

Bias Assessment in Local Regression - Supplementary Material

Wenkai Ma¹ and W. John Braun²

1 Simulation Study

In this document, the finite-sample behavior for the proposed local polynomial bias estimator with pointwise confidence interval for the estimate of bias is investigated. We use the `BiasVarPlot()` function in the *MPV* (Braun, 2019) package in R (R Core Team, 2019).

The simulation study design is summarized in Table 1. Target functions from He and Huang (2009) are employed since they represent different noise to signal ratio magnitude. Noise to signal is defined as $\sigma/(\text{Var}(g(x)) + \sigma^2)^{1/2}$. The higher noise to ratio statistic, the harder the estimation will be in terms of choosing the optimal degree and bandwidth.

Example	$g(x)$	σ	Noise/Signal
1	$x + 2e^{-16x^2}$	0.4	1/3
2	$\sin(2x) + 2e^{-16x^2}$	0.3	1/3
3	$0.3e^{-4(x+1)^2} + 0.7e^{-16*(x-1)^2}$	0.1	1/2

Table 1: Target functions, σ and noise to signal ratios considered in the simulation study.

The simulation study follows these steps:

1. Randomly generate x and y according to target functions on the predefined support $[-2, 2]$.
2. Fit LPR to the simulated ordered pairs, using degree k_1 and bandwidth h_1 .
3. Calculate the true bias at each grid point.
4. Compute the confidence interval for the bias according to the methodology described in the main paper. using degree k_2 and bandwidth h_2 .
5. Compute the confidence interval for the uncorrected local linear estimator and for the bias-corrected estimator.
6. Check whether the confidence intervals for the bias contain the true bias at each x point.

¹Workplace Safety and Insurance Board, Toronto, Ontario.

²Department of Computer Science, Mathematics, Physics and Statistics, University of British Columbia, Kelowna, BC

7. Check whether the confidence intervals for the regression function contain the true function at each x point.
8. Repeat steps 1 through 7. We used 1000 repetitions.
9. Obtain the true confidence interval coverage, by computing the proportion of confidential intervals containing the true bias. We also computed the coverage for the uncorrected and bias-corrected confidence intervals.

According to the four scenarios in Table 2, simulation results for the first target function are shown in Figures 2 to 5, for the second target function, results are in Figures 6 to 9 and for the second target function, results are in Figures 10 to 13. In particular, we plotted the pointwise coverages for each of the values of x in the domain of interest. We observed that almost all of these simulation worked well, in that the actual coverage randomly fluctuates about the nominal coverage, though in a few cases, it might be somewhat conservative in that actual coverage slightly exceeds the nominal level.

Scenario	Nominal Level	h_2	Sample Size	k_2
s1	0.95	h	100	$k + 1$
s2	0.95	h	200	$k + 1$
s3	0.95	$h/2$	100	k
s4	0.95	$h/2$	200	k

Table 2: Simulation design table containing nominal confidence level, relative bandwidth, the number of simulated observations per sample and relative degree. Each row represents a scenario.

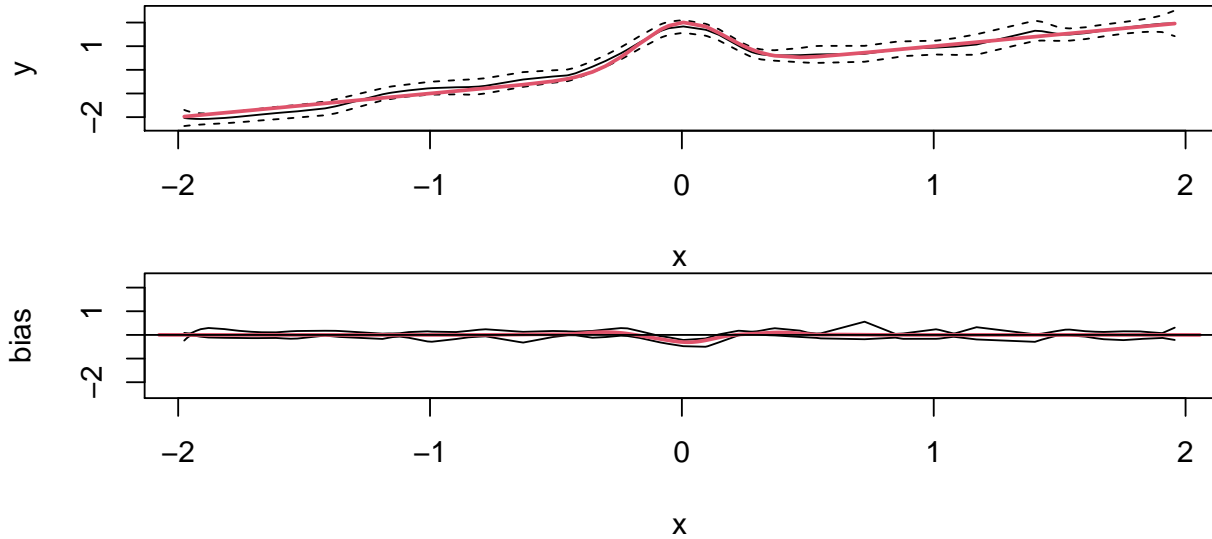


Figure 1: Pointwise Bias assessment tool for Example 1 according to the setup in Scenario 1. Top Panel: Target function (red curve), Local Polynomial regression (solid black) and confidence limits (dashed black). Bottom Panel: True bias (red curve), confidence interval for bias in one simulation (black curve)

References

- Braun, W.J. (2019). MPV: Data Sets from Montgomery, Peck and Vining. R package version 1.55.
- He, H. and Huang, L.-S. (2009). Double-smoothing for bias reduction in local linear regression. *Journal of Statistical Planning and Inference*, 139(3):1056–1072.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

