The Bootstrap's Finite Sample Distribution: An Analytical Approach

Lawrence C. Marsh[1]
Department of Economics and Econometrics
University of Notre Dame, Notre Dame, IN 46556-5611

Abstract

This paper provides a method for determining the exact 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 derive the variances of some selected bootstrap estimators. Other exact 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 (including ratios and matrix inverses) 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: Monday, September 20, 2004

---

[1] I thank Tom Cosimano, Ken Kelley, Nelson Mark, Ke-Hai Yuan and my colleagues in the Department of Economics and Econometrics for helpful comments and suggestions.

1. INTRODUCTION

The purpose of this paper is to propose an analytical approach to understanding the exact 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.[2] 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 exactly 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.[3] 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.

[2] The term "directly analyzable" will be defined in section 2 of this paper.

[3] 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.

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 determining the exact analytical (and numerical) value of the variance of the statistic of interest implied by the bootstrap procedure without actually having to generate the simulated bootstrap samples. 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. The first formula tells the applied econometrician how to obtain the exact number needed for an application, while the second expression tells the econometric theorist what that first number represents in terms of the variances, covariances and other moments of the underlying, unknown distribution.

The plan of the paper is as follows. In section 2 we define the directly analyzable class of bootstrap estimators and, as an example of deriving moments, we 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 Hall-Carlestein (non-overlapping) and Hall-Künsch (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. The proofs for the lemmas and theorems are provided in the Appendix.

2. The Standard Bootstrap Procedure

Consider a set of n real numbers $\{X_i : i = 1,\ldots,n\}$ where each X_i may be scalar or vector valued and where n is a finite, positive integer. Typically bootstrap applications involve 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 $\{X_i : i = 1,\ldots,n\}$ 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 $\{X_j^* : j = 1,\ldots,m\}$. 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 $\{M_i : i = 1,\ldots,n\}$ records the frequency with which each of the corresponding elements of the original set $\{X_i : i = 1,\ldots,n\}$ 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 $\sum_{i=1}^{n} m_i = m$.

Denote the set of any function, f, of each of the original sample values as $\{f(X_i) : i = 1,\ldots,n\}$. Denote the set of that same function of each of the individual bootstrapped values as $\{f(X_j^*) : j = 1,\ldots,m\}$. An 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, $\{X_i : i = 1,\ldots,n\}$, 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* $Cov_X[\cdot,\cdot]$.

***Definition*:** *Any bootstrap statistic,* θ_n^**, that is a function of the elements of the set* $\{f(X_j^*): j = 1,\ldots,m\}$ *and satisfies the separability condition*

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

where $g(M_i)$ *and* $h(f(X_i))$ *are independent functions and where 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. Consequently, we will restrict our examples to estimators that are either directly analyzable or that can be made analyzable by a Taylor series expansion.

We begin with the standard bootstrap, which will provide a benchmark or starting point for the use of this analytical approach for deriving any of the moments (in this example the variance) of the exact finite sample distribution of various bootstrap estimators under a variety of bootstrap procedures. 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 the moments of the original underlying distribution.

We provide *Assumption 1* because each bootstrap procedure is different in the way in which the bootstrap sample is drawn or transformed upon being drawn. Later sections of this paper will define other assumptions that define other bootstrap procedures. This assumption is the simplest assumption and holds only for the standard bootstrap as analyzed in this section.

***Assumption 1*:** *For the standard bootstrap, each of the m bootstrapped values for the set* $\{X_j^*: j = 1,\ldots,m\}$ *is drawn with replacement with equal probability from the original set of n sample values* $\{X_i : i = 1,\ldots,n\}$ *having corresponding random frequencies* $\{M_i : i = 1,\ldots,n\}$.

Under *Assumption 1* and any function, *f*, the empirical bootstrap process can be transformed into the corresponding analytical process using the following equality.

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

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 $\{M_i = m_i : i = 1,\ldots,n\}$ may be expressed as

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

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 Assumption 1,*

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

This result is already well understood and accepted from simulations in the bootstrap literature.

Lemma 2: *Under Assumption 1,*

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

For any moment of the bootstrap-induced distribution represented by the function, f, these two lemmas provide the results needed to produce the corresponding variance of that moment.

Theorem 1: *Under Assumption 1,*

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

The role of the original sample size n is kept separate from that of the bootstrap sample size m. This analytical formula equals the empirical average over an infinite number of bootstrap samples, each of size m. The usual empirical bootstrap simulations are attempting to approximate this analytical quantity. Often, in common practice, m is set equal to n.

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 exact 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 general, we will do this in terms of the unknown moments of the underlying random variable using our expectation notation, $E_X[.]$, $\mathrm{Var}_X[.]$, and $\mathrm{Cov}_X[. , .]$. 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 Assumption 1,*

$$Var_{M,X}\left[\frac{1}{m}\sum_{j=1}^{m} f\left(X_j^*\right)\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var_X\left[f\left(X_i\right)\right] + \frac{n-1}{m\,n^2}\sum_{i=1}^{n} E_X\left[\left(f\left(X_i\right)\right)^2\right]$$

$$+ \frac{2}{n^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)} Cov_X\left[f\left(X_i\right),\ f\left(X_k\right)\right] - \frac{2}{m\,n^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)} E_X\left[f\left(X_i\right) f\left(X_k\right)\right].$$

In addition to the variance of the sum being the sum of the variances plus two times the covariances, the standard bootstrap adds on some extra terms. These extra terms in *Theorem 2* go away when drawing an infinite number of times with replacement in creating each bootstrap sample ($m \to \infty$). Alternative bootstrap methods that don't generate these extra 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 $f\left(X_j^*\right)=\left(X_j^*-\overline{X}\right)^2$ where $\overline{X}$ 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

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

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

By *Theorem 1* we have

$$Var_M\left[\frac{1}{m}\sum_{j=1}^{m}\left(X_j^*-\overline{X}\right)^2\right]=\frac{n-1}{m\,n^2}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^4-\frac{2}{m\,n^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)}\left(X_i-\overline{X}\right)^2\left(X_k-\overline{X}\right)^2,$$

which gives the formula for the variance of the bootstrap variance estimator under the EDF.

By *Theorem 2* we have

$$Var_{M,X}\left[\frac{1}{m}\sum_{j=1}^{m}\left(X_j^*-\overline{X}\right)^2\right]=\frac{1}{n^2}\sum_{i=1}^{n}Var_X\left[\left(X_i-\overline{X}\right)^2\right]+\frac{n-1}{m\,n^2}\sum_{i=1}^{n}E_X\left[\left(X_i-\overline{X}\right)^4\right]$$

$$+\frac{2}{n^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)}Cov_X\left[\left(X_i-\overline{X}\right)^2,\left(X_k-\overline{X}\right)^2\right]-\frac{2}{m\,n^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)}E_X\left[\left(X_i-\overline{X}\right)^2\left(X_k-\overline{X}\right)^2\right]$$

which tells us what object the formula in *Theorem 1* is actually delivering under the joint distribution. One interesting aspect of this result is that the bootstrap is adding positive values to the variances and subtracting positive values from the covariances and apparently influencing the stability of the corresponding variance-covariance matrix. This ridge-regression type effect is greatest when *m* is small and goes away as *m* approaches infinity. Similar insights might be available for skewness, kurtosis and other features of the finite sample distribution.

