A Semiparametric Approach for Analyzing Nonignorable Missing Data

Hui Xie [1], Yi Qian, [2] and Leming Qu [3]

Supplementary Material

S1 An Extended Analysis of the Wage Offer Dataset

The ISNI analysis conducted in Section 6 of the main article assumes that a logit transformation of the probability of missingness depends on *lwage* in a linear form. Assuming such a linear form for a continuous outcome in sensitivity analysis for nonignorable missingness is prevalent in the literature. In practice, however, there may exist important cases where, after controlling for the observed predictors for missingness, both unusually small and large ideal outcomes lead to similar increased or decreased probabilities of missingness. In such cases, a quadratic relation in the outcome y_i would be of relevance. Although this is unlikely to be the case for our wage offer dataset, we will use the dataset to illustrate the application of the ISNI method with a quadratic functional form. That is, we assume a quadratic functional form for $\eta_{\gamma_1}(y_i) = \gamma_{11} y_i + \gamma_{12} y_i^2 = \gamma_{11}(r_1 y_i + r_2 y_i^2)$, where y is used to denote *lwage* for notational simplicity, $r_1 = 1$, $r_2 = \frac{\gamma_{12}}{\gamma_{11}}$, and r_2 is a user-specified value to conduct sensitivity analysis for a specific form. For a given value of r_2, the ISNIs for the regression parameters are shown in the Appendix of the main article to be

$$\text{ISNI}_r \quad = \quad -\hat{\sigma}^2 \Big(\sum_{i:g_i=1} x_i x_i^T\Big)^{-1} \sum_{i:g_i=0} (1 + 2r_2 \hat{\mu}_i) x_i h_i.$$

where $\hat{\mu}_i = x_i^T \hat{\beta}(0)$. The column "ISNI$_r$" under "GAM+Quadratic $\eta_{\gamma_1}(y)$" in Table 3 reports the ISNI values when r_2 is set to be $-\frac{1}{2\mu_0}$ and μ_0 is the mean of the outcome in the sample where the missing outcome is replaced with the predicted value under the MAR assumption.

One needs to take care in interpreting these ISNI values based on the quadratic function for $\eta_{\gamma_1}(y)$ because the mechanism of nonignorable missingness becomes more complicated than in the linear case. To help the interpretation, the column "MAR Est +ISNI$_r * \mu_0/\sigma^2$" gives $\hat{\theta}(\gamma_{11})$ when $\gamma_{11} = \mu_0/\sigma^2$. This corresponds to a scenario

[1] Department of Epidemiology and Biostatistics, University of Illinois, Chicago, IL 60612. Email: huixie@uic.edu.

[2] Kellogg School of Management, Northwestern University, 2001 Sheridan Rd, Evanston, IL 60208. Email: yiqian@northwestern.edu.

[3] Department of Mathematics, Boise State University, 1910 University Dr., Boise ID 83725. Email: lqu@boisestate.edu.

that $\eta_{\gamma_1}(y_i) = -\frac{1}{2}(\frac{y_i-\mu_0}{\sigma})^2$. The form of $\eta_{\gamma_1}(y_i)$ implies that when y is 2-σ below the μ_0, the odds of being observed is about e^{-2} given the same values of the other predictors for missingness. Therefore, when Y is between μ_0 and $\mu_0 - 2\sigma$, the magnitude of nonignorability is approximately the same size as when $\gamma_1 = 1/\sigma$ in the sensitivity analysis conducted in the main article that assumes $\eta_{\gamma_1}(y) = \gamma_1 y$. When y is further below $\mu_0 - 2\sigma$, the magnitude of nonignorability is larger than in the linear case, which tends to cause larger parameter changes (i.e., larger sensitivity). On the other hand, the assumed quadratic function of $\eta_{\gamma_1}(y)$ implies that when Y is unusually large, the probability of being observed tends to be more similar to those with unusually small Y than if one assumes a linear function of $\eta_{\gamma_1}(y)$. This tends to cause smaller parameter changes (i.e., smaller sensitivity). Therefore, it is hard to predict how the sensitivity changes, as one moves from the linear $\eta_{\gamma_1}(y)$ to the quadratic $\eta_{\gamma_1}(y)$.

When we compare the values listed in the ninth and the last columns of Table 3 with those listed in the second column of the same table, we find that the parameter changes from the MAR estimates under the above quadratic function are either similar to or smaller than those from the linear case. In particular, the parameter changes are substantially smaller in the quadratic case than in the linear case for the last three parameters in Table 3. Therefore, it seems that in the wage offer dataset, the above quadratic case for nonignorability tends to give a smaller assessment of sensitivity than the linear case. In the wage offer dataset, it is unlikely that the relationship between acceptance of the job and the wage offered by the job would be quadratic. Thus, we believe that the ISNI analysis based on the linear function for $\eta_{\gamma_1}(y)$ is more plausible. Nonetheless, the analysis illustrates a potential use of ISNI with a quadratic function for $\eta_{\gamma_1}(y)$.

