

Decompositions

Sophie Woodward

2023-06-08

Simulation Study

Denote X as exposure X , Z as unmeasured confounder, and Y as outcome.

Simulation 1: Nested DGM

We simulate the following scenario 100 times. Across $n_1 = 25$ states of 5×5 grids,

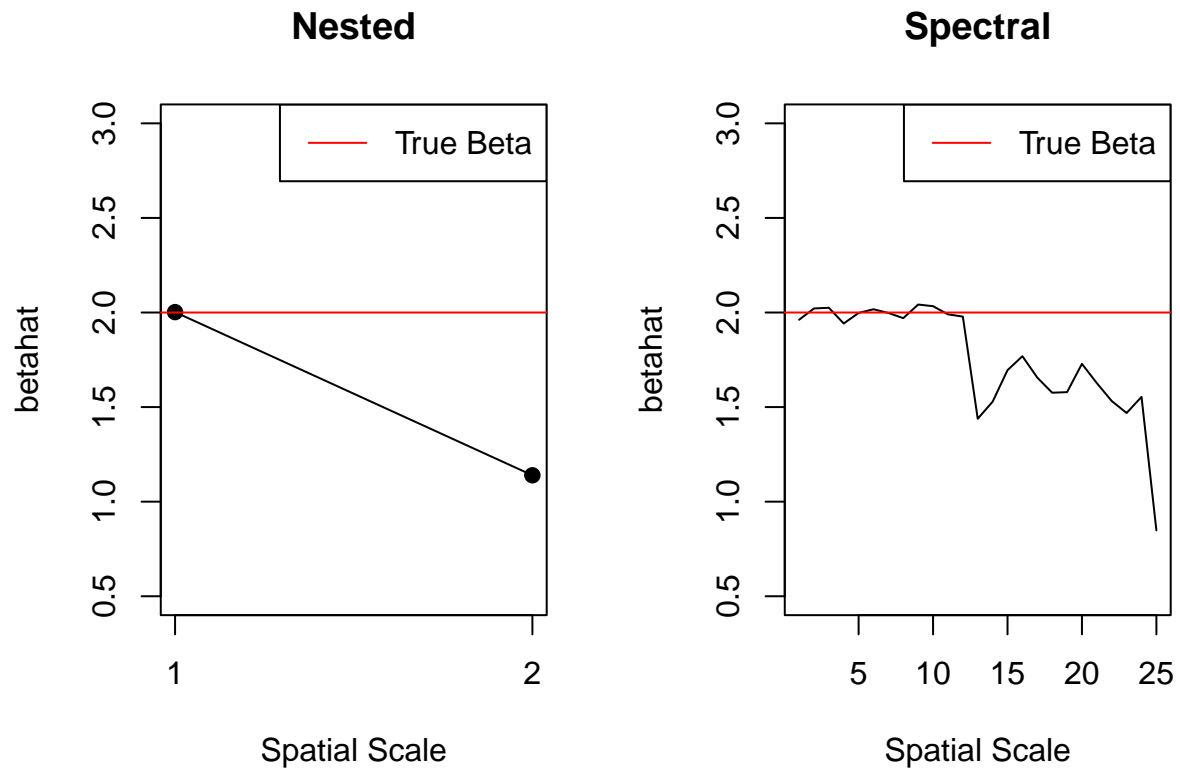
$$\begin{aligned}X_{1,s} &\sim \text{Exp}(1) \\Z_{1,s} &\sim 0.9X_2 + \sqrt{1 - 0.9^2}\text{Exp}(1)\end{aligned}$$

for $s = 1, \dots, n_1 = 25$. Across $n_2 = 625$ counties of 1×1 grids, let

$$\begin{aligned}X_{2,i} &\sim \text{Exp}(1) \\Z_{2,i} &\sim 0.001X_2 + \sqrt{1 - 0.001^2}\text{Exp}(1)\end{aligned}$$

and $X_i = X_{1,s(i)} + X_{2,i}$, $Z_i = Z_{1,s(i)} + Z_{2,i}$ for $i = 1, \dots, n_2 = 625$. By construction, X, Z are nearly uncorrelated within the states of 5×5 , but correlated across states. (Note that this DGM exactly follows the nested decomposition.) We let $Y_i = 2X_i - Z_i + \epsilon$ where $\epsilon_i \sim \mathcal{N}(0, 1)$ independently across i .

