Chapter 7: Asymptotic theory for the semiparametric one-step estimator

May 11, 2022

- Consider a semiparametric model for the law F of a random vector Z,

$$\mathcal{F} = \{F_{\eta,\vartheta} : \eta \in \Xi, \vartheta \in O\}$$

where both η and ϑ can be infinite dimensional.

- Suppose for each $(\eta, \vartheta) \in \Xi \times O$, $\beta(F)$ is a pathwise differentiable parameter at $F_{\eta,\vartheta}$ with respect to some class $\mathcal{A}$ in model $\mathcal{F}$.
- Let $\psi_{F_{\eta,\vartheta}}(Z)$ be a gradient of $\beta\left(F_{\eta,\vartheta}\right)$ at $F_{\eta,\vartheta}$.
- Suppose that both $\beta\left(F_{\eta,\vartheta}\right)$ and $\psi_{F_{\eta,\vartheta}}(Z)$ depend on (η, ϑ) only through η, so for short, we write them respectively, as $\beta(\eta)$ and $\psi(Z;\eta)$.

- **Example:** To fix ideas consider the missing data example of ch 6. In this example $Z = \left(Y, R, X^T\right)^T$ where R is a binary r.v., Y is a scalar r.v. which to simplify we will assume is continuous, and X is a random vector with discrete and/or continuous components. The parameter is

$$\beta(F) \equiv E_F\left[E_F(Y|R=1,X)\right]$$

Note this parameter depends only on

$$b_F(X) \equiv E_F(Y|R=1,X=\cdot) \text{ and } f_X(\cdot)$$

In ch 3 we considered inference in three models, namely

  - $$\mathcal{F}_{np} = \left\{F : E_F\left[\left\{E_F(Y|R=1,X)\right\}^2\right] < \infty, P_F(R=1|X) > \sigma_F > 0\right\},$$
  - $$\mathcal{F}_{sem,fixed} = \{F \in \mathcal{F}_{np} : P_F(R=1|X) = \pi^*(X)\}$$
    where $\pi^*(x)$ is specified, i.e. known , and
  - $$\mathcal{F}_{sem,par} = \{F \in \mathcal{F}_{np} : P_F(R=1|X) = \pi(X;\alpha),\ \alpha \in \Xi \subseteq \mathbb{R}^r\}$$
    where $\pi(x;\alpha)$ is a specified, i.e. known, function of x and α which is differentiable wrt α at every x.

- We saw that in the three models, the efficient influence function at F was

$$\psi_F(Y,R,X) = E_F(Y|R=1,X) + \frac{R\{Y - E_F(Y|R=1,X)\}}{P_F(R=1|X)} - \beta(F)$$

Note that $\psi_F(Y,R,X)$ depends on F only through

$$b_F(\cdot) \equiv E_F(Y|R=1,X=\cdot) \text{ and } \pi_F(\cdot) \equiv P_F(R=1|X=\cdot)$$

- Thus, if we define

$$\eta \equiv (b_F, \pi_F, F_X) \text{ and } \vartheta \equiv \left(F_{\varepsilon|R=1,X}\right)$$

where $\varepsilon \equiv Y - E_F(Y|R=1,X)$, then the assumptions in the previous slide hold for the model $\mathcal{F}_{np}$, the parameter $\beta(F)$ and the gradient $\psi_F(Y,R,X)$.

- Returning now to the general formulation, we will now study a general estimation strategy for $\beta(\eta)$. We will evaluate in particular, a set of conditions under which our strategy yields a RAL estimator of $\beta(\eta)$ and in particular, a set of condition under which the influence function is equal to a given gradient of $\beta(\eta)$, say $\psi(Z;\eta)$.
- We will consider a generalization to semiparametric inference, of the so-called one-step estimator.
- The one step procedure will require a preliminary estimator of η.
- To avoid certain regularity conditions requirements, we will consider a sample split estimation strategy for computing η.
- Specifically, given $Z_1, ..., Z_n \overset{iid}{\sim} f_{\eta,\vartheta}$ if

$$\widetilde{\eta}_n \equiv \widetilde{\eta}_n(Z_1, ..., Z_n)$$

is a chosen procedure for estimating η based on n observations then we define, with $m = \lfloor n/2 \rfloor$,

$$\widetilde{\eta}_1 \equiv \widetilde{\eta}_m(Z_1, ..., Z_m)$$
$$\widetilde{\eta}_2 \equiv \widetilde{\eta}_{n-m}(Z_{m+1}, ..., Z_n)$$

- That is, $\widetilde{\eta}_1$ and $\widetilde{\eta}_2$ are the results of applying our estimation procedure to the first and second halfs of the data.

- Now, define

$$\widehat{\beta}_1 \equiv \beta(\widetilde{\eta}_2) + \frac{1}{m}\sum_{i=1}^{m}\psi(Z_i;\widetilde{\eta}_2)$$

  - $\beta(\widetilde{\eta}_2)$ acts as our preliminary, plug-in, estimator of $\beta(\eta)$ based the second half of the sample $Z_{m+1}, ..., Z_n$ (called the "training sample")
  - The term $\frac{1}{m}\sum_{i=1}^{m}\varphi(Z_i;\widetilde{\eta}_2)$ is an estimator of $e(\widetilde{\eta}_2)$ based on the first half of the sample $Z_1, ..., Z_m$ (called the "validation sample") where for any η'

$$e(\eta') \equiv E_{\eta,\vartheta}[\psi(Z;\eta')] \equiv \int \psi(z;\eta') f_{\eta,\vartheta}(z)\,dz$$

- Likewise define

$$\widehat{\beta}_2 \equiv \beta(\widetilde{\eta}_1) + \frac{1}{n-m}\sum_{i=m+1}^{n}\psi(Z_i;\widetilde{\eta}_1)$$

- Finally, define

$$\widehat{\beta} \equiv \frac{m}{n}\widehat{\beta}_1 + \frac{n-m}{n}\widehat{\beta}_2$$

- Now, suppose that we can show that for some $\varphi(z;\eta)$ verifying $E_{\eta,\vartheta}[\psi(Z;\eta)] = 0$, it holds that

$$\sqrt{m}\left\{\widehat{\beta}_1 - \beta(\eta)\right\} = \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\varphi(Z_i;\eta) + \frac{1}{\sqrt{n-m}}\sum_{i=m+1}^{n}\phi(Z_i;\eta) + o_p(1) \quad (1)$$

- Then, reversing the roles of the training and validation data, it also holds that

$$\sqrt{n-m}\left\{\widehat{\beta}_2 - \beta(\eta)\right\} = \frac{1}{\sqrt{n-m}}\sum_{i=m+1}^{n}\varphi(Z_i;\eta) + \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\phi(Z_i;\eta) + o_p(1)$$

- The last two displays imply that (prove it in the privacy of your own room)

$$\sqrt{n}\left\{\widehat{\beta} - \beta(\eta)\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\varphi(Z_i;\eta) + \phi(Z_i;\eta)] + o_p(1)$$

- So, we will just focus on the estimator $\widehat{\beta}_1$ from one of the two sample-partitions and study under which conditions will
  - $\widehat{\beta}_1$ satisfies the expansion (1)
  - $\varphi(Z;\eta) + \phi(Z;\eta)$ coincides with a given gradient $\psi(Z;\eta)$

- **Example:** (continuation of the missing data example). Suppose that we assume just the non-parametric model $\mathcal{F}_{np}$ but nevertheless, to come up with our procedure for estimating η we assume "working parametric models" for b and for π. Specifically, suppose $\widetilde{\eta}_2 = \left(\widetilde{b}_2, \widetilde{\pi}_2, \widetilde{F}_{2,X}\right)$, where
  - $\widetilde{F}_{2,X}$ is the empirical marginal distribution of X in the training sample,
  - $\widetilde{b}_2(X) \equiv \widetilde{\gamma}^T\widetilde{X}$ where $\widetilde{X} = \left[1, X^T\right]^T$ and $\widetilde{\gamma}_2$ solves the least squares equation

$$\sum_{i=m+1}^{n} R_i\widetilde{X}_i\left[Y_i - X_i^T\gamma\right] = 0$$

  - $\widetilde{\pi}_2(X) \equiv \pi(X;\widetilde{\alpha}_2)$ where $\widetilde{\alpha}_2$ is the ML estimator of a "working model" $\pi(X;\alpha)$ for α, say a logistic regression model $\pi(X;\alpha) = \exp\left[\alpha^T\widetilde{X}\right] / \left\{1 + \exp\left[\alpha^T\widetilde{X}\right]\right\}$, i.e. solving

$$\sum_{i=m+1}^{n}\widetilde{X}_i\left[R_i - \exp\left[\alpha^T\widetilde{X}_i\right] / \left\{1 + \exp\left[\alpha^T\widetilde{X}_i\right]\right\}\right] = 0$$

- Then, with $\psi(Z_i;\eta)$ being the unique mean zero gradient for $\beta(F)$ (and hence the efficient influence function) in model $\mathcal{F}_{np}$ we obtain

$$\begin{aligned}\widehat{\beta}_1 &\equiv \beta(\widetilde{\eta}_2)+\frac{1}{m}\sum_{i=1}^{m}\psi(Z_i;\widetilde{\eta}_2)\\ &= \beta(\widetilde{\eta}_2)+\frac{1}{m}\sum_{i=1}^{m}\left[\widetilde{b}_2(X_i)+\frac{R_i\left\{Y_i-\widetilde{b}_2(X_i)\right\}}{\widetilde{\pi}_2(X_i)}-\beta(\widetilde{\eta}_2)\right]\\ &= \frac{1}{m}\sum_{i=1}^{m}\left[\widetilde{b}_2(X_i)+\frac{R_i\left\{Y_i-\widetilde{b}_2(X_i)\right\}}{\widetilde{\pi}_2(X_i)}\right]\end{aligned}$$

- Now, for any η^*, not necessarily the true η, write

$$\begin{aligned}\sqrt{m}\left\{\widehat{\beta}_1-\beta(\eta)\right\} &= \underbrace{\frac{\sqrt{m}}{m}\sum_{i=1}^{m}\left[\psi(Z_i;\widetilde{\eta}_2)-e(\widetilde{\eta}_2)\right]-\frac{\sqrt{m}}{m}\sum_{i=1}^{m}\left[\psi(Z_i;\eta^*)-e(\eta^*)\right]}_{A_m}\\ &\quad+\underbrace{\frac{\sqrt{m}}{m}\sum_{i=1}^{m}\left[\psi(Z_i;\eta^*)-e(\eta^*)\right]}_{B_m}\\ &\quad+\underbrace{\sqrt{m}\left\{\beta(\widetilde{\eta}_2)-\beta(\eta)\right\}+\sqrt{m}\,e(\widetilde{\eta}_2)}_{C_m}\end{aligned}$$

