Calculation of the Asymptotic Distribution of Semiparametric M-Estimators

Hidehiko Ichimura and Sokbae Lee

Department of Economics, UCL

CeMMaP, UCL and IFS

Plan

1. Model and Motivation

2. An overview

3. An illustration of our approach

4. Conclusion

1 Model and Motivation

We characterize asymptotic distribution of 2 step M-estimators of θ_0 where θ_0 minimizes $E[m(Z, \theta, f_0(\cdot, \theta))]$. The first step is an estimator of an unknown function f_0.

Assume that for each θ, a nonparametric estimator $\hat{f}_n(\cdot, \theta)$ of $f_0(\cdot, \theta)$ is available.

The observed data $\{Z_i : i = 1, \ldots, n\}$ are a random sample of Z.

We study an M-estimator of θ_0 that minimizes

$$\hat{S}_n(\theta) \equiv n^{-1} \sum_{i=1}^{n} m(Z_i, \theta, \hat{f}_n(\cdot, \theta)).$$

Examples: Partially Linear, Single Index, Klein-Spady, Semiparametric Likelihood, Average Derivative, Average Density.

Four Motivations:

- Works by Andrews (1994), Newey (1994), Chen and Shen (1998), Ai and Chen (2003), Chen, Linton, and Van Keilegom (2003) clarified the structure of the asymptotic analysis but given a problem, we still cannot carry out the asymptotic analysis in the same way we can for the regular 2 step case.

- The asymptotic variance has been characterized by Newey (1994) when nonparametric component does not have any restriction but from empirical standpoint we want to allow for some restrictions on the nonparametric component perhaps for the dimension reduction purpose.

- Even for the case Newey's result applies, the contribution of the first stage estimator to the overall asymptotic distribution is captured by the term expressed as the solution to an integral equation and not immediately clear what it is.

- Common structure needs to be clarified further to obtain systematic results on smoothing parameter choice problem.

- We provide a simple formula for semiparametric M-estimators under regularity conditions that are relatively straightforward to verify and also weaker than those available in the literature.

- We illustrate the use of the formula by applying it to
  - profiled estimation of a single index quantile regression model and
  - semiparametric least squares estimator under model misspecification.

2 Overview

- Our approach is analogous to the standard analysis of the 2-step estimators when the objective function is not smooth.

- The basic result appeals to the Taylor's series expansion of the expectation of the objective function.

- Since the objective function is a functional we need to use the Taylor's series expansion based on the Fréchet differentiability.

- Since the first stage estimator is an estimator of a function, we need to suitably modify the concept of asymptotic linearity.

- We identify two cases to be distinguished:

  - When f does not depend on θ.

  - When $D_f E\left[m\left(Z, \theta_0, f_0\left(\cdot, \theta_0\right)\right)\right](h) = 0$ for any h.

- When f does not depend on θ, we can allow less smoothness in function m with respect to f.

- When $D_f E\left[m\left(Z, \theta_0, f_0\left(\cdot, \theta\right)\right)\right](h) = 0$ for any h, we can allow less smoothness in the estimator of f_0.

3 An illustration of our approach

When f does not depend on θ:

- $\Delta_{1i}(\theta_0, f)$ denote the L_2 partial derivative of $m_i(\theta, f)$ with respect to θ evaluated at $\theta = \theta_0$

- Define

  $$R_i(\theta, f) = m_i(\theta, f) - m_i(\theta_0, f) - \Delta_{1i}(\theta_0, f)(\theta - \theta_0).$$

  and

  $$\begin{aligned} & S_n(\theta, f) \\ = \; & n^{-1} \sum_{i=1}^{n} [m_i(\theta, f) - m_i(\theta_0, f)] \\ = \; & n^{-1} \sum_{i=1}^{n} \{R_i(\theta, f) - E[R_i(\theta, f)]\} \\ & + n^{-1} \sum_{i=1}^{n} \{\Delta_{1i}(\theta_0, f) - E[\Delta_{1i}(\theta_0, f)]\}(\theta - \theta_0) \\ & + E\{m_i(\theta, f) - m_i(\theta_0, f)\}. \end{aligned}$$