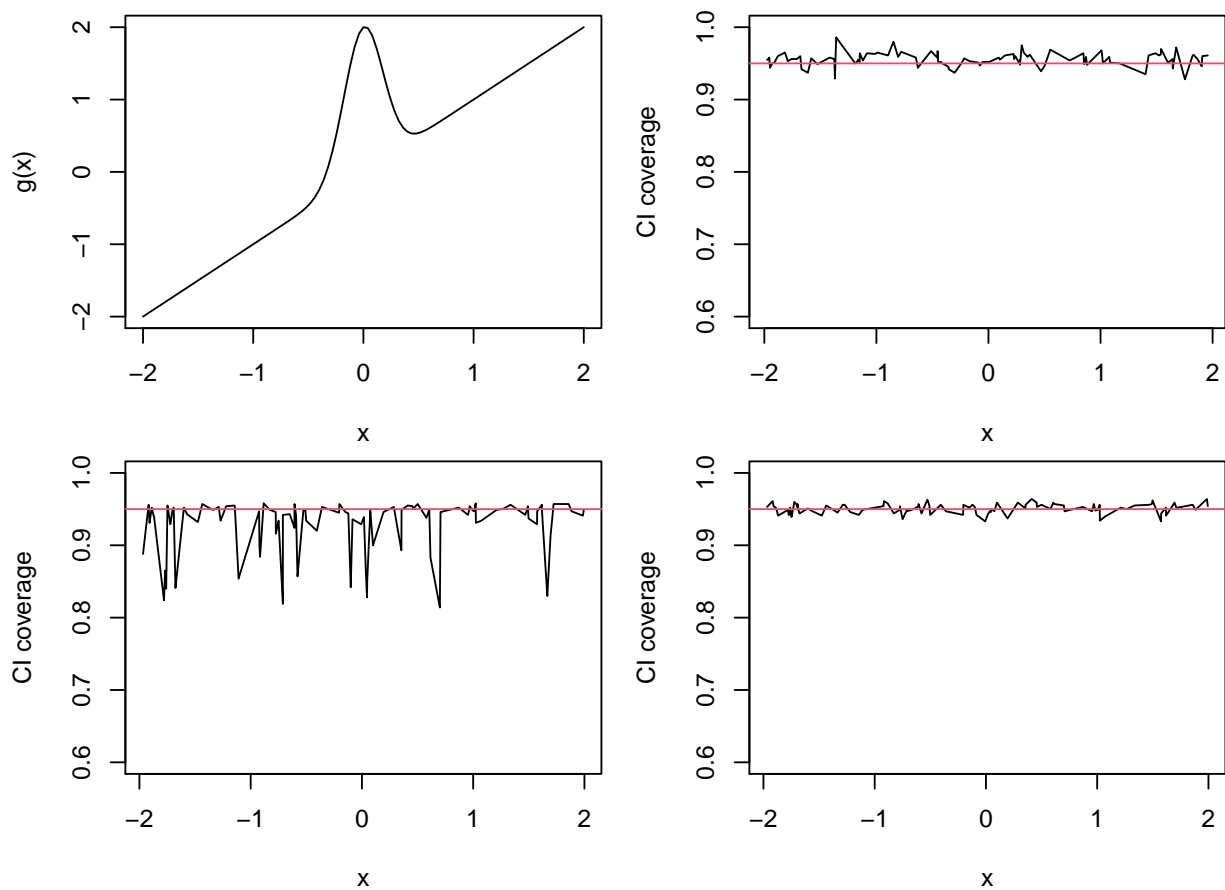


Figure 2: Pointwise Bias assessment tool performance for local linear regression for Example 1 according to the setup in Scenario 1. For reference, the target function is pictured in the top left panel. Confidence interval coverage for the bias $B(x)$ is plotted as a function of x in the top right panel, with the nominal level plotted as a horizontal line in red. The panels on the bottom show confidence interval coverage for the regression function $g(x)$, using the uncorrected estimator (left) and for the bias-corrected estimator (right).

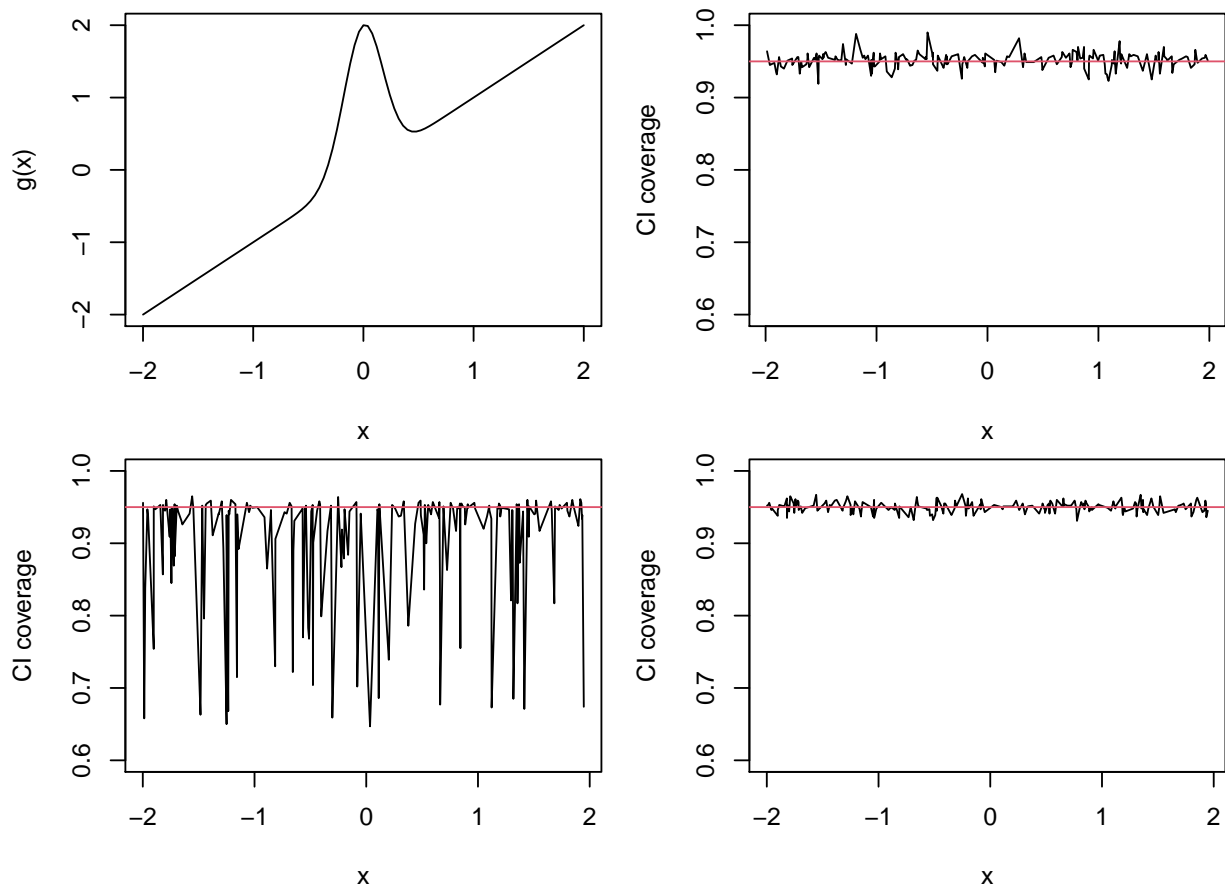


Figure 3: Pointwise Bias assessment tool performance for local linear regression for Example 1 according to the setup in Scenario 2. For reference, the target function is pictured in the top left panel. Confidence interval coverage for the bias $B(x)$ is plotted as a function of x in the top right panel, with the nominal level plotted as a horizontal line in red. The panels on the bottom show confidence interval coverage for the regression function $g(x)$, using the uncorrected estimator (left) and for the bias-corrected estimator (right).

