Supplementary Materials for

Constrained Maximum Likelihood Estimation for Model Calibration Using Summary-level Information from External Big Data Sources

S.1 Additional Simulation Results

S.1.1 Simulation Results for the Logistic Model with $N = 400$

Tables S.1–S.4 provide simulation results supplementary for those of Tables 1–4 in the main text, respectively. Here, the sample size of the internal study is $N = 400$. All the other details are the same as the corresponding simulation studies in the main text.

S.1.2 Simulation Results for the Probit Model and Further Data Application

Tables S.5–S.7 provide simulation results when the full model is given by the probit model. In these simulations, the reduced model is still given by a logistic model. In the *under-specification* and *missing covariate* scenarios, the parameter values of the true model $(\beta_0, \beta_X, \beta_Z, \beta_{XZ}) = (-1, 0.4, 0.4, 0.2)$; in the measurement error scenario, they are $(\beta_0, \beta_Z) = (-1, 0.4)$. Such specifications lead to a population disease prevalence around 20%. All the other details are the same as the simulation studies in the main text.

S.1.3 Further Data Applications

Table S.8 presents further analysis results of BCDDP plus BPC3 data. In the this analysis, the CML and GR methods are implemented assuming that the external model parameters come from a dataset that is so large that uncertainty can be ignored.

Table S.1: Simulation results for the *under-specification* setting; results multiplied by 10^3 are presented, and the coverage probabilities (CP) are reported as percents

	β_0			β_X			β_Z			β_{XZ}		
	Int	GR/ mGR	CML	Int	GR/ mGR	CML	Int	GR/ mGR	CML	Int	GR/ mGR	CML
simple random; $N = 400$												
Bias	-25.4	-5.56	-6.12	9.87	4.97	5.53	15.2	4.69	5.34	-0.41	-0.24	1.05
SE	152	40.7	40.8	151	36.3	35.0	154	37.0	35.6	149	151	149
ESE	147	35.9	37.7	148	34.8	34.3	148	34.5	34.3	137	135	139
MSE	23.6	1.69	1.70	22.8	1.34	1.25	24.0	1.39	1.30	22.1	22.7	22.1
CP	945	862	878	956	936	950	947	936	958	929	927	933
case-control; $N = 400$												
Bias	-	-	-1.17	10.2	29.6	3.89	16.3	29.4	3.68	3.03	2.92	-0.77
SE	-	-	26.9	121	18.8	27.4	120	19.5	26.8	116	117	114
ESE	-	-	26.1	117	20.3	27.7	117	20.7	28.0	114	114	112
MSE	-	-	0.72	14.8	1.23	0.77	14.8	1.24	0.73	13.4	13.6	12.9
CP	-	-	907	941	618	950	935	623	950	948	947	959

Int: internal-data only method
GR: generalized regression, mGR: modified GR for case-control sampling
CML: constrained maximum likelihood
ESE: estimated standard error
MSE: mean squared error
CP: coverage probability of a 95% confidence interval interval

Table S.2: Simulation results for the *missing covariate* setting; results multiplied by 10^3 are presented, and the coverage probabilities (CP) are reported as percents

	β_0			β_X			β_Z			β_{XZ}		
	Int	GR/ mGR	CML	Int	GR/ mGR	CML	Int	GR/ mGR	CML	Int	GR/ mGR	CML
simple random; $N = 400$												
Bias	-25.4	-12.3	-12.8	9.87	2.18	2.05	15.2	15.8	14.4	-0.41	-0.86	2.18
SE	152	57.8	57.7	151	67.7	67.7	154	154	154	149	150	149
ESE	147	53.3	54.2	148	64.4	65.1	148	148	148	137	136	139
MSE	23.6	3.49	3.49	22.8	4.58	4.58	24.0	24.1	24.0	22.1	22.5	22.2
CP	945	944	950	956	936	943	947	941	942	929	929	936
case-control; $N = 400$												
Bias	-	-	0.85	10.2	16.2	2.57	16.3	16.4	16.5	3.03	2.47	2.59
SE	-	-	36.1	121	42.2	45.3	120	120	120	116	116	116
ESE	-	-	36.9	117	42.6	45.4	117	117	117	114	114	115
MSE	-	-	1.30	14.8	2.04	2.06	14.8	14.8	14.8	13.4	13.5	13.4
CP	-	-	952	941	914	947	935	936	936	948	945	961

Int: internal-data only method
GR: generalized regression, mGR: modified GR for case-control sampling
CML: constrained maximum likelihood
ESE: estimated standard error
MSE: mean squared error
CP: coverage probability of a 95% confidence interval

Table S.3: Simulation results for the *measurement error* setting; results presented are multiplied by 10^3, and the coverage probability (CP) is in percents

	β_0			β_Z		
	Int	GR/mGR	CML	Int	GR/mGR	CML
simple random; $N = 400$						
Bias	-12.5	-10.8	-0.84	7.33	7.06	2.72
SE	141	41.8	25.6	142	138	66.5
ESE	139	38.7	24.4	137	131	61.3
MSE	20.1	1.87	0.66	20.2	19.2	4.43
CP	947	913	952	954	942	912
case-control; $N = 400$						
Bias	-	-	1.01	8.33	7.99	7.13
SE	-	-	20.4	104	99.4	59.3
ESE	-	-	20.7	106	101	58.3
MSE	-	-	0.42	10.9	9.93	3.56
CP	-	-	955	961	956	948

Int: internal-data only method
GR: generalized regression, mGR: modified GR for case-control sampling
CML: constrained maximum likelihood
ESE: estimated standard error
MSE: mean squared error
CP: coverage probability of a 95% confidence interval.

Table S.4: Simulation results for the *missing covariate* setting when the covariate distributions are different between internal and external populations; results multiplied by 10^3 are presented, and the coverage probabilities (CP) are reported as percents

	β_0			β_X			β_Z			β_{XZ}		
	Int	CML	SCML	Int	CML	SCML	Int	CML	SCML	Int	CML	SCML
simple random; $N = 400$												
Bias	-15.0	-38.0	-8.16	19.1	-80.1	8.02	9.79	8.24	6.10	-9.96	-5.68	-13.4
SE	149	55.6	42.8	150	62.9	50.3	147	147	144	147	149	143
ESE	147	55.2	44.4	147	64.6	50.2	147	148	145	136	137	136
MSE	22.5	4.53	1.90	22.9	10.4	2.60	21.7	21.8	20.7	21.8	22.1	20.5
CP	948	950	940	952	843	900	948	945	950	925	933	938
Bias	-	-26.3	-6.42	15.6	-86.0	4.02	6.48	8.02	6.09	2.27	-4.45	-3.37
SE	-	37.9	39.1	122	45.0	44.3	119	119	117	117	118	116
ESE	-	37.3	38.6	117	45.7	44.3	117	116	115	115	112	114
MSE	-	2.13	1.57	15.2	9.42	1.98	14.2	14.2	13.8	13.7	13.9	13.5
CP	-	934	939	945	542	920	944	937	944	948	943	953