- where, recall,

$$e(\eta')\equiv\int\psi(z;\eta')\,f_{\eta,\vartheta}(z)\,dz$$

- Consider the term A_m.
- If $\beta(\eta)$ is R^k- valued then $\psi(z;\eta)$ is also R^k- valued. Write $\psi=(\psi_1,...,\psi_k)^T$ and let $\|\psi(z;\widetilde{\eta}_2)-\psi(z;\eta^*)\|^2\equiv\sum_{j=1}^{k}\{\psi_j(z;\widetilde{\eta}_2)-\psi_j(z;\eta^*)\}^2$.
- We will now show that if η^* is the probability limit of $\widetilde{\eta}_2$ in the sense that

$$\int\|\psi(z;\widetilde{\eta}_2)-\psi(z;\eta^*)\|^2 f(z;\eta,\vartheta)\,dz \underset{m\to 0}{\overset{P_{F_{\eta,\vartheta}}}{\to}} 0 \tag{2}$$

then

$$A_m \underset{m\to 0}{\overset{P_{F_{\eta,\vartheta}}}{\to}} 0$$

- Recall that $\widetilde{\eta}_2$ depends on the training data $Z_{m+1},...,Z_n$ which is independent of the validation data $Z_1,...,Z_m$.
- Now,

$$\begin{aligned}&E_{\eta,\vartheta}\left[A_m|Z_{m+1},...,Z_n\right]=\\ &= \frac{\sqrt{m}}{m}\sum_{i=1}^{m}\left[E_{\eta,\vartheta}\left[\psi(Z_i;\widetilde{\eta}_2)-\psi\left(Z_i;\eta^*\right)|Z_{m+1},...,Z_n\right]\right]-\left[e(\widetilde{\eta}_2)-e\left(\eta^*\right)\right]\\ &= E_{\eta,\vartheta}\left[\psi(Z;\widetilde{\eta}_2)-\psi\left(Z;\eta^*\right)|Z_{m+1},...,Z_n\right]-\left[e(\widetilde{\eta}_2)-e\left(\eta^*\right)\right]\\ &= 0\end{aligned}$$

- Furthermore, letting $A_{j,m}$ denote the j^{th} entry of the R^k- valued vector A_m, we have

$$\begin{aligned}var_{\eta,\vartheta}\left[A_{j,m}|Z_{m+1},...,Z_n\right] &= var_{\eta,\vartheta}\left[\psi_j(Z;\widetilde{\eta}_2)-\psi_j\left(Z;\eta^*\right)|Z_{m+1},...,Z_n\right]\\ &\leq E_{\eta,\vartheta}\left[\left\{\psi_j(Z;\widetilde{\eta}_2)-\psi_j\left(Z;\eta^*\right)\right\}^2|Z_{m+1},...,Z_n\right]\\ &= \int\left\{\psi_j(z;\widetilde{\eta}_2)-\psi_j\left(z;\eta^*\right)\right\}^2 f(z;\eta,\vartheta)\,dz\\ &\leq \int\left\|\psi(z;\widetilde{\eta}_2)-\psi\left(z;\eta^*\right)\right\|^2 f(z;\eta,\vartheta)\,dz \underset{m\to 0}{\overset{P_{F_{\eta,\vartheta}}}{\to}} 0\end{aligned}$$

- Consequently,

$$E_{\eta,\vartheta}\left[A_{j,m}^{2}|Z_{m+1},...,Z_{n}\right]=var_{\eta,\vartheta}\left[A_{j,m}|Z_{m+1},...,Z_{n}\right]\underset{m\to0}{\overset{P_{F_{\eta,\vartheta}}}{\to}}0$$

- Then, by Tchebichev's inequality, for any $\delta>0$

$$\begin{aligned}Q_{m} &\equiv q_{m}\left(Z_{m+1},...,Z_{n}\right)\\ &\equiv P_{\eta,\vartheta}\left[|A_{j,m}|>\delta|Z_{m+1},...,Z_{n}\right]\\ &\leq \frac{E_{\eta,\vartheta}\left[A_{j,m}^{2}|Z_{m+1},...,Z_{n}\right]}{\delta^{2}}\underset{m\to0}{\overset{P_{F_{\eta,\vartheta}}}{\to}}0\end{aligned}$$

- Furthermore, Q_m, being a conditional probability, satisfies $|Q_m|\leq1$, so Q_m is a bounded sequence that converges to 0 in probability. Then,

$$E_{\eta,\vartheta}\left[Q_{m}\right]\underset{m\to0}{\to}0$$

- But $E_{\eta,\vartheta}\left[Q_{m}\right]=P_{\eta,\vartheta}\left[|A_{j,m}|>\delta\right]$, which then shows that $A_{j,m}\underset{m\to0}{\overset{P_{F_{\eta,\vartheta}}}{\to}}0$. Since this is true for all $j=1,...,k$, then $A_m\underset{m\to0}{\overset{P_{F_{\eta,\vartheta}}}{\to}}0$.

- Consider next the term B_m. This term is just

$$B_{m}=\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\varphi\left(Z_{i};\eta\right)$$

where

$$\varphi\left(Z;\eta\right)\equiv\psi\left(Z;\eta^{*}\right)-\int\psi\left(z;\eta^{*}\right)f\left(z;\eta,\vartheta\right)dz$$

- Finally, consider the term C_m.

$$\boxed{C_{m}=\sqrt{m}\left\{\beta\left(\widetilde{\eta}_{2}\right)-\beta\left(\eta^{*}\right)\right\}+\sqrt{m}\,e\left(\widetilde{\eta}_{2}\right)}$$

- Define for any η'

$$\chi\left(\eta'\right)\equiv\beta\left(\eta'\right)-\beta\left(\eta\right)+E_{\eta}\left[\psi\left(Z;\eta'\right)\right]$$

- Then, noticing that $\chi\left(\eta\right)=0$, we can write C_m as

$$\boxed{C_{m}=\sqrt{m}\left\{\chi\left(\widetilde{\eta}_{2}\right)-\chi\left(\eta\right)\right\}}$$

- Thus, C_m is $\sqrt{m}$ times the difference of the plug-in estimator $\chi\left(\widetilde{\eta}_{2}\right)$ of $\chi\left(\eta\right)$ where $\widetilde{\eta}_{2}$ depends only on the training sample data.
- The term C_m must be analyzed individually in each estimation problem and its asymptotic behavior will depend on the nature of the estimator $\widetilde{\eta}_{2}$.
- At the level of generality presented here, we can nevertheless investigate the implications of different asymptotic behaviors of C_m.
- Clearly, if $\sqrt{m}\left\{\chi\left(\widetilde{\eta}_{2}\right)-\chi\left(\eta\right)\right\}$ diverges as $m\to\infty$ then $\sqrt{m}\left\{\widehat{\beta}_{1}-\beta\left(\eta\right)\right\}$ necessarily diverges, and so does $\sqrt{n}\left\{\widehat{\beta}-\beta\left(\eta\right)\right\}$.

- Suppose instead we could show that $\chi\left(\widetilde{\eta}_{2}\right)$ is an asymptotically linear estimator of $\chi\left(\eta\right)$, say with influence function $\phi\left(Z;\eta\right)$, that is

$$\sqrt{n-m}\left\{\chi\left(\widetilde{\eta}_{2}\right)-\chi\left(\eta\right)\right\}=\frac{1}{\sqrt{n-m}}\sum_{i=m+1}^{n}\phi\left(Z_{i};\eta\right)+o_{p}\left(1\right)$$

- then we would conclude that

$$\begin{aligned}\sqrt{m}\left\{\widehat{\beta}_{1}-\beta\left(\eta\right)\right\} &= A_{m}+B_{m}+C_{m}+o_{p}\left(1\right)\\ &= o_{p}\left(1\right)+\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\varphi\left(Z_{i};\eta\right)+\sqrt{m}\left\{\chi\left(\widetilde{\eta}_{2}\right)-\chi\left(\eta\right)\right\}\\ &= o_{p}\left(1\right)+\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\varphi\left(Z_{i};\eta\right)+\underbrace{\left(\frac{\sqrt{m}}{\sqrt{n-m}}\right)}_{1+o(1)}\underbrace{\sqrt{n-m}\left\{\chi\left(\widetilde{\eta}_{2}\right)-\chi\left(\eta\right)\right\}}_{O(1)}\\ &= o_{p}\left(1\right)+\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\varphi\left(Z_{i};\eta\right)+\sqrt{n-m}\left\{\chi\left(\widetilde{\eta}_{2}\right)-\chi\left(\eta\right)\right\}\\ &= o_{p}\left(1\right)+\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\varphi\left(Z_{i};\eta\right)+\frac{1}{\sqrt{n-m}}\sum_{i=m+1}^{n}\phi\left(Z_{i};\eta\right)\end{aligned}$$

- As we have argued before, the expansion

$$\sqrt{m}\left\{\widehat{\beta}_1-\beta\left(\eta\right)\right\}=\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\varphi\left(Z_i;\eta\right)+\frac{1}{\sqrt{n-m}}\sum_{i=m+1}^{n}\phi\left(Z_i;\eta\right)+o_p\left(1\right)$$

in turn, implies

$$\sqrt{n}\left\{\widehat{\beta}-\beta\left(\eta\right)\right\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\varphi\left(Z_i;\eta\right)+\phi\left(Z_i;\eta\right)\right]+o_p\left(1\right)$$

- Thus, recalling that

$$\varphi\left(Z_i;\eta\right)=\psi\left(Z_i;\eta^*\right)-E_\eta\left[\psi\left(Z_i;\eta^*\right)\right]$$

we would conclude that

$$\sqrt{n}\left\{\widehat{\beta}-\beta\left(\eta\right)\right\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\psi\left(Z_i;\eta^*\right)-E_\eta\left[\psi\left(Z_i;\eta^*\right)\right]+\phi\left(Z_i;\eta\right)\right]+o_p\left(1\right)$$

- That is, we would conclude that $\widehat{\beta}$ is an asymptotically linear of $\beta\left(\eta\right)$ at $F_{\eta,\vartheta}$ with influence function

$$\psi\left(Z;\eta^*\right)-E_\eta\left[\psi\left(Z;\eta^*\right)\right]\ \ +\phi\left(Z;\eta\right)$$

where η^* satisfies condition (2) of slide 11 (essentially η^* is a type of probability limit of $\widetilde{\eta}_n\left(Z_1,...,Z_n\right)$ under $F_{\eta,\vartheta}$).