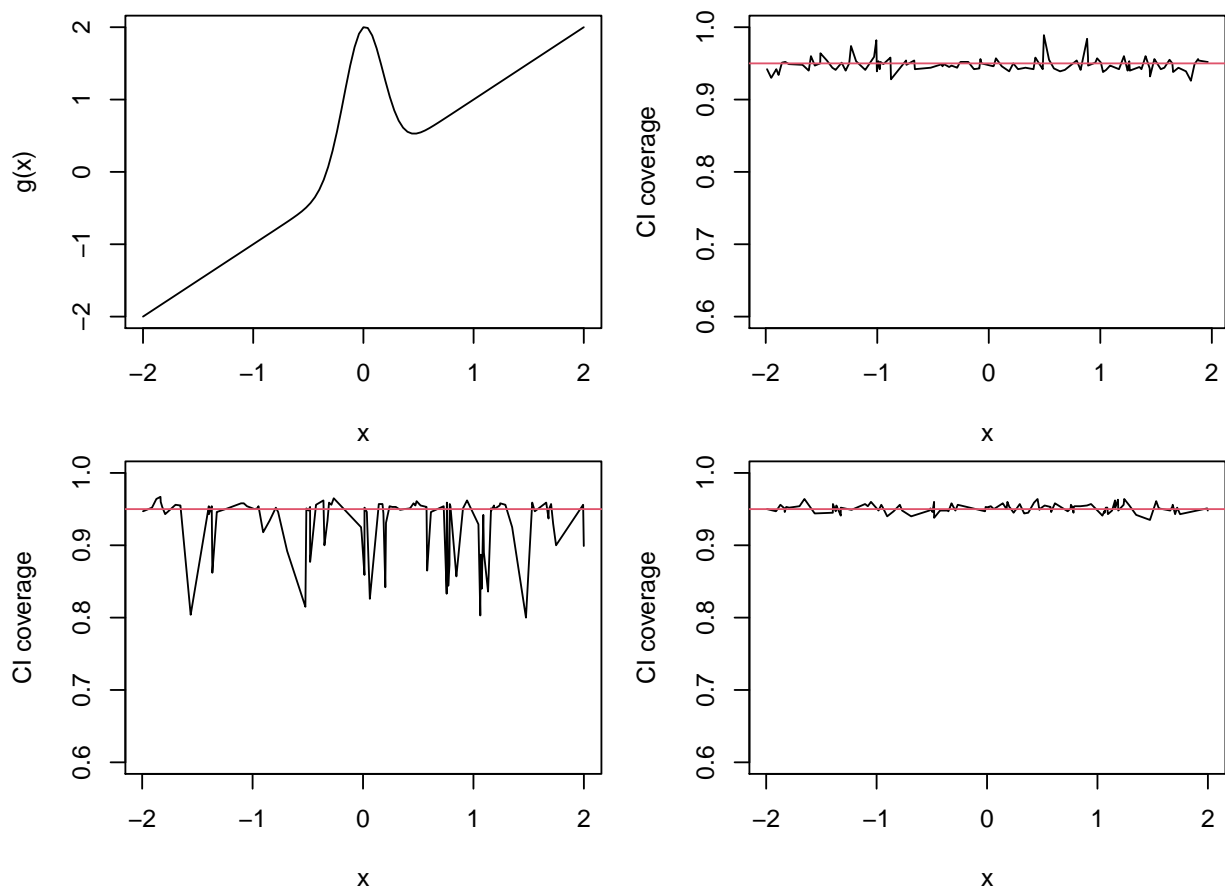


Figure 4: Pointwise Bias assessment tool performance for local linear regression for Example 1 according to the setup in Scenario 3. For reference, the target function is pictured in the top left panel. Confidence interval coverage for the bias $B(x)$ is plotted as a function of x in the top right panel, with the nominal level plotted as a horizontal line in red. The panels on the bottom show confidence interval coverage for the regression function $g(x)$, using the uncorrected estimator (left) and for the bias-corrected estimator (right).

