CHAPTER 4

Regression Analysis

This chapter is divided into three parts and a historical notes section. Section 4.1 deals with linear regression and Section 4.2 deals with the nonlinear regression problems. Section 4.3 deals with nonparametric regression models. In Section 4.4 we provide historical notes regarding the development of the bootstrap procedures in both the linear and nonlinear cases.

In Section 4.1.1 we will briefly review the well-known Gauss–Markov theory, which applies to least-squares estimation in the linear regression problem. A natural question for the practitioner is to ask "Why bootstrap in the linear regression case? Isn't least-squares a well-established approach that has served us well in countless applications?" The answer is that for many problems, least-squares regression has served us well and is always useful as a first approach but is problematic when the residuals have heavy-tailed distributions or if even just a few outliers are present.

The difficulty is that in some applications, certain key assumptions may be violated. These assumptions are as follows: (1) The error term in the model has a probability distribution that is the same for each observation and does not depend on the predictor variables (i.e., independence and homoscedasticity); (2) the predictor variables are observed without error; and (3) the error term has a finite variance.

Under these three assumptions, the least-squares procedure provides the best linear unbiased estimate of the regression parameters. However, if assumption 1 is violated because the variance of the residuals varies as the predictor variables change, a weighted least-squares approach may be more appropriate.

The strongest case for least-squares estimation can be made when the error term has a Gaussian or approximately a Gaussian distribution. Then the theory of maximum likelihood also applies and confidence intervals and

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition,
by Michael R. Chernick

hypothesis tests for the parameters can be applied using the standard theory and the standard statistical packages.

However, if the error distribution is non-Gaussian and particularly if the error distribution is heavy-tailed, least-squares estimation may not be suitable (robust regression methods may be better). When the error distribution is non-Gaussian, regardless of what estimation procedure is used, it is difficult to determine confidence intervals for the parameters or to obtain prediction intervals for the response variable.

This is where the bootstrap can help, and we will illustrate it for both the linear and nonlinear cases. In the nonlinear case, even standard errors for the estimates are not easily obtained, but bootstrap estimates are fairly straightforward.

There are two basic approaches to bootstrapping in the regression problem. One is to first fit the model and bootstrap the residuals. The other is to bootstrap the vector of the response variables and the associated predictor variable. Bootstrapping the residuals requires that the residuals be independent and identically distributed (or at least exchangeable).

In a quasi-optical experiment (Shimabukuro, Lazar, Dyson, and Chernick, 1984), I used the bootstrap to estimate the standard errors for two of the parameters in the nonlinear regression model. Results are discussed in Section 4.2.2. The residuals appear to be correlated with the incident angle of the measurement. This invalidates the exchangeability assumption, but how does it affect the standard errors of the parameters?

Our suspicion is that bootstrapping the residuals makes the bootstrap sample more variable and consequently biases the estimated standard errors on the high side. This, however, remains an open question. Clearly, from the intuitive point of view the bootstrapping is not properly mimicking the variation in the actual residuals and the procedure can be brought into question.

A second method with more general applicability is to bootstrap the vector of the observed response variable and the associated predictor variables. This only requires that the vectors are exchangeable and does not place explicit requirements on the residuals from the model.

However, some statisticians, particularly from the British school, view the second method philosophically as an inappropriate approach. To them, the regression problem requires that the predictor variables be fixed for the experiment and not selected at random from a probability distribution. The bootstrapping of the vector of response and predictor variables implicitly assumes a joint probability distribution for the vector of predictor variables and response. From their point of view, this is an inappropriate model and hence the vector approach is not an option.

However, from the practical point of view, if the approach of bootstrapping the vector has nice robustness properties related to model specification, it is justified. This was suggested by Efron and Tibshirani (1993, p. 113) for the case of a single predictor variable. Since it is robust, it is not important whether or not the method closely mimics the assumed but not necessarily correct

regression model. Presumably their observation extends to the case of more than one predictor variable.

On the other hand, some might argue that bootstrapping the residuals is only appropriate when the predictor variables are not fixed. This comes down to another philosophical issue that only statisticians care about. The question is one of whether conditional inference is valid when the experiment really involves an unconditional joint distribution for the predictor and response variables.

This is a familiar technical debate for statisticians because it is the same issue regarding the appropriateness of conditioning on the marginal totals in a 2×2 contingency table. Conditioning on ancillary information in the data (i.e., information in the data that does not have any affect on the "best" estimate of a parameter is a principle used by Sir Ronald Fisher in his theory of inference and is best known to be applied in Fisher's exact permutation test, which is most commonly used in applications involving categorical data).

For the practitioner, I repeat the sage advice of my friend and former colleague, V. K. Murthy, who often said "the proof of the pudding is in the eating." This applies here to these bootstrap regression methods as it does in the comparison of variants of the bootstrap in discriminant analysis. If we simulate the process under accepted modeling assumptions, the method that performs best in the simulation is the one to use regardless of how much you believe or like some particular theory.

These two methods for bootstrapping in regression are given by Efron (1982a, pp. 35–36). These methods are very general. They apply to linear and nonlinear regression models and can be used for least-squares or for any other estimation procedure. We shall now describe these bootstrap methods.

A general regression model can be given by

$$Y_i = g_i(\boldsymbol{\beta}) + \varepsilon_i \qquad \text{for } i = 1, 2, \ldots, n.$$

The functions g_i are of known form and may depend on a fixed vector of covariates c_i. The vector $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and the ε_i are independent and identically distributeed with some distribution F.

We assume that F is "centered" at zero. Usually this means that the expected or average value of ε_i is zero. However, in cases where the expected value does not exist, we may use the criterion that $P(\varepsilon < 0) = 0.50$.

Given the observed vector

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

where the ith component y_i is the observed value of the random variable Y_i, we find the estimate of $\boldsymbol{\beta}$, which minimizes the distance measure between $\boldsymbol{y}$ and $\boldsymbol{\lambda}(\boldsymbol{\beta})$ where

$$\boldsymbol{\lambda}(\boldsymbol{\beta}) = \begin{pmatrix} g_1(\beta) \\ g_2(\beta) \\ \vdots \\ g_n(\beta) \end{pmatrix}.$$

Denote the distance measure by $D(\boldsymbol{y}, \boldsymbol{\lambda}, (\boldsymbol{\beta}))$. If

$$\mathbf{D}(\boldsymbol{y}, \boldsymbol{\lambda}, (\boldsymbol{\beta})) = \sum_1^n [(y_i - g_i(\beta)]^2,$$

we get the usual least-squares estimates. For least absolute deviations, we would choose

$$\mathbf{D}(\boldsymbol{y}, \boldsymbol{\lambda}, (\boldsymbol{\beta})) = \sum_1^n \|[y_i - g_i(\beta)]\|.$$

Now by taking $\hat{\beta} = \min_{\beta} \mathbf{D}(\boldsymbol{y}, \boldsymbol{\lambda}, (\boldsymbol{\beta}))$, we have our parameter estimate of $\boldsymbol{\beta}$. The residuals are then obtained as $\hat{\varepsilon}_i = y_i - g_i(\hat{\beta})$.

The first bootstrap approach is to simply bootstrap the residuals. This is accomplished by constructing the distribution F_n that places probability $1/n$ at each $\hat{\varepsilon}_i$ We then generate bootstrap residuals ε_i^* for $i = 1, 2, \ldots, n$, where the ε_i^* are obtained by sampling independently from F_n (i.e., we sample with replacement from $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$). We then have a bootstrap sample data set;

$$y_i^* = g_i(\hat{\beta}) + \varepsilon_i^* \qquad \text{for } i = 1, 2, \ldots, n.$$

For each such bootstrap data set y^*, we obtain

$$\hat{\beta}^* = \min_{\beta} D[y^*, \lambda, \beta].$$

The procedure is repeated B times and the covariance matrix for $\hat{\beta}$ is estimated as $\hat{\Sigma} = \frac{1}{B-1} \sum_{j=1}^{B} \left(\hat{\beta}_j^* - \hat{\beta}^*\right)\left(\hat{\beta}_j^* - \hat{\beta}^*\right)^T$, where $\hat{\beta}_j^*$ is the bootstrap estimate from the jth bootstrap sample and $\hat{\beta}_*^* = \frac{1}{B} \sum_{j=1}^{B} \hat{\beta}_j^*$. This is the covariance estimate suggested by Efron (1982a, p. 36).

We note that bootstrap theory suggests simply using $\hat{\beta}$ in place of $\hat{\beta}_*^*$. The resulting covariance estimate should be close to that suggested by Efron. Confidence intervals for β can be obtained by the methods described in Chapter 3, but with the bootstrap samples for the $\hat{\beta}$ values.

The second approach is to bootstrap the vector

$$Z = \begin{pmatrix} y_i \\ c_i \end{pmatrix}$$

of the observations y_i and the covariates or predictor variables $\boldsymbol{c}_i$ for $i = 1, 2, \ldots, n$. The bootstrap samples are then z_i^* for $i = 1, 2, \ldots, n$ obtained by giving probability of selection $1/n$ to each z_i. Taking $z_i^* = \begin{pmatrix} y_i^* \\ c_i^* \end{pmatrix}$, we use y_i^* to obtain the $\hat{\beta}^*$ just as before.

Efron claims that although the two approaches are asymptotically equivalent for the given model, the second approach is less sensitive to model misspecification. It also appears that since we do not bootstrap the residuals, the second approach may be less sensitive to the assumptions concerning independence or exchangeability of the error terms.

4.1. LINEAR MODELS

In the case of the linear regression model, if the least-squares estimation procedure is used, there is nothing to be gained by bootstrapping. As long as the error terms are independent and identically distributed with mean zero and common variance σ^2, the least-squares estimates of the regression parameters will be the best among all linear unbiased estimators. The covariance matrix corresponding to the least-squares estimate $\hat{\beta}$ of the parameter vector $\boldsymbol{\beta}$ is given by

$$\sum = \sigma^2 (X^T X)^{-1},$$

where X is called the design matrix and $(X^T X)^{-1}$ is well-defined if X is a full-rank matrix. If $\hat{\sigma}^2$ is the least-squares estimate of the residual variance σ^2, then

$$\hat{\sum} = \hat{\sigma}^2 (X^T X)^{-1}$$

is the commonly used estimate of the parameter covariance matrix.

For more details see Draper and Smith (1981). These least-squares estimates are the standard estimates that can be found in all the standard statistical computer programs.

