1 PLR model, residualization and population moment

We consider the partially linear regression (PLR) model. For $i = 1, \ldots, n$, we observe

$$W_i = (Y_i, D_i, X_i),$$

with $Y_i \in \mathbb{R}$, $D_i \in \mathbb{R}$, and $X_i \in \mathbb{R}^p$. The structural model is

$$Y = \theta_0 D + g_0(X) + \varepsilon, \qquad \mathbb{E}[\varepsilon \mid D, X] = 0, \tag{1}$$

where $\theta_0 \in \mathbb{R}$ is the parameter of interest and $g_0 : \mathbb{R}^p \to \mathbb{R}$ is an unknown nuisance function.

Define the nuisance regression functions

$$m_0(X) := \mathbb{E}[D \mid X], \tag{2}$$

$$\ell_0(X) := \mathbb{E}[Y \mid X]. \tag{3}$$

Using (1), we express $\ell_0(X)$ explicitly. For each x,

$$\begin{aligned}
\ell_0(x) &= \mathbb{E}[Y \mid X = x] \\
&= \mathbb{E}\big[\theta_0 D + g_0(X) + \varepsilon \mid X = x\big] \\
&= \theta_0 \mathbb{E}[D \mid X = x] + \mathbb{E}[g_0(X) \mid X = x] + \mathbb{E}[\varepsilon \mid X = x].
\end{aligned} \tag{4}$$

We have

$$\mathbb{E}[D \mid X = x] = m_0(x), \qquad \mathbb{E}[g_0(X) \mid X = x] = g_0(x),$$

and, using iterated expectations,

$$\mathbb{E}[\varepsilon \mid X] = \mathbb{E}\big[\mathbb{E}[\varepsilon \mid D, X] \mid X\big] = \mathbb{E}[0 \mid X] = 0. \tag{5}$$

Substituting into (4),

$$\ell_0(x) = \theta_0 m_0(x) + g_0(x). \tag{6}$$

Residualization

Define the residualized variables

$$\tilde{D} := D - m_0(X), \qquad \tilde{Y} := Y - \ell_0(X). \tag{7}$$

Substitute (1) and (6) into $\tilde{Y}$:

$$
\begin{aligned}
\tilde{Y} &= Y - \ell_0(X) \\
&= (\theta_0 D + g_0(X) + \varepsilon) - (\theta_0 m_0(X) + g_0(X)) \\
&= \theta_0(D - m_0(X)) + \varepsilon \\
&= \theta_0 \tilde{D} + \varepsilon.
\end{aligned}
\tag{8}
$$

Thus, after residualization,

$$
\tilde{Y} - \theta_0 \tilde{D} = \varepsilon.
\tag{9}
$$

Population moment condition

Multiply both sides of (9) by $\tilde{D}$:

$$
(\tilde{Y} - \theta_0 \tilde{D})\tilde{D} = \varepsilon \tilde{D}.
\tag{10}
$$

Take expectations:

$$
\mathbb{E}\big[(\tilde{Y} - \theta_0 \tilde{D})\tilde{D}\big] = \mathbb{E}[\varepsilon \tilde{D}].
\tag{11}
$$

We now show $\mathbb{E}[\varepsilon \tilde{D}] = 0$. Using iterated expectations,

$$
\begin{aligned}
\mathbb{E}[\varepsilon \tilde{D}] &= \mathbb{E}\big[\mathbb{E}[\varepsilon \tilde{D} \mid X]\big] \\
&= \mathbb{E}\big[\mathbb{E}[\varepsilon(D - m_0(X)) \mid X]\big] \\
&= \mathbb{E}\Big(\mathbb{E}[\varepsilon D \mid X] - m_0(X)\mathbb{E}[\varepsilon \mid X]\Big).
\end{aligned}
\tag{12}
$$

Consider each term separately.

First term:

$$
\begin{aligned}
\mathbb{E}[\varepsilon D \mid X] &= \mathbb{E}\big[D\varepsilon \mid X\big] \\
&= \mathbb{E}\big[\mathbb{E}[D\varepsilon \mid D, X] \mid X\big] \\
&= \mathbb{E}\big[D\mathbb{E}[\varepsilon \mid D, X] \mid X\big] \\
&= \mathbb{E}[D \cdot 0 \mid X] \\
&= 0.
\end{aligned}
\tag{13}
$$

Second term:

$$
m_0(X)\mathbb{E}[\varepsilon \mid X] = m_0(X) \cdot 0 = 0.
\tag{14}
$$

Substituting (13) and (14) into (12),

$$\mathbb{E}[\varepsilon \tilde{D}] = \mathbb{E}[0 - 0] = 0. \tag{15}$$

Therefore, from (11) and (15),

$$\mathbb{E}\big[(\tilde{Y} - \theta_0 \tilde{D})\tilde{D}\big] = 0. \tag{16}$$

Remark 1. Equation (16) is the population normal equation for the PLR parameter θ_0 after residualizing both Y and D on X.

