

Partially linear additive hazards regression for clustered and right censored data

Wei Chen · Fengling Ren

Received: date / Accepted: date

Abstract For analyzing clustered survival data, a flexible partially linear additive hazards model is proposed. To accommodate the nonlinear effects, the unknown regression function is approximated by B-splines. All regression coefficients are estimated through a system of pseudo-score functions. Under certain conditions, the proposed estimators are shown to be asymptotically normal, where a consistent estimator of the covariance matrix is given. Simulation studies are also conducted to evaluate the finite sample performance of the proposed method, which is illustrated using a real data set from an AIDS clinical trial.

Keywords B-spline · Clustered data · Marginal model · Partially linear additive hazards · Right censored data

Mathematics Subject Classification (2020) 62G05

1 Simulation results in Example 1

2 Simulation results in Example 2

Wei Chen
School of Zhangjiagang, Jiangsu University of Science and Technology, Zhangjiagang, P.R. China
Tel.: +086-0512-56731140
E-mail: chenweixiyang@sina.com

Fengling Ren
School of Computer Science and Engineering, Xinjiang University of Finance and Economics, Urumqi, P.R. China

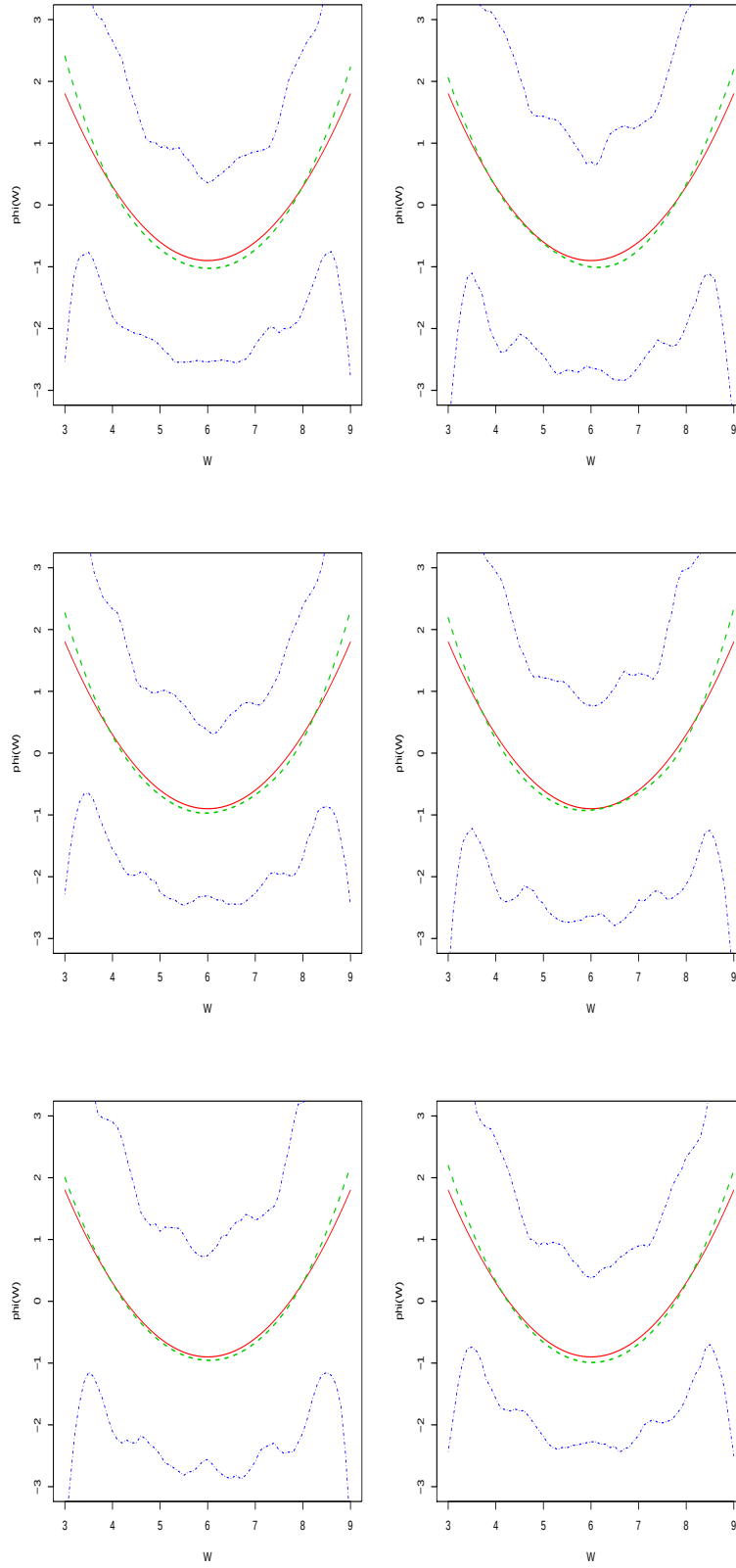


Fig. 1 The solid line (red) is the true curve. The dashed lines (green) are average estimated curves for $n = 100$ with 1000 duplications for Example 1, under $(\theta, CR) = (0, 20\%), (0, 50\%), \dots, (3, 50\%)$ by row, respectively. The dash-dotted lines (blue) are the 95% point-wise confidence intervals