- Also

$$
\begin{aligned}
& n^{-1} \sum_{i=1}^{n} \{\Delta_{1i}(\theta_0, f) - E[\Delta_{1i}(\theta_0, f)]\} \\
= \; & n^{-1} \sum_{i=1}^{n} \{\Delta_{1i}(\theta_0, f_0) - E[\Delta_{1i}(\theta_0, f_0)]\} \\
& + \Bigg[n^{-1} \sum_{i=1}^{n} \{\Delta_{1i}(\theta_0, f) - E[\Delta_{1i}(\theta_0, f)]\} \\
& - n^{-1} \sum_{i=1}^{n} \{\Delta_{1i}(\theta_0, f_0) - E[\Delta_{1i}(\theta_0, f_0)]\} \Bigg]
\end{aligned}
$$

$$
\begin{aligned}
& E\left\{m_i(\theta, f) - m_i(\theta_0, f)\right\} \\
= \; & D_\theta E\left\{m_i(\theta_0, f)\right\}(\theta - \theta_0) \\
& + \frac{1}{2}(\theta - \theta_0)^T D_{\theta\theta} E\left\{m_i(\theta_0, f_0)\right\}(\theta - \theta_0) \\
& + o\left(\|\theta - \theta_0\|^2\right) \\
= \; & D_\theta E\left\{m_i(\theta_0, f_0)\right\}(\theta - \theta_0) \\
& + D_{\theta f} E\left\{m_i(\theta_0, f_0)\right\}(f - f_0)(\theta - \theta_0) \\
& + \Big[D_{\theta f} E\Big\{m_i(\theta_0, \bar{f})\Big\} \\
& - D_{\theta f} E\left\{m_i(\theta_0, f_0)\right\}(f - f_0)(\theta - \theta_0)\Big] \\
& + \frac{1}{2}(\theta - \theta_0)^T D_{\theta\theta} E\left\{m_i(\theta_0, f_0)\right\}(\theta - \theta_0) \\
& + o\left(\|\theta - \theta_0\|^2\right).
\end{aligned}
$$

Assume that $\hat{f}$ has the asymptotic linear form:

$$\left\|\hat{f}(\cdot) - f_0(\cdot) - n^{-1}\sum_{i=1}^{n}\psi_{ni}(\cdot)\right\| = o_p\left(n^{-1/2}\right).$$

Then

$$\begin{aligned}
& D_{\theta f}E\left\{m_i(\theta_0, f_0)\right\}\left(\hat{f} - f_0\right)(\theta - \theta_0) \\
= \; & D_{\theta f}E\left\{m_i(\theta_0, f_0)\right\}\left(n^{-1}\sum_{i=1}^{n}\psi_{ni}(\cdot)\right)(\theta - \theta_0) \\
& + o_p\left(n^{-1/2}\|\theta - \theta_0\|\right) \\
= \; & n^{-1}\sum_{i=1}^{n} D_{\theta f}E\left\{m_i(\theta_0, f_0)\right\}\left(\psi_{ni}(\cdot)\right)(\theta - \theta_0) \\
& + o_p\left(n^{-1/2}\|\theta - \theta_0\|\right)
\end{aligned}$$

Putting these calculations together, we have

$$
\begin{aligned}
& S_n(\theta, f) \\
= \; & n^{-1} \sum_{i=1}^{n} [\{\Delta_{1i}(\theta_0, f) - E[\Delta_{1i}(\theta_0, f)]\} \\
& + D_{\theta f} E\{m_i(\theta_0, f_0)\}(\psi_{ni}(\cdot))](\theta - \theta_0) \\
& + \frac{1}{2}(\theta - \theta_0)^T D_{\theta\theta} E\{m_i(\theta_0, f_0)\}(\theta - \theta_0) \\
& + o_p\left(n^{-1}\right).
\end{aligned}
$$

When f depends on θ:

- The centering used earlier is no longer appropriate.