2 Orthogonal score and Neyman orthogonality

Score definition

We introduce generic nuisance functions

$$g : \mathbb{R}^p \to \mathbb{R}, \qquad m : \mathbb{R}^p \to \mathbb{R},$$

and set $\eta := (g, m)$. The PLR score is defined as

$$\psi(W; \theta, \eta) := (D - m(X))(Y - g(X) - \theta(D - m(X))). \tag{17}$$

Denote the associated population moment by

$$\Psi(\theta, \eta) := \mathbb{E}\big[\psi(W; \theta, \eta)\big]. \tag{18}$$

At the truth, we set

$$\eta_0 := (g_0^\star, m_0), \qquad g_0^\star(X) := \ell_0(X) = \mathbb{E}[Y \mid X], \qquad m_0(X) := \mathbb{E}[D \mid X].$$

(Here $g_0^\star$ is the conditional mean of Y given X; it is not the structural g_0 from (1) but satisfies $\ell_0(X) = \theta_0 m_0(X) + g_0(X)$.)

Substitute (θ_0, η_0) into (17):

$$\psi(W; \theta_0, \eta_0) = (D - m_0(X))\Big(Y - g_0^\star(X) - \theta_0\big(D - m_0(X)\big)\Big). \tag{19}$$

Use $g_0^\star(X) = \ell_0(X)$ and (8),

$$Y - \ell_0(X) = \theta_0(D - m_0(X)) + \varepsilon.$$

3

Therefore,

$$Y - g_0^\star(X) - \theta_0(D - m_0(X)) = (Y - \ell_0(X)) - \theta_0(D - m_0(X))$$
$$= \varepsilon. \tag{20}$$

Substituting (20) into (19),

$$\psi(W; \theta_0, \eta_0) = (D - m_0(X))\varepsilon. \tag{21}$$

Thus the population moment at (θ_0, η_0) is

$$\Psi(\theta_0, \eta_0) = \mathbb{E}[(D - m_0(X))\varepsilon] = 0, \tag{22}$$

using exactly the same argument as in (15).

Neyman orthogonality

Let $\eta = (g, m)$ belong to a function space. For a direction $h = (h_g, h_m)$, consider the path

$$\eta_t := \eta_0 + th = \left(g_0^\star + th_g, \ m_0 + th_m\right), \qquad t \in \mathbb{R}, \tag{23}$$

and define

$$\Phi(t) := \Psi(\theta_0, \eta_t). \tag{24}$$

[Neyman orthogonality] The score ψ is Neyman-orthogonal at (θ_0, η_0) if

$$\left.\frac{d}{dt}\Psi(\theta_0, \eta_t)\right|_{t=0} = 0 \quad \text{for all directions } h = (h_g, h_m).$$

We now compute the derivative explicitly.

For brevity, write

$$U_t := D - m_t(X) = D - m_0(X) - th_m(X), \tag{25}$$

$$V_t := Y - g_t(X) - \theta_0(D - m_t(X)). \tag{26}$$

Then

$$\psi(W; \theta_0, \eta_t) = U_t V_t. \tag{27}$$

We first expand V_t:

$$V_t = Y - (g_0^\star(X) + th_g(X)) - \theta_0\Big(D - (m_0(X) + th_m(X))\Big)$$
$$= (Y - g_0^\star(X)) - th_g(X) - \theta_0(D - m_0(X) - th_m(X))$$
$$= (Y - g_0^\star(X) - \theta_0(D - m_0(X))) - th_g(X) + \theta_0 th_m(X). \tag{28}$$

Using (20), the term in parentheses is ε, so

$$V_t = \varepsilon - th_g(X) + \theta_0 th_m(X). \tag{29}$$

Differentiate (25) and (29) with respect to t:

$$\frac{d}{dt}U_t = -h_m(X), \tag{30}$$

$$\frac{d}{dt}V_t = -h_g(X) + \theta_0 h_m(X). \tag{31}$$

Using the product rule on (27),

$$\frac{d}{dt}\psi(W; \theta_0, \eta_t) = \frac{dU_t}{dt}V_t + U_t\frac{dV_t}{dt}. \tag{32}$$

Evaluate at $t = 0$:

$$\frac{d}{dt}\psi(W; \theta_0, \eta_t)\Big|_{t=0} = (-h_m(X))\,V_0 + U_0\big(-h_g(X) + \theta_0 h_m(X)\big)$$
$$= -h_m(X)\varepsilon + (D - m_0(X))\big(-h_g(X) + \theta_0 h_m(X)\big), \tag{33}$$

since $U_0 = D - m_0(X)$ and $V_0 = \varepsilon$ by (29).

