Supporting Information to "Efficient Semiparametric Inference for Two-Phase Studies with Outcome and Covariate Measurement Errors"

Ran Tao[1,2], Sarah C. Lotspeich[1], Gustavo Amorim[1], Pamela A. Shaw[3], and Bryan E. Shepherd[1]

[1]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee 37232, U.S.A.

[2]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, Tennessee 37232, U.S.A.

[3]Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A.

Table S1. Simulation results for LSE under additive errors in Y^* and X^* when simple random sampling is used in the second phase. Bias and SE are, respectively, the empirical bias and standard error of the parameter estimator; SEE is the empirical mean of the standard error estimator; CP is the coverage probability of the 95% confidence interval; and RE is the efficiency of LSE relative to that of the other three methods. Each entry is based on 10,000 replicates

						RE		
r	p	Bias	SE	SEE	CP	MBE	CCE	SMLE
−0.5	0.1	0.000	0.050	0.050	0.949	0.665	0.592	0.497
	0.3	0.000	0.050	0.050	0.949	1.059	0.735	0.592
	0.6	0.000	0.050	0.050	0.949	1.552	0.818	0.700
	1.0	0.000	0.050	0.050	0.950	2.094	0.860	0.824
−0.3	0.1	0.000	0.050	0.050	0.949	0.622	0.573	0.498
	0.3	0.000	0.050	0.050	0.949	0.983	0.721	0.600
	0.6	0.000	0.050	0.050	0.949	1.473	0.811	0.713
	1.0	0.000	0.050	0.050	0.950	2.061	0.859	0.845
0.0	0.1	0.000	0.050	0.050	0.949	0.570	0.544	0.497
	0.3	0.000	0.050	0.050	0.949	0.892	0.698	0.603
	0.6	0.000	0.050	0.050	0.949	1.340	0.802	0.733
	1.0	0.000	0.050	0.050	0.950	1.957	0.859	0.862
0.3	0.1	0.000	0.050	0.050	0.949	0.539	0.524	0.486
	0.3	0.000	0.050	0.050	0.949	0.809	0.680	0.599
	0.6	0.000	0.050	0.050	0.949	1.165	0.789	0.720
	1.0	0.000	0.050	0.050	0.950	1.783	0.850	0.847
0.5	0.1	0.000	0.050	0.050	0.949	0.524	0.512	0.478
	0.3	0.000	0.050	0.050	0.949	0.760	0.662	0.585
	0.6	0.000	0.050	0.050	0.949	1.075	0.770	0.699
	1.0	0.000	0.050	0.050	0.950	1.625	0.836	0.824

Table S2. Simulation results under additive errors in Y^* and X^* for $r = 0.3$ and $p = 0.6$ with varying second phase sample sizes. Bias and SE are, respectively, the empirical bias and standard error of the parameter estimator; SEE is the empirical mean of the standard error estimator; CP is the coverage probability of the 95% confidence interval; and RE is the efficiency relative to that of the SMLE. Each entry is based on 10,000 replicates

	MBE					CCE					SMLE			
n_2	Bias	SE	SEE	CP	RE	Bias	SE	SEE	CP	RE	Bias	SE	SEE	CP
25	−0.027	0.193	0.169	0.944	0.625	−0.003	0.181	0.141	0.868	0.711	−0.005	0.152	0.138	0.823
50	−0.013	0.125	0.117	0.950	0.819	−0.001	0.123	0.107	0.905	0.856	−0.008	0.114	0.115	0.915
100	−0.006	0.089	0.086	0.949	0.820	−0.001	0.084	0.079	0.931	0.911	−0.005	0.080	0.083	0.951
200	−0.002	0.067	0.066	0.949	0.718	−0.001	0.061	0.058	0.938	0.873	−0.001	0.057	0.056	0.944
300	−0.001	0.059	0.059	0.952	0.648	0.000	0.050	0.049	0.945	0.907	0.000	0.047	0.046	0.943