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

In the context of bootstrapping regression residuals, one difficulty with using the standard bootstrap is that it fails to enforce some 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 $-\left(\sqrt{5}-1\right)/2$ with probability p = $\left(\sqrt{5}+1\right)/\left(2\sqrt{5}\right)$ and $\left(\sqrt{5}+1\right)/2$ with probability (1-p) = $\left(\sqrt{5}-1\right)/\left(2\sqrt{5}\right)$. We will analyze the Rademacher version, but a similar analysis of the Mammen version follows easily by analogy.

To clearly differentiate the Liu-Davidson-Flachaire wild bootstrap from the standard bootstrap, we provide *Assumption 2* in contrast to *Assumption 1* which served as the basis for the standard bootstrap.

***Assumption 2*:** *For the Liu-Davidson-Flachaire wild bootstrap, draw a sample of size m with replacement from a set of n real numbers* $\{X_i : i = 1,\ldots,n\}$ *with equal probability, and, by doing so, produce a corresponding set of random frequencies* $\{M_i : i = 1,\ldots,n\}$ *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* $\{X_j^* : j = 1,\ldots,m\}$.

We continue with the expected value notation of section 2, but add the following. 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 Assumption 2,* $E_{W,M}\left(\frac{1}{m}\sum_{j=1}^{m} f\left(X_j^*\right)\right)=0$.

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 *Assumption 2*.

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

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

Q.E.D.

Lemma 3 shows that the Rademacher distribution centers the bootstrap statistic around zero.

We will now look to see what it does to the bootstrap statistic's variance.

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

Thus, once again, the analytical process replaces the empirical process in producing a formula that is equivalent to averaging over an infinite number of bootstrap samples, each of size *m*. This wild bootstrap has eliminated the cross-product term that was part of the standard bootstrap. The variance of the bootstrap estimator under the joint distribution of the bootstrap-induced randomness and the randomness of the underlying distribution has also dropped the covariances that were present in the corresponding formula for the standard bootstrap as well as the two extra terms (see *Theorem 3*) as shown in *Theorem 4*.

Theorem 4: *Under Assumption 2,*

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

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

This provides us with a much clearer picture of what the formula in *Theorem 3* is delivering. The Liu-Davidson-Flachaire wild bootstrap not only eliminates the "extra" terms in *Theorem 2*, but also eliminates all the covariances as well. Also, now the variance goes to zero for infinite *m*. Analogously, under the Mammen (1993) wild bootstrap, the probability of W_i successes is

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

and the needed expected values are

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

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 boostrap we get

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

and by analogy with *Theorem 4* we get

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

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

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 Mammem (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 boostrap 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 $\{\mathbf{Y}_t, \mathbf{X}_t : t=1,\ldots,T\}$, form k matrix blocks, $\{\boldsymbol{B}_g\colon g=1,\ldots,k\}$, each with l consecutive time-specific matrix observations.

For the non-overlapping block bootstrap, for some given l, choose k so that $k\,l = T$ and order the blocks such that the first block, $\boldsymbol{B}_1$, covers observations from 1 through l, the second block, $\boldsymbol{B}_2$, covers observations from $l+1$ through $2\,l$, and so forth until the last block, $\boldsymbol{B}_k$, which covers observations from $T-l+1$ through T.

For the overlapping block bootstrap the first block, $\boldsymbol{B}_1$, covers observations from 1 through l, the second block, $\boldsymbol{B}_2$, covers observations from 2 through $l+1$, and so forth until the last block, $\boldsymbol{B}_k$, which covers observations from $T-l+1$ through T. The overlapping block bootstrap has $k = T-l+1$ blocks available for sampling.

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, $k = T - T^{1/3} + 1$.

In addition to the expected value notation used in the preceding sections, we add the following. 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 $Var_M[\cdot]$, $Var_{M,B}[\cdot]$, $Var_B[\cdot]$, $Cov_M[\cdot,\cdot]$ and $Cov_B[\cdot,\cdot]$.

We provide *Assumption 3* to specify the random sampling procedure for the fixed block bootstrap.

***Assumption 3*:** *For the fixed block bootstrap, randomly select with replacement b blocks from the k blocks,* $\{\boldsymbol{B}_g: g=1,\ldots,k\}$*, with corresponding random frequencies* $\{M_g : g = 1,\ldots,k\}$ *to obtain the bootstrap resampled blocks:* $\{\boldsymbol{B}_j^*: j=1,\ldots,b\}$.

Note that under the bootstrap procedure defined under *Assumption 3,* the integer *b* may be smaller than, equal to, or greater than the integer *k*. Under this fixed block bootstrap procedure, we average some known function, θ, of each of these randomly selected blocks.

(4.1)
$$\frac{1}{b}\sum_{j=1}^{b}\theta\left(B_j^*\right) = \frac{1}{b}\sum_{g=1}^{k}M_g\,\theta\left(B_g\right) .$$

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

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

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

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

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

(4.4)
$$E_M\left(M_g\, M_h\right) = \frac{b\left(b-1\right)}{k^2} \quad \textit{and} \quad Cov_M\left(M_g\, , M_h\right) = \frac{-b}{k^2} .$$

The following theorem provides the variance of the block bootstrap estimator in this context.

***Theorem 5*:** Under *Assumption 3,*

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

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

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

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.

From the point of view of econometric theory the next theorem reveals what the formula in *Theorem 5* is providing.

***Theorem 6*:** *Under Assumption 3,*

$$Var_{M,B}\left[\frac{1}{b}\sum_{j=1}^{b}\theta\left(B_j^*\right)\right] = \frac{1}{k^2}\sum_{g=1}^{k}Var_B\left[\theta\left(B_g\right)\right] + \frac{k-1}{b\,k^2}\sum_{g=1}^{k}E_B\left[\left(\theta\left(B_g\right)\right)^2\right]$$
$$+ \frac{2}{k^2}\sum_{g<h}^{\left(\frac{k^2-k}{2}\right)}Cov_B\left[\theta\left(B_g\right),\theta\left(B_h\right)\right] - \frac{2}{b\,k^2}\sum_{g<h}^{\left(\frac{k^2-k}{2}\right)}E_B\left[\theta\left(B_g\right)\theta\left(B_h\right)\right] .$$

This result is similar to that of *Theorem 2* except here the data from the blocks are being combined using some known function $\theta(.)$ that is specified by the user of the block bootstrap. 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 $\{M_i : i = 1,\ldots,n\}$, are directly analyzable functions, then nonlinear bootstrap functions become directly analyzable when expressed as Taylor series expansions.