Taking expectations,

$$\frac{d}{dt}\Psi(\theta_0, \eta_t)\Big|_{t=0} = -\mathbb{E}\big[h_m(X)\varepsilon\big] - \mathbb{E}\big[(D - m_0(X))h_g(X)\big] + \theta_0\mathbb{E}\big[(D - m_0(X))h_m(X)\big]. \tag{34}$$

We now show that each expectation is zero.

First term:

$$\mathbb{E}[h_m(X)\varepsilon] = \mathbb{E}\big[\mathbb{E}[h_m(X)\varepsilon \mid X]\big]$$
$$= \mathbb{E}\big[h_m(X)\mathbb{E}[\varepsilon \mid X]\big]$$
$$= \mathbb{E}[h_m(X) \cdot 0] = 0, \tag{35}$$

using (5).

Second and third terms: for any measurable function $a(X)$,

$$
\begin{aligned}
\mathbb{E}\big[(D - m_0(X))a(X)\big] &= \mathbb{E}\big[\mathbb{E}[(D - m_0(X))a(X) \mid X]\big] \\
&= \mathbb{E}\big[a(X)\mathbb{E}[D - m_0(X) \mid X]\big] \\
&= \mathbb{E}\big[a(X)(\mathbb{E}[D \mid X] - m_0(X))\big] \\
&= \mathbb{E}\big[a(X)(m_0(X) - m_0(X))\big] \\
&= 0.
\end{aligned}
\tag{36}
$$

Hence both $\mathbb{E}[(D - m_0(X))h_g(X)]$ and $\mathbb{E}[(D - m_0(X))h_m(X)]$ are zero.

Substituting (35) and (36) into (34),

$$
\frac{d}{dt}\Psi(\theta_0, \eta_t)\bigg|_{t=0} = 0.
\tag{37}
$$

Lemma 1 (Neyman orthogonality)**.** *The score* (17) *is Neyman-orthogonal at* (θ_0, η_0)*, i.e.* (37) *holds for all directions* $h = (h_g, h_m)$*.*

3 Empirical score, DML estimator, Jacobian, and condition number

Let $\hat{g}$ and $\hat{m}$ be *cross-fitted* estimators of $g_0^\star$ and m_0 (trained on folds not containing observation i). Define the empirical residuals

$$
\hat{U}_i := D_i - \hat{m}(X_i), \qquad \hat{V}_i := Y_i - \hat{g}(X_i).
\tag{38}
$$

The empirical score average is

$$
\Psi_n(\theta, \eta) := \frac{1}{n}\sum_{i=1}^{n} \psi(W_i; \theta, \eta).
\tag{39}
$$

For $\hat{\eta} := (\hat{g}, \hat{m})$, use (17) and (38):

$$
\begin{aligned}
\psi(W_i; \theta, \hat{\eta}) &= (D_i - \hat{m}(X_i))\big(Y_i - \hat{g}(X_i) - \theta(D_i - \hat{m}(X_i))\big) \\
&= \hat{U}_i(\hat{V}_i - \theta\hat{U}_i).
\end{aligned}
\tag{40}
$$

Therefore,

$$
\begin{aligned}
\Psi_n(\theta, \hat{\eta}) &= \frac{1}{n}\sum_{i=1}^{n} \hat{U}_i(\hat{V}_i - \theta\hat{U}_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} \hat{U}_i\hat{V}_i - \theta\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i^2.
\end{aligned}
\tag{41}
$$

DML estimator as a Z-estimator

The DML estimator $\hat{\theta}$ solves the empirical moment condition

$$\Psi_n(\hat{\theta}, \hat{\eta}) = 0. \tag{42}$$

Substitute (41) with $\theta = \hat{\theta}$:

$$0 = \frac{1}{n} \sum_{i=1}^{n} \hat{U}_i \hat{V}_i - \hat{\theta} \frac{1}{n} \sum_{i=1}^{n} \hat{U}_i^2. \tag{43}$$

