Why it is different.

Suppose $X_1, X_2, ..., X_T$ is a sequence of stationary weakly dependent variables.

- Weak dependence means, roughly, that random variables far apart from one another are nearly independent.
- Mixing.

Note that the distribution of this data sequence can no longer be described by a single marginal distribution function.

- We need the entire joint distribution.

Consider a situation where the parameter of interest is the population mean, such that $\theta = E(X_1)$.

Suppose we estimate $\theta$ by $\hat{\theta} = \frac{1}{T} \sum_{t=1}^{T} X_t$ and let $T_n = \hat{\theta} - \theta$.

Then $T_n$ depends on the entire joint distribution of $X_1, X_2, ..., X_T$. 
For example:

\[
V(T_n) = \frac{1}{T} \left[ V(X_1) + 2 \sum_{i=1}^{T-1} \left( 1 - \frac{i}{T} \right) Cov(X_1, X_{1+i}) \right]
\]

This depends on all the bivariate distributions of \(X_1\) and \(X_i\) for \(i = 1, \ldots, T\).

Now suppose we carry out the standard iid bootstrap in this situation. We create a sample \(X^*_1, X^*_2, \ldots, X^*_T\).

- Note that conditional on the original data, \(X^*_1, X^*_2, \ldots, X^*_T\) are actually iid variables.
- As a result, we would still have \(\hat{\theta}^* = \frac{1}{T} \sum_{t=1}^{T} X_t^*\) is a valid estimator of the mean.
- However, \(V(\hat{\theta}^*)\) is wrong.
Suppose \( \{X_t\}_{t \geq 1} \) is a sequence of random variables such that

\[
X_t = h(X_{t-1}, \ldots, X_{t-p}; \beta) + \varepsilon_t
\]

where:
- \( p < T \)
- \( \beta \) is a \( q \times 1 \) vector of parameters
- \( h: \mathbb{R}^{p+q} \rightarrow \mathbb{R} \) is a known function
- \( \{\varepsilon_t\}_{t>p} \) is an iid sequence of random variables that are independent of \( X_1, \ldots, X_p \).

Examples?

Let \( \hat{\beta}_n \) be the estimator of \( \beta \), and assume that we are interested in some statistic \( T_n \).
When the innovations are iid

Now, define the residuals:

$$\hat{\epsilon}_t = X_t - h (X_{t-1}, ..., X_{t-p}; \hat{\beta}_n), \ p < t \leq T$$

Note that in general

$$\bar{\epsilon}_T = \frac{1}{T - p} \sum_{t=1}^{T-p} \hat{\epsilon}_{t+p} \neq 0$$

For this reason, define the centered residuals:

$$\tilde{\epsilon}_t = \hat{\epsilon}_t - \bar{\epsilon}_T, \ p < t \leq T$$

It has been shown that if you forget to center, a non-disappearing random bias is introduced in the bootstrap.
Now, draw a random sample of size \((T - p), \varepsilon_p^*, \varepsilon_{p+1}^*, \ldots, \varepsilon_T^*\) from \(\{\tilde{\varepsilon}_t\}_{p < t \leq T}\).

Finally, create your bootstrap sample as:

\[
X_t^* = \begin{cases} 
  X_t & 1 \leq t \leq p \\
  h(X_{t-1}, \ldots, X_{t-p}; \hat{\beta}_n) + \varepsilon_t^* & p < t \leq T
\end{cases}
\]

Note that

- this method is similar to the residual bootstrap for regression models.
- The \(\varepsilon_t^*\) are iid and mean 0 by construction

There are some different versions of this method, but all with the same principle.

Once you have the bootstrap data, it is possible to calculate \(T^*\) and use the methods we covered under the iid bootstrap.
To use the previous method, we needed very specific information about the model.

We do not always have that information.

Return to the initial example:

- $X_1, X_2, \ldots, X_T$ is a sequence of stationary weakly dependent variables
- the parameter of interest is the population mean, $\theta$
- we estimate $\theta$ by $\hat{\theta} = \frac{1}{T} \sum_{t=1}^{T} X_t$ and let $T_n = \hat{\theta} - \theta$ and

\[
V(T_n) = \frac{1}{T} \left[ V(X_1) + 2 \sum_{i=1}^{T-1} \left(1 - \frac{i}{T}\right) \text{Cov}(X_1, X_{1+i}) \right]
\]

Because of weak dependence, the lower-order lag autocovariances are more important for $V(T_n)$ than the higher-order lag autocovariances.
The principle of the Block Bootstrap II

