Introduction to Statistical Time Series
WAYNE A. FULLER

CHAPTER 9

Regression, Trend, and Seasonality

The majority of the theory presented to this point assumes that the time series under investigation is stationary. Many time series encountered in practice are not stationary. They may fail for any of several reasons:

1. The mean is a function of time, other than the constant function.
2. The variance is a function of time, other than the constant function.
3. The time series is generated by a nonstationary stochastic mechanism.

We consider processes of the third type in Chapter 10. Examples of hypothesized nonstationarity of the first kind occur most frequently in the applied literature, but there are also many examples of heterogeneous variances.

The traditional model for economic time series is

$$Y_t = T_t + S_t + Z_t, \tag{9.0.1}$$

where T_t is the "trend" component, S_t is the "seasonal" component, and Z_t is the "irregular" or "random" component. In our terminology, Z_t is a stationary time series. Often T_t is further decomposed into "cyclical" and "long-term" components. Casual inspection of many economic time series leads one to conclude that the mean is not constant through time, and that monthly or quarterly time series display a type of "periodic" behavior wherein peaks and troughs occur at "nearly the same" time each year. However, these two aspects of the time series typically do not exhaust the variability, and therefore the random component is included in the representation.

While the model (9.0.1) is an old one indeed, a precise definition of the components has not evolved. This is not necessarily to be viewed as a weakness of the representation. In fact, the terms acquire meaning only when a procedure is used to estimate them, and the meaning is determined by the procedure. The reader should not be disturbed by this. An example from another area might serve to clarify the issue. The "intelligence quotient" of a person is the person's score on an I.Q. test, and I.Q. acquires meaning only in the context of the procedure used to

determine it. Although the test may be based on a theory of mental behavior, the I.Q. test score should not be taken to be the only estimator of that attribute of humans we commonly call intelligence.

For a particular economic time series and a particular objective, one model and estimation procedure for trend and seasonality may suffice; for a different time series or a different objective, an alternative specification may be preferred.

We shall now study some of the procedures used to estimate trend and seasonality and (or) to reduce a nonstationary time series to stationarity. Since the mean function of a time series may be a function of other time series or of fixed functions of time, we are led to consider the estimation of regression equations wherein the error is a time series.

9.1. GLOBAL LEAST SQUARES

In many situations we are able to specify the mean of a time series to be a simple function of time, often a low order polynomial in t or trigonometric polynomial in t. A sample of n observations can then be represented by

$$\mathbf{y} = \mathbf{\Phi}\boldsymbol{\beta} + \mathbf{z}, \tag{9.1.1}$$

where $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_r)$ is a vector of unknown parameters,

$$\mathbf{y}' = (Y_1, Y_2, \ldots, Y_n), \qquad \mathbf{z}' = (Z_1, Z_2, \ldots, Z_n),$$

$$\mathbf{\Phi} = (\boldsymbol{\varphi}_{\cdot 1}, \boldsymbol{\varphi}_{\cdot 2}, \ldots, \boldsymbol{\varphi}_{\cdot r}),$$

$\boldsymbol{\varphi}_{\cdot i}$, $i = 1, 2, \ldots, r$, are n-dimensional column vectors, and Z_t is a zero mean time series. In many situations, we may be willing to assume that Z_t is stationary, but we also consider estimation under weaker assumptions.

The elements of $\boldsymbol{\varphi}_{\cdot i}$, say φ_{ti}, are functions of time. For example, we might have $\varphi_{t1} \equiv 1$, $\varphi_{t2} = t$, $\varphi_{t3} = t^2$. The elements of $\boldsymbol{\varphi}_{\cdot i}$ may also be random functions of time, for example, a stationary time series. In the random case, we shall assume that $\mathbf{z}$ is independent of $\mathbf{\Phi}$ and investigate the behavior of the estimators conditional on a particular realization of φ_{ti}. Thus, in this section, all φ_{ti} will be treated as fixed functions of time. Notice that $\mathbf{y}$, $\mathbf{\Phi}$, $\boldsymbol{\varphi}_{\cdot i}$, and $\mathbf{z}$ might properly be subscripted by n. To simplify the notation, we have omitted the subscript.

The simple least squares estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_S = (\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{y}. \tag{9.1.2}$$

In (9.1.2) and throughout this section, we assume $\mathbf{\Phi}'\mathbf{\Phi}$ is nonsingular. Assume that the time series is such that the matrix $\mathbf{V}_{zz} = E\{\mathbf{z}\mathbf{z}'\}$ is nonsingular. Then the generalized least squares (best linear unbiased) estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_G = [\mathbf{\Phi}'\mathbf{V}_{zz}^{-1}\mathbf{\Phi}]^{-1}\mathbf{\Phi}'\mathbf{V}_{zz}^{-1}\mathbf{y}. \tag{9.1.3}$$

The covariance matrix of $\hat{\boldsymbol{\beta}}_S$ is

$$E\{(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta})'\} = (\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\mathbf{V}_{zz}\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}, \tag{9.1.4}$$

while that of $\hat{\boldsymbol{\beta}}_G$ is

$$E\{(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})'\} = (\boldsymbol{\Phi}'\mathbf{V}_{zz}^{-1}\boldsymbol{\Phi})^{-1}.$$

It is well known that $\hat{\boldsymbol{\beta}}_G$ is superior to $\hat{\boldsymbol{\beta}}_S$ in that the variance of any linear contrast $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}_G$ is no larger than the variance of the corresponding linear contrast $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}_S$. However, the construction of $\hat{\boldsymbol{\beta}}_G$ requires knowledge of $\mathbf{V}_{zz}$, and generally $\mathbf{V}_{zz}$ is not known. In fact, one may wish to estimate the mean function of Y_t prior to investigating the covariance structure of the error time series. Therefore, the properties of the simple least squares estimator are of interest.

To investigate the large sample behavior of least squares estimators, let

$$\begin{aligned}\mathbf{D}_n &= \operatorname{diag}\{d_{n11}, d_{n22}, \ldots, d_{nrr}\} \\ &\stackrel{\text{def}}{=} \operatorname{diag}\{(\boldsymbol{\varphi}_{\cdot 1}'\boldsymbol{\varphi}_1)^{1/2}, (\boldsymbol{\varphi}_{\cdot 2}'\boldsymbol{\varphi}_2)^{1/2}, \ldots, (\boldsymbol{\varphi}_{\cdot r}'\boldsymbol{\varphi}_r)^{1/2}\},\end{aligned}$$

where $\boldsymbol{\varphi}_{\cdot i}$ is the ith column of the matrix $\boldsymbol{\Phi}$. The least squares estimator is consistent for $\boldsymbol{\beta}$ under mild assumptions.

Proposition 9.1.1. Let the model (9.1.1) hold where Z_t is a zero mean time series and φ_{ti}, $i = 1, 2, \ldots, r$, $t = 1, 2, \ldots$, are fixed functions of time. Assume

$$\lim_{n\to\infty} d_{nii} = \infty, \qquad i = 1, 2, \ldots, r, \tag{9.1.5}$$

$$\lim_{n\to\infty} \mathbf{D}_n^{-1}\boldsymbol{\Phi}'\boldsymbol{\Phi}\mathbf{D}_n^{-1} = \mathbf{A}_0, \tag{9.1.6}$$

$$\lim_{n\to\infty} \mathbf{D}_n^{-1}\boldsymbol{\Phi}'\mathbf{V}_{zz}\boldsymbol{\Phi}\mathbf{D}_n^{-1} = \mathbf{B}, \tag{9.1.7}$$

where $\mathbf{A}_0$ is nonsingular. Then $\hat{\beta}_{Sj} - \beta_{Sj}^0 = O_p(d_{njj}^{-1})$, where $\hat{\beta}_{Sj}$ is the jth element of $\hat{\boldsymbol{\beta}}_S$ and β_{Sj}^0 is the true parameter value.

Proof. The covariance matrix of the least squares estimator is given in (9.1.4), and the variance of the normalized estimator is

$$\mathbf{V}\{\mathbf{D}_n(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta})\} = \mathbf{D}_n(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\mathbf{D}_n\mathbf{D}_n^{-1}\boldsymbol{\Phi}'\mathbf{V}_{zz}\boldsymbol{\Phi}\mathbf{D}_n^{-1}\mathbf{D}_n(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\mathbf{D}_n.$$

It follows that $\mathbf{D}_n(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}^0) = O_p(1)$, because

$$\lim_{n\to\infty} \mathbf{V}\{\mathbf{D}_n(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}^0)\} = \mathbf{A}_0^{-1}\mathbf{B}\mathbf{A}_0^{-1}. \qquad \blacktriangle$$

Under stronger assumptions, the limiting distribution of the least squares estimator can be obtained. Assume that the real valued functions φ_{tt} of t satisfy

$$\lim_{n\to\infty} \sup_{1\le t\le n} \left(\sum_{j=1}^{n} \varphi_{ji}^2\right)^{-1} \varphi_{ti}^2 = 0, \qquad i = 1, 2, \ldots, r, \tag{9.1.8}$$

and

$$\lim_{n\to\infty} d_{nii}^{-1} d_{njj}^{-1} \sum_{t=1}^{n-h} \varphi_{ti}\varphi_{t+h,j} = a_{hij} = a_{-h,ij}, \tag{9.1.9}$$

$h = 0, 1, 2, \ldots$, $i, j = 1, 2, \ldots, r$, where $\mathbf{A}_0$ is nonsingular and $\mathbf{A}_h$ is the matrix with typical element a_{hij}. The assumptions on the φ_{ti} are quite modest and, for example, are satisfied by polynomial, trigonometric polynomial, and logarithmic functions of time. The following theorem demonstrates that the limiting distribution of the vector of standardized estimators is normal for a wide class of stationary time series.

Theorem 9.1.1. Let Z_t be a stationary time series defined by

$$Z_t = \sum_{j=0}^{\infty} v_j e_{t-j}, \qquad t = 0, \pm 1, \pm 2, \ldots,$$

where $\{v_j\}$ is absolutely summable and the e_t are independent $(0, \sigma^2)$ random variables with distribution functions $F_t(e)$ such that

$$\lim_{\delta\to\infty} \sup_{t=1,2,\ldots} \int_{|e|>\delta} e^2 \, dF_t(e) = 0.$$

Let the φ_{ti}, $i = 1, 2, \ldots, r$, $t = 1, 2, \ldots$, be fixed functions of time satisfying the assumptions (9.1.8) and (9.1.9). Then

$$\mathbf{D}_n(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}) \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{A}_0^{-1}\mathbf{B}\mathbf{A}_0^{-1}), \tag{9.1.10}$$

where $\mathbf{A}_0$ is the matrix with elements a_{0ij} defined in (9.1.9), $\mathbf{B}$ is defined in (9.1.7), and $\hat{\boldsymbol{\beta}}_S$ is defined in (9.1.2). The covariance matrix is nonsingular if $\mathbf{B}$ of (9.1.7) is nonsingular.

Proof. Consider the linear combination $\Sigma_{t=1}^{n} c_t Z_t$, where

$$c_t = \sum_{i=1}^{r} \lambda_i d_{nii}^{-1} \varphi_{ti}$$

and the λ_i are arbitrary real numbers. Now, by our assumptions,

$$\begin{aligned}\lim_{n\to\infty} \sum_{t=1}^{n-h} c_t c_{t+h} &= \lim_{n\to\infty} \sum_{t=1}^{n-h} \sum_{i=1}^{r} \sum_{j=1}^{r} \lambda_i \lambda_j d_{nii}^{-1} d_{njj}^{-1} \varphi_{ti}\varphi_{t+h,j} \\ &= \sum_{i=1}^{r} \sum_{j=1}^{r} \lambda_i \lambda_j a_{hij} \overset{\text{(say)}}{=} g_\lambda(h),\end{aligned}$$

and c_t is completely analogous to $(\Sigma_{j=1}^{n} C_j^2)^{-1/2} C_t$ of Theorem 6.3.4. By Theorem 6.3.4, $\Sigma_{t=1}^{n} c_t Z_t$ converges to a normal random variable with variance

$$\sum_{h=-\infty}^{\infty} g_\lambda(h)\gamma_Z(h) = \sum_{i=1}^{r}\sum_{j=1}^{r} \lambda_i \lambda_j b_{ij} ,$$

where $b_{ij} = \Sigma_{h=-\infty}^{\infty} a_{hij}\gamma_Z(h)$ is the ijth element of $\mathbf{B}$ of (9.1.7). Since $\lim_{n\to\infty} \mathbf{D}_n^{-1}\mathbf{\Phi}'\mathbf{\Phi}\mathbf{D}_n^{-1} = \mathbf{A}_0$ and since λ was arbitrary, the result follows from Theorem 5.3.3. ▲

The matrix $\mathbf{A}_h$ with ijth element equal to a_{hij} of (9.1.9) is analogous to the matrix $\mathbf{\Gamma}(h)$ of Section 4.4, and therefore the spectral representation

$$\mathbf{A}_h = \int_{-\pi}^{\pi} e^{i\omega h}\, d\mathbf{M}(\omega) \tag{9.1.11}$$

holds, where $\mathbf{M}(\omega_2) - \mathbf{M}(\omega_1)$ is a positive semidefinite Hermitian matrix for all $-\pi \le \omega_1 < \omega_2 \le \pi$, and $\mathbf{A}_0 = \mathbf{M}(\pi) - \mathbf{M}(-\pi)$. We state without proof some of the results of Grenander and Rosenblatt (1957), which are based on this representation.

Theorem 9.1.2. Let the assumptions (9.1.1), (9.1.8), and (9.1.9) hold, and let the spectral density of the stationary time series Z_t be positive for all ω. Then

$$\lim_{n\to\infty} \mathbf{D}_n^{-1}\mathbf{\Phi}'\mathbf{V}_{zz}\mathbf{\Phi}\mathbf{D}_n^{-1} = 2\pi \int_{-\pi}^{\pi} f_Z(\omega)\, d\mathbf{M}(\omega) ,$$

$$\lim_{n\to\infty} \mathbf{D}_n^{-1}\mathbf{\Phi}'\mathbf{V}_{zz}^{-1}\mathbf{\Phi}\mathbf{D}_n^{-1} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_Z^{-1}(\omega)\, d\mathbf{M}(\omega) ,$$

$$\lim_{n\to\infty} \mathbf{E}\{\mathbf{D}_n(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta})'\mathbf{D}_n\} = 2\pi\mathbf{A}_0^{-1}\left[\int_{-\pi}^{\pi} f_Z(\omega)\, d\mathbf{M}(\omega)\right]\mathbf{A}_0^{-1} , \tag{9.1.12}$$

$$\lim_{n\to\infty} \mathbf{E}\{\mathbf{D}_n(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})'\mathbf{D}_n\} = \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} f_Z^{-1}(\omega)\, d\mathbf{M}(\omega)\right]^{-1} , \tag{9.1.13}$$

where $f_Z(\omega)$ is the spectral density function of the process Z_t.

Proof. See Grenander and Rosenblatt (1957). ▲

The asymptotic efficiency of the simple least squares and the generalized least squares is the same if the two covariance matrices (9.1.12) and (9.1.13) are equal. Following Grenander and Rosenblatt, we denote the set of points of increase of $\mathbf{M}(\omega)$ by S. That is, the set S contains all ω such that for any interval (ω_1, ω_2)

where $\omega_1 < \omega < \omega_2$, $\mathbf{M}(\omega_2) - \mathbf{M}(\omega_1)$ is a positive semidefinite matrix and not the null matrix.

Theorem 9.1.3. Let the assumptions (9.1.1), (9.1.8), and (9.1.9) hold, and let the spectral density of the stationary time series Z_t be positive for all ω. Then the simple least squares estimator and the generalized least squares estimator have the same asymptotic efficiency if and only if the set S is composed of q distinct points, $\omega_1, \omega_2, \ldots, \omega_q$, $q \leq r$.

Proof. See Grenander and Rosenblatt (1957). ▲

It can be shown that polynomials and trigonometric polynomials satisfy the conditions of Theorem 9.1.3. For example, we established the result for the special case of a constant mean function and autoregressive Y_t in Section 6.1. One may easily establish for the linear trend that

$$\lim_{n\to\infty} \frac{1}{2\pi n}\left(\frac{6}{n(n+1)(2n+1)}\right)\left(\sum_{s=1}^{n}\sum_{t=1}^{n} te^{-\epsilon\omega t}se^{-\epsilon\omega s}\right) = \begin{cases} 0, & \omega \neq 0, \\ \dfrac{3}{8\pi}, & \omega = 0, \end{cases}$$

from which it follows that the set S of points of increase is $\omega_1 = 0$.

The reader should not forget that these are asymptotic results. If the sample is of moderate size, it may be desirable to estimate $\mathbf{V}$ and transform the data to obtain final estimates of the trend function. Also, the simple least squares estimators may be asymptotically efficient, but it does not follow that the simple formulas for the estimated variances of the coefficients are consistent. In fact, the estimated variances may be badly biased. See Section 9.7.

9.2. GRAFTED POLYNOMIALS

In many applications the mean function is believed to be a "smooth" function of time but the functional form is not known. While it is difficult to define the term "smooth" in this context, several aspects of functional behavior can be identified. For a function defined on the real line, the function would be judged to be continuous and, in most situations, to have a continuous first derivative. This specification is incomplete in that one also often expects few changes in the sign of the first derivative.

Obviously low order polynomials satisfy the stated requirements. Also, by the Weierstrass approximation theorem, we know that any continuous function defined on a compact interval of the real line can be uniformly approximated by a polynomial. Consequently, polynomials have been heavily used to approximate the mean function. However, if the mean function is such that higher order polynomials are required, the approximating function may be judged unsatisfactory in that it contains a large number of changes in sign of the derivative. An alternative

approximation that generally overcomes this problem is to approximate segments of the function by low order polynomials and then join the segments to form a continuous function. The segments may be joined together so that the derivatives of a desired order are continuous.

Our presentation is based on the fact that, on the real line, the functions

$$\varphi_{ti} = \begin{cases} (t - A_i)^k, & t > A_i, \\ 0 & \text{otherwise}, \end{cases}$$

where $i = 1, 2, \ldots, M$, k and M are positive integers, and $A_i > A_{i-1}$, are continuous with continuous $(k-1)$st derivatives. It follows that

$$g(t) = b_0 + b_1 t + \cdots + b_k t^k + \sum_{i=1}^{M} b_{k+i} \varphi_{ti}$$

is also continuous with continuous $(k-1)$st derivative. We call the function $g(t)$ a *grafted polynomial of degree k.*

To illustrate the use of grafted polynomials in the estimation of the mean function of a time series let us make the following assumption: "The time series may be divided into periods of length A such that the mean function in each period is adequately approximated by a quadratic in time. Furthermore the mean function possesses a continuous first derivative." Let n observations indexed by $t = 1, 2, \ldots, n$ be available. We construct the functions φ_{ti}, $i = 1, 2, \ldots, M$, for $k = 2$ and $A_i = Ai$, where M is an integer such that $|A(M+1) - n| < A$. In this formulation we assume that, if $n \neq A(M+1)$, the last interval is the only one whose length is not A. At the formal level all we need do is regress our time series upon $t, t^2, \varphi_{t1}, \varphi_{t2}, \ldots, \varphi_{tM}$ to obtain estimates of b_j, $j = 0, 1, \ldots, M+2$. However, if M is at all large, we can expect to encounter numerical problems in obtaining the inverse and regression coefficients. To reduce the possibility of numerical problems, we suggest that the φ's be replaced by the linear combinations

$$w_{ti} = \varphi_{ti} - 3\varphi_{t,i+1} + 3\varphi_{t,i+2} - \varphi_{t,i+3}, \qquad i = 1, 2, \ldots, M,$$

where, for convenience, we define

$$\varphi_{t,M+1} = \varphi_{t,M+2} = \varphi_{t,M+3} = 0$$

for all t. Note that

$$w_{ti} = \begin{cases} (t - Ai)^2, & Ai < t < (i+1)A, \\ (t - Ai)^2 - 3[t - (i+1)A]^2, & (i+1)A \leq t < (i+2)A, \\ [t - (i+3)A]^2, & (i+2)A \leq t < (i+3)A, \\ 0 & \text{otherwise}. \end{cases}$$

Since the function w_{ti} is symmetric about $(i + 1.5)A$, the w-variables can be

written down immediately. The w_{ti} remain correlated, but there should be little trouble in obtaining the inverse.