Multiply both sides of (43) by n:

$$0 = \sum_{i=1}^{n} \hat{U}_i \hat{V}_i - \hat{\theta} \sum_{i=1}^{n} \hat{U}_i^2. \tag{44}$$

Rearrange to solve for $\hat{\theta}$:

$$\hat{\theta} = \frac{\sum_{i=1}^{n} \hat{U}_i \hat{V}_i}{\sum_{i=1}^{n} \hat{U}_i^2}. \tag{45}$$

Empirical Jacobian

Differentiate (40) with respect to θ:

$$\begin{aligned}
\partial_\theta \psi(W_i; \theta, \hat{\eta}) &= \partial_\theta \left(\hat{U}_i (\hat{V}_i - \theta \hat{U}_i) \right) \\
&= \hat{U}_i (-\hat{U}_i) \\
&= -\hat{U}_i^2.
\end{aligned} \tag{46}$$

Then

$$\begin{aligned}
\partial_\theta \Psi_n(\theta, \hat{\eta}) &= \frac{1}{n} \sum_{i=1}^{n} \partial_\theta \psi(W_i; \theta, \hat{\eta}) \\
&= \frac{1}{n} \sum_{i=1}^{n} (-\hat{U}_i^2) \\
&= -\frac{1}{n} \sum_{i=1}^{n} \hat{U}_i^2.
\end{aligned} \tag{47}$$

This derivative does not depend on θ, so we define the *empirical Jacobian*

$$\hat{J}_\theta := \partial_\theta \Psi_n(\theta, \hat{\eta}) = -\frac{1}{n} \sum_{i=1}^{n} \hat{U}_i^2. \tag{48}$$

Since $\hat{U}_i^2 \geq 0$, we have $\hat{J}_\theta \leq 0$.

DML condition number

We define the DML condition number as

$$\kappa_{\mathrm{DML}} := -\frac{1}{\hat{J}_\theta} = \frac{1}{\left|\hat{J}_\theta\right|} = \frac{n}{\sum_{i=1}^n \hat{U}_i^2}. \tag{49}$$

By construction, $\kappa_{\mathrm{DML}} > 0$ whenever $\sum \hat{U}_i^2 > 0$.

Remark 2. Small $\sum \hat{U}_i^2$ (weak residual variation in D) implies large κ_{DML}, which corresponds to a nearly flat score and a sensitive estimator.

4 Refined linearization of the DML estimator

Define the empirical moment

$$\Psi_n(\theta, \eta) := \frac{1}{n} \sum_{i=1}^n \psi(W_i; \theta, \eta).$$

At the truth and the estimated nuisances, define

$$S_n := \Psi_n(\theta_0, \eta_0) = \frac{1}{n} \sum_{i=1}^n \psi(W_i; \theta_0, \eta_0), \tag{50}$$

$$B_n := \Psi_n(\theta_0, \hat{\eta}) - \Psi_n(\theta_0, \eta_0), \tag{51}$$

and let R_n denote the higher-order remainder to be specified.

Regularity conditions

Assumption 1 (Regularity)**.**

(i) **(Score regularity)** For $\tilde{\theta}$ between $\hat{\theta}$ and θ_0,

$$\partial_\theta \Psi_n(\tilde{\theta}, \hat{\eta}) = \hat{J}_\theta + o_P(1).$$

In PLR with (17), this holds exactly with $\partial_\theta \Psi_n(\tilde{\theta}, \hat{\eta}) = \hat{J}_\theta$.

(ii) **(Non-degeneracy)** There exist $c_J > 0$ and $\delta_J \in (0,1)$ such that

$$\mathbb{P}(|\hat{J}_\theta| \geq c_J) \geq 1 - \delta_J.$$

(iii) **(Nuisance rate)** The nuisance estimators satisfy

$$\|\hat{m} - m_0\|_{L^2} \cdot \|\hat{g} - g_0^\star\|_{L^2} = o_P(n^{-1/2}).$$

(iv) **(Moment bounds)** For the score at the truth,

$$\mathbb{E}[\psi(W; \theta_0, \eta_0)^2] =: \sigma_\psi^2 < \infty, \qquad \mathbb{E}[|\psi(W; \theta_0, \eta_0)|^3] \le M_3 < \infty.$$