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

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

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

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

$$E_M\left[\left(\overline{X}^* - \overline{X}\right)' G_2\left(\overline{X}\right)\left(\overline{X}^* - \overline{X}\right)\right] =$$

$$E_M\left[\overline{X}^{*\prime} G_2\left(\overline{X}\right)\overline{X}^*\right] - E_M\left[\overline{X}^{*\prime} G_2\left(\overline{X}\right)\overline{X}\right] - E_M\left[\overline{X}' G_2\left(\overline{X}\right)\overline{X}^*\right] + E_M\left[\overline{X}' G_2\left(\overline{X}\right)\overline{X}\right]$$

where the third term is equal to

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

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

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

Consequently, the primary bias term simplifies to

$$E_M\left[\left(\overline{X}^* - \overline{X}\right)' G_2\left(\overline{X}\right)\left(\overline{X}^* - \overline{X}\right)\right] = E_M\left[\overline{X}^{*\prime} G_2\left(\overline{X}\right)\overline{X}^*\right] - \overline{X}' G_2\left(\overline{X}\right)\overline{X} .$$

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

$$E_M\left[\overline{X}^{*}{}'G_2\left(\overline{X}\right)\overline{X}^{*}\right] = \frac{1}{m^2}E_M\left\{\left[\sum_{i=1}^{n}X_i'M_i'\right]G_2\left(\overline{X}\right)\left[\sum_{i=1}^{n}M_iX_i\right]\right\}$$

$$= \frac{1}{m^2}E_M\left\{\sum_{i=1}^{n}X_i'M_i'G_2\left(\overline{X}\right)M_iX_i + 2\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)}X_i'M_i'G_2\left(\overline{X}\right)M_kX_k\right\}$$

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

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

$$E_M\left[M_i'G_2\left(\overline{X}\right)M_k\right] = E_M\left\{\left[\sum_{g=1}^{p}\left(c_g'\,r_g'\right)\otimes\left(M_{ig}\right)\right]\left[\sum_{s=1}^{p}\sum_{t=1}^{p}\left(r_s\,c_t\right)\otimes\left(G_2\left(\overline{X}\right)_{st}\right)\right]\left[\sum_{h=1}^{p}\left(r_h\,c_h\right)\otimes\left(M_{kh}\right)\right]\right\}$$

$$= E_M\left\{\left[\sum_{s=1}^{p}\sum_{t=1}^{p}\left(r_s\,c_t\right)\otimes\left(M_{is}\,M_{kt}\,G_2\left(\overline{X}\right)_{st}\right)\right]\right\}$$

$$= \sum_{s=1}^{p}\sum_{t=1}^{p}\left(r_s\,c_t\right)\otimes\left(\frac{m(m-1)}{n^2}G_2\left(\overline{X}\right)_{st}\right)$$

$$= \frac{m(m-1)}{n^2}\;G_2\left(\overline{X}\right).$$

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

$$E_M\left[M_i'G_2\left(\overline{X}\right)M_i\right] = \sum_{s=1}^{p}\sum_{t=1}^{p}\left(r_s\,c_t\right)\otimes\left(E_M\left[M_{is}^2\right]G_2\left(\overline{X}\right)_{st}\right)$$

$$= \sum_{s=1}^{p}\sum_{t=1}^{p}\left(r_s\,c_t\right)\otimes\left(\frac{m(n+m-1)}{n^2}G_2\left(\overline{X}\right)_{st}\right)$$

$$= \frac{m(n+m-1)}{n^2}\;G_2\left(\overline{X}\right).$$

Substituting these expressions into

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

yields the following expression.

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

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

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

In general, start with an n x 1 vector of original sample values X and randomly draw rows from X to form an m x 1 vector of bootstrapped values X^*. Define an m x n matrix H such that $X^*=HX$ where the rows of H are all zeros except for a one in the position corresponding to the element of X that was randomly drawn. If the elements of X were drawn with equal probability then the expected value of H will be $E_H[H] = (1/n)\, 1_m 1_n'$ where 1_m is an m x 1 column vector of ones and 1_n is an n x 1 column vector of ones. Consider some bootstrap statistic $\theta_m^* = g(X^*) = g(HX)$. If the first and second derivatives of $g(.)$ with respect to X^* exist, then a Taylor series expanded around some m x 1 vector X_o^* can be written as

(5.2) $$\theta_m^* = g(X_o^*) + [G_1(X_o^*)]'(X^*-X_o^*) + (1/2)\,(X^*-X_o^*)'[G_2(X_o^*)](X^*-X_o^*) + R^*$$

where $G_1(.)$ is the m x 1 vector of first derivatives, $G_2(.)$ is the m x m matrix of second derivatives and R^* is the remainder term. A natural quantity $X_o^* = H_o X$ to expand around under the bootstrap-induced distribution would be the mean value of $X^* = HX$ under that distribution which is $E_H[X^*] = E_H[H]\, X = (1/n)\, 1_m 1_n' X$. Substituting this value for X_o^* into (5.2) yields

(5.3) $$\theta_m^* = g((1/n)1_m 1_n' X) + [G_1((1/n)1_m 1_n' X)]'(H-(1/n)1_m 1_n')\, X$$
$$+ (1/2)\, X'(H-(1/n)1_m 1_n')'[G_2((1/n)1_m 1_n' X)]\,(H-(1/n)1_m 1_n')\, X + R^* .$$

As a polynomial in H, the finite sample properties of this bootstrap statistic can now be determined analytically and compared with those of alternative bootstrap statistics.

6. ANALYTICAL SOLUTION FOR BOOTSTRAP REGRESSION COEFFICIENT COVARIANCE MATRIX

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

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, $\hat{\beta} = \left(X'X\right)^{-1} X'Y$, 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 $e = (I_n - X (X'X)^{-1} X')\in$ and *X* is considered to be a nonrandom matrix, to provide a set of bootstrapped residuals $e^* = \{ e_1^*, e_2^*, \ldots, e_n^* \}'$ that are centered around zero we introduce the *n* x *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 *Assumption 4* as a set of procedural assumptions and definitions that characterize this least squares regression bootstrap.

Assumption 4:

Take n draws with replacement from the set of least squares residuals $e = \{ e_1, e_2, \ldots, e_n \}'$ *to obtain a corresponding set of bootstrapped residuals* $e^* = \{ e_1^*, e_2^*, \ldots, e_n^* \}'$. *For* $Y^* = X\hat{\beta} + Ae^*$ *define the bootstrap regression estimator as* $\hat{\beta}^* = \left(X'X\right)^{-1} X'Y^*$.

Define an n x n matrix, H, with elements, $\{H_{ij} : i,j=1,\ldots,n\}$, such that $e^* = He$, where the element in the i^{th} row and j^{th} column of H is equal to one if the j^{th} element of e is randomly drawn (with replacement) on the i^{th} draw and is equal to zero otherwise. For notation, I_n is an n x n identity matrix, 1_n is an n x 1 column vector of all ones, 1_{n^2} is an n^2 x 1 column vector of all ones, "*vec*" is the standard vectorization operator, "$\otimes$" means Kronecker product, and the n x 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 $A\, 1_n1_n' = 0 = 1_n1_n' A$ and the last term drops out.

