On the effect of confounding in linear regression models: an approach based on the theory of quadratic forms

Martina Narcisi[1,2] · **Fedele Greco**[1] · **Carlo Trivisano**[1]

Abstract

In the last two decades, significant research efforts have been dedicated to addressing the issue of spatial confounding in linear regression models. Confounding occurs when the relationship between the covariate and the response variable is influenced by an unmeasured confounder associated with both. This results in biased estimators for the regression coefficients reduced efficiency, and misleading interpretations. This article aims to understand how confounding relates to the parameters of the data generating process. The sampling properties of the regression coefficient estimator are derived as ratios of dependent quadratic forms in Gaussian random variables: this allows us to obtain exact expressions for the marginal bias and variance of the estimator, that were not obtained in previous studies. Moreover, we provide an approximate measure of the marginal bias that gives insights of the main determinants of bias. Applications in the framework of geostatistical and areal data modeling are presented. Particular attention is devoted to the difference between smoothness and variability of random vectors involved in the data generating process. Results indicate that marginal covariance between the covariate and the confounder, along with marginal variability of the covariate, play the most relevant role in determining the magnitude of confounding, as measured by the bias.

Keywords Confounding · Spatial statistics · Quadratic form · Carlson's function

✉ Martina Narcisi
martina.narcisi2@unibo.it

Fedele Greco
fedele.greco@unibo.it

Carlo Trivisano
carlo.trivisano@unibo.it

1 Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Via Belle Arti 41, 40126 Bologna, Italy

2 Department of Biological, Geological, and Environmental Sciences, University of Bologna, P.za Porta S. Donato 1, 40126 Bologna, Italy

🍃 Springer

1 Introduction

The issue of spatial confounding within linear regression models is particularly relevant in environmental applications, when the aim is to assess the effect of spatially varying environmental variables, as for example pollutants, on other environmental outcomes. In this context, we delve into the issue of confounding and its far-reaching implications, that are useful to overcome such problem in the framework of complex environmental data.

The main focus of this paper is to evaluate the impact of unobserved significant information on the estimation of regression parameters, an issue known in the literature as confounding. It arises when the relationship between the covariate and the response variable is influenced by an unmeasured confounder associated with both (Fig. 1). This can result in severely biased estimates for the regression coefficients of the measured covariates. Such deviation from the true values takes place because the posited statistical model, assuming a random effect correlated with the covariate, tries to model data from a generative mechanism characterized by confounding. It is crucial to understand the consequences of confounding when the main objective is to estimate the relationship between the response and the covariates through regression coefficients.

Confounding can occur in various statistical models and research areas, such as epidemiology, environmental sciences, public health and physics. Spatial models are often employed in these fields, leading to extensive research on spatial confounding. Previous studies have discussed the impact of spatially varying covariates and the introduction of spatial random effects on estimation (Clayton et al. 1993; Bernardinelli et al. 1995; Reich et al. 2006). However, it is now widely recognized that confounding can still exist even with suitable correlation structures for the residuals. The literature on spatial confounding has focused on assessing the strength of spatial association between the covariate, the confounder and their interaction, to understand its impact on the sampling properties of regression coefficient estimators. The parameters influencing spatial autocorrelation of the covariate and confounder are found to be of major relevance. In line with Paciorek (2010), it is commonly asserted that a confounder that is smoother than the covariates leads to lower bias and less confounding. In this paper we discuss this topic, investigating the dependence of

Fig. 1 Inter-dependencies among variables that characterize the data generating process (1)-(2)
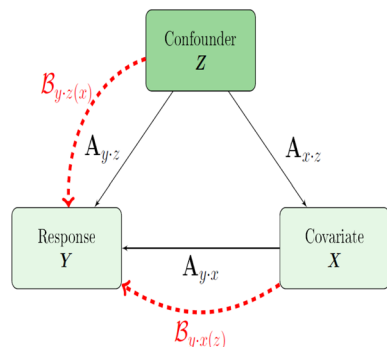
confounding on several features of the data generative mechanism. The existing literature can be divided into two strands. The first strand aims to quantify and evaluate the impact that spatial confounding has on regression coefficients (Paciorek 2010; Page et al. 2017; Nobre et al. 2021), while the second one focuses on developing methods that account for spatial confounding to obtain accurate estimates of the parameters of interest (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. 2023; Dupont et al. 2022; Yang 2021; Reich et al. 2021; Marques et al. 2022; Hui and Bondell 2022). The main objective of this article is to evaluate the sampling properties of parameter estimators in a simple setup with one covariate. A formal study of these sampling properties in the presence of confounding is conducted by defining the Data Generating Process (DGP) and a separate model for parameter estimation. 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 features of the DGP. Hypotheses 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) and more generally maximum likelihood estimators in mixed linear regression models. These estimators can be cast as the same linear function of the data with appropriate weighting matrices: this allows a unified treatment of the sampling distribution of all these estimators, which is provided in our Sect. 3. The estimator 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 (QFs) in Gaussian random variables (Provost and Mathai 1992). Following Paolella (2018), it is possible to obtain their expected value, providing an analytic expression of the marginal sampling properties of the estimator by means of the Carlson's function (Carlson 1963; Lauricella 1893). These sampling properties serve as indicators of the effect that confounding has on target parameters.

Confounding has been a focal point in numerous environmental applications where different methods have been proposed for adjusting estimates of the regression coefficients. To name a few, Dominici et al. (2004) assess the pollution-mortality relationship due to unmeasured time-varying factors, Paciorek (2010) examines the association between ambient air pollution and birthweight in eastern Massachusetts, Papadogeorgou et al. (2018) investigate the effects of power plant emission reduction technologies on ambient ozone pollution, Dupont et al. (2022) explore the temperature's impact on the crown defoliation of trees, and Marques et al. (2022) analyze the monthly precipitation variations in Germany. A shared characteristic among these methods is the reliance on unverifiable assumptions about the data generating process when information regarding confounding variables is absent. Although the issues addressed in this paper may not offer direct guidance for estimating regression coefficients in the presence of unobserved relevant information, we believe that the theoretical insights presented in the following sections, elucidating the connections between parameters of the data generating process and the

bias of estimates, can contribute to the development of novel methods. This will be object of future research.

The article is structured as follows: in Sect. 2, we introduce the data generating process and the statistical model adopted for inference. In Sect. 3, we elucidate the fundamental aspects of confounding in terms of quadratic forms, which enable us to furnish the marginal sampling properties of the estimator in a closed-form expression. Section 4 is dedicated to the introduction of measures concerning the marginal variability and smoothness of Gaussian random vectors. To show the relationship between these measures and the structure of the DGP, an application is presented in Sect. 5. The application concerns two models that are commonly applied in geostatistical analysis and areal data modeling.

2 Analytic framework

Informally, confounding occurs when the regression coefficient of a response variable on a covariate is estimated with lack of information on a third covariate, the confounder, which is associated with both. In this section we introduce the data generating process and the posited statistical model for estimation: sampling properties of regression coefficient estimators are derived, conditionally on the covariate distribution. A formal definition of confounding concludes the section.

2.1 The data generating process

To introduce the problem of confounding, a stochastic generative model, i.e. the DGP, is considered. Specification of the DGP starts by the following conditional distribution of the n-dimensional response vector Y on the covariates X and Z:

$$Y|X, Z \sim \mathcal{N}_n\big(\mathcal{B}_{y \cdot 0(xz)}\mathbf{1}_n + \mathcal{B}_{y \cdot x(z)}X + \mathcal{B}_{y \cdot z(x)}Z, \quad \Sigma_{y|x,z}\big), \tag{1}$$

where $\mathbf{1}_n$ is the n-dimensional unit vector and $\Sigma_{y|x,z}$ is the covariance matrix expressing the variability of the dependent variable Y that is not explained by the linear relationship with the regressors X and Z. Moreover, $\mathcal{B}_{y \cdot 0(xz)}$ denotes the intercept term, while $\mathcal{B}_{y \cdot x(z)}$ 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 subscripts of the partial regression coefficients aim 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 the other variable (within brackets) in the conditional mean.

This paper aims at discussing how the features of the joint distribution of X and Z affect the sampling distribution of the estimators of $\mathcal{B}_{y \cdot x(z)}$. Without loss of generality, we consider zero mean processes for both the covariate X and the confounder Z, so that $\big(X^\top, Z^\top\big)^\top \sim \mathcal{N}_{2n}\big(\mathbf{0}, \Sigma_{x,z}\big)$ where

$$\boldsymbol{\Sigma}_{x,z} = \begin{bmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}_{zx} & \boldsymbol{\Sigma}_z \end{bmatrix} = \begin{bmatrix} \sigma_x^2 \mathbf{R}_x & \sigma_{xz} \mathbf{R}_{xz} \\ \sigma_{xz} \mathbf{R}_{zx} & \sigma_z^2 \mathbf{R}_z \end{bmatrix} \qquad (2)$$

is the $2n$-dimensional joint covariance matrix of X and Z, $\boldsymbol{\Sigma}_x$ is the marginal covariance matrix of X and $\boldsymbol{\Sigma}_{xz} = \boldsymbol{\Sigma}_{zx}^\top$ is the cross-covariance matrix. Expression of each block of $\boldsymbol{\Sigma}_{x,z}$ as the product of a scalar times a structure matrix $\mathbf{R}$ is useful for the following developments. As observed by Paciorek (2010) and Page et al. (2017), the treatment of X and Z as random processes allows for the derivation of analytic results that can give insights on confounding: in this paper we extend such analytic results leveraging on the theory of quadratic forms in Gaussian variables.