- Finally, suppose that we could prove that $\sqrt{m}\left\{\chi\left(\widetilde{\eta}_2\right)-\chi\left(\eta\right)\right\}=o_p\left(1\right)$ then we would conclude that

$$\sqrt{m}\left\{\widehat{\beta}_1-\beta\left(\eta\right)\right\}=\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left\{\psi\left(Z_i;\eta^*\right)-E_\eta\left[\psi\left(Z_i;\eta^*\right)\right]\right\}+o_p\left(1\right)$$

- But, in fact, typically when $\sqrt{m}\left\{\chi\left(\widetilde{\eta}_2\right)-\chi\left(\eta\right)\right\}=o_p\left(1\right)$ it is also the case that the η^* in condition (2) of slide 11 is equal to η. In such case, the last display would be the same as

$$\sqrt{m}\left\{\widehat{\beta}_1-\beta\left(\eta\right)\right\}=\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left\{\psi\left(Z_i;\eta\right)-\underbrace{E_\eta\left[\psi\left(Z_i;\eta\right)\right]}_{=0}\right\}$$

- The last display, in turn, would imply that

$$\sqrt{n}\left\{\widehat{\beta}-\beta\left(\eta\right)\right\}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi\left(Z_i;\eta\right)+o_p\left(1\right)$$

thus yielding $\widehat{\beta}$ to be an asymptotically linear estimator of $\beta\left(\eta\right)$ at $F_{\eta,\vartheta}$ with influence function

$$\psi\left(Z;\eta\right)$$

- I want to call your attention to some intuitive heuristic point about what is to be generally expected from the behavior of $\sqrt{m}\left\{\chi\left(\widetilde{\eta}_2\right)-\chi\left(\eta\right)\right\}$ when $\psi\left(z;\eta\right)$ is a mean zero gradient of $\beta\left(\eta\right)$.
- Recall $\beta\left(\eta'\right)$ is pathwise differentiable at $F_{\eta,\vartheta}$ wrt to a given class $\mathcal{A}$ in model $\mathcal{F}$. Now, suppose that $\chi\left(\eta'\right)$ is a pathwise differentiable parameter at $F_{\eta,\vartheta}$ wrt to the class $\mathcal{A}$ in model $\mathcal{F}$. Then, for any parametric sumodel $F_{sub}=\left\{F_{\eta_\theta,\vartheta_\theta}:\theta\in\Theta\right\}$ such that $\eta_{\theta^*}=\eta$ we have

$$\begin{aligned}\left.\frac{d}{d\theta}\chi\left(\eta_\theta\right)\right|_{\theta=\theta^*}&=\left.\frac{d}{d\theta}\beta\left(\eta_\theta\right)\right|_{\theta=\theta^*}+\left.\frac{d}{d\theta}E_\eta\left[\psi\left(Z;\eta_\theta\right)\right]\right|_{\theta=\theta^*}\\&=E_\eta\left[\psi\left(Z;\eta\right)S_\theta\left(\theta^*\right)\right]+\left.\frac{d}{d\theta}E_\eta\left[\psi\left(Z;\eta_\theta\right)\right]\right|_{\theta=\theta^*}\end{aligned}$$

- But since $E_{\eta_\theta}\left[\psi\left(Z;\eta_\theta\right)\right]=0$ for all θ then

$$\begin{aligned}0&=\left.\frac{d}{d\theta}E_{\eta_\theta}\left[\psi\left(Z;\eta_\theta\right)\right]\right|_{\theta=\theta^*}\\&=\left.\frac{d}{d\theta}E_\eta\left[\psi\left(Z;\eta_\theta\right)\right]\right|_{\theta=\theta^*}+\left.\frac{d}{d\theta}E_{\eta_\theta}\left[\psi\left(Z;\eta\right)\right]\right|_{\theta=\theta^*}\\&=\left.\frac{d}{d\theta}E_\eta\left[\psi\left(Z;\eta_\theta\right)\right]\right|_{\theta=\theta^*}+E_\eta\left[\psi\left(Z;\eta\right)S_\theta\left(\theta^*\right)\right]\end{aligned}$$

from where we deduce that

$$\boxed{\left.\tfrac{d}{d\theta}\chi\left(\eta_\theta\right)\right|_{\theta=\theta^*}=0}$$

- Intuitively, because the maps

$$\theta\rightarrow-E_\eta\left[\psi\left(Z;\eta_\theta\right)\right]$$

and

$$\theta\rightarrow\beta\left(\eta_\theta\right)$$

have the same derivative at θ^* [where $\eta_{\theta^*}=\eta$], for all paths $\theta\rightarrow\eta_\theta$, then one would expect that for η' close to η,

$$\beta\left(\eta'\right)-\beta\left(\eta\right)\ \ \text{and}\ \ -E_\eta\left[\psi\left(Z;\eta'\right)\right]-\underbrace{\left[-E_\eta\left[\psi\left(Z;\eta\right)\right]\right]}_{=0}$$

would have roughly the same magnitude up to second order, i.e. that if η lives in some normed space Ξ , then

$$\begin{aligned}&\left\{\beta\left(\eta'\right)-\beta\left(\eta\right)\right\}+E_\eta\left[\psi\left(Z;\eta'\right)\right]\\=&\left\{\beta\left(\eta'\right)-\beta\left(\eta\right)\right\}-\left\{-E_\eta\left[\psi\left(Z;\eta'\right)\right]-\underbrace{\left[-E_\eta\left[\psi\left(Z;\eta\right)\right]\right]}_{=0}\right\}\\=&\,O\left(\left\|\eta'-\eta\right\|^2\right)\end{aligned}$$

where $\|\cdot\|$ is the norm in the space Ξ

- This, is stated somewhat more formally as in the next slide.

- If the (possibly, infinite dimensional) space Ξ where η lies is normed with norm denoted $\|\cdot\|_{\Xi}$, and if the map $\eta' \to \chi(\eta')$ admits the expansion

$$\chi(\eta') = \chi(\eta) + \chi_\eta(\eta' - \eta) + O\left(\|\eta' - \eta\|^2\right)$$

where $\chi_\eta(\cdot)$ is the Frechet derivative of $\chi(\cdot)$ and the class $\mathcal{A}$ has tangent space equal to the maximal tangent space of model $\mathcal{F}$ at $F_{\eta,\vartheta}$, then it must hold that $\chi_\eta(\eta' - \eta) = 0$ for all η', because it can be shown that if the pathwise derivatives $\frac{d}{d\theta}\chi(\eta_\theta)\Big|_{\theta=\theta^*}$ for all models in class $\mathcal{A}$ are 0, then the Frechet derivative is also 0.

- Then, we conclude that

$$\chi(\eta') - \chi(\eta) = O\left(\|\eta' - \eta\|^2\right)$$

- If we now replace η' with $\widetilde{\eta}_2$ and replace χ with its definition, then recalling that $E_\eta[\psi(Z;\eta)] = 0$ because $\psi(Z;\eta)$ is a mean zero gradient, we obtain

$$\boxed{\beta(\widetilde{\eta}_2) - \beta(\eta) + E_\eta[\psi(Z;\widetilde{\eta}_2)] = \chi(\widetilde{\eta}_2) = O\left(\|\widetilde{\eta}_2 - \eta\|^2\right)}$$

- Now, if $\widetilde{\eta}_2$ is an estimator constructed assuming a *correctly specified* parametric working model for η, then typically it holds that $\sqrt{m}\|\widetilde{\eta}_2 - \eta\| = O_p(1)$ under $F_{\eta,\vartheta}$. In such case,

$$\begin{aligned}
\sqrt{m}O\left(\|\widetilde{\eta}_2 - \eta\|^2\right) &= \sqrt{m}\|\widetilde{\eta}_2 - \eta\|^2 O(1) \\
&= \frac{1}{\sqrt{m}}\left\{\sqrt{m}\|(\widetilde{\eta}_2 - \eta)\|\right\}^2 O(1) \\
&= \frac{1}{\sqrt{m}}O_p(1)O(1) \\
&= o_p(1)
\end{aligned}$$

- So

$$\begin{aligned}
\sqrt{m}\left\{\beta(\widetilde{\eta}_2) - \beta(\eta) + E_\eta[\psi(Z;\widetilde{\eta}_2)]\right\} &= \sqrt{m}O\left(\|\widetilde{\eta}_2 - \eta\|^2\right) \\
&= o_p(1)
\end{aligned}$$

- In some instances, as in our missing data example, η involves two regression functions say $\eta = (\nu, \kappa)$ and it just happens that

$$\beta(\widetilde{\eta}_2) - \beta(\eta) + E_\eta[\psi(Z;\widetilde{\eta}_2)] = O(\|(\widetilde{\nu}_2 - \nu)(\widetilde{\kappa}_2 - \kappa)\|)$$

- In such case, if $\widetilde{\nu}_2$ and $\widetilde{\kappa}_2$ are estimated under parametric working models for ν and κ, then if one of the models, say the model for ν, is correctly specified but the other is incorrectly specified, it will typically be the case that the $O(\|(\widetilde{\nu}_2 - \nu)(\widetilde{\kappa}_2 - \kappa)\|)$ term is asymptotically linear and has an influence function $\phi(Z;\eta)$. We will see this scenario in the missing data example next.

- So, in such case as discussed in slide 7, $\widehat{\beta}$ is asymptotically linear with influence function

$$\psi(Z_i;\eta^*) - E_\eta[\psi(Z_i;\eta^*)] + \phi(Z;\eta)$$

- If $\widetilde{\eta}_2$ involves an ordinary smoothing estimator of a regression function and X is $d-$dimensional, then typically

$$\|\widetilde{\eta}_2 - \eta\| = O_p\left(m^{-\frac{\delta/d}{1+2\delta/d}}\right)$$

where δ is the number of derivatives of the true regression function η.

- So, if $\delta/d > 1/2$ it follows that $\frac{\delta/d}{1+2\delta/d} > 1/4$, so $\frac{1}{2} - \frac{2\delta/d}{1+2\delta/d} < 0$ and consequently

$$\sqrt{m}\|\widetilde{\eta}_2 - \eta\|^2 = m^{1/2}O_p\left(m^{-\frac{2\delta/d}{1+2\delta/d}}\right) = O_p\left(m^{\frac{1}{2}-\frac{2\delta/d}{1+2\delta/d}}\right) = o_p(1)$$

- Thus, we would typically expect that when $\delta/d > 1/2$, $\sqrt{m}\{\chi(\widetilde{\eta}_2) - \chi(\eta)\} = o_p(1)$, and consequently, as argued in slide 18, that $\widehat{\beta}$ is asymptotically linear estimator of $\beta(\eta)$ with influence function $\psi(Z_i;\eta)$.