If, in addition, the error terms are Gaussian or approximately Gaussian, the least-squares estimates are also the maximum likelihood estimates. Also, the confidence intervals for the regression parameters, hypotheses tests about the parameters, and prediction intervals for a new observation based on known values of the regression variables can be determined in a straightforward way.

In the non-Gaussian case, even though we can estimate the parameter covariance matrix, we will not know the probability distribution for $\hat{\beta}$ and so we cannot determine confidence intervals and prediction intervals or perform hypothesis tests using the standard methods. The bootstrap approach does, however, provide a method for approximating the distribution of $\hat{\beta}$ through bootstrap sample estimates $\hat{\beta}^*$.

First we review the Gauss–Markov theory of least-squares estimation in Section 4.1.1. In Section 4.1.2 we discuss, in more detail, situations where we might prefer to use other estimates of β such as the least absolute deviation estimates or M-estimates.

In Section 4.1.3 we discuss bootstrap residuals and the possible problems that can arise. If we bootstrap the vector of response and predictor variables, we can avoid some of the problems of bootstrapping residuals.

4.1.1. Gauss–Markov Theory

The least-squares estimator of the regression parameters are maximum likelihood when the error terms is assumed to be Gaussian. Consequently, the least-squares estimates have the usual optimal properties under the Gaussian model. They are unbiased and asymptotically efficient. In fact, they have the minimum variance among unbiased estimators.

The Gauss–Markov theorem is a more general result in that it applies to linear regression models with general error distributions. All that is assumed is that the error distribution has mean zero and variance σ^2. The theorem states that among all estimators that are both unbiased and a linear function of the responses y_i for $i = 1, 2, \ldots, n$ the least-squares estimate has the smallest possible variance.

The result was first shown by Carl Friedrich Gauss in 1821. For more details about the theory, see the *Encyclopedia of Statistical Science*, Vol. 3, pp. 314–316.

4.1.2. Why Not Just Use Least Squares?

In the face of all these optimal properties, one should ask why least squares shouldn't always be the method of choice? The basic answer is that the least-squares estimates are very sensitive to violations in the modeling assumptions. If the error distribution has heavy tails or the data contain a few "outliers," the least-squares estimates will not be very good.

This is particularly true if these outliers are located at high leverage points (i.e., points that will have a large influence on the slope parameters). High leverage points occur at or near the extreme values of the predictor variables. In cases of heavy tails or outliers, the method of least absolute deviations or other robust regression procedures such as M-estimation or the method of repeated medians provide better solutions though analytically they are more complex.

Regardless of the procedure used, we may be interested in confidence regions for the regression parameters or prediction intervals for future cases. Under the Gaussian theory for least squares, this is possible. However, if the error distribution is non-Gaussian and unknown, the bootstrap provides a method for computing standard errors for the regression parameters or prediction intervals for future values, regardless of the method of estimation.

There are many other complications to the regression problem that can be handled by bootstrapping. These include the problem of heteroscedasticity of the variance of the error term, nonlinearity in the model terms, and bias adjustment when transformation of variables is used.

For a bootstrap-type approach to the problem of retransformation bias, see Duan (1983). Bootstrap approaches to the problem of heteroscedasticity are covered in Carroll and Ruppert (1988).

An application of bootstrapping residuals for a nonlinear regression problem is given in Shimbukuro et al. (1984) and will be discussed later. When procedures other than least-squares are used, confidence intervals and prediction intervals are still available by bootstrapping.

Both editions of a book by Miller (1986, 1997) deal with linear models. These are very excellent references for the understanding of the importance of modeling assumptions. They also demonstrate when and why the methods are robust to departures from basic assumptions. These texts also point out when robust and bootstrap statistical procedures are more appropriate.

4.1.3. Should I Bootstrap the Residuals from the Fit?

From Efron (1979a, Section 7), the bootstrap estimate of the covariance matrix for the coefficients in a linear regression model is shown to be

$$\hat{\Sigma} = \sigma^2 \left(\sum_{i=1}^{n} c_i c_i^T \right)^{-1},$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2.$$

The model is given by $y_i = c_i\boldsymbol{\beta} + \varepsilon_i$ for $i = 1, 2, \ldots, n$ and $\hat{\varepsilon}_i$ is the residual estimate obtained by least-squares. The only difference between this estimate and the standard one from the Gauss–Markov theory is use of n in the denominator of the estimate for σ^2. The standard theory would use $n - p$, where p is the number of the covariates in the model (i.e., the dimension of the vector

$\boldsymbol{\beta}$). So we see that at least when the linear least-squares model is an appropriate method bootstrapping the residuals gives nearly the same answer as the Gauss–Markov theory for larger. Of course, in such a case, we do not need to bootstrap since we already have an adequate model.

It is important to ask how well this approach to bootstrapping residuals works when there is not an adequate theory for estimating the covariance matrix for the regression parameters. There are many situations that we would like to consider: (1) heteroscedasticity in the residual variance; (2) correlation structure in the residuals; (3) nonlinear models; (4) non-Gaussian error distributions; and (5) more complex econometric and time series models.

Unfortunately, the theory has not quite reached the level of maturity to give complete answers in these cases. There are still many open research questions to be answered. In this section and in Section 4.2, we will try to give partial answers to (1) through (4); (5) is being deferred to Chapter 5, which covers time series methods.

A second approach to bootstrapping in a regression problem is to bootstrap the entire vector

$$Z_i = \begin{pmatrix} y_i \\ c_i \end{pmatrix}$$

that is a $(p + 1)$-dimensional vector of the response variable and the covariate values. A bootstrap sample is obtained by choosing integers at random with replacement from the set $1, 2, 3, \ldots, n$ until n integers have been chosen. If, on the first selection, say, integer j is chosen, then the bootstrap observation is $Z_i^* = Z_j$. After a bootstrap sample has been chosen, the regression model is fit to the bootstrap samples producing an estimate $\boldsymbol{\beta}^*$. By repeating this B times, we get $\beta_1^*, \beta_2^*, \ldots, \beta_B^*$ the bootstrap sample estimates of β. The usual sample estimates of variance and covariance can then be applied to $\beta_1^*, \beta_2^*, \ldots, \beta_B^*$.

Efron and Tibshirani (1986) claim that the two approaches are asymptotically equivalent (presumably when the covariates are assumed to be chosen from a probability distribution), but can perform differently in small sample situations.

The latter method does not take full advantage of the special structure of the regression problem. Whereas bootstrapping the residuals leads to the estimates $\hat{\Sigma}$ and $\hat{\sigma}^2$ as defined earlier when $B \to \infty$, this latter procedure does not.

The advantage is that it provides better estimates of the variability in the regression parameters when the model is not correct. We recommend it over bootstrapping the residuals when (1) there is heteroscedasticity in the residual variance, (2) there is correlation structure in the residuals, or (3) we

suspect that there may be other important parameters missing from the model.

Wu (1986) discusses the use of a jackknife approach in regression analysis which he views to be superior to the bootstrap approaches we have mentioned. His approach works particularly well in the case of heteroscedasticity of residual variances.

There are several discussants to Wu's paper. Some strongly support the bootstrap approach and point out modifications for heteroscedastic models, Wu claims that even such modifications to the bootstrap will not work for nonlinear and binary regression problems. The issues are far from settled.

The two bootstrap methods described in this section apply equally to nonlinear (homoscedastic, i.e., constant variance) models as well as the linear (homoscedastic) models. In the next section, we will give some examples of nonlinear models. We will then consider a particular experiment where we bootstrap the residuals.

4.2. NONLINEAR MODELS

The theory of nonlinear regression models has advanced greatly in the 1970s and 1980s. Much of this development has been well-documented in recent textbooks devoted strictly to nonlinear models. Two such books are Bates and Watts (1988) and Gallant (1987).

The nonlinear models can be broken up into two categories. In the first category, local linear approximations can be made using Taylor series, for example. When this can be done, approximate confidence or prediction intervals can be generated based on asymptotic theory.

Much of this theory is covered in Gallant (1987). In the aerospace industry, there has been great success applying local linearization methods in the construction of Kalman filters for missiles, satellites and other orbiting objects.

The second category is the highly nonlinear model for which the linear approximation will not work. Bates and Watts (1988) provide methods for diagnosing the severity of the nonlinearity.

The bootstrap method can be applied to any type or nonlinear model. The two methods as described in Efron (1982a) can be applied to fairly general problems. To bootstrap, we do not need to have a differentiable functional form. The nonlinear model could even be a computer algorithm rather than an analytical expression. We do not need to restrict the residual variance to have a Gaussian distribution. The only requirements are that the residuals should be independent and identically distributed (exchangeable may be sufficient) and their distribution should have a finite variance. The distribution of the residuals should not change as the predictor variables are changed. This requirement imposes homoscedasticity on the residual variance.

The distribution of the residuals should not change because the predictor variables changed. This requirement imposes homoscedasticity on the residual variance.

For models with heteroscedastic variance, modifications to the bootstrap are available. We shall not discuss these modifications here. To learn more about it, look at the discussion to Wu (1986).

4.2.1. Examples of Nonlinear Models

In Section 4.2.2, we discuss a quasi-optical experiment that was performed to determine the accuracy of a new measurement technique for the estimation of optical properties of materials used to transmit and/or receive millimeter wavelength signals. This experiment was conducted at the Aerospace Laboratory.

As a statistician in the engineering group, I was asked to determine the standard errors of their estimates. The statistical model was nonlinear and I chose to use the bootstrap to estimate the standard error. Details on the model and the results of the analysis are given in Section 4.2.2.

Many problems that arise in practice can be solved by approximate models that are linear in the parameters (remember that in statistical models the distinction between linear and nonlinear is in the parameters and not in the predictor variables). The scope of applicability of linear models can, at times, be extended by including transformations of the variables.

However, there are limits to what can adequately be approximated by linear models. In many practical scientific endeavors, the model may arise from a solution to a differential equation. A nonlinear model that could arise as the solution of a simple differential equation might be the function

$$f(x, \sigma)=\sigma_1+\sigma_2 \exp(\sigma_3 x),$$

where x is a predictor variable and

$$\sigma=\begin{pmatrix}\sigma_1\\ \sigma_2\\ \sigma_3\end{pmatrix}$$

is a three-dimensional parameter vector.

