term. Note that both $\boldsymbol{\varepsilon}$ and $\boldsymbol{\gamma}$ are not observable. Compared with the linear regression model in Section 2.2.1, it is assumed that some of the model's coefficients are random. Moreover, as previously mentioned, such introduction of random effects comes from the need to accommodate more complex error structures such as those arising in hierarchical models in order to take into account that observations are correlated. The random part, $\mathbf{H}\boldsymbol{\gamma}$, may take many different forms, giving rise to a rich class of models characterized by some assumptions. The staple ones for (2.25) are that the error vector and the random effects are uncorrelated, mean zero and marked by finite variance. There are different types of linear mixed models, depending on how these models are classified. Here we assume both random effects and error in (2.25) to be normally distributed, $\boldsymbol{\gamma} \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{\Sigma}_\gamma)$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$. Moreover, from the random effects viewpoint, when confounding occurs a mistake is made by assuming that the random effect are independent of the covariate, violating one of the basic assumptions of mixed models (Breslow and Clayton, 1993). This leads to biased estimate of fixed effect on which our attention is concerned.

A standard method of estimation of fixed effects in Gaussian mixed models is the maximum likelihood brought in this framework by Hartley and Rao (1967). Marginalizing with respect to the random effects, a mixed model can be expressed as

$$\boldsymbol{Y} \sim \mathcal{N}_n\left(\tilde{\mathbf{X}}\boldsymbol{\beta}, \boldsymbol{V}\right) \qquad \text{with} \qquad \mathbf{V} = \mathbf{H}\boldsymbol{\Sigma}_\gamma\mathbf{H}^\top + \boldsymbol{\Sigma}_\varepsilon. \tag{2.26}$$

By differentiating its log-likelihood function with respect to the unknown fixed effects and using the standard procedure of finding the ML estimator we get:

$$\hat{\boldsymbol{\beta}}^{ML} = \left(\tilde{\mathbf{X}}^{\top}\mathbf{V}^{-1}\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}^{\top}\mathbf{V}^{-1}\boldsymbol{Y}. \tag{2.27}$$

2.3 Conditional sampling properties of the estimator

Since estimators (2.24) and (2.27) share the same functional form, the sampling properties of the estimators of the regression coefficients can be studied independently on the posited model by defining the following general estimator:

$$\hat{\boldsymbol{\beta}} = \mathbf{J}\boldsymbol{Y}, \tag{2.28}$$

where

$$\mathbf{J} = \left(\tilde{\mathbf{X}}^{\top}\mathbf{S}^{-1}\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}^{\top}\mathbf{S}^{-1} \tag{2.29}$$

in which the weighting matrix $\mathbf{S}$ varies with the assumed model. For example, $\mathbf{S} = \mathbf{I}_n$ delivers the OLS estimator.

Investigating the effect of confounding on the estimation of $\mathcal{B}_{y\cdot x(z)}$, namely the partial effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$, means studying the sampling properties of the estimators $\hat{\beta}_x$, under the models described above. In this task, the effect of $\boldsymbol{Z}$ on $\boldsymbol{X}$ and $\boldsymbol{Y}$ is crucial. In line with the works of Paciorek (2010) and Page et al. (2017), we start presenting the sampling distribution of the estimator (2.28) conditionally on $\boldsymbol{X}$, and then, in the next chapter, we will treat the obtained properties as random variables in order to obtain their expected value.

As a first step, the following proposition introduces a well known result, i.e. the distribution of $\hat{\boldsymbol{\beta}}|\boldsymbol{X}$, that is re-written under our notation and as a function of quantities useful for further developments.

Proposition 2.3.1. *The estimator $\hat{\boldsymbol{\beta}}$ in (2.28) conditional on $\boldsymbol{X}$ has the following sampling distribution:*

$$\hat{\boldsymbol{\beta}}|\boldsymbol{X} \sim \mathcal{N}_2\left(\boldsymbol{J}(\mathcal{B}_{y\cdot 0(xz)}\boldsymbol{1}_n + \boldsymbol{A}_{y\cdot x}\boldsymbol{X}), \boldsymbol{J}\boldsymbol{\Sigma}_{y|x}\boldsymbol{J}^{\top}\right). \tag{2.30}$$

Proof. Using the main properties of normal distribution, the moments of $\boldsymbol{Z}|\boldsymbol{X}$ can be obtained from Equation (2.10) and due to the result in Proposition 2.1.4, the following conditional sampling properties of estimator $\hat{\boldsymbol{\beta}}$ defined in (2.28) are obtained as

$$\mathbb{E}_Y\left[\hat{\boldsymbol{\beta}}|\boldsymbol{X}\right] = \mathbb{E}_Y\left[\mathbf{J}\boldsymbol{Y}|\boldsymbol{X}\right]$$

$$
\begin{aligned}
&= \mathbf{J}\,\mathbb{E}_{Z,\varepsilon}\left[\mathcal{B}_{y\cdot 0(xz)}\mathbf{1}_n + \mathcal{B}_{y\cdot x(z)}\boldsymbol{X} + \mathcal{B}_{y\cdot z(x)}\boldsymbol{Z} + \boldsymbol{\varepsilon}_{y|x,z}\middle|\boldsymbol{X}\right] \\
&= \mathbf{J}\left[\mathcal{B}_{y\cdot 0(xz)}\mathbf{1}_n + \mathcal{B}_{y\cdot x(z)}\boldsymbol{X} + \mathcal{B}_{y\cdot z(x)}\mathbb{E}_Z\left[\boldsymbol{Z}|\boldsymbol{X}\right] + \mathbb{E}_\varepsilon[\boldsymbol{\varepsilon}_{y|x,z}]\right] \\
&= \mathbf{J}\left[\mathcal{B}_{y\cdot 0(xz)}\mathbf{1}_n + (\mathcal{B}_{y\cdot x(z)} + \mathcal{B}_{y\cdot z(x)}\mathbf{A}_{z\cdot x})\boldsymbol{X}\right] \\
&= \mathbf{J}[\mathcal{B}_{y\cdot 0(xz)}\mathbf{1}_n + \mathbf{A}_{y\cdot x}\boldsymbol{X}]
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{V}_Y\left[\hat{\boldsymbol{\beta}}|\boldsymbol{X}\right] &= \mathbb{V}_Y\left[\mathbf{J}\boldsymbol{Y}|\boldsymbol{X}\right] \\
&= \mathbf{J}\left(\mathcal{B}^2_{y\cdot z(x)}\mathbb{V}_Z\left[\boldsymbol{Z}|\boldsymbol{X}\right] + \mathbb{V}_\varepsilon\left[\boldsymbol{\varepsilon}_{y|x,z}|\boldsymbol{X}\right]\right)\mathbf{J}^\top \\
&= \mathbf{J}\left(\mathcal{B}^2_{y\cdot z(x)}\boldsymbol{\Sigma}_z - \mathcal{B}^2_{y\cdot z(x)}\boldsymbol{\Sigma}_{zx}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xz} + \sigma^2_{y|x,z}\mathbf{I}_n\right)\mathbf{J}^\top \\
&= \mathbf{J}\boldsymbol{\Sigma}_{y|x}\mathbf{J}^\top
\end{aligned}
$$

□

Since our target parameter is $\mathcal{B}_{y\cdot x(z)}$, we are particularly interested in the sampling properties of the second element of $\hat{\boldsymbol{\beta}}$, that is $\hat{\beta}_x$. From the result in Proposition 2.3.1, it immediately follows that:

$$
\text{Bias}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right] = \mathcal{B}_{y\cdot z(x)}\mathbf{J}_{2\bullet}\boldsymbol{\Sigma}_{zx}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{X}, \tag{2.31}
$$

and

$$
\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right] = \mathbf{J}_{2\bullet}\boldsymbol{\Sigma}_{y|x}\mathbf{J}_{2\bullet}^\top, \tag{2.32}
$$

where $\mathbf{J}_{2\bullet}$ indicates the second row of the matrix $\mathbf{J}$. We emphasize that the adopted notation is aimed at stressing that we are evaluating moments in the sampling space with respect to the response process $\boldsymbol{Y}$, whereas we still consider $\boldsymbol{X}$ as a random variable. Furthermore, a crucial aspect of these results is that they confirm that when the response and confounder are correlated given the covariate and the assumption of independence or exogeneity in Item 3 is not met, i.e. $\boldsymbol{\Sigma}_{xz} \neq \mathbf{0}$, the estimator $\hat{\beta}_x$ is biased, as previously formalized in Section 2.1.3.

As anticipated, similar results are well-known in the widely studied spatial framework and regarding particular assumptions about the cross-correlation structure between covariate and confounder (see Paciorek (2010), Page et al. (2017), Nobre et al. (2021), Marques et al. (2022)). Due to the fact that our primary scope is evaluating the impact of confounding of regression coefficients more in depth and within general framework, i.e. reporting results as unconstrained as possible from the restrictions coming from the building assumptions regarding the structure of covariance matrices made in the DGP (2.3), we consider Proposition 2.3.1 and considerations already present in literature to be the starting point of this dissertation.

Chapter 3

Main features of confounding in terms of quadratic forms

After the presentation of theoretical framework underlying this thesis, in this chapter we provide exact formulas for the marginal sampling properties of the estimator defined in (2.28). Using the law of total expectation and the law of total variance, these quantities can be found noting that the conditional sampling properties reported in Section 2.3 can be expressed as ratios of dependent quadratic forms in Gaussian random variables. In Section 3.1 some key concepts about quadratic forms in Gaussian variables are briefly outlined. In Section 3.2 we obtain the marginal sampling properties of the estimator of the parameter of interest. In Section 3.3 some remarks on estimator variance and mean square error are exhibited. In particular, we bring out how these quantities may be decomposed onto a part confounding-dependent and the remaining unrelated with it. In Section 3.4 we focus on the unstructured DGP providing the exact marginal distribution of the OLS estimator. Finally, in Section 3.5, some relevant features of structured DGPs in terms of covariance and cross-covariance matrix's eigenvalues are described. In particular, we study two relevant random variables, the sampling variance and the inverse smoothness of a random vector, whose expected values quantify the variability and the inverse of the smoothing level of a random process. The behaviour of these quantities when varying the structure parameters characterizing the processes is also shown, assuming three different covariance functions. In this way, the fundamental role played by the eigenvalues of the chosen covariance matrix is emphasized. Section 3.6 proposes some measures aimed at quantifying confounding. Specifically, we put forward a proxy for the estimator bias, its first order Laplace approximation, as a representative scalar quantity for the magnitude of confounding.

As a starting point, we proof that the sampling properties of $\hat{\beta}_x|\boldsymbol{X}$ reported in Proposition 2.3.1 may be expressed in terms of QFs. Indeed, they are random variables giving rise to ratios of dependent quadratic forms in Gaussian random variables. This fact is formalized in the following Lemma.

Lemma 3.0.1. *Considering the data generating process in* (2.3) *and the model in* (2.23), *where* $\boldsymbol{X}$ *and* $\boldsymbol{Z}$ *are jointly distributed as is* (2.11)*:*

(i)

$$\mathrm{Bias}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right] = \mathcal{B}_{y\cdot z(x)}\frac{\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{A}_{z\cdot x}\boldsymbol{X}}{\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{X}}, \tag{3.1}$$

(ii)

$$\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right] = \frac{\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{\Sigma}_{y|x}\boldsymbol{\Delta}\boldsymbol{X}}{(\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{X})^2}, \tag{3.2}$$

where $\boldsymbol{\Delta} = \boldsymbol{S}^{-1} - \dfrac{\boldsymbol{S}^{-1}\boldsymbol{1}_n\boldsymbol{1}_n^\top\boldsymbol{S}^{-1}}{\boldsymbol{1}_n^\top\boldsymbol{S}^{-1}\boldsymbol{1}_n}$ *is the weighted centering matrix in which* $\boldsymbol{S}$ *is the covariance matrix depending upon the posited model.*

Proof. We can rewrite the term $\mathbf{J}$ defined in Equation (2.29) as follows:

$$\mathbf{J} = \begin{bmatrix} \mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n & \mathbf{1}_n^\top\mathbf{S}^{-1}\boldsymbol{X} \\ \boldsymbol{X}^\top\mathbf{S}^{-1}\mathbf{1}_n & \boldsymbol{X}^\top\mathbf{S}^{-1}\boldsymbol{X}\end{bmatrix}^{-1}\tilde{\mathbf{X}}^\top\mathbf{S}^{-1}$$

$$= \frac{1}{\mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n}\begin{bmatrix} 1 & \dfrac{\mathbf{1}_n^\top\mathbf{S}^{-1}\boldsymbol{X}}{\mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n} \\ \dfrac{\boldsymbol{X}^\top\mathbf{S}^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n} & \dfrac{\boldsymbol{X}^\top\mathbf{S}^{-1}\boldsymbol{X}}{\mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n}\end{bmatrix}^{-1}\tilde{\mathbf{X}}^\top\mathbf{S}^{-1}.$$

Defining $\tilde{\delta} = \mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n = \sum_{i=1}^n \delta_{ij}$ where δ_{ij} is one element of the matrix $\mathbf{S}^{-1}$, we get

$$\mathbf{J} = \frac{1}{\tilde{\delta}}\left(\frac{\boldsymbol{X}^\top\mathbf{S}^{-1}\boldsymbol{X}}{\tilde{\delta}} - \frac{\boldsymbol{X}^\top\mathbf{S}^{-1}\mathbf{1}_n\mathbf{1}_n^\top\mathbf{S}^{-1}\boldsymbol{X}}{\tilde{\delta}^2}\right)^{-1}\begin{bmatrix} \dfrac{\boldsymbol{X}^\top\mathbf{S}^{-1}\boldsymbol{X}}{\mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n} & -\dfrac{\mathbf{1}_n^\top\mathbf{S}^{-1}\boldsymbol{X}}{\mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n} \\ -\dfrac{\boldsymbol{X}^\top\mathbf{S}^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n} & 1\end{bmatrix}\begin{bmatrix}\mathbf{1}_n^\top\mathbf{S}^{-1} \\ \boldsymbol{X}^\top\mathbf{S}^{-1}\end{bmatrix}$$

$$= \frac{1}{\boldsymbol{X}^\top\mathbf{S}^{-1}\boldsymbol{X} - \dfrac{\boldsymbol{X}^\top\mathbf{S}^{-1}\mathbf{1}_n\mathbf{1}_n^\top\mathbf{S}^{-1}\boldsymbol{X}}{\tilde{\delta}}}\begin{bmatrix}\dfrac{\boldsymbol{X}^\top\mathbf{S}^{-1}\boldsymbol{X}\mathbf{1}_n^\top\mathbf{S}^{-1} - \mathbf{1}_n^\top\mathbf{S}^{-1}\boldsymbol{X}\boldsymbol{X}^\top\mathbf{S}^{-1}}{\tilde{\delta}} \\ \boldsymbol{X}^\top\mathbf{S}^{-1} - \dfrac{\boldsymbol{X}^\top\mathbf{S}^{-1}\mathbf{1}_n\mathbf{1}_n^\top\mathbf{S}^{-1}}{\tilde{\delta}}\end{bmatrix}$$

$$= \frac{1}{\boldsymbol{X}^\top \left(\mathbf{S}^{-1} - \frac{\mathbf{S}^{-1}\mathbf{1}_n\mathbf{1}_n^\top\mathbf{S}^{-1}}{\tilde{\delta}}\right)\boldsymbol{X}} \begin{bmatrix} \dfrac{\boldsymbol{X}^\top\mathbf{S}^{-1}\boldsymbol{X}\mathbf{1}_n^\top\mathbf{S}^{-1} - \mathbf{1}_n^\top\mathbf{S}^{-1}\boldsymbol{X}\boldsymbol{X}^\top\mathbf{S}^{-1}}{\tilde{\delta}} \\ \boldsymbol{X}^\top\left(\mathbf{S}^{-1} - \dfrac{\mathbf{S}^{-1}\mathbf{1}_n\mathbf{1}_n^\top\mathbf{S}^{-1}}{\tilde{\delta}}\right) \end{bmatrix}.$$

The second row of $\mathbf{J}$, denoted as $\mathbf{J}_{2\bullet}$, is

$$\mathbf{J}_{2\bullet} = \frac{\boldsymbol{X}^\top\left(\mathbf{S}^{-1} - \dfrac{\mathbf{S}^{-1}\mathbf{1}_n\mathbf{1}_n^\top\mathbf{S}^{-1}}{\tilde{\delta}}\right)}{\boldsymbol{X}^\top\left(\mathbf{S}^{-1} - \frac{\mathbf{S}^{-1}\mathbf{1}_n\mathbf{1}_n^\top\mathbf{S}^{-1}}{\tilde{\delta}}\right)\boldsymbol{X}} = \frac{\boldsymbol{X}^\top\boldsymbol{\Delta}}{\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{X}}.$$

Plugging it into (2.31), we get

$$\mathrm{Bias}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right] = \mathcal{B}_{y\cdot z(x)}\frac{\boldsymbol{X}^\top\boldsymbol{\Delta}}{\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{X}}\boldsymbol{\Sigma}_{zx}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{X},$$

which proofs point (i). In addition, since $\boldsymbol{\Delta} = \boldsymbol{\Delta}^\top$, from (2.32) it is trivial to prove (ii). □

Regarding the weighted centering matrix, it is significant to highlight that it is the product of a symmetric and an idempotent matrices. In fact, through the following representation

$$\boldsymbol{\Delta} = \mathbf{S}^{-1}\left(\mathbf{I}_n - \frac{\mathbf{1}_n\mathbf{1}_n^\top\mathbf{S}^{-1}}{\mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n}\right) = \mathbf{S}^{-1}\tilde{\boldsymbol{\Delta}}$$

it is possible to verify the idempotent condition for $\tilde{\boldsymbol{\Delta}}$:

$$\tilde{\boldsymbol{\Delta}}\tilde{\boldsymbol{\Delta}} = \mathbf{I}_n + \frac{\mathbf{1}_n\left(\mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n\right)\mathbf{1}_n^\top\mathbf{S}^{-1}}{\left(\mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n\right)^2} - 2\frac{\mathbf{1}_n\mathbf{1}_n^\top\mathbf{S}^{-1}}{\mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n} = \mathbf{I}_n - \frac{\mathbf{1}_n\mathbf{1}_n^\top\mathbf{S}^{-1}}{\mathbf{1}_n^\top\mathbf{S}^{-1}\mathbf{1}_n} = \tilde{\boldsymbol{\Delta}}.$$

An idempotent matrix has one zero eigenvalue: this means that $\tilde{\boldsymbol{\Delta}}\mathbf{1}_n = \mathbf{0}$ and since the rank of a matrix does not change when we multiply it by a full-rank matrix, then, assuming $\mathbf{S} \succ 0$, $\mathrm{rank}\,\boldsymbol{\Delta} = \mathrm{rank}\,\tilde{\boldsymbol{\Delta}} = n-1$, with the minimum eigenvalue of the weighted centering matrix equal to zero.

The marginal moments of $\hat{\beta}_x$ can be retrieved using the law of iterated expectation and the law of the total variance (Billingsley, 1995, Chapter 6). As noted by Aronow and Miller (2019), these laws are relevant because of their ability to express unconditional expectation and variance in term of conditional ones, allowing us to make some calculations more tractable. Indeed, we seek the following quantities:

$$\mathbb{E}_{Y,X}\left[\hat{\beta}_x\right] = \mathbb{E}_X\left[\mathbb{E}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right], \tag{3.3}$$

$$\mathbb{V}_{Y,X}\left[\hat{\beta}_x\right] = \mathbb{E}_X\left[\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] + \mathbb{V}_X\left[\mathbb{E}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right]. \tag{3.4}$$

The developments in what follows will be based on the theory of QF: some basic concepts on the topic are reported in the next section.

3.1 Quadratic forms in Gaussian random variables

Considering the random process $\boldsymbol{X} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{\Sigma}_x)$, it is possible to define the quadratic form (QF, see Provost and Mathai, 1992, for a comprehensive overview of the topic) associated to a symmetric matrix $\mathbf{A} \in \mathcal{S}^n$ as:

$$Q_A(\boldsymbol{X}) = \boldsymbol{X}^\top \mathbf{A} \boldsymbol{X}.$$

We remark that considering $\mathbf{A}$ as symmetric does not imply a loss of generality. Indeed, if $\mathbf{A}$ is not symmetric, the quadratic form $Q_A(\boldsymbol{X})$ is equivalent to the quadratic form $Q_{A^s}(\boldsymbol{X})$ where $\mathbf{A}^s = \frac{\mathbf{A}+\mathbf{A}^\top}{2} \in \mathcal{S}^n$ is the symmetric part of $\mathbf{A}$. In fact, $\mathbf{A}$ can be written as:

$$\mathbf{A} = \frac{\mathbf{A}+\mathbf{A}^\top}{2} + \frac{\mathbf{A}-\mathbf{A}^\top}{2},$$

and $\boldsymbol{X}^\top(\mathbf{A} - \mathbf{A}^\top)\boldsymbol{X} = 0$, getting $Q_A(\boldsymbol{X}) = Q_{A^s}(\boldsymbol{X})$. Then, there is one-to-one correspondence between quadratic forms and symmetric matrices that determine them.

Decomposing the covariance matrix as follows $\mathbf{\Sigma}_x = \mathbf{\Sigma}_x^{1/2}\mathbf{\Sigma}_x^{1/2}$, we note that $Q_A(\boldsymbol{X})$ can be expressed as a function of a standard multivariate normal vector $\boldsymbol{\nu} = \mathbf{\Sigma}_x^{-1/2}\boldsymbol{X}$ such that:

$$\begin{aligned} Q_A(\boldsymbol{X}) &= \boldsymbol{X}^\top \mathbf{\Sigma}_x^{-1/2}\mathbf{\Sigma}_x^{1/2}\mathbf{A}\mathbf{\Sigma}_x^{1/2}\mathbf{\Sigma}_x^{-1/2}\boldsymbol{X}^\top \\ &= \boldsymbol{\nu}^\top \tilde{\mathbf{A}} \boldsymbol{\nu} \\ &= Q_{\tilde{A}}(\boldsymbol{\nu}), \end{aligned}$$

where $\tilde{\mathbf{A}} = \mathbf{\Sigma}_x^{1/2}\mathbf{A}\mathbf{\Sigma}_x^{1/2}$.

Many properties of $Q_A(\boldsymbol{X})$, such as moments and distribution function, are strictly related to the eigenvalues of the matrix $\tilde{\mathbf{A}}$. We indicate them with

$$\boldsymbol{\lambda}(\tilde{\mathbf{A}}) = \left(\lambda(\tilde{\mathbf{A}})_1, \ldots, \lambda(\tilde{\mathbf{A}})_n\right)^\top,$$

such that $\lambda\big(\tilde{\mathbf{A}}\big)_1 \geq \lambda\big(\tilde{\mathbf{A}}\big)_2 \geq \cdots \geq \lambda\big(\tilde{\mathbf{A}}\big)_n$. Indeed, the expected value is

$$\mathbb{E}_X[Q_A(\boldsymbol{X})] = \sum_{i=1}^{n} \lambda(\mathbf{A}\boldsymbol{\Sigma}_x)_i = \sum_{i=1}^{n} \lambda\big(\tilde{\mathbf{A}}\big)_i = \operatorname{tr}\big(\tilde{\mathbf{A}}\big), \tag{3.5}$$

and the moment generating function is

$$\phi_{Q_A(\boldsymbol{X})}(t) = \mathbb{E}_X\left[e^{tQ_A(\boldsymbol{X})}\right] = |\mathbf{I}_n - 2t\mathbf{A}\boldsymbol{\Sigma}_x|^{-1/2} = \prod_{i=1}^{n}\left(1 - 2t\lambda\big(\tilde{\mathbf{A}}\big)_i\right)^{-1/2},$$

where $|\cdot|$ denotes the determinant. Furthermore, if $\tilde{\mathbf{A}}$ is symmetric and idempotent, $Q_A(\boldsymbol{X})$ has a Chi-square distribution with $\operatorname{tr}\big(\tilde{\mathbf{A}}\big)$ degree of freedom indicated with $\chi^2_{\operatorname{tr}(\tilde{A})}$. When $\tilde{\mathbf{A}} = \mathbf{M} = \mathbf{I}_n - \frac{\mathbf{1}_n\mathbf{1}_n^\top}{n}$, where $\mathbf{M}$ is the centering matrix, $Q_M(\boldsymbol{\nu}) \sim \chi^2_{n-1}$.

Once QFs are defined, our attention moves to ratios of powers of dependent QFs. Let us consider a further positive semidefinite matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$, then we introduce the following ratio of QFs

$$\begin{aligned} R^{p,q}_{A,B}(\boldsymbol{X}) &= \frac{(\boldsymbol{X}^\top\mathbf{A}\boldsymbol{X})^p}{(\boldsymbol{X}^\top\mathbf{B}\boldsymbol{X})^q} \\ &= \frac{(\boldsymbol{\nu}^\top\tilde{\mathbf{A}}\boldsymbol{\nu})^p}{(\boldsymbol{\nu}^\top\tilde{\mathbf{B}}\boldsymbol{\nu})^q} \\ &= R^{p,q}_{\tilde{A},\tilde{B}}(\boldsymbol{\nu}), \end{aligned} \tag{3.6}$$

where $p \geq 0$, $q \geq 0$ are integer and $\tilde{\mathbf{B}} = \boldsymbol{\Sigma}_x^{1/2}\mathbf{B}\boldsymbol{\Sigma}_x^{1/2}$.

Computing the expectation of such random variable is of primary interest for the developments in the chapter. It represents a well known problem of numerical probability and it is faced in several works, such as Magnus (1986), Roberts (1995) and Bao and Kan (2013). The latter provides an up-to-date review and includes most of the exploited results. Firstly, $\mathbb{E}_X\left[R^{p,q}_{A,B}(\boldsymbol{X})\right]$ exists if and only if $\operatorname{rank}\tilde{\mathbf{B}} > 2q$, and it can be numerically evaluated as

$$\mathbb{E}_X\left[R^{p,q}_{A,B}(\boldsymbol{X})\right] = \frac{1}{\Gamma(q)}\int_0^\infty t^{q-1}\left.\frac{\partial^p}{\partial t_1^p}\phi(t_1,t_2)\right|_{t_1=0,t_2=-t} \mathrm{d}t, \tag{3.7}$$

where $\phi(t_1,t_2) = |\mathbf{I}_n - 2t_1\tilde{\mathbf{A}} - 2t_2\tilde{\mathbf{B}}|^{-1/2}$ is the joint moment generating function (jmgf) of $\boldsymbol{X}^\top\mathbf{A}\boldsymbol{X}$ and $\boldsymbol{X}^\top\mathbf{B}\boldsymbol{X}$ (Sawa, 1978).

Since many statistical quantities can be written as ratios of quadratic forms, the computation of the expected value has been of great interest to statisticians and econometricians.

Bao and Kan (2013) remind us that there are two approaches for evaluating it: the first is by integration and it starts with Sawa (1978). This method is by far the most popular one in the literature and Xiao-Li (2005) provides a very good review of the literature on the subject. The second approach relies on some infinite series expansion of the ratio involving the invariant polynomials of matrix argument (Smith, 1989). We are concerned about obtaining computationally efficient expression of the expectation of the ratio of dependent QFs defined in (3.6) using the first approach. Relatively straightforward expressions are available for moments of a QF in normal variables in the case in which the variable is spherical. These moments appear as simple integrals which can be evaluated numerically in a straightforward manner. The most popular method for its numerical evaluation is to make use of the results in Sawa (1978) and Cressie et al. (1981).

3.2 Marginal sampling properties of $\hat{\beta}_x$ in terms of quadratic forms

Concerning the expected bias of $\hat{\beta}_x$, note that Equation (3.1) can be cast in the form of Equation (3.6) by posing $\mathbf{A} = \mathbf{\Delta A}_{z\cdot x}$ and $\mathbf{B} = \mathbf{\Delta}$:

$$
\begin{aligned}
R^{1,1}_{A,B}(\boldsymbol{X}) &= \frac{\boldsymbol{X}^\top \mathbf{\Delta A}_{z\cdot x}\boldsymbol{X}}{\boldsymbol{X}^\top \mathbf{\Delta}\boldsymbol{X}} \\
&= \frac{(\mathbf{\Sigma}_x^{-1/2}\boldsymbol{X})^\top \mathbf{\Sigma}_x^{1/2}\mathbf{\Delta\Sigma}_{zx}\mathbf{\Sigma}_x^{-1}\mathbf{\Sigma}_x^{1/2}(\mathbf{\Sigma}_x^{-1/2}\boldsymbol{X})}{(\mathbf{\Sigma}_x^{-1/2}\boldsymbol{X})^\top \mathbf{\Sigma}_x^{1/2}\mathbf{\Delta\Sigma}_x^{1/2}(\mathbf{\Sigma}_x^{-1/2}\boldsymbol{X})} \\
&= \frac{\boldsymbol{\nu}^\top \mathbf{\Sigma}_x^{1/2}\mathbf{\Delta\Sigma}_{zx}\mathbf{\Sigma}_x^{-1/2}\boldsymbol{\nu}}{\boldsymbol{\nu}^\top \mathbf{\Sigma}_x^{1/2}\mathbf{\Delta\Sigma}_x^{1/2}\boldsymbol{\nu}} \\
&= R^{1,1}_{\tilde{A},\tilde{B}}(\boldsymbol{\nu})
\end{aligned} \tag{3.8}
$$

where $\tilde{\mathbf{A}} = \mathbf{\Sigma}_x^{1/2}\mathbf{\Delta\Sigma}_{zx}\mathbf{\Sigma}_x^{-1/2}$ and $\tilde{\mathbf{B}} = \mathbf{\Sigma}_x^{1/2}\mathbf{\Delta\Sigma}_x^{1/2}$. The joint moment generating function of the numerator $\boldsymbol{\nu}^\top\tilde{\mathbf{A}}\boldsymbol{\nu}$ and the denominator $\boldsymbol{\nu}^\top\tilde{\mathbf{B}}\boldsymbol{\nu}$ in (3.8) is given by

$$
\phi(t_1, t_2) = |\mathbf{I}_n - 2t_1\mathbf{\Sigma}_x^{1/2}\mathbf{\Delta\Sigma}_{zx}\mathbf{\Sigma}_x^{-1/2} - 2t_2\mathbf{\Sigma}_x^{1/2}\mathbf{\Delta\Sigma}_x^{1/2}|^{-1/2}.
$$

Let consider the following spectral decomposition

$$
\tilde{\mathbf{B}} = \mathbf{\Sigma}_x^{1/2}\mathbf{\Delta\Sigma}_x^{1/2} = \mathbf{P\Lambda P}^\top,
$$

where

$$
\mathbf{\Lambda} = \operatorname{diag}(\boldsymbol{\lambda}) = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \tag{3.9}
$$

contains the eigenvalues of $\tilde{\mathbf{B}}$ and $\mathbf{P}$ is the orthogonal eigenvectors matrix such that $\mathbf{P}\mathbf{P}^\top = \mathbf{P}^\top\mathbf{P} = \mathbf{I}_n$. As a result, the jmgf takes the following form

$$\begin{aligned}
\phi(t_1,t_2) &= |\mathbf{P}^\top\mathbf{P}|^{-1/2}|\mathbf{I}_n - 2t_1\mathbf{\Sigma}_x^{1/2}\mathbf{\Delta}\mathbf{\Sigma}_{zx}\mathbf{\Sigma}_x^{-1/2} - 2t_2\mathbf{\Sigma}_x^{1/2}\mathbf{\Delta}\mathbf{\Sigma}_x^{1/2}|^{-1/2} \\
&= |\mathbf{I}_n - 2t_1\mathbf{P}^\top\mathbf{\Sigma}_x^{1/2}\mathbf{\Delta}\mathbf{\Sigma}_{zx}\mathbf{\Sigma}_x^{-1/2}\mathbf{P} - 2t_2\mathbf{P}^\top\mathbf{\Sigma}_x^{1/2}\mathbf{\Delta}\mathbf{\Sigma}_x^{1/2}\mathbf{P}|^{-1/2} \\
&= |\mathbf{I}_n - 2t_1\boldsymbol{C}_1 - 2t_2\mathbf{\Lambda}|^{-1/2} \\
&= |\mathbf{F}|^{-1/2}
\end{aligned}$$

in which

$$\begin{aligned}
\mathbf{C}_1 &= \mathbf{P}^\top\tilde{\mathbf{A}}\mathbf{P} \\
&= \mathbf{P}^\top\mathbf{\Sigma}_x^{1/2}\mathbf{\Delta}\mathbf{\Sigma}_{zx}\mathbf{\Sigma}_x^{-1/2}\mathbf{P},
\end{aligned}$$

with (i,j)-th element $c_{1,ij}$. The expected value of (3.8) is now obtainable using Equation (3.7) with $p = q = 1$:

$$\mathbb{E}_X\left[R_{A,B}^{1,1}(\boldsymbol{X})\right] = \int_0^\infty \frac{\partial}{\partial t_1}\phi(t_1,t_2)\bigg|_{t_1=0,t_2=-t} dt. \tag{3.10}$$

Since, $\frac{\partial}{\partial t_1}\mathbf{F} = -2\mathbf{C}_1$, for $p = 1$

$$\begin{aligned}
\frac{\partial}{\partial t_1}\phi(t_1,t_2) &= -\frac{1}{2}|\mathbf{F}|^{-3/2}\left(\frac{\partial}{\partial t_1}|\mathbf{F}|\right) \\
&= |\mathbf{F}|^{-1/2}\mathrm{tr}\left(\mathbf{F}^{-1}\boldsymbol{C}_1\right),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial t_1}\phi(t_1,t_2)\bigg|_{t_1=0,t_2=-t} &= |\mathbf{I}_n + 2t\mathbf{\Lambda}|^{-1/2}\mathrm{tr}\left[(\mathbf{I}_n + 2t\mathbf{\Lambda})^{-1}\boldsymbol{C}_1\right] \\
&= \prod_{i=1}^n (1+2\lambda_i t)^{-1/2}\sum_{j=1}^n \frac{c_{1,jj}}{1+2\lambda_j t}.
\end{aligned}$$

From Equation (3.3), the marginal expectation of the estimator $\hat{\beta}_x$ is

$$\begin{aligned}
\mathbb{E}_{Y,X}\left[\hat{\beta}_x\right] &= \mathbb{E}_X\left[\mathbb{E}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] \\
&= \mathcal{B}_{y\cdot x(z)} + \mathcal{B}_{y\cdot z(x)}\int_0^\infty |\mathbf{I}_n + 2t\mathbf{\Lambda}|^{-1/2}\mathrm{tr}\left[(\mathbf{I}_n + 2t\mathbf{\Lambda})^{-1}\boldsymbol{C}_1\right] dt \\
&= \mathcal{B}_{y\cdot x(z)} + \mathcal{B}_{y\cdot z(x)}\sum_{j=1}^n c_{1,jj}\int_0^\infty \prod_{i=1}^n (1+2\lambda_i t)^{-1/2}\frac{1}{1+2\lambda_j t} dt
\end{aligned}$$

$$= \mathcal{B}_{y\cdot x(z)} + \mathrm{Bias}_{Y,X}\left[\hat{\beta}_x\right], \tag{3.11}$$

in which the estimator marginal bias is highlighted.

Concerning the sampling variance, it is possible to re-write Equation (3.4) as

$$\begin{aligned}
\mathbb{V}_{Y,X}\left[\hat{\beta}_x\right] &= \mathbb{E}_X\left[\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] + \mathbb{V}_X\left[\mathrm{Bias}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] \\
&= \mathbb{E}_X\left[\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] + \mathbb{E}_X\left[\mathrm{Bias}^2_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] - \mathrm{Bias}^2_{Y,X}\left[\hat{\beta}_x\right] \qquad (3.12)\\
&= \mathbb{E}_X\left[\frac{\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{\Sigma}_{y|x}\boldsymbol{\Delta}\boldsymbol{X}}{(\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{X})^2}\right] + \\
&\quad + \mathcal{B}^2_{y\cdot z(x)}\mathbb{E}_X\left[\left(\frac{\boldsymbol{X}^\top\boldsymbol{\Delta}\mathbf{A}_{z\cdot x}\boldsymbol{X}}{\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{X}}\right)^2\right] - \mathrm{Bias}^2_{Y,X}\left[\hat{\beta}_x\right] \qquad (3.13)
\end{aligned}$$

in which using the Formula (3.7) with $p = 1$ and $q = 2$ allows to determine

$$\begin{aligned}
\mathbb{E}_X\left[\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] &= \mathbb{E}_X\left[\frac{\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{\Sigma}_{y|x}\boldsymbol{\Delta}\boldsymbol{X}}{(\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{X})^2}\right] \\
&= \sum_{j=1}^{n} c_{2,jj}\int_0^\infty t\prod_{i=1}^{n}(1+2\lambda_i t)^{-1/2}\frac{1}{1+2\lambda_j t}\,dt,
\end{aligned} \tag{3.14}$$

where $c_{2,ij}$ is the(i,j)-th element of the matrix $\mathbf{C}_2 = \mathbf{P}^\top\boldsymbol{\Sigma}_x^{1/2}\boldsymbol{\Delta}\boldsymbol{\Sigma}_{y|x}\boldsymbol{\Delta}\boldsymbol{\Sigma}_x^{1/2}\mathbf{P}$.

The computation of the quantity $\mathbb{E}_X\left[\mathrm{Bias}^2_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right]$ requires $p = 2$ in Formula (3.7). Following the line of Paolella (2018) and using the properties of the trace reported in Theorem A.0.1, we first observe that

$$\begin{aligned}
\frac{\partial^2|\mathbf{F}|}{\partial t_1^2} &= \frac{\partial}{\partial t_1}\left(-2|\mathbf{F}|\,\mathrm{tr}(\mathbf{F}^{-1}\mathbf{C}_1)\right) \\
&= \frac{\partial\left(-2|\mathbf{F}|\right)}{\partial t_1}\mathrm{tr}(\mathbf{F}^{-1}\mathbf{C}_1) - 2|\mathbf{F}|\frac{\partial}{\partial t_1}\left(\mathrm{tr}(\mathbf{F}^{-1}\mathbf{C}_1)\right) \\
&= 4|\mathbf{F}|\mathrm{tr}^2(\mathbf{F}^{-1}\mathbf{C}_1) - 2|\mathbf{F}|\,\mathrm{tr}\left(\mathbf{F}^{-1}\frac{\partial\mathbf{C}_1}{\partial t_1} + \frac{\partial\mathbf{F}^{-1}}{\partial t_1}\mathbf{C}_1\right) \\
&= 4|\mathbf{F}|\left(\mathrm{tr}^2(\mathbf{F}^{-1}\mathbf{C}_1) - \mathrm{tr}(\mathbf{F}^{-1}\mathbf{C}_1)^2\right)
\end{aligned}$$

and

$$\frac{\partial^2}{\partial t_1^2}\phi(t_1,t_2) = \frac{\partial}{\partial t_1}\left(-\frac{1}{2}|\mathbf{F}|^{-3/2}\frac{\partial|\mathbf{F}|}{\partial t_1}\right)$$

$$
\begin{aligned}
&= \frac{3}{4}|\mathbf{F}|^{-5/2}\left(\frac{\partial|\mathbf{F}|}{\partial t_1}\right)^2 - \frac{1}{2}|\mathbf{F}|^{-3/2}\frac{\partial^2|\mathbf{F}|}{\partial t_1^2} \\
&= \frac{3}{4}|\mathbf{F}|^{-5/2}\left(4|\mathbf{F}|^2\operatorname{tr}^2(\mathbf{F}^{-1}\mathbf{C}_1)\right) - \frac{1}{2}|\mathbf{F}|^{-3/2}4|\mathbf{F}|\left[\operatorname{tr}^2(\mathbf{F}^{-1}\mathbf{C}_1) - \operatorname{tr}(\mathbf{F}^{-1}\mathbf{C}_1)^2\right] \\
&= |\mathbf{F}|^{-1/2}(\operatorname{tr}^2(\mathbf{F}^{-1}\mathbf{C}_1) + 2\operatorname{tr}(\mathbf{F}^{-1}\mathbf{C}_1)^2).
\end{aligned}
$$

Thus,

$$
\left.\frac{\partial^2}{\partial t_1^2}\phi(t_1,t_2)\right|_{t_1=0,t_2=-t} = \prod_{i=1}^n(1+2\lambda_i t)^{-1/2}\sum_{i=1}^n\sum_{j=1}^n\frac{c_{1,ii}c_{1,jj}+2c_{1,ij}^2}{(1+2\lambda_i t)(1+2\lambda_j t)}.
$$

Next, using the Formula (3.7) with $p=q=2$, we obtain

$$
\begin{aligned}
\mathbb{E}_X\left[\operatorname{Bias}_Y^2\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] = \mathcal{B}_{y\cdot z(x)}^2 &\sum_{i=1}^n\sum_{j=1}^n(c_{1,ii}c_{1,jj}+2c_{1,ij}^2)\cdot \\
&\cdot\int_0^\infty t\prod_{i=1}^n(1+2\lambda_i t)^{-1/2}\frac{1}{(1+2\lambda_i t)(1+2\lambda_j t)}\,dt.
\end{aligned} \tag{3.15}
$$

Finally, from Equation (3.13) we get the expression for the marginal variance of the estimator $\hat{\beta}_x$ as:

$$
\begin{aligned}
\mathbb{V}_{Y,X}\left[\hat{\beta}_x\right] = &\sum_{j=1}^n c_{2,jj}\int_0^\infty t\prod_{i=1}^n(1+2\lambda_i t)^{-1/2}\frac{1}{1+2\lambda_j t}\,dt + \\
&+\mathcal{B}_{y\cdot z(x)}^2\sum_{i=1}^n\sum_{j=1}^n(c_{1,ii}c_{1,jj}+2c_{1,ij}^2)\cdot \\
&\cdot\int_0^\infty t\prod_{i=1}^n(1+2\lambda_i t)^{-1/2}\frac{1}{(1+2\lambda_i t)(1+2\lambda_j t)}\,dt + \\
&-\left(\mathcal{B}_{y\cdot z(x)}\sum_{j=1}^n c_{1,jj}\int_0^\infty\prod_{i=1}^n(1+2\lambda_i t)^{-1/2}\frac{1}{1+2\lambda_j t}\,dt\right)^2.
\end{aligned} \tag{3.16}
$$

The next statement formalizes what has been exposed so far regarding the expected value and variance of the estimator $\hat{\beta}_x$ expressed in terms of integrals.The mean square error (MSE) is also obtained.

Theorem 3.2.1. *The estimator $\hat{\beta}_x$ defined in (2.28) has expected value and variance provided in (3.11) and (3.16), respectively. Additionally, its mean square error may be* *expressed as:*

$$MSE_{X,Y}\left[\hat{\beta}_x\right] = \mathbb{E}_X\left[\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] + \mathbb{E}_X\left[\text{Bias}^2_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] \tag{3.17}$$

Proof. The main results are derived along the section. Equation (3.17) is achieved following the definition of mean square error and (3.12) as follows:

$$\begin{aligned}
\text{MSE}_{X,Y}\left[\hat{\beta}_x\right] &= \mathbb{E}_X\left[\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] + \mathbb{E}_X\left[\text{Bias}^2_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] \\
&= \mathbb{V}_{X,Y}\left[\hat{\beta}_x\right] - \mathbb{V}_X\left[\mathbb{E}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] + \mathbb{E}_X\left[\text{Bias}^2_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] \\
&= \mathbb{V}_{X,Y}\left[\hat{\beta}_x\right] - \mathbb{V}_X\left[\text{Bias}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] + \mathbb{E}_X\left[\text{Bias}^2_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] \\
&= \mathbb{V}_{X,Y}\left[\hat{\beta}_x\right] + \text{Bias}^2_{X,Y}\left[\hat{\beta}_x\right]
\end{aligned}$$

□

Results obtained so far on the marginal properties of $\hat{\beta}_x$ need further explorations in order to deliver customary computational tools for evaluating the involved integrals. To this aim, we find it convenient to represent them in terms of an hypergeometric function: the Carlson's R function, $R(a;\boldsymbol{b},\boldsymbol{z})$ (see Carlson (1963) for more details). It is a natural outcome of a procedure for generalizing the Gauss hypergeometric function. Moreover, on m complex variables $z_1,\ldots,z_m$ and $m+1$ complex parameters $a,b_1,\ldots,b_m$ it is the same as Lauricella's F_D (Lauricella, 1893) except for small important modification. The Carlson's R function is defined as follows

$$R(a;b_1,\ldots,b_m;z_1,\ldots,z_m) = F_D(a;b_1,\ldots,b_m;b_1+\ldots b_m;1-z_1,\ldots,1-z_m).$$

with $||1-z_i||<1$, for $i=1,\ldots,m$. The R function is often used to make unified statement of a property of several integrals (Olver et al., 2010). Its extensive use is justified by two distinctive properties, symmetry and homogeneity. The former means that it is invariant under permutation of the subscript $1,\ldots,m$ and the latter implies:

$$R(a;b_1,\ldots,b_m;sz_1,\ldots,sz_m) = s^{-a}R(a;b_1,\ldots,b_m;z_1,\ldots,z_m).$$

For our purpose we deem the particular case in which

$$\begin{aligned}
\int_0^\infty t^{a-1}\prod_{i=1}^m(1+sz_it)^{-b_i}\,dt &= s^{-a}B(a,a')R(a;b_1,\ldots,b_m;z_1,\ldots,z_m) \\
&= B(a,a')R(a;\boldsymbol{b},s\boldsymbol{z}),
\end{aligned} \tag{3.18}$$

where $B(a, a') = \Gamma(a)\Gamma(a')/\Gamma(a+a')$ is the beta function, $\Gamma(\cdot)$ is the gamma function and a' is defined by

$$a + a' = b = \sum_{i=1}^{m} b_i \in \mathbb{Q} \setminus \{0\}.$$

Let consider the n-dimensional vector of eigenvalues $\boldsymbol{\lambda}$ defined in (3.9). Assuming $s\boldsymbol{z} = 2\boldsymbol{\lambda}$, $b_i = \frac{1}{2}\ \forall\, i = 1, \ldots, m$, and posing:

$$\begin{aligned} a &= q \\ a' &= \frac{n}{2} + p - q \\ m &= n + 2p, \end{aligned}$$

the right-hand side of Equation (3.18) can be re-written as follows:

$$\begin{aligned} \int_0^\infty t^{q-1} \prod_{i=1}^{n+2p} (1 + 2\lambda_i t)^{-1/2}\, dt &= B\left(q, \frac{n}{2} + p - q\right) R\left(q; \frac{1}{2}\mathbf{1}_{n+2p}, 2\boldsymbol{\lambda}\right) \\ &= I^{p,q}(\boldsymbol{\lambda}), \end{aligned} \tag{3.19}$$

where $I^{p,q}(\boldsymbol{\lambda})$ denotes the integral characterized by the powers of the QFs' ratio, p and q, and the n-dimensional vector of denominator matrix eigenvalues.