Linearization

Since $\hat{\theta}$ solves $\Psi_n(\hat{\theta}, \hat{\eta}) = 0$, perform a first-order Taylor expansion around θ_0:

$$\Psi_n(\hat{\theta}, \hat{\eta}) = \Psi_n(\theta_0, \hat{\eta}) + \partial_\theta \Psi_n(\tilde{\theta}, \hat{\eta})(\hat{\theta} - \theta_0), \tag{52}$$

for some random $\tilde{\theta}$ between $\hat{\theta}$ and θ_0. Using Assumption 1(i), write

$$\partial_\theta \Psi_n(\tilde{\theta}, \hat{\eta}) = \hat{J}_\theta + r_{n,\theta}, \qquad r_{n,\theta} = o_P(1). \tag{53}$$

Thus (52) becomes

$$\begin{aligned} 0 &= \Psi_n(\theta_0, \hat{\eta}) + (\hat{J}_\theta + r_{n,\theta})(\hat{\theta} - \theta_0) \\ &= \Psi_n(\theta_0, \hat{\eta}) + \hat{J}_\theta(\hat{\theta} - \theta_0) + r_{n,\theta}(\hat{\theta} - \theta_0). \end{aligned} \tag{54}$$

Rearrange (54):

$$\hat{J}_\theta(\hat{\theta} - \theta_0) = -\Psi_n(\theta_0, \hat{\eta}) - r_{n,\theta}(\hat{\theta} - \theta_0). \tag{55}$$

Assumption 1(ii) ensures that $\hat{J}_\theta$ is bounded away from zero with high probability, so we can divide by $\hat{J}_\theta$:

$$\hat{\theta} - \theta_0 = -\hat{J}_\theta^{-1} \Psi_n(\theta_0, \hat{\eta}) - \hat{J}_\theta^{-1} r_{n,\theta}(\hat{\theta} - \theta_0). \tag{56}$$

Define the remainder

$$R_n := -\hat{J}_\theta^{-1} r_{n,\theta}(\hat{\theta} - \theta_0). \tag{57}$$

Then (56) is

$$\hat{\theta} - \theta_0 = -\hat{J}_\theta^{-1} \Psi_n(\theta_0, \hat{\eta}) + R_n. \tag{58}$$

Next, decompose the score at θ_0:

$$\begin{aligned} \Psi_n(\theta_0, \hat{\eta}) &= \Psi_n(\theta_0, \eta_0) + \big(\Psi_n(\theta_0, \hat{\eta}) - \Psi_n(\theta_0, \eta_0)\big) \\ &= S_n + B_n, \end{aligned} \tag{59}$$

where S_n and B_n were defined in (50)–(51). Substitute (59) into (58):

$$\hat{\theta} - \theta_0 = -\hat{J}_\theta^{-1}(S_n + B_n) + R_n. \tag{60}$$

Using the definition (49),

$$-\hat{J}_\theta^{-1} = \kappa_{\mathrm{DML}},$$

so (60) becomes

$$\hat{\theta} - \theta_0 = \kappa_{\mathrm{DML}}(S_n + B_n) + R_n. \tag{61}$$

We now describe orders of magnitude. From (21),

$$S_n = \frac{1}{n} \sum_{i=1}^{n} (D_i - m_0(X_i))\varepsilon_i,$$

so by the central limit theorem and Assumption 1(iv),

$$S_n = O_P(n^{-1/2}).$$

By Neyman orthogonality and Assumption 1(iii), the nuisance perturbation satisfies

$$B_n = o_P(n^{-1/2}).$$

Under standard DML results, $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$, so from (57),

$$R_n = O_P\big((\hat{\theta} - \theta_0)^2\big) = O_P(n^{-1}) = o_P(n^{-1/2}).$$

Lemma 2 (Linearization). *Under Assumption 1,*

$$\hat{\theta} - \theta_0 = \kappa_{\mathrm{DML}}(S_n + B_n) + R_n,$$

with $S_n = O_P(n^{-1/2})$, $B_n = o_P(n^{-1/2})$, and $R_n = o_P(n^{-1/2})$.

