Supporting Information to Instrument Assisted Regression for Errors in Variables Models with Binary Response

KUN XU
Department of Statistics, Texas A&M University
YANYUAN MA
Department of Statistics, Texas A&M University
LIQUN WANG
Department of Statistics, University of Manitoba

March 5, 2014

Appendix

A.1 Derivation of Λ

If we consider the parametric submodel,

$$\begin{aligned} & \mathrm{pr}(Y=y,\mathbf{S}=\mathbf{s},\mathbf{Z}=\mathbf{z};\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\eta}_1,\boldsymbol{\eta}_2) \\ = & \int \mathrm{pr}(Y=y\mid \mathbf{S},\mathbf{Z},\boldsymbol{\epsilon};\boldsymbol{\beta},\gamma) f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}\mid \mathbf{s},\mathbf{z},\boldsymbol{\eta}_2) f_{\mathbf{S},\mathbf{Z}}(\mathbf{s},\mathbf{z};\boldsymbol{\eta}_1) d\mu(\boldsymbol{\epsilon}), \end{aligned}$$

the nuisance score vectors with respect to $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ are $\partial \log f_{\mathbf{S},\mathbf{Z}}(\mathbf{s},\mathbf{z};\boldsymbol{\eta}_1)/\partial\boldsymbol{\eta}_1$ and $E\{\partial \log f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}\mid \mathbf{s},\mathbf{z},\boldsymbol{\eta}_2)/\partial\boldsymbol{\eta}_2 \mid Y,\mathbf{S},\mathbf{Z}\}$ respectively. The former has the property

$$E\{\partial \log f_{\mathbf{S},\mathbf{Z}}(\mathbf{s},\mathbf{z};\boldsymbol{\eta}_1)/\partial\boldsymbol{\eta}_1\} = \mathbf{0},$$

and $\partial \log f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}\mid \mathbf{s},\mathbf{z},\boldsymbol{\eta}_2)/\partial\boldsymbol{\eta}_2$ satisfies

$$\begin{aligned} & E\{\partial \log f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}\mid \mathbf{s},\mathbf{z},\boldsymbol{\eta}_2)/\partial\boldsymbol{\eta}_2 \mid \mathbf{S},\mathbf{Z}\} = \mathbf{0}, \\ & E[\{\partial \log f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}\mid \mathbf{s},\mathbf{z},\boldsymbol{\eta}_2)/\partial\boldsymbol{\eta}_2\}\boldsymbol{\epsilon}^{\mathrm{T}} \mid \mathbf{S},\mathbf{Z}] = \mathbf{0}. \end{aligned}$$

The last equation comes from the condition that $E(\boldsymbol{\epsilon}\mid \mathbf{S},\mathbf{Z}) = 0$. This completes the nuisance tangent space derivation for a parametric submodel. Since the nuisance tangent space of our original model is the mean square closure of the nuisance tangent space of all parametric submodels, the conjecture for the desired nuisance tangent space is the direct sum of two subspaces Λ_1 and Λ_2, where

$$\begin{aligned}\Lambda_1 &= \{\mathbf{f}(\mathbf{S},\mathbf{Z}) : \mathbf{f} \in \mathbb{R}^p, E(\mathbf{f}) = \mathbf{0}, E(\mathbf{f}^{\mathrm{T}}\mathbf{f}) < \infty\} \\ \Lambda_2 &= [E\{\mathbf{f}(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}) \mid Y,\mathbf{S},\mathbf{Z}\} : \mathbf{f} \in \mathbb{R}^p, E(\mathbf{f} \mid \mathbf{S},\mathbf{Z}) = \mathbf{0}, E(\boldsymbol{\epsilon}\mathbf{f}^{\mathrm{T}} \mid \mathbf{S},\mathbf{Z}) = \mathbf{0}, E(\mathbf{f}^{\mathrm{T}}\mathbf{f}) < \infty].\end{aligned}$$