- For example, choose $\ell$ large but $\ell < T$. Then

$$\left| \sum_{i=\ell}^{T-1} \left(1 - \frac{i}{T}\right) \text{Cov} \left(X_1, X_{1+i}\right) \right| \leq \sum_{i=\ell}^{\infty} \left| \text{Cov} \left(X_1, X_{1+i}\right) \right| \xrightarrow{\ell \to \infty} 0$$

- That means we can approximate

$$V \left(T_n\right) \approx \frac{1}{\ell} \left[ V \left(X_1\right) + 2 \sum_{i=1}^{\ell-1} \left(1 - \frac{i}{T}\right) \text{Cov} \left(X_1, X_{1+i}\right) \right]$$

- This only depends on the joint distribution of the shorter series $X_1, X_2, ..., X_{\ell}$.

- The block bootstrap makes use of this fact.
Consider now a more general case:

- We are interested in $T_n (X_1, X_2, ..., X_T, \theta)$ where $\theta$ is a level one parameter and $T_n$ is invariant to permutations of $X_1, X_2, ..., X_T$.
- $T_n = \frac{1}{T} \sum_{t=1}^{T} X_t - \theta$ satisfies this.

Now suppose that $\ell$ is some integer such that both $\ell$ and $\frac{T}{\ell}$ are large.

- Example: $\ell = \lfloor T^{\delta} \rfloor$, $0 < \delta < 1$
- To make notation easier, assume $b = \frac{T}{\ell}$ is an integer.
- For the non-overlapping block bootstrap (NBB) we partition the data into $b$ blocks:

$$Y_1 = \{X_1, ..., X_{\ell}\}, ..., Y_b = \{X_{(b-1)\ell + 1}, ..., X_T\}$$

We then sample (with replacement) $b$ blocks $Y_1^*, ..., Y_b^*$ and, putting them together, get the bootstrap dataset $X_1^*, X_2^*, ..., X_T^*$. 
This method, like the others, attempts to re-create the relationship between the population and the sample.

Notation: Let $P_k$ denote the joint distribution of $(X_1, \ldots, X_k)$, $k \geq 1$.

Then, because of stationarity, $Y_1, \ldots, Y_b$ are all identically distributed with distribution $P_\ell$.

Also, because of the weak dependence assumption, $Y_1, \ldots, Y_b$ are approximately independent.
The basic principle is therefore:

- Assume stationarity and weak dependence.
- Then the data is approximately distributed as $P^b_\ell = P_\ell \otimes P_\ell \otimes \ldots \otimes P_\ell$.
- Specifically, the empirical distribution of $Y_1, \ldots, Y_b$ is $\tilde{P}^b_\ell$, which is an estimate of $P^b_\ell$ which is "close" to the true distribution.
- The bootstrap sample $Y^*_1, \ldots, Y^*_b = X^*_1, X^*_2, \ldots, X^*_T$ has the exact distribution $\tilde{P}^b_\ell$ (which is close to $P^b_\ell$, which is close to the true distribution)

Underlying all these "close to", are a lot of approximations, which require assumptions to hold.
The Block Bootstrap I
An Example

Suppose $X_1, \ldots, X_T$ are generated by a stationary $ARMA(1, 1)$ process:

$$X_t = \beta_1 X_{t-1} + \epsilon_t + \alpha_1 \epsilon_{t-1}$$

where $|\alpha_1| < 1$ and $|\beta_1| < 1$ are parameters at $\{\epsilon_t\}$ is a sequence of standard iid random variables.

Suppose we have 100 observations and we wish to use the NBB with block length $\ell = 5$.

Then we create 20 blocks:

$$B_1 = (X_1, \ldots, X_5), \quad B_2 = (X_6, \ldots, X_{10}), \ldots, B_{20} = (X_{96}, \ldots, X_{100})$$

We now re-sample 20 blocks $\{B_1^*, B_2^*, \ldots, B_{20}^*\}$. Use these to create the bootstrap dataset $X_1^*, \ldots, X_{100}^*$.

Now consider estimation of the variance of $\hat{\mu} = \bar{X}_T$. 
The Block Bootstrap II

An Example

- The bootstrap estimator is given by
  \[ \hat{V}(\overline{X}_T) = E_* (\overline{X}_T^* - E_* (\overline{X}_T^*))^2 = E_* (\overline{X}_T^* - \overline{X}_T)^2 \]

- It turns out that in the case we do NOT need to draw many bootstrap samples to evaluate this expression.

