

# **The Bootstrap's Finite Sample Distribution: An Analytical Approach**

Lawrence C. Marsh<sup>1</sup>  
Department of Economics and Econometrics  
University of Notre Dame, Notre Dame, IN 46556-5611

## **Abstract**

This paper provides a method for determining the finite sample properties of the bootstrap. Previously, information about bootstrap performance has been primarily limited to asymptotic properties and Monte Carlo experiments. The exact small sample properties of most bootstrap procedures have not been determined. We show how to transform the empirical process into an analytical process and separate the bootstrap-induced randomness from the randomness of the underlying random variable. We provide examples of deriving the variances of some selected bootstrap estimators. Other properties such as bias, skewness, kurtosis, and mean squared error could be readily derived using this approach. Bootstrap estimators that are nonlinear functions of the bootstrap sample values can be represented by Taylor series as polynomials in their bootstrap-induced frequencies. Consequently, their finite sample distributions can be analyzed up to any desired degree of accuracy.

KEYWORDS: multinomial, wild bootstrap, block bootstrap, nonlinear, variance, bias

Initial Version: May 2003

Current Version: Tuesday, September 14, 2004

---

<sup>1</sup> I thank Tom Cosimano, Ken Kelley, Nelson Mark, Ke-Hai Yuan and my colleagues in the Department of Economics and Econometrics as well as . . . . . for helpful comments and suggestions.

## 1. INTRODUCTION

The purpose of this paper is to demonstrate an analytical approach to understanding the finite sample distribution induced by particular bootstrap procedures. We define the class of bootstrap estimators whose finite sample distributions are “directly analyzable” with this approach.<sup>2</sup> A Taylor series expansion can be used to transform nonlinear bootstrap structures into polynomials which are easily analyzed using this method. Examples are provided that give an analytical derivation of the variances of some selected bootstrap estimators. Our analytical approach enables one to determine mean squared error, skewness, kurtosis and a host of other properties of the finite sample distribution. We leave hypothesis testing and confidence interval issues for another paper.

From the beginning the bootstrap literature has recognized that bootstrap-induced randomness generally follows the multinomial distribution.<sup>3</sup> Efron (1979) used the analytical approach in some simple examples, but little has been done to develop this approach or the Taylor series approximation to it as a general analytical strategy for deriving the finite sample distributions of bootstrap estimators under the many alternative bootstrap procedures. Neither the traditional bootstrap literature nor the internet bootstrap literature has made much use of the multinomial distribution. In order to reach more general conclusions for the broader class of estimators including nonparametric estimators, the bootstrap literature has been almost exclusively focused on asymptotic theory and Monte Carlo results. With the increasing popularity of the bootstrap, more and more research is being performed using bootstrap simulations in situations where an analytical approach could provide more precise solutions.

---

<sup>2</sup> The term “directly analyzable” will be defined later in this section.

<sup>3</sup> An exception is the balanced bootstrap of Davison, Hinkley and Schechtman (1986) which is based on the hypergeometric distribution. Our approach is just as easy to use under the hypergeometric distribution as it is under the multinomial distribution. Other examples include the wild bootstrap which uses the multinomial and binomial distributions and the stationary bootstrap which uses the multinomial, binomial and geometric distributions.

In addition to the standard bootstrap, the examples offered herein of using the analytical approach include the wild bootstrap of Liu (1988) and Davidson and Flachaire (2001), the wild bootstrap of Mammen (1993), the non-overlapping, fixed-block bootstrap of Hall (1985) and Carlstein (1986), and the overlapping fixed-block bootstrap of Hall (1985), Künsch (1989), and Politis and Romano (1993). Bootstrap estimators that are nonlinear functions of the bootstrap sample values can be transformed by a Taylor series expansion into directly analyzable bootstrap estimators. We provide an example of this approach based on Horowitz (2001). Finally, the analytical solution for the covariance matrix of the bootstrap estimator of the coefficients of a linear regression model with nonstochastic regressors is derived. This paper was motivated by the Horowitz (2003) survey paper calling for work on the bootstrap's finite sample distribution.

Consider a set of real numbers  $\{X_i : i = 1, \dots, n\}$  where each  $X_i$  may be scalar or vector valued and where  $n$  is a finite, positive integer. Starting with Efron (1979) most past bootstrap applications have involved drawing (with replacement) a sample of size  $n$ . However, a more general bootstrap procedure involves randomly drawing a bootstrap sample of size  $m$  from the set  $\{X_i : i = 1, \dots, n\}$  where  $m$  is a finite, positive integer, which may be smaller than, greater than, or equal to  $n$ . Without loss of generality we will treat  $X_i$  as a scalar for notational convenience.

Each bootstrap sample consists of the set of real numbers  $\{X_j^* : j = 1, \dots, m\}$ . Since the sampling is done with replacement, the elements of the original sample may occur in the bootstrap sample once, more than once, or not at all. Consequently, the set  $\{M_i : i = 1, \dots, n\}$  records the frequency with which each of the corresponding elements of the original set  $\{X_i : i = 1, \dots, n\}$  occur in the bootstrap sample where the random variables  $M_i$  draw the corresponding realized values  $m_i$  from the set of non-negative integers such that  $\sum_{i=1}^n m_i = m$ .

**Definition:** Consider a set of real numbers  $\{X_i: i = 1, \dots, n\}$  where each  $X_i$  is drawn with corresponding frequencies  $\{M_i: i = 1, \dots, n\}$  in a sample of size  $m$  that produces a set of bootstrap values  $\{X_j^*: j = 1, \dots, m\}$ . Denote the set of any function,  $f$ , of each of the original sample values as  $\{f(X_i): i = 1, \dots, n\}$ . Denote the set of that same function of each of the individual bootstrapped values as  $\{f(X_j^*): j = 1, \dots, m\}$ . An expected value under the bootstrap-induced distribution will be denoted by the subscript  $M$ , as in  $E_M[\cdot]$ . Any statistic,  $\theta_n^*$ , that is a function of the elements of the set  $\{f(X_j^*): j = 1, \dots, m\}$  and satisfies the separability condition

$$\theta_n^* \left( \{f(X_j^*): j = 1, \dots, m\} \right) = \sum_{i=1}^n g(M_i) h(f(X_i))$$

where  $g(M_i)$  and  $h(f(X_i))$  are independent functions and the expected value  $E_M[g(M_i)]$  exists, is a “**directly analyzable**” bootstrap statistic.

This definition is given simply as a way to define the class of statistics addressed by this paper. Other bootstrap statistics might be directly analyzable after a suitable approximation.

In practice the bootstrap procedure might involve approximating the standard error or variance of a sample statistic. As an example, we will derive the finite sample variance of the bootstrap estimator under various bootstrap procedures. Since the bootstrap simulations treat the sample values of the underlying random variables as if they were the fixed set of population values, we will first follow this approach and calculate the variance of the statistic with these values as given. This produces a formula for precisely calculating the variance of the statistic of interest implied by the bootstrap procedure without actually having to generate the simulated bootstrap samples. Needless to say, this provides a simpler, faster and more precise determination of the variance, and, therefore, the corresponding standard error, than approximating this quantity by computer simulations.

However, while this formula for the variance tells us the formula that is implied by the repeated bootstrap simulations, it does not tell us the variance under the joint distribution of both

the bootstrap-induced randomness and the randomness of the underlying random variables associated with the original sample (as drawn from the true overall population). Consequently, we will also derive this quantity as well.

As a benchmark or reference we will start by analytically deriving the variance of any bootstrap sample moment over  $j = 1, \dots, m$  represented by the function,  $f$ . We will do this with the original sample values treated as population values, and then under the joint distribution of the bootstrap-induced randomness and the randomness that generated the original sample. To produce results that are as general as possible the latter will be expressed in terms of its moments. We will restrict our attention to estimators that are within the class of estimators that are directly analyzable as defined above or that can be made analyzable by a Taylor series expansion.

In section 2 we will derive the variance for the standard bootstrap estimator. Section 3 provides the variance for Liu-Davidson-Flachaire wild bootstrap estimator based on the Rademacher distribution as well as the variance for Mammen's version of the wild bootstrap. Section 4 derives the variances of both overlapping and non-overlapping versions of the fixed block bootstrap. Section 5 extends the analysis to bootstrap procedures for statistics that are nonlinear functions of the bootstrap sample values. Section 6 derives the covariance matrix of the bootstrap estimator of the coefficients of a linear regression model with nonstochastic regressors. Finally, section 7 concludes.

## 2. DERIVING THE VARIANCE UNDER THE STANDARD BOOTSTRAP PROCEDURE

The simplest case is, of course, the standard bootstrap, which will provide a benchmark or starting point for the use of this analytical approach for deriving the moments (in this case the variance) of the finite sample distribution of various bootstrap estimators under a variety of bootstrap procedures.

To simplify and expedite the exposition throughout this section, we provide *Condition 1* as a set of procedural assumptions and definitions that characterize the standard bootstrap.

**Condition 1:** Draw a sample of size  $m$  with replacement with equal probability from the original set of sample values corresponding to the variables  $\{X_i: i = 1, \dots, n\}$  having corresponding random frequencies  $\{M_i: i = 1, \dots, n\}$  in obtaining a set of bootstrapped values  $\{X_j^*: j = 1, \dots, m\}$ . Denote the set of any function,  $f$ , of each of the original sample values as  $\{f(X_i): i = 1, \dots, n\}$ . Denote the set of that same function of each of the individual bootstrapped values as  $\{f(X_j^*): j = 1, \dots, m\}$ . Any expected value under the bootstrap-induced distribution will be denoted by the subscript  $M$ , as in  $E_M[\cdot]$ . Any expected value under the joint distribution of the bootstrap-induced randomness and the randomness that generated the original sample,  $\{X_i: i = 1, \dots, n\}$ , will be denoted by the subscripts  $M$  and  $X$ , as in  $E_{M,X}[\cdot]$ . Conditional expectations will be denoted  $E_{M|X}[\cdot]$  and  $E_{X|M}[\cdot]$ . Variances and covariances will be denoted with corresponding subscripts such as  $Var_M[\cdot]$ ,  $Var_{M,X}[\cdot]$ ,  $Var_X[\cdot]$  and  $Cov_X[\cdot, \cdot]$ .

For example, using the set  $\{X_j^*: j = 1, \dots, m\}$  specify the standard bootstrap estimator of any population moment over  $j = 1, \dots, m$  represented by the function,  $f$ , of an individual bootstrap value,  $X_j^*$ , and re-express it in terms of the original set  $\{X_i: i = 1, \dots, n\}$  and the set of corresponding frequencies  $\{M_i: i = 1, \dots, n\}$  as

$$(2.1) \quad \frac{1}{m} \sum_{j=1}^m f(X_j^*) = \frac{1}{m} \sum_{i=1}^n M_i f(X_i).$$

For the standard bootstrap the  $M_i$ 's are distributed according to a multinomial distribution with  $m$  draws with replacement from an original sample of size  $n$  and with all of the underlying probabilities equal to  $1/n$ . The probability of observing a particular realization of the set  $\{M_i=m_i : i = 1, \dots, n\}$  may be expressed as

$$(2.2) \quad P(M_1=m_1, M_2=m_2, \dots, M_n=m_n) = \frac{n^{-m} m!}{m_1! m_2! \dots m_n!} \quad \text{where } m = \sum_{i=1}^n m_i.$$

We will now produce a lemma for any bootstrap sample moment defined by the function,  $f$ , and a second lemma for its square to produce the theorem for the variance of that moment.

