THE BOOTSTRAP IN REGRESSION

First consider multiple linear regression.

Let $(\underset{\sim}{x}_1, Y_1), \ldots, (\underset{\sim}{x}_n, Y_n)$ be data pairs such that

$$Y_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta}_0 + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n \in \mathbb{R}^p$ are deterministic, $\varepsilon_1, \ldots, \varepsilon_n$ are iid with $\mathbb{E}\varepsilon_1 = 0$ and $\mathbb{E}\varepsilon_1^2 = \sigma^2 < \infty$.

Let $\mathbb{X}_n = (\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n)^T$ be the design matrix.

The least squares estimator of $\underset{\sim}{\beta}$ is given by

$$\hat{\underset{\sim}{\beta}}_n = \underset{\underset{\sim}{\beta} \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n \left[Y_i - \underset{\sim}{x}_i^T \underset{\sim}{\beta} \right]^2 = (\mathbb{X}_n^T \mathbb{X}_n)^{-1} \mathbb{X}_n^T \underset{\sim}{Y},$$

where $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)^T$.

By a multivariate version of the Lindeberg central limit theorem, we have

$$\sqrt{n}\,(n^{-1}\mathbb{X}_n^T\mathbb{X}_n)^{\frac{1}{2}} (\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}_0)/\sigma = \frac{1}{\sigma}\left(\sum_{i=1}^n \underset{\sim}{x}_i \underset{\sim}{x}_i^T\right)^{-\frac{1}{2}} \sum_{i=1}^n \underset{\sim}{x}_i\, \varepsilon_i \xrightarrow{D} N(0, I_p),$$

as $n \to \infty$, provided

$$\max_{1 \le i \le n} h_{ii} \longrightarrow 0 \quad \text{as} \quad n \to \infty,$$

where h_{ii} is the i^{th} diagonal entry of the matrix $\mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T$.

Moreover, for any $\underset{\sim}{c} \in \mathbb{R}^p$, $\sqrt{n}\left[\underset{\sim}{c}^T (n^{-1}\mathbb{X}_n^T\mathbb{X}_n)^{-1} \underset{\sim}{c}\right]^{-1/2} \underset{\sim}{c}^T (\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}_0)/\sigma \xrightarrow{D} N(0,1)$ as $n \to \infty$.

A residual bootstrap version of $\hat{\underset{\sim}{\beta}}_n$:

To construct a bootstrap version $\hat{\underset{\sim}{\beta}}_n^*$ of $\hat{\underset{\sim}{\beta}}_n$, define the residuals

$$\hat{\varepsilon}_i = Y_i - \underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}}_n \qquad i=1,\dots,n,$$

and introduce bootstrap residuals

$$\varepsilon_1^*,\dots,\varepsilon_n^* \mid Y_1,\dots,Y_n \overset{iid}{\sim} \frac{1}{n}\sum_{i=1}^n \delta_{\hat{\varepsilon}_i}.$$

Then set $Y_i^* = \underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}}_n + \varepsilon_i^*$, $i=1,\dots,n$ and

$$\hat{\underset{\sim}{\beta}}_n^* = \underset{\underset{\sim}{\beta}\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n \left[Y_i^* - \underset{\sim}{x}_i^T\underset{\sim}{\beta}\right]^2 = (\mathbb{X}_n^T\mathbb{X}_n)^{-1}\mathbb{X}_n^T \underset{\sim}{Y}^*,$$

where $\underset{\sim}{Y}^* = (Y_1^*,\dots,Y_n^*)^T$.

Now, letting $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i^2$, we have the following result:

Result (Bootstrap consistency for multiple linear regression):

Provided

$$\max_{1\le i\le n} h_{ii} \to 0 \text{ as } n\to\infty$$

we have, for any $\underset{\sim}{c}\in\mathbb{R}^p$ specifying a linear contrast,

$$\sup_{x\in\mathbb{R}} \Big| P_*\Big(\sqrt{n}\,\left[\underset{\sim}{c}^T(n^{-1}\mathbb{X}_n^T\mathbb{X}_n)^{-1}\underset{\sim}{c}\right]^{-1/2} \underset{\sim}{c}^T(\hat{\underset{\sim}{\beta}}_n^* - \hat{\underset{\sim}{\beta}}_n)/\hat{\sigma}_n \le x\Big)$$
$$- P\Big(\sqrt{n}\left[\underset{\sim}{c}^T(n^{-1}\mathbb{X}_n^T\mathbb{X}_n)^{-1}\underset{\sim}{c}\right]^{-1/2} \underset{\sim}{c}^T(\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}_0)/\sigma \le x\Big)\Big| \to 0$$

in probability as $n \to \infty$, where P_* denotes probability conditional on $Y_1, \ldots, Y_n$.

That is, the bootstrap "works" for multiple linear regression.
In the above, $\mathcal{B}(\mathbb{R}^p)$ is the collection of all Borel sets in $\mathbb{R}^p$.

A detailed proof is given in the Lec 11 supplement.
For studentized version replace σ with $\hat{\sigma}_n$ and $\hat{\sigma}_n$ with $\hat{\sigma}_n^*$, where

$$\hat{\sigma}_n^{*2} = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^{*2}, \qquad \hat{\varepsilon}_i^* = Y_i^* - \underset{\sim}{x}_i^T \underset{\sim}{\hat{\beta}}_n^*, \qquad i = 1, \ldots, n.$$

We can show that for any $\underset{\sim}{c} \in \mathbb{R}^p$, we have

$$\sup_{x \in \mathbb{R}} \Big| P_*\Big(\sqrt{n} \left[\underset{\sim}{c}^T (n^{-1} \mathbb{X}_n^T \mathbb{X}_n)^{-1} \underset{\sim}{c} \right]^{-1/2} \underset{\sim}{c}^T (\underset{\sim}{\hat{\beta}}_n^* - \underset{\sim}{\hat{\beta}}_n) / \hat{\sigma}_n^* \le x \Big)$$

$$- P\Big(\sqrt{n} \left[\underset{\sim}{c}^T (n^{-1} \mathbb{X}_n^T \mathbb{X}_n)^{-1} \underset{\sim}{c} \right]^{-1/2} \underset{\sim}{c}^T (\underset{\sim}{\hat{\beta}}_n - \underset{\sim}{\beta}_0) / \hat{\sigma}_n \le x \Big) \Big| \to 0$$

in probability as $n \to \infty$.