- Let \( V_j = \frac{1}{\ell} \sum_{i=1}^{\ell} X_{(j-1)\ell+i} = \overline{B}_j \) Then
  \[ \hat{V}(\overline{X}_T) = E_* \left( \frac{1}{b} \sum_{j=1}^{b} \overline{B}_j^* - \overline{X}_T \right)^2 = E_* \left( \frac{1}{b} \sum_{j=1}^{b} \overline{B}_j^* \right)^2 - \overline{X}_T^2 \]
  \[ = E_* \left( \frac{1}{b^2} \sum_{j=1}^{b} (\overline{B}_j^*)^2 \right) - \overline{X}_T^2 = \frac{1}{b^2} \sum_{j=1}^{b} (\overline{B}_j)^2 - \overline{X}_T^2 \]
Now suppose we wish to estimate the distribution of

\[ T_n = \frac{\sqrt{T} (\bar{X}_T - \mu)}{\hat{\sigma}_T}, \]

that is

\[ G_n (u) = P (T_n \leq u) \]

We then create the bootstrap version of \( T_n \):

\[ T^*_n = \frac{\sqrt{T} (\bar{X}^*_T - \bar{X}_T)}{\hat{\sigma}^*_T} \]

and we can find

\[ G^*_n (u) = P (T^*_n \leq u) \]
Here we are back to a situation where we need to draw many bootstrap samples to find $G_n^*(u)$.

A way of looking at this is that we are using Monte Carlo simulations to find our estimators.

Once we have $G_n^*(u)$, we can proceed to find critical values or create confidence intervals using the methods discussed under the iid bootstrap.
The block bootstrap we have seen is just one version. That one is called Non-overlapping Block Bootstrap or NBB.

Next we will describe the Moving Block Bootstrap, or MBB.

Let the data be $X_1, \ldots, X_n$ and consider estimators of the form $\hat{\theta} = T(F_n)$, where $F_n$ is the empirical distribution function of $X_1, \ldots, X_n$.

Let $\ell$ be an integer such that $\ell \to \infty$ and $\frac{\ell}{n} \to 0$ as $n \to \infty$.

Let $B_i = (X_i, \ldots, X_{i+\ell-1})$ be the block of length $\ell$ starting at observation $i$, $1 \leq i \leq N = n - \ell + 1$. 

Helle Bunzel (ISU) Bootstrap and Subsampling Dependent Data March 13, 2009 16 / 26
We are forming $N$ blocks and drawing each with probability $\frac{1}{N}$.

We then randomly sample $k$ blocks from $B_1, \ldots, B_N$ denoted by $B_1^*, \ldots, B_k^*$ and from here obtain data $X_1^*, \ldots, X_m^*$, where $m = k\ell$. 

**FIGURE 2.1.** The collection $\{B_1, \ldots, B_N\}$ of overlapping blocks under the MBB.
The bootstrap version of the estimator is then denoted by:

$$\theta_{m,n}^* = T(F_{m,n}^*)$$

where $F_{m,n}^*$ is the empirical distribution of $X_1^*, ..., X_m^*$.

Note that:

- Typically $k$ is picked such that the bootstrap sample is approximately the same size as the original sample.
- Each block is picked with probability $\frac{1}{N}$.
- If $\ell = 1$ this method reduces to the iid bootstrap.

Why do we eventually get independent bootstrap samples?
For timeseries we also need more general estimators. Consider for example the estimator of the autocovariance function at lag $k \geq 0$:

$$
\hat{\gamma}_n(k) = \frac{1}{n-k} \sum_{j=1}^{n-k} (X_{j+k} - \bar{X}_{n,k}) (X_j - \bar{X}_{n,k})
$$

This is not a simple function of $F_n$. Instead we can write this as a function of

$$
F_{n,p} = \frac{1}{n - p + 1} \sum_{j=1}^{n-p+1} F_{X_j,X_{j+1},...,X_{j+p-1}}
$$

such that

$$
\hat{\theta} = T(F_{n,p})
$$

For $\hat{\gamma}_n(k)$, we’d use $p = k + 1$
Many other estimators of value for timeseries can be written like this as well.