For each of the 100 scenarios, we decompose X, Z, Y at different spatial scales using 1) nested decomposition and 2) spectral decomposition. At each spatial scale ω , we obtain an estimate $\hat{\beta}(\omega)$ of $\beta = 2$ from a linear regression of $Y(\omega)$ on $X(\omega)$. I hypothesize that $\hat{\beta}(\omega)$ is unbiased for high ω (finer spatial scales) since by construction confounding dissipates locally.



Looking at the nested plot, it's nice to see that the bias of the coefficient is 0 at low spatial scales when the DGM is indeed nested. Looking at the spectral plot, we see that the spectral decomposition also results in a bias of 0 for low spatial scales. The curve here takes on an almost-stepwise shape.

Simulation 2: Spectral DGM

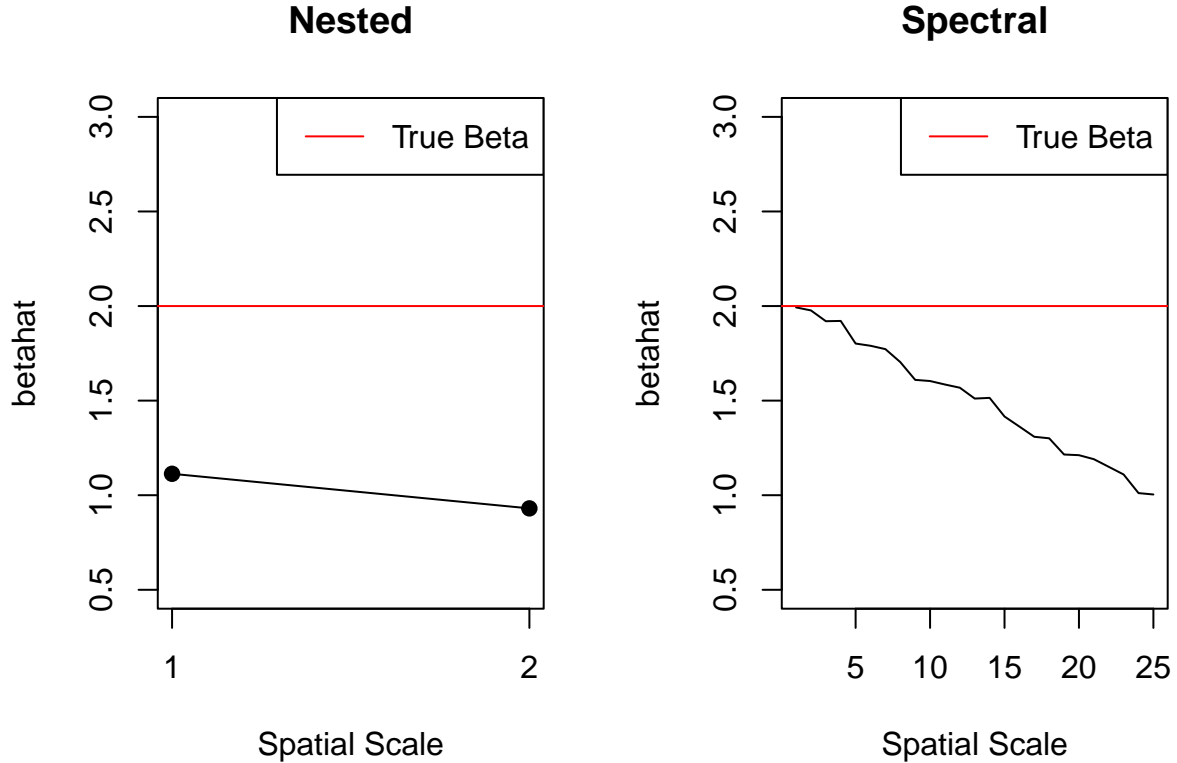
We repeat simulation 1 but now the data-generating model originates from the spectral decomposition rather than the nested. In particular, we use the graph Fourier transform to project X and Z into the spectral domain. In the spectral domain,

$$X_i^* \sim \text{Exp}(1)$$

$$Z_i^* \sim \rho_i X_2 + \sqrt{1 - \rho_i^2} \text{Exp}(1)$$

for $i = 1, \dots, 625$ where $\rho_i = 0, 1/624, 2/624, \dots, 1$. So the covariance between X and Z dissipates for smaller i , which corresponds to larger eigenvalues ω of the graph Laplacian, or smaller spatial scales.