Exercise: Construct

(i) a C.I. for β_{0j} based on the Normal distribution

(ii) residual bootstrap C.I. for β_{0j} using the above.

Solution: Letting $\underset{\sim}{c} = \underset{\sim}{e}_j^T = \left(\mathbb{1}(k=j) \right)_{1 \le k \le p}^T$, we have

$$\underset{\sim}{c}^T (\underset{\sim}{\hat{\beta}}_n - \underset{\sim}{\beta}_0) = \hat{\beta}_{nj} - \beta_{0j}$$

and

$$\underset{\sim}{c}^T \left(n^{-1} \mathbb{X}_n^T \mathbb{X}_n \right)^{-1} \underset{\sim}{c} = \hat{\Omega}_{jj},$$

where $\hat{\Omega} = \left(n^{-1} \mathbb{X}_n^T \mathbb{X}_n \right)^{-1}$.

For (i) it can be shown by the multivariate Lindeberg C.L.T. that

$$\sqrt{n} \left[\underset{\sim}{c}^T \left(n^{-1} \mathbb{X}_n^T \mathbb{X}_n \right)^{-1} \underset{\sim}{c} \right]^{-1/2} \underset{\sim}{c}^T \left(\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}_0 \right) / \hat{\sigma}_n \xrightarrow{D} N(0, I_m) \quad \text{as} \quad n \to \infty,$$

so

$$\frac{\sqrt{n} \left(\hat{\beta}_{nj} - \beta_{0j} \right)}{\sqrt{\hat{\Omega}_{jj}} \; \hat{\sigma}_n} \xrightarrow{D} N(0,1) \quad \text{as} \quad n \to \infty.$$

Therefore

$$\hat{\beta}_{nj} \pm z_{\alpha/2} \sqrt{\hat{\Omega}_{jj}} \frac{\hat{\sigma}_n}{\sqrt{n}}$$

contains β_{0j} with probability converging to $1-\alpha$ as $n \to \infty$.

For (ii), let $\hat{G}_{nj}$ be the cdf

$$\hat{G}_{nj}(x) = P_* \left(\frac{\sqrt{n} \left(\hat{\beta}^*_{nj} - \hat{\beta}_{nj} \right)}{\sqrt{\hat{\Omega}_{jj}} \; \hat{\sigma}^*_n} \le x \right).$$

Then a $(1-\alpha) \cdot 100\%$ residual bootstrap C.I. for β_{0j} is

$$\left(\hat{\beta}_{nj} - \hat{G}_{nj}^{-1}(1-\alpha/2) \sqrt{\hat{\Omega}_{jj}} \frac{\hat{\sigma}_n}{\sqrt{n}}, \quad \hat{\beta}_{nj} - \hat{G}_{nj}^{-1}(\alpha/2) \sqrt{\hat{\Omega}_{jj}} \frac{\hat{\sigma}_n}{\sqrt{n}} \right).$$

In practice, given B Monte Carlo realizations of the pivot

$$T_j^* = \frac{\sqrt{n}\,(\hat{\beta}^*_{nj} - \hat{\beta}_{nj})}{\sqrt{\hat{\Omega}_{jj}}\;\hat{\sigma}_n^*},$$

and ordering them as, say $T_j^{*(1)} \le \dots \le T_j^{*(B)}$, we can put

$$\hat{G}_{nj}^{-1}(\alpha/2) = T_j^{*(\lceil(\alpha/2)\cdot n\rceil)}, \qquad \hat{G}_{nj}^{-1}(1-\alpha/2) = T_j^{*(\lceil(1-\alpha/2)\cdot n\rceil)}.$$

Refinements via Edgeworth expansion have been explored.
See Peter Hall's book.

Wild bootstrap for heteroskedastic error terms:

Let $(\underset{\sim}{x}_1, Y_1), \dots, (\underset{\sim}{x}_n, Y_n)$ be data pairs such that

$$Y_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta}_0 + \varepsilon_i, \qquad i = 1, \dots, n,$$

where $\underset{\sim}{x}_1, \dots, \underset{\sim}{x}_n \in \mathbb{R}^p$ are deterministic, and

$\varepsilon_1, \dots, \varepsilon_n$ are independent with mean zero and $\mathbb{E}\,\varepsilon_i^2 = \sigma_i^2 \in (0,\infty)$, $i = 1, \dots, n$.

So the error terms are allowed to have different variances.

Consider estimating or making inferences on a linear contrast $\underset{\sim}{c}^T \underset{\sim}{\beta}_0$
for some $\underset{\sim}{c} \in \mathbb{R}^p$.

We have

$$\mathrm{Var}\left(\sqrt{n}\cdot \underset{\sim}{c}^T \hat{\underset{\sim}{\beta}}_n\right) = n \cdot \underset{\sim}{c}^T (\mathbb{X}_n^T \mathbb{X}_n)^{-1} \mathbb{X}_n^T V_n \mathbb{X}_n (\mathbb{X}_n^T \mathbb{X}_n)^{-1} \underset{\sim}{c} =: \sigma_{c,n}^2$$

where

$$V_n = \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_n^2\right),$$

and, under some conditions,

$$\sqrt{n}\ \underset{\sim}{c}^T\left(\underset{\sim}{\hat{\beta}}_n - \underset{\sim}{\hat{\beta}}_0\right)/\sigma_{c,n} \xrightarrow{D} N(0,1)$$

as $n \to \infty$.