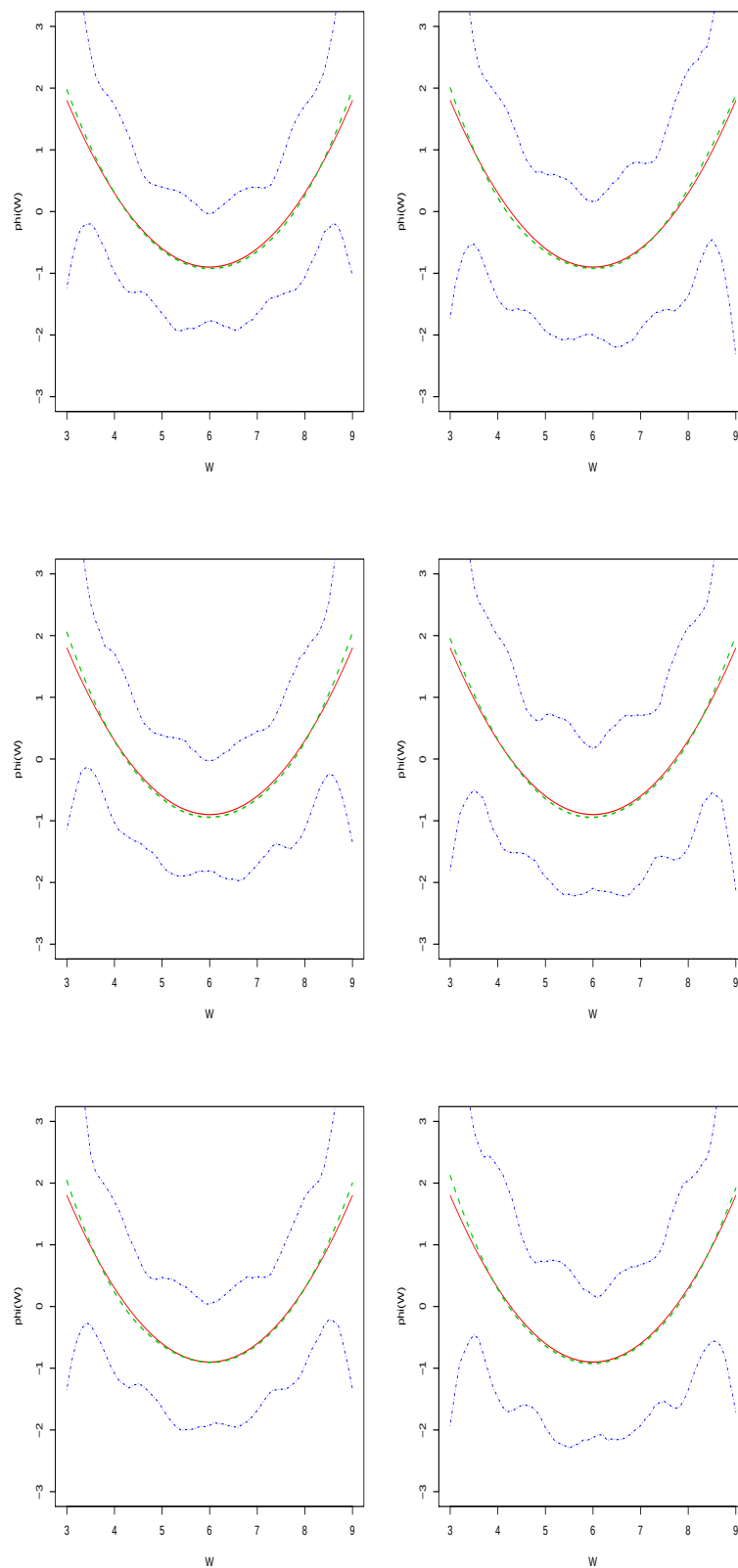


Fig. 2 The solid line (red) is the true curve. The dashed lines (green) are average estimated curves for $n = 200$ with 1000 duplications for Example 1, under $(\theta, CR) = (0, 20\%), (0, 50\%), \dots, (3, 50\%)$ by row, respectively. The dash-dotted lines (blue) are the 95% point-wise confidence intervals