- Note however that if $\delta/d < 1/2$, then $\sqrt{m}\,\|\widetilde{\eta}_2 - \eta\|^2$ diverges, and thus $\sqrt{m}\left\{\widehat{\beta} - \beta(\eta)\right\}$ does not converge in law.
- Note also that for the plug-in estimator $\beta(\widetilde{\eta}_2)$ it holds that

$$\beta(\widetilde{\eta}_2) - \beta(\eta) = \dot{\beta}_\eta(\widetilde{\eta}_2 - \eta) + O\left(\|\widetilde{\eta}_2 - \eta\|^2\right)$$

- For several usual estimators $\widetilde{\eta}_2$, the term $\dot{\beta}_\eta(\widetilde{\eta}_2 - \eta)$ does not vanish, so

$$\beta(\widetilde{\eta}_2) - \beta(\eta) = O(\|\widetilde{\eta}_2 - \eta\|)$$

- For such settings even if $\widetilde{\eta}_2$ were to be converge to η at the optimal rate, i.e. $\|\widetilde{\eta}_2 - \eta\| = O_p\left(m^{-\frac{\delta/d}{1+2\delta/d}}\right)$, it would happen that $\sqrt{m}\{\beta(\widetilde{\eta}_2) - \beta(\eta)\}$ would diverge.

- **Example:** (continuation of the missing data example). If $\widetilde{\eta}_2 = \left(\widetilde{b}_2, \widetilde{\pi}_2, \widetilde{F}_{2,X}\right)$, then, if $\psi(Z_i;\eta)$ denotes the unique mean zero gradient for $\beta(F)$ (and hence the efficient influence function) in model $\mathcal{F}_{np}$ recall that we obtain

$$\begin{aligned}
\widehat{\beta}_1 &\equiv \beta(\widetilde{\eta}_2) + \frac{1}{m}\sum_{i=1}^{m}\psi(Z_i;\widetilde{\eta}_2) \\
&= \beta(\widetilde{\eta}_2) + \frac{1}{m}\sum_{i=1}^{m}\left[\widetilde{b}_2(X_i) + \frac{R_i\left\{Y_i - \widetilde{b}_2(X_i)\right\}}{\widetilde{\pi}_2(X_i)} - \beta(\widetilde{\eta}_2)\right] \\
&= \frac{1}{m}\sum_{i=1}^{m}\left[\widetilde{b}_2(X_i) + \frac{R_i\left\{Y_i - \widetilde{b}_2(X_i)\right\}}{\widetilde{\pi}_2(X_i)}\right]
\end{aligned}$$

- In this example, $\eta' = \left(b', \pi', F'_x\right)$ and

$$\begin{aligned}
\chi(\eta') &= \beta(\eta') - \beta(\eta) + E_\eta\left[\psi(Z;\eta')\right] \\
&= \beta(\eta') - E_{\eta,\vartheta}\left[b_{F_{\eta,\vartheta}}(X)\right] + E_{\eta,\vartheta}\left[b'(X) + \frac{R\{Y - b'(X)\}}{\pi'(X)} - \beta(\eta')\right] \\
&= E_{\eta,\vartheta}\left[\left\{b'(X) - b_{F_{\eta,\vartheta}}(X)\right\} + \frac{R\{Y - b'(X)\}}{\widetilde{\pi}_2(X)}\right] \\
&= E_{\eta,\vartheta}\left[\left\{b'(X) - b_{F_{\eta,\vartheta}}(X)\right\} + \frac{R\left\{b_{F_{\eta,\vartheta}}(X) - b'(X)\right\}}{\pi'(X)}\right] \\
&= E_{\eta,\vartheta}\left[\left\{b'(X) - b_{F_{\eta,\vartheta}}(X)\right\} + \frac{\pi_{F_{\eta,\vartheta}}(X)\left\{b_{F_{\eta,\vartheta}}(X) - b'(X)\right\}}{\pi'(X)}\right] \\
&= E_{\eta,\vartheta}\left[\left\{b'(X) - b_{F_{\eta,\vartheta}}(X)\right\}\left\{1 - \frac{\pi_{F_{\eta,\vartheta}}(X)}{\pi'(X)}\right\}\right]
\end{aligned}$$

- Note that in this problem $\chi(\eta')$ depends on $\eta' = \left(b', \pi', F'_x\right)$ only through $\left(b', \pi'\right)$.
- So, in what follows we will write $\chi\left(b', \pi'\right)$ or $\chi(\eta')$ indistinctively.

- Now, suppose
  - we estimate $b(X) = E(Y|R=1,X)$ assuming a "working" linear regression model for Y on X among subjects with $R = 1$. That is $\widetilde{b}_2(X) \equiv \widetilde{\gamma}^T\widetilde{X}$ where $\widetilde{X} = \left[1, X^T\right]^T$ and $\widetilde{\gamma}_2$ solves the least squares equation

$$\sum_{i=m+1}^{n} u(Z_i;\gamma) \equiv \sum_{i=m+1}^{n} R_i\widetilde{X}_i\left[Y_i - X_i^T\gamma\right] = 0$$

  - $\widetilde{\pi}_2(X) \equiv \pi(X;\widetilde{\alpha}_2)$ where $\widetilde{\alpha}_2$ is the ML estimator of a "working model" $\pi(X;\alpha)$ for α, say a logistic regression model $\pi(X;\alpha) = \exp\left[\alpha^T\widetilde{X}\right]/\left\{1 + \exp\left[\alpha^T\widetilde{X}\right]\right\}$, i.e. solving

$$\sum_{i=m+1}^{n} S_\alpha(Z_i;\alpha) \equiv \sum_{i=m+1}^{n} \widetilde{X}_i\left[R_i - \exp\left[\alpha^T\widetilde{X}_i\right]/\left\{1 + \exp\left[\alpha^T\widetilde{X}_i\right]\right\}\right] = 0$$

- Note that neither working model need be correct.
- Regardless of whether or not the models are correct, suppose that the equation in γ

$$E_{\eta,\vartheta}\left[u(Z;\gamma)\right] = 0$$

has a unique solution, say $\gamma(\eta)$,
- and the equation in α,

$$E_\eta\left[S_\alpha(Z;\alpha)\right] = 0$$

also has a unique solution, say $\alpha(\eta)$.

- Under regularity conditions, it follows from the asymptotic theory for $Z-$ estimators that we briefly discussed in Ch 3, that

$$\boxed{\sqrt{n-m}\left\{\left[\begin{array}{c}\widetilde{\gamma}_2\\ \widetilde{\alpha}_2\end{array}\right]-\left[\begin{array}{c}\gamma(\eta)\\ \alpha(\eta)\end{array}\right]\right\}=-\frac{1}{\sqrt{n-m}}\sum_{i=m+1}^{n}\left[\begin{array}{c}\phi^{\gamma}(Z_i;\eta)\\ \phi^{\alpha}(Z_i;\eta)\end{array}\right]+o_p(1)}$$

where

$$\phi^{\gamma}(Z;\eta)=-\left\{\partial E_{\eta,\vartheta}\left[u(Z;\gamma)\right]/\partial\gamma^T\Big|_{\gamma=\gamma(\eta)}\right\}^{-1}u(Z;\gamma(\eta))$$

and

$$\phi^{\alpha}(Z;\eta)=-\left\{\partial E_{\eta,\vartheta}\left[S_{\alpha}(Z;\alpha)\right]/\partial\alpha^T\Big|_{\alpha=\alpha(\eta)}\right\}^{-1}S_{\alpha}(Z;\alpha(\eta))$$

- In particular, $\widetilde{\gamma}_2 \overset{P_{F_{\eta,\vartheta}}}{\to} \gamma(\eta)$ and $\widetilde{\alpha}_2 \overset{P_{F_{\eta,\vartheta}}}{\to} \gamma(\eta)$.

- Now, if $\widetilde{\eta}_2=\left(\widetilde{b}_2,\widetilde{\pi}_2,\widehat{F}_{n-m,X}\right)$, then

$$\begin{aligned}\psi(Z;\widetilde{\eta}_2) &= \widetilde{X}^T\widetilde{\gamma}_2+\frac{R}{\text{expit}\left[\widetilde{X}^T\widetilde{\alpha}_2\right]}\left(Y-\widetilde{X}^T\widetilde{\gamma}_2\right)-\frac{1}{n-m}\sum_{i=m+1}^{n}\widetilde{X}^T\widetilde{\gamma}_2\\ &= \widetilde{X}^T\widetilde{\gamma}_2+\frac{R}{\text{expit}\left[\widetilde{X}^T\widetilde{\alpha}_2\right]}\left(Y-\widetilde{X}^T\widetilde{\gamma}_2\right)-\widetilde{E}_2\left(\widetilde{X}\right)^T\widetilde{\gamma}_2\end{aligned}$$

- Letting $\widetilde{E}_2\left(\widetilde{X}\right)$ denote the sample mean of X in the training sample, and

$$\eta^*\equiv\left(b(\cdot,\gamma(\eta)),\pi(\cdot,\alpha(\eta)),F_{X,\eta,\vartheta}\right)$$

- we then have that for each fixed z,

$$\begin{aligned}&\psi(z;\widetilde{\eta}_2)-\psi(z;\eta^*)=\\ =\ &\widetilde{x}^T\left\{\widetilde{\gamma}_2-\gamma(\eta)\right\}+\frac{r\left(y-\widetilde{x}^T\widetilde{\gamma}_2\right)}{\text{expit}\left[\widetilde{x}^T\widetilde{\alpha}_2\right]}-\frac{r\left(y-\widetilde{x}^T\gamma(\eta)\right)}{\text{expit}\left[\widetilde{x}^T\alpha(\eta)\right]}-\widetilde{E}_2\left(\widetilde{X}\right)^T\widetilde{\gamma}_2+E_{F_X}\left(\widetilde{X}\right)^T\gamma(\eta)\\ &\overset{P_{F_{\eta,\vartheta}}}{\to} 0\end{aligned}$$

- If all the components of the random vector Z are bounded, then the preceding convergence in probability for each fixed z, implies that

$$\int\left(\psi(z;\widetilde{\eta}_2)-\psi(z;\eta^*)\right)^2 f(z;\eta,\vartheta)\,dz \overset{P_{F_{\eta,\vartheta}}}{\to} 0$$

- This last displayed convergence is precisely the condition (2) in slide 11.