Table S3. Simulation results under the model $Y = 0.3 + 0.4X_a + 0.5X_b + \epsilon$ when there are additive errors in Y^* and X_a^* and simple random sampling is used in the second phase. Bias and SE are, respectively, the empirical bias and standard error of the parameter estimator; SEE is the empirical mean of the standard error estimator; CP is the coverage probability of the 95% confidence interval; and RE is the efficiency relative to that of the SMLE. Each entry is based on 10,000 replicates

				MBE					CCE					SMLE			
τ	r	p	Covariate	Bias	SE	SEE	CP	RE	Bias	SE	SEE	CP	RE	Bias	SE	SEE	CP
0.5	0.3	0.6	X_a	0.000	0.042	0.042	0.950	0.810	0.000	0.040	0.040	0.947	0.888	−0.002	0.038	0.038	0.948
			X_b	0.000	0.081	0.082	0.954	1.081	0.000	0.084	0.084	0.945	1.010	0.000	0.084	0.084	0.945
		1.0	X_a	0.001	0.044	0.044	0.949	0.764	0.000	0.041	0.041	0.948	0.887	−0.002	0.039	0.038	0.948
			X_b	0.000	0.086	0.087	0.951	1.033	−0.001	0.087	0.087	0.950	1.005	−0.001	0.087	0.087	0.950
	0.5	0.6	X_a	0.000	0.042	0.042	0.949	0.811	0.000	0.040	0.040	0.948	0.883	−0.001	0.038	0.038	0.950
			X_b	0.000	0.080	0.081	0.953	1.079	0.000	0.083	0.083	0.946	1.006	0.000	0.083	0.083	0.946
		1.0	X_a	0.001	0.044	0.043	0.948	0.769	0.000	0.041	0.041	0.948	0.882	−0.001	0.039	0.038	0.948
			X_b	0.000	0.084	0.085	0.950	1.042	−0.001	0.086	0.086	0.950	1.004	−0.001	0.086	0.086	0.950
1.0	0.3	0.6	X_a	0.000	0.046	0.046	0.949	0.721	0.000	0.042	0.041	0.947	0.866	−0.002	0.039	0.039	0.947
			X_b	0.000	0.087	0.089	0.953	1.043	0.000	0.089	0.088	0.946	1.006	0.000	0.089	0.088	0.945
		1.0	X_a	0.001	0.050	0.049	0.948	0.626	0.000	0.043	0.042	0.947	0.849	−0.002	0.040	0.039	0.945
			X_b	0.000	0.096	0.097	0.950	0.951	−0.001	0.093	0.092	0.950	1.015	−0.001	0.093	0.093	0.951
	0.5	0.6	X_a	0.000	0.045	0.045	0.949	0.735	0.000	0.042	0.041	0.948	0.866	−0.001	0.039	0.038	0.948
			X_b	0.000	0.085	0.087	0.953	1.058	0.000	0.088	0.087	0.947	1.004	0.000	0.088	0.087	0.946
		1.0	X_a	0.001	0.049	0.048	0.950	0.648	0.000	0.043	0.042	0.948	0.852	−0.001	0.039	0.039	0.947
			X_b	0.000	0.092	0.093	0.949	0.981	−0.001	0.091	0.091	0.949	1.011	−0.001	0.092	0.091	0.952

Table S4. Simulation results under the model $Y = 0.3 + 0.4X_a + 0.5X_b + \epsilon$ when there are additive errors in Y^* and X_a^* and stratified simple random sampling is used in the second phase. Bias and SE are, respectively, the empirical bias and standard error of the parameter estimator; SEE is the empirical mean of the standard error estimator; CP is the coverage probability of the 95% confidence interval; and RE is the efficiency relative to that of the SMLE. Each entry is based on 10,000 replicates