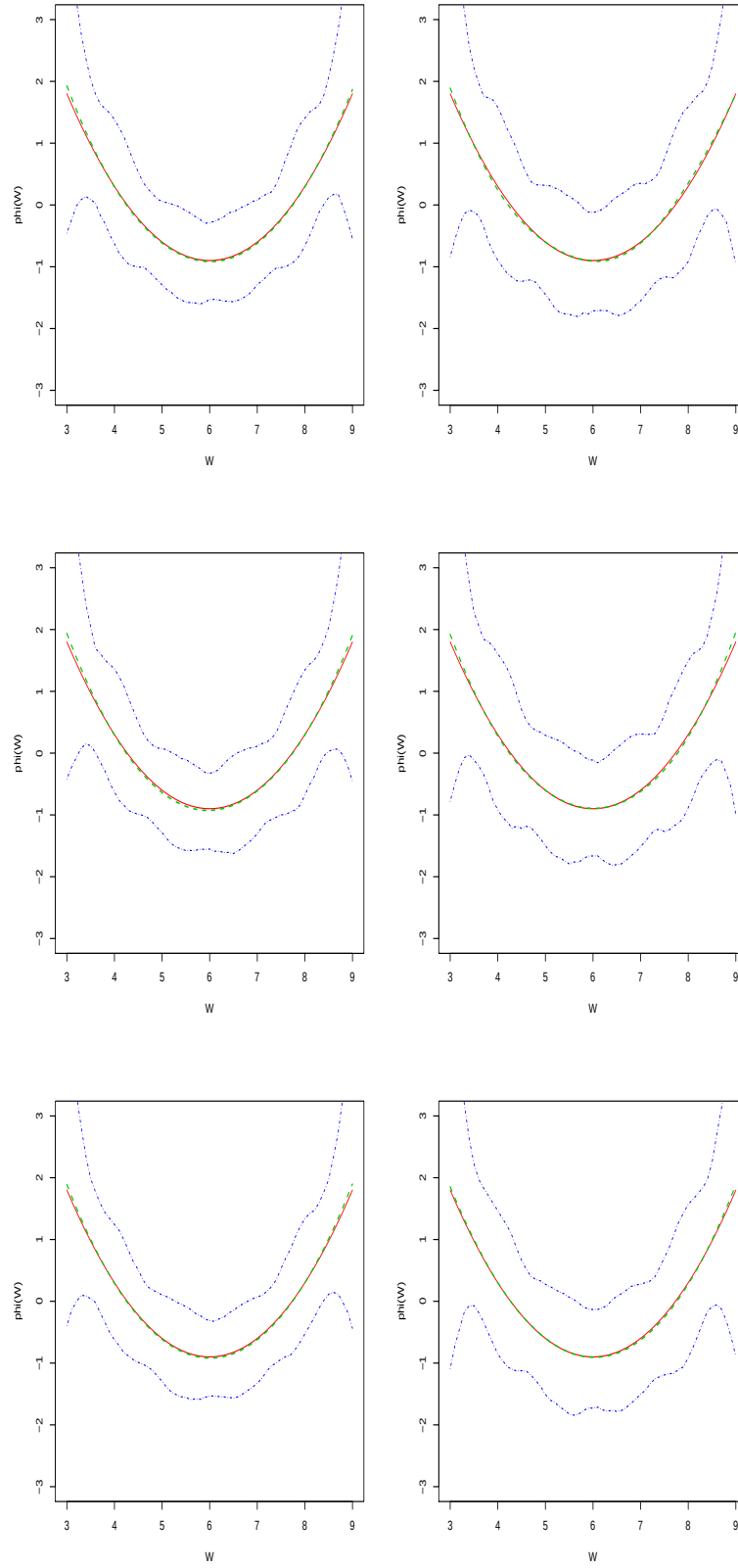


Fig. 3 The solid line (red) is the true curve. The dashed lines (green) are average estimated curves for $n = 400$ with 1000 duplications for Example 1, under $(\theta, CR) = (0, 20\%), (0, 50\%), \dots, (3, 50\%)$ by row, respectively. The dash-dotted lines (blue) are the 95% point-wise confidence intervals