A common problem in time series analysis is the so-called harmonic regression problem. We may know that the response function is periodic or the sum of a few periodic functions, but we do not know the amplitude or the frequency of the periodic components. Here it is the fact that the frequencies are among the unknown parameters that makes the model nonlinear. The simple case of a single periodic function can be described by the following function.

$$f(t, \varphi) = \varphi_0 + \varphi_1 \sin(\varphi_2 t + \varphi_3)$$

where t is the time since a specific epoch and

$$\varphi = \begin{pmatrix} \varphi_0 \\ \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{pmatrix}$$

is a vector of unknown parameters. The parameters φ_1, φ_2, and φ_3 all have physical interpretations. φ_1 is called the amplitude, φ_2 is the frequency, and φ_3 is the phase delay.

Because of the trigonometric identity

$$\sin(A + B) = \sin A \cos B + \cos A \sin B$$

we can reexpress

$$\varphi_1 \sin(\varphi_2 t \varphi_3)$$

as

$$\varphi_1 \cos \varphi_3 \sin \varphi_2 t + \varphi_1 \sin \varphi_3 \cos \varphi_2 t.$$

The problem can then be reparameterized as

$$f(t, A) = A_0 + A_1 \sin A_2 t + A_3 \cos A_2 t,$$

where

$$A = \begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{pmatrix}$$

and $A_0 = \varphi_0$, $A_1 = \varphi_1 \cos\varphi_3$, $A_2 = \varphi_2$, and $A_3 = \varphi_1 \sin \varphi_3$.

This reparameterized form of the model is the form given by Gallant (1987, p. 3) with slightly different notation.

There are many other examples where nonlinear models are solutions to differential equations or systems of differential equations. Even in the case of linear differential equations or systems of linear differential equations, the solutions involve exponential functions (both real- and complex-valued). The results are then real-valued functions that are periodic or exponential or a combination of both.

If constants involved in the differential equation are unknown, then their estimates will be obtained through the solution of a nonlinear model. As a simple example, consider the equation

$$\frac{d}{dx} y(x) = -\sigma_1 y(x)$$

subject to the initial condition y(0) = 1. The solution is then

$$y(x) = e^{-\varphi_1 x}.$$

Since φ_1 is an unknown parameter, the function $y(x)$ is nonlinear in φ_1.

For a commonly used linear system of differential equations whose solution involves a nonlinear model, see Gallant (1987, pp. 5–8). Such systems of differential equations arise in compartmental analysis commonly used in chemical kinetics problems.

4.2.2. A Quasi-optical Experiment

In this experiment, I was asked as a consulting statistician to determine estimates of two parameters that were of interest to the experimenters. More importantly, they needed a "good" estimate of the standard errors of these estimates since they were proposing a new measurement technique that they believed would be more accurate than previous methods.

Since the model was nonlinear and I was given a computer program rather than an analytic expression, I chose to bootstrap the residuals. The results were published in Shimabukuro, Lazar, Dyson, and Chernick (1984).

The experimenters were interested in the relative permittivity and the loss tangent (two material properties related to the transmission of signals at millimeter wavelengths through a dielectric slab). The experimental setup is graphically depicted in Figure 4.1. Measurements are taken to compute $|T|^2$, where T is a complex number called the transmission coefficient. An expression for T is given by

$$T = \frac{(1-r^2)e^{-(\beta_1-\beta_0)di}}{1-r^2 e^{-2\beta_1 di}},$$

where

$$\beta_1 = \frac{2\pi}{\lambda_0}\sqrt{\varepsilon_1/\varepsilon_0 - \sin^2\varphi},$$

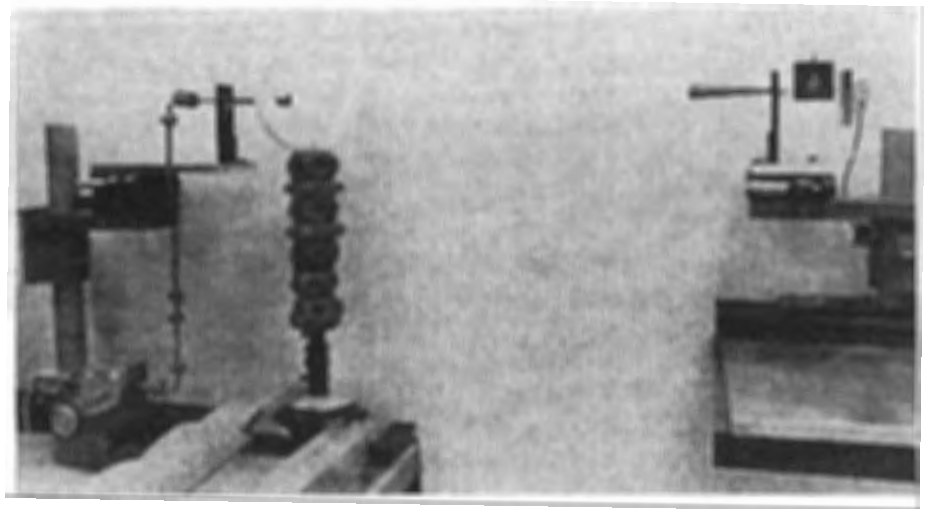

Figure 4.1 Photograph of experimental setup. The dielectric sample is mounted in the teflon holder. [From Shimabukuro et al. (1984).]

$$\beta_0 = \frac{2\pi}{\lambda_0}\cos\varphi,$$

$$\varepsilon_1 = \varepsilon_r\varepsilon_0\left(1 - \frac{i\sigma}{\omega\varepsilon_r\varepsilon_0}\right),$$

and

ε_0 = permittivity of free space
ε_r = relative permittivity
σ = conductivity
λ_0 = free-space wavelength
$\dfrac{\sigma}{\omega\varepsilon_r\varepsilon_0}$ = tan δ = loss tangent
d = thickness of the slab
r = reflection coefficient of a plane wave incident to a dielectric boundary
ω = free-space frequency
$i = \sqrt{-1}$

For more details on the various conditions of the experiment, see Shimbukuro et al. (1984).

We applied the bootstrap to the residuals using the nonlinear model

$$y_i = g_i(\mathbf{v}) + \varepsilon_i \qquad \text{for } i = 1, 2, \ldots, N$$

where y_i is the power transmission measurement at incident angle φ_i with $\varphi_i = i - 1$ degrees. The nonlinear function $g_i(u)$ is $|T|^2$ and $\mathbf{v}$ is a vector of two parameters, ε_r (relative permittivity) and tan δ (loss tangent). For simplicity the wavelength λ, the slab thickness d and the angle of incidence φ_i are all assumed to be known for each observation. The experimenters believe that measurement error in these variables would be relatively small and have little effect on the parameter estimates. Some checking of these assumptions was made.

For most of the materials, 51 observations were taken. We chose to do 20 bootstrap replications for each model. Results were given for eight materials and are shown in Table. 4.1.

The actual least-squares fit to the eight materials are shown in Figure 4.2. We notice that the fit is generally better at the higher-incidence angles. This suggests a violation of the assumption of independent and identically distributed residuals. There may be a bias at the low incidence angles indicative of either model inadequacy or poorer measurements.

Looking back on the experiment, there are several possible ways we might have improved the bootstrap procedure. Since bootstrapping residuals is more sensitive to the correctness of the model, it may have been better to bootstrap the vector.

Recent advances in bootstrapping in heteroscedastic models may also have helped. A rule of thumb for estimating standard errors is to take 100–200 bootstrap replications, whereas we only did 20 replications in this research.

Table 4.1 Estimates of Permittivities and Loss Tangents (f = 93.7888 GHz)

	Least-Squares Estimate		Bootstrap Estimates with Standard Error	
Material	ε_r	tan δ	ε_r	tan δ
Teflon	2.065	0.0002	2.065 ± 0.004	0.00021 ± 0.00003
Rexolite	2.556	0.0003	2.556 ± 0.005	0.00026 ± 0.00006
TPX	2.150	0.0010	2.149 ± 0.005	0.0009 ± 0.0001
Herasil (fused quartz)	3.510	0.0010	3.511 ± 0.005	0.0010 ± 0.0001
36D	2.485 (2.45)	0.0012 (<0.0007)	2.487 ± 0.008	0.0011 ± 0.0002
36DA	3.980 (3.7)	0.0012 (<0.0007)	3.980 ± 0.009	0.0014 ± 0.0001
36DK	5.685 (5.4)	0.0040 (<0.0008)	5.685 ± 0.009	0.0042 ± 0.0001
36DS	1.765 (1.9)	0.0042 (0.001)	1.766 ± 0.006	0.0041 ± 0.0001

Source: Shimabukuro et al. (1984).