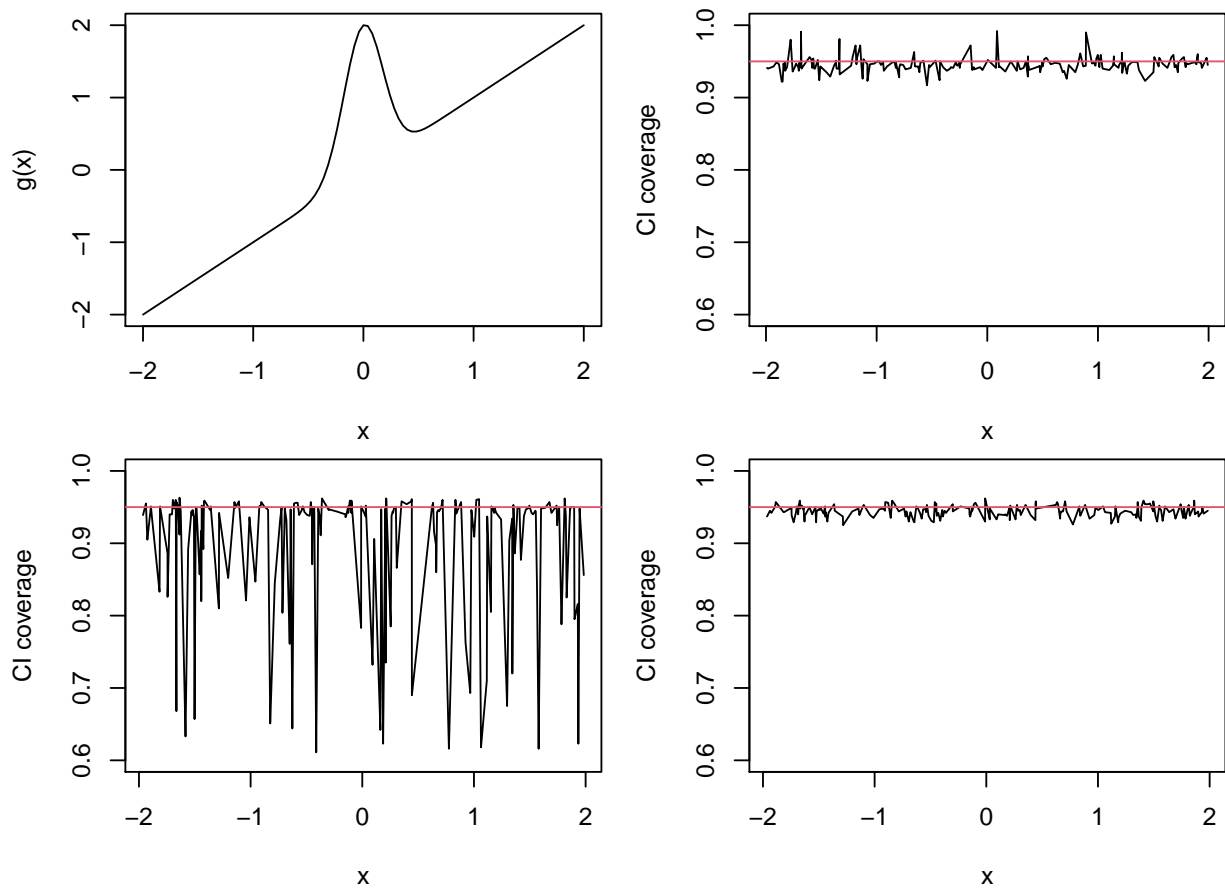


Figure 5: Pointwise Bias assessment tool performance for local linear regression for Example 1 according to the setup in Scenario 4. For reference, the target function is pictured in the top left panel. Confidence interval coverage for the bias $B(x)$ is plotted as a function of x in the top right panel, with the nominal level plotted as a horizontal line in red. The panels on the bottom show confidence interval coverage for the regression function $g(x)$, using the uncorrected estimator (left) and for the bias-corrected estimator (right).

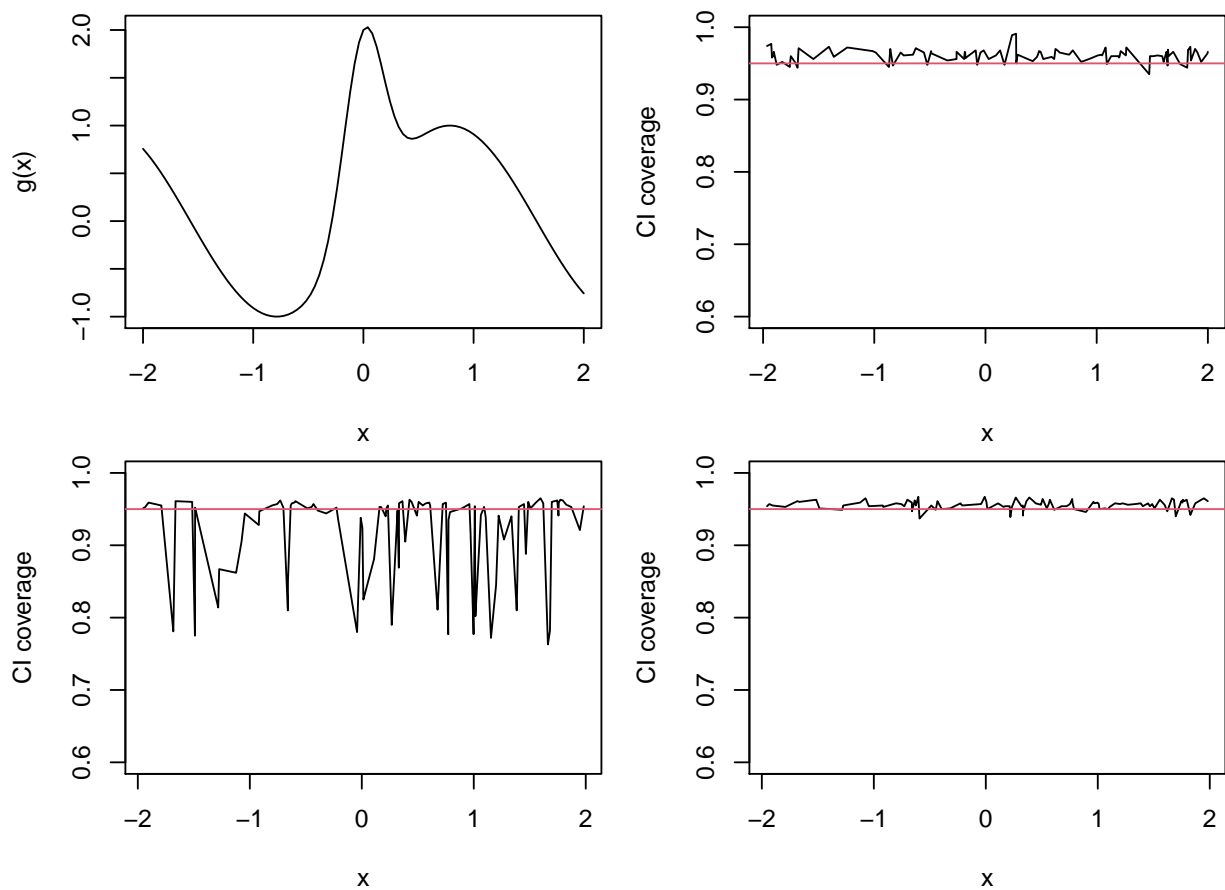


Figure 6: Pointwise Bias assessment tool performance for local linear regression for Example 2 according to the setup in Scenario 1. For reference, the target function is pictured in the top left panel. Confidence interval coverage for the bias $B(x)$ is plotted as a function of x in the top right panel, with the nominal level plotted as a horizontal line in red. The panels on the bottom show confidence interval coverage for the regression function $g(x)$, using the uncorrected estimator (left) and for the bias-corrected estimator (right).