Carlson (1963) states that the R function reduces to another function of the same type with one less variable if one of its variables z_i vanishes. Because of the symmetry property of R and if $Re(a' - b_{-1:k}) > 0$ with $b_{-1:k} = b_{k+1} + \cdots + b_m$ we have:

$$B(a, a')R(a; b_1, \ldots, b_m; z_1, \ldots, z_k, \mathbf{0}_{b_{-1:k}}) = B(a, a' - b_{-1:k})R(a; b_1, \ldots, b_k; z_1, \ldots, z_k).$$

In this way, defining $\boldsymbol{\lambda}_+ = \operatorname{diag}(\boldsymbol{\Lambda}_+)$, the vector of positive eigenvalues, the Formula in (3.19) may be express as:

$$\begin{aligned} \int_0^\infty t^{q-1} \prod_{i=1}^{n+2p} (1 + 2\lambda_i t)^{-1/2}\, dt &= B\left(q, \frac{n}{2} + p - q - \frac{1}{2}h\right) R\left(q; \frac{1}{2}\mathbf{1}_{n+2p-h}, 2\boldsymbol{\lambda}_+\right) \\ &= I_h^{p,q}(\boldsymbol{\lambda}), \end{aligned}$$

where h is the number of zero eigenvalues in $\boldsymbol{\lambda}$.

Based on this, we can use the Carlson's R function for the evaluation of integrals in (3.11), (3.14), (3.15) in order to use efficient available algorithms to compute the marginal sampling properties of estimator $\hat{\beta}_x$.

Regarding Equation (3.11) in which $p = q = 1$ and exploiting the symmetric property of R, for all $j = 1, \dots, n$ we define

$$\mathbf{\Lambda}_j^{'} = \text{diag}\left(\lambda_1, \dots, \lambda_n, \lambda_j, \lambda_j\right) = \text{diag}\left(\boldsymbol{\lambda}_j^{'}\right)$$

and observe that

$$\begin{aligned}\int_0^\infty \prod_{i=1}^n (1+2\lambda_i t)^{-1/2} \frac{1}{1+2\lambda_j t} dt &= B\left(1, \frac{n}{2} - \frac{1}{2}h_j\right) R_j\left(1; \frac{1}{2}\mathbf{1}_{n+2p-h_j}, 2\boldsymbol{\lambda}_{j,+}^{'}\right) \\ &= I_{h_j}^{1,1}\left(\boldsymbol{\lambda}^{'}\right),\end{aligned} \tag{3.20}$$

where $h_j = 1$ for $j = 1, \dots, n-1$ and $h_n = 3$. Similarly, referring to Equation (3.14) in which $p = 1$, $q = 2$, and for all $j = 1, \dots, n$ we get:

$$\begin{aligned}\int_0^\infty t \prod_{i=1}^n (1+2\lambda_i t)^{-1/2} \frac{1}{1+2\lambda_j t}\, dt &= B\left(2, \frac{n}{2} - 1 - \frac{1}{2}h_j\right) R_j\left(2; \frac{1}{2}\mathbf{1}_{n+2p-h_j}, 2\boldsymbol{\lambda}_{j,+}^{'}\right) \\ &= I_{h_j}^{1,2}\left(\boldsymbol{\lambda}^{'}\right),\end{aligned}$$

Finally, concerning Equation (3.15) in which $p = q = 2$ and, for all $i, j = 1, \dots, n$ we define the following diagonal matrix

$$\mathbf{\Lambda}_{ij}^{''} = \text{diag}\left(\lambda_1, \dots, \lambda_n, \lambda_j, \lambda_j, \lambda_i, \lambda_i\right) = \text{diag}\left(\boldsymbol{\lambda}_{ij}^{''}\right)$$

and observe that

$$\begin{aligned}\int_0^\infty t \prod_{i=1}^n \frac{(1+2\lambda_i t)^{-1/2}}{(1+2\lambda_i t)(1+2\lambda_j t)}\, dt &= B\left(2, \frac{n}{2} - \frac{1}{2}h_{ij}\right) R_{ij}\left(2; \frac{1}{2}\mathbf{1}_{n+2p-h_{ij}}, 2\boldsymbol{\lambda}_{ij,+}^{''}\right) \\ &= I_{h_{ij}}^{2,2}\left(\boldsymbol{\lambda}^{''}\right)\end{aligned}$$

where

$$h_{ij} = \begin{cases} 1 & i, j = 1, \dots, n-1, \\ 3 & i = n,\ j = 1, \dots, n-1 \text{ and } j = n,\ i = 1, \dots, n-1, \\ 5 & i = j = n. \end{cases} \tag{3.21}$$

The next statement collects and formalizes the above analytical results enabling to compute the marginal sampling properties of estimator $\hat{\beta}_x$ with no use of simulation study. In practice, the computation of these quantities exploits a function from the `R`-package `QF` (Gardini et al., 2022).

Theorem 3.2.2. *The expected value and variance of the estimator defined in* (2.28) *may* *be expressed in terms of Carlson's R function as follows:*

$$\mathbb{E}_{Y,X}\left[\hat{\beta}_x\right] = \mathcal{B}_{y\cdot x(z)} + \mathcal{B}_{y\cdot z(x)} \sum_{j=1}^{n} c_{1,jj} I^{1,1}_{h_j}\left(\boldsymbol{\lambda}^{'}\right) \tag{3.22}$$

$$\begin{aligned}\mathbb{V}_{Y,X}\left[\hat{\beta}_x\right] &= \sum_{j=1}^{n} c_{2,jj} I^{1,2}_{h_j}\left(\boldsymbol{\lambda}^{'}\right) + \mathcal{B}^2_{y\cdot z(x)} \sum_{i=1}^{n}\sum_{j=1}^{n}(c_{1,ii}c_{1,jj} + 2c^2_{1,ij}) I^{2,2}_{h_{ij}}\left(\boldsymbol{\lambda}^{''}\right) + \\ &\quad - \left(\mathcal{B}_{y\cdot z(x)} \sum_{j=1}^{n} c_{1,jj} I^{1,1}_{h_j}\left(\boldsymbol{\lambda}^{'}\right)\right)^2 . \end{aligned} \tag{3.23}$$

where $h_j = 1$ *for* $j = 1, \ldots, n-1$ *and* $h_n = 3$ *and* h_{ij} *as in* (3.21).

Proof. The proof derives from the discussion along this section. □

Afterwards, we discuss the results provided in the current section.

3.3 Some remarks on the variance and mean square error

Unlike the bias, the conditional variance of $\hat{\beta}_x$ (3.2) depends on the structure of all processes included in the DGP. About it we highlight two results. Firstly, from Proposition 2.1.4, Equation (3.2) can be re-written as

$$\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right] = \underbrace{\underbrace{\frac{\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{\Sigma}_{y|x,z}\boldsymbol{\Delta}\boldsymbol{X}}{(\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{X})^2}}_{\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X},\mathcal{B}_{y\cdot z(x)}=0,\boldsymbol{A}_{z\cdot x}=\boldsymbol{0}\right]} + \mathcal{B}^2_{y\cdot z(x)}\frac{\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{\Sigma}_z\boldsymbol{\Delta}\boldsymbol{X}}{(\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{X})^2}}_{\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X},\boldsymbol{A}_{z\cdot x}=\boldsymbol{0}\right]} + \tag{3.24}$$

$$- \mathcal{B}^2_{y\cdot z(x)}\frac{\boldsymbol{X}^\top\boldsymbol{\Delta}\left(\boldsymbol{A}_{z\cdot x}\boldsymbol{\Sigma}_x\boldsymbol{A}^\top_{z\cdot x}\right)\boldsymbol{\Delta}\boldsymbol{X}}{(\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{X})^2}. \tag{3.25}$$

The first line (3.24) expresses the conditional variance when the covariate and the confounder are independent, that is indicated with $\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X},\boldsymbol{A}_{z\cdot x}=\boldsymbol{0}\right]$. It includes the conditional variance $\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X},\mathcal{B}_{y\cdot z(x)}=0,\boldsymbol{A}_{z\cdot x}=\boldsymbol{0}\right]$ when $\boldsymbol{X}\perp\boldsymbol{Z}$ and $\boldsymbol{Y}\perp\boldsymbol{Z}|\boldsymbol{X}$. Secondly, next theorem shows that, regardless the choice of relationship characterizing the joint distribution in (2.11), the conditional variance is always bigger compared to the case in which $\boldsymbol{X}$ and $\boldsymbol{Z}$ are independent.

Theorem 3.3.1. *The conditional variance of the estimator* $\hat{\beta}_x$ *assumes the maximum* *value when $\boldsymbol{X}$ and $\boldsymbol{Z}$ are independent. In other words:*

$$\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right] \geq \mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}, \boldsymbol{A}_{z\cdot x} = \boldsymbol{0}\right]. \tag{3.26}$$

Proof. From Equation (3.25), we have

$$\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right] = \mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}, \boldsymbol{A}_{z\cdot x} = \mathbf{0}\right] - \mathcal{B}^2_{y\cdot z(x)}\frac{\boldsymbol{X}^\top\boldsymbol{\Delta}\left(\boldsymbol{A}_{z\cdot x}\boldsymbol{\Sigma}_x\boldsymbol{A}^\top_{z\cdot x}\right)\boldsymbol{\Delta}\boldsymbol{X}}{(\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{X})^2}.$$

Both QFs' ratios on the right-side are positive valued random variables since $\boldsymbol{\Sigma}_{y|x,z}$, $\boldsymbol{\Sigma}_z$ and $\boldsymbol{\Sigma}_x$ are positive definite (see (A.9)). Hence, $\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right] - \mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}, \boldsymbol{A}_{z\cdot x} = \mathbf{0}\right] \geq 0$. □

This result implies that the conditional variance under the assumption of dependence between $\boldsymbol{X}$ and $\boldsymbol{Z}$, i.e. $\mathbf{A}_{x\cdot z} \neq \mathbf{0}$, is always inflated. Theorem 3.3.1 shows that greater strength of dependence between $\boldsymbol{X}$ and $\boldsymbol{Z}$ leads to less variance. A similar result has been already shown from Paciorek (2010) and Page et al. (2017) for some specific type of dependence structure for the joint distribution of $(\boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top$. Therefore, the crucial aspect of Theorems 3.0.1 and 3.3.1 is that the absence of confounding carries out the maximum value for the variance and the minimum absolute bias. As the dependences characterizing the confounding expressed via $\mathbf{A}_{x\cdot z}$ and $\mathcal{B}_{y\cdot z(x)}$ increases, the variance decreases and the bias increases. These considerations made on conditionals sampling properties remain valid marginally.

Regarding the marginal variance, Equation (3.12) can be re-written as

$$\begin{aligned}\mathbb{V}_{Y,X}\left[\hat{\beta}_x\right] = &\underbrace{\mathbb{E}_X\left[\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}, \boldsymbol{A}_{z\cdot x} = \mathbf{0}\right]\right]}_{\mathbb{V}_{Y,X}\left[\hat{\beta}_x|\boldsymbol{A}_{z\cdot x}=\mathbf{0}\right]} + \\ &- \mathcal{B}^2_{y\cdot z(x)}\mathbb{E}_X\left[\frac{\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{A}_{z\cdot x}\boldsymbol{\Sigma}_x\boldsymbol{A}^\top_{z\cdot x}\boldsymbol{\Delta}\boldsymbol{X}}{(\boldsymbol{X}^\top\boldsymbol{\Delta}\boldsymbol{X})^2}\right] + \mathbb{E}_X\left[\mathrm{Bias}^2_Y\left[\hat{\beta}_x|\boldsymbol{X}\right]\right] - \mathrm{Bias}^2_{Y,X}\left[\hat{\beta}_x\right].\end{aligned}$$

Hence, the part of estimator variance that changes when confounding occurs is

$$\mathbb{V}^{cd}_{Y,X}\left[\hat{\beta}_x\right] = \mathbb{V}_{Y,X}\left[\hat{\beta}_x\right] - \mathbb{V}_{Y,X}\left[\hat{\beta}_x|\boldsymbol{A}_{z\cdot x} = \mathbf{0}\right], \tag{3.27}$$

where the superscript *cd* stays for confounding-dependent. The same rationale applies to the the marginal mean square error: the confounding-dependent part is

$$\mathrm{MSE}^{cd}_{Y,X}\left[\hat{\beta}_x\right] = \mathrm{MSE}_{Y,X}\left[\hat{\beta}_x\right] - \mathbb{V}_{Y,X}\left[\hat{\beta}_x|\boldsymbol{A}_{z\cdot x} = \mathbf{0}\right]$$

$$= \mathbb{V}_{Y,X}\left[\hat{\beta}_x\right] - \mathbb{V}_{Y,X}\left[\hat{\beta}_x | \boldsymbol{A}_{z\cdot x} = \mathbf{0}\right] + \text{Bias}^2_{Y,X}\left[\hat{\beta}_x\right]. \tag{3.28}$$

Next result provides a condition on how structure of $\mathbf{S}$ need to be in order to minimize the estimator variance. In fact, in the following statement we show that the estimator conditionally on $\boldsymbol{X}$ has a lowest sampling variance if the errors in linear regression model have the structure of the covariance matrix of the marginalized DGP in (2.5), $\boldsymbol{\Sigma}_{y|x}$.

Theorem 3.3.2. *Assuming a generic non-singular error covariance matrix $\boldsymbol{S}$ for the estimator in (2.28) such that $\boldsymbol{S} = \boldsymbol{\Sigma}_{y|x}$, the estimator's conditional variance is minimal.*

Proof. Let us consider a generic non-singular covariance matrix as a sum of two matrices, $\mathbf{S} = \boldsymbol{\Sigma}_{y|x} + \mathbf{C}$, $\mathbf{C} \in \mathcal{S}^n$. If we suppose that the covariance matrix in the estimator is $\boldsymbol{\Sigma}_{y|x}$, i.e. $\mathbf{C}$ is null matrix, then we can define the estimator $\boldsymbol{J}_0\boldsymbol{Y}$, otherwise, from Equation (2.28) we specify it by $\boldsymbol{J}_C\boldsymbol{Y}$.

Below we express the matrix $\mathbf{J}_C$ in dependence of $\mathbf{J}_0$. The following computations use matrix inversion rules reported in (A.1) and (A.2):

$$\begin{aligned}
\boldsymbol{J}_C &= \left[\tilde{\mathbf{X}}^\top \left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1} \tilde{\mathbf{X}}\right]^{-1} \tilde{\mathbf{X}}^\top \left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1} \\
&= \left[\tilde{\mathbf{X}}^\top \left(\boldsymbol{\Sigma}^{-1}_{y|x} - \boldsymbol{\Sigma}^{-1}_{y|x}\mathbf{C}\left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1}\right) \tilde{\mathbf{X}}\right]^{-1} \tilde{\mathbf{X}}^\top \left(\boldsymbol{\Sigma}^{-1}_{y|x} - \boldsymbol{\Sigma}^{-1}_{y|x}\mathbf{C}\left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1}\right) \\
&= \left(\tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x}\tilde{\mathbf{X}} - \tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x}\mathbf{C}\left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1}\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x} + \\
&\qquad - \left(\tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x}\tilde{\mathbf{X}} - \tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x}\mathbf{C}\left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1}\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x}\mathbf{C}\left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1} \\
&= \boldsymbol{J}_0 + \left(\tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x}\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x}\mathbf{C}\left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1}\tilde{\mathbf{X}}\left[\tilde{\mathbf{X}}^\top \left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1}\tilde{\mathbf{X}}\right]^{-1} \cdot \\
&\qquad \tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x} - \left[\tilde{\mathbf{X}}^\top \left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1}\tilde{\mathbf{X}}\right]^{-1} \tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x}\mathbf{C}\left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1} \\
&= \boldsymbol{J}_0 + \left(\tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x}\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^\top \left[\boldsymbol{\Sigma}^{-1}_{y|x} - \left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1}\right]\tilde{\mathbf{X}}\left[\tilde{\mathbf{X}}^\top \left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1}\tilde{\mathbf{X}}\right]^{-1} \cdot \\
&\qquad \cdot \tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x} - \left[\tilde{\mathbf{X}}^\top \left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1}\tilde{\mathbf{X}}\right]^{-1} \tilde{\mathbf{X}}^\top \left[\boldsymbol{\Sigma}^{-1}_{y|x} - \left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1}\right] \\
&= \boldsymbol{J}_0 + \left[\tilde{\mathbf{X}}^\top \left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1}\tilde{\mathbf{X}}\right]^{-1} \tilde{\mathbf{X}}^\top \left(\boldsymbol{\Sigma}_{y|x} + \mathbf{C}\right)^{-1} - \left(\tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x}\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x}.
\end{aligned}$$

Then, we get $\boldsymbol{J}_C = \boldsymbol{J}_0 + \boldsymbol{D}$ such that $\boldsymbol{D}\boldsymbol{X} = \mathbf{0}$ and $\text{Bias}_Y[\boldsymbol{J}_C\boldsymbol{Y}|\boldsymbol{X}] = \mathcal{B}_{y\cdot z(x)}(\boldsymbol{J}_0\boldsymbol{A}_{z\cdot x}\boldsymbol{X} + \boldsymbol{D}\boldsymbol{A}_{z\cdot x}\boldsymbol{X})$.

The covariance matrix of the estimator $\boldsymbol{J}_C\boldsymbol{Y}$ is

$$\mathbb{V}_Y[\boldsymbol{J}_C\boldsymbol{Y}|\boldsymbol{X}] = \left[\left(\tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x}\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x} + \boldsymbol{D}\right] \boldsymbol{\Sigma}_{y|x} \left[\left(\tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x}\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}^{-1}_{y|x} + \boldsymbol{D}\right]^\top$$

$$
\begin{aligned}
&= \left(\tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}_{y|x}^{-1} \tilde{\mathbf{X}}\right)^{-1} + \left(\tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}_{y|x}^{-1} \tilde{\mathbf{X}}\right)^{-1} (\boldsymbol{D}\boldsymbol{X})^\top + \boldsymbol{D}\boldsymbol{X} \left(\tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}_{y|x}^{-1} \tilde{\mathbf{X}}\right)^{-1} + \boldsymbol{D}\boldsymbol{\Sigma}_{y|x}\boldsymbol{D}^\top \\
&= \left(\tilde{\mathbf{X}}^\top \boldsymbol{\Sigma}_{y|x}^{-1} \tilde{\mathbf{X}}\right)^{-1} + \boldsymbol{D}\boldsymbol{\Sigma}_{y|x}\boldsymbol{D}^\top
\end{aligned}
$$

Since $\boldsymbol{D}\boldsymbol{\Sigma}_{y|x}\boldsymbol{D}^\top$ is positive semidefinite, the variance of $\hat{\beta}_x$ is minimal if $\mathbf{C} = \mathbf{0}$, and consequently $\boldsymbol{D} = \mathbf{0}$. □

Corollary 3.3.3. *If $\boldsymbol{S} = \boldsymbol{\Sigma}_{y|x}$,*

$$
\mathbb{E}_X \left[\mathbb{V}_Y \left[\hat{\beta}_x | \boldsymbol{X}\right]\right] = I_1^{0,1}(\boldsymbol{\lambda}). \tag{3.29}
$$

Proof. If $\mathbf{S} = \boldsymbol{\Sigma}_{y|x}$, following Provost and Mathai (1992), one has:

$$
\begin{aligned}
\mathbb{E}_X \left[\frac{\boldsymbol{X}^\top \boldsymbol{\Delta}\boldsymbol{\Sigma}_{y|x}\boldsymbol{\Delta}\boldsymbol{X}}{(\boldsymbol{X}^\top \boldsymbol{\Delta}\boldsymbol{X})^2}\right] &= \mathbb{E}_X \left[\frac{\boldsymbol{X}^\top \mathbf{S}^{-1}\tilde{\boldsymbol{\Delta}}\boldsymbol{\Sigma}_{y|x}\mathbf{S}^{-1}\tilde{\boldsymbol{\Delta}}\boldsymbol{X}}{(\boldsymbol{X}^\top \boldsymbol{\Delta}\boldsymbol{X})^2}\right] \\
&= \mathbb{E}_X[(\boldsymbol{X}^\top \boldsymbol{\Delta}\boldsymbol{X})^{-1}] = \int_0^\infty \prod_{i=1}^n (1 + 2\lambda_i t)^{-1/2}\, dt.
\end{aligned}
$$

Thus, from plugging $p = 0$ and $q = 1$ in (3.19), we obtain the result

$$
\begin{aligned}
\mathbb{E}_X \left[\mathbb{V}_Y \left[\hat{\beta}_x | \boldsymbol{X}\right]\right] &= B\left(1, \frac{n-3}{2}\right) R\left(1; \frac{1}{2}\mathbf{1}_{n-1}, 2\lambda_1, \dots, 2\lambda_{n-1}\right) \\
&= I_1^{0,1}(\boldsymbol{\lambda}).
\end{aligned}
$$

□

3.4 Marginal distribution of $\hat{\beta}_x$ under spherical DGP

In the next chapter we seek to describe confounding under several choices of data generating process focusing on the case in which all the covariance matrices of the random processes are structured, but now we examine the basic case of an unstructured data generating process.

Considering unstructured DGP means assuming $\boldsymbol{\Sigma}_{y|x,z} = \sigma^2_{y|x,z}\mathbf{I}_n$ and

$$
\boldsymbol{\Sigma}_{x,z} = \begin{pmatrix} \sigma_x^2 \mathbf{I}_n & \rho_{xz}\sigma_x\sigma_z\mathbf{I}_n \\ \rho_{xz}\sigma_x\sigma_z\mathbf{I}_n & \sigma_z^2\mathbf{I}_n \end{pmatrix}, \tag{3.30}
$$

where σ_x^2 and σ_z^2 are the marginal variances of the spherical processes $\boldsymbol{X}$ and $\boldsymbol{Z}$, respectively, and ρ_{xz} (along this section also simply denoted ρ) is the correlation parameter used to quantify the strength of association between the covariate and confounder.

In line with Definition 2.1.1, if both $\mathcal{B}_{y\cdot z(x)}$ and ρ_{xz} are not null the regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$ is confounded by $\boldsymbol{Z}$. For any choice of $\mathbf{S}$, the variance of $\hat{\beta}_x|\boldsymbol{X}$ is always stochastic, instead the bias is deterministic, since it does not depend on $\boldsymbol{X}$. The next result follows immediately from Lemma 3.0.1.

$$
\begin{aligned}
\operatorname{Bias}_{Y,X}\left[\hat{\beta}_x\right] &= \operatorname{Bias}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right] \\
&= \mathcal{B}_{y\cdot z(x)}\mathcal{B}_{z\cdot x} \\
&= \mathcal{B}_{y\cdot z(x)}\frac{\sigma_{zx}}{\sigma_x^2} \\
&= \mathcal{B}_{y\cdot z(x)}\rho_{xz}\frac{\sigma_z}{\sigma_x}.
\end{aligned} \tag{3.31}
$$

As expected, there is no bias when either ρ_{xz} or $\mathcal{B}_{y\cdot z(x)}$ are null. From the previous formula and as illustrated in Figure 3.1, it is evident that the correlation and variability of the covariate and confounder contribute to the intensity of the bias. Such estimator bias rises as the correlation parameter or the confounder marginal variance increases. Small σ_z and ρ_{xz} means that the confounder slightly disturbs the relationship between $\boldsymbol{Y}$ and $\boldsymbol{X}$. Here, the estimator expected value is simply the regression coefficient, $\mathcal{B}_{y\cdot x}$, of $\boldsymbol{Y}$ on $\boldsymbol{X}$. Equation (3.31) is a baseline well-known result found also by Paciorek (2010) and Page et al. (2017) but for different setting. In order to obtain (3.31), the authors do not consider an unstructured DGP but assume specific choices of cross-covariance and covariance function that lead to a scalar $\mathbf{A}_{z\cdot x}$. In both works, (3.31) is derived assuming the same covariance function and structure parameters for the processes.

According to the bias formula in Theorem 3.0.1, it is clear that such simplification of the regression matrix is the key point that enables the deterministic bias in (3.31). However, as it will be shown later, even assuming a structured data generating mechanism may get (3.31). In the remaining cases, when such simplification does not occur, it is not immediate to measure the sources of confounding. Certainly the kind of marginal and cross structure that binds the two covariates plays a principal role.

In this trivial setting, assuming an OLS estimator $\hat{\beta}_x$, i.e. taking $\mathbf{S} = \mathbf{I}_n$, or more in general a scalar matrix $\mathbf{S}$, we provide the marginal sampling distribution of estimator $\hat{\beta}_x$ in closed-form. It follows a non-centered and scaled Student-t obtained as a normal-inverse gamma mixture.

Theorem 3.4.1. *Under an unstructured data generating process as in* (3.30)*, the marginal*

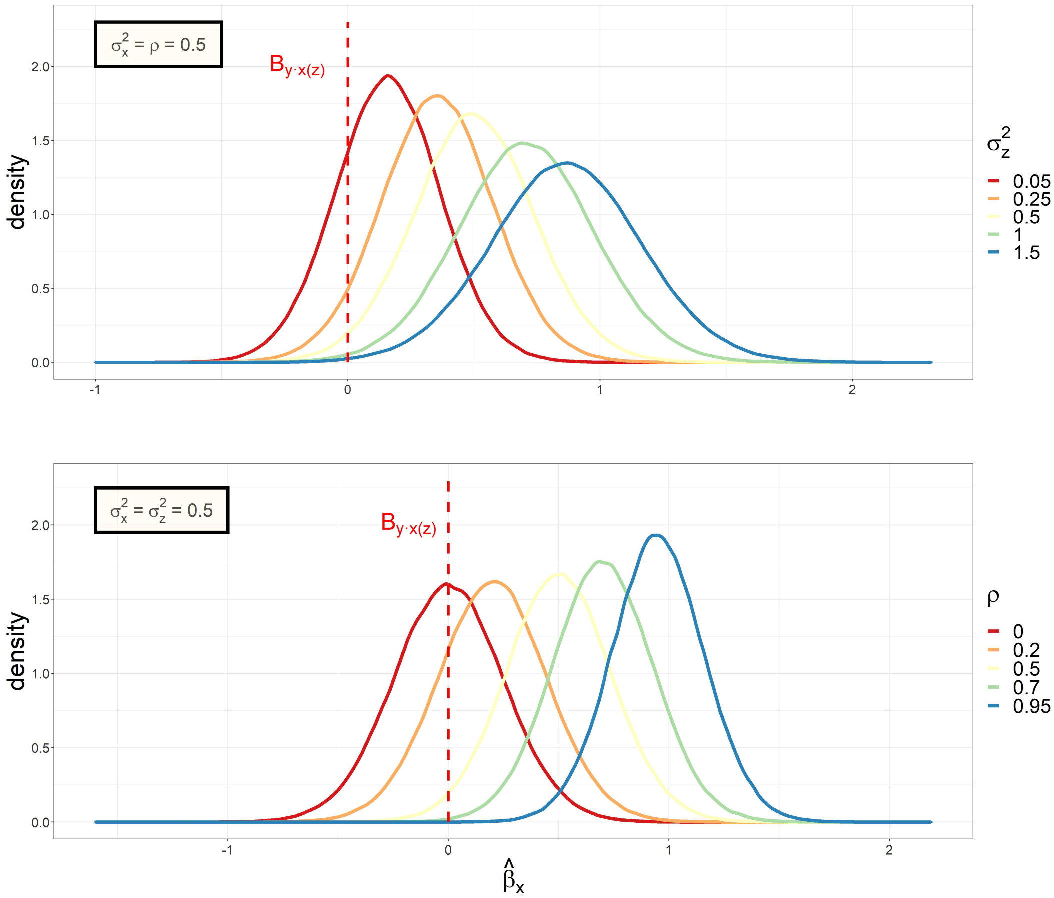

Figure 3.1: Under the model assumption $\mathbf{S} = \mathbf{\Sigma}_{y|x}$, different behaviour of sampling distributions of estimator bias as the correlation parameter (bottom) and confounder marginal variance parameter (top) vary.

sampling distribution of the OLS estimator is

$$\hat{\beta}_x \sim t_{n-1}\left(\mathcal{B}_{y\cdot x(z)} + \mathcal{B}_{y\cdot z(x)}\rho_{xz}\frac{\sigma_z}{\sigma_x},\ \frac{\sigma^2_{y|x}}{(n-1)\sigma^2_x}\right). \tag{3.32}$$

Proof. Under spherical DGP, from Proposition 2.1.4 one gets

$$\begin{aligned}\mathbf{\Sigma}_{y|x} &= \mathbf{\Sigma}_{y|x,z} + \mathbf{A}_{y\cdot z(x)}\mathbf{\Sigma}_{z|x}\mathbf{A}^{\top}_{y\cdot z(x)} \\ &= \left[\sigma^2_{y|x,z} + \mathcal{B}^2_{y\cdot z(x)}(1-\rho^2_{xz})\sigma^2_z\right]\mathbf{I}_n = \sigma^2_{y|x}\mathbf{I}_n,\end{aligned} \tag{3.33}$$

since $\mathbf{\Sigma}_{z|x} = (1-\rho^2_{xz})\sigma^2_z\mathbf{I}_n$ and $\mathbf{A}_{y\cdot z(x)} = \mathcal{B}_{y\cdot z(x)}\mathbf{I}_n$. Moreover, the weighted centering matrix $\mathbf{\Delta}$ coincides with the centering matrix $\mathbf{M}$. This leads to the following estimator's conditional sampling variance

$$\begin{aligned}\mathbb{V}_Y[\hat{\beta}_x|\boldsymbol{X}] &= \left(\boldsymbol{X}^\top \mathbf{M}\boldsymbol{X}\right)^{-1} \\ &= \sigma^2_{y|x} Q_M^{-1}(\boldsymbol{X}) \\ &= \frac{\sigma^2_{y|x}}{\sigma^2_x} Q_M^{-1}(\boldsymbol{\nu}),\end{aligned}$$

where $\boldsymbol{\nu}$ is distributed as a standard multivariate normal. As introduced in Section 3.1, $Q_M(\boldsymbol{\nu})$ follows a Chi-square distribution with $n-1$ degree of freedom. Hence, for the properties of Gamma distributions, $W = \mathbb{V}_Y[\hat{\beta}_x|\boldsymbol{X}]$ is distributed as an Inverse Gamma with shape $\alpha = (n-1)/2$ and scale $\beta = \sigma^2_{y|x}/(2\sigma^2_x)$.

Defining

$$m = \mathbb{E}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right] = \mathbb{E}_{X,Y}\left[\hat{\beta}_x\right] = \mathcal{B}_{y\cdot x(z)} + \mathcal{B}_{y\cdot z(x)}\rho_{xz}\frac{\sigma_z}{\sigma_x},$$

and having

$$\begin{aligned}\hat{\beta}_x|W &\sim \mathcal{N}(m, W), \\ W &\sim \mathcal{IG}(\alpha, \beta),\end{aligned}$$

the marginal distribution of $\hat{\beta}_x$ is obtained as

$$\begin{aligned}f_{\hat{\beta}_x}(b) &= \int_{\mathcal{W}} f_{\hat{\beta}_x|W}(b) f_W(w)\, dw \\ &= \int_{\mathcal{W}} \frac{1}{\sqrt{2\pi}w^{\frac{1}{2}}} \exp\left(-\frac{(b-m)^2}{2w}\right) \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \exp\left(-\frac{\beta}{w}\right) w^{-(\alpha+1)}\, dw \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{1}{\sqrt{2\pi}} \int_{\mathcal{W}} \exp\left(-\frac{(b-m)^2/2+\beta}{w}\right) w^{-\left(\alpha+\frac{1}{2}+1\right)}\, dw \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{1}{\sqrt{2\pi}} \frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{[(b-m)^2/2+\beta]^{\alpha+\frac{1}{2}}}.\end{aligned}$$

Recalling that the Student-t distribution with $k > 0$ degree of freedom, location parameter l and scale parameter s of a generic random variable X has the following density

$$f_X(x;\, k, l, s) = \frac{1}{\sqrt{ks}\, B(\frac{k}{2}, \frac{1}{2})} \left(1 + \frac{(x-l)^2}{ks}\right)^{-\frac{k+1}{2}},$$

the above marginal density can be expressed as

$$f_{\hat{\beta}_x}(b) = \frac{1}{\sqrt{\pi}}\frac{1}{\sqrt{2\beta}}\frac{\Gamma\left(\alpha+\frac{1}{2}\right)}{\Gamma(\alpha)}\left[1+\frac{(b-m)^2}{2\beta}\right]^{-\alpha-\frac{1}{2}}.$$

This is a Student-t distribution with $k = 2\alpha = n-1$, $l = m$ and $s = \frac{\beta}{\alpha} = \frac{\sigma^2_{y|x}}{(n-1)\sigma^2_x}$. $\square$

Summarizing the main points of this section are that, under a spherical DGP:

- the bias is deterministic (i.e. it does not depend on $\boldsymbol{X}$), and it is an increasing function of $\mathcal{B}_{y\cdot z(x)}$ and the covariance σ_{xz}: it is important to note that the variance of the confounder σ^2_z and the correlation parameter ρ_{xz} contribute to the covariance through their product, so that their effect cannot be identified separately. In our opinion, a customary procedure aimed at understanding the marginal sampling properties of β_x consists in keeping σ^2_x and σ^2_z fixed and letting σ_{xz} vary in the interval $(-\sigma_x\sigma_z; \sigma_x\sigma_z)$: by doing this, we are actually varying the strength of the correlation between $\boldsymbol{X}$ and $\boldsymbol{Z}$ in a controlled setting that clearly separate the marginal variances of the covariate and confounder from the relationship between them;
- the variance depends on $\boldsymbol{X}$ (i.e. it is a random variable) and its marginal expected value is an increasing function of $\mathcal{B}_{y\cdot z(x)}$, σ_z and a decreasing function of ρ_{xz}: thus, in this case, the effect of σ_z and ρ_{xz} can be identified separately.

All the considerations above are based on the interpretation of parameters σ^2_x and σ^2_z as marginal variances. This is a clear-cut interpretation of these parameters only under a spherical DGP. To see this, we introduce the random variable

$$V_x = \frac{\boldsymbol{X}^\top \mathbf{M}\boldsymbol{X}}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{3.34}$$

i.e. the *sampling variance* of $\boldsymbol{X}$. V_x is a quadratic form in Gaussian random variables whose expected value, that we dub the *expected sampling variance*, is

$$\begin{aligned} EV_x &= \mathbb{E}_X[V_x] \\ &= \frac{1}{n-1}\sum_{i=1}^{n}\lambda(\mathbf{M}\boldsymbol{\Sigma}_x)_i = \bar{\lambda}_{M\Sigma_x} \\ &= \sigma^2_x\bar{\lambda}_+(\mathbf{M}\mathbf{R}_x), \end{aligned} \tag{3.35}$$

where $\bar{\lambda}_+(\mathbf{M}\mathbf{R}_x)$ is the mean of positive eigenvalues of matrix $\mathbf{M}\mathbf{R}_x$. The matrix $\mathbf{R}_x$ reflects the covariance structure of the random vector $\boldsymbol{X}$, while the term σ^2_x acts as scaler of the marginal variability of the process. When $\mathbf{R}_x = \mathbf{I}_n$, i.e. when the process is spherical, $\bar{\lambda}_+(\mathbf{MR}_x) = 1$ and $\mathbb{E}_X[V_x] = \sigma_x^2$. When $\mathbf{R}_x \neq I_n$, $\mathbb{E}_X[V_x] \neq \sigma_x^2$: this must be taken into account when studying the effect of confounding under non-spherical (or structured) DGPs.

Moreover, the same logic holds for the covariance between two processes $\boldsymbol{X}$ and $\boldsymbol{Z}$, trough the expected value of their *sampling covariance*, that we call also co-variability, defined as:

$$
\begin{aligned}
EV_{xz} &= EV_{zx} \\
&= \mathbb{E}_{X,Z}[V_{xz}] = \mathbb{E}_{X,Z}\left[\frac{\boldsymbol{X}^\top \mathbf{M}\boldsymbol{Z}}{n-1}\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \lambda(\mathbf{M}\boldsymbol{\Sigma}_{xz})_i = \bar{\lambda}_{M\Sigma_{xz}} \\
&= \sigma_{xz}\bar{\lambda}_+(\mathbf{MR}_{xz}).
\end{aligned} \tag{3.36}
$$

This expected value equals σ_{xz} only when $\mathbf{R}_{xz} = \mathbf{I}_n$. These definitions allow us to note that the eigenvalues of a covariance matrix are decisive to take into account the variability of a processes in the evaluation of confounding.
In addition, the relevance of these QFs can be emphasized by noting that, in a linear regression model, the variability of the response variable $\boldsymbol{Y}$ is decomposed as follows:

$$
\begin{aligned}
EV_y &= \mathbb{E}_Y\left[\frac{\boldsymbol{Y}^\top \mathbf{M}\boldsymbol{Y}}{n-1}\right] \\
&= \mathbb{E}_Y\left[\frac{[\mathcal{B}_{y\cdot x(z)}\boldsymbol{X} + \mathcal{B}_{y\cdot z(x)}\boldsymbol{Z} + \boldsymbol{\varepsilon}_{y|x,z}]^\top \mathbf{M}[\mathcal{B}_{y\cdot x(z)}\boldsymbol{X} + \mathcal{B}_{y\cdot z(x)}\boldsymbol{Z} + \boldsymbol{\varepsilon}_{y|x,z}]}{n-1}\right] \\
&= \frac{1}{n-1}\Bigg(\mathcal{B}^2_{y\cdot x(z)}\mathbb{E}_X\left(\boldsymbol{X}^\top \mathbf{M}\boldsymbol{X}\right) + \mathcal{B}^2_{y\cdot z(x)}\mathbb{E}_X\left(\boldsymbol{Z}^\top \mathbf{M}\boldsymbol{Z}\right) + \\
&\qquad + 2\mathcal{B}_{y\cdot x(z)}\mathcal{B}_{y\cdot z(x)}\mathbb{E}_{X,Z}\left(\boldsymbol{X}^\top \mathbf{M}\boldsymbol{Z}\right)\Bigg) + \sigma^2_{y|x,z} \\
&= \mathcal{B}^2_{y\cdot x(z)}\bar{\lambda}_{M\Sigma_x} + \mathcal{B}^2_{y\cdot z(x)}\bar{\lambda}_{M\Sigma_z} + 2\mathcal{B}_{y\cdot x(z)}\mathcal{B}_{y\cdot z(x)}\bar{\lambda}_{M\Sigma_{xz}} + \sigma^2_{y|x,z}.
\end{aligned}
$$

In the next section, we outline some relevant features of structured DGPs in terms of covariance matrix eigenvalues, while Section 3.6 proposes some measures aimed at quantifying confounding.

3.5 Marginal variability and smoothness of a structured DGP

Another distinctive property of a random vector is its *smoothness*. It is important to remark that the relation between smoothness and marginal variability needs to be taken into account in order to avoid misleading conclusion about confounding. Smoothness of a random process depends upon the correlation structure of the covariance matrix. For this reason, we recall that the correlation matrix $\mathbf{C}_x$ of a process $\boldsymbol{X}$ is obtained from the covariance or the structure matrix as follows:

$$\begin{aligned}\mathbf{C}_x &= \operatorname{diag}(\boldsymbol{\Sigma}_x)^{-1/2}\,\boldsymbol{\Sigma}_x \operatorname{diag}(\boldsymbol{\Sigma}_x)^{-1/2} \\ &\propto \operatorname{diag}(\mathbf{R}_x)^{-1/2}\,\mathbf{R}_x \operatorname{diag}(\mathbf{R}_x)^{-1/2}.\end{aligned}$$

To inspect the smoothness of a random process $\boldsymbol{X}$, we define the random vector $\boldsymbol{s} = \operatorname{diag}(\boldsymbol{\Sigma}_x)^{-1/2}\boldsymbol{X}$ and the random variable

$$IS_x = \frac{\boldsymbol{s}^\top \mathbf{M}\boldsymbol{s}}{n-1}, \tag{3.37}$$

i.e. the *sampling inverse smoothness* of $\boldsymbol{X}$. It is a quadratic form in Gaussian random variables too and accordingly with the previous lines, we provide its expected value as the mean of the positive eigenvalues of the matrix $\mathbf{MC}_x$. That means

$$\downarrow \bar{\lambda}_{MC_x} \qquad \Longleftrightarrow \qquad \uparrow \text{smoothness}.$$

To illustrate how the marginal variability and the smoothness of a random process change with respect to the choice of the covariance function, we consider three representative type of functions generating the covariance matrices of the covariate and confounder processes in (2.3): in a spatial framework, we consider the *Matérn* covariance function, widely used in geostatistical data analysis and the *conditional autoregressive process*, widely used in areal data analysis. The *autoregressive process of order 1* is presented as a relevant example in time series analysis.

3.5.1 Matérn covariance function

For the spatial setup, we adopt a class of isotropic covariance functions, the Matérn family (MF) (Matérn, 1986; Guttorp and Gneiting, 2006), which specifies the covariance function for $\boldsymbol{\Sigma}$ as $\sigma^2 \cdot M_\nu(d; r)$, where $\sigma^2 > 0$ is the marginal variance and

$$M(\boldsymbol{\theta}) = M_\nu(d; r) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{2d\sqrt{\nu}}{r}\right)^\nu \mathcal{K}_\nu\left(\frac{2d\sqrt{\nu}}{r}\right) \tag{3.38}$$

is the spatial correlation at Euclidean distance d between two locations. In this case, by construction, the structure matrix corresponds to the correlation matrix, namely

$$\mathbf{\Sigma} = \sigma^2 \mathbf{R} = \sigma^2 \mathbf{C} = \sigma^2 M_\nu(d; r).$$

In addition, $\boldsymbol{\theta} = (\nu, r)^\top$ is the vector of structure parameters with $\nu > 0$ defining the order of differentiability of the function, r being the range parameter, and $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν. As a special case, if ν equals 0.5 the exponential function arises. We denote with d_0 the distance at which the correlation is negligible, conventionally taken as having dropped to only 0,05 (Banerjee et al., 2014). It is obtained setting $\exp(-d_0/r) = 0.05$, getting $d_0 \approx 3r$, since $\log(0.05) \approx -3$. Thus, this remark will be used in Chapter 4 for the choice of r's range on a unit square grid.

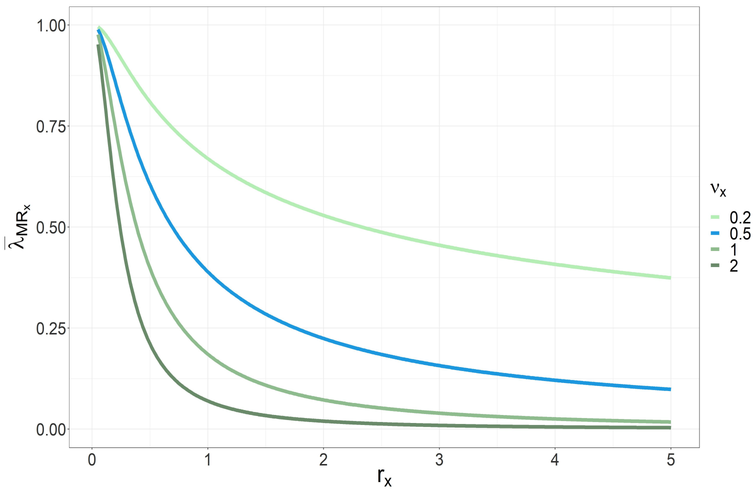

Figure 3.2: Decay of variability/inverse smoothness trend of $\boldsymbol{X}$ generated as Gaussian process with Matérn spatial correlation function $M_{\nu_x}(d_x, r_x)$, as a function of r_x.

Considering a random vector $\boldsymbol{X}$, a space in the form of regular square grid and $\nu \in \{0.2, 0.5, 1, 2\}$ in (3.38), Figure 3.2 shows that the range parameter r_x is inversely proportional to the mean of the positive eigenvalues of the matrix $\mathbf{MR}_x$. Hence, for different value of ν it is shown the decay (monotonic behaviour) of sampling variability as r_x increase. This means that the smoothness of the process grows with the range parameter.

The curves' monotonicity displayed in Figure 3.2 may be demonstrated by determining that the sign of the first derivative of the sampling variance function with respect to the range parameter remains unchanged. The following theorem states that in the case of exponential correlation function there is a decreasing monotony shown by the negativity of the first derivative over the entire domain of the range parameter.

Theorem 3.5.1. *Considering a n-dimensional Gaussian process generated with exponen-*

tial covariance function, the sampling variance of the process is strictly decreasing function of the range parameter r.