- $\Delta_{2i}\left(\theta_0, f_0\left(\cdot, \theta_0\right)\right)$ denote the L_2 partial derivative of $m_i(\theta, f)$ with respect to f evaluated at $\theta = \theta_0$ and $f = f_0\left(\cdot, \theta_0\right)$.

- We need additional smoothness of function m not needed when f did not depend on θ.

- Define

$$\begin{aligned} R_i(\theta, f) \;=\;& m_i(\theta, f\left(\cdot, \theta\right)) - m_i(\theta_0, f_0\left(\cdot, \theta_0\right)) \\ & -\Delta_{1i}\left(\theta, f_0\left(\cdot, \theta_0\right)\right)\left(\theta - \theta_0\right) \\ & -\Delta_{2i}\left(\theta_0, f_0\left(\cdot, \theta_0\right)\right)\left(f\left(\cdot, \theta\right) - f_0\left(\cdot, \theta_0\right)\right). \end{aligned}$$

$$
\begin{aligned}
& S_n(\theta, f) \\
= \; & n^{-1} \sum_{i=1}^{n} \left[m_i(\theta, f(\cdot, \theta)) - m_i(\theta_0, f_0(\cdot, \theta_0)) \right] \\
= \; & n^{-1} \sum_{i=1}^{n} \left\{ R_i(\theta, f) - E\left[R_i(\theta, f) \right] \right\} \\
& + n^{-1} \sum_{i=1}^{n} \{ \Delta_{1i}(\theta_0, f_0(\cdot, \theta_0)) \\
& - E\left[\Delta_{1i}(\theta_0, f_0(\cdot, \theta_0)) \right] \} (\theta - \theta_0) \\
& + n^{-1} \sum_{i=1}^{n} \{ \Delta_{2i}(\theta_0, f_0(\cdot, \theta_0)) \\
& - E\left[\Delta_{2i}(\theta_0, f_0(\cdot, \theta_0)) \right] \} (f(\cdot, \theta) - f_0(\cdot, \theta_0)) \\
& + E\left\{ m_i(\theta, f(\cdot, \theta)) - m_i(\theta_0, f_0(\cdot, \theta_0)) \right\}.
\end{aligned}
$$

- We assume

$$
\begin{aligned}
& \left\|f(\cdot,\theta)-f\left(\cdot,\theta_0\right)-\partial f_0\left(\cdot,\theta_0\right)/\partial\theta\left(\theta-\theta_0\right)\right\| \\
= \ & o\left(\left\|\theta-\theta_0\right\|\right).
\end{aligned}
$$

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\{\Delta_{2i}\left(\theta_0, f_0\left(\cdot,\theta_0\right)\right) \\
& -E\left[\Delta_{2i}\left(\theta_0, f_0\left(\cdot,\theta_0\right)\}\right)\right]\}\left(f(\cdot,\theta)-f_0\left(\cdot,\theta_0\right)\right) \\
= \ & n^{-1}\sum_{i=1}^{n}\{\Delta_{2i}\left(\theta_0, f_0\left(\cdot,\theta_0\right)\right) \\
& -E\left[\Delta_{2i}\left(\theta_0, f_0\left(\cdot,\theta_0\right)\}\right)\right]\}\partial f_0\left(\cdot,\theta_0\right)/\partial\theta\left(\theta-\theta_0\right) \\
& +n^{-1}\sum_{i=1}^{n}\{\Delta_{2i}\left(\theta_0, f_0\left(\cdot,\theta_0\right)\right) \\
& -E\left[\Delta_{2i}\left(\theta_0, f_0\left(\cdot,\theta_0\right)\}\right)\right]\}\left(f\left(\cdot,\theta_0\right)-f_0\left(\cdot,\theta_0\right)\right) \\
& +n^{-1}\sum_{i=1}^{n}\{\Delta_{2i}\left(\theta_0, f_0\left(\cdot,\theta_0\right)\right) \\
& -E\left[\Delta_{2i}\left(\theta_0, f_0\left(\cdot,\theta_0\right)\}\right)\right]\}o_p\left(\left\|\theta-\theta_0\right\|\right).
\end{aligned}
$$