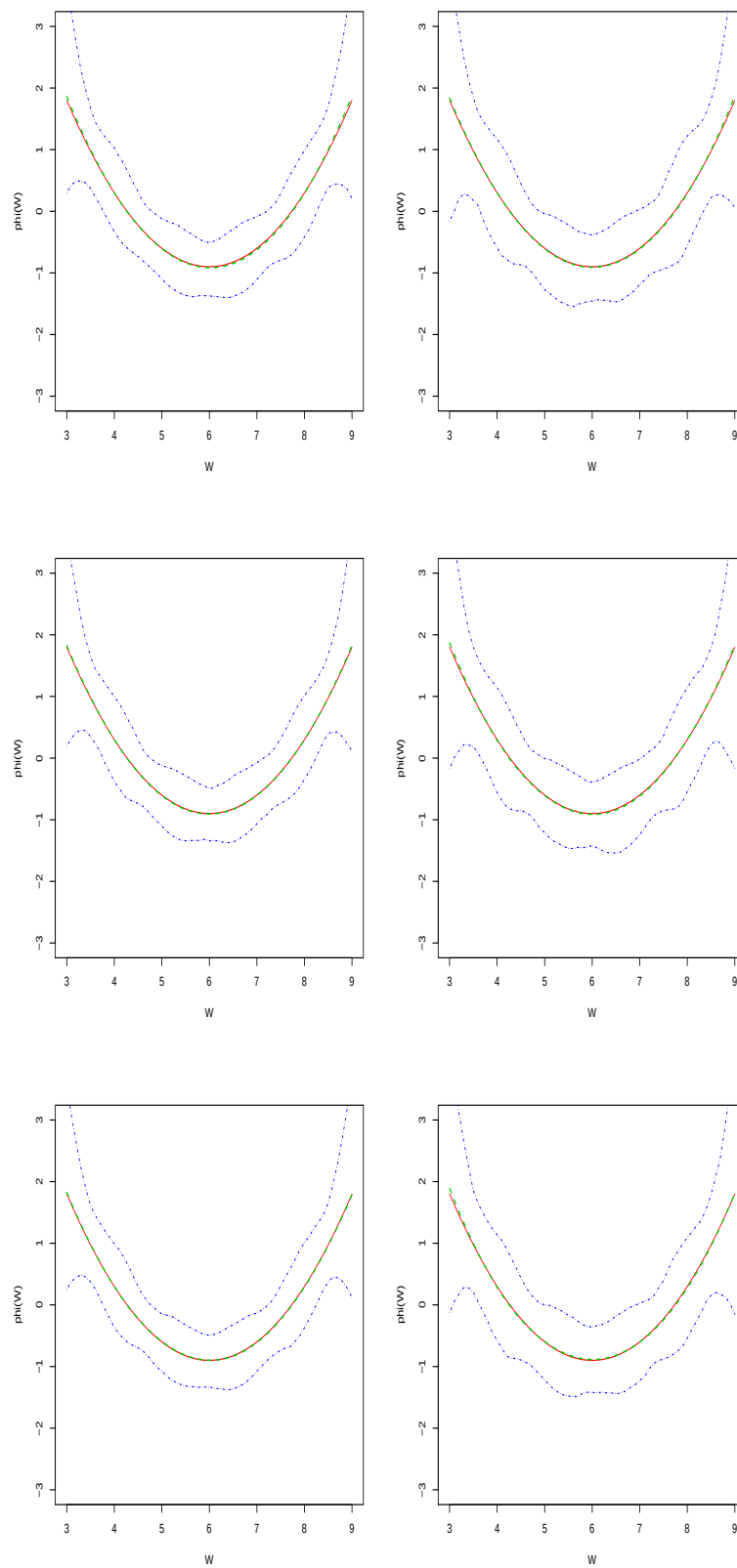


Fig. 4 The solid line (red) is the true curve. The dashed lines (green) are average estimated curves for $n = 800$ with 1000 duplications for Example 1, under $(\theta, CR) = (0, 20\%), (0, 50\%), \dots, (3, 50\%)$ by row, respectively. The dash-dotted lines (blue) are the 95% point-wise confidence intervals