- Now, write

$$\begin{aligned}\chi(\widetilde{\eta}_2) &= \chi\left(\widetilde{b}_2,\widetilde{\pi}_2\right)\\ &= \chi\left(b(;\widetilde{\gamma}_2),\pi(\cdot;\widetilde{\alpha}_2)\right)\\ &\equiv \tau(\widetilde{\gamma}_2,\widetilde{\alpha}_2)\end{aligned}$$

- If X and Y are bounded, it can be shown that $\tau(\gamma,\alpha)$ is a differentiable function of (γ,α) at $(\gamma(\eta),\alpha(\eta))$. Then, by problem 1 of hmw 3, we have that

$$\sqrt{n-m}\left\{\tau(\widetilde{\gamma}_2,\widetilde{\alpha}_2)-\tau(\gamma(\eta),\alpha(\eta))\right\}=-\frac{1}{\sqrt{n-m}}\sum_{i=m+1}^{n}\phi(Z_i;\eta)+o_p(1)$$

where

$$\boxed{\phi(Z;\eta)=\frac{\partial\tau(\gamma,\alpha)}{\partial(\gamma^T,\alpha^T)}\Big|_{(\gamma,\alpha)=(\gamma(\eta),\alpha(\eta))}\left[\begin{array}{c}\phi^{\gamma}(Z;\eta)\\ \phi^{\alpha}(Z;\eta)\end{array}\right]}$$

- So we conclude that

$$\sqrt{n-m}\left\{\chi\left(\widetilde{\eta}_2\right)-\chi\left(\eta^*\right)\right\}=\frac{1}{\sqrt{n-m}}\sum_{i=m+1}^{n}\phi\left(Z_i;\eta\right)+o_p\left(1\right)$$

- Equivalently, since $\sqrt{m}/\sqrt{n-m}=1+o\left(1\right)$,

$$\boxed{\sqrt{m}\left\{\chi\left(\widetilde{\eta}_2\right)-\chi\left(\eta^*\right)\right\}=\tfrac{1}{\sqrt{n-m}}\sum_{i=m+1}^{n}\phi\left(Z_i;\eta\right)+o_p\left(1\right)}$$

- Now, let us consider four possible scenarios:
  - Both models for b and π are incorrect
  - Both models for b and π are correct
  - The model for π is correct but the model for b is incorrect
  - The model for b is correct but the model for π is incorrect

- In scenario (1), i.e. when the models for b and π are both incorrect we have

$$b\left(\cdot,\gamma\left(\eta\right)\right)\neq b_{F_{\eta,\vartheta}}\left(\cdot\right)\text{ and }\pi\left(\cdot,\alpha\left(\eta\right)\right)\neq\pi_{F_{\eta,\vartheta}}\left(\cdot\right)$$

- Now, recalling that

$$\eta^*\equiv\left(b\left(\cdot,\gamma\left(\eta\right)\right),\pi\left(\cdot,\alpha\left(\eta\right)\right),F_{X,\eta,\vartheta}\right)$$

  we conclude that

$$\chi\left(\eta^*\right)=E_{\eta,\vartheta}\left[\left\{b\left(X,\gamma\left(\eta\right)\right)-b_{F_{\eta,\vartheta}}\left(X\right)\right\}\left\{1-\frac{\pi_{F_{\eta,\vartheta}}\left(X\right)}{\pi\left(X,\alpha\left(\eta\right)\right)}\right\}\right]$$

- The expectation in the right hand side of the last display is not equal to zero, except if miraculaously $b\left(X,\gamma\left(\eta\right)\right)$ and $\pi\left(X,\alpha\left(\eta\right)\right)$ just happen to co-vary in such a way that the expectation in the right hand side cancels.

- But if, as is nearly always the case, $\chi\left(\eta^*\right)\neq 0$, then

$$\begin{aligned}\sqrt{m}\left\{\chi\left(\widetilde{\eta}_2\right)-\underbrace{\chi\left(\eta\right)}_{=0}\right\} &= \sqrt{m}\left\{\chi\left(\widetilde{\eta}_2\right)-\chi\left(\eta^*\right)\right\}+\sqrt{m}\chi\left(\eta^*\right)\\ &= \frac{1}{\sqrt{n-m}}\sum_{i=m+1}^{n}\phi\left(Z_i;\eta\right)+o_p\left(1\right)+\sqrt{m}\chi\left(\eta^*\right)\end{aligned}$$

  diverges to $+\infty$ if $\chi\left(\eta^*\right)>0$ and to $-\infty$ if $\chi\left(\eta^*\right)<0$.

- Then, in scenario 1, $\sqrt{m}\left\{\widehat{\beta}_1-\beta\left(\eta\right)\right\}$, nearly always, diverges.

- In fact, when both models for b and π are wrong, $\widehat{\beta}_1$ does not even converge in probability to $\beta\left(\eta\right)$. To see this, write

$$\begin{aligned}\widehat{\beta}_1-\beta\left(\eta\right) &= \underbrace{\frac{1}{\sqrt{m}}A_m}_{o_p(1)}+\underbrace{\frac{1}{\sqrt{m}}B_m}_{=o_p(1)}+\frac{1}{\sqrt{m}}C_m\\ &= o_p\left(1\right)+\chi\left(\widetilde{\eta}_2\right)\overset{P_{F_{\eta,\vartheta}}}{\to}\chi\left(\eta^*\right)\neq 0\end{aligned}$$

- To study the other three scenarios, we first note that in all three scenarios,

$$\begin{aligned}\chi(\eta^*) &= \tau(\gamma(\eta), \alpha(\eta)) \\ &= E_{\eta,\vartheta}\left[\left\{b(X;\gamma(\eta)) - b_{F_{\eta,\vartheta}}(X)\right\}\left\{1 - \frac{\pi_{F_{\eta,\vartheta}}(X)}{\pi(X;\alpha(\eta))}\right\}\right] \\ &= 0\end{aligned}$$

because

- when the model for write π is correct, $\pi_{F_{\eta,\vartheta}}(X) = \pi(X;\alpha(\eta))$, so

$$\left\{1 - \frac{\pi_{F_{\eta,\vartheta}}(X)}{\pi(X;\alpha(\eta))}\right\} = 0,$$

and

- when the model for b is correct, $b(X;\gamma(\eta)) = b_{F_{\eta,\vartheta}}(X)$,

- So, we are already in the position to conclude that $\widehat{\beta}_1$ is asymptotically linear and converges to $\beta(\eta)$ in all three scenarios.
- In particular, $\widehat{\beta}_1$ and consequently $\widehat{\beta}$, is consistent and asymptotically normal for $\beta(\eta)$ so long as one of the working models for either π or b is correct, but not necessarily both are correct.

- More precisely, define the following two semiparametric models

$$\begin{aligned}\mathcal{F}_{sub,1} &= \{F \in \mathcal{F}_{np} : \text{ the working model for } b \text{ holds}\} \\ \mathcal{F}_{sub,2} &= \{F \in \mathcal{F}_{np} : \text{ the working model for } \pi \text{ holds}\}\end{aligned}$$

- Then $\widehat{\beta}$ is consistent for $\beta(\eta)$ under any F in the union model $\mathcal{F}_{sub,1} \cup \mathcal{F}_{sub,2}$.
- Even more, $\sqrt{n}\left\{\widehat{\beta} - \beta(\eta)\right\}$ converges to a mean zero normal distribution under any F in the union model $\mathcal{F}_{sub,1} \cup \mathcal{F}_{sub,2}$.
- **Definition:** given a semiparametric model, $\mathcal{F}$ and two (possibly also semiparametric) submodels of $\mathcal{F}$, say $\mathcal{F}_{sub,1}$ and $\mathcal{F}_{sub,2}$ an estimator $\widehat{\beta}$ is said to be
  - double-robust consistent for $\beta(F)$ in the union model $\mathcal{F}_{sub,1} \cup \mathcal{F}_{sub,2}$ if $\widehat{\beta}$ converges in probability to $\beta(F)$ under any F in $\mathcal{F}_{sub,1} \cup \mathcal{F}_{sub,2}$
  - double-robust asymptotically normal and unbiased for $\beta(F)$ if $\sqrt{n}\left\{\widehat{\beta} - \beta(\eta)\right\}$ converges to a mean zero normal distribution under any F in $\mathcal{F}_{sub,1} \cup \mathcal{F}_{sub,2}$

- Let us explore what drives quite generally double-robustness.
- In the missing data example, denoting $\nu' \equiv b_{F_{\eta',\vartheta}}(\cdot)$ and $\kappa' \equiv \pi_{F_{\eta',\vartheta}}(\cdot)$ we saw that

$$\chi(\eta') \equiv \beta(\eta') - \beta(\eta) + E_\eta\left[\psi(Z;\eta')\right]$$

has the following special property that

$$\chi(\eta') \text{ depends on } \eta' \text{ only through the product } \left(\nu' - \nu\right)\left(\kappa' - \kappa\right)$$

where ν' and κ' are components of η'.

- This special feature actually was the reason for the double-robustness consistency and asymptotic normality of $\widehat{\beta}_1$
- In general, a one-step sample split estimator will be double robust, every time the "drift" $\chi(\eta') = \beta(\eta') - \beta(\eta) + E_\eta[\psi(Z;\eta')]$ verifies the property in the preceding display.

- Returning to the missing data example, let us explore the behaviour of $\sqrt{m}\{\chi(\widetilde{\eta}_2) - \chi(\eta)\}$ in each of the three scenarios. Recall that

$$\boxed{\sqrt{m}\{\chi(\widetilde{\eta}_2) - \chi(\eta^*)\} = \frac{1}{\sqrt{n-m}}\sum_{i=m+1}^n \phi(Z_i;\eta) + o_p(1)}$$

where

$$\boxed{\phi(Z;\eta) = \left.\frac{\partial\tau(\gamma,\alpha)}{\partial(\gamma^T,\alpha^T)}\right|_{(\gamma,\alpha)=(\gamma(\eta),\alpha(\eta))}\begin{bmatrix}\phi^\gamma(Z;\eta) \\ \phi^\alpha(Z;\eta)\end{bmatrix}}$$

- Now, to calculate $\left.\frac{\partial\tau(\gamma,\alpha)}{\partial(\gamma^T,\alpha^T)}\right|_{(\gamma,\alpha)=(\gamma(\eta),\alpha(\eta))}$ we will decompose η as (λ,κ) where

$$\lambda \equiv \left(b_{F_{\eta,\vartheta}}, F_{\eta,\vartheta,X}\right) \text{ and } \kappa \equiv \pi_{F_{\eta,\vartheta}}.$$

- If we let $\lambda^* \equiv (b(\cdot,\gamma(\eta)), F_{\eta,\vartheta,X})$, $\kappa^* \equiv \pi(\cdot,\alpha(\eta))$, $\lambda_\gamma \equiv (b(\cdot,\gamma), F_{\eta,\vartheta,X})$ and $\kappa_\alpha \equiv \pi(\cdot;\alpha)$, then

$$\left.\frac{\partial\tau(\gamma,\alpha)}{\partial(\gamma,\alpha)}\right|_{(\gamma,\alpha)=(\gamma(\eta),\alpha(\eta))} = \begin{bmatrix}\left.\frac{\partial}{\partial\gamma^T}\chi(\lambda_\gamma,\kappa^*)\right|_{\gamma=\gamma(\eta)} \\ \left.\frac{\partial}{\partial\alpha^T}\chi(\lambda^*,\kappa_\alpha)\right|_{\alpha=\alpha(\eta)}\end{bmatrix}$$

- The following three facts hold in the missing data example and as we will see are the essense for the asymptotic behaviour of $\sqrt{m}\left\{\chi\left(\widetilde{\eta}_2\right)-\chi\left(\eta\right)\right\}$.
  - The assumed model $\mathcal{F}$, namely $\mathcal{F}_{np}, \mathcal{F}_{par}$ or $\mathcal{F}_{fix}$ is a factorized likelihood model where the first factor of the likelihood depends on λ and the second likelihood factor depends on κ.
  - $$\boxed{\beta\left(\eta\right) \text{ depends on } \eta \text{ only through } \lambda}$$
  - $$\boxed{\chi\left(\lambda,\kappa'\right)\equiv\beta\left(\lambda\right)-\beta\left(\lambda\right)+E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda,\kappa'\right)\right]=0 \text{ for all } \lambda \text{ and all } \kappa'}$$
    $$\boxed{\chi\left(\lambda',k\right)\equiv\beta\left(\lambda'\right)-\beta\left(\lambda\right)+E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda',\kappa\right)\right]=0 \text{ for all } \lambda' \text{ and all } \kappa}$$
- Note that in the missing data example this holds because $\chi\left(\eta'\right)$ is a function of $\left(b'-b\right)\left(\pi'-\pi\right)$.

