Problem Set 7: M-Estimation

Review the Concepts and Proofs

1. Explain in an intuitive way what identification of nonlinear models means. Try to find examples of NLS functions $m(\mathbf{x}, \boldsymbol{\theta})$ where identification fails.

2. What is a loss function in the estimation context? Which loss function do you minimize when you apply OLS? Why is an estimator based on the Huber loss function less sensitive to outliers than the OLS estimator?

3. Explain the scalar version of the mean value theorem.

4. What is the gradient of the general objective function $q(\mathbf{w}, \boldsymbol{\theta})$ with P parameters $\boldsymbol{\theta}$? What is the score? What are the dimensions of gradient and score?

5. What is the Hessian of the general objective function $q(\mathbf{w}, \boldsymbol{\theta})$ with P parameters $\boldsymbol{\theta}$? What is the dimension of the Hessian?

6. Use a mean value expansion of the scores and a central limit theorem applied to the scores to show that the M-estimator is asymptotically normally distributed. Does the Huber loss function satisfy all assumptions?

7. Show that under homoscedasticity, the generalized information matrix equality holds for NLS.

8. Use a mean value expansion to find the asymptotic distribution of a possibly nonlinear function $\mathbf{c}(\cdot)$ of the M-estimator $\hat{\boldsymbol{\theta}}$.

Exercises

1. Show that the asymptotic distribution of the M-estimator collapses to that of the OLS estimator if $q(\mathbf{w}, \boldsymbol{\theta}) = (y - \mathbf{x}\boldsymbol{\theta})^2$. Explain.

2. Consider the exponential regression with $m(\mathbf{x}, \boldsymbol{\theta}) = \exp(\theta_0 + \theta_1 x_1)$.

   (a) Find the score and Hessian.

   (b) Find $\mathbf{A}_o$, $\mathbf{B}_o$ and their estimators.

   (c) Find the marginal effect of x_1 and the elasticity of $\mathrm{E}(y|\mathbf{x})$ with respect to x_1.

   (d) Suppose estimation using a sample of 49 observations yields the estimate $\hat{\boldsymbol{\theta}} = (0.693147, 1)'$ with the estimated variance matrix

$$\hat{\mathbf{V}} = \begin{pmatrix} 0.15 & 0.07 \\ 0.07 & 0.2 \end{pmatrix}.$$

Construct a 95% confidence interval for the marginal effect conditional on the observation $x_1 = 0$ (this might be an interesting individual or the average of all x_{1i}).

(e) For the estimates in (d), find the asymptotic distribution of $\gamma = \theta_0^2 - \theta_1^2$ and construct a 95% confidence interval for γ.

(f) Test the two-sided null hypothesis $\gamma = 0$ using both a t test and a Wald test.

3. Consider the M-estimator $q(\mathbf{w}, \boldsymbol{\theta}) = 0.5 \log[1 + u^2/k]$, $u = y - \mathbf{x}\boldsymbol{\theta}$, $k > 0$.

   (a) Find the score and Hessian.

   (b) Use the score to show that compared to OLS this estimator downweights the influence of large residuals $\hat{u}_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\theta}}$.

   (c) Show that as k increases, the M-estimator becomes closer and closer to the OLS estimator. Hint: think of a quadratic approximation to $q(\mathbf{w}, \boldsymbol{\theta})$.

4. Consider the Huber estimator

$$q(\mathbf{w}, \boldsymbol{\theta}) = \begin{cases} 2ku - k^2 & \text{if } u > k \\ u^2 & \text{if } -k \leq u \leq k \\ -2ku - k^2 & \text{if } u < -k \end{cases},$$

   where $u = y - \mathbf{x}\boldsymbol{\theta}$ and $k > 0$.

   (a) Show that $q(\mathbf{w}, \boldsymbol{\theta})$ is once continuously differentiable.

   (b) Show that a piecewise defined Hessian is discontinuous at k and $-k$.

   (c) Argue why the Newton-Raphson algorithm may fail when the piecewise defined Hessian is used. Hint: simplify the model to $\mathbf{x} = 1$ (so that you are estimating the mean of y). Then for the sample size $N = 1$ show graphically what happens if the starting value is chosen far away from y_1. Finally, generalize to $N > 1$.