**Lemma 1:** Under Condition 1,  $E_M \left[ \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right] = \frac{1}{n} \sum_{i=1}^n f(X_i).$

Proof of **Lemma 1:**

Substitute from (2.1) and note that a frequency,  $M_i$ , resulting from  $m$  random draws with replacement from a multinomial distribution with equal probabilities of  $1/n$  has a mean,

$E_M(M_i) = m/n$ , to obtain the result

$$\begin{aligned} E_M \left[ \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right] &= \sum_{M_1+\dots+M_n=m} \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \frac{n^{-m} m!}{(M_1! \dots M_n!)} = \sum_{M_1+\dots+M_n=m} \left( \frac{1}{m} \sum_{i=1}^n M_i f(X_i) \right) \frac{n^{-m} m!}{(M_1! \dots M_n!)} \\ &= \frac{1}{m} \sum_{i=1}^n \left[ \sum_{M_1+\dots+M_n=m} M_i \frac{n^{-m} m!}{(M_1! \dots M_n!)} \right] f(X_i) = \frac{1}{m} \sum_{i=1}^n [E_M(M_i)] f(X_i) = \frac{1}{m} \sum_{i=1}^n \left[ \frac{m}{n} \right] f(X_i) = \frac{1}{n} \sum_{i=1}^n f(X_i) \end{aligned}$$

*Q.E.D.*

**Lemma 2:** Under Condition 1,

$$E_M \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right)^2 \right] = \frac{(m+n-1)}{m n^2} \sum_{i=1}^n (f(X_i))^2 + \frac{2(m-1)}{m n^2} \sum_{i < k} \binom{n^2-n}{2} f(X_i) f(X_k) .$$

Proof of **Lemma 2:**

Substitute from (2.1) and note that the multinomial distribution with equal probabilities has

$$E_M(M_i) = \frac{m}{n}, \quad E_M(M_i^2) = \frac{m(m-1) + nm}{n^2} \quad \text{and} \quad E_M(M_i M_k) = \frac{m^2 - m}{n^2} \quad \text{for } i \neq k \text{ to obtain}$$

$$\begin{aligned} E_M \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right)^2 \right] &= \sum_{M_1 + \dots + M_n = m} \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right)^2 \frac{n^{-m} m!}{(M_1! \dots M_n!)} \\ &= \sum_{M_1 + \dots + M_n = m} \left( \frac{1}{m} \sum_{i=1}^n M_i f(X_i) \right)^2 \frac{n^{-m} m!}{(M_1! \dots M_n!)} \\ &= \sum_{M_1 + \dots + M_n = m} \left( \frac{1}{m^2} \sum_{i=1}^n M_i^2 (f(X_i))^2 + \frac{2}{m^2} \sum_{i < k} \binom{n^2-n}{2} M_i M_k f(X_i) f(X_k) \right) \frac{n^{-m} m!}{(M_1! \dots M_n!)} \\ &= \frac{1}{m^2} \sum_{i=1}^n \left[ \sum_{M_1 + \dots + M_n = m} M_i^2 \frac{n^{-m} m!}{(M_1! \dots M_n!)} \right] (f(X_i))^2 + \frac{2}{m^2} \sum_{i < k} \binom{n^2-n}{2} \left[ \sum_{M_1 + \dots + M_n = m} M_i M_k \frac{n^{-m} m!}{(M_1! \dots M_n!)} \right] f(X_i) f(X_k) \\ &= \frac{1}{m^2} \sum_{i=1}^n [E_M(M_i^2)] (f(X_i))^2 + \frac{2}{m^2} \sum_{i < k} \binom{n^2-n}{2} [E_M(M_i M_k)] f(X_i) f(X_k) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{m^2} \sum_{i=1}^n \left[ \frac{m(m-1) + n m}{n^2} \right] (f(X_i))^2 + \frac{2}{m^2} \sum_{i < k}^{\binom{n^2-n}{2}} \left[ \frac{m^2 - m}{n^2} \right] f(X_i) f(X_k) \\
&= \frac{(m+n-1)}{m n^2} \sum_{i=1}^n (f(X_i))^2 + \frac{2(m-1)}{m n^2} \sum_{i < k}^{\binom{n^2-n}{2}} f(X_i) f(X_k) . \qquad \qquad \qquad Q.E.D.
\end{aligned}$$

**Theorem 1:** Under Condition 1,

$$\text{Var}_M \left[ \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right] = \frac{n-1}{m n^2} \sum_{i=1}^n (f(X_i))^2 - \frac{2}{m n^2} \sum_{i < k}^{\binom{n^2-n}{2}} f(X_i) f(X_k) .$$

See the Appendix for the proof of **Theorem 1**:

For the standard bootstrap this completes our analysis under the bootstrap-induced distribution when the original sample is treated as the population. Next we will look at the broader problem of carrying out finite sample analysis when also taking into consideration the inherent randomness implied by the original sample as drawn from some unknown overall population. In other words, we will now look at the joint distribution between the bootstrap-induced randomness and the randomness of the underlying random variable. To keep our results as general as we can, we will do this in terms of the unknown moments of the underlying random variable using the expectation notation,  $E_X$ . This will reveal the minimum information about the underlying distribution that would be needed to precisely determine the variance under this joint distribution. A corresponding analysis could be carried out using this approach to reveal the characteristics of any of the other moments of the finite sample distribution.

**Theorem 2:** Under Condition 1,

$$\begin{aligned} \text{Var}_{M,X} \left[ \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right] &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}_X [f(X_i)] + \frac{n-1}{m n^2} \sum_{i=1}^n (E_X [(f(X_i))^2]) \\ &\quad + \frac{2}{n^2} \sum_{i < k}^{\binom{n^2-n}{2}} \text{Cov}_X [f(X_i), f(X_k)] - \frac{2}{m n^2} \sum_{i < k}^{\binom{n^2-n}{2}} E_X [f(X_i) f(X_k)]. \end{aligned}$$

See the Appendix for the proof of **Theorem 2**.

Obviously the “extra” terms in **Theorem 2** could be eliminated by drawing an infinite number of times with replacement in creating each bootstrap sample. Alternative methods of eliminating these terms include using the wild bootstrap, which will be discussed in the next section.

An example of using these results might be an analysis of the function

$f(X_j^*) = (X_j^* - \bar{X})^2$  where  $\bar{X}$  is the mean of the original sample data. This function, when averaged over the  $m$  bootstrap draws in (2.1), is the bootstrap estimator of the variance under the bootstrap’s empirical distribution function (EDF).

By **Lemma 1** we have

$$E_M \left[ \frac{1}{m} \sum_{j=1}^m (X_j^* - \bar{X})^2 \right] = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

which is just the usual maximum likelihood estimator of the variance.

By **Theorem 1** we have

$$\text{Var}_M \left[ \frac{1}{m} \sum_{j=1}^m (X_j^* - \bar{X})^2 \right] = \frac{n-1}{m n^2} \sum_{i=1}^n ((X_i - \bar{X})^2)^2 - \frac{2}{m n^2} \sum_{i < k}^{\binom{n^2-n}{2}} (X_i - \bar{X})^2 (X_k - \bar{X})^2,$$

which tells us the formula for the bootstrap estimator of the variance under the EDF.

By **Theorem 2** we have

$$\begin{aligned} \text{Var}_{M,X} \left[ \frac{1}{m} \sum_{j=1}^m (X_j^* - \bar{X})^2 \right] &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}_X \left[ (X_i - \bar{X})^2 \right] + \frac{n-1}{m n^2} \sum_{i=1}^n \left( E_X \left[ \left( (X_i - \bar{X})^2 \right)^2 \right] \right) \\ &+ \frac{2}{n^2} \sum_{i < k}^{\binom{n^2-n}{2}} \text{Cov}_X \left[ (X_i - \bar{X})^2, (X_k - \bar{X})^2 \right] - \frac{2}{m n^2} \sum_{i < k}^{\binom{n^2-n}{2}} E_X \left[ (X_i - \bar{X})^2 (X_k - \bar{X})^2 \right] \end{aligned}$$

which tells us what object the formula in **Theorem 1** is actually delivering under the joint distribution.

### 3. THE ANALYTICAL SOLUTION FOR THE VARIANCE OF THE WILD BOOTSTRAP

One immediate difficulty with the standard bootstrap is that it fails to enforce even the most elementary requirements, such as mean-zero errors. To rectify this shortcoming, Wu (1986) proposed and Liu (1988) formulated a weighted bootstrap known as the wild bootstrap.

The wild bootstrap draws randomly from the least squares residuals with equal probability to produce the standard multinomial distribution for the frequencies, but, in addition, multiplies the chosen residual by a special factor. In the simplest case as proposed by Davidson and Flachaire (2001), the factor, as determined from random draws from the Rademacher distribution, is plus one with a probability of one-half and minus one with a probability of one-half. This was earlier referred to by Liu (1988) as a special lattice distribution called the two points distribution. A more widely used version proposed by Mammen (1993) specifies the factor as  $-(\sqrt{5}-1)/2$  with probability  $p = (\sqrt{5}+1)/(2\sqrt{5})$  and  $(\sqrt{5}+1)/2$  with probability  $(1-p) = (\sqrt{5}-1)/(2\sqrt{5})$ . We will analyze the Rademacher version, but a similar analysis of the Mammen version follows easily by analogy.

To simplify and expedite the exposition throughout this section we provide **Condition 2** as a set of procedural assumptions and definitions that characterize the Liu-Davidson-Flachaire wild bootstrap.

**Condition 2:** *Draw a sample of size  $m$  with replacement from a set of  $n$  real numbers  $\{X_i: i = 1, \dots, n\}$  with equal probability, and, by doing so, produce a corresponding set of random frequencies  $\{M_i: i = 1, \dots, n\}$  that follow the multinomial distribution. Next multiply each value in the bootstrap sample by a value of plus one with a probability of one-half and a value of minus one with a probability of one-half. These values then constitute the set of bootstrap sample values  $\{X_j^*: j = 1, \dots, m\}$ . Denote the set of any function,  $f$ , of each of the original sample values as  $\{f(X_i): i = 1, \dots, n\}$ . Denote the set of that same function of each of the individual bootstrapped values as  $\{f(X_j^*): j = 1, \dots, m\}$ . Any expected value under the bootstrap-induced distribution will be denoted by the subscript  $M$ , as in  $E_M[\cdot]$  while expectations under the randomness that generated the original sample will be denoted by the subscript  $X$ , as in  $E_X[\cdot]$  and variance  $Var_X[\cdot]$ . Any expected value under the joint distribution of the multinomial randomness and the binomial randomness will be denoted by the subscripts  $M$  and  $W$  (for wild),*

as in  $E_{W,M}[\cdot]$  with variance  $Var_{W,M}[\cdot]$  and conditional expectation  $E_{W|M}[\cdot]$ . Any expected value under the joint distribution of the bootstrap-induced randomness and the randomness that generated the original sample will be denoted by the subscripts  $W$ ,  $M$  and  $X$ , as in  $E_{W,M,X}[\cdot]$  and variance  $Var_{W,M,X}[\cdot]$  and with a corresponding conditional expectation  $E_{X|M,W}[\cdot]$ .

We now introduce **Lemma 3** and its proof to demonstrate the analytical procedure for setting up the problem for analysis.

**Lemma 3:** Under Condition 2,  $E_{W,M}\left(\frac{1}{m}\sum_{j=1}^m f(X_j^*)\right) = 0$ .

Proof of **Lemma 3**:

Here there is a joint bootstrap-induced distribution that can be expressed as a multinomial distribution multiplied times a binomial distribution conditional on that multinomial.

Let  $W_i$  be the number of positive one values randomly drawn for each multinomial frequency,  $M_i$ . Conditional on  $M_i$  then  $W_i$  is a random variable with a binomial distribution having a probability of 0.5, a mean of  $0.5 M_i$  and a variance of  $0.25 M_i$ . Since  $W_i$  is the number of positive ones assigned to the bootstrap draws, then  $(M_i - W_i)$  is the number of zeros (which generate negative ones in this setup). Consequently, we have the following expectation under Condition 2.

$$E_{W,M}\left(\frac{1}{m}\sum_{j=1}^m f(X_j^*)\right) = E_M E_{W|M}\left(\frac{1}{m}\sum_{j=1}^m f(X_j^*)\right) = E_M E_{W|M}\left(\frac{1}{m}\sum_{i=1}^n [W_i - (M_i - W_i)]f(X_i)\right).$$

$$\begin{aligned}
E_{W,M} \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) &= E_M E_{W|M} \left( \frac{1}{m} \sum_{i=1}^n [2W_i - M_i] f(X_i) \right) \\
&= E_M \left( \frac{1}{m} \sum_{i=1}^n [2E_{W|M} W_i - M_i] f(X_i) \right) \\
&= E_M \left( \frac{1}{m} \sum_{i=1}^n [M_i - M_i] f(X_i) \right) = 0
\end{aligned}$$

*Q.E.D.*

**Theorem 3:** Under Condition 2,  $Var_{W,M} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] = \frac{1}{mn} \sum_{i=1}^n [(f(X_i))^2]$ .