				MBE					CCE					SMLE			
τ	r	p	Covariate	Bias	SE	SEE	CP	RE	Bias	SE	SEE	CP	RE	Bias	SE	SEE	CP
0.5	0.3	0.6	X_a	0.000	0.044	0.043	0.945	0.795	0.000	0.041	0.040	0.945	0.932	−0.002	0.039	0.039	0.947
			X_b	0.000	0.083	0.082	0.946	0.943	0.000	0.080	0.073	0.923	1.005	0.000	0.080	0.079	0.947
		1.0	X_a	0.000	0.045	0.045	0.947	0.785	0.001	0.042	0.041	0.945	0.884	−0.002	0.040	0.039	0.946
			X_b	0.001	0.087	0.087	0.950	0.856	0.000	0.080	0.074	0.929	1.001	0.000	0.080	0.080	0.950
	0.5	0.6	X_a	0.000	0.044	0.043	0.945	0.799	−0.009	0.041	0.040	0.936	0.936	−0.001	0.039	0.039	0.946
			X_b	0.000	0.082	0.081	0.947	0.966	0.000	0.080	0.072	0.921	1.005	0.000	0.080	0.079	0.948
		1.0	X_a	−0.001	0.044	0.044	0.947	0.791	−0.012	0.042	0.041	0.934	0.889	−0.002	0.039	0.039	0.946
			X_b	0.001	0.085	0.085	0.950	0.890	0.000	0.080	0.074	0.927	1.000	0.000	0.080	0.080	0.950
1.0	0.3	0.6	X_a	0.000	0.046	0.045	0.943	0.739	0.002	0.043	0.042	0.945	0.850	−0.002	0.040	0.039	0.946
			X_b	0.000	0.089	0.089	0.948	0.834	0.000	0.081	0.075	0.930	1.006	0.000	0.081	0.080	0.948
		1.0	X_a	0.000	0.048	0.048	0.949	0.697	0.003	0.045	0.044	0.945	0.797	−0.002	0.040	0.039	0.945
			X_b	0.001	0.097	0.097	0.950	0.703	0.000	0.081	0.078	0.937	1.000	0.000	0.081	0.081	0.950
	0.5	0.6	X_a	0.000	0.046	0.045	0.944	0.748	−0.011	0.043	0.042	0.935	0.858	−0.001	0.040	0.039	0.946
			X_b	0.000	0.087	0.087	0.948	0.869	0.000	0.081	0.075	0.928	1.003	0.000	0.081	0.080	0.948
		1.0	X_a	−0.001	0.047	0.047	0.948	0.708	−0.013	0.044	0.044	0.935	0.804	−0.002	0.040	0.039	0.945
			X_b	0.001	0.094	0.093	0.950	0.751	0.000	0.081	0.077	0.935	1.000	0.000	0.081	0.081	0.950

Table S5. Simulation results under additive errors in Y^* and X^* for $r = 0.3$ and $p = 0.6$ when the distribution of ϵ is not normal. Bias and SE are, respectively, the empirical bias and standard error of the parameter estimator; SEE is the empirical mean of the standard error estimator; CP is the coverage probability of the 95% confidence interval; and RE is the efficiency relative to that of the SMLE. Each entry is based on 10,000 replicates