To understand the consequences of lack of information concerning the unobserved variable Z, it is customary to consider the conditional distribution $Y|X$ marginalized over Z,

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

where

$$\mathbf{A}_{y \cdot x} = \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_x^{-1} = \mathcal{B}_{y \cdot x(z)} \mathbf{I}_n + \mathcal{B}_{y \cdot z(x)} \mathbf{A}_{z \cdot x} \qquad (3)$$

is the regression matrix of Y on X which depends on Z through the partial regression coefficient of Y on Z, $\mathcal{B}_{y \cdot z(x)}$, and through the regression matrix of Z on X, $\mathbf{A}_{z \cdot x}$. A well-understood case in terms of confounding is the spherical DGP, obtained by considering diagonal structure matrices $\mathbf{R} = \mathbf{I}_n$ in Equation (2), i.e.:

$$\boldsymbol{\Sigma}_{x,z} = \begin{bmatrix} \sigma_x^2 \mathbf{I}_n & \sigma_{xz} \mathbf{I}_n \\ \sigma_{xz} \mathbf{I}_n & \sigma_z^2 \mathbf{I}_n \end{bmatrix}. \qquad (4)$$

Under this DGP, the regression matrix $\mathbf{A}_{z \cdot x}$ corresponds to the scalar matrix $\mathcal{B}_{z \cdot x} \mathbf{I}_n$. This is due to the fact that cross-correlations $\mathrm{cor}(X_i, Z_j) = 0 \ \forall i \neq j$. As a consequence, the regression matrix in Equation (3) reduces to the scalar matrix:

$$\mathbf{A}_{y \cdot x} = \left(\mathcal{B}_{y \cdot x(z)} + \mathcal{B}_{y \cdot z(x)} \mathcal{B}_{z \cdot x} \right) \mathbf{I}_n.$$

The conditional distribution defined in Equation (1), combined with the joint distribution (2), delivers the following joint distribution of Y, X and Z:

$$\begin{pmatrix} Y \\ X \\ Z \end{pmatrix} \sim \mathcal{N}_{3n} \left(\begin{pmatrix} \mathcal{B}_{y \cdot 0(xz)} \mathbf{I}_n \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_y & \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yz} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{xz} \\ \boldsymbol{\Sigma}_{zy} & \boldsymbol{\Sigma}_{zx} & \boldsymbol{\Sigma}_z \end{bmatrix} \right)$$

where the joint covariance matrix $\boldsymbol{\Sigma}_{y,x,z}$ can be expressed as a function of the regression matrices $\mathbf{A}_{y \cdot z}$ and partial regression coefficients as

$$\begin{bmatrix} \boldsymbol{\Sigma}_{y|x,z} + \left[\mathcal{B}_{y \cdot x(z)} \mathbf{I}_n \ : \ \mathcal{B}_{y \cdot z(x)} \mathbf{I}_n \right] \boldsymbol{\Sigma}_{x,z} \left[\mathcal{B}_{y \cdot x(z)} \mathbf{I}_n \ : \ \mathcal{B}_{y \cdot z(x)} \mathbf{I}_n \right]^\top & \mathbf{A}_{y \cdot x} \boldsymbol{\Sigma}_x & \mathbf{A}_{y \cdot z} \boldsymbol{\Sigma}_z \\ \mathbf{A}_{x \cdot y} \boldsymbol{\Sigma}_y & \boldsymbol{\Sigma}_x & \mathbf{A}_{x \cdot z} \boldsymbol{\Sigma}_z \\ \mathbf{A}_{z \cdot y} \boldsymbol{\Sigma}_y & \mathbf{A}_{z \cdot x} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_z \end{bmatrix}.$$

The sufficient conditions that ensure its positive definiteness are $\Sigma_{x,z} > 0$ and $\Sigma_{y|x,z} > 0$. Hence, regression matrices $\mathbf{A}_{y \cdot x}$ and $\mathbf{A}_{y \cdot z}$ do not impact positive definiteness of $\Sigma_{y,x,z}$.

Developments proposed in this paper are aimed at illustrating how the degree of confounding changes with varying levels and structure of interdependence between the observed and unobserved variables.

2.2 The statistical model

After defining the DGP, this subsection introduces the posited statistical model for parameter estimation when only the realizations of Y and X are observed.

Given a phenomenon of interest, different model specifications can be proposed that reflect the researcher beliefs and assumptions. Starting from the following model:

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

one obtains the generalized least squares estimator

$$\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_0, \hat{\beta}_x\right)^\top = \mathbf{J}Y, \tag{5}$$

where $\mathbf{J} = \left(\tilde{\mathbf{X}}^\top \mathbf{S}^{-1} \tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}^\top \mathbf{S}^{-1}$ and $\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{1}_n & : & \mathbf{X} \end{bmatrix}$ is the design matrix. When $\mathbf{S} = \mathbf{I}_n$ estimator (5) corresponds to the OLS estimator. More complex models, i.e. linear mixed effect models, lead to different estimators that, conditionally on other model parameters, share the same functional form of estimator (5). We do not considered such estimators in this paper for the sake of simplicity. In line with the works of Paciorek (2010) and Page et al. (2017), we study the sampling properties of $\hat{\beta}_x$ as an estimator of $\mathcal{B}_{y \cdot x(z)}$ to investigate confounding.

2.3 Conditional sampling properties of $\hat{\beta}_x$

We start by presenting the sampling distribution of estimator (5) conditionally on X. In Sect. 3 the marginal sampling properties with respect to X will be obtained. As a first step, the following proposition introduces a well-known result adopted to study confounding in a spatial framework (see Paciorek (2010), Page et al. (2017), Nobre et al. (2021), Marques et al. (2022)).

Proposition 1 *The estimator $\hat{\boldsymbol{\beta}}$ in (5) conditional on X has the following sampling distribution*:

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

Proof See the Appendix. $\square$

With reference to the estimator $\hat{\beta}_x$, it follows from Proposition 1 that:

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

and

$$\mathbb{V}_Y\big[\hat{\beta}_x|X\big] = \mathbf{J}_{2\bullet}\boldsymbol{\Sigma}_{y|x}\mathbf{J}_{2\bullet}^\top.$$

where $\mathbf{J}_{2\bullet}$ indicates the second row of $\mathbf{J}$. We emphasize that the subscript Y aims at stressing that bias and variance are obtained by integrating over Y, conditionally on X.

The main novelties introduced in this paper are based on the expression of conditional bias and variance as ratios of quadratic forms in Gaussian random variables. This is formalized in the following proposition.

Proposition 2 *Considering the DGP in* (1)-(2)*, bias and variance of $\hat{\beta}_x$ can be expressed in terms of ratios of quadratic forms as*:

$$\text{Bias}_Y\big[\hat{\beta}_x|X\big] = \mathcal{B}_{y\cdot z(x)}\frac{X^\top\boldsymbol{\Delta}\mathbf{A}_{z\cdot x}X}{X^\top\boldsymbol{\Delta}X}, \tag{6}$$

$$\mathbb{V}_Y\big[\hat{\beta}_x|X\big] = \frac{X^\top\boldsymbol{\Delta}\boldsymbol{\Sigma}_{y|x}\boldsymbol{\Delta}X}{(X^\top\boldsymbol{\Delta}X)^2}, \tag{7}$$

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

Proof See the Appendix. $\square$

When $\mathbf{S} = \mathbf{I}_n$, $\boldsymbol{\Delta}$ reduces to the centering matrix $\mathbf{M} = \mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^\top/n$. Equation (6) highlights that when *both* $\mathcal{B}_{y\cdot z(x)} \neq 0$ and $\mathbf{A}_{x\cdot z} \neq \mathbf{0}$, the estimator $\hat{\beta}_x$ is biased. Moreover, it is worth noting that the conditional variance of $\hat{\beta}_x$ depends on the structure of all processes included in the DGP. Indeed, given that

$$\boldsymbol{\Sigma}_{y|x} = \boldsymbol{\Sigma}_{y|x,z} + \mathcal{B}_{y\cdot z(x)}^2\boldsymbol{\Sigma}_{z|x} = \boldsymbol{\Sigma}_{y|x,z} + \mathcal{B}_{y\cdot z(x)}^2\big(\boldsymbol{\Sigma}_z - \boldsymbol{A}_{z\cdot x}\boldsymbol{\Sigma}_x\boldsymbol{A}_{z\cdot x}^\top\big),$$

Equation (7) can be re-written as

$$\mathbb{V}_Y\big[\hat{\beta}_x|X\big] = \frac{X^\top \Delta \Sigma_{y|x,z} \Delta X}{(X^\top \Delta X)^2} + \mathcal{B}^2_{y \cdot z(x)} \left(\frac{X^\top \Delta \Sigma_z \Delta X}{(X^\top \Delta X)^2} - \frac{X^\top \Delta \big(A_{z \cdot x} \Sigma_x A_{z \cdot x}^\top\big) \Delta X}{(X^\top \Delta X)^2} \right).$$

The first term expresses the conditional variance when Y does not depend on Z. When Y depends on Z, i.e. $\mathcal{B}_{y \cdot z(x)} \neq 0$, the first term between brackets inflates the variance with a magnitude depending on Σ_z while the second term, which is non-null when $A_{z \cdot x} \neq 0$, deflates the conditional variance. In fact, correlation between X and Z produces a reduction of $\mathbb{V}_Y\big[\hat{\beta}_x|X\big]$, as stated in the following theorem.

Theorem 1 *The conditional variance of the estimator $\hat{\beta}_x$ assumes the maximum value when X and Z are independent*:

$$\mathbb{V}_Y\big[\hat{\beta}_x|X\big] \geq \mathbb{V}_Y\big[\hat{\beta}_x|X, A_{z \cdot x} = 0\big].$$

Proof See the Appendix. □

Summarizing, as the dependences expressed via $A_{x \cdot z}$ and $\mathcal{B}_{y \cdot z(x)}$ increase, the conditional variance decreases and the conditional bias increases. Interpretation of bias and variance of $\hat{\beta}_x$ as a ratio of quadratic forms allows for a formal treatment of the relationship between the magnitude of confounding and some features of the DGP. In particular, Sect. 3 frames the confounding problem within the context of the theory of quadratic forms, delivering analytic results on the marginal bias that, to the best of our knowledge, have been obtained only by simulation in previous papers.

2.4 Formal definition of confounding

Both Propositions 1 and 2 show that the bias of the estimator $\hat{\beta}_x$ depends on the correlation structure of the DGP through the relationship that ties the confounder to the response variable and to the observed covariate. The following definition of confounding is coherent with the one proposed by Thaden and Kneib (2018), with adapted notation.