Again, for each of the 100 scenarios we decompose X, Z, Y at different spatial scales using 1) nested decomposition and 2) spectral decomposition. At each spatial scale ω , we obtain an estimate $\hat{\beta}(\omega)$ of $\beta = 2$ from a linear regression of $Y(\omega)$ on $X(\omega)$. I hypothesize that $\hat{\beta}(\omega)$ is unbiased for high ω (finer spatial scales) since by construction confounding dissipates locally.



Looking at the spectral plot, it's nice to see that the bias of the coefficient is 0 at low spatial scales when the DGM is indeed spectral. The estimates obtained from the nested decomposition are biased at both of the two grid levels.

Simulation 3: Local Confounding

We repeat simulations 1 and 2 but now confounding dissipates at larger scales.

For the nested DGM,

$$X_{1,s} \sim \text{Exp}(1)$$

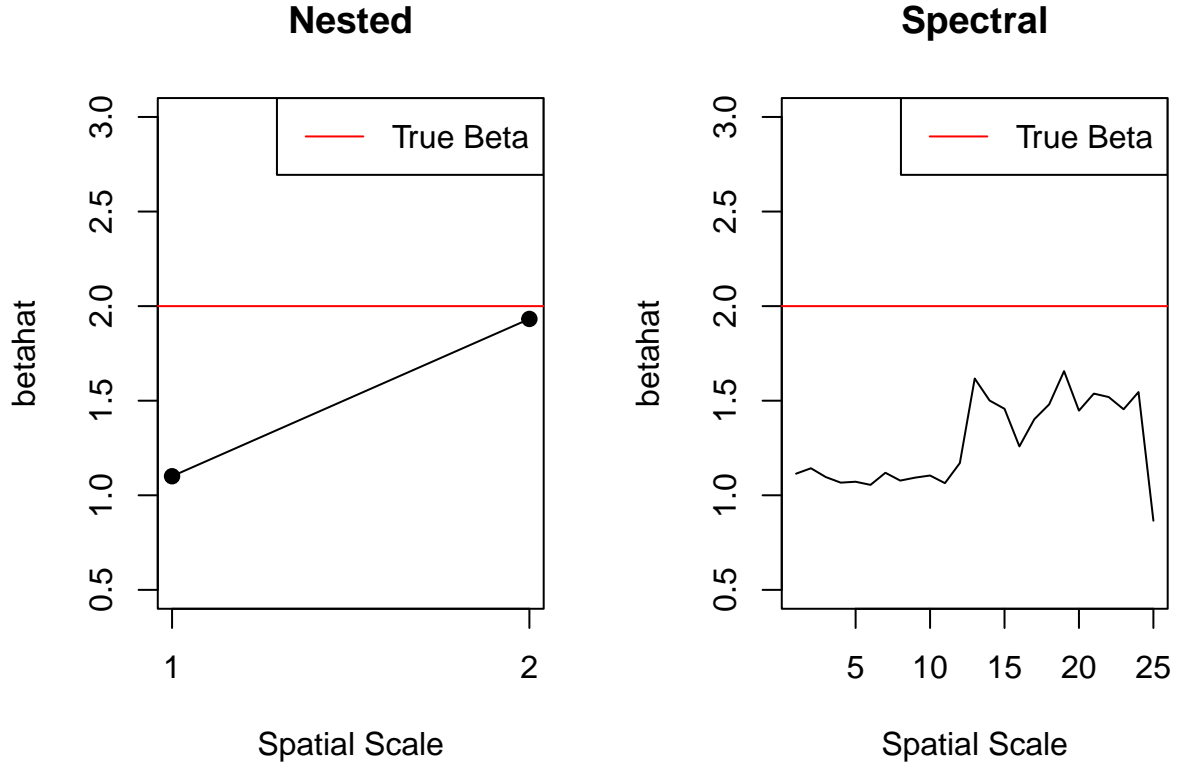
$$Z_{1,s} \sim 0.001X_2 + \sqrt{1 - 0.001^2}\text{Exp}(1)$$

for each 5×5 state, $s = 1, \dots, n_1 = 25$. Across $n_2 = 625$ counties of 1×1 grids, let

$$X_{2,i} \sim \text{Exp}(1)$$

$$Z_{2,i} \sim 0.9X_2 + \sqrt{1 - 0.9^2}\text{Exp}(1)$$

and $X_i = X_{1,s(i)} + X_{2,i}$, $Z_i = Z_{1,s(i)} + Z_{2,i}$ for $i = 1, \dots, n_2 = 625$. By construction, X, Z are nearly uncorrelated across the states of 5×5 , but correlated within states.