$$
\begin{aligned}
& E\left\{m_i(\theta, f(\cdot,\theta)) - m_i(\theta_0, f_0(\cdot,\theta_0))\right\} \\
= \; & E\left\{m_i(\theta_0, f(\cdot,\theta)) - m_i(\theta_0, f_0(\cdot,\theta_0))\right\} \\
& + D_\theta E\left\{m_i(\theta_0, f(\cdot,\theta))\right\}(\theta - \theta_0) \\
& + \frac{1}{2}(\theta - \theta_0)^T D_{\theta\theta} E\left\{m_i(\theta_0, f(\cdot,\theta))\right\}(\theta - \theta_0) \\
& + o\left(\|\theta - \theta_0\|^2\right)
\end{aligned}
$$

$$
\begin{aligned}
& E\left\{m_i(\theta_0, f(\cdot,\theta)) - m_i(\theta_0, f_0(\cdot,\theta_0))\right\} \\
= \; & E\left\{m_i(\theta_0, f(\cdot,\theta_0)) - m_i(\theta_0, f_0(\cdot,\theta_0))\right\} \\
& + D_f E\left\{m_i(\theta_0, f(\cdot,\theta_0))\right\}\left(f(\cdot,\theta) - f(\cdot,\theta_0)\right) \\
& + \frac{1}{2} D_{ff} E\left\{m_i(\theta_0, \bar{f}(\cdot,\theta))\right\} \\
& \left(f(\cdot,\theta) - f(\cdot,\theta_0), f(\cdot,\theta) - f(\cdot,\theta_0)\right).
\end{aligned}
$$

Note that

$$
\begin{aligned}
& f(\cdot,\theta) - f(\cdot,\theta_0) \\
= \; & \partial f_0(\cdot,\theta_0)/\partial\theta\,(\theta - \theta_0) + o\left(\|\theta - \theta_0\|\right).
\end{aligned}
$$

Thus

$$
\begin{aligned}
& \frac{1}{2} D_{ff} E\left\{m_i(\theta_0, \bar{f}(\cdot,\theta))\right\} \\
& \left(f(\cdot,\theta) - f(\cdot,\theta_0), f(\cdot,\theta) - f(\cdot,\theta_0)\right) \\
= \; & \frac{1}{2} D_{ff} E\left\{m_i(\theta_0, f_0(\cdot,\theta_0))\right\} \\
& \left(\partial f_0(\cdot,\theta_0)/\partial\theta\,(\theta - \theta_0), \partial f_0(\cdot,\theta_0)/\partial\theta\,(\theta - \theta_0)\right) \\
& + o\left(\|\theta - \theta_0\|^2\right)
\end{aligned}
$$