- Consider scenario (2) in which both models for b and π are correct. In such case,

$$\lambda^*\equiv\left(b\left(\cdot,\gamma\left(\eta\right)\right),F_{\eta,\vartheta,X}\right) \text{ is equal to } \lambda\equiv\left(b_{F_{\eta,\vartheta}}\left(\cdot\right),F_{\eta,\vartheta,X}\right)$$

and

$$\kappa^*\equiv\pi\left(\cdot,\alpha\left(\eta\right)\right) \text{ is equal to } \kappa=\pi_{F_{\eta,\vartheta}}\left(\cdot\right).$$

- Then, from

$$\boxed{\chi\left(\lambda',\kappa\right)\equiv\beta\left(\lambda'\right)-\beta\left(\lambda\right)+E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda',\kappa\right)\right]=0 \text{ for all } \lambda' \text{ and all } \kappa}$$

we conclude that

$$\chi\left(\lambda_\gamma,\kappa^*\right)=0 \text{ for all } \gamma$$

- Likewise, from

$$\boxed{\chi\left(\lambda,\kappa'\right)\equiv\beta\left(\lambda\right)-\beta\left(\lambda\right)+E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda,\kappa'\right)\right]=0 \text{ for all } \lambda \text{ and all } \kappa'}$$

we conclude that

$$\chi\left(\lambda^*,\kappa_\alpha\right)=0 \text{ for all } \alpha$$

- From where we obtain that

$$\boxed{\left.\frac{\partial\tau\left(\gamma,\alpha\right)}{\partial\left(\gamma,\alpha\right)}\right|_{(\gamma,\alpha)=(\gamma(\eta),\alpha(\eta))}=\left[\begin{array}{c}\left.\frac{\partial}{\partial\gamma^T}\chi\left(\lambda_\gamma,\kappa^*\right)\right|_{\gamma=\gamma(\eta)}\\ \left.\frac{\partial}{\partial\alpha^T}\chi\left(\lambda^*,\kappa_\alpha\right)\right|_{\alpha=\alpha(\eta)}\end{array}\right]=\left[\begin{array}{c}0\\0\end{array}\right]}$$

- Then,

$$\phi\left(Z;\eta\right)=\left.\frac{\partial\tau\left(\gamma,\alpha\right)}{\partial\left(\gamma^T,\alpha^T\right)}\right|_{(\gamma,\alpha)=(\gamma(\eta),\alpha(\eta))}\left[\begin{array}{c}\phi^\gamma\left(Z;\eta\right)\\ \phi^\alpha\left(Z;\eta\right)\end{array}\right]=0$$

- and therefore

$$\boxed{\sqrt{m}\left\{\chi\left(\widetilde{\eta}_2\right)-\chi\left(\eta^*\right)\right\}=o_p\left(1\right)}$$

- Consequently

$$\boxed{\sqrt{m}\left\{\widehat{\beta}_1-\beta\left(\eta\right)\right\}=\frac{1}{\sqrt{m}}\sum_{i=m+1}^n\psi\left(Z;\eta\right)+o_p\left(1\right)}$$

- We thus arrive at the conclusion that if under $F_{\eta,\vartheta}$ the working models for b and π are both correct, then

$$\boxed{\sqrt{n}\left\{\widehat{\beta}-\beta\left(\eta\right)\right\}=\frac{1}{\sqrt{n}}\sum_{i=m+1}^n\psi\left(Z;\eta\right)+o_p\left(1\right)}$$

- where

$$\psi\left(Z;\eta\right)=E_{\eta,\vartheta}\left[Y|R=1,X\right]+\frac{R}{P_{\eta,\vartheta}\left(R=1|X\right)}\left\{Y-E_{\eta,\vartheta}\left[Y|R=1,X\right]\right\}-\beta\left(F_{\eta,\vartheta}\right)$$

is the efficient influence function for $\beta\left(\eta\right)$ in models $F_{np}, \mathcal{F}_{sem,fixed}$ and $\mathcal{F}_{sem,par}$.

- Now, consider scenario (3) in which the model for π is correct but the model for b is incorrect . In such case,

$$\kappa^* \equiv \pi\left(\cdot, \alpha\left(\eta\right)\right) \text{ is equal to } \kappa = \pi_{F_{\eta,\vartheta}}\left(\cdot\right).$$

so

$$\chi\left(\lambda_\gamma, \kappa^*\right) = 0 \text{ for all } \gamma$$

- and consequently,

$$\left.\frac{\partial}{\partial\gamma^T}\chi\left(\lambda_\gamma, \kappa^*\right)\right|_{\gamma=\gamma(\eta)}$$

- Thus,

$$\begin{aligned}
\phi\left(Z;\eta\right) &= \left.\frac{\partial\tau\left(\gamma,\alpha\right)}{\partial\left(\gamma^T,\alpha^T\right)}\right|_{(\gamma,\alpha)=(\gamma(\eta),\alpha(\eta))}\begin{bmatrix}\phi^\gamma\left(Z;\eta\right)\\ \phi^\alpha\left(Z;\eta\right)\end{bmatrix}\\
&= \left[0, \left.\frac{\partial}{\partial\alpha^T}\chi\left(\lambda^*, \kappa_\alpha\right)\right|_{\alpha=\alpha(\eta)}\right]\begin{bmatrix}\phi^\gamma\left(Z;\eta\right)\\ \phi^\alpha\left(Z;\eta\right)\end{bmatrix}\\
&= \left.\frac{\partial}{\partial\alpha^T}\chi\left(\lambda^*, \kappa_\alpha\right)\right|_{\alpha=\alpha(\eta)}\phi^\alpha\left(Z;\eta\right)
\end{aligned}$$

- Now, recalling that $\chi\left(\lambda^*, \kappa_\alpha\right) \equiv \beta\left(\lambda^*\right) - \beta\left(\lambda\right) + E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda', \kappa_\alpha\right)\right]$ we obtain

$$\left.\frac{\partial}{\partial\alpha}\chi\left(\lambda^*, \kappa_\alpha\right)\right|_{\alpha=\alpha(\eta)} = \left.\frac{\partial}{\partial\alpha}E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda^*, \kappa_\alpha\right)\right]\right|_{\alpha=\alpha(\eta)}$$

- To calculate the derivative on the right hand side we note that

$$\beta\left(\lambda^*\right) - \beta\left(\lambda\right) + E_{\lambda,\kappa_\alpha,\vartheta}\left[\psi\left(Z;\lambda^*, \kappa_\alpha\right)\right] = 0 \text{ for all } \alpha$$

- Then, taking derivativative wrt α at $\alpha\left(\eta\right)$ in both sides of the last display, we have (under regularity conditions)

$$\begin{aligned}
0 &= \left.\frac{\partial}{\partial\alpha}E_{\lambda,\kappa_\alpha,\vartheta}\left[\psi\left(Z;\lambda^*, \kappa_\alpha\right)\right]\right|_{\alpha=\alpha(\eta)}\\
&= \left.\frac{\partial}{\partial\alpha}E_{\lambda,\kappa_\alpha,\vartheta}\left[\psi\left(Z;\lambda^*, \kappa\right)\right]\right|_{\alpha=\alpha(\eta)} + \left.\frac{\partial}{\partial\alpha}E_{\lambda,\kappa,,\vartheta}\left[\psi\left(Z;\lambda^*, \kappa_\alpha\right)\right]\right|_{\alpha=\alpha(\eta)}\\
&= E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda^*, \kappa\right)S_\alpha\left(Z;\alpha^*\right)\right] + \left.\frac{\partial\tau\left(\gamma\left(\eta\right),\alpha\right)}{\partial\alpha}\right|_{\alpha=\alpha(\eta)}
\end{aligned}$$

- Thus

$$\boxed{\left.\frac{\partial\tau(\gamma(\eta),\alpha)}{\partial\alpha^T}\right|_{\alpha=\alpha(\eta)} \equiv -E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda^*, \kappa\right)S_\alpha\left(Z;\alpha\left(\eta\right)\right)\right]}$$

- On the other hand, because $\widetilde{\alpha}_2$ is the maximum likelihood estimator of α under the logistic regression model for $P\left(R=1|X\right)$ then the influence function $\widetilde{\alpha}_2$ is

$$\begin{aligned}
\phi^\alpha\left(Z;\eta\right) &= -\left\{\partial E_{\eta,\vartheta}\left[S_\alpha\left(Z;\alpha\right)\right]/\partial\alpha^T\Big|_{\alpha=\alpha(\eta)}\right\}^{-1}S_\alpha\left(Z;\alpha\left(\eta\right)\right)\\
&= \left\{E_{\eta,\vartheta}\left[S_\alpha\left(Z;\alpha\left(\eta\right)\right)S_\alpha\left(Z;\alpha\left(\eta\right)\right)\right]\right\}^{-1}S_\alpha\left(Z;\alpha\left(\eta\right)\right)
\end{aligned}$$

- So finally,

$$\begin{aligned}
&\phi\left(Z;\eta\right) = \left.\frac{\partial\tau\left(\gamma\left(\eta\right),\alpha\right)}{\partial\alpha^T}\right|_{\alpha=\alpha(\eta)}\phi^\alpha\left(Z;\eta\right)\\
=\ & -E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda^*, \kappa\right)S_\alpha\left(Z;\alpha\left(\eta\right)\right)^T\right]\left\{E_{\lambda,\kappa,\vartheta}\left[S_\alpha\left(Z;\alpha\left(\eta\right)\right)S_\alpha\left(Z;\alpha\left(\eta\right)\right)\right]\right\}^{-1}S_\alpha\left(Z;\alpha\left(\eta\right)\right)\\
=\ & -\Pi\left[\psi\left(Z;\lambda^*, \kappa\right)|S_\alpha\left(Z;\alpha\left(\eta\right)\right)\right]
\end{aligned}$$