See the Appendix for proof of **Theorem 3**.

**Theorem 4:** Under Condition 2,  $Var_{W,M,X} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] = \frac{1}{mn} \sum_{i=1}^n E_X [(f(X_i))^2]$ .

See the Appendix for proof of **Theorem 4**.

Note that if  $\frac{1}{m n} \sum_{i=1}^n E_X [f(X_i)] = 0$ , then  $Var_{W,M,X} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] = \frac{1}{mn} \sum_{i=1}^n Var_X [f(X_i)]$ .

Analogously, under the Mammen (1993) wild bootstrap, the probability of  $W_i$  successes is

$$(3.1) \quad P(W_i | M_i) = \binom{M_i}{W_i} \left( \frac{\sqrt{5}-1}{2\sqrt{5}} \right)^{W_i} \left( \frac{\sqrt{5}+1}{2\sqrt{5}} \right)^{M_i - W_i}$$

and the needed expected values are

$$(3.2) \quad EW_i = \left( \frac{\sqrt{5}-1}{2\sqrt{5}} \right) M_i, \quad EW_i^2 = \frac{2}{\sqrt{5}} M_i + \left( \frac{3-\sqrt{5}}{10} \right) M_i^2 \quad \text{and} \quad EW_i W_k = \left( \frac{3-\sqrt{5}}{10} \right) M_i M_k$$

and the variance of  $W_i$  is easily shown to be  $VarW_i = \frac{2}{\sqrt{5}} M_i$ .

Consequently, by analogy with **Theorem 3**, for the Mammen wild bootstrap we get

$$Var_{w,M} \left[ \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right] = \frac{10}{mn\sqrt{5}} \sum_{i=1}^n [(f(X_i))^2]$$

and by analogy with **Theorem 4** we get

$$Var_{w,M,X} \left[ \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right] = \frac{10}{mn\sqrt{5}} \sum_{i=1}^n E_X [(f(X_i))^2].$$

Note that if  $\frac{1}{m} \sum_{i=1}^n E_X [f(X_i)] = 0$ , then  $Var_{w,M,X} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] = \frac{10}{mn\sqrt{5}} \sum_{i=1}^n Var_X [f(X_i)]$ .

In the case of regression analysis under the i.i.d. assumption the population errors have zero means while the sample residuals average to zero by construction (as long as there is an intercept term). Since the Mammen (1993) wild bootstrap and the Liu (1988) and Davidson and Flachaire (2001) wild bootstrap both have expected values of zero, and the former has a larger variance than the latter, then it implies that, in the context of regression analysis with zero-mean errors, the population finite sample mean squared error of the Mammen wild bootstrap will be larger than the mean squared error of the Liu-Davidson-Flachaire wild bootstrap.

#### 4. ANALYTIC SOLUTION FOR THE VARIANCE OF FIXED BLOCK BOOTSTRAPS

Initial attempts to use bootstrap methods to capture consecutive time sequences involved randomly selecting non-overlapping fixed blocks of time series data. See Hall (1985) and Carlstein (1986) for the presentation, discussion and analysis of these methods. Alternatively, Hall (1985), Künsch (1989), and Politis and Romano (1993) have presented analyses of the overlapping block bootstrap. A very helpful overview and useful discussion of the research on block bootstrapping can be found in Härdle, Horowitz and Kreiss (2003).

Let  $\mathbf{Y}_t$  refer to a matrix of endogenous variables and  $\mathbf{X}_t$  refer to a matrix of predetermined variables, which are all observed at time  $t$ . For example, for panel/longitudinal data these  $\mathbf{Y}_t$  and  $\mathbf{X}_t$  matrices could represent the data for a simultaneous equations model of cross-sectional units all observed at time  $t$ , where the  $\mathbf{X}_t$  matrix could include lagged values of some or all of the variables in the  $\mathbf{Y}_t$  matrix. Using a total of  $T$  consecutive observations from this set of time-specific matrices  $\{\mathbf{Y}_t, \mathbf{X}_t : t=1, \dots, T\}$ , form  $k$  matrix blocks,  $\{\mathbf{B}_g : g=1, \dots, k\}$ , each with  $l$  consecutive time-specific matrix observations. The size of these blocks,  $l$ , is a key issue. Hall, Horowitz and Jing (1995) have shown that an optimal block size for the non-overlapping block bootstrap for variance estimation is  $l = T^{1/3}$ , which would imply  $k = T^{2/3}$ , and for the overlapping block bootstrap,  $k = T - T^{1/3} + 1$ .

To simplify and expedite the exposition in this section we provide **Condition 3** as a set of procedural assumptions and definitions that characterize the block bootstrap.

**Condition 3:** Form  $k$  matrix blocks,  $\{\mathbf{B}_g : g=1, \dots, k\}$ , each with  $l$  consecutive time-specific matrix observations from a total of  $T$  consecutive time periods. For a non-overlapping block bootstrap choose  $l$  so that  $kl = T$  and order the blocks such that the first block,  $\mathbf{B}_1$ , covers observations

from 1 through  $l$ , the second block,  $\mathbf{B}_2$ , covers observations from  $l+1$  through  $2l$ , and so forth until the last block,  $\mathbf{B}_k$ , which covers observations from  $T-l+1$  through  $T$ . For an overlapping block bootstrap the first block,  $\mathbf{B}_1$ , covers observations from 1 through  $l$ , the second block,  $\mathbf{B}_2$ , covers observations from 2 through  $l+1$ , and so forth until the last block,  $\mathbf{B}_k$ , which covers observations from  $T-l+1$  through  $T$ . The overlapping block bootstrap has  $k = T-l+1$  blocks available for sampling. Next randomly select with replacement  $b$  blocks from the  $k$  blocks,  $\{\mathbf{B}_g: g=1, \dots, k\}$ , with corresponding random frequencies  $\{M_g: g=1, \dots, k\}$  to obtain the bootstrap resampled blocks:  $\{\mathbf{B}_j^*: j=1, \dots, b\}$ . Note that the integer  $b$  may be smaller than, equal to, or greater than the integer  $k$ . Finally, average some known function,  $\theta$ , of each of these randomly selected blocks.

$$(4.1) \quad \frac{1}{b} \sum_{j=1}^b \theta(\mathbf{B}_j^*) = \frac{1}{b} \sum_{g=1}^k M_g \theta(\mathbf{B}_g).$$

The probability of obtaining a particular realization of the set  $\{M_g: g=1, \dots, k\}$  is given by the multinomial distribution as:

$$(4.2) \quad p(M_1 = m_1, \dots, M_k = m_k) = \frac{k^{-b} b!}{m_1! \dots m_k!} \quad \text{where} \quad b = \sum_{i=1}^k m_i.$$

The expected values of the  $g^{\text{th}}$  block for  $g = 1, \dots, k$  for the first and second moments around zero and the second moment around the mean (the variance) are given by:

$$(4.3) \quad E_M(M_g) = \frac{b}{k}, \quad E_M(M_g^2) = \frac{b(b-1) + kb}{k^2} \quad \text{and} \quad \text{Var}_M(M_g) = \frac{b(k-1)}{k^2}.$$

The expected cross products and covariances between frequencies  $M_g$  for  $g = 1, \dots, k$  and  $M_h$  for  $h = 1, \dots, k$  are given by:

$$(4.4) \quad E_M(M_g M_h) = \frac{b(b-1)}{k^2} \quad \text{and} \quad \text{Cov}_M(M_g, M_h) = \frac{-b}{k^2}.$$

Any expected value under the bootstrap-induced distribution will be denoted by the subscript  $M$ , as in  $E_M[\cdot]$ . Any expected value under the joint distribution of the bootstrap-induced randomness and the randomness that generated the data in the blocks will be denoted by the subscripts  $M$  and  $B$ , as in  $E_{M,B}[\cdot]$ . Variances and covariances will be denoted with corresponding subscripts such as  $\text{Var}_M[\cdot]$ ,  $\text{Var}_{M,B}[\cdot]$ ,  $\text{Var}_B[\cdot]$ ,  $\text{Cov}_M[\cdot, \cdot]$  and  $\text{Cov}_B[\cdot, \cdot]$ .

**Theorem 5:** Under *Condition 3*,

$$\text{Var}_M \left[ \frac{1}{b} \sum_{j=1}^b \theta(B_j^*) \right] = \frac{k-1}{b k^2} \sum_{g=1}^k [(\theta(B_g))^2] - \frac{2}{b k^2} \sum_{g < h}^{\binom{k^2-k}{2}} \theta(B_g) \theta(B_h) .$$

See the Appendix for the proof of **Theorem 5**.

Note that a special case of **Theorem 5** could be  $\theta(B_g) = \frac{1}{l} \sum_{i=1}^l f(X_{gi})$  which just produces the grand mean over all the blocks such that

$$\text{Var}_M \left[ \frac{1}{b} \sum_{j=1}^b \theta(B_j^*) \right] = E_M \left\{ \left[ \frac{1}{b l} \sum_{g=1}^k \sum_{i=1}^l M_g f(X_{gi}) \right]^2 \right\} - \left\{ E_M \left[ \frac{1}{b l} \sum_{g=1}^k \sum_{i=1}^l M_g f(X_{gi}) \right] \right\}^2 .$$

For example, this could then be used to evaluate the small sample properties of any bootstrap sample moment estimator or some function of residuals from a regression.

**Theorem 6:** Under *Condition 3*,

$$\begin{aligned} \text{Var}_{M,B} \left[ \frac{1}{b} \sum_{j=1}^b \theta(B_j^*) \right] &= \frac{1}{k^2} \sum_{g=1}^k \text{Var}_B [\theta(B_g)] + \frac{k-1}{b k^2} \sum_{g=1}^k E_B [(\theta(B_g))^2] . \\ &+ \frac{2}{k^2} \sum_{g < h}^{\binom{k^2-k}{2}} \text{Cov}_B [\theta(B_g), \theta(B_h)] - \frac{2}{b k^2} \sum_{g < h}^{\binom{k^2-k}{2}} E_B [\theta(B_g) \theta(B_h)] . \end{aligned}$$

See the Appendix for the proof of **Theorem 6**.

Note that as  $b$  goes to infinity, once again the “extra” terms disappear.

## 5. ANALYTICAL APPROACH TO NONLINEAR BOOTSTRAP FUNCTIONS

Since polynomials in the elements of the set  $\{M_i : i = 1, \dots, n\}$ , are directly analyzable functions, then nonlinear bootstrap functions become directly analyzable when expressed as Taylor series expansions.

For example, given a random sample  $\{X_i : i = 1, \dots, n\}$  of a random  $p \times 1$  vector  $X$  with a sample vector of means  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ , we define  $\bar{X}^* = \frac{1}{m} \sum_{j=1}^m X_j^*$  as a vector of bootstrap sample means of a bootstrap sample  $\{X_j^* : j = 1, \dots, m\}$  whose  $m$  values were drawn randomly with replacement from the original sample  $\{X_i : i = 1, \dots, n\}$ . Horowitz (2001) approximates the bias of  $\theta_n = g(\bar{X})$  as an estimator of  $\theta_o = g(\mu)$  where  $\mu = E(X)$  for a smooth nonlinear function  $g$  as

$$(5.1) \quad B_n^* = \frac{1}{2} E_M [(\bar{X}^* - \bar{X})' G_2(\bar{X})(\bar{X}^* - \bar{X})] + O(n^{-2})$$

almost surely, where  $G_2(\bar{X})$  is the matrix of second partial derivatives of  $g$ .