Table 1: Simulation study of the performance of four estimates for case 1.
$\hat{\beta}_1^0 = \hat{\beta}_1(0)$ is the MAR estimate of β_1; $\hat{\beta}_{1L}^{\gamma_1} = \hat{\beta}_{1L}(\gamma_1) = \hat{\beta}_1(0) + ISNI_L \times \gamma_1$; $\hat{\beta}_{1P}^{\gamma_1} = \hat{\beta}_{1P}(\gamma_1) = \hat{\beta}_1(0) + ISNI_P \times \gamma_1$; $\hat{\beta}_{1G}^{\gamma_1} = \hat{\beta}_{1G}(\gamma_1) = \hat{\beta}_1(0) + ISNI_G \times \gamma_1$.

	Linear				Quadratic				Cubic				Sine			
	$\hat{\beta}_1^0$	$\hat{\beta}_{1L}^{\gamma_1}$	$\hat{\beta}_{1P}^{\gamma_1}$	$\hat{\beta}_{1G}^{\gamma_1}$	$\hat{\beta}_1^0$	$\hat{\beta}_{1L}^{\gamma_1}$	$\hat{\beta}_{1P}^{\gamma_1}$	$\hat{\beta}_{1G}^{\gamma_1}$	$\hat{\beta}_1^0$	$\hat{\beta}_{1L}^{\gamma_1}$	$\hat{\beta}_{1P}^{\gamma_1}$	$\hat{\beta}_{1G}^{\gamma_1}$	$\hat{\beta}_1^0$	$\hat{\beta}_{1L}^{\gamma_1}$	$\hat{\beta}_{1P}^{\gamma_1}$	$\hat{\beta}_{1G}^{\gamma_1}$
$\beta = \rho = 0$																
$\sqrt{MSE}$ ($\times 10$)																
$\gamma_1 = -1$	1.33	0.56	0.56	0.56	0.69	0.46	0.46	0.46	1.76	0.60	0.60	0.59	0.90	0.71	0.70	0.65
$\gamma_1 = -0.5$	0.96	0.59	0.59	0.59	0.56	0.48	0.48	0.48	1.12	0.61	0.61	0.61	0.76	0.67	0.67	0.65
$\gamma_1 = -0.1$	0.64	0.63	0.63	0.63	0.49	0.49	0.49	0.49	0.67	0.66	0.66	0.65	0.67	0.67	0.67	0.66
$\gamma_1 = 0.1$	0.58	0.57	0.57	0.57	0.46	0.46	0.46	0.46	0.64	0.63	0.63	0.63	0.60	0.61	0.61	0.60
$\gamma_1 = 0.5$	0.96	0.53	0.53	0.53	0.52	0.45	0.45	0.45	1.11	0.62	0.62	0.62	0.79	0.69	0.68	0.67
$\gamma_1 = 1$	1.36	0.59	0.59	0.59	0.68	0.50	0.49	0.49	1.75	0.65	0.65	0.65	0.87	0.74	0.73	0.67
$\lvert\text{Bias}\rvert \times 10$																
$\gamma_1 = -1$	1.22	0.12	0.12	0.11	0.52	0.07	0.04	0.04	1.67	0.17	0.17	0.12	0.64	0.23	0.21	0.04
$\gamma_1 = -0.5$	0.76	0.00	0.00	0.00	0.29	0.00	0.01	0.00	0.94	0.02	0.02	0.01	0.38	0.12	0.10	0.01
$\gamma_1 = -0.1$	0.11	0.05	0.05	0.05	0.03	0.03	0.03	0.03	0.13	0.07	0.06	0.06	0.07	0.04	0.03	0.01
$\gamma_1 = 0.1$	0.11	0.06	0.06	0.06	0.00	0.06	0.06	0.06	0.16	0.04	0.04	0.03	0.03	0.08	0.08	0.05
$\gamma_1 = 0.5$	0.80	0.02	0.02	0.02	0.27	0.02	0.01	0.00	0.93	0.04	0.04	0.00	0.40	0.11	0.09	0.00
$\gamma_1 = 1$	1.25	0.10	0.10	0.09	0.50	0.09	0.06	0.06	1.65	0.18	0.17	0.13	0.59	0.27	0.25	0.09
SD$\times$10																
$\gamma_1 = -1$	0.53	0.55	0.55	0.55	0.46	0.46	0.46	0.46	0.56	0.58	0.58	0.58	0.63	0.67	0.66	0.66
$\gamma_1 = -0.5$	0.58	0.59	0.59	0.59	0.48	0.48	0.48	0.48	0.61	0.61	0.61	0.61	0.67	0.66	0.66	0.66
$\gamma_1 = -0.1$	0.63	0.63	0.63	0.63	0.49	0.49	0.49	0.49	0.66	0.66	0.66	0.66	0.67	0.67	0.67	0.67
$\gamma_1 = 0.1$	0.57	0.57	0.57	0.57	0.46	0.46	0.46	0.46	0.63	0.63	0.63	0.63	0.60	0.60	0.60	0.60
$\gamma_1 = 0.5$	0.53	0.53	0.53	0.53	0.45	0.45	0.45	0.45	0.61	0.62	0.62	0.62	0.68	0.68	0.68	0.67
$\gamma_1 = 1$	0.55	0.58	0.58	0.58	0.47	0.49	0.49	0.49	0.59	0.63	0.63	0.63	0.64	0.68	0.68	0.66