This would give a C.I. for $\underset{\sim}{c}^T\underset{\sim}{\beta}_0$ of the form

$$\underset{\sim}{c}^T\underset{\sim}{\hat{\beta}}_n \pm z_{\alpha/2}\,\frac{\sigma_{c,n}}{\sqrt{n}}.$$

However, we do not know $\sigma_1^2, \ldots, \sigma_n^2$.

Under some conditions, the following timator is consistent for σ_c^2:

$$\hat{\sigma}_{c,n}^2 = n \cdot \underset{\sim}{c}^T\left(\mathbb{X}_n^T\mathbb{X}_n\right)^{-1}\mathbb{X}_n^T \operatorname{diag}\left(\hat{\varepsilon}_1^2, \ldots, \hat{\varepsilon}_n^2\right)\ \mathbb{X}_n\left(\mathbb{X}_n^T\mathbb{X}_n\right)^{-1}\underset{\sim}{c}.$$

This leads to the feasible interval

$$\underset{\sim}{c}^T\underset{\sim}{\hat{\beta}}_n \pm z_{\alpha/2}\,\frac{\hat{\sigma}_{c,n}}{\sqrt{n}}.$$

If we use the residual bootstrap in the heteroskedastic setting, sampling from the residuals with replacement, we will scramble the structure of the heteroskedasticity.

Here is where the Wild bootstrap comes in.

Wild bootstrap:

For each $i = 1, \dots, n$, the wild bootstrap prescribes a bootstrap residual ε_i^{*W} which satisfies the three conditions

$$(i) \quad \mathbb{E}_*\left[\varepsilon_i^{*W}\right] = 0$$

$$(ii) \quad \mathbb{E}_*\left[\left(\varepsilon_i^{*W}\right)^2\right] = \hat{\varepsilon}_i^{\,2}$$

$$(iii) \quad \mathbb{E}_*\left[\left(\varepsilon_i^{*W}\right)^3\right] = \hat{\varepsilon}_i^{\,3},$$

when $\mathbb{E}_*$ is expectation conditional on the observed data.

Having drawn $\varepsilon_1^{*W}, \dots, \varepsilon_n^{*W}$ from distributions satisfying (i), (ii), and (iii), let

$$Y_i^{*W} = \underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}}_n + \varepsilon_i^{*W}, \qquad i = 1, \dots, n.$$

Then the wild bootstrap version $\hat{\underset{\sim}{\beta}}_n^{*W}$ of $\hat{\underset{\sim}{\beta}}_n$ is given by

$$\hat{\underset{\sim}{\beta}}_n^{*W} = (\mathbb{X}_n^T \mathbb{X}_n)^{-1} \mathbb{X}_n^T \underset{\sim}{Y}^{*W},$$

where $\underset{\sim}{Y}^{*W} = (Y_1^{*W}, \dots, Y_n^{*W})^T$.

Then the conditional cdf, given the observed data, of

$\sqrt{n}\, \underset{\sim}{c}^T (\hat{\underset{\sim}{\beta}}_n^{*W} - \hat{\underset{\sim}{\beta}}_n)$ is the wild bootstrap version of $\sqrt{n}\, \underset{\sim}{c}^T (\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}_0)$

$\sqrt{n}\, \dfrac{\underset{\sim}{c}^T (\hat{\underset{\sim}{\beta}}_n^{*W} - \hat{\underset{\sim}{\beta}}_n)}{\hat{\sigma}_{c,n}^*}$ is the wild bootstrap version of $\sqrt{n}\, \dfrac{\underset{\sim}{c}^T (\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}_0)}{\hat{\sigma}_{c,n}}$, where

$$\hat{\sigma}^{*2}_{c,n} = n \cdot \underset{\sim}{c}^T (\mathbb{X}_n^T \mathbb{X}_n)^{-1} \mathbb{X}_n^T \operatorname{diag}\left[(\hat{\varepsilon}_1^{*W})^2, \ldots, (\hat{\varepsilon}_n^{*W})^2 \right] \mathbb{X}_n (\mathbb{X}_n^T \mathbb{X}_n)^{-1} \underset{\sim}{c}$$

with

$$\hat{\varepsilon}_i^{*W} = Y_i^{*W} - \underset{\sim}{x}_i^T \underset{\sim}{\hat{\beta}}_n^{*W}, \qquad i = 1, \ldots, n.$$

Generating the wild bootstrap residuals:

Mammen (1993):

For each $i = 1, \ldots, n$, generate $V_{i,1}, V_{i,2} \overset{\text{ind}}{\sim} N(0,1)$ and set

$$U_i = (\delta_1 + V_{i,1}/\sqrt{2})(\delta_2 + V_{i,2}/\sqrt{2}) - \delta_1 \delta_2,$$

where

$$\delta_1 = \left(\frac{3}{4} + \frac{\sqrt{17}}{12} \right)^{1/2} \quad \text{and} \quad \delta_2 = \left(\frac{3}{4} - \frac{\sqrt{17}}{12} \right)^{1/2}.$$

Then let

$$\varepsilon_i^{*W} = \hat{\varepsilon}_i \cdot U_i$$

Das (2019):

For each $i = 1, \ldots, n$, generate $U_i \sim \text{Beta}(1/2, 3/2)$. Then set

$$\varepsilon_i^{*W} = \hat{\varepsilon}_i \cdot 4(U_i - 1/4).$$

Result: Performance of Wild Bootstrap (From Mammen (1993))