Also

$$
\begin{aligned}
& D_f E\left\{m_i(\theta_0, f(\cdot, \theta_0))\right\}(f(\cdot, \theta) - f(\cdot, \theta_0)) \\
= \; & D_f E\left\{m_i(\theta_0, f_0(\cdot, \theta_0))\right\}(f(\cdot, \theta) - f(\cdot, \theta_0)) \\
& + D_{ff} E\left\{m_i(\theta_0, \bar{f}(\cdot, \theta_0))\right\} \\
& (f(\cdot, \theta) - f(\cdot, \theta_0), f(\cdot, \theta) - f_0(\cdot, \theta_0)) \\
= \; & D_f E\left\{m_i(\theta_0, f_0(\cdot, \theta_0))\right\} \\
& (\partial f(\cdot, \theta_0) - \partial f_0(\cdot, \theta_0))(\theta - \theta_0) \\
& + D_f E\left\{m_i(\theta_0, f_0(\cdot, \theta_0))\right\} \partial f_0(\cdot, \theta_0)(\theta - \theta_0) \\
& + D_f E\left\{m_i(\theta_0, f_0(\cdot, \theta_0))\right\} \partial^2 f_0(\cdot, \theta_0)(\theta - \theta_0, \theta - \theta_0) \\
& + o\left(\|\theta - \theta_0\|^2\right) \\
& + D_{ff} E\left\{m_i(\theta_0, \bar{f}(\cdot, \theta_0))\right\} \\
& (\partial f_0(\cdot, \theta_0) / \partial\theta (\theta - \theta_0) + o(\|\theta - \theta_0\|) \\
& , f(\cdot, \theta_0) - f_0(\cdot, \theta_0) + o(\|\theta - \theta_0\|))
\end{aligned}
$$

To recap,

$$
\begin{aligned}
& S_n(\theta, f) \\
= \; & \text{terms not depending on } \theta \\
& + \text{terms smaller than } n^{-1} \\
& + + n^{-1} \sum_{i=1}^{n} \{ \Delta_{1i} (\theta_0, f_0 (\cdot, \theta_0)) \\
& - E \left[\Delta_{1i} (\theta_0, f_0 (\cdot, \theta_0)) \right] \} (\theta - \theta_0) \\
& + n^{-1} \sum_{i=1}^{n} \{ \Delta_{2i} (\theta_0, f_0 (\cdot, \theta_0)) \\
& - E \left[\Delta_{2i} (\theta_0, f_0 (\cdot, \theta_0) \}) \right] \} \partial f_0 (\cdot, \theta_0) / \partial \theta (\theta - \theta_0) \\
& + D_\theta E \{ m_i(\theta_0, f(\cdot, \theta)) \} (\theta - \theta_0) \\
& + \frac{1}{2} (\theta - \theta_0)^T T D_{\theta\theta} E \{ m_i(\theta_0, f_0 (\cdot, \theta_0)) \} (\theta - \theta_0) \\
& + D_f E \{ m_i(\theta_0, f_0 (\cdot, \theta_0)) \} \\
& (\partial f (\cdot, \theta_0) - \partial f_0 (\cdot, \theta_0)) (\theta - \theta_0) \\
& + D_{ff} E \left\{ m_i(\theta_0, \bar{f} (\cdot, \theta_0)) \right\} \\
& (\partial f_0 (\cdot, \theta_0) / \partial \theta (\theta - \theta_0), f (\cdot, \theta_0) - f_0 (\cdot, \theta_0))
\end{aligned}
$$

- The last two terms are the additional contribution when f depends on θ. In most cases the first of the two terms are zero because $D_f E\left\{m_i(\theta_0, f_0(\cdot, \theta_0))\right\}$ is identically 0.

4 Conclusion

- Main findings so far are:

  1. Asymptotic Variance calculation is reduced to knowing the linear term for the nonparametric estimators used and the first two Frechet derivatives.

  2. When f does not depend on θ, smoothness of m can be relaxed considerably.

  3. In most cases $D_f E\left\{m_i(\theta_0, f_0(\cdot, \theta_0))\right\}$ is a zero or the estimator does not depend on θ so that

the derivative of the estimator with respect to θ does not contribute to the overall asymptotic variance.

4. Semiparametric estimators under misspecification is an example where $D_f E\left\{m_i(\theta_0, f_0(\cdot, \theta_0))\right\}$ is not zero.

5. This can be used to construct a specification test. This is in contrast to the standard tests where Pitman's drift is the source of the power of the tests. The test will gain power from a random 'drift'.