Table 2: Simulation study of the performance of four estimates for case 2.
$\hat\beta_1^0 = \hat\beta_1(0)$ is the MAR estimate of β_1; $\hat\beta_{1L}^{\gamma_1} = \hat\beta_{1L}(\gamma_1) = \hat\beta_1(0) + ISNI_L \times \gamma_1$; $\hat\beta_{1P}^{\gamma_1} = \hat\beta_{1P}(\gamma_1) = \hat\beta_1(0) + ISNI_P \times \gamma_1$; $\hat\beta_{1G}^{\gamma_1} = \hat\beta_{1G}(\gamma_1) = \hat\beta_1(0) + ISNI_G \times \gamma_1$.

	Linear				Quadratic				Cubic				Sine			
	$\hat\beta_1^0$	$\hat\beta_{1L}^{\gamma_1}$	$\hat\beta_{1P}^{\gamma_1}$	$\hat\beta_{1G}^{\gamma_1}$	$\hat\beta_1^0$	$\hat\beta_{1L}^{\gamma_1}$	$\hat\beta_{1P}^{\gamma_1}$	$\hat\beta_{1G}^{\gamma_1}$	$\hat\beta_1^0$	$\hat\beta_{1L}^{\gamma_1}$	$\hat\beta_{1P}^{\gamma_1}$	$\hat\beta_{1G}^{\gamma_1}$	$\hat\beta_1^0$	$\hat\beta_{1L}^{\gamma_1}$	$\hat\beta_{1P}^{\gamma_1}$	$\hat\beta_{1G}^{\gamma_1}$
$\beta = \rho = 0$																
$\sqrt{MSE}$ $(\times 10)$																
$\gamma_1 = -1$	1.39	0.57	0.57	0.57	1.71	1.22	1.04	1.04	2.03	1.69	1.24	1.21	0.91	0.69	0.68	0.63
$\gamma_1 = -0.5$	0.90	0.55	0.55	0.55	1.25	1.11	1.03	1.03	1.49	1.33	1.15	1.15	0.76	0.69	0.69	0.66
$\gamma_1 = -0.1$	0.64	0.63	0.63	0.63	1.05	1.03	1.03	1.03	1.19	1.17	1.17	1.17	0.72	0.71	0.71	0.70
$\gamma_1 = 0.1$	0.60	0.59	0.59	0.59	1.07	1.06	1.06	1.06	1.27	1.28	1.27	1.27	0.66	0.65	0.65	0.65
$\gamma_1 = 0.5$	0.97	0.61	0.61	0.61	1.25	1.08	0.99	0.99	1.58	1.44	1.26	1.26	0.76	0.68	0.67	0.65
$\gamma_1 = 1$	1.37	0.60	0.60	0.60	1.75	1.11	0.98	0.99	2.02	1.77	1.24	1.23	0.95	0.72	0.71	0.67
$\lvert\text{Bias}\rvert \times 10$																
$\gamma_1 = -1$	1.28	0.06	0.06	0.05	1.37	0.60	0.05	0.03	1.68	1.18	0.06	0.03	0.68	0.21	0.22	0.05
$\gamma_1 = -0.5$	0.72	0.05	0.04	0.04	0.74	0.42	0.05	0.05	0.97	0.65	0.02	0.01	0.37	0.14	0.13	0.03
$\gamma_1 = -0.1$	0.16	0.00	0.00	0.00	0.24	0.01	0.07	0.07	0.22	0.12	0.02	0.01	0.14	0.03	0.03	0.04
$\gamma_1 = 0.1$	0.16	0.00	0.00	0.00	0.15	0.09	0.01	0.01	0.11	0.22	0.08	0.09	0.09	0.02	0.02	0.00
$\gamma_1 = 0.5$	0.76	0.01	0.01	0.01	0.78	0.39	0.01	0.01	0.96	0.69	0.00	0.03	0.38	0.14	0.13	0.02
$\gamma_1 = 1$	1.25	0.08	0.08	0.07	1.48	0.50	0.14	0.13	1.62	1.25	0.01	0.05	0.67	0.19	0.18	0.03
SD$\times$10																
$\gamma_1 = -1$	0.54	0.56	0.56	0.57	1.02	1.06	1.04	1.04	1.14	1.21	1.24	1.21	0.62	0.66	0.65	0.63
$\gamma_1 = -0.5$	0.54	0.55	0.55	0.55	1.01	1.03	1.02	1.03	1.14	1.16	1.16	1.15	0.67	0.68	0.67	0.67
$\gamma_1 = -0.1$	0.63	0.63	0.63	0.63	1.03	1.03	1.03	1.03	1.17	1.17	1.17	1.17	0.71	0.71	0.71	0.71
$\gamma_1 = 0.1$	0.58	0.59	0.59	0.59	1.06	1.06	1.06	1.06	1.26	1.26	1.26	1.26	0.65	0.65	0.65	0.65
$\gamma_1 = 0.5$	0.60	0.61	0.61	0.61	0.98	1.01	0.99	0.99	1.26	1.27	1.26	1.26	0.65	0.67	0.66	0.65
$\gamma_1 = 1$	0.56	0.60	0.60	0.60	0.94	0.99	0.98	0.98	1.20	1.26	1.24	1.23	0.67	0.70	0.69	0.68