5   Coverage error bound for the standard DML confidence interval

Define the estimated standard error

$$\widehat{\mathrm{SE}}_{\mathrm{DML}} := \frac{\kappa_{\mathrm{DML}}}{\sqrt{n}} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \hat{U}_i^2 \hat{\varepsilon}_i^2}, \tag{62}$$

where

$$\hat{\varepsilon}_i := Y_i - \hat{g}(X_i) - \hat{\theta}(D_i - \hat{m}(X_i)). \tag{63}$$

The standard $(1 - \alpha)$ confidence interval is

$$\mathrm{CI}_{\mathrm{std}} := [\hat{\theta} \pm z_{1-\alpha/2} \, \widehat{\mathrm{SE}}_{\mathrm{DML}}], \tag{64}$$

where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ quantile of $N(0,1)$. Define the t-statistic

$$T_n := \frac{\hat{\theta} - \theta_0}{\widehat{\mathrm{SE}}_{\mathrm{DML}}}. \tag{65}$$

Then

$$\theta_0 \in \mathrm{CI}_{\mathrm{std}} \quad \Longleftrightarrow \quad |T_n| \leq z_{1-\alpha/2},$$

so

$$\mathbb{P}(\theta_0 \in \mathrm{CI}_{\mathrm{std}}) = \mathbb{P}\big(|T_n| \leq z_{1-\alpha/2}\big). \tag{66}$$

Concentration assumptions

Assumption 2 (Concentration). For some $\delta \in (0,1)$, with probability at least $1 - \delta$,

(i) **(Sampling fluctuation)** There exists $a_n(\delta)$ with $a_n(\delta) = O(\sigma_\psi/\sqrt{n})$ such that

$$|S_n| \leq a_n(\delta).$$

(ii) **(Nuisance and remainder)** There exists $r_n(\delta)$ with $r_n(\delta) = O(n^{-1/2-\gamma})$ for some $\gamma > 0$ such that

$$|B_n| + |R_n| \leq r_n(\delta).$$

(iii) **(SE consistency)** There exist $s_n > 0$ and $c_\xi < 1/2$ such that

$$\left| \widehat{\mathrm{SE}}_{\mathrm{DML}} - s_n \right| \leq c_\xi s_n.$$

On the event of Assumption 2(iii),

$$(1 - c_\xi)s_n \leq \widehat{\mathrm{SE}}_{\mathrm{DML}} \leq (1 + c_\xi)s_n,$$

so $\widehat{\mathrm{SE}}_{\mathrm{DML}}$ is bounded away from zero and of order s_n. As we will formalize below, a natural choice is

$$s_n := \frac{\kappa_{\mathrm{DML}}\sigma_\psi}{\sqrt{n}},$$

since S_n has variance σ_ψ^2/n and $\hat{\theta} - \theta_0 \approx \kappa_{\mathrm{DML}}S_n$.

Coverage error bound

We compare the distribution of T_n with that of a standard normal Z. The key point is that the *parameter-scale* error $\hat{\theta} - \theta_0$ is amplified by the condition number κ_{DML}, while the t-statistic is normalized to remove this scaling. The coverage bound is therefore "κ-free" in t-scale but implicitly *driven* by κ_{DML} in θ-scale.

Step 1 (Ideal statistic and Berry–Esseen). Recall the decomposition

$$T_n = T_{n,0} + \Delta_n, \qquad T_{n,0} := \frac{\kappa_{\mathrm{DML}} S_n}{\widehat{\mathrm{SE}}_{\mathrm{DML}}}, \qquad \Delta_n := \frac{\kappa_{\mathrm{DML}} B_n + R_n}{\widehat{\mathrm{SE}}_{\mathrm{DML}}}.$$

At (θ_0, η_0) we have $\Psi(\theta_0, \eta_0) = 0$ by (22), so $\mathbb{E}[\psi(W_i; \theta_0, \eta_0)] = 0$. Together with Assumption 1(iv), the score $\psi_i := \psi(W_i; \theta_0, \eta_0)$ satisfies

$$\mathbb{E}[\psi_i] = 0, \qquad \mathrm{Var}(\psi_i) = \sigma_\psi^2 \in (0, \infty), \qquad \mathbb{E}\big[|\psi_i|^3\big] \le M_3 < \infty.$$

Define the nonrandom target scale

$$s_n := \frac{\kappa_{\mathrm{DML}} \sigma_\psi}{\sqrt{n}}.$$

Then

$$\tilde{T}_{n,0} := \frac{\kappa_{\mathrm{DML}} S_n}{s_n} = \frac{\sqrt{n}}{\sigma_\psi} S_n = \frac{1}{\sqrt{n}\, \sigma_\psi} \sum_{i=1}^n \psi_i$$

is the usual standardized average of the score. In particular, s_n is precisely the *asymptotic standard deviation of $\hat{\theta}$*, so the typical θ-scale fluctuation is of order

$$\hat{\theta} - \theta_0 \approx \kappa_{\mathrm{DML}} S_n = O_P\left(\frac{\kappa_{\mathrm{DML}}}{\sqrt{n}}\right).$$

Thus κ_{DML} directly inflates the parameter-scale error and the confidence interval length, even though we will normalize it away in t-statistic scale.