Definition 1 Let $(Y^\top, X^\top, Z^\top)^\top$ follow the DGP of Equations (1)-(2). Then, the regression of Y on X is defined *confounded* by Z if both the following conditions are verified:

   (i) Y and Z are conditionally dependent given X ($Y \not\perp Z|X$), i.e. $\mathcal{B}_{y \cdot z(x)} \neq 0$
  (ii) X and Z are dependent ($X \not\perp Z$), i.e. $A_{z \cdot x} \neq 0$

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

Definition 2 Let $(Y^\top, X^\top, Z^\top)^\top$ follow the DGP of Equations (1)-(2). The regression of Y on X is confounded by Z if both $\Sigma_{yz} \neq 0$ and $\Sigma_{xz} \neq 0$.

In other words, confounding occurs if the unobserved variable is related with both the response and the covariate. As shown in Fig. 1, the confounder $\mathbf{Z}$ influences the response and covariate simultaneously via $\mathcal{B}_{y \cdot z(x)}$ and $\mathbf{A}_{z \cdot x}$, respectively. In the same spirit of Thaden and Kneib (2018), we highlight that the overall effect of the confounder can be expressed through the regression matrix $\mathbf{A}_{y \cdot z}$. The so-called indirect effect is linked to the regression matrix $\mathbf{A}_{x \cdot z}$, while the direct one is linked to the partial regression coefficient $\mathcal{B}_{y \cdot z(x)}$.

Confounding is not related to either the sole DGP or the posited model itself, but rather pertains to the characteristics of the DGP that the model fails to capture. The following section frames the confounding problem within the context of the theory of quadratic forms, offering some marginal results concerning bias and variance of $\hat{\beta}_x$ that are exploited to investigate the links between DGP, posited model and confounding.

3 Main features of confounding in terms of quadratic forms

Given the importance of quadratic forms in this work, a brief introduction to the subject is given below.

3.1 Quadratic forms in Gaussian random variables

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

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

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

$$Q_A(X) = \boldsymbol{v}^\top \tilde{\mathbf{A}} \boldsymbol{v} = Q_{\tilde{A}}(\boldsymbol{v}),$$

where $\tilde{\mathbf{A}} = \boldsymbol{\Sigma}_x^{1/2} \mathbf{A} \boldsymbol{\Sigma}_x^{1/2}$. Many properties of $Q_A(X)$, such as moments and distribution function, are strictly related to the eigenvalues of the matrix $\tilde{\mathbf{A}}$. We indicate them with

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

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

$$\mathbb{E}_X[Q_A(X)] = \sum_{i=1}^n \lambda(\mathbf{A}\boldsymbol{\Sigma}_x)_i = \sum_{i=1}^n \lambda(\tilde{\mathbf{A}})_i = \text{tr}(\tilde{\mathbf{A}}), \tag{8}$$

and the moment generating function is

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

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

$$R_{A,B}^{p,q}(X) = \frac{(X^\top \mathbf{A} X)^p}{(X^\top \mathbf{B} X)^q} = \frac{(\boldsymbol{\nu}^\top \tilde{\mathbf{A}} \boldsymbol{\nu})^p}{(\boldsymbol{\nu}^\top \tilde{\mathbf{B}} \boldsymbol{\nu})^q} = R_{\tilde{A},\tilde{B}}^{p,q}(\boldsymbol{\nu}) \tag{9}$$

where $p \geq 0$, $q \geq 0$ are integers 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 paper. It represents a well-known problem of numerical probability and 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, $\mathbb{E}_X\left[R_{A,B}^{p,q}(X)\right]$ exists if and only if rank$(\tilde{\mathbf{B}}) > 2q$, and it can be numerically evaluated as

$$\frac{1}{\Gamma(q)} \int_0^\infty t^{q-1} \left.\frac{\partial^p}{\partial t_1^p} \phi(t_1, t_2)\right|_{t_1=0, t_2=-t} \mathrm{d}t, \tag{10}$$

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 of $X^\top \mathbf{A} X$ and $X^\top \mathbf{B} X$, and $|\cdot|$ denotes the determinant.

Since many statistical quantities can be written as ratios of quadratic forms, the computation of their expected value has been very important for statisticians. The most popular method for its numerical evaluation is to make use of the results in Sawa (1978) and Cressie et al. (1981). This method is by far the most widespread one in the literature and Xiao-Li (2005) provides a very good review of the literature on the subject. We are concerned about obtaining computationally efficient expressions of the expectation of the ratio of dependent QFs defined in (9). Relatively straightforward expressions are available for moments of a QF in spherical normal variables. These moments appear as simple integrals which can be evaluated numerically in a straightforward manner.

3.2 Marginal sampling properties

The marginal moments of $\hat{\beta}_x$ can be retrieved using the law of iterated expectation and the law of the total variance. We find it convenient to represent the estimator marginal sampling properties in terms of a hypergeometric function: Carlson's function $R(a;b,z)$ (see Carlson (1963) for more details). Its extensive use is justified by two distinctive properties, symmetry and homogeneity. The former

means that it is invariant under permutations of the subscript $1, \ldots, m$ and the latter implies:

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

We consider the special case in which

$$\int_0^\infty t^{a-1} \prod_{i=1}^m (1 + sz_i t)^{-b_i} \, dt = B(a, a')R(a;\mathbf{b}, sz), \tag{11}$$

where $B(a, a') = \Gamma(a)\Gamma(a')/\Gamma(a + a')$ is the beta function expressed via a gamma function, $\Gamma(\cdot)$, and a' is defined by $a + a' = b = \sum_{i=1}^m b_i \in \mathbb{Q}\backslash\{0\}$.

Let us consider an n-dimensional vector λ. Assuming $sz = 2\lambda$, $b_i = \frac{1}{2}$ $\forall\, i = 1, \ldots, m$, and posing:

$$a = q, \quad a' = \frac{n}{2} + p - q, \quad m = n + 2p,$$

the right-hand side of Equation (11) can be re-written as follows:

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

where $I^{p,q}(\lambda)$ 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. This property is helpful for the operative computation. The next statement formalizes the analytical results enabling to compute the exact marginal sampling properties of estimator $\hat{\beta}_x$ with no use of simulation study.

Theorem 2 *The expected value and variance of the estimator $\hat{\beta}_x$ defined in (5) may be expressed in terms of Carlson's R functions as follows*:

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

$$\mathbb{V}_{Y,X}\left[\hat{\beta}_x\right] = \sum_{j=1}^n c_{2,jj} I_{h_j}^{1,2}(\lambda) - \text{Bias}_{Y,X}^2\left[\hat{\beta}_x\right] +$$

$$+ \mathcal{B}_{y \cdot z(x)}^2 \sum_{i=1}^n \sum_{j=1}^n (c_{1,ii}c_{1,jj} + 2c_{1,ij}^2) I_{h_{ij}}^{2,2}(\lambda),$$

where

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

The eigenvalues $(\lambda_1, \dots, \lambda_n)$, such that $\mathbf{\Lambda} = \mathrm{diag}(\lambda)$, are derived from the spectral decomposition $\mathbf{\Sigma}_x^{1/2} \mathbf{\Delta} \mathbf{\Sigma}_x^{1/2} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^\top$ and $c_{1,ij}$, $c_{2,ij}$ are the (i,j)-th entries of the matrices

$$\mathbf{C}_1 = \mathbf{P}^\top \mathbf{\Sigma}_x^{1/2} \mathbf{\Delta} \mathbf{\Sigma}_{zx} \mathbf{\Sigma}_x^{-1/2} \mathbf{P}$$

and

$$\mathbf{C}_2 = \mathbf{P}^\top \mathbf{\Sigma}_x^{1/2} \mathbf{\Delta} \mathbf{\Sigma}_{y|x} \mathbf{\Delta} \mathbf{\Sigma}_x^{1/2} \mathbf{P},$$

respectively.

Proof See the Appendix. □

The evaluation of the integrals involved in the previous theorem is carried out by computing Carlson's *R* functions: such computations exploit algorithms implemented in the R-package QF (Gardini et al. 2022).

To highlight the effect of confounding on the marginal variance, we rewrite it as:

$$\mathbb{V}_{Y,X}[\hat{\beta}_x] = \mathbb{V}_{Y,X}[\hat{\beta}_x | \mathbf{A}_{z\cdot x} = \mathbf{0}] + \mathbb{V}_{Y,X}^{cd}[\hat{\beta}_x],$$

where $\mathbb{V}_{Y,X}[\hat{\beta}_x | \mathbf{A}_{z\cdot x} = \mathbf{0}]$ denotes the marginal variance in absence of confounding. It can be shown (Narcisi 2023) that the confounding-dependent (*cd*) part of the marginal variance, $\mathbb{V}_{Y,X}^{cd}[\hat{\beta}_x]$, corresponds to

$$\mathbb{E}_X\left[\mathrm{Bias}_Y^2[\hat{\beta}_x | X]\right] - \mathrm{Bias}_{Y,X}^2[\hat{\beta}_x] - \mathcal{B}_{y\cdot z(x)}^2 \mathbb{E}_X\left[\frac{X^\top \mathbf{\Delta} \mathbf{A}_{z\cdot x} \mathbf{\Sigma}_x \mathbf{A}_{z\cdot x}^\top \mathbf{\Delta} X}{(X^\top \mathbf{\Delta} X)^2}\right].$$

Note that this component can take both negative and positive values, i.e. confounding can increase or decrease the marginal variance of the estimator with respect to the case $X \perp Z$. Following the results above, the marginal Mean Square Error (MSE) can be expressed as

$$\mathrm{MSE}_{Y,X}[\hat{\beta}_x] = \mathrm{Bias}_{Y,X}^2[\hat{\beta}_x] + \mathbb{V}_{Y,X}^{cd}[\hat{\beta}_x] + \mathbb{V}_{Y,X}[\hat{\beta}_x | \mathbf{A}_{z\cdot x} = \mathbf{0}]. \tag{12}$$

Exact computation of these quantities allows to study the impact of some relevant features of the DGP on the sampling properties of $\hat{\beta}_x$, as will be shown in Sect. 5. In the following section, we provide some approximations to exact formulae aiming to highlight the role of correlation and cross-correlation structures.