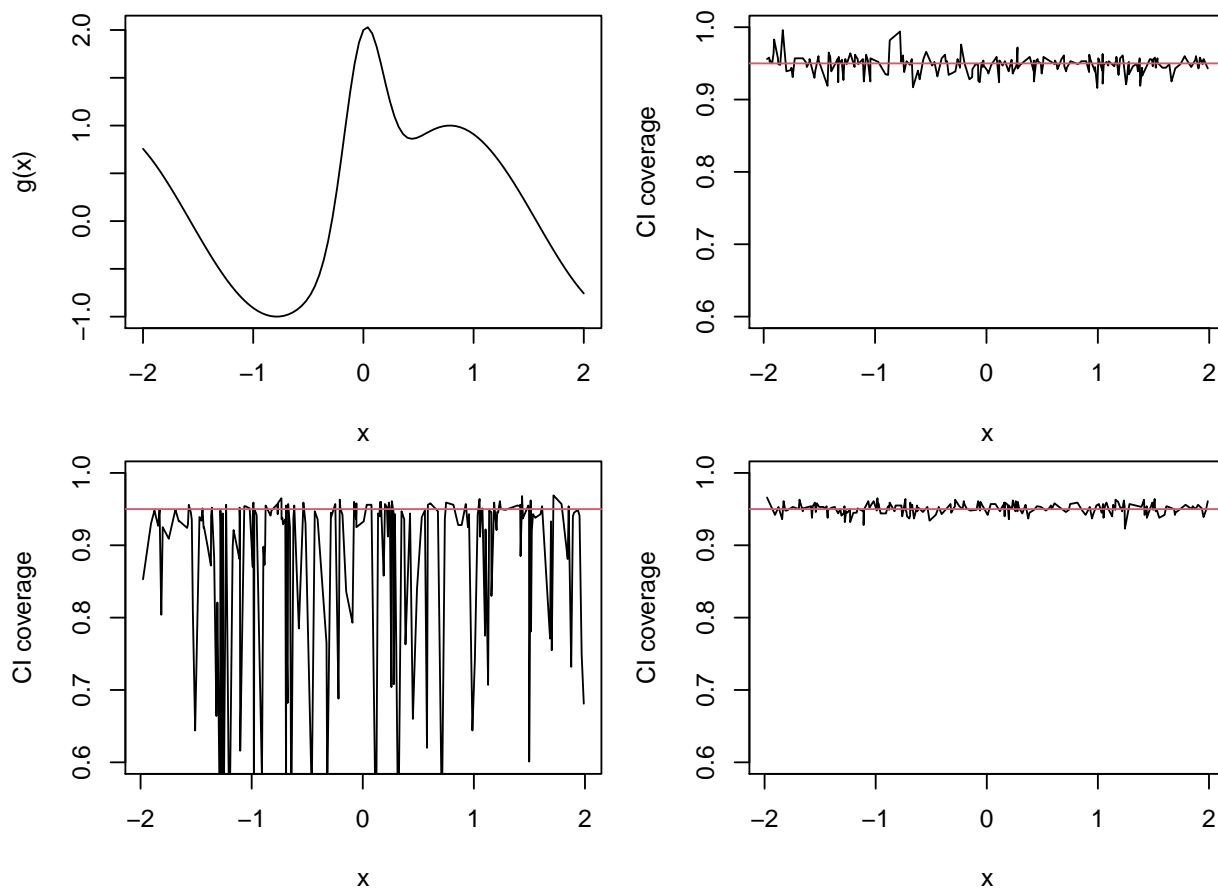


Figure 7: Pointwise Bias assessment tool performance for local linear regression for Example 2 according to the setup in Scenario 2. For reference, the target function is pictured in the top left panel. Confidence interval coverage for the bias $B(x)$ is plotted as a function of x in the top right panel, with the nominal level plotted as a horizontal line in red. The panels on the bottom show confidence interval coverage for the regression function $g(x)$, using the uncorrected estimator (left) and for the bias-corrected estimator (right).

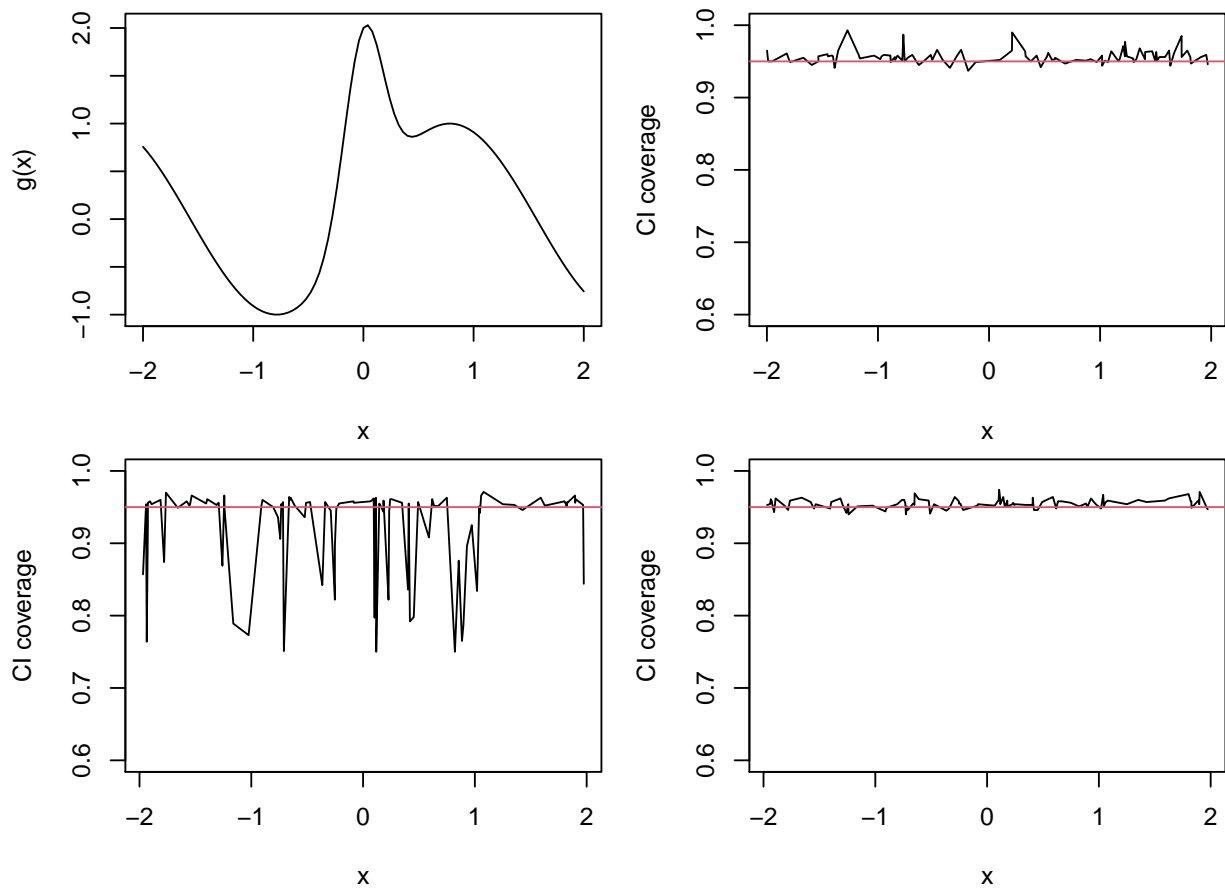


Figure 8: Pointwise Bias assessment tool performance for local linear regression for Example 2 according to the setup in Scenario 3. For reference, the target function is pictured in the top left panel. Confidence interval coverage for the bias $B(x)$ is plotted as a function of x in the top right panel, with the nominal level plotted as a horizontal line in red. The panels on the bottom show confidence interval coverage for the regression function $g(x)$, using the uncorrected estimator (left) and for the bias-corrected estimator (right).