The classical Berry–Esseen theorem applied to the sum $\sum_{i=1}^n \psi_i$ yields

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}(\tilde{T}_{n,0} \le z) - \Phi(z) \right| \le \frac{C_1}{\sqrt{n}}, \tag{67}$$

for some constant $C_1 > 0$ depending only on M_3 and σ_ψ. Since the concentration event $\mathcal{G}$ in Assumption 2 has probability at least $1 - \delta$, the same bound holds conditional on $\mathcal{G}$ up to a change in constants, and we continue to denote the constant by C_1:

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}(\tilde{T}_{n,0} \le z \mid \mathcal{G}) - \Phi(z) \right| \le \frac{C_1}{\sqrt{n}}.$$

Why introduce s_n? Berry–Esseen controls the distance to normality for sums normalized by a *deterministic* scale (here s_n). We cannot apply it directly to $T_{n,0}$ because $\widehat{\mathrm{SE}}_{\mathrm{DML}}$ is random. Instead, we first work with the "ideal" statistic $\tilde{T}_{n,0}$, which uses the nonrandom scale s_n, and then control the error from replacing s_n by $\widehat{\mathrm{SE}}_{\mathrm{DML}}$ in Step 2.

Step 2 (Replacing $\widehat{\mathrm{SE}}_{\mathrm{DML}}$ by s_n). On the event $\mathcal{G}$ where Assumption 2(i) and (iii) hold, we have

$$T_{n,0} - \tilde{T}_{n,0} = \kappa_{\mathrm{DML}} S_n \Big(\frac{1}{\widehat{\mathrm{SE}}_{\mathrm{DML}}} - \frac{1}{s_n} \Big) = \kappa_{\mathrm{DML}} S_n \frac{s_n - \widehat{\mathrm{SE}}_{\mathrm{DML}}}{\widehat{\mathrm{SE}}_{\mathrm{DML}} \, s_n}.$$

Using the concentration bounds on S_n and on the standard error, we obtain

$$
\begin{aligned}
|T_{n,0} - \tilde{T}_{n,0}| &\leq \kappa_{\mathrm{DML}} \, |S_n| \frac{|\widehat{\mathrm{SE}}_{\mathrm{DML}} - s_n|}{\widehat{\mathrm{SE}}_{\mathrm{DML}} \, s_n} \\
&\leq \kappa_{\mathrm{DML}} \, a_n(\delta) \frac{c_\xi s_n}{(1 - c_\xi) s_n^2} \quad \text{(by Assumption 2(i),(iii))} \\
&= \frac{c_\xi}{1 - c_\xi} \frac{\kappa_{\mathrm{DML}} a_n(\delta)}{s_n}.
\end{aligned}
$$

By construction of s_n and Assumption 2(i), $a_n(\delta) = O(\sigma_\psi / \sqrt{n})$ and $s_n = \kappa_{\mathrm{DML}} \sigma_\psi / \sqrt{n}$, so

$$\frac{\kappa_{\mathrm{DML}} a_n(\delta)}{s_n} = O(1),$$

and therefore, on $\mathcal{G}$,

$$|T_{n,0} - \tilde{T}_{n,0}| \leq C_2' c_\xi, \tag{68}$$

for some $C_2' > 0$ independent of n.

In particular, if the standard error estimator is $\sqrt{n}$-consistent in the sense that $c_\xi = c_\xi(n) = O(n^{-1/2})$, then the right-hand side in (68) is $O(n^{-1/2})$ and can be absorbed into the Berry–Esseen term by adjusting the constant C_1.

Step 3 (Bounding the bias and remainder term). On $\mathcal{G}$, Assumption 2(ii) gives

$$|B_n| + |R_n| \leq r_n(\delta), \qquad r_n(\delta) = O(n^{-1/2 - \gamma}), \ \gamma > 0.$$

Using the lower bound on $\widehat{\mathrm{SE}}_{\mathrm{DML}}$ and the definition of s_n,

$$
\begin{aligned}
|\Delta_n| = \left| \frac{\kappa_{\mathrm{DML}} B_n + R_n}{\widehat{\mathrm{SE}}_{\mathrm{DML}}} \right| &\leq \frac{\kappa_{\mathrm{DML}} |B_n| + |R_n|}{(1 - c_\xi) s_n} \\
&\leq \frac{(\kappa_{\mathrm{DML}} + 1) \, r_n(\delta)}{(1 - c_\xi) s_n} = \frac{(\kappa_{\mathrm{DML}} + 1) \, r_n(\delta) \, \sqrt{n}}{(1 - c_\xi) \, \kappa_{\mathrm{DML}} \sigma_\psi}.
\end{aligned}
$$