In the second part of the proof, we must show that for any bounded random functions $\mathbf{f}_1(\mathbf{S},\mathbf{Z}) \in \Lambda_1$ and $E\{\mathbf{f}_2(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}) \mid Y,\mathbf{S},\mathbf{Z}\} \in \Lambda_2$, they are the nuisance score vectors of a particular parametric submodel. When the true models for $f_{\mathbf{S},\mathbf{Z}}(\mathbf{s},\mathbf{z})$ and $f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon} \mid \mathbf{s},\mathbf{z})$ are $f_0(\mathbf{s},\mathbf{z})$ and $f_0(\boldsymbol{\epsilon} \mid \mathbf{s},\mathbf{z})$ respectively, we define new functions with the aid of $\mathbf{f}_1(\mathbf{S},\mathbf{Z})$ and $\mathbf{f}_2(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z})$ such that

$$\begin{aligned}f_{\mathbf{S},\mathbf{Z}}(\mathbf{s},\mathbf{z};\boldsymbol{\eta}_1) &= f_0(\mathbf{s},\mathbf{z})\{1 + \boldsymbol{\eta}_1^{\mathrm{T}}\mathbf{f}_1(\mathbf{S},\mathbf{Z})\} \\ f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon} \mid \mathbf{s},\mathbf{z},\boldsymbol{\eta}_2) &= f_0(\boldsymbol{\epsilon} \mid \mathbf{s},\mathbf{z})[1 + \boldsymbol{\eta}_2^{\mathrm{T}}\mathbf{f}_2(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z})].\end{aligned}$$

$\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ must be sufficiently small such that

$$1 + \boldsymbol{\eta}_1^{\mathrm{T}}\mathbf{f}_1(\mathbf{S},\mathbf{Z}) \geq 0, \text{ and } 1 + \boldsymbol{\eta}_2^{\mathrm{T}}\mathbf{f}_2(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}) \geq 0.$$

Both $f_{\mathbf{S},\mathbf{Z}}(\mathbf{s},\mathbf{z};\boldsymbol{\eta}_1)$ and $f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon} \mid \mathbf{s},\mathbf{z},\boldsymbol{\eta}_2)$ are valid probability density function because they are positive and their integration from negative infinity to infinity are 1, as can be seen below.

$$\begin{aligned}\int\int f_{\mathbf{S},\mathbf{Z}}(\mathbf{s},\mathbf{z};\boldsymbol{\eta}_1)d\mu(\mathbf{s},\mathbf{z}) &= \int\int f_0(\mathbf{s},\mathbf{z})d\mu(\mathbf{s},\mathbf{z}) + \int\int f_0(\mathbf{s},\mathbf{z})\boldsymbol{\eta}_1^{\mathrm{T}}\mathbf{f}_1(\mathbf{S},\mathbf{Z})d\mu(\mathbf{s},\mathbf{z}) \\ &= 1 + \boldsymbol{\eta}_1^{\mathrm{T}}E\{\mathbf{f}_1(\mathbf{S},\mathbf{Z})\} = 1, \\ \int f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon} \mid \mathbf{s},\mathbf{z},\boldsymbol{\eta}_2)d\mu(\boldsymbol{\epsilon}) &= \int f_0(\boldsymbol{\epsilon} \mid \mathbf{s},\mathbf{z})d\mu(\boldsymbol{\epsilon}) + \int f_0(\boldsymbol{\epsilon} \mid \mathbf{s},\mathbf{z})\boldsymbol{\eta}_2^{\mathrm{T}}\mathbf{f}_2(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z})d\mu(\boldsymbol{\epsilon}) \\ &= 1 + \boldsymbol{\eta}_2^{\mathrm{T}}E\{\mathbf{f}_2(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}) \mid \mathbf{S},\mathbf{Z}\} = 1.\end{aligned}$$