Any expected value under the randomness that generated the original sample, $e_i : i = 1,\ldots,n\}$, will be denoted by the subscript $\in$, as in $E_{\in}[\cdot]$ and covariance, $Cov_{\in}[\cdot,\cdot]$. 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 $\in$ and H, as in $E_{\in,H}[\cdot]$ and covariance, $Cov_{\in,H}[\cdot,\cdot]$. Conditional expectations will be denoted $E_{H|\in}[\cdot]$ and covariance, $Cov_{H|\in}[\cdot,\cdot]$.

Under *Assumption 4*, since each of the n draws will take place for sure, the relevant distribution is the conditional distribution of drawing j on the ith 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 $Var_{j|i}(H_{ij}) = (n-1)/n^2$. Given an n x 1 <u>column</u> vector r_i whose elements are all zeros except for a one in its ith "row" position, and a 1 x n <u>row</u> vector c_j also made up of all zeros except for a one in its jth "column" position, we have $H = \sum_{i=1}^{n}\sum_{j=1}^{n} r_i c_j H_{ij}$, $E_{H|\in}(H) = \sum_{i=1}^{n}\sum_{j=1}^{n} r_i c_j E_{j|i,\in}(H_{ij})$ and $E_{j|i,\in}(H_{ij}) = \frac{1}{n}$ so that $E_{H|\in}(H) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} r_i c_j = \frac{1}{n} 1_n 1_n'$ where 1_n is an n x 1 column vector of all ones. Since $e^*=He$, it follows that $E_{H|\in}(e^*) = \frac{1}{n} 1_n 1_n' e$. Since $A = (I_n - (1/n)1_n1_n')$, we have $A1_n1_n' = 0$ and, therefore, $E_{H|\in}(Ae^*) = 0$. Also, we have $1_n1_n'A = 0$ and, therefore, $E_{H|\in}(HAe) = 0$ so centering the original residuals is just as effective as centering the bootstrapped residuals.

Theorem 7:

Under Assumption 4,

$$Cov_{H|\in}\left(\hat{\beta}^*\right) = (X'X)^{-1}X'A\left\{\frac{1}{n}\left[I_n \otimes (e'e)\right] + \frac{1}{n^2}\left[\left(1_n1_n' - I_n\right)\otimes\left(1_{n^2}' \, vec(ee')\right)\right]\right\}AX\,(X'X)^{-1}$$

$$-\frac{1}{n^2}(X'X)^{-1}X'A\,1_n1_n'\,ee'\,1_n1_n'\,AX\,(X'X)^{-1}.$$

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 Assumption 4,

$$Cov_{\in,H}\left(\hat{\beta}^*\right) = Cov_{\in}\left(\hat{\beta}\right) + (X'X)^{-1}X'A\,Cov_{\in}(\in)\left(I_n - X\,(X'X)^{-1}X'\right)\left(\frac{1}{n}1_n1_n'\right)AX\,(X'X)^{-1}$$

$$+ (X'X)^{-1}X'A\left(\frac{1}{n}1_n1_n'\right)\left(I_n - X\,(X'X)^{-1}X'\right)Cov_{\in}(\in)\,AX\,(X'X)^{-1}$$

$$+ (X'X)^{-1}X'A\left\{\frac{1}{n}\left[I_n \otimes tr(\Phi)\right] + \frac{1}{n^2}\left[\left(1_n1_n' - I_n\right)\otimes\left(1_{n^2}' \, vec(\Phi)\right)\right]\right\}AX\,(X'X)^{-1}$$

$$- (X'X)^{-1}X'A\left(\frac{1}{n}1_n1_n'\right)\left(I_n - X\,(X'X)^{-1}X'\right)E_{\in}(\in)E_{\in}(\in')\left(I_n - X\,(X'X)^{-1}X'\right)\left(\frac{1}{n}1_n1_n'\right)AX\,(X'X)^{-1}$$

where $\Phi = \left(I_n - X\,(X'X)^{-1}X'\right)E_{\in}(\in\in')\left(I_n - X\,(X'X)^{-1}X'\right).$

If the bootstrapped residuals are not centered, this provides a general result for the linear model for any variance-covariance matrix, $Cov_{\in}(\in)$, for the population errors. However, with $A = (I_n - (1/n)1_n1_n')$ the bootstrapped residuals sum to zero, and this reduces down to

$$Cov_{\in,H}\left(\hat{\beta}^*\right) = Cov_{\in}\left(\hat{\beta}\right) + (X'X)^{-1}X'\left\{\left(I_n - \frac{1}{n}1_n1_n'\right)\otimes\left(\frac{1}{n}tr(\Phi) - \frac{1}{n^2}1_{n^2}'\,vec(\Phi)\right)\right\}X\,(X'X)^{-1}.$$

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. Also, we provided a general formula for expressing the Taylor series expansion of any nonlinear bootstrap statistic which would include bootstrap statistics in the form of ratios (e.g. likelihood ratio or t-statistics) or matrix inverses (e.g. Wald statistics, J-statistics, Lagrange multiplier statistics) of the bootstrap sample values. 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 statistics can be analyzed in this manner to determine their bias, variance, mean squared error, skewness, kurtosis and many other finite sample properties and characteristics.

APPENDIX

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

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

$$= \frac{1}{m}\sum_{i=1}^{n}\left[\sum_{M_1+...+M_n=m} M_i\frac{n^{-m}m!}{(M_1!...M_n!)}\right] f\left(X_i\right) = \frac{1}{m}\sum_{i=1}^{n}\left[E_M\left(M_i\right)\right] f\left(X_i\right) = \frac{1}{m}\sum_{i=1}^{n}\left[\frac{m}{n}\right] f\left(X_i\right) = \frac{1}{n}\sum_{i=1}^{n} f\left(X_i\right)$$

.

Q.E.D.

Proof of ***Lemma 2*:**

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

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

$$E_M\left[\left(\frac{1}{m}\sum_{j=1}^{m} f\left(X_j^*\right)\right)^2\right] = \sum_{M_1+...+M_n=m}\left(\frac{1}{m}\sum_{j=1}^{m} f\left(X_j^*\right)\right)^2\frac{n^{-m}m!}{(M_1!...M_n!)}$$

$$= \sum_{M_1+...+M_n=m}\left(\frac{1}{m}\sum_{i=1}^{n} M_i\ f\left(X_i\right)\right)^2\frac{n^{-m}m!}{(M_1!...M_n!)}$$

$$= \sum_{M_1+...+M_n=m}\left(\frac{1}{m^2}\sum_{i=1}^{n} M_i^2\left(f\left(X_i\right)\right)^2 + \frac{2}{m^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)} M_i M_k\ f\left(X_i\right) f\left(X_k\right)\right)\frac{n^{-m}m!}{(M_1!...M_n!)}$$

$$= \frac{1}{m^2}\sum_{i=1}^{n}\left[\sum_{M_1+\ldots+M_n=m} M_i^2 \frac{n^{-m}m!}{(M_1!\ldots M_n!)}\right]\left(f(X_i)\right)^2 + \frac{2}{m^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)}\left[\sum_{M_1+\ldots+M_n=m} M_i M_k \frac{n^{-m}m!}{(M_1!\ldots M_n!)}\right] f(X_i) f(X_k)$$