This procedure is felt to have merit when one is called on to extrapolate a time series. Since polynomials tend to plus or minus infinity as t increases, practitioners typically hesitate to use high order polynomials in extrapolation. Although a time series may display a nonlinear trend, we might wish to extrapolate on the basis of a linear trend. To accomplish this, we approximate the trend of the last K periods of a time series of n observations by a straight line tangent to a higher degree trend for the earlier portion of the time series. If the mean function for the first part of the sample is to be approximated by a grafted quadratic one could construct the regression variables

$$\varphi_{t1} = t,$$

$$\varphi_{t,1+i} = \begin{cases} [n - K - (i-1)A - t]^2, & t < n - K - (i-1)A, \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \ldots, M$, where $|K + AM - n| < A$. Using these variables, the estimated regression equation can be written as

$$\hat{b}_0 + \sum_{j=1}^{M+1} \hat{b}_j \varphi_{tj},$$

where the $\hat{b}$'s are the estimated regression coefficients. The forecast equation for the mean function is $\hat{b}_0 + \hat{b}_1 t$, $t = n+1, n+2, \ldots$.

We can also estimate a mean function that is linear for both the first and last portions of the observational period. Using a grafted quadratic in periods of length A for the remainder of the function, the required regression variables would be, for example,

$$\varphi_{t1} = t,$$

$$\varphi_{t,1+i} = \begin{cases} [t - K - A(i-1)]^2, & K + A(i-1) < t \leq K + Ai, \\ A^2 + 2A(t - K - Ai), & t > K + Ai, \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \ldots, M$, where $n > K + MA$ and the mean function is linear for $t < K$ and $t > K + MA$. If M is large, a transformation similar to that discussed above can be used to reduce computational problems.

Example 9.2.1. Yields of wheat in the United States for the period 1908 through 1991 are given in Table 9.2.1. In this example, we use only the data from 1908 through 1971. Yields for the first 20 to 30 years of the period are relatively constant, but there is a definite increase in yields from the late 1930s through 1971. As an approximation to the trend line, we fit a function that is constant for the first 25 years, increases at a quadratic rate until 1961, and is linear from 1961

Table 9.2.1. United States Wheat Yields (1908 to 1991)

Year	Yield	Year	Yield	Year	Yield
1908	14.3	1936	12.8	1964	25.8
1909	15.5	1937	13.6	1965	26.5
1910	13.7	1938	13.3	1966	26.3
1911	12.4	1939	14.1	1967	25.9
1912	15.1	1940	15.3	1968	28.4
1913	14.4	1941	16.8	1969	30.6
1914	16.1	1942	19.5	1970	31.0
1915	16.7	1943	16.4	1971	33.9
1916	11.9	1944	17.7	1972	32.7
1917	13.2	1945	17.0	1973	31.7
1918	14.8	1946	17.2	1974	27.3
1919	12.9	1947	18.2	1975	30.6
1920	13.5	1948	17.9	1976	30.3
1921	12.7	1949	14.5	1977	30.7
1922	13.8	1950	16.5	1978	31.4
1923	13.3	1951	16.0	1979	34.2
1924	16.0	1952	18.4	1980	33.5
1925	12.8	1953	17.3	1981	34.5
1926	14.7	1954	18.1	1982	35.5
1927	14.7	1955	19.8	1983	39.4
1928	15.4	1956	20.2	1984	38.8
1929	13.0	1957	21.8	1985	37.5
1930	14.2	1958	27.5	1986	34.4
1931	16.3	1959	21.6	1987	37.7
1932	13.1	1960	26.1	1988	34.1
1933	11.2	1961	23.9	1989	32.7
1934	12.1	1962	25.0	1990	39.5
1935	12.2	1963	25.2	1991	34.3

SOURCE: U.S. Department of Agriculture, *Agricultural Statistics*, various issues.

to 1971. A continuous function with continuous derivative that satisfies these requirements is

$$\varphi_{t1} = \begin{cases} 0, & 1 \leq t \leq 25, \\ (t-25)^2, & 25 \leq t \leq 54, \\ 841 + 58(t-54), & 54 \leq t, \end{cases} \tag{9.2.1}$$

where $t = 1$ for 1908.

The trend line obtained by regressing the 64 yield observations on the vector $(1, \varphi_{t1})$ is displayed in Figure 9.2.1. The equation for the estimated trend line is

$$\hat{Y}_t = 13.97 + 0.0123\varphi_{t1}$$

and the residual mean square is 2.80.

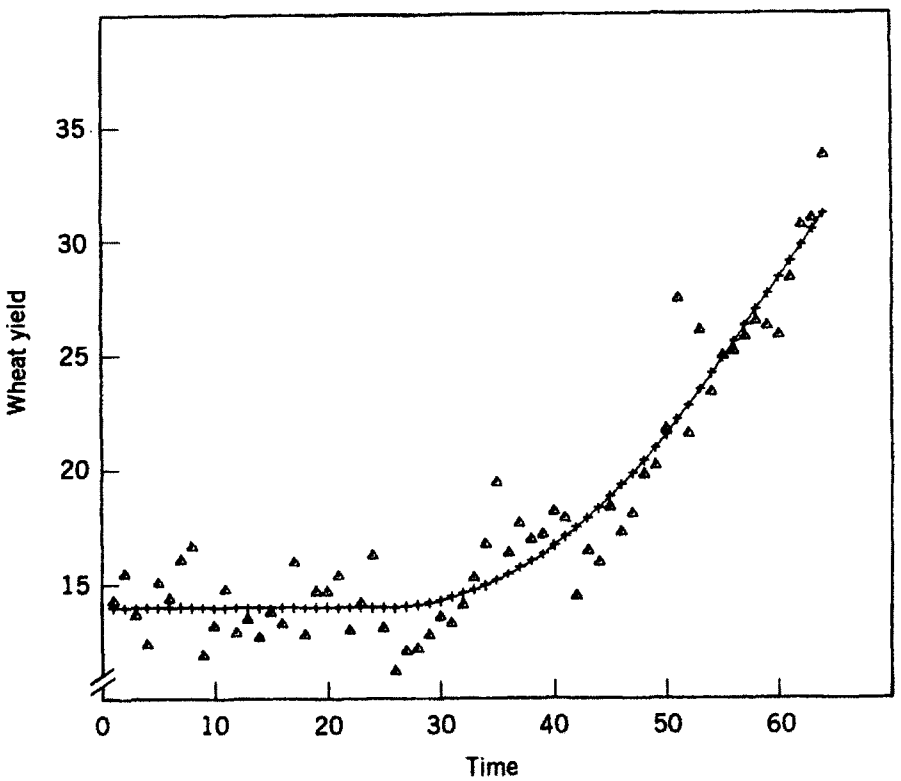

Figure 9.2.1. U.S. wheat yields 1908–1971 and grafted quadratic trend.

The trend in the data and the properties of the time series will be investigated further in Example 9.5.1 and Example 9.7.2. ▲▲

In the example and in our presentation we assumed that one was willing to specify the points at which the different polynomials are joined. The problem of using the data to estimate the join points has been considered by Gallant and Fuller (1973).

Grafted polynomials passing through a set of given points and joined in such a way that they satisfy certain restrictions on the derivatives are called *spline functions* in approximation theory. A discussion of such functions and their properties is contained in Ahlberg, Nilson, and Walsh (1967), Greville (1969), and Wahba (1990).

9.3. ESTIMATION BASED ON LEAST SQUARES RESIDUALS

9.3.1. Estimated Autocorrelations

One of the reasons for estimating the trend function is to enable one to investigate the properties of the time series Z_t. Let the regression model be that defined in (9.1.1):

$$\mathbf{y} = \boldsymbol{\Phi}\boldsymbol{\beta} + \mathbf{z}. \tag{9.3.1}$$

The calculated residuals using the simple least squares estimator are

$$\hat{Z}_t = Y_t - \boldsymbol{\varphi}_{t\cdot}\hat{\boldsymbol{\beta}}_S = Z_t - \boldsymbol{\varphi}_{t\cdot}(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}), \tag{9.3.2}$$

where

$$\hat{\boldsymbol{\beta}}_S = (\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\mathbf{y}$$

and $\boldsymbol{\varphi}_{t\cdot}$ is the tth row of $\boldsymbol{\Phi}$. The estimated autocovariance of Z_t computed using the estimated Z_t is

$$\hat{\gamma}_{\hat{Z}}(h) = \frac{1}{n}\sum_{t=1}^{n-h}\hat{Z}_t\hat{Z}_{t+h}, \qquad h = 0, 1, 2, \ldots, n-1. \tag{9.3.3}$$

Theorem 9.3.1. Let Y_t be defined by the model (9.3.1) where $Z_t = \sum_{j=0}^{\infty}\alpha_j e_{t-j}$, $\{\alpha_j\}$ is absolutely summable, $\{e_t\}$ is a sequence of independent $(0, \sigma^2)$ random variables with bounded $2+\delta$ $(\delta > 0)$ moments, and the φ_{ti}, $i = 1, 2, \ldots, r$, $t = 1, 2, \ldots$, are fixed functions of time satisfying the assumptions (9.1.5) and (9.1.6). Then

$$\hat{\gamma}_{\hat{Z}}(h) - \hat{\gamma}_Z(h) = O_p(n^{-1}),$$

where $\hat{\gamma}_{\hat{Z}}(h)$ is defined in (9.3.3) and

$$\hat{\gamma}_Z(h) = \frac{1}{n}\sum_{t=1}^{n-h}Z_tZ_{t+h}, \qquad h = 0, 1, 2, \ldots, n-1.$$

Proof. We have, for $h = 0, 1, \ldots, n-1$,

$$\begin{aligned}\sum_{t=1}^{n-h}\hat{Z}_t\hat{Z}_{t+h} = \sum_{t=1}^{n-h}Z_tZ_{t+h} &- \sum_{t=1}^{n-h}Z_t\boldsymbol{\varphi}_{t+h,\cdot}\mathbf{D}_n^{-1}\mathbf{D}_n(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}) \\ &- \sum_{t=1}^{n-h}Z_{t+h}\boldsymbol{\varphi}_{t\cdot}\mathbf{D}_n^{-1}\mathbf{D}_n(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}) \\ &+ \sum_{t=1}^{n-h}(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta})'\mathbf{D}_n\mathbf{D}_n^{-1}\boldsymbol{\varphi}_{t\cdot}'\boldsymbol{\varphi}_{t+h,\cdot}\mathbf{D}_n^{-1}\mathbf{D}_n(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}),\end{aligned}$$

where $\mathbf{D}_n$ is the diagonal matrix whose elements are the square roots of the elements of the principal diagonal of $\boldsymbol{\Phi}'\boldsymbol{\Phi}$. By the absolute summability of $\gamma_Z(h)$, the assumption (9.1.5), and the assumption (9.1.6)

$$\sum_{t=1}^{n-h}Z_t\boldsymbol{\varphi}_{t+h,\cdot}\mathbf{D}_n^{-1} = O_p(1),$$

$$\mathbf{D}_n(\hat{\boldsymbol{\beta}}_S - \boldsymbol{\beta}) = O_p(1).$$

Since each element of the $r \times r$ matrix

$$\sum_{t=1}^{n-h} \mathbf{D}_n^{-1} \boldsymbol{\varphi}_{t\cdot}' \boldsymbol{\varphi}_{t+h,\cdot} \mathbf{D}_n^{-1}$$

is less than one in absolute value, the result follows. ▲

It is an immediate consequence of Theorem 9.3.1 that the limiting behavior of the estimated autocovariances and autocorrelations defined by Theorem 6.3.5 and Corollary 6.3.5 also holds for autocovariances and autocorrelations computed using $\hat{Z}_t$ in place of Z_t. The bias in the estimated autocorrelations is $O(n^{-1})$ and therefore can be ignored in large samples. However, if the sample is small and if several φ_{ti} are included in the regression, the bias in $\hat{\gamma}_2(h)$ may be sizable.

To briefly investigate the nature of the bias in $\hat{\gamma}_2(h)$ we use the statistic studied by Durbin and Watson (1950, 1951). They suggested the von Neumann ratio (see Section 6.2) computed from the calculated residuals as a test of the hypothesis of independent errors. The statistic is

$$d = \frac{\sum_{t=2}^{n} (\hat{Z}_t - \hat{Z}_{t-1})^2}{\sum_{t=1}^{n} \hat{Z}_t^2} = \frac{\hat{\mathbf{z}}'\mathbf{H}\hat{\mathbf{z}}}{\hat{\mathbf{z}}'\hat{\mathbf{z}}}, \tag{9.3.4}$$

where

$$\mathbf{H} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix},$$

$\hat{\mathbf{z}} = [\mathbf{I} - \boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}']\mathbf{z}$ is the vector of calculated residuals, and $\mathbf{z}$ is the vector of observations on the original time series. Recall that

$$d = 2 - 2\hat{r}_{\hat{z}}(1) - \left(\sum_{t=1}^{n} \hat{Z}_t^2\right)^{-1} (\hat{Z}_1^2 + \hat{Z}_n^2),$$

where

$$\hat{r}_{\hat{z}}(1) = \left(\sum_{t=1}^{n} \hat{Z}_t^2\right)^{-1} \sum_{t=2}^{n} \hat{Z}_t \hat{Z}_{t-1}.$$

The expected values of the numerator and denominator of d are given by

$$\begin{aligned} E\{\hat{\mathbf{z}}'\mathbf{H}\hat{\mathbf{z}}\} &= E\{\mathbf{z}'[\mathbf{I} - \boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}']\mathbf{H}[\mathbf{I} - \boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}']\mathbf{z}\} \\ &= \operatorname{tr}(\mathbf{V}_{zz}\mathbf{H}) - \operatorname{tr}[\mathbf{V}_{zz}\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\mathbf{H}] - \operatorname{tr}[\mathbf{V}_{zz}\mathbf{H}\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'] \\ &\quad + \operatorname{tr}[\mathbf{V}_{zz}\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\mathbf{H}\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'] \end{aligned} \tag{9.3.5}$$

and

$$E\{\hat{\mathbf{z}}'\hat{\mathbf{z}}\} = \operatorname{tr}[\mathbf{V}_{zz} - \mathbf{V}_{zz}\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'], \tag{9.3.6}$$

where $\mathbf{V}_{zz} = E\{\mathbf{z}\mathbf{z}'\}$, and $\operatorname{tr}\mathbf{B}$ is the trace of the matrix $\mathbf{B}$.

It is easy to see that the expectation (9.3.5) may differ considerably from $2(n-1)[\gamma_Z(0) - \gamma_Z(1)]$. The expectation of the ratio is not necessarily equal to the ratio of the expectations, and this may further increase the bias in some cases (e.g., see Exercise 9.3).

We now turn to a consideration of d as a test of independence. For normal independent errors Durbin and Watson have shown that the ratio can be reduced to a canonical form by a simultaneous diagonalization of the numerator and denominator quadratic forms. Therefore, in this case, we can write

$$d = \frac{\sum_{j=1}^{n} \lambda_j \epsilon_j^2}{\sum_{j=1}^{n} \delta_j \epsilon_j^2} = \frac{\sum_{j=1}^{n} \lambda_j \epsilon_j^2}{\sum_{j=1}^{n-k'-1} \epsilon_j^2},$$

where it is assumed that the model contains an intercept and k' other independent variables, the λ_j are the roots of $[\mathbf{I} - \boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}']\mathbf{H}[\mathbf{I} - \boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}']$, the δ_j are the roots of $\mathbf{I} - \boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'$, and the ϵ_j are normal independent random variables with mean zero and variance $\sigma_\epsilon^2 = \sigma_Z^2$. Furthermore, the distribution of the ratio is independent of the distribution of the denominator. The exact distribution of d, depending on the roots λ and δ, can be obtained. For normal independent errors Z_t, Durbin and Watson were able to obtain upper and lower bounds for the distribution, and they tabled the percentage points of these bounding distributions.

For small sample sizes and (or) a large number of independent variables the bounding distributions may differ considerably. Durbin and Watson suggested that the distribution of $\frac{1}{4}d$ for a particular $\boldsymbol{\Phi}$-matrix could be approximated by a beta distribution with the same first two moments as $\frac{1}{4}d$.

As pointed out in Section 6.2, in the null case the t-statistic associated with the first order autocorrelation computed from the von Neumann ratio has percentage points approximately equal to those of Student's t with $n+3$ degrees of freedom. For the d-statistic computed from the regression residuals it is possible to use the moments of the d-statistic to develop a similar approximation. The expected value of d under the assumption of independent errors is

$$\begin{aligned} E\{d\} &= \{\operatorname{tr}\mathbf{H} - \operatorname{tr}[\boldsymbol{\Phi}'\mathbf{H}\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}]\}(n-k'-1)^{-1} \\ &= \{2(n-1) - \operatorname{tr}[(\Delta\boldsymbol{\Phi})'(\Delta\boldsymbol{\Phi})(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}]\}(n-k'-1)^{-1}, \end{aligned} \tag{9.3.7}$$

where $(\Delta\boldsymbol{\Phi})'(\Delta\boldsymbol{\Phi})$ is the matrix of sums of squares and products of the first differences. Consider the statistics

$$\hat{\rho} = r_d + \tfrac{1}{2}(E\{d\} - 2)\frac{n-k'+1}{n-k'}(1 - r_d^2) \tag{9.3.8}$$

and

$$t_d = (n - k' + 1)^{1/2}\hat{\rho}(1 - \hat{\rho}^2)^{-1/2}, \tag{9.3.9}$$

where $r_d = \frac{1}{2}(2 - d)$. The factor $1 - r_d^2$ keeps $\hat{\rho}$ in the interval $[-1, 1]$ for most samples. For normal independent $(0, \sigma^2)$ errors the expected value of $\hat{\rho}$ is $O(n^{-2})$ and the variance of $\hat{\rho}$ differs from that of an estimated autocorrelation computed from a sample of size $n - k'$ by terms of order n^{-2}. This suggests that the calculated t_d be compared with the tabular value of Student's t for $n - k' + 3$ degrees of freedom. By using the second moment associated with the particular Φ-matrix, the t-statistic could be further modified to improve the approximation, though there is little reason to believe that the resulting approximation would be as accurate as the beta approximation suggested by Durbin and Watson.

9.3.2. Estimated Variance Functions

The residuals can be used to investigate the nature of the variance of Z_t as well as to study the autocorrelation structure. By Proposition 9.1.1, the least squares estimator of $\boldsymbol{\beta}$ is consistent in the presence of variance heterogeneity. A simple form of variance heterogeneity is that in which the variance of the errors depends on fixed functions of time.

In Proposition 9.3.1, we show that the errors in the least squares estimated autoregressive coefficients are $O_p(n^{-1/2})$ in the presence of variance heterogeneity.

Proposition 9.3.1. Let Z_t be a time series satisfying

$$\begin{aligned} \sum_{i=0}^{p} \alpha_i Z_{t-i} &= a_t, \\ a_t &= \sigma_{at} e_t \end{aligned} \tag{9.3.10}$$

for $t = 0, \pm 1, \pm 2, \ldots$, where e_t are independent $(0, 1)$ random variables with bounded $2 + \delta$ moments, $\delta > 0$, $\{\sigma_{at}\}$ is a fixed sequence bounded above by M_a and bounded away from zero, and the roots of the characteristic equation associated with (9.3.10) are less than one in absolute value. Define the least squares estimator of $\boldsymbol{\alpha}$ by

$$\hat{\boldsymbol{\alpha}} = -\left(\sum_{t=p+1}^{n} \mathbf{X}_t' \mathbf{X}_t \right)^{-1} \sum_{t=p+1}^{n} \mathbf{X}_t' Z_t,$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$ and $\mathbf{X}_t = (Z_{t-1}, Z_{t-2}, \ldots, Z_{t-p})$. Then

$$\mathbf{G}_n^{-1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{I}),$$

where

$$\mathbf{G}_n = \left(\sum_{t=p+1}^{n} \mathbf{X}_t' \mathbf{X}_t \right)^{-1} \sum_{t=p+1}^{n} \mathbf{X}_t' \mathbf{X}_t \sigma_{at}^2 \left(\sum_{t=p+1}^{n} \mathbf{X}_t' \mathbf{X}_t \right)^{-1}. \tag{9.3.11}$$

Proof. The Z_t can be written as $\sum_{j=0}^{\infty} w_j a_{t-j}$, where the w_j are defined in (2.6.4). Now,

$$(n-h)^{-1} \sum_{t=1}^{n-h} Z_t Z_{t+h} = (n-h)^{-1} \sum_{t=1}^{n-h} \sum_{j=0}^{\infty} w_j w_{j+h} a_{t-j}^2 + (n-h)^{-1} \sum_{t=0}^{n-h} \sum_{j=0}^{\infty} \sum_{\substack{i=0 \\ i \neq j}}^{\infty} w_j w_{i+h} a_{t-j} a_{t-i}. \tag{9.3.12}$$

The a_t are independent and have bounded $2+\delta$ moments. Therefore,

$$p\lim n^{-1} \sum_{t=1}^{n} (a_t^2 - \sigma_{at}^2) = 0$$

and

$$p\lim (n-h)^{-1} \sum_{t=1}^{n} \sum_{j=0}^{t-h} w_j w_{j+h} (a_t^2 - \sigma_{at}^2) = 0$$

by Theorem 6.3.2.