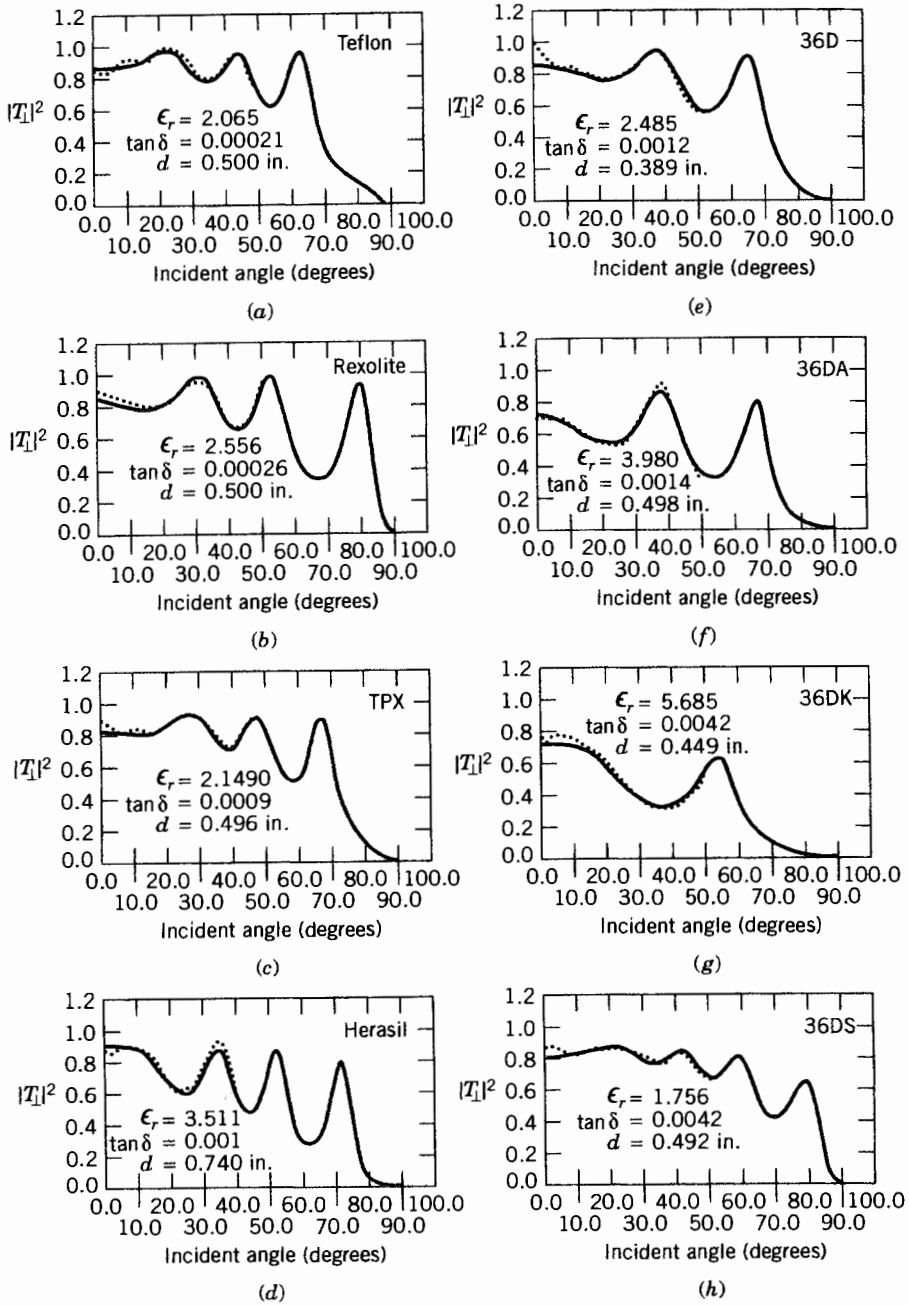

Figure 4.2 The measured power transmission for different dielectric samples is shown by the dotted lines. The line curves are the calculated $|T_\perp|^2$ using the best-fit estimates of ε_r and tan δ. [From Shimabukuro et al. (1984).]

From a data analytic point of view, it may have been helpful to delete the low-angle observations and see the effect on the fit. We might then have decided to fit the parameters and bootstrap only for angle greater than, say, 15 degrees.

By bootstrapping the residuals, the large residuals at the low angles would be added at the higher angles for some of the bootstrap samples. We believed that this would tend to increase the variability in the parameter estimates of the bootstrap sample and hence lead to an overestimate of their standard errors.

Since the estimated standard errors were judged to be good enough by the experimenters, we felt that our approach was adequate. The difficulty with the residual assumptions was recognized at the time.

4.3. NONPARAMETRIC MODELS

Given a vector $\boldsymbol{X}$, the regression function $E(y|\boldsymbol{X})$ is often a smooth function in $\boldsymbol{X}$. In Sections 4.1 and 4.2, we considered specific linear and nonlinear forms for the regression function. Nonparametric regression is an approach that allows more general smooth functions as possibilities for the regression function. The nonparametric regression model for an observed data set (y_i,x_i) for $1 \le i \le n$ is

$$y_i = g(x_i) + \varepsilon_i, \qquad 1 \le i \le n,$$

where $g(\boldsymbol{x}) = \boldsymbol{E}(y|\mathbf{x})$ is the function we wish to estimate. We assume that the ε_i are independent and identically distributed with mean zero and variance σ^2.

In the regression model, $\boldsymbol{x}$ is assumed to be given as in a designed experiment. One approach to the estimation of the function g is kernel smoothing [see Hardle (1990a,b) or Hall (1992a, pp. (257–269)]. The bootstrap is used to help determine the degree of smoothing (i.e., determine the tradeoff between variance and bias analogous to its use in nonparametric density estimation).

Cox's proportional hazards model is a standard regression method for dealing with censored data [see Cox (1972)]. The hazard function $h(t|\boldsymbol{x})$ is the derivative of the survival function $\boldsymbol{S}(t|\boldsymbol{x})$ = probability of surviving t or more time units given predictor variables $\boldsymbol{x}$. In Cox's model $h(t|\boldsymbol{x}) = h_0(t)e^{(\beta x)}$, where $h_0(t)$ is an arbitrary unspecified function assumed to depend solely on t.

Through the use of the "partial likelihood" function, the regression parameters β can be estimated independently of the function $h_0(t)$. Because of the form of $h(t|\boldsymbol{x})$, the method is sometimes referred to as semi parametric.

Efron and Tibshirani (1986) apply the bootstrap to leukemia data for mice in order to assess the effectiveness of a treatment. See their article for more details.

Without going into the details, we mention projection pursuit regression and alternating conditional expectation (ACE) as two other "nonparametric" regression techniques which have been studied recently. Efron and Tibshirani (1986) provide examples of applications of both methods and show how the bootstrap can be applied when using these techniques.

The interested reader can consult Friedman and Stuetzle (1981) for the original source on project pursuit. The original work describing ACE (or alternating conditional expectation) is Breiman and Friedman (1985).

Briefly, projection pursuit searches for linear combinations of the predictor variables and takes smooth functions of those linear combinations to form the prediction equation. ACE generalizes the Box–Cox regression model by transforming the response variable with an unspecified smooth function as opposed to a simple power transformation.

4.4. HISTORICAL NOTES

Although regression analysis is one of the most widely used statistical techniques, application of the bootstrap to regression problems has only appeared fairly recently. The many fine books on regression analysis including Draper and Smith (1981) for linear regression and Gallant (1987) and Bates and Watts (1988) do not mention or pay much attention to bootstrap methods. A recent exception is Sen and Srivastava (1990).

Draper and Smith (1998) also incorporate a discussion of the bootstrap. Early discussion of the two methods of bootstrapping in the nonlinear regression model with homoscedastic errors can be found in Efron (1982a). Carroll, Ruppert, and Stefanski (1995) deal with the bootstrap applied to the nonlinear calibration problem (measurement error models and other nonlinear regression problems, pp. 273–279, Appendix A.6).

Efron and Tibshirani (1986) provide a variety of interesting applications and some insightful discussion of bootstrap applications in regression problems. They go on to discuss nonparametric regression applications including projection pursuit regression and methods for deciding on transformations for the response variable such as the alternating conditional expectation method (ACE) of Breiman and Friedman (1985). Texts devoted to nonparametric regression and smoothing methods include Hardle (1990a,b), Hart (1997), and Simonoff (1996). Belsley, Kuh, and Welsch (1980) cover multicollinearity and related regression diagnostics.

Bootstrapping the residuals is an approach that also can be applied to time series models. We shall discuss time series applications in the next chapter. An example of a time series application to the famous Wolfer sunspot numbers is given in Efron and Tibshirani (1986, p. 65).

Shimabukuro et al. (1984) was an early example of a practical application of a nonlinear regression problem. The first major study of the bootstrap as applied to the problem of estimating the standard errors of the regression coefficients by constrained least squares with an unknown, but estimated, residual covariance matrix can be found in Freedman and Peters (1984a). Similar analyses for econometric models can be found in Freedman and Peters (1984b).

Peters and Freedman (1984b) also deals with issues related to bootstrapping in regression problems. Their study is very interesting because it shows that the conventional asymptotic formulas that are correct for very large samples do not work well in small-to-moderate sample size problems. They show that these standard errors can be too small by a factor of nearly three! On the other hand the bootstrap method gives accurate answers. The motivating example is an econometric equation for the energy demand by industry.

In Freedman and Peters (1984b) the bootstrap is applied to a more complex econometric model. Here the authors show that the three-stage least-squares estimates and the conventional estimated standard errors of the coefficients are good. However, conventional prediction intervals based on the model are too small due to forecast bias and underestimation of the forecast variance.

The bootstrap approach given by Freedman and Peters (1984b) seems to provide better prediction intervals in their example. The authors point out that there is unfortunately no good rule of thumb to apply to determine when the conventional formulas will work of when it may be necessary to resort to the bootstrap. They suggest that the development of such a rule of thumb could be a result of additional research. Even the bootstrap procedure has problems in this context.

Theoretical work on the use of bootstrap in regression is given in Freedman (1981), Bickel and Freedman (1983), Weber (1984), Wu (1986), and Shao (1988a,b). Another application to an econometric model appears in Daggett and Freedman (1985).

Theoretical work related to robust regression is given in Shorack (1982). Rousseeuw (1984) applies the bootstrap to the least median of squares algorithm. Efron (1992a) discusses the application of bootstrap to estimating percentiles of a regression function.

Jeong and Maddala (1993) review various resampling tests for econometric models. Hall (1989c) shows that the bootstrap applied to regression problems can lead to confidence interval estimates that are unusually accurate.

Various recent regression applications include Breiman (1992) for model selection related to x-fixed prediction, Brownstone (1992) regarding admissibility of linear model selection techniques, Bollen and Stine (1993) regarding fitting of structural equation models, and Cao-Abad (1991) regarding rates of convergence for a bootstrap variation called the "wild" bootstrap. The wild bootstrap is useful in nonparametric regression [see also Mammen (1993), who applies the wild bootstrap in linear models], DeAngelis, Hall, and Young (1993a) related to L^1 regression, Lahiri (1994c) for M-estimation in multiple linear regression problems, Dikta (1990) for nearest-neighbor regression, and Green, Hahn, and Rocke (1987) for an economic application to the estimation of elasticities.

Wu (1986) gives a detailed theoretical treatment of jackknife methods applied to regression problems. He deals mainly with the problem of heteroscedastic errors. He is openly critical of the blind application of bootstrap methods and illustrates that certain bootstrap approaches will give incorrect results when applied to data for which heterosecedastic models are appropriate. A number of the discussants including Beran, Efron, Freedman, and Tibshirani defend the appropriate use of the "right" bootstrap in this context. The issue is a complex one which even today is not completely settled.

It is fair to say that Jeff Wu's criticism of the bootstrap in regression problems was a reaction to the "euphoria" expressed for the bootstrap in some of the earlier works such as Efron and Gong (1983, Section 1) or Diaconis and Efron (1983).

Although enthusiasm for the bootstrap approach is justified, some statements could leave naive users of statistical methods with the idea that it is easy to just apply the bootstrap to any problem they might have. I think that every bootstrap researcher would agree that careful analysis of the problem is a necessary step in any applied problem and that if bootstrap methods are appropriate, one must be careful to choose the "right" bootstrap method from the many possible bootstraps.