Under conditions given in Mammen (1993),

$$\sup_{x \in \mathbb{R}} \left| P_*\left(\sqrt{n}\, \underset{\sim}{c}^T \left(\hat{\underset{\sim}{\beta}}_n^{*W} - \hat{\underset{\sim}{\beta}}_n \right) \le x \right) - P\left(\sqrt{n}\, \underset{\sim}{c}^T \left(\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}_0 \right) \le x \right) \right| = O_p\left(n^{-1/2} + p n^{-1}\right)$$

and

$$\sup_{x \in \mathbb{R}} \left| P_*\left(\sqrt{n}\, \frac{\underset{\sim}{c}^T \left(\hat{\underset{\sim}{\beta}}_n^{*W} - \hat{\underset{\sim}{\beta}}_n \right)}{\hat{\sigma}^*_{c,n}} \le x \right) - P\left(\sqrt{n}\, \frac{\underset{\sim}{c}^T \left(\hat{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}_0 \right)}{\hat{\sigma}_{c,n}} \le x \right) \right| = O_p\left(n^{-1} + p n^{-3/2}\right)$$

as $n \to \infty$.

Resampling pairs:

One can also construct a bootstrap version $\hat{\underset{\sim}{\beta}}_n^*$ of $\hat{\beta}_n$ by resampling from the data pairs $(\underset{\sim}{x}_1, Y_1), \dots, (\underset{\sim}{x}_n, Y_n)$

That is, drawing $(\underset{\sim}{x}_1^*, Y_1^*), \dots, (\underset{\sim}{x}_n^*, Y_n^*)$ from $(\underset{\sim}{x}_1, Y_1), \dots, (\underset{\sim}{x}_n, Y_n)$ with replacement and defining

$$\hat{\beta}_n^* = (\mathbb{X}_n^{*T} \mathbb{X}_n^*)^{-1} \mathbb{X}_n^{*T} \underset{\sim}{Y}^*,$$

where $\mathbb{X}_n^* = (\underset{\sim}{x}_1^*, \dots, \underset{\sim}{x}_n^*)^T$ and $\underset{\sim}{Y}^* = (Y_1^*, \dots, Y_n^*)$.

This has been proposed for the setting in which we consider the design points $\underset{\sim}{x}_1, \dots, \underset{\sim}{x}_n$ to be random instead of fixed.

Mammen (1993) argued, however, that the Wild bootstrap is appropriate for this setting.

Resampling pairs is much more computationally expensive, as one must compute the inverse of $\mathbb{X}_n^{*T}\mathbb{X}_n^*$ for every bootstrap sample.

Mammen (1993) compared the resampling pairs bootstrap with the Wild bootstrap, while tracking the effect of the dimension p, and gave this table:

Estimation of	$\mathcal{L}(\sqrt{n}\, c^T(\hat{\beta}-\beta))$	$\mathcal{L}(\sqrt{n}\, c^T(\hat{\beta}-\beta)/\hat{\sigma}_c)$
Normal approximation $N(0,\hat{\sigma}_c^2)$	$O_P(n^{-1/2} + p\, n^{-1})$	-
Wild bootstrap	$O_P(n^{-1/2} + p\, n^{-1})$	$\boxed{O_P(n^{-1} + p\, n^{-3/2})}$
Bootstrap (resampling pairs)	$O_P(n^{-1/2} + p\, n^{-1})$	$O_P(p\, n^{-1})$

less adversely affected by large p.

Table 1. *Rates of convergence of the bootstrap procedures and the mean zero normal approximation under the assumption* $E(\varepsilon_i \mid X_i) = 0$.

This argues in favor of the wild bootstrap over the resampling pairs bootstrap.

BOOTSTRAP CONFIDENCE BANDS IN NONPARAMETRIC REGRESSION

Now consider the nonparametric setting

$$Y_i = m(x_i) + \varepsilon_i, \quad i = 1, \ldots, n$$

where $x_1, \ldots, x_n \in [0,1]$ are deterministic and $\varepsilon_1, \ldots, \varepsilon_n$ are independent with

$$\mathbb{E}\,\varepsilon_i = 0, \quad \mathbb{E}\,\varepsilon_i^2 = \sigma_i^2 \in (0, \infty)$$

for $i = 1, \ldots, n$, so that we can accomodate heteroscedastic error term variances.

We consider linear estimators of m; that is, we consider estimators of the form

$$\hat{m}_n(x) = \sum_{i=1}^{n} W_{ni}(x)\, Y_i$$

for some weights $W_{n1}(x), \ldots, W_{nn}(x)$, $x \in [0,1]$.

We consider two methods based on undersmoothing — deliberately choosing an estimator that is too wiggly in regards to MSE, in order that the bias be of smaller order than the standard deviation. This is necessary to have asymptotically correct confidence intervals in nonparametric regression.

We proceed, given undersmoothing of the estimator with the assumption that the bias is negligible, treating $\mathbb{E}\hat{m}_n(x)$ as though it were equal to $m(x)$.

Constant-variance case:

Let $\mathbb{E}\,\varepsilon_i^2 = \sigma^2 \in (0,\infty)$ for all $i=1,\dots,n$, and define the cdf

$$G_n(z) = P\left(\sup_{x\in[0,1]} \left| \frac{\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x)}{\hat{\sigma}_n \sqrt{\sum_{i=1}^n W_{ni}^2(x)}} \right| \le z \right)$$

where $\hat{\sigma}_n$ is some consistent estimator of σ.

We consider $(1-\alpha)100\%$ confidence bands for m which have the form

$$\left\{ (x,y) : \hat{m}_n(x) - G_n^{-1}(1-\alpha)\,\hat{\sigma}_n \sqrt{\sum_{i=1}^n W_{ni}^2(x)} \le y \le \hat{m}_n(x) + G_n^{-1}(1-\alpha)\,\hat{\sigma}_n \sqrt{\sum_{i=1}^n W_{ni}^2(x)},\ x\in[0,1] \right\},$$

where $G_n^{-1}(1-\alpha)$ is replaced by an approximation or an estimate.

Residual bootstrap

A residual bootstrap estimator of the cdf G_n is

$$\hat{G}_n(z) = P_*\left(\sup_{x\in[0,1]} \left| \frac{\hat{m}_n^*(x) - \mathbb{E}_*\,\hat{m}_n^*(x)}{\hat{\sigma}_n^* \sqrt{\sum_{i=1}^n W_{ni}^2(x)}} \right| \le z \right),$$

where we construct $\hat{m}_n^*$ and $\hat{\sigma}_n^*$ as follows:

Draw $\varepsilon_1^*,\dots,\varepsilon_n^*$ with replacement from $\hat{\varepsilon}_1,\dots,\hat{\varepsilon}_n$.

Set $Y_i^* = \hat{m}_n(X_i) + \varepsilon_i^*$, $i=1,\dots,n$.

Then $\hat{m}_n^*(x) = \sum_{i=1}^n W_{ni}(x)\,Y_i^*$ and $\hat{\sigma}_n^*$ is computed with $(X_1,Y_1^*),\dots,(X_n,Y_n^*)$.

Note that

$$
\begin{aligned}
\mathbb{E}_* \hat{m}_n^*(x) &= \mathbb{E}_* \sum_{i=1}^n W_{ni}(x) Y_i^* \\
&= \sum_{i=1}^n W_{ni}(x) \left[\hat{m}_n(X_i) + \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i \right] \\
&= \sum_{i=1}^n W_{ni}(x) \hat{m}_n(X_i),
\end{aligned}
$$

provided $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ sum to zero, which gives

$$
\hat{m}_n^*(x) - \mathbb{E}_* \hat{m}_n^*(x) = \sum_{i=1}^n W_{ni}(x) Y_i^* - \sum_{i=1}^n W_{ni}(x) \hat{m}_n(X_i) = \sum_{i=1}^n W_{ni}(x) \varepsilon_i^* .
$$

Therefore, we can construct a Monte Carlo approximation to $\hat{G}_n$ by generating many realizations of

$$
T_n^* = \max_{1 \le j \le N} \left| \frac{\sum_{i=1}^n W_{ni}(x_j) \varepsilon_i^*}{\hat{\sigma}_n^* \sqrt{\sum_{i=1}^n W_{ni}^2(x_j)}} \right|
$$

for some grid of values $x_1, \dots, x_N \in [0,1]$.

For $x_1, \dots, x_N$ you could choose the design points $X_1, \dots, X_n$ if n large.

Given realizations $T_n^{*(1)} \le \dots \le T_n^{*(B)}$, let

$$
\hat{G}_n(z) = \frac{1}{B} \sum_{b=1}^B \mathbb{1}\left(T^{*(b)} \le z \right).
$$

$$
\hat{G}_n^{-1}(\alpha) = T_n^{*(\lfloor \alpha B \rfloor)} .
$$

A good reference is Neumann & Polzehl (1998).

Asymptotic approximation (Sun & Loader):

Noting that

$$\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x) = \sum_{i=1}^n W_{ni}(x)\,\varepsilon_i,$$

we may write

$$\sup_{x\in[0,1]} \left| \frac{\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x)}{\hat{\sigma}_n \sqrt{\sum_{i=1}^n W_{ni}^2(x)}} \right| = \sup_{x\in[0,1]} \left| \frac{\sum_{i=1}^n M_{ni}(x)\,\varepsilon_i}{\hat{\sigma}_n} \right|,$$

where

$$M_{ni}(x) = \frac{W_{ni}(x)}{\sqrt{\sum_{i=1}^n W_{ni}^2(x)}} \qquad i=1,\dots,n.$$

Sun & Loader give

$$P\left(\sup_{x\in[0,1]} \left| \frac{\sum_{i=1}^n M_{ni}(x)\,\varepsilon_i}{\sigma} \right| > c \right) \approx 2\left(1 - \Phi(c)\right) + \frac{K_0}{\pi} e^{-c^2/2}$$

for large c and large enough n, where

$$K_0 = \int_0^1 \sqrt{\sum_{i=1}^n \left[\frac{\partial}{\partial x} M_{ni}(x)\right]^2}\, dx.$$

Provided $\hat{\sigma}_n$ is a consistent estimator for σ, an asymptotic $(1-\alpha)100\%$ confidence band for m based on this result is given by

$$\left\{ (x,y) : \hat{m}_n(x) - c\cdot\hat{\sigma}_n \sqrt{\sum_{i=1}^n W_{ni}^2(x)} \le y \le \hat{m}_n(x) + c\cdot\hat{\sigma}_n \sqrt{\sum_{i=1}^n W_{ni}^2(x)},\quad x\in[0,1] \right\},$$

where c solves the equation

$$2\left(1 - \Phi(c)\right) + \frac{K_0}{\pi} e^{-c^2/2} = \alpha.$$

To compute (approximate) K_0 in practice, just find

$$M_{n1}(x_j), \ldots, M_{nn}(x_j), \qquad j = 1, \ldots, N$$

for a grid of values $x_1, \ldots, x_N \in [0,1]$. Then

$$K_0 = \int_0^1 \sqrt{\sum_{i=1}^{n} \left[\frac{\partial}{\partial x} M_{ni}(x)\right]^2} \, dx$$

$$= \sum_{j=1}^{N-1} \int_{x_j}^{x_{j+1}} \sqrt{\sum_{i=1}^{n} \left[\frac{\partial}{\partial x} M_{ni}(x)\right]^2} \, dx$$

$$\approx \sum_{j=1}^{N-1} \int_{x_j}^{x_{j+1}} \sqrt{\sum_{i=1}^{n} \left[\frac{M_{ni}(x_{j+1}) - M_{ni}(x_j)}{x_{j+1} - x_j}\right]^2} \, dx$$

$$= \sum_{j=1}^{N-1} \sqrt{\sum_{i=1}^{n} \left[M_{ni}(x_{j+1}) - M_{ni}(x_j)\right]^2} .$$

Thus, if we set

$$\underset{\sim}{M}_{ni}(x) = \frac{1}{\sqrt{\sum_{i=1}^{n} W_{ni}^2(x)}} \left(W_{n1}(x), \ldots, W_{nn}(x)\right)^T,$$

then

$$K_0 \approx \sum_{j=1}^{N-1} \left\| \underset{\sim}{M}_{ni}(x_{j+1}) - \underset{\sim}{M}_{ni}(x_j) \right\|_2 .$$

Heteroscedastic case:

Now allow $\mathbb{E}\,\varepsilon_i^2 = \sigma_i^2$, $0 < \sigma_1^2, \dots, \sigma_n^2 < \infty$. We consider the cdf

$$H_n(z) = P\left(\sup_{x \in [0,1]} \left| \frac{\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x)}{\sqrt{\sum_{i=1}^n W_{ni}^2(x)\,\hat{\varepsilon}_i^2}} \right| \le z \right)$$

We consider $(1-\alpha)100\%$ confidence bands for m which have the form

$$\left\{ (x,y) : \hat{m}_n(x) - H_n^{-1}(1-\alpha)\sqrt{\sum_{i=1}^n W_{ni}^2(x)\,\hat{\varepsilon}_i^2} \le y \le \hat{m}_n(x) + H_n^{-1}(1-\alpha)\sqrt{\sum_{i=1}^n W_{ni}^2(x)\,\hat{\varepsilon}_i^2},\quad x \in [0,1] \right\},$$

where $H_n^{-1}(1-\alpha)$ is replaced by an asymptotic approximation or a bootstrap estimator.

Wild Bootstrap

A wild bootstrap estimator of the cdf H_n is

$$\hat{H}_n(z) = P_*\left(\sup_{x \in [0,1]} \left| \frac{\hat{m}_n^*(x) - \mathbb{E}_*\,\hat{m}_n^*(x)}{\sqrt{\sum_{i=1}^n W_{ni}^2(x)\,\hat{\varepsilon}_i^{*2}}} \right| \le z \right),$$

where we construct $\hat{m}_n^*$ and $\hat{\sigma}_n^*$ as follows:

Draw $\varepsilon_1^*, \dots, \varepsilon_n^*$ based on $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ with the Mammen or Das methods.

Set $Y_i^* = \hat{m}_n(X_i) + \varepsilon_i^*$, $i = 1, \dots, n$.

Then $\hat{m}_n^*(x) = \sum_{i=1}^n W_{ni}(x)\, Y_i^*$ and $\hat{\varepsilon}_i^* = Y_i^* - \hat{m}_n^*(X_i)$, $i = 1, \dots, n$.

Asymptotic approximation:

We again use the asymptotic approximation from Sun & Loader. We write

$$\frac{\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x)}{\sqrt{\mathrm{Var}\{\hat{m}_n(x)\}}} = \frac{\sum_{i=1}^{n} W_{ni}(x)\,\varepsilon_i}{\sqrt{\sum_{j=1}^{n} W_{nj}(x)\,\sigma_j^2}} = \frac{\sum_{i=1}^{n} W_{ni}(x)\,\sigma_i \left(\varepsilon_i/\sigma_i\right)}{\sqrt{\sum_{j=1}^{n} W_{nj}^2(x)\,\sigma_j^2}} = \sum_{i=1}^{n} \bar{M}_{ni}(x)\left(\varepsilon_i/\sigma_i\right),$$

where

$$\bar{M}_{ni}(x) = \frac{W_{ni}(x)\,\sigma_i}{\sqrt{\sum_{j=1}^{n} W_{nj}^2(x)\,\sigma_j^2}}, \qquad i = 1, \dots, n,$$

noting that

$$P\left(\sup_{x\in[0,1]} \left| \sum_{i=1}^{n} \bar{M}_{ni}(x)\left(\varepsilon_i/\sigma_i\right)\right| > c\right) \approx 2\left(1 - \Phi(c)\right) + \frac{\kappa_0}{\pi} e^{-c^2/2}$$

Now we basically use the squared residuals as estimators of $\sigma_1^2, \dots, \sigma_n^2$, like

$$\hat{\sigma}_1^2 = \hat{\varepsilon}_1^2, \quad \dots, \quad \hat{\sigma}_n^2 = \hat{\varepsilon}_n^2.$$

Then we write

replace with $|\hat{\varepsilon}_j|$

$$\frac{\hat{m}_n(x) - \mathbb{E}\,\hat{m}_n(x)}{\sqrt{\sum_{j=1}^{n} W_{nj}^2(x)\,\hat{\varepsilon}_j^2}} = \frac{\sum_{i=1}^{n} W_{ni}(x)\,\sigma_i \left(\varepsilon_i/\sigma_i\right)}{\sqrt{\sum_{j=1}^{n} W_{nj}^2(x)\,\hat{\varepsilon}_j^2}} \approx \frac{\sum_{i=1}^{n} W_{ni}(x)\,|\hat{\varepsilon}_i| \left(\varepsilon_i/\sigma_i\right)}{\sqrt{\sum_{j=1}^{n} W_{nj}^2(x)\,\hat{\varepsilon}_j^2}} = \sum_{i=1}^{n} \hat{M}_{ni}(x)\left(\frac{\varepsilon_i}{\sigma_i}\right),$$

where

$$\hat{M}_{ni}(x) = \frac{W_{ni}(x)\,|\hat{\varepsilon}_i|}{\sqrt{\sum_{j=1}^{n} W_{nj}^2(x)\,\hat{\varepsilon}_j^2}}, \qquad i = 1, \dots, n.$$

An asymptotic $(1-\alpha)100\%$ confidence band for m based on this result is given by

$$\left\{ (x,y) : \hat{m}_n(x) - c \cdot \sqrt{\sum_{i=1}^n W_{ni}^2(x)\,\hat{\varepsilon}_i^2} \le y \le \hat{m}_n(x) + c \cdot \sqrt{\sum_{i=1}^n W_{ni}^2(x)\,\hat{\varepsilon}_i^2}, \quad x \in [0,1] \right\},$$

where c is obtained as in the constant-variance case, but with K_0 computed using $\hat{\mu}_{n1}(x), \dots, \hat{\mu}_{nn}(x)$.

We can also try to estimate the function $\sigma^2 : [0,1] \to (0,\infty)$ explicitly. See Wasserman, pg 92, for details.

Recall that these methods are based on undersmoothing. However, it is difficult in practice to know what amount of undersmoothing to use. How do you choose the tuning parameters?

For a method not based on undersmoothing, see Hall & Horowitz (2013).

ADDENDUM:

Result: Under iid errors with variance σ^2, we have

$$\frac{\sqrt{n}\ \underset{\sim}{c}^T(\hat{\beta}_n - \beta)}{[\underset{\sim}{c}^T(\frac{1}{n}X^TX)^{-1}\underset{\sim}{c}]^{1/2}\cdot\sigma} \to N(0,1) \quad \text{in distribution}$$

as $n \to \infty$, provided $\max_{1\le i\le n} h_{ii} \to 0$ as $n \to \infty$.

Proof: We write

$$\frac{\sqrt{n}\ \underset{\sim}{c}^T(\hat{\beta}_n - \beta)}{[\underset{\sim}{c}^T(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n)^{-1}\underset{\sim}{c}]^{1/2}\cdot\sigma} = \frac{\sqrt{n}\ \underset{\sim}{c}^T(\mathbb{X}_n^T\mathbb{X}_n)^{-1}\mathbb{X}_n^T\underset{\sim}{\varepsilon}}{[\underset{\sim}{c}^T(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n)^{-1}\underset{\sim}{c}]^{1/2}\cdot\sigma} = \frac{\underset{\sim}{a}^T\underset{\sim}{\varepsilon}}{\sigma\sqrt{\underset{\sim}{a}^T\underset{\sim}{a}}}$$

where $\underset{\sim}{a} = \sqrt{n}\ \mathbb{X}_n(\mathbb{X}_n^T\mathbb{X}_n)^{-1}\underset{\sim}{c}$, noting that

$$\begin{aligned}
\underset{\sim}{a}^T\underset{\sim}{a} &= n\ \underset{\sim}{c}^T(\mathbb{X}_n^T\mathbb{X}_n)^{-1}\mathbb{X}_n^T\mathbb{X}_n(\mathbb{X}_n^T\mathbb{X}_n)^{-1}\underset{\sim}{c} \\
&= \underset{\sim}{c}^T(\tfrac{1}{n}\mathbb{X}_n^T\mathbb{X}_n)^{-1}(\tfrac{1}{n}\mathbb{X}_n^T\mathbb{X}_n)(\tfrac{1}{n}\mathbb{X}_n^T\mathbb{X}_n)^{-1}\underset{\sim}{c} \\
&= \underset{\sim}{c}^T(\tfrac{1}{n}\mathbb{X}_n^T\mathbb{X}_n)^{-1}\underset{\sim}{c}.
\end{aligned}$$

According to the corollary to the Lindeberg C.L.T. from lecture 4, it is sufficient to show

$$\|\underset{\sim}{a}\|_\infty / \|\underset{\sim}{a}\|_2 \to 0 \quad \text{as} \quad n \to \infty.$$

We have

$$\frac{\|\underset{\sim}{a}\|_\infty}{\|\underset{\sim}{a}\|_2} = \frac{\max_{1\le i\le n}\left|\sqrt{n}\ \underset{\sim}{x}_i^T(\mathbb{X}_n^T\mathbb{X}_n)^{-1}\underset{\sim}{c}\right|}{\sqrt{\underset{\sim}{c}^T(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n)^{-1}\underset{\sim}{c}}}$$

$$= \frac{1}{\sqrt{n}} \frac{\max_{1\le i\le n} \left| \underset{\sim}{x}_i^T \left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)^{-1/2} \left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)^{-1/2} \underset{\sim}{c} \right|}{\left\| \left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)^{-1/2} \underset{\sim}{c} \right\|_2}$$

$$\le \frac{1}{\sqrt{n}} \frac{\max_{1\le i\le n} \left\| \underset{\sim}{x}_i^T \left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)^{-1/2} \right\|_2 \left\| \left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)^{-1/2} \underset{\sim}{c} \right\|_2}{\left\| \left(\frac{1}{n}\mathbb{X}^T\mathbb{X}\right)^{-1/2} \underset{\sim}{c} \right\|_2}$$

$$\le \frac{1}{\sqrt{n}} \max_{1\le i\le n} \sqrt{\underset{\sim}{x}_i^T \left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)^{-1} \underset{\sim}{x}_i}$$

$$= \sqrt{\max_{1\le i\le n} h_{ii}}$$

which goes to zero as $n \to \infty$ by assumption.

__Result:__ If $\mathbb{E}\,\varepsilon_i^2 = \sigma_i^2 \in (0,\infty)$, $i=1,\dots,n$, we have

$$\frac{\sqrt{n}\, \underset{\sim}{c}^T (\hat{\beta}_n - \beta)}{\left[n\, \underset{\sim}{c}^T (\mathbb{X}_n^T\mathbb{X}_n)^{-1} \mathbb{X}_n^T \mathbb{V}_n \mathbb{X}_n (\mathbb{X}_n^T\mathbb{X}_n)^{-1} \underset{\sim}{c}\right]^{1/2}} \longrightarrow N(0,1) \quad \text{in distribution}$$

as $n \to \infty$, where $\mathbb{V}_n = \operatorname{diag}(\sigma_1^2, \dots, \sigma_n^2)$, provided

$$\max_{1\le i\le n} \frac{\vartheta_{ii}}{\sigma_i} \longrightarrow 0 \quad \text{as} \quad n \to \infty,$$

where ϑ_{ii}, $i=1,\dots,n$ is the i^{th} diagonal entry of $\mathbb{X}(\mathbb{X}^T\mathbb{V}\mathbb{X})^{-1}\mathbb{X}^T$.

Proof: We have

$$\frac{\sqrt{n}\ \mathbf{c}^T(\hat{\beta}_n - \beta)}{\left[n\ \mathbf{c}^T(\mathbb{X}_n^T\mathbb{X}_n)^{-1}\mathbb{X}_n^T\mathbb{V}_n\mathbb{X}_n(\mathbb{X}_n^T\mathbb{X}_n)^{-1}\mathbf{c}\right]^{1/2}} = \frac{\sqrt{n}\ \mathbf{c}^T(\mathbb{X}_n^T\mathbb{X}_n)^{-1}\mathbb{X}_n^T\mathbb{V}_n^{1/2}\mathbb{V}_n^{-1/2}\boldsymbol{\varepsilon}}{\mathbf{c}^T\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)^{-1}\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{V}_n\mathbb{X}_n\right)\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)^{-1}\mathbf{c}}$$

$$= \frac{\mathbf{a}^T(\mathbb{V}_n^{-1/2}\boldsymbol{\varepsilon})}{\sqrt{\mathbf{a}^T\mathbf{a}}},$$

where $\mathbf{a} = \sqrt{n}\,\mathbb{V}_n^{1/2}\mathbb{X}_n(\mathbb{X}_n^T\mathbb{X}_n)^{-1}\mathbf{c}$ and $\mathrm{Cov}(\mathbb{V}^{-1/2}\boldsymbol{\varepsilon}) = I_n$, noting that

$$\mathbf{a}^T\mathbf{a} = n\ \mathbf{c}^T(\mathbb{X}_n^T\mathbb{X}_n)^{-1}\mathbb{X}_n^T\mathbb{V}_n\mathbb{X}_n(\mathbb{X}_n^T\mathbb{X}_n)\mathbf{c}$$

$$= \mathbf{c}^T\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)^{-1}\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{V}_n\mathbb{X}_n\right)\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)\mathbf{c}$$

$$= \left\|\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{V}_n\mathbb{X}_n\right)^{1/2}\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)\mathbf{c}\right\|_2^2.$$

It suffices to show $\|\mathbf{a}\|_\infty / \|\mathbf{a}\|_2 \to 0$ as $n \to \infty$. We have

$$\frac{\|\mathbf{a}\|_\infty}{\|\mathbf{a}\|_2} = \frac{\max_{1\le i\le n}\left|\sqrt{n}\ \frac{1}{\sigma_i}\mathbf{x}_i^T(\mathbb{X}_n^T\mathbb{X}_n)\mathbf{c}\right|}{\left\|\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{V}_n\mathbb{X}_n\right)^{1/2}\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)\mathbf{c}\right\|_2}$$

$$= \frac{\max_{1\le i\le n}\left|\frac{1}{\sqrt{n}}\ \frac{1}{\sigma_i}\mathbf{x}_i^T\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{V}_n\mathbb{X}_n\right)^{-1/2}\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{V}_n\mathbb{X}_n\right)^{1/2}\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)\mathbf{c}\right|}{\left\|\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{V}_n\mathbb{X}_n\right)^{1/2}\left(\frac{1}{n}\mathbb{X}_n^T\mathbb{X}_n\right)\mathbf{c}\right\|_2}$$

$$\leq \frac{1}{\sqrt{n}} \frac{\max\limits_{1 \leq i \leq n} \left\| \frac{1}{\sigma_i} \left(\frac{1}{n} \mathbb{X}_n^T \mathbb{V}_n \mathbb{X}_n\right)^{-1/2} \underset{\sim}{x}_i \right\|_2 \left\| \left(\frac{1}{n} \mathbb{X}_n^T \mathbb{V}_n \mathbb{X}_n\right)^{1/2} \left(\frac{1}{n} \mathbb{X}_n^T \mathbb{X}_n\right) \underset{\sim}{c} \right\|_2}{\left\| \left(\frac{1}{n} \mathbb{X}_n^T \mathbb{V}_n \mathbb{X}_n\right)^{1/2} \left(\frac{1}{n} \mathbb{X}_n^T \mathbb{X}_n\right) \underset{\sim}{c} \right\|_2}$$

$$\leq \sqrt{\max_{1 \leq i \leq n} \frac{\underset{\sim}{x}_i^T \left(\mathbb{X}_n^T \mathbb{V}_n \mathbb{X}_n\right)^{-1} \underset{\sim}{x}_i}{\sigma_i^2}}$$

$$= \sqrt{\max_{1 \leq i \leq n} \frac{v_{ii}}{\sigma_i^2}} \rightarrow 0$$

which goes to zero by assumption.

$$\mathbb{X} \left(\mathbb{X} \mathbb{V} \mathbb{X}\right)^{-1} \mathbb{X}^T \mathbb{V} \; Y$$