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

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

$$= \frac{(m+n-1)}{m\,n^2}\sum_{i=1}^{n}\left(f(X_i)\right)^2 + \frac{2(m-1)}{m\,n^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)} f(X_i) f(X_k) \quad . \qquad \textit{Q.E.D.}$$

Proof of ***Theorem 1*:**

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

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

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

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

Proof of ***Theorem 2*:**

Apply the definition of variance in this context as

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

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

$$E_{M,X}\left[\left(\frac{1}{m}\sum_{j=1}^{m} f\left(X_j^*\right)\right)^2\right] \;=\; E_{M,X}\left[\left(\frac{1}{m}\sum_{i=1}^{n} M_i\, f\left(X_i\right)\right)^2\right]$$

$$= \;E_{M,X}\left[\frac{1}{m^2}\sum_{i=1}^{n} M_i^2\left(f\left(X_i\right)\right)^2 + \frac{2}{m^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)} M_i M_k\, f\left(X_i\right) f\left(X_k\right)\right]$$

$$= \;\frac{1}{m^2}\sum_{i=1}^{n}\left[E_M\left(M_i^2\right)\right]E_X\left[\left(f\left(X_i\right)\right)^2\right] + \frac{2}{m^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)}\left[E_M\left(M_i\, M_k\right)\right]E_X\left[f\left(X_i\right)f\left(X_k\right)\right]$$

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

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

$$\left[E_{M,X}\left(\frac{1}{m}\sum_{j=1}^{m} f\left(X_j^*\right)\right)\right]^2 = \left[E_{M,X}\left(\frac{1}{m}\sum_{i=1}^{n} M_i\, f\left(X_i\right)\right)\right]^2 = \left[\frac{1}{n}\sum_{i=1}^{n} E_X\left[f\left(X_i\right)\right]\right]^2$$

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

Now substitute in to the formula for the variance.

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

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

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

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

Q.E.D.

Proof of ***Theorem 3*:**

We have established by *Lemma 3* that the mean is zero so the second term below for the variance (the negative term) will be zero.

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

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

$$=E_M E_{W|M}\left[\frac{1}{m^2}\sum_{i=1}^{n}(2W_i-M_i)^2\left[f(X_i)\right]^2+\frac{2}{m^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)}(2W_i-M_i)(2W_k-M_k)f(X_i)f(X_k)\right]$$

$$=E_M E_{W|M}\left[\frac{1}{m^2}\sum_{i=1}^{n}\left(4W_i^2+M_i^2-4M_iW_i\right)\left[f(X_i)\right]^2+\frac{2}{m^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)}\left(4W_iW_k-2W_iM_k-2M_iW_k+M_iM_k\right)f(X_i)f(X_k)\right]$$

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

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

The following expected values follow immediately:

(A2) $E_{W/M}W_i = 0.5M_i$, $E_{W|M}W_i^2 = 0.25M_i(M_i+1)$, and $E_{W|M}W_iW_k = 0.25M_iM_k$.

$$Var_{W,M}\left[\left(\frac{1}{m}\sum_{j=1}^{m} f\left(X_j^*\right)\right)\right] = E_M\left[\frac{1}{m^2}\sum_{i=1}^{n}\left(4E_{W|M}\left(W_i^2\right)+M_i^2-4M_iE_{W|M}\left(W_i\right)\right)\left[f\left(X_i\right)\right]^2\right]$$

$$+E_M\left[\frac{2}{m^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)}\left(4E_{W|M}\left(W_iW_k\right)-2E_{W|M}\left(W_i\right)M_k-2M_iE_{W|M}\left(W_k\right)+M_iM_k\right)f\left(X_i\right)f\left(X_k\right)\right]$$

$$Var_{W,M}\left[\left(\frac{1}{m}\sum_{j=1}^{m} f\left(X_j^*\right)\right)\right] = E_M\left[\frac{1}{m^2}\sum_{i=1}^{n}\left(M_i^2+M_i+M_i^2-2M_i^2\right)\left[f\left(X_i\right)\right]^2\right]$$

$$+E_M\left[\frac{2}{m^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)}\left(M_iM_k-M_iM_k-M_iM_k+M_iM_k\right)f\left(X_i\right)f\left(X_k\right)\right]$$

$$Var_{W,M}\left[\left(\frac{1}{m}\sum_{j=1}^{m} f\left(X_j^*\right)\right)\right] = E_M\left[\frac{1}{m^2}\sum_{i=1}^{n}M_i\left[\left(f\left(X_i\right)\right)^2\right]\right]$$

$$= \frac{1}{m^2}\sum_{i=1}^{n}E_M\left(M_i\right)\left[\left(f\left(X_i\right)\right)^2\right] = \frac{1}{m\,n}\sum_{i=1}^{n}\left[\left(f\left(X_i\right)\right)^2\right].$$ *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.

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

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

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

$$= E_M E_{W|M}\left[\frac{1}{m^2}\sum_{i=1}^{n}\left(4W_i^2 + M_i^2 - 4M_i W_i\right) E_X\left[\left(f\left(X_i\right)\right)^2\right]\right]$$

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

Now consider the binomial distribution for W_i conditional on M_i.

The probability of W_i successes (positive ones) is:

(A3) $$P\left(W_i \mid M_i\right) = \binom{M_i}{W_i} 0.5^{W_i}\left(1 - 0.5\right)^{M_i - W_i}$$

The following expected values follow immediately:

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

$$Var_{W,M,X}\left[\left(\frac{1}{m}\sum_{j=1}^{m} f\left(X_j^*\right)\right)\right] = E_M\left[\frac{1}{m^2}\sum_{i=1}^{n}\left(4E_{W|M}\left(W_i^2\right) + M_i^2 - 4M_i E_{W|M}\left(W_i\right)\right) E_X\left[\left(f\left(X_i\right)\right)^2\right]\right]$$

$$+ \quad E_M\left[\frac{2}{m^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)}\left(4E_{W|M}\left(W_i W_k\right) - 2E_{W|M}\left(W_i\right) M_k - 2M_i E_{W|M}\left(W_k\right) + M_i M_k\right) E_X\left[f\left(X_i\right) f\left(X_k\right)\right]\right]$$

$$Var_{W,M,X}\left[\left(\frac{1}{m}\sum_{j=1}^{m} f\left(X_j^*\right)\right)\right] = E_M\left[\frac{1}{m^2}\sum_{i=1}^{n}\left(M_i^2 + M_i + M_i^2 - 2M_i^2\right) E_X\left[\left(f\left(X_i\right)\right)^2\right]\right]$$

$$+ E_M\left[\frac{2}{m^2}\sum_{i<k}^{\left(\frac{n^2-n}{2}\right)}\left(M_i M_k - M_i M_k - M_i M_k + M_i M_k\right) E_X\left[f\left(X_i\right) f\left(X_k\right)\right]\right]$$

$$Var_{W,M,X}\left[\left(\frac{1}{m}\sum_{j=1}^{m} f\left(X_j^*\right)\right)\right] = E_M\left[\frac{1}{m^2}\sum_{i=1}^{n} M_i\, E_X\left[\left(f\left(X_i\right)\right)^2\right]\right]$$

$$= \frac{1}{m^2}\sum_{i=1}^{n} E_M\left(M_i\right)E_X\left[\left(f\left(X_i\right)\right)^2\right] = \frac{1}{m\,n}\sum_{i=1}^{n} E_X\left[\left(f\left(X_i\right)\right)^2\right] \qquad Q.E.D.$$