Proof. Let denote with $\boldsymbol{\Sigma}$ the covariance matrix characterizing the process. From some properties of the trace reported in Theorem A.0.1, we may observe that

$$\begin{aligned}
\frac{d\,\bar{\lambda}_{M\Sigma}(r)}{dr} &= \frac{1}{n-1}\frac{d\,\mathrm{tr}(\mathbf{M}\boldsymbol{\Sigma}(r))}{dr} \\
&= \frac{1}{n-1}\mathrm{tr}\left(\frac{d\mathbf{M}\boldsymbol{\Sigma}(r)}{dr}\right) \\
&= \frac{1}{n-1}\mathrm{tr}\left(\mathbf{M}\frac{d\boldsymbol{\Sigma}(r)}{dr}\right) \\
&= \frac{1}{n-1}\left[\mathrm{tr}\left(\frac{d\boldsymbol{\Sigma}(r)}{dr}\right) - \mathrm{tr}\left(\frac{\mathbf{1}\mathbf{1}^\top}{n}\frac{d\boldsymbol{\Sigma}(r)}{dr}\right)\right],
\end{aligned}$$

where $\bar{\lambda}_{M\Sigma}(r) = \bar{\lambda}_{M\Sigma}$. This notation is used to stress the dependence of the value on r. In addition,

$$\left(\frac{d\boldsymbol{\Sigma}(r)}{dr}\right)_{ij} = \exp\left\{-\frac{d_{ij}}{r}\right\}\frac{d_{ij}}{r^2}$$

is always zero when $i = j$ because the Euclidean distance d_{ii} is zero. Thus, we obtain:

$$\begin{aligned}
\frac{d\,\mathrm{tr}(\mathbf{M}\boldsymbol{\Sigma}(r))}{dr} &= -\frac{1}{n-1}\mathrm{tr}\left(\frac{\mathbf{1}\mathbf{1}^\top}{n}\frac{d\boldsymbol{\Sigma}(r)}{dr}\right) \\
&= -\frac{1}{n(n-1)r^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\exp\left\{-\frac{d_{ij}}{r}\right\}d_{ij} < 0, \qquad \forall r.
\end{aligned}$$

□

Generally, regarding MF class, the larger ν, the smoother the process (Gneiting et al., 2010). Moreover, we bring to mind that the MF covariance function provides a marginally specified distribution of the process. In this case, the marginal variance σ^2 does not depend upon the parameters involved in the function. As being so, the equality

$$\bar{\lambda}_{M\Sigma} = \bar{\lambda}_{MC}$$

is verified. For this reason Figure 3.2 shows simultaneously the trend of the variability and the inverse smoothness of the stochastic vector leading to the following scheme:

$$\uparrow r \qquad \Longleftrightarrow \qquad \downarrow \text{variability} \qquad \Longleftrightarrow \qquad \uparrow \text{smoothness.}$$

3.5.2 First order autoregressive process

For the temporal setup, we consider a random vector $\boldsymbol{X}$ to be a zero-mean autoregressive process of order one (AR1) (See appendix C for more details about the process and for a new proposal of good approximation of eigenvalues of its precision matrix) with normal errors that can be expressed in conditional form

$$X_t|X_1,\ldots,X_{t-1} \sim \mathcal{N}(\phi X_{t-1},\sigma^2), \tag{3.39}$$

where the index $t \in \{1,\ldots,n\}$ represents time, σ^2 is the conditional variance and ϕ, the autocorrelation, is the structure parameter. This requires $|\phi| < 1$ for stationary to hold. The covariance matrix of an AR1 process is characterized by the following entries

$$\begin{aligned}\Sigma_{ij} &= \sigma^2 R_{ij} \\ &= \sigma^2 \frac{1}{1-\phi^2}\phi^{|i-j|} = \sigma^2\frac{1}{1-\phi^2}C_{ij} \qquad\qquad \forall i,j = 1,\ldots,n,\end{aligned} \tag{3.40}$$

where $C_{ij} = \phi^{|i-j|}$ is the correlation function of the process and the remaining part is the marginal variance expressed in function of the conditional and structure parameters. For different values of $n \in \{5, 50, 100, 300\}$, Figure 3.3 shows the different trend of the process' sampling variance (bottom) and inverse smoothness (top).

The covariance function in (3.40) arises from a conditionally specified model (see equation (3.39)). As a result, the marginal variance varies with structure parameter. This is the reason why, in Figure 3.3, trends of curves regarding variability and smoothness are different, unlike in the Matérn case. Indeed, for different values of n, Figure 3.3 (bottom) shows that the variability increases when the absolute value of ϕ_x increases. Actually, the trend of the sampling variance varies not only with the autocorrelation parameter but also with n. In the limiting case of $n \to \infty$, considering the covariance matrix characterizing an AR(1) process reported in (C.2) and in order to find out the trend of the sampling variance, we compute the following derivative of $\mathrm{tr}(\mathbf{M\Sigma})$ with respect to the ϕ:

$$\begin{aligned}\frac{d\,\bar{\lambda}_{M\Sigma}(\phi)}{d\phi} &= \frac{n}{n-1}\sigma^2\frac{d}{d\phi}\frac{1}{1-\phi^2} \\ &= \frac{n}{n-1}\sigma^2\frac{2\phi}{(1-\phi^2)^2}.\end{aligned}$$

This demonstrates the result illustrated for large n in Figure 3.3 (bottom): the sampling variance of the process in function of ϕ is increasing for positive values of ϕ, decreasing otherwise.

With regard to the smoothness of the AR(1) process, it can be shown that the inverse smoothness characterizing such process is strictly decreasing function of the autocorre-

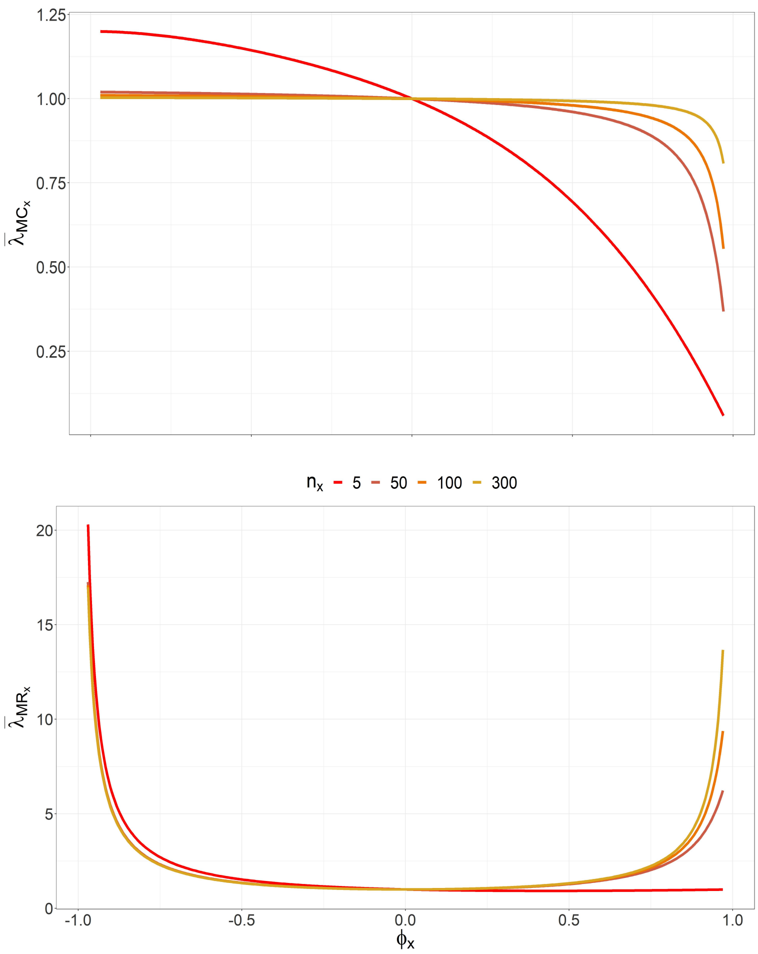

Figure 3.3: Decay of inverse smoothness (top) and variability (bottom) trends of $\boldsymbol{X}$ generated as Gaussian process with AR(1) covariance function, as the autocorrelation parameter increases.

lation parameter ϕ. Indeed, as highlighted in Figure 3.3, for an AR(1) process only the previous term in the process and the noise term contribute to the realization. Increasing ϕ from 0 toward 1 makes the process smoother because the output gets a larger contribution from the previous term relative to the noise. This results in a "smoothing" of the output (Siegel and Wagner, 2022). If ϕ is close to 0, then the process still looks like white noise, but as ϕ approaches -1 from 0, process is more jagged than a white noise process.

3.5.3 Conditional autoregressive process

The case of AR1 shows that when the marginal variance depends on the structure parameters, the variability and smoothness of the process have different patterns. It is interesting to show what happens to the trends of the variability and smoothness of the process when considering another type of covariance function that provide a marginal variance depending upon the structure parameters, in particular in the spatial domain.

For this purpose, we assume $\boldsymbol{X}$ follows a conditionally specified process in the spatial setup, the conditional autoregressive (CAR) process. It is defined by a full set of site specific univariate Gaussian conditionals (Besag, 1974; Rue and Held, 2005) with mean

$$\mathbb{E}[X_i|\boldsymbol{X}_{-i}] = \sum_{k=1}^{n} \kappa w_{ik}\, X_k = \sum_{k\sim i} \kappa w_{ik}\, X_k,$$

where $\boldsymbol{X}_{-i}$ is the $(n-1)$-vector such that $\boldsymbol{X}_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)^\top$ and κ is the structure parameter that can also be interpreted as a spatial smoothness parameter (local interaction) because its value determines the spatial smoothness of the process (see Figure 3.4 (top)). The above expression defines a unique joint multivariate Gaussian distribution with zero mean and covariance matrix

$$\boldsymbol{\Sigma} = \sigma^2(\mathbf{I}_n - \kappa \boldsymbol{W})^{-1}$$

such that

$$\mathbb{V}[X_i|\boldsymbol{X}_{-i}] = \Sigma_{ii} \qquad\qquad i = 1, \ldots, n.$$

Here, a spatial lattice structure is accompanied by the $n \times n$ adjacency or neighborhood matrix $\boldsymbol{W}$, with elements $w_{ii} = 0$, $w_{ik} = 1$ if the site k is neighbor to site i (denoted $k \sim i$ hereafter) and $w_{ik} = 0$ otherwise. The sufficient condition ensuring positive definiteness of the covariance matrix is $\kappa \in (\kappa_{\min}, \kappa_{\max})$, where $\kappa_{\min} = \lambda(\boldsymbol{W})_n^{-1}$ and $\kappa_{\max} = \lambda(\boldsymbol{W})_1^{-1}$. Moreover, κ is a direct measure of (conditional) spatial autocorrelation.

We now explore two geographical regions: the first contains the boundaries of Missouri's 115 counties and the second comprises the borders of Texas' 255 counties. To guarantee that both covariance matrices are positive definite, we restrict κ to be in the interval $(-0.3467, 0.1702)$ and $(-0.3169, 0.1503)$, respectively. Figure 3.4 (bottom) confirms, as for the AR1 case, a decreasing trend of the sampling variance as κ_x varies from $\kappa_{\min}$ to 0, and an increasing trend of the sampling variance when κ_x varies from 0 to $\kappa_{\max}$. Moreover, Figure 3.4 exhibits a monotonic decreasing trend of sampling inverse smoothness as a function of the structure parameter, less steep for negative values ok κ_x and rapidly descending otherwise. This highlights that smoothness is an increasing function of κ_x. The differences observed for Missouri and Texas show that smoothness and variability are influenced by the spatial layout of the lattice.

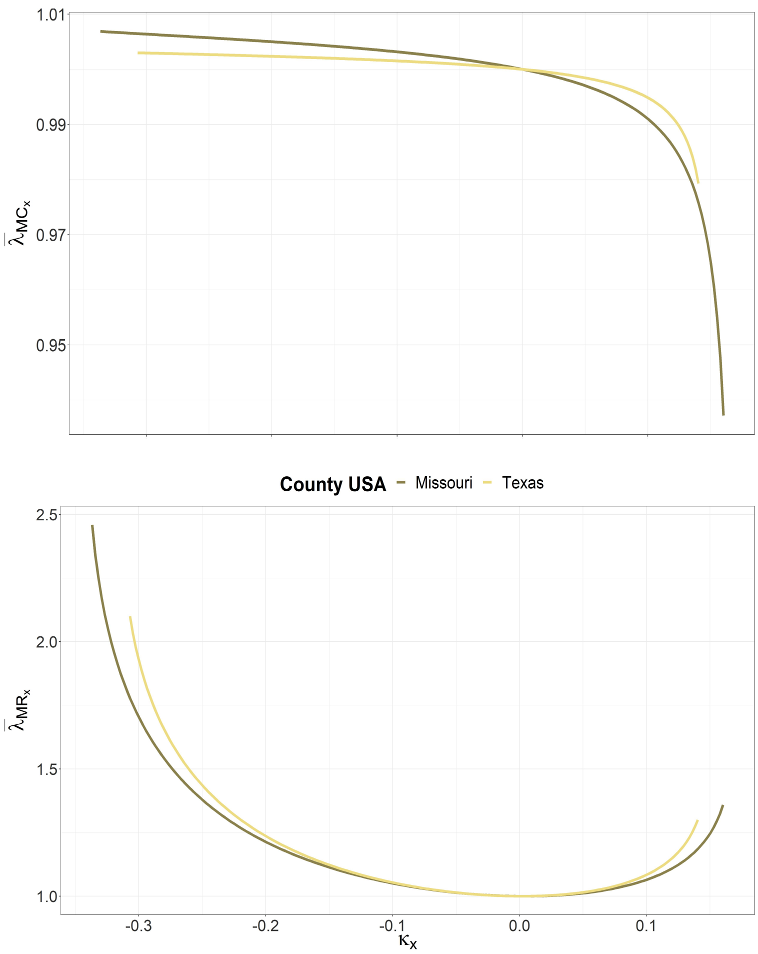

Figure 3.4: Decay of inverse smoothness (top) and variability (bottom) trends of $\boldsymbol{X}$ generated as Gaussian CAR spatial process as function of the structure parameter.

3.6 Quantification of confounding

In this section we provide some measures aimed at approximating the expected bias and at highlighting how the structure of the DGP impacts the regression of $\boldsymbol{Y}$ on the covariates. The latter point deserves some investigation: the structure of the DGP impacts the variance explained by the covariates in a non-trivial way, but the theory concerning the distribution of Gaussian QFs is exploited to give some useful insights about this topic.

3.6.1 Laplace approximation

Although in Theorem (3.2.2) we provided the exact formula for calculating the estimator bias, in this section we suggest an approximation in order to underline how the structure

of the DGP impacts the regression of $\boldsymbol{Y}$ on the covariates.
The Laplace method for approximating integrals is applied to give a general approximation for the first moment of a ratio of quadratic forms in random variables (Lieberman, 1994). This simple approximation, which only entails basic algebraic operations, has evident practical appeal for the evaluation of confounding. Actually, it is a Taylor approximation at the point $(\operatorname{tr}(\boldsymbol{\Delta}\Sigma_{zx}), \operatorname{tr}(\boldsymbol{\Delta}\Sigma_{x}))$. Then, the Laplace expansion up to $O(n^{-2})$ for $\mathbb{E}_X\left[R^{1,1}_{\Delta A_{z.x},\Delta}(\boldsymbol{X})\right]$ is:

$$\mathbb{E}_X\left[R^{1,1}_{\Delta A_{z.x},\Delta}(\boldsymbol{X})\right] = E_L + e_1 + O(n^{-2}), \tag{3.41}$$

where

$$\begin{aligned}
E_L &= \frac{\mathbb{E}_X\left[\boldsymbol{X}^\top \boldsymbol{\Delta}\mathbf{A}_{z\cdot x}\boldsymbol{X}\right]}{\mathbb{E}_X\left[\boldsymbol{X}^\top \boldsymbol{\Delta}\boldsymbol{X}\right]} = \frac{\operatorname{tr}(\boldsymbol{\Delta}\boldsymbol{\Sigma}_{zx})}{\operatorname{tr}(\boldsymbol{\Delta}\boldsymbol{\Sigma}_{x})} = \frac{\bar{\lambda}_{\Delta\Sigma_{zx}}}{\bar{\lambda}_{\Delta\Sigma_{x}}}. \qquad (3.42)\\
e_1 &= \frac{\mathbb{E}_X\left[\boldsymbol{X}^\top \boldsymbol{\Delta}\mathbf{A}_{z\cdot x}\boldsymbol{X}\right]}{\mathbb{E}^3_X\left[\boldsymbol{X}^\top \boldsymbol{\Delta}\boldsymbol{X}\right]}\mathbb{V}_X\left[\boldsymbol{X}^\top \boldsymbol{\Delta}\boldsymbol{X}\right] + \\
&\quad - \frac{\operatorname{Cov}_X\left(\boldsymbol{X}^\top \boldsymbol{\Delta}\mathbf{A}_{z\cdot x}\boldsymbol{X}, \boldsymbol{X}^\top \boldsymbol{\Delta}\boldsymbol{X}\right)}{\mathbb{E}^2_X\left[\boldsymbol{X}^\top \boldsymbol{\Delta}\boldsymbol{X}\right]}\\
&= 2\left[\frac{\operatorname{tr}(\boldsymbol{\Delta}\boldsymbol{\Sigma}_{zx})\operatorname{tr}(\boldsymbol{\Delta}\boldsymbol{\Sigma}_{x}) - \operatorname{tr}(\boldsymbol{\Delta}\boldsymbol{\Sigma}_{zx}\cdot\boldsymbol{\Delta}\boldsymbol{\Sigma}_{x})}{\operatorname{tr}^2(\boldsymbol{\Delta}\boldsymbol{\Sigma}_{x})}\right]\\
&= O(n^{-1}).
\end{aligned}$$

In some special case, the approximate mean collapses to the exact formula. The sufficient conditions to achieve it are established in Lieberman (1994). Indeed, in the case of spherical DGP, E_L is the exact value for the bias. Moreover it happens also when a spherical process is assumed only for the covariate.
When a OLS model is posited, i.e. when $\boldsymbol{\Delta} = \mathbf{M}$, the approximation of the marginal expectation of the estimator bias depends upon the cross-covariance between $\boldsymbol{X}$ and $\boldsymbol{Z}$ and the covariance matrix of the covariate. This shows that the estimator bias is larger when the variability of the expected sampling covariance of confounder and the covariate, the co-variability EV_{xz}, is large with respect to the variability of the covariate, i.e.

$$\gg \text{Bias} \qquad \Longleftrightarrow \qquad \bar{\lambda}_{M\Sigma_{zx}} \gg \bar{\lambda}_{M\Sigma_{x}},$$

while the marginal variability of the confounder, $\bar{\lambda}_{M\Sigma_z}$, has no direct impact on bias, with the only caveat that Σ_z must satisfy the conditions that guarantee positive definiteness of the DGP joint covariance matrix.

3.6.2 Portion of explained variability

In line with Paciorek (2010), the magnitude of confounding also depends on two further quantities suggested by Definition 2.1.1 and Lemma 3.0.1: the portion of covariate and response variability explained by the confounder process. To capture these contributions, we introduce the random variables denoted by $\mathrm{PV}_{x(z)}$ and $\mathrm{PV}_{y(z)}$, respectively. These quantities are a sort of coefficient of determination.

In order to determine $\mathrm{PV}_{x(z)}$, we consider the following regression model:

$$\boldsymbol{X} = \mathbf{a}_{x\cdot 0(z)} + \mathbf{A}_{x\cdot z}\boldsymbol{Z} + \boldsymbol{\varepsilon}_{x|z}, \tag{3.43}$$

where the normally distributed error term, $\boldsymbol{\varepsilon}_{x|z}$, is characterized by the covariance matrix $\boldsymbol{\Sigma}_{x|z}$ introduced in Equation (2.10). The sampling variance of the covariate $\boldsymbol{X}$ can be re-written as follows:

$$\begin{aligned} V_x &= \frac{\boldsymbol{X}^\top \mathbf{M}\boldsymbol{X}}{n-1} \\ &= \frac{1}{n-1}\left(\boldsymbol{Z}^\top \mathbf{A}_{x\cdot z}^\top \mathbf{M}\mathbf{A}_{x\cdot z}\boldsymbol{Z} + \boldsymbol{\varepsilon}_{x|z}^\top \mathbf{M}\boldsymbol{\varepsilon}_{x|z} + 2\boldsymbol{\varepsilon}_{x|z}^\top \mathbf{M}\mathbf{A}_{x\cdot z}\boldsymbol{Z}\right) \\ &= \frac{1}{n-1}\begin{pmatrix}\boldsymbol{Z}\\ \boldsymbol{X}|\boldsymbol{Z}\end{pmatrix}^\top \begin{bmatrix}\mathbf{A}_{x\cdot z}^\top \mathbf{M}\mathbf{A}_{x\cdot z} & \mathbf{A}_{x\cdot z}^\top \mathbf{M}\\ \mathbf{M}\mathbf{A}_{x\cdot z} & \mathbf{M}\end{bmatrix}\begin{pmatrix}\boldsymbol{Z}\\ \boldsymbol{X}|\boldsymbol{Z}\end{pmatrix}. \end{aligned}$$

The expected value is:

$$\mathrm{EV}_x = \frac{\mathrm{tr}(\mathbf{M}\boldsymbol{\Sigma}_x)}{n-1} = \bar{\lambda}_{M(A_{x\cdot z}\Sigma_z A_{x\cdot z}^\top)} + \bar{\lambda}_{M\Sigma_{x|z}}.$$

Thus, posing

$$\mathbf{A} = \begin{bmatrix}\mathbf{A}_{x\cdot z}^\top \mathbf{M}\mathbf{A}_{x\cdot z} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} \qquad \text{and} \qquad \mathbf{B} = \begin{bmatrix}\mathbf{A}_{x\cdot z}^\top \mathbf{M}\mathbf{A}_{x\cdot z} & \mathbf{A}_{x\cdot z}^\top \mathbf{M}\\ \mathbf{M}\mathbf{A}_{x\cdot z} & \mathbf{M}\end{bmatrix},$$

the portion of the covariate variability explained by the confounder process $\mathrm{PV}_{x(z)}$ is defined by the following ratio of quadratic forms in Gaussian random variables

$$\begin{aligned} \mathrm{PV}_{x(z)} &= \frac{\begin{pmatrix}\boldsymbol{Z}\\ \boldsymbol{X}|\boldsymbol{Z}\end{pmatrix}^\top \begin{bmatrix}\mathbf{A}_{x\cdot z}^\top \mathbf{M}\mathbf{A}_{x\cdot z} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{pmatrix}\boldsymbol{Z}\\ \boldsymbol{X}|\boldsymbol{Z}\end{pmatrix}}{\begin{pmatrix}\boldsymbol{Z}\\ \boldsymbol{X}|\boldsymbol{Z}\end{pmatrix}^\top \begin{bmatrix}\mathbf{A}_{x\cdot z}^\top \mathbf{M}\mathbf{A}_{x\cdot z} & \mathbf{A}_{x\cdot z}^\top \mathbf{M}\\ \mathbf{M}\mathbf{A}_{x\cdot z} & \mathbf{M}\end{bmatrix}\begin{pmatrix}\boldsymbol{Z}\\ \boldsymbol{X}|\boldsymbol{Z}\end{pmatrix}} \\ &= R_{A,B}^{1,1}\left((\boldsymbol{Z}^\top, (\boldsymbol{X}|\boldsymbol{Z})^\top)^\top\right). \end{aligned}$$

We obtain the matrix $\boldsymbol{\Sigma}_{z,x|z}^{1/2} \succ 0$ such that the block diagonal covariance matrix of the two independent processes $\boldsymbol{Z}$ and $\boldsymbol{X}|\boldsymbol{Z}$, $\boldsymbol{\Sigma}_{z,x|z} \in \mathcal{S}_{++}^{2n}$, can be decomposed as $\boldsymbol{\Sigma}_{z,x|z}^{1/2}\boldsymbol{\Sigma}_{z,x|z}^{1/2} = \boldsymbol{\Sigma}_{z,x|z}$, so that $PV_{x(z)}$ can be written as a function of a standard normal vector $\boldsymbol{\nu}$, that is

$$\mathrm{PV}_{x(z)} = R^{1,1}_{\tilde{A},\tilde{B}}(\boldsymbol{\nu}),$$

where $\tilde{\mathbf{A}} = \boldsymbol{\Sigma}_{z,x|z}^{1/2}\mathbf{A}\boldsymbol{\Sigma}_{z,x|z}^{1/2}$ and $\tilde{\mathbf{B}} = \boldsymbol{\Sigma}_{z,x|z}^{1/2}\mathbf{B}\boldsymbol{\Sigma}_{z,x|z}^{1/2}$. With the aim to provide the expected $\mathrm{PV}_{x(z)}$, let us consider the following spectral decomposition

$$\tilde{\mathbf{B}} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^\top = \mathbf{P}\,\mathrm{diag}\,(\boldsymbol{\lambda})\,\mathbf{P}^\top,$$

and define $\mathbf{C}_x = \mathbf{P}^\top\tilde{\mathbf{A}}\mathbf{P}$. From Equations (3.10) and (3.20), we obtain the expected value of $\mathrm{PV}_{x(z)}$ in term of Carlson's R function

$$\begin{aligned}
\mathbb{E}_X\left[\mathrm{PV}_{x(z)}\right] &= \sum_{j=1}^{2n} c_{x,jj}\int_0^\infty \prod_{i=1}^{2n}(1+2\lambda_i t)^{-1/2}\frac{1}{1+2\lambda_j t}\,dt \\
&= \sum_{j=1}^{2n} c_{x,jj} B\left(1, n-\frac{1}{2}h_j\right) R_j\left(1; \frac{1}{2}\mathbf{1}_{2n+2-h_j}, 2\boldsymbol{\lambda}'_{j,+}\right) \\
&= I^{1,1}_{h_j}\left(\boldsymbol{\lambda}'\right),
\end{aligned} \tag{3.44}$$

where $\boldsymbol{\lambda}'_j = (\lambda_1,\ldots,\lambda_{2n},\lambda_j,\lambda_j)$ and $h_j = \mathrm{rankdef}\,\tilde{\mathbf{B}}$[1] for $j = 1,\ldots,\mathrm{rank}\,\tilde{\mathbf{B}}$ and $h_j = \mathrm{rankdef}\,\tilde{\mathbf{B}}+2$ for $j = 1,\ldots,n$.

Applying Theorem 3.3.1, the variance is supposed to assume the maximum value when $\mathrm{PV}_{x(z)}$ approaches 0 and the minimum when it is close to 1. It is immediate to see that the quantity in (3.44) is a sort of multivariate version of a linear coefficient of determination R^2 of the regression of $\boldsymbol{X}$ on $\boldsymbol{Z}$, that is the proportion of the variation of the dependent variable $\boldsymbol{X}$ explained by $\boldsymbol{Z}$. In our opinion, this is a possible alternative to R^2-like measure of association proposed in the literature. For example, Rencher (2002), in order to reflect the amount of association between the variables, proposes the largest squared canonical correlation, i..e. the maximum eigenvalue of the matrix $\mathbf{A}_{z\cdot x}\mathbf{A}_{x\cdot z}$. Another alternative measure proposed in this framework is the *RV coefficient* suggested by Robert and Escoufier (1976):

$$RV = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{xz})}{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_x)\mathrm{tr}(\boldsymbol{\Sigma}_z)}}.$$

In a similar manner we provide the expression for $\mathrm{PV}_{y(z)}$. Considering the DGP defined in (2.3), the covariate $\boldsymbol{X}$ and the residual structure expressd via the latent variable $\boldsymbol{Z}$ participate to explicate the variability of the outcome (Waller and Gotway, 2004). Because

[1] A matrix is said to be rank-deficient if it does not have full rank. The rank deficiency of a matrix $\mathbf{A}$, denoted by $\mathrm{rankdef}\,\mathbf{A}$, is the difference between the lesser of the number of rows and columns, and the rank.

of this, now let us consider the sampling variance of $\boldsymbol{Y}$ decomposed as follows:

$$
\begin{aligned}
V_y &= \frac{\boldsymbol{Y}^\top \mathrm{M} \boldsymbol{Y}}{n-1} \\
&= \frac{\mathcal{B}^2_{y\cdot x(z)} \boldsymbol{X}^\top \mathrm{M} \boldsymbol{X} + \mathcal{B}^2_{y\cdot z(x)} \boldsymbol{Z}^\top \mathrm{M} \boldsymbol{Z} + 2\mathcal{B}_{y\cdot x(z)}\mathcal{B}_{y\cdot z(x)} \boldsymbol{X}^\top \mathrm{M} \boldsymbol{Z}}{n-1} + \\
&\quad + \frac{\boldsymbol{\varepsilon}^\top_{y|x,z} \mathrm{M} \boldsymbol{\varepsilon}_{y|x,z} + 2\mathcal{B}_{y\cdot x(z)} \boldsymbol{X}^\top \mathrm{M} \boldsymbol{\varepsilon}_{y|x,z} + 2\mathcal{B}_{y\cdot z(x)} \boldsymbol{Z}^\top \mathrm{M} \boldsymbol{\varepsilon}_{y|x,z}}{n-1} \\
&= \frac{1}{n-1} \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Z} \\ \boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z} \end{pmatrix}^\top \begin{bmatrix} \mathcal{B}_{y\cdot x(z)}\mathrm{M} & \mathcal{B}_{y\cdot x(z)}\mathcal{B}_{y\cdot z(x)}\mathrm{M} & \mathcal{B}_{y\cdot x(z)}\mathrm{M} \\ \mathcal{B}_{y\cdot x(z)}\mathcal{B}_{y\cdot z(x)}\mathrm{M} & \mathcal{B}^2_{y\cdot z(x)}\mathrm{M} & \mathcal{B}_{y\cdot z(x)}\mathrm{M} \\ \mathcal{B}_{y\cdot x(z)}\mathrm{M} & \mathcal{B}_{y\cdot z(x)}\mathrm{M} & \mathrm{M} \end{bmatrix} \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Z} \\ \boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z} \end{pmatrix},
\end{aligned}
\tag{3.45}
$$

in which we call $\mathbf{B}$ the matrix characterizing the quadratic form expressed in the last line. The expected value of V_y is

$$
\begin{aligned}
\mathrm{EV}_y &= \frac{\mathrm{tr}(\boldsymbol{M}\boldsymbol{\Sigma}_y)}{n-1} \\
&= \frac{\mathrm{tr}(\boldsymbol{M}\left[\mathcal{B}_{y\cdot x(z)}\mathbf{I}_n \ : \ \mathcal{B}_{y\cdot z(x)}\mathbf{I}_n\right] \boldsymbol{\Sigma}_{x,z} \left[\mathcal{B}_{y\cdot x(z)}\mathbf{I}_n \ : \ \mathcal{B}_{y\cdot z(x)}\mathbf{I}_n\right])}{n-1} + \frac{\mathrm{tr}(\boldsymbol{M}\boldsymbol{\Sigma}_{y|x,z})}{n-1} \\
&= \mathcal{B}^2_{y\cdot x(z)}\bar{\lambda}_{M\Sigma_x} + \mathcal{B}^2_{y\cdot z(x)}\bar{\lambda}_{M\Sigma_z} + 2\mathcal{B}_{y\cdot x(z)}\mathcal{B}_{y\cdot z(x)}\bar{\lambda}_{M\Sigma_{xz}} + \sigma^2_{y|x,z}.
\end{aligned}
$$

Because of the covariates dependence we need to take into account also the bilinear term $\boldsymbol{X}^\top \mathrm{M}\boldsymbol{Z}$, that is proportional to the sampling covariance of the covariate and the confounder. It depends on the regression coefficients, the variance parameters and the structure matrices $\boldsymbol{R}_x$ and $\boldsymbol{R}_z$ through their respective eigenvalues: different structure matrices lead to different expected values of $\boldsymbol{V}_y$.

Let

$$
\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathcal{B}_{y\cdot x(z)}\mathcal{B}_{y\cdot z(x)}\mathrm{M} & \mathbf{0} \\ \mathcal{B}_{y\cdot x(z)}\mathcal{B}_{y\cdot z(x)}\mathrm{M} & \mathcal{B}^2_{y\cdot z(x)}\mathrm{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix},
$$

the portion of the response variability explained by confounder process, $\mathrm{PV}_{y(z)}$, can be expressed as the following ratio of quadratic forms in Gaussian random variables

$$
\mathrm{PV}_{y(z)} = R^{1,1}_{A,B}\left((\boldsymbol{X}^\top, \boldsymbol{Z}^\top, (\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z})^\top)^\top\right). \tag{3.46}
$$

Along the line that leads to the Formula (3.44), the expected value of $\mathrm{PV}_{y(z)}$ is

$$\mathbb{E}_Y\left[\mathrm{PV}_{y(z)}\right] = \sum_{j=1}^{3n} c_{y,jj} B\left(1, \frac{3n}{2} - \frac{1}{2}h_j\right) R_j\left(1; \frac{1}{2}\mathbf{1}_{3n+2-h_j}, 2\boldsymbol{\lambda}'_{j,+}\right) = I^{1,1}_{h_j}\left(\boldsymbol{\lambda}'\right),$$

where $\boldsymbol{\lambda}'_j = (\lambda_1, \ldots, \lambda_{3n}, \lambda_j, \lambda_j)$ contains the eigenvalues of $\tilde{\mathbf{B}} = \boldsymbol{\Sigma}^{1/2}_{x,z,y|x,z}\mathbf{B}\boldsymbol{\Sigma}^{1/2}_{x,z,y|x,z} = \mathbf{P}^\top\boldsymbol{\Lambda}\mathbf{P}$ in which $\boldsymbol{\Sigma}_{x,z,y|x,z}$ is the block diagonal covariance matrix of the two independent vectors $(\boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top$ and $\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{Z}$. In addition, $\mathbf{C}_y = \mathbf{P}^\top\tilde{\boldsymbol{\Lambda}}\mathbf{P}$ and $h_j = \operatorname{rankdef}\tilde{\mathbf{B}}$ for $j = 1, \ldots, \operatorname{rank}\tilde{\mathbf{B}}$ and $h_j = \operatorname{rankdef}\tilde{\mathbf{B}} + 2$ for $j = 1, \ldots, n$.

In the next chapter we will provide a discussion of the effect of the portion of explained variability on confounding, taking also into account the variability and smoothness of $\boldsymbol{X}$ and $\boldsymbol{Z}$.

Chapter 4

Applications

In this chapter we present a brief review of the main results of the spatial statistics literature regarding the evaluation of spatial confounding. Subsequently, we study the marginal sampling properties of $\hat{\beta}_x$ for DGPs employed in geostatistical , areal and temporal frameworks.

In the following we indicate how all scenarios of data generating process posited in the study will be specified. We undertake several type of covariance functions to construct valid covariance matrices of the underlying processes $\boldsymbol{\xi} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_{\xi})$ and $\boldsymbol{\psi} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_{\psi})$ in order to depict different dependence structures of the random vector $(\boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top$. The three kinds of covariance function used along the dissertation have been presented in Section 3.5, and these are the Matérn correlation function (geostatistical data, Section 3.5.1), the conditional autoregressive covariance function (areal data, Section 3.5.3) and the autoregressive of order one covariance function (temporal data, Section 3.5.2).

A critical step is to identify a suitable dependence structure not only within each variable, but between variables as well. In the multivariate literature, several different approaches have been developed for modeling cross-covariance among the variables of interest, such as the linear coregionalization model (LMC) (Matheron, 1982; Goulard and Voltz, 1992; Grzebyk and Wackernagel, 1994; Wackernagel, 2003) and its generalization, the modeling techniques based on the latent dimensions, the convolution methods and the copula-based model. It is difficult to specify non-trivial, valid parametric models for cross-covariance functions, because of the notorious requirement of positive definiteness. In this work, we use the linear model of coregionalization, where each component is represented as a linear combination of a set of independent underlying variables, $\boldsymbol{\xi}$ and $\boldsymbol{\psi}$, because under a certain condition it guarantees a valid joint covariance matrix for the vector of main processes, $(\boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top$. Gelfand et al. (2004) propose to construct a multivariate process by linear transformation of independent processes, giving rise to the most basic coregionalization model. In order to express a complete joint distribution for the vector of main processes, we need to model not only the dependence of measurements across units but also the dependence of the two variables at each unit. The latter is modeled by the coregionalization matrix $\mathbf{T} = \mathbf{F}\mathbf{F}^\top$ which needs to be positive definite to guarantee validity of joint covariance matrix. Aware of it, we consider

$$\mathbf{T} = \begin{bmatrix} \sigma_\xi^2 & \sigma_{\xi\psi} \\ \sigma_{\psi\xi} & \sigma_\psi^2 \end{bmatrix}, \tag{4.1}$$

where $\sigma_\xi^2 > 0$ and $\sigma_\psi^2 > 0$ are the marginal variance of the variables on subscript and $\sigma_{\xi\psi} = \rho\sigma_\psi\sigma_\xi$ is the covariance that can be expressed also in term of correlation parameter ρ, subject to $|\rho| < 1$, describing the strength of dependence between variables. Then, the LMC takes the following form:

$$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Z} \end{pmatrix} = (\mathbf{F} \otimes \mathbf{I}_n) \begin{pmatrix} \boldsymbol{\xi}^* \\ \boldsymbol{\psi}^* \end{pmatrix},$$

where $\boldsymbol{\xi}^* \sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}_\xi)$ and $\boldsymbol{\psi}^* \sim \mathcal{N}_n(\mathbf{0}, \mathbf{R}_\psi)$ are independent processes specified by some covariance function and $\otimes$ denotes the Kronecker product. We indicate with $\mathbf{R}_\xi$ the structure matrix that reflects the structure of the random vector $\boldsymbol{\xi}$ such that $\boldsymbol{\Sigma}_\xi = \sigma_\xi^2\mathbf{R}_\xi$. If $\boldsymbol{\xi}$ and $\boldsymbol{\psi}$ are identically distributed with covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_\xi = \boldsymbol{\Sigma}_\psi$, one obtains a separable cross-covariance function that allows to express the joint covariance matrix as

$$\boldsymbol{\Sigma}_{x,z} = \mathbf{T} \otimes \boldsymbol{\Sigma}. \tag{4.2}$$

A more general LMC for stationary processes arises when the underlying variables are independent but not identically distributed, giving rise to the following joint covariance matrix

$$\boldsymbol{\Sigma}_{x,z} = \mathbf{f}_1\mathbf{f}_1^\top \otimes \boldsymbol{\Sigma}_\xi + \mathbf{f}_2\mathbf{f}_2^\top \otimes \boldsymbol{\Sigma}_\psi$$

where $\mathbf{f}_k$ indicates the k-th column vector of $\mathbf{F}$, with $k = \{1, 2\}$.

The LCM implies symmetric cross-covariances by construction. We consider a further approach able to build valid asymmetric cross-covariance structures that we define "revised" linear model of coregionalization (rLMC) represented as

$$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Z} \end{pmatrix} = \begin{bmatrix} \mathbf{R}_\xi^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_\psi^{1/2} \end{bmatrix} \left[\begin{bmatrix} 1 & \sqrt{\rho}_{xz} \\ \sqrt{\rho}_{xz} & 1 \end{bmatrix} \otimes \mathbf{I}_n \right] \begin{pmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \end{pmatrix},$$

where $\mathbf{R}_\xi^{1/2}$ is such that $\boldsymbol{\Sigma}_\xi = \mathbf{R}_\xi^{1/2}\left(\mathbf{R}_\xi^{1/2}\right)^\top$. In particular, when $\mathbf{R}_\xi^{1/2}$ is the lower triangular matrix obtained through the Cholesky decomposition, we denote it by $\mathbf{L}_\xi$. The rLMC has been employed by Page et al. (2017) for the first time, and later by Nobre et al. (2021) and Marques et al. (2022). Additionally, ρ_{xz} is the correlation parameter such that

$|\rho_{xz}| < 1$ and $\boldsymbol{\nu}_1$, $\boldsymbol{\nu}_2$ are standard multivariate Gaussian processes. Then, the covariance matrix that characterizes the joint distribution of $(\boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top$ is always positive definite by construction and it is

$$\boldsymbol{\Sigma}_{x,z} = \begin{bmatrix} \boldsymbol{\Sigma}_\xi & \rho_{xz}\sigma_\xi\sigma_\psi \mathbf{R}_\xi^{1/2}\left(\mathbf{R}_\psi^{1/2}\right)^\top \\ \rho_{xz}\sigma_\xi\sigma_\psi \mathbf{R}_\psi^{1/2}\left(\mathbf{R}_\xi^{1/2}\right)^\top & \boldsymbol{\Sigma}_\psi \end{bmatrix}. \tag{4.3}$$

Under the rLMC $\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_\xi$ and $\boldsymbol{\Sigma}_z = \boldsymbol{\Sigma}_\psi$.

4.1 Main literature results in our perspective

Throughout the thesis, several references have already been made to the three papers that most inspired and led this work: Paciorek (2010), Page et al. (2017) and Nobre et al. (2021). In this section we provide a brief review of these papers set out in our notation, in which we will point out the assumptions made, the aims and the conclusions deduced by authors adding our own considerations.

In each work the authors suppose a stochastic data generative mechanism and study the properties of the estimator defined in (2.28) for the parameters of interest. By proposing their application settings we wish to corroborate our thesis, i.e. that the bias is closely related to the variability of the covariate and the co-variability between the covariate and the confounder rather than to the smoothness of these main processes.

4.1.1 Paciorek (2010)

Paciorek (2010) aims to investigate how the "spatial scales", what may be called smoothness in a broader scope of application, of the residual and the covariate affects inference. Consequently, a smaller (or local) spatial scale is associated to less smoothness characterizing a process.

The analytic and simulation results of Paciorek (2010) illustrate how the bias depends on the structure parameters characterizing the covariate and residual processes. The Matérn correlation function introduced in (3.38) is adopted. It is characterized by the fact that correlation parameters are directly related to the smoothness of the processes. As it will be discussed later, this can lead to a misleading link between the processes smoothness and the bias of the estimator.

The starting result concerns the situation where the covariate and the unmeasured confounder are characterized by the same level of smoothness. It means assuming that they share the same structure parameters, i.e. $\boldsymbol{\theta} = \boldsymbol{\theta}_x = \boldsymbol{\theta}_z$, for the construction of the marginal covariance matrices, $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_z$. In turn, a single structure matrix $\mathbf{R}(\boldsymbol{\theta}) = \mathbf{R}(\boldsymbol{\theta}_x) = \mathbf{R}(\boldsymbol{\theta}_z)$ is obtained. In order to get a valid joint distribution for the bivariate vector of covariate and confounder, a separable LMC as defined in (4.2) is considered. It brings the joint covariance matrix

$$\begin{aligned}\mathbf{\Sigma}_{x,z} &= \mathbf{T} \otimes \mathbf{R}(\boldsymbol{\theta}) \\ &= \begin{bmatrix} \sigma_x^2 & \rho_{xz}\sigma_x\sigma_z \\ \rho_{xz}\sigma_x\sigma_z & \sigma_z \end{bmatrix} \otimes \mathbf{R}(\boldsymbol{\theta}).\end{aligned}$$

In this setting, the bias for the estimator $\hat{\beta}_x$ conditionally on $\boldsymbol{X}$ reported in Equation (5) of the paper is the same deterministic value we have presented in Section 3.4 under the unstructured DGP assumption in Equation (3.31). Thus, whenever a separable cross-correlation model is assumed, this is equivalent to posit the same structure parameters, leading to a deterministic conditional bias of $\hat{\beta}_x$ that is function of marginal variance and covariance parameters. Another setup that leads to the same result consists in assuming a linear model of coregionalization in which the $\mathbf{F}$ matrix in (4.1) is a lower Cholesky triangle with the following structure

$$\mathbf{F} = \begin{bmatrix} \sigma_\xi & 0 \\ \rho_{xz}\sigma_\psi & \sqrt{1-\rho_{xz}^2}\sigma_\psi \end{bmatrix},$$

such that the joint covariance matrix of $(\boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top$ is obtained as follows

$$\mathbf{\Sigma}_{x,z} = \begin{bmatrix} \sigma_\xi^2\mathbf{\Sigma}_\xi & \rho_{xz}\sigma_\xi\sigma_\psi\mathbf{\Sigma}_\xi \\ \rho_{xz}\sigma_\xi\sigma_\psi\mathbf{\Sigma}_\xi & \rho_{xz}^2\sigma_\psi^2\mathbf{\Sigma}_\xi + (1-\rho_{xz}^2)\sigma_\psi^2\mathbf{\Sigma}_\psi \end{bmatrix}.$$

Thereby, in order to explore the situation in which the covariate varies "at two scales", meaning the covariate and the confounder has two different covariance structure where the covariate one is partially characterized by the confounder one, Paciorek (2010) uses a sort of LMC with $\mathbf{F}$ upper triangle Cholesky of $\mathbf{T}$ matrix in (4.1). He considers two underlying processes, $\boldsymbol{\xi} \sim \mathcal{N}_n(\mathbf{0}, \sigma_\xi^2\mathbf{R}(\boldsymbol{\theta}_\xi))$ and $\boldsymbol{\psi} \sim \mathcal{N}_n(\mathbf{0}, \sigma_\psi^2\mathbf{R}(\boldsymbol{\theta}_\psi))$, such that $\boldsymbol{X} = \boldsymbol{\psi} + \boldsymbol{\xi}$ is decomposed into a component, $\boldsymbol{\psi}$, that has the same structure parameters of the confounder $\boldsymbol{Z}$, and a component $\boldsymbol{\xi}$, which is independent of $\boldsymbol{\psi}$ and $\boldsymbol{Z}$. Specifically, this leads to the following joint covariance matrix

$$\mathbf{\Sigma}_{x,z} = \begin{bmatrix} \sigma_\psi^2\mathbf{R}(\boldsymbol{\theta}_\psi) + \sigma_\xi^2\mathbf{R}(\boldsymbol{\theta}_\xi) & \rho_{xz}\sigma_\psi\sigma_z\mathbf{R}(\boldsymbol{\theta}_\psi) \\ \rho_{xz}\sigma_\psi\sigma_z\mathbf{R}(\boldsymbol{\theta}_\psi) & \sigma_z^2\mathbf{R}(\boldsymbol{\theta}_\psi) \end{bmatrix}. \tag{4.4}$$

According with Lemma 3.0.1, we reproduce Equation (6) of the paper expressed in our notation:

$$\begin{aligned}\mathbb{E}_Y\left[\hat{\beta}_x|\boldsymbol{X}\right] &= \mathcal{B}_{y\cdot x(z)} + \mathcal{B}_{y\cdot z(x)}\rho_{xz}\sigma_\psi\sigma_z\frac{\boldsymbol{X}^T\boldsymbol{\Delta}\mathbf{R}(\boldsymbol{\theta}_\psi)\left(\sigma_\psi^2\mathbf{R}(\boldsymbol{\theta}_\psi)+\sigma_\xi^2\mathbf{R}(\boldsymbol{\theta}_\xi)\right)^{-1}\boldsymbol{X}}{\boldsymbol{X}^T\boldsymbol{\Delta}\boldsymbol{X}} \\ &= \mathcal{B}_{y\cdot x(z)} + \mathcal{B}_{y\cdot z(x)}\rho_{xz}\frac{\sigma_z}{\sigma_\psi}\frac{\boldsymbol{X}^T\boldsymbol{\Delta}\left(\mathbf{I}_n+\frac{\sigma_\xi^2}{\sigma_\psi^2}\mathbf{R}(\boldsymbol{\theta}_\xi)\mathbf{R}(\boldsymbol{\theta}_\psi)^{-1}\right)^{-1}\boldsymbol{X}}{\boldsymbol{X}^T\boldsymbol{\Delta}\boldsymbol{X}}, \end{aligned} \tag{4.5}$$

where the QF is denoted by the author with $k(\boldsymbol{X})$, that is what he calls "bias modification term". Besides, Paciorek (2010) focuses the attention on the following ratios detectable within the bias Formula (4.5) via trivial algebra and matrix manipulation

$$\begin{aligned}p_c &= \frac{\sigma_\psi^2}{\sigma_\psi^2+\sigma_\xi^2}, \\ p_z &= \frac{\mathcal{B}_{y\cdot z(x)}^2\sigma_z^2}{\mathcal{B}_{y\cdot z(x)}^2\sigma_z^2+\sigma_{y|x,z}^2},\end{aligned}$$

where p_z should be the portion of the residual variation being the contribution of the confounder and p_c should quantify the magnitude of the confounded component of $\boldsymbol{X}$ relative to its total variation. This interpretation is legitimate only when $\boldsymbol{X}$ and $\boldsymbol{Z}$ are spherical. In this regard, in Section 3.6, the two quantities that succeed in this respect were presented considering the structure of the processes, $\mathrm{PV}_{y(z)}$ and $\mathrm{PV}_{x(z)}$, respectively. About that, Figure 4.2 shows the relation between the expected value of the bias modification term and the two aforementioned portions.