Moreover,

$$\begin{aligned}\int f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon} \mid \mathbf{s},\mathbf{z},\boldsymbol{\eta}_2)\boldsymbol{\epsilon}^{\mathrm{T}}d\mu(\boldsymbol{\epsilon}) &= \int f_0(\boldsymbol{\epsilon} \mid \mathbf{s},\mathbf{z})\boldsymbol{\epsilon}^{\mathrm{T}}d\mu(\boldsymbol{\epsilon}) + \int f_0(\boldsymbol{\epsilon} \mid \mathbf{s},\mathbf{z})\boldsymbol{\eta}_2^{\mathrm{T}}\mathbf{f}_2(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z})\boldsymbol{\epsilon}^{\mathrm{T}}d\mu(\boldsymbol{\epsilon}) \\ &= 0 + \boldsymbol{\eta}_2^{\mathrm{T}}E\{\mathbf{f}_2(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z})\boldsymbol{\epsilon}^{\mathrm{T}} \mid \mathbf{S},\mathbf{Z}\} = 0.\end{aligned}$$

So the density for $\boldsymbol{\epsilon}$ given $\mathbf{S}$ and $\mathbf{Z}$ also satisfies $E(\boldsymbol{\epsilon} \mid \mathbf{S},\mathbf{Z}) = \mathbf{0}$. One the other hand, the score vectors for the parametric submodel are

$$\begin{aligned}
\mathcal{S}_{\eta_1} &= \frac{\partial \log f_{\mathbf{S},\mathbf{Z}}(\mathbf{s},\mathbf{z};\boldsymbol{\eta}_1)}{\partial \boldsymbol{\eta}_1} = \mathbf{f}_1(\mathbf{S},\mathbf{Z}),\\
\mathcal{S}_{\eta_2} &= \int \frac{\partial \log f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon} \mid \mathbf{s},\mathbf{z},\boldsymbol{\eta}_2)}{\partial \boldsymbol{\eta}_2} f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon} \mid Y=y,\mathbf{s},\mathbf{z})d\mu(\boldsymbol{\epsilon}) = E\{\mathbf{f}_2(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}) \mid Y,\mathbf{S},\mathbf{Z}\}.
\end{aligned}$$

This leads to the result.

A.2 Derivation of $\Lambda^{\perp}$

Using the form of the nuisance tangent space, it can be shown that

$$\Lambda = [E\{\mathbf{f}(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}) \mid Y,\mathbf{S},\mathbf{Z}\} : \mathbf{f} \in \mathbb{R}^p, E(\boldsymbol{\epsilon}\mathbf{f}^{\mathrm{T}} \mid \mathbf{S},\mathbf{Z}) = \mathbf{0}, E(\mathbf{f}^{\mathrm{T}}\mathbf{f}) < \infty].$$

Therefore, any element $g(Y,\mathbf{S},\mathbf{Z}) \in \Lambda^{\perp}$ must satisfy

$$\begin{aligned}
0 &= E[g^{\mathrm{T}}(Y,\mathbf{S},\mathbf{Z})E\{\mathbf{f}(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}) \mid Y,\mathbf{S},\mathbf{Z}\}]\\
&= E[E\{g^{\mathrm{T}}(Y,\mathbf{S},\mathbf{Z})\mathbf{f}(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}) \mid \boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}\}]\\
&= E[E\{g^{\mathrm{T}}(Y,\mathbf{S},\mathbf{Z}) \mid \boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}\}f(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z})]
\end{aligned}$$

for any $f(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z})$ such that $E(\boldsymbol{\epsilon}\mathbf{f}^{\mathrm{T}} \mid \mathbf{S},\mathbf{Z}) = \mathbf{0}$. Therefore, $E\{g(Y,\mathbf{S},\mathbf{Z}) \mid \boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}\}$ must have the form $\mathbf{a}(\mathbf{S},\mathbf{Z})\boldsymbol{\epsilon}$ such that $E(\mathbf{a}^{\mathrm{T}}\mathbf{a}) < \infty$. This yields the desired result.