Proof of ***Theorem 5*:**

Start with a common representation for the variance:

$$Var_M\left[\frac{1}{b}\sum_{j=1}^{b}\theta\left(B_j^*\right)\right] = Var_M\left[\frac{1}{b}\sum_{g=1}^{k} M_g\,\theta\left(B_g\right)\right] = E_M\left\{\left[\frac{1}{b}\sum_{g=1}^{k} M_g\,\theta\left(B_g\right)\right]^2\right\} - \left\{E_M\left[\frac{1}{b}\sum_{g=1}^{k} M_g\,\theta\left(B_g\right)\right]\right\}^2$$

$$= \frac{1}{b^2}\sum_{g=1}^{k} E_M\left(M_g^2\right)\left[\left(\theta\left(B_g\right)\right)^2\right] + \frac{2}{b^2}\sum_{g<h}^{\left(\frac{k^2-k}{2}\right)} E_M\left(M_g M_h\right)\theta\left(B_g\right)\theta\left(B_h\right)$$

$$- \frac{1}{b^2}\sum_{g=1}^{k}\left(E_M\left(M_g\right)\right)^2\left[\left(\theta\left(B_g\right)\right)^2\right] - \frac{2}{b^2}\sum_{g<h}^{\left(\frac{k^2-k}{2}\right)} E_M\left(M_g\right)E_M\left(M_h\right)\theta\left(B_g\right)\theta\left(B_h\right)$$

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

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

$$- \frac{1}{b^2}\sum_{g=1}^{k}\left(\frac{b}{k}\right)^2\left[\left(\theta\left(B_g\right)\right)^2\right] - \frac{2}{b^2}\sum_{g<h}^{\left(\frac{k^2-k}{2}\right)}\left(\frac{b}{k}\right)\left(\frac{b}{k}\right)\theta\left(B_g\right)\theta\left(B_h\right)$$

which, in turn, reduces to:

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

Proof of ***Theorem 6*:**

Start with a standard formula for variance.

$$Var_{M,B}\left[\frac{1}{b}\sum_{j=1}^{b}\theta\left(B_j^*\right)\right] = Var_{M,B}\left[\frac{1}{b}\sum_{g=1}^{k}M_g\,\theta\left(B_g\right)\right]$$

$$= E_{M,B}\left\{\left[\frac{1}{b}\sum_{g=1}^{k}M_g\,\theta\left(B_g\right)\right]^2\right\}-\left\{E_{M,B}\left[\frac{1}{b}\sum_{g=1}^{k}M_g\,\theta\left(B_g\right)\right]\right\}^2$$

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.

$$Var_{M,B}\left[\frac{1}{b}\sum_{j=1}^{b}\theta\left(B_j^*\right)\right] = \frac{1}{b^2}\sum_{g=1}^{k}E_M\left(M_g^2\right)E_B\left[\left(\theta\left(B_g\right)\right)^2\right] + \frac{2}{b^2}\sum_{g<h}^{\left(\frac{k^2-k}{2}\right)}E_M\left(M_gM_h\right)E_B\left[\theta\left(B_g\right)\theta\left(B_h\right)\right]$$

$$- \frac{1}{b^2}\sum_{g=1}^{k}\left(E_M\left(M_g\right)\right)^2\left[\left(E_B\left[\theta\left(B_g\right)\right]\right)^2\right] - \frac{2}{b^2}\sum_{g<h}^{\left(\frac{k^2-k}{2}\right)}E_M\left(M_g\right)E_M\left(M_h\right)E_B\left[\theta\left(B_g\right)\right]E_B\left[\theta\left(B_h\right)\right].$$

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

$$Var_{M,B}\left[\frac{1}{b}\sum_{j=1}^{b}\theta\left(B_j^*\right)\right] = \frac{1}{b^2}\sum_{g=1}^{k}\frac{b(b-1)+kb}{k^2}E_B\left[\left(\theta\left(B_g\right)\right)^2\right] + \frac{2}{b^2}\sum_{g<h}^{\left(\frac{k^2-k}{2}\right)}\frac{b(b-1)}{k^2}E_B\left[\theta\left(B_g\right)\theta\left(B_h\right)\right]$$

$$- \frac{1}{b^2}\sum_{g=1}^{k}\left(\frac{b}{k}\right)^2\left[\left(E_B\left[\theta\left(B_g\right)\right]\right)^2\right] - \frac{2}{b^2}\sum_{g<h}^{\left(\frac{k^2-k}{2}\right)}\left(\frac{b}{k}\right)\left(\frac{b}{k}\right)E_B\left[\theta\left(B_g\right)\right]E_B\left[\theta\left(B_h\right)\right].$$

which, in turn, reduces to:

$$Var_{M,B}\left[\frac{1}{b}\sum_{j=1}^{b}\theta\left(B_j^*\right)\right] = \frac{b+k-1}{b\ k^2}\sum_{g=1}^{k}E_B\left[\left(\theta\left(B_g\right)\right)^2\right] + \frac{2\left(b-1\right)}{b\ k^2}\sum_{g<h}^{\left(\frac{k^2-k}{2}\right)}E_B\left[\theta\left(B_g\right)\theta\left(B_h\right)\right]$$

$$- \frac{1}{k^2}\sum_{g=1}^{k}\left[\left(E_B\left[\theta\left(B_g\right)\right]\right)^2\right] - \frac{2}{k^2}\sum_{g<h}^{\left(\frac{k^2-k}{2}\right)}E_B\left[\theta\left(B_g\right)\right]E_B\left[\theta\left(B_h\right)\right]$$

$$= \frac{1}{k^2}\sum_{g=1}^{k} Var_B\left[\theta\left(B_g\right)\right] + \frac{k-1}{b\ k^2}\sum_{g=1}^{k} E_B\left[\left(\theta\left(B_g\right)\right)^2\right]$$

$$+ \frac{2}{k^2}\sum_{g<h}^{\left(\frac{k^2-k}{2}\right)} Cov_B\left[\theta\left(B_g\right), \theta\left(B_h\right)\right] - \frac{2}{b\ k^2}\sum_{g<h}^{\left(\frac{k^2-k}{2}\right)} E_B\left[\theta\left(B_g\right)\theta\left(B_h\right)\right]. \qquad Q.E.D.$$

Proof of ***Theorem 7*****:**

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

$$E_{H|\in}\hat{\beta}^* = E_{H|\in}\hat{\beta} + E_{H|\in}\left(X'X\right)^{-1} X'Ae^*$$

$$= \hat{\beta} + E_{H|\in}\left(X'X\right)^{-1} X'AH\,e$$

$$= \hat{\beta} + \left(X'X\right)^{-1} X'AE_{H|\in}\left(H\right) e$$

$$= \hat{\beta} + \frac{1}{n}\left(X'X\right)^{-1} X'A\,1_n 1_n'\ e$$

Consequently, we have

$$Cov_{H|\in}\left(\hat{\beta}^*\right) = E_{H|\in}\left\{\left(\hat{\beta}^* - E_{H|\in}\hat{\beta}^*\right)\left(\hat{\beta}^* - E_{H|\in}\hat{\beta}^*\right)'\right\}$$

$$= E_{H|\in}\left\{\left(X'X\right)^{-1} X'A\left(H - \frac{1}{n}1_n 1_n'\right) ee'\left(H' - \frac{1}{n}1_n 1_n'\right) AX\left(X'X\right)^{-1}\right\}.$$