The primary part of this bias can be evaluated analytically as follows.

$$E_M [(\bar{X}^* - \bar{X})' G_2(\bar{X})(\bar{X}^* - \bar{X})] = E_M [\bar{X}^{*'} G_2(\bar{X}) \bar{X}^*] - E_M [\bar{X}^{*'} G_2(\bar{X}) \bar{X}] - E_M [\bar{X}' G_2(\bar{X}) \bar{X}^*] + E_M [\bar{X}' G_2(\bar{X}) \bar{X}]$$

where the third term is equal to

$$E_M [\bar{X}' G_2(\bar{X}) \bar{X}^*] = E_M \left[ \bar{X}' G_2(\bar{X}) \frac{1}{m} \sum_{j=1}^m X_j^* \right] = E_M \left[ \bar{X}' G_2(\bar{X}) \frac{1}{m} \sum_{i=1}^n M_i X_i \right] \\ = \bar{X}' G_2(\bar{X}) \frac{1}{m} \sum_{i=1}^n E_M [M_i] X_i = \bar{X}' G_2(\bar{X}) \frac{1}{m} \sum_{i=1}^n \frac{m}{n} X_i = \bar{X}' G_2(\bar{X}) \bar{X} .$$

By analogy,  $E_M [\bar{X}^{*'} G_2(\bar{X}) \bar{X}] = \bar{X}' G_2(\bar{X}) \bar{X}$  as well.

Consequently, the primary bias term simplifies to

$$E_M [(\bar{X}^* - \bar{X})' G_2(\bar{X})(\bar{X}^* - \bar{X})] = E_M [\bar{X}^{*'} G_2(\bar{X}) \bar{X}^*] - \bar{X}' G_2(\bar{X}) \bar{X} .$$

Now evaluate the first term on the right-hand side of this expression as

$$\begin{aligned}
E_M[\bar{X}^*{}'G_2(\bar{X})\bar{X}^*] &= \frac{1}{m^2}E_M\left\{\left[\sum_{i=1}^n X_i M_i'\right]G_2(\bar{X})\left[\sum_{i=1}^n M_i X_i\right]\right\} \\
&= \frac{1}{m^2}E_M\left\{\sum_{i=1}^n X_i M_i' G_2(\bar{X}) M_i X_i + 2 \sum_{i < k}^{\binom{n^2-n}{2}} X_i M_i' G_2(\bar{X}) M_k X_k\right\} \\
&= \frac{1}{m^2}\left\{\sum_{i=1}^n X_i' E_M\left[M_i' G_2(\bar{X}) M_i\right] X_i + 2 \sum_{i < k}^{\binom{n^2-n}{2}} X_i' E_M\left[M_i' G_2(\bar{X}) M_k\right] X_k\right\}
\end{aligned}$$

where, when  $r_s$  is a  $p \times 1$  column vector of all zeros except for a one in the  $s^{\text{th}}$  row and  $c_t$  is a  $1 \times p$  row vector of all zeros except for a one in its  $t^{\text{th}}$  column ( $r_g, c_g$  and  $r_h, c_h$  by analogy), we have

$$\begin{aligned}
E_M\left[M_i' G_2(\bar{X}) M_k\right] &= E_M\left\{\left[\sum_{g=1}^p (c_g' r_g') \otimes (M_{ig})\right] \left[\sum_{s=1}^p \sum_{t=1}^p (r_s c_t) \otimes (G_2(\bar{X})_{st})\right] \left[\sum_{h=1}^p (r_h c_h) \otimes (M_{kh})\right]\right\} \\
&= E_M\left\{\left[\sum_{s=1}^p \sum_{t=1}^p (r_s c_t) \otimes (M_{is} M_{kt} G_2(\bar{X})_{st})\right]\right\} \\
&= \sum_{s=1}^p \sum_{t=1}^p (r_s c_t) \otimes \left(\frac{m(m-1)}{n^2} G_2(\bar{X})_{st}\right) \\
&= \frac{m(m-1)}{n^2} G_2(\bar{X}).
\end{aligned}$$

The analysis of  $E_M\left[M_i' G_2(\bar{X}) M_i\right]$  is similar except that  $M_{is} = M_{it}$  for all  $s$  and  $t$  so we have

$$\begin{aligned}
E_M\left[M_i' G_2(\bar{X}) M_i\right] &= \sum_{s=1}^p \sum_{t=1}^p (r_s c_t) \otimes (E_M[M_{is}^2] G_2(\bar{X})_{st}) \\
&= \sum_{s=1}^p \sum_{t=1}^p (r_s c_t) \otimes \left(\frac{m(n+m-1)}{n^2} G_2(\bar{X})_{st}\right) \\
&= \frac{m(n+m-1)}{n^2} G_2(\bar{X}).
\end{aligned}$$

Substituting these expressions into

$$E_M[\bar{X}^*{}'G_2(\bar{X})\bar{X}^*] = \frac{1}{m^2}\left\{\sum_{i=1}^n X_i' E_M\left[M_i' G_2(\bar{X}) M_i\right] X_i + 2 \sum_{i < k}^{\binom{n^2-n}{2}} X_i' E_M\left[M_i' G_2(\bar{X}) M_k\right] X_k\right\}$$

yields the following expression.

$$E_M[\bar{X}^*{}' G_2(\bar{X}) \bar{X}^*] = \frac{(n+m-1)}{mn^2} \sum_{i=1}^n X_i' G_2(\bar{X}) X_i + \frac{2(m-1)}{mn^2} \sum_{i < k}^{\binom{n^2-n}{2}} X_i' G_2(\bar{X}) X_k .$$

Finally, substituting this all back into (5.1) yields an analytical expression for the bias.

$$B_n^* = \frac{1}{2} \left[ \frac{(n+m-1)}{mn^2} \sum_{i=1}^n X_i' G_2(\bar{X}) X_i + \frac{2(m-1)}{mn^2} \sum_{i < k}^{\binom{n^2-n}{2}} X_i' G_2(\bar{X}) X_k - \bar{X}' G_2(\bar{X}) \bar{X} \right] + O(n^{-2}).$$

In general, the elements of any  $m \times m$  bootstrapped Hessian matrix  $G_{m2}$  are, in effect, randomly selected from the elements of the  $n \times n$  Hessian matrix  $G_{n2}$ , which is based on the original set of sample values  $\{X_i: i = 1, \dots, n\}$ . If the  $m \times 1$  vector of bootstrapped values  $X^* = (X_1^*, \dots, X_m^*)'$  has a corresponding set of unspecified initial values  $X_o^* = (X_{1o}^*, \dots, X_{mo}^*)'$  and the  $n \times 1$  vector of original sample values  $X = (X_1, \dots, X_n)'$  has a corresponding  $n \times n$  diagonal frequency matrix  $M = \text{diag}(M_1, \dots, M_n)$ , then the random bootstrap selection process results in

$$(5.2) \quad (X^* - X_o^*)' G_{m2} (X^* - X_o^*) = \left( MX - \frac{m}{n} X \right)' G_{n2} \left( MX - \frac{m}{n} X \right)$$

where the expected values under the bootstrap-induced distribution are  $E_M(M_i) = \frac{m}{n}$ ,

$$E_M(M_i^2) = \frac{m(m-1) + nm}{n^2} \quad \text{and} \quad E_M(M_i M_k) = \frac{m^2 - m}{n^2} \quad \text{for } i \neq k. \quad \text{Under this bootstrap-induced}$$

random selection process, the initial values for the Taylor series,  $X_o^* = (X_{1o}^*, \dots, X_{mo}^*)'$ , are not uniquely identified *a priori*, since, in effect, the Taylor series is being initialized around the original set of sample values<sup>4</sup>,  $\{X_i: i = 1, \dots, n\}$ , which does not have a one-to-one correspondence with the set of bootstrapped values  $\{X_j^*: j = 1, \dots, m\}$ . Note that (5.2) extends the (2.1) approach.

---

<sup>4</sup> actually around each of their mean-frequency weighted values,  $(m/n)X_i$  for  $i=1, \dots, n$ , to be more precise.

## 6. ANALYTICAL SOLUTION FOR BOOTSTRAP REGRESSION COEFFICIENT COVARIANCE MATRIX

Begin with the traditional regression model with an  $n \times 1$  dependent variable vector,  $Y$ , an  $n \times k$  regressor matrix,  $X$ , a  $k \times 1$  regression coefficient vector,  $\beta$ , and an  $n \times 1$  error term,  $\epsilon$ , in standard linear form  $Y = X\beta + \epsilon$ .

There are two alternative approaches to bootstrapping the estimator,  $\hat{\beta}$ , of the regression coefficient vector. One approach treats the  $X$  matrix as nonstochastic so that the standard error of  $\hat{\beta}$  can be derived from estimates of the conditional variance of the dependent variable,  $Y$ . Another approach is to view the  $X$  matrix as random with the random matched pairs of  $(X_t, Y_t)$  values bootstrapped together as  $(X_t^*, Y_t^*)$ . The first approach essentially involves bootstrapping the regression residuals or an equivalent procedure. The second approach is more challenging in that it complicates the problem by producing randomness in the bootstrapped  $X'X$  inverse matrix as well as in the  $X'Y$  vector. We present the nonstochastic- $X$  case here. We present the stochastic- $X$  case (the paired bootstrap) in a separate paper since it requires a much more extensive analysis.

When appropriate, the analytical bootstrap solutions presented in the preceding sections can be applied to the bootstrapping of residuals from a regression to calculate the standard errors of the regression coefficient estimators. Some of those methods are more useful for models with independent and identically distributed (i.i.d.) random errors while others (e.g. block bootstrap) are more appropriate for dependent error patterns. This subsection will provide a simple matrix version of bootstrapping which can be altered to represent many different variations of bootstrapping methods.

Bootstrapping residuals can be interpreted as inducing a bootstrap distribution on the dependent variable,  $Y$ , while maintaining the nonstochastic nature of the regressor matrix,  $X$ . The least squares estimator,  $\hat{\beta} = (X'X)^{-1}X'Y$ , will generate a corresponding set of least squares residuals,  $e = Y - X\hat{\beta}$ . Freedman (1981) has pointed out that bootstrap procedures can fail unless the residuals are centered at zero. A constant term in the regression will automatically center the residuals at zero, but when no constant term is present, the residuals will need to be centered as part of the bootstrap procedure.

Since  $e = (I_n - X(X'X)^{-1}X')e$  and  $X$  is considered to be a nonrandom matrix, to provide a set of bootstrapped residuals  $e^* = \{e_1^*, e_2^*, \dots, e_n^*\}'$  that are centered around zero we introduce the  $n \times n$  matrix  $A = (I_n - (1/n)1_n1_n')$  and replace  $e^*$  with  $Ae^*$  to ensure that the bootstrap residuals are centered at zero. Alternatively, one could redefine  $e$  to be  $Ae$  to center the original set of least squares residuals. By inserting  $A$  in the equations we can simultaneously derive the results for the uncentered case,  $A = I_n$ , and the centered case,  $A = (I_n - (1/n)1_n1_n')$ .

To simplify and expedite the exposition in this section we provide **Condition 4** as a set of procedural assumptions and definitions that characterize this least squares regression bootstrap.