Table 3: ISNIs for Parameter Estimates in the Wage Offer Dataset.

	MAR		Linear Logistic			GAM + Linear $\eta_{\gamma_1}(y)$			GAM + Quadratic $\eta_{\gamma_1}(y)$	
Predictor	Est.	S.E.	ISNI	c	MAR Est. +ISNI/σ	ISNI	c	MAR Est. +ISNI/σ	ISNI$_r$	MAR Est. +ISNI$_r * \mu_0/\sigma^2$
Intercept	-0.36	0.32	-0.26	0.87	-0.72	-0.19	1.20	-0.62	-0.18	-0.75
educ	0.10	0.015	0.019	0.57	0.126	0.018	0.61	0.124	0.010	0.120
exper	0.04	0.013	0.024	0.40	0.073	0.016	0.57	0.064	0.006	0.054
expersq	-0.00075	0.00039	-0.00050	0.60	-0.0015	-0.00022	1.29	-0.0010	-0.00012	-0.0010
nwifeinc	0.0057	0.0033	-0.0035	0.67	0.00079	-0.0017	1.43	0.0034	0.0013	0.0084
kidslt6	-0.056	0.088	-0.12	0.53	-0.221	-0.12	0.53	-0.220	0.0078	-0.039
kidsge6	-0.018	0.028	0.007	2.76	-0.0077	0.004	4.92	-0.012	-0.0013	-0.020
age	-0.0035	0.005	-0.006	0.67	-0.011	-0.007	0.57	-0.013	-0.0005	-0.0046