This expression expands to become

$$Cov_{H|\in}\left(\hat{\beta}^*\right) = \left(X'X\right)^{-1} X'AE_{H|\in}\left(Hee'H'\right)AX\left(X'X\right)^{-1} - \frac{1}{n}\left[\left(X'X\right)^{-1} X'AE_{H|\in}\left(H\right)ee'1_n 1_n' AX\left(X'X\right)^{-1}\right]$$

$$- \frac{1}{n}\left[\left(X'X\right)^{-1} X'A1_n 1_n' ee'E_{H|\in}\left(H'\right)AX\left(X'X\right)^{-1}\right] + \frac{1}{n^2}\left[\left(X'X\right)^{-1} X'A1_n 1_n'\ ee'1_n 1_n' AX\left(X'X\right)^{-1}\right].$$

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

$$Cov_{H|\in}\left(\hat{\beta}^{*}\right) = \left(X'X\right)^{-1} X'AE_{H|\in}\left(Hee'H'\right)AX\left(X'X\right)^{-1} - \frac{1}{n^{2}}\left[\left(X'X\right)^{-1} X'A1_{n}1_{n}' ee'1_{n}1_{n}' AX\left(X'X\right)^{-1}\right]$$

.

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 $\in$. The *Hee'H'* matrix can be decomposed as follows:

$$H\,ee'H' = \left[\sum_{i=1}^{n}\sum_{j=1}^{n}\left(v_{i}\right)\otimes\left(H_{ij}\,e_{j}\right)\right]\left[\sum_{k=1}^{n}\sum_{l=1}^{n}\left(v_{k}'\right)\otimes\left(e_{l}\,H_{kl}\right)\right]$$

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

$$H\,ee'H' = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\left(v_{i}v_{k}'\right)\otimes\left(e_{j}\,e_{l}\,H_{ij}\,H_{kl}\right)$$

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

$$E_{H|\in}\left(H\,ee'H'\right) = E_{H|\in}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\left(v_{i}v_{i}'\right)\otimes\left(e_{j}^{2}\,H_{ij}^{2}\right)\right] + E_{H|\in}\left[\sum_{i\neq k}\sum_{j=1}^{n}\sum_{l=1}^{n}\left(v_{i}v_{k}'\right)\otimes\left(e_{j}\,e_{l}\,H_{ij}\,H_{kl}\right)\right]$$

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

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

$$E_{H|\in}\left(H\,ee'H'\right) = \frac{1}{n}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\left(v_{i}v_{i}'\right)\otimes\left(e_{j}^{2}\right)\right] + \frac{1}{n^{2}}\left[\sum_{i\neq k}\sum_{j=1}^{n}\sum_{l=1}^{n}\left(v_{i}v_{k}'\right)\otimes\left(e_{j}\,e_{l}\right)\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}\left(v_{i}v_{i}'\right)\otimes\left(\sum_{j=1}^{n}e_{j}^{2}\right)\right] + \frac{1}{n^{2}}\left[\sum_{i\neq k}\left(v_{i}v_{k}'\right)\otimes\left(\sum_{j=1}^{n}\sum_{l=1}^{n}e_{j}\,e_{l}\right)\right]$$

$$= \frac{1}{n}\left[I_{n}\otimes\left(e'e\right)\right] + \frac{1}{n^{2}}\left[\left(1_{n}1_{n}' - I_{n}\right)\otimes\left(1_{n^{2}}'\,vec\left(ee'\right)\right)\right]$$

where I_n is an n x n identity matrix, 1_n is an n x 1 column vector of all ones, 1_{n^2} is an n^2 x 1 column vector of all ones, and "vec" is the standard vectorization operator.

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

$$Cov_{H|\in}\left(\hat{\beta}^*\right) = \left(X'X\right)^{-1} X'A\left\{\frac{1}{n}\left[I_n \otimes (e'e)\right] + \frac{1}{n^2}\left[\left(1_n 1_n' - I_n\right) \otimes \left(1_{n^2}' \, vec(ee')\right)\right]\right\} AX\left(X'X\right)^{-1}$$

$$- \frac{1}{n^2}\left(X'X\right)^{-1} X'A 1_n 1_n' \; ee' \, 1_n 1_n' AX\left(X'X\right)^{-1}. \qquad \textit{Q.E.D.}$$

Proof of ***Theorem 8*:**

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

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

so that $\hat{\beta}^* = \hat{\beta} + \left(X'X\right)^{-1} X'A\,H\left(I_n - X\left(X'X\right)^{-1} X'\right)\in.$

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

$$E_{\in,H}\hat{\beta}^* = E_{\in,H}\hat{\beta} + E_{\in,H}\left(X'X\right)^{-1} X'A\,e^*$$

$$= E_{\in,H}\hat{\beta} + E_{\in,H}\left(X'X\right)^{-1} X'AH\,e$$

$$= E_{\in}\hat{\beta} + E_H E_{\in|H}\left(X'X\right)^{-1} X'AH\left(I_n - X\left(X'X\right)^{-1} X'\right)\in$$

$$= E_{\in}\hat{\beta} + \left(X'X\right)^{-1} X'AE_H(H)\left(I_n - X\left(X'X\right)^{-1} X'\right)E_{\in}(\in) .$$

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.

$$E_{\in,H}\hat{\beta}^{*} \;=\; E_{\in}\hat{\beta} \;+\; \frac{1}{n}(X'X)^{-1}X'A\;1_n1_n'\left(I_n - X(X'X)^{-1}X'\right)E_{\in}(\in)\;.$$