**Condition 4:**

*Given an  $n \times 1$  random dependent variable vector  $Y$  and an  $n \times k$  nonrandom regressor matrix  $X$ , take  $n$  draws with replacement from the set of least squares residuals  $e = \{e_1, e_2, \dots, e_n\}'$  corresponding to the least squares estimator,  $\hat{\beta} = (X'X)^{-1}X'Y$ , where  $e = Y - X\hat{\beta}$ , to obtain a corresponding set of bootstrapped residuals  $e^* = \{e_1^*, e_2^*, \dots, e_n^*\}'$ . For  $Y^* = X\hat{\beta} + Ae^*$  define the bootstrap regression estimator as  $\hat{\beta}^* = (X'X)^{-1}X'Y^*$ . Define an  $n \times n$  matrix,  $H$ , with elements,  $\{H_{ij} : i, j=1, \dots, n\}$ , such that  $e^* = He$ , where the element in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of  $H$  is equal to one if the  $j^{\text{th}}$  element of  $e$  is randomly drawn (with replacement) on the  $i^{\text{th}}$  draw and is*

equal to zero otherwise. For notation,  $I_n$  is an  $n \times n$  identity matrix,  $1_n$  is an  $n \times 1$  column vector of all ones,  $1_{n^2}$  is an  $n^2 \times 1$  column vector of all ones, “vec” is the standard vectorization operator, “ $\otimes$ ” means Kronecker product, and the  $n \times n$  matrix  $A$  is  $A = I_n$ , the identity matrix, if the residuals are not centered and is  $A = (I_n - (1/n)1_n1_n')$ , if the residuals are centered, in which case  $A1_n1_n' = 0 = 1_n1_n'A$  and the last term drops out. Any expected value under the randomness that generated the original sample,  $\{e_i : i = 1, \dots, n\}$ , will be denoted by the subscript  $\epsilon$ , as in  $E_\epsilon[\cdot]$  and covariance,  $Cov_\epsilon[\cdot, \cdot]$ . Any expected value under the joint distribution of the randomness that generated the original sample and the bootstrap-induced randomness will be denoted by the subscripts  $\epsilon$  and  $H$ , as in  $E_{\epsilon, H}[\cdot]$  and covariance,  $Cov_{\epsilon, H}[\cdot, \cdot]$ . Conditional expectations will be denoted  $E_{H|\epsilon}[\cdot]$  and covariance,  $Cov_{H|\epsilon}[\cdot, \cdot]$ .

Under **Condition 4**, since each of the  $n$  draws will take place for sure, the relevant distribution is the conditional distribution of drawing  $j$  on the  $i^{\text{th}}$  draw. Analytically this yields the expectations  $E_{j|i}(H_{ij}) = 1/n$  and  $E_{j|i}(H_{ij}^2) = 1/n$ , and, therefore, the variance is  $\text{Var}_{j|i}(H_{ij}) = (n-1)/n^2$ . Given an  $n \times 1$  column vector  $r_i$  whose elements are all zeros except for a one in its  $i^{\text{th}}$  “row” position, and a  $1 \times n$  row vector  $c_j$  also made up of all zeros except for a one in its  $j^{\text{th}}$  “column” position, we have 
$$H = \sum_{i=1}^n \sum_{j=1}^n r_i c_j H_{ij}, \quad E_{H|\epsilon}(H) = \sum_{i=1}^n \sum_{j=1}^n r_i c_j E_{j|i, \epsilon}(H_{ij}) \quad \text{and}$$
 
$$E_{j|i, \epsilon}(H_{ij}) = \frac{1}{n} \quad \text{so that} \quad E_{H|\epsilon}(H) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n r_i c_j = \frac{1}{n} 1_n 1_n'$$
 where  $1_n$  is an  $n \times 1$  column vector of all ones. Since  $e^* = He$ , it follows that  $E_{H|\epsilon}(e^*) = \frac{1}{n} 1_n 1_n' e$ . Since  $A = (I_n - (1/n)1_n1_n')$ , we have  $A1_n1_n' = 0$  and, therefore,  $E_{H|\epsilon}(Ae^*) = 0$ . Also, we have  $1_n1_n'A = 0$  and, therefore,  $E_{H|\epsilon}(HAe) = 0$  so centering the original residuals is just as effective as centering the bootstrapped residuals.

**Theorem 7:**

Under Condition 4,

$$\begin{aligned} Cov_{H|\epsilon}(\hat{\beta}^*) &= (X'X)^{-1} X' A \left\{ \frac{1}{n} [I_n \otimes (e'e)] + \frac{1}{n^2} \left[ (1_n 1_n' - I_n) \otimes (1_n' vec(ee')) \right] \right\} A X (X'X)^{-1} \\ &- \frac{1}{n^2} (X'X)^{-1} X' A 1_n 1_n' ee' 1_n 1_n' A X (X'X)^{-1}. \end{aligned}$$

See the Appendix for the proof of **Theorem 7**. Note that this theorem allows for the situation where the residuals do not necessarily sum to zero. However, the last term will be zero if the residuals sum to zero. This specifically addresses the problem of uncentered residuals raised by Friedman (1981) by allowing for this possibility and revealing its effect in finite samples on the covariance matrix of the bootstrap estimator.

**Theorem 8:**

Under Condition 4,

$$\begin{aligned} Cov_{\epsilon,H}(\hat{\beta}^*) &= Cov_{\epsilon}(\hat{\beta}) + (X'X)^{-1} X' A Cov_{\epsilon}(\epsilon) (I_n - X(X'X)^{-1} X') \left( \frac{1}{n} 1_n 1_n' \right) A X (X'X)^{-1} \\ &+ (X'X)^{-1} X' A \left( \frac{1}{n} 1_n 1_n' \right) (I_n - X(X'X)^{-1} X') Cov_{\epsilon}(\epsilon) A X (X'X)^{-1} \\ &+ (X'X)^{-1} X' A \left\{ \frac{1}{n} [I_n \otimes tr(\Phi)] + \frac{1}{n^2} \left[ (1_n 1_n' - I_n) \otimes (1_n' vec(\Phi)) \right] \right\} A X (X'X)^{-1} \\ &- (X'X)^{-1} X' A \left( \frac{1}{n} 1_n 1_n' \right) (I_n - X(X'X)^{-1} X') E_{\epsilon}(\epsilon) E_{\epsilon}(\epsilon') (I_n - X(X'X)^{-1} X') \left( \frac{1}{n} 1_n 1_n' \right) A X (X'X)^{-1} \end{aligned}$$

where  $\Phi = (I_n - X(X'X)^{-1} X') E_{\epsilon}(\epsilon \epsilon') (I_n - X(X'X)^{-1} X')$ .

See the Appendix for the proof of **Theorem 8**.

If the bootstrapped residuals are not centered, this provides a general result for the linear model for any variance-covariance matrix,  $Cov_{\epsilon}(\epsilon)$ , for the population errors. However, with  $A = (I_n - (1/n)1_n 1_n')$  the bootstrapped residuals sum to zero, and this reduces down to

$$Cov_{\epsilon,H}(\hat{\beta}^*) = Cov_{\epsilon}(\hat{\beta}) + (X'X)^{-1} X' \left\{ \left( I_n - \frac{1}{n} 1_n 1_n' \right) \otimes \left( \frac{1}{n} tr(\Phi) - \frac{1}{n^2} 1_n' vec(\Phi) \right) \right\} X (X'X)^{-1}.$$

## 7. CONCLUSION

The objective of this paper has been to demonstrate an analytical approach to understanding the finite sample distribution of bootstrap estimators. In each example we showed how to replace the empirical process with the corresponding analytical process. Since not all bootstrap estimators are readily analyzed using this approach, a definition of a “directly analyzable” bootstrap estimator was given as a guide and as a way of characterizing the class of bootstrap estimators that can be most easily analyzed using this method. Often bootstrap estimators that are not directly analyzable can be made analyzable by use of an appropriate approximation such as a Taylor series expansion.

In this paper we provided examples of determining the variance of the standard bootstrap, two wild bootstraps and two fixed-block bootstraps. In addition, an example was provided of using this analytical approach to estimate the bias of a nonlinear bootstrap estimator. Finally, the analytical solution for the covariance matrix of the bootstrap estimator of the coefficients of a linear regression model with nonstochastic regressors was derived. The finite sample distributions of many of the most frequently used bootstrap procedures can be analyzed in this manner to determine bias, variance, mean squared error, skewness, kurtosis and many other finite sample properties and characteristics.

## APPENDIX

Proof of **Theorem 1**:

Substitute  $E_M \left[ \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right] = \frac{1}{n} \sum_{i=1}^n f(X_i)$  from **Lemma 1** and

$$E_M \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right)^2 \right] = \frac{(m+n-1)}{m n^2} \sum_{i=1}^n (f(X_i))^2 + \frac{2(m-1)}{m n^2} \sum_{i < k}^{\binom{n^2-n}{2}} f(X_i) f(X_k) \quad \text{from Lemma 2}$$

$$\text{into } \text{Var}_M \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] = E_M \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right)^2 \right] - \left[ E_M \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right]^2$$

$$\text{to get } \text{Var}_M \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] = \frac{n-1}{m n^2} \sum_{i=1}^n (f(X_i))^2 - \frac{2}{m n^2} \sum_{i < k}^{\binom{n^2-n}{2}} f(X_i) f(X_k). \quad \text{Q.E.D.}$$

Proof of **Theorem 2**:

Apply the definition of variance in this context as

$$\text{Var}_{M,X} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] = E_{M,X} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right)^2 \right] - \left[ E_{M,X} \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right]^2$$

We start by evaluating the first term on the right-hand side.

$$\begin{aligned} E_{M,X} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right)^2 \right] &= E_{M,X} \left[ \left( \frac{1}{m} \sum_{i=1}^n M_i f(X_i) \right)^2 \right] \\ &= E_{M,X} \left[ \frac{1}{m^2} \sum_{i=1}^n M_i^2 (f(X_i))^2 + \frac{2}{m^2} \sum_{i < k}^{\binom{n^2-n}{2}} M_i M_k f(X_i) f(X_k) \right] \\ &= \frac{1}{m^2} \sum_{i=1}^n [E_M(M_i^2)] E_X [(f(X_i))^2] + \frac{2}{m^2} \sum_{i < k}^{\binom{n^2-n}{2}} [E_M(M_i M_k)] E_X [f(X_i) f(X_k)] \end{aligned}$$

$$= \frac{(m+n-1)}{m n^2} \sum_{i=1}^n E_X [(f(X_i))^2] + \frac{2(m-1)}{m n^2} \binom{n^2-n}{2} \sum_{i < k} E_X [f(X_i)f(X_k)].$$

This completes the evaluation of the first term on the right-hand side.

Next evaluate the second term on the right-hand side. Using the work of **Lemma 1** we obtain

$$\begin{aligned} \left[ E_{M,X} \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right]^2 &= \left[ E_{M,X} \left( \frac{1}{m} \sum_{i=1}^n M_i f(X_i) \right) \right]^2 = \left[ \frac{1}{n} \sum_{i=1}^n E_X [f(X_i)] \right]^2 \\ &= \frac{1}{n^2} \sum_{i=1}^n (E_X [f(X_i)])^2 + \frac{2}{n^2} \binom{n^2-n}{2} \sum_{i < k} E_X [f(X_i)] E_X [f(X_k)]. \end{aligned}$$

Now substitute in to the formula for the variance.

$$\begin{aligned} \text{Var}_{M,X} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] &= E_{M,X} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right)^2 \right] - \left[ E_{M,X} \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right]^2 = \\ &= \frac{m+n-1}{m n^2} \sum_{i=1}^n E_X [(f(X_i))^2] - \frac{1}{n^2} \sum_{i=1}^n (E_X [f(X_i)])^2 + \frac{2(m-1)}{m n^2} \binom{n^2-n}{2} \sum_{i < k} E_X [f(X_i)f(X_k)] - \frac{2}{n^2} \binom{n^2-n}{2} \sum_{i < k} E_X [f(X_i)] E_X [f(X_k)] \end{aligned}$$

to obtain  $\text{Var}_{M,X} \left[ \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right] =$

$$\frac{1}{n^2} \sum_{i=1}^n \text{Var}_X [f(X_i)] + \frac{n-1}{m n^2} \sum_{i=1}^n (E_X [(f(X_i))^2]) + \frac{2}{n^2} \binom{n^2-n}{2} \sum_{i < k} \text{Cov}_X [f(X_i), f(X_k)] - \frac{2}{m n^2} \binom{n^2-n}{2} \sum_{i < k} E_X [f(X_i)f(X_k)].$$

*Q.E.D.*

Proof of **Theorem 3**:

We have established by **Lemma 3** that the mean is zero so the last term below for the variance

will be zero.