13

In the interesting regime where $\kappa_{\mathrm{DML}} \geq 1$, we have $(\kappa_{\mathrm{DML}} + 1)/\kappa_{\mathrm{DML}} \leq 2$, and therefore

$$|\Delta_n| \leq C_2 \sqrt{n} \, r_n(\delta), \tag{69}$$

for some $C_2 > 0$ depending only on $(1 - c_\xi)$ and σ_ψ.

Crucially, in θ-scale the corresponding contribution is

$$\kappa_{\mathrm{DML}}(B_n + R_n) = O_P(\kappa_{\mathrm{DML}} \, r_n(\delta)),$$

so even if $r_n(\delta)$ is of order $n^{-1/2-\gamma}$, a large κ_{DML} can make $\kappa_{\mathrm{DML}} r_n(\delta)$ non-negligible relative to the nominal $\kappa_{\mathrm{DML}}/\sqrt{n}$ variance. This is where ill-conditioning enters the finite-sample behavior: the same score-scale error $r_n(\delta)$ is *magnified* in the parameter scale by κ_{DML}.

Step 4 (From shifts in T_n to coverage error). We have

$$\theta_0 \in \mathrm{CI}_{\mathrm{std}} \quad \Longleftrightarrow \quad |T_n| \leq z_{1-\alpha/2},$$

so

$$\mathbb{P}(\theta_0 \in \mathrm{CI}_{\mathrm{std}}) = \mathbb{P}(|T_n| \leq z_{1-\alpha/2}).$$

On the event $\mathcal{G}$,

$$T_n = T_{n,0} + \Delta_n, \qquad T_{n,0} = \tilde{T}_{n,0} + (T_{n,0} - \tilde{T}_{n,0}).$$

Anti-concentration inequalities for the normal distribution imply that shifting a statistic by an amount δ changes $\mathbb{P}(|\cdot| \leq z_{1-\alpha/2})$ by at most a constant multiple of $|\delta|$. Combining this with (67), (68), and (69), and adding the probability of the complement event $\mathcal{G}^c$ (which is at most δ), yields a bound of the form

$$\left| \mathbb{P}(\theta_0 \in \mathrm{CI}_{\mathrm{std}}) - (1 - \alpha) \right| \leq \frac{C_1}{\sqrt{n}} + C_2 \sqrt{n} \, r_n(\delta) + C_3 \delta + C_4 c_\xi, \tag{70}$$

for suitable constants $C_1, C_2, C_3, C_4 > 0$.

If, in addition, the standard error estimator is $\sqrt{n}$-consistent so that $c_\xi = O(n^{-1/2})$, the last term in (70) is also $O(n^{-1/2})$ and can be absorbed into the first term by adjusting C_1. In that case, the coverage error bound simplifies to

$$\left| \mathbb{P}(\theta_0 \in \mathrm{CI}_{\mathrm{std}}) - (1 - \alpha) \right| \leq \frac{\tilde{C}_1}{\sqrt{n}} + C_2 \sqrt{n} \, r_n(\delta) + C_3 \delta,$$

for a suitably modified constant $\tilde{C}_1 > 0$.

Remark 3 (How κ_{DML} enters the picture). The bound (70) is expressed in t-statistic scale, where the leading Berry–Esseen term is of order $1/\sqrt{n}$ and does not display κ_{DML} explicitly because we have normalized by the correct scale $s_n \propto \kappa_{\mathrm{DML}}/\sqrt{n}$. However, combining this bound with the

linearization

$$\hat{\theta} - \theta_0 = \kappa_{\mathrm{DML}}(S_n + B_n) + R_n,$$

shows that the *parameter-scale* error and the confidence interval length are both of order $\kappa_{\mathrm{DML}}/\sqrt{n}$, while the bias term is of order $\kappa_{\mathrm{DML}}r_n(\delta)$. Thus, even if the t-statistic is well approximated by a normal distribution, a large κ_{DML} magnifies nuisance error and leads to regimes where $\kappa_{\mathrm{DML}}r_n(\delta)$ is non-negligible compared to $\kappa_{\mathrm{DML}}/\sqrt{n}$, causing systematic coverage failures in θ-space. This is the sense in which finite-sample reliability of DML is governed by the condition number.