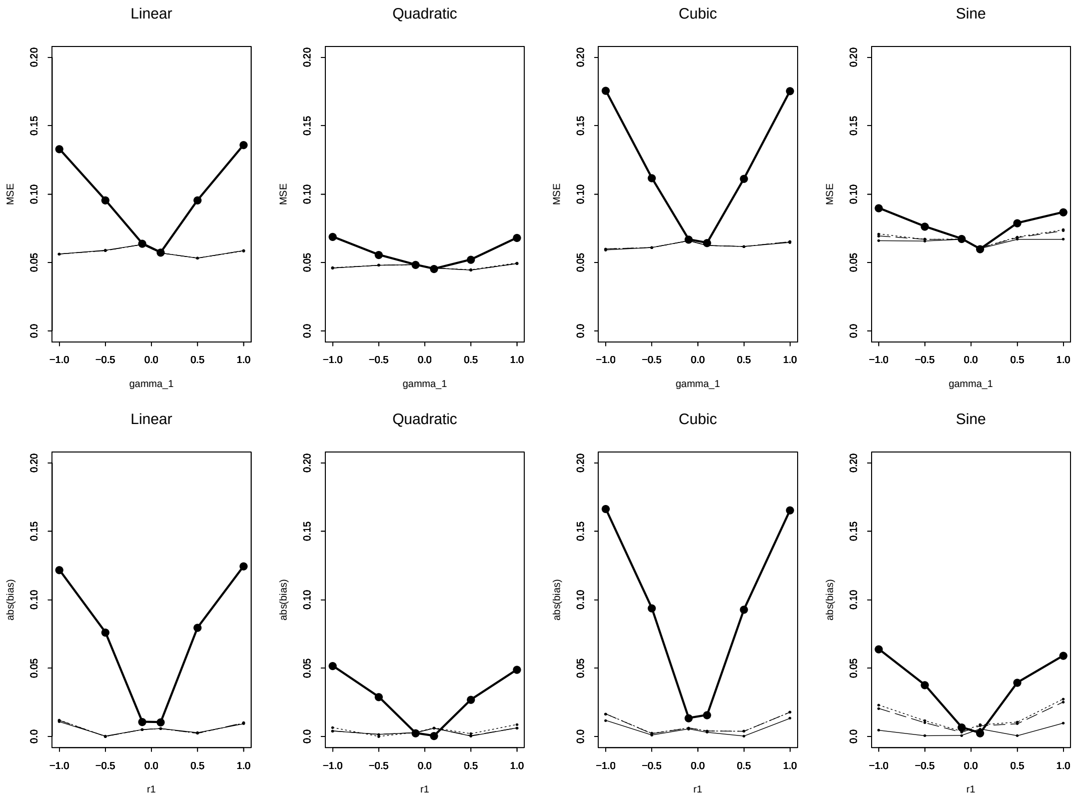

Figure 1: Plot of MSE and bias of four estimates for case 1 and $\beta_1 = \rho = 0$.
The thick solid line: the MAR estimate $\hat{\beta}_1(0)$.
The dotted line: the adjusted estimate using linear predictor, $\hat{\beta}_{1L}(\gamma_1)$.
The dashed line: the adjusted estimate using polynomial predictor, $\hat{\beta}_{1P}(\gamma_1)$.
The thin solid line: the adjusted estimate using smoothing spline predictor, $\hat{\beta}_{1G}(\gamma_1)$.

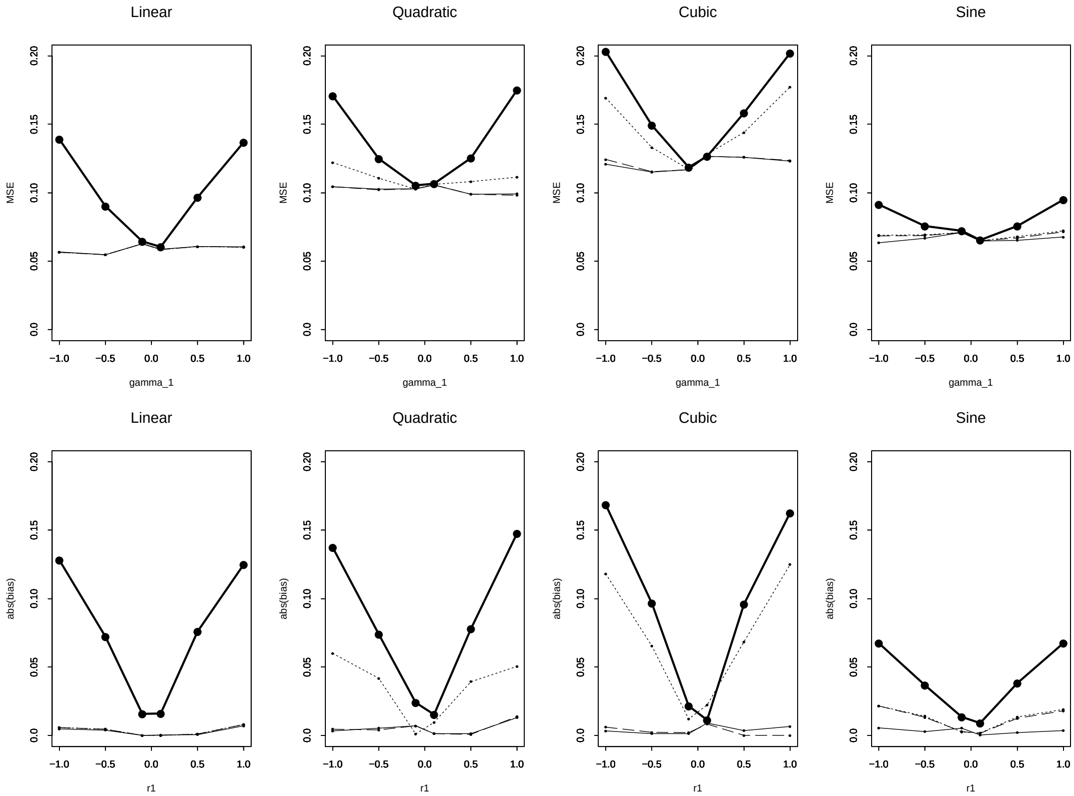

Figure 2: Plot of MSE and bias of four estimates for case 2 and $\beta_1 = \rho = 0$.
The thick solid line: the MAR estimate $\hat{\beta}_1(0)$.
The dotted line: the adjusted estimate using linear predictor, $\hat{\beta}_{1L}(\gamma_1)$.
The dashed line: the adjusted estimate using polynomial predictor, $\hat{\beta}_{1P}(\gamma_1)$.
The thin solid line: the adjusted estimate using smoothing spline predictor, $\hat{\beta}_{1G}(\gamma_1)$.