$$\begin{aligned}
& \text{Var}_{W,M} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] = E_{W,M} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right)^2 \right] - \left[ E_{W,M} \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right]^2 \\
& = E_M E_{W|M} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right)^2 \right] = E_M E_{W|M} \left[ \left( \frac{1}{m} \sum_{i=1}^n [W_i - (M_i - W_i)] f(X_i) \right)^2 \right] \\
& = E_M E_{W|M} \left[ \frac{1}{m^2} \sum_{i=1}^n (2W_i - M_i)^2 [f(X_i)]^2 + \frac{2}{m^2} \sum_{i < k}^{\binom{n^2-n}{2}} (2W_i - M_i)(2W_k - M_k) f(X_i) f(X_k) \right] \\
& = E_M E_{W|M} \left[ \frac{1}{m^2} \sum_{i=1}^n (4W_i^2 + M_i^2 - 4M_i W_i) [f(X_i)]^2 + \frac{2}{m^2} \sum_{i < k}^{\binom{n^2-n}{2}} (4W_i W_k - 2W_i M_k - 2M_i W_k + M_i M_k) f(X_i) f(X_k) \right]
\end{aligned}$$

Now consider the binomial distribution for  $W_i$  conditional on  $M_i$ . The probability of  $W_i$  successes (positive ones) is:

$$(A1) \quad P(W_i | M_i) = \binom{M_i}{W_i} 0.5^{W_i} (1-0.5)^{M_i-W_i} .$$

The following expected values follow immediately:

$$(A2) \quad E_{W|M} W_i = 0.5 M_i , \quad E_{W|M} W_i^2 = 0.25 M_i (M_i + 1) , \quad \text{and} \quad E_{W|M} W_i W_k = 0.25 M_i M_k .$$

$$\begin{aligned} \text{Var}_{W,M} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] &= E_M \left[ \frac{1}{m^2} \sum_{i=1}^n (4E_{W|M}(W_i^2) + M_i^2 - 4M_i E_{W|M}(W_i)) [f(X_i)]^2 \right] \\ &+ E_M \left[ \frac{2}{m^2} \sum_{i < k}^{\binom{n^2-n}{2}} (4E_{W|M}(W_i W_k) - 2E_{W|M}(W_i) M_k - 2M_i E_{W|M}(W_k) + M_i M_k) f(X_i) f(X_k) \right] \end{aligned}$$

$$\begin{aligned} \text{Var}_{W,M} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] &= E_M \left[ \frac{1}{m^2} \sum_{i=1}^n (M_i^2 + M_i + M_i^2 - 2M_i^2) [f(X_i)]^2 \right] \\ &+ E_M \left[ \frac{2}{m^2} \sum_{i < k}^{\binom{n^2-n}{2}} (M_i M_k - M_i M_k - M_i M_k + M_i M_k) f(X_i) f(X_k) \right] \end{aligned}$$

$$\begin{aligned} \text{Var}_{W,M} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] &= E_M \left[ \frac{1}{m^2} \sum_{i=1}^n M_i [(f(X_i))^2] \right] \\ &= \frac{1}{m^2} \sum_{i=1}^n E_M(M_i) [(f(X_i))^2] = \frac{1}{m n} \sum_{i=1}^n [(f(X_i))^2]. \end{aligned} \quad \text{Q.E.D.}$$

**Proof of *Theorem 4*:**

We have established by *Lemma 3* that the mean is zero, and this does not change if we first take the expected value with respect to  $X$  (the randomness from the original sample data), so the last term below for the variance will be zero.

$$\begin{aligned} \text{Var}_{W,M,X} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] &= E_{W,M,X} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right)^2 \right] - \left[ E_{W,M,X} \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right]^2 \\ &= E_M E_{W|M} E_{X|M,W} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right)^2 \right] = E_M E_{W|M} E_{X|M,W} \left[ \left( \frac{1}{m} \sum_{i=1}^n [W_i - (M_i - W_i)] f(X_i) \right)^2 \right] \end{aligned}$$

$$\begin{aligned}
&= E_M E_{W|M} \left[ \frac{1}{m^2} \sum_{i=1}^n (2W_i - M_i)^2 E_{X|M,W} [(f(X_i))^2] + \frac{2}{m^2} \sum_{i < k}^{\binom{n^2-n}{2}} (2W_i - M_i)(2W_k - M_k) E_{X|M,W} [f(X_i)f(X_k)] \right] \\
&= E_M E_{W|M} \left[ \frac{1}{m^2} \sum_{i=1}^n (4W_i^2 + M_i^2 - 4M_i W_i) E_X [(f(X_i))^2] \right. \\
&\quad \left. + E_M E_{W|M} \left[ \frac{2}{m^2} \sum_{i < k}^{\binom{n^2-n}{2}} (4W_i W_k - 2W_i M_k - 2M_i W_k + M_i M_k) E_X [f(X_i)f(X_k)] \right] \right].
\end{aligned}$$

Now consider the binomial distribution for  $W_i$  conditional on  $M_i$ .

The probability of  $W_i$  successes (positive ones) is:

$$(A3) \quad P(W_i | M_i) = \binom{M_i}{W_i} 0.5^{W_i} (1-0.5)^{M_i-W_i}$$

The following expected values follow immediately:

$$(A4) \quad E_{W/M} W_i = 0.5 M_i, \quad E_{W/M} W_i^2 = 0.25 M_i (M_i + 1), \quad \text{and} \quad E_{W/M} W_i W_k = 0.25 M_i M_k.$$

$$\begin{aligned}
\text{Var}_{W,M,X} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] &= E_M \left[ \frac{1}{m^2} \sum_{i=1}^n (4E_{W|M}(W_i^2) + M_i^2 - 4M_i E_{W|M}(W_i)) E_X [(f(X_i))^2] \right] \\
&+ E_M \left[ \frac{2}{m^2} \sum_{i < k}^{\binom{n^2-n}{2}} (4E_{W|M}(W_i W_k) - 2E_{W|M}(W_i) M_k - 2M_i E_{W|M}(W_k) + M_i M_k) E_X [f(X_i)f(X_k)] \right]
\end{aligned}$$

$$\begin{aligned} \text{Var}_{W,M,X} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] &= E_M \left[ \frac{1}{m^2} \sum_{i=1}^n (M_i^2 + M_i + M_i^2 - 2M_i^2) E_X \left[ (f(X_i))^2 \right] \right] \\ &+ E_M \left[ \frac{2}{m^2} \sum_{i < k}^{\binom{n^2-n}{2}} (M_i M_k - M_i M_k - M_i M_k + M_i M_k) E_X \left[ f(X_i) f(X_k) \right] \right] \end{aligned}$$

$$\begin{aligned} \text{Var}_{W,M,X} \left[ \left( \frac{1}{m} \sum_{j=1}^m f(X_j^*) \right) \right] &= E_M \left[ \frac{1}{m^2} \sum_{i=1}^n M_i E_X \left[ (f(X_i))^2 \right] \right] \\ &= \frac{1}{m^2} \sum_{i=1}^n E_M(M_i) E_X \left[ (f(X_i))^2 \right] = \frac{1}{m n} \sum_{i=1}^n E_X \left[ (f(X_i))^2 \right] \quad \text{Q.E.D.} \end{aligned}$$

Proof of **Theorem 5**:

Start with a common representation for the variance:

$$\begin{aligned} \text{Var}_M \left[ \frac{1}{b} \sum_{j=1}^b \theta(B_j^*) \right] &= \text{Var}_M \left[ \frac{1}{b} \sum_{g=1}^k M_g \theta(B_g) \right] = E_M \left\{ \left[ \frac{1}{b} \sum_{g=1}^k M_g \theta(B_g) \right]^2 \right\} - \left\{ E_M \left[ \frac{1}{b} \sum_{g=1}^k M_g \theta(B_g) \right] \right\}^2 \\ &= \frac{1}{b^2} \sum_{g=1}^k E_M(M_g^2) \left[ (\theta(B_g))^2 \right] + \frac{2}{b^2} \sum_{g < h}^{\binom{k^2-k}{2}} E_M(M_g M_h) \theta(B_g) \theta(B_h) \\ &\quad - \frac{1}{b^2} \sum_{g=1}^k (E_M(M_g))^2 \left[ (\theta(B_g))^2 \right] - \frac{2}{b^2} \sum_{g < h}^{\binom{k^2-k}{2}} E_M(M_g) E_M(M_h) \theta(B_g) \theta(B_h) \end{aligned}$$

Then substitute in from (4.3) and (4.4) to obtain:

$$\begin{aligned} &= \frac{1}{b^2} \sum_{g=1}^k \frac{b(b-1) + kb}{k^2} \left[ (\theta(B_g))^2 \right] + \frac{2}{b^2} \sum_{g < h}^{\binom{k^2-k}{2}} \frac{b(b-1)}{k^2} \theta(B_g) \theta(B_h) \\ &\quad - \frac{1}{b^2} \sum_{g=1}^k \left( \frac{b}{k} \right)^2 \left[ (\theta(B_g))^2 \right] - \frac{2}{b^2} \sum_{g < h}^{\binom{k^2-k}{2}} \left( \frac{b}{k} \right) \left( \frac{b}{k} \right) \theta(B_g) \theta(B_h) \end{aligned}$$

which, in turn, reduces to:

$$\text{Var}_M \left[ \frac{1}{b} \sum_{j=1}^b \theta(B_j^*) \right] = \frac{k-1}{b k^2} \sum_{g=1}^k [(\theta(B_g))^2] - \frac{2}{b k^2} \sum_{g < h}^{\binom{k^2-k}{2}} \theta(B_g) \theta(B_h). \quad Q.E.D.$$

**Proof of *Theorem 6*:**

Start with a standard formula for variance.

$$\begin{aligned} & \text{Var}_{M,B} \left[ \frac{1}{b} \sum_{j=1}^b \theta(B_j^*) \right] \\ &= \text{Var}_{M,B} \left[ \frac{1}{b} \sum_{g=1}^k M_g \theta(B_g) \right] \\ &= E_{M,B} \left\{ \left[ \frac{1}{b} \sum_{g=1}^k M_g \theta(B_g) \right]^2 \right\} - \left\{ E_{M,B} \left[ \frac{1}{b} \sum_{g=1}^k M_g \theta(B_g) \right] \right\}^2 \end{aligned}$$

where  $E_{M,B}$  is the expected value under the joint distribution of the bootstrap-induced randomness and the randomness implied by the underlying blocks of data values.