Using the arguments of the proof of Theorem 6.3.5, it can be shown that the second term of (9.3.12) converges to zero in probability. Therefore,

$$p\lim \left[n^{-1} \sum_{t=1}^{n-h} Z_t Z_{t+h} - n^{-1} \sum_{t=1}^{n-h} \sum_{j=0}^{\infty} w_j w_{j+h} \sigma_{a,t-j}^2 \right] = 0.$$

The error in the least squares estimator is

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = -\left(\sum_{t=p+1}^{n} \mathbf{X}_t' \mathbf{X}_t \right)^{-1} \sum_{t=p+1}^{n} \mathbf{X}_t' a_t \tag{9.3.13}$$

and by the assumption (9.3.10),

$$E\{(\mathbf{X}_t' \mathbf{a}_t, \mathbf{X}_t' \mathbf{X}_t \mathbf{a}_t^2) \mid \mathscr{A}_{t-1}\} = (\mathbf{0}, \mathbf{X}_t' \mathbf{X}_t \sigma_{at}^2) \tag{9.3.14}$$

for $t = 0, \pm 1, \pm 2, \ldots$, where $\mathscr{A}_{t-1}$ is the sigma-field generated by $Z_{t-1}, Z_{t-2}, \ldots$. Because $a_{t-j}^2 \sigma_{at}^2$ has absolute moment greater than one, the arguments associated with obtaining the probability limit of (9.3.12) can be used to show that

$$p\lim_{n \to \infty} (n-p)^{-1} \left[\sum_{t=p+1}^{n} \mathbf{X}_t' \mathbf{X}_t \sigma_{at}^2 - \sum_{t=p+1}^{n} E\{\mathbf{X}_t' \mathbf{X}_t\} \sigma_{at}^2 \right] = \mathbf{0}.$$

It follows that

$$\left(\sum_{t=p+1}^{n} \mathbf{X}_t'\mathbf{X}_t\sigma_{at}^2\right)^{-1/2} \sum_{t=p+1}^{n} \mathbf{X}_t' a_t \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}) \tag{9.3.15}$$

by Theorem 5.3.4. ▲

The result of Proposition 9.3.1 can be extended to the estimated autoregressive parameters computed from least squares residuals.

Proposition 9.3.2. Let the model (9.1.1) hold, where Z_t is the autoregressive time series of (9.3.10) and φ_{ti}, $i = 1, 2, \ldots, r$, $t = 1, 2, \ldots$, are fixed functions of time. Let the assumptions (9.1.5) and (9.1.6) of Proposition 9.1.1 and the assumptions of Proposition 9.3.1 hold. Let $\hat{Z}_t$ be the least squares residuals of (9.3.2), and let

$$\tilde{\boldsymbol{\alpha}} = -\left(\sum_{t=p+1}^{n} \hat{\mathbf{W}}_t'\hat{\mathbf{W}}_t\right)^{-1} \sum_{t=p+1}^{n} \hat{\mathbf{W}}_t'\hat{Z}_t,$$

where $\hat{\mathbf{W}}_t = (\hat{Z}_{t-1}, \hat{Z}_{t-2}, \ldots, \hat{Z}_{t-p})$. Then

$$\mathbf{G}_n^{-1/2}(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}),$$

where $\mathbf{G}_n$ is defined in (9.3.11).

Proof. Because

$$|E\{Z_t Z_{t+h}\}| < (1 - \lambda^2)\lambda^{|h|} M_a^2$$

for some $0 < \lambda < 1$, one can use the arguments of the proof of Theorem 9.3.1 to show that

$$\sum_{t=1}^{n-h} \hat{Z}_t\hat{Z}_{t+h} = \sum_{t=1}^{n-h} Z_t Z_{t+h} + O_p(1).$$ ▲

Example 9.3.1. The 150 observations in Table 9.3.1 were computer generated. A plot of the data suggests that the mean function is approximately a linear function of time. If we fit a linear trend by ordinary least squares, we obtain

$$\hat{Y}_t = 15.472 + 0.3991t. \tag{9.3.16}$$

If a second order autoregression is fitted to the residuals by ordinary least squares, we obtain

$$\hat{Z}_t = \underset{(0.067)}{1.524}\,\hat{Z}_{t-1} - \underset{(0.067)}{0.616}\,\hat{Z}_{t-2}, \tag{9.3.17}$$

Table 9.3.1. Data for Example 9.3.1

16.45	25.13	46.49	49.57	73.07
16.39	26.21	47.78	51.03	72.83
16.25	26.98	48.05	50.80	70.05
16.23	28.60	46.54	51.49	66.30
17.16	30.80	43.15	52.91	61.47
17.55	32.65	40.34	56.95	60.47
17.48	32.92	41.30	59.32	61.30
18.32	34.41	42.49	59.76	62.38
18.31	35.65	43.00	59.29	63.46
19.09	37.15	43.75	58.74	62.38
19.53	37.49	45.86	57.80	61.61
18.08	37.64	45.95	56.27	61.82
19.21	37.03	44.87	52.98	61.03
19.96	36.78	44.22	52.27	62.18
20.69	37.26	43.95	56.17	62.60
21.15	36.83	43.52	61.33	61.83
19.69	36.38	41.51	64.69	64.23
19.36	36.79	39.41	68.49	65.62
18.83	35.87	36.07	70.19	67.17
19.84	35.89	37.23	71.47	68.61
22.20	36.21	40.14	70.98	69.36
22.42	36.95	40.41	68.63	69.91
22.18	37.93	41.06	67.47	70.28
22.38	36.55	39.56	67.68	71.92
22.45	33.94	40.33	68.47	72.48
23.47	34.28	43.90	69.52	71.59
25.03	37.51	47.28	71.47	70.19
25.41	40.42	48.71	73.33	69.69
24.91	41.70	49.46	72.17	73.55
25.49	43.39	50.61	72.72	77.55

where $s^2 = 1.646$ and the standard errors are those of the ordinary regression program.

If the absolute values of the residuals from the autoregression are plotted against time, there is an increase in the mean of the absolute values. Several models can be considered for the variance of the errors. We postulate a model in which the variance is related to the mean. Our model for the variance of a_t is

$$\sigma_{at}^2 = E\{a_t^2\} = e^{\kappa_1}[E\{Y_t\}]^{\kappa_2}, \tag{9.3.18}$$

where a_t is the error in the autoregression estimated by (9.3.17). We estimated (κ_1, κ_2), replacing a_t^2 with $\hat{a}_t^2$ from regression (9.3.17) and replacing $E\{Y_t\}$ with $\hat{Y}_t$ from (9.3.16), by maximum likelihood under the assumption that a_t are NI$(0, \sigma_{at}^2)$. The estimates are

$$(\hat{\kappa}_1, \hat{\kappa}_2) = \underset{(0.558)}{(-3.533}\ ,\ \underset{(0.148)}{1.051}),$$

where the standard errors are from the inverse of the estimated information matrix.

We have estimated the variance function for the a_t, but if the variance function is a smooth function, the variance of the autoregressive process Z_t will be nearly a multiple of the variance of the a_t. Therefore, we define new variables

$$(\hat{\sigma}_{at}^{-1} Y_t, \hat{\sigma}_{at}^{-1}, \hat{\sigma}_{at}^{-1} t) = (L_t, \varphi_{t1}, \varphi_{t2}),$$

where $\hat{\sigma}_{at}$ is the function (9.3.18) evaluated at $E\{Y_t\} = \hat{Y}_t$ and $(\hat{\kappa}_1, \hat{\kappa}_2) = (-3.533, 1.051)$. The parameters of the model

$$\begin{aligned} L_t &= \varphi_{t1}\beta_1 + \varphi_{t2}\beta_2 + u_t, \\ u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} &= e_t, \end{aligned} \tag{9.3.19}$$

estimated by Gaussian maximum likelihood, are

$$(\beta_1, \beta_2, \alpha_1, \alpha_2, \sigma_e^2) = \underset{(1.465)}{(15.113}, \underset{(0.021)}{0.407}, \underset{(0.068)}{-1.473}, \underset{(0.068)}{0.570}, \underset{(0.090)}{0.995}).$$

Estimation for a model such as (9.3.19) is discussed in Section 9.7. The estimate of $(\beta_1, \beta_2, \alpha_1, \alpha_2)$ is not greatly different from that obtained in (9.3.16) and (9.3.17) by ordinary least squares. However, the confidence interval for a prediction is different under the model in which the variance is a function of the mean. The standard errors for predictions one, two, and three periods ahead are (1.283, 2.328, 3.178) for the model computed under the assumption of homogeneous variances. The corresponding prediction standard errors are (1.673, 2.985, 4.023) for the model with the variance function of (9.3.18). ▲

In Example 9.3.1, the variance was assumed to be a fixed function of time. Models in which the variance is a random function have also been considered in the literature. Most often the variance at the current time is expressed as a function of the past behavior of the time series. Under mild assumptions, it can be shown that the least squares estimator of the parameter vector of the autoregressive process has a normal distribution in the limit.

Proposition 9.3.3. Let (Z_t, σ_{at}), $t = 0, \pm 1, \pm 2, \ldots$, be a covariance stationary time series satisfying

$$\begin{aligned} Z_t &= \sum_{j=0}^{\infty} w_j a_{t-j}, \\ a_t &= \sigma_{at} e_t, \end{aligned}$$

where the w_j are absolutely summable, e_t are iid(0, 1) random variables with $2+\delta_2$ ($\delta_2 > 0$) moments, e_{t+j} is independent of σ_{at} for $j \geq 0$, and

$$\operatorname*{plim}_{n\to\infty} n^{-1} \sum_{t=1}^{n} \sigma_{at}^2 = \sigma_a^2 ,$$

where $\sigma_a^2 = E\{\sigma_{at}^2\}$. Then

$$\operatorname*{plim}_{n\to\infty} n^{-1} \sum_{t=1}^{n-h} (Z_t - \bar{z}_n)(Z_{t+h} - \bar{z}_n) = \gamma_Z(h) ,$$

where $\bar{z}_n = n^{-1} \sum_{t=1}^{n} Z_t$ and

$$\gamma_Z(h) = \sum_{j=0}^{\infty} w_j w_{j+h} \sigma_a^2 .$$

If, in addition, Z_t is an autoregressive process satisfying

$$Z_t + \sum_{i=1}^{p} \alpha_i Z_{t-i} = a_t ,$$

where the roots of the characteristic equation are less than one in absolute value, and if

$$\operatorname*{plim}_{n\to\infty} n^{-1} \sum_{t=p+1}^{n} \mathbf{X}_t' \mathbf{X}_t \sigma_{at}^2 = E\{\mathbf{X}_t' \mathbf{X}_t \sigma_{at}^2\} , \tag{9.3.20}$$

then

$$\mathbf{G}_n^{-1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{I}) ,$$

where $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\alpha}}$, $\mathbf{G}_n$, and $\mathbf{X}_t$ are defined in Proposition 9.3.1.

Proof. The sequence $\{a_t\}$ is a martingale difference sequence and

$$n^{1/2} \bar{a}_n \xrightarrow{\mathscr{L}} N(0, \sigma_a^2)$$

by Theorem 5.3.4. Hence, $n^{1/2}\bar{z}_n$ also converges in distribution to a normal random variable by Theorem 6.3.3. Thus,

$$(n-h)^{-1} \sum_{t=1}^{n} (Z_t - \bar{z}_n)(Z_{t+h} - \bar{z}_n) = (n-h)^{-1} \sum_{t=1}^{n} Z_t Z_{t+h} + O_p(n^{-1}) .$$

We expand $n^{-1} \sum_{t=1}^{n-h} Z_t Z_{t+h}$ as in (9.3.12) and observe that

$$\operatorname*{plim}_{n\to\infty} (n-h)^{-1} \sum_{t=1}^{n-h} \sum_{j=0}^{t} w_j w_{j+h} a_{t-j}^2 = \gamma_Z(h)$$

by Theorem 6.3.2. It can be shown that

$$\operatorname*{plim}_{n\to\infty} \sum_{t=1}^{n-h} \sum_{\substack{j=0 \\ i\neq j}}^{\infty} \sum_{i=0}^{\infty} w_j w_{i+h} a_{t-i} a_{t-j} = 0$$

by applying the arguments of the proof of Theorem 6.3.5. Therefore, the sample covariance is consistent for the population covariance.

The error in the least squares estimator is given in (9.3.13). Under the present assumptions, (9.3.14) holds and, by the assumption (9.3.20) and Theorem 5.3.4, we obtain the limiting normal distribution result of (9.3.15). The limiting result for $\mathbf{G}_n^{-1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ then follows. ▲

Engle (1982) introduced a variance model called the autoregressive conditional heteroscedastic model (ARCH). An example is

$$\begin{aligned} Z_t &= \theta Z_{t-1} + a_t, \\ a_t &= (\kappa_1 + \kappa_2 a_{t-1}^2)^{1/2} e_t, \end{aligned} \tag{9.3.21}$$

where the e_t are NI(0, 1) random variables and $0 \leq \kappa_2 < 3^{-1/2}$. Some properties of this model are developed in Exercise 2.43 of Chapter 2. Clearly, many extensions of the simple model are possible, and a number have been considered in the literature. See Engle and Bollerslev (1986), Weiss (1984, 1986), Pantula (1986), Bollerslev, Chou, and Kroner (1992), Higgins and Bera (1992), and Granger and Teräsvirta (1993).

A model closely related to the ARCH model postulates the variances to be an unobserved stochastic process. See Melino and Turnbull (1990), Jacquier, Polson, and Rossi (1994), and Harvey, Ruiz, and Shephard (1994). An example model is

$$Z_t = \theta Z_{t-1} + a_t, \tag{9.3.22}$$

$$a_t = \sigma_{at} e_t, \tag{9.3.23}$$

$$h_t - \mu_h = \psi(h_{t-1} - \mu_h) + b_t, \tag{9.3.24}$$

where $h_t = \log \sigma_{at}$ and $(e_t, b_t) \sim \text{II}[(0, 0), \text{diag}(1, \sigma_b^2)]$. In (9.3.24), the logarithm of the unobserved variance process is a first order autoregressive process. Thus, σ_{at} is always positive.

If the a_t are such that $\log a_t^2$ has second moments, it is natural to take logarithms of (9.3.23) to create a linear model. Then,

$$\log a_t^2 = 2h_t + \log e_t^2. \tag{9.3.25}$$

If e_t is normally distributed, then $\log e_t^2$ is distributed as log chi square which has mean of about -1.27 and variance of about 4.93. The distribution of $\log e_t^2$ is heavily skewed to the left. If (9.3.24) is a stationary autoregressive process and e_t is normally distributed, then $\log a_t^2$ is a stationary autoregressive moving average process. Nonlinear estimation methods can be used to estimate the parameters of

(9.3.24) treating the variance of $\log e_t^2$ as known or unknown. In Example 9.3.2, we consider a modification of the log transformation in the estimation.

Example 9.3.2. As an example of the fitting of a stochastic volatility model, we consider the prices of June futures contracts for the Standard and Poor's 500 index, collected as the average price per minute from 8:30 a.m. to 3:15 p.m. on April 13, 1987. Our basic observations are the first differences of the prices. The data are considered by Breidt and Carriquiry (1994). If a first order autoregression is fitted to the data, the estimated coefficient is 0.340 with a standard error of 0.047. The residuals from the first order autoregression, denoted by $\hat{a}_t$, are plotted against time in Figure 9.3.1.

The observations in Figure 9.3.1 show evidence of a serially correlated variance. At the end of the series, large deviations tend to be followed by large deviations, while in the middle, small deviations tend to be followed by small deviations. This dependence is reflected in the sample autocorrelations for the $\hat{a}_t^2$. The first five autocorrelations of the $\hat{a}_t^2$ are 0.204, 0.501, 0.205, 0.346, and 0.263. These correlations are significantly different from zero because the standard error under zero correlation is about 0.05.

We wish to fit a variance model of the form (9.3.23), (9.3.24) to the residuals from the autoregressive fit. In practice, there are some disadvantages to a logarithmic transformation of the a_t^2 as defined in (9.3.25). There may be some zero values for a_t^2, and the original distribution may not be normal. Therefore, we

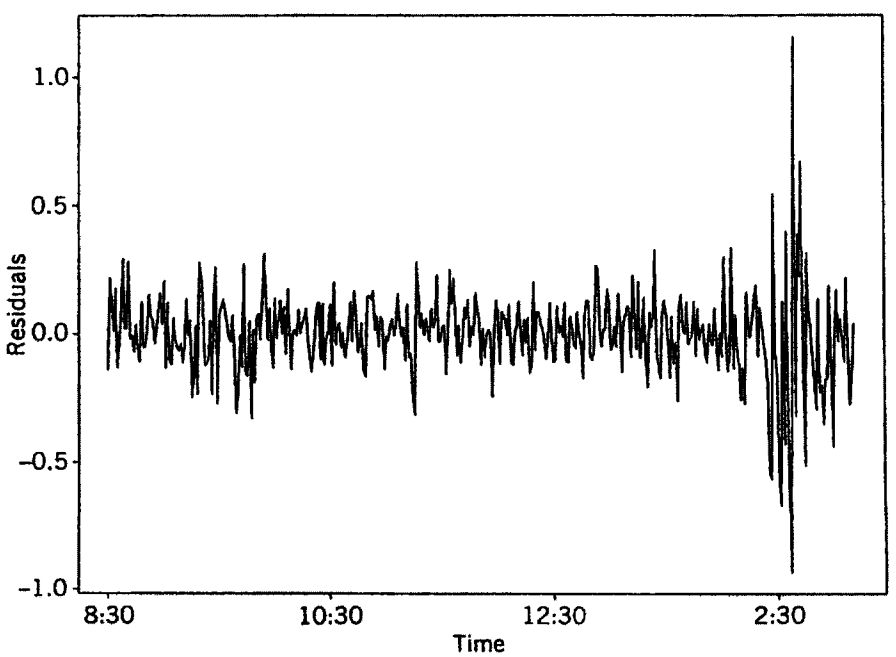

Figure 9.3.1. Residuals from autoregression of price differences.

suggest the following estimation scheme that is more robust against nonnormality. Let ξ be a small fraction, say 0.02, of the average of the a_t^2. Let

$$\begin{aligned} U_t &= \log(a_t^2 + \xi) - (a_t^2 + \xi)^{-1}\xi \\ &= 2h_t + w_t , \end{aligned} \tag{9.3.26}$$

where

$$w_t = \log(e_t^2 + \sigma_{at}^{-2}\xi) - (e_t^2 + \sigma_{at}^{-2}\xi)^{-1}\sigma_{at}^{-2}\xi . \tag{9.3.27}$$

If e_t^2 is large, w_t differs little from $\log e_t^2$. For example, the difference is less than 0.016 if $\sigma_{at}^{-2}\xi = 0.02$ and $e_t^2 > 0.10$. For normal $(0, 1)$ random variables and $\sigma_{at}^{-2}\xi = \xi = 0.02$, the mean of w_t is about -1.09 and the variance is about 3.19. Thus, the ξ-modification of the log transformation produces variables with smaller variance than $\log e_t^2$ because the function (9.3.26) is bounded below by $\log \xi - 1$. Also, the distribution of w_t is much less skewed than that of $\log e_t^2$.

The U_t of (9.3.26) can be used to estimate the parameters of the h_t-process up to an additive constant. For example, if our specification for the h_t-process is a pth order autoregressive process, we write

$$\begin{aligned} U_t &= \mu_U + X_t + g_t , \\ X_t &= \sum_{i=1}^{p} \psi_i X_{t-i} + b_t , \end{aligned} \tag{9.3.28}$$

where $\mu_U = E\{2h_t\} + E\{w_t\}$, $g_t = w_t - E\{w_t\}$, $X_t = 2h_t - E\{2h_t\}$, w_t is defined in (9.3.27), $\{b_t\}$ is independent of $\{g_t\}$, and $b_t \sim \text{II}(0, \sigma_b^2)$. Because of the nature of our transformation, the conditional distributions of the g_t given h_t are only approximately equal.