Stine (1985) deals with bootstrapping for prediction intervals, and Bai and Olshen as discussants to the paper by Hall (1988b) provide some elementary asymptotic theory for prediction intervals. Olshen, Biden, Wyatt, and Sutherland (1989) provide a very interesting application to gait analysis.

A theoretical treatment of nonparametric kernel methods in regression problems is given in Hall (1992a). His development is based on asymptotic expansions (i.e., Edgeworth expansions). Other key articles related to bootstrap applications to nonparametric regression include Hardle and Bowman (1988) and Hardle and Marron (1991).

The reader may first want to consult Silverman (1986) for a treatment of kernel density methods and some applications of the bootstrap in density estimation. Devorye and Gyorfi (1985) also deals with kernel density methods as does Hand (1982), and for multivariate densities see Scott (1992). Hardle (1990a) provides an account of nonparametric regression techniques.

Hayes, Perl, and Efron (1989) have extended bootstrap methods to the case of several unrelated samples with application to estimating contrasts in particle physics problems. Hastie and Tibshirani (1990) treat a general class of models called generalized additive models. These include both the linear and the generalized linear models as special cases. It can be viewed as a form of curve fitting but is not quite as general as nonparametric regression.

Bailer and Oris (1994) provide regression examples for toxicity testing and compare bootstrap methods with likelihood and Poisson regression models (a particular class of generalized linear models). One of their examples appears in Davison and Hinkley (1997, practical number 6, pp. 383–384).

CHAPTER 5

Forecasting and Time Series Analysis

5.1. METHODS OF FORECASTING

One of the most common problems in the "real world" is forecasting. We try to forecast tomorrow's weather or when the next big earthquake will hit. When historical data are available and models can be developed which fit the historical data well, we may be able to produce accurate forecasts. For certain problems (e.g., earthquake predictions or the Dow Jones Industrial Average) the lack of a good statistical model makes forecasting problematic (i.e., no better than crystal ball gazing).

Among the most commonly used forecasting techniques are exponential smoothing and autoregressive integrated moving average (ARIMA) modeling. The ARIMA models are often referred to as the Box–Jenkins models after George Box and Gwilym Jenkins, who popularized the approach in Box and Jenkins (1970, 1976). The autoregressive models, which are a subset of the ARIMA models, actually go back to Yule (1927).

Exponential smoothing is an approach that provides forecasts future values using exponentially decreasing weights on the past values. The weights are determined by smoothing constants that are estimated from the data. The simplest form—single exponential smoothing—is a special case of the ARIMA models namely the IMA (1, 1) model. The smoothing constant in the model can be determined from the moving average parameter of the IMA (1, 1) model. The smoothing constant can be determined from the moving average parameter of the IMA (1, 1) model.

Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition,
by Michael R. Chernick

5.2. TIME SERIES MODELS

ARIMA models are attractive because they provide good empirical approximations to a large class of time series. There is a body of statistical theory showing that "most" stationary stochastic processes can be well-approximated by high-order autoregressive processes.

The term *stationary stochastic process* generally means strictly stationary. A stochastic process is said to be strictly stationary if the joint probability distribution of k consecutive observations does not depend on the time parameter t for all choices of $k = 1, 2, 3, 4, 5, \ldots, \infty$. Informally, this means that if we are looking at the first k observations in a time series, the statistical properties of that set of observations wouldn't change if we took any other set of k consecutive observations in the time series.

A weaker form of stationarity is second-order (or weak) stationarity. Second-order stationarity requires only that the second-order moments exist and that the first- and second-order moments, the mean function and the autocorrelation function, respectively, do not depend on time (i.e., they are constant over time).

Strict stationarity implies weak stationarity, but there are weakly stationary, processes that are not strictly stationary. For Gaussian processes, second-order (weakly) stationary processes are strictly stationary because they have the property that the joint distribution, for any choice of k consecutive observations, depends only on the first and second moments of their joint distribution.

Box and Jenkins used the mixed autoregressive moving average model to provide a parsimonious representation for these high-order autoregressive processes (i.e., by including just a few moving average terms an equivalent model is found with only a small number of parameters to estimate). To generalize this further to handle trends and seasonal variations (i.e., nonstationarity), Box and Jenkins (1976) include differencing and seasonal differences of the series. Using mathematical operator notation, let

$$W_t = \Delta^d Y_t,$$

where Y_t is the original observation at time t and the operation Δ^d applies the difference operation Δd times where Δ is defined by $\Delta y_t = y_t - y_{t-1}$.

So,

$$\Delta^2 y_t = \Delta(y_t - y_{t-1}) = \Delta y_t - \Delta y_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}.$$

In general,

$$\Delta^d y_t = \Delta^{d-1}(\Delta y_t) = \Delta^{d-1}(y_t - y_{t-1}) = \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1}.$$

After differencing the times series, W_t is a stationary ARMA (p, q) model given by the equation

$$W_t = b_1 W_{t-1} + b_2 W_{t-2} + \cdots + b_p W_{t-p} + e_t + a_0 e_{t-2} + \cdots + a_q e_{t-q},$$

where e_t, $e_{t-1}, \ldots, e_{t-q}$ are the assumed random innovations and W_{t-1}, $W_{t-2}, \ldots, W_{t-p}$ are past values of the dth difference of the Y_t series.

These ARIMA models handle polynomial trends in the time series. Additional seasonal components can be handled by seasonal differences [see Box and Jenkins (1976) for details].

Although the Box–Jenkins models cover a large class of time series and provide very useful forecasts and prediction intervals, they have drawbacks for some cases. The models are linear and the least-squares or maximum likelihood parameter estimates are good only if the innovation series e_t is nearly Gaussian.

If the innovation series e_t has heavy tails or there are a few spurious observations in the data, the estimates can be distorted and the prediction intervals are not valid. In fact, the Box–Jenkins methodology for choosing the order of the model (i.e., deciding on the values for p, d, and q) will not work if outliers are present. This is because estimates for the autocorrelation and partial autocorrelation functions are very sensitive to outliers [see, for example, Chernick, Downing, and Pike (1982) or Martin (1980)].

One approach to overcoming the difficulty is to detect and remove the outliers and then fit the Box–Jenkins model with some missing observations. Another approach is to use robust estimation procedures for parameters [see Rousseeuw and Leroy (1987)].

In the 1980s there were also a number of interesting theoretical developments in bilinear and other nonlinear time series models which may help to extend the applicability of statistical time series modeling [see Tong (1983, 1990)].

Even if an ARIMA model is appropriate and the innovations e_t are uncorrelated but not Gaussian, it may be appropriate to bootstrap the residuals to obtain appropriate standard errors for the model parameters and the predictions. Bootstrap prediction intervals may also be appropriate.

The approach is the same as we have discussed in Chapter 4, which covers regression analysis. The confidence interval methods of Chapter 3 may be appropriate for the prediction intervals. We shall discuss this further in the next section.

5.3. WHEN DOES BOOTSTRAPPING HELP WITH PREDICTION INTERVALS?

Some results are available on the practical application of the bootstrap to time series models. These results apply to stationary autoregressive (AR)

processes, a subset of the stationary autoregressive-moving average (ARMA) models discussed in the previous section.

To illustrate how the bootstrap can be applied to an autoregressive model, we will illustrate the approach with the simple first-order autoregressive process. This model is sufficient to illustrate the key points. For the first-order autoregression (AR (1) model) the model is given by

$$y_t = b_1 y_{t-1} + e_t,$$

where y_t is the observation at time t (possible centered to have zero mean) and e_t are the innovations.

If the average of the observed series is not zero, a sample estimate of the mean is subtracted from each observation in order to center the data. In practice, if the original series appears to be nonstationary, differencing methods or other forms of trend removal would be applied first.

For Gaussian processes, least-squares or maximum likelihood estimates for b_1 are computed along with standard errors for the estimates. If y_{t_m} is the last observation, then a one-step-ahead prediction is obtained at $t_m + 1$ by using $\hat{b}_1 y_{t_m}$ as the prediction, where $\hat{b}_1$ is the estimate of b_1. Statistical software packages (e.g., SAS/ETS, BMDP, and IMSL) provide such estimates of parameters and also produce forecast intervals.

These procedures work well when the e_t have approximately a Gaussian distribution with mean zero. Stine (1987) provides forecasts and prediction intervals with the classical Gaussian model but using a bootstrap approach. He shows that although the bootstrap is not as efficient as the classical estimate when the Gaussian approximation is valid, it provides much better prediction intervals for non-Gaussian cases.

In order to apply the bootstrap to the AR (1) model, we need to generate a bootstrap sample. First we need an estimate $\hat{b}_1$. We may take the Gaussian maximum likelihood estimate generated by a software program such as PROC ARIMA from SAS. We then generate the estimated residuals, namely,

$$\hat{e}_r = y_t - \hat{b} y_{t-1} \qquad \text{for } t = 2, 3, \ldots, t_m.$$

Note that we cannot compute a residual $\hat{e}_1$ since y_0 is not available to us. A bootstrap sample $y_1^*, y_2^*, \ldots, y_{t_m}^*$ is then generated by bootstrapping the residuals. We simply generate $e_2^*, e_3^*, \ldots, e_{t_m}^*$ by sampling with replacement from $\hat{e}_2, \hat{e}_3, \ldots, \hat{e}_{t_m}$ and defining by recursion:

$$y_2^* = \hat{b}_1 y_1^* + e_2^*, \qquad y_3^* = \hat{b}_1 y_2^* + e_3^*, \ldots, y_{t_m}^* = \hat{b}_1 y_{t_m-1}^* + e_{t_m-1}^*.$$

Efron and Tibshirani (1986) take $y_1^* = y_1$ for each bootstrap sample. With autoregressive processes, since we have a first time point that we denote as $t = 1$, we need initial values. In the AR (1) example, we see that we need a single initial value to start the process. In this case we let $y_1^* = y_1$.

In general for the pth-order autoregression, we will need p initial values. Stine (1987) and Thombs and Schucany (1990) provide alternative methods for obtaining starting values for the bootstrap samples.

Now for each bootstrap sample, an estimate $\hat{b}_1^*$ is obtained by applying the estimation procedure to $y_1^*, y_2^*, \ldots, y_{t_m}^*$. Efron and Tibshirani illustrate this on the Wolfer sunspot data. They obtain the standard errors for $\hat{b}_1$ by this procedure. They then go on to fit an AR (2) (second-order autoregressive model) to the sunspot data and obtain bootstrap estimates of the standard errors for the two parameters in the AR (2) model. They did not go on to consider prediction intervals.