Int: internal-data only method
CML: constrained maximum likelihood method
SCML: synthetic constrained maximum likelihood method
ESE: estimated standard error
MSE: mean squared error
CP: coverage probability of a 95% confidence interval

Table S.5: Simulation results for the *under-specification* setting; results multiplied by 10^3 are presented, and the coverage probabilities (CP) are reported in percents

	β_0			β_X			β_Z			β_{XZ}		
	Int	GR	CML	Int	GR	CML	Int	GR	CML	Int	GR	CML
simple random; $N = 400$												
Bias	-17.1	-5.45	-4.81	11.9	5.12	4.22	13.9	8.17	7.08	-2.03	-3.35	-1.96
SE	89.2	18.6	18.3	93.6	26.9	23.4	92.1	270	23.7	98.4	100	99.4
ESE	86.3	16.2	16.4	89.5	24.5	23.7	89.3	245	23.7	87.4	84.9	91.6
MSE	8.25	0.37	0.36	8.90	0.75	0.57	8.68	0.79	0.61	9.69	0.79	0.61
CP	930	927	856	938	925	962	946	912	960	917	900	935
simple random; $N = 1000$												
Bias	-3.78	-1.44	-1.18	2.94	1.10	0.75	1.78	3.46	3.51	0.09	0.19	0.31
SE	53.5	10.7	10.7	59.4	15.0	13.7	56.6	15.2	13.6	59.0	58.7	59.0
ESE	53.6	10.0	9.93	55.9	14.4	13.5	55.9	14.4	13.4	55.9	55.1	57.1
MSE	2.88	0.12	0.12	3.53	0.22	0.19	3.21	0.23	0.20	3.47	3.44	3.47
CP	95.2	90.6	88.0	93.5	94.1	95.6	95.6	93.3	94.1	93.9	93.6	95.0

Int: internal-data only method

CML: constrained maximum likelihood

ESE: estimated standard error

MSE: mean squared error

CP: coverage probability of 95% confidence interval

Table S.6: Simulation results for the *missing covariate* setting; results multiplied by 10^3 are presented, and the coverage probabilities (CP) are reported in percents

	β_0			β_X			β_Z			β_{XZ}		
	Int	GR	CML	Int	GR/	CML	Int	GR	CML	Int	GR	CML
simple random; $N = 400$												
Bias	-17.1	-8.95	-8.50	11.9	3.66	2.03	13.9	15.5	14.4	-2.03	-3.33	-0.95
SE	89.2	37.7	37.6	93.6	44.3	43.0	92.1	92.8	92.6	98.4	99.1	99.6
ESE	86.3	35.9	36.3	89.5	42.4	42.7	89.3	88.8	92.2	87.4	86.4	92.2
MSE	8.25	1.50	1.49	8.90	1.97	1.85	8.68	8.84	8.77	9.69	9.82	9.91
CP	930	946	952	938	929	948	946	939	947	917	909	938
simple random; $N = 1000$												
Bias	-3.78	-1.77	-1.42	2.94	2.29	1.75	1.78	2.33	1.87	0.09	0.01	0.66
SE	53.5	21.6	21.6	59.4	26.5	25.9	56.6	56.8	56.7	59.0	58.7	59.2
ESE	53.6	21.6	21.6	55.9	25.6	25.5	55.9	55.7	56.1	55.9	55.6	57.2
MSE	2.88	0.47	0.47	3.53	0.71	0.67	3.21	3.23	3.21	3.47	3.44	3.50
CP	95.2	95.1	95.0	93.5	93.3	94.5	95.6	95.9	95.9	93.9	93.8	94.7

Int: internal-data only method

GR: generalized regression

CML: constrained maximum likelihood

ESE: estimated standard error

MSE: mean squared error

CP: coverage probability of 95% confidence interval

Table S.7: Simulation results for the *measurement error* setting; results multiplied by 10^3 are presented, and the coverage probability (CP) is in percents

	β_0			β_Z		
	Int	GR	CML	Int	GR	CML
simple random; $N = 400$						
Bias	-6.04	-7.13	-1.59	5.10	5.00	2.57
SE	82.2	34.7	27.6	83.8	82.0	58.7
ESE	80.0	32.9	26.2	81.5	77.9	55.0
MSE	6.79	1.26	0.77	7.05	6.74	3.45
CP	942	936	946	945	944	916
simple random; $N = 1000$						
Bias	-0.51	-2.11	-0.48	-0.45	-0.05	0.90
SE	51.1	21.2	16.5	52.4	50.0	35.7
ESE	50.3	20.6	16.5	51.2	49.2	34.9
MSE	2.61	0.45	0.27	2.74	2.50	1.28
CP	95.7	93.8	94.8	94.9	95.4	94.4

Int: internal-data only method

GR: generalized regression

CML: constrained maximum likelihood

ESE: estimated standard error

MSE: mean squared error

CP: coverage probability of 95% confidence interval

Table S.8: Analysis results of BCDDP data assuming that the uncertainty associated with the parameters in the external mmodel (θ) can be ignored. The variables in the model include: number of first-degree relatives with breast cancer (numrel), age at menarche (two dummy variables agemen1–agemen2), age at first live birth (three dummy variables ageflb1–ageflb3), weight, number of previous biopsies (nbiops), and mammographic density (MD)†

	Internal data	mGR	CML
variable	Est. (SE)	Est. (SE)	Est. (SE)
numrel	0.648 (0.090)	0.346 (0.030)	0.297 (0.020)
agemen1	0.083 (0.091)	0.079 (0.019)	0.077 (0.019)
agemen2	0.468 (0.124)	0.167 (0.028)	0.167 (0.028)
ageflb1	-0.018 (0.146)	-0.117 (0.030)	-0.117 (0.031)
ageflb2	0.086 (0.144)	-0.005 (0.033)	-0.005 (0.033)
ageflb3	0.251 (0.137)	0.165 (0.040)	0.163 (0.040)
weight	0.020 (0.004)	0.022 (0.002)	0.024 (0.001)
nbiops	0.180 (0.070)	0.178 (0.069)	0.165 (0.073)
MD	0.430 (0.044)	0.428 (0.045)	0.441 (0.043)