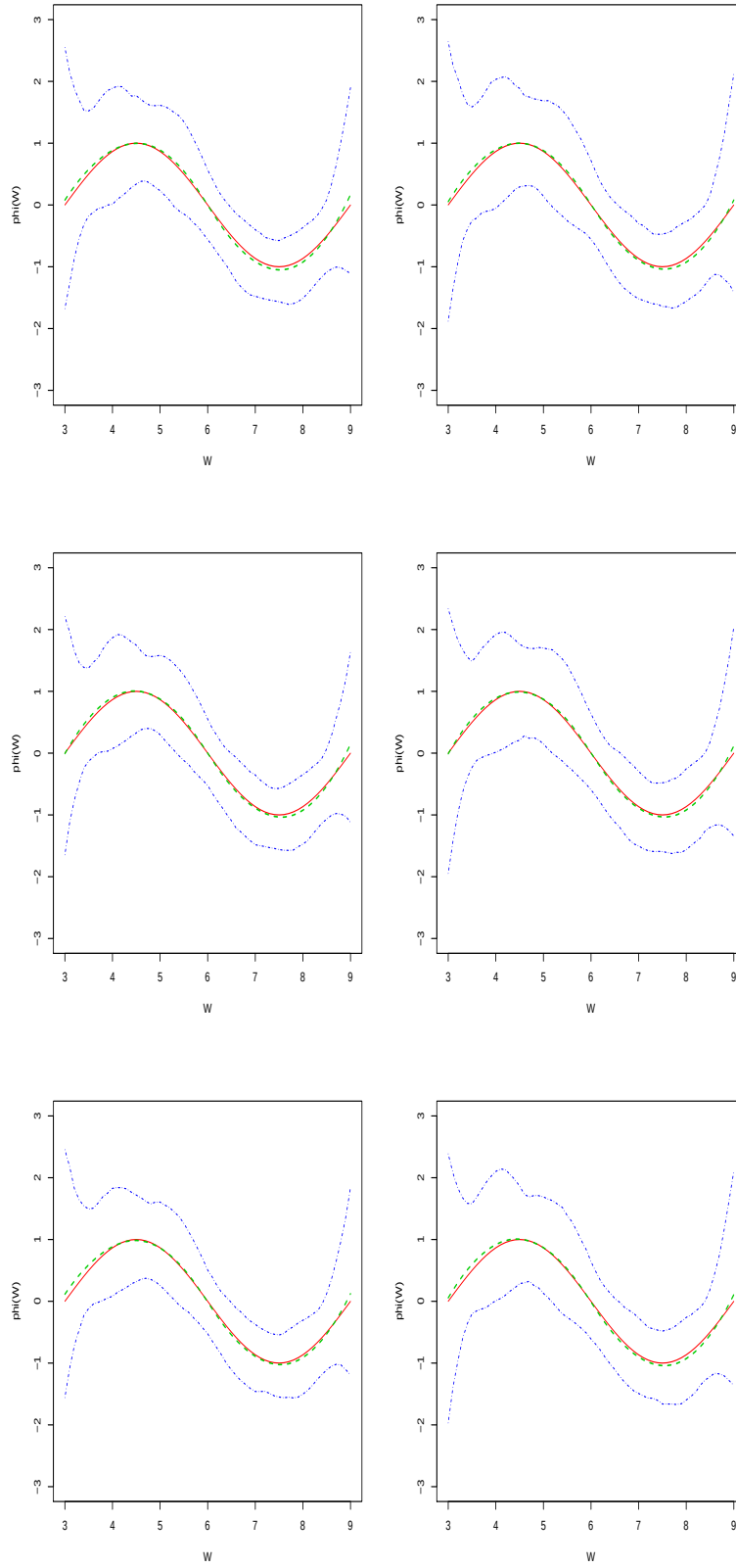


Fig. 5 The solid line (red) is the true curve. The dashed lines (green) are average estimated curves for $n = 100$ with 1000 duplications for Example 2, under $(\theta, CR) = (0, 20\%), (0, 50\%), \dots, (3, 50\%)$ by row, respectively. The dash-dotted lines (blue) are the 95% point-wise confidence intervals