Under the same structure parameters setting, i.e. for $\boldsymbol{\theta}_\xi = \boldsymbol{\theta}_\psi$, $\mathbb{E}_X\left[k(\boldsymbol{X})\right] = p_c$ and the resulting bias is equal to the previous case, that is

$$\mathcal{B}_{y\cdot z(x)}\rho_{xz}\frac{\sigma_z\sigma_\psi}{\sigma_\psi^2+\sigma_\xi^2} = \mathcal{B}_{y\cdot z(x)}\frac{\sigma_{xz}}{\sigma_x^2}.$$

In the figures reported in what follows, the values relative to the marginal sampling properties of estimator $\hat{\beta}_x$ are computed using the formulas expressed in Theorem 3.2.2. Values reported in these figures match with the ones obtained through the simulation study implemented by Paciorek (2010).

For a regular grid of $n = 100$ locations on the unit square and using an exponential correlation function for the processes, i.e. $\boldsymbol{\theta} = (0.5, r)^\top$, Figure 4.1 shows the expected value of $k(\boldsymbol{X})$. Since it is proportional to the bias, it is possible to deduce that when $r_\xi \ll r_\psi$ there is less bias than under the same structure parameter setting. Above the diagonal, for $r_\xi \gg r_\psi$, there is more bias. The patterns in Figure 4.1 are similar regardless of the values of p_c and p_z. For larger values of p_c the bias is larger, while for larger values of p_z the effects of the structure parameters are weaker.

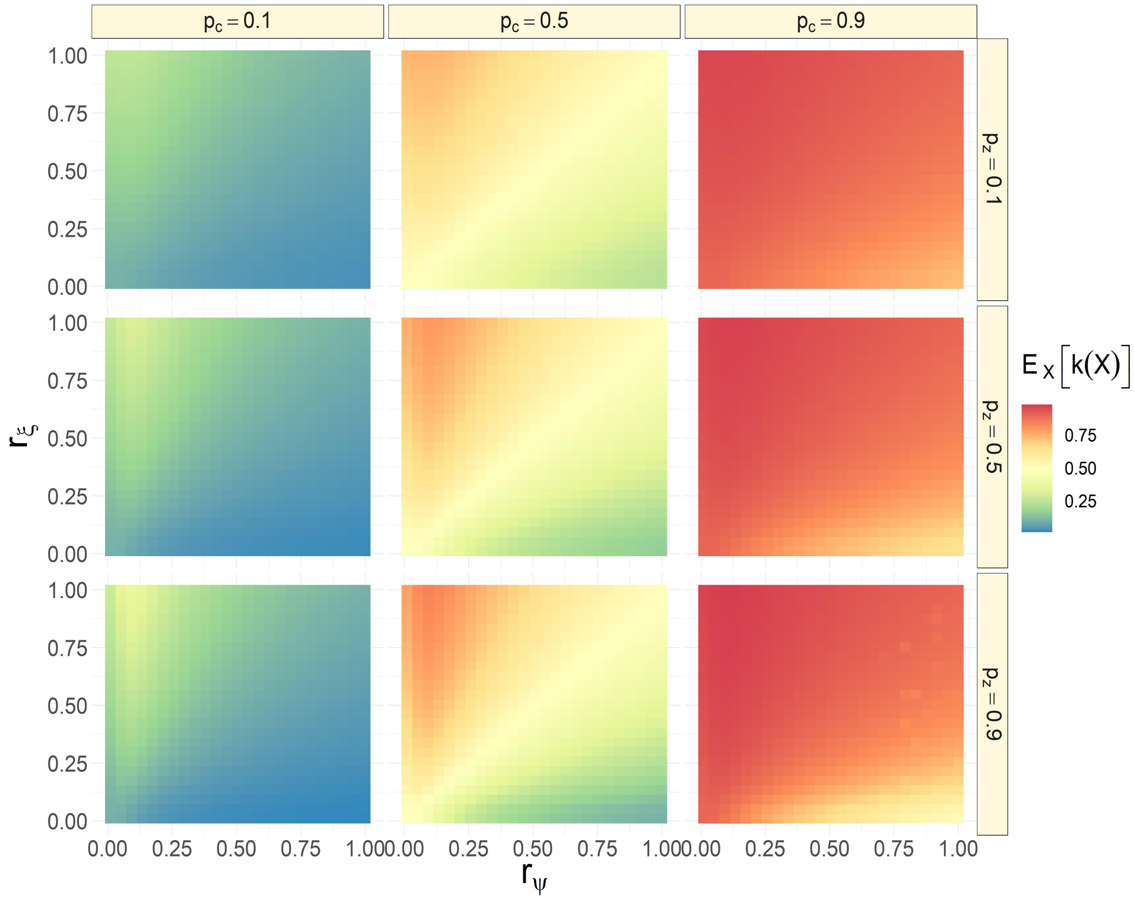

Figure 4.1: The expected value of the bias modification term as a function of the structure parameters of two underlying processes generating the covariate and confounder structure for a selection of values of p_c and p_z. Regarding the model assumption, it is posited $\mathbf{S} = \mathbf{\Sigma}_{y|x}$.

From the above, Paciorek (2010) concludes that the inclusion of the spatial residual term in the model accounts for spatial correlation reducing the bias from spatial confounder only "when there is unconfounded variability in the exposure at a scale smaller than the scale of confounding". It means that there is less confounding when the confounder is smoother than the covariate.

As mentioned in Section 3.5, the expected inverse smoothness of a process, and so the smoothness, is related to the eigenvalues of the correlation matrix of a process. In particular as in this case, when one posits a covariance function that actually is also a correlation function, e.g. the Matérn function, it can be ensured that the estimator bias depends upon them and consequently on the smoothness of processes involved. More broadly, it is function of the covariance matrix eigenvalues. In fact, our theoretical explanation to what is illustrated in Figures 4.1 is linked to that. Using the first order Laplace approximation for the bias defined in Equation (3.42), we state that a lower bias occurs when

$$\bar{\lambda}_{M\Sigma_{xz}} \ll \bar{\lambda}_{M\Sigma_x},$$

that is

$$\begin{aligned} \rho_{xz}\sigma_z\sigma_\psi\bar{\lambda}_{MR_\psi} &\ll \sigma_\xi^2\bar{\lambda}_{MR_\xi} + \sigma_\psi^2\bar{\lambda}_{MR_\psi} \\ (\rho_{xz}\sigma_z\sigma_\psi - \sigma_\psi^2)\bar{\lambda}_{MR_\psi} &\ll \sigma_\xi^2\bar{\lambda}_{MR_\xi}. \end{aligned} \tag{4.6}$$

Hence, holding constant all the variance and correlation parameters in (4.6) and since Paciorek (2010) uses the Matérn correlation function characterized by the following relation

$$r_\xi \ll r_\psi \Longleftrightarrow \bar{\lambda}_{MR_\psi} \ll \bar{\lambda}_{MR_\xi}$$

(see Figure 3.2), our idea coincides with what has been gathered by the author.

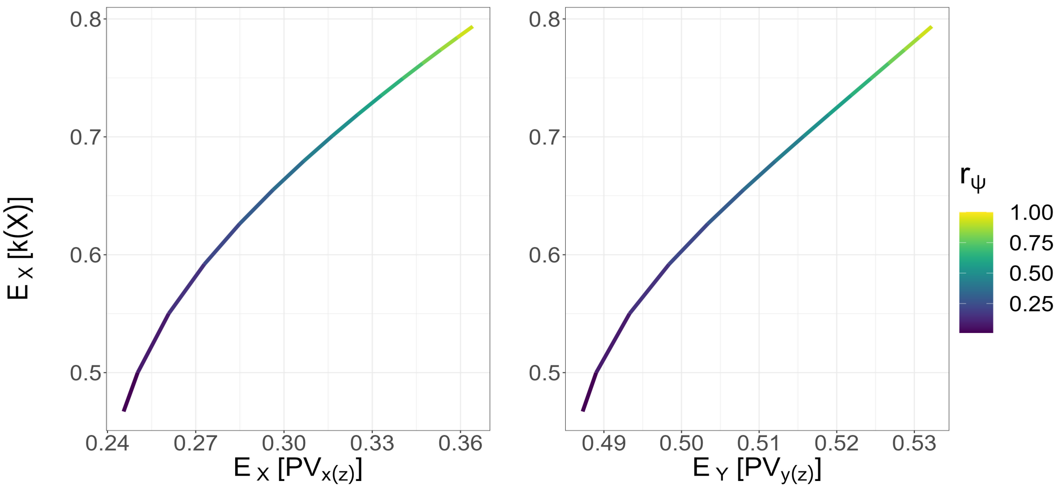

Figure 4.2: On the right (left) the expected value of the bias modification term as a function of the portion of the response (covariate) variability explained by the confounder processes is shown with respect to r_ψ. Regarding the model assumption, it is posited $\mathbf{S} = \mathbf{\Sigma}_{y|x}$.

In our view, the estimator bias in Formula (3.22) may not be regulated only by what Paciorek (2010), Page et al. (2017) and Guan et al. (2022) call the "spatial scale" expressed through the range parameter but it depends on the variability of processes involved. In this regard it will be interesting to study the effect that different choices of covariance functions has on confounding.

4.1.2 Page et al. (2017)

This paper provides a more detailed study of regression coefficient estimation from spatial models when covariate $\boldsymbol{X}$ and unmeasured confounder $\boldsymbol{Z}$ are correlated and, in addition, it contains a formal study regarding spatial prediction.

The data generating process is specified as $\boldsymbol{X} \sim \mathcal{N}_n(\mathbf{0}, \sigma_x^2 \mathbf{R}_x(\boldsymbol{\theta}_x))$ and $\boldsymbol{Z} \sim \mathcal{N}_n(\mathbf{0}, \sigma_z^2 \mathbf{R}_z(\boldsymbol{\theta}_z))$ jointly normal with cross-covariance structure determined from rLMC (see Equation (4.3)). The authors derive the analytic formulas of conditional expected value of bias, variance and mean square error of the estimator defined in (2.28) in which it is assumed $\mathbf{S} = \boldsymbol{\Sigma}_{y|x}$. Hence, denoting $\boldsymbol{\Omega} = \sigma^2_{y|x,z}\mathbf{I}_n + \boldsymbol{\Sigma}_z$ yields to the following sampling properties:

$$
\begin{aligned}
\mathbb{E}_Y\left[\hat{\boldsymbol{\beta}}|\boldsymbol{X}\right] &= \begin{pmatrix}\mathcal{B}_{y\cdot 0(xz)}\\ \mathcal{B}_{y\cdot x(z)}\end{pmatrix} + \rho_{xz}\frac{\sigma_z}{\sigma_x}\mathbf{J}\mathbf{L}_z\mathbf{L}_x^{-1}\boldsymbol{X}, && (4.7)\\
\mathbb{V}_Y\left[\hat{\boldsymbol{\beta}}|\boldsymbol{X}\right] &= \mathbf{J}\boldsymbol{\Omega}\mathbf{J}^\top - \mathbf{J}\boldsymbol{\Sigma}_{zx}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xz}\mathbf{J}^\top\\
&= \left(\tilde{\boldsymbol{X}}^\top\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{X}}\right)^{-1} - \sigma_z^2\rho_{xz}^2\mathbf{J}\mathbf{R}_z\mathbf{J}^\top, && (4.8)\\
\mathrm{MSE}_Y\left[\hat{\boldsymbol{\beta}}|\boldsymbol{X}\right] &= \rho_{xz}^2\frac{\sigma_z^2}{\sigma_x^2}\mathrm{tr}\left(\mathbf{J}\mathbf{L}_z\mathbf{L}_x^{-1}\boldsymbol{X}\boldsymbol{X}^\top\mathbf{L}_x^{-\top}\mathbf{L}_z^\top\mathbf{J}^\top\right)\\
&\quad + \sigma^2_{y|x,z}\mathrm{tr}\left(\tilde{\boldsymbol{X}}^\top\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{X}}\right)^{-1} - \sigma_z^2\rho_{xz}^2\mathrm{tr}\left(\mathbf{J}\mathbf{R}_z\mathbf{J}^\top\right). && (4.9)
\end{aligned}
$$

They are a particular case of the result in Proposition 2.3.1, where

$$
\begin{aligned}
\mathbf{A}_{z\cdot x} &= \rho_{xz}\frac{\sigma_z}{\sigma_x}\mathbf{L}_z\mathbf{L}_x^{-1},\\
\boldsymbol{\Sigma}_{y|x} &= \sigma^2_{y|x,z}\mathbf{I}_n + \sigma_z^2(1-\rho_{xz}^2)\mathbf{R}_z.
\end{aligned}
$$

Moreover, regarding the coefficient related to the covariate, the previous assumptions allows the matching also with the result in Lemma 3.0.1.

In this paper two kind of data are employed: point-referenced or geostatistical data using the exponential correlation function belonging to MF class and areal data using a conditional autoregressive covariance function to model $\boldsymbol{X}$ and $\boldsymbol{Z}$. The values relative to the marginal sampling properties of estimator $\hat{\beta}_x$ are computed using the formulas expressed in Theorem 3.2.2. The values match with the ones obtained through the simulation study implemented by Page et al. (2017).

Regarding the geostatistical data, Figure 4.3 shows that increasing ρ_{xz} impacts the all sampling properties of $\hat{\beta}_x$. As expected by theoretical result in (3.25) and as noted by the authors, the influence that r_z and r_x have on the estimator bias does not vary when

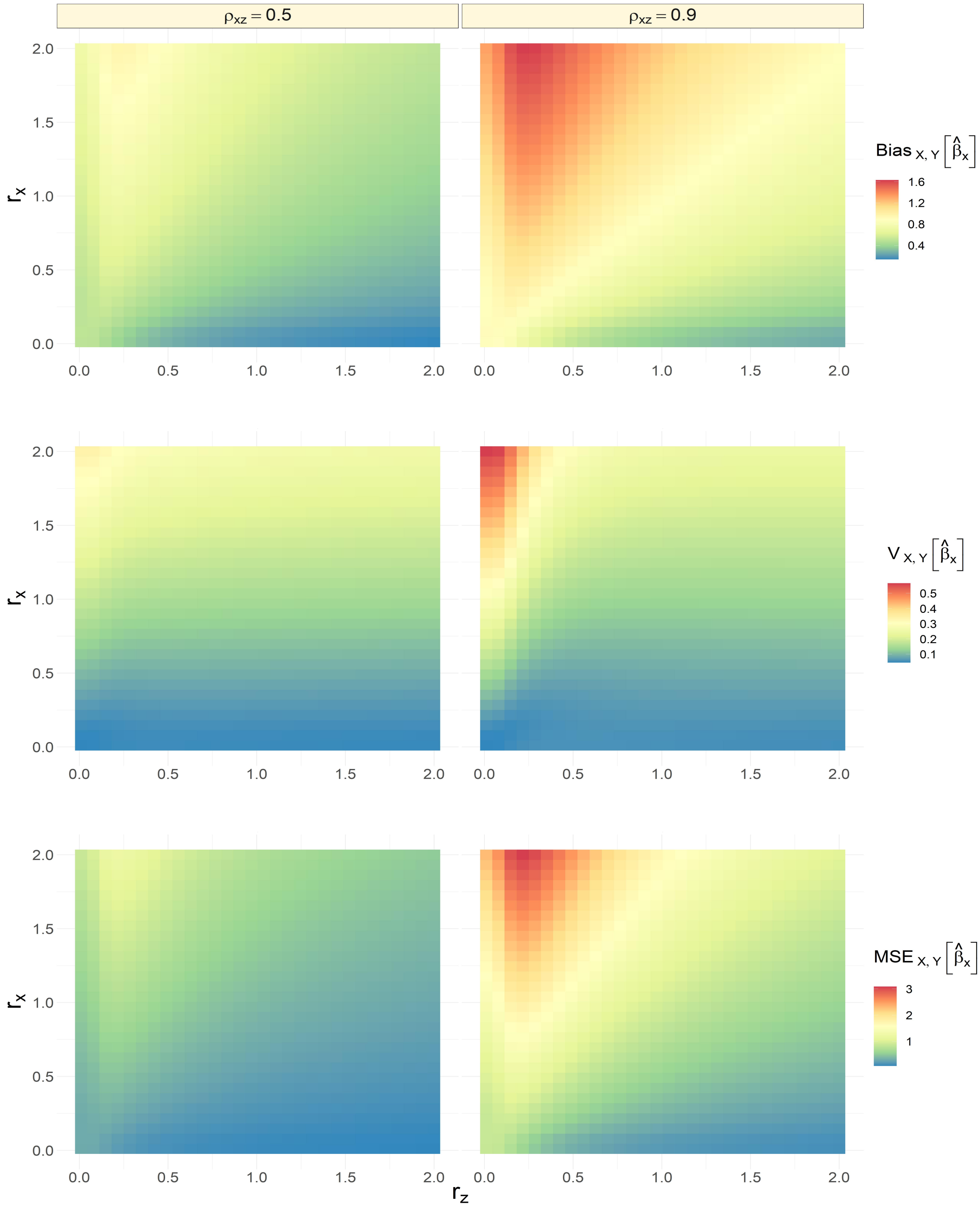

Figure 4.3: Bias, variance and MSE values associated with $\hat{\beta}_x$. They are evaluated for values of range parameter $r_x = r_z \in (0, 2)$. Additionally, $\rho \in \{0.5, 0.9\}$ while all other variance components are fixed at $\sigma^2_{y|x,z} = \sigma^2_x = \sigma^2_z = 1$. Regarding the model assumption, it is posited $\mathbf{S} = \mathbf{\Sigma}_{y|x}$.

ρ_{xz} increases, while that for the variance, and consequently for MSE, seems to change. That because the correlation parameter is proportional to the bias, unlike the others. The explanation for the patterns its the crucial point. In our opinion, there is no reason to bind it to the range parameters of $\boldsymbol{X}$ and $\boldsymbol{Z}$, but rather on their variability and co-variability. It seems that it may depend upon the portion of the response or covariate variability explained by the confounder.

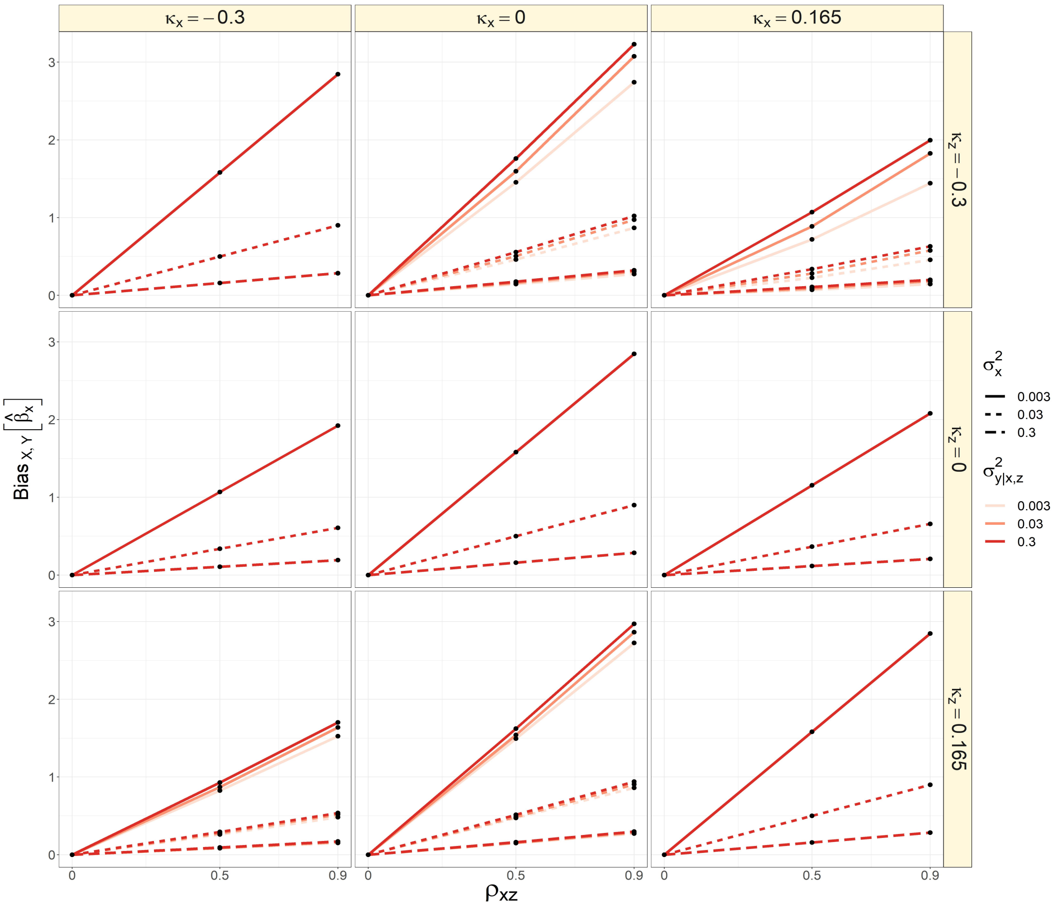

Figure 4.4: Estimator bias evaluated for values of the structure parameters $\kappa_x, \kappa_z \in \{-0.3, 0, 0.165\}$ and $\rho_{xz} \in \{0, 0.5, 0.9\}$. Regarding the model assumption, it is posited $\mathbf{S} = \boldsymbol{\Sigma}_{y|x}$.

However, what is apparent when considering the results obtained by Paciorek (2010) and Page et al. (2017) for the same geostatistical setting in which the only difference is the assumption regarding the cross-covariance function that gives $\boldsymbol{\Sigma}_{x,z}$ is that it possible to identify a similar pattern for the estimator bias but the explanation cannot be related to the smoothness of the processes for both cases. That because also the co-variability of the main processes plays an important role.

Concerning the areal data, Figure 4.4 shows the same result illustrated by the authors and the consequent difficulty in interpreting the results concerning the estimator bias as the structure parameters vary. What is clear is the relationship between ρ_{xz} and the bias.

4.1.3 Nobre et al. (2021)

The authors investigate the effects of spatial confounding in hierarchical spatial models (Gelfand et al., 2007). In particular, in a random intercept model considering unit and cluster-level covariates, respectively $\boldsymbol{X}_1 = \boldsymbol{X}_{1*} + \mathbf{C}\boldsymbol{\kappa}$ and $\boldsymbol{X}_2 = \mathbf{C}\boldsymbol{W}$, where $\mathbf{C}$ is a $n \times m$ matrix of 0's and 1's connecting the cluster with each of its units and $\boldsymbol{X}_{1*}$ is the part of $\boldsymbol{X}_1$ assumed to be known. They define locations as clusters and observations within clusters as units. Observations are accessible over m spatial locations, and at each location $i = 1, \dots, m$, n_i units are observed. Indicating with $\boldsymbol{Z} = \mathbf{C}\boldsymbol{\nu}$ the spatially structured latent process, the study considers the following data generating process

$$\boldsymbol{Y} = \mathcal{B}_{y\cdot 0(x_1x_2z)}\mathbf{1}_n + \mathcal{B}_{y\cdot x_1(z)}\boldsymbol{X}_1 + \mathcal{B}_{y\cdot x_2(z)}\boldsymbol{X}_2 + \boldsymbol{Z} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_{y|x_1,x_2,z}),$$

where $\boldsymbol{Y}$ is the n-dimensional response process, with $n = \sum_{i=1}^m n_i$. Here the two target parameters are $\mathcal{B}_{y\cdot x_1(z)}$ and $\mathcal{B}_{y\cdot x_2(z)}$. Then, the conditional distribution marginalized with respect to $\boldsymbol{Z}$ is given by

$$\boldsymbol{Y}|\boldsymbol{X}_1, \boldsymbol{X}_2 \sim \mathcal{N}_n(\mathcal{B}_{y\cdot 0(x_1x_2)}\mathbf{1}_n + \mathbf{A}_{y\cdot x_1}\boldsymbol{X}_1 + \mathbf{A}_{y\cdot x_2}\boldsymbol{X}_2,\ \mathcal{B}^2_{y\cdot z(x_1x_2)}\boldsymbol{\Sigma}_{z|x_1,x_2} + \boldsymbol{\Sigma}_{y|x_1,x_2,z}).$$

Regarding the statistical model assumed, the authors consider the estimator defined in (2.28) in which $\mathbf{S} = \boldsymbol{\Sigma}_{y|x}$. To settle the formalization of DGP, they assume further that $(\boldsymbol{\nu}^\top, \boldsymbol{\kappa}^\top, \boldsymbol{W}^\top)^\top$, and consequently $(\boldsymbol{Z}^\top, \boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top)^\top$, follows a LMC such that:

$$\begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\kappa} \\ \boldsymbol{W} \end{pmatrix} = \begin{pmatrix} f_{11} & 0 & 0 \\ f_{21} & f_{22} & 0 \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} \boldsymbol{\omega}_1 \\ \boldsymbol{\omega}_2 \\ \boldsymbol{\omega}_3 \end{pmatrix},$$

and so

$$\begin{pmatrix} \boldsymbol{Z} \\ \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{X}_{1*} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} f_{11} & 0 & 0 \\ f_{21} & f_{22} & 0 \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} \mathbf{C}\boldsymbol{\omega}_1 \\ \mathbf{C}\boldsymbol{\omega}_2 \\ \mathbf{C}\boldsymbol{\omega}_3 \end{pmatrix},$$

where $\boldsymbol{\omega}_k$, $k = 1, 2, 3$, are independent, each following a zero mean Gaussian process with unit marginal variance and covariance matrix $\mathbf{R}_k$. This leads to the following covariance matrix

$$\boldsymbol{\Sigma}_{\nu,\kappa,W} = \begin{bmatrix} f_{11}^2\mathbf{R}_1 & f_{11}f_{21}\mathbf{R}_1 & f_{11}f_{31}\mathbf{R}_1 \\ f_{11}f_{21}\mathbf{R}_1 & \sum_{i=1}^2 f_{2i}^2\mathbf{R}_i & \sum_{i=1}^2 f_{2i}f_{3i}\mathbf{R}_i \\ f_{11}f_{31}\mathbf{R}_1 & \sum_{i=1}^2 f_{2i}f_{3i}\mathbf{R}_i & \sum_{i=1}^3 f_{3i}^2\mathbf{R}_i \end{bmatrix}.$$

In order to express the sampling distribution of the estimator $\hat{\boldsymbol{\beta}}$ reported in Proposition 1 of the paper in our notation and following the Formula (2.12), we compute the conditional mean $\boldsymbol{\mu}_{z|x_1x_2}$ as:

$$\begin{aligned}
\mathbb{E}_Z\left[\boldsymbol{Z}|\boldsymbol{X}_1,\boldsymbol{X}_2\right] &= \mathbf{A}_{z\cdot x_1x_2}\begin{pmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{pmatrix} - \mathbf{A}_{z\cdot x_1(x_2)}\boldsymbol{X}_{1*}\\
&= \mathbf{A}_{z\cdot x_1(x_2)}\mathbf{C}\boldsymbol{\kappa} + \mathbf{A}_{z\cdot x_2(x_1)}\mathbf{C}\boldsymbol{W}\\
&= \left[\boldsymbol{\Sigma}_{zx_1} \;:\; \boldsymbol{\Sigma}_{zx_2}\right]\begin{bmatrix}\boldsymbol{\Sigma}_{x_1} & \boldsymbol{\Sigma}_{x_1x_2}\\ \boldsymbol{\Sigma}_{x_2x_1} & \boldsymbol{\Sigma}_{x_2}\end{bmatrix}^{-1}\begin{pmatrix}\mathbf{C}\boldsymbol{\kappa}\\ \mathbf{C}\boldsymbol{W}\end{pmatrix}\\
&= \mathbf{C}\left[\boldsymbol{\Sigma}_{\nu\kappa} \;:\; \boldsymbol{\Sigma}_{\nu W}\right]\begin{bmatrix}\mathbf{C}^\top & \mathbf{0}\\ \mathbf{0} & \mathbf{C}^\top\end{bmatrix}\begin{bmatrix}\mathbf{C}\boldsymbol{\Sigma}_{\kappa}\mathbf{C}^\top & \mathbf{C}\boldsymbol{\Sigma}_{\kappa W}\mathbf{C}^\top\\ \mathbf{C}\boldsymbol{\Sigma}_{W\kappa}\mathbf{C}^\top & \mathbf{C}\boldsymbol{\Sigma}_{W}\mathbf{C}^\top\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}\end{bmatrix}\begin{pmatrix}\boldsymbol{\kappa}\\ \boldsymbol{W}\end{pmatrix}.
\end{aligned}$$

Considering that there is at least one observation per location ($n \geq m$), the pseudo inverse[1] of the diagonal block matrices allows to get the simplification:

$$\begin{aligned}
\mathbb{E}_Z\left[\boldsymbol{Z}|\boldsymbol{X}_1,\boldsymbol{X}_2\right] &= \mathbf{C}\left[\begin{bmatrix}\mathbf{C} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{C}\boldsymbol{\Sigma}_{\kappa}\mathbf{C}^\top & \mathbf{C}\boldsymbol{\Sigma}_{\kappa W}\mathbf{C}^\top\\ \mathbf{C}\boldsymbol{\Sigma}_{W\kappa}\mathbf{C}^\top & \mathbf{C}\boldsymbol{\Sigma}_{W}\mathbf{C}^\top\end{bmatrix}\begin{bmatrix}\mathbf{C}^\top & \mathbf{0}\\ \mathbf{0} & \mathbf{C}^\top\end{bmatrix}^{-1}\right]^{-1}\begin{pmatrix}\boldsymbol{\kappa}\\ \boldsymbol{W}\end{pmatrix}\\
&= \mathbf{C}\mathbf{A}_{\nu\cdot\kappa W}\begin{pmatrix}\boldsymbol{\kappa}\\ \boldsymbol{W}\end{pmatrix} = \mathbf{C}\mathbb{E}_{\nu}\left[\boldsymbol{\nu}|\boldsymbol{\kappa},\boldsymbol{W}\right].
\end{aligned}$$

Thus, the mean and the variance of the estimator conditionally on $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are:

$$\begin{aligned}
\mathbb{E}_Y\left[\hat{\boldsymbol{\beta}}|\boldsymbol{X}_1,\boldsymbol{X}_2\right], &= \begin{pmatrix}\mathcal{B}_{y\cdot 0(x_1x_zz)}\\ \mathcal{B}_{y\cdot x_1(z)}\\ \mathcal{B}_{y\cdot x_2(z)}\end{pmatrix} + \mathbf{C}\boldsymbol{\mu}_{\nu|\kappa W}\\
\mathbb{V}_Y\left[\hat{\boldsymbol{\beta}}|\boldsymbol{X}_1,\boldsymbol{X}_2\right] &= \left(\tilde{\boldsymbol{X}}^\top\boldsymbol{\Omega}^{-1}\tilde{\boldsymbol{X}}\right)^{-1} - \mathbf{J}\mathbf{C}\boldsymbol{\Sigma}_{\nu\cdot\kappa W}\boldsymbol{\Sigma}^{-1}_{\kappa,W}\boldsymbol{\Sigma}^\top_{\nu\cdot\kappa W}\mathbf{C}^\top\mathbf{J}^\top,
\end{aligned}$$

where $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_z + \sigma^2_{y|x_1,x_2,z}\mathbf{I}_n$.

Unfortunately, working with two covariates setting does not allow to compute the marginal sampling properties of the estimator with the expressions provided in Theorem 3.2.1. This may be a natural extension for a future work.

4.2 Several scenarios for structured DGP

After a brief review of the literature, we focus on the main aim of the thesis: the evaluation of confounding under different kind of structured generative mechanism. To this end,

[1] Let a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ be any matrix with rank $n \leq m$, the generalized inverse is $\mathbf{A}^- \in \mathbb{R}^{n\times m}$ such that $\mathbf{A}^- = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top$.

we fix the variance parameters σ_x and σ_z, and we let the structure vector parameter $\boldsymbol{\theta}$ setting up the covariance matrices and the correlation parameter ρ_{xz} differ. We focus on the study of the sampling properties of the ordinary least square estimator obtained by setting $\mathbf{S} = \mathbf{I}_n$ in (2.28). We have also employed the generalized least square implementing the optimal choice $\mathbf{S} = \boldsymbol{\Sigma}_{y|x}$. Since this did not lead to substantially different results, they were not reported.

Let us build up the structured DGP. Fixing the inferential target parameter $\mathcal{B}_{y\cdot x(z)} = 1$ and $\boldsymbol{\Sigma}_{y|x,z} = \sigma^2_{y|x,z}\mathbf{I}_n$ with $\sigma^2_{y|x,z} = 1$, we address situations in which confounding takes place. Thus, we assume $\mathcal{B}_{y\cdot z(x)} = 1$ and non-null cross-covariance between the covariate and confounder, i.e. $\boldsymbol{\Sigma}_{xz} \neq \mathbf{0}$ (see Definition 2.1.2). To explore how the choice concerning DGP changes the impact that confounding has on inference, we consider the following three cases which differ from each other by the type of joint covariance matrix $\boldsymbol{\Sigma}_{x,z}$ expressing the correlation structure of the main processes.

Case A: The following joint covariance matrix of $\boldsymbol{X}$ and $\boldsymbol{Z}$ is assumed:

$$\boldsymbol{\Sigma}_{x,z} = \begin{bmatrix} \mathbf{R}(\boldsymbol{\theta}_x) & \rho_{xz}\mathbf{R}(\boldsymbol{\theta}_{xz}) \\ \rho_{xz}\mathbf{R}(\boldsymbol{\theta}_{xz})^\top & \mathbf{R}(\boldsymbol{\theta}_z) \end{bmatrix}, \tag{4.10}$$

which is positive definite when $\mathbf{R}(\boldsymbol{\theta}_x) - \rho^2_{xz}\mathbf{R}(\boldsymbol{\theta}_{xz})\mathbf{R}(\boldsymbol{\theta}_z)^{-1}\mathbf{R}(\boldsymbol{\theta}_{xz}) \succ 0$. From a consequence of Definition A.0.1, this is equivalent to requiring

$$\frac{\boldsymbol{v}^\top\mathbf{R}(\boldsymbol{\theta}_x)\boldsymbol{v}}{\boldsymbol{v}^\top\mathbf{R}(\boldsymbol{\theta}_{xz})\mathbf{R}(\boldsymbol{\theta}_z)^{-1}\mathbf{R}(\boldsymbol{\theta}_{zx})\boldsymbol{v}} \geq \rho^2_{xz} \qquad \forall \boldsymbol{v} \neq \mathbf{0}$$
$$\frac{\tilde{\boldsymbol{v}}^\top\mathbf{R}(\boldsymbol{\theta}_{xz})^{-1}\mathbf{R}(\boldsymbol{\theta}_x)\mathbf{R}(\boldsymbol{\theta}_{zx})^{-1}\mathbf{R}(\boldsymbol{\theta}_z)\tilde{\boldsymbol{v}}}{\tilde{\boldsymbol{v}}^\top\tilde{\boldsymbol{v}}} \geq \rho^2_{xz} \qquad \forall \tilde{\boldsymbol{v}} \neq \mathbf{0},$$

where $\tilde{\boldsymbol{v}} = \mathbf{R}(\boldsymbol{\theta}_z)^{-1/2}\mathbf{R}(\boldsymbol{\theta}_{zx})\boldsymbol{v}$. Since the left-hand term in the last inequality is a Rayleigh quotient, it can be observed that

$$\lambda_{\min}\left(\mathbf{R}(\boldsymbol{\theta}_{xz})^{-1}\mathbf{R}(\boldsymbol{\theta}_x)\mathbf{R}(\boldsymbol{\theta}_{zx})^{-1}\mathbf{R}(\boldsymbol{\theta}_z)\right) \leq \frac{\tilde{\boldsymbol{v}}^\top\mathbf{R}(\boldsymbol{\theta}_{xz})^{-1}\mathbf{R}(\boldsymbol{\theta}_x)\mathbf{R}(\boldsymbol{\theta}_{zx})^{-1}\mathbf{R}(\boldsymbol{\theta}_z)\tilde{\boldsymbol{v}}}{\tilde{\boldsymbol{v}}^\top\tilde{\boldsymbol{v}}}.$$

Hence, the sufficient condition to ensure the positive definiteness of (4.10) is

$$\lambda_{\min}\left(\mathbf{R}(\boldsymbol{\theta}_{xz})^{-1}\mathbf{R}(\boldsymbol{\theta}_x)\mathbf{R}(\boldsymbol{\theta}_{zx})^{-1}\mathbf{R}(\boldsymbol{\theta}_z)\right) \geq \rho^2_{xz}. \tag{4.11}$$

In this case, the cross-covariance matrix is parameterized independently on the parameters characterizing the marginal covariance matrices of $\boldsymbol{X}$ and $\boldsymbol{Z}$. This construction is intended to show that, as expected form Theorem 3.0.1, the bias relies on the structure of $\boldsymbol{Z}$ only when the cross-covariance matrix is obtained as function of $\boldsymbol{\Sigma}_z$ by construction. Under this setting, we obtain a deterministic conditional bias as in (3.31), when $\boldsymbol{\theta}_x = \boldsymbol{\theta}_{zx}$.

Case B: In order to explore the situation in which the covariate varies with two different structure parameters, let $\boldsymbol{X} = \boldsymbol{\xi} + \boldsymbol{\psi}$ be decomposed into a component, $\boldsymbol{\psi}$, that has the same structure parameters of the confounder $\boldsymbol{Z}$, and a component with different structure, $\boldsymbol{\xi}$, which is independent from $\boldsymbol{\psi}$ and $\boldsymbol{Z}$. In this way, using a LMC, we consider the setting implemented by Paciorek (2010) expressed in (4.4) in which $\sigma_\xi^2 = \sigma_\psi^2 = \sigma_z^2/2$. This leads to the joint covariance matrix:

$$\boldsymbol{\Sigma}_{x,z} = \sigma_z^2 \begin{bmatrix} \dfrac{\mathbf{R}(\boldsymbol{\theta}_\psi) + \mathbf{R}(\boldsymbol{\theta}_\xi)}{2} & \dfrac{\rho_{xz}}{\sqrt{2}}\mathbf{R}(\boldsymbol{\theta}_\psi) \\ \dfrac{\rho_{xz}}{\sqrt{2}}\mathbf{R}(\boldsymbol{\theta}_\psi) & \mathbf{R}(\boldsymbol{\theta}_\psi) \end{bmatrix} \tag{4.12}$$

which is positive definite $\forall \rho_{xz} \in (-1, 1)$. In this setup, the regression matrix is:

$$\mathbf{A}_{z\cdot x} = \frac{\rho_{xz}}{\sqrt{2}} \left[\mathbf{R}(\boldsymbol{\theta}_\xi)\mathbf{R}(\boldsymbol{\theta}_\psi)^{-1} + \mathbf{I}_n\right]^{-1}. \tag{4.13}$$

Note that also in this case a deterministic conditional bias expressed as in (3.31) is provided when $\boldsymbol{\theta}_\xi = \boldsymbol{\theta}_\psi$ (in agreement with Paciorek (2010)).

Case C: This case refers to the revised LMC in which $\sigma_x^2 = \sigma_z^2 = 1$, that can generate asymmetric cross-covariance. The joint covariance matrix is

$$\begin{bmatrix} \mathbf{R}(\boldsymbol{\theta}_x) & \rho_{xz}\mathbf{R}(\boldsymbol{\theta}_x)^{1/2}\left(\mathbf{R}(\boldsymbol{\theta}_z)^{1/2}\right)^\top \\ \rho_{xz}\mathbf{R}(\boldsymbol{\theta}_z)^{1/2}\left(\mathbf{R}(\boldsymbol{\theta}_x)^{1/2}\right)^\top & \mathbf{R}(\boldsymbol{\theta}_z) \end{bmatrix}, \tag{4.14}$$

and regression matrix $\mathbf{A}_{z\cdot x} = \rho_{xz}\mathbf{R}(\boldsymbol{\theta}_z)^{1/2}\mathbf{R}(\boldsymbol{\theta}_x)^{-1/2}$. The matrix $\mathbf{R}^{1/2}$ is the lower triangle from Cholesky decomposition, thus $\mathbf{R}^{1/2}$ can be indicated with $\mathbf{L}$. For future studies it might be interesting to use $\mathbf{R}^{1/2} = \mathbf{U}\boldsymbol{\Lambda}^{1/2}$ provided by the spectral decomposition $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$. This assumption allows to relate the eigenvalues of $\boldsymbol{\Sigma}$ and $\mathbf{R}^{1/2}$ simplifying, for example, the computation of bias bounds planned for future research work. Besides, according to Page et al. (2017), we get the deterministic bias expressed in (3.31) when the structure parameters of the main processes are the same.

We remark that in each case all $\mathbf{R}$ marginal covariance matrices in the $\boldsymbol{\Sigma}_{x,z}$ are derived under the same covariance function. Next, for Case A-C, we build marginal matrices adopting the Matérn covariance function, the first order autoregressive process and the conditional autoregressive process. In every framework and for each combination of parameters $\boldsymbol{\theta}_\xi$, $\boldsymbol{\theta}_\psi$, ρ_{xz}, we compute the marginal sampling properties of the estimator $\hat{\beta}_x$ using the fundamental analytic results obtained in Chapter 3.

4.2.1 Matérn covariance function

The point-referenced analysis is carried out by considering a regular grid of 100 spatial units that are located on the unit square. We employ the Matérn covariance function introduced in Section 3.5.1 with $\nu = 0.5$ i.e. the exponential correlation function characterized by the structure vector parameter $\boldsymbol{\theta} = (0.5, r)^\top$. We adopt a sequence of values $r \in (0, 1)$ for both underlying processes, $\boldsymbol{\xi}$ and $\boldsymbol{\psi}$, that contributes to construct $\boldsymbol{\Sigma}_{x,z}$ (see Section 2.1.2.1).

Case A: Let $\boldsymbol{X} = \boldsymbol{\xi}$ and $\boldsymbol{Z} = \boldsymbol{\psi}$ be the main processes in DGP. Free from restriction due to the choice of cross-covariance function depending on the structure parameters of the main processes, we adopt a cross-covariance matrix deriving from exponential function with structure parameter r_{zx}. As already anticipated, this case can generate non-positive definite $\boldsymbol{\Sigma}_{x,z}$ if inequality (4.11) does not hold.

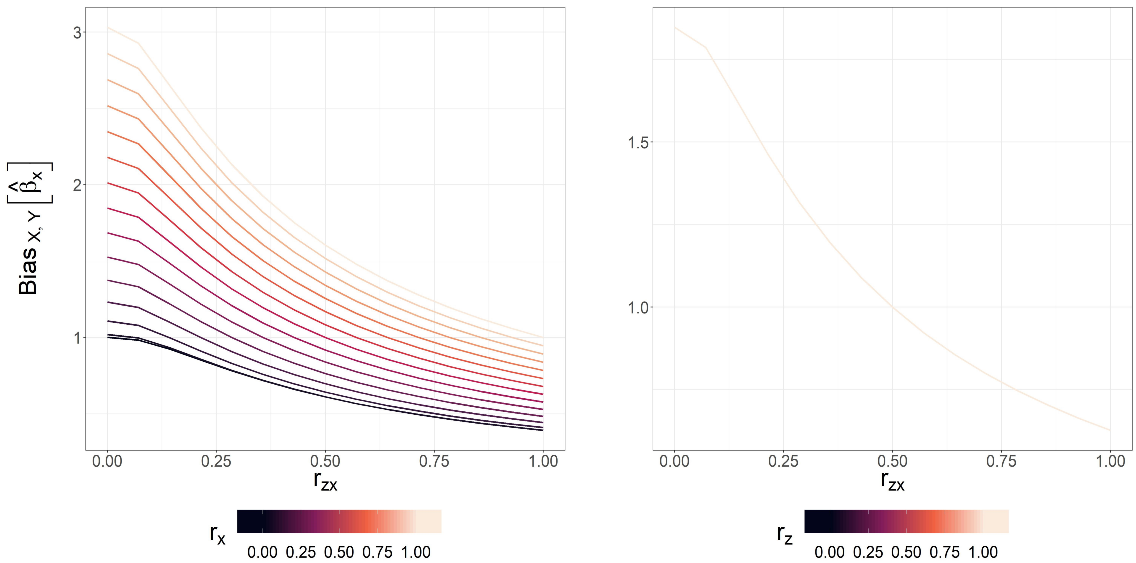

Figure 4.5: Marginal bias of $\hat{\beta}_x$ decay as the range parameter of the cross-covariance matrix r_{zx} increases respect to r_x (left) and r_z (right) (geostatistical data, case A).