We use a first order autoregressive model for the h_t of stock prices. Then the U_t of (9.3.26) is an autoregressive moving average, which we write as

$$U_t - \mu_U = \psi(U_{t-1} - \mu_U) + \eta_t + \beta\eta_{t-1} . \tag{9.3.29}$$

We replace a_t^2 with $\hat{a}_t^2$ in the definition of U_t of (9.3.26) and set $\xi = 0.00058$, where the average of the $\hat{a}_t^2$ is 0.0290. The estimates of the parameters are

$$\begin{aligned} (\hat{\psi}, \hat{\beta}, \hat{\sigma}_\eta^2) = (&0.939, -0.816, 2.961) . \\ &(0.033)\quad (0.055) \end{aligned}$$

From (9.3.29) and (9.3.28), $\beta\sigma_\eta^2 = -\psi\sigma_g^2$ and

$$(1 + \beta^2)\sigma_\eta^2 = \sigma_b^2 + (1 + \psi^2)\sigma_g^2 .$$

It follows that

$$\begin{aligned} \hat{\sigma}_g^2 &= -\hat{\psi}^{-1}\hat{\beta}\hat{\sigma}_\eta^2 = 2.5742 , \\ \hat{\sigma}_b^2 &= \hat{\sigma}_\eta^2[1 + \hat{\beta}^2 + \hat{\psi}^{-1}\hat{\beta}(1 + \hat{\psi}^2)] = 0.0903 . \end{aligned}$$

The fact that $\hat{\sigma}_g^2$ differs considerably from 3.19 suggests that the original e_t are not normally distributed. Given an estimate of the parameter ψ of the X_t-process, we can construct smoothed estimates of X_t, $t = 1, 2, \ldots, n$ and, hence, estimates of $2h_t$, up to an additive constant. Then the constant can be estimated as a multiplicative constant in the original (nonlogarithmic) scale. Let ζ be defined by

$$\sigma_{at}^2 = \zeta \exp\{2(h_t - \mu_h)\}, \tag{9.3.30}$$

where $\mu_h = E\{h_t\}$. Then an estimator of ζ is

$$\hat{\zeta} = n^{-1} \sum_{t=1}^{n} [\exp\{\hat{X}_t\}]^{-1} \hat{a}_t^2, \tag{9.3.31}$$

where $\hat{X}_t$ is the smoothed estimator of X_t constructed from the fitted model (9.3.28). The quantities

$$\hat{\sigma}_{at}^2 = \hat{\zeta} \exp\{\hat{X}_t\} \tag{9.3.32}$$

are smoothed estimates of the unknown σ_{at}^2.

If the $\hat{\sigma}_{at}^2$ are quite variable over the sample, a second round of calculations can be carried out in which the initial transformation is

$$U_t = \log(\hat{a}_t^2 + \hat{\sigma}_{at}^2 k) - (\hat{a}_t^2 + \hat{\sigma}_{at}^2 k)^{-1} \hat{\sigma}_{at}^2 k,$$

where k is a constant, such as 0.02, and $\hat{\sigma}_{at}^2 = \hat{\zeta} \exp\{\hat{X}_t\}$. In this example, the estimates are changed very little by the second round of computations. However, in some situations there is a considerable change, and in general a second round of computation is recommended. See Breidt and Carriquiry (1994).

Given the estimates of σ_g^2, σ_b^2, and ψ for the price data, we constructed smoothed estimates $\hat{X}_t$ using the Kalman filter recursions for fixed interval smoothing described, for example, in Harvey (1989, p. 154) and Anderson and Moore (1979, Chapter 7). The estimate of ζ of (9.3.30) is $\hat{\zeta} = 0.01550$. The smoothed estimates of the standard deviations σ_{at} from (9.3.32) are plotted against time in Figure 9.3.2. They reflect the pattern observed in the data. That is, there is a period of high volatility near the end of the series and low volatility in the middle of the observation period. ▲▲

9.4. MOVING AVERAGES—LINEAR FILTERING

9.4.1. Moving Averages for the Mean

In the previous sections we considered methods of estimating the mean function for the entire period of observation. One may be interested in a simple approximation to the mean as a part of a preliminary investigation, where one is not willing to specify the mean function for the entire period, or one may desire a relatively simple method of removing the mean function that will permit simple

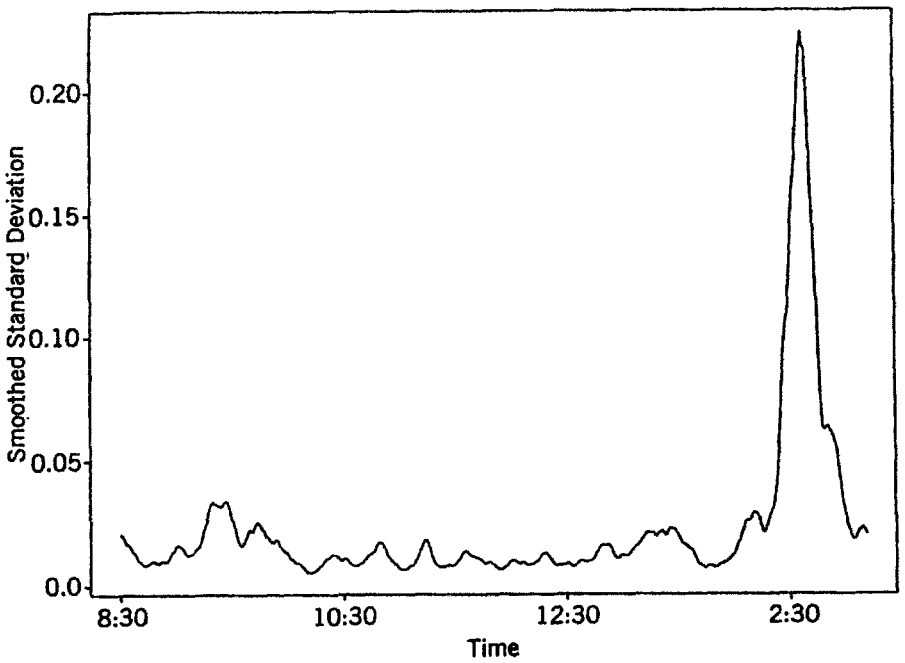

Figure 9.3.2. Smoothed estimates of σ_{st} for price differences.

extrapolation of the time series. In such cases the method of moving averages may be appropriate.

One basis for the method of moving averages is the presumption that for a period of M observations, the mean is adequately approximated by a specified function. The function is typically linear in the parameters and most commonly is a polynomial in t. Thus, the specification is

$$Y_{t+j} = g(j; \beta_t) + Z_{t+j}, \qquad j = -M_1, -M_1 + 1, \ldots, M_2 - 1, M_2, \quad (9.4.1)$$

where Z_t is a stationary time series with zero expectation, β_t is a vector of parameters, and $M = M_1 + M_2 + 1$. The form of the approximating function g is assumed to hold for all t, but the parameters are permitted to be a function of t. Both the "local" and "approximate" nature of the specification should now be clear. If the functional form for the expectation of Y_t held exactly in the interval, then it would hold for the entire realization, and the constants M_1 and M_2 would become $t - 1$ and $n - t$, respectively.

Given specification (9.4.1), a set of weights, $w_j(s)$, are constructed that, when applied to the Y_{t+j}, $j = -M_1, -M_1 + 1, \ldots, M_2 - 1, M_2$, furnish an estimator of $g(s; \beta_t)$ for the specified s. It follows that an estimator of Z_{t+s} is given by

$$Y_{t+s} - \hat{g}(s; \beta_t),$$

where

$$\hat{g}(s; \boldsymbol{\beta}_t) = \sum_{j=-M_1}^{M_2} Y_{t+j} w_j(s) .$$

In the terminology of Section 4.3 the set of weights $\{w_j(s)\}$ is a linear filter.

Let us consider an example. Assume that for a period of five observations the time series is adequately represented by

$$Y_{t+j} = \beta_{0t} + \beta_{1t} j + \beta_{2t} j^2 + Z_{t+j} , \qquad j = -2, -1, 0, 1, 2 , \tag{9.4.2}$$

where Z_t is a stationary time series with zero expectation. Using this specification, we calculate the least squares estimator for the trend value of the center observation, $s = 0$, as a linear function of the five observations. The model may be written in matrix form as

$$\mathbf{y}_t = \boldsymbol{\Phi} \boldsymbol{\beta}_t + \mathbf{z}_t ,$$

where

$$\boldsymbol{\beta}_t' = (\beta_{0t}, \beta_{1t}, \beta_{2t}) ,$$

$\mathbf{y}_t = (Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2})'$ is the vector in the first column of Table 9.4.1, $\boldsymbol{\Phi}$ is the matrix defined by the second, third, and fourth columns of Table 9.4.1 (those columns headed by β_{0t}, β_{1t}, and β_{2t}), and $\mathbf{z}_t = (Z_{t-2}, Z_{t-1}, Z_t, Z_{t+1}, Z_{t+2})'$ is the vector of (unobservable) elements of the stationary time series. The least squares estimator of the mean at $j = 0$, $g(0; \boldsymbol{\beta}_t)$, is given by

$$\begin{aligned} \hat{\beta}_{0t} &= (1, 0, 0)(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\mathbf{y}_t \\ &= \sum_{j=-2}^{2} w_j(0) Y_{t+j} = \mathbf{w}'\mathbf{y}_t , \end{aligned}$$

where the vector $(1, 0, 0)$ is the value of the vector of independent variables associated with the third observation. That is, $(1, 0, 0)$ is the third row of the matrix $\boldsymbol{\Phi}$ and is associated with $j = 0$. The vector of weights

Table 9.4.1. Calculation of Weights for a Five Period Quadratic Moving Average

	Weights for Trend						Weights for Trend Adjusted Series
$\mathbf{y}_t$	β_{0t}	β_{1t}	β_{2t}	$g(0; \boldsymbol{\beta}_t)$	$g(1; \boldsymbol{\beta}_t)$	$g(2; \boldsymbol{\beta}_t)$	$(s = 0)$
Y_{t-2}	1	−2	4	−6/70	−10/70	6/70	6/70
Y_{t-1}	1	−1	1	24/70	12/70	−10/70	−24/70
Y_t	1	0	0	34/70	24/70	−6/70	36/70
Y_{t+1}	1	1	1	24/70	26/70	18/70	−24/70
Y_{t+2}	1	2	4	−6/70	18/70	62/70	6/70

$$\mathbf{w}' = (1, 0, 0)(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}',$$

once computed, can be applied to any vector $\mathbf{y}_t$ of five observations. These weights are given in the fifth column of Table 9.4.1 under the heading "$g(0; \beta_t)$." The least squares estimator of Z_t for the third observation ($s = 0$) is given by

$$Y_t - \hat{\beta}_{0t} = Y_t - (1, 0, 0)(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{y}_t.$$

The vector of weights giving the least squares estimator of Z_t is presented in the last column of Table 9.4.1. One may readily verify that the inner product of this vector of weights with each of the columns of the original matrix $\mathbf{\Phi}$ is zero. Thus, if the original time series satisfies the specification (9.4.2), the time series of estimated residuals created by applying the weights in the last column will be a stationary time series with zero mean. The created time series X_t is given by

$$\begin{aligned} X_t &= \sum_{j=-2}^{2} r_j Y_{t+j} = \sum_{j=-2}^{2} r_j(\beta_{0t} + \beta_{1t}j + \beta_{2t}j^2 + Z_{t+j}) \\ &= \tfrac{1}{70}(6Z_{t-2} - 24Z_{t-1} + 36Z_t - 24Z_{t+1} + 6Z_{t+2}), \end{aligned}$$

where r_j denotes the weights in the last column of Table 9.4.1. Although X_t can be thought of as an estimator of Z_t, X_t is a linear combination of the five original Z_t included in the moving average.

A regression program that computes the estimated value and deviation from fit for each observation on the dependent variable is a convenient method of computation. Form a regression problem with the independent variables given by the trend specification (e.g., constant, linear, and quadratic) and the dependent variable defined by

$$y_{t+j} = \begin{cases} 1, & j = s, \\ 0 & \text{otherwise}, \end{cases}$$

where the estimated mean is to be computed for $j = s$. The vector of weights for the estimated mean is then given by

$$\mathbf{w} = \hat{\mathbf{v}} = \mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{v} = \mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\boldsymbol{\varphi}_{s\cdot}', \tag{9.4.3}$$

where $\boldsymbol{\varphi}_{s\cdot}$ is the vector of observations on the independent variables associated with $j = s$. Similarly, the vector of weights for computing the trend adjusted time series at $j = s$ is given by

$$\mathbf{v} - \hat{\mathbf{v}} = \mathbf{v} - \mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\mathbf{v} = \mathbf{v} - \mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\boldsymbol{\varphi}_{s\cdot}'. \tag{9.4.4}$$

The computation of (9.4.3) is recognized as the computation of $\hat{\mathbf{y}}$ associated with the regression of the vector $\mathbf{v}$ on $\mathbf{\Phi}$. The weights in (9.4.4) are then the deviations from fit of the same regression.

The moving average estimator of the trend and of the trend adjusted time series are frequently calculated for $j = 0$ and $M_1 = M_2$. This configuration is called a *centered moving average*. The popularity of the centered moving average is perhaps due to the following theorem.

Theorem 9.4.1. The least squares weights constructed for the estimator of trend for a centered moving average of $2M + 1$ observations, M a positive integer, under the assumption of a pth degree polynomial, p nonnegative and even, are the same as those computed for the centered moving average of $2M + 1$ observations under the assumption of a $(p + 1)$st degree polynomial.

Proof. We construct our matrix Φ in a manner analogous to that of Table 9.4.1, that is, with columns given by j^k, $k = 0, 1, 2, \ldots, p$, $j = 0, \pm 1, \pm 2, \ldots, \pm M$.

Since we are predicting trend for $j = 0$, the regression coefficients of all odd powers are multiplied by zero. That is, the elements of φ_s' in (9.4.3) associated with the odd powers are all zero. It remains only to show that the regression coefficients of the even powers remain unchanged by the presence or absence of the $(p + 1)$st polynomial in the regression. But

$$\sum_{j=-M}^{M} j^{2k+1} = 0$$

for k a positive integer. Therefore, the presence or absence of odd powers of j in the matrix Φ leaves the coefficients of the even powers unchanged, and the result follows. ▲

The disadvantage of a centered moving average is the loss of observations at the beginning and at the end of the realization. The loss of observations at the end of the observed time series is particularly critical if the objective of the study is to forecast future observations. If the trend and trend adjusted time series are computed for the end observations using the same model, the variance–covariance structure of these estimates differs from those computed for the center of the time series. We illustrate with our example.

Assume that the moving average associated with equation (9.4.2) and Table 9.4.1 is applied to a sequence of independent identically distributed random variables. That is, the trend value for the first observation is given by applying weights $g(-2, \beta_t)$ to the first five observations, and the trend adjusted value for the first observation is obtained by subtracting the trend value from Y_1. The weights $g(-2, \beta_t)$ are the weights $g(2, \beta_t)$ arranged in reverse order. For the second observation the weights $g(-1, \beta_t)$ are used. For $t = 3, 4, \ldots, n - 2$, the weights $g(0, \beta_t)$ are used. Denote the trend adjusted time series by X_t, $t = 1, 2, \ldots, n$. If $3 \leq t \leq n - 2$ and $3 \leq t + h \leq n - 2$, then

$$\text{Cov}\{X_t, X_{t+h}\} = \begin{cases} 2520/4900, & h = 0, \\ -2016/4900, & h = \pm 1, \\ 1008/4900, & h = \pm 2, \\ -288/4900, & h = \pm 3, \\ 36/4900, & h = \pm 4, \\ 0 & \text{otherwise}. \end{cases}$$

For the first observation,

$$\text{Cov}\{X_1, X_t\} = \begin{cases} 560/4900, & t = 1, \\ -1260/4900, & t = 2, \\ 420/4900, & t = 3, \\ 252/4900, & t = 4, \\ -420/4900, & t = 5, \\ 204/4900, & t = 6, \\ -36/4900, & t = 7, \\ 0 & \text{otherwise}. \end{cases}$$

9.4.2. Moving Averages of Integrated Time Series

In Chapter 10 we shall study nonstationary time series that can be represented as an autoregressive process with root of unit absolute value. We have been constructing moving average weights to remove a mean that is polynomial in time. We shall see that weights constructed to remove such a mean will also eliminate the nonstationarity arising from an autoregressive component with unit root. The time series $\{W_t, t \in (0, 1, 2, \ldots)\}$ is called an *integrated time series of order s* if it is defined by

$$W_t = \sum_{j_s=0}^{t} \cdots \sum_{j_2=0}^{j_3} \sum_{j_1=0}^{j_2} Z_{j_1},$$

where $\{Z_t, t \in (0, 1, 2, \ldots)\}$ is a stationary time series with zero mean and positive spectral density at $\omega = 0$.

Theorem 9.4.2. A moving average constructed to remove the mean will reduce a first order integrated time series to stationarity, and a moving average constructed to remove a linear trend will reduce a second order integrated time series to stationarity.

Proof. We first prove that a moving average constructed to remove the mean (zero degree polynomial) will reduce a first order integrated time series $W_t = \Sigma_{s=0}^{t} Z_s$ to stationarity. Define the time series created by applying a moving average to remove the mean by

$$X_t = \sum_{j=1}^{M} c_j W_{t+j}, \qquad t = 0, 1, 2, \ldots,$$

where the weights satisfy

$$\sum_{j=1}^{M} c_j = 0.$$

Then

$$\begin{aligned} X_t &= \sum_{j=1}^{M} c_j \sum_{s=0}^{t+j} Z_s \\ &= \sum_{j=1}^{M} c_j \left(W_t + \sum_{r=1}^{j} Z_{t+r} \right) \\ &= \sum_{j=1}^{M} c_j \sum_{r=1}^{j} Z_{t+r}, \end{aligned}$$

which is a finite moving average of a stationary time series and therefore is stationary.

Consider next the second order integrated time series

$$U_t = \sum_{r=0}^{t} \sum_{s=0}^{r} Z_s = \sum_{r=0}^{t} W_r = \sum_{j=1}^{t+1} jZ_{t-j+1},$$

where W_r is a first order integrated time series. The weights d_j constructed to remove a linear trend satisfy $\Sigma_{j=1}^{M} d_j = 0$ and $\Sigma_{j=1}^{M} jd_j = 0$. Therefore, the time series created by applying such weights is given by

$$\begin{aligned} X_t &= \sum_{j=1}^{M} d_j U_{t+j} = \sum_{j=1}^{M} d_j \left(U_t + \sum_{r=1}^{j} W_{t+r} \right) \\ &= \sum_{j=1}^{M} d_j \sum_{r=1}^{j} \left(W_t + \sum_{s=1}^{r} Z_{t+s} \right) \\ &= \sum_{j=1}^{M} d_j j W_t + \sum_{j=1}^{M} d_j \sum_{s=1}^{j} sZ_{t+j-s+1} \\ &= \sum_{j=1}^{M} \sum_{s=1}^{j} d_j s Z_{t+j-s+1}, \end{aligned}$$

which, once again, is a finite moving average of a stationary time series. ▲

The reader may extend this theorem to higher orders.

Moving averages are sometimes repeatedly applied to the same time series.

Theorem 9.4.3 can be used to show that the repeated application of a moving average constructed to remove a low order polynomial trend will remove a high order polynomial trend.

Theorem 9.4.3. Let p and q be integers, $p \geq 0$, $q \geq 1$. A moving average constructed to remove a pth degree polynomial trend will reduce a $(p+q)$th degree polynomial trend to degree $q-1$.