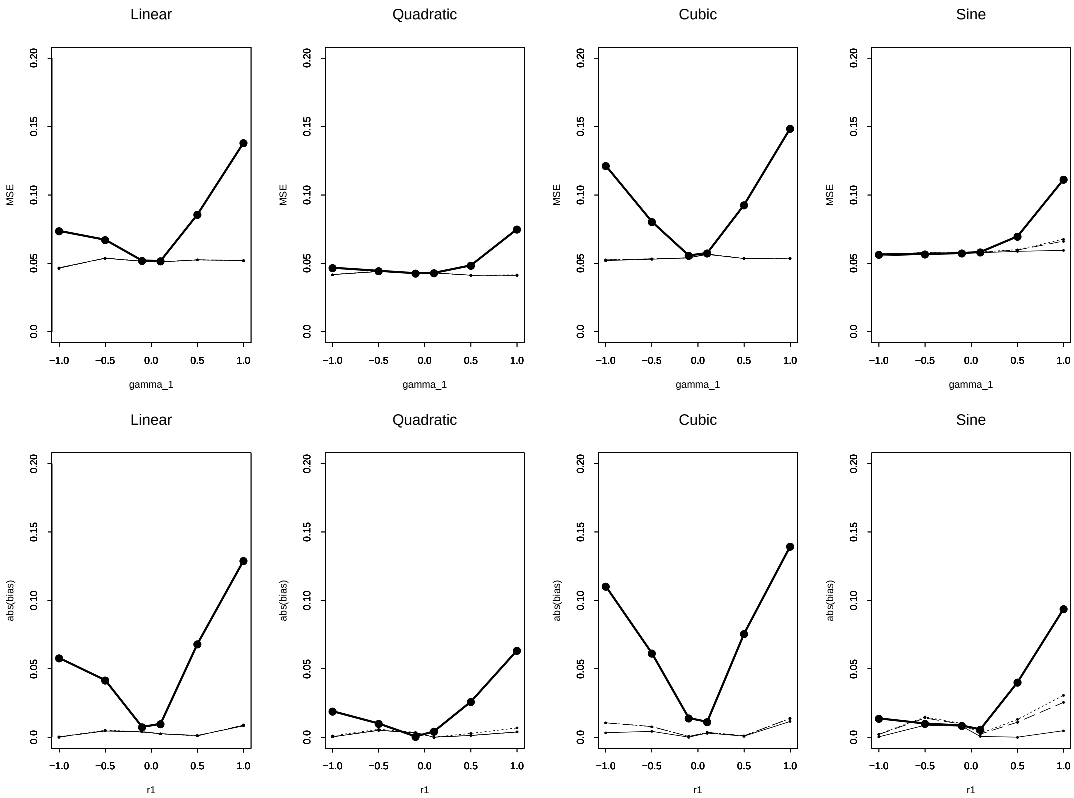

Figure 3: Plot of MSE and bias of four estimates for case 1 and $\beta_1 = \rho = 0.5$.
The thick solid line: the MAR estimate $\hat{\beta}_1(0)$.
The dotted line: the adjusted estimate using linear predictor, $\hat{\beta}_{1L}(\gamma_1)$.
The dashed line: the adjusted estimate using polynomial predictor, $\hat{\beta}_{1P}(\gamma_1)$.
The thin solid line: the adjusted estimate using smoothing spline predictor, $\hat{\beta}_{1G}(\gamma_1)$.