Follow a series of steps analogous to those in the proof of *Theorem 2* to obtain the following interim result.

$$\begin{aligned} \text{Var}_{M,B} \left[ \frac{1}{b} \sum_{j=1}^b \theta(B_j^*) \right] &= \frac{1}{b^2} \sum_{g=1}^k E_M(M_g^2) E_B[(\theta(B_g))^2] + \frac{2}{b^2} \sum_{g < h}^{\binom{k^2-k}{2}} E_M(M_g M_h) E_B[\theta(B_g) \theta(B_h)] \\ &\quad - \frac{1}{b^2} \sum_{g=1}^k (E_M(M_g))^2 [E_B[\theta(B_g)]]^2 - \frac{2}{b^2} \sum_{g < h}^{\binom{k^2-k}{2}} E_M(M_g) E_M(M_h) E_B[\theta(B_g)] E_B[\theta(B_h)]. \end{aligned}$$

Then substitute in from (4.3) and (4.4) to obtain

$$\begin{aligned}
\text{Var}_{M,B} \left[ \frac{1}{b} \sum_{j=1}^b \theta(B_j^*) \right] &= \frac{1}{b^2} \sum_{g=1}^k \frac{b(b-1)+kb}{k^2} E_B [(\theta(B_g))^2] + \frac{2}{b^2} \sum_{g<h}^{\binom{k^2-k}{2}} \frac{b(b-1)}{k^2} E_B [\theta(B_g)\theta(B_h)] \\
&\quad - \frac{1}{b^2} \sum_{g=1}^k \left(\frac{b}{k}\right)^2 [(E_B[\theta(B_g)])^2] - \frac{2}{b^2} \sum_{g<h}^{\binom{k^2-k}{2}} \left(\frac{b}{k}\right)\left(\frac{b}{k}\right) E_B [\theta(B_g)] E_B [\theta(B_h)].
\end{aligned}$$

which, in turn, reduces to:

$$\begin{aligned}
\text{Var}_{M,B} \left[ \frac{1}{b} \sum_{j=1}^b \theta(B_j^*) \right] &= \frac{b+k-1}{b k^2} \sum_{g=1}^k E_B [(\theta(B_g))^2] + \frac{2(b-1)}{b k^2} \sum_{g<h}^{\binom{k^2-k}{2}} E_B [\theta(B_g)\theta(B_h)] \\
&\quad - \frac{1}{k^2} \sum_{g=1}^k [(E_B[\theta(B_g)])^2] - \frac{2}{k^2} \sum_{g<h}^{\binom{k^2-k}{2}} E_B [\theta(B_g)] E_B [\theta(B_h)] \\
&= \frac{1}{k^2} \sum_{g=1}^k \text{Var}_B [\theta(B_g)] + \frac{k-1}{b k^2} \sum_{g=1}^k E_B [(\theta(B_g))^2] \\
&\quad + \frac{2}{k^2} \sum_{g<h}^{\binom{k^2-k}{2}} \text{Cov}_B [\theta(B_g), \theta(B_h)] - \frac{2}{b k^2} \sum_{g<h}^{\binom{k^2-k}{2}} E_B [\theta(B_g)\theta(B_h)]. \quad Q.E.D.
\end{aligned}$$

**Proof of Theorem 7:**

By substituting in  $Y^* = X\hat{\beta} + Ae^*$  we get  $\hat{\beta}^* = \hat{\beta} + (X'X)^{-1}X'Ae^*$  so that the mean of the bootstrapped regression coefficient estimator with respect to the distribution of  $H$  conditional on  $\epsilon$  becomes

$$\begin{aligned}
E_{H|\epsilon} \hat{\beta}^* &= E_{H|\epsilon} \hat{\beta} + E_{H|\epsilon} (X'X)^{-1} X'Ae^* \\
&= \hat{\beta} + E_{H|\epsilon} (X'X)^{-1} X'AH e \\
&= \hat{\beta} + (X'X)^{-1} X'AE_{H|\epsilon}(H) e
\end{aligned}$$

$$= \hat{\beta} + \frac{1}{n}(X'X)^{-1}X'A1_n1_n'e$$

Consequently, we have

$$\begin{aligned} \text{Cov}_{H|\epsilon}(\hat{\beta}^*) &= E_{H|\epsilon} \left\{ (\hat{\beta}^* - E_{H|\epsilon}\hat{\beta}^*)(\hat{\beta}^* - E_{H|\epsilon}\hat{\beta}^*)' \right\} \\ &= E_{H|\epsilon} \left\{ (X'X)^{-1}X'A \left( H - \frac{1}{n}1_n1_n' \right) ee' \left( H' - \frac{1}{n}1_n1_n' \right) AX (X'X)^{-1} \right\}. \end{aligned}$$

This expression expands to become

$$\begin{aligned} \text{Cov}_{H|\epsilon}(\hat{\beta}^*) &= (X'X)^{-1}X'AE_{H|\epsilon}(Hee'H')AX(X'X)^{-1} - \frac{1}{n} \left[ (X'X)^{-1}X'AE_{H|\epsilon}(H)ee'1_n1_n'AX(X'X)^{-1} \right] \\ &\quad - \frac{1}{n} \left[ (X'X)^{-1}X'A1_n1_n'ee'E_{H|\epsilon}(H')AX(X'X)^{-1} \right] + \frac{1}{n^2} \left[ (X'X)^{-1}X'A1_n1_n'ee'1_n1_n'AX(X'X)^{-1} \right]. \end{aligned}$$

Substituting in  $E_{H|\epsilon}(H) = \frac{1}{n}1_n1_n'$  and collecting terms this reduces to

$$\text{Cov}_{H|\epsilon}(\hat{\beta}^*) = (X'X)^{-1}X'AE_{H|\epsilon}(Hee'H')AX(X'X)^{-1} - \frac{1}{n^2} \left[ (X'X)^{-1}X'A1_n1_n'ee'1_n1_n'AX(X'X)^{-1} \right]$$

As already noted, the last term is needed in case the original set of residuals do not sum up to zero (e.g. a regression that has no intercept term).

We now need to examine  $Hee'H'$  and determine its expected value under the distribution of  $H$  conditional on  $\epsilon$ . The  $Hee'H'$  matrix can be decomposed as follows:

$$Hee'H' = \left[ \sum_{i=1}^n \sum_{j=1}^n (v_i) \otimes (H_{ij} e_j) \right] \left[ \sum_{k=1}^n \sum_{l=1}^n (v'_k) \otimes (e_l H_{kl}) \right]$$

where  $v_i$  and  $v_k$  are  $n \times 1$  column vectors whose elements are all zeros except for a one in their  $i^{\text{th}}$  and  $k^{\text{th}}$  positions, respectively.

$$Hee'H' = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n (v_i v'_k) \otimes (e_j e_l H_{ij} H_{kl})$$

We can now separate the diagonal elements from the off-diagonal elements of  $Hee'H'$  and take the expected value.

$$E_{H|\epsilon}(Hee'H') = E_{H|\epsilon} \left[ \sum_{i=1}^n \sum_{j=1}^n (v_i v_i') \otimes (e_j^2 H_{ij}^2) \right] + E_{H|\epsilon} \left[ \sum_{i \neq k}^n \sum_{j=1}^n \sum_{l=1}^n (v_i v_k') \otimes (e_j e_l H_{ij} H_{kl}) \right]$$

with the diagonal elements in the first term and the off-diagonal in the second term.

Since  $E_{H|\epsilon}(H_{ij}) = \frac{1}{n}$ ,  $E_{H|\epsilon}(H_{ij}^2) = \frac{1}{n}$  and  $E_{H|\epsilon}(H_{ij} H_{kl}) = E_{H|\epsilon}(H_{ij}) E_{H|\epsilon}(H_{kl}) = \frac{1}{n^2}$ , we have

$$\begin{aligned} E_{H|\epsilon}(Hee'H') &= \frac{1}{n} \left[ \sum_{i=1}^n \sum_{j=1}^n (v_i v_i') \otimes (e_j^2) \right] + \frac{1}{n^2} \left[ \sum_{i \neq k}^n \sum_{j=1}^n \sum_{l=1}^n (v_i v_k') \otimes (e_j e_l) \right] \\ &= \frac{1}{n} \left[ \sum_{i=1}^n (v_i v_i') \otimes \left( \sum_{j=1}^n e_j^2 \right) \right] + \frac{1}{n^2} \left[ \sum_{i \neq k}^n (v_i v_k') \otimes \left( \sum_{j=1}^n \sum_{l=1}^n e_j e_l \right) \right] \\ &= \frac{1}{n} [I_n \otimes (e'e)] + \frac{1}{n^2} \left[ (1_n 1_n' - I_n) \otimes (1_{n^2}' \text{vec}(ee')) \right] \end{aligned}$$

where  $I_n$  is an  $n \times n$  identity matrix,  $1_n$  is an  $n \times 1$  column vector of all ones,  $1_{n^2}$  is an  $n^2 \times 1$  column vector of all ones, and “vec” is the standard vectorization operator.

Now substitute this in for  $E_{H|\epsilon}(Hee'H')$  in the covariance matrix to get

$$\begin{aligned} \text{Cov}_{H|\epsilon}(\hat{\beta}^*) &= (X'X)^{-1} X'A \left\{ \frac{1}{n} [I_n \otimes (e'e)] + \frac{1}{n^2} \left[ (1_n 1_n' - I_n) \otimes (1_{n^2}' \text{vec}(ee')) \right] \right\} AX (X'X)^{-1} \\ &\quad - \frac{1}{n^2} (X'X)^{-1} X'A 1_n 1_n' ee' 1_n 1_n' AX (X'X)^{-1}. \end{aligned} \quad \text{Q.E.D.}$$

Proof of **Theorem 8**:

Starting with  $\hat{\beta}^* = (X'X)^{-1} X'Y^*$  and then substituting in  $Y^* = X\hat{\beta} + Ae^*$

we get  $\hat{\beta}^* = \hat{\beta} + (X'X)^{-1} X'Ae^*$  where  $e^* = He$  and  $e = (I_n - X(X'X)^{-1}X')\epsilon$

so that  $\hat{\beta}^* = \hat{\beta} + (X'X)^{-1} X'AH(I_n - X(X'X)^{-1}X')\epsilon$ .

Then the mean of the bootstrapped regression coefficient estimator with respect to the joint distribution of  $\epsilon$  and  $H$  becomes

$$\begin{aligned} E_{\epsilon, H} \hat{\beta}^* &= E_{\epsilon, H} \hat{\beta} + E_{\epsilon, H} (X'X)^{-1} X'Ae^* \\ &= E_{\epsilon, H} \hat{\beta} + E_{\epsilon, H} (X'X)^{-1} X'AH e \\ &= E_{\epsilon} \hat{\beta} + E_H E_{\epsilon|H} (X'X)^{-1} X'AH (I_n - X(X'X)^{-1}X') \epsilon \\ &= E_{\epsilon} \hat{\beta} + (X'X)^{-1} X'AE_H(H)(I_n - X(X'X)^{-1}X')E_{\epsilon}(\epsilon) . \end{aligned}$$

We can substitute in  $E_H(H) = \frac{1}{n} 1_n 1_n'$ , but do not wish to restrict the expected errors to be zero

in order to make our results apply even when the zero-mean condition does not hold.

$$E_{\epsilon, H} \hat{\beta}^* = E_{\epsilon} \hat{\beta} + \frac{1}{n} (X'X)^{-1} X'A 1_n 1_n' (I_n - X(X'X)^{-1}X') E_{\epsilon}(\epsilon) .$$

Therefore, given  $\hat{\beta}^* = \hat{\beta} + (X'X)^{-1} X'AH(I_n - X(X'X)^{-1}X')\epsilon$ , it follows that

$$\begin{aligned} (\hat{\beta}^* - E_{\epsilon, H} \hat{\beta}^*) &= (\hat{\beta} - E_{\epsilon} \hat{\beta}) + (X'X)^{-1} X'AH(I_n - X(X'X)^{-1}X')\epsilon \\ &\quad - (X'X)^{-1} X'A \left( \frac{1}{n} 1_n 1_n' \right) (I_n - X(X'X)^{-1}X') E_{\epsilon}(\epsilon) \end{aligned}$$

so we have

$$(\hat{\beta}^* - E_{\epsilon, H} \hat{\beta}^*) (\hat{\beta}^* - E_{\epsilon, H} \hat{\beta}^*)'$$

$$\begin{aligned}
&= (\hat{\beta} - E_{\epsilon} \hat{\beta})(\hat{\beta} - E_{\epsilon} \hat{\beta})' + (\hat{\beta} - E_{\epsilon} \hat{\beta})\epsilon'(I_n - X(X'X)^{-1}X')H'A X(X'X)^{-1} \\
&- (\hat{\beta} - E_{\epsilon} \hat{\beta})E_{\epsilon}(\epsilon')(I_n - X(X'X)^{-1}X')\left(\frac{1}{n}1_n1_n'\right)AX(X'X)^{-1} \\
&+ (X'X)^{-1}X'A H(I_n - X(X'X)^{-1}X')\epsilon(\hat{\beta} - E_{\epsilon} \hat{\beta})' \\
&- (X'X)^{-1}X'A \left(\frac{1}{n}1_n1_n'\right)(I_n - X(X'X)^{-1}X')E_{\epsilon}(\epsilon)(\hat{\beta} - E_{\epsilon} \hat{\beta})' \\
&+ (X'X)^{-1}X'A H(I_n - X(X'X)^{-1}X')\epsilon\epsilon'(I_n - X(X'X)^{-1}X')H'A X(X'X)^{-1} \\
&- (X'X)^{-1}X'AH(I_n - X(X'X)^{-1}X')\epsilon E_{\epsilon}(\epsilon')(I_n - X(X'X)^{-1}X')\left(\frac{1}{n}1_n1_n'\right)AX(X'X)^{-1} \\
&- (X'X)^{-1}X'A \left(\frac{1}{n}1_n1_n'\right)(I_n - X(X'X)^{-1}X')E_{\epsilon}(\epsilon)\epsilon'(I_n - X(X'X)^{-1}X')H'A X(X'X)^{-1} \\
&+ (X'X)^{-1}X'A \left(\frac{1}{n}1_n1_n'\right)(I_n - X(X'X)^{-1}X')E_{\epsilon}(\epsilon)E_{\epsilon}(\epsilon')(I_n - X(X'X)^{-1}X')\left(\frac{1}{n}1_n1_n'\right)AX(X'X)^{-1} \\
&.
\end{aligned}$$

