Homework 3

DSO603, due by the end of Oct 16, 2025

Question 1

Let θ_0 and η_0 be the true target parameter and nuisance parameter/function, respectively. Suppose (θ_0, η_0) solves the estimation equation

$$\mathbb{E}[\psi(D; \theta_0, \eta_0)] = 0.$$

1. State the Neyman orthogonality condition for a score function $\psi(D; \theta, \eta)$ at (θ_0, η_0) for all admissible directions $(\eta - \eta_0)$. Explain in words what this condition means.

2. For the partially linear regression model discussed in class

$$\psi(D; \theta, \eta) = (W - e(X))\big(Y - g(X) - W\theta\big),$$

   verify the orthogonality condition at (θ_0, η_0) by computing the derivative with respect to perturbations $g \mapsto g_0 + th_g$ and $e \mapsto e_0 + th_e$ and showing the result equals zero.

3. Starting from

$$\sqrt{n}(\hat{\theta} - \theta_0) = -J_0^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(D_i; \theta_0, \eta_0) + J_0^{-1} \sqrt{n}\, \partial_\eta \mathbb{E}[\psi]\big|_{\eta_0} (\hat{\eta} - \eta_0) + o_p(1),$$

   explain why orthogonality removes the first-order nuisance effect and why it suffices that $\|\hat{\eta} - \eta_0\| = o_p(n^{-1/4})$ to achieve valid $\sqrt{n}$ inference.

Question 2

Consider the following IV model:

$$Y = W\theta_0 + g_0(X) + \zeta, \qquad W = Z\pi_0 + m_0(X) + v,$$

where $\eta_0 = (g_0, e_0, r_0)$ with $g_0(X) = \mathbb{E}[Y|X]$, $e_0(X) = \mathbb{E}[W|X]$, and $r_0(X) = \mathbb{E}[Z|X]$. Let $D = (Y, W, X, Z)$. The orthogonal score is

$$\psi(D; \theta, \eta) = \big(Y - g(X) - \theta\{W - e(X)\}\big)\big(Z - r(X)\big),$$

1. Verify $\mathbb{E}[\psi(D; \theta_0, \eta_0)] = 0$.

2. Compute $\frac{\partial}{\partial t}\mathbb{E}[\psi(D; \theta_0, \eta_0 + t(h_g, h_e, h_r))]\big|_{t=0}$ and show it equals 0 using the law of iterated expectations.

Question 3

Consider the partially linear regression (PLR) model

$$Y = W\theta_0 + g_0(X) + U, \qquad W = e_0(X) + V,$$

with $X \in \mathbb{R}^p$, $\mathbb{E}[U \mid X, W] = 0$, and $\mathbb{E}[V \mid X] = 0$. Set $\theta_0 = 2$. Generate (U, V) jointly normal with $\mathrm{Cov}(U, V) = \rho \in \{0, 0.3\}$ to control endogeneity.

Model setting

Setting 1 (Low-dimensional, nonlinear nuisances). Let $n = 500$, $p = 10$, $X \sim \mathcal{N}(0, I_p)$.

$$g_0(x) = \sin(x_1) + x_2^2, \qquad e_0(x) = 0.5x_3 - 0.3x_4.$$

Setting 2 (High-dimensional, linear sparse nuisances). Let $n \in \{500, 1000\}$, $p \in \{800, 1200\}$ with $p > n$ (ignore the combination where $p < n$), $X \sim \mathcal{N}(0, \Sigma_X)$ where $(\Sigma_X)_{jk} = 0.5^{|j-k|}$ (mild collinearity). Define $s \in \{10, 30\}$ active coordinates $\mathcal{S} \subset \{1, \dots, p\}$ with $|\mathcal{S}| = s$.

$$g_0(x) = x^\top \beta_0, \quad e_0(x) = x^\top \gamma_0,$$

where β_0, γ_0 are s-sparse (nonzeros on $\mathcal{S}$ with i.i.d. signs and magnitudes over Uniform[0.3,0.6]). This setting is designed so that linear controls are correct but *high dimensional*. Classical OLS with all X is infeasible; use Lasso-type methods.

Estimators

We will consider the following estimators:

i. *Naïve plug-in OLS:* Fit $\hat{g}(X)$ via ML (RF in model setting 1; Lasso in model setting 2), then regress $Y - \hat{g}(X)$ on W.

ii. *DML (cross-fitting, partialling out):* Fit $\hat{g}(X) \approx \mathbb{E}[Y \mid X]$ and $\hat{e}(X) \approx \mathbb{E}[W \mid X]$ with K-fold cross-fitting (use RF/GBM in Model setting 1; Lasso/Post-Lasso in Model setting 2). Form residuals $\tilde{Y} = Y - \hat{g}(X)$, $\tilde{W} = W - \hat{e}(X)$, and regress $\tilde{Y}$ on $\tilde{W}$.

Conduct the following simulations:

1. For each model setting, run $R = 100$ replications for each (ρ, n, p, s) setting (as applicable). Keep the same random seed grid across methods for fair comparison.

2. For model setting 2, use cross-validated Lasso (e.g., 10-fold) within each training fold of cross-fitting. Report selected sparsity $\hat{s}_g, \hat{s}_e$ per replication.

3. For each method and parameter design setting, report:

    - Bias, variance, RMSE of $\hat{\theta}$ relative to θ_0;

- Empirical coverage of nominal 95% CIs and average CI length;
- For model setting 2: average selected model size $(\hat{s}_g, \hat{s}_e)$ and proportion of times $\mathcal{S} \subseteq \widehat{\mathcal{S}}$ (screening property).

If you need to estimate the varaince, use the plug-in estimator:

$$\widehat{\mathrm{Var}}(\hat{\theta}) = \frac{\hat{\Omega}}{\hat{J}^2}, \qquad \hat{J} = \frac{1}{n}\sum_{i=1}^{n} \tilde{W}_i^2, \qquad \hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} \tilde{W}_i^2(\tilde{Y}_i - \hat{\theta}\tilde{W}_i)^2.$$

4. (a) For *Model setting 1*, compare naïve plug-in vs. DML.
   (b) For *Model setting 2*, compare DML with Lasso/Post-Lasso against naïve plug-in (Lasso-only).