

**Design and results from the application of all three methods to three simulated datasets**

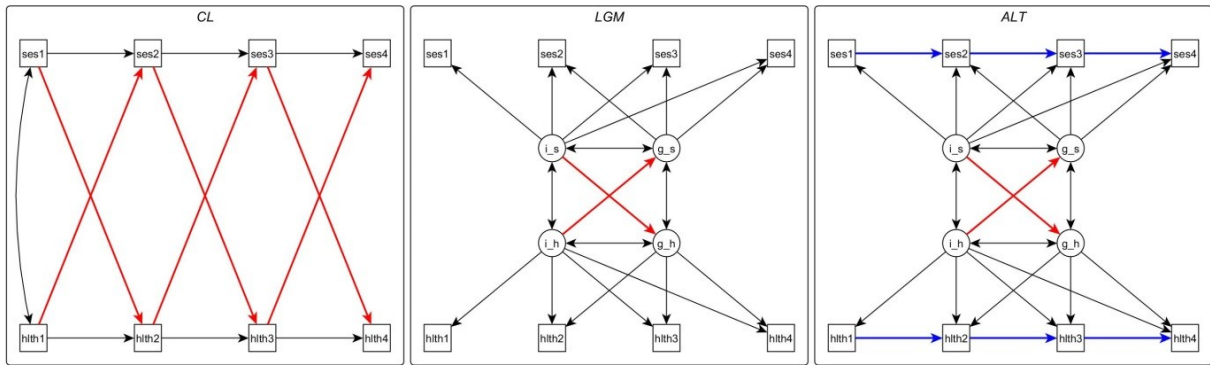


Figure 1: The design of the three simulations.

We have four observed variables, SES1 represents the socioeconomic status at time 1; HEALTH1 represents the health status at time 1, etc. Our focus in this simulation is on the ratio between social causation (SC) and health selection (HS). For all three simulations, we generate N=5000 with 1000 replications, constant autoregression parameters (AR) and with health that declines with age. The ratio between SC and HS =2 (fixed) which means SC is stronger than HS. Specifically, in CL simulation (see Figure 1, CL part) we define the parameters as follows: AR=0.5; SC=0.1; HS=0.05; all variances=1; and the correlation between SES1 and HEALTH1=0.2. In LGM the parameters of all covariances of the trajectories are 0.1; all variances of observed variables are 1; SC is 0.1, and HS is 0.05. Finally, in ALT all parameters are exactly like LGM simulation above plus the ARs = 0.5.

In Table 1 below we present the results. It represents a 3-by-3 table where the columns show the simulated data following Figure 1 above and the rows show the model used to analyse the data. Logically, the diagonal in grey shows the correct results close to the imposed ratio of 2, because here the same model is used to create the data and to analyse it. For example, in the CL simulation, when we analyse the CL data using CL method, as expected the estimated ratios equal 2. However, the ratio is different when the CL data is analysed using LGM. It is 0.49, while using ALT is 14.48. For ALT we set SES1 and HEALTH1 as exogenous covariates, otherwise the estimation will not be convergent.

		Simulation					
		CL		LGM		ALT	
Model	CL	P12	2.00	P12	12.80	P12	1.53
		P23	2.00	P23	7.23	P23	1.58
		P34	2.00	P34	4.85	P34	2.71
	LGM	0.49	2.04		2.06		
	ALT	14.48	1.83	2.10			

Table 1. The comparison of CL, LGM and ALT with simulated data (the main diagonal is when the data is properly analysed with its method, while the off diagonal is when it is analysed using other methods). P12 refers to “between Time 1 and Time 2”, etc.

In the LGM simulation, when we use the CL method we obtain ratios for the three periods far above 2, and using the ALT method we obtain a ratio similar to 2. In the ALT simulation using the CL method we obtain ratios for the three periods that are different from 2, but their average (1.924) is close to 2. Using the LGM or the ALT model we again obtain results close to 2.

What we conclude from these simulations is that within the group of LGM and ALT it does not matter greatly which model is used to simulate or analyse the data, confirming the similarity of these two models. However, data simulated with either LGM or ALT cannot be analysed with CL or vice versa. These simulation results can only partly guide the use of these models with real data because the latter is not created by any of these models and thus the correctness of the results will not depend so much on the use of the right model. However, our results illustrate the sensitivity and specificity of the models in the resulting ratio between SC and HS with respect to the underlying data structure.