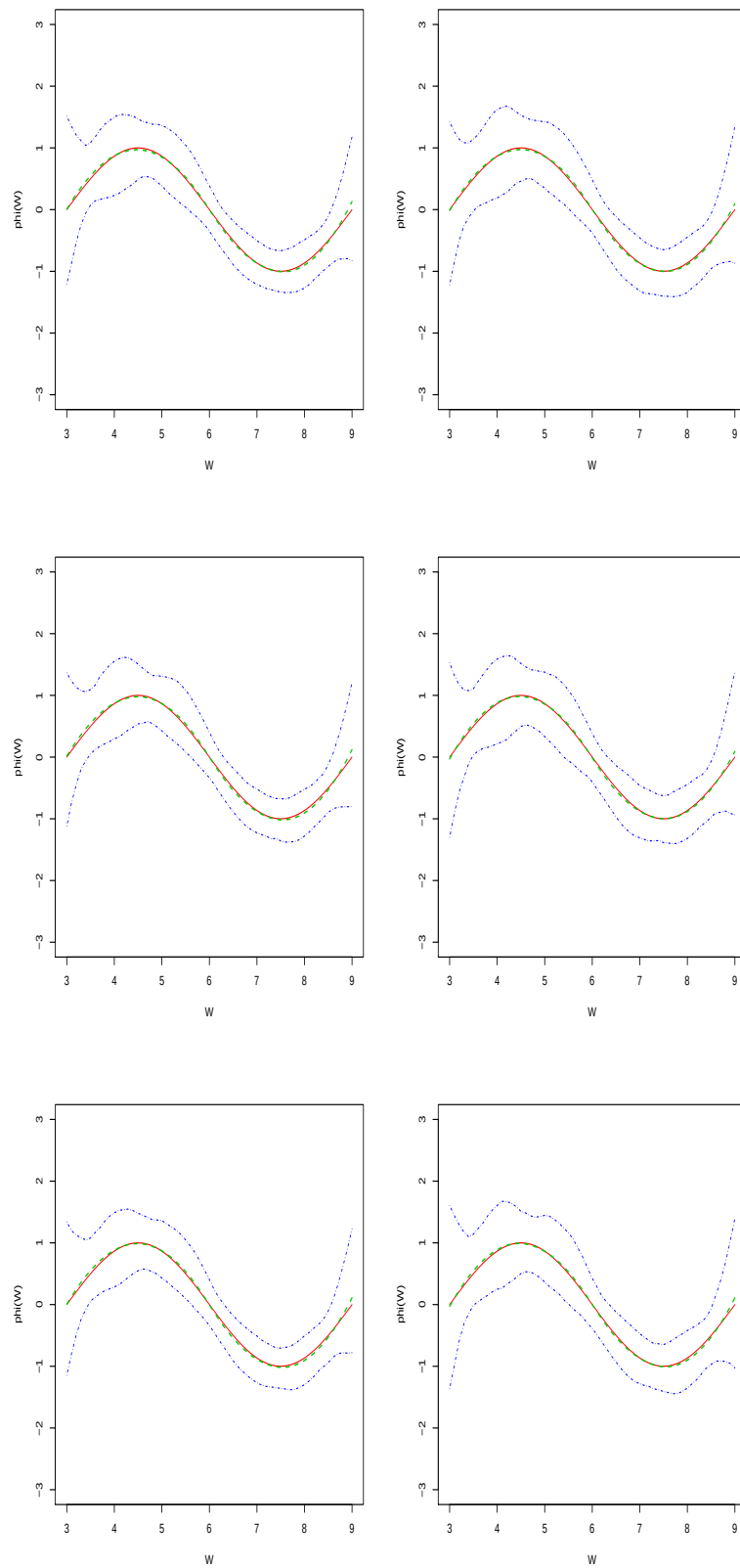


Fig. 6 The solid line (red) is the true curve. The dashed lines (green) are average estimated curves for $n = 200$ with 1000 duplications for Example 2, under $(\theta, CR) = (0, 20\%), (0, 50\%), \dots, (3, 50\%)$ by row, respectively. The dash-dotted lines (blue) are the 95% point-wise confidence intervals

