Simulation study 2 with fixed sample size

Tetiana Gorbach

2/23/2021

Contents

The data generating process is adapted from the simulation study in Fulcher et al, 2020 to the DAG in Figure 1a:

$$C \sim \text{Bern}(0.5)$$
$$A|C \sim \text{Bern}\{expit(0.5C)\}$$
$$Z|A \sim N(\beta_1 A, 1)$$
$$Y|Z, C \sim N(1 + \gamma_1 Z + \gamma_2 C, 1).$$

Parameter values in the simulations are

- sample size = 1000

- number sample of replicates K = 1000

- β_1 = -1.5, -1, -0.5, -0.1, 0, 0.1, 0.5, 1, 1.5

- γ_1 = -2, -1, -0.5, -0.1, 0, 0.1, 0.5, 1, 2

- γ_2 = -2, -1, -0.5, -0.1, 0, 0.1, 0.5, 1, 2

- $\sigma_z = 1$

- $\sigma_y = 1$

Expectations from the simulations study:

- Since considered data model satisfies all the criteria, the EIF constructed under the assumption that all the criteria are fulfilled (called later "TD BD FD") should have the lowest empirical variance

- The variance of the influence function constructed under TD and BD assumptions simultaneously (called "TD BD") should be smaller than the variance of the IFs under only TD or only BD assumptions

- The variance of the influence function constructed under TD and FD assumptions simultaneously (called "TD FD") should be smaller than the variance of the IFs under only TD or only FD assumptions

- The variance of the influence function constructed under FD and BD assumptions simultanesouly (equal to "TD BD FD") should be smaller than the variance of the IFs under only FD or under only BD assumptions

Scaled empirical variance, all methods together

Plot all the results. Not possible to see anything, except that the results for positive and negative values of the parameters seem to be approximatelly symmetrical. That is why I will later construct plots only for positive values of the parameters.
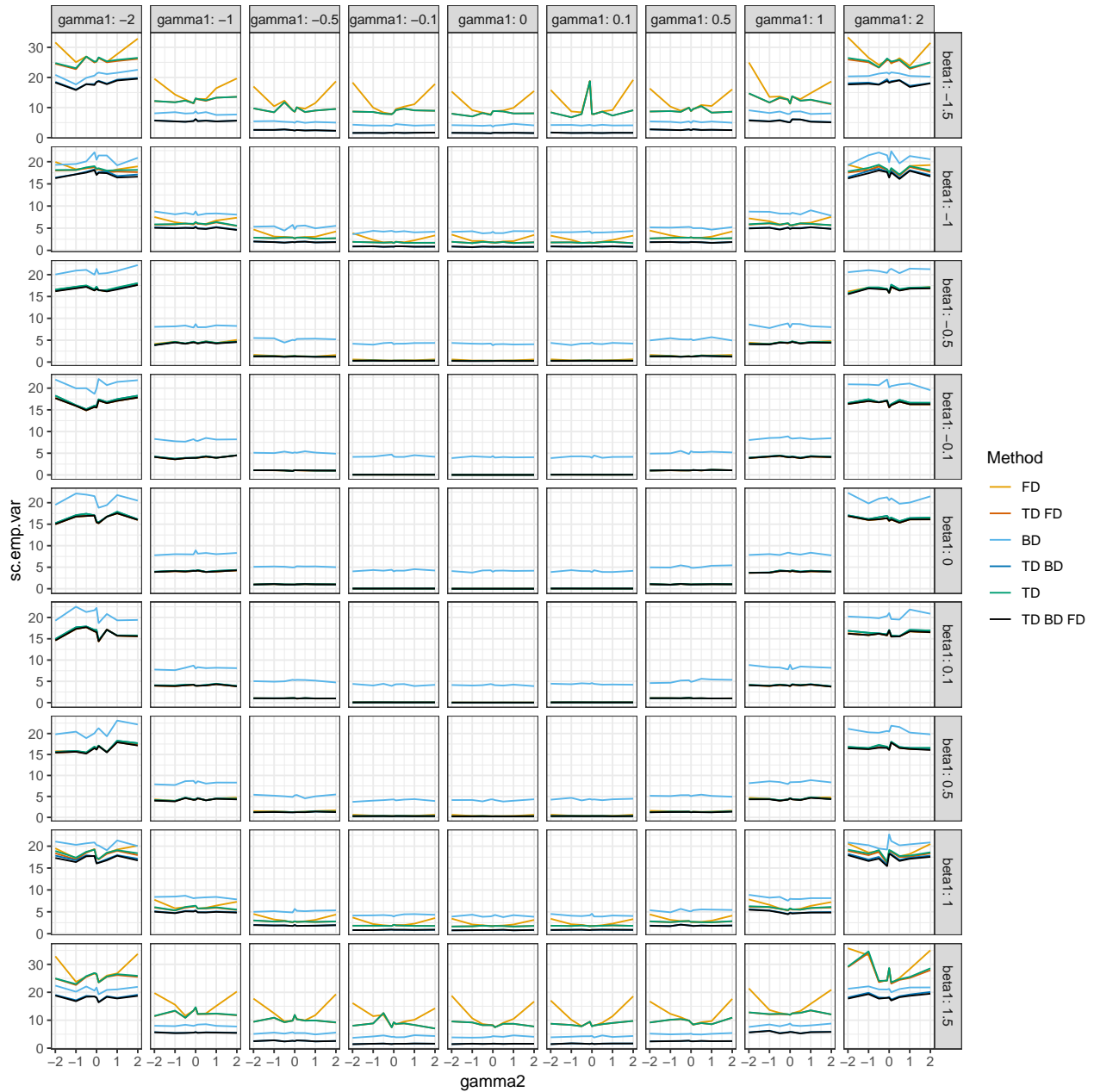
Plot for only positive values of the parameters. Conclusions from the plot:

- Still hard to distinquish between the lines for low β_1

- Empirical variances increase with increasing γ_1 (effect of Z on Y).

- The emp. variance of the BD estimator is higher than the others unless β_1 is big (last row). For big β_1 FD, TD, and TD+FD have higher variances.
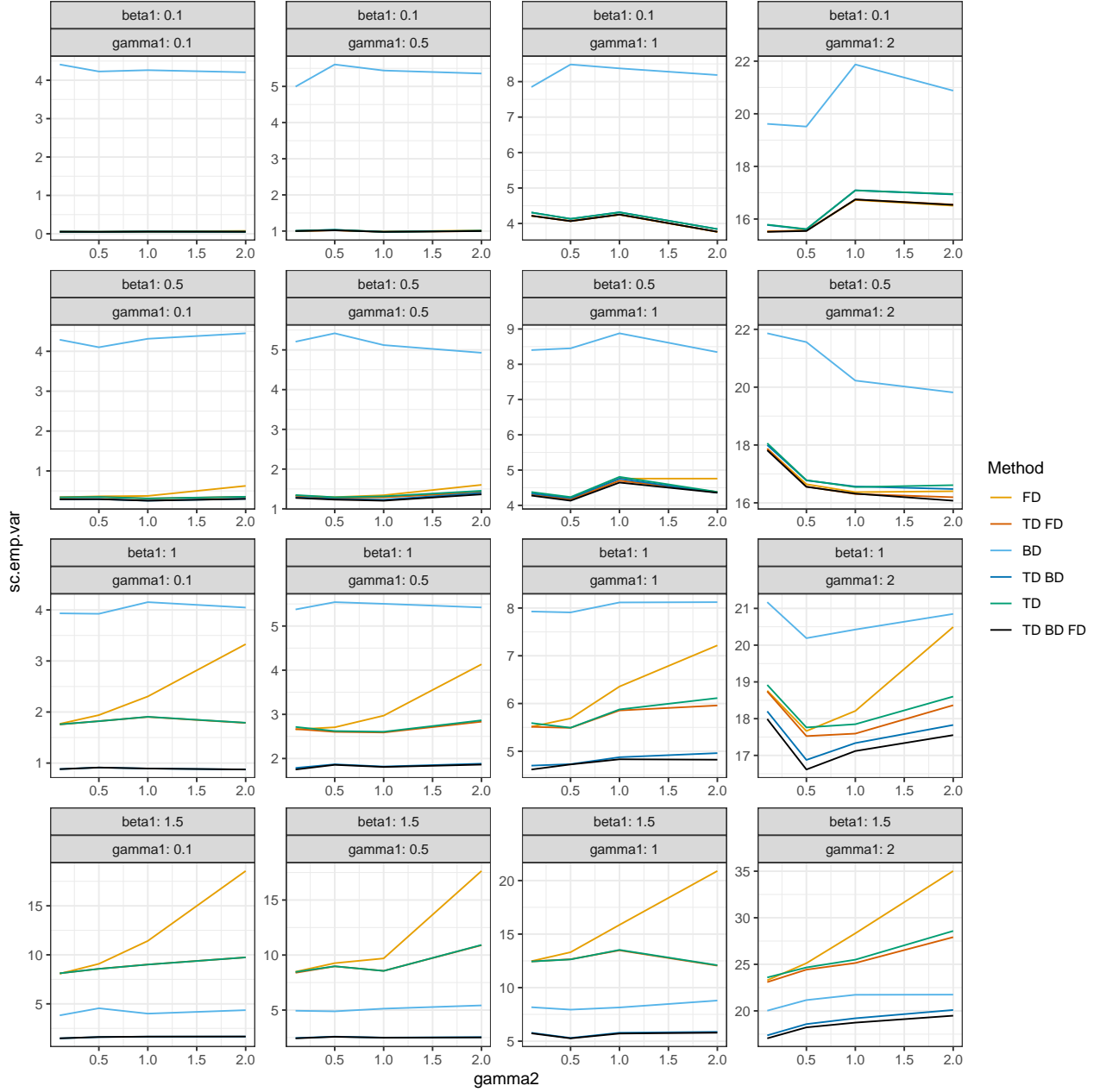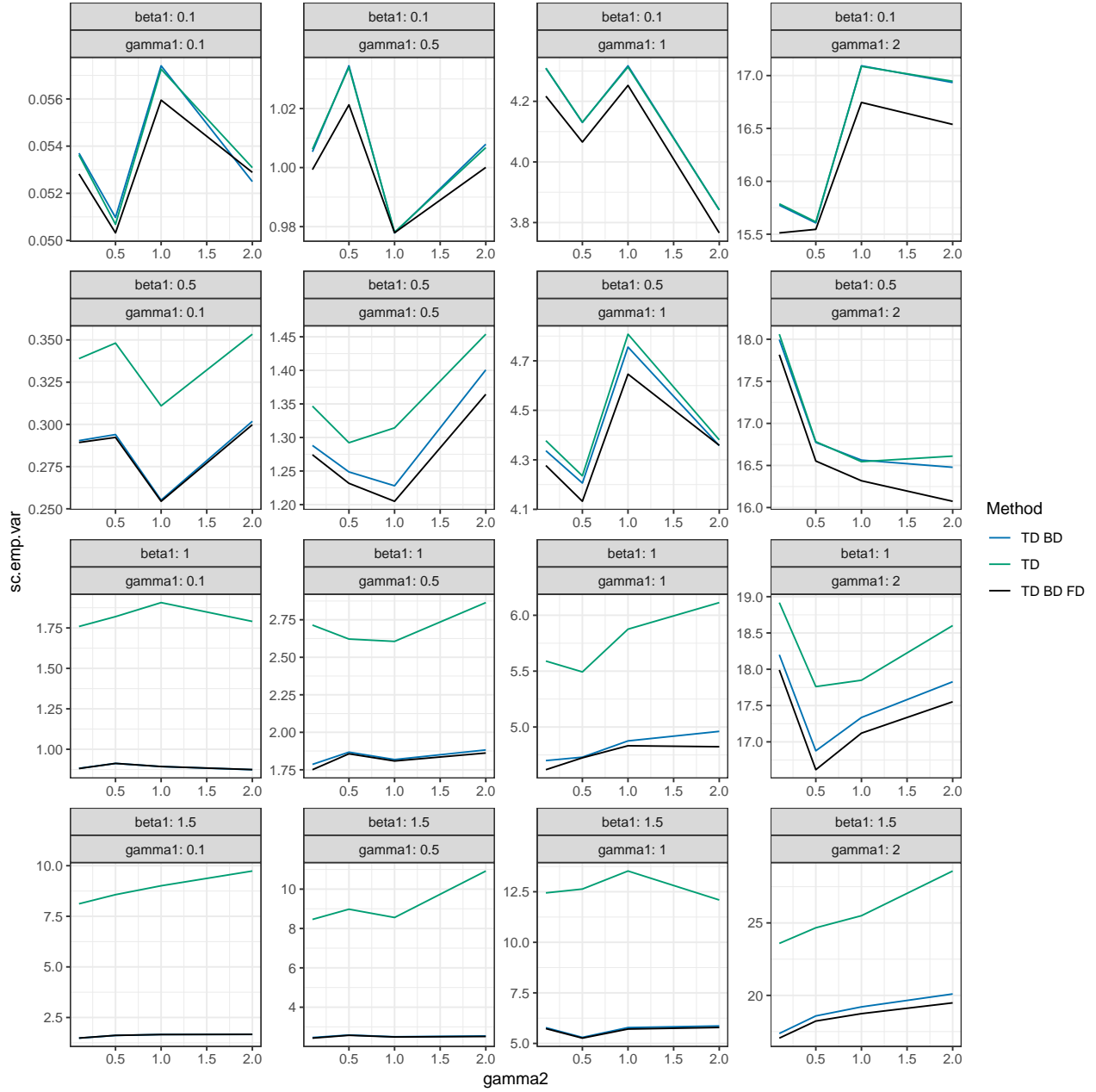
Scaled empirical variance for TD, BD and TD+BD

