9 Semiparametric Methods

9.1 Parametric vs. partly linear test statistic is a degenerate U-statistic

Recall that to show $\widehat{I}_n$ is a degenerate U-statistic we have to demonstrate that the kernel of our statistic has mean 0 and is symmetric.

$$I_n = \frac{1}{n(n-1)|\mathbf{h}|} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} u_i u_j K_h(\mathbf{z}_i, \mathbf{z}_j) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j>i}^{n} H_n(\psi_i, \psi_j),$$

where $H_n(\psi_i, \psi_j) = |\mathbf{h}|^{-1} u_i u_j K_h(\mathbf{z}_i, \mathbf{z}_j)$. It is immediately clear that $H_n(\cdot, \cdot)$ is symmetric given that our kernel function is symmetric. Taking a conditional expectation we have

$$\begin{aligned} E\left[H_n(\psi_i, \psi_j)|\psi_i\right] =& E\left[|\mathbf{h}|^{-1} u_i u_j K_h(\mathbf{z}_i, \mathbf{z}_j)|(u_i, \mathbf{z}_i)\right] \\ =& |\mathbf{h}|^{-1} u_i E\left[u_j K_h(\mathbf{z}_i, \mathbf{z}_j)|\psi_i\right] \\ =& |\mathbf{h}|^{-1} u_i E\left[K_h(\mathbf{z}_i, \mathbf{z}_j) E\left[u_j|\mathbf{z}_j\right]\right] = 0, \end{aligned}$$

where the second to last equality holds via the conditional law of iterated expectations and the last equality is true under the null hypothesis.

9.2 Partly linear vs. nonparametric test statistic is degenerate U-statistic

Recall that to show $\widehat{I}_n$ is a degenerate U-statistic we have to demonstrate that the kernel of our statistic has mean 0 and is symmetric.

$$I_n = \frac{1}{n(n-1)|\mathbf{h}|} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} u_i u_j K_h(\mathbf{w}_i, \mathbf{w}_j) = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j>i}^{n} H_n(\psi_i, \psi_j),$$

where $H_n(\psi_i, \psi_j) = |\mathbf{h}|^{-1} u_i u_j K_h(\mathbf{w}_i, \mathbf{w}_j)$. It is immediately clear that $H_n(\cdot, \cdot)$ is symmetric given that our kernel function is symmetric. Taking a conditional expectation we have

$$\begin{aligned} E\left[H_n(\psi_i, \psi_j)|\psi_i\right] =& E\left[|\mathbf{h}|^{-1} u_i u_j K_h(\mathbf{w}_i, \mathbf{w}_j)|(u_i, \mathbf{w}_i)\right] \\ =& |\mathbf{w}|^{-1} u_i E\left[u_j K_h(\mathbf{w}_i, \mathbf{w}_j)|\psi_i\right] \\ =& |\mathbf{h}|^{-1} u_i E\left[K_h(\mathbf{w}_i, \mathbf{w}_j) E\left[u_j|\mathbf{w}_j\right]\right] = 0, \end{aligned}$$

where the second to last equality holds via the conditional law of iterated expectations and the last equality is true under the null hypothesis.

9.3 Single index cross-validation function

First, define

$$\widetilde{S}(\beta) = \sum_{i=1}^{n} \left(y_i - m(\mathbf{x}_i \beta)\right)^2$$

and

$$\widehat{S}(\beta) = \sum_{i=1}^{n} \left(y_i - \widehat{m}_{-i}(\mathbf{x}_i \beta)\right)^2 .$$

Further, let

$$\begin{aligned} D_i =& \widehat{m}_{-i}(\mathbf{x}_i \beta_0) - m(\mathbf{x}_i \beta_0) \\ \delta_i =& m(\mathbf{x}_i \beta) - m(\mathbf{x}_i \beta_0) \\ \triangle_i =& \left(\widehat{m}_{-i}(\mathbf{x}_i \beta) - m(\mathbf{x}_i \beta)\right) - \left(\widehat{m}_{-i}(\mathbf{x}_i \beta_0) - m(\mathbf{x}_i \beta_0)\right) . \end{aligned}$$

D_i is the difference between the known and estimated function, evaluated at the true parameter vector. δ_i is the known function evaluated at the true parameter vector and some other value.

Next, recalling that the model is $y_i = m(\mathbf{x}_i\beta_0) + \varepsilon_i$, we write $\widetilde{S}(\beta)$ in terms of δ_i as

$$\begin{aligned}
\widetilde{S}(\beta) &= \sum_{i-1}^{n} [y_i - m(\mathbf{x}_i\beta)]^2 \\
&= \sum_{i-1}^{n} [m(\mathbf{x}_i\beta_0) - m(\mathbf{x}_i\beta) + \varepsilon_i]^2 \\
&= \sum_{i-1}^{n} [-(m(\mathbf{x}_i\beta) - m(\mathbf{x}_i\beta_0)) + \varepsilon_i]^2 \\
&= \sum_{i-1}^{n} [-\delta_i + \varepsilon_i]^2 \\
&= \sum_{i-1}^{n} [\varepsilon_i - \delta_i]^2 \\
&= \sum_{i=1}^{n} \delta_i^2 - 2\sum_{i=1}^{n} \varepsilon_i\delta_i + \sum_{i=1}^{n} \varepsilon_i^2.
\end{aligned}$$

We now write $\widehat{S}(\beta)$ as

$$\begin{aligned}
\widehat{S}(\beta) &= \sum_{i=1}^{n} [y_i - \widehat{m}_{-i}(\mathbf{x}_i\beta)]^2 \\
&= \sum_{i=1}^{n} [m(\mathbf{x}_i\beta_0) - \widehat{m}_{-i}(\mathbf{x}_i\beta) + \varepsilon_i]^2 \\
&= \sum_{i=1}^{n} [\varepsilon_i - (\widehat{m}_{-i}(\mathbf{x}_i\beta) - m(\mathbf{x}_i\beta_0))]^2 \\
&= \sum_{i=1}^{n} [\varepsilon_i - (\widehat{m}_{-i}(\mathbf{x}_i\beta) - m(\mathbf{x}_i\beta) + m(\mathbf{x}_i\beta) - \widehat{m}_{-i}(\mathbf{x}_i\beta_0) + \widehat{m}_{-i}(\mathbf{x}_i\beta_0) - m(\mathbf{x}_i\beta_0))]^2 \\
&= \sum_{i=1}^{n} [\varepsilon_i - (\widehat{m}_{-i}(\mathbf{x}_i\beta) - m(\mathbf{x}_i\beta) - \widehat{m}_{-i}(\mathbf{x}_i\beta_0) + m(\mathbf{x}_i\beta_0) \\
&\qquad\qquad + m(\mathbf{x}_i\beta) + \widehat{m}_{-i}(\mathbf{x}_i\beta_0) - m(\mathbf{x}_i\beta_0) - m(\mathbf{x}_i\beta_0))]^2 \\
&= \sum_{i=1}^{n} [\varepsilon_i - (\triangle_i + (m(\mathbf{x}_i\beta) - m(\mathbf{x}_i\beta_0)) + (\widehat{m}_{-i}(\mathbf{x}_i\beta_0) - m(\mathbf{x}_i\beta_0)))]^2 \\
&= \sum_{i=1}^{n} [\varepsilon_i - (\triangle_i + \delta_i + D_i)]^2 .
\end{aligned}$$

This can be further written as

$$
\begin{aligned}
\widehat{S}(\beta) &= \sum_{i=1}^{n} \left[\varepsilon_i - (\triangle_i + \delta_i + D_i)\right]^2 \\
&= \sum_{i=1}^{n} \varepsilon_i^2 - 2\sum_{i=1}^{n} \varepsilon_i \left(\triangle_i + \delta_i + D_i\right) + \sum_{i=1}^{n} \left[\triangle_i + \delta_i + D_i\right]^2 .
\end{aligned}
$$

Lastly, consider the difference between $\widetilde{S}(\beta)$ and $\widehat{S}(\beta)$, which is

$$
\begin{aligned}
\widehat{S}(\beta) - \widetilde{S}(\beta) =& \sum_{i=1}^{n} \varepsilon_i^2 - 2\sum_{i=1}^{n} \varepsilon_i \left(\triangle_i + \delta_i + D_i\right) + \sum_{i=1}^{n} \left[\triangle_i + \delta_i + D_i\right]^2 \\
& - \left[\sum_{i=1}^{n} \delta_i^2 - 2\sum_{i=1}^{n} \varepsilon_i\delta_i + \sum_{i=1}^{n} \varepsilon_i^2\right] \\
=& -2\sum_{i=1}^{n} \varepsilon_i\triangle_i - 2\sum_{i=1}^{n} \varepsilon_i\delta_i - 2\sum_{i=1}^{n} \varepsilon_i D_i + \sum_{i=1}^{n} \triangle_i^2 + \sum_{i=1}^{n} \delta_i^2 + \sum_{i=1}^{n} D_i^2 \\
& + 2\sum_{i=1}^{n} \left(\triangle_i\delta_i + \triangle_i D_i + \delta_i D_i\right) + 2\sum_{i=1}^{n} \varepsilon_i\delta_i - \sum_{i=1}^{n} \delta_i^2 \\
=& \sum_{i=1}^{n} D_i^2 + \sum_{i=1}^{n} \triangle_i^2 + 2\sum_{i=1}^{n} \left[\triangle_i\delta_i + \triangle_i D_i + \delta_i D_i - \varepsilon_i\triangle_i - \varepsilon_i D_i\right].
\end{aligned}
$$

Härdle, Hall and Ichimura (1993) demonstrate that the terms $\sum_{i=1}^{n} \triangle_i^2$, $\sum_{i=1}^{n} \triangle_i\delta_i$, $\sum_{i=1}^{n} \triangle_i D_i$, $\sum_{i=1}^{n} \delta_i D_i$, $\sum_{i=1}^{n} \varepsilon_i\triangle_i$, and $\sum_{i=1}^{n} \varepsilon_i D_i$ are all asymptotically negligible. This leaves us with

$$
\widehat{S}(\beta) = \widetilde{S}(\beta) + \sum_{i=1}^{n} D_i^2 + \text{smaller order terms},
$$

Or,

$$
\begin{aligned}
CV(\widehat{\beta}, h) = \widehat{S}(\widehat{\beta}) &= \sum_{i=1}^{n} \left(y_i - \widehat{m}_{-i}(\mathbf{x}_i \widehat{\beta}) \right)^2 \\
&= \sum_{i=1}^{n} \left(y_i - m(\mathbf{x}_i \widehat{\beta}) \right)^2 + \sum_{i=1}^{n} \left[\widehat{m}_{-i}(\mathbf{x}_i \beta_0) - m(\mathbf{x}_i \beta_0) \right]^2 + \text{smaller order terms} \\
&\approx \sum_{i=1}^{n} \left(y_i - m(\mathbf{x}_i \widehat{\beta}) \right)^2 + \sum_{i=1}^{n} \left[\widehat{m}_{-i}(\mathbf{x}_i \beta_0) - m(\mathbf{x}_i \beta_0) \right]^2 ,
\end{aligned}
$$

as stated in the text.