4 Links between DGP structure and confounding

We start by considering the spherical DGP (4): in this case, the conditional bias is constant with respect to X. As a consequence, marginal and conditional biases coincide:

$$\text{Bias}_{Y,X}\big[\hat{\beta}_x\big] = \text{Bias}_Y\big[\hat{\beta}_x|X\big] = \mathcal{B}_{y \cdot z(x)} \mathcal{B}_{z \cdot x} = \mathcal{B}_{y \cdot z(x)} \frac{\sigma_{zx}}{\sigma_x^2}. \tag{13}$$

This simplification with respect to the bias Formula (6) is due to the fact that the regression matrix of Z on X is scalar, i.e. $\mathbf{A}_{z \cdot x} = \mathcal{B}_{z \cdot x} \mathbf{I}_n$. This is the key point that enables the *deterministic bias* in (13) and allows a clear-cut interpretation of the effect of the DGP structure on the bias, which is an increasing function of the marginal covariance σ_{zx} and a decreasing function of the marginal variance σ_x^2 of the random vector X. Note that the marginal variance of Z, σ_z^2, does not contribute explicitly to Equation (13).

When $\mathbf{A}_{z \cdot x}$ is not a scalar matrix, it is not immediate to measure the sources of confounding. Understanding the links between the DGP structure and the bias is a more complicated task that can be addressed by leveraging on the theory of quadratic forms: in Sect. 4.1 we introduce a measure of marginal variability of a Gaussian random vector and a measure of marginal covariance between Gaussian random vectors that deliver the simple counterparts σ_x^2 and σ_{xz} as special cases when the DGP is spherical. Additionally, we introduce a measure of similarity between the components of a Gaussian random vectors, aimed at describing the smoothness of the DGP.

4.1 Variability and smoothness of Gaussian random vectors

Let $X \sim \mathcal{N}_n\big(\mathbf{0}, \boldsymbol{\Sigma}_x\big)$, with $\boldsymbol{\Sigma}_x = \sigma_x^2 \mathbf{R}_x$ where $\mathbf{R}_x$ reflects the covariance structure of the random vector X, while σ_x^2 acts as a scaler. The random variable

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

i.e. the *sampling variance* of X, is a quadratic form in Gaussian random variables whose expected value, which we dub the *expected sampling variance* is, as indicated in Equation (8)

$$EV_x = \mathbb{E}_X[V_x] = \bar{\lambda}_{M\Sigma_x} = \sigma_x^2 \bar{\lambda}_{MR_x}, \tag{14}$$

where $\bar{\lambda}_{MR_x}$ is the mean of positive eigenvalues of $\mathbf{M}\mathbf{R}_x$. When $\mathbf{R}_x = \mathbf{I}_n$, as in the case of spherical DGP, $\bar{\lambda}_{MR_x} = 1$ and $\mathbb{E}_X[V_x] = \sigma_x^2$. When $\mathbf{R}_x \neq \mathbf{I}_n$, $\mathbb{E}_X[V_x] \neq \sigma_x^2$ depends on the eigenvalues of $\mathbf{M}\mathbf{R}_x$: this must be taken into account when studying the effect of confounding under non-spherical (or structured) DGPs.

The same logic holds for the covariance between two random vectors X and Z, through the expected value of their *sampling covariance* defined as:

$$EV_{xz} = \mathbb{E}_{X,Z}\left[V_{xz}\right] = \mathbb{E}_{X,Z}\left[\frac{X^\top M Z}{n-1}\right] = \bar{\lambda}_{M\Sigma_{xz}} = \sigma_{xz}\bar{\lambda}_{MR_{xz}} \qquad (15)$$

where V_{xz} is the *sampling covariance* of X and Z. It equals σ_{xz} only when $\mathbf{R}_{xz} = \mathbf{I}_n$.

The relevance of these QFs can be emphasized by noting that, in a linear regression model, the variability of the response variable Y is decomposed as follows:

$$EV_y = \mathcal{B}^2_{y \cdot x(z)}EV_x + \mathcal{B}^2_{y \cdot z(x)}EV_z + 2\mathcal{B}_{y \cdot x(z)}\mathcal{B}_{y \cdot z(x)}EV_{zx} + \sigma^2_{y|x,z}. \qquad (16)$$

Thus, the variability of the response variable can be obtained as a function of expected sampling variances and covariances of the explanatory variables. As they are defined, such quantities show that the eigenvalues of variance and covariance matrices of the DGP are suitable quantities for quantifying the explanatory power of the covariates.

As a further tool to investigate the features of the DGP that determine the magnitude of confounding, we introduce an indicator for measuring the level of similarity between components of a random vector, which depends upon the correlation structure implied by the covariance matrix: this is strictly related to the level of smoothness of the DGP. Recalling that the correlation matrix $\mathbf{C}_x$ of a random vector X is obtained from the covariance matrix as

$$\mathbf{C}_x = \mathrm{diag}\left(\mathbf{\Sigma}_x\right)^{-\frac{1}{2}}\mathbf{\Sigma}_x\,\mathrm{diag}\left(\mathbf{\Sigma}_x\right)^{-\frac{1}{2}} = \mathrm{diag}\left(\mathbf{R}_x\right)^{-\frac{1}{2}}\mathbf{R}_x\,\mathrm{diag}\left(\mathbf{R}_x\right)^{-\frac{1}{2}},$$

we define the random vector $s = \mathrm{diag}\left(\mathbf{\Sigma}_x\right)^{-\frac{1}{2}}X$ and the random variable

$$IS_x = \frac{s^\top M s}{n-1}.$$

This stochastic variable is designated as the *sampling inverse smoothness* of X, a terminology rooted in its expected value $\mathbb{E}_X\left[IS_x\right] = \bar{\lambda}_{MC_x}$ being inversely related the smoothness of the underlying process. Note that $\mathbb{E}_X\left[IS_x\right] = 1$ if $\mathbf{R}_x = \mathbf{I}_n$.

In order to illustrate how marginal variability and inverse smoothness are related to the DGP structure, we consider two widely adopted models in spatial statistics: the Gaussian Random Field with exponential covariance function, which is a special case of the *Matérn* covariance function (Matérn 1986), largely used in geostatistical analysis and the Conditional Auto Regressive (CAR) model (Besag 1974; Rue and Held 2005), widely used for areal data modeling. Both models are parameterized by a spatial correlation parameter denoted as θ in what follows.

With regard to the first model, $\mathbf{\Sigma}_x = \sigma^2_x\mathbf{R}_x$, with the ij-th entry of $\mathbf{R}_x$ being

$$\mathbf{R}_x(i,j) = \exp\left(-\frac{d_{ij}}{\theta}\right)$$

where d_{ij} is the Euclidean distance between points i and j. The range parameter θ regulates the decay of spatial correlation as a function of the distance. The smoothness of the process, intended as the strength of correlation between nearby points in space, grows with θ.
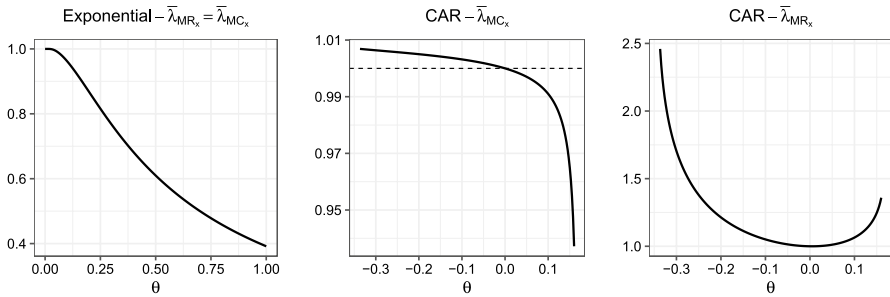
Fig. 2 Left panel: marginal inverse smoothness and variance, exponential model. Middle panel: marginal inverse smoothness, CAR model. Right panel: marginal variance, CAR model

In this case $\mathbf{R}_x$ is a correlation matrix ($\mathbf{R}_x = \mathbf{C}_x$): as a consequence, $\bar{\lambda}_{MR_x} = \bar{\lambda}_{MC_x}$ and $EV_x = \sigma_x^2 \mathbb{E}_X[IS_x]$, i.e. expected inverse smoothness and expected sampling variance are proportional, with proportionality constant σ_x^2.

Consider now the CAR model, where $\boldsymbol{\Sigma}_x = \sigma_x^2 \mathbf{R}_x$, with $\mathbf{R}_x = (\mathbf{I}_n - \theta W)^{-1}$. Here, a spatial lattice structure is described by the $n \times n$ neighborhood matrix W specified as $w_{ii} = 0$, $w_{ij} = 1$ if area i is neighbor to area j and $w_{ij} = 0$ otherwise. The sufficient condition ensuring positive definiteness of the covariance matrix is $\theta \in (\theta_{\min}, \theta_{\max})$, where $\theta_{\min}$ and $\theta_{\max}$ are the inverse of the smallest and largest eigenvalues of W respectively (Cressie 1993). Moreover, θ is a measure of (conditional) spatial autocorrelation. In what follows we adopt the lattice of the 115 Missouri's counties to build W.

In the CAR case, $\mathbf{R}_x \neq \mathbf{C}_x$, hence expected sampling variance and expected inverse smoothness show different behaviors with respect to θ.

Figure 2, left panel, shows the behavior of $\mathbb{E}_X[IS_x]$ as a function of θ in the Exponential model: coherently with the well-known theory concerning the exponential correlation function, smoothness is an increasing function ($\mathbb{E}_X[IS_x]$ is a decreasing function) of both θ. On the other hand, the expected sampling variance EV_x is a decreasing function of θ. Figure 2, middle panel, displays the relationship between θ and $\mathbb{E}_X[IS_x]$ when considering the CAR model: it can be noticed that smoothness increases with θ (inverse smoothness decreases), confirming the interpretation of θ as a spatial correlation parameter. Moreover, when $\theta < 0$, implying negative spatial correlation, $\mathbb{E}_X[IS_x] > 1$, meaning that the DGP is less smooth than the spherical DGP, while positive values of θ deliver a smoother process with respect to the spherical DGP. In the right panel of Fig. 2 it can be noticed that the expected sampling variance EV_x decreases when $\theta < 0$ and increases when $\theta > 0$. Hence, in the case of the CAR model, expected smoothness and expected sampling variance show different behaviors as functions of the smoothness parameter θ.