Figure 4.5 (left) shows that increasing the range parameter of the cross-covariance matrix while holding the one of $\boldsymbol{X}$ constant produces a reduction in the bias. In particular, the decrease becomes steeper for higher r_x values. It seems to be the consequence of the product matrix characterizing the quadratic form of conditional bias in Theorem 3.0.1, $\boldsymbol{\Sigma}_{zx}\boldsymbol{\Sigma}_x^{-1}$. According to the theory, the bias increases when the expected sampling covariance of $\boldsymbol{X}$ and $\boldsymbol{Z}$ raises or the variability (smoothness) of the covariate decreases (increases). Indeed, considering the first order of Laplace approximation of the estimator bias E_L defined

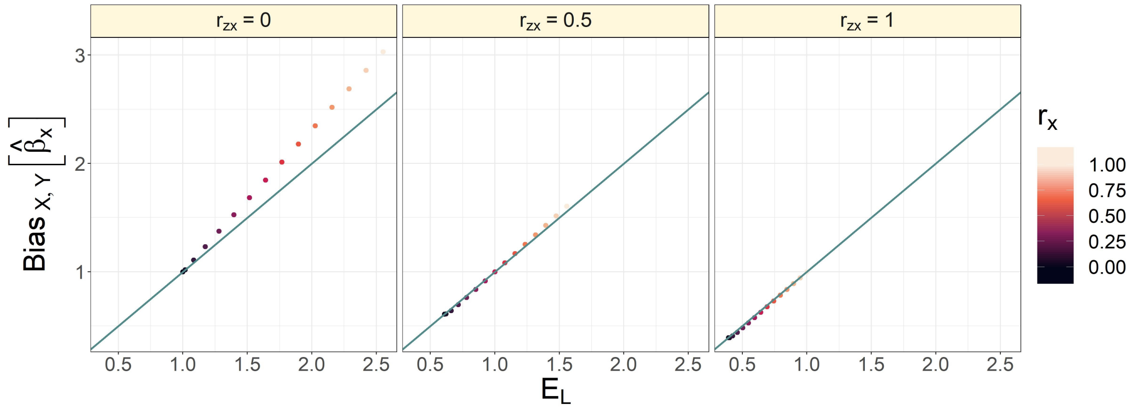

Figure 4.6: Linear relationship between the Laplace approximation of the estimator bias E_L and its exact value respect to $r_x \in (0, 1)$ and $r_{zx} \in \{0, 0.5, 1\}$. The green line is the intercept used to underline the strong relation between quantities.

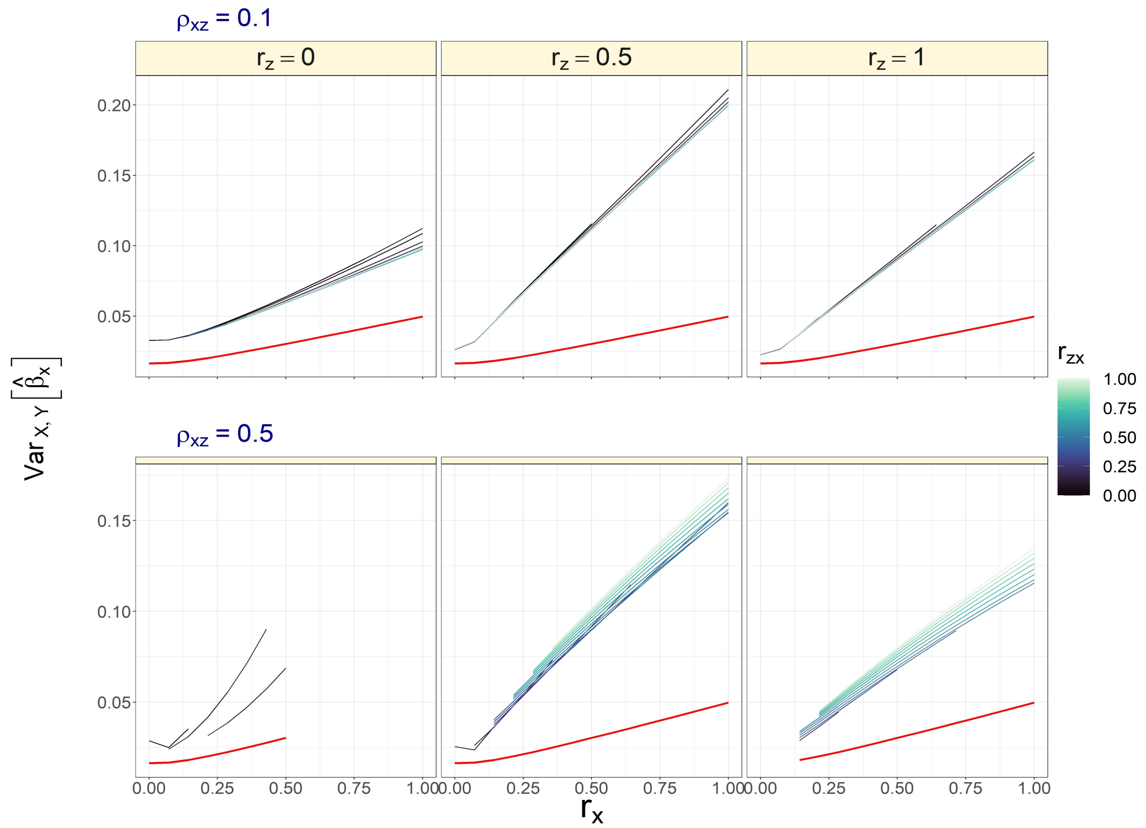

Figure 4.7: Marginal variance trends of $\hat{\beta}_x$ under positive definite conditions for $\mathbf{\Sigma}_{x,z}$ and for $\rho_{xz} \in \{0.1, 0.5\}$.

in (3.42) a meaningful proxy of the bias, Figure 4.6 illustrates the linear relationship between it and the exact bias as r_{zx} and r_x vary. In addition, it is interesting to underline that the confounding appears to depend upon the covariate smoothness and not that of confounder. Indeed, as it is evident from Figure 4.5 (right), it is crucial to highlight that the structure of confounder does not affect the bias. To our knowledge, it is a result that is in contrast with what has been claimed in the literature. These misguided conclusions are caused by the fact that to date the cross-covariance function usually has been assumed in dependence of the structure parameters defining the main processes in the DGP.

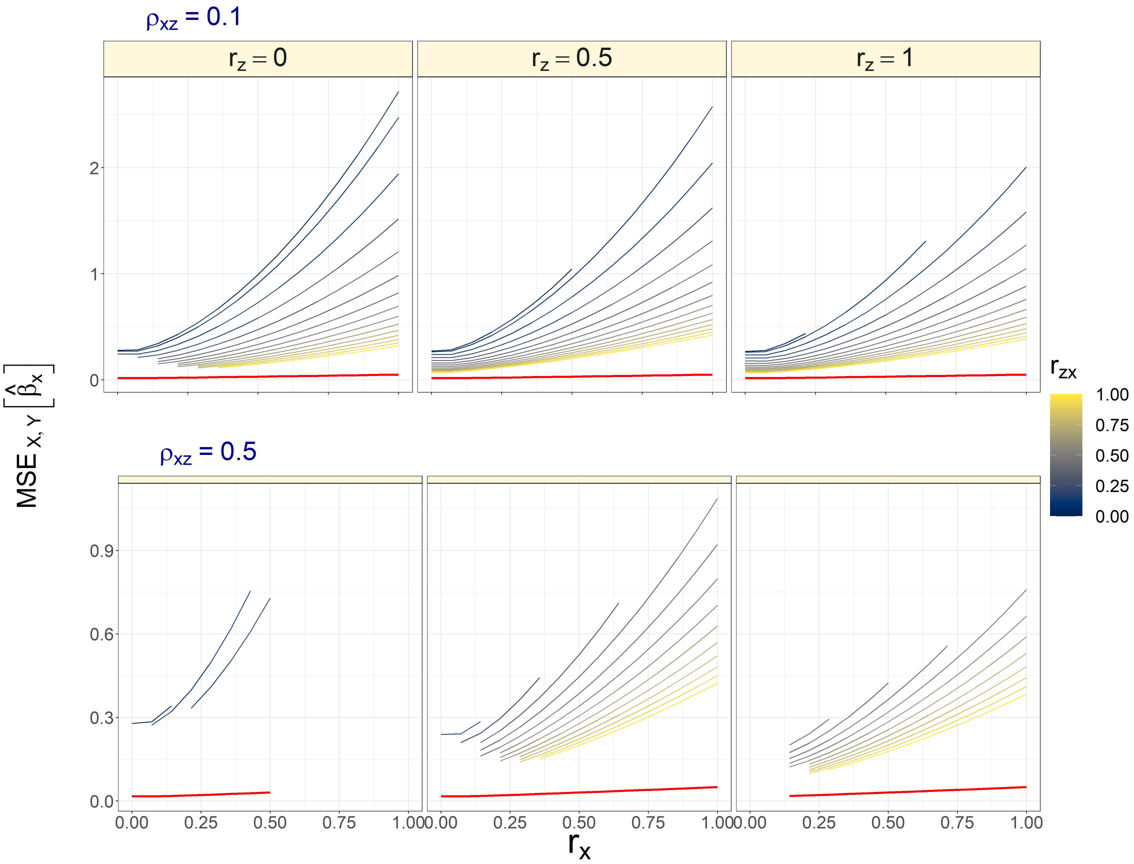

Figure 4.8: Marginal mean square error of $\hat{\beta}_x$ under positive definite conditions for $\mathbf{\Sigma}_{x,z}$ and $\rho_{xz} \in \{0.1, 0.5\}$.

On the other hand, as expected from Formula (3.24), the estimator variance and mean square error are also influenced by the confounder structure. Figures 4.7 and 4.8 show the estimator variance and MSE considering only the range parameter values that lead to positive definite joint covariance of the covariate and confounder when $\rho_{xz} \in \{0.1, 0.5\}$. It is no surprising that as the correlation parameter increases there are less combinations of range parameters that produce valid $\boldsymbol{\Sigma}_{x,z}$. The red curve in Figure 4.7 (4.8) represents the trend of $\mathbb{E}_X\left[\mathbb{V}_Y\left[\hat{\beta}_x|\boldsymbol{X}, \mathcal{B}_{y\cdot z(x)} = 0, \boldsymbol{A}_{z\cdot x} = \boldsymbol{0}\right]\right]$, i.e. the estimator variance (MSE) when there is no omitted variable $\boldsymbol{Z}$ in the model, denoted in (3.24) and exclusively dependent on the covariate process. Additionally, as axpected from (3.24), Figure 4.7 highlights that the variance is always inflated when there is an omitted variable correlated with the observed one. How it rises is mainly related to the type and strength of correlation (r_{zx} and ρ_{xz}) that links the main processes. Furthermore, it might seems that the magnitude of the variance and the MSE increases as ρ_{xz} decreases but in a different way to the variation of r_{zx}. About the variance in Figure 4.7, we note that as ρ_{xz} decreases r_{zx} loses its relevance. Instead, the bias is proportional to the correlation parameter, thus it increases as ρ_{xz} increases regardless of other parameters (not shown because it follows by Formula (3.1)). Moreover, fixing r_z whose changing seems minor matter, Figure 4.9 compares the part of the estimator sampling properties that are confounding-dependent (see Equations (3.27) and (3.28)) showing their trend as the other range parameters increase. It comes into view that for all quantities the values rise considerably faster when the main processes' co-variability is higher.

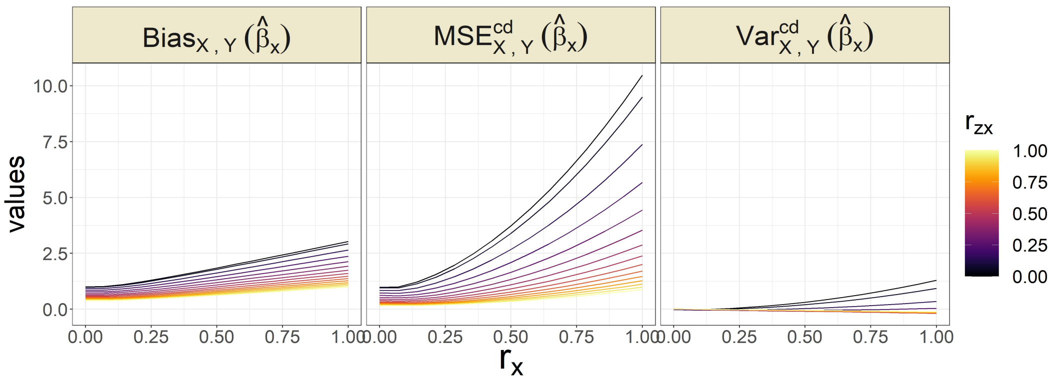

Figure 4.9: Comparison of the part of marginal bias, mean square error and variance varying when confounding occurs in geostatistical setting, case A. Note that the whole bias is confounding-dependent.

Finally, Table 4.1 emphasizes that the estimator sampling properties change also in dependence of the expected variability of all processes, $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{Z}$ (see Section 3.5). In particular, since the process variability is inversely proportional to the level of smoothness in the MF case, such investigated quantities depend also on the processes smoothness. To date this fact has not been taken into account, but we would like to study the behavior of the target quantities, for example, fixing the sampling variance of the response process.

r_x	r_z	r_{zx}	$\text{Bias}_{Y,X}\left[\hat{\beta}_x\right]$	$\mathbb{V}_{Y,X}\left[\hat{\beta}_x\right]$	$\text{MSE}_{Y,X}\left[\hat{\beta}_x\right]$	EV_x	EV_z	EV_y
0	0	0	0.10	0.03	0.04	1.00	1.00	5.00
0.5	0	0	0.18	0.06	0.10	0.61	1.00	4.61
1	0	0	0.30	0.11	0.20	0.39	1.00	4.39
0	0.5	0	0.10	0.03	0.04	1.00	0.61	4.61
0.5	0.5	0	0.18	0.12	0.15	0.61	0.61	4.22
1	0.5	0	0.30	0.21	0.31	0.39	0.61	4.00
0	1	0	0.10	0.02	0.03	1.00	0.39	4.39
0.5	1	0	0.18	0.09	0.13	0.61	0.39	4.00
1	1	0	0.30	0.18	0.27	0.39	0.39	3.78
0	0	0.5	0.06	0.03	0.04	1.00	1.00	4.22
0.5	0	0.5	0.10	0.06	0.07	0.61	1.00	3.83
1	0	0.5	0.16	0.10	0.12	0.39	1.00	3.61
0	0.5	0.5	0.06	0.03	0.03	1.00	0.61	3.83
0.5	0.5	0.5	0.10	0.11	0.12	0.61	0.61	3.44
1	0.5	0.5	0.16	0.20	0.23	0.39	0.61	3.22
0	1	0.5	0.06	0.02	0.03	1.00	0.39	3.61
0.5	1	0.5	0.10	0.09	0.10	0.61	0.39	3.22
1	1	0.5	0.16	0.16	0.19	0.39	0.39	3.00
0	0	1	0.04	0.03	0.03	1.00	1.00	3.78
0.5	0	1	0.06	0.06	0.06	0.61	1.00	3.39
1	0	1	0.10	0.10	0.11	0.39	1.00	3.18
0	0.5	1	0.04	0.03	0.03	1.00	0.61	3.39
0.5	0.5	1	0.06	0.11	0.12	0.61	0.61	3.00
1	0.5	1	0.10	0.20	0.21	0.39	0.61	2.78
0	1	1	0.04	0.02	0.02	1.00	0.39	3.18
0.5	1	1	0.06	0.09	0.09	0.61	0.39	2.78
1	1	1	0.10	0.16	0.17	0.39	0.39	2.57

Table 4.1: Marginal sampling properties of $\hat{\beta}_x$ and variability of processes $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{Z}$ for a subset of scenarios in which $r_x, r_z, r_{zx} \in \{0, 0.5, 1\}$ and $\rho_{xz} = 0.1$ (geostatistical data, case A).

Case B: Since in this case the covariate process is the sum of two underlying processes, $\boldsymbol{\xi}$ and $\boldsymbol{\psi}$, it is interesting to note how the expected sampling variance of $\boldsymbol{X}$, EV_x, varies with respect to the range parameters r_ξ and r_ψ. In this regard, Figure 4.10 (left panel) displays how the covariate variability depends on both structure parameters of the underlying processes: it decreases as both increase.

In this case, the expected sampling covariance, EV_{zx}, is equivalent to the confounder variability that is function of the sole r_ψ whose distinctive trend is shown in Figure 3.2. Since we are using an OLS estimator, these two quantities identify the Laplace approximation of the bias E_L presented in Figure 4.10

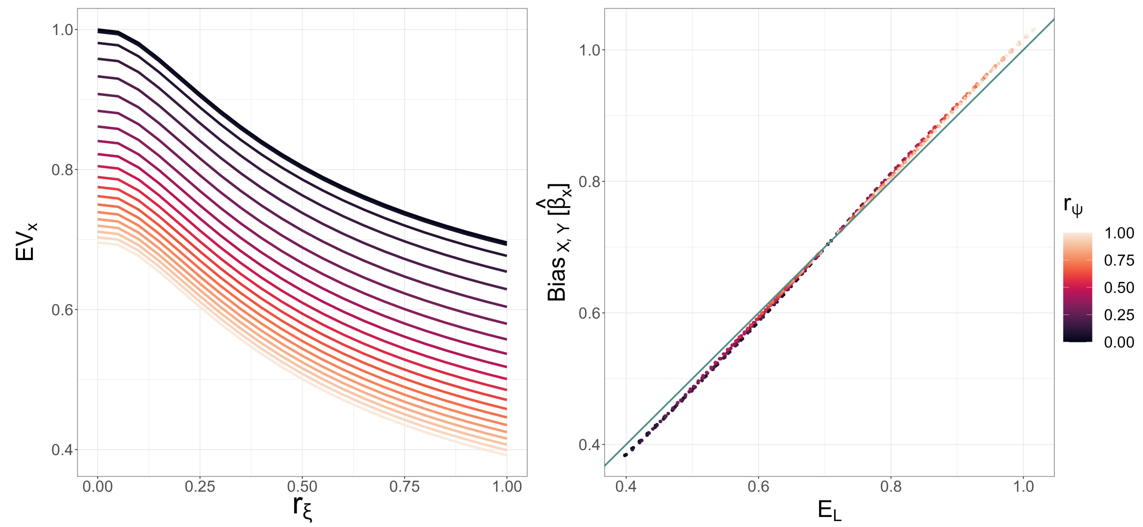

Figure 4.10: On the left it is shown the decay of the covariate variability, EV_x, as a function of the range parameter r_ξ and r_ψ and, on the right, the relationship between the first order Laplace approximation of the estimator bias and its exact value with respect to confounder range parameter (geostatistical data, case B).

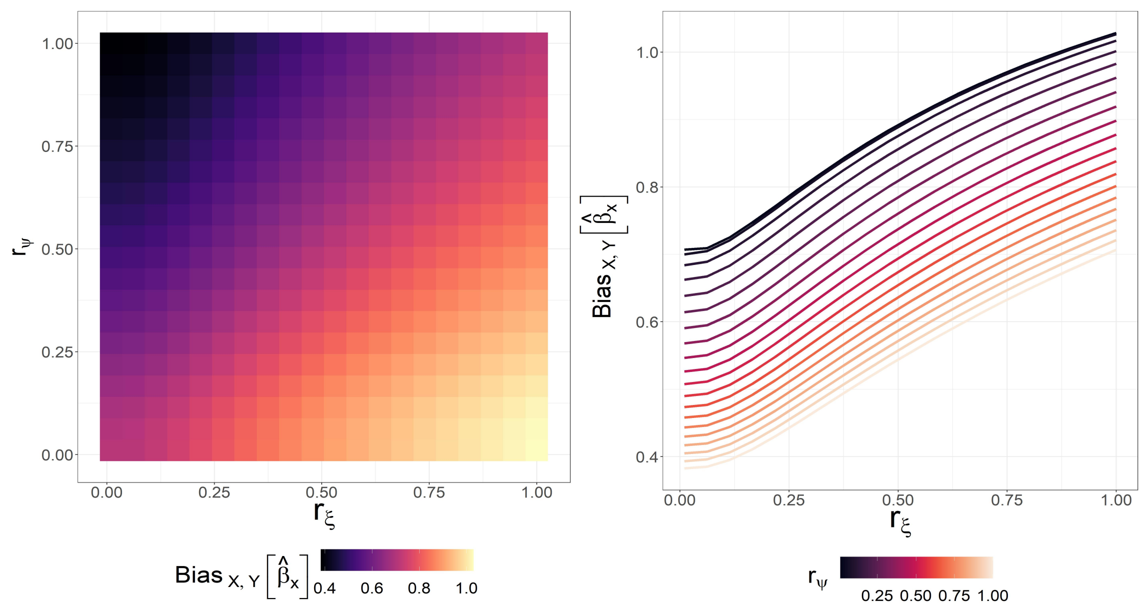

Figure 4.11: Estimator bias from two different perspectives as the range parameter r_ξ and r_ψ vary.

which confirms to be a good proxy. Furthermore, Figures 4.11 and 4.12 show the pattern of the estimator bias and variance, respectively, from two different perspectives. In particular, from both plots in Figure 4.11 it is evident that the bias increases as r_ξ increases and r_ψ decreases. According to Paciorek (2010),

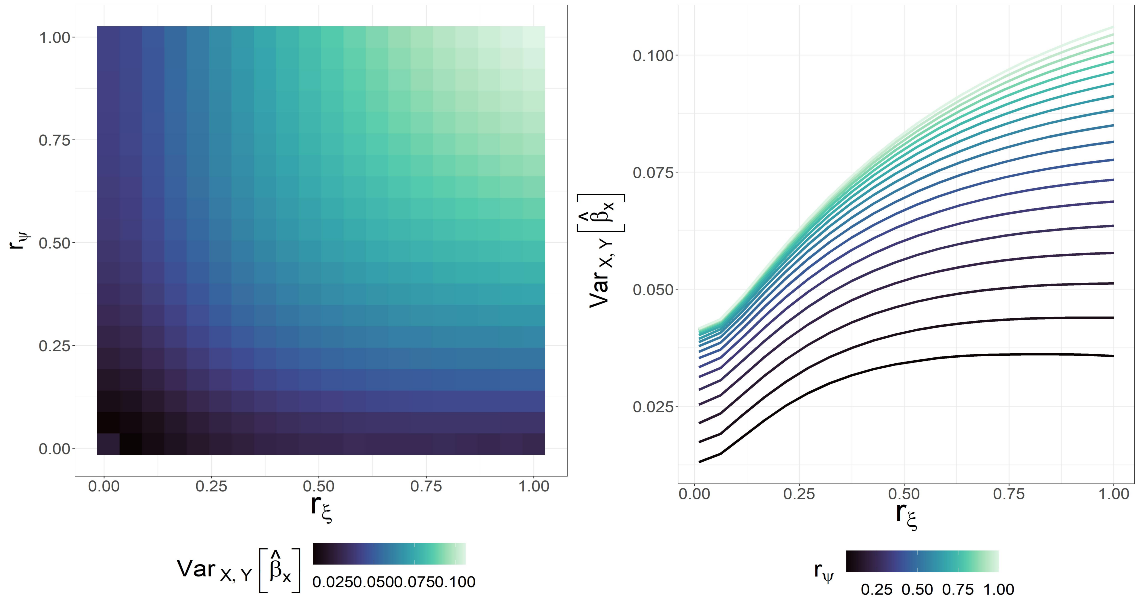

Figure 4.12: Estimator variance from two different perspectives as the range parameter r_ξ and r_ψ vary.

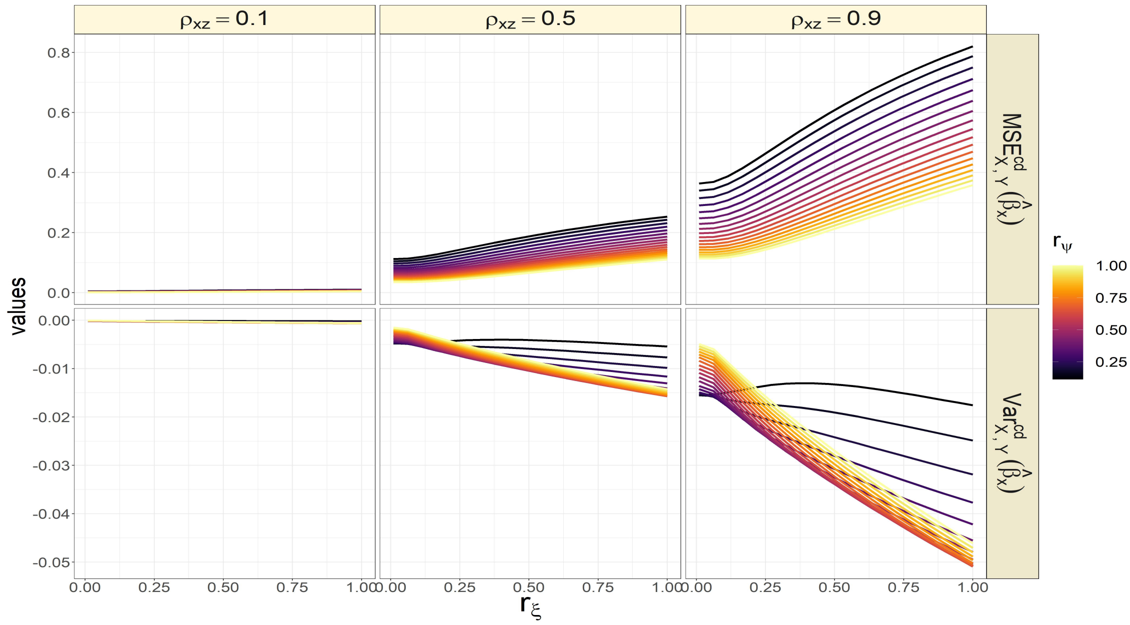

Figure 4.13: Comparison of the part of mean square error and variance varying when confounding occurs in geostatistical setting, case B.

when $r_\psi \ll r_\xi$ do we see more bias than in the case in which $r_\psi = r_\xi$ (deterministic bias (3.31)). In this case of MF covariance function it is equivalent to request $EV_\xi \ll EV_\psi$ as seen in (4.6). In turn, it means $EV_x \ll EV_{zx}$.

Additionally, Figure 4.12 allows to notice that the estimator variance increases when both range parameters increase. Then, concerning the part of sampling properties that depends on confounding, Figure 4.13 shows how the variance is negligible but still has its decreasing pattern as both range parameters increase and the MSE, increases as ρ_{xz} increases regardless of other parameters values.

Case C: In line with Page et al. (2017), plots in Figure 4.14 show from two different point of view that the bias increases as the covariate range parameter increases and the confounder range parameter decreases.

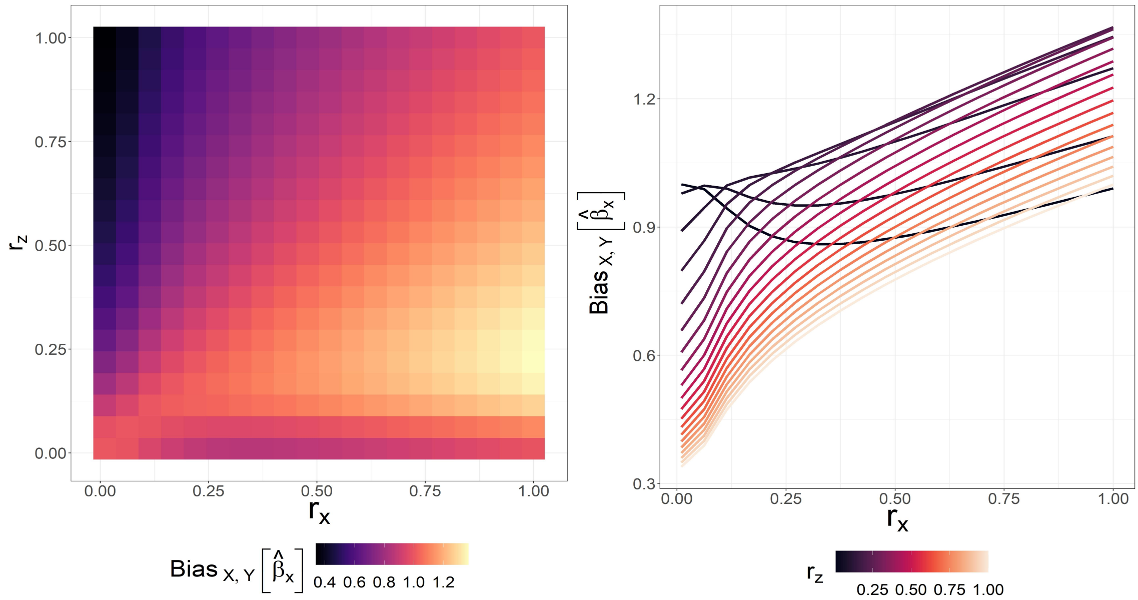

Figure 4.14: Estimator bias from two different perspectives as the range parameter r_x and r_z vary .

As above, an explanation for this might be related the regression matrix $\mathbf{A}_{z \cdot x}$ employed in DGP featured the case (4.14). Indeed, it supports our idea about the fact that the estimator bias is very closely related to the eigenvalues of $\mathbf{\Sigma}_{zx}$ and $\mathbf{\Sigma}_x$. Hence, it is linked to the spectrum of triangle matrices obtained by Cholesky decomposition used to achieve matrix defined in (4.14). In Figure 4.15 it is shown the relation between the approximated bias E_L and its exact value that takes into account the eigenvalues of the matrices in $\mathbf{A}_{z \cdot x}$. This time, the match is less noticeable, especially for smaller value of confounder parameter. Lastly, Figure 4.16 displays how the MSE and variance depending on confounding change over $\rho_{xz} \in \{0.1, 0.5, 0.9\}$.

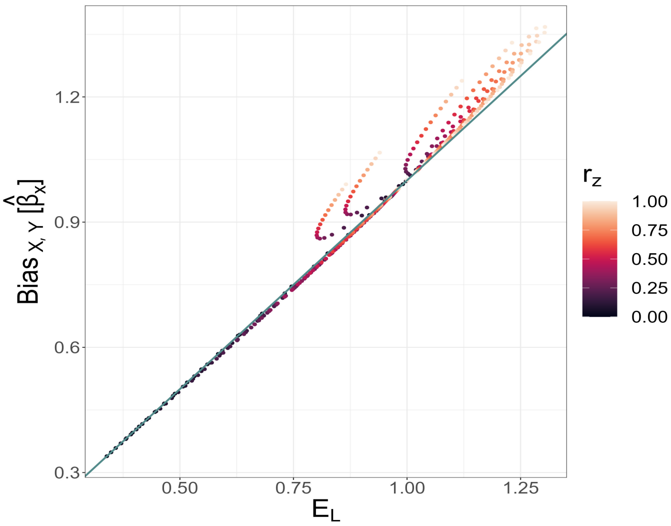

Figure 4.15: The relationship between the Laplace approximation and the exact estimator bias respect to r_z.

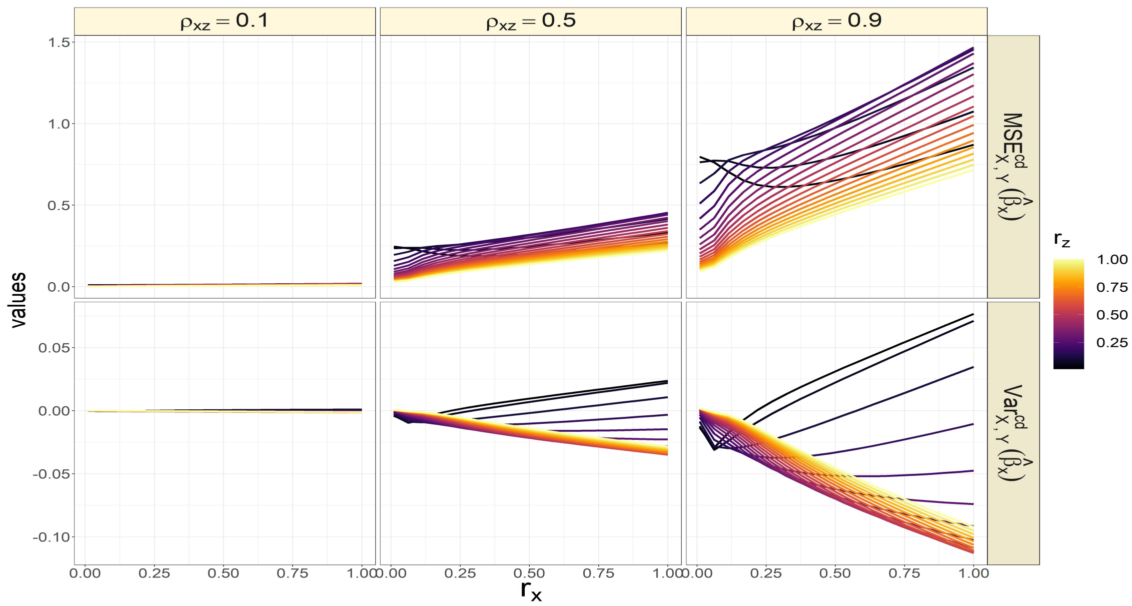

Figure 4.16: Comparison of the part of mean square error and variance varying when confounding takes place in geostatistical setting, case C.

4.2.2 Conditional Autoregressive process

In an areal spatial framework, to see how different neighborhood structures might influence the estimator sampling properties, we examine also the areal data modeling setup introduced in Section 3.5.3 and used by Page et al. (2017) and Nobre et al. (2021). We assume that the underlying processes $\boldsymbol{\xi}$ and $\boldsymbol{\psi}$ follow a conditional autoregressive model. Our geographical region contains the boundaries of Missouri's 115 counties. To ensure that the covariance matrices are positive definite we restrict the structure parameter $\kappa \in [-0.34, 0.16]$ where $\mathbf{W}$ is the neighbouring matrix of Missouri. Also this time, the results are presented divided for cases listed at the beginning of the current section.

Case A: Figure 4.17 presents the trend of the estimator bias of $\hat{\beta}_x$ as the correlation parameter of the cross-covariance matrix κ_{zx} increases respect to the range of κ_x from two perspectives. The 3D version, on the right side, clarify the bias pattern on the left. Here the curves are justified by the variability trend that characterizes the CAR process (see Figure 3.4) used to build the main processes.

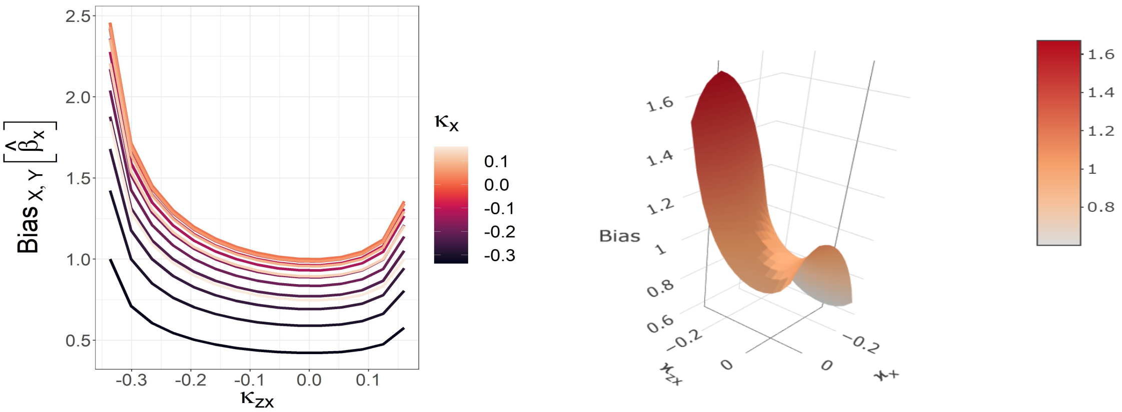

Figure 4.17: Marginal bias of $\hat{\beta}_x$ as the range parameter of the cross-covariance matrix κ_{zx} increases respect to κ_x from two perspectives in areal setting, case A.

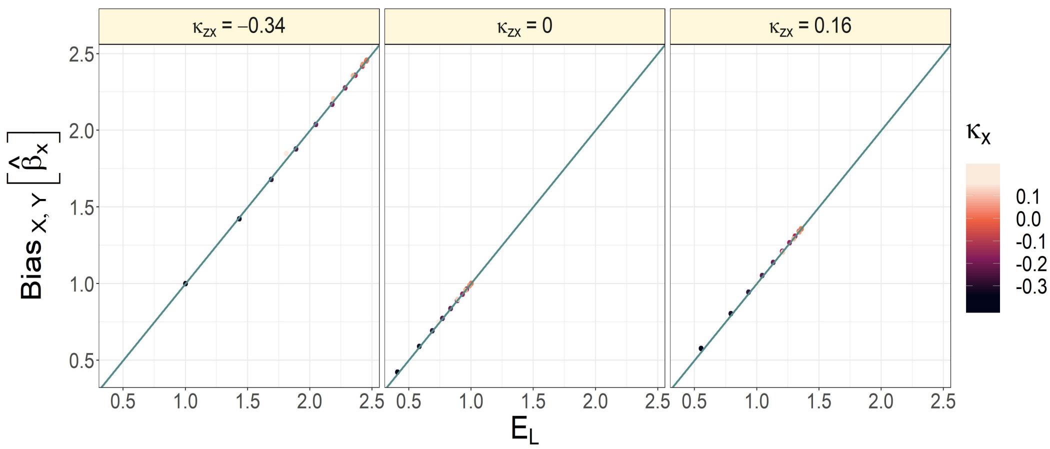

Figure 4.18: Laplace approximation of the estimator bias vs its exact value respect to κ_x and $\kappa_{zx} \in \{-0.34, 0, 0.16\}$.

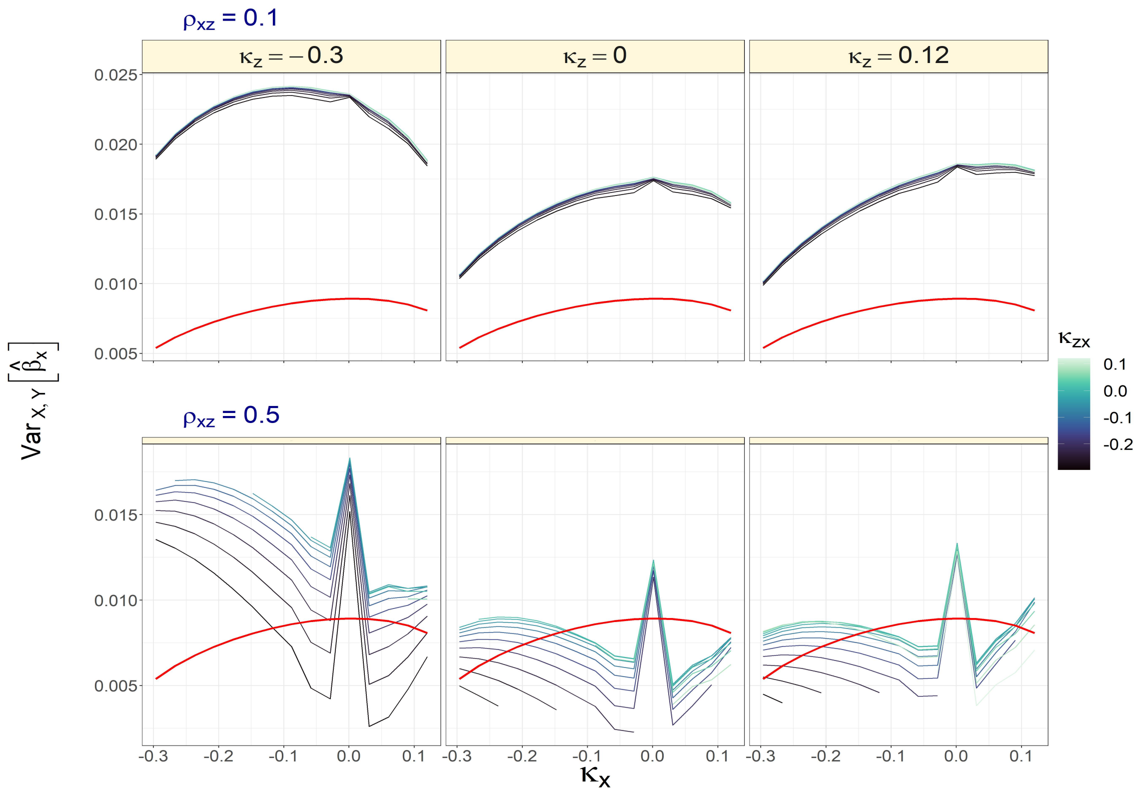

Figure 4.19: Marginal estimator variance under positive definite conditions for $\mathbf{\Sigma}_{x,z}$ and $\rho_{xz} \in \{0.1, 0.5\}$. The red curve indicates the variance when there is no omitted variable $\boldsymbol{Z}$ in the model.

By looking at Figure 4.18, it emerges the accuracy of first order Laplace approximation to the bias. It confirms what ensured above, the trend that can be seen in Figure 4.17 relies on the expected sampling covariance of the covariate and confounder modified for the effect of covariate variability. With respect to the estimator variance, it seems that κ_x and κ_z produce large changes. As ρ_{xz} decreases the impact of κ_{zx} on the variance decreases, instead for higher ρ_{xz} the variance seems to reduce drastically (Figure 4.19). Regarding the mean square error (not shown), it is dominated by the bias, especially for high values of the correlation parameter.

Case B: From Figure 4.20 we can see how the estimator bias depends on both structure parameters of the underlying processes however it is not so clear how the behaviour of $\boldsymbol{X}$ and $\boldsymbol{Z}$ influences the confounding if we focus on such structure parameters. For this reason, we display Figure 4.21. Through the bias approximation E_L it illustrates that the co-variability of the main processes and the variability of the covariate are the most relevant quantities that allow to explain confounding.

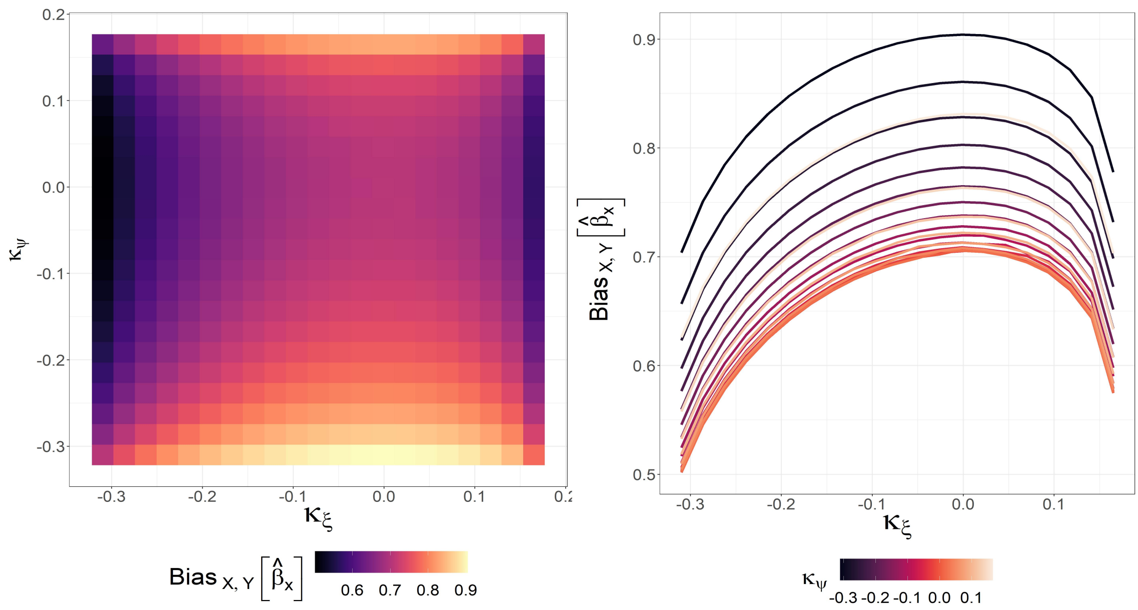

Figure 4.20: Estimator bias from two different perspectives as the range parameter κ_ξ and κ_ψ vary.

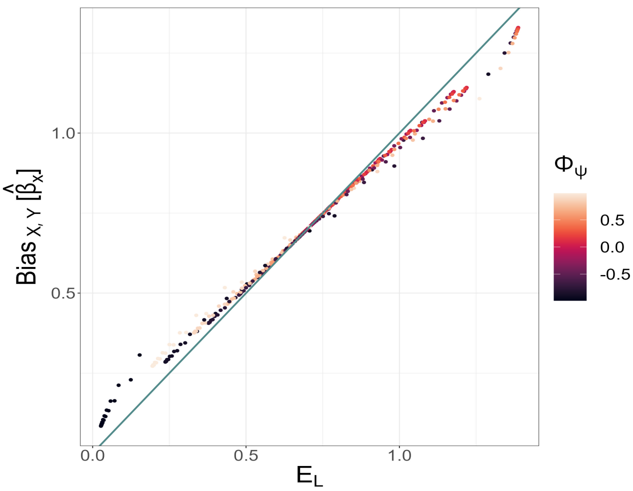

Figure 4.21: The accuracy of first order Laplace approximation to the bias respect to confounder range parameter.