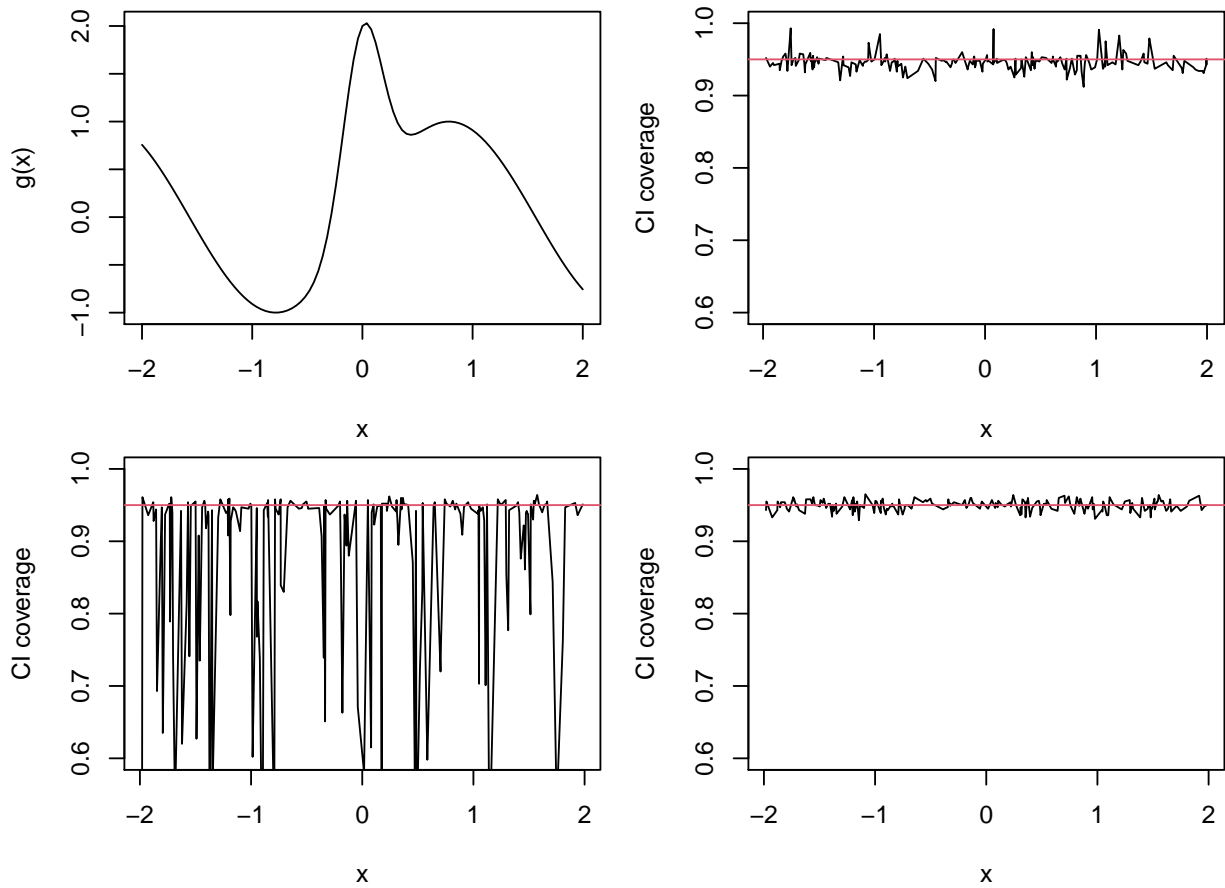


Figure 9: Pointwise Bias assessment tool performance for local linear regression for Example 2 according to the setup in Scenario 4. For reference, the target function is pictured in the top left panel. Confidence interval coverage for the bias $B(x)$ is plotted as a function of x in the top right panel, with the nominal level plotted as a horizontal line in red. The panels on the bottom show confidence interval coverage for the regression function $g(x)$, using the uncorrected estimator (left) and for the bias-corrected estimator (right).