- BD variance is higher than TD for small β_1 (this is expected from the Examples in the paper).

- As expected, the empirical variance of TD+BD seems to be smaller than individual variance of TD or BD for any combination of parameters (maybe hard to evaluate for small β_1, see the plot on the next page)

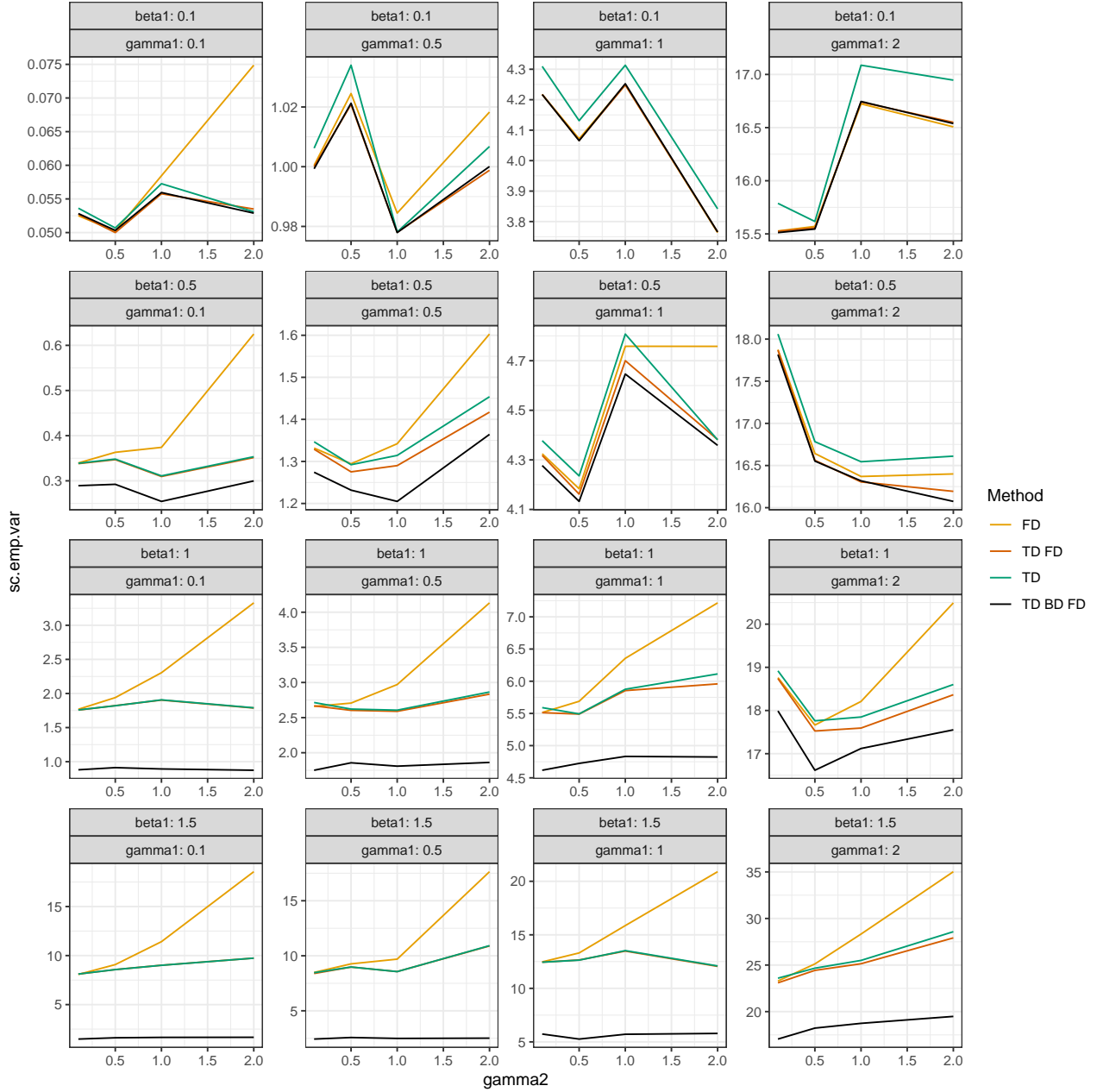- TD BD FD has the lowest variance (see the plot on the next page).
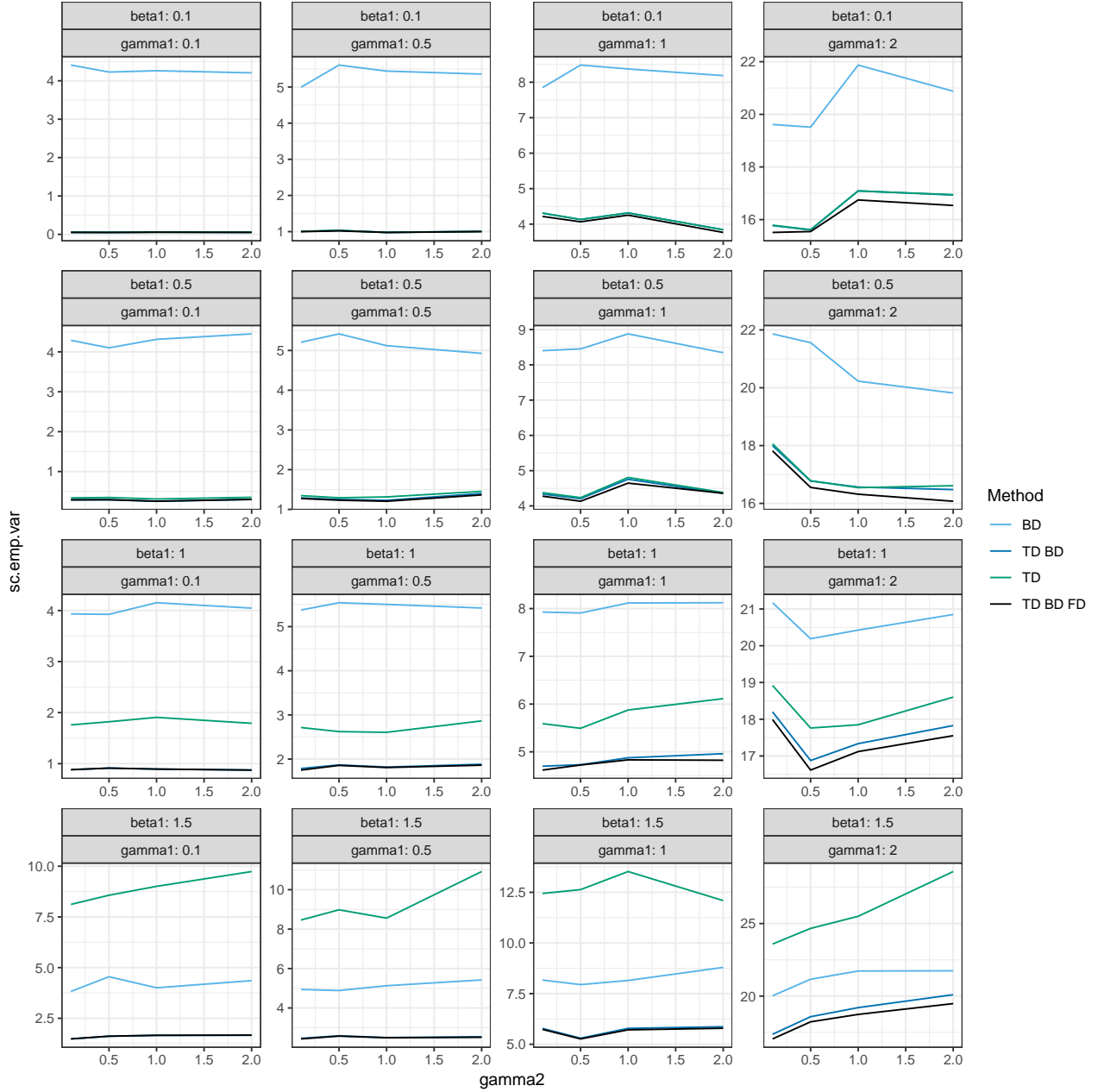
Scaled empirical variance for TD, BD and TD+BD (without BD)