This provides the basis for an expression for the covariance matrix for this bootstrap regression estimator.

$$\begin{aligned}
Cov_{\epsilon, H}(\hat{\beta}^*) &= E_{\epsilon, H} \left\{ (\hat{\beta}^* - E_{\epsilon, H} \hat{\beta}^*)(\hat{\beta}^* - E_{\epsilon, H} \hat{\beta}^*)' \right\} \\
Cov_{\epsilon, H}(\hat{\beta}^*) &= Cov_{\epsilon}(\hat{\beta}) + (X'X)^{-1}X'AE_H(H\Phi H')AX(X'X)^{-1} \\
&+ (X'X)^{-1}X'A Cov_{\epsilon}(\epsilon)(I_n - X(X'X)^{-1}X')\left(\frac{1}{n}1_n1_n'\right)AX(X'X)^{-1} \\
&+ (X'X)^{-1}X'A \left(\frac{1}{n}1_n1_n'\right)(I_n - X(X'X)^{-1}X')Cov_{\epsilon}(\epsilon)AX(X'X)^{-1} \\
&- (X'X)^{-1}X'A \left(\frac{1}{n}1_n1_n'\right)(I_n - X(X'X)^{-1}X')E_{\epsilon}(\epsilon)E_{\epsilon}(\epsilon')(I_n - X(X'X)^{-1}X')\left(\frac{1}{n}1_n1_n'\right)AX(X'X)^{-1}
\end{aligned}$$

where  $\Phi = (I_n - X(X'X)^{-1}X')E_{\epsilon}(\epsilon\epsilon')E_{\epsilon}(I_n - X(X'X)^{-1}X')$ .

Now we need to determine  $E_H(H\Phi H')$  in order to complete our derivation of the variance-covariance matrix of the bootstrap estimator without imposing the zero-means assumption.

Given an  $n \times 1$  column vector  $r_i$  whose elements are all zeros except for a one in its  $i^{\text{th}}$  “row” position, and a  $1 \times n$  row vector  $c_j$  also made up of all zeros except for a one in its  $j^{\text{th}}$

“column” position, we have  $H = \sum_{i=1}^n \sum_{j=1}^n r_i c_j H_{ij}$ ,  $E_{H|\epsilon}(H) = \sum_{i=1}^n \sum_{j=1}^n r_i c_j E_{j|\epsilon}(H_{ij})$  and

$E_{j|\epsilon}(H_{ij}) = \frac{1}{n}$  so that  $E_{H|\epsilon}(H) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n r_i c_j = \frac{1}{n} 1_n 1_n'$  where  $1_n$  is an  $n \times 1$  column vector of

all ones. In order to determine  $E_H(H\Phi H')$ , we need to decompose the  $H$ , the  $\Phi$  and the  $H'$

matrices as follows:

$$\begin{aligned} H e e' H' &= \left[ \sum_{i=1}^n \sum_{j=1}^n (r_i c_j) \otimes (H_{ij}) \right] \left[ \sum_{k=1}^n \sum_{l=1}^n (r_k c_l) \otimes (\Phi_{kl}) \right] \left[ \sum_{s=1}^n \sum_{t=1}^n (c'_t r'_s) \otimes (H_{st}) \right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \sum_{s=1}^n \sum_{t=1}^n (r_i c_j r_k c_l c'_t r'_s) \otimes (H_{ij} \Phi_{kl} H_{st}) . \\ H e e' H' &= \sum_{k=1}^n \sum_{l=1}^n \sum_{i=s}^n \sum_{j=t}^n (r_i c_j r_k c_l c'_t r'_s) \otimes (\Phi_{kl} H_{ij} H_{st}) \\ &+ \sum_{k=1}^n \sum_{l=1}^n \sum_{i=s}^n \sum_{j \neq t}^n (r_i c_j r_k c_l c'_t r'_s) \otimes (\Phi_{kl} H_{ij} H_{st}) \\ &+ \sum_{k=1}^n \sum_{l=1}^n \sum_{i \neq s}^n \sum_{j=t}^n (r_i c_j r_k c_l c'_t r'_s) \otimes (\Phi_{kl} H_{ij} H_{st}) \\ &+ \sum_{k=1}^n \sum_{l=1}^n \sum_{i \neq s}^n \sum_{j \neq t}^n (r_i c_j r_k c_l c'_t r'_s) \otimes (\Phi_{kl} H_{ij} H_{st}) . \end{aligned}$$

The first two of these four terms strictly refer to the diagonal elements which are generated whenever  $i=s$  since the dot product of row “ $i$ ” of the  $H$  matrix is formed with column “ $s$ ” of the  $H'$  matrix with the corresponding elements of the  $\Phi$  matrix serving as weights.

The second term is always zero, because on the  $i^{\text{th}}$  draw only one element of  $e$  is drawn so the corresponding element of  $H$  is a one while all the other elements of  $H$  for the  $i^{\text{th}}$  draw are zeros. Therefore, any product of two elements of  $H$  for the same draw (i.e. the  $i^{\text{th}}$  draw) will be either a one times a zero, a zero times a one, or a zero times a zero.

The last two terms then refer to the off-diagonal elements of the  $H\Phi H'$  matrix since for these last two terms “ $i$ ” is not equal to “ $s$ ”.

By defining  $c_i = r_i'$  and  $c_s = r_s'$ , and then combining the last two terms while recognizing that  $c_j r_k$  equals one when  $j=k$  and is zero otherwise, and  $c_l c_t'$  is one when  $l=t$  and is zero otherwise, then  $H\Phi H'$  may be re-expressed as

$$H e e' H' = \sum_{i=1}^n \sum_{j=1}^n (r_i c_i) \otimes (\Phi_{jj} H_{ij}^2) + \sum_{i \neq s} \sum_{j=1}^n \sum_{t=1}^n (r_i c_s) \otimes (\Phi_{jt} H_{ij} H_{st})$$

where the first term of this new expression represents the diagonal elements of  $H\Phi H'$  and the second term represents the off-diagonal elements.

Recall that  $E_{H,\epsilon}(H_{ij}) = \frac{1}{n}$  and  $E_{H|\epsilon}(H) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n r_i c_j = \frac{1}{n} \mathbf{1}_n \mathbf{1}_n'$ , and since  $H_{ij}$  is either zero or one, then  $E_{H,\epsilon}(H_{ij}^2) = \frac{1}{n}$  and  $E_{H,\epsilon}(H_{ij} H_{st}) = E_{H,\epsilon}(H_{ij}) E_{H,\epsilon}(H_{st}) = \frac{1}{n^2}$  for  $i \neq s$ .

Therefore,

$$E_{H|\epsilon}(H\Phi H') = \frac{1}{n} [I_n \otimes \text{tr}(\Phi)] + \frac{1}{n^2} \left[ (\mathbf{1}_n \mathbf{1}_n' - I_n) \otimes (\mathbf{1}_n^2 \text{vec}(\Phi)) \right]$$

where “ $\text{tr}$ ” means trace, “ $\text{vec}$ ” is the usual vectorization operator, and

$$\Phi = (I_n - X(X'X)^{-1}X') E_{\epsilon}(\epsilon \epsilon') (I_n - X(X'X)^{-1}X')$$

Therefore, substituting into our expression for the bootstrap variance-covariance matrix, we have

$$\begin{aligned}
Cov_{\epsilon, H}(\hat{\beta}^*) &= Cov_{\epsilon}(\hat{\beta}) + (X'X)^{-1}X'A Cov_{\epsilon}(\epsilon)(I_n - X(X'X)^{-1}X')\left(\frac{1}{n}1_n1_n'\right)AX(X'X)^{-1} \\
&+ (X'X)^{-1}X'A\left(\frac{1}{n}1_n1_n'\right)(I_n - X(X'X)^{-1}X')Cov_{\epsilon}(\epsilon)AX(X'X)^{-1} \\
&+ (X'X)^{-1}X'A\left\{\frac{1}{n}[I_n \otimes tr(\Phi)] + \frac{1}{n^2}\left[(1_n1_n' - I_n) \otimes (1_n' vec(\Phi))\right]\right\}AX(X'X)^{-1} \\
&- (X'X)^{-1}X'A\left(\frac{1}{n}1_n1_n'\right)(I_n - X(X'X)^{-1}X')E_{\epsilon}(\epsilon)E_{\epsilon}(\epsilon')(I_n - X(X'X)^{-1}X')\left(\frac{1}{n}1_n1_n'\right)AX(X'X)^{-1}
\end{aligned}$$

where  $\Phi = (I_n - X(X'X)^{-1}X')E_{\epsilon}(\epsilon\epsilon')(I_n - X(X'X)^{-1}X')$ . *Q.E.D.*

## REFERENCES

- Beran, R. (1986). Discussion of “Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis” by C.F.J. Wu. *Annals of Statistics*, 14, 1295-1298.
- Carlestein, E. (1986). The Use of Subseries Methods for Estimating the Variance of a General Statistic from a Stationary Times Series. *Annals of Statistics*, 14, 1171-1179.
- Davidson, R. and E. Flachaire (2001). The Wild Bootstrap, Tamed at Last. STICERD working paper, London School of Economics, Darp58.
- Davison, A.C., D.V. Hinkley and E. Schechtman (1986). Efficient Bootstrap Simulation. *Biometrika*, 73 (3), 555-566.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics*, 7, 1-26.
- Freedman, D.A. (1981). Bootstrapping Regression Models. *Annals of Statistics*, 9 (6), 1218-1228.
- Hall, P. (1985). Resampling a Coverage Process. *Stochastic Process Applications*, 19, 259-269.
- Hall, P., J. L. Horowitz and B.-Y. Jing (1995). On Blocking Rules for the Bootstrap with Dependent Data. *Biometrika*, 82 (3), 561-574.
- Härdle, W., J. L. Horowitz and J.-P. Kreiss (2003). Bootstrap Methods for Time Series. *International Statistical Review*, 71 (2), 435-459.
- Horowitz, J. L. (2001). The Bootstrap. In *Handbook of Econometrics*, vol. 5, Eds. J. J. Heckman and E. E. Leamer. Amsterdam: North Holland Publishing Co.
- Horowitz, J. L. (2003). The Bootstrap in Econometrics. *Statistical Science*, 18 (2), 211-218.
- Künsch (1989). The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics*, 17, 1217-1241.
- Liu, R. Y. (1988). Bootstrap Procedure Under Some Non-i.i.d. Models. *Annals of Statistics*, 16, 1696-1708.
- Mammen, E. (1993). Bootstrap and Wild Bootstrap for High Dimensional Linear Models. *Annals of Statistics*, 21, 255-285.
- Politis, D.N. (2003). The Impact of Bootstrap Methods on Time Series Analysis. *Statistical Science*, 18 (2), 219-230.
- Politis, D.N. and J.P. Romano (1993). Estimating the Distribution of a Studentized Statistics by Subsampling. *Bulletin of the International Statistical Institute*, 2, 315-316.
- Wu, C.F.J. (1986). Jackknife Bootstrap and Other Resampling Methods in Regression Analysis. *Annals of Statistics*, 14, 1261-1295.