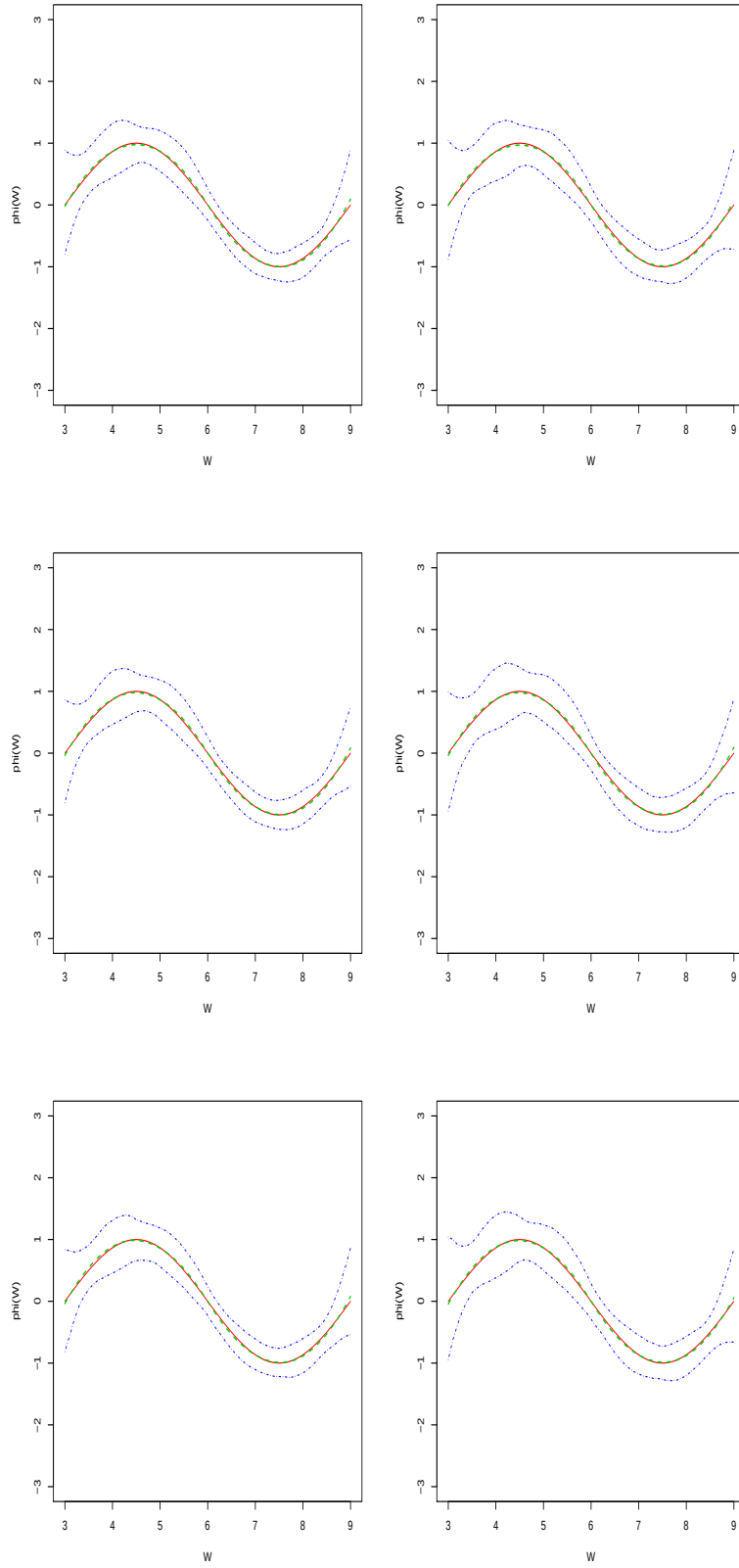


Fig. 7 The solid line (red) is the true curve. The dashed lines (green) are average estimated curves for $n = 400$ with 1000 duplications for Example 2, under $(\theta, CR) = (0, 20\%), (0, 50\%), \dots, (3, 50\%)$ by row, respectively. The dash-dotted lines (blue) are the 95% point-wise confidence intervals