Proof. We write the trend function as

$$T_t = \sum_{r=0}^{p+q} b_r t^r .$$

If a moving average with weights $\{w_j: j = -M_1, -M_1+1, \ldots, M_2\}$ is applied to this function, we obtain

$$\sum_{j=-M_1}^{M_2} w_j T_{t+j} = \sum_{j=-M_1}^{M_2} w_j \sum_{r=0}^{p+q} b_r (t+j)^r$$
$$= \sum_{j=-M_1}^{M_2} w_j \sum_{r=0}^{p+q} b_r \sum_{k=0}^{r} \binom{r}{k} j^k t^{r-k} .$$

Interchanging the order of summation and using the fact that a filter constructed to remove a pth degree polynomial trend satisfies

$$\sum_{j=-M_1}^{M_2} w_j j^r = 0 , \qquad r = 0, 1, 2, \ldots, p ,$$

we obtain the conclusion. ▲

9.4.3. Seasonal Adjustment

Moving averages have been heavily used in the analysis of time series displaying seasonal variation. Assume that Y_t is a monthly time series that can be represented locally by

$$Y_{t+j} = \alpha_t + \beta_t j + \sum_{m=1}^{12} \delta_{tm} D_{t+j,m} + Z_{t+j} , \tag{9.4.5}$$

where α_t, β_t, δ_{tm} are parameters,

$$D_{tm} = \begin{cases} 1 & \text{if } Y_t \text{ is observed in month } m, \\ 0 & \text{otherwise} , \end{cases}$$

$$\sum_{m=1}^{12} \delta_{tm} = 0,$$

and Z_t is a stationary time series with zero mean. In this representation $\alpha_t + \beta_t j$ is the trend component and $\Sigma_{m=1}^{12} \delta_{tm} D_{t+j,m}$ is the seasonal component. Since the sum of the seasonal effects is zero over a period of 12 observations, it follows that a moving average such as

$$\hat{\tau}_t = \tfrac{1}{12}(\tfrac{1}{2}Y_{t-6} + Y_{t-5} + Y_{t-4} + \cdots + Y_{t+4} + Y_{t+5} + \tfrac{1}{2}Y_{t+6}) \tag{9.4.6}$$

or

$$\tfrac{1}{36}(\tfrac{1}{2}Y_{t-18} + Y_{t-17} + Y_{t-16} + \cdots + Y_{t+16} + Y_{t+17} + \tfrac{1}{2}Y_{t+18})$$

will not contain the seasonal component. Letting

$$\begin{aligned} d_0 &= \tfrac{11}{12}, \\ d_j &= -\tfrac{1}{12}, \qquad j = \pm 1, \pm 2, \ldots, \pm 5, \\ d_{-6} &= d_6 = -\tfrac{1}{24}, \end{aligned}$$

the difference

$$Y_t - \hat{\tau}_t = \sum_{m=1}^{12} \delta_{tm} D_{tm} + \sum_{j=-6}^{6} d_j Z_{t+j} \tag{9.4.7}$$

constructed for the time series (9.4.5) contains the original seasonal component but no trend. A moving average of the time series $Y_t - \hat{\tau}_t$ can then be used to estimate the seasonal component of the time series. For example,

$$\frac{1}{5} \sum_{i=-2}^{2} (Y_{t+12i} - \hat{\tau}_{t+12i}) = \hat{S}_t \tag{9.4.8}$$

furnishes an estimator of the seasonal component for time t based on 5 years of data. The 12 values of $\hat{S}_t$ computed for a year by this formula do not necessarily sum to zero. Therefore, one may modify the estimators to achieve a zero sum for the year by defining

$$\tilde{S}_{t(k)} = \hat{S}_t - \frac{1}{12} \sum_{j=-k+1}^{12-k} \hat{S}_{t+j}, \tag{9.4.9}$$

where $\tilde{S}_{t(k)}$ is the seasonal component at time t that is associated with the kth month. The seasonally adjusted time series is then given by the difference $Y_t - \tilde{S}_{t(k)}$.

The seasonal adjustment procedure described above was developed from an additive model and the adjustment was accomplished with a difference. It is quite common to specify a multiplicative model and use ratios in the construction.

It is also possible to construct directly a set of weights using regression procedures and a model such as (9.4.5). To illustrate the procedure of seasonal adjustment based directly on a regression model, we assume that a quarterly time series can be represented for a period of 21 quarters by

$$Y_{t+j} = \sum_{r=0}^{3} \beta_r j^r + \sum_{k=1}^{4} \alpha_k D_{t+j,k} + Z_{t+j}, \qquad j = 0, \pm 1, \ldots, \pm 10,$$

where

$$D_{tk} = \begin{cases} 1 & \text{if } Y_t \text{ is observed in quarter } k, \\ 0 & \text{otherwise} \end{cases}$$

and

$$\sum_{k=1}^{4} \alpha_k = 0.$$

We call $\Sigma_{k=1}^{4} \alpha_k D_{tk}$ the quarter (or seasonal) effect and $\Sigma_{r=0}^{3} \beta_r j^r$ the trend effect. The first seven columns of Table 9.4.2 give a Φ-matrix for this problem. We have incorporated the restriction $\Sigma_{k=1}^{4} \alpha_k = 0$ by setting $\alpha_4 = -\alpha_1 - \alpha_2 - \alpha_3$.

With our coding of the variables, the trend value for t, the center observation, is given by β_0, and the seasonal value for t is given by α_1. To compute the weights needed for the trend value at t, we regress the β_0-column on the remaining six columns, compute the deviations, and divide each deviation by the sum of squares of the deviations. It is readily verified that this is equivalent to computing the vector of weights by

$$\mathbf{w}'_{\beta_0} = \mathbf{J}_{\beta_0}(\Phi'\Phi)^{-1}\Phi',$$

where

$$\mathbf{J}_{\beta_0} = (1, 0, 0, 0, 0, 0, 0).$$

The weights for the seasonal component at time t can be calculated by regressing the column associated with α_1 on the remaining columns, computing the deviations, and dividing each deviation by the sum of squares. These operations are equivalent to computing

$$\mathbf{w}'_{\alpha_1} = \mathbf{J}_{\alpha_1}(\Phi'\Phi)^{-1}\Phi',$$

where

$$\mathbf{J}_{\alpha_1} = (0, 0, 0, 0, 1, 0, 0).$$

Let $\mathbf{v}$ be a vector with 1 in the 11th (tth) position and zeros elsewhere. Then the weights for the trend and seasonally adjusted time series are given by

Table 9.4.2. Calculation of Weights for the Trend and Seasonal Components of a Quarterly Time Series

Index	Φ							Weights		
j	β_0	β_1	β_2	β_3	α_1	α_2	α_3	$\mathbf{w}_{\beta_0}$	$\mathbf{w}_{\alpha_1}$	$\mathbf{w}_A$
−10	1	−10	100	−1000	0	0	1	−0.0477	−0.0314	0.0791
−9	1	−9	81	−729	−1	−1	−1	−0.0304	−0.0407	0.0711
−8	1	−8	64	−512	1	0	0	−0.0036	0.1562	−0.1526
−7	1	−7	49	−343	0	1	0	0.0232	−0.0469	0.0237
−6	1	−6	36	−216	0	0	1	0.0595	−0.0437	−0.0158
−5	1	−5	25	−125	−1	−1	−1	0.0634	−0.0515	−0.0119
−4	1	−4	16	−64	1	0	0	0.0768	0.1469	−0.2237
−3	1	−3	9	−27	0	1	0	0.0902	−0.0546	−0.0356
−2	1	−2	4	−8	0	0	1	0.1131	−0.0499	−0.0632
−1	1	−1	1	−1	−1	−1	−1	0.1036	−0.0562	−0.0474
0	1	0	0	0	1	0	0	0.1036	0.1438	0.7526
1	1	1	1	1	0	1	0	0.1036	−0.0562	−0.0474
2	1	2	4	8	0	0	1	0.1131	−0.0499	−0.0632
3	1	3	9	27	−1	−1	−1	0.0902	−0.0546	−0.0356
4	1	4	16	64	1	0	0	0.0768	0.1469	−0.2237
5	1	5	25	125	0	1	0	0.0634	−0.0515	−0.0119
6	1	6	36	216	0	0	1	0.0595	−0.0437	−0.0158
7	1	7	49	343	−1	−1	−1	0.0232	−0.0469	0.0237
8	1	8	64	512	1	0	0	−0.0036	0.1562	−0.1526
9	1	9	81	729	0	1	0	−0.0304	−0.0407	0.0711
10	1	10	100	1000	0	0	1	−0.0477	−0.0314	0.0791

$$\mathbf{w}_A = \mathbf{v} - \mathbf{w}_{\beta_0} - \mathbf{w}_{\alpha_1}.$$

The weights $\mathbf{w}_A$ can also be obtained by regressing $\mathbf{v}$ on Φ and computing the deviations from regression.

9.4.4. Differences

Differences have been used heavily when the objective is to reduce a time series to stationarity and there is little interest in estimating the mean function of the time series. Differences of the appropriate order will remove nonstationarity associated with locally polynomial trends in the mean and will reduce to stationarity integrated time series. For example, if Y_t is defined by

$$Y_t = \alpha + \beta t + W_t, \tag{9.4.10}$$

where

$$W_t = \sum_{j=0}^{t} Z_{t-j}$$

and Z_t is a stationary time series with zero mean, then

$$\Delta Y_t = Y_t - Y_{t-1} = \beta + Z_t$$

is a stationary time series.

The first difference operator can be viewed as a multiple of the moving average constructed to remove a zero degree polynomial trend from the second of two observations. The second difference is a multiple of the moving average constructed to remove a linear trend from the third of three observations. The second difference can also be viewed as a multiple of the moving average constructed to remove a linear trend from the second of three observations, and so forth. Therefore, the results of the previous subsections and of Section 2.4 may be combined in the following lemma.

Lemma 9.4.1. Let Y_t be a time series that is the sum of a polynomial trend of degree r and an integrated time series of order p. Then the qth difference of Y_t, where $r \leq q$ and $p \leq q$, is a stationary time series. The mean of $\Delta^q Y_t$ is zero for $r \leq q-1$.

Proof. See Corollary 2.4.1 and Theorems 9.4.2 and 9.4.3. ▲

Differences of lag other than one are important in transforming time series. For the function $f(t)$ defined on the integers, we define the *difference of lag H* by

$$\Delta_{(H)} f(t) = f(t) - f(t-H), \tag{9.4.11}$$

where H is a positive integer.

In Section 1.6 we defined a periodic function of period H with domain T to be a function satisfying

$$f(t+H) = f(t) \qquad \forall t, \quad t+H \in T.$$

For T the set of integers and H a positive integer, the difference of lag H of a periodic function of period H is identically zero. Therefore, differences of lag H have been used to remove seasonal and other periodic components from time series. For example, if the expected value of a monthly time series is written as

$$E\{Y_t\} = \mu + \beta t + \sum_{i=1}^{12} \alpha_i D_{it}, \tag{9.4.12}$$

where

$$D_{it} = \begin{cases} 1 & \text{if } Y_t \text{ is observed in month } i, \\ 0 & \text{otherwise}, \end{cases}$$

$$\sum_{i=1}^{12} \alpha_i = 0,$$

then the expected value of $\Delta_{(12)}Y_t$ is $E\{Y_t - Y_{t-12}\} = 12\beta$. The difference of lag 12 removed the periodic component and reduced the linear trend to a constant as well. Also, a difference of any finite lag will reduce a first order integrated time series to stationarity.

Mixtures of differences of different lags can be used. For example, if we take the first difference of the difference of lag 12 (or the difference of lag 12 of the first difference) of the time series Y_t of (9.4.12), the expected value of the resultant time series is zero; that is, $E\{\Delta_{(12)}\,\Delta Y_t\} = 0$.

The effect of repeated application of differences of different lags is summarized in Lemma 9.4.2.

Lemma 9.4.2. Let the time series Y_t be the sum of (1) a polynomial trend of degree r, (2) an integrated time series of order p, and (3) a sum of periodic functions of order $H_1, H_2, \ldots, H_q$, where $r \leq q$ and $p \leq q$. Then the difference $X_t = \Delta_{(H_1)}\Delta_{(H_2)}\cdots\Delta_{(H_q)}Y_t$, where $1 \leq H_i < \infty$ for all i, is a stationary time series. If $r \leq q - 1$, the mean of X_t is zero.

Proof. Reserved for the reader. ▲

9.5. STRUCTURAL MODELS

In Sections 9.1 and 9.2, we demonstrated the use of specified functions of time to approximate a mean function. In Section 9.4, we studied moving averages as local approximations to the mean function. Stochastic functions are also used as models for the means of time series. Models with stochastic means are sometimes called structural models in econometrics. See, for example, Harvey (1989). To introduce the approach, consider the simple model

$$\begin{aligned} Y_t &= \mu_t + e_t, \qquad t = 1, 2, \ldots, \\ \mu_t &= \mu_{t-1} + a_t, \end{aligned} \tag{9.5.1}$$

where $(e_t, a_t)' \sim \mathrm{NI}(\mathbf{0}, \mathrm{diag}\{\sigma_e^2, \sigma_a^2\})$. Models such as (9.5.1) are also called *unobserved components* models and can be considered a special case of (9.0.1). It follows from (9.5.1) that

$$X_t = \Delta Y_t = a_t + e_t - e_{t-1} \tag{9.5.2}$$

and hence X_t is a normally distributed stationary process with mean zero and autocovariances $\gamma_X(0) = \sigma_a^2 + 2\sigma_e^2$, $\gamma_X(1) = -\sigma_e^2$, and $\gamma_X(h) = 0$ for $|h| > 1$. That is, X_t is a first order moving average time series with representation

$$X_t = u_t + \beta u_{t-1}, \tag{9.5.3}$$

where $\beta(1 + \beta^2)^{-1} = -(\sigma_a^2 + 2\sigma_e^2)^{-1}\sigma_e^2$, $\beta \in [-1, 0)$, and u_t are uncorrelated

with mean zero and variance $\sigma_u^2 = (1+\beta^2)^{-1}(2\sigma_e^2 + \sigma_a^2)$. If we let $\kappa = \sigma_e^{-2}\sigma_a^2$, then $\kappa = 2 - \beta^{-1}(1+\beta)^2$ and $\sigma_e^2 = (\kappa+2)^{-1}(1+\beta^2)\sigma_u^2$. When $\sigma_a^2 = 0$, the model (9.5.1) reduces to the constant mean model and $X_t = e_t - e_{t-1}$ is a noninvertible moving average.

Given a sample segment from a realization of Y_t, we can estimate the unknown parameters (σ_e^2, σ_a^2) by fitting a first order moving average to $X_t = \Delta Y_t$, with the parameter β restricted to $[-1, 0)$. The resulting estimates of β and σ_u^2 are used to construct estimates of (σ_e^2, σ_a^2). Given an estimate of (σ_e^2, σ_a^2), estimates of the μ_t can be obtained with filtering methods. Because of the form (9.5.1), it is natural to use the Kalman filter procedures of Section 4.6 to estimate the μ_t.

It is also clear that a prediction constructed with estimates of σ_a^2 and σ_e^2 is identical to the prediction obtained by using the moving average representation for X_t.

A more general trend model is obtained by including a random change component. The expanded model is

$$\begin{aligned} Y_t &= \mu_t + e_{t1}\,, \\ \mu_t &= \zeta_{t-1} + \mu_{t-1} + e_{t2}\,, \\ \zeta_t &= \zeta_{t-1} + e_{t3}\,, \end{aligned} \tag{9.5.4}$$

where $(e_{t1}, e_{t2}, e_{t3}) \sim \mathrm{NI}(0, \mathrm{diag}\{\sigma_1^2, \sigma_2^2, \sigma_3^2\})$. The ζ-component is a local linear trend.

If we take second differences of Y_t, we obtain

$$X_t = \Delta^2 Y_t = e_{t-1,3} + \Delta e_{t2} + \Delta^2 e_{t1}\,. \tag{9.5.5}$$

By the assumptions on e_t, we have

$$[\gamma_X(0), \gamma_X(1), \gamma_X(2)] = [\sigma_3^2 + 2\sigma_2^2 + 6\sigma_1^2, -\sigma_2^2 - 4\sigma_1^2, \sigma_1^2]\,,$$

and $\gamma_X(h) = 0$ for $h > 2$. Thus, $X_t = \Delta^2 Y_t$ can be represented as a second order moving average

$$X_t = u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2}\,. \tag{9.5.6}$$

If we let $\kappa_2 = \sigma_1^{-2}\sigma_2^2$ and $\kappa_3 = \sigma_1^{-2}\sigma_3^2$, then

$$\begin{aligned} \kappa_2 &= -4 - \rho_X^{-1}(2)\rho_X(1)\,, \\ \kappa_3 &= \rho_X^{-1}(2) - 2\kappa_2 - 6\,, \\ \sigma_1^2 &= (\kappa_3 + 2\kappa_2 + 6)^{-1}\sigma_u^2(1 + \beta_1^2 + \beta_2^2)\,. \end{aligned}$$

As with the simple model, the fact that variances are nonnegative restricts the possible values for (β_1, β_2).

Example 9.5.1. To illustrate the use of the structural model to estimate the

trend, we use the data on wheat yields of Example 9.2.1. We consider the simple model

$$\begin{aligned} Y_t &= \mu_t + e_t, \\ \mu_t &= \mu_{t-1} + a_t, \end{aligned} \tag{9.5.7}$$

where $(e_t, a_t)' \sim \text{II}(0, \text{diag}\{\sigma_e^2, \sigma_a^2\})$. We let $X_t = \Delta Y_t$ and fit the first order moving average of (9.5.3) to the first differences of the data to obtain

$$\underset{(0.1010)}{X_t = -\ 0.4109\ u_{t-1}} + u_t$$

with $\hat{\sigma}_u^2 = 4.4019$. Eighty-three differences were used, and the zero mean model is estimated. Using

$$\begin{aligned} (1+\beta^2)^{-1}\beta &= -(2\sigma_e^2 + \sigma_a^2)^{-1}\sigma_e^2, \\ (1+\beta^2)\sigma_u^2 &= 2\sigma_e^2 + \sigma_a^2, \end{aligned}$$

we obtain $(\hat{\sigma}_e^2, \hat{\sigma}_a^2) = (1.8086, 1.5278)$. Using $\hat{V}\{\hat{\sigma}_u^2\} = 0.4726$ and Taylor approximation methods, the estimated standard errors are 0.5267 and 0.5756 for $\hat{\sigma}_e^2$ and $\hat{\sigma}_a^2$, respectively.

Given the estimates and the model, one can use filtering procedures to estimate the individual μ_t. We use the covariance structure to construct a linear filter. If the process begins at time one with a fixed unknown initial value, denoted by μ_1, the covariance matrix of the vector $(Y_1, Y_2, \ldots, Y_n)' = \mathbf{Y}$ is

$$\mathbf{V}_{nn} = \mathbf{I}\sigma_e^2 + \mathbf{L}\mathbf{L}'\sigma_a^2,$$

where $\mathbf{L}$ is an $n \times n$ lower triangular matrix with $L_{ij} = 1$ for $j < i$ and $L_{ij} = 0$ for $j \geq i$. The best linear unbiased estimator of the unknown initial value is

$$\tilde{\mu}_1 = (\mathbf{J}'\mathbf{V}_{nn}^{-1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{V}_{nn}^{-1}\mathbf{Y}, \tag{9.5.8}$$

where $\mathbf{J}'$ is an n-dimensional row vector composed of all ones. Now, from (9.5.7),

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)' = \mu_1\mathbf{J} + \mathbf{L}\mathbf{a},$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_n)'$. Therefore, the best unbiased estimator of $\boldsymbol{\mu}$ is

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \mathbf{J}\hat{\mu}_1 + \sigma_a^2\mathbf{L}\mathbf{L}'\mathbf{V}_{nn}^{-1}(\mathbf{Y} - \mathbf{J}\hat{\mu}_1) \\ &= \{\mathbf{J}(\mathbf{J}'\mathbf{V}_{nn}^{-1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{V}_{nn}^{-1} + \sigma_a^2\mathbf{L}\mathbf{L}'\mathbf{V}_{nn}^{-1}[\mathbf{I} - \mathbf{J}(\mathbf{J}'\mathbf{V}_{nn}^{-1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{V}_{nn}^{-1}]\}\mathbf{Y}. \end{aligned} \tag{9.5.9}$$

For fixed $\sigma_a^2\sigma_e^{-2}$, the estimator is a linear function of $\mathbf{Y}$, say $\mathbf{K}\mathbf{Y}$, and the weights to be applied to the elements of Y_j to estimate any particular μ_t decline rapidly as

the distance between j and t increases. In equation (9.5.9), $\hat{\boldsymbol{\mu}}$ is an n-dimensional vector, but any subset of the original vector of observations can be used to construct a smaller vector of estimates. Using our estimates of σ_a^2 and σ_e^2, the vector of optimal weights for estimating μ_m using $(Y_{m-5}, Y_{m-4}, \ldots, Y_{m+4}, Y_{m+5})$ is

$$\mathbf{H}_m = (0.007, 0.013, 0.029, 0.071, 0.171, 0.418, 0.171, 0.071, 0.029, 0.013, 0.007).$$

Notice that the sum of the weights is one and that the weights are symmetric about the center weight. If $\mathbf{K}$ denotes the matrix multiplying $\mathbf{Y}$ in (9.5.9), the vector $\mathbf{H}_m$ is the sixth row of the eleven by eleven matrix $\mathbf{K}$.