A.3 Proof of Theorem 1

Note that even under a possibly incorrect working model, we have

$$\begin{aligned}
& E\{\mathcal{S}^*_{\mathrm{eff}}(Y_i,\mathbf{S}_i,\mathbf{Z}_i,\boldsymbol{\theta})\}\\
=\ & E[E\{\mathcal{S}^*_{\mathrm{eff}}(Y_i,\mathbf{S}_i,\mathbf{Z}_i,\boldsymbol{\theta}) \mid \boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}\}]\\
=\ & E(E[\mathcal{S}^*_{\theta}(Y,\mathbf{S},\mathbf{Z}) - E^*\{\mathbf{b}(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}) \mid Y,\mathbf{S},\mathbf{Z}\} \mid \boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}])\\
=\ & E\left(E\left[\mathcal{S}^*_{\theta}(Y,\mathbf{S},\mathbf{Z})\boldsymbol{\epsilon}^{\mathrm{T}} - E^*\{\mathbf{b}(\boldsymbol{\epsilon},\mathbf{S},\mathbf{Z}) \mid Y,\mathbf{S},\mathbf{Z}\}\boldsymbol{\epsilon}^{\mathrm{T}} \mid \mathbf{S},\mathbf{Z}\right]\left\{E^*(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}} \mid \mathbf{S},\mathbf{Z})\right\}^{-1}\boldsymbol{\epsilon}\right)\\
=\ & \mathbf{0},
\end{aligned}$$

which implies that the corresponding estimator $\boldsymbol{\theta}$ is consistent. In the above display, the second equality is due to the construction of $\mathcal{S}_{\mathrm{eff}}$, the third equality is because $\mathbf{b}$ satisfies the integral equation (7), and the last equality is because $E(\boldsymbol{\epsilon} \mid \mathbf{S},\mathbf{Z}) = \mathbf{0}$. Standard Taylor expansion then yields the desired result. Namely,

$$\begin{aligned} n^{-1/2}\sum_{i=1}^{n} \mathcal{S}^*_{\rm eff}(Y_i,\mathbf{S}_i,\mathbf{Z}_i,\widehat{\boldsymbol\theta}) &= n^{-1/2}\sum_{i=1}^{n} \mathcal{S}^*_{\rm eff}(Y_i,\mathbf{S}_i,\mathbf{Z}_i,\boldsymbol\theta) \\ &+ n^{-1/2}\frac{\partial}{\partial\boldsymbol\theta^{\rm T}}\left\{\sum_{i=1}^{n} \mathcal{S}^*_{\rm eff}(Y_i,\mathbf{S}_i,\mathbf{Z}_i,\boldsymbol\theta)\right\}(\widehat{\boldsymbol\theta}-\boldsymbol\theta)+o_p(1). \end{aligned}$$

The left side of the equation is zero by observing since $\widehat{\boldsymbol\theta}$ is the solution of the estimating equation. It implies,

$$\begin{aligned} &\sqrt{n}(\widehat{\boldsymbol\theta}-\boldsymbol\theta) \\ =& \left\{-\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol\theta^{\rm T}} \mathcal{S}^*_{\rm eff}(Y_i,\mathbf{S}_i,\mathbf{Z}_i,\boldsymbol\theta)\right\}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathcal{S}^*_{\rm eff}(Y_i,\mathbf{S}_i,\mathbf{Z}_i,\boldsymbol\theta)\right\}+o_p(1) \\ =& -\mathbf{A}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathcal{S}^*_{\rm eff}(Y_i,\mathbf{S}_i,\mathbf{Z}_i,\boldsymbol\theta)\right\}+o_p(1). \end{aligned}$$

The asymptotic result in Theorem 1 is deduced by implementing the central limit theorem. Furthermore, if the true model $f_{\boldsymbol\epsilon}(\boldsymbol\epsilon \mid \mathbf{S},\mathbf{Z})$ is used, the variance will achieve the minimum semiparametric bound because

$$\begin{aligned} E\left\{\frac{\partial}{\partial\boldsymbol\theta^{\rm T}} \mathcal{S}_{\rm eff}(Y_i,\mathbf{S}_i,\mathbf{Z}_i)\right\} &= -E\left\{ \mathcal{S}_{\rm eff}(Y_i,\mathbf{S}_i,\mathbf{Z}_i)\ \mathcal{S}^{\rm T}_{\boldsymbol\theta}(Y_i,\mathbf{S}_i,\mathbf{Z}_i)\right\} \\ &= -E\left\{ \mathcal{S}_{\rm eff}(Y_i,\mathbf{S}_i,\mathbf{Z}_i)\ \mathcal{S}^{\rm T}_{\rm eff}(Y_i,\mathbf{S}_i,\mathbf{Z}_i)\right\}. \end{aligned}$$