To be able to compare TD, TD+BD, TD+BD+FD for small β_1, I do not plot BD:
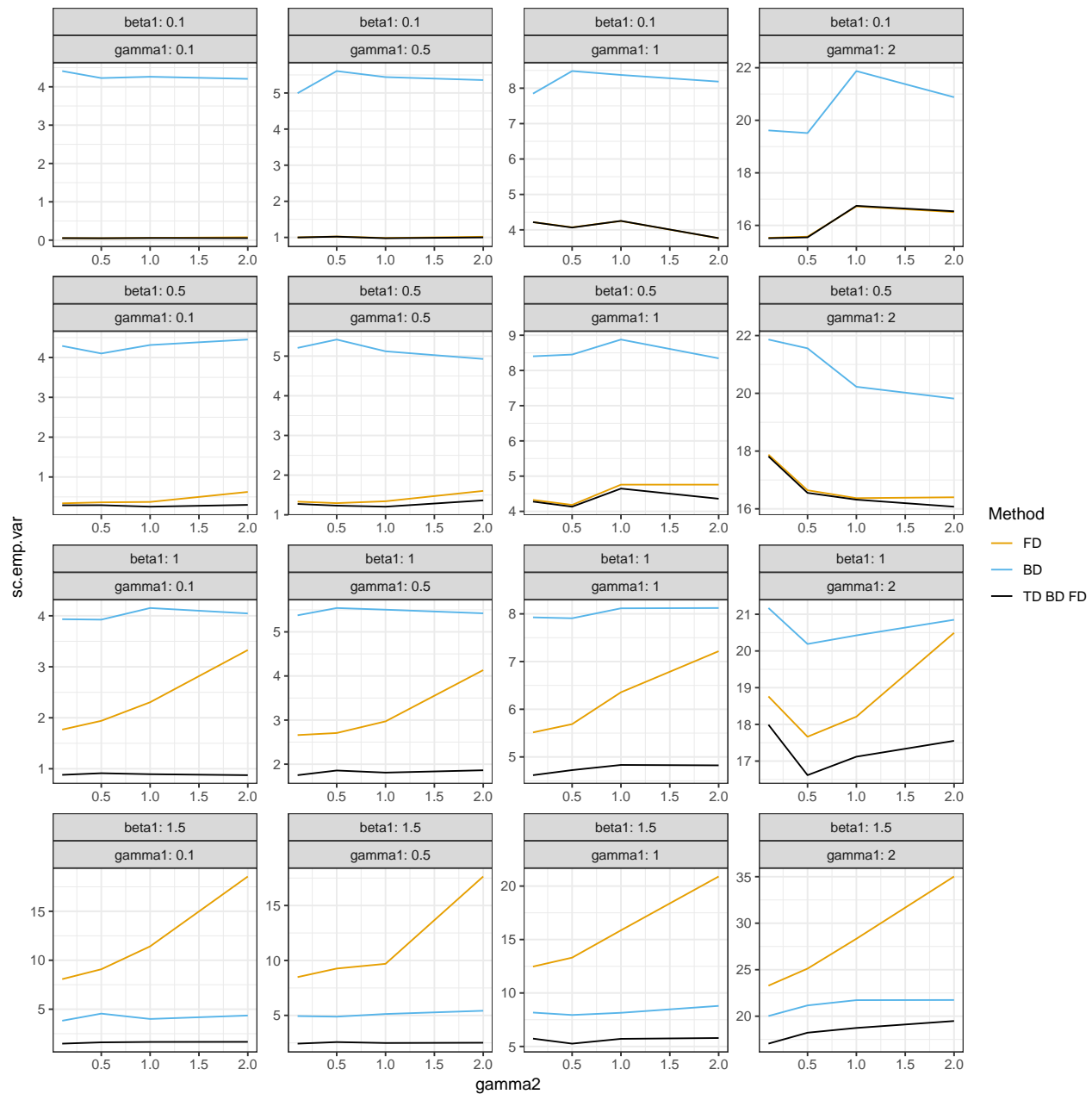
Scaled empirical variance for TD, FD and TD+FD