Figure 9.5.1 contains a plot of the estimated μ_t values. Values for 1913 through 1986 were computed using $\mathbf{H}_m$. That is, the estimates are centered moving averages with weights $\mathbf{H}_m$. The end values were computed using the optimal estimator based on the first eleven and last eleven observations. The vector of estimated variances of the errors $\hat{\mu}_t - \mu_t$ for the last six observations is

$$(0.755, 0.755, 0.757, 0.764, 0.808, 1.066).$$

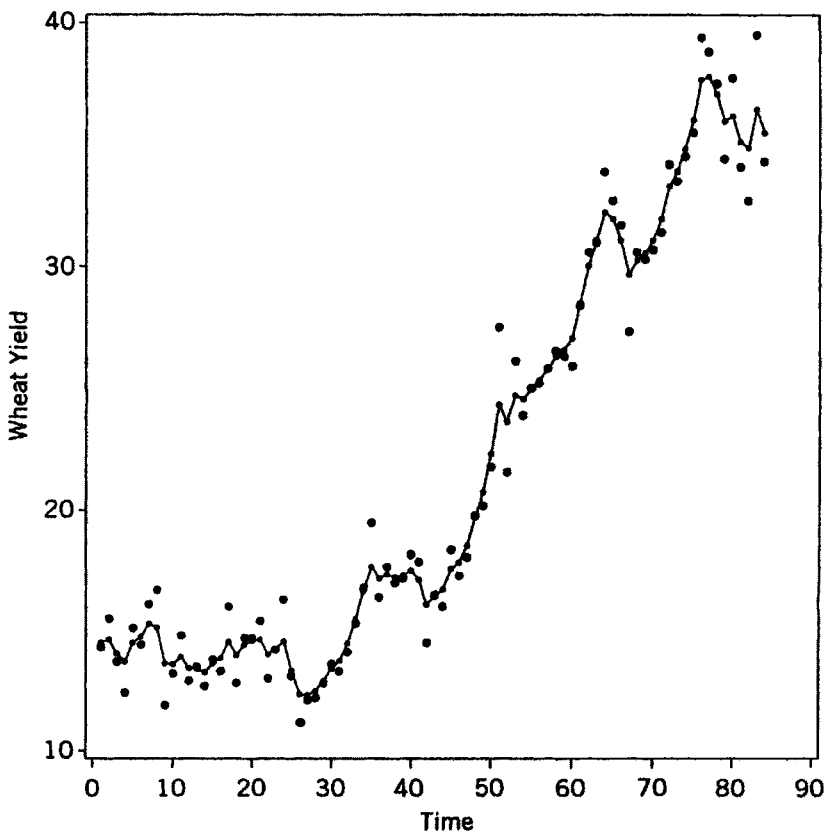

Figure 9.5.1. U.S. wheat yields 1908–1991 and structural trend estimates from model (9.5.7).

The variance for Y_{n-5} is the variance for all estimators in the center of the observation vector, that is, more than 5 years from the end points. The estimated variances are the last six diagonal elements of the matrix

$$\mathbf{LL}'\hat{\sigma}_a^2 - \mathbf{KLL}'\hat{\sigma}_a^2 - \mathbf{LL}'\mathbf{K}'\hat{\sigma}_a^2 + \mathbf{K}\hat{\mathbf{V}}_{nn}\mathbf{K}' .$$

The random walk model for μ_t provides a rather flexible model for the mean function. As a result, the $\hat{\mu}_t$ of Figure 9.5.1 form a much more variable estimated mean function than the grafted polynomial function of Example 9.2.1. The random walk model might be questioned for these data because of the rather extended period of increasing yields. Although one may not be satisified with the final estimates of μ_t obtained from the simple structural model, the estimates furnish useful information about the general movement of the time series.

Under the model (9.5.7), the predictor for μ_{n+s}, $s > 0$, is $\hat{\mu}_n$. The variance of the prediction error is $V\{\hat{\mu}_{n+s} - \mu_{n+s}\} = V\{\hat{\mu}_n - \mu_n\} + s\sigma_a^2$. Thus, the predictions of the mean for 1992, 1993, and 1994 are (35.46, 35.46, 35.46), respectively, and the estimated standard errors of the estimated means are (1.61, 2.03, 2.38). The predictions of the individual yields for 1992, 1993, and 1994 are also (35.46, 35.46, 35.46). The variances of the prediction errors are the variances of the estimated means increased by σ_e^2. The estimated standard errors of the prediction errors are (2.03, 2.38, 2.68). ▲▲

Example 9.5.1 illustrates the fact that the structural model can be used for preliminary investigation of a time series. We discussed the construction of moving averages of arbitrary length based on polynomial models in Section 9.4. By specifying a structural model for the trend, we also obtain a moving average estimator of the trend. The "length" of the moving average is based on estimates constructed with the data.

9.6. SOME EFFECTS OF MOVING AVERAGE OPERATORS

Linear moving averages have well-defined effects on the correlation and spectral properties of stationary time series. These were discussed earlier (see Theorems 2.1.1 and 4.3.1), but the effects of trend-removal filters are of sufficient importance to merit special investigation.

Proposition 9.6.1. Let X_t be a stationary time series with absolutely summable covariance function, and let $Y_t = \sum_{j=-L}^{M} a_j X_{t-j}$, where L and M are nonnegative integers and the weights a_j satisfy $\sum_{j=-L}^{M} a_j = 0$. Then the spectral density of Y_t evaluated at zero is zero; that is, $f_Y(0) = 0$.

Proof. By Theorem 4.3.1, the spectral density of Y_t is given by

$$f_Y(\omega) = (2\pi)^2 f_a(\omega) f_a^*(\omega) f_X(\omega) ,$$

where

$$2\pi f_a(\omega) = \sum_{j=-L}^{M} a_j e^{-i\omega j}.$$

Since $\Sigma_{j=-L}^{M} a_j = 0$, it follows that $f_a(0) = 0$. ▲

Since $f_Y(0) = (2\pi)^{-1} \Sigma_{h=-\infty}^{\infty} \gamma_Y(h)$, it follows that the sum of the autocovariances is zero for a stationary time series that has been filtered to remove the mean. For example, one may check that the covariances of X_t and X_{t+h}, $h = -4, -3, \ldots, 4$, for the example of Table 9.4.1 sum to zero.

The weights in the last column of Table 9.4.1 were constructed to remove a quadratic trend. The transfer function of that filter is

$$2\pi f_a(\omega) = (70)^{-1}(36 - 48\cos\omega + 12\cos 2\omega),$$

and the squared gain is

$$|2\pi f_a(\omega)|^2 = (70)^{-2}(36 - 48\cos\omega + 12\cos 2\omega)^2.$$

The squared gain is plotted in Figure 9.6.1. The squared gain is zero at zero and rises very slowly. Since the weights remove a quadratic trend, the filter removes much of the power from the spectral density at low frequencies. Since the spectral density of white noise is constant, the squared gain is a multiple of the spectral density of a moving average of uncorrelated random variables where the coefficients are given in Table 9.4.1.

The squared gain of the first difference operator is displayed in Figure 9.6.2. While it is of the same general appearance as the function of Figure 9.6.1, the squared gain of the first difference operator rises from zero more rapidly.

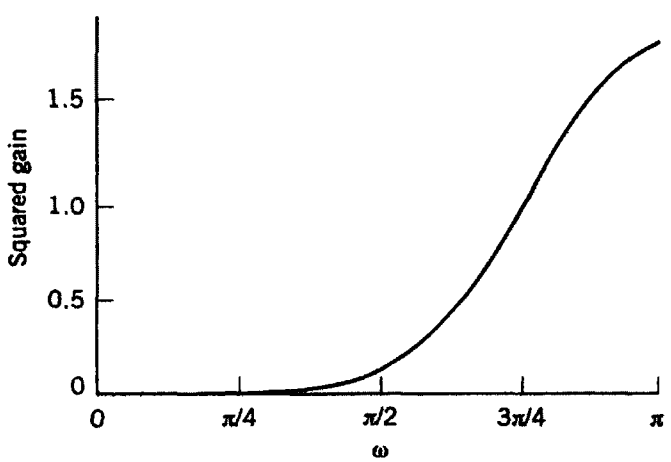

Figure 9.6.1. Squared gain of filter in Table 9.4.1.

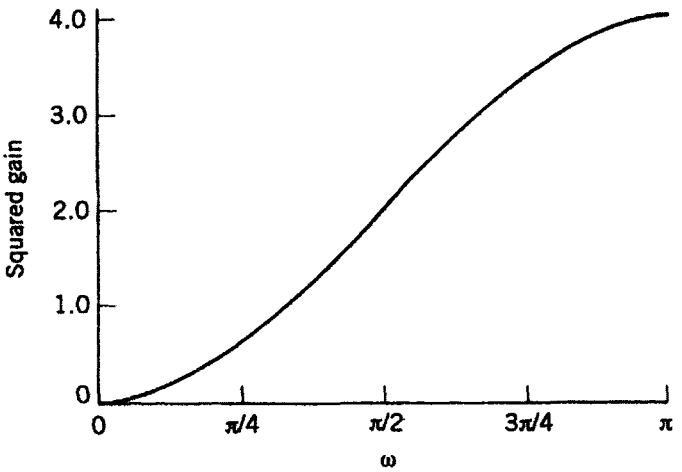

Figure 9.6.2. Squared gain of first difference operator.

If a time series contains a perfect sine component, the difference of the time series will also contain a perfect sine component of the same period, but of different amplitude and phase. That is, if $Y_t = \sin \omega t$, the difference is

$$\begin{aligned} \Delta Y_t &= \sin \omega t - \sin \omega(t-1) \\ &= 2 \sin \tfrac{1}{2}\omega \cos \omega(t - \tfrac{1}{2}) . \end{aligned}$$

We note that the amplitude of the sine wave is changed. If ω is such that $|\sin \frac{1}{2}\omega| < \frac{1}{2}$, the amplitude will be reduced. This will be true for long period waves. On the other hand, for short period waves, $\pi/3 < \omega < \pi$, the amplitude will be increased. Note that this agrees with Figure 9.6.2, which shows the transfer function to be greater than one for $\omega > \pi/3$.

Filtering a stationary time series with least squares weights to remove the seasonal effects will reduce the power of the spectral density to zero at the seasonal frequencies. For a time series with p observations per period of interest, the seasonal frequencies are defined to be $2mp^{-1}\pi$, $m = 1, 2, \ldots, L[p]$, where $L[p]$ is the largest integer less than or equal $p/2$. For example, with a monthly time series, the seasonal frequencies are $\pi/6$, $\pi/3$, $\pi/2$, $2\pi/3$, $5\pi/6$, and π. We have not included the zero frequency because most seasonal adjustment schemes are not constructed to remove the mean.

Proposition 9.6.2. Let a_j be a least squares linear filter of length R constructed to remove seasonal variation from a time series with p $(R \geq p)$ observations per period of interest. Let the time series have an absolutely summable covariance function. Then the spectral density of the filtered time series is zero at the seasonal frequencies.

Proof. A linear least squares filter constructed to remove the seasonal effects

satisfies

$$\sum_{j=1}^{R} a_j \sin \frac{2\pi m}{p} j = 0,$$

$$\sum_{j=1}^{R} a_j \cos \frac{2\pi m}{p} j = 0, \qquad m = 1, 2, \ldots, L[p].$$

This is so because any periodic function of period p defined on the integers can be represented as a sum of p sines and cosines. We have

$$2\pi f_a(\omega) = \sum_{j=1}^{R} a_j e^{-i\omega j}$$

and, setting $\omega = 2\pi m/p$,

$$2\pi f_a\left(\frac{2\pi m}{p}\right) = \sum_{j=1}^{R} a_j \left(\cos \frac{2\pi m}{p} j - i \sin \frac{2\pi m}{p} j\right) = 0$$

for $m = 1, 2, \ldots, L[p]$. ▲

The first difference is a filter satisfying the conditions of Proposition 9.6.1, and the difference of lag p satisfies the conditions of Proposition 9.6.2. We now consider the effects of difference operators on autoregressive moving average time series.

Proposition 9.6.3. Let X_t be an autoregressive moving average time series of order (p, q) expressible as

$$\sum_{j=0}^{p} \alpha_j X_{t-j} = \sum_{r=0}^{q} \beta_r e_{t-r},$$

where $\{e_t\}$ is a sequence of uncorrelated $(0, \sigma^2)$ random variables. Let the roots of $m^p + \alpha_1 m^{p-1} + \cdots + \alpha_p = 0$ be less than one in absolute value, and let the roots of $s^q + \beta_1 s^{q-1} + \cdots + \beta_q = 0$ be $s_1, s_2, \ldots, s_q$. Then the first difference $Y_t = X_t - X_{t-1}$ is an autoregressive moving average $(p, q+1)$ with the autoregressive portion unchanged and the roots of the moving average portion given by s_1, $s_2, \ldots, s_q$, 1.

Proof. The spectral density of X_t is given by

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{\left(\sum_{j=0}^{q} \beta_j e^{ij\omega}\right)\left(\sum_{j=0}^{q} \beta_j e^{-ij\omega}\right)}{\left(\sum_{j=0}^{p} \alpha_j e^{ij\omega}\right)\left(\sum_{j=0}^{p} \alpha_j e^{-ij\omega}\right)}$$

$$= \frac{\sigma^2}{2\pi} \frac{\prod_{j=1}^{q} (e^{i\omega} - s_j) \prod_{j=1}^{q} (e^{-i\omega} - s_j)}{\prod_{j=1}^{p} (e^{i\omega} - m_j) \prod_{j=1}^{p} (e^{-i\omega} - m_j)}.$$

The spectral density of Y_t is given by

$$\begin{aligned} f_Y(\omega) &= (1 - e^{i\omega})(1 - e^{-i\omega}) f_X(\omega) \\ &= (e^{i\omega} - 1)(e^{-i\omega} - 1) f_X(\omega) \\ &= \frac{\sigma^2}{2\pi} \frac{\prod_{j=1}^{q+1} (e^{i\omega} - s_j) \prod_{j=1}^{q+1} (e^{-i\omega} - s_j)}{\prod_{j=1}^{p} (e^{i\omega} - m_j) \prod_{j=1}^{p} (e^{-i\omega} - m_j)}, \end{aligned}$$

where $s_{q+1} = 1$. ▲

It follows immediately from Proposition 9.6.3 that the kth difference of the stationary autoregressive moving average time series of order (p, q) is an autoregressive moving average time series of order $(p, q + k)$, where at least k of the roots of the auxiliary equation associated with the moving average are one. Differences of lag r have similar effects on a time series.

Proposition 9.6.4. Given a time series X_t satisfying the assumptions of Proposition 9.6.3, the difference of lag r of X_t is an autoregressive moving average time series of order $(p, q + r)$ where the autoregressive portion is unchanged and the roots of the moving average portion are $s_1, s_2, \ldots, s_q$ plus the r roots of the equation $s^r = 1$.

Proof. The spectral density of $Y_t = X_t - X_{t-r}$ is given by

$$f_Y(\omega) = (1 - e^{-i\omega r})(1 - e^{i\omega r}) f_X(\omega)$$

and, for example, the factor $e^{-i\omega r} - 1$ can be written as

$$\prod_{j=q+1}^{q+r} (e^{-i\omega} - s_j),$$

where the s_j, $j = q + 1, q + 2, \ldots, q + r$, are the r roots of $s^r = 1$. ▲

These results have important practical implications. If one is attempting to identify an autoregressive moving average model for a time series that has been differenced, it is wise to consider a moving average of order at least as large as the order of differences applied to the original time series. If the time series was stationary before differencing, the characteristic polynomial of the moving average

portion of the differenced time series contains at least one unit root and the estimation theory of Section 8.3 is not applicable.

9.7. REGRESSION WITH TIME SERIES ERRORS

In this section we treat the problem of obtaining estimates of the parameters of regression equations with time series errors. We investigated the properties of the ordinary least squares estimators in Section 9.1 and found that for some special types of independent variables the ordinary least squares estimators are asymptotically fully efficient. For other independent variables, including most stationary time series, the ordinary least squares estimators are not efficient. Moreover, in most situations, the variance estimators associated with ordinary least squares are biased.

To illustrate some of these ideas, consider the simple model

$$\begin{aligned} Y_t &= \beta X_t + Z_t, \\ (Z_t, X_t) &= (\rho Z_{t-1}, \lambda X_{t-1}) + (e_t, u_t), \end{aligned} \tag{9.7.1}$$

where the e_t are normal independent $(0, \sigma_e^2)$ random variables, the u_t are normal independent $(0, \sigma_u^2)$ random variables, e_t is independent of u_j for all t, j, $|\rho| < 1$, and $|\lambda| < 1$. Under the model (9.7.1), the expected value of

$$\hat{\beta}_S = \left(\sum_{t=1}^{n} X_t^2 \right)^{-1} \sum_{t=1}^{n} X_t Y_t$$

conditional on the observed X_t, $t = 1, 2, \ldots, n$, is β, and hence $\hat{\beta}_S$ is unbiased.

The variance of the least squares estimator of β conditional on $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is given by equation (9.1.4),

$$V\{\hat{\beta}_S \mid \mathbf{X}\} = \left(\sum_{t=1}^{n} X_t^2 \right)^{-2} \sum_{t=1}^{n} \sum_{j=1}^{n} X_t X_j \rho^{|t-j|} \sigma_z^2,$$

where $\sigma_z^2 = \gamma_Z(0)$. Now the sample autocovariance of X_t converges to the population autocovariance, and

$$\begin{aligned} \operatorname*{plim}_{n\to\infty} n^{-1} \sum_{t=1}^{n} \sum_{j=1}^{n} X_t X_j \rho^{|t-j|} \sigma_z^2 &= \lim_{n\to\infty} \sum_{h=-(n-1)}^{n-1} n^{-1}(n - |h|)\lambda^{|h|}\rho^{|h|}\sigma_x^2\sigma_z^2 \\ &= (1 - \rho\lambda)^{-1}(1 + \rho\lambda)\sigma_x^2\sigma_z^2, \end{aligned}$$

where $\sigma_x^2 = \gamma_X(0)$. Hence,

$$\operatorname*{plim}_{n\to\infty} nV\{\hat{\beta}_S \mid \mathbf{X}\} = [(1 - \rho\lambda)\sigma_x^2]^{-1}(1 + \rho\lambda)\sigma_z^2.$$

If ρ is known, the variance of the generalized least squares estimator computed with known ρ, conditional on $\mathbf{X}$, is

$$V\{\hat{\beta}_G \mid \mathbf{X}\} = \left[(1-\rho^2)X_1^2 + \sum_{t=2}^{n} (X_t - \rho X_{t-1})^2 \right]^{-1} \sigma_e^2 .$$

It follows that the large sample relative efficiency of generalized least squares to ordinary least squares is

$$\operatorname*{plim}_{n\to\infty} \frac{V\{\hat{\beta}_S \mid X_1, X_2, \ldots, X_n\}}{V\{\hat{\beta}_G \mid X_1, X_2, \ldots, X_n\}} = \frac{(1+\rho\lambda)(1-2\rho\lambda+\rho^2)}{(1-\rho\lambda)(1-\rho^2)} .$$

This expression is greater than one for all nonzero (ρ, λ) less than one in absolute value. For $\rho = \lambda = (0.5)^{1/2}$, the relative efficiency is 300%!

The estimated variance of the ordinary least squares estimator obtained by the usual formulas is

$$\hat{V}_S\{\hat{\beta}_S\} = \left(\sum_{t=1}^{n} X_t^2 \right)^{-1} s^2 ,$$

where $s^2 = (n-1)^{-1} \Sigma_{t=1}^{n} (Y_t - \hat{\beta}_S X_t)^2$.