The last equation is true since $\mathcal{S}_{\rm eff}$ is the projection of $\mathcal{S}_{\boldsymbol\theta}$ onto the space $\Lambda^{\perp}$. It means that $\mathbf{A}=-\mathbf{B}$, and the variance becomes $\mathbf{B}=[E\{ \mathcal{S}_{\rm eff}(Y,\mathbf{S},\mathbf{Z})^{\otimes 2}\}]^{-1}$. □

A.4 Proof of Theorem 2

Consider the joint estimating equation

$$\begin{aligned} \sum_{i=1}^{n} \mathcal{S}^*_{\rm eff}(Y_i,\mathbf{S}_i,\mathbf{Z}_i,\boldsymbol\theta,\boldsymbol\alpha) &= \mathbf{0} \\ \sum_{i=1}^{n} \mathcal{S}_{\boldsymbol\alpha}(Y_i,\mathbf{S}_i,\mathbf{Z}_i,\boldsymbol\alpha) &= \mathbf{0} \end{aligned}$$

for estimating $\boldsymbol{\alpha}, \boldsymbol{\theta}$ simultaneous, the Taylor expansion yields

$$\begin{aligned}
\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \\ \widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \end{pmatrix} &= -\left\{\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \frac{\partial}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\mathcal{S}_{\mathrm{eff}}^{*} & \frac{\partial}{\partial\boldsymbol{\alpha}^{\mathrm{T}}}\mathcal{S}_{\mathrm{eff}}^{*} \\ \frac{\partial}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\mathcal{S}_{\boldsymbol{\alpha}} & \frac{\partial}{\partial\boldsymbol{\alpha}^{\mathrm{T}}}\mathcal{S}_{\boldsymbol{\alpha}} \end{pmatrix}\right\}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \mathcal{S}_{\mathrm{eff}}^{*} \\ \mathcal{S}_{\boldsymbol{\alpha}} \end{pmatrix}\right\} + o_p(1) \\
&= -\begin{pmatrix} \mathbf{A} & \mathbf{A}_1 \\ \mathbf{0} & \mathbf{A}_2 \end{pmatrix}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \mathcal{S}_{\mathrm{eff}}^{*} \\ \mathcal{S}_{\boldsymbol{\alpha}} \end{pmatrix}\right\} + o_p(1).
\end{aligned}$$

It indicates the normal limiting distribution with variance

$$\begin{pmatrix} \mathbf{A} & \mathbf{A}_1 \\ \mathbf{0} & \mathbf{A}_2 \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{B} & \mathbf{B}_1^{\mathrm{T}} \\ \mathbf{B}_1 & \mathbf{B}_2 \end{pmatrix}\begin{pmatrix} \mathbf{A} & \mathbf{A}_1 \\ \mathbf{0} & \mathbf{A}_2 \end{pmatrix}^{-\mathrm{T}},$$

by the central limit theorem. The $(1,1)$th cell of the resulting matrix by expanding the above expression is $\mathbf{V} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^{\mathrm{T}} + \mathbf{V}_{\alpha}$ where

$$\mathbf{V}_{\alpha} = \mathbf{A}^{-1}\left\{\mathbf{A}_1\mathbf{A}_2^{-1}\mathbf{B}_2(\mathbf{A}_1\mathbf{A}_2^{-1})^{\mathrm{T}} - \mathbf{A}_1\mathbf{A}_2^{-1}\mathbf{B}_1 - (\mathbf{A}_1\mathbf{A}_2^{-1}\mathbf{B}_1)^{\mathrm{T}}\right\}(\mathbf{A}^{-1})^{\mathrm{T}}.$$

When $f_{\boldsymbol{\epsilon}}^{*}(\boldsymbol{\epsilon} \mid \mathbf{S}, \mathbf{Z}) = f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon} \mid \mathbf{S}, \mathbf{Z})$, $-\mathbf{A} = \mathbf{B}$. And the resulting estimation variance is minimized. □