Case C: Plots in Figure 4.22 illustrate the bias trend as the range parameters κ_x and κ_z vary. It appears to be smaller when the variability of $\boldsymbol{X}$ is higher (namely to the extremes of the κ_x range) and the one of $\boldsymbol{Z}$ is lower. This is affine with what it is shown in Figure 4.23 via bias approximation E_L. In the regard of a generic case C independently from the kind of covariance function used, it is important to emphasize that the confounder structure parameter influences the

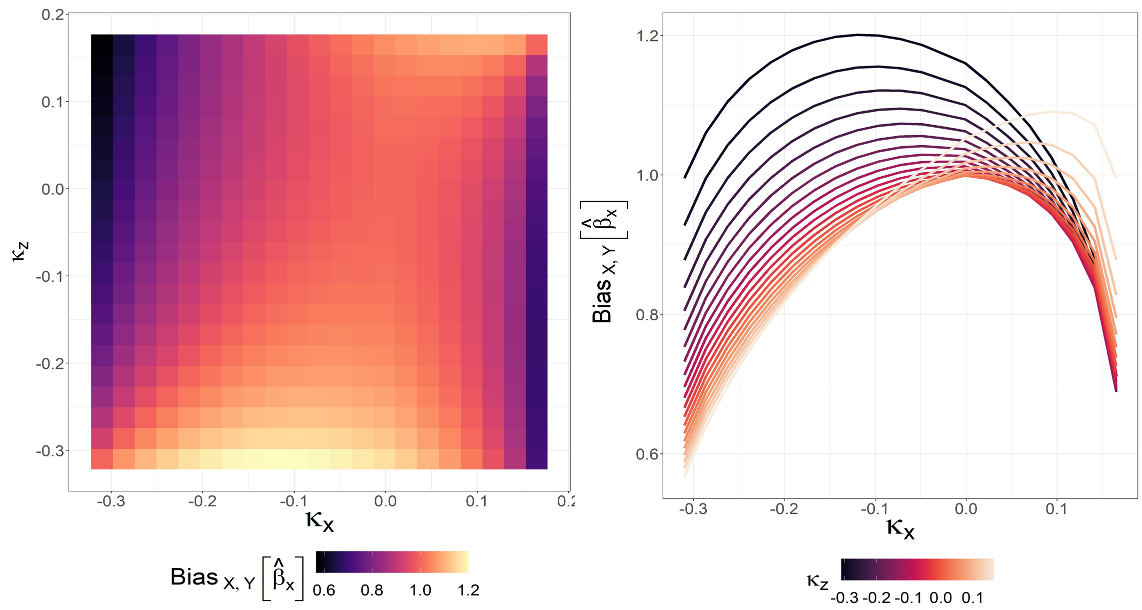

Figure 4.22: Estimator bias from two different angles as the range parameter κ_x and κ_z vary.

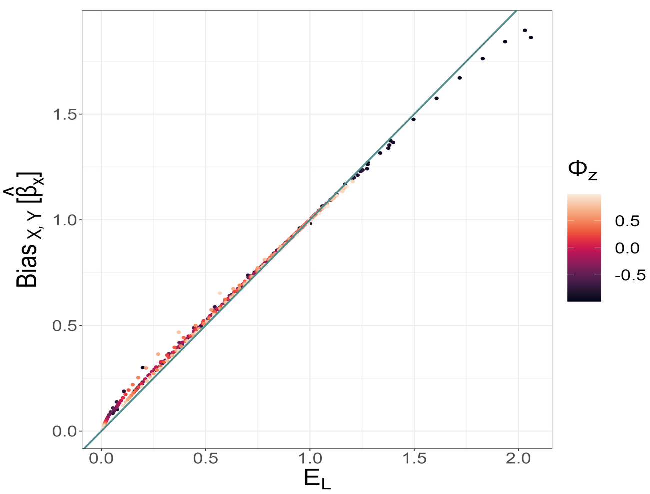

Figure 4.23: The accuracy of first order Laplace approximation to the bias respect to κ_z.

bias just because, in this specific case, it contributes to the construction of the cross-covariance matrix of $\boldsymbol{X}$ and $\boldsymbol{Z}$. In addition, Figure 4.24 displays how the MSE and variance depending on confounding changes over $\rho_{xz} \in \{0.1, 0.5, 0.9\}$. Regarding the mean square error, it increases with the correlation parameter, instead the variance decreases. In particular, as ρ_{xz} increases both quantities are more affected by the variation of κ_z.

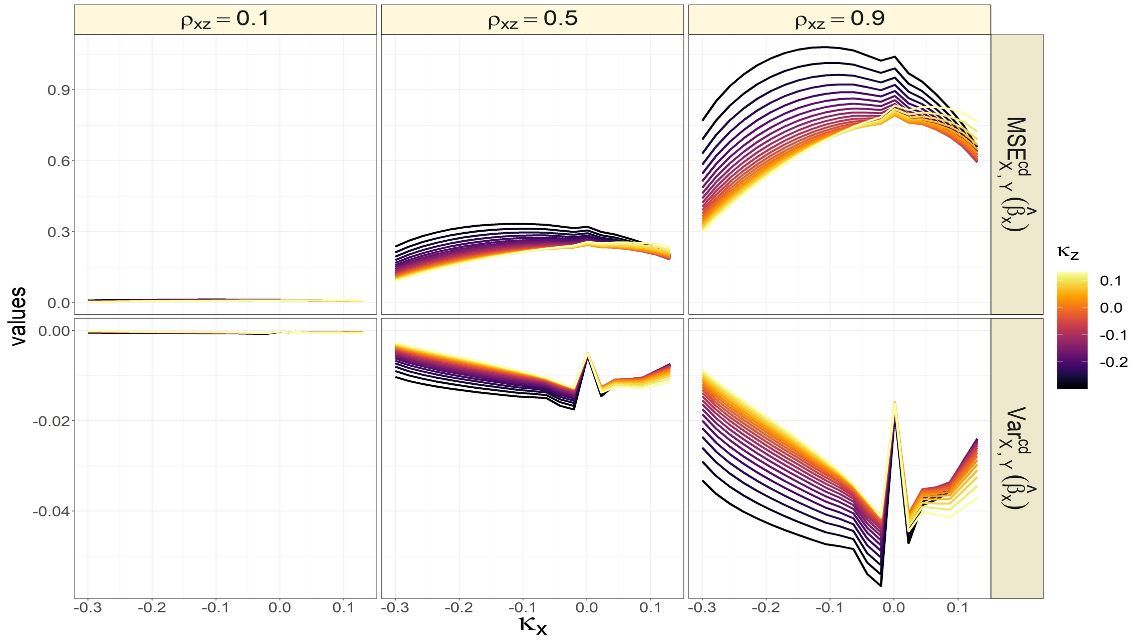

Figure 4.24: Comparison of the part of mean square error and variance varying when confounding occurs assuming case C for areal data as the correlation parameter $\rho_{xz} \in \{0.1, 0.5, 0.9\}$ changes.

4.2.3 Autoregressive process of order 1

The previous two sections have revealed that it is possible to get an overall insight about confounding focusing on the variability and co-variability characterizing the main processes even facing with different kind of data in the spatial field. Now, having in mind the idea that the confounding issue it is not related just to the spatial setting but it can be generalized to every field, we discuss a case relevant to temporal data. Hence, we consider the case where the processes building up the DGP are temporally auto-correlated (see Section 3.5.2) assuming that the underlying random vectors $\boldsymbol{\xi}$ and $\boldsymbol{\psi}$ are autoregressive processes of order one. Fixing $n = 50$, we consider the structure parameter $\phi \in [-0.98, 0.98]$ for all processes.

Case A: Figure 4.25 presents the trend of the estimator bias of $\hat{\beta}_x$ as the autocorrelation parameter of the cross-covariance matrix ϕ_{zx} increases with respect to ϕ_x. As for the previous application, it appears to follow the co-variability of $\boldsymbol{X}$ and $\boldsymbol{Z}$ amended for the effect of the covariate variability (see Figure 3.3).

In this regard, Figure 4.26 shows the relation between the exact estimator bias and its Laplace approximation up to the first order. It is not a linear association, however it is impressive how the values are close to the bisector (green line). Additionally, the division into the different values of ϕ_{zx} highlights that the bias

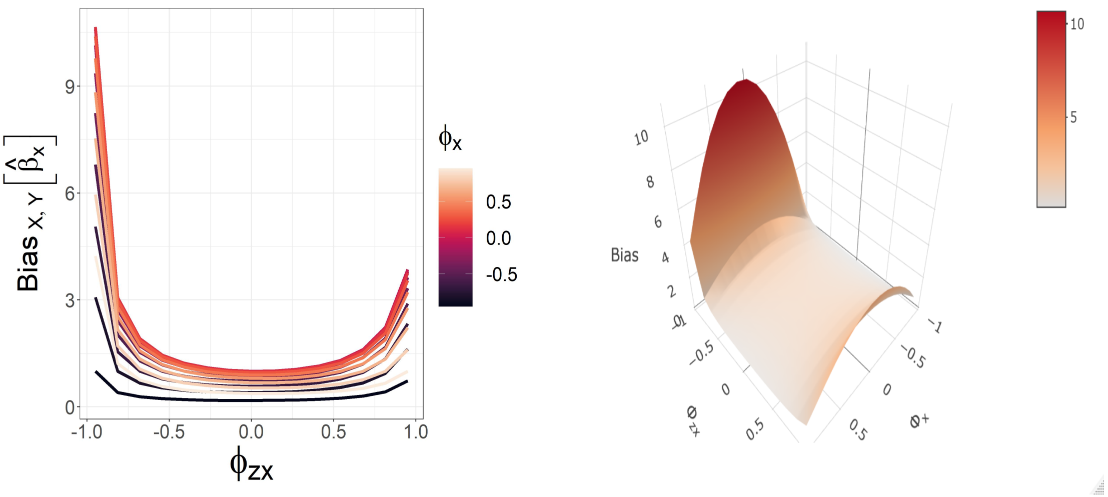

Figure 4.25: Marginal bias of $\hat{\beta}_x$ as the autocorrelation parameter of the cross-covariance matrix ϕ_{zx} increases respect to ϕ_x from two perspectives under temporal framework (case A).

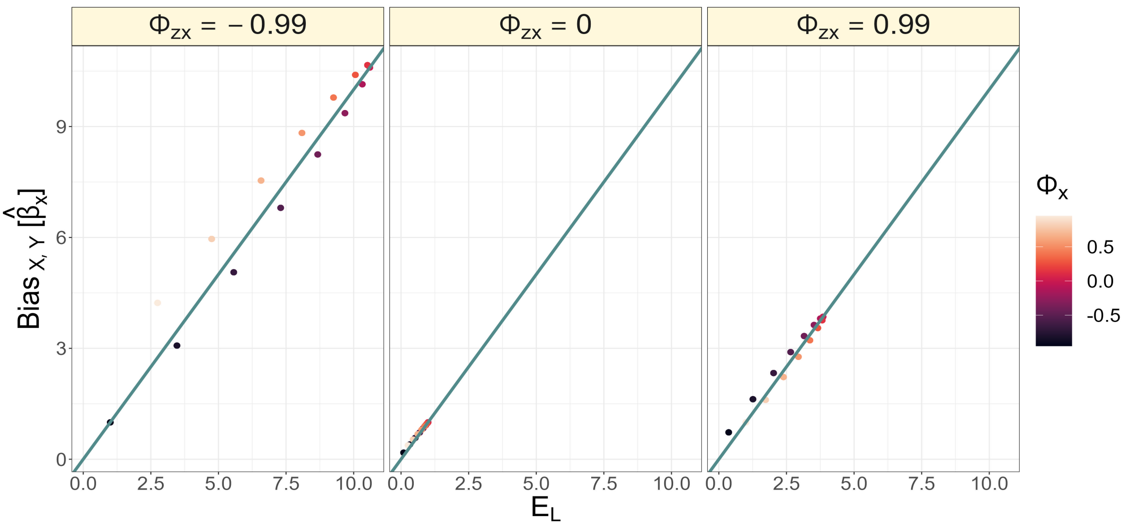

Figure 4.26: The accuracy of first order Laplace approximation to the bias respect to ϕ_x and $\phi_{zx} \in \{-0.99, 0, 0.99\}$.

takes the minimum values when the cross-correlation structure is simplified to be the identity matrix and it confirms that as the cross-covariance structure gets more complex (or stronger) the bias increases. Concerning the estimator variance and mean square error (see Figure 4.27,), it seems that ϕ_x and ϕ_z produce notable changes. While ρ_{xz} and ϕ_{zx} have little significant influence on the variance (except for the positive definiteness condition), as opposed to what happens for MSE in Figure 4.28. In addition, the values of ϕ_z change drastically their pattern, in particular for $\phi_z \approx 0$ the variance is quite stable.

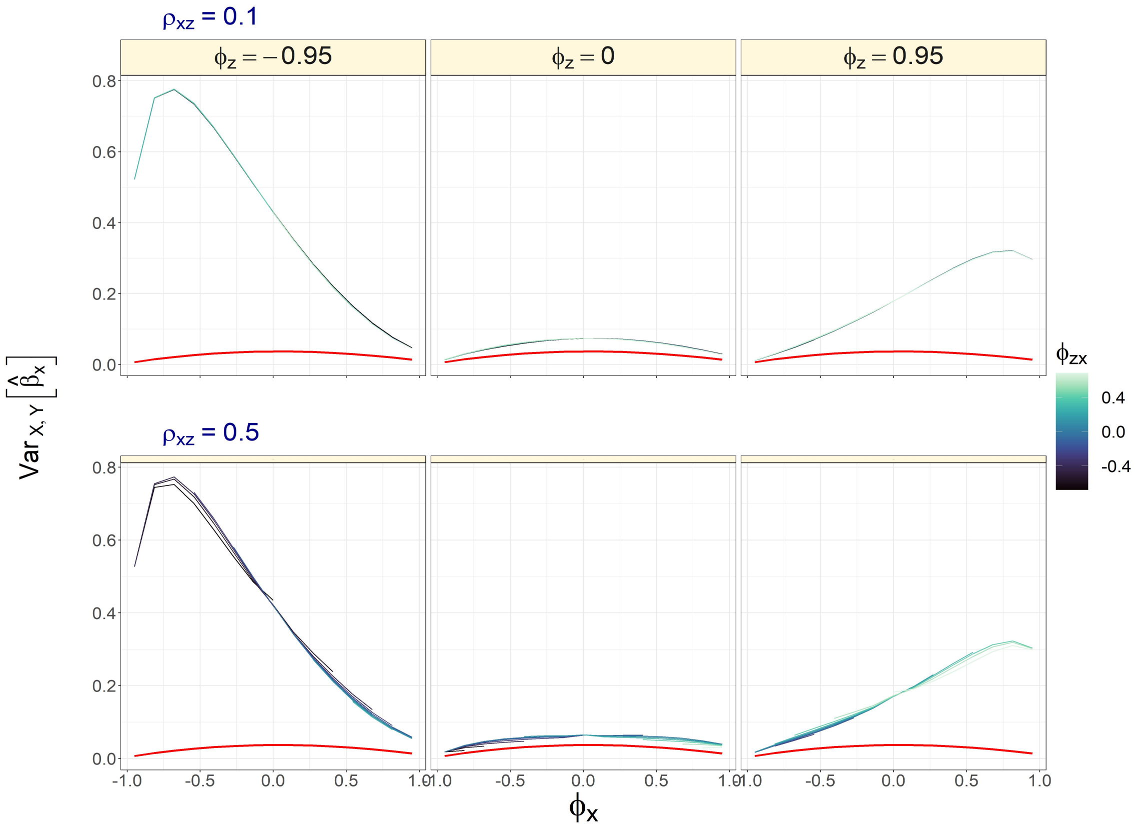

Figure 4.27: Marginal variance trends of $\hat{\beta}_x$ when $\mathbf{\Sigma}_{x,z} \succ 0$ and $\rho_{xz} \in \{0.1, 0.5\}$. The red curve stands for the estimator variance when there is no omitted variable $\boldsymbol{Z}$ in the model.

Case B: From Figure 4.29 (left) it is evident how EV_x is the sum of the underlying processes variability. The plot on the right shows also in this case that E_L is an accurate approximation of the bias. In order to see the variation bias in function of the structure parameters we present Figure 4.29. It reveals higher bias when $EV_\xi \ll EV_\psi$, namely $|\phi_\xi| \ll |\phi_\psi|$, confirming the strong relationship that the estimator bias has with the variability quantities. Eventually, Figure 4.31 shows how the confounding-dependent part of MSE and variance changes over $\rho_{xz} \in \{0.1, 0.5, 0.9\}$. Regarding the mean square error, its magnitude increases with the correlation parameter, instead the variance decreases. In particular, as ρ_{xz} increases both quantities are more affected by the variation of ϕ_ψ.

Case C: Panels in Figure 4.32 present the bias trend as the range parameters ϕ_x and ϕ_z vary. As for the areal data case, the bias appears to be smaller when the

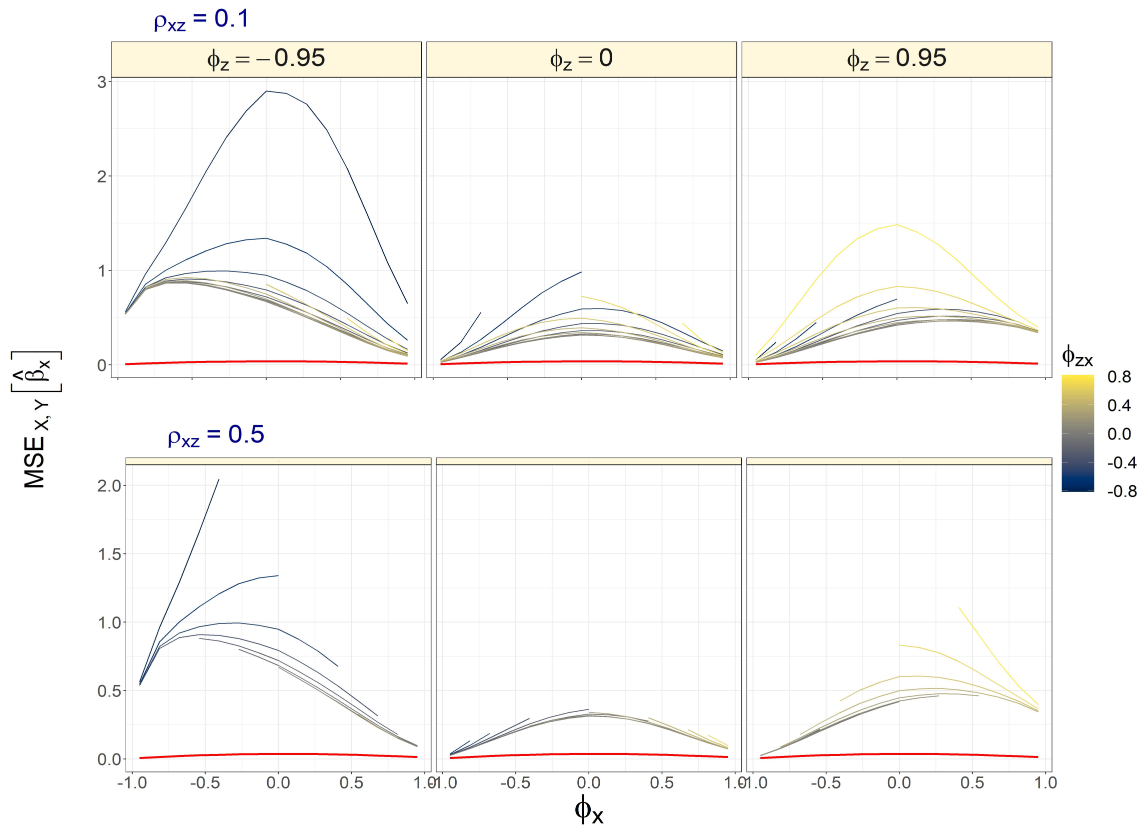

Figure 4.28: Marginal mean square error trends of $\hat{\beta}_x$ when $\mathbf{\Sigma}_{x,z} \succ 0$ and $\rho_{xz} \in \{0.1, 0.5\}$. The red curve stands for the estimator MSE when there is no omitted variable $\boldsymbol{Z}$ in the model (case A).

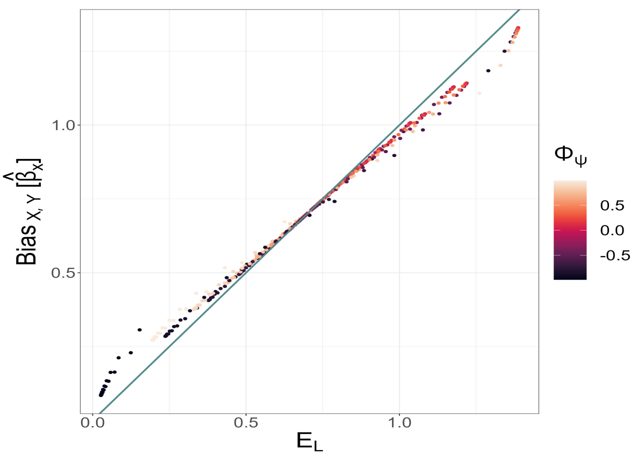

Figure 4.29: The accuracy of first order Laplace approximation to the bias respect to the confounder parameter (case B).

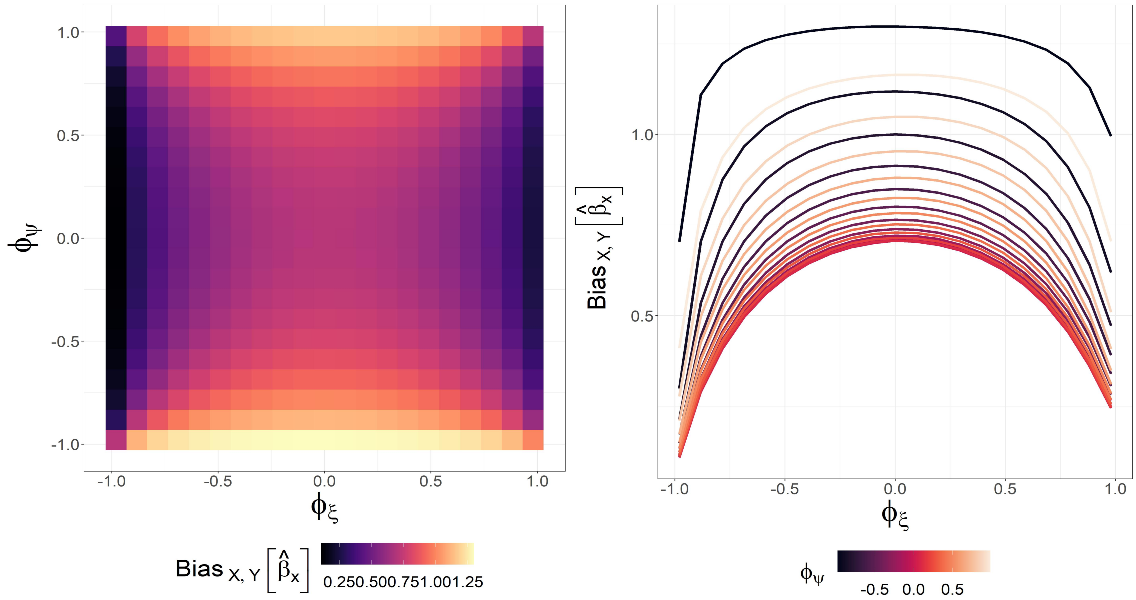

Figure 4.30: Estimator bias from two different perspectives as the range parameter ϕ_ξ and ϕ_ψ vary (case B).

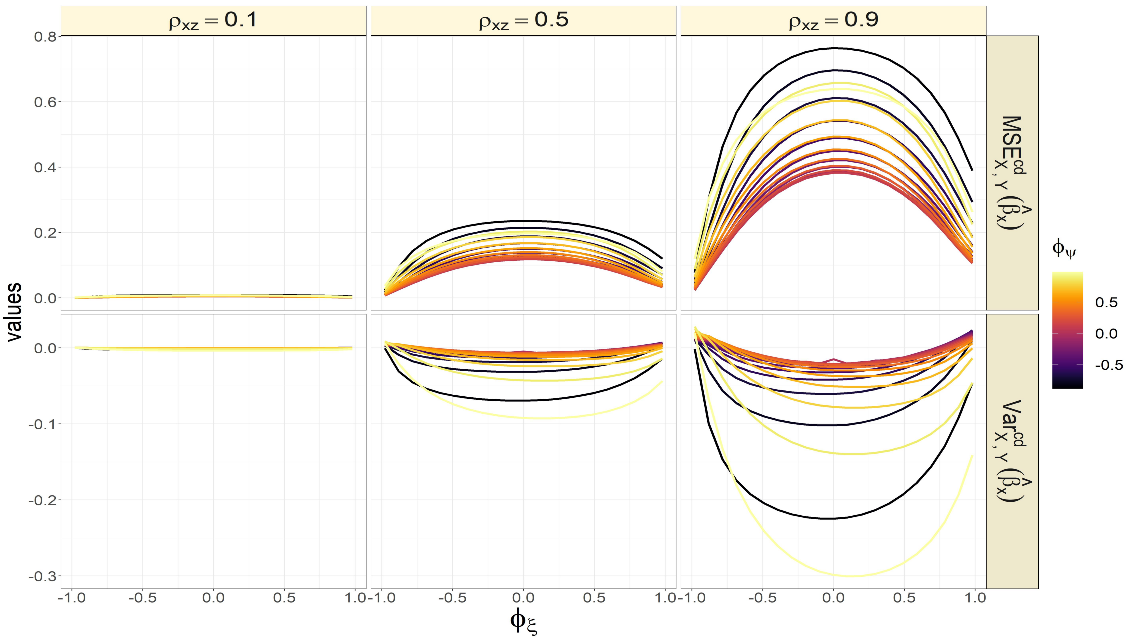

Figure 4.31: Comparison of the part of mean square error and variance varying when confounding occurs as the correlation parameter $\rho_{xz} \in \{0.1, 0.5, 0.9\}$ changes (case B).

covariate variability is higher (on the edge of the ϕ_x range) and the one of confounder is lower. This is in line with what it is shown in Figure 4.33 exhibiting

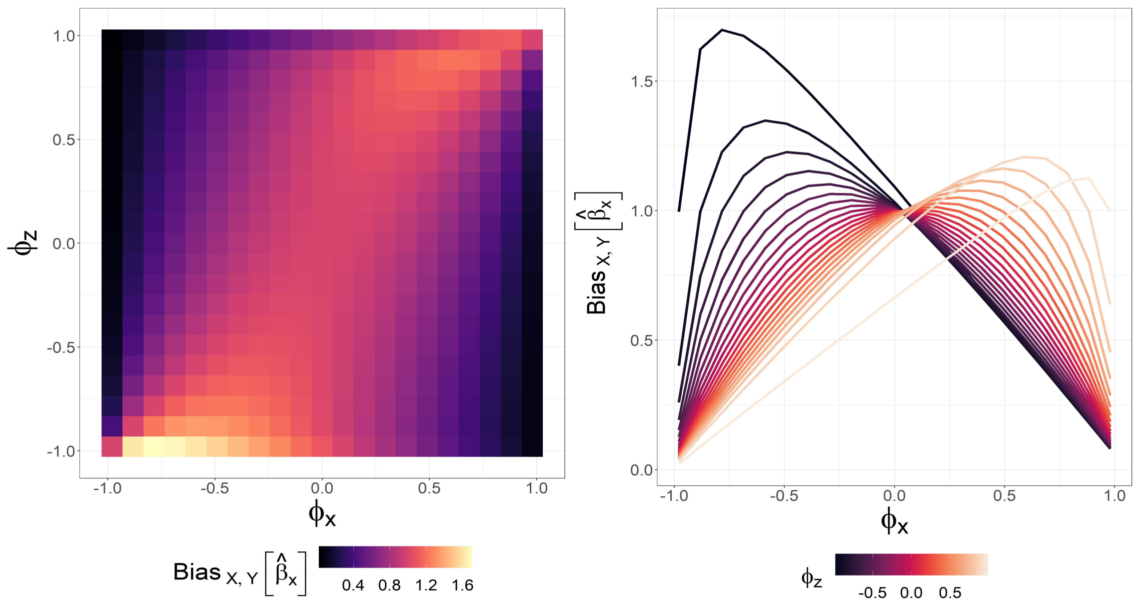

Figure 4.32: Estimator bias from two different perspectives as the range parameter ϕ_x and ϕ_z vary (case C).

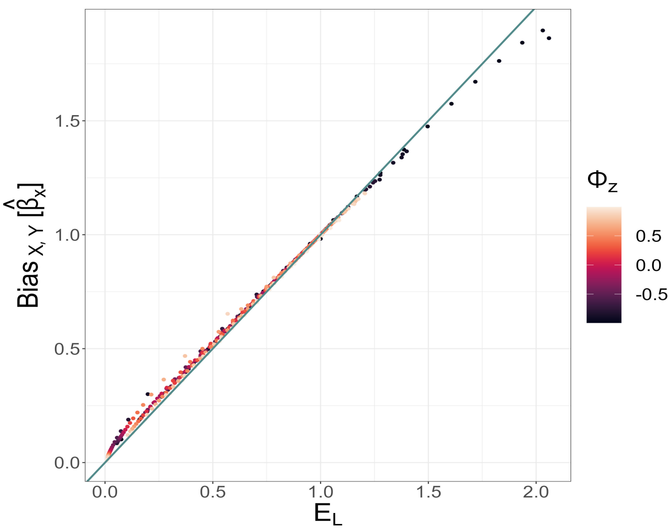

Figure 4.33: The accuracy of first order Laplace approximation to the bias respect to confounder parameter (case C).

an association between E_L and estimator bias next to the linearity. Furthermore, Figure 4.34 displays how the MSE and variance depend on confounding. With respect to the mean square error, its scale increases with the correlation parameter. In particular, as ρ_{xz} increases, the variation of ϕ_z has more impact on the confounding-dependent part of the estimator sampling properties.

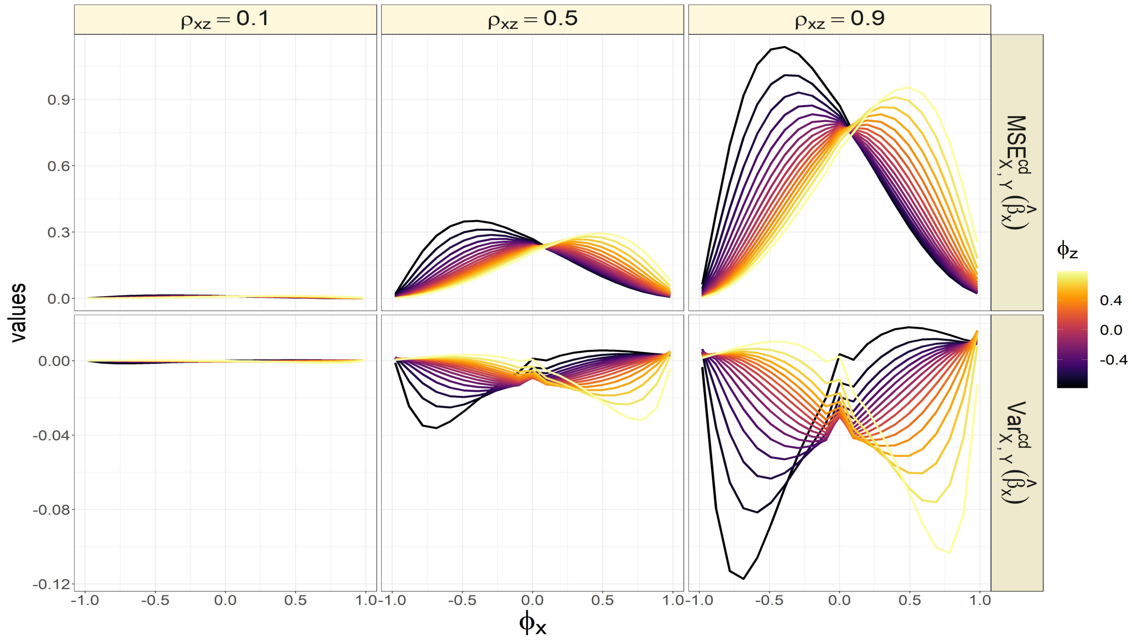

Figure 4.34: Comparison of the part of mean square error and variance varying when confounding occurs as the correlation parameter $\rho_{xz} \in \{0.1, 0.5, 0.9\}$ changes.

Finally, extending the spatial confounding framework to general structured setting characterized by specific auto-correlation structures allow us to reach interesting conclusions. Indeed, through the results achieved in this section we have shown how the issue of confounding does not depend upon the application field of the linear regression model in which it occurs and not even by the structural parameters that characterize the data generating process, at least not directly. Focusing on the estimator bias as the principal marker of confounding, it is evident how the relationship between the structure parameters of the DGP and the bias changes upon the kind of covariance and cross-covariance function used to build up the generative mechanism. This makes it difficult to outline a global picture that correctly understands and explains the relationships influencing the confounding. What remains unchanged is the connection that the ratio of the confounder-covariate co-variability and the covariate variability, i.e. $\bar{\lambda}_{M\Sigma_{xz}}/\bar{\lambda}_{M\Sigma_x}$, has with the estimator bias. It has been suggested by the first order Laplace approximation of such estimator sampling property. Despite the literature asserts that the level of covariate and confounder processes' smoothness is strictly linked to the estimator bias, in our opinion such sampling property is directly connected with the variability of the covariate process and the expected sampling covariance of the covariate and confounder processes. The link with the smoothness arise only when the cross-correlation function depends on the parameter of the marginal distribution of $\boldsymbol{Z}$.

In particular, the results show that the Laplace approximation of the estimator bias up to the first order E_L is accurate under the different scenarios. When an OLS estimator is posited, this approximation matches with the ratio $\bar{\lambda}_{M\Sigma_{xz}}/\bar{\lambda}_{M\Sigma_x}$. Moreover, with the objective to understand confounding, another relevant insight concerns the fact that, as expected from theory developed in Chapter 3, the confounder structure does not affect the distortion of the estimator $\hat{\beta}_x$, rather it influences its further sampling properties, the variance and mean square error, that depends on it because the confounder is an omitted variable by definition. Of course, if one assumes a data generative mechanism in which the cross-covariance of confounder and covariate is function of the confounder structure (e.g. cases B and C), unavoidably all sampling properties are depending on it.

Conclusions

Summary and conclusion

In this thesis, the problem of confounding in linear regression models is addressed. In particular, we study, through the evaluation of estimator sampling properties, how confounding affects the estimate of the inferential target, i.e. the regression parameter of the covariate on the outcome.

The spatial literature has extensively dealt with this issue. To assess the impact of confounding on the sampling properties of the regression coefficient estimators, the research focused on the strength of the auto-correlation characterizing covariate and confounder, both spatially varying. To date, what is clear from the previous studies is that the parameters influencing the spatial auto-correlation of the covariate and confounder processes are undoubtedly of great importance. This consideration remains true also for a wide range of applications. Since extending the spatial confounding framework to general structured settings is one of the objectives of the thesis, we consider what has been asserted by the spatial literature as our starting point in order to move towards broader settings.

We provide more awareness regarding the effect of confounding on coefficient estimates by generalizing the theory initialized by Paciorek (2010). According to him and Page et al. (2017), our main references along the dissertation, the increasingly accepted idea is that the smoothness of covariate and confounder processes is an important factor that produces relevant changes in estimator sampling properties. In particular, Paciorek (2010) affirms that a confounder smoother than the covariate leads to a lower bias, and subsequently to less confounding. Actually, one may agree only in specific situations, when the parameters governing the confounder covariance matrix contribute to the cross-covariance matrix, e.g. assuming a Matérn correlation function and an LMC or rLMC model for the cross-covariance structure. Besides, moving away from particular cases, the connection to smoothness would not hold. In this regard, we introduce the expected sampling variance and covariance, expressing the variability of a process and the variability of the interaction between two processes. In situations other than those just described, the smoothness and the variability of a random vector do not match. For example, when an autoregressive of order one or a conditional autoregressive process is posited for the data generating process.

Being in the one-covariate setting allows us to compute the marginal sampling properties of the estimator without carrying out any simulation study. This is possible because bias, variance and mean square error conditionally on the covariate process are derived as ratios of dependent quadratic forms in Gaussian random variables. It is then achievable to provide marginal analytic results by means of Carlson's R function. This development enables us to work nimbly under several workflows that consider spatial and temporal data with specific assumptions regarding the covariance and cross-covariance functions used in the generative mechanism.

To conclude, in our opinion and in light of the evidence coming from the application study, the estimator sampling properties depend upon the aforementioned variability quantities and on the portion of the response and covariate variability expressed by the confounder. Considering the estimator bias as the principal marker of confounding, we point out that the confounder smoothness is not the most relevant measure determining bias. Indeed, the cross-covariance matrix characterizing the covariate-confounder interaction plays the most prominent role in detecting it. Specifically, the bias mainly hinges on the covariate variability and on the expected sampling covariance of covariate and confounder. Based on this fact, we propose a representative quantity for the extent of confounding as a proxy of the estimator bias, its first order Laplace approximation E_L, defined as the ratio of the expected sampling covariance and the expected sampling variance of the covariate. Moreover, we show theoretically and empirically that the confounder structure does not affect the estimator bias, rather it influences the variance of the estimator.

Future developments

The research reported in this thesis contributes to the literature with a wider understanding of confounding in linear regression models. In particular, it allows to manage several framework featured by different fields of application and all kinds of data generating mechanisms based on the stochastic relation between the response, the covariate and the confounder. Future studies may consider other workflows to examine the sources and consequences of confounding in more depth.

Regarding our application study in Chapter 4, it has been developed under the model assumption $\mathbf{S} = \mathbf{I}_n$, namely by considering $\hat{\beta}_x$ to be the OLS estimator. Nonetheless, such analysis has been reproduced also positing $\mathbf{S} = \boldsymbol{\Sigma}_{y|x}$. That choice corresponds to the best option concerning the error covariance matrix, because in this way we suppose to know

the variance structure that characterizes the latent part of the phenomenon. However, as remarked along the dissertation, this assumption yields the minimum variance estimator, but does not remove the bias. In fact, the results are not included since they were not so different from the ones gathered under the OLS assumption.
Additionally, it is recommended to put the attention also to the response variability. We could go back over our analysis, fix the response variability in each scenario and look into the relationship between the estimator sampling properties and the portion of response and covariate variability explained by the confounder processes.

In our work, we present the first order Laplace approximation E_L as a significant proxy for the bias of $\hat{\beta}_x$. Besides, one of the future goals of the research is to investigate more deeply the key aspects of all estimator sampling properties. Starting with the estimator bias and aware about the important role of the eigenvalues of the matrices appearing in the ratio of quadratic forms that defines it, we supply boundaries for the bias.

It is possible to revise the estimator bias conditionally on $\boldsymbol{X}$ as follows

$$\text{Bias}\left[\hat{\beta}_x|\boldsymbol{X}\right] \quad = \mathcal{B}_{y\cdot z(x)}\tilde{\lambda}_u \frac{\boldsymbol{X}^\top \tilde{\lambda}_u^{-1} \boldsymbol{\Delta}\mathbf{A}_{z\cdot x}\boldsymbol{X}}{\boldsymbol{X}^\top \boldsymbol{\Delta}\boldsymbol{X}},$$

namely in dependence of an upper bound $\tilde{\lambda}_u$ such that

$$\frac{\boldsymbol{X}^\top \tilde{\lambda}_u^{-1} \boldsymbol{\Delta}\mathbf{A}_{z\cdot x}\boldsymbol{X}}{\boldsymbol{X}^\top \boldsymbol{\Delta}\boldsymbol{X}} < 1.$$

To find out such bound, we observe that

$$Pr\left(\frac{\boldsymbol{X}^\top \boldsymbol{\Delta}\mathbf{A}_{z\cdot x}\boldsymbol{X}}{\boldsymbol{X}^\top \boldsymbol{\Delta}\boldsymbol{X}} < t\right) = Pr\left(\boldsymbol{X}^\top \boldsymbol{\Delta}\left(\mathbf{A}_{z\cdot x} - t\mathbf{I}_n\right)\boldsymbol{X} < 0\right) \tag{4.15}$$

implies that $t = t^*$ is an upper bound if (4.15) is equal to 1. Letting Formula (4.15) take the value 1 means that the QF $\boldsymbol{X}^\top \boldsymbol{\Delta}\left(\mathbf{A}_{z\cdot x} - t\mathbf{I}_n\right)\boldsymbol{X} \leq 0$ for all $\boldsymbol{X} \neq \mathbf{0}$, i.e. that is negative definite. Then,

$$\lambda\left[\boldsymbol{\Delta}\left(\mathbf{A}_{z\cdot x} - t\mathbf{I}_n\right)\right]_{\max} \leq \lambda(\boldsymbol{\Delta})_{\max}\lambda\left(\mathbf{A}_{z\cdot x} - t\mathbf{I}_n\right)_{\max}.$$

Since it is guaranteed that $\lambda(\boldsymbol{\Delta})_{\max} > 0$ and that

$$\boldsymbol{\lambda}\left(\mathbf{A}_{z\cdot x} - t\mathbf{I}\right) = \boldsymbol{\lambda}\left(\mathbf{A}_{z\cdot x}\right) - t\mathbf{1},$$

we get

$$\lambda\left[\boldsymbol{\Delta}\left(\mathbf{A}_{z\cdot x} - t\mathbf{I}_n\right)\right]_i < 0 \qquad \forall i = 1, \ldots, n,$$

if an only if $t > \lambda(\boldsymbol{A}_{z \cdot x})_{\max}$. Consequently, $t^* = \lambda(\boldsymbol{A}_{z \cdot x})_{\max}$ is an upper bound of $\text{Bias}\left[\hat{\beta}_x | \boldsymbol{X}\right]$ that enables us to denote $\mathcal{B}_{y \cdot z(x)} \lambda(\boldsymbol{A}_{z.x})_{\max}$ as an upper bound for the bias. A similar direction could also be pursued for the estimator variance and mean square error.

Moreover, when the final objective is to obtain a point prediction for the response, it may be sufficient to just estimate the model, since the potential bias in estimating the target parameter will be compensated by estimating random effects (Page et al., 2017). This does not hold for the quantification of prediction uncertainty. It could be interesting to focus our study on predictions, although it could be considered a second order problem.

In this dissertation, the assessment of the impact that confounding has on regression coefficients is faced but, regarding spatial framework, the statistics literature dealt also with the development of methods to account for spatial confounding. As possible future outgrowth, it would be desirable finding a way to adjust for confounding in a Bayesian framework. A brief review of this topic follows.
First of all, Reich et al. (2006) and Hodges and Reich (2010) propose a method called *restricted spatial regression* (RSR). It provides fitting a model in which the random effects are constrained to a subspace orthogonal to the column space of the fixed effects design matrix heeding the spatial correlation without changing the estimates of the fixed effects. Its target is deconfounding the two types of effects reducing the variance rise and improving the inference of the fixed effect. One of the weakness is that this solution assigns all the variability explained by measured and unmeasured covariates to the observed ones. Hefty discussion about the worthiness of restricted spatial regression has been produced since it was first proposed, and several alternatives to the original idea have been reported (Hughes and Haran, 2013; Hanks et al., 2015; Hughes, 2017; Guan and Haran, 2018; Prates et al., 2019; Dupont et al., 2021; Adin et al., 2021). In this regard, Khan and Calder (2022) and Zimmerman and Hoef (2021) debunk the RSR approach in the two statistical context, Bayesian and frequentist, respectively. Then, several alternatives to it have been brought forward but other methods for avoiding the problem in object are limited, and with theoretical bases often intractable since methodology tends to rely on simulations alone. In the causal inference font there are some interesting results (Osama et al., 2019; Davis et al., 2019, 2021). Papadogeorgou et al. (2018) and Schnell and Papadogeorgou (2020) consider a joint model for response and covariate based on Gaussian Markov random field developing a new method, *termed distance adjusted propensity score matching* (DAPSm) that incorporates informations on proximity of spatial units into a propensity score matching procedure (Rosenbaum and Rubin, 1983). In addition, Thaden and Kneib (2018) propose a *geoadditive structural equation model* (gSEM) consists of three stages to regress away the spatial structure from both the response and the covariate leading to

unbiased estimates provided by simulation study. It not so clear how this method work and it seems unpreferable to eliminate all spatial information from the model. Aware of this limit, Dupont et al. (2021) suggest a novel approach, the *Spatial+* model, that is a simple modification of the spatial model where the covariates are replaced by their residuals after spatial dependence has been regressed away using a thin spline formulation. A practical advantage over gSEM is that, as the response variable is unchanged from the spatial model, standard model selection criteria can be used for comparisons with the spatial and no spatial models.

Therefore, proper adjustment for spatial confounding is a important and open issue as evidenced by many context in which it is being tackled (Lee and Sarran, 2015; Bradley et al., 2015; Murakami and Griffith, 2015; Hefley et al., 2017; Pereira et al., 2020; Azevedo et al., 2020; Reich et al., 2021; Azevedo et al., 2022; Hui and Bondell, 2022). Donegan et al. (2020) propose a Bayesian method for spatial regression using eigenvector spatial filtering and regularized horseshoe prior. Likewise Paciorek (2010), Page et al. (2017) and Keller and Szpiro (2020), Guan et al. (2022) address the importance of spatial scale of the treatments and missing confounder variable developing a model in the spectral domain and studying their correlations at different scales. They show that the optimal adjustment for confounding is not possible without further assumption, and so they assume that the correlation at different spatial scales dissipates at high frequencies of spectral domain. Recently, Marques et al. (2022) develop a new prior structure able to deal with spatial confounding managing to increase computational efficiency of it by exploring the sparsity of GMRF's in SPDE approach (Lindgren et al., 2011). They also extend the prior structure to the case of multiple covariates being correlated with the spatial random effect.

In conclusion, facing the issue of confounding from a Bayesian point of view means wondering what conjectures about the confounder (latent) variable must be done in order to learn about the target parameter $\mathcal{B}_{y \cdot x(z)}$ and, in turn, also about bias, starting from the knowledge of $\boldsymbol{Y}$ and $\boldsymbol{X}$ (Eberly and Carlin, 2000). That means finding out a reliable posterior distribution of the parameter. The shortcoming lies in the fact that one cannot learn about the bias without observing $\boldsymbol{Z}$. Nonetheless, it would be desirable to develop a Bayesian framework capable to adjust the biased estimates due to confounding in Gaussian models and to elaborate a novel prior structure to deal with such problem. Allocating informative priors to the observed confounder-response association may reduce the bias in that parameter.