Since $V\{\hat{\beta}_S \mid \mathbf{X}\} = O_p(n^{-1})$, we have $\hat{\beta}_S - \beta = O_p(n^{-1/2})$ and s^2 converges to the variance of Z_t. Therefore,

$$\operatorname{plim} n\hat{V}_S\{\hat{\beta}_S\} = \operatorname{plim} n\left(\sum_{t=1}^{n} X_t^2 \right)^{-1} s^2 = \sigma_x^{-2}\sigma_z^2$$

and

$$\operatorname{plim}\left\{ \frac{\hat{V}_S\{\hat{\beta}_S\} - V\{\hat{\beta}_S \mid \mathbf{X}\}}{V\{\hat{\beta}_S \mid \mathbf{X}\}} \right\} = \frac{-2\rho\lambda}{1+\rho\lambda} .$$

The common least squares estimator of variance will be an over- or underestimate, depending on the signs of ρ and λ. If $\rho = \lambda = (0.5)^{1/2}$, the ordinary least squares estimator will be estimating a quantity approximately one-third of the true variance. Given these results on the efficiency of ordinary least squares and on the bias in the ordinary least squares estimator of variance, it is natural to consider an estimator such as (9.1.3), where the estimated autocorrelations are used in place of the unknown parameters.

In Section 9.3 we established that the autocorrelations of the error time series could be estimated from the calculated regression residuals. We apply those results to our current model.

Proposition 9.7.1. Let Y_t satisfy

$$Y_t = \sum_{i=1}^{r} \varphi_{ti}\beta_i + Z_t , \tag{9.7.2}$$

where Z_t is a stationary autoregressive process,

$$Z_t + \sum_{j=1}^{p} \alpha_j Z_{t-j} = e_t ,$$

the roots of the characteristic equation are less than one in absolute value, the φ_{ti} satisfy (9.1.7), (9.1.8), and (9.1.9), and $\{e_t\}$ is a sequence of independent $(0, \sigma^2)$ random variables with $E\{e_t^4\} = \eta\sigma^4$. Denote the simple least squares estimator of $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k)'$ and the simple least squares residuals by $\hat{Z}_t$.

Let $\tilde{\boldsymbol{\alpha}}$ be the estimator of $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_p)'$ obtained by regressing $\hat{Z}_t$ on $(\hat{Z}_{t-1}, \hat{Z}_{t-2}, \ldots, \hat{Z}_{t-p})$, $t = p+1, p+2, \ldots, n$, and let $\hat{\boldsymbol{\alpha}}$ be the estimator obtained by regressing Z_t on $(Z_{t-1}, Z_{t-2}, \ldots, Z_{t-p})$. Then

$$\tilde{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}} = O_p(n^{-1}) .$$

Proof. The result is an immediate consequence of Theorem 9.3.1. ▲

Given an estimator of the correlational structure of the error time series, we wish to construct the estimated generalized least squares estimator. Often the most expeditious way to obtain this estimator is to transform the data. In the present case, the Gram–Schmidt orthogonalization leads one to the transformed variables

$$\begin{aligned}
\epsilon_1 &= \delta_{11} Z_1 , \\
\epsilon_2 &= \delta_{22} Z_2 - \delta_{21} Z_1 , \\
&\vdots \\
\epsilon_p &= \delta_{pp} Z_p - \sum_{j=1}^{p-1} \delta_{p,p-j} Z_{p-j} , \\
\epsilon_t &= e_t = Z_t + \sum_{j=1}^{p} \alpha_j Z_{t-j} , \qquad t = p+1,\ p+2, \ldots, n ,
\end{aligned} \tag{9.7.3}$$

where $\delta_{11} = \gamma_Z^{-1/2}(0)\sigma$, $\delta_{22} = \{[1 - \rho_Z^2(1)]\gamma_Z(0)\}^{-1/2}\sigma$, $\delta_{21} = \rho_Z(1)\{[1 - \rho_Z^2(1)]\gamma_Z(0)\}^{-1/2}\sigma$, etc. The ϵ_t are uncorrelated with constant variance σ^2. For the first order autoregressive time series the transformed variables are

$$\begin{aligned}
\epsilon_1 &= (1 - \rho^2)^{1/2} Z_1 , \\
\epsilon_t &= Z_t - \rho Z_{t-1} , \qquad t = 2, 3, \ldots, n .
\end{aligned}$$

As a second example of the transformation, consider the second order autoregressive process

$$Z_t = 1.53 Z_{t-1} - 0.66 Z_{t-2} + e_t ,$$

where the variance of e_t is σ^2. Then by (2.5.8) the variance of Z_t is $11.77\sigma^2$. The correlations for the time series are given in Table 6.4.2. The transformation is

$$\epsilon_1 = (11.77)^{-1/2} Z_1 ,$$
$$\epsilon_2 = (1.7718)^{-1/2}(Z_2 - 0.9217 Z_1) ,$$
$$\epsilon_t = e_t = Z_t - 1.53 Z_{t-1} + 0.66 Z_{t-2} , \qquad t = 3, 4, \ldots, n .$$

To define the estimated generalized least squares estimator, we can express our original model (9.7.1) in the matrix notation of Section 9.1 as

$$\mathbf{y} = \mathbf{\Phi}\boldsymbol{\beta} + \mathbf{z} .$$

We let $\mathbf{V}_{zz} = E\{\mathbf{z}\mathbf{z}'\}$ and $\boldsymbol{\epsilon} = \mathbf{T}\mathbf{z}$, where $\mathbf{T}$ is the $n \times n$ transformation matrix defined in (9.7.3). Then the estimated generalized least squares estimator is obtained by regressing $\hat{\mathbf{T}}\mathbf{y}$ on $\hat{\mathbf{T}}\mathbf{\Phi}$,

$$\tilde{\boldsymbol{\beta}} = [\mathbf{\Phi}'\hat{\mathbf{T}}'\hat{\mathbf{T}}\mathbf{\Phi}]^{-1}\mathbf{\Phi}'\hat{\mathbf{T}}'\hat{\mathbf{T}}\mathbf{y} = [\mathbf{\Phi}'\hat{\mathbf{V}}_{zz}^{-1}\mathbf{\Phi}]^{-1}\mathbf{\Phi}'\hat{\mathbf{V}}_{zz}^{-1}\mathbf{y} , \tag{9.7.4}$$

where $\hat{\mathbf{V}}_{zz}^{-1} = \hat{\mathbf{T}}'\hat{\mathbf{T}}\hat{\sigma}^{-2}$, $\hat{\sigma}^2 = (n-2p)^{-1}\sum_{t=p+1}^{n}(\hat{Z}_t + \sum_{j=1}^{p}\tilde{\alpha}_j\hat{Z}_{t-j})^2$, and $\hat{\mathbf{T}}$ is obtained from $\mathbf{T}$ by replacing α_j with $\tilde{\alpha}_j$, σ^2 with $\hat{\sigma}^2$, and so forth.

In Section 5.7 we demonstrated that the limiting distributions of the generalized least squares estimator and of the estimated generalized least squares estimator are the same under mild assumptions. By Theorem 5.7.4 and Proposition 9.7.1, the estimated generalized least squares estimator for an error process that is an autoregressive moving average will have the same limiting distribution as the generalized least squares estimator based on the true parameters, under mild assumptions. In Theorem 9.7.1, we give a direct proof of the result for autoregressive errors using the transformation $\hat{\mathbf{T}}$.

Theorem 9.7.1. Let Y_t satisfy the model (9.7.2), where the φ_{ti} satisfy the assumptions (9.1.7), (9.1.8), and (9.1.9). Then

$$\mathbf{D}_n(\hat{\boldsymbol{\beta}}_G - \tilde{\boldsymbol{\beta}}) = O_p(n^{-1/2}) ,$$

where $\hat{\boldsymbol{\beta}}_G = [\mathbf{\Phi}'\mathbf{V}_{zz}^{-1}\mathbf{\Phi}]^{-1}\mathbf{\Phi}'\mathbf{V}_{zz}^{-1}\mathbf{y}$, $\tilde{\boldsymbol{\beta}}$ is defined in (9.7.4), $d_{nii} = (\sum_{t=1}^{n}\varphi_{ti}^2)^{1/2}$, and

$$\mathbf{D}_n = \operatorname{diag}\{d_{n11}, d_{n22}, \ldots, d_{nrr}\} .$$

Proof. We have, for $\hat{\mathbf{T}}_{t\cdot}$ the tth row of $\hat{\mathbf{T}}$ and $\boldsymbol{\varphi}_{\cdot i}$ the ith column of $\mathbf{\Phi}$,

$$\hat{\epsilon}_t = \hat{\mathbf{T}}_{t\cdot}\mathbf{z} = \epsilon_t + \sum_{j=1}^{p}(\tilde{\alpha}_j - \alpha_j)Z_{t-j} ,$$

$$\hat{w}_{ti} = \hat{\mathbf{T}}_{t\cdot}\boldsymbol{\varphi}_{\cdot i} = \varphi_{ti} + \sum_{j=1}^{p}\alpha_j\varphi_{t-j,i} + \sum_{j=1}^{p}(\tilde{\alpha}_j - \alpha_j)\varphi_{t-j,i}$$

$$\overset{\text{def}}{=} w_{ti} + \sum_{j=1}^{p}(\tilde{\alpha}_j - \alpha_j)\varphi_{t-j,i}$$

for $t = p+1,\ p+2, \ldots, n$, with similar expressions holding for $t = 1, 2, \ldots, p$. By Proposition 9.7.1 and Theorem 8.2.1, $\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = O_p(n^{-1/2})$, and it follows that

$$d_{nii}^{-1} \sum_{t=1}^{n} \hat{\epsilon}_t \hat{w}_{it} = d_{nii}^{-1} \sum_{t=1}^{n} \epsilon_t w_{it} + O_p(n^{-1/2}),$$

$$(d_{nii} d_{njj})^{-1} \sum_{t=1}^{n} \hat{w}_{it} \hat{w}_{tj} = (d_{nii} d_{njj})^{-1} \sum_{t=1}^{n} w_{it} w_{tj} + O_p(n^{-1/2}).$$

Therefore,

$$\begin{aligned} \mathbf{D}_n(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= (\mathbf{D}_n^{-1}\boldsymbol{\Phi}'\hat{\mathbf{T}}'\hat{\mathbf{T}}\boldsymbol{\Phi}\mathbf{D}_n^{-1})^{-1}\mathbf{D}_n^{-1}\boldsymbol{\Phi}'\hat{\mathbf{T}}'\hat{\mathbf{T}}\mathbf{z} \\ &= (\mathbf{D}_n^{-1}\boldsymbol{\Phi}'\mathbf{V}_{zz}^{-1}\boldsymbol{\Phi}\mathbf{D}_n^{-1})^{-1}\mathbf{D}_n^{-1}\boldsymbol{\Phi}'\mathbf{V}_{zz}^{-1}\mathbf{z} + O_p(n^{-1/2}) \\ &= \mathbf{D}_n(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta}) + O_p(n^{-1/2}). \end{aligned}$$

▲

Since the residual mean square for the regression of $\hat{\mathbf{T}}\mathbf{y}$ on $\hat{\mathbf{T}}\boldsymbol{\Phi}$ converges in probability to σ^2, one can use the ordinary regression statistics for approximate tests and confidence intervals.

Because of the infinite autoregressive nature of invertible moving average time series, an exact transformation of the form (9.7.3) is cumbersome. However, the approximate difference equation transformation will be adequate for many purposes. For example, if

$$Z_t = e_t + be_{t-1}, \qquad |b| < 1,$$

the transformation

$$\begin{aligned} \epsilon_1 &= (1+b^2)^{-1/2} Z_1, \\ \epsilon_2 &= (1+b^2)^{1/2}(1+b^2+b^4)^{-1/2}[Z_2 - (1+b^2)^{-1} b Z_1], \\ \epsilon_t &= Z_t - b\epsilon_{t-1}, \qquad t = 3, 4, \ldots, n, \end{aligned} \tag{9.7.5}$$

will generally be satisfactory when b is not too close to one in absolute value.

Example 9.7.1. To illustrate the fitting of a regression model with time series errors, we use the data studied by Prest (1949). The data were discussed by Durbin and Watson (1951), and we use the data as they appear in that article. The data, displayed in Table 9.7.1, pertain to the consumption of spirits in the United Kingdom from 1870 to 1938. The dependent variable Y_t is the annual per capita consumption of spirits in the United Kingdom. The explanatory variables φ_{t1} and φ_{t2} are per capita income and price of spirits, respectively, both deflated by a general price index. All data are in logarithms. The model suggested by Prest can be written

$$Y_t = \beta_0 + \beta_1 \varphi_{t1} + \beta_2 \varphi_{t2} + \beta_3 \varphi_{t3} + \beta_4 \varphi_{t4} + Z_t,$$

Table 9.7.1. Annual Consumption of Spirits in the United Kingdom from 1870 to 1938

Year	Consumption Y_t	Income φ_{t1}	Price φ_{t2}	Year	Consumption Y_t	Income φ_{t1}	Price φ_{t2}
1870	1.9565	1.7669	1.9176	1905	1.9139	1.9924	1.9952
1871	1.9794	1.7766	1.9059	1906	1.9091	2.0117	1.9905
1872	2.0120	1.7764	1.8798	1907	1.9139	2.0204	1.9813
1873	2.0449	1.7942	1.8727	1908	1.8886	2.0018	1.9905
1874	2.0561	1.8156	1.8984	1909	1.7945	2.0038	1.9859
1875	2.0678	1.8083	1.9137	1910	1.7644	2.0099	2.0518
1876	2.0561	1.8083	1.9176	1911	1.7817	2.0174	2.0474
1877	2.0428	1.8067	1.9176	1912	1.7784	2.0279	2.0341
1878	2.0290	1.8166	1.9420	1913	1.7945	2.0359	2.0255
1879	1.9980	1.8041	1.9547	1914	1.7888	2.0216	2.0341
1880	1.9884	1.8053	1.9379	1915	1.8751	1.9896	1.9445
1881	1.9835	1.8242	1.9462	1916	1.7853	1.9843	1.9939
1882	1.9773	1.8395	1.9504	1917	1.6075	1.9764	2.2082
1883	1.9748	1.8464	1.9504	1918	1.5185	1.9965	2.2700
1884	1.9629	1.8492	1.9723	1919	1.6513	2.0652	2.2430
1885	1.9396	1.8668	2.0000	1920	1.6247	2.0369	2.2567
1886	1.9309	1.8783	2.0097	1921	1.5391	1.9723	2.2988
1887	1.9271	1.8914	2.0146	1922	1.4922	1.9797	2.3723
1888	1.9239	1.9166	2.0146	1923	1.4606	2.0136	2.4105
1889	1.9414	1.9363	2.0097	1924	1.4551	2.0165	2.4081
1890	1.9685	1.9548	2.0097	1925	1.4425	2.0213	2.4081
1891	1.9727	1.9453	2.0097	1926	1.4023	2.0206	2.4367
1892	1.9736	1.9292	2.0048	1927	1.3991	2.0563	2.4284
1893	1.9499	1.9209	2.0097	1928	1.3798	2.0579	2.4310
1894	1.9432	1.9510	2.0296	1929	1.3782	2.0649	2.4363
1895	1.9569	1.9776	2.0399	1930	1.3366	2.0582	2.4552
1896	1.9647	1.9814	2.0399	1931	1.3026	2.0517	2.4838
1897	1.9710	1.9819	2.0296	1932	1.2592	2.0491	2.4958
1898	1.9719	1.9828	2.0146	1933	1.2635	2.0766	2.5048
1899	1.9956	2.0076	2.0245	1934	1.2549	2.0890	2.5017
1900	2.0000	2.0000	2.0000	1935	1.2527	2.1059	2.4958
1901	1.9904	1.9939	2.0048	1936	1.2763	2.1205	2.4838
1902	1.9752	1.9933	2.0048	1937	1.2906	2.1205	2.4636
1903	1.9494	1.9797	2.0000	1938	1.2721	2.1182	2.4580
1904	1.9332	1.9772	1.9952				

SOURCE: Durbin and Watson (1951). Reproduced with permission of the Trustees of Biometrika and of the authors.

where 1869 is the origin for t, $\varphi_{t3} = 10^{-2}t$, $\varphi_{t4} = 10^{-4}(t-35)^2$, and we assume Z_t is a stationary time series.

The ordinary least squares regression equation is

$$\hat{Y}_t = \underset{}{2.14} + \underset{(0.28)}{0.69}\,\varphi_{t1} - \underset{(0.14)}{0.63}\,\varphi_{t2} - \underset{(0.05)}{0.95}\,\varphi_{t3} - \underset{(0.09)}{1.15}\,\varphi_{t4} - \underset{(0.16)}{}.$$

The residual mean square is 0.00098. The numbers in parentheses are the standard errors computed by the ordinary regression formulas and would be a part of the output in any standard regression computer program. They are *not* proper estimators of the standard errors when the error in the equation is autocorrelated. We shall return to this point.

The Durbin–Watson d is 0.5265. Under the null hypothesis of normal independent errors, the expected value of d is

$$E\{d\} = [2(68) - 0.67](64)^{-1} = 2.1145,$$

where

$$tr\{(\Delta\Phi)'(\Delta\Phi)(\Phi'\Phi)^{-1}\} = 0.67.$$

Therefore, from equations (9.3.8) and (9.3.9),

$$\begin{aligned}\hat{\rho} &= 0.7367 + (0.5)(0.1145)(66)(65)^{-1}(1 - 0.5427)\\ &= 0.7633\end{aligned}$$

and

$$t_d = (66)^{1/2}(0.7633)(0.4174)^{-1/2} = 9.60.$$

Obviously, the hypothesis of independent errors is rejected. To investigate the nature of the autocorrelation, we fit several autoregressive models to the regression residuals. The results are summarized in Table 9.7.2. The statistics of that table are consistent with the hypothesis that the error time series is a first order autoregressive process. Therefore, we construct the transformation of (9.7.3) for a first order autoregressive process with $\hat{\rho} = 0.7633$. The transformation matrix is

$$\hat{\mathbf{T}} = \begin{pmatrix} 0.6460 & 0 & 0 & \cdots & 0 \\ -0.7633 & 1 & 0 & \cdots & 0 \\ 0 & -0.7633 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

where $\hat{\mathbf{T}}$ is a 69×69 matrix and $0.6460 = (1 - \hat{\rho}^2)^{1/2} = [1 - (0.7633)^2]^{1/2}$.

Regressing the transformed Y_t on the transformed φ_{ti}, $i = 0, 1, \ldots, 4$, we obtain the estimated generalized least squares equation

$$\begin{aligned}\hat{Y}_t = {} & 2.36 + 0.72\,\varphi_{t1} - 0.80\,\varphi_{t2} - 0.81\,\varphi_{t3} - 0.92\,\varphi_{t4}.\\ & (0.30)\quad (0.15)\qquad (0.07)\qquad (0.11)\qquad (0.27)\end{aligned}$$

The numbers in parentheses are consistent estimates of the standard errors of the

Table 9.7.2. Analysis of Variance for Residuals from Spirit Consumption Regression

Source	Degrees of Freedom	Mean Square
$\hat{Z}_{t-1}$	1	0.03174
$\hat{Z}_{t-2}$ after $\hat{Z}_{t-1}$	1	0.00036
$\hat{Z}_{t-3}$ after $\hat{Z}_{t-1}, \hat{Z}_{t-2}$	1	0.00013
$\hat{Z}_{t-4}$ after $\hat{Z}_{t-1}, \hat{Z}_{t-2}, \hat{Z}_{t-3}$	1	0.00004
Error	56	0.00049

estimators. They are computed from the transformed data by the usual regression formulas. The error mean square for the transformed regression is 0.000417.

Every one of the estimated standard errors for the generalized least squares estimator based on $\hat{\rho} = 0.7633$ is greater than the standard error calculated by the usual formulas for the simple least squares estimates. Of course, we know that the standard formulas give biased estimators for simple least squares when the errors are autocorrelated. To demonstrate the magnitude of this bias, we estimate the variance of the simple least squares estimators, assuming that the error time series is

$$Z_t = 0.7633 Z_{t-1} + e_t ,$$

where e_t is a sequence of uncorrelated $(0, 0.000417)$ random variables. The covariance matrix of the simple least squares estimator is estimated by $(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\hat{\mathbf{V}}_{zz}\mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}$, where $\mathbf{\Phi}$ is the 69×5 matrix of observations on the independent variables and $\hat{\mathbf{V}}_{zz}$ is the estimated covariance matrix for 69 observations from a first order autoregressive time series with $\hat{\rho} = 0.7633$.