- As expected, the empirical variance of TD+FD seems to be smaller than individual variance of TD or FD for any combination of parameters.

- TD BD FD has the lowest variance (I blame the result for $\beta_1 = 0.1$ and $gamma_1 = 2$ on sampling variablility, see next page).
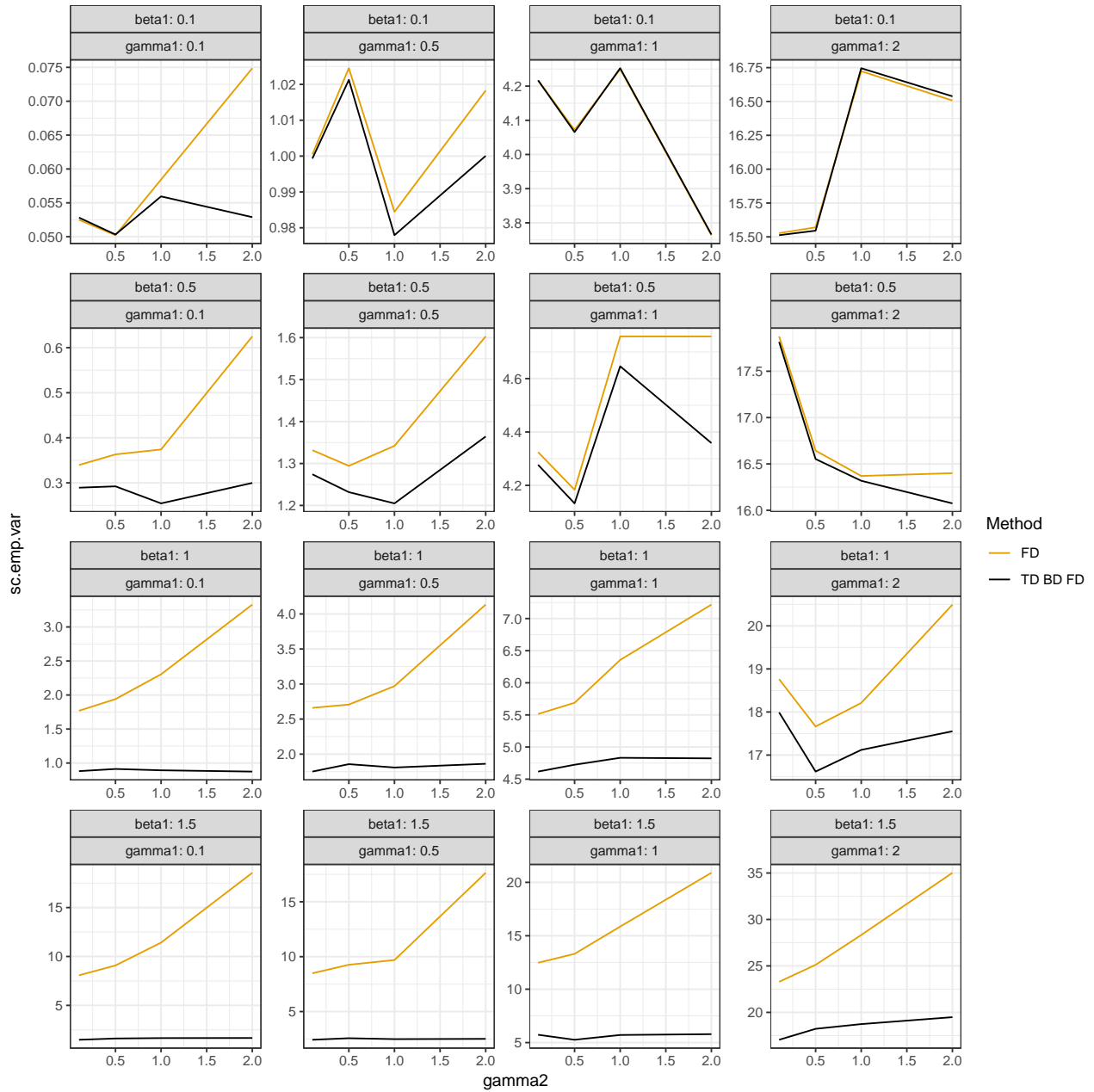
Scaled empirical variance for FD, BD and BD+FD (=TD BD FD)

- TD BD FD has the lowest variance.

Scaled empirical variance for FD, BD and BD+FD (=TD BD FD) (without Bd)

- As expected, the empirical variance of TD BD FD seems to be smaller than individual variance of FD or BD for any combination of parameters.

Summary

- The empirical results from the simulation study correspond to the expectations.
- Show bounds?