- Thus,

$$\boxed{\phi\left(Z;\eta\right) = -\Pi\left[\psi\left(Z;\lambda^*, \kappa\right)|S_\alpha\left(Z;\alpha\left(\eta\right)\right)\right]}$$

- We then arrive at

$$\begin{aligned}
&\sqrt{n}\left\{\widehat{\beta} - \beta\left(\eta\right)\right\} = \sum_{i=1}^n\left\{\varphi\left(Z_i;\eta\right) + \phi\left(Z_i;\eta\right)\right\}\\
=\ & \frac{1}{\sqrt{n}}\sum_{i=1}^n\left\{\psi\left(Z_i;\lambda^*, \kappa\right) - E_\eta\left[\psi\left(Z;\lambda^*, \kappa\right)\right] + \phi\left(Z;\eta\right)\right\}\\
=\ & \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\left\{\psi\left(Z_i;\lambda^*, \kappa\right) - E_\eta\left[\psi\left(Z;\lambda^*, \kappa\right)\right]\right\}\right]\\
& -\Pi\left[\left\{\psi\left(Z_i;\lambda^*, \kappa\right) - E_\eta\left[\psi\left(Z;\lambda^*, \kappa\right)\right]\right\}|S_\alpha\left(Z_i;\alpha\left(\eta\right)\right)\right]\\
=\ & \frac{1}{\sqrt{n}}\sum_{i=m+1}^n RESID\left\{\psi\left(Z_i;\lambda^*, \kappa\right) - E_\eta\left[\psi\left(Z;\lambda^*, \kappa\right)\right]\right\}
\end{aligned}$$

where

$$RESID\left(W\right) = W - \Pi\left[W|S_\alpha\left(Z;\alpha\left(\eta\right)\right)\right]$$

- Recalling the form of the efficient influence function in the missing data example we obtain that

$$\psi\left(Z;\lambda^{*},\kappa\right)=b\left(X;\gamma\left(\eta\right)\right)+\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}\left(Y-b\left(X;\gamma\left(\eta\right)\right)\right)-E_{F_{\eta,\vartheta}}\left[b\left(X;\gamma\left(\eta\right)\right)\right]$$

then

$$\begin{aligned}E_{\eta}\left[\psi\left(Z;\lambda^{*},\kappa\right)\right] &= E_{\eta}\left[\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}\left(Y-b\left(X;\gamma\left(\eta\right)\right)\right)\right]\\ &= E_{\eta}\left[b_{F_{\eta,\vartheta}}\left(X\right)-b\left(X;\gamma\left(\eta\right)\right)\right]\\ &= \beta\left(\eta\right)-E_{\eta}\left[b\left(X;\gamma\left(\eta\right)\right)\right]\end{aligned}$$

- We thus have

$$\begin{aligned}&\psi\left(Z;\lambda^{*},\kappa\right)-E_{\eta}\left[\psi\left(Z;\lambda^{*},\kappa\right)\right]\\ =\ & b\left(X;\gamma\left(\eta\right)\right)+\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}\left(Y-b\left(X;\gamma\left(\eta\right)\right)\right)-\beta\left(\eta\right)\\ =\ & \frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}\left\{Y-\beta\left(\eta\right)\right\}-\left\{\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}-1\right\}\left\{b\left(X;\gamma\left(\eta\right)\right)-\beta\left(\eta\right)\right\}\end{aligned}$$

- To summarize: the influence function of $\widehat{\beta}$ is
  - when the model for π is correct but the model for b is incorrect the influence function of $\widehat{\beta}$ is

$$\begin{aligned}&\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}\left\{Y-\beta\left(\eta\right)\right\}-\left\{\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}-1\right\}\left\{b\left(X;\gamma\left(\eta\right)\right)-\beta\left(\eta\right)\right\}\\ &-\Pi\left[\left\{\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}\left\{Y-\beta\left(\eta\right)\right\}-\left\{\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}-1\right\}\left\{b\left(X;\gamma\left(\eta\right)\right)-\beta\left(\eta\right)\right\}\right\}|S_{\alpha}\left(Z;\alpha\left(\eta\right)\right)\right]\end{aligned}$$

  - when the models for π and b are correct the influence function of $\widehat{\beta}$ is

$$\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}\left\{Y-\beta\left(\eta\right)\right\}-\left\{\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}-1\right\}\left\{b\left(X;\gamma\left(\eta\right)\right)-\beta\left(\eta\right)\right\}$$

- The preceding result has the following consequence, which at first sight, may appear counterintuitive.
- Suppose two investigators will analyze the same data from the following two stage study design. At the first stage of the study X was measured on random sample and at the second stage a random subsample was selected with prob. $\pi^{*}\left(X\right)$ and Y was measured on this subsample.
- The investigators will effectively be analyzing the data under model $\mathcal{F}_{sem,fix}$.
- The first investigator will compute the one step estimator of $\beta\left(\eta\right)=E_{\eta}\left[E_{\eta}\left(Y|R=1,X\right)\right]$ assuming an incorrect model for b and a model for π that specifies that $\pi_{F_{\eta,\vartheta}}$ is known an equal to $\pi^{*}\left(X\right)$, that is, the model for $\pi_{F_{\eta,\vartheta}}$ includes just one probability $\pi^{*}\left(X\right)$. Call $\widehat{\beta}^{fix}$ to this investigator's estimator.
- The second investigator will compute the one step estimator of $\beta\left(\eta\right)=E_{\eta}\left[E_{\eta}\left(Y|R=1,X\right)\right]$ assuming the same incorrect model for b but will assume a correctly specified model for $\pi_{F_{\eta,\vartheta}}$ that is of the form

$$\log\left[\frac{\pi\left(X;\alpha\right)}{1-\pi\left(X;\alpha\right)}\right]=\log\left[\frac{\pi^{*}\left(X\right)}{1-\pi^{*}\left(X\right)}\right]+\alpha^{T}\widetilde{X}$$

- Note that this model is correctly specified with true parameter value $\alpha=0$. Call $\widehat{\beta}^{par}$ to this investigator's estimator $\widehat{\beta}^{par}$.

- In view of the preceding discussion,

$$\sqrt{n}\left\{\widehat{\beta}^{fix}-\beta\left(\eta\right)\right\}\overset{D\left(F_{\eta,\vartheta}\right)}{\to}N\left(0,V_{fix}\left(\eta\right)\right)$$

and

$$\sqrt{n}\left\{\widehat{\beta}^{par}-\beta\left(\eta\right)\right\}\overset{D\left(F_{\eta,\vartheta}\right)}{\to}N\left(0,V_{par}\left(\eta\right)\right)$$

where

$$\begin{aligned}V_{par}\left(\eta\right) &= var_{F_{\eta,\vartheta}}\left[\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}\left\{Y-\beta\left(\eta\right)\right\}-\left\{\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}-1\right\}\left\{b\left(X;\gamma\left(\eta\right)\right)-\beta\left(\eta\right)\right\}\right.\\ &\left.-\Pi\left[\left\{\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}\left\{Y-\beta\left(\eta\right)\right\}-\left\{\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}-1\right\}\left\{b\left(X;\gamma\left(\eta\right)\right)-\beta\left(\eta\right)\right\}\right\}|S_{\alpha}\left(Z;\alpha\left(\eta\right)\right)\right]\right]\\ &\leq var_{F_{\eta,\vartheta}}\left[\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}\left\{Y-\beta\left(\eta\right)\right\}-\left\{\frac{R}{\pi_{F_{\eta,\vartheta}}\left(X\right)}-1\right\}\left\{b\left(X;\gamma\left(\eta\right)\right)-\beta\left(\eta\right)\right\}\right]\\ &= V_{fix}\left(\eta\right)\end{aligned}$$

Here

$$S_{\alpha}\left(Z;\alpha\left(\eta\right)\right)=\widetilde{X}\left(R-\pi\left(X;\alpha\left(\eta\right)\right)\right)$$

- So, estimating the missingness probability $\pi\left(X\right)$ even when it is known, can never decrease the asymptotic precision with which one estimates $\beta\left(\eta\right)$

- This counterintuitive result is justified by the following remarks
  - The general belief that estimation of nuisance parameters cannot decrease the variance with which one estimates a parameter of interest, is correct ONLY when one compares the asymptotic variance of **efficient estimators of the parameter of interest** under models that assume that the nuisance parameter is known vs unknwon
  - **Neither $\widehat{\beta}^{fix}$ is efficient, i.e be RAL with efficient influence function, in the model F_{fix} nor $\widehat{\beta}^{par}$ is efficient in model $\mathcal{F}_{par}$**. We can see this immediately by noticing that neither have efficient influence function. But in fact, we could have seen this even without any calculation by just noticing that the consistence of $\widehat{\beta}^{fix}$ and $\widehat{\beta}^{par}$ depends on the correct specification of the respective models for π [see in our earlier discussion of the scenario (1)] and any estimator whose consistency depends on the correct specification of the model for the ancillary parameter π cannot be efficient. Recall that the three models $\mathcal{F}_{np}, \mathcal{F}_{par}$ and $\mathcal{F}_{fix}$ are factorized likelihood models, the parameter of interest $\beta(F)$ depends on the first factor of the likelihood and the parameter π enters into the second factor only. We know that the efficient influence function is the same regardless of whether the missingness model is known or unknown, thus any estimator whose consistency depends on the correct specification of the missingness probabilities must must fail to have efficient influence function.
  - Note that if the model for b is correct, $\widehat{\beta}^{par}$ is not asymptotically more precise than $\widehat{\beta}^{fix}$, in fact, both estimators have the same asymptotic variance because they both have the same influence function $\psi(Z;\eta)$. But this does not represent a contradiction to remarks 1 and 2, because when b is correctly modeled, $\widehat{\beta}^{fix}$ and $\widehat{\beta}^{par}$ are indeed RAL with **efficient influence function.**

- To develop intuition about this result, consider a two - stage sampling design with $\pi^*(X) = 1/2$.
- Suppose that X is just a binary variable that is highly predictive of the possibly missing outcome Y^f.
- For didactic reasons, we will consider, the "in-sample" one-step estimators $\widetilde{\beta}^{fix}$ and $\widetilde{\beta}^{par}$,

$$\widetilde{\beta}^{fix} = \beta\left(\widetilde{\eta}^{fix}\right) + \frac{1}{n}\sum_{i=1}^{n}\psi\left(Z_i;\widetilde{\eta}^{fix}\right) \text{ and } \widetilde{\beta}^{par} = \beta\left(\widetilde{\eta}^{par}\right) + \frac{1}{n}\sum_{i=1}^{n}\psi\left(Z_i;\widetilde{\eta}^{par}\right)$$

where $\widetilde{\eta}^{fix} = \left(\widetilde{b}, \pi^*, \widehat{F}_{n,X}\right)$, $\widetilde{\eta}^{por} = \left(\widetilde{b}, \widetilde{\pi}, \widehat{F}_{n,X}\right)$, $\widetilde{\pi}(X) = \pi(X;\widetilde{\alpha})$ where $\pi(X;\alpha)$ =expit$\left(\widetilde{X}^T\alpha\right)$, $\widetilde{\alpha}$ solves $\sum_{i=1}^{n} X_i\left(R_i - expit\left(\widetilde{X}_i^T\alpha\right)\right) = 0$ and for simplicity we will take $\widetilde{b}(X) = 0$.