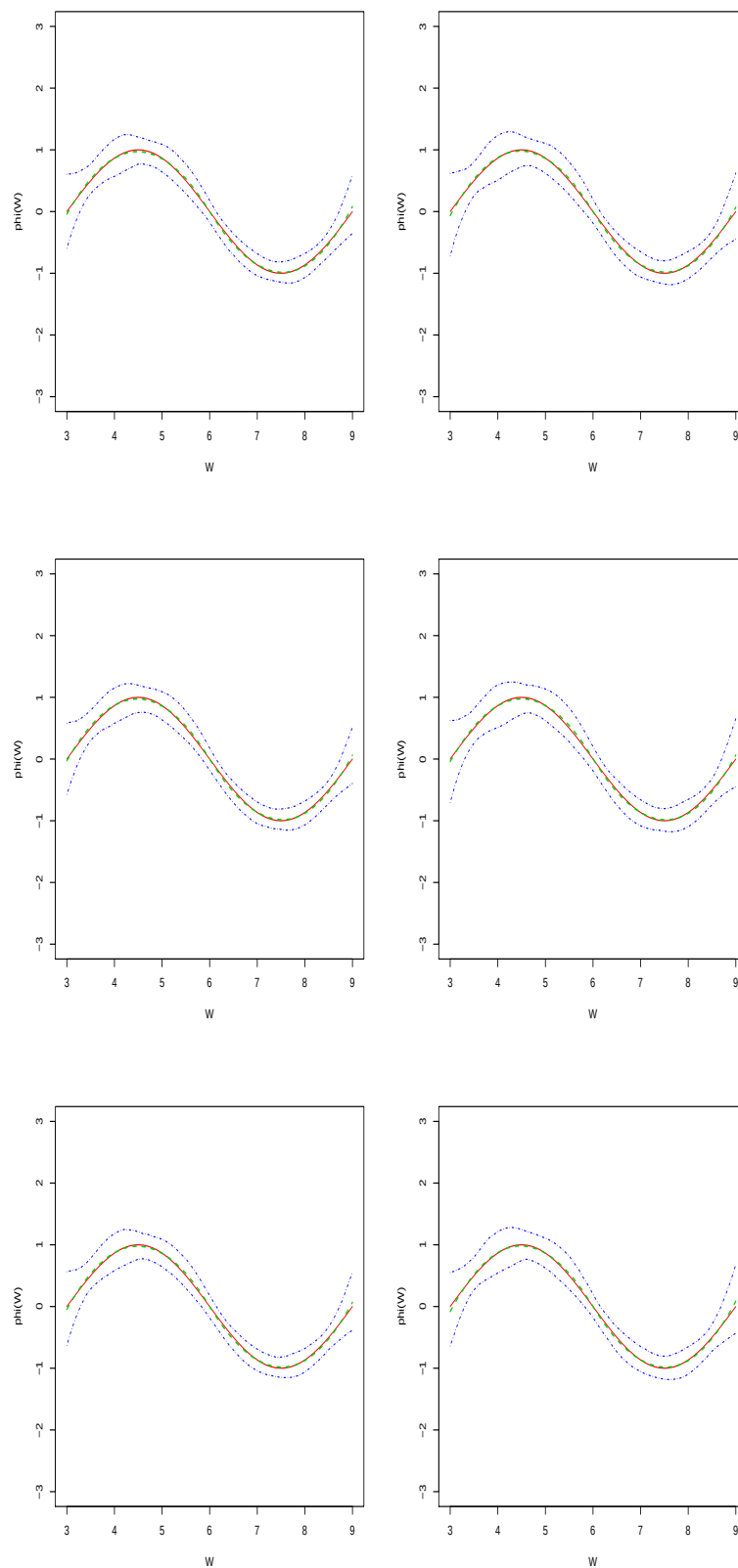


Fig. 8 The solid line (red) is the true curve. The dashed lines (green) are average estimated curves for $n = 800$ with 1000 duplications for Example 2, under $(\theta, CR) = (0, 20\%), (0, 50\%), \dots, (3, 50\%)$ by row, respectively. The dash-dotted lines (blue) are the 95% point-wise confidence intervals