For the Gaussian case, the theory has been developed to obtain the minimum mean-square error predictions based on "known" autoregressive parameters. Formulas for the predictions and their mean-square errors can be found in Box and Jenkins (1976) or Fuller (1976). Stine (1987) shows that when the autoregressive parameter b_1 is replaced by the estimate $\hat{b}_1$ in the forecasting equations, the prediction mean-square error increases.

Stine (1987) provides a Taylor series expansion to estimate the mean-square error of the prediction that works well for Gaussian data. The bootstrap estimates of mean-square error are biased, but his bootstrap approach does provide good prediction intervals. We shall describe this approach, which we recommend when the residuals do not fit well to the Gaussian model.

Stine (1987) assumes that the innovations have a continuous and strictly increasing distribution with finite moments. He also assumes that the distribution is symmetric about zero. The key difference between Stine's approach and that of Efron and Tibshirani is the introduction of the symmetric error distribution. Instead of sampling with replacement from the empirical distribution for the estimated residuals (the method of Efron and Tibshirani previously described), Stine does the following:

Let

$$\begin{aligned} F_T(x) &= \frac{1}{2} + (L(x)/[2(T-p)]), \quad x \geq 0, t = p+1, \ldots, T \\ &= 1 - F_T(-x), \qquad\qquad x < 0, \end{aligned}$$

where $L(x)$ = number of t such that $k|\hat{\varepsilon}_t| \leq x$, and

$$k = [(T-p)/(T-2p)]^{1/2}.$$

This choice of F_T produces bootstrap residuals that are symmetric about zero and have a variance that is the same as the original set of residuals.

A bootstrap approximation to the prediction error distribution is easily obtained given the bootstrap estimates of the autoregressive parameters and the bootstrap observations $y_1^*, y_2^*, \ldots, y_{t_m}^*$. The prediction formulas are used to obtain bootstrap prediction $\hat{y}_{t_m+f}^*$ for the time $t_m + f$, f time steps in the future. The variable $\hat{y}_{t_m+f}^* - \hat{y}_{t_m+f}$ provides the bootstrap sample estimate of prediction

error f steps ahead, where $\hat{y}^*_{t_m+f}$ is the original prediction based on the original estimates of the autoregressive parameters and the observations $y_1, y_2, \ldots, y_{t_m}$. Actually, Stine uses a more sophisticated approach based on the structure of the forecast equation [see Stine (1987) for details].

Another difference between Stine's approach and that of Efron and Tibshirani is that Efron and Tibshirani fix the first p values of the process in generating the bootstrap sample whereas Stine chooses a block of p consecutive observations at random to initiate the bootstrap sample.

In practice, we will know the last p observations when making future predictions. Autoregressive forecasts for $1, 2, \ldots, f$ steps ahead depend only on the autoregressive parameters and the last p observations. Consequently, it makes sense to condition on the last p observations when generating the bootstrap predictions.

Thombs and Schucany (1990) use a time-reversal property for autoregressive processes to fix the last p observations and generate bootstrap samples for the earlier observations. They apply the backward representation (Box and Jenkins, 1976, pp. 197–200) to express values of the process at time t as a function of future values. This representation is based on generating the process backward in time, which is precisely what we want to do with the bootstrap samples. The correlation structure for the reversed process is the same as for the forward process.

For Gaussian processes, this means that the two series are distributionally equivalent. Weiss (1975) has shown that for linear processes (including autoregressions) the time-reversed version is distributional equivalent to the original only if the process is Gaussian.

Chernick, Daley, and Littlejohn (1988) provide an example of a first-order autoregression with exponential marginal distributions whose reversed version also has exponential marginals, is first-order Markov, and has a special structure. The process is not time-reversible (in the strict sense where reversibility means distribution equivalence of the two stochastic processes, original and time-reversed) as can be seen by looking at sample paths.

Thombs and Schucany (1990) also present simulation results that show that their method has promise. They did not use the symmetrized distribution for the residuals. In small samples, they concede that some refinements such as the bias-corrected percentile method might be helpful.

Unfortunately, we cannot recommend a particular bootstrap procedure as a "best" approach to bootstrapping time series even for generating prediction intervals for autoregressive time series. The method of Stine (1987) is recommended for use when the distributions are non-Gaussian. For nearly Gaussian time series, the standard methods available in most statistical time series programs are more efficient.

These methods are called model-based and the results do not work well when the form of the model is misspecified. Künsch (1989) was the first to develop the block bootstrap method in the context of stationary time series. It turns out to be a general approach that can be applied in many dependent

data situations including spatial data, M-dependent data, and time series. Lahiri has developed a theory for bootstrapping dependent data predominantly for classes of block bootstrap methods including (1) moving block bootstrap, (2) nonoverlapping block bootstrap, and (3) generalized block bootstrap that includes the circular block bootstrap and the stationary block bootstrap. This work is well-summarized along with other bootstrap methods for dependent data in the text by Lahiri (2003a). Block-based versus model-based bootstrap methods are considered in the next section. Alternative approaches to time series problems are described in Sections 5.4 and 5.5, with block resampling methods contrasted to model-based methods in Section 5.4.

5.4. MODEL-BASED VERSUS BLOCK RESAMPLING

The methods described thus far all fall under the category of model-based resampling methods, because the residuals are generated and resampled based on a time series model [i.e., the AR(1) model in the earlier illustration]. Refinements to the above approach are described in Davison and Hinkley (1997, pp. 389–391).

There they center the residuals by subtracting the average of the residuals. They then use a prescription just as we have described above. However, they point out that the generated series is not stationary. This is due to the initial values. This could be remedied by starting the series in equilibrium or more practically by allowing a "burn-in" period of k observations that are discarded. We choose k so that the series has "reached" stationarity.

To use the model-based approach, we need to know the parameters and the structure of the model, and this is not always easy to discern from the data. If we choose an incorrect structure, the resampled series will have a different structure (which we incorrectly thrust upon it) from the original data and hence will have different statistical properties. So if we know that we have a stationary series but we don't know the structure, analogous to the nonparametric alternative to the distributional assumptions for the observations, we would like a bootstrap resampling structure that doesn't depend on this unknown structure. But what is the time series analog to nonparametric models?

Bose (1988) showed that if an autoregressive process is a "correct" model (or for practical use at least approximately correct) there is an advantage to using the model-based resampling approach, namely, good higher-order asymptotic properties for a wide variety of statistics that can be derived from the model. On the other hand, we could pay a heavy price, in that the estimates could be biased and/or grossly inaccurate if the model structure is wrong. This is very much like the tradeoff we have between parametric and nonparametric inference where the model is the assumed parametric family of distributions for the observations.

A remedy, the block bootstrap, which was first introduced by Carlstein (1986), was further developed by Künsch (1989) and is a method that resamples the time series, in blocks (possibly overlapping blocks). For uncorrelated exchangeable sequences, the original nonparametric bootstrap that resamples the individual observations is appropriate. For stationary time series, successive observations are correlated but observations separated by a large time gap are nearly uncorrelated. This can be seen by the exponentially declining autocorrelation function for a stationary AR (1) model.

A key idea in the development and success of block resampling is that, for stationary series, individual blocks of observations that are separated far enough in time will be approximately uncorrelated and can be treated as exchangeable. So suppose the time series has length $n = bl$. We can generate b nonoverlapping blocks each of length l.

The key idea that underlies this approach is that if the blocks are sufficiently long, each block preserves, in the resampled series, the dependence present in the original data sequence. The resampling or bootstrap scheme here is to resample with replacement from the set of b blocks.

There are several variants on this idea. One is to allow the blocks to overlap. This was one of Künsch's proposals, and it allows for more blocks than if they are required not to overlap.

Suppose we take the first block to be (y_1, y_2, y_3, y_4), the second to be (y_2, y_3, y_4, y_5), the third to be (y_3, y_4, y_5, y_6), and so on. The effect of this approach is that the first $l - 1$ observations from the original series appear in fewer blocks than the rest.

Note that observation y_1 appears in only one block, y_2 appears in only two blocks, and so on. This effect can be overcome by wrapping the data around in a circle (i.e., the last observation in the series is followed again by the first, etc.)

At the time of the writing of the first edition the block bootstrap approach was the subject of much additional research. Many theoretical results and applications have occurred from 1999 to the present (2007).

Professor Lahiri, from Iowa State University, has been one of the prime contributors and has nicely summarized the theoretical properties and (through examples) the applications of the various types of block bootstrap methods for time series and other models of dependent data (including spatial data) in his text (Lahiri, 2003a). I will not cover these topics in depth but rather refer the reader to the literature and the chapters in the Lahiri text as they are discussed.

The various block bootstraps discussed in Lahiri (2003a, Chapter 2) are (1) the moving block bootstrap (MBB), (2) nonoverlapping block bootstrap (NBB), (3) circular block bootstrap (CBB), and (4) the stationary block bootstrap (SBB). I will give formal definitions and discuss these methods in detail later in this section.

Lahiri (2003a) also compares various block methods based on both theory and empirical simulation results (Chapter 5), covers methods for selecting the

block size for the moving block bootstrap (Chapter 7), pointing out how it can be generalized to other block methods, and covers model-based methods (Chapter 8) including the ones discussed in this chapter and more, frequency domain methods (Chapter 9) such as the ones we will discuss in Section 5.5, long-range-dependent models (Chapter 10), heavy-tailed distributions and the estimation of extreme values (Chapter 11), and spatial data (Chapter 12). In Chapter 8 we will cover some of the results for spatial data and in Chapter 9 we will cover situations where the naïve bootstrap fails, which includes the estimation of extreme values. Lahiri shows that for dependent data, the moving block bootstrap also fails if the resample size m is the same as the original sample size n. But as we will see for the independent case in Chapter 9 of this text, an m-out-of-n bootstrap remedies the situation. Lahiri derives the same result for MBB.

Some of the drawbacks of block methods in general are as follows: (1) Resampled blocks do not quite mimic the behavior of the time series, and (2) they have a tendency to weaken the dependency in the series.

Two methods, postblackening and resampling blocks of blocks, both help to remedy these problems. The interested reader should consult Davison and Hinkley (1997, pp. 397–398) for some discussion of these methods.