To summarize, a sample from the exponential model is expected to be smoother and less variable as θ increases, on the other hand a sample from the CAR model is expected to be smoother as θ increases, while variability increases as θ approaches the boundaries $(\theta_{\min}, \theta_{\max})$.

A thorough discussion of the relationships between smoothness and variability is out of the scope of this paper: here, we are interested in showing how these features of the DGP relate to confounding. In particular, in the following subsection, we provide an approximation to the marginal bias that is strictly related to the quantities introduced in this section and that will be useful to discuss the relevance of variability and smoothness as determinants of bias.

4.2 Approximation of estimator bias

Equations (14)-(15) can be exploited for obtaining the first-order Taylor series approximation of the marginal bias, i.e. the expected value of a ratio is approximated by the ratio of expected values:

$$\text{Bias}_{Y,X}\left[\hat{\beta}_x\right] \approx \mathcal{B}_{y \cdot z(x)} \frac{\mathbb{E}_X\left[X^\top \Delta A_{z \cdot x} X\right]}{\mathbb{E}_X\left[X^\top \Delta X\right]} = \mathcal{B}_{y \cdot z(x)} \frac{\bar{\lambda}_{\Delta \Sigma_{zx}}}{\bar{\lambda}_{\Delta \Sigma_x}} = E_T. \tag{17}$$

In the case of spherical DGP, E_T is the exact value for the bias.

When $\Delta = M$, i.e. when the OLS estimator is considered, the approximation coincides with the ratio between EV_{zx} and EV_x: approximate bias is an increasing function of the expected sampling covariance and a decreasing function of the expected sampling variance. The marginal variability of the confounder, EV_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. Note that the smoothness of the DGP does not enter explicitly in Equation (17), suggesting that such feature of the DGP is not a relevant determinant of bias, as will be discussed in Sect. 5.

Table 1 Joint confounder-covariate covariance matrices and regression matrices of Z on X specified for DGPs A-C

DGP	$\Sigma_{x,z}$	$A_{z \cdot x}$
A	$\begin{bmatrix} R(\theta_x) & \rho R(\theta_z) \\ \rho R(\theta_z)^\top & R(\theta_z) \end{bmatrix}$	$\rho R(\theta_z) R(\theta_x)^{-1}$
B	$\begin{bmatrix} \dfrac{R(\theta_z) + R(\theta_x)}{2} & \dfrac{\rho}{\sqrt{2}} R(\theta_z) \\ \dfrac{\rho}{\sqrt{2}} R(\theta_z) & R(\theta_z) \end{bmatrix}$	$\rho \sqrt{2} \left[R(\theta_x) R(\theta_z)^{-1} + I_n \right]^{-1}$
C	$\begin{bmatrix} R(\theta_x) & \rho R(\theta_x)^{\frac{1}{2}} \left(R(\theta_z)^{\frac{1}{2}} \right)^\top \\ \rho R(\theta_z)^{\frac{1}{2}} \left(R(\theta_x)^{\frac{1}{2}} \right)^\top & R(\theta_z) \end{bmatrix}$	$\rho R(\theta_z)^{\frac{1}{2}} \left(R(\theta_x)^{-\frac{1}{2}} \right)^\top$

5 Applications

This section provides a study of the marginal sampling properties of $\hat{\beta}_x$ under several DGPs: concerning the conditional distribution $Y|X, Z$ in Equation (1), we set $\mathcal{B}_{y \cdot x(z)} = \mathcal{B}_{y \cdot z(x)} = 1$ and $\boldsymbol{\Sigma}_{y|x,z} = \mathbf{I}_n$. As for the specification of the joint covariance matrix $\boldsymbol{\Sigma}_{x,z}$ in Equation (2), we consider the three DGPs summarised in Table 1, where regression matrices $\mathbf{A}_{z \cdot x}$ are reported as well. Moreover, we fix $\sigma_x^2 = \sigma_z^2 = \sigma_{zx} = 1$.

Parameter ρ governs the strength of cross-correlation between X and Z and determines positive definiteness of $\boldsymbol{\Sigma}_{x,z}$. A sufficient condition to ensure the positive definiteness in DGP-A is $\lambda_{\min}\left(\mathbf{R}(\theta_z)^{-1}\mathbf{R}(\theta_x)\mathbf{R}(\theta_z)^{-1}\mathbf{R}(\theta_{\bar{z}})\right) \geq \rho^2$, while the condition $|\rho| < 1$ is sufficient for positive definiteness in DGPs B and C.

In DGP-A, the cross-covariance matrix is parameterized independently on the parameters of the marginal covariance matrices of X and Z. This construction is intended to show that, as expected from Theorem 2, the bias relies on the structure of Z only when the cross-covariance matrix is obtained as a function of $\boldsymbol{\Sigma}_z$ by construction. This happens in both DGP-B and DGP-C.

In DGP-B, which is a modification of the DGP discussed in Paciorek (2010), parameter θ_z governs both the covariance matrix of Z and the cross-covariance matrix, while the covariance matrix of X is a function of both θ_z and θ_x.

In DGP-C, which corresponds to the specification adopted in Page et al. (2017), the cross-covariance matrix is obtained as the product of the lower Cholesky factors of the covariance matrices of X and Z, indexed by spatial correlation parameters θ_x and θ_z respectively. Note that this is the only DGP that can cover the case of asymmetric cross-covariances.

In what follows, we study the marginal sampling properties of $\hat{\beta}_x$ in the cases where the structure matrices $\mathbf{R}$ are specified via the exponential and the CAR model. Results provided in Theorem 2 allow to compute exact quantities with no need for simulation. We focus on the study of the sampling properties of the OLS estimator: results obtained using the GLS estimator (not shown) are very similar.

5.1 Results

With regard to the exponential model, a regular grid of $n = 64$ spatial units located on the unit square is considered and the sampling properties of $\hat{\beta}_x$ are studied by letting all the spatial correlation parameters vary in the interval $[0, 1]$. Figure 3 reports, for each considered DGP, the expected sampling covariance (EV_{zx}, first line), the expected sampling variance EV_x (second line) and the *exact* marginal bias $\text{Bias}_{Y,X}\left[\hat{\beta}_x\right]$ (third line). The relative error of approximation (17) is reported in the last line. It is worth noting that, with reference to bias, parameter ρ constitutes a mere scaling constant: Fig. 3 displays the behavior of the marginal bias as a function

of the spatial correlation parameters, rather than its magnitude; to this aim, ρ values are irrelevant.

It can be noticed that, with reference to DGPs A and B, EV_{zx} does not depend on θ_x and is a decreasing function of θ_z, while, with reference to DGPs A and C, EV_x does not depend on θ_z and is a decreasing function of θ_x. Moreover, in DGP-B, EV_x is a decreasing function of both spatial correlation parameters, while, in DGP-C, EV_{zx} is a non-monotonic function of both θ_x and θ_z. These behaviors are reflected by the marginal bias toward the ratio of expected sampling covariance and expected sampling variance of the observed covariate X, which delivers the approximation provided in Equation (17). Last line of Fig. 3 shows a satisfactory accuracy of the approximation for all considered DGPs: the relative error is mostly less than $\pm 5\%$, higher errors are observed for low values of θ_z and high θ_x with reference to DGPs A and C.

Regarding DGPs A and B, it turns out that the marginal bias is an increasing function of θ_x (maintaining θ_z constant) and a decreasing function of θ_z (maintaining θ_x constant). Note that the spatial correlation parameter indexing the covariance matrix of Z in DGP-A, ($\theta_{\tilde{z}}$ in Table 1), has no impact on bias: dependency of bias on the spatial correlation parameter of the confounder is actually an artifact generated
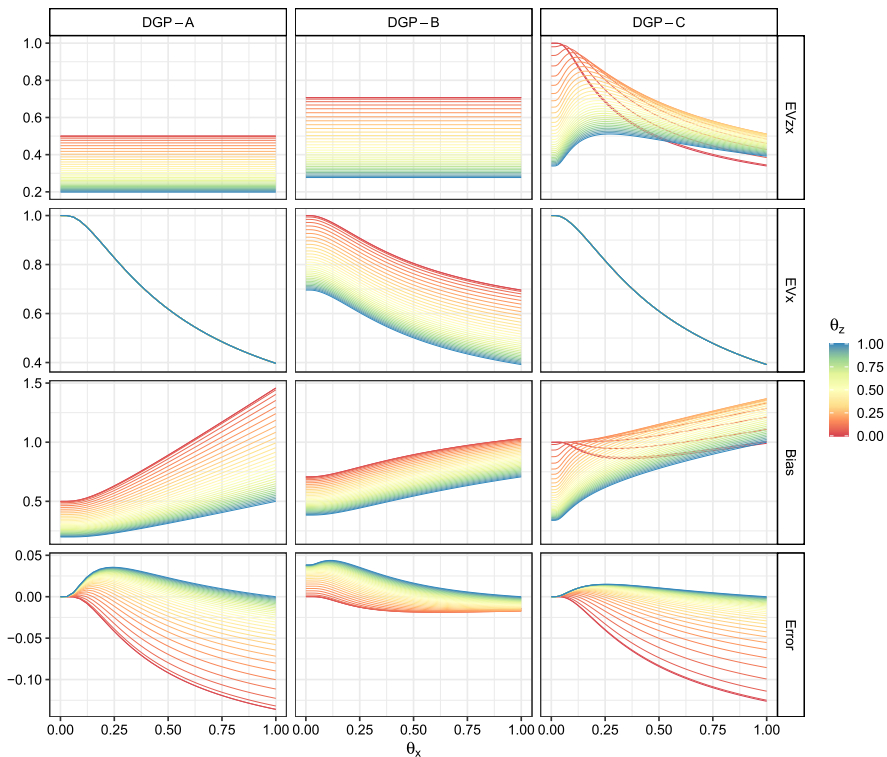
Fig. 3 Expected sampling covariance, expected sampling variance, marginal bias and relative error of the Taylor approximation with respect to θ_x and θ_z for DGPs A-C. Exponential model

when the same parameter appears both in the cross-covariance matrix and in the marginal covariance matrix of Z, as observed in DGPs B and C. Interpretation of the bias trend in Case C is less immediate because of the product involved in the construction of the cross-covariance matrix. This gives rise to the non-monotonic behavior of the marginal bias with respect to θ_x and θ_z, such behavior is inherited from EV_{zx}.