To get a MBB version of this estimator we:

1. Fix a block size \( \ell \) such that \( 1 < \ell \leq n - p \).
2. Define the blocks in terms of \( Y_i \)'s as
   \[
   \tilde{B}_j = (Y_j, \ldots, Y_{j+1-1}), \ 1 \leq j \leq n - p - \ell + 2
   \]
3. Now sample \( k \geq 1 \) blocks to generate bootstrap observations
   \[
   Y_1^*, \ldots, Y_{\ell}^*, Y_{\ell+1}^*, \ldots, Y_{2\ell}^*, \ldots, Y_m^*
   \]
   where \( m = \ell \cdot k \).
4. We can now define the bootstrap estimator as
   \[
   \theta_{m,n}^* = T(\tilde{F}_{m,n}^*)
   \]
   where \( \tilde{F}_{m,n}^* \) is the emirical distribution function of \( Y_1^*, \ldots, Y_m^* \).
Because this involves sampling blocks of $Y$'s, which are actually themselves blocks of $X$'s, this method is also called blocks of blocks.

There are variations on this method.
There are pros and cons of these two methods.

- The MBB has more possible blocks.
- The NBB has less finite sample dependence.

Compare a simple estimator calculated under the two different methods.

- Consider the sample mean: \( \hat{\theta}_n = \frac{1}{n} \sum_{j=1}^{n} X_j \)

Let the MBB estimator be denoted by

\[
\theta_{m,n} = \frac{1}{m} \sum_{j=1}^{m} X_j^*
\]
Comparing MBB and NBB II

- And the NBB estimator:

\[ \theta_{m,n}^{* (2)} = \frac{1}{m} \sum_{j=1}^{m} X_{2,j}^* \]

where the ”2” is used to signify the NBB estimator.

- Let us calculate the means of these estimators:

  - Note that we are calculating the mean when we’ve carried out ONE draw.

\[
E^* [\theta_{m,n}^*] = E^* \left[ \frac{1}{\ell} \sum_{i=1}^{\ell} X_i^* \right] = \frac{1}{N} \sum_{j=1}^{N} \left( \frac{1}{\ell} \sum_{i=1}^{\ell} X_{j+i-1} \right)
\]

\[
= \frac{1}{N \ell} \sum_{j=1}^{N} \left( \sum_{i=1}^{\ell} X_{j+i-1} \right)
\]

\[
= \frac{1}{N \ell} \left\{ (X_1 + X_2 + \ldots + X_\ell) + (X_2 + X_3 + \ldots + X_{\ell+1}) + \ldots + (X_{N+2} + \ldots + X_{N+\ell-1}) \right\}
\]
Recall that $N = n - \ell + 1$

$$E_* [\theta_{m,n}^*] = \frac{1}{N\ell} \left\{ (X_1 + X_2 + \ldots + X_\ell) + (X_2 + X_3 + \ldots + X_{\ell+1}) + \ldots + (X_{N+2} + \ldots + X_n) \right\}$$

$$= \frac{1}{N\ell} \left\{ n\bar{X}_n - \sum_{j=1}^{\ell-1} (\ell - j) (X_j + X_{n-j+1}) \right\}$$

$$= \frac{1}{N} \left\{ n\bar{X}_n - \frac{1}{\ell} \sum_{j=1}^{\ell-1} (\ell - j) (X_j + X_{n-j+1}) \right\}$$
Comparing MBB and NBB IV

- Now for the NBB estimator:

\[ E_\star \left[ \theta_{m,n}^{(2)} \right] = E_\star \left[ \frac{1}{\ell} \sum_{i=1}^{\ell} X_{2,i}^* \right] = \frac{1}{b} \sum_{j=1}^{b} \left[ \frac{1}{\ell} \sum_{i=1}^{\ell} X_{(j-1)\ell+i} \right] \]

\[ = \frac{1}{b\ell} \sum_{j=1}^{b} \sum_{i=1}^{\ell} X_{(j-1)\ell+i} = \frac{1}{b\ell} \sum_{i=1}^{b\ell} X_i \]

\[ = \frac{1}{b\ell} \left\{ n\bar{X}_n - \sum_{i=b\ell+1}^{n} X_i \right\} \]

- Note that these means are different. Under certain conditions, however, it is possible to show that

\[ E \left\{ \left( E_\star \left[ \theta_{m,n}^* \right] - E_\star \left[ \theta_{m,n}^{(2)} \right] \right)^2 \right\} = O \left( \frac{\ell}{n^2} \right) \]

- Therefore the differences between the two are small for large sample sizes.
One issue with both the MBB and the NBB are that the first and last observations have lower probability of appearing in the sample.

Methods proposed to fix this:

- **Circular Block Bootstrap.** Arrange data in a circle and then proceed as MBB.
  - It turns out that here the expected value of the bootstrap estimator is the original sample mean.

- **Stationary Block Bootstrap.** Uses stochastic blocklengths.

- **Subsampling:** Similar to MBB, except take only 1 sample.
  - Less computationally demanding
  - No automatic second order accuracy.