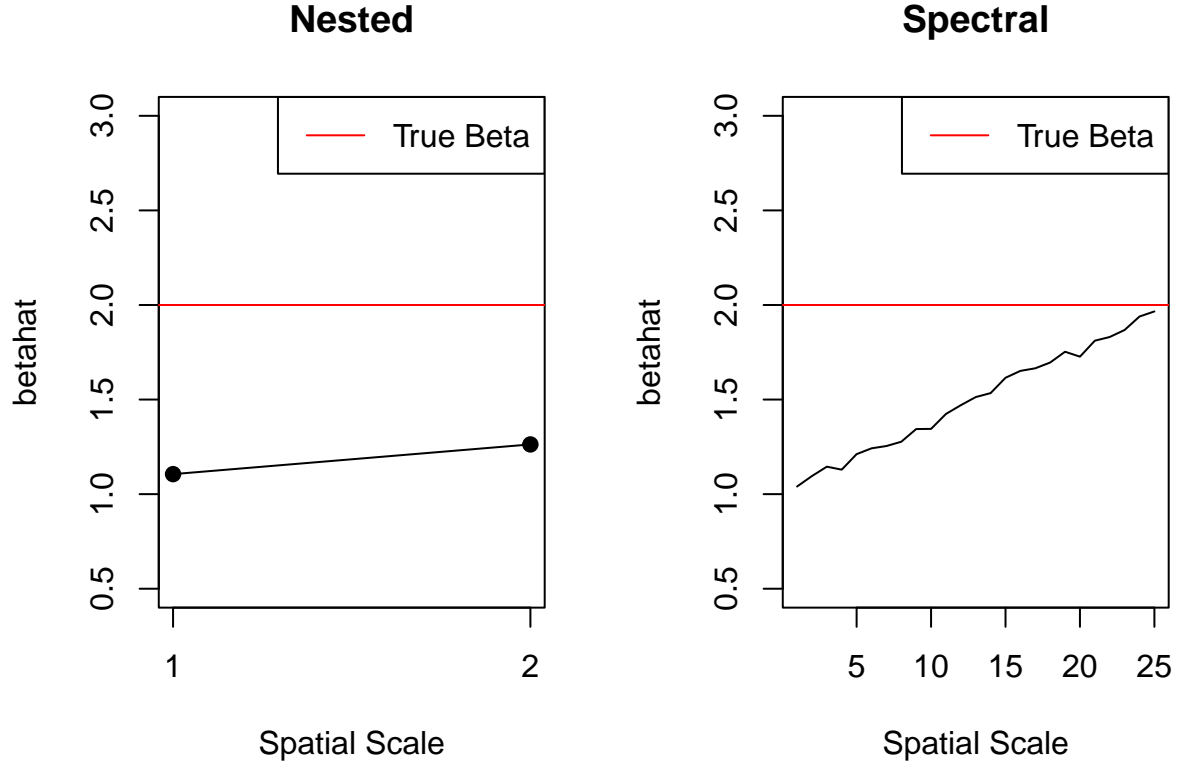


For the spectral DGM,

$$X_i^* \sim \text{Exp}(1)$$

$$Z_i^* \sim \rho_i X_2 + \sqrt{1 - \rho_i^2} \text{Exp}(1)$$

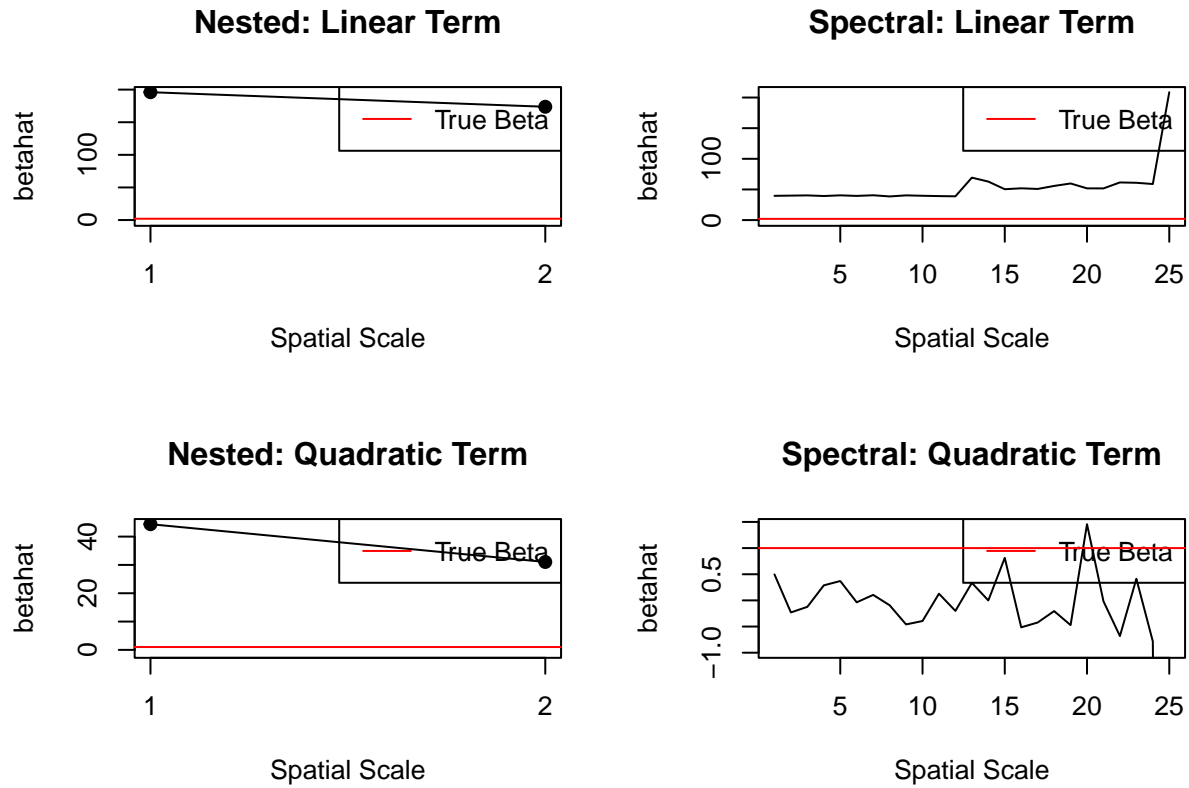
for $i = 1, \dots, 625$ where $\rho_i = 1, 623/624, 622/624, \dots, 0$.



Simulation 4: Nonlinear Outcome Model

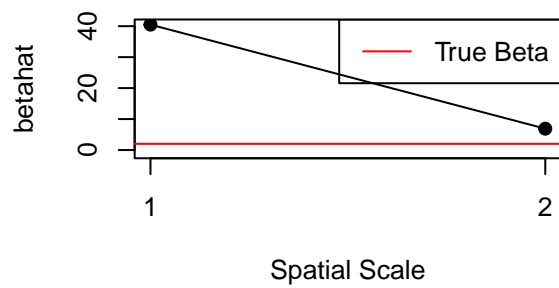
We repeat simulation 1 and 2 but the outcome model now includes a quadratic