	MBE					CCE					SMLE			
Distribution	Bias	SE	SEE	CP	RE	Bias	SE	SEE	CP	RE	Bias	SE	SEE	CP
t_3	−0.002	0.078	0.077	0.953	0.732	0.000	0.072	0.069	0.945	0.874	−0.002	0.067	0.065	0.944
t_5	−0.001	0.064	0.064	0.949	0.665	0.000	0.055	0.055	0.949	0.896	−0.001	0.052	0.051	0.946
t_{10}	0.000	0.058	0.058	0.949	0.641	0.000	0.049	0.048	0.944	0.907	0.001	0.047	0.046	0.945
t_{15}	−0.003	0.057	0.057	0.951	0.651	0.001	0.047	0.047	0.947	0.931	−0.001	0.046	0.044	0.942
t_{20}	−0.002	0.056	0.056	0.949	0.623	0.000	0.047	0.046	0.944	0.897	−0.001	0.044	0.044	0.945
t_{30}	0.000	0.056	0.056	0.952	0.611	0.000	0.046	0.045	0.946	0.936	−0.001	0.044	0.043	0.947
Uniform$(-1, 1)$	−0.002	0.044	0.044	0.953	0.396	0.000	0.027	0.027	0.949	1.007	0.000	0.027	0.027	0.941
Uniform$(-2, 2)$	−0.002	0.059	0.059	0.951	0.691	0.000	0.051	0.050	0.948	0.951	−0.001	0.049	0.048	0.944

Table S6. Simulation results for LSE under additive errors in Y^* and X^* when residual-dependent sampling is used in the second phase. Bias and SE are, respectively, the empirical bias and standard error of the parameter estimator; SEE is the empirical mean of the standard error estimator; and CP is the coverage probability of the 95% confidence interval. Each entry is based on 10,000 replicates

r	p	Bias	SE	SEE	CP
−0.5	0.1	0.090	0.049	0.072	0.854
	0.3	0.190	0.053	0.068	0.142
	0.6	0.232	0.051	0.063	0.017
	1.0	0.222	0.050	0.058	0.016
−0.3	0.1	0.071	0.049	0.073	0.930
	0.3	0.153	0.053	0.069	0.368
	0.6	0.191	0.053	0.064	0.107
	1.0	0.180	0.052	0.060	0.113
0.0	0.1	0.041	0.048	0.073	0.983
	0.3	0.091	0.053	0.070	0.810
	0.6	0.118	0.054	0.066	0.587
	1.0	0.111	0.054	0.062	0.571
0.3	0.1	0.011	0.048	0.073	0.997
	0.3	0.023	0.053	0.071	0.983
	0.6	0.032	0.055	0.068	0.963
	1.0	0.030	0.055	0.064	0.954
0.5	0.1	−0.010	0.047	0.073	0.997
	0.3	−0.023	0.053	0.071	0.986
	0.6	−0.032	0.055	0.069	0.965
	1.0	−0.034	0.056	0.065	0.950

Table S7. Simulation results under additive errors in Y^*, X_1^*, and X_2^* when simple random sampling is used in the second phase. Bias and SE are, respectively, the empirical bias and standard error of the parameter estimator; SEE is the empirical mean of the standard error estimator; and CP is the coverage probability of the 95% confidence interval. Each entry is based on 10,000 replicates

		X_1				X_2			
r	p	Bias	SE	SEE	CP	Bias	SE	SEE	CP
0.0	0.1	−0.003	0.035	0.034	0.945	−0.002	0.036	0.035	0.941
	0.3	−0.005	0.039	0.038	0.940	−0.005	0.039	0.038	0.943
	0.6	−0.009	0.042	0.041	0.938	−0.010	0.042	0.041	0.935
	1.0	−0.012	0.045	0.044	0.935	−0.012	0.044	0.044	0.939
0.3	0.1	−0.001	0.035	0.034	0.942	−0.001	0.035	0.034	0.943
	0.3	−0.003	0.038	0.037	0.948	−0.003	0.039	0.037	0.940
	0.6	−0.005	0.041	0.041	0.942	−0.005	0.042	0.041	0.941
	1.0	−0.006	0.044	0.043	0.940	−0.005	0.044	0.043	0.944
0.5	0.1	−0.001	0.035	0.034	0.946	−0.001	0.034	0.034	0.945
	0.3	−0.001	0.038	0.037	0.945	−0.001	0.037	0.037	0.948
	0.6	−0.002	0.040	0.040	0.946	−0.002	0.041	0.040	0.943
	1.0	−0.002	0.042	0.042	0.946	−0.001	0.043	0.042	0.946