†: adjusted for 5-year age strata

mGR: modified GR for case-control sampling

CML: constrained likelihood method

Est.: estimated coefficient

SE: estimated standard error

S.2 The Score Function and Negative Hessian Matrix for Pseudo-Loglikelihood

S.2.1 Simple Random Sampling Design

Let $s_\beta(Y, X, Z) = \partial \log\{f_\beta(Y|X, Z)\}/\partial\beta$ be the score function of the likelihood for the full model $f_\beta(\cdot)$. Based on the pseudo-loglikelihood (equation (4) in the main text) for the constrained likelihood, the score function for (β, λ) of the pseudo-loglikelihood can be obtained as

$$s^*_\beta = \frac{\partial\, l^*_{\beta,\lambda}}{\partial\beta} = \sum_{i=1}^{N} s_\beta(Y_i, X_i, Z_i) + \widetilde{s}_\beta(X_i, Z_i) \equiv \sum_{i=1}^{N} s^*_\beta(Y_i, X_i, Z_i),$$

with

$$\widetilde{s}_\beta(X_i, Z_i) = \frac{c_\beta(X_i, Z_i; \theta)\lambda}{1 - \lambda^T u_\beta(X_i, Z_i; \theta)}, \tag{S.1}$$

$$c_\beta(x, z; \theta) = \int_Y s_\beta(Y, x, z) U^T(Y, x, \theta) f_\beta(Y|x, z) dY, \tag{S.2}$$

and

$$s_\lambda^* = \frac{\partial\, l_{\beta,\lambda}^*}{\partial \lambda} = \sum_{i=1}^N \frac{u_\beta(X_i, Z_i; \theta)}{1 - \lambda^T u_\beta(X_i, Z_i; \theta)} \equiv \sum_{i=1}^N s_\lambda^*(X_i, Z_i). \tag{S.3}$$

The estimator $(\widehat{\beta}, \widehat{\lambda})$ for (β, λ) is obtained as the solution to $\{s_\beta^{*T}, s_\lambda^{*T}\}^T = 0$.

Let

$$i_{\beta\beta}(Y, X, Z) = -\partial^2 \log\{f_\beta(Y|X, Z)\}/\partial\beta\partial\beta^T, \tag{S.4}$$

and $d_{\beta\beta}(Y, X, Z) = i_{\beta\beta}(Y, X, Z) - s_\beta(Y, X, Z)^{\otimes 2}$, with $A^{\otimes 2} = AA^T$ for a vector A. The negative Hessian matrix $I^* = -\partial^2\, l_{\beta,\lambda}^* / \partial(\beta^T, \lambda^T)^T \partial(\beta^T, \lambda^T)$ derived from the pseudo-loglikelihood is given by the following expressions for the component matrices:

$$\begin{aligned} I_{\beta\beta}^* &= -\frac{\partial^2\, l_{\beta,\lambda}^*}{\partial\beta\partial\beta^T} \\ &= \sum_{i=1}^N i_{\beta\beta}(Y_i, X_i, Z_i) + \frac{d_\beta(X_i, Z_i; \theta)}{1 - \lambda^T u_\beta(X_i, Z_i; \theta)} - \widetilde{s}_\beta(X_i, Z_i)^{\otimes 2}, \end{aligned}$$

where

$$d_\beta(x, z; \theta) = \int_Y \lambda^T U(Y|x, \theta) d_{\beta\beta}(Y, x, z) f_\beta(Y|x, z) dY; \tag{S.5}$$

$$\begin{aligned} I_{\beta\lambda}^* = I_{\beta\lambda}^{*T} &= -\frac{\partial^2\, l_{\beta,\lambda}^*}{\partial\beta\partial\lambda^T} \\ &= -\sum_{i=1}^N \left\{ \frac{c_\beta(X_i, Z_i; \theta)}{1 - \lambda^T u_\beta(X_i, Z_i; \theta)} + \widetilde{s}_\beta(X_i, Z_i) s_\lambda^{*T}(X_i, Z_i) \right\}; \end{aligned}$$

and

$$I_{\lambda\lambda}^* = -\frac{\partial^2\, l_{\beta,\lambda}^*}{\partial\lambda\partial\lambda^T} = -\sum_{i=1}^N s_\lambda^*(X_i, Z_i)^{\otimes 2}.$$

S.2.2 Case-Control Sampling Design

As defined in equation (8) in the main text,

$$p_{\beta,\alpha}(y|x,z) = \frac{\mu_y f_\beta(y|x,z)}{\sum_y \mu_y f_\beta(y|x,x)} = \frac{\exp(\alpha y) f_\beta(y|x,z)}{\sum_y \exp(\alpha y) f_\beta(y|x,z)},$$

with $\alpha = \log(\mu_1/\mu_0)$. We can then rewrite the pseudo-loglikelihood (equation (6) in the main text) as

$$l^{*,cc}_{\beta,\lambda,\alpha} = \sum_{i=1}^{N} \log\left\{\frac{p_{\beta,\alpha}(Y_i, X_i, Z_i)}{1-\lambda^T u_{\beta,\alpha}(X_i, Z_i;\theta)}\right\},$$

where

$$u_{\beta,\alpha}(X,Z;\theta) = \sum_y U(y|X,\theta)\exp(-\alpha y) p_{\beta,\alpha}(y|X,Z),$$

and λ is redefined as $\lambda := \lambda/\mu_0$, With the above expression the pseudo-loglikelihood under the case-control sampling design has the same form as that under the simple random sampling (equation (4) in the main text), except that now the additional nuisance parameter α appears. In fact, we can extend the results to more general settings where the internal study is performed under the stratified case-control sampling design, namely the sampling is done through strata formed by the cross-classification of the case-control status as well as levels of some covariates (or their crude surrogates). We denote by W the vector of indicators for the covariate levels defining the strata. In this general setting, the pseudo-loglikelihood takes the form

$$l^{*,cc}_{\beta,\lambda,\alpha} = \sum_{i=1}^{N} \log\left\{\frac{p_{\beta,\alpha}(Y_i, X_i, Z_i)}{1-\lambda^T u_{\beta,\alpha}(X_i, Z_i;\theta)}\right\},$$

where

$$p_{\beta,\alpha}(Y|X,Z) = \frac{\exp(YW^T\alpha) f_\beta(Y|X,Z)}{\sum_y \exp(yW^T\alpha) f_\beta(y|X,Z)} \tag{S.6}$$

and

$$u_{\beta,\alpha}(X,Z;\theta) = \sum_y U(y|X,\theta) e^{-yW^T\alpha} p_{\beta,\alpha}(y|X,Z).$$

