Ec2147, Spring 2021

Professor Bryan Graham

Problem Set 1: part (a)

Due: March 16th, 2021

Problem sets are due by 5PM EST. You may work in groups (indeed I encourage you to all work together), but each student should turn in their own write-up.

Helpful reading: Newey (1990), Brown and Newey (1998) and Graham (2011). For an alternative, and highly elegant approach, see Chamberlain (1987).

Let $Z = (X', Y)'$ be a vector of modeling variables and $\{Z_i\}_{i=1}^{\infty}$ an independent and identically distribution random sequence drawn from unknown distribution F_0. The sole prior restriction on F_0 is that, for some known function $\psi(Z, \beta)$ and unknown parameter $\beta_0 \in \mathbb{B} \subset \mathbb{R}^K$

$$\mathbb{E}[\psi(Z, \beta_0)] = 0.$$

Assume that $\dim(\psi(Z, \beta_0)) = J > \dim(\beta_0) = K$.

Let $a(Z, \beta)$ be some known function of Z and β. The goal is to efficiently estimate the mean

$$\mu_0 = \mathbb{E}[a(Z, \beta_0)].$$

1. Let $X = \left(\tilde{X}', W\right)$' with W a binary treatment indicator (also partition $\beta = \left(\tilde{\beta}, \alpha'\right)'$) and assume that

   $$\Pr(Y = 1|\, X) = \Phi\left(X'\tilde{\beta} + W\alpha\right).$$

   Let $\delta_l = \mathbb{E}[Y|\, X \in \mathbb{X}_l]$ for $l = 1, \ldots, L$. Assume that these conditional means are known to the econometrician (e.g., from register data). The target estimand is

   $$\mu_0 = \mathbb{E}\left[\Phi\left(X'\tilde{\beta} + \alpha\right) - \Phi\left(X'\tilde{\beta}\right)\right].$$

   Outline a plausible empirical setting to which you could adapt the above components. Show how your setting is accommodated by the general setup outlined above.

2. Calculate the semiparametric variance bound, say $\mathcal{I}(\beta_0)^{-1}$, for β_0. You may use either the approach outlined in Newey (1990) (and also lecture) or that of Chamberlain (1987).

3. Show that the semiparametric variance bound, say $\mathcal{I}(\mu_0)^{-1}$, for μ_0 is

   $$\mathcal{I}(\mu_0)^{-1} = \Sigma_{aa} - \Sigma_{a\psi}\Omega_0^{-1}\Sigma_{a\psi}' + \left[\Sigma_{a\psi}\Omega_0^{-1}\Gamma_0 - \Xi_0\right]\mathcal{I}(\beta_0)^{-1}\left[\Gamma_0'\Omega_0^{-1}\Sigma_{a\psi}' - \Xi_0'\right]$$

   where $\Sigma_{aa} = \mathbb{E}\left[(a(Z, \beta_0) - \mu_0)(a(Z, \beta_0) - \mu_0)'\right]$, $\Sigma_{a\psi} = \mathbb{E}\left[(a(Z, \beta_0) - \mu_0)\psi(Z, \beta_0)'\right]$, $\Omega_0 = \mathbb{E}\left[\psi(Z, \beta_0)\psi(Z, \beta_0)'\right]$, $\Gamma_0 = \mathbb{E}\left[\frac{\partial \psi(Z, \beta_0)}{\partial \beta'}\right]$ and $\Xi_0 = \mathbb{E}\left[\frac{\partial a(Z, \beta_0)}{\partial \beta'}\right]$. As before, you may use either the approach of Chamberlain (1987) or Newey (1990) (in this case I also recommend Brown and Newey (1998)). Interpret each of the three components of the inverse information expression above.

4. Assume that $\hat{\beta}$ is an efficient estimate of β_0.

   (a) Is $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} a\left(Z_i, \hat{\beta}\right)$ efficient? Explain.

(b) Let $\hat{b}_1$ and $\hat{b}_\psi$ be the intercept and slope coefficients associated with the least squares fit of $a\left(Z_i, \hat{\beta}\right)$ onto a constant and $\psi\left(Z_i, \hat{\beta}\right)$. Let $\hat{\psi}_N = \frac{1}{N}\sum_{i=1}^{N}\psi\left(Z_i, \hat{\beta}\right)$.

i. Is $\hat{\mu} = \hat{b}_1 + \hat{\psi}_N' \hat{b}_\psi$ consistent for μ_0? Is it efficient? Explain.

ii. Is $\hat{\mu} = \hat{b}_1$ consistent for μ_0? Is it efficient? Explain.

(c) Assume that β_0 is known (such that the analyst computes the least squares fit of $a\left(Z_i, \beta_0\right)$ onto a constant and $\psi\left(Z_i, \beta_0\right)$). Show that, in this case,

$$\hat{b}_1 = \sum_{i=1}^{N}\left[a\left(Z_i, \beta_0\right) - \Sigma_{a\psi}\Omega_0^{-1}\psi\left(Z_i, \beta_0\right)\right]/N + o_p\left(N^{-1/2}\right)$$

and hence that $\hat{b}_1$ is asymptotically uncorrelated with $\psi_N = \frac{1}{N}\sum_{i=1}^{N}\psi\left(Z_i, \beta_0\right)$. Connect this observation to your analysis in part (b) above.

References

Brown, B. W. and Newey, W. K. (1998). Efficient semiparametric estimation of expectations. *Econometrica*, 66(2):453 – 464.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305 – 334.

Graham, B. S. (2011). Efficiency bounds for missing data models with semiparametric restrictions. *Econometrica*, 79(2):437 – 452.

Newey, W. K. (1990). Efficient instrumental variables estimation of nonlinear models. *Econometrica 58*, 58(4):809 – 837.