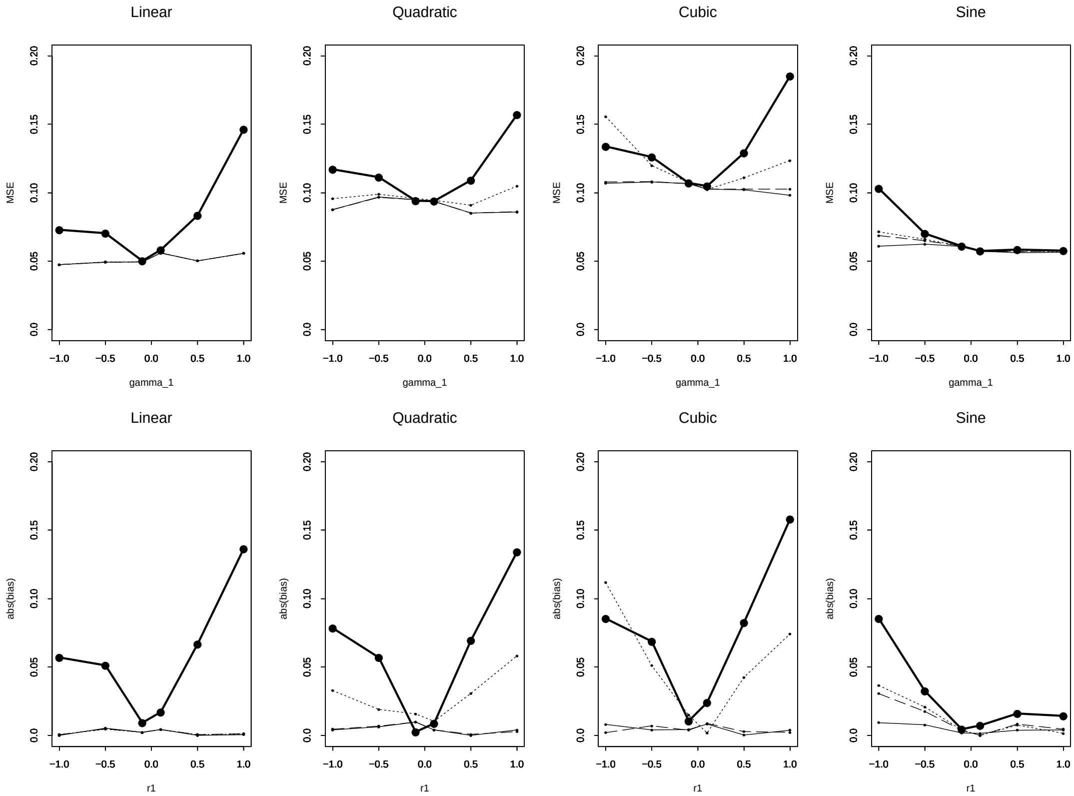

Figure 4: Plot of MSE and bias of four estimates for case 2 and $\beta_1 = \rho = 0.5$.
The thick solid line: the MAR estimate $\hat{\beta}_1(0)$.
The dotted line: the adjusted estimate using linear predictor, $\hat{\beta}_{1L}(\gamma_1)$.
The dashed line: the adjusted estimate using polynomial predictor, $\hat{\beta}_{1P}(\gamma_1)$.
The thin solid line: the adjusted estimate using smoothing spline predictor, $\hat{\beta}_{1G}(\gamma_1)$.