Another simple way to overcome this difficulty is what is called the stationary block bootstrap, SBB as referred to by Lahiri (2003a) and described in Section 2.7.2 of Lahiri (2003a) with statistical properties for the sample mean given in Section 3.3 of Lahiri (2003a). The stationary block bootstrap is a block bootstrap scheme that instead of having fixed length blocks has a random block length size. The distribution for block length is given using the random length L, where

$$\Pr(L = j) = (1-p)^{j-l}p, \qquad \text{for } j = 1, 2, 3, \ldots, \infty.$$

This length distribution is the geometric distribution with parameter p. The mean block length for L is $\lambda = p^{-1}$. We may choose λ as one might choose the length of a fixed block length. Since $\lambda = 1/p$, determining λ also determines p. The stationary block bootstrap was first described by Politis and Romano (1994a).

It appears that the block resampling method has desirable properties of robustness to model specification in that it applies to a broad class of stationary series. Other variations and some theory related to block resampling can be found in Davison and Hinkley (1997, pp. 401–403) for choice of block length and pp. 405–408 for the underlying theory. Hall (1998) provides an overview of the subject. A very detailed and up-to-date coverage of block resampling can be found in the text Lahiri (2003a) and in the summary article Lahiri (2006) in the book "Frontiers in Statistics" Fan and Koul (2006).

Davison and Hinkley (1997) illustrate the application of block resampling using data on the river heights over time for the Rio Negro. A concern of the study was that there is a trend for heights of the river near Manasas to increase

over time due to deforestation. A test for trend was applied, and there is some evidence that a trend may be present but the statistical test was inconclusive. The trend test was based on a test statistic that was a linear combination of the observations, namely,

$$T = \sum a_i Y_i,$$

where, for $i = 1, 2, 3, \ldots, n$, Y_i is the sequence for river levels at Manasas and

$$a_i = (-1)[1-((i-1)/(n+1))]^{1/2} - i[1-i/(n+1)]^{1/2} \qquad \text{for } i = 1, 2, 3, \ldots, n.$$

The test based on this statistic is optimal for detecting a monotonic trend when the observations are independent and identically distributed (i.e., IID under the null hypothesis). However, the time series data show clear autocorrelation at time lags i. A smoothed version of the Rio Negro river heights (i.e., a centered ten-year moving average) is shown in Figure 5.1 taken from Davison and Hinkley (1997).

The test statistic T above is still used, and its value in the example turns out to be 7.908. But is this statistically significantly large based on the null hypothesis? Instead of using the distribution of the test statistic under the null hypothesis, Davison and Hinkley choose to estimate its null distribution using block resampling. This is a more realistic approach for the Rio Negro data.

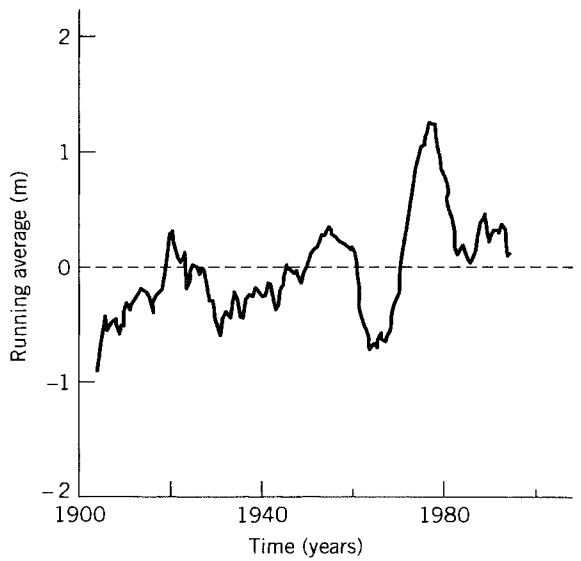

Figure 5.1 Ten-year running average of the Manasas data. [From Davison and Hinkley (1997, Figure. 8.9, p. 403), with permission from Cambridge University Press.]

They compare the stationary bootstrap to a fixed block length method. The purpose is to use the bootstrap to estimate the variance of T under the null hypothesis that the series is stationary but uncorrelated (as opposed to an IID null hypothesis). The asymptotic normality of T is used to do the statistical inference.

Many estimates were obtained using these two methods because various block sizes were used. For the fixed block length method, various fixed block sizes were chosen; for the stationary bootstrap, several average block lengths were specified. The bottom line is that the variance of T is about 25 based on the first 120 time points, but the lowest "reasonable" estimate for the variance of T based on the entire series is approximately 45! This gives us a p-value of 0.12 for the test statistic, indicating a lack of strong evidence for a trend.

When considering autoregressive processes, there are three cases to consider that involve the roots of the characteristic polynomial associated with the time series. See Box and Jenkins (1976) for details about the characteristic polynomial and the relationship of its roots to stationarity. The roots of the characteristic polynomial are found in the complex plane. If all the roots fall inside the unit circle, the time series is stationary. When one or more of the roots lies on the boundary of the unit circle, the time series is nonstationary and called unstable. If all the roots of the characteristic polynomial lie outside the unit circle, the time series is nonstationary and called explosive. In the first case the model-based method that Lahiri calls the autoregressive bootstrap (ARB) can be used. In the case of unstable processes the ARB bootstrap is not consistent but can be made consistent by an m-out-of-n modification that is one of the two. In the case of explosive processes, another remedy is required. The details are given in Chapter 8 of Lahiri (2003a) and the remedies are also covered in Chapter 9, where we cover remedies when the ordinary bootstrap methods fail.

5.5. EXPLOSIVE AUTOREGRESSIVE PROCESSES

An explosive autoregressive process is simply an autoregressive time series whose characteristic polynomial has all its roots outside the unit circle. As such, it is nonstationary process with unusual properties.

Datta (1995) showed that the normalized least-squares estimator of the autoregressive parameters in the explosive case converges to a nonnormal limiting distribution that is dependent on the initial p-observations. As a result, in the explosive case, any bootstrap method needs to use a consistent estimate of joint distribution of the first p-observations. Or alternatively, one can consider the distribution of the parameter by conditioning on the first p-observations. This is how Lahiri (2003a) constructs a consistent ARB estimate. In the explosive case the innovation series may not have a finite expectation; so although in the stationary case the innovations are centered, they cannot be in the explosive case.

The bootstrap observations are generated by the following bootstrap recursion relationship:

$$X_i^* = \hat{\beta}_{1n} X_{i-1}^* + \cdots + \hat{\beta}_{pn} X_{i-p}^* + \varepsilon_i^*, \qquad i \geq p+1.$$

This is well-defined when because of the conditioning argument we set $(X_1^*, \ldots, X_p^*)' \equiv (X_1, \ldots, X_p)'$. The bootstrap error variables ε_i^* are generated at random with replacement from the residuals, $\{\varepsilon_i \equiv X_i - \sum_{j=1}^{p} \beta_{jn} X_{i-j}$: $p+1 \leq i \leq n\}$. Datta has proven [Theorem 3.1 of Datta (1995)] that this ARB is consistent. This result may seem surprising since in the unstable case a similar ARB is not consistent and requires an m-out-of-n bootstrap to be consistent.

5.6. BOOTSTRAPPING STATIONARY ARMA PROCESSES

The stationary ARMA process was first popularized by Box and Jenkins (1970) as a representation that is parsimonious in terms of parameters. The process could also be represented as an infinite moving average process or possibly even an infinite autoregressive process. In practice, since the process is stationary, the series could be approximated by a finite AR process or a finite moving average process. But in either case the number of parameters required in the truncated process is much more than the few AR and MA parameters that appear in the ARMA representation.

Now let $\{X_i\}$, $i \in Z$, be a stationary ARMA (p, q) process satisfying the equation

$$X_i = \sum_{j=1}^{p} \beta_j X_{i-j} + \sum_{j=1}^{q} \alpha_j \varepsilon_{i-j} + \varepsilon_i, \qquad i \in Z,$$

where p and q are integers greater than or equal to 1. The formal description of this model-based bootstrap is involved but can be found in Lahiri (2003a, pp. 214–217). He invokes the standard stationarity and invertibility conditions that Box and Jenkins (1970) generally assume for an ARMA process. Given these conditions, the ARMA process admits both an infinite moving average and an infinite autoregressive representation. The resulting bootstrap is called ARMAB by Lahiri.

5.7. FREQUENCY-BASED APPROACHES

As we have mentioned before, second-order stationary Gaussian processes are strictly stationary as well and are characterized by their mean value function and their autocovariance (or autocorrelation) function. The Fourier transformation of the autocorrelation function is a function of frequency called the spectral density function.

Since a mean zero stationary Gaussian process is characterized by its autocorrelation function and the Fourier transform of the autocorrelation function is invertible, the spectral density function also characterizes the process. This helps explain the importance of the autocorrelation function and the spectral density function in the theory of stationary time series (especially for stationary Gaussian time series). Time series methods based on knowledge or estimates of the autocorrelation function are called time domain methods, and time series methods based on the spectral density function are called frequency domain methods. Brillinger (1981) gives a nice theoretical account of the frequency domain approach to time series.

The periodogram, the sample analog to the spectral density function, and smoothed versions of the periodogram that are estimates of the spectral density function have many interesting and useful properties, which are covered in detail in Brillinger (1981). The Fourier transform of the time series data itself is a complex function called the empirical Fourier transform.

From the theory of stationary processes, it is known that if the process has a well-defined spectral density function and can be represented by an infinite moving average process (representation), then as the series length $n \to \infty$ the real and imaginary parts of this empirical Fourier transform at the Fourier frequencies $\omega_\kappa = 2\pi\kappa/n$ are approximately independent and normally distributed with mean zero and variance $ng(\omega_\kappa)/2$, where $g(\omega_\kappa)$ is the true spectral density function at ω_κ.

This asymptotic result is important and practically useful. The empirical Fourier transform is easy to compute thanks to a technique known as the fast Fourier transform (FFT), and independent normal random variables are easier to deal with than nonnormal correlated variables.

So we use these ideas to construct a bootstrap. Instead of bootstrapping, the original series we can use a parametric bootstrap on the empirical Fourier transformed data. In the frequency domain we have basically an uncorrelated series of observations on the set of Fourier frequencies. The parametric bootstrap samples the indices of the Fourier frequencies with replacement, and then at each sampled frequency a bootstrap observation is generated from the estimated normal distribution. This generates a bootstrap version of the empirical Fourier transform, and then a bootstrap sample for the original series is obtained by inverting this Fourier transform. This idea has been exploited in what Davison and Hinkley (1997) call the phase scrambling algorithm. Although the concept is easy to understand, the actual algorithm is somewhat complicated. The interested reader can see more detail and examples in Davison and Hinkley (1997, pp. 408–409).