- Note also, that because we are taking $\widetilde{b}(X) = 0$, the estimators $\widetilde{\beta}^{fix}$ and $\widetilde{\beta}^{par}$ are the so-called inverse probability weighted estimators (see slide 25)

$$\widetilde{\beta}^{fix} = \frac{1}{n}\sum_{i=1}^{n} Y_i R_i/(1/2) \text{ and } \widetilde{\beta}^{par} = \frac{1}{n}\sum_{i=1}^{n} Y_i R_i/\widetilde{\pi}(X_i)$$

- Also, because X is binary, $\widetilde{\pi}(x=j) = \#\{i : R_i = 1, X_i = j\}/n_j$ where $n_j = \#\{i : X_i = j\}$, $j = 0, 1$.

- Now, it can be shown that $\widetilde{\beta}^{fix} - \widehat{\beta}^{fix} = o_p(1)$ and $\widetilde{\beta}^{par} - \widehat{\beta}^{par} = o_p(1)$. (in sample estimation does not affect asymptotic behaviour essentially because in this problem, $\psi(\cdot;\widetilde{\eta})$ and $\psi(\cdot;\eta)$ fall in a Donsker class), so the empirical process term A_n is $o_p(1)$ even if $\widetilde{\eta}$ is from the same sample).
- Recall that $\pi^*(X) = 1/2$. However, suppose that by the luck of the draw, of the n_0 subjects with $X = 0$ only 40% are selected into the second stage and of the n_1 subjects with $X = 1$, 60% are selected into the second stage. Then, $\widetilde{\pi}(x=0) = 0.4$ and $\widetilde{\pi}(x=1) = 0.6$. Thus, in $\widetilde{\beta}^{par}$ the $0.4n_0$ subjects with $X = 0$ selected to the second stage are weighted by 1/0.4 to account for those $n_0 - 0.4n_0$ subjects that were not selected, so effectively creating a "pseudo-sample" of $0.4n_0 \times 1/0.4 = n_0$ subjects with $X = 0$. Likewise, in $\widetilde{\beta}^{par}$ the subjects with $X = 1$ are weighted so as to effectively create a "pseudo-sample" with n_1 subjects with $X = 1$. Thus, weighting by $\widetilde{\pi}$ balance out chance imbalances in the covariate X.
- In contrast, $\widetilde{\beta}^{fix}$ effectively creates pseudo samples of $0.4n_0/(1/2) = 0.8n_0$ subjects with $X = 0$ and $0.6n_1/(1/2) = 1.2n_1$ subjects with $X = 1$, so chance imbalances are not corrected.

- Turn now to the last scenario, i.e. scenario (4) in which the model for b is correct but the model for π is incorrect.
- In this case, we have, in complete symmetry to scenario (3)

$$\lambda^* \equiv (b(\cdot, \gamma(\eta)), F_{\eta,\vartheta}) \text{ is equal to } \lambda \equiv \left(b_{F_{\eta,\vartheta}}(\cdot), F_{\eta,\vartheta}\right).$$

so

$$\chi(\lambda^*, \kappa_\alpha) = 0 \text{ for all } \alpha$$

- and consequently,

$$\frac{\partial}{\partial \alpha^T} \chi(\lambda^*, \kappa_\alpha)\Big|_{\alpha=\alpha(\eta)} = 0$$

- Thus,

$$\begin{aligned}
\phi(Z;\eta) &= \frac{\partial \tau(\gamma,\alpha)}{\partial(\gamma^T, \alpha^T)}\Big|_{(\gamma,\alpha)=(\gamma(\eta),\alpha(\eta))} \begin{bmatrix} \phi^\gamma(Z;\eta) \\ \phi^\alpha(Z;\eta) \end{bmatrix} \\
&= \left[\frac{\partial}{\partial \gamma^T} \chi(\lambda_\gamma, \kappa^*)\Big|_{\gamma=\gamma(\eta)}, 0\right] \begin{bmatrix} \phi^\gamma(Z;\eta) \\ \phi^\alpha(Z;\eta) \end{bmatrix} \\
&= \frac{\partial}{\partial \gamma^T} \chi(\lambda_\gamma, \kappa^*)\Big|_{\gamma=\gamma(\eta)} \phi^\gamma(Z;\eta)
\end{aligned}$$

- However, unlike the case (3) there is not more structure to $\frac{\partial}{\partial \gamma^T} \chi(\lambda_\gamma, \kappa^*)\Big|_{\gamma=\gamma(\eta)} \phi^\gamma(Z;\eta)$ so we can't elaborate further on the form of the influence function.

- We thus conclude that

$$\begin{aligned}
\sqrt{n}\left\{\widehat{\beta} - \beta(\eta)\right\} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[\{\psi(Z_i;\lambda,\kappa^*) - E_\eta[\psi(Z;\lambda,\kappa^*)]\} + \\
&\qquad \frac{\partial}{\partial \gamma^T} \chi(\lambda_\gamma, \kappa^*)\Big|_{\gamma=\gamma(\eta)} \phi^\gamma(Z_i;\eta)\Big]
\end{aligned}$$

where $\phi(Z;\eta)$ is the influence function of $\widetilde{\gamma}$

- Close attention to our derivations in scenarios (2) and (3) reveals that we can draw conclussion not just for the missing data example but more generally for one step sample split estimators. Specifically, the following result holds:
- Proposition 8.1: Suppose that
  - $\mathcal{F} = \{f(z;\eta,\vartheta) = g_1(z;\lambda,\vartheta)\, g_1(z;\kappa) : \eta = (\lambda,\kappa), \lambda \in \mathbb{L}, \kappa \in \mathbb{K}, \vartheta \in O\}$
  - the parameter of interest $\beta\left(F_{\eta,\vartheta}\right)$ satisfies

$$\boxed{\beta\left(F_{\eta,\vartheta}\right) \text{ depends on } (\eta,\vartheta) \text{ only through } \lambda}$$

  - a gradient $\psi_{F_{\eta,\vartheta}}(Z)$ of $\beta\left(F_{\eta,\vartheta}\right)$ satisfies

$$\psi_{F_{\eta,\vartheta}}(Z) \text{ depends on } (\eta,\vartheta) \text{ only through } \eta, \text{ so we write it as } \psi(Z;\lambda,\kappa)$$

  - $\chi\left(\lambda',\kappa'\right) \equiv \beta\left(\lambda'\right) - \beta(\lambda) + E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda',\kappa'\right)\right]$ satisfies

$$\boxed{\chi\left(\lambda,\kappa'\right) \equiv \beta(\lambda) - \beta(\lambda) + E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda,\kappa'\right)\right] = 0 \text{ for all } \lambda \text{ and all } \kappa'}$$

$$\boxed{\chi\left(\lambda',k\right) \equiv \beta\left(\lambda'\right) - \beta(\lambda) + E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda',\kappa\right)\right] = 0 \text{ for all } \lambda' \text{ and all } \kappa}$$

  - $\widetilde{\eta} = (\lambda_{\widetilde{\gamma}}, \kappa_{\widetilde{\alpha}})$ where $\widetilde{\gamma}$ is an estimator of a parameter γ indexing a parametric model λ_γ for λ and $\widetilde{\alpha}$ is the MLE of α indexing a parametric model κ_α for κ.
  - $\widetilde{\gamma}$ is asymptotically linear under $F_{\eta,\vartheta}$

- Let, $\widehat{\beta}$ be the one step sample split estimator, that is, with $m = \lfloor n/2 \rfloor$

$$\widehat{\beta} \equiv \frac{m}{n}\left[\beta\left(\lambda_{\widetilde{\gamma}_2}\right) + \frac{1}{m}\sum_{i=1}^{m} \psi\left(Z_i;\lambda_{\widetilde{\gamma}_2},\kappa_{\widetilde{\alpha}_2}\right)\right] + \frac{n-m}{n}\left[\beta\left(\lambda_{\widetilde{\gamma}_1}\right) + \frac{1}{m}\sum_{i=1}^{m} \psi\left(Z_i;\lambda_{\widetilde{\gamma}_1},\kappa_{\widetilde{\alpha}_1}\right)\right]$$

where $\widetilde{\gamma}_j$ and $\widetilde{\alpha}_j, j = 1, 2$ denote the estimators $\widetilde{\gamma}$ and $\widetilde{\alpha}$ computed from the j^{th} half of the sample.

- Under (1)-(6) it holds that
  - if the models λ_γ and κ_α are both correct under $F_{\eta,\vartheta}$, then

$$\sqrt{n}\left\{\widehat{\beta} - \beta(\lambda)\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi(Z_i;\lambda,\kappa) + o_p(1)$$

  - if the model for κ_α is correct but the model for λ_γ is incorrect, then with $\alpha(\eta) \equiv$plim$\widetilde{\alpha}$, $\lambda^* \equiv \lambda_{\gamma(\eta)}$ where $\gamma(\eta) \equiv$plim$\widetilde{\gamma}$ and $S_\alpha(Z;\alpha(\eta))$ denoting the score for α at $\alpha(\eta)$, it holds that

$$\begin{aligned}
\sqrt{n}\left\{\widehat{\beta} - \beta(\lambda)\right\} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\left\{\psi\left(Z_i;\lambda^*,\kappa\right) - E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda^*,\kappa\right)\right]\right\} \\
&\quad -\Pi\left[\left\{\psi\left(Z_i;\lambda^*,\kappa\right) - E_{\lambda,\kappa,\vartheta}\left[\psi\left(Z;\lambda^*,\kappa\right)\right]\right\} | S_\alpha(Z;\alpha(\eta))\right]\Big] + o_p(1)
\end{aligned}$$

  - if the model for λ_γ is correct but the model for κ_α is incorrect, then with $\alpha(\eta) \equiv$plim$\widetilde{\alpha}$, $\lambda^* \equiv \lambda_{\gamma(\eta)}$ where $\gamma(\eta) \equiv$plim$\widetilde{\gamma}$ and $\phi^\gamma(Z;\eta)$ denoting the influence function of $\widetilde{\gamma}$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\left\{\psi\left(Z_i;\lambda,\kappa^*\right) - E_\eta\left[\psi\left(Z;\lambda,\kappa^*\right)\right]\right\} + \frac{\partial}{\partial \gamma^T}\chi\left(\lambda_\gamma,\kappa^*\right)\Big|_{\gamma=\gamma(\eta)} \phi^\gamma(Z_i;\eta)\right]$$