term of X . In particular, $Y_i = 2X_i + X_i^2 - Z_i + \epsilon$. The scale-specific analyses attempt to estimate both coefficients of X . Neither method does a good job at recovering either of the coefficients.

Nested DGM

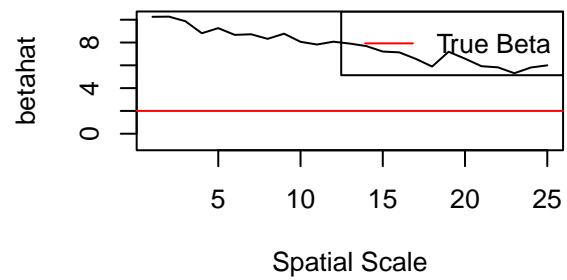


Spectral DGM

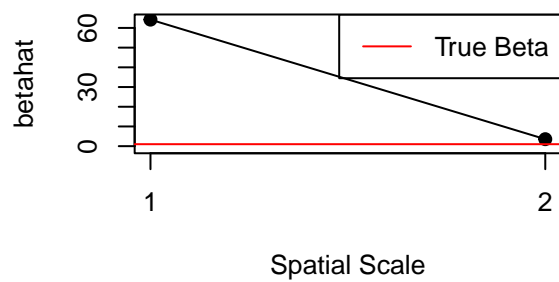
Nested: Linear Term



Spectral: Linear Term



Nested: Quadratic Term



Spectral: Quadratic Term