The score functions s^*_β and s^*_λ for (β,λ), and the corresponding component submatrices $I^*_{\beta\beta}$, $I^*_{\beta\lambda}$ and $I^*_{\lambda\lambda}$ of the negative Hessian matrix of the pseudo-loglikelihood thus have the same forms as those given in the case of simple random sampling, with $f_\beta(Y|X,Z)$ now replaced by $p_{\beta,\alpha}(Y|X,Z)$, and $u_\beta(X,Z;\theta)$ replaced by $u_{\beta,\alpha}(X,Z;\theta)$. The score function for the nuisance parameter α is given by

$$s^*_\alpha = \frac{\partial\, l^*_{\beta,\lambda,\alpha}}{\partial \alpha} = \sum_{i=1}^{N} s_\alpha(Y_i, X_i, Z_i) + \widetilde{s}_\alpha(X_i, Z_i) \equiv \sum_{i=1}^{N} s^*_\alpha(Y_i, X_i, Z_i),$$

where $s_\alpha(Y,X,Z) = \partial \log\{p_{\beta,\alpha}(Y|X,Z)\}/\partial\alpha$, and

$$\widetilde{s}_\alpha(X_i,Z_i) = \frac{c_{\beta,\alpha}(X_i,Z_i;\theta)\lambda}{1-\lambda^T u_{\beta,\alpha}(X_i,Z_i;\theta)},$$

with

$$c_{\beta,\alpha}(X,Z;\theta) = \sum_y \{s_\alpha(y,X,Z) - yW\}U^T(y|X,\theta)e^{-yW^T\alpha}p_{\beta,\alpha}(y|X,Z;\beta).$$

The component submatrices of the negative Hessian matrix corresponding to α are given as

$$\begin{aligned} I^*_{\alpha\alpha} &= -\frac{\partial^2 l^{*,cc}_{\beta,\lambda,\alpha}}{\partial\alpha\partial\alpha^T} \\ &= \sum_{i=1}^N i_{\alpha\alpha}(Y_i,X_i,Z_i) + \frac{d_{\beta,\alpha}(X_i,Z_i;\theta)}{1-\lambda^T u_{\beta,\alpha}(X_i,Z_i;\theta)} - \widetilde{s}_\alpha(X_i,Z_i)^{\otimes 2}, \end{aligned}$$

where $i_{\alpha\alpha}(Y,X,Z) = -\partial^2\log\{p_{\beta,\alpha}(Y|X,Z)\}/\partial\alpha\partial\alpha^T$, and

$$d_{\beta,\alpha}(x,z;\theta) = \sum_y \lambda^T U(y|x,\theta)\{d_{\alpha\alpha}(y,X,Z) - yWW^T + yW s_\alpha^T(y,X,Z)\}e^{-yW^T\alpha}p_{\beta,\alpha}(y|X,Z),$$

with $d_{\alpha\alpha}(Y,X,Z) = i_{\alpha\alpha}(Y,X,Z) - s_\alpha(Y,X,Z)^{\otimes 2}$;

$$\begin{aligned} I^*_{\alpha\beta} = I^{*T}_{\beta\alpha} &= -\frac{\partial^2 l^{*,cc}_{\beta,\lambda,\alpha}}{\partial\alpha\partial\beta^T} \\ &= \sum_{i=1}^N \left\{ i_{\alpha\beta}(Y_i,X_i,Z_i) + \frac{e_{\beta,\alpha}(X_i,Z_i;\theta)}{1-\lambda^T u_{\beta,\alpha}(X_i,Z_i;\theta)} - \widetilde{s}_\alpha(X_i,Z_i)\widetilde{s}^T_\beta(X_i,Z_i) \right\}, \end{aligned}$$

where $i_{\alpha\beta}(Y,X,Z) = -\partial^2 \log\{p_{\beta,\alpha}(Y|X,Z)\}/\partial\alpha\partial\beta^T$, and

$$e_{\beta,\alpha}(x,z;\theta) = \sum_y \lambda^T U(y|x,\theta)\{d_{\alpha\beta}(y,X,Z) + yW s_\beta^T(y,X,Z)\}e^{-yW^T\alpha}p_{\beta,\alpha}(y|X,Z),$$

with $d_{\alpha\beta}(Y,X,Z) = i_{\alpha\beta}(Y,X,Z) - s_\alpha(Y,X,Z)s_\beta(Y,X,Z)^T$; and

$$\begin{aligned} I^*_{\alpha\lambda} = I^{*T}_{\lambda\alpha} &= -\frac{\partial^2 l^{*,cc}_{\beta,\lambda,\alpha}}{\partial\alpha\partial\lambda^T} \\ &= -\sum_{i=1}^N \left\{ \frac{c_{\beta,\alpha}(X_i,Z_i;\theta)}{1-\lambda^T u_{\beta,\alpha}(X_i,Z_i;\theta)} + \widetilde{s}_\alpha(X_i,Z_i)s^{*T}_\lambda(X_i,Z_i) \right\}. \end{aligned}$$

Note that in Proposition 1 we have absorbed the nuisance parameter α into β so that the expressions provided there can be unified under both the settings of simple random and case-control designs.

S.3 The Score Function and Negative Hessian Matrix for Synthetic Constrained Likelihood

In synthetic constrained likelihood method, the covariate distribution $F^{\dagger}(X, Z)$ in the external study, different from the distribution $F(X, Z)$ in the internal study, is estimated by a reference sample $(X_j^{\dagger}, Z_j^{\dagger})$, $j = 1, \ldots, N_r$, where N_r is the size of the reference sample. When the internal sample is obtained under the simple random sampling design, the synthetic constrained likelihood is defined as $l_{\beta,\lambda}^{\dagger} = \log(L_{\beta,F}) + \lambda^T \int u_\beta(X, Z; \theta) d\widetilde{F}^{\dagger}(X, Z)$, with $\widetilde{F}^{\dagger}$ the empirical distribution of $(X^{\dagger}, Z^{\dagger})$ in the reference sample, and the synthetic constrained maximum likelihood (SCML) estimator $(\widetilde{\beta}, \widetilde{\lambda})$ for (β, λ) is obtained by solving the estimating equations $\partial\, l_{\beta,\lambda}^{\dagger}/\partial\beta = 0$ and $\partial\, l_{\beta,\lambda}^{\dagger}/\partial\lambda = 0$. The expressions for $s_\beta^{\dagger} = l_{\beta,\lambda}^{\dagger}/\partial\beta$ and $s_\lambda^{\dagger} = l_{\beta,\lambda}^{\dagger}/\partial\lambda$ are:

$$\begin{aligned} s_\beta^{\dagger} = \frac{\partial l_{\beta,\lambda}^{\dagger}}{\partial\beta} &= \sum_{i=1}^{N} s_\beta(Y_i, X_i, Z_i) + \sum_{j=1}^{N_r} \widetilde{s}_\beta^{\dagger}(X_j^{\dagger}, Z_j^{\dagger}), \\ \widetilde{s}_\beta^{\dagger}(X_j, Z_j) &= c_\beta(X_j^{\dagger}, Z_j^{\dagger}; \theta)\lambda, \end{aligned} \tag{S.7}$$

with $s_\beta(Y, X, Z) = \partial\log\{f_\beta(Y|X, Z)\}/\partial\beta$ and $c_\beta(x, z; \theta)$ defined in (S.2), and

$$s_\lambda^{\dagger} = \frac{\partial l_{\beta,\lambda}^{\dagger}}{\partial\lambda} = \sum_{j=1}^{N_r} u_\beta(X_j^{\dagger}, Z_j^{\dagger}; \theta) \equiv \sum_{j=1}^{N_r} s_\lambda^{\dagger}(X_j^{\dagger}, Z_j^{\dagger}). \tag{S.8}$$

The component matrices for $I^{\dagger} = -\partial^2\, l_{\beta,\lambda}^{\dagger}/\partial(\beta^T, \lambda^T)^T \partial(\beta^T, \lambda^T)$ are given as

$$\begin{aligned} I_{\beta\beta}^{\dagger} &= -\frac{\partial^2\, l_{\beta,\lambda}^{\dagger}}{\partial\beta\partial\beta^T} \\ &= \sum_{i=1}^{N} i_{\beta\beta}(Y_i, X_i, Z_i) + \sum_{j=1}^{N_r} \lambda^T d_\beta(X_j^{\dagger}, Z_j^{\dagger}; \theta), \end{aligned}$$

where $i_{\beta\beta}(Y, X, Z) = -\partial^2\log\{f_\beta(Y|X, Z)\}/\partial\beta\partial\beta^T$ and $d_\beta(x, z; \theta)$ is defined in (S.5);

$$I_{\beta\lambda}^{\dagger} = I_{\beta\lambda}^{\dagger T} = -\frac{\partial^2\, l_{\beta,\lambda}^{\dagger}}{\partial\beta\partial\lambda^T} = -\textstyle\sum_{j=1}^{N_r} c_\beta(X_j^{\dagger}, Z_j^{\dagger}; \theta);$$

and $I_{\lambda\lambda}^{\dagger} = -\partial^2\, l_{\beta,\lambda}^{\dagger}/\partial\lambda\partial\lambda^T$ is a $\ell \times \ell$ zero matrix with ℓ the dimension of λ.

When the internal sample is obtained by the case-control sampling design, following Prentice and Pyke (1979), the likelihood contribution

$$L_{\beta,F}^{cc} = \prod_{i=1}^{N} f_\beta(Y_i|X_i, Z_i) dF(X_i, Z_i),$$

with $dF(X_i, Z_i)$ treated nonparametrically is equivalent to the pseudo-likelihood

$$L^{cc}_{\beta,\alpha} = \prod_{i=1}^{N} p_{\beta,\alpha}(Y_i|X_i, Z_i),$$

with α and $p_{\beta,\alpha}(y|x, z)$ defined in equation (8) in the main text. The synthetic constrained likelihood is then defined as $l^{\dagger}_{\beta,\lambda,\alpha} = \log(L^{cc}_{\beta,\alpha}) + \lambda^T \int u_\beta(X, Z; \theta)d\widetilde{F}^{\dagger}(X, Z)$. The nuisance parameter α appears only in the likelihood contribution from the internal sample (first term in $l^{\dagger}_{\beta,\lambda,\alpha}$) and not in the contribution from the reference sample (second term in $l^{\dagger}_{\beta,\lambda,\alpha}$). Therefore, all the results for parameters β and λ derived under the simple random sampling design directly apply, except that in the expressions involving the likelihood contribution from the internal sample, such as $s_\beta(Y, X, Z)$ and $i_{\beta\beta}(Y, X, Z)$, the quantity $f_\beta(Y|X, Z)$ is replaced by $p_{\beta,\alpha}(Y|X, Z)$. The nuisance parameter α is directly solved from

$$0 = s^{\dagger}_{\alpha} = \sum_{i=1}^{N} \frac{\partial}{\partial\alpha}\log\{p_{\beta,\alpha}(Y_i|X_i, Z_i)\} \equiv \sum_{i=1}^{N} s^{\dagger}_{\alpha}(Y_i, X_i, Z_i).$$

The negative Hessian matrix is now defined as $I^{\dagger} = -\partial^2\, l^{\dagger}_{\beta,\lambda,\alpha}/\partial(\beta^T, \lambda^T, \alpha^T)^T\partial(\beta^T, \lambda^T, \alpha^T)$, where

$$I^{\dagger}_{\alpha\alpha} = -\frac{\partial^2\, l^{\dagger}_{\beta,\lambda,\alpha}}{\partial\alpha\partial\alpha^T} = \sum_{i=1}^{N} i_{\alpha\alpha}(Y_i, X_i, Z_i),$$

with $i_{\alpha\alpha}(Y, X, Z) = -\partial^2\log\{p_{\beta,\alpha}(Y|X, Z)\}/\partial\alpha\partial\alpha^T$, and

$$I^{\dagger}_{\alpha\beta} = -\frac{\partial^2\, l^{\dagger}_{\beta,\lambda,\alpha}}{\partial\alpha\partial\beta^T} = \sum_{i=1}^{N} i_{\alpha\beta}(Y_i, X_i, Z_i),$$

with $i_{\alpha\beta}(Y, X, Z) = -\partial^2\log p_{\beta,\alpha}(Y|X, Z)/\partial\alpha\partial\beta^T$. The components

$$I^{\dagger}_{\alpha\lambda} = I^{\dagger T}_{\lambda\alpha} = -\frac{\partial^2\, l^{\dagger}_{\beta,\lambda,\alpha}}{\partial\alpha\partial\beta^T}$$

are zero matrices.

The extension to the stratified case-control design in the internal study, as discussed in Appendix S.2.2, is straightforward by redefining $p_{\beta,\alpha}(Y|X, Z)$ as in (S.6), and generalizing α to a vector of parameters corresponding to the levels of covariates W defining the strata.

S.4 Conditions for the Theoretical Results

Below is the list of conditions employed in the theoretical results.

(i) The true parameter value β_0 is an interior point of the parameter space which is compact.

(ii) The function $u_\beta(x,z;\theta)$ defined in (3) in the main text is twice continuously differentiable in β in a neighborhood $\mathcal{N}$ of β_0, and $E\|u_\beta(X,Z;\theta)\|^3 < \infty$ in $\mathcal{N}$.

(iii) $E\{u_\beta(X,Z;\theta)u_\beta^T(X,Z;\theta)\}$ is positive definite at $\beta=\beta_0$, and

$$N^{-1}\sum_i u_\beta(X_i,Z_i;\theta)u_\beta^T(X_i,Z_i;\theta) \to E\{u_\beta(X,Z;\theta)u_\beta^T(X,Z;\theta)\}$$

in probability uniformly in $\mathcal{N}$.

(iv) $E\|c_\beta(X,Z;\theta)\| < \infty$ in $\mathcal{N}$ with $c_\beta(x,z;\theta)$ defined in (S.2).

(v) $E\|d_\beta(X,Z;\theta)\| < \infty$ in $\mathcal{N}$ with $d_\beta(x,z;\theta)$ defined in (S.5).

S.5 Proofs of Lemmas and Propositions

Proof of Lemma 1:

Let $s_\beta(y,x,z) = \partial \log\{f_\beta(y|x,z)\}/\partial\beta$, and $i_{\beta\beta}(y,x,z) = -\partial^2 \log\{f_\beta(y|x,z)\}/\partial\beta\partial\beta^T$. Write $u_\beta(X_i,Z_i) = u_\beta(X_i,Z_i;\theta)$ since the value of θ is always fixed. Consider the ball $\mathcal{B} = \{\beta : \|\beta-\beta_0\| \le n^{-1/k}\}$ for some integer $k \ge 3$ such that $\mathcal{B} \subset \mathcal{N}$. For a fixed β within $\mathcal{B}$, let $\lambda(\beta)$ be the solution to the $\partial l^*_{\beta,\lambda}/\partial\lambda = 0$ or $\partial l^{*,cc}_{\beta,\lambda}/\partial\lambda = 0$, namely

$$\sum_{i=1}^N \frac{u_\beta(X_i,Z_i)}{1-\lambda(\beta)^T u_\beta(X_i,Z_i)} = 0.$$

Applying Taylor expansion to the left-hand side of the above equation, we obtain

$$\begin{aligned}\lambda(\beta) &= -\left\{N^{-1}\sum_{i=1}^N u_\beta(X_i,Z_i)u_\beta^T(X_i,Z_i)\right\}^{-1}\left\{N^{-1}\sum_{i=1}^N u_\beta(X_i,Z_i)\right\} + o_p(N^{-1/k}) \\ &= O_p(N^{-1/k})\end{aligned}$$

uniformly over $\mathcal{B}$ with probability tending to one as $N\to\infty$. Also, by Taylor expansion, with probability one

$$\sum_{i=1}^N s_\beta(Y_i,X_i,Z_i) = -\left\{N^{-1}\sum_{i=1}^N i_{\beta\beta}(Y_i,X_i,Z_i)\right\}^{-1} N(\beta-\beta_0) + o_p(N^{1-1/k}).$$

Let $Q(\beta) = l^*_{\beta,\lambda(\beta)}$. By Taylor expansion again and using the results above, we have uniformly for β on the surface of $\mathcal{B}$,

$$
\begin{aligned}
& -\sum_{i=1}^{N} \log\{1-\lambda^T(\beta)u_\beta(X_i,Z_i)\} \\
= & \sum_{i=1}^{N} \lambda^T(\beta)u_\beta(X_i,Z_i) + \sum_{i=1}^{N}\{\lambda^T(\beta)u_\beta(X_i,Z_i)\}^2 + o_p(N^{1-2/k}) \\
= & -\frac{N}{2}\left\{N^{-1}\sum_{i=1}^{N} u_\beta(X_i,Z_i)\right\}^T \left\{N^{-1}\sum_{i=1}^{N} u_{\beta_0}(X_i,Z_i)u^T_{\beta_0}(X_i,Z_i)\right\}^{-1} \left\{N^{-1}\sum_{i=1}^{N} u_\beta(X_i,Z_i)\right\} \\
& +o_p(N^{1-2/k}) \\
= & -K_1 N^{1-2/k};
\end{aligned}
$$

$$
\begin{aligned}
& \sum_{i=1}^{N} \log\{f_\beta(Y_1|X_i,Z_i)\} \\
= & \sum_{i=1}^{N} \log\{f_{\beta_0}(Y_1|X_i,Z_i)\} + \sum_{i=1}^{N}(\beta-\beta_0)^T s_\beta(Y_i,X_i,Z-i) \\
& +\frac{N}{2}(\beta-\beta_0)^T\left\{N^{-1}\sum_{i=1}^{N} i_{\beta\beta}(Y_i,X_i,Z_i)\right\}^{-1}(\beta-\beta_0) + o_p(N^{1-2/k}) \\
= & \sum_{i=1}^{N} \log\{f_{\beta_0}(Y_1|X_i,Z_i)\} - \frac{N}{2}(\beta-\beta_0)^T\left\{N^{-1}\sum_{i=1}^{N} i_{\beta\beta}(Y_i,X_i,Z_i)\right\}^{-1}(\beta-\beta_0) + o_p(N^{1-2/k}) \\
= & \sum_{i=1}^{N} \log\{f_{\beta_0}(Y_1|X_i,Z_i)\} - K_2 N^{1-2/k},
\end{aligned}
$$

where K_1 and K_2 are positive constants. On the other hand, it can similarly be shown that $-\sum_i \log\{1-\lambda^T(\beta)u_\beta(X_i,Z_i)\}$ evaluated at β_0 is negative of order $O_p(\text{loglog}N)$. All these results imply that $Q(\beta)$ is maximized at $\widehat{\beta}$ in the interior of $\mathcal{B}$, and $\widehat{\beta}$ and $\widehat{\lambda} = \lambda(\widehat{\beta})$ satisfy $\partial l^*_{\beta,\lambda}/\partial\eta = 0$. The proof for $Q(\beta) = l^{*,cc}_{\beta,\lambda(\beta)}$, i.e. for the case-control setting, is fully parallel.

Proof of Proposition 1: Using the notation given in Appendix S.2, $\widehat{\eta}$ satisfies $\sum_i s^*_i(\beta,\lambda) = 0$ with $s^*_i(\beta,\lambda) = \{s^{*T}_\beta(Y_i,X_i,Z_i), s^{*T}_\lambda(X_i,Z_i)\}^T$, and

$$
I^*(\beta,\lambda) = -\sum_{i=1}^{N} \partial s^*_i(\beta,\lambda)/\partial\eta = \begin{pmatrix} I^*_{\beta\beta} & I^*_{\beta\lambda} \\ I^{*T}_{\beta\lambda} & I^*_{\lambda\lambda} \end{pmatrix}.
$$

Under regularity conditions for $f_\beta(y|x,z)$ and conditions (i)-(v), we have by Taylor expansion and Lemma 1,

$$
\begin{aligned}
0 = N^{-1}\sum_i s_i^*(\widehat{\beta},\widehat{\lambda}) &= N^{-1}\sum_i s_i^*(\beta_0,0) - N^{-1}I^*(\beta_0,0)(\widehat{\eta}-\eta_0) + o_p(N^{-1/2}) \\
&= N^{-1}\sum_i s_i^*(\beta_0,0) - \mathcal{I}^*(\widehat{\eta}-\eta_0) + o_p(N^{-1/2}),
\end{aligned}
$$

where

$$
\mathcal{I}^* = \begin{pmatrix} E\{i_{\beta\beta}(Y,X,Z)\} & -E\{c_\beta(X,Z;\theta)\} \\ -E\{c_\beta^T(X,Z;\theta)\} & -E\{u_\beta(X,Z)u_\beta^T(X,Z)\} \end{pmatrix}
$$

is evaluated at $\beta=\beta_0$, with $i_{\beta\beta}(Y,X,Z)$, $c_\beta(X,Z;\theta)$, and $u_\beta(X,Z;\theta)$ defined in (S.4), (S.2), and (3) of the main text, respectively. Hence,

$$
\sqrt{N}(\widehat{\eta}-\eta_0) = \mathcal{I}^{*-1}\sqrt{N}\left\{N^{-1}\sum_i s_i^*(\beta_0,0)\right\} + o_p(1).
$$

By the fact that $E\{s_{\beta_0}(Y,X,Z)\}=0$ and $E\{u_{\beta_0}(X,Z)\}=0$, $E\{s_i^*(\beta_0,0)\}=0$. Also, from expressions given in Appendix S.2, $\mathrm{var}\{s_i^*(\beta_0,0)\} = E\{s_i^*(\beta_0,0)s_i^{*T}(\beta_0,0)\} = \mathcal{J}^*$, where

$$
\mathcal{J}^* = \begin{pmatrix} E\{i_{\beta\beta}(Y,X,Z)\} & O \\ O & E\{u_\beta(X,Z)u_\beta^T(X,Z)\} \end{pmatrix},
$$

evaluated at $\beta=\beta_0$, By central limit theorem we then have, as $N\to\infty$, $\sqrt{N}(\widehat{\eta}-\eta_0)$ follows asymptotically the normal distribution with mean zero and covariance matrix $\mathcal{I}^{*-1}\mathcal{J}^*\mathcal{I}^{*-1}$. Further matrix calculation simplifies the expression for the asymptotic variance given in Proposition 1.

When the uncertainty in the parameter θ of the external model cannot be ignored because it is estimated from a finite external sample, the variance estimator for $\widehat{\eta}$ needs to be modified to account for such additional uncertainty, which can be simply achieved using the conventional delta method. Let $\sigma_{\eta\theta} = E\{\partial s_i^*(\beta,\lambda)/\partial\theta^T\} = E\{\partial\widetilde{s}_\beta^T(X_i,Z_i)/\partial\theta, \partial s_\lambda^{*T}(X_i,Z_i)/\partial\theta\}^T$, which can be readily obtained given the explicit formulas in (S.1) and (S.3) for $\widetilde{s}_\beta(X_i,Z_i)$ and $s_\lambda^*(X_i,Z_i)$. Let v_θ/N_e be the variance of the estimate of θ, with N_e the size of the external sample. Then, by δ-method, $\sqrt{N}(\widehat{\eta}-\eta_0)$ has the asymptotic covariance matrix $\mathcal{I}^{*-1}(\mathcal{J}^*+\rho\mathcal{V}^*)\mathcal{I}^{*-1}$, where $\mathcal{V}^*=\sigma_{\eta\theta}v_\theta\sigma_{\eta\theta}^T$ with $\rho=\lim N/N_e$. Accordingly, an estimator for the variance of $\widehat{\eta}$ can be obtained as

$$
\widehat{\mathcal{I}}^{*-1}(\widehat{\mathcal{J}}^*/N + \widehat{\mathcal{V}}^*/N_e)\widehat{\mathcal{I}}^{*-1},
$$

where $\widehat{\mathcal{A}}$ denotes the empirical analogue of $\mathcal{A}$ evaluated at $\widehat{\eta}$.

Proof of Lemma 2: Since the proofs for the simple random and case-control designs are essentially the same, we detail the former one only. Write $\partial l^{\dagger}_{\beta,\lambda}/\partial\beta = \sum_{i=1}^{N} s^{\dagger}_{i}(\beta) + \sum_{j=1}^{N_r} \widetilde{s}^{\dagger}_{j}(\beta,\lambda)$, where $s^{\dagger}_{i}(\beta) = \{s^{\dagger T}_{\beta}(Y_i, X_i, Z_i), 0^T\}^T$, with 0 a ℓ-dimensional vector of zeros, and $\widetilde{s}^{\dagger}_{j}(\beta,\lambda) = \{\widetilde{s}^{\dagger T}_{\beta}(X^{\dagger}_{j}, Z^{\dagger}_{j}), s^{\dagger T}_{\lambda}(X^{\dagger}_{j}, Z^{\dagger}_{j})\}^T$. Explicit expressions for these functions are given in Appendix S.3. Thus $0 = \sum_{i=1}^{N} s^{\dagger}_{i}(\widetilde{\beta}) + \sum_{j=1}^{N_r} \widetilde{s}^{\dagger}_{j}(\widetilde{\beta}, \widetilde{\lambda})$.

Consider the ball $\mathcal{B}$ contained in the neighborhood of β_0 as defined in the proof of Lemma 1. In the following we suppress the dependence on θ in $c_\beta(X, Z; \theta)$ since θ is always fixed. By Taylor expansion around the point $(\beta^*, 0)$ with $\beta^* \in \mathcal{B}$ and 0 a ℓ-vector of zeros (ℓ = dimension of λ), we have

$$0 = \sum_{i=1}^{N} s^{\dagger}_{i}(\beta^*) + \sum_{j=1}^{N_r} \widetilde{s}^{\dagger}_{j}(\beta^*, 0) - I^{\dagger}_{\beta\beta}(\beta - \beta^*) - I^{\dagger T}_{\beta\lambda}(\beta - \beta^*) - I^{\dagger}_{\beta\lambda}\lambda,$$

with (β, λ) lying between $(\widetilde{\beta}, \widetilde{\lambda})$ and $(\beta^*, 0)$, where $I^{\dagger}_{\beta\beta}$ and $I^{\dagger}_{\beta\lambda}$ are negative Hessian matrices whose expressions are given in Appendix S.3, and they are evaluated at (β, λ). Accordingly, we have

$$\begin{pmatrix} \widehat{\beta} - \beta_0 \\ \widehat{\lambda} \end{pmatrix} = \begin{pmatrix} H_{\beta\beta} & H_{\beta\lambda} \\ H^{T}_{\beta\lambda} & O \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{N} s^{\dagger T}_{\beta}(Y_i, X_i, Z_i) \\ \sum_{j=1}^{N_r} s^{\dagger}_{\lambda}(X^{\dagger}_{j}, Z^{\dagger}_{j}) \end{pmatrix} + o_p(N^{-1/k}),$$

with $H_{\beta\beta} = \sum_{i=1}^{N} i_{\beta\beta}(Y_i, X_i, Z_i)$ and $H_{\beta\lambda} = -\sum_{j=1}^{N_r} c_\beta(X^{\dagger}_{j}, Z^{\dagger}_{j})$. So,

$$\lambda = \mathcal{H} H^{T}_{\beta\lambda} H^{-1}_{\beta\beta} \sum_{i=1}^{N} s^{\dagger T}_{\beta}(Y_i, X_i, Z_i) - \mathcal{H} \sum_{j=1}^{N_r} \widetilde{s}^{\dagger}_{\lambda}(X^{\dagger}_{j}, Z^{\dagger}_{j}),$$

with $\mathcal{H} = (H^{T}_{\beta\lambda} H^{-1}_{\beta\beta} H_{\beta\lambda})^{-1}$. As $N \to \infty$, the limiting value of the matrix $\mathcal{H}$ is of order $O_p(N_r^{-2} N)$, and that of $H^{T}_{\beta\lambda} H^{-1}_{\beta\beta}$ is $O_p(N_r N^{-1})$ and hence λ is of order $O(N^{-(1/k)}\kappa^{-1})$ with probability one uniformly for $\beta \in \mathcal{B}$.

Recalling the sufficient conditions for constrained maximality in the Lagrange multiplier theory, together with the condition $q > \ell$ (Chiang and Wainwright, 1984, p. 385), we can then conclude the results of this lemma by noting that, with probability tending to one, $\partial^2 l^{\dagger}_{\beta,\lambda}/\partial\beta\partial\beta^T = -I^{\dagger}_{\beta\beta}$ is negative definite by the fact that $\lambda \to 0$ and the regularity condition that $\partial^2 \log\{f_\beta(y|x, z)\}/\partial\beta\partial\beta^T$ is negative definite for $\beta \in \mathcal{B}$.

Proof of Proposition 2: Under the conditions specified, the Taylor expansion yields

$$0 = \sum_{i=1}^{N} s^{\dagger}_{i}(\beta_0) + \sum_{j=1}^{N_r} \widetilde{s}^{\dagger}_{j}(\beta_0, 0) - I^{\dagger}_{\beta\beta}(\widehat{\beta} - \beta_0) - I^{\dagger T}_{\beta\lambda}(\widehat{\beta} - \beta_0) - I^{\dagger}_{\beta\lambda}\widehat{\lambda} + o_p(N^{1/2} + N_r^{1/2}),$$

with $I^\dagger_{\beta\beta}$ and $I^\dagger_{\beta\lambda}$ evaluated at $(\beta_0, 0)$. Hence,

$$\begin{aligned}
\sqrt{N}\begin{pmatrix} \widehat{\beta}-\beta_0 \\ \widehat{\lambda} \end{pmatrix} &= \begin{pmatrix} \frac{1}{N}\sum_{i=1}^N i_{\beta\beta}(Y_i, X_i, Z_i) & -\frac{N_r}{N}\frac{1}{N_r}\sum_{j=1}^{N_r} c_\beta(X_j^\dagger, Z_j^\dagger) \\ -\frac{N_r}{N}\frac{1}{N_r}\sum_{j=1}^{N_r} c_\beta^T(X_j^\dagger, Z_j^\dagger) & O \end{pmatrix}^{-1} \\
&\quad \times \begin{pmatrix} \frac{1}{\sqrt{N}}\sum_{i=1}^N s_\beta^{\dagger T}(Y_i, X_i, Z_i) \\ \sqrt{\frac{N_r}{N}}\frac{1}{\sqrt{N_r}}\sum_{j=1}^{N_r} s_\lambda^\dagger(X_j^\dagger, Z_j^\dagger) \end{pmatrix} + o_p(1) \\
&= \mathcal{I}^{\dagger -1}\mathcal{Z}_N + o_p(1),
\end{aligned}$$

with

$$\mathcal{I}^\dagger = \begin{pmatrix} B & -\kappa C \\ -\kappa C^T & O \end{pmatrix},$$

and

$$\mathcal{Z}_N = \begin{pmatrix} \frac{1}{\sqrt{N}}\sum_{i=1}^N s_\beta^{\dagger T}(Y_i, X_i, Z_i) \\ \frac{\sqrt{\kappa}}{\sqrt{N_r}}\sum_{j=1}^{N_r} s_\lambda^\dagger(X_j^\dagger, Z_j^\dagger) \end{pmatrix},$$

which, following the fact $E\{s_\beta^{\dagger T}(Y, X, Z)\} = 0$ and $E^\dagger\{s_\lambda^\dagger(X^\dagger, Z^\dagger)\} = 0$, converges in distribution to a mean-zero normal distribution with covariance matrix

$$\mathcal{J}^\dagger = \begin{pmatrix} B & O \\ O & \kappa L \end{pmatrix}.$$

Accordingly, as $N \to \infty$, $\sqrt{N}(\widetilde{\eta} - \eta_0)$ follows a normal distribution with mean zero and covariance matrix $\mathcal{I}^{\dagger -1}\mathcal{J}^\dagger\mathcal{I}^{\dagger -1}$, which by matrix calculation reduces to the expression given in Proposition 2.

When the uncertainty in the parameter θ of the external model needs to be accounted for, we modify the variance estimator for $\widetilde{\eta}$ using the δ-method. Let $\sigma^\dagger_{\eta\theta} = E\{\partial\widetilde{s}_j^\dagger(\beta, \lambda)/\partial\theta^T\} = E\{\partial\widetilde{s}_\beta^{\dagger T}(X_j^\dagger, Z_j^\dagger)/\partial\theta, \partial s_\lambda^{\dagger T}(X_j^\dagger, Z_j^\dagger)/\partial\theta\}^T$, which can be readily obtained given the formulas in (S.7) and (S.8) for $\widetilde{s}_\beta^\dagger(X_j^\dagger, Z_j^\dagger)$ and $s_\lambda^\dagger(X_j^\dagger, Z_j^\dagger)$. Let v_θ/N_e be the variance of the estimate of θ, with N_e the size of the external sample. By the delta method, the asymptotic covarinace matrix of $\mathcal{Z}_N$ is

$$\mathcal{K}^\dagger = \begin{pmatrix} B & O \\ O & \kappa(L + \rho^\dagger v_\theta) \end{pmatrix},$$

where $\rho^\dagger = \lim N_r/N_e$. Therefore, $\sqrt{N}(\widetilde{\eta} - \eta_0)$ has the asymptotic covariance matrix $\mathcal{I}^{\dagger -1}\mathcal{K}^\dagger\mathcal{I}^{\dagger -1}$. Accordingly, an estimator for the variance of $\widetilde{\eta}$ can be obtained as

$$N^{-1}\widetilde{\mathcal{I}}^{\dagger -1}\widetilde{\mathcal{K}}^\dagger\widetilde{\mathcal{I}}^{\dagger -1},$$

where $\widetilde{\mathcal{A}}$ denotes the empirical analogue of $\mathcal{A}$ evaluated at $\widetilde{\eta}$.

REFERENCES

Chiang, A. C., Wainwright, K. (1984), *Fundamental Methods of Mathematical Economics*, 3rd ed., New York: McGraw-Hill.