Appendix A

Linear algebra tools

In this appendix we include a few algebra tools that are used to provide some results of the thesis. Most of them can be read in Gentle (2007), Banerjee and Roy (2014) and Horn and Johnson (2013).

The trace of square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ of dimension n is defined as $tr(\mathbf{A}) = \sum_{i=1}^{n} A_{ii}$. In the ensuing theorem we report its main properties (Searle and Khuri, 2017)[Chapter 12].

Theorem A.0.1. *Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$. Then, the following holds:*

(i) $tr(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i(\boldsymbol{A})$;

(ii) $tr(\boldsymbol{A}) = tr(\boldsymbol{A}^\top)$;

(iii) $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$;

(iv) If $\boldsymbol{A}$ is a square matrix whose diagonal elements are differentiable functions of x, then
$$\frac{\partial \mathrm{tr}(\boldsymbol{A})}{\partial x} = \mathrm{tr}\left(\frac{\partial \boldsymbol{A}}{\partial x}\right);$$

(v) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices whose elements are differentiable functions of x and such that the product $\boldsymbol{AB}$ is defined, then
$$\frac{\partial \boldsymbol{AB}}{\partial x} = \frac{\partial \boldsymbol{A}}{\partial x}\boldsymbol{B} + \boldsymbol{A}\frac{\partial \boldsymbol{B}}{\partial x};$$

(vi) If $\boldsymbol{A}$ is a symmetric matrix whose elements are differentiable functions of x, then
$$\frac{\partial |\boldsymbol{A}(x)|}{\partial x} = |\boldsymbol{A}|\,\mathrm{tr}\left(\boldsymbol{A}^{-1}\frac{\partial \boldsymbol{A}}{\partial x}\right)$$

and

$$\frac{\partial \boldsymbol{A}^{-1}(x)}{\partial x} = -\boldsymbol{A}^{-1}\left(\frac{\partial \boldsymbol{A}}{\partial x}\right)\boldsymbol{A}^{-1}.$$

Given that a covariance matrix to be valid needs to be symmetric and positive definite, we discuss the definiteness of a matrix and some of its properties.

Definition A.0.1. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$. Then $\mathbf{A}$ is positive definite i.e. $\boldsymbol{x}^T\mathbf{A}\boldsymbol{x} > 0$ (positive semidefinite i.e $\boldsymbol{x}^T\mathbf{A}\boldsymbol{x} \geq 0$) for all non-zero $\boldsymbol{x} \in \mathbb{R}^n$ if and only if every eigenvalue of $\mathbf{A}$ is positive (non-negative).

Regarding the inverse of the sum of two matrices, Henderson and Searle (1981) report an interesting result. Considering a non-singular matrix $\mathbf{A}$ and $\mathbf{U}$, $\mathbf{B}$ and $\mathbf{V}$ that may be rectangular. It is known that

$$\begin{aligned}(\mathbf{A}+\mathbf{UBV})^{-1} &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{B}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1}, &\text{(A.1)}\\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{I}+\mathbf{BVA}^{-1}\mathbf{U})^{-1}\mathbf{BVA}^{-1}. &\text{(A.2)}\end{aligned}$$

The identity (A.1) is called *Woodbury matrix identity* and the (A.2) is its simple form.

The authors also discuss about the inverse of a generic block matrix ensuring further versions. A square matrix $\mathbf{A}$ that can be partitioned as

$$\mathbf{A} = \begin{bmatrix}\mathbf{A}_{11} & \mathbf{A}_{12}\\ \mathbf{A}_{21} & \mathbf{A}_{22}\end{bmatrix} \tag{A.3}$$

where $\mathbf{A}_{11}$ is non-singular, has interesting properties that depend on the matrix

$$\mathbf{A}/\mathbf{A}_{11} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$$

which is called the *Schur complement of* $\boldsymbol{A}_{11}$ in $\mathbf{A}$.

Proposition A.0.2 (Schur-Barachiewicz inverse formula)**.** *Suppose* $\boldsymbol{A}$ *is non-singular and can be partitioned as above with* $\boldsymbol{A}_{11}$ *non-singular. The inverse of* $\boldsymbol{A}$ *is given by*

$$\boldsymbol{A}^{-1} = \begin{bmatrix}\boldsymbol{A}_{11}^{-1} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}/\boldsymbol{A}_{11})^{-1}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}(\boldsymbol{A}/\boldsymbol{A}_{11})^{-1}\\ -(\boldsymbol{A}/\boldsymbol{A}_{11})^{-1}\boldsymbol{A}_{21}\ \boldsymbol{A}_{11}^{-1} & (\boldsymbol{A}/\boldsymbol{A}_{11})^{-1}\end{bmatrix} \tag{A.4}$$

and if $\boldsymbol{A}_{11}$ *is also square, the determinant of* $\boldsymbol{A}$ *is the product of the determinant of the principal submatrix and the determinant of its Schur complement:*

$$|\boldsymbol{A}| = |\boldsymbol{A}_{11}| \cdot |\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}|$$

and for the properties of the determinant we have that:

$$|\boldsymbol{A}| = |\boldsymbol{A}_{22}| \cdot |\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}|. \tag{A.5}$$

Moreover, if $\boldsymbol{A}_{22}$ is non-singular the Schur complement of $\boldsymbol{A}_{22}$ in $\boldsymbol{A}$ is

$$\boldsymbol{A}/\boldsymbol{A}_{22} = \boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} \tag{A.6}$$

the inverse is given by

$$\boldsymbol{A}^{-1} = \begin{bmatrix} (\boldsymbol{A}/\boldsymbol{A}_{22})^{-1} & -(\boldsymbol{A}/\boldsymbol{A}_{22})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \\ -\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}(\boldsymbol{A}/\boldsymbol{A}_{22})^{-1} & \boldsymbol{A}_{22}^{-1} + \boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}(\boldsymbol{A}/\boldsymbol{A}_{22})^{-1}\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1} \end{bmatrix}. \tag{A.7}$$

The Schur complement arises when performing a block Gaussian elimination on matrix $\mathbf{A}$ that leads to lower-upper decomposition. If $\mathbf{A}_{11}$ is invertible, then it is possible to use the Schur complement $\mathbf{A}/\mathbf{A}_{11}$ to obtain the following factorization of $\mathbf{A}$:

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}/\mathbf{A}_{11} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}.$$

This factorization gives us useful ways to express positive semidefiniteness of matrices with a block structure (Horn and Johnson, 2013). Moreover, if we assume that $\mathbf{A}$ is symmetric, so that $\mathbf{A}_{11}$, $\mathbf{A}_{22}$ are symmetric and $\mathbf{A}_{21} = \mathbf{A}_{12}^\top$, then we can express $\mathbf{A}$ as

$$\begin{aligned} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^\top & \mathbf{A}_{22} \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}/\mathbf{A}_{11} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}_{11}^{-1}\mathbf{A}_{12} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}/\mathbf{A}_{11} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{bmatrix}^\top \end{aligned}$$

and also,

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^\top & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}/\mathbf{A}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}^\top \tag{A.8}$$

which shows that $\mathbf{A}$ is similar to a block-diagonal matrix, and so, they have the same spectrum, i.e. the same eigenvalues. As a consequence, we have the following version of "Schur's trick" to check if a symmetric matrix $\mathbf{A}$ is positive definite (Zhang, 2005):

Theorem A.0.3. *For any symmetric matrix $\boldsymbol{A} \in \mathcal{S}^n$ that has the form as in (A.3), if $\boldsymbol{A}_{22}$ is invertible then the coming properties hold:*

(i) $\boldsymbol{A} \succ 0 \iff \boldsymbol{A}_{22} \succ 0$ *and* $\boldsymbol{A}/\boldsymbol{A}_{22} \succ 0$.

(ii) If $\boldsymbol{A}_{22} \succ 0$*, then* $\boldsymbol{A} \succeq 0 \Longleftrightarrow \boldsymbol{A}/\boldsymbol{A}_{22} \succeq 0$.

Instead, if $\boldsymbol{A}_{22}$ *is invertible then the following properties hold:*

(i) $\boldsymbol{A} \succ 0 \Longleftrightarrow \boldsymbol{A}_{11} \succ 0$ *and* $\boldsymbol{A}/\boldsymbol{A}_{11} \succ 0$.

(ii) If $\boldsymbol{A}_{11} \succ 0$*, then* $\boldsymbol{A} \succeq 0 \Longleftrightarrow \boldsymbol{A}/\boldsymbol{A}_{11} \succeq 0$.

As intermediate consequence of Definition A.0.1 we have that

$$\mathbf{A} \succeq 0 \Longleftrightarrow \mathbf{B}\mathbf{A}\mathbf{B}^\top \succeq 0, \tag{A.9}$$

for any given non-singular $\mathbf{B} \in \mathbb{R}^{n \times n}$. Another implication is that a block diagonal matrix is positive (semi)definite if and only if each of its diagonal blocks is positive (semi)definite.

From Theorem A.0.3 and its consequence (A.9), it is possible to state that every positive (semi)definite symmetric matrix $\mathbf{A} \in \mathcal{S}^n$ can be represented as

$$\begin{aligned}
\mathbf{A} &= \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^\top & \mathbf{A}_{22} \end{pmatrix} && \text{(A.10)} \\
&= \begin{pmatrix} \mathbf{A}/\mathbf{A}_{22} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^\top & \mathbf{A}_{22} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{A}/\mathbf{A}_{22} + \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{22}\mathbf{A}_{22}^{-1}\mathbf{A}_{22}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{22} \\ (\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{22})^\top & \mathbf{A}_{22} \end{pmatrix}. && \text{(A.11)}
\end{aligned}$$

Defining $\mathbf{A}_{1\cdot 2} = \mathbf{A}_{12}\mathbf{A}_{22}^{-1}$, the matrix can be expressed as

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}/\mathbf{A}_{22} + \mathbf{A}_{1\cdot 2}\mathbf{A}_{22}\mathbf{A}_{1\cdot 2}^\top & \mathbf{A}_{1\cdot 2}\mathbf{A}_{22} \\ (\mathbf{A}_{1\cdot 2}\mathbf{A}_{22})^\top & \mathbf{A}_{22} \end{pmatrix}, \tag{A.12}$$

where $\mathbf{A}_{22} \succ 0$ ($\mathbf{A}_{22} \succeq 0$) and $\mathbf{A}/\mathbf{A}_{22} \succ 0$ ($\mathbf{A}/\mathbf{A}_{22} \succeq 0$). Furthermore if $\mathbf{A}_{22} \succ 0$ ($\mathbf{A}_{22} \succeq 0$) and $\mathbf{A}/\mathbf{A}_{22} \succ 0$ ($\mathbf{A}/\mathbf{A}_{22} \succeq 0$), then $\mathbf{A} \succ 0$ ($\mathbf{A} \succeq 0$).

The positive definiteness of the matrices on the principal diagonal is not a sufficient condition for the positive definiteness of the matrix. For this reason we are interested to find some additional conditions to achieve it. For that purpose, first of all let us introduce the *Gershgorin's Theorem*, useful to figure out what range the eigenvalues of a certain matrix would be in (Horn and Johnson, 2013):

Theorem A.0.4. *Let* $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ *and let* $d_i = \sum_{i \neq j} |A_{ij}|$, $i = 1, \ldots, n$. *The set* $D_i = \{z \in \mathbb{C} : |z - A_{ii}| \leq d_i\}$ *is called an Gershgorin disk of the matrix* $\boldsymbol{A}$ *and the union of disks* $G(\boldsymbol{A}) = \cup_{i=1}^n D_i$ *is called Gershgorin domain. Every eigenvalue* λ_A *of matrix* $\boldsymbol{A}$ *belongs* *to $G(\boldsymbol{A})$ and satisfies:*

$$|\lambda_A - A_{ii}| \leq d_i \qquad \text{for } i = 1, \ldots, n. \tag{A.13}$$

Furthermore the following properties hold:

(i) If $\boldsymbol{A}$ is strictly row diagonally dominant, that is

$$|A_{ii}| > \sum_{j \neq i}^{n} |A_{ij}|, \qquad \text{for } i = 1, \ldots, n,$$

then $\boldsymbol{A}$ is invertible.

(ii) If $\boldsymbol{A}$ is strictly row diagonally dominant and $A_{ii} > 0$ for $i = 1, \ldots, n$, then every eigenvalue of $\boldsymbol{A}$ has a strictly positive real part.

In particular, Theorem A.0.4 implies the following statements:

Corollary A.0.5. *A matrix $\boldsymbol{A} \in \mathcal{S}^n$ is called (strictly) diagonally dominat if*

$$|A_{ii}| - \sum_{i \neq j} |A_{ij}| \geq 0 \; (> 0) \qquad \text{for } i \text{ or } j \in \{1, \ldots, n\}. \tag{A.14}$$

If $\boldsymbol{A}$ is (strictly) diagonally dominant with all diagonal entries positive , then $\boldsymbol{A} \in \mathcal{S}^n_+$ ($\boldsymbol{A} \in \mathcal{S}^n_{++}$)

Being interested to the positive definiteness of difference between two matrices we state what follows

Theorem A.0.6. *Given $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{S}^n$, a sufficient condition for the difference matrix $\boldsymbol{B} - \boldsymbol{A}$ to be positive (semi)definite is that*

$$B_{ii} > A_{ii} + \sum_{i \neq j}^{n} |B_{ij} - A_{ij}| \quad \left(B_{ii} \geq A_{ii} + \sum_{i \neq j}^{n} |B_{ij} - A_{ij}| \right), \qquad \text{for } i = 1, \ldots, n.$$

Proof. In analyzing the theorem A.0.4 we see that every eigenvalues of a matrix lies within one of its Gershgorin disk. Thus, if we consider the symmetric difference matrix $\mathbf{C} = \mathbf{B} - \mathbf{A}$, each disk is centered at $C_{ii} = B_{ii} - A_{ii}$ and the eigenvalues of $\boldsymbol{C}$ lies in the disk

$$\left(C_{ii}, \sum_{i \neq j} |C_{ij}| \right) = \left(B_{ii} - A_{ii}, \sum_{i \neq j} |B_{ij} - A_{ij}| \right) \qquad \text{for } i = 1, \ldots, n.$$

From (A.13), we can say that each eigenvalue λ_C of the matrix $\mathbf{C}$ verify the following inequality:

$$C_{ii} - \sum_{i \neq j} |C_{ij}| \leq \lambda_C \leq C_{ii} + \sum_{i \neq j} |C_{ij}| \qquad \text{for } i = 1, \dots, n$$

and, thus, if we want to request the matrix $\mathbf{C}$ to be positive (semi)definite it is sufficient to require $\lambda_C > 0$ (≥ 0), i.e.:

$$C_{ii} - \sum_{i \neq j} |C_{ij}| > 0 \ (\geq 0), \tag{A.15}$$

that is

$$B_{ii} > A_{ii} + \sum_{i \neq j}^{n} |B_{ij} - A_{ij}| \quad \left(B_{ii} \geq A_{ii} + \sum_{i \neq j}^{n} |B_{ij} - A_{ij}| \right), \qquad \text{for } i = 1, \dots, n.$$

Equation (A.15) is the definition of diagonally dominant with positive diagonal entries for matrix $\mathbf{C} \in \mathcal{S}^n$, and so, as a result of Corollary A.0.5, $\mathbf{B} - \mathbf{A} \in \mathcal{S}^n_{++}$ $(\mathbf{B} - \mathbf{A} \in \mathcal{S}^n_{+})$. $\square$

Appendix B

Properties of multivariate normal distribution

In the initial part of this appendix we report the main properties of the multivariate normal distribution. For a review on the topic see Rencher (2002) and Rencher and Schaalje (2007). Afterwards, we report some remarkable results that allow to specify valid covariance matrices through the specification of simpler conditional and marginal forms.

B.1 Multivariate normal distributions

The density of a normal random vector $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ with mean $\boldsymbol{\mu} \in \mathbb{R}^n$ and covariance matrix $\boldsymbol{\Sigma} \in \mathcal{S}^n_{++}$, is

$$f_X(\boldsymbol{x}) = (2\pi)^{n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\}$$

Here, $\mu_i = \mathbb{E}[X_i]$, $\Sigma_{ij} = Cov(X_i, X_j)$, $\Sigma_{ii} = Var(X_i)$ and $Corr(X_i, X_j) = \Sigma_{ij}/(\Sigma_{ii}\Sigma_{jj})^{-1/2}$. This is analogous to write $\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We now divide the vector $\boldsymbol{X}$ into three parts, $\boldsymbol{X} = (\boldsymbol{X}_A^\top, \boldsymbol{X}_B^\top, \boldsymbol{X}_C^\top)^\top$, and split $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ accordingly:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \\ \boldsymbol{\mu}_C \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_A & \boldsymbol{\Sigma}_{AB} & \boldsymbol{\Sigma}_{AC} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_B & \boldsymbol{\Sigma}_{BC} \\ \boldsymbol{\Sigma}_{CA} & \boldsymbol{\Sigma}_{CB} & \boldsymbol{\Sigma}_C \end{pmatrix}.$$

Then for one of the main property of the normal distribution we have:

$$\boldsymbol{X}_A \sim \mathcal{N}_{n_A}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_A), \quad \boldsymbol{X}_B \sim \mathcal{N}_{n_B}(\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B) \quad \text{and} \quad \boldsymbol{X}_C \sim \mathcal{N}_{n_C}(\boldsymbol{\mu}_C, \boldsymbol{\Sigma}_C).$$

We can consider three normal random vectors $\boldsymbol{Y} \in \mathbb{R}^{n_y}, \boldsymbol{X} \in \mathbb{R}^{n_x}$ and $\boldsymbol{Z} \in \mathbb{R}^{n_z}$ such that $\boldsymbol{Y} \sim \mathcal{N}_{n_y}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$, $\boldsymbol{X} \sim \mathcal{N}_{n_x}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$, $\boldsymbol{Z} \sim \mathcal{N}_{n_z}(\boldsymbol{\mu}_z, \boldsymbol{\Sigma}_z)$ and, we assume that the random vector $(\boldsymbol{Y}^\top, \boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top$ is normally distributed such that

$$(\boldsymbol{Y}^\top, \boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top \sim \mathcal{N}_{n_y+n_x+n_z}\left(\boldsymbol{\mu}_{y,x,z}, \boldsymbol{\Sigma}_{y,x,z}\right),$$

where

$$\boldsymbol{\mu}_{y,x,z} = \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_z \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma}_{y,x,z} = \begin{pmatrix} \boldsymbol{\Sigma}_y & \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yz} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}_{zy} & \boldsymbol{\Sigma}_{zx} & \boldsymbol{\Sigma}_z \end{pmatrix}. \tag{B.1}$$

Positing that also every couple of random vector is characterized by the Gaussian distributions

$$\begin{aligned}
(\boldsymbol{Y}^\top, \boldsymbol{X}^\top)^\top &\sim \mathcal{N}_{n_y+n_x}\left(\boldsymbol{\mu}_{y,x}, \boldsymbol{\Sigma}_{y,x}\right), \\
(\boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top &\sim \mathcal{N}_{n_x+n_z}\left(\boldsymbol{\mu}_{x,z}, \boldsymbol{\Sigma}_{x,z}\right), \\
(\boldsymbol{Y}^\top, \boldsymbol{Z}^\top)^\top &\sim \mathcal{N}_{n_y+n_z}\left(\boldsymbol{\mu}_{y,z}, \boldsymbol{\Sigma}_{y,z}\right),
\end{aligned}$$

we specify some vectors and matrices that are the moments of them:

$$\boldsymbol{\mu}_{y,x} = \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \qquad \boldsymbol{\mu}_{x,z} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_z \end{pmatrix}, \qquad \boldsymbol{\mu}_{y,z} = \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_z \end{pmatrix},$$

$$\boldsymbol{\Sigma}_{y,x} = \begin{pmatrix} \boldsymbol{\Sigma}_y & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_x \end{pmatrix}, \qquad \boldsymbol{\Sigma}_{x,z} = \begin{pmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}_{zx} & \boldsymbol{\Sigma}_z \end{pmatrix}, \qquad \boldsymbol{\Sigma}_{y,z} = \begin{pmatrix} \boldsymbol{\Sigma}_y & \boldsymbol{\Sigma}_{yz} \\ \boldsymbol{\Sigma}_{zy} & \boldsymbol{\Sigma}_z \end{pmatrix}.$$

Furthermore, we define the following matrices:

$$\boldsymbol{\Sigma}_{y\cdot xz} = \begin{pmatrix} \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yz} \end{pmatrix}, \qquad \boldsymbol{\Sigma}_{x\cdot yz} = \begin{pmatrix} \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xz} \end{pmatrix}, \qquad \boldsymbol{\Sigma}_{z\cdot yx} = \begin{pmatrix} \boldsymbol{\Sigma}_{zy} & \boldsymbol{\Sigma}_{zx} \end{pmatrix} \tag{B.2}$$

and, in general, $\boldsymbol{\Sigma}_{xz\cdot y} = \boldsymbol{\Sigma}_{y\cdot xz}^\top$. According to this notation, we want to recall how to find the conditional distributions of the different combinations of the random vectors $\boldsymbol{Y}$,$\boldsymbol{X}$ and $\boldsymbol{Z}$.

Using some notion of normal distribution theory stated in Casella and Berger (2002) we show the following known results using the notation adopted for the thesis.

Proposition B.1.1. *Let $(\boldsymbol{Y}^\top, \boldsymbol{X}^\top)^\top \in \mathbb{R}^{n_y \times n_x}$ be a normally distributed random vector, $(\boldsymbol{Y}^\top, \boldsymbol{X}^\top)^\top \sim \mathcal{N}_{n_y+n_x}(\boldsymbol{\mu}_{y,x}, \boldsymbol{\Sigma}_{y,x})$, then the conditional random vector $\boldsymbol{Y}|\boldsymbol{X}$ is also* *normal,*

$$\boldsymbol{Y}|\boldsymbol{X} \sim \mathcal{N}_{n_y}\left(\boldsymbol{\mu}_{y|x}, \boldsymbol{\Sigma}_{y|x}\right) \tag{B.3}$$

where

$$\begin{aligned}\boldsymbol{\mu}_{y|x} &= \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x)\\ \boldsymbol{\Sigma}_{y|x} &= \boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}.\end{aligned}$$

Proof.

$$\begin{aligned}f_{Y|X}(\boldsymbol{y}|\boldsymbol{x}) &= \frac{f_{Y,X}(\boldsymbol{x},\boldsymbol{y})}{f_X(\boldsymbol{x})}\\ &= \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^{n_y+n_x}\frac{1}{|\boldsymbol{\Sigma}_{y,x}|^{1/2}}\exp\left\{-\frac{1}{2}\begin{pmatrix}\boldsymbol{y}-\boldsymbol{\mu}_y\\ \boldsymbol{x}-\boldsymbol{\mu}_x\end{pmatrix}^\top\boldsymbol{\Sigma}_{y,x}^{-1}\begin{pmatrix}\boldsymbol{y}-\boldsymbol{\mu}_y\\ \boldsymbol{x}-\boldsymbol{\mu}_x\end{pmatrix}\right\}}{\left(\frac{1}{\sqrt{2\pi}}\right)^{n_x}\frac{1}{|\boldsymbol{\Sigma}_x|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_x)^\top\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x)\right\}}.\end{aligned}$$

Using the Schur's formula (A.5) for the determinant of $\boldsymbol{\Sigma}_{y,x}$, the Schur-Barachiewicz inverse formula (A.7) we get:

$$\boldsymbol{\Sigma}_{y,x}^{-1} = \begin{bmatrix}(\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x)^{-1} & -(\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x)^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}\\ -\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}(\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x)^{-1} & \boldsymbol{\Sigma}_x^{-1} + \boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}(\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x)^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}\end{bmatrix}$$

where the Schur complement (A.6) is $\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}$, symmetric, then we get:

$$\begin{aligned}f_{Y|X}(\boldsymbol{y}|\boldsymbol{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^{n_y}\frac{1}{|\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x|}\exp\Bigg\{ &-\frac{1}{2}\Bigg[(\boldsymbol{y}-\boldsymbol{\mu}_y)^\top(\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x)^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_y)(\boldsymbol{x}-\boldsymbol{\mu}_x)^\top\cdot\\ &\cdot\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}(\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x)^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x)+\\ &-2(\boldsymbol{y}-\boldsymbol{\mu}_y)^\top(\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x)^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x)\Bigg]\Bigg\}\\ \propto \exp\Bigg\{ &-\frac{1}{2}\Bigg[(\boldsymbol{y}-\boldsymbol{\mu}_y)^\top(\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x)^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_y)+\\ &-2(\boldsymbol{y}-\boldsymbol{\mu}_y)^\top(\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x)^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x)\Bigg]\Bigg\}\\ \propto \exp\Bigg\{ &-\frac{1}{2}\Bigg[(\boldsymbol{y}-\boldsymbol{\mu}_y-\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x))^\top(\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x)^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_y-\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}\cdot\\ &\cdot(\boldsymbol{x}-\boldsymbol{\mu}_x)) - (\boldsymbol{x}-\boldsymbol{\mu}_x)^\top\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}(\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x)^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x)\Bigg]\Bigg\}\end{aligned}$$

$$\propto \exp\Bigg\{-\frac{1}{2}\Bigg[(\boldsymbol{y}-\boldsymbol{\mu}_y-\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x))^\top(\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x)^{-1}\cdot$$
$$\cdot(\boldsymbol{y}-\boldsymbol{\mu}_y-\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x))\Bigg]\Bigg\}.$$

This is the kernel of a multivariate Gaussian distribution of mean

$$\boldsymbol{\mu}_{y|x}=\boldsymbol{\mu}_y+\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x)$$

and covariance matrix

$$\boldsymbol{\Sigma}_{y|x}=\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x,$$

and so, the conditional distribution $f_{Y|X}(\boldsymbol{y}|\boldsymbol{x})$ is $\mathcal{N}_{n_y}(\boldsymbol{\mu}_{y|x},\boldsymbol{\Sigma}_{y|x})$. □

Consequently,

$$\boldsymbol{X}|\boldsymbol{Y}\sim\mathcal{N}_{n_x}\left(\boldsymbol{\mu}_x+\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_y),\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_x\right)$$

with the covariance matrix $\boldsymbol{\Sigma}_{x|y}=\boldsymbol{\Sigma}_{y,x}/\boldsymbol{\Sigma}_y=\boldsymbol{\Sigma}_x-\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{yx}$.

Proposition B.1.2. *Let* $(\boldsymbol{Y}^\top,\boldsymbol{X}^\top,\boldsymbol{Z}^\top)^\top\in\mathbb{R}^{n_y+n_x+n_z}$ *be a normally distributed random vector,* $(\boldsymbol{Y}^\top,\boldsymbol{X}^\top,\boldsymbol{Z}^\top)^\top\sim\mathcal{N}_{n_y+n_x+n_z}(\boldsymbol{\mu}_{y,x,z},\boldsymbol{\Sigma}_{y,x,z})$*, then the conditional random vectors* $\boldsymbol{Y}|(\boldsymbol{X}^\top,\boldsymbol{Z}^\top)^\top$ *and* $(\boldsymbol{X}^\top,\boldsymbol{Z}^\top)^\top|\boldsymbol{Y}$ *are also Gaussian:*

(i) $\boldsymbol{Y}|(\boldsymbol{X}^\top,\boldsymbol{Z}^\top)^\top\sim\mathcal{N}_{n_y}\left(\boldsymbol{\mu}_{y|x,z},\boldsymbol{\Sigma}_{y|x,z}\right)$*, where*

$$\boldsymbol{\mu}_{y|x,z}=\boldsymbol{\mu}_y+\boldsymbol{\Sigma}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}^{-1}\begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix},$$
$$\boldsymbol{\Sigma}_{y|x,z}=\boldsymbol{\Sigma}_y-\boldsymbol{\Sigma}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}^{-1}\boldsymbol{\Sigma}_{y\cdot xz}^\top;$$

(ii) $(\boldsymbol{X}^\top,\boldsymbol{Z}^\top)^\top|\boldsymbol{Y}\sim\mathcal{N}_{n_x+n_z}\left(\boldsymbol{\mu}_{x,z|y},\boldsymbol{\Sigma}_{x,z|y}\right)$*, where*

$$\boldsymbol{\mu}_{x,z|y}=\begin{pmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_z\end{pmatrix}+\boldsymbol{\Sigma}_{y\cdot xz}^\top\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_y),$$
$$\boldsymbol{\Sigma}_{x,z|y}=\boldsymbol{\Sigma}_{x,z}-\boldsymbol{\Sigma}_{y\cdot xz}^\top\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{y\cdot xz}.$$

Proof. (i)

$$f_{Y|X,Z}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z})=\frac{f_{Y,X,Z}(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})}{f_{X,Z}(\boldsymbol{x},\boldsymbol{z})}$$

$$= \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^{n_y+n_x+n_z} \frac{1}{|\boldsymbol{\Sigma}_{y,x,z}|^{1/2}} \exp\left\{-\frac{1}{2}\begin{pmatrix}\boldsymbol{y}-\boldsymbol{\mu}_y\\ \boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix}^\top \boldsymbol{\Sigma}_{y,x,z}^{-1}\begin{pmatrix}\boldsymbol{y}-\boldsymbol{\mu}_y\\ \boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix}\right\}}{\left(\frac{1}{\sqrt{2\pi}}\right)^{n_x+n_z} \frac{1}{|\boldsymbol{\Sigma}_{x,z}|^{1/2}} \exp\left\{-\frac{1}{2}\begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix}^\top \boldsymbol{\Sigma}_{x,z}^{-1}\begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix}\right\}}$$

Using the Schur's formula (A.5) for the determinant of $\boldsymbol{\Sigma}_{y,x,z}$ and the Schur-Barachiewicz inverse formula (A.7) we gain:

$$\boldsymbol{\Sigma}_{y,x,z}^{-1} = \begin{bmatrix} (\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_{x,z})^{-1} & -(\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_{x,z})^{-1}\boldsymbol{\Sigma}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}^{-1} \\ -\boldsymbol{\Sigma}_{x,z}^{-1}\boldsymbol{\Sigma}_{y\cdot xz}^\top(\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_{x,z})^{-1} & \boldsymbol{\Sigma}_{x,z}^{-1} + \boldsymbol{\Sigma}_{x,z}^{-1}\boldsymbol{\Sigma}_{y\cdot xz}^\top(\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_{x,z})^{-1}\boldsymbol{\Sigma}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}^{-1}\end{bmatrix}$$

where $\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_{x,z} = \boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}^{-1}\boldsymbol{\Sigma}_{y\cdot xz}^\top$ is the Schur component of the block $\boldsymbol{\Sigma}_{x,z}$, like in (A.6), we obtain:

$$\begin{aligned} f_{Y|X,Z}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z}) &= \left(\frac{1}{\sqrt{2\pi}}\right)^{n_y} \frac{1}{|\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_{x,z}|} \exp\left\{-\frac{1}{2}\Big[(\boldsymbol{y}-\boldsymbol{\mu}_y)^\top(\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_{x,z})^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_y)+\right. \\ &\quad + \begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix}^\top \boldsymbol{\Sigma}_{x,z}^{-1}\boldsymbol{\Sigma}_{y\cdot xz}^\top(\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_{x,z})^{-1}\boldsymbol{\Sigma}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}^{-1}\begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix} + \\ &\quad \left. - 2(\boldsymbol{y}-\boldsymbol{\mu}_y)^\top(\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_{x,z})^{-1}\boldsymbol{\Sigma}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}^{-1}\begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix}\Big]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\Big[(\boldsymbol{y}-\boldsymbol{\mu}_y)^\top(\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_{x,z})^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_y)+\right. \\ &\quad \left. - 2\left[\boldsymbol{\Sigma}_{x,z}^{-1}\boldsymbol{\Sigma}_{y\cdot xz}^\top(\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_y)^{-T}(\boldsymbol{y}-\boldsymbol{\mu}_y)\right]^\top \begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix}\Big]\right\}. \end{aligned}$$

Now, using $\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_y = \boldsymbol{\Sigma}_{x,z} - \boldsymbol{\Sigma}_{y\cdot x,z}^\top\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{y\cdot x,z}$, the Schur component of the block $\boldsymbol{\Sigma}_y$, and recalling the Woodbury matrix identity in (A.1), we just keep the part depending on $\boldsymbol{y}$:

$$\begin{aligned} f_{Y|X,Z}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z}) = \exp\Bigg\{-\frac{1}{2}\Bigg[\Bigg(&\boldsymbol{y}-\boldsymbol{\mu}_y - \left(\boldsymbol{\Sigma}_y^{-1}+\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{y\cdot xz}(\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_y)^{-1}\boldsymbol{\Sigma}_{y\cdot xz}^\top\boldsymbol{\Sigma}_y^{-1}\right)^{-1}\cdot \\ &\cdot \boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{y\cdot xz}(\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_y)^{-1}\begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix}\Bigg)^\top \cdot \\ &\cdot \left[\boldsymbol{\Sigma}_y^{-1}+\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{y\cdot x,z}(\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_y)^{-1}\boldsymbol{\Sigma}_{y\cdot x,z}^\top\boldsymbol{\Sigma}_y^{-1}\right]\cdot \\ &\cdot \Bigg(\boldsymbol{y}-\boldsymbol{\mu}_y - \left(\boldsymbol{\Sigma}_y^{-1}+\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{y\cdot xz}(\boldsymbol{\Sigma}_{y,x,z}/\boldsymbol{\Sigma}_y)^{-1}\boldsymbol{\Sigma}_{y\cdot xz}^\top\boldsymbol{\Sigma}_y^{-1}\right)^{-1}\cdot \end{aligned}$$

$$\cdot\, \mathbf{\Sigma}_y^{-1}\mathbf{\Sigma}_{y\cdot x,z}(\mathbf{\Sigma}_{y,x,z}/\mathbf{\Sigma}_y)^{-1}\begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix}\Bigg)\Bigg]\Bigg\}.$$

In addition, noting that

$$\begin{aligned}
\left[\mathbf{\Sigma}_y^{-1}+\mathbf{\Sigma}_y^{-1}\mathbf{\Sigma}_{y\cdot xz}(\mathbf{\Sigma}_{y,x,z}/\mathbf{\Sigma}_y)^{-1}\mathbf{\Sigma}_{y\cdot xz}^{\top}\mathbf{\Sigma}_y^{-1}\right]^{-1}&\mathbf{\Sigma}_y^{-1}\mathbf{\Sigma}_{y\cdot xz}(\mathbf{\Sigma}_{y,x,z}/\mathbf{\Sigma}_y)^{-1}=\\
&=\left(\mathbf{\Sigma}_y-\mathbf{\Sigma}_{y\cdot x,z}\mathbf{\Sigma}_{x,z}^{-1}\mathbf{\Sigma}_{y\cdot xz}^{\top}\right)\mathbf{\Sigma}_y^{-1}\mathbf{\Sigma}_{y\cdot xz}(\mathbf{\Sigma}_{y,x,z}/\mathbf{\Sigma}_y)^{-1}\\
&=\mathbf{\Sigma}_{y\cdot x,z}(\mathbf{\Sigma}_{y,x,z}/\mathbf{\Sigma}_y)^{-1}-\mathbf{\Sigma}_{y\cdot xz}\mathbf{\Sigma}_{x,z}^{-1}\mathbf{\Sigma}_{y\cdot x,z}^{\top}\mathbf{\Sigma}_y^{-1}\mathbf{\Sigma}_{y\cdot xz}(\mathbf{\Sigma}_{y,x,z}/\mathbf{\Sigma}_y)^{-1}\\
&=\mathbf{\Sigma}_{y\cdot xz}\left(\mathbf{I}_{n_y}-\mathbf{\Sigma}_{x,z}^{-1}\mathbf{\Sigma}_{y\cdot xz}^{\top}\mathbf{\Sigma}_y^{-1}\mathbf{\Sigma}_{y\cdot xz}\right)(\mathbf{\Sigma}_{y,x,z}/\mathbf{\Sigma}_y)^{-1}\\
&=\mathbf{\Sigma}_{y\cdot xz}\mathbf{\Sigma}_{x,z}^{-1}\left(\mathbf{\Sigma}_{x,z}-\mathbf{\Sigma}_{y\cdot xz}^{\top}\mathbf{\Sigma}_y^{-1}\mathbf{\Sigma}_{y\cdot xz}\right)(\mathbf{\Sigma}_{y,x,z}/\mathbf{\Sigma}_y)^{-1}\\
&=\mathbf{\Sigma}_{y\cdot xz}\mathbf{\Sigma}_{x,z}^{-1},
\end{aligned}$$

the conditional distribution takes the following form:

$$\begin{aligned}
f_{Y|X,Z}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z})\propto\exp\Bigg\{-\frac{1}{2}\Bigg[\Bigg(\boldsymbol{y}-\boldsymbol{\mu}_y-\mathbf{\Sigma}_{y\cdot xz}\mathbf{\Sigma}_{x,z}^{-1}\begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix}\Bigg)^{\top}\big[\mathbf{\Sigma}_y+\\
-\mathbf{\Sigma}_{y\cdot xz}\mathbf{\Sigma}_{x,z}^{-1}\mathbf{\Sigma}_{y\cdot xz}^{\top}\big]^{-1}\Bigg(\boldsymbol{y}-\boldsymbol{\mu}_y-\mathbf{\Sigma}_{y\cdot xz}\mathbf{\Sigma}_{x,z}^{-1}\begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix}\Bigg)\Bigg]\Bigg\},
\end{aligned}$$

that is the kernel of multivariate normal distribution of mean

$$\boldsymbol{\mu}_{y|x,z}=\boldsymbol{\mu}_y+\mathbf{\Sigma}_{y\cdot xz}\mathbf{\Sigma}_{x,z}^{-1}\begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix}$$

and covariance matrix

$$\mathbf{\Sigma}_{y|x,z}=\mathbf{\Sigma}_y-\mathbf{\Sigma}_{y\cdot xz}\mathbf{\Sigma}_{x,z}^{-1}\mathbf{\Sigma}_{y\cdot xz}^{\top},$$

and so, the conditional distribution $f_{Y|X,Z}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z})$ is $\mathcal{N}_{n_y}(\boldsymbol{\mu}_{y|x,z},\mathbf{\Sigma}_{y|x,z})$.

(ii) through a similar manner, it is possible to proof also this point.

□

Accordingly, it can be expressed the distributions of the remaining combinations of conditional vectors:

$$\boldsymbol{Z}|(\boldsymbol{X}^{\top},\boldsymbol{Y}^{\top})^{\top}\sim\mathcal{N}_{n_z}\left(\boldsymbol{\mu}_z+\mathbf{\Sigma}_{z\cdot xy}\mathbf{\Sigma}_{x,y}^{-1}\begin{pmatrix}\boldsymbol{x}-\boldsymbol{\mu}_x\\ \boldsymbol{y}-\boldsymbol{\mu}_y\end{pmatrix},\mathbf{\Sigma}_z-\mathbf{\Sigma}_{z\cdot xy}\mathbf{\Sigma}_{x,y}^{-1}\mathbf{\Sigma}_{z\cdot xy}^{\top}\right),$$

$$\boldsymbol{X}|(\boldsymbol{Y}^{\top},\boldsymbol{Z}^{\top})^{\top}\sim\mathcal{N}_{n_x}\left(\boldsymbol{\mu}_x+\mathbf{\Sigma}_{x\cdot yz}\mathbf{\Sigma}_{y,z}^{-1}\begin{pmatrix}\boldsymbol{y}-\boldsymbol{\mu}_y\\ \boldsymbol{z}-\boldsymbol{\mu}_z\end{pmatrix},\mathbf{\Sigma}_x-\mathbf{\Sigma}_{x\cdot yz}\mathbf{\Sigma}_{y,z}^{-1}\mathbf{\Sigma}_{x\cdot yz}^{\top}\right);$$

$$(\boldsymbol{Y}^\top, \boldsymbol{X}^\top)^\top | \boldsymbol{Z} \sim \mathcal{N}_{n_y+n_x}\left(\begin{pmatrix}\boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x\end{pmatrix} + \boldsymbol{\Sigma}_{z\cdot yx}^\top \boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_z), \boldsymbol{\Sigma}_{y,x} - \boldsymbol{\Sigma}_{z\cdot yx}^\top \boldsymbol{\Sigma}_z^{-1} \boldsymbol{\Sigma}_{z\cdot yx}\right),$$

$$(\boldsymbol{Y}^\top, \boldsymbol{Z}^\top)^\top | \boldsymbol{X} \sim \mathcal{N}_{n_y+n_z}\left(\begin{pmatrix}\boldsymbol{\mu}_y \\ \boldsymbol{\mu}_z\end{pmatrix} + \boldsymbol{\Sigma}_{x\cdot yz}^\top \boldsymbol{\Sigma}_x^{-1}(\boldsymbol{z} - \boldsymbol{\mu}_x), \boldsymbol{\Sigma}_{y,z} - \boldsymbol{\Sigma}_{x\cdot yz}^\top \boldsymbol{\Sigma}_x^{-1} \boldsymbol{\Sigma}_{x\cdot yz}\right).$$

B.2 Conditions for positive definiteness of a covariance matrix

It is often difficult specifying a valid joint covariance matrix for multivariate random vectors. To avoid this difficulty, we introduce an approach in which we directly specify the joint distribution for a process through the specification of simpler conditional and marginals forms. Having in mind the essential features of the normal distribution from the Section B.1, the novel notation introduced in the thesis allows to clearly express valid joint distribution of a random vector consisting of two and three normal sub-vectors depending on the regression, marginal covariance and conditional covariance matrices.

In this section, we are interested to find the conditions for which the matrix $\boldsymbol{\Sigma}_{y,x,z} \in \mathbb{R}^{3n\times 3n}$ defined in (2.14) is a variance matrix, i.e. symmetric and positive definite, and so $\boldsymbol{\Sigma}_{y,x,z} \in \mathcal{S}^{3n}_{++}$. We state the ensuing Theorem to achieve the positive definiteness of covariance matrix of joint normal distribution of the vector $(\boldsymbol{Y}^\top, \boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top \in \mathbb{R}^{3n}$.

Theorem B.2.1. *Let $\boldsymbol{\Sigma}_{y,x,z} \in \mathcal{S}^{3n}$ expressed as in (2.14). If $\boldsymbol{\Sigma}_z \succ 0$, $\boldsymbol{\Sigma}_{x|z} \succ 0$ and $\boldsymbol{\Sigma}_{y|x,z} \succ 0$, then, for all possible simple regression matrices $\boldsymbol{A}_{x\cdot z}$, $\boldsymbol{A}_{y\cdot x}$, $\boldsymbol{A}_{y\cdot z}$,*

$$\boldsymbol{\Sigma}_{y,x,z} \succ 0.$$

Proof. Considering the covariance matrix $\boldsymbol{\Sigma}_{x,z}$, $\boldsymbol{\Sigma}_{x|z}$ is the Schur complement of the block $\boldsymbol{\Sigma}_z$. Hence, according to Theorem A.0.3, $\boldsymbol{\Sigma}_{x,z} \succ 0$.
Now, considering the matrix $\boldsymbol{\Sigma}_{y,x,z}$ as the form suggested in (A.12), we get

$$\begin{pmatrix}\boldsymbol{\Sigma}_y & \boldsymbol{\Sigma}_{y\cdot xz} \\ \boldsymbol{\Sigma}_{y\cdot xz}^\top & \boldsymbol{\Sigma}_{x,z}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma}_{y|x,z} + \mathbf{A}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}\mathbf{A}_{y\cdot xz}^\top & \boldsymbol{\Sigma}_{y\cdot xz} \\ \boldsymbol{\Sigma}_{y\cdot xz}^\top & \boldsymbol{\Sigma}_{x,z}\end{pmatrix},$$

and, given that $\boldsymbol{\Sigma}_{y|x,z}$, that is the Schur complement of $\boldsymbol{\Sigma}_{x,z}$, is positive definite, in the light of the aforementioned theorem, $\boldsymbol{\Sigma}_{y,x,z} \succ 0$. □

From what has been said so far, if we have a matrix $\mathbf{A} \in \mathcal{S}^n$ expressed in block form like in (A.10) in which $\mathbf{A}_{11} \succ 0$ and $\mathbf{A}_{22} \succ 0$, it is not obvious that it is positive definite unless the Schur complement is. However, assuming the principal block matrices positive definite, thanks to the Gershgorin's Theorem (A.13) and Theorem A.0.6, it is sufficient to require additionally the condition of strictly diagonally dominant for symmetric matrices in (A.14) verified for the Schur complement $\mathbf{A}/\mathbf{A}_{11}$ or $\mathbf{A}/\mathbf{A}_{22}$.

As a result, we propose a further theorem that yields to obtain the positive definiteness of covariance matrix characterizing the joint Gaussian distribution of the vector $(\boldsymbol{Y}^\top, \boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top$.