Therefore, given $\hat{\beta}^{*} = \hat{\beta} + (X'X)^{-1}X'A\,H\left(I_n - X(X'X)^{-1}X'\right)\in$, it follows that

$$\left(\hat{\beta}^{*} - E_{\in,H}\hat{\beta}^{*}\right) \;=\; \left(\hat{\beta} - E_{\in}\hat{\beta}\right) \;+\; (X'X)^{-1}X'A\,H\left(I_n - X(X'X)^{-1}X'\right)\in$$

$$-\;(X'X)^{-1}X'A\left(\frac{1}{n}1_n1_n'\right)\left(I_n - X(X'X)^{-1}X'\right)E_{\in}(\in)$$

so we have

$$\left(\hat{\beta}^{*} - E_{\in,H}\hat{\beta}^{*}\right)\left(\hat{\beta}^{*} - E_{\in,H}\hat{\beta}^{*}\right)'$$

$$= \left(\hat{\beta} - E_{\in}\hat{\beta}\right)\left(\hat{\beta} - E_{\in}\hat{\beta}\right)' + \left(\hat{\beta} - E_{\in}\hat{\beta}\right)\in'\left(I_n - X(X'X)^{-1}X'\right)H'A\,X(X'X)^{-1}$$

$$-\left(\hat{\beta} - E_{\in}\hat{\beta}\right)E_{\in}(\in')\left(I_n - X(X'X)^{-1}X'\right)\left(\frac{1}{n}1_n1_n'\right)AX(X'X)^{-1}$$

$$+\quad (X'X)^{-1}X'A\,H\left(I_n - X(X'X)^{-1}X'\right)\in\left(\hat{\beta} - E_{\in}\hat{\beta}\right)'$$

$$-\quad (X'X)^{-1}X'A\left(\frac{1}{n}1_n1_n'\right)\left(I_n - X(X'X)^{-1}X'\right)E_{\in}(\in)\left(\hat{\beta} - E_{\in}\hat{\beta}\right)'$$

$$+\;(X'X)^{-1}X'A\,H\left(I_n - X(X'X)^{-1}X'\right)\in\in'\left(I_n - X(X'X)^{-1}X'\right)H'A\,X\,(X'X)^{-1}$$

$$-\;(X'X)^{-1}X'AH\left(I_n - X(X'X)^{-1}X'\right)\in E_{\in}(\in')\left(I_n - X(X'X)^{-1}X'\right)\left(\frac{1}{n}1_n1_n'\right)AX\,(X'X)^{-1}$$

$$-\;(X'X)^{-1}X'A\left(\frac{1}{n}1_n1_n'\right)\left(I_n - X(X'X)^{-1}X'\right)E_{\in}(\in)\in'\left(I_n - X(X'X)^{-1}X'\right)H'AX\,(X'X)^{-1}$$

$$+\;(X'X)^{-1}X'A\left(\frac{1}{n}1_n1_n'\right)\left(I_n - X(X'X)^{-1}X'\right)E_{\in}(\in)E_{\in}(\in')\left(I_n - X(X'X)^{-1}X'\right)\left(\frac{1}{n}1_n1_n'\right)AX\,(X'X)^{-1}$$

.

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

$$Cov_{\in,H}\left(\hat{\beta}^*\right) \; = \; E_{\in,H}\left\{\left(\hat{\beta}^* - E_{\in,H}\hat{\beta}^*\right)\left(\hat{\beta}^* - E_{\in,H}\hat{\beta}^*\right)'\right\}$$

$$Cov_{\in,H}\left(\hat{\beta}^*\right) \; = \; Cov_{\in}\left(\hat{\beta}\right) \; + \; \left(X'X\right)^{-1}X'A\,E_H\left(H\,\Phi\,H'\right)AX\left(X'X\right)^{-1}$$

$$+ \; \left(X'X\right)^{-1}X'A\,Cov_{\in}(\in)\left(I_n - X\left(X'X\right)^{-1}X'\right)\left(\frac{1}{n}1_n1_n'\right)AX\left(X'X\right)^{-1}$$

$$+ \; \left(X'X\right)^{-1}X'A\left(\frac{1}{n}1_n1_n'\right)\left(I_n - X\left(X'X\right)^{-1}X'\right)Cov_{\in}(\in)\,AX\left(X'X\right)^{-1}$$

$$- \; \left(X'X\right)^{-1}X'A\left(\frac{1}{n}1_n1_n'\right)\left(I_n - X\left(X'X\right)^{-1}X'\right)E_{\in}(\in)E_{\in}(\in')\left(I_n - X\left(X'X\right)^{-1}X'\right)\left(\frac{1}{n}1_n1_n'\right)AX\left(X'X\right)^{-1}$$

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

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

Given an n x 1 *column* vector r_i whose elements are all zeros except for a one in its i^{th} "row" position, and a 1 x n *row* vector c_j also made up of all zeros except for a one in its j^{th} "column" position, we have $H = \sum_{i=1}^{n}\sum_{j=1}^{n} r_i\,c_j\,H_{ij}$, $E_{H|\in}\left(H\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} r_i\,c_j\,E_{j|i,\in}\left(H_{ij}\right)$ and

$E_{j|i,\in}\left(H_{ij}\right) = \dfrac{1}{n}$ so that $E_{H|\in}\left(H\right) = \dfrac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} r_i\,c_j = \dfrac{1}{n}\;1_n\,1_n'$ where 1_n is an n x 1 column vector of all ones. In order to determine $E_H\left(H\,\Phi\,H'\right)$, we need to decompose the *H*, the Φ and the *H'* matrices as follows:

$$\begin{aligned} H\,ee'H' \; &= \; \left[\sum_{i=1}^{n}\sum_{j=1}^{n}\left(r_i\,c_j\right)\otimes\left(H_{ij}\right)\right]\left[\sum_{k=1}^{n}\sum_{l=1}^{n}\left(r_k\,c_l\right)\otimes\left(\Phi_{kl}\right)\right]\left[\sum_{s=1}^{n}\sum_{t=1}^{n}\left(c_t'r_s'\right)\otimes\left(H_{st}\right)\right] \\ &= \; \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{s=1}^{n}\sum_{t=1}^{n}\left(r_i c_j\; r_k c_l\; c_t' r_s'\right)\otimes\left(H_{ij}\,\Phi_{kl}\,H_{st}\right) . \end{aligned}$$

$$H\,ee'H' \;=\; \sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{i=s}\sum_{j=t}\left(r_i c_j\; r_k c_l\; c_t' r_s'\right)\otimes\left(\Phi_{kl}\; H_{ij} H_{st}\right)$$
$$+\;\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{i=s}\sum_{j\neq t}\left(r_i c_j\; r_k c_l\; c_t' r_s'\right)\otimes\left(\Phi_{kl}\; H_{ij}\; H_{st}\right)$$
$$+\;\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{i\neq s}\sum_{j=t}\left(r_i c_j\; r_k c_l\; c_t' r_s'\right)\otimes\left(\Phi_{kl}\; H_{ij}\; H_{st}\right)$$
$$+\;\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{i\neq s}\sum_{j\neq t}\left(r_i c_j\; r_k c_l\; c_t' r_s'\right)\otimes\left(\Phi_{kl}\; H_{ij}\; H_{st}\right).$$

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 Φ matrix serving as weights.

The second term is always zero, because on the i^{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^{th} draw are zeros. Therefore, any product of two elements of *H* for the same draw (i.e. the i^{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 *HΦH'* 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 *HΦH'* may be re-expressed as

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

where the first term of this new expression represents the diagonal elements of *HΦH'* and the second term represents the off-diagonal elements.

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

Therefore,

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

where “tr” means trace, “vec” is the usual vectorization operator, and

$\Phi = \left(I_n - X\left(X'X\right)^{-1}X'\right)E_\in\left(\in\in'\right)\left(I_n - X\left(X'X\right)^{-1}X'\right)$.

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

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

where $\Phi = \left(I_n - X\left(X'X\right)^{-1}X'\right)E_\in\left(\in\in'\right)\left(I_n - X\left(X'X\right)^{-1}X'\right)$. *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 Statistic 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.