Davison and Hinkley (1997) then apply the phase scrambling algorithm to the Rio Negro data. This allows them to compare their previous time domain bootstrapping approach (SSB) with this frequency domain approach. For the null hypothesis, they again assume that the series is an AR (2) process and get an estimate of the variance of the trend estimator T. Using the frequency domain approach, they again determine T to be close to 51. So this result is very close to the result from the previous time domain SSB approach.

Now under the conditions described above, the periodogram has its values at the Fourier frequencies, and they are well-approximated as independent identically distributed exponential random variables. If one is interested only in confidence intervals for the spectral density at certain frequencies or to access variability of estimates that are based on the periodogram values, it is only necessary to resample the periodogram values and you don't have to bother with the empirical Fourier transform or the original time series. This method is called periodogram resampling, and details about the method and its applications to inference about the spectral density function are given by Davison and Hinkley (1997, pp. 412–414).

These frequency domain bootstraps are part of a general category of methods called transformation-based bootstraps where the bootstrapping all takes place on the transformed data and analysis can then be done in the time domain after taking the inverse transform. Lahiri (2003a) covers a number of these approaches on pages 40–41 of the text and uses the acronym TBB for transformation-based bootstrap. Lahiri provides a generalization of a method due originally to Hurvich and Zeger (1987) which is similar conceptually but still different from the method described above from the Davison and Hinkley (1997) text.

Hurvich and Zeger (1987) consider the discrete Fourier transform (DFT) of the data and bootstrap the transformed data rather the original series and then apply the IID nonparametric bootstrap to this transformed data. In this way, they also take advantage of the result in time series analysis that the Fourier transform of the series at distinct frequencies λ_i, where $-\pi < \lambda_i \leq \pi$, are approximately distributed as complex normal and are independent [see Brillinger (1981) or Brockwell and Davis (1991, Chapter 10) for more details].

In Lahiri (2003a), he generalized the approach of Hurvich and Zeger. His development now follows. We let $\theta = \theta(P)$ be the parameter of interest and P the probability measure that generates the observed series. Let T_n be an estimator of θ based on the observed series up to time n. The goal is to approximate the sampling distribution of a studentized statistic R_n that is used to draw inference about θ. The bootstrapping is done on R_n and will be used to get estimates of θ. See Lahiri (2003a, pp. 40–41) and Lahiri (2003a, Chapter 9) for further discussion of the Hurvich and Zeger approach along with more detail about the use of frequency domain bootstraps (FDBs).

5.8. THE SIEVE BOOTSTRAP

Another time domain approach to bootstrap from a stationary stochastic process is called the sieve bootstrap. We let P be the unknown joint probability distribution of the "infinite time series sequence" $\{X_1, X_2, X_3, \ldots, X_n, \ldots\}$. In the IID case we use the empirical distribution F_n or some other estimate of the marginal distribution F and the joint distribution for the first n observations is the product of the F_n s by independence. In this case, because the observations in the time series are dependent, the joint distribution is not the product of the marginal distributions.

The idea of the sieve bootstrap is to choose a sequence of joint distributions $[\tilde{P}_n]_{n>0}$ called a sieve that approximates P. This sequence is such that for each n the probability measure $\tilde{P}_{n+1}$ is a finer approximation to P than the previous member of the sequence $\tilde{P}_n$. This sequence of measures converges to P as $n \to \infty$ in an appropriate sense.

For a large class of stationary processes, Bühlmann (1997) presents a sieve bootstrap method based on a sieve of increasing order, a p^{th}-order autoregressive process. Read Bühlmann (1997) for more details. We will give a brief description similar to the description in Lahiri (2003a). Another approach suggested in Bühlmann (2002a) is based on a variable-length Markov chain. When considering the choice of a sequence of approximating distributions for the sieve, there is a tradeoff between the accuracy of the approximating distribution and its range of validity. This tradeoff is discussed in Lahiri (2002b).

Now let us consider a stationary sequence $[X_n]$ with $n \in Z$, where Z is the set of positive integers with $EX_1 = \mu$ that admits a one-sided infinite moving average representation given by

$$X_i - \mu = \sum_{j=0}^{+\infty} \alpha_j \varepsilon_{i-j}, \qquad i \in Z$$

with $\sum_{j=1}^{+\infty} \beta_j^2 < \infty$. This representation indicates that for autoregressive processes of finite order: $p_n \to \infty$ as $n \to \infty$, but $n^{-1} p_n \to 0$ as $n \to \infty$. The autoregressive representation is given by

$$X_i - \mu = \sum_{j=1}^{p_n} \beta_j (X_{i-j} - \mu) + \varepsilon_i, \qquad i \in Z.$$

Using the autoregressive representation above, we fit the parameters β_j to an AR (p_n) model. The sieve is then based on the sequence of probability measures associated with the fitted AR (p_n) model. For more details see Lahiri (2003a, pp. 41–43). In his paper, Bühlmann (1997) establishes the consistency of this autoregressive sieve bootstrap.

5.9. HISTORICAL NOTES

The use of ARIMA and seasonal ARIMA models for forecasting and control problems was first popularized by Box and Jenkins (1970, 1976). This work was recently updated in Box, Jenkins, and Reinsel (1994). A classic theoretical text on time series analysis is Anderson (1971).

A popular common theoretical account of time series analysis is Brockwell and Davis (1991), which covers both time domain and frequency domain

analysis. Fuller (1976) is another excellent text at the high undergraduate or graduate school level that also covers both domains well. A couple of articles by Tong (1983, 1990) deal with nonlinear time series models.

Bloomfield (1976), Brillinger (1981), and Priestley (1981) are all time series texts that concentrate strictly on the frequency domain approach. Hamilton (1994) is another major text on time series. Braun and Kulperger (1997) did some work on the Fourier transform approach to bootstrapping.

The idea of bootstrapping residuals was described in Efron (1982a) in the context of regression. It is not clear who was the first to make the obvious extension of this to ARMA time series models. Findley (1986) was probably the first to point out some difficulties with the bootstrap approach particularly regarding the estimation of mean-square error.

Efron and Tibshirani (1986) showed how bootstrapping residuals provided improved standard error estimates for the autoregressive parameter estimates for the Wolfer sunspot data. Stine (1987) and Thombs and Schucany (1990) provide refinements to obtain better prediction intervals. Other empirical studies are Chatterjee (1986) and Holbert and Son (1986). McCullough (1994) provides an application of bootstrapping prediction intervals for AR(p) models.

Results for nonstationary autoregressions appear in Basawa, Mallik, McCormick, and Taylor (1989) and Basawa, Mallik, McCormick, Reeves, and Taylor (1991a,b). Theoretical developments are given in Bose (1988) and Künsch (1989).

Künsch (1989) is an attempt to develop a general theory for bootstrapping stationary time series. Bose (1988) also shows good asymptotic higher-order properties when applying model-based resampling to a wide class of statistics used with autoregressive processes.

Shao and Yu (1993) apply the bootstrap for the sample mean in a general class of time series, namely, stationary mixing processes. Hall and Jing (1996) apply resampling methods to general dependent data situations. Lahiri (2003a) is the new authoritative and up-to-date text to cover time series and other dependent data problems, including extremes in stationary processes and spatial data models.

Model-based resampling for time series was discussed by Freedman (1984), Freedman and Peters (1984a,b), Swanepoel and van Wyk (1986), and Efron and Tibshirani (1986). Li and Maddala (1996) provide a survey of related time domain literature on bootstrapping with emphasis on econometric applications.

Peters and Freedman (1985) deal with bootstrapping for the purpose of comparing competing forecasting equations. Tsay (1992) provides an applied account of parametric bootstrapping of time series.

Kabaila (1993a) discusses prediction in time series. Stoffer and Wall (1991) apply the bootstrap to state space models for time series. Chen, Davis, Brockwell, and Bai (1993) use model-based resampling to determine the appropriate order for an autoregressive model.

Good higher-order asymptotic properties for block resampling [similar to the work in Bose (1988)] have been demonstrated by Lahiri (1991) and Götze and Künsch (1996). Davison and Hall (1993) show that good asymptotic properties for the bootstrap generally depend crucially on the choice of a variance estimate. Lahiri (1992b) applies an Edgeworth correction in using the moving block bootstrap for both stationary and nonstationary time series models.

Block resampling was introduced by Carlstein (1986). The key breakthrough with the block resampling approach came later when Künsch (1989) provided many of the important theoretical developments on the block bootstrap idea and introduced the idea of overlapping blocks.

The stationary bootstrap was introduced by Politis and Romano (1994a). They also proposed the circular block bootstrap in an earlier work, Politis and Romano (1992a). Liu and Singh (1992b) obtain general results for moving block jackknife and bootstrap approaches to general types of weak dependence. Liu (1988) and Liu and Singh (1995) deal with bootstrap approaches to general data sets that are not IID. For the most recent developments in block bootstrap theory and methods see Lahiri (2003a) and Lahiri (2006).

Theoretical developments for general block resampling schemes followed the work of Künsch, in the articles Politis and Romano (1993a, 1994b), Bühlmann and Künsch (1995), and Lahiri (1995). Issues of block length are addressed by Hall, Horowitz, and Jing (1995). Lahiri (2003a, pp. 175–186) covers optimal block sizes for estimating bias, variance, and distribution quantiles. He covers much of the research from Hall, Horowitz, and Jing (1995).

Fan and Hung (1997) use balanced resampling (a variance reduction technique that is covered in Chapter 7) to bootstrap finite Markov chains. Liu and Tang (1996) use bootstrap method for control charting in both the independent and dependent situations.

Frequency domain resampling has been discussed by Franke and Hardle (1992) with an analogy to nonparametric regression. Janas (1993) and Dahlhaus and Janas (1996) extended these results. Politis, Romano, and Lai (1992) provide bootstrap confidence bands for spectra and cross-spectra (frequency domain analog, respectively, autocorrelation and cross-correlation functions in the time domain).

The sieve bootstrap was introduced for a class of stationary stochastic processes (that admit an infinite moving average representation) by Bühlmann (1997). It is also covered in Section 2.10 of Lahiri (2003a).