Alternative estimates of the standard errors of the coefficients are compared in Table 9.7.3. The consistent estimates of the standard errors of simple least squares obtained from the matrix $(\mathbf{\Phi}'\mathbf{\Phi})^{-1}\mathbf{\Phi}'\hat{\mathbf{V}}_{zz}\mathbf{\Phi}(\mathbf{\Phi}'\mathbf{\Phi})^{-1}$ are approximately twice those obtained from the simple formulas. This means that the variance estimates

Table 9.7.3. Comparison of Alternative Estimators of Standard Errors of Estimators for Spirits Example

Coefficient	Estimator for Simple Least Squares Assuming $\rho = 0$	Consistent Estimator for Simple Least Squares	Consistent Estimator for Generalized Least Squares
β_0	0.282	0.545	0.304
β_1	0.138	0.262	0.146
β_2	0.054	0.112	0.072
β_3	0.088	0.179	0.108
β_4	0.161	0.313	0.266

computed under the assumption of zero correlation are approximately one-fourth of the consistent estimates. The estimated standard errors for the generalized least squares estimators range from about one-half (0.56 for income) to nine-tenths (0.85 for time squared) of those for simple least squares. This means that the estimated efficiency of simple least squares relative to generalized least squares ranges from about three-tenths to about seven-tenths.

The computations for regression models with autoregressive errors are available in several computer programs. The procedure AUTOREG of SAS Institute, Inc. (1989) is one convenient program. If we specify a first order autoregressive error and the Yule–Walker procedure of estimating the autoregressive process, we obtain an estimate $\hat{\rho} = 0.7160$ with a standard error of 0.0879. The estimate of ρ differs from ours because of the different estimator used. The estimated regression equation is nearly identical to the one given above. ▲▲

In Example 9.7.1 we gave details of a two-step computational procedure for estimation of a regression equation with time series errors. Computer programs offer one the option of specifying and estimating the complete model, either by the two-step procedure or as a single maximization problem.

Assume the time series error in a linear regression model to be an autoregressive process. Thus, we write

$$Y_t = \boldsymbol{\varphi}_{t\cdot}\boldsymbol{\beta} + Z_t, \qquad t = 1, 2, \ldots, n, \tag{9.7.6a}$$

$$Z_t + \sum_{i=1}^{p} \alpha_i Z_{t-i} = e_t, \tag{9.7.6b}$$

where $\{e_t\}$ is a sequence of NI(0, σ^2) random variables, and $\boldsymbol{\varphi}_{t\cdot}$, $t = 1, 2, \ldots, n$, are k-dimensional row vectors of explanatory variables. If the process Z_t is a stationary normal process, the log likelihood is

$$L(\mathbf{Y}; \boldsymbol{\theta}) = -0.5[n \log 2\pi + \log|\mathbf{V}_{zz}| + (\mathbf{Y} - \boldsymbol{\Phi}\boldsymbol{\beta})'\mathbf{V}_{zz}^{-1}(\mathbf{Y} - \boldsymbol{\Phi}\boldsymbol{\beta})], \tag{9.7.7}$$

where $\mathbf{V}_{zz} = \mathbf{V}_{zz}(\boldsymbol{\alpha}, \sigma^2)$ is the $n \times n$ covariance matrix of $(Z_1, Z_2, \ldots, Z_n)$, $\mathbf{Y}$ is the n-dimensional column vector $(Y_1, Y_2, \ldots, Y_n)'$, $\boldsymbol{\Phi}$ is the $n \times k$ matrix $(\boldsymbol{\varphi}_{1\cdot}', \boldsymbol{\varphi}_{2\cdot}', \ldots, \boldsymbol{\varphi}_{n\cdot}')'$, and $\boldsymbol{\theta}' = (\beta_1, \beta_2, \ldots, \beta_r, \alpha_1, \alpha_2, \ldots, \alpha_p, \sigma^2)$.

The theorems of Section 5.5 can be used to obtain the limiting distribution of the maximum likelihood estimator. The limiting distribution of the maximum likelihood estimator of $\boldsymbol{\beta}$ is the same as the limiting distribution of the generalized least squares estimator, if the latter exists.

Theorem 9.7.2. Let Y_t satisfy equation (9.7.6a), where Z_t is a stationary process with the roots of the characteristic equation associated with (9.7.6b) less than one in absolute value. Assume the e_t are iid(0, σ^2) random variables or that the e_t satisfy the martingale difference conditions of Corollary 5.3.4. Let $\{\boldsymbol{\varphi}_{t\cdot}\}$ and $\{\mathbf{M}_n\}$ be sequences such that

$$\lim_{n\to\infty} \mathbf{M}_n^{-1}\mathbf{\Phi}'\mathbf{V}_{zz}^{-1}\mathbf{\Phi}\mathbf{M}_n^{-1\prime} = \mathbf{I}$$

and

$$\lim_{n\to\infty} \sup_{1\le t\le n} |\mathbf{M}_n^{-1}\boldsymbol{\varphi}_{t\cdot}|^2 = 0 .$$

Let $\hat{\boldsymbol{\theta}}$ be the estimator that maximizes the log Gaussian likelihood (9.7.7), and let $\boldsymbol{\theta}^0$ be the true parameter vector. Then

$$\mathbf{M}_n'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}) ,$$
$$n^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{A}^{-1}\sigma^2) ,$$

$\hat{\sigma}^2 \xrightarrow{P} \sigma^2$, and $\mathbf{M}_n'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)$ is independent of $n^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0)$ in the limit, where $\mathbf{A} = E\{(Z_{t-1}, Z_{t-2}, \ldots, Z_{t-p})'(Z_{t-1}, Z_{t-2}, \ldots, Z_{t-p})\}$.

Proof. Omitted. ▲

Example 9.7.2. We use the wheat yield data of Example 9.2.1 for the period 1908 to 1991 to illustrate the nonlinear fitting. The $\hat{\mu}_t$ from Example 9.5.1 and a plot of the data suggest that wheat yields were nearly constant for the first 25 years and for the last several, say five, years. Therefore, we consider a mean function that is constant for the first 25 years and for the last five years. A function that is continuous with continuous first derivative that meets these requirements is

$$\varphi_{t2} = \begin{cases} 0, & 1 \le t \le 25, \\ (t-25)^2, & 25 \le t \le 54, \\ 841 + 58(t-54), & 54 \le t \le 70, \\ 841 + 58(t-54) - 2.9(t-70)^2, & 70 \le t \le 80, \\ 2059, & t \ge 80, \end{cases}$$

where $t = 1$ for 1908. The first part of the function was used in Example 9.2.1. We assume the errors follow a first order autoregressive process. Then our model is

$$Y_t = \beta_0 + \beta_1 \varphi_{t2} + Z_t \tag{9.7.8}$$

for $t = 1, 2, \ldots, n$, where Z_t is the stationary process satisfying

$$Z_t + \alpha_1 Z_{t-1} = e_t ,$$

and we assume the e_t are iid$(0, \sigma_e^2)$.

If we use procedure AUTOREG of SAS/ETS® to compute the Gaussian maximum likelihood estimates, we obtain

$$\underset{(0.380)}{\hat{\beta}_0 = 14.257}, \quad \underset{(0.00037)}{\hat{\beta}_1 = 0.01075}, \quad \underset{(0.107)}{\hat{\alpha}_1 = -0.292}, \quad \hat{\sigma}_e^2 = 3.427 .$$

The model predictions for the next three observations are

$$(\hat{Y}_{1992}, \hat{Y}_{1993}, \hat{Y}_{1994}) = \underset{(1.95)}{(35.78,} \quad \underset{(2.05)}{36.21,} \quad \underset{(2.07)}{36.34)}.$$

The prediction for 1992 is similar to that obtained in Example 9.5.1. The model of Example 9.5.1 postulates a random walk for the mean function and uncorrelated deviations from the mean function. Hence, the predictions for all future periods are the best estimate of the yield for 1991, and the estimate for 1991 is heavily influenced by the actual observation for 1991. The predictions under the model (9.7.8) are based on the estimated trend associated with the function φ_{t2}. Because the function is constant for $t \geq 80$, the predictions ultimately fall on a horizontal line. Because the yield for 1991 is about 2.1 bushels below the trend line, the prediction for 1992 is about $0.29 \times 2.1 = 0.61$ bushels below the trend line, the prediction for 1993 is about $(0.29)^2 \times 2.1 = 0.18$ bushels below the trend line, and the prediction for 1994 is about $(0.29)^3 \times 2.1 = 0.05$ bushels below the trend line. Because the 1991 yield is below the trend line and the estimated autocorrelation is positive, the predictions for the next three years increase toward the trend line.

The standard errors of the predictions constructed in this example are smaller than those of the predictions of Example 9.5.1. In this example, the mean is assumed to be of a known functional form. In Example 9.5.1, the mean is assumed to be random walk. If the mean function is known, then it is easier to predict than the random walk. However, in using the estimated trend function for prediction, one projects the trend function into the future. Often practitioners hesitate to project polynomial trends very far into the future. In our current example, the most recent past of the mean function is assumed to be constant, and we might be less worried about our trend assumption than we would have been in constructing predictions in 1971 using the function of Example 9.2.1. To project linearly in 1971, one would need to believe that the improvements in varieties and pesticides and the use of fertilizer that led to increased yields during the observation period would continue at the same rate during the prediction period. In the case of wheat yields the linear trend predictions would have peformed well for about fifteen years, from 1971 until about 1986. ▲▲

Example 9.7.3. As a second illustration of nonlinear estimation, we use the data on consumption of spirits of Table 9.7.1. If we use nonlinear least squares and fit the model with a first order autoregressive error to the last $n - 1$ observations, we obtain

$$\hat{Y}_t = \underset{(0.304)}{2.421} + \underset{(0.149)}{0.716}\,\varphi_{t1} - \underset{(0.074)}{0.818}\,\varphi_{t2} - \underset{(0.119)}{0.846}\,\varphi_{t3} - \underset{(0.380)}{0.560}\,\varphi_{t4},$$

$$\hat{Z}_t = \underset{(0.078)}{0.788}\,Z_{t-1},$$

and $\hat{\sigma}^2 = 0.00041$, where $\varphi_{t3} = 10^{-2}t$ and $\varphi_{t4} = 10^{-4}(t-35)^2$. If we fit the full model by maximum likelihood, we obtain

$$\hat{Y}_t = \underset{(0.308)}{2.391} + \underset{(0.150)}{0.727}\,\varphi_{t1} - \underset{(0.075)}{0.822}\,\varphi_{t2} - \underset{(0.113)}{0.773}\,\varphi_{t3} - \underset{(0.300)}{0.932}\,\varphi_{t4},$$

$$\hat{Z}_t = \underset{(0.076)}{0.812}\,Z_{t-1},$$

and $\hat{\sigma}^2 = 0.00042$. As expected, the results are very similar to those obtained in Example 9.7.1 because the three procedures have the same asymptotic efficiency.

If we fit the model with a second order autoregressive error by maximum likelihood, we obtain

$$\hat{Y}_t = \underset{(0.333)}{2.473} + \underset{(0.156)}{0.704}\,\varphi_{t1} - \underset{(0.077)}{0.832}\,\varphi_{t2} - \underset{(0.127)}{0.845}\,\varphi_{t3} - \underset{(0.420)}{0.470}\,\varphi_{t4},$$

$$\hat{Z}_t = \underset{(0.140)}{0.742}\,Z_{t-1} + \underset{(0.139)}{0.054}\,Z_{t-2},$$

and $\hat{\sigma}^2 = 0.00042$. The second order coefficient in the autoregressive model is small relative to its standard error, agreeing with the results of Table 9.7.2. ▲▲

Regression models in which the current value of a time series Y_t is a function of current and lagged values of a time series X_t are sometimes called *transfer function* models. The models are often written in terms of the backshift operator. Thus, the simple model

$$Y_t = \beta_1 X_t + \beta_2 X_{t-1} + e_t$$

can be written as

$$Y_t = (\beta_1 + \beta_2 \mathcal{B})X_t + e_t.$$

More complicated models may contain ratios of polynomials in the backshift operator. For example,

$$\begin{aligned} Y_t &= [A(\mathcal{B})]^{-1}B(\mathcal{B})X_t + e_t \\ &= \sum_{i=0}^{\infty} c_i X_{t-i} + e_t, \end{aligned}$$

where $A(\mathcal{B}) = 1 - a_1\mathcal{B} - \cdots - a_r\mathcal{B}^r$, $B(\mathcal{B}) = b_1 - b_2\mathcal{B} - \cdots - b_s\mathcal{B}^s$, the c_i are the coefficients defined by $[A(\mathcal{B})]^{-1}B(\mathcal{B})$, and it is assumed that the roots of $A(\mathcal{B}) = 0$ are greater than one so that the coefficients converge to zero. The set of coefficients $(c_0, c_1, \ldots)$ is called the *impulse response function*. Such models fall into the class of nonlinear regression models covered by the theory of Section 5.5, provided X_t is reasonably well behaved.

9.8 REGRESSION EQUATIONS WITH LAGGED DEPENDENT VARIABLES AND TIME SERIES ERRORS

In the classical regression problem the error in the equation is assumed to be distributed independently of the independent variables in the equation. Independent variables with this property we call *ordinary*. In our investigation of the estimation of the autoregressive parameters in Chapter 8, we noted that the autoregressive representation had the appearance of a regression problem, but that the vector of errors was not independent of the vector of lagged values of the time series. A generalization of the autoregressive model is one containing ordinary independent variables and also lagged values of the dependent variable as explanatory variables.

Consider the model

$$\begin{aligned} Y_t &= \sum_{i=1}^{k} \varphi_{ti}\zeta_i - \sum_{j=1}^{p} \alpha_j Y_{t-j} + e_t \\ &= \mathbf{X}_t\boldsymbol{\theta} + e_t \end{aligned} \tag{9.8.1}$$

for $t = 1, 2, \ldots,$ where $\mathbf{X}_t = (\varphi_{t1}, \varphi_{t2}, \ldots, \varphi_{tk}, Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})$, $\boldsymbol{\theta} = (\zeta_1, \zeta_2, \ldots, \zeta_k, -\alpha_1, -\alpha_2, \ldots, -\alpha_p)$, the φ_{ti} are fixed constants, and e_t are uncorrelated $(0, \sigma^2)$ random variables. The characteristic equation associated with equation (9.8.1) is

$$m^p + \sum_{j=1}^{p} \alpha_j m^{p-j} = 0. \tag{9.8.2}$$

Such models are common in economics. See Judge *et al.* (1985, Chapter 10). Ordinary least squares is a natural way to estimate the parameters of equation (9.8.1). Let

$$\hat{\boldsymbol{\theta}} = \left(\sum_{t=1}^{n} \mathbf{X}_t'\mathbf{X}_t\right)^{-1} \sum_{t=1}^{n} \mathbf{X}_t'\mathbf{Y}_t \tag{9.8.3}$$

denote the least squares estimator. The limiting distribution of the properly standardized $\hat{\boldsymbol{\theta}}$ is normal under mild assumptions.

Theorem 9.8.1. Let Y_t satisfy equation (9.8.1), where $Y_0, Y_{-1}, \ldots, Y_{-p+1}$ are fixed. Assume

$$E\{(e_t, e_t^2) \mid \mathcal{A}_{t-1}\} = (0, \sigma^2) \qquad \text{a.s.}$$

and

$$E\{|e_t|^{2+\nu} \mid \mathcal{A}_{t-1}\} < L < \infty \qquad \text{a.s.}$$

for all t and some $\nu > 0$, where $\mathcal{A}_{t-1}$ is the sigma-field generated by

$(\mathbf{X}_t, \mathbf{X}_{t-1}, \ldots, \mathbf{X}_1)$. Assume there exists a sequence $\{\mathbf{M}_n\}$ of fixed nonsingular matrices such that

$$\operatorname*{plim}_{n\to\infty} \mathbf{M}_n^{-1} \sum_{t=1}^{n} \mathbf{X}_t'\mathbf{X}_t\mathbf{M}_n^{-1\prime} = \lim_{n\to\infty} E\left\{\mathbf{M}_n^{-1} \sum_{t=1}^{n} \mathbf{X}_t'\mathbf{X}_t\mathbf{M}_n^{-1\prime}\right\} = \mathbf{I} \tag{9.8.4}$$

and

$$\lim_{n\to\infty} \sup_{1\le t\le n} E\{|\mathbf{M}_n^{-1}\mathbf{X}_t'|^2\} = 0\,. \tag{9.8.5}$$

Let $\hat{\boldsymbol{\theta}}$ be defined by (9.8.3). Then

$$\mathbf{M}_n'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{\mathscr{L}} N(\mathbf{0}, \mathbf{I})\,.$$

Proof. Let $\boldsymbol{\eta}'$ be an arbitrary $(k+p)$-dimensional row vector, and define

$$Z_{tn} = \boldsymbol{\eta}'\mathbf{M}_n^{-1}\mathbf{X}_t'e_t\,,$$

$$\delta_{tn}^2 = E\{Z_{tn}^2 \mid \mathscr{A}_{t-1}\} = \boldsymbol{\eta}'\mathbf{M}_n^{-1}\mathbf{X}_t'\mathbf{X}_t\mathbf{M}_n^{-1\prime}\boldsymbol{\eta}\sigma^2\,,$$

$$s_{nn}^2 = E\{V_{nn}^2\} = E\left\{\boldsymbol{\eta}'\mathbf{M}_n^{-1}\sum_{t=1}^{n}\mathbf{X}_t'\mathbf{X}_t\mathbf{M}_n^{-1\prime}\boldsymbol{\eta}\sigma^2\right\},$$

where $V_{nn}^2 = \Sigma_{t=1}^n \delta_{tn}^2$. By assumption,

$$\operatorname{plim} s_{nn}^{-2}V_{nn}^2 = 1$$

and condition (ii) of Theorem 5.3.4 is satisfied. We have

$$s_{nn}^{-2}\sum_{t=1}^{n}\int_{R_t} |\boldsymbol{\eta}'\mathbf{M}_n^{-1}\mathbf{X}_t'e|^2\, dF_t(e) \le s_{nn}^{-2}\sum_{t=1}^{n}|\boldsymbol{\eta}'\mathbf{M}_n^{-1}\mathbf{X}_t'|^2\int_{R_0} e^2\, dF_t(e)\,,$$

where $R_t = \{e: |e| > |\boldsymbol{\eta}'\mathbf{M}_n^{-1}\mathbf{X}_t'|^{-1}\epsilon\}$ for $t = 1, 2, \ldots, F_t(e)$ is the distribution function of e_t, and

$$R_0 = \left\{e: |e| > \left(\sup_{1\le t\le n}|\boldsymbol{\eta}'\mathbf{M}_n^{-1}\mathbf{X}_t'|\right)^{-1}\epsilon\right\}.$$

Because $s_{nn}^{-2}V_{nn}^2 \xrightarrow{P} 1$ and because the supremum of the integrals of e^2 over R_0 goes to zero as n increases, condition (iii) of Theorem 5.3.4 is satisfied. Because $\boldsymbol{\eta}$ is arbitrary, the joint normality is established. ▲

In Theorem 9.8.1, conditions on the nature of the difference equation associated with (9.8.1) are imposed through the assumptions (9.8.4) and (9.8.5). If $\{\boldsymbol{\Phi}_t\}$ is a fixed sequence, where $\boldsymbol{\Phi}_t = (\varphi_{t1}, \varphi_{t2}, \ldots, \varphi_{tk})$, then the existence of a sequence of fixed matrices $\{\mathbf{M}_{11n}\}$ such that

$$\lim_{n\to\infty} \mathbf{M}_{11n}^{-1} \sum_{t=1}^{n} \mathbf{\Phi}_t' \mathbf{\Phi}_t \mathbf{M}_{11n}^{-1\prime} = \mathbf{I},$$

$$\lim_{n\to\infty} \sup_{1\le t\le n} |\mathbf{M}_{11n}^{-1} \mathbf{\Phi}_t'|^2 = 0,$$

together with the assumption that the roots of (9.8.2) are less than one in absolute value, is sufficient for (9.8.4) and (9.8.5). See Fuller, Hasza, and Goebel (1981).

The conclusion of Theorem 9.8.1 can be obtained with martingale difference errors or with iid$(0, \sigma^2)$ errors. The martingale difference property is critical to the proof. If the e_t are autocorrelated, $E\{Y_{t-1}e_t\}$ may not be zero, in which case the least squares estimator is no longer consistent.

To obtain consistent estimators for the parameters in the presence of autocorrelated errors, we use the method of instrumental variables introduced in Section 5.6. The lagged values of φ_{ti} are the natural instrumental variables. To simplify the presentation, we consider a model with two independent variables and two lagged values of Y_t,

$$Y_t = \beta_1 \varphi_{t1} + \beta_2 \varphi_{t2} + \lambda_1 Y_{t-1} + \lambda_2 Y_{t-2} + Z_t, \qquad t = 1, 2, \ldots, \tag{9.8.6