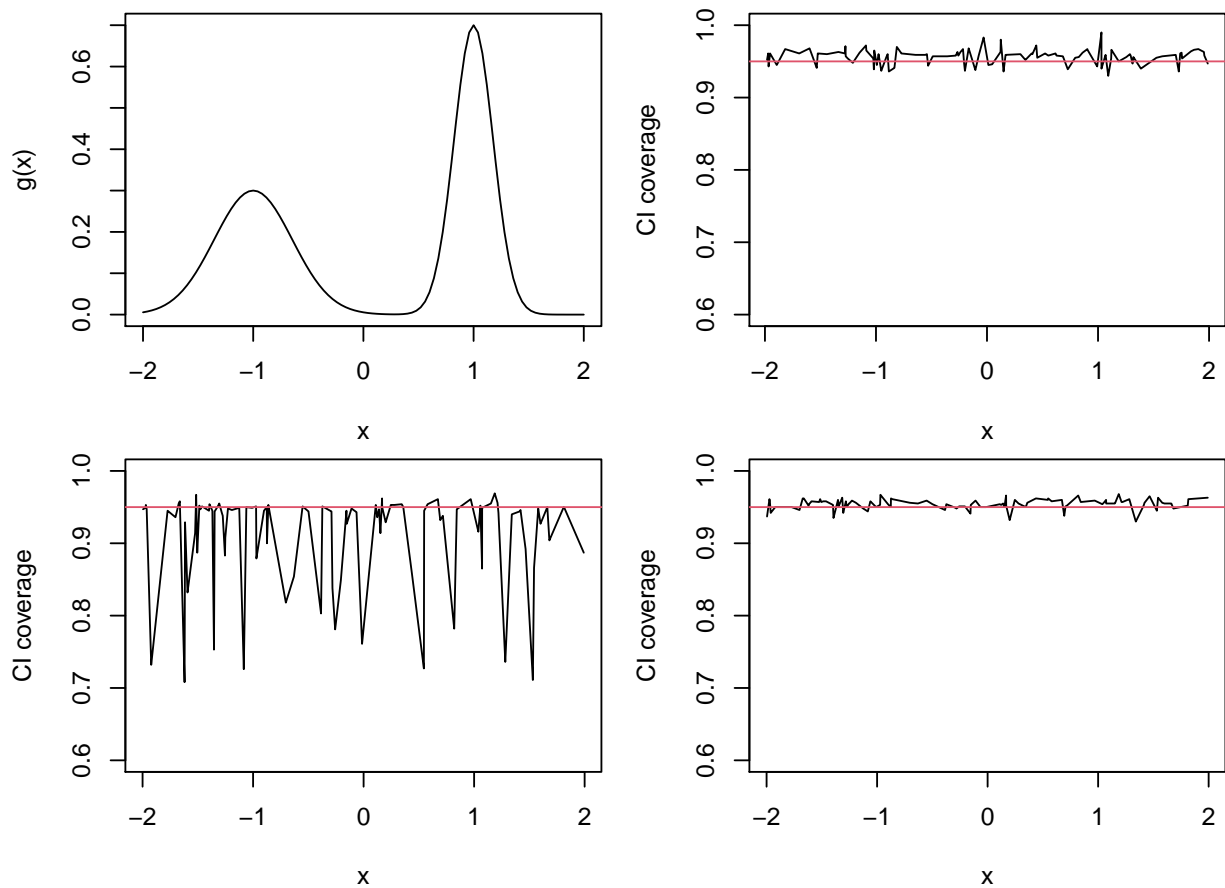


Figure 10: Pointwise Bias assessment tool performance for local linear regression for Example 3 according to the setup in Scenario 1. For reference, the target function is pictured in the top left panel. Confidence interval coverage for the bias $B(x)$ is plotted as a function of x in the top right panel, with the nominal level plotted as a horizontal line in red. The panels on the bottom show confidence interval coverage for the regression function $g(x)$, using the uncorrected estimator (left) and for the bias-corrected estimator (right).

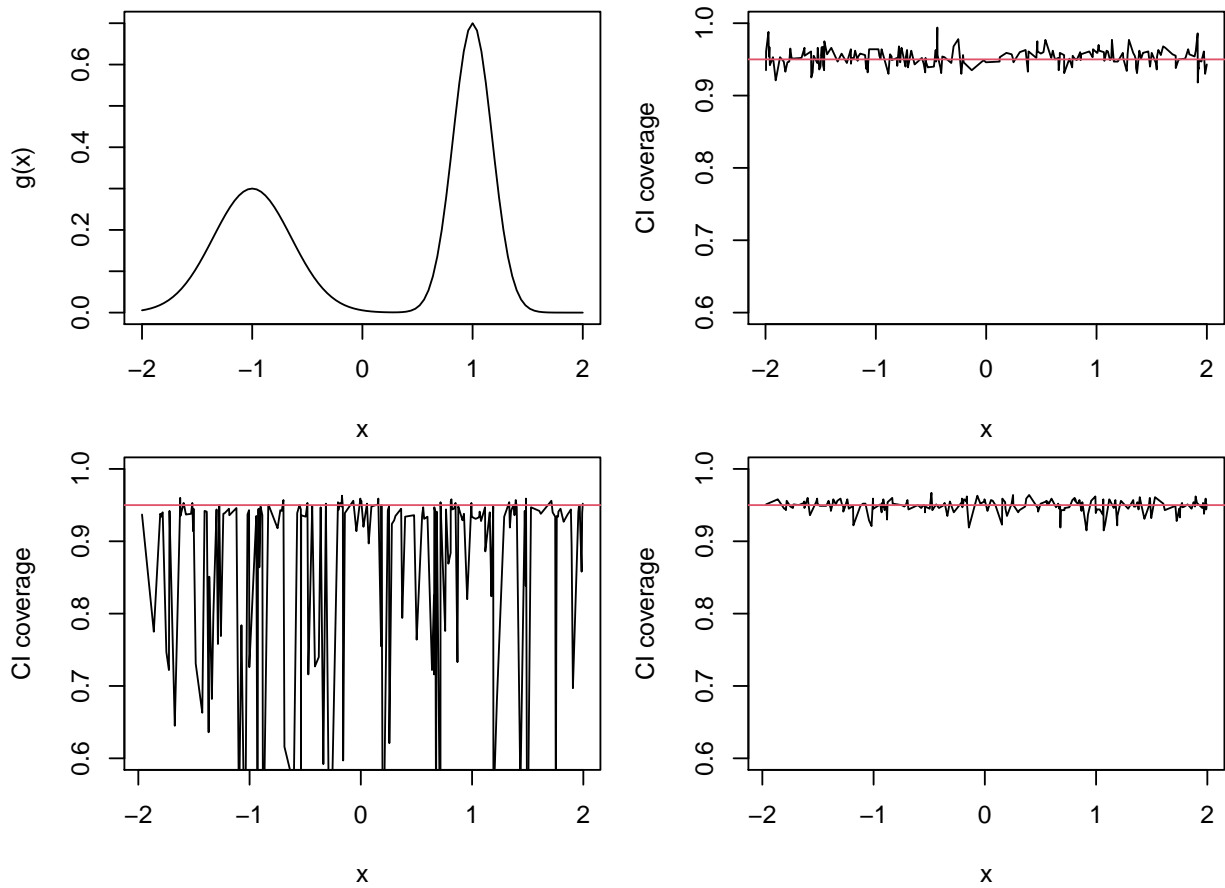


Figure 11: Pointwise Bias assessment tool performance for local linear regression for Example 3 according to the setup in Scenario 2. For reference, the target function is pictured in the top left panel. Confidence interval coverage for the bias $B(x)$ is plotted as a function of x in the top right panel, with the nominal level plotted as a horizontal line in red. The panels on the bottom show confidence interval coverage for the regression function $g(x)$, using the uncorrected estimator (left) and for the bias-corrected estimator (right).