In summary, the findings presented in Fig. 3 indicate that the primary contributors to bias are the expected sampling variance and covariance. Notably, the similarity in behavior between smoothness and expected sampling variance when employing an exponential correlation function (as depicted in the left panel of Fig. 2) complicates our ability to discern whether the key drivers of bias are predominantly associated with the strength of spatial autocorrelation (smoothness) or the marginal variability. A more in-depth analysis of the case involving CAR model provides enlightening insights into this matter.

With regard to the CAR model, the lattice of the 115 Missouri's counties is considered. The spatial correlation parameters vary in the interval $[\theta_{\min} + 0.01, \theta_{\max} - 0.01]$ to ensure positive definiteness of variance and covariance matrices. Figure 4 is the counterpart of Fig. 3 in the case of CAR model.

It can be noticed that, with reference to DGPs A and B, EV_{zx} does not depend on θ_x and is a convex function of θ_z, while, with reference to DGPs A and C, EV_x does not depend on θ_z and is a convex function of θ_x, coherently with what was observed in Fig. 2, right panel. Moreover, EV_x in DGP-B and EV_{zx} in DGP-C are convex functions of both spatial correlation parameters. In this instance, the Taylor approximation given by Equation (17) exhibits a high degree of accuracy, as evidenced by the relative error falling within the range of 1% to -4% across all scenarios.

The key insight gleaned from Fig. 4 is the lack of significance of DGP smoothness in influencing bias. Despite the fact that smoothness rises as a function of spatial autocorrelation parameters (as shown in Fig. 2, middle panel), the bias trends exhibit a convex pattern, primarily shaped by the characteristics of EV_x and EV_{zx}. In conclusion, marginal variance and covariance are the main determinants of bias. This is strictly tied with the decomposition of the response marginal variance reported in Equation (16): when EV_x increases, keeping other factors fixed, the explanatory power of X increases and bias decreases.

Figure 5 shows, in the case of the exponential model, the contribution of each component constituting the estimator mean square error in (12) to the total as a function of the parameter θ_x and θ_z fixing $\rho = 0.5$.

It is evident how the mean squared error is primarily determined by bias. It can be seen that the bias share is a decreasing function of θ_z, while the variance component not dependent on confounding shows an opposite behavior which is related to the role of θ_z in the distribution of the confounder. Actually, the most relevant role of smoothness and marginal variability of the confounder is related to the marginal variance, rather than to the bias of the estimator. The confounding-dependent share of the variance predominantly assumes negative values, showing that confounding tends to generate a reduction in variance. Similar results (not shown) are obtained in the case of the CAR model.
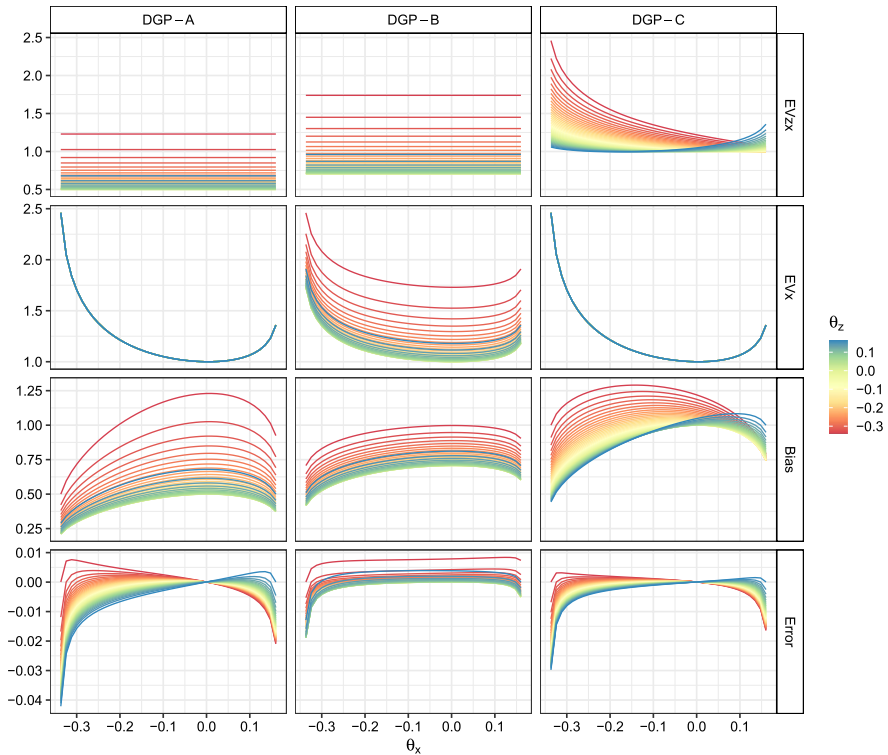
Fig. 4 Expected sampling covariance, expected sampling variance, marginal bias and relative error of the Taylor approximation with respect to θ_x and θ_z for DGPs A-C. CAR model

6 Conclusions

In this work the problem of confounding in linear regression models is addressed. In particular, we study, through the evaluation of the estimator sampling properties, how confounding affects the estimate of the inferential target.

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. We provide more awareness regarding the effect of confounding on coefficient estimates by generalizing the theory discussed by Paciorek (2010) and Page et al. (2017), who introduced the widely accepted idea that the smoothness of covariate and confounder processes is an important factor that impacts the estimator sampling properties. In particular,
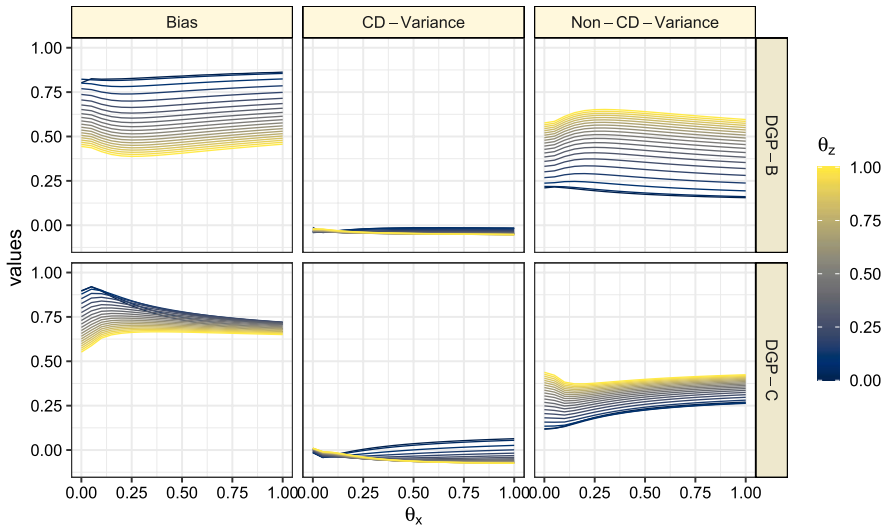
Fig. 5 The influence of the three components of the estimator mean square error on it as function of the range parameters θ_x, θ_z with $\rho = 0.5$ (exponential function)

Paciorek (2010) affirms that a confounder smoother than the covariate leads to a lower bias, and subsequently to less confounding. Actually, one may agree with this belief only in specific situations, such as when the parameters governing the confounder covariance matrix contribute to the cross-covariance matrix, e.g. assuming an exponential correlation function and DGPs B-C. In other cases, as demonstrated by the application involving the CAR model, 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, respectively. When 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, as the bias mainly hinges on the covariate variability and on the expected sampling covariance between covariate and confounder. Moreover, we note that the confounder structure does not affect the estimator bias; rather, it influences the variance of the estimator.

While the primary focus of the paper is theoretical and doesn't offer immediate implications for practical applications, certain aspects of the presented theory could potentially guide advancements in environmental contexts. For instance, one could attempt to estimate bias resulting from confounding by devising an estimator for the approximate bias (17). Given the unobservable nature of the confounder, such estimations would necessitate relying on assumptions about the DGP, which may be challenging to verify.

Appendix

Proof of Proposition 1 Using the main properties of the normal distribution, the conditional sampling properties of the estimator $\hat{\beta}_x$ defined in (5) are obtained as follows

$$
\begin{aligned}
\mathbb{E}_Y\big[\hat{\beta}|X\big] &= \mathbb{E}_Y\big[\mathbf{J}Y|X\big] \\
&= \mathbf{J}\,\mathbb{E}_{Z,\varepsilon}\big[\mathcal{B}_{y\cdot 0(xz)}\mathbf{1}_n + \mathcal{B}_{y\cdot x(z)}X + \mathcal{B}_{y\cdot z(x)}Z + \varepsilon_{y|x,z}\big|X\big] \\
&= \mathbf{J}\big[\mathcal{B}_{y\cdot 0(xz)}\mathbf{1}_n + \mathcal{B}_{y\cdot x(z)}X + \mathcal{B}_{y\cdot z(x)}\mathbb{E}_Z\big[Z|X\big]\big] \\
&= \mathbf{J}\big[\mathcal{B}_{y\cdot 0(xz)}\mathbf{1}_n + \mathcal{B}_{y\cdot x(z)}X + \mathcal{B}_{y\cdot z(x)}\mathbf{A}_{z\cdot x}X\big] \\
&= \mathbf{J}\big[\mathcal{B}_{y\cdot 0(xz)}\mathbf{1}_n + (\mathcal{B}_{y\cdot x(z)} + \mathcal{B}_{y\cdot z(x)}\mathbf{A}_{z\cdot x})X\big] \\
&= \mathbf{J}[\mathcal{B}_{y\cdot 0(xz)}\mathbf{1}_n + \mathbf{A}_{y\cdot x}X] \\
\mathbb{V}_Y\big[\hat{\beta}|X\big] &= \mathbb{V}_Y[\mathbf{J}Y|X] \\
&= \mathbf{J}\Sigma_{y|x}\mathbf{J}^\top.
\end{aligned}
$$

$\square$

Proof of Proposition 2 We can rewrite the matrix $\mathbf{J}$ as follows:

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

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

$$\mathbf{J} = \frac{1}{\tilde{\delta}} \left(\frac{X^\top \mathbf{S}^{-1} X}{\tilde{\delta}} - \frac{X^\top \mathbf{S}^{-1} \mathbf{1}_n \mathbf{1}_n^\top \mathbf{S}^{-1} X}{\tilde{\delta}^2} \right)^{-1} \left[\begin{array}{cc} \dfrac{X^\top \mathbf{S}^{-1} X}{\mathbf{1}_n^\top \mathbf{S}^{-1} \mathbf{1}_n} & -\dfrac{\mathbf{1}_n^\top \mathbf{S}^{-1} X}{\mathbf{1}_n^\top \mathbf{S}^{-1} \mathbf{1}_n} \\ -\dfrac{X^\top \mathbf{S}^{-1} \mathbf{1}_n}{\mathbf{1}_n^\top \mathbf{S}^{-1} \mathbf{1}_n} & 1 \end{array} \right] \left[\begin{array}{c} \mathbf{1}_n^\top \mathbf{S}^{-1} \\ X^\top \mathbf{S}^{-1} \end{array} \right]$$

$$= \frac{1}{X^\top \mathbf{S}^{-1} X - \dfrac{X^\top \mathbf{S}^{-1} \mathbf{1}_n \mathbf{1}_n^\top \mathbf{S}^{-1} X}{\tilde{\delta}}} \left[\begin{array}{c} \dfrac{X^\top \mathbf{S}^{-1} X \mathbf{1}_n^\top \mathbf{S}^{-1} - \mathbf{1}_n^\top \mathbf{S}^{-1} X X^\top \mathbf{S}^{-1}}{\tilde{\delta}} \\ X^\top \mathbf{S}^{-1} - \dfrac{X^\top \mathbf{S}^{-1} \mathbf{1}_n \mathbf{1}_n^\top \mathbf{S}^{-1}}{\tilde{\delta}} \end{array} \right]$$

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

Hence,

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

$$= \frac{X^\top \boldsymbol{\Delta}}{X^\top \boldsymbol{\Delta} X}.$$

Consequently, we get

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

and we come by to the result in (6). In addition, from (A.1) and since $\boldsymbol{\Delta} = \boldsymbol{\Delta}^\top$, the proof of (7) follows accordingly.

Proof of Theorem 1 From

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

we can observe that 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. Hence, $\mathbb{V}_Y \left[\hat{\beta}_x | X \right] - \mathbb{V}_Y \left[\hat{\beta}_x | X, A_{z \cdot x} = \mathbf{0} \right] \geq 0$. □

Proof of Theorem 2 Concerning the expected bias of $\hat{\beta}_x$, note that Equation (6) can be cast in the form of (9) by posing $\mathbf{A} = \boldsymbol{\Delta} A_{z \cdot x}$ and $\mathbf{B} = \boldsymbol{\Delta}$:

$$R_{A,B}^{1,1}(X) = \frac{X^\top \Delta A_{z\cdot x} X}{X^\top \Delta X}$$

$$= \frac{(\Sigma_x^{-1/2}X)^\top \Sigma_x^{1/2} \Delta \Sigma_{zx} \Sigma_x^{-1} \Sigma_x^{1/2}(\Sigma_x^{-1/2}X)}{(\Sigma_x^{-1/2}X)^\top \Sigma_x^{1/2} \Delta \Sigma_x^{1/2}(\Sigma_x^{-1/2}X)} \qquad (A.2)$$

$$= \frac{v^\top \Sigma_x^{1/2} \Delta \Sigma_{zx} \Sigma_x^{-1/2} v}{v^\top \Sigma_x^{1/2} \Delta \Sigma_x^{1/2} v} = R_{\tilde{A},\tilde{B}}^{1,1}(v),$$

where $\tilde{A} = \Sigma_x^{1/2} \Delta \Sigma_{zx} \Sigma_x^{-1/2}$ and $\tilde{B} = \Sigma_x^{1/2} \Delta \Sigma_x^{1/2}$. The joint moment generating function of the numerator $v^\top \tilde{A} v$ and the denominator $v^\top \tilde{B} v$ in (A.2) is given by

$$\phi(t_1, t_2) = |I_n - 2t_1 \Sigma_x^{1/2} \Delta \Sigma_{zx} \Sigma_x^{-1/2} - 2t_2 \Sigma_x^{1/2} \Delta \Sigma_x^{1/2}|^{-1/2}.$$

Let consider the following spectral decomposition

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

where $\Lambda = \text{diag}\lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ contains the eigenvalues of $\tilde{B}$, and P is the orthogonal eigenvectors matrix such that $PP^\top = P^\top P = I_n$. As a result, the joint moment generating function takes the following form:

$$\phi(t_1, t_2) = |P^\top P|^{-1/2} |I_n - 2t_1 \Sigma_x^{1/2} \Delta \Sigma_{zx} \Sigma_x^{-1/2} - 2t_2 \Sigma_x^{1/2} \Delta \Sigma_x^{1/2}|^{-1/2}$$

$$= |I_n - 2t_1 P^\top \Sigma_x^{1/2} \Delta \Sigma_{zx} \Sigma_x^{-1/2} P - 2t_2 P^\top \Sigma_x^{1/2} \Delta \Sigma_x^{1/2} P|^{-1/2}$$

$$= |I_n - 2t_1 C_1 - 2t_2 \Lambda|^{-1/2}$$

$$= |F|^{-1/2}$$

in which

$$C_1 = P^\top \tilde{A} P$$

$$= P^\top \Sigma_x^{1/2} \Delta \Sigma_{zx} \Sigma_x^{-1/2} P,$$

with (i, j)-th element $c_{1,ij}$. The expected value of (A.2) is now obtainable using Equation (10) with $p = q = 1$:

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

Since, $\frac{\partial}{\partial t_1} F = -2C_1$, for $p = 1$

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

$$= |F|^{-1/2} \text{tr}\left(F^{-1} C_1\right),$$

and

$$\frac{\partial}{\partial t_1} \phi(t_1, t_2)\Big|_{t_1=0, t_2=-t} = |\mathbf{I}_n + 2t\mathbf{\Lambda}|^{-1/2} \text{tr}\big[(\mathbf{I}_n + 2t\mathbf{\Lambda})^{-1}\mathbf{C}_1\big]$$

$$= \prod_{i=1}^{n}(1 + 2\lambda_i t)^{-1/2} \sum_{j=1}^{n} \frac{c_{1,jj}}{1 + 2\lambda_j t}.$$

Using the law of total expectation, the marginal expectation of the estimator $\hat{\beta}_x$ is

$$
\begin{aligned}
\mathbb{E}_{Y,X}\big[\hat{\beta}_x\big] &= \mathbb{E}_X\big[\mathbb{E}_Y\big[\hat{\beta}_x|X\big]\big] \\
&= \mathcal{B}_{y \cdot x(z)} + \mathcal{B}_{y \cdot z(x)} \int_0^\infty |\mathbf{I}_n + 2t\mathbf{\Lambda}|^{-1/2} \text{tr}\big[(\mathbf{I}_n + 2t\mathbf{\Lambda})^{-1}\mathbf{C}_1\big]\, dt \\
&= \mathcal{B}_{y \cdot x(z)} + \mathcal{B}_{y \cdot z(x)} \sum_{j=1}^{n} c_{1,jj} \int_0^\infty \prod_{i=1}^{n}(1 + 2\lambda_i t)^{-1/2} \frac{1}{1 + 2\lambda_j t}\, dt \\
&= \mathcal{B}_{y \cdot x(z)} + \text{Bias}_{Y,X}\big[\hat{\beta}_x\big],
\end{aligned}
\tag{A.3}
$$

in which the estimator for marginal bias is highlighted.

Concerning the sampling variance, the law of total variance allows to get:

$$
\begin{aligned}
\mathbb{V}_{Y,X}\big[\hat{\beta}_x\big] &= \mathbb{E}_X\big[\mathbb{V}_Y\big[\hat{\beta}_x|X\big]\big] + \mathbb{V}_X\big[\text{Bias}_Y\big[\hat{\beta}_x|X\big]\big] \\
&= \mathbb{E}_X\big[\mathbb{V}_Y\big[\hat{\beta}_x|X\big]\big] + \mathbb{E}_X\big[\text{Bias}_Y^2\big[\hat{\beta}_x|X\big]\big] - \text{Bias}_{Y,X}^2\big[\hat{\beta}_x\big] \\
&= \mathbb{E}_X\left[\frac{X^\top \Delta \Sigma_{y|x} \Delta X}{(X^\top \Delta X)^2}\right] + \\
&\quad + \mathcal{B}_{y \cdot z(x)}^2 \mathbb{E}_X\left[\left(\frac{X^\top \Delta \mathbf{A}_{z \cdot x} X}{X^\top \Delta X}\right)^2\right] - \text{Bias}_{Y,X}^2\big[\hat{\beta}_x\big]
\end{aligned}
\tag{A.4}
$$

in which, using (10) with $p = 1$ and $q = 2$, it is possible to determine

$$
\begin{aligned}
\mathbb{E}_X\big[\mathbb{V}_Y\big[\hat{\beta}_x|X\big]\big] &= \mathbb{E}_X\left[\frac{X^\top \Delta \Sigma_{y|x} \Delta X}{(X^\top \Delta X)^2}\right] \\
&= \sum_{j=1}^{n} c_{2,jj} \int_0^\infty t \prod_{i=1}^{n}(1 + 2\lambda_i t)^{-1/2} \frac{1}{1 + 2\lambda_j t}\, dt,
\end{aligned}
\tag{A.5}
$$