Theorem B.2.2. *Let $\boldsymbol{\Sigma}_{y,x,z} \in \mathcal{S}^{3n}$ expressed as in (2.14). If $\boldsymbol{\Sigma}_y \succ 0$, $\boldsymbol{\Sigma}_x \succ 0$ and $\boldsymbol{\Sigma}_z \succ 0$, then, for all possible simple regression matrices $\boldsymbol{A}_{x\cdot z}$, $\boldsymbol{A}_{y\cdot x}$, $\boldsymbol{A}_{y\cdot z}$,*

$$\boldsymbol{\Sigma}_{y,x,z} \succ 0,$$

if the following conditions are verified:

(i) $(\boldsymbol{\Sigma}_x)_{ii} > \left(\boldsymbol{A}_{x\cdot z}\boldsymbol{\Sigma}_z\boldsymbol{A}_{x\cdot z}^\top\right)_{ii} + \sum_{i\neq j}\left|\left(\boldsymbol{\Sigma}_x - \boldsymbol{A}_{x\cdot z}\boldsymbol{\Sigma}_z\boldsymbol{A}_{x\cdot z}^\top\right)_{ij}\right| \qquad \forall i = 1, \ldots, n$

(ii) $(\boldsymbol{\Sigma}_y)_{ii} > \left(\boldsymbol{A}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}\boldsymbol{A}_{y\cdot xz}^\top\right)_{ii} + \sum_{i\neq j}\left|\left(\boldsymbol{\Sigma}_y - \boldsymbol{A}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}\boldsymbol{A}_{y\cdot xz}^\top\right)_{ij}\right| \qquad \forall i = 1, \ldots, n.$

Proof. Given that for hypothesis $\boldsymbol{\Sigma}_z \succ 0$, from Theorem B.2.1 $\boldsymbol{\Sigma}_{y,x,z}$ will be positive definite if $\boldsymbol{\Sigma}_{x|z} \succ 0$ and $\boldsymbol{\Sigma}_{y|x,z} \succ 0$. Writing the matrix $\boldsymbol{\Sigma}_{x|z}$ in the following form

$$\boldsymbol{\Sigma}_x - \mathbf{A}_{x\cdot z}\boldsymbol{\Sigma}_z\mathbf{A}_{x\cdot z}^\top,$$

we can see that both matrices are symmetric and positive definite, the first one by assumption and the second because of the result in Equation (A.9). In turn, according to Theorem A.0.6, if condition (i) is true, then $\boldsymbol{\Sigma}_{x|z} \succ 0$. In the same way is possible to show that verify condition (ii) in the theorem's setting is equivalent to assume $\boldsymbol{\Sigma}_{y|x,z} \succ 0$. □

By using an alternative representation of joint covariance matrix expressed on the form (B.1), we can provide the following corollary:

Corollary B.2.3. *Let $\boldsymbol{\Sigma}_{y,x,z} \in \mathcal{S}^{3n}$ expressed as in (2.14). If $\boldsymbol{\Sigma}_y \succ 0$, $\boldsymbol{\Sigma}_x \succ 0$ and $\boldsymbol{\Sigma}_z \succ 0$, then, for all possible cross-covariance matrices $\boldsymbol{\Sigma}_{yx}$, $\boldsymbol{\Sigma}_{yz}$ and $\boldsymbol{\Sigma}_{xz}$,*

$$\boldsymbol{\Sigma}_{y,x,z} \succ 0,$$

if the following conditions are verified

(i) $\boldsymbol{\Sigma}_x - \boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_z^{-1}\boldsymbol{\Sigma}_{xz}^\top \succ 0,$

(ii) $\boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}^{-1}\boldsymbol{\Sigma}_{y\cdot xz}^\top \succ 0.$

Proof. The results follow from Theorem B.2.1 and Theorem A.0.3. □

Afterwards, the next lemma gives a tool to build a valid joint covariance matrix of $(\boldsymbol{Y}^\top, \boldsymbol{X}^\top, \boldsymbol{Z}^\top)^\top$ assuming scalar cross-correlation matrix of $\boldsymbol{X}$ and $\boldsymbol{Z}$.

Lemma B.2.4. *Let $\boldsymbol{A}, \boldsymbol{B}$ be two positive definite matrices, i.e. $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{S}_+^n$. Every eigenvalue of $\boldsymbol{A} - \boldsymbol{B}$ is contained in the interval*

$$[\lambda(\boldsymbol{A})_1 - \lambda(\boldsymbol{B})_n,\ \lambda(\boldsymbol{A})_n - \lambda(\boldsymbol{B})_1],$$

where $\lambda(\boldsymbol{A})_1$ is the minimum eigenvalue of the matrix $\boldsymbol{A}$ and $\lambda(\boldsymbol{B})_n$ is the maximum eigenvalue of the matrix $\boldsymbol{B}$.

Proof. Recalling one of the Weyl Inequalities (Weyl, 1912):

$$\lambda(\mathbf{A})_i + \lambda(\mathbf{B})_{j-i+1} \leq \lambda(\mathbf{A}+\mathbf{B})_j \leq \lambda(\mathbf{A})_k + \lambda(\mathbf{B})_{j-k+n},$$

for every integer $1 \leq i \leq j \leq k \leq n$. A conseguence of this inequalities is the following:

$$\lambda(\mathbf{A})_1 + \lambda(\mathbf{B})_1 \leq \lambda(\mathbf{A}+\mathbf{B})_j \leq \lambda(\mathbf{A})_n + \lambda(\mathbf{B})_n \qquad \text{for } j = 1, \dots, n.$$

Thanks to the positive definiteness of the matrices, if we consider the sum of $\mathbf{A}$ and $-\mathbf{B}$, the expected result arises:

$$\lambda(\mathbf{A})_1 - \lambda(\mathbf{A})_n \leq \lambda(\mathbf{A} - \mathbf{B})_j \leq \lambda(\mathbf{A})_n - \lambda(\mathbf{B})_1 \qquad \text{for } j = 1, \dots, n.$$

$\square$

Corollary B.2.5. *Let $\boldsymbol{\Sigma}_{y,x,z} \in \mathcal{S}^{3n}$ expressed as in (2.14). If $\boldsymbol{\Sigma}_x \succ 0$, $\boldsymbol{\Sigma}_z \succ 0$ and $\boldsymbol{\Sigma}_{y|x,z} \succ 0$, then, for all possible cross-covariance matrices $\boldsymbol{\Sigma}_{yx}$, $\boldsymbol{\Sigma}_{yz}$ and*

$$\boldsymbol{\Sigma}_{xz} = \rho_{xz} \boldsymbol{I}_n,$$

$\boldsymbol{\Sigma}_{y,x,z} \succ 0$, if

$$-\sqrt{\min\left(\boldsymbol{\lambda}(\boldsymbol{\Sigma}_x) \circ \boldsymbol{\lambda}(\boldsymbol{\Sigma}_z)\right)} < \rho_{xz} < \sqrt{\min\left(\boldsymbol{\lambda}(\boldsymbol{\Sigma}_x) \circ \boldsymbol{\lambda}(\boldsymbol{\Sigma}_z)\right)},$$

where $\circ$ represents the Hadamard product[1].

Proof. According to Theorem B.2.1 $\boldsymbol{\Sigma}_{y,x,z} \succ 0$ if $\boldsymbol{\Sigma}_{x,z} \succ 0$. From Theorem A.0.3, it is positive definite if $\boldsymbol{\Sigma}_{x|z} \succ 0$. In this setting, $\boldsymbol{\Sigma}_{x|z} = \boldsymbol{\Sigma}_x - \rho_{xz}^2 \boldsymbol{\Sigma}_z^{-1}$. From Lemma B.2.4 we gain:

$$\lambda(\boldsymbol{\Sigma}_x)_1 - \rho_{xz}^2 \lambda(\boldsymbol{\Sigma}_z^{-1})_n \leq \lambda(\boldsymbol{\Sigma}_{x|z})_j \leq \lambda(\boldsymbol{\Sigma}_x)_n - \rho_{xz}^2 \lambda(\boldsymbol{\Sigma}_z^{-1})_1.$$

[1] Also known as the element-wise product.

Hence, $\mathbf{\Sigma}_{x,z} \succ 0$ if $\lambda(\mathbf{\Sigma}_x)_1 - \rho_{xz}^2 \lambda(\mathbf{\Sigma}_z^{-1})_n > 0$, that is:

$$\rho_{xz}^2 < \lambda(\mathbf{\Sigma}_x)_1 \lambda(\mathbf{\Sigma}_z)_1,$$

and then it emerges that

$$-\sqrt{\lambda(\mathbf{\Sigma}_x)_1 \lambda(\mathbf{\Sigma}_z)_1} < \rho_{xz} < \sqrt{\lambda(\mathbf{\Sigma}_x)_1 \lambda(\mathbf{\Sigma}_z)_1}.$$

Considering that $\min\big(\boldsymbol{\lambda}(\mathbf{\Sigma}_x) \circ \boldsymbol{\lambda}(\mathbf{\Sigma}_z)\big) = \lambda(\mathbf{\Sigma}_x)_1 \lambda(\mathbf{\Sigma}_z)_1$, we get the statement. □

In order to consider different kind of structured cross-covariance matrices, in the following theorem we provide a sufficient condition to build a valid joint covariance matrix $\mathbf{\Sigma}_{y,x,z}$ by fixing the ones of the marginal random vectors, $\mathbf{\Sigma}_y$, $\mathbf{\Sigma}_x$ and $\mathbf{\Sigma}_z$ and using the Cholesky decomposed matrices to construct the cross-correlation matrices.

Theorem B.2.6. *Let $\mathbf{\Sigma}_y$, $\mathbf{\Sigma}_x$, $\mathbf{\Sigma}_z \in \mathcal{S}_{++}^n$ be the covariance matrices of the random vectors $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{Z}$ and let ρ_{yx}, ρ_{yz} and ρ_{xz} the simple correlations parameters.*

Fixing $\rho_{yx}, \rho_{xz} \in (-1, 1)$ and considering the Cholesky factorization of the marginal covariance matrices such that $\mathbf{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$. The cross-covariance matrices $\mathbf{\Sigma}_{yx}$, $\mathbf{\Sigma}_{yz}$ and $\mathbf{\Sigma}_{xz}$ are build as follows

$$\mathbf{\Sigma}_{yx} = \rho_{yx} \boldsymbol{L}_y \boldsymbol{L}_x^\top, \qquad \mathbf{\Sigma}_{yz} = \rho_{yz} \boldsymbol{L}_y \boldsymbol{L}_z^\top \qquad \text{and} \qquad \mathbf{\Sigma}_{xz} = \rho_{xz} \boldsymbol{L}_x \boldsymbol{L}_z^\top.$$

Then

$$\mathbf{\Sigma}_{y,x,z} \succ 0,$$

if

$$\rho_{yz} \in \left(\rho_{yx}\rho_{xz} - \sqrt{(1-\rho_{xz}^2)(1-\rho_{yx}^2)},\, \rho_{yx}\rho_{xz} + \sqrt{(1-\rho_{xz}^2)(1-\rho_{yx}^2)}\right) \tag{B.4}$$

Proof. Given that, for assumption $\mathbf{\Sigma}_y$, $\mathbf{\Sigma}_x$, $\mathbf{\Sigma}_z$ are positive definite, they have unique Cholesky decompositions as follow:

$$\mathbf{\Sigma}_y = \mathbf{L}_y \mathbf{L}_y^\top, \quad \mathbf{\Sigma}_x = \mathbf{L}_x \mathbf{L}_x^\top \quad \text{and} \quad \mathbf{\Sigma}_z = \mathbf{L}_z \mathbf{L}_z^\top.$$

We obtain that the last two conditions in Corollary B.2.3 are verified if $\rho_{xz} \in (-1, 1)$ (for assumption) and (B.4) holds:

(i) $\mathbf{\Sigma}_x - \mathbf{\Sigma}_{xz} \mathbf{\Sigma}_z^{-1} \mathbf{\Sigma}_{xz}^\top = \mathbf{\Sigma}_x - \rho_{xz}^2 \mathbf{L}_x \mathbf{L}_z^\top \mathbf{L}_z^{-T} \mathbf{L}_z^{-1} \mathbf{L}_z \mathbf{L}_x^\top = (1-\rho_{xz}^2)\mathbf{\Sigma}_x \succ 0$;

(ii) Noting that $(1-\rho_{xz}^2)\mathbf{\Sigma}_x$ is the Schur complement of the block $\mathbf{\Sigma}_z$ of the matrix $\mathbf{\Sigma}_{x,z}$, we verify the second condition.

$$
\begin{aligned}
&\boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}^{-1}\boldsymbol{\Sigma}_{y\cdot xz}^{\top} = \\
&= \boldsymbol{\Sigma}_y - \left[\boldsymbol{\Sigma}_{yx} : \boldsymbol{\Sigma}_{yz}\right] \begin{bmatrix} \frac{1}{1-\rho_{xz}^2}\boldsymbol{\Sigma}_x^{-1} & -\frac{1}{1-\rho_{xz}^2}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_z^{-1} \\ -\frac{1}{1-\rho_{xz}^2}\boldsymbol{\Sigma}_z^{-1}\boldsymbol{\Sigma}_{zx}\boldsymbol{\Sigma}_x^{-1} & \boldsymbol{\Sigma}_z^{-1} + \frac{1}{1-\rho_{xz}^2}\boldsymbol{\Sigma}_z^{-1}\boldsymbol{\Sigma}_{zx}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xz}\boldsymbol{\Sigma}_z^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{yx}^{\top} \\ \boldsymbol{\Sigma}_{yz}^{\top} \end{bmatrix} \\
&= \boldsymbol{\Sigma}_y - \left[\rho_{yx}\mathbf{L}_y\mathbf{L}_x^{\top} \quad : \quad \rho_{yz}\mathbf{L}_y\mathbf{L}_z^{\top}\right] \cdot \\
&\cdot \begin{bmatrix} \frac{1}{1-\rho_{xz}^2}\mathbf{L}_x^{-T}\mathbf{L}_x^{-1} & -\frac{\rho_{xz}}{1-\rho_{xz}^2}\mathbf{L}_x^{-T}\mathbf{L}_x^{-1}\mathbf{L}_x\mathbf{L}_z^{\top}\mathbf{L}_z^{-T}\mathbf{L}_z^{-1} \\ -\frac{\rho_{xz}}{1-\rho_{xz}^2}\mathbf{L}_z^{-T}\mathbf{L}_z^{-1}\mathbf{L}_z\mathbf{L}_x^{\top}\mathbf{L}_x^{-T}\mathbf{L}_x^{-1} & \mathbf{L}_z^{-T}\mathbf{L}_z^{-1} + \frac{\rho_{xz}^2}{1-\rho_{xz}^2}\mathbf{L}_z^{-T}\mathbf{L}_z^{-1}\mathbf{L}_z\mathbf{L}_x^{\top}\mathbf{L}_x^{-T}\mathbf{L}_x^{-1}\mathbf{L}_x\mathbf{L}_z^{\top}\mathbf{L}_z^{-T}\mathbf{L}_z^{-1} \end{bmatrix} \cdot \\
&\cdot \begin{bmatrix} \rho_{yx}\mathbf{L}_x\mathbf{L}_y^{\top} \\ \rho_{yz}\mathbf{L}_z\mathbf{L}_y^{\top} \end{bmatrix} \\
&= \boldsymbol{\Sigma}_y - \left[\rho_{yx}\mathbf{L}_y\mathbf{L}_x^{\top} \quad : \quad \rho_{yz}\mathbf{L}_y\mathbf{L}_z^{\top}\right] \begin{bmatrix} \frac{1}{1-\rho_{xz}^2}\mathbf{L}_x^{-T}\mathbf{L}_x^{-1} & -\frac{\rho_{xz}}{1-\rho_{xz}^2}\mathbf{L}_x^{-T}\mathbf{L}_z^{-1} \\ -\frac{\rho_{xz}}{1-\rho_{xz}^2}\mathbf{L}_z^{-T}\mathbf{L}_x^{-1} & \left(1 + \frac{\rho_{xz}^2}{1-\rho_{xz}^2}\right)\mathbf{L}_z^{-T}\mathbf{L}_z^{-1} \end{bmatrix} \begin{bmatrix} \rho_{yx}\mathbf{L}_x\mathbf{L}_y^{\top} \\ \rho_{yz}\mathbf{L}_z\mathbf{L}_y^{\top} \end{bmatrix} \\
&= \boldsymbol{\Sigma}_y - \left[\left(\frac{\rho_{yx}-\rho_{yz}\rho_{xz}}{1-\rho_{xz}^2}\right)\mathbf{L}_y\mathbf{L}_x^{-1} \qquad \left(\rho_{yz} + \frac{\rho_{yz}\rho_{xz}^2-\rho_{yx}\rho_{xz}}{1-\rho_{xz}^2}\right)\mathbf{L}_y\mathbf{L}_z^{-1}\right] \begin{bmatrix} \rho_{yx}\mathbf{L}_x\mathbf{L}_y^{\top} \\ \rho_{yz}\mathbf{L}_z\mathbf{L}_y^{\top} \end{bmatrix} \\
&= \boldsymbol{\Sigma}_y - \left[\left(\frac{\rho_{yx}^2-\rho_{yx}\rho_{yz}\rho_{xz}}{1-\rho_{xz}^2}\right)\mathbf{L}_y\mathbf{L}_y^{T} + \left(\rho_{yz}^2 + \frac{\rho_{yz}^2\rho_{xz}^2-\rho_{yz}\rho_{yx}\rho_{xz}}{1-\rho_{xz}^2}\right)\mathbf{L}_y\mathbf{L}_y^{T}\right] \\
&= \left[1 - \frac{\rho_{yz}^2 - \rho_{yz}^2\rho_{xz}^2}{1-\rho_{xz}^2} - \frac{\rho_{yx}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz} + \rho_{yz}^2\rho_{xz}^2}{1-\rho_{xz}^2}\right]\boldsymbol{\Sigma}_y \\
&= \left[1 - \left(\rho_{yz}^2 + \frac{(\rho_{yx}-\rho_{yz}\rho_{xz})^2}{1-\rho_{xz}^2}\right)\right]\boldsymbol{\Sigma}_y.
\end{aligned}
$$

Hence, $\boldsymbol{\Sigma}_{y|xz} = \boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{y\cdot xz}\boldsymbol{\Sigma}_{x,z}^{-1}\boldsymbol{\Sigma}_{y\cdot xz}^{\top} \succ 0$ if $\left[1 - \left(\rho_{yz}^2 + \frac{(\rho_{yx}-\rho_{yz}\rho_{xz})^2}{1-\rho_{xz}^2}\right)\right] > 0$, that is:

$$
\begin{aligned}
\boldsymbol{\Sigma}_{y|xz} \succ 0 \iff & 1 - \frac{\rho_{yx}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz} + \rho_{yz}^2}{1-\rho_{xz}^2} > 0 \\
& \rho_{yx}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz} + \rho_{yz}^2 < 1 - \rho_{xz}^2 \\
& \rho_{yx}\rho_{xz} - \sqrt{(1-\rho_{xz}^2)(1-\rho_{yx}^2)} < \rho_{yz} < \rho_{yx}\rho_{xz} + \sqrt{(1-\rho_{xz}^2)(1-\rho_{yx}^2)}.
\end{aligned}
$$

Moreover, it is trivial to proof that

$$\rho_{yz} \in \left(\rho_{yx}\rho_{xz} - \sqrt{(1-\rho_{xz}^2)(1-\rho_{yx}^2)},\ \rho_{yx}\rho_{xz} + \sqrt{(1-\rho_{xz}^2)(1-\rho_{yx}^2)}\right) \subseteq (-1,1).$$

In fact, it is sufficient to show that the inequality $\rho_{yx}\rho_{xz} - \sqrt{(1-\rho_{xz}^2)(1-\rho_{yx}^2)} \geq 0$ is always true:

$$\begin{aligned}\sqrt{(1-\rho_{xz}^2)(1-\rho_{yx}^2)} &\leq 1 + \rho_{yx}\rho_{xz}\\ (1-\rho_{xz}^2)(1-\rho_{yx}^2) &\leq 1 + 2\rho_{yx}\rho_{xz} + \rho_{yx}^2\rho_{xz}^2\\ (\rho_{yx} + \rho_{xz})^2 &\geq 0.\end{aligned}$$

Among the assumptions of the theorem there is the requirement that the definition interval of ρ_{yx} is open to avoid degenerate covariance matrices. □

A more simplified version of this theorem as a way to build a valid cross covariance matrix is used for the fist time by Page et al. (2017) in spatial literature regarding confounding, in order to work in a solid set up to study the problem. That is what we have indicated with rLMC model. Moreover, it is interesting to note that, in a regression model as in (2.1), the quantity $\rho_{yz}^2 + \frac{(\rho_{yx}-\rho_{yz}\rho_{xz})^2}{1-\rho_{xz}^2}$ is the *coefficient of determination* R^2 of a linear regression model with regressors X and Z. It can be express trough the semipartial correlation $\rho_{yz|x}$:

$$\begin{aligned}R^2 &= \rho_{yz}^2 + \frac{(\rho_{yx}-\rho_{yz}\rho_{xz})^2}{1-\rho_{xz}^2} = \rho_{yx}^2 + \frac{(\rho_{yz}-\rho_{yx}\rho_{xz})^2}{1-\rho_{xz}^2}\\ &= \rho_{yz}^2 + \rho_{yx|z}^2 = \rho_{yx}^2 + \rho_{yz|x}^2.\end{aligned}$$

Thus, requesting the positive definiteness of $\mathbf{\Sigma}_y$ is equivalent to ask for $R^2 \in [0,1]$ and this means that $\rho_{yx}^2 + \rho_{yz|x}^2 \in [0,1]$.

Remark B.2.1. If $\mathbf{\Sigma}_{y,x,z}$ is positive definite, then equivalently $\hat{\mathbf{\Sigma}}_{y,x,z} = \mathbf{\Pi}^\top \mathbf{\Sigma}_{y,x,z}\mathbf{\Pi}$ is positive definite for $\mathbf{\Pi}$ being a permutation matrix. Therefore, according to the chosen permutation, everything we said for the sub-matrices $\mathbf{\Sigma}_z$,$\mathbf{\Sigma}_{x|z}$ and $\mathbf{\Sigma}_{y|x,z}$ can be modified for other sub-matrix combinations.

Appendix C

Eigenvalue of autoregressive process of first order

In this appendix, according to the work of Yueh (2005), an alternative proposal to Stroeker (1983) for the approximation of precision matrix eigenvalues of an autoregressive of order one process are provided.

A zero-mean autoregressive process of order one with normal errors can be expressed as

$$X_t = \phi X_{t-1} + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2), \tag{C.1}$$

where the index $t \in \{1, \dots, n\}$ represents time. We consider the zero-mean AR1 process here because a mean term can be added in any time. With these assumptions, the distribution of the process is normal with mean 0 and variance $\sigma^2/(1-\phi^2)$, where σ^2 is the conditional variance and ϕ the autocorrelation. This implies the importance of ϕ to be less than 1 (in absolute value) in order that the process be stable (Siegel and Wagner, 2022). We can express (C.1) in the direct conditional form

$$X_t | X_1, \dots, X_{t-1} \sim \mathcal{N}(\phi X_{t-1}, \sigma^2)$$

for $t \in \{2, \dots, n\}$. As shown in Lindsey (2004) and Rue and Held (2005), if stationarity is assumed, an alternative specification of an AR1 can also be used: direct construction of a multivariate Gaussian distribution. The joint distribution of $\boldsymbol{X} = (X_1, \dots, X_n)^T$ has zero mean and a covariance matrix given by

$$\boldsymbol{\Sigma} = \frac{\sigma^2}{1-\phi^2}\begin{bmatrix} 1 & \phi & \dots & \phi^{n-2} & \phi^{n-1} \\ \phi & 1 & \dots & \phi^{n-3} & \phi^{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi^{n-2} & \phi^{n-3} & \dots & 1 & \phi \\ \phi^{n-1} & \phi^{n-2} & \dots & \phi & 1 \end{bmatrix}. \tag{C.2}$$

It is a dense matrix with entries

$$\Sigma_{ij} = \frac{\sigma^2}{1-\phi^2}\phi^{|i-j|},$$

where $\phi^{|i-j|}$ is the correlation function of the AR1 process. The precision matrix has a special form, with zeros everywhere except on the main and first minor diagonals:

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2}\begin{bmatrix} 1 & -\phi & \dots & 0 & 0 \\ -\phi & 1+\phi^2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1+\phi^2 & -\phi \\ 0 & 0 & \dots & -\phi & 1 \end{bmatrix}. \tag{C.3}$$

The tridiagonal form is due to the fact that X_s and X_t with $1 \leq s < t \leq n$ are conditionally independent given $\{X_{s+1}, \dots, X_{t-1}\}$ if $|t-s| > 1$ (Rue and Held, 2005). The autoregressive process of order 1 is a special case of a Gaussian Markov random field in which, in general, it is possible notice the relationship between conditional independence and the sparse structure of the precision matrix.

Aware about the important role of covariate matrix eigenvalues in the evaluation of confounding, it might be useful to have information on the eigenvalues of $\boldsymbol{\Sigma}^{-1}$ in (C.3) in explicit form. This matrix is "nearly" a tridiagonal Toeplitz matrix for which the eigenvalues and eigenvectors are known in closed form (Grenander and Szegö, 1958). In fact, considering $\tilde{\boldsymbol{\Sigma}}^{-1} = \boldsymbol{\Sigma}^{-1} + \phi^2\boldsymbol{E}_n$ where $\boldsymbol{E}_n = \operatorname{diag}(1, 0, \dots, 0, 1)$ is a $n \times n$ diagonal matrix, $\tilde{\boldsymbol{\Sigma}}^{-1}$ is a Toeplitz matrix of finite order and then, by Stroeker (1983) and Yang (2021), we obtain an explicit form for the entire set of eigenvalues

$$\lambda_k(\tilde{\boldsymbol{\Sigma}}^{-1}) = 1 + \phi^2 - 2\phi\cos\left(\frac{k\pi}{n+1}\right) \qquad k = 1, \dots, n,$$

and the corresponding normalized eigenvector $\tilde{\boldsymbol{u}}_k = (\tilde{u}_{1k}, \dots, \tilde{u}_{nk})$ is

$$\tilde{u}_{jk} = \sqrt{\frac{2}{n+1}}\sin\left(\frac{kj\pi}{n+1}\right) \qquad j, k = 1, \dots, n.$$

The above expression is for $\phi > 0$, and when $\phi < 0$, the eigenstructure has the same expressions but is arranged in the reverse order.

For example, the precision matrix of a Random Walk process of order one, i.e. autoregressive process of order 1 with $\phi = 1$, is a Toeplitz matrix with 2 and -1 on main and first minor diagonals, respectively. Consequently, and as also previously demonstrated by Elliott (1953) and Gregory and Karney (1969), its eigenvalues are

$$\lambda_k = 2 - 2\phi \cos\left(\frac{k\pi}{n+1}\right) \qquad\qquad k = 1, \ldots, n.$$

Moreover Stroeker (1983) gives certain approximations of eigenvalues of $\mathbf{\Sigma}^{-1}$, which are expecially useful for large n, that are:

$$\lambda_k(\mathbf{\Sigma}^{-1}) = 1 + \phi^2 - 2\phi \cos\left(\frac{k\pi}{n+1}\right) - \frac{4\phi^2}{n+1} \sin^2\left(\frac{k\pi}{n+1}\right) \qquad\qquad k = 1, \ldots, n.$$

The following error bounds for each λ_k

$$\epsilon_k = \phi^2 \sqrt{\frac{4}{n+1}} \sin\left(\frac{k\pi}{n+1}\right) \qquad\qquad k = 1, \ldots, n$$

show that for increasing n the approximations improve.

In order to find more precise approximations for such eigenvalues, we follow Yueh (2005) deriving the eigenvalues of a generic tridiagonal matrix of the form

$$\boldsymbol{M} = \begin{bmatrix} -\alpha + b & c & 0 & \ldots & 0 & 0 \\ a & b & c & 0 & \ldots & 0 \\ 0 & a & b & c & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & 0 & a & b & c \\ 0 & 0 & \ldots & 0 & a & -\beta + b \end{bmatrix}.$$

The author determines them by the method of symbolic calculus in Cheng (2003), providing the necessary condition to obtain them. In this article the author deal with the eigenvalue problem $\boldsymbol{Mu} = \lambda \boldsymbol{u}$, where a, b, c and α, β are number in $\mathbb{C}$, instead, we consider the real domain in which $a = c = -\phi$, $b = \phi^2 + 1$ and $\alpha = \beta = \phi^2$, in order to work on the precision matrix of AR1 process (C.3).

Theorem C.0.1. *Considering an autoregressive process of the first order, a good approx-*

imation of each precision matrix $\mathbf{\Sigma}^{-1}$ *eigenvalues and* $\sigma^2 = 1$ *is:*

$$\lambda_k(\mathbf{\Sigma}^{-1}) = 1 + \phi^2 + 2\phi \cos\left(\frac{k\pi}{n - \frac{\phi-1}{\phi+1}}\right) \qquad k = 1, \dots, n.$$

Proof. According to Yueh (2005), we consider the quantity

$$\gamma_{\pm} = \frac{-(b-\lambda) \pm \sqrt{(b-\lambda)^2 - 4ac}}{2a}.$$

Let $\gamma_{\pm} = p \pm iq$ where $p, q \in \mathbb{C}$ and $q \neq 0$. It becomes

$$\begin{aligned}\gamma_{\pm} &= \frac{-(\phi^2+1-\lambda) \pm \sqrt{(\phi^2+1-\lambda)^2 - 4\phi^2}}{-2\phi} \\ &= \frac{\lambda - (\phi^2+1)}{-2\phi} \pm i^2 \frac{\sqrt{(\phi^2+1-\lambda)^2 - 4\phi^2}}{2\phi}.\end{aligned}$$

Since, $\gamma_+ + \gamma_- = 2p = [\lambda - (\phi^2+1)]/(-\phi)$ and

$$\begin{aligned}\gamma_+\gamma_- &= p^2 + q^2 \\ &= \left(\frac{\lambda - (\phi^2+1)}{-2\phi}\right)^2 + \left(i\frac{\sqrt{(\phi^2+1-\lambda)^2 - 4\phi^2}}{2\phi}\right)^2 = 1 = \frac{c}{a},\end{aligned}$$

we may write

$$\gamma_{\pm} = \sqrt{p^2+q^2}(\cos\theta \pm i \sin\theta)$$

where

$$\cos\theta = \frac{p}{\sqrt{p^2+q^2}} = \frac{\lambda - (\phi^2+1)}{2\phi}, \qquad \theta \in \mathbb{R}. \tag{C.4}$$

Given that $\gamma_+ \neq \gamma_-$, using the Chebyshev polynomials of the second kind for θ, it is achieved the necessary condition to gain the eigenvalues of the precision matrix:

$$\phi^2 \sin[(n+1)\theta] + 2\phi^3 \sin(n\theta) + \phi^4 \sin[(n-1)\theta] = 0.$$

Using the arbitrary phase shift for more than two sinusoids from the Harmonic Addition Theorem (Oo and Gan, 2012) yields to the following equality

$$\sqrt{1 + 2\phi^2 + \phi^4 + 2\phi[\phi\cos\theta - 2(1+\phi^2)cos\theta]}\, \sin\left(n\theta + \tan^{-1}\left(\frac{\sin\theta(1-\phi^2)}{(1+\phi^2)\cos\theta + 2\phi}\right)\right) = 0.$$

Assuming $\phi \neq 0$, since on the contrary case is easy, the last becomes

$$\sin\left(n\theta + \tan^{-1}\left(\frac{\sin\theta(1-\phi^2)}{(1+\phi^2)\cos\theta + 2\phi}\right)\right) = 0.$$

Through the Maclaurin series of $\tan^{-1}(\cdot)$ up to the second order, we find the solution

$$\theta = \frac{k\pi}{n - \frac{\phi-1}{\phi+1}},$$

solving the coming equation

$$n\theta - \theta\left(\frac{\phi-1}{\phi+1}\right) = k\pi.$$

Then, by (C.4) we have the formula for the k-th eigenvalue:

$$\lambda_k = 1 + \phi^2 + 2\phi\cos\theta \qquad\qquad \theta \neq m\pi,\ m \in \mathbb{Z}.$$

$\square$

Bibliography

A. Adin, T. Goicoa, J. S. Hodges, P. M. Schnell, and M. D. Ugarte. Alleviating confounding in spatio-temporal areal models with an application on crimes against women in India. *Statistical Modelling*, 2021.

M. P. Allen. *Understanding Regression Analysis.* Springer US, Boston, MA, 1997.

T. W. Anderson. *An introduction to multivariate statistical analysis.* John Wiley & Sons Inc, New York, third edition, 1984.

P. Aronow and B. Miller. *Foundations of Agnostic Statistics.* Cambridge University Press, 2019.

D. R. M. Azevedo, D. Bandyopadhyay, M. Prates, A.-S. G. Abdel-Salam, and D. Garcia. Assessing spatial confounding in cancer disease mapping using R. *Cancer Report (Hoboken)*, 3(4), 2020.

D. R. M. Azevedo, M. O. Prates, and D. Bandyopadhyay. Alleviating spatial confounding in frailty models. *Biostatistics*, 2022.

S. Banerjee and A. Roy. *Linear Algebra and Matrix Analysis for Statistics.* Texts in statistical science. CRC Press, 2014.

S. Banerjee, B. Carlin, and A. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data.* Chapman and Hall/CRC, second edition, 2014.

Y. Bao and R. Kan. On the moments of ratios of quadratic forms in normal random variables. *Journal of Multivariate Analysis*, 117:229–245, 2013.

L. Bernardinelli, D. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi, and M. Songini. Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*, 14 (21-22):2433–2443, 1995.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 36(2):192–236, 1974.

P. Billingsley. *Probability and Measure.* John Wiley and Sons, New York, third edition, 1995.

J. R. Bradley, S. H. Holan, and C. K. Wikle. Multivariate spatio-temporal models for high-dimensional areal data with application to Longitudinal Employer-Household Dynamics. *The Annals of Applied Statistics*, 9(4):1761 – 1791, 2015.

N. E. Breslow and D. G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, 1993.

B. Carlson. Lauricella's hypergeometric function fd. *Journal of Mathematical Analysis and Applications*, 7(3):452–470, 1963.

G. Casella and R. L. Berger. *Statistical inference.* Belmont, CA, Duxbury, 2002.

S. Cheng. *Partial Difference Equations.* Advances in discrete mathematics and applications. Taylor & Francis, 2003.

D. G. Clayton, L. Bernardinelli, and C. Montomoli. Spatial correlation in ecological analysis. *International journal of epidemiology*, 22(6):1193–1202, 1993.

N. Cressie. *Statistics for Spatial Data.* Probability and Mathematical Statistics. Wiley-Interscience, revised edition, 1993.

N. Cressie, A. S. Davis, J. L. Folks, and G. E. Policello. The moment-generating function and negative integer moments. *The American Statistician*, 35(3):148–150, 1981.

C. Dandine-Roulland and H. Perdry. The use of the linear mixed model in human geneticss. *Human Heredity*, 80(4):196–206, 2019.

M. L. Davis, B. Neelon, P. J. Nietert, K. J. Hunt, L. F. Burgette, A. B. Lawson, and L. E. Egede. Addressing geographic confounding through spatial propensity scores: a study of racial disparities in diabetes. *Statistical Methods in Medical Research*, 28:734 – 748, 2019.

M. L. Davis, B. Neelon, P. J. Nietert, L. F. Burgette, K. J. Hunt, A. B. Lawson, and L. E. Egede. Propensity score matching for multilevel spatial data: accounting for geographic confounding in health disparity studies. *International Journal of Health Geographics*, 20, 2021.

P. Diggle, P. Diggle, and P. Heagerty. *Analysis of Longitudinal Data.* Oxford University Press, Oxford, 2002.

C. Donegan, Y. Chun, and A. E. Hughes. Bayesian estimation of spatial filters with Moran's eigenvectors and hierarchical shrinkage priors. Technical report, 2020.

E. Dupont, S. N. Wood, and N. Augustin. Spatial+: a novel approach to spatial confounding. *Biometrics*, 2021.

L. E. Eberly and B. P. Carlin. Identifiability and convergence issues for markov chain monte carlo fitting of spatial models. *Statistic in Medicine*, 19:2279–2294, 2000.

J. F. Elliott. The characteristic roots of certain real symmetric matrices, 1953.

A. Gardini, F. Greco, and C. Trivisano. *QF: Density, Cumulative and Quantile Functions of Quadratic Forms*. CRAN, 2022. R package version 0.0.6.

A. Gelfand, A. Schmidt, S. Banerjee, and C. Sirmans. Nonstationary multivariate process modeling through spatially varying coregionalization. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, 13(2):263–312, 2004.

A. E. Gelfand, S. Banerjee, C. Sirmans, Y. Tu, and S. Eng Ong. Multilevel modeling using spatial processes: Application to the singapore housing market. *Computational Statistics & Data Analysis*, 51(7):3567–3579, 2007.

J. E. Gentle. *Matrix Algebra: Theory, Computations, and Applications in Statistics*. Springer Texts in Statistics, 2007.

B. A. Gilbert, A. Datta, J. A. Casey, and E. L. Ogburn. Approaches to spatial confounding in geostatistics. 2021.

T. Gneiting, W. Kleiber, and M. Schlather. Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491):1167–1177, 2010.

A. B. Golgher and P. R. Voss. How to interpret the coefficients of spatial models: Spillovers, direct and indirect effects. *Spatial Demography*, 4(3):175–205, 2015.

M. Goulard and M. Voltz. Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24:269–286, 1992.

R. T. Gregory and D. Karney. A collection of matrices for testing computational algorithms. 1969.

U. Grenander and G. Szegö. *Toeplitz Forms and Their Applications*. 1958.

M. Grzebyk and H. Wackernagel. Multivariate analysis and spatial/temporal scales: Real and complex models. pages 19–33. Proceeding of the XVIIth International Biometrics Conference, 1994.

Y. Guan and M. Haran. A computationally efficient projection-based approach for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics*, 27 (4):701–714, 2018.

Y. Guan, G. L. Page, B. J. Reich, M. Ventrucci, and S. Yang. Spectral adjustment for spatial confounding. *Biometrika*, 12 2022.

P. Guttorp and T. Gneiting. Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika*, 93(4):989–995, 2006.

E. M. Hanks, E. M. Schliep, M. B. Hooten, and J. A. Hoeting. Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics*, 26(4):243–254, 2015.

H. Hartley and J. Rao. Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, 54:93–108, 1967.

D. Harville. *Matrix Algebra from a Statistician Perspective.* Springer, New York, 1997.

F. Hayashi. *Econometrics.* Princeton University Press, Princeton, N.J., 2000.

T. J. Hefley, M. B. Hooten, E. M. Hanks, R. E. Russell, and D. P. Walsh. The Bayesian Group Lasso for Confounded Spatial Data. *Journal of Agricultural, Biological and Environmental Statistics*, 22(1):42–59, 2017.

H. V. Henderson and S. R. Searle. On deriving the inverse of a sum of matrices. *Siam Review*, 23:53–60, 1981.

J. S. Hodges and B. J. Reich. Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love. *The American Statistician*, 64(4):325–334, 2010.

R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, second edition, 2013.

J. Hughes. Spatial regression and the Bayesian filter. *arXiv: Methodology*, 2017.

J. Hughes and M. Haran. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 75(1):139–159, 2013.

F. K. C. Hui and H. D. Bondell. Spatial confounding in generalized estimating equations. *The American Statistician*, 76(3):238–247, 2022.

J. Jiang and T. Nguyen. *Linear and Generalized Linear Mixed Models and Their Applications.* Springer Series in Statistics. Springer New York, NY, second edition, 2021.

J. P. Keller and A. A. Szpiro. Selecting a scale for spatial confounding adjustment. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(3):1121–1143, 2020.

K. Khan and C. Berrett. Re-thinking spatial confounding in spatial linear mixed models, 2023.

K. Khan and C. A. Calder. Restricted spatial regression methods: Implications for inference. *Journal of the American Statistical Association*, 117(537):482–494, 2022.

G. Lauricella. Sulle funzioni ipergeometriche a piu variabili. *Rendiconti del Circolo Matematico di Palermo*, 7:111–158, 1893.

D. Lee and C. Sarran. Controlling for unmeasured confounding and spatial misalignment in long-term air pollution and health studies. *Environmetrics*, 26(7):477–487, 2015.

J. LeSage. An introduction to spatial econometrics. *Revue d'économie industrielle*, 123 (3):19–44, 2008.

O. Lieberman. A laplace approximation to the moments of a ratio of quadratic forms. *Biometrika*, 81(4):681–690, 1994.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

J. K. Lindsey. *Statistical Analysis of Stochastic Processes in Time*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2004.

J. R. Magnus. The exact moments of a ratio of quadratic forms in normal variables. *Annales d'Économie et de Statistique*, (4):95–109, 1986.

I. Marques, T. Kneib, and N. Klein. Mitigating spatial confounding by explicitly correlating Gaussian random fields. *Environmetrics*, 2022.

G. Matheron. Pour une analyse krigeante des données régionalisées. Internal note N° 732, 1982.

B. Matérn. *Spatial Variation*. Springer-Verlag, second edition, 1986. Berlin.

D. Murakami and D. Griffith. Random effects specifications in eigenvector spatial filtering: a simulation study. *Journal of Geographical Systems*, 17, 2015.

W. S. Nobre, A. M. Schmidt, and J. B. M. Pereira. On the effects of spatial confounding in hierarchical models. *International Statistical Review*, 89(2):302–322, 2021.

F. W. J. Olver, , D. W. Lozier, R. F. Boisvert, and C. W. Clark. *The NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

N. Oo and W. S. Gan. On harmonic addition theorem. *International Journal of Computer and Communication Engineering*, 1(3):200–202, 2012.

M. Osama, D. Zachariah, and T. B. Schön. Inferring heterogeneous causal effects in presence of spatial confounding. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 4942–4950. PMLR, 2019.

C. J. Paciorek. The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science*, 25(1):107 – 125, 2010.

G. L. Page, Y. Liu, Z. He, and D. Sun. Estimation and prediction in the presence of spatial confounding for spatial linear models. *Scandinavian Journal of Statistics*, 44(3): 780–797, 2017.

M. Paolella. *Linear Models and Time-Series Analysis: Regression, ANOVA, ARMA and GARCH*. Wiley Series in Probability and Statistics. Wiley, 2018.

G. Papadogeorgou, C. Choirat, and C. M. Zigler. Adjusting for unmeasured spatial confounding with distance adjusted propensity score matching. *Biostatistics*, 20(2): 256–272, 2018.

J. B. M. Pereira, W. S. Nobre, I. F. L. Silva, and A. M. Schmidt. Spatial confounding in hurdle multilevel beta models: the case of the brazilian mathematical olympics for public schools. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(3):1051–1073, 2020.

M. O. Prates, R. M. Assunção, and E. C. Rodrigues. Alleviating Spatial Confounding for Areal Data Problems by Displacing the Geographical Centroids. *Bayesian Analysis*, 14 (2):623 – 647, 2019.

S. Provost and A. Mathai. *Quadratic Forms in Random Variables: Theory and Applications*. Statistics : textbooks and monographs. Marcel Dek ker, 1992.

N. H. Pusponegoro, R. N. Rachmawati, K. A. Notodiputro, and B. Sartono. Linear mixed model for analyzing longitudinal data: A simulation study of children growth differences. *Procedia Computer Science*, 116:284–291, 2017. Discovery and innovation of computer science technology in artificial intelligence era: The 2nd International Conference on Computer Science and Computational Intelligence (ICCSCI 2017).

J. Rao and I. Molina. *Small Area Models*. John Wiley & Sons, Ltd, 2015.

B. J. Reich, J. S. Hodges, and V. Zadnik. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*, 62(4):1197–1206, 2006.

B. J. Reich, S. Yang, Y. Guan, A. B. Giffin, M. J. Miller, and A. Rappold. A review of spatial causal inference methods for environmental and epidemiological applications. *International Statistical Review*, 89(3):605–634, 2021.

A. C. Rencher. *Methods of Multivariate Analysis.* John Wiley & Sons Inc, New York, second edition, 2002.

A. C. Rencher and B. Schaalje. *Linear Models in Statistics.* John Wiley & Sons Inc, 2007.

P. Robert and Y. Escoufier. A unifying tool for linear multivariate statistical methods: The rv- coefficient. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 25(3):257–265, 1976.

L. A. Roberts. On the existence of moments of ratios of quadratic forms. *Econometric Theory*, 11(4):750–774, 1995.

P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

H. Rue and L. Held. *Gaussian Markov random fields: theory and applications.* CRC Press, 2005.

T. Sawa. The exact moments of the least squares estimator for the autoregressive model. *Journal of Econometrics*, 8(2):159–172, 1978.

P. Schnell and G. Papadogeorgou. Mitigating unobserved spatial confounding when estimating the effect of supermarket access on cardiovascular disease deaths. *Annals of Applied Statistics*, 14:2069–2095, 2020.

S. Searle and A. Khuri. *Matrix Algebra Useful for Statistics.* Wiley Series in Probability and Statistics. Wiley, 2017.

A. F. Siegel and M. R. Wagner. *Practical Business Statistics.* Academic Press, eighth edition edition, 2022.

M. D. Smith. On the expectation of a ratio of quadratic forms in normal variables. *Journal of Multivariate Analysis*, 31(2):244–257, 1989.

R. Stroeker. Approximations of the eigenvalues of the covariance matrix of a first-order autoregressive process. *Journal of Econometrics*, 22(3):269–279, 1983.

H. Thaden and T. Kneib. Structural equation models for dealing with spatial confounding. *The American Statistician*, 72(3):239–252, 2018.

H. Wackernagel. *Multivariate Geostatistics: An Introduction with Applications.* Springer Berlin Heidelberg, third edition, 2003.

L. Waller and C. Gotway. *Applied Spatial Statistics for Public Health Data.* Wiley Series in Probability and Statistics. Wiley, 2004.

R. Weiber and D. Mühlhaus. *Strukturgleichungsmodellierung.* Springer Gabler Berlin, Heidelberg, second edition, 2014.

H. Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen*, 71:441–479, 1912.

J. M. Wooldridge. *Introductory econometrics: a modern approach.* Mason, OH: Thomson/South-Western, seventh edition, 2019.

M. Xiao-Li. From Unit Root to Stein's Estimator to Fisher's K Statistics: If You Have a Moment, I can tell you more. *Statistical Science*, 20(2):141–162, 2005.

J. Yang. On the extreme eigenvalues of the precision matrix of the nonstationary autoregressive process and its applications to outlier estimation of panel time series. 2021.

J. Yang, B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, and P. M. Visscher. Common SNPs explain a large proportion of the heritability for human heights. *Nature Genetics*, 42, 2010.

W. C. Yueh. Eigenvalues of several tridiagonal matrices. *Applied Mathematics E-Notes*, 5:66–74, 2005.

F. Zhang. *The Schur Complement and Its Applications (Numerical Methods and Algorithms).* first edition, 2005.

D. L. Zimmerman and J. M. V. Hoef. On deconfounding spatial confounding in linear models. *The American Statistician*, 76(2):159–167, 2021.