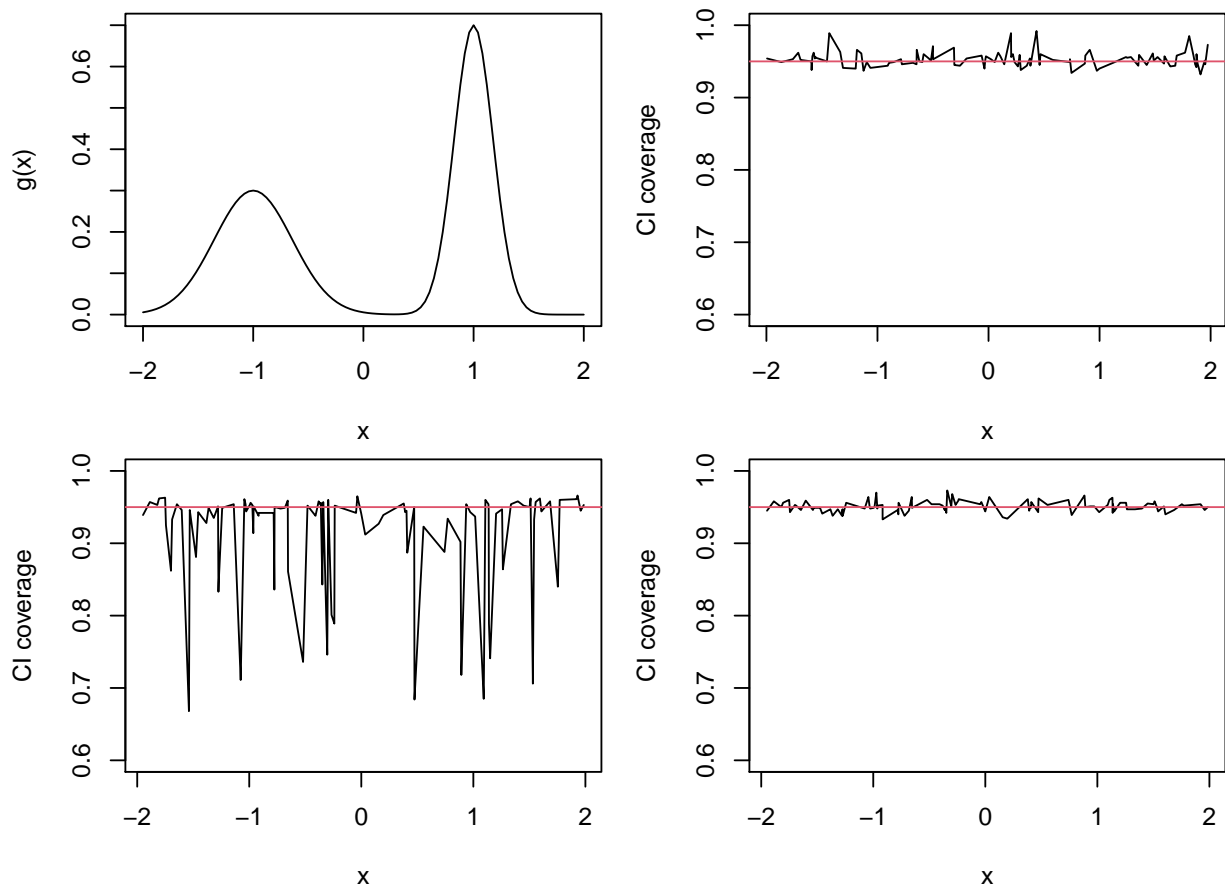


Figure 12: Pointwise Bias assessment tool performance for local linear regression for Example 3 according to the setup in Scenario 3. For reference, the target function is pictured in the top left panel. Confidence interval coverage for the bias $B(x)$ is plotted as a function of x in the top right panel, with the nominal level plotted as a horizontal line in red. The panels on the bottom show confidence interval coverage for the regression function $g(x)$, using the uncorrected estimator (left) and for the bias-corrected estimator (right).

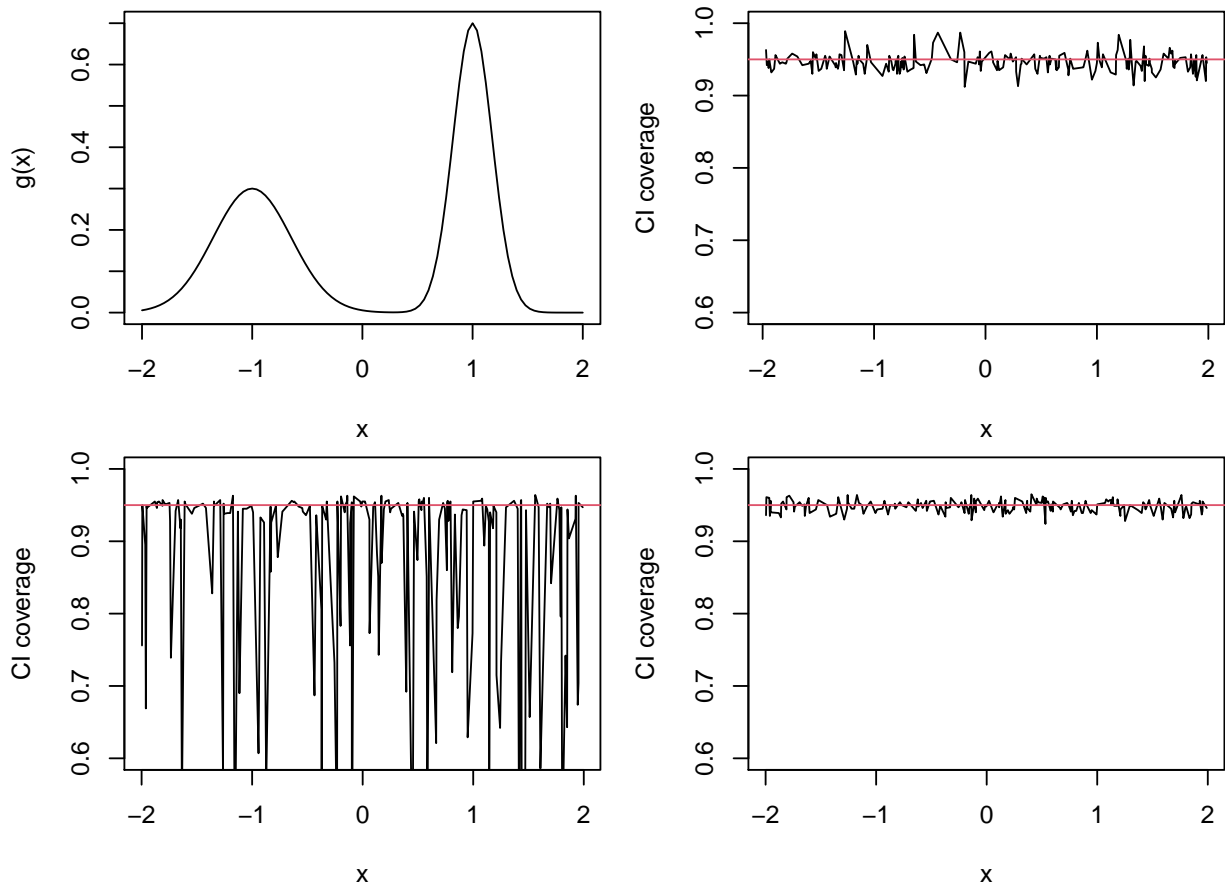


Figure 13: Pointwise Bias assessment tool performance for local linear regression for Example 3 according to the setup in Scenario 4. For reference, the target function is pictured in the top left panel. Confidence interval coverage for the bias $B(x)$ is plotted as a function of x in the top right panel, with the nominal level plotted as a horizontal line in red. The panels on the bottom show confidence interval coverage for the regression function $g(x)$, using the uncorrected estimator (left) and for the bias-corrected estimator (right).