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

The computation of the quantity $\mathbb{E}_X\big[\text{Bias}_Y^2\big[\hat{\beta}_x|X\big]\big]$ requires $p = 2$ in (10). Following Paolella (2018), we first observe that

$$\frac{\partial^2 |\mathbf{F}|}{\partial t_1^2} = \frac{\partial}{\partial t_1}\left(-2|\mathbf{F}|\operatorname{tr}(\mathbf{F}^{-1}\mathbf{C}_1)\right)$$

$$= \frac{\partial(-2|\mathbf{F}|)}{\partial t_1}\operatorname{tr}(\mathbf{F}^{-1}\mathbf{C}_1) - 2|\mathbf{F}|\frac{\partial}{\partial t_1}\left(\operatorname{tr}(\mathbf{F}^{-1}\mathbf{C}_1)\right)$$

$$= 4|\mathbf{F}|\operatorname{tr}^2(\mathbf{F}^{-1}\mathbf{C}_1) - 2|\mathbf{F}|\operatorname{tr}\left(\mathbf{F}^{-1}\frac{\partial \mathbf{C}_1}{\partial t_1} + \frac{\partial \mathbf{F}^{-1}}{\partial t_1}\mathbf{C}_1\right)$$

$$= 4|\mathbf{F}|\left(\operatorname{tr}^2(\mathbf{F}^{-1}\mathbf{C}_1) - \operatorname{tr}(\mathbf{F}^{-1}\mathbf{C}_1)^2\right)$$

and

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

$$= \frac{3}{4}|\mathbf{F}|^{-5/2}\left(\frac{\partial |\mathbf{F}|}{\partial t_1}\right)^2 - \frac{1}{2}|\mathbf{F}|^{-3/2}\frac{\partial^2 |\mathbf{F}|}{\partial t_1^2}$$

$$= \frac{3}{4}|\mathbf{F}|^{-5/2}\left(4|\mathbf{F}|^2\operatorname{tr}^2(\mathbf{F}^{-1}\mathbf{C}_1)\right) - \frac{1}{2}|\mathbf{F}|^{-3/2}4|\mathbf{F}|\left[\operatorname{tr}^2(\mathbf{F}^{-1}\mathbf{C}_1) - \operatorname{tr}(\mathbf{F}^{-1}\mathbf{C}_1)^2\right]$$

$$= |\mathbf{F}|^{-1/2}(\operatorname{tr}^2(\mathbf{F}^{-1}\mathbf{C}_1) + 2\operatorname{tr}(\mathbf{F}^{-1}\mathbf{C}_1)^2).$$

Thus,

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

Next, using (10) with $p = q = 2$, we obtain

$$\mathbb{E}_X\left[\operatorname{Bias}_Y^2\left[\hat{\beta}_x|X\right]\right] = \mathcal{B}_{y\cdot z(x)}^2\sum_{i=1}^{n}\sum_{j=1}^{n}(c_{1,ii}c_{1,jj} + 2c_{1,ij}^2)\cdot$$
$$\cdot\int_0^{\infty}t\prod_{i=1}^{n}(1+2\lambda_i t)^{-1/2}\frac{1}{(1+2\lambda_i t)(1+2\lambda_j t)}\,dt. \tag{A.6}$$

Finally, from Equation (A.4) we get the expression for the marginal variance of the estimator $\hat{\beta}_x$ as:

$$\mathbb{V}_{Y,X}\big[\hat{\beta}_x\big] = \sum_{j=1}^{n} c_{2,jj} \int_0^{\infty} t \prod_{i=1}^{n} (1+2\lambda_i t)^{-1/2} \frac{1}{1+2\lambda_j t}\, dt +$$

$$+ \mathcal{B}^2_{y \cdot z(x)} \sum_{i=1}^{n} \sum_{j=1}^{n} (c_{1,ii} c_{1,jj} + 2c^2_{1,ij}) \cdot$$

$$\cdot \int_0^{\infty} t \prod_{i=1}^{n} (1+2\lambda_i t)^{-1/2} \frac{1}{(1+2\lambda_i t)(1+2\lambda_j t)}\, dt +$$

$$- \left(\mathcal{B}_{y \cdot z(x)} \sum_{j=1}^{n} c_{1,jj} \int_0^{\infty} \prod_{i=1}^{n} (1+2\lambda_i t)^{-1/2} \frac{1}{1+2\lambda_j t}\, dt \right)^2.$$

Based on the Carlson's properties introduced in Sect. 3.2, we can use the Carlson's R function for the evaluation of integrals in (A.3), (A.5), (A.6) in order to use efficient available algorithms to compute the marginal sampling properties of estimator $\hat{\beta}_x$.

Considering Equation (A.3) with $p = q = 1$ and exploiting the symmetric property of R, for all $j = 1, \ldots, n$ we define

$$\Lambda'_j = \mathrm{diag}(\lambda_1, \ldots, \lambda_n, \lambda_j, \lambda_j) = \mathrm{diag}\lambda'_j$$

and observe that

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

$$= I^{1,1}_{h_j}(\lambda),$$

where $\lambda'_{j,+}$ is the j-th vector of positive eigen values in Λ' and $h_j = 1$ for $j = 1, \ldots, n-1$ and $h_n = 3$. Similarly, referring to Equation (A.5) in which $p = 1$, $q = 2$, and for all $j = 1, \ldots, n$ we get:

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

$$= I^{1,2}_{h_j}(\lambda),$$

Finally, taking Equation (A.6) in which $p = q = 2$ and, for all $i, j = 1, \ldots, n$ we define the following diagonal matrix

$$\Lambda''_{ij} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n, \lambda_j, \lambda_j, \lambda_i, \lambda_i) = \mathrm{diag}\lambda''_{ij}$$

and observe that

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

$$= I^{2,2}_{h_{ij}}(\lambda)$$

where

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

$\square$

References

Bao Y, Kan R (2013) On the moments of ratios of quadratic forms in normal random variables. J Multivar Anal 117:229–245

Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M (1995) Bayesian analysis of space-time variation in disease risk. Stat Med 14(21–22):2433–2443

Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. J Roy Stat Soc 36(2):192–236

Carlson B (1963) Lauricella's hypergeometric function fd. J Math Anal Appl 7(3):452–470

Clayton DG, Bernardinelli L, Montomoli C (1993) Spatial correlation in ecological analysis. Int J Epidemiol 22(6):1193–1202

Cressie N (1993) Statistics for spatial data, Revised. Wiley Interscience, Hoboken, New Jersey

Cressie N, Davis AS, Folks JL, Policello GE (1981) The moment-generating function and negative integer moments. Am Stat 35(3):148–150

Dominici F, McDermott A, Hastie TJ (2004) Improved semiparametric time series models of air pollution and mortality. J Am Stat Assoc 99(468):938–948

Dupont E, Wood SN, Augustin N (2022) Spatial+: a novel approach to spatial confounding. Biometrics 78(4):1279–1290. https://doi.org/10.1111/biom.13656

Gardini A, Greco F, Trivisano C (2022) The mellin transform to manage quadratic forms in normal random variables. J Comput Graph Stat 31(4):1416–1425. https://doi.org/10.1080/10618600.2022.2034639

Guan Y, Page GL, Reich BJ, Ventrucci M, Yang S (2023) Spectral adjustment for spatial confounding. Biometrika 110(3):699–719. https://doi.org/10.1093/biomet/asac069

Hanks EM, Schliep EM, Hooten MB, Hoeting JA (2015) Restricted spatial regression in practice: geo-statistical models, confounding, and robustness under model misspecification. Environmetrics 26(4):243–254

Hefley TJ, Hooten MB, Hanks EM, Russell RE, Walsh DP (2017) The Bayesian group lasso for confounded spatial data. J Agric Biol Environ Stat 22(1):42–59

Hodges JS, Reich BJ (2010) Adding spatially-correlated errors can mess up the fixed effect you love. Am Stat 64(4):325–334

Hughes J, Haran M (2013) Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. J Royal Stat Soc Ser B 75(1):139–159

Hui FKC, Bondell HD (2022) Spatial confounding in generalized estimating equations. Am Stat 76(3):238–247

Lauricella G (1893) Sulle funzioni ipergeometriche a piu variabili. Rendiconti del Circolo Matematico di Palermo 7:111–158

Magnus JR (1986) The exact moments of a ratio of quadratic forms in normal variables. Annal dÉconomie et de Statistique 4:95–109

Marques I, Kneib T, Klein N (2022) Mitigating spatial confounding by explicitly correlating Gaussian random fields. Environmetrics 33(5):e2727. https://doi.org/10.1002/env.2727

Matérn B (1986) Spatial variation, 2nd edn. Springer, Berlin

Narcisi M (2023) On the effect of confounding in linear regression model: an approach based on the theory of quadratic forms. (Dissertation thesis, University of Bologna)

Nobre WS, Schmidt AM, Pereira JBM (2021) On the effects of spatial confounding in hierarchical models. Int Stat Rev 89(2):302–322

Paciorek CJ (2010) The importance of scale for spatial-confounding bias and precision of spatial regression estimators. Stat Sci 25(1):107–125

Page GL, Liu Y, He Z, Sun D (2017) Estimation and prediction in the presence of spatial confounding for spatial linear models. Scand J Stat 44(3):780–797

Paolella M (2018) Linear models and time-series analysis: regression, anova, arma and garch. Wiley, Hoboken, New Jersey

Papadogeorgou G, Choirat C, Zigler CM (2018) Adjusting for unmeasured spatial confounding with distance adjusted propensity score matching. Biostatistics 20(2):256–272

Provost S, Mathai A (1992) Quadratic forms in random variables: theory and applications. Marcel Dekker, New York

Reich BJ, Hodges JS, Zadnik V (2006) Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. Biometrics 62(4):1197–1206

Reich BJ, Yang S, Guan Y, Giffin AB, Miller MJ, Rappold A (2021) A review of spatial causal inference methods for environmental and epidemiological applications. Int Stat Rev 89(3):605–634

Roberts LA (1995) On the existence of moments of ratios of quadratic forms. Economet Theor 11(4):750–774

Rue H, Held L (2005) Gaussian markov random fields: theory and applications. CRC Press, New York

Sawa T (1978) The exact moments of the least squares estimator for the autoregressive model. J Econ 8(2):159–172

Thaden H, Kneib T (2018) Structural equation models for dealing with spatial confounding. Am Stat 72(3):239–252

Xiao-Li M (2005) From unit root to Stein's estimator to Fisher's K statistics: if you have a moment, I can tell you more. Stat Sci 20(2):141–162

Yang J (2021) On the extreme eigenvalues of the precision matrix of the nonstationary autoregressive process and its applications to outlier estimation of panel time series. arXiv preprint arXiv:2109.02204