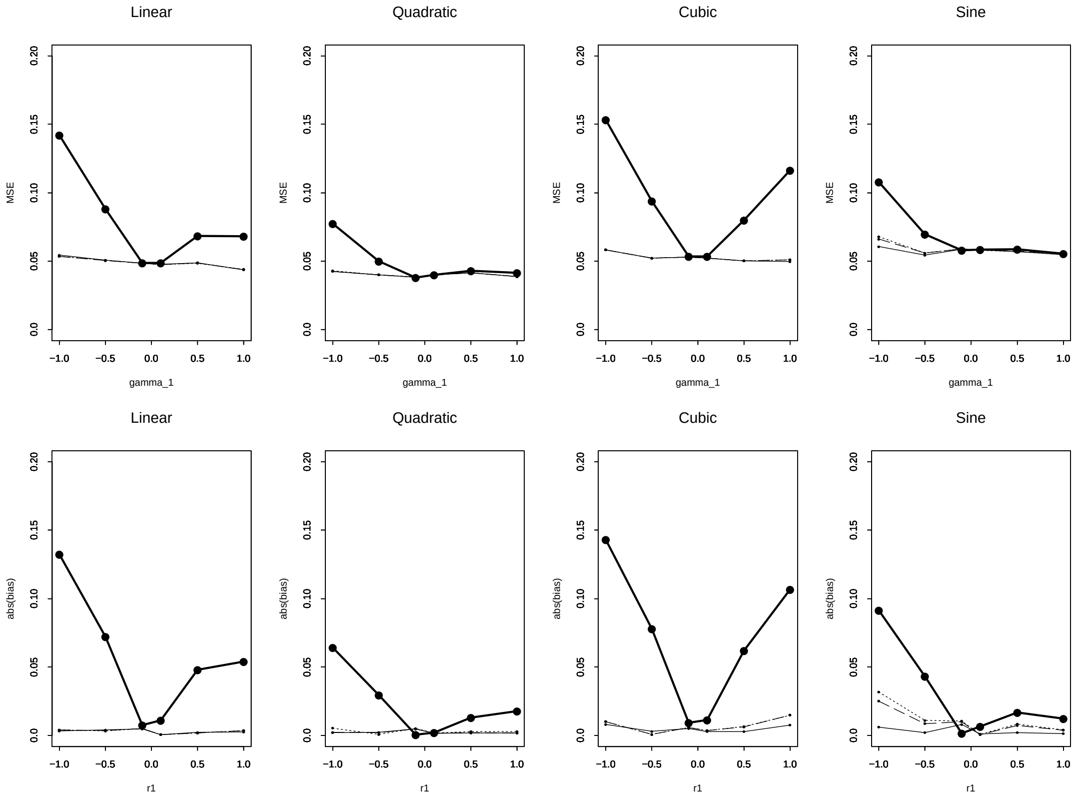

Figure 5: Plot of MSE and bias of four estimates for case 1 and $\beta_1 = \rho = -0.5$.
The thick solid line: the MAR estimate $\hat{\beta}_1(0)$.
The dotted line: the adjusted estimate using linear predictor, $\hat{\beta}_{1L}(\gamma_1)$.
The dashed line: the adjusted estimate using polynomial predictor, $\hat{\beta}_{1P}(\gamma_1)$.
The thin solid line: the adjusted estimate using smoothing spline predictor, $\hat{\beta}_{1G}(\gamma_1)$.

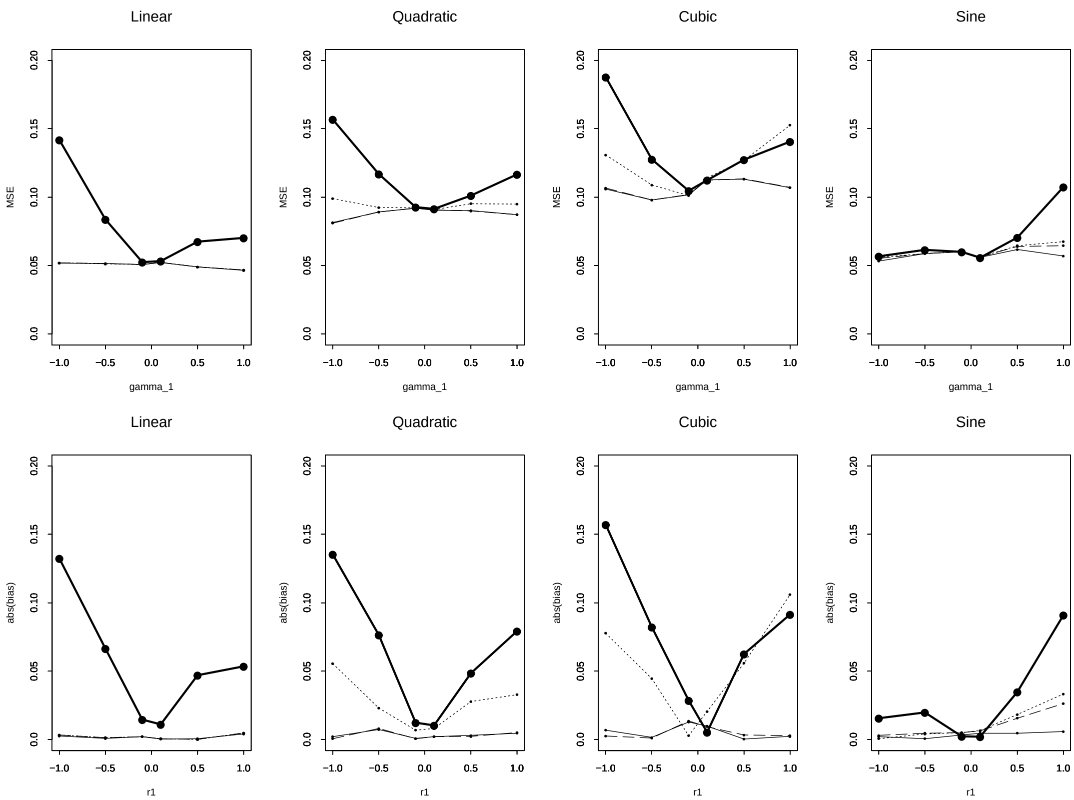

Figure 6: Plot of MSE and bias of four estimates for case 2 and $\beta_1 = \rho = -0.5$.
The thick solid line: the MAR estimate $\hat{\beta}_1(0)$.
The dotted line: the adjusted estimate using linear predictor, $\hat{\beta}_{1L}(\gamma_1)$.
The dashed line: the adjusted estimate using polynomial predictor, $\hat{\beta}_{1P}(\gamma_1)$.
The thin solid line: the adjusted estimate using smoothing spline predictor, $\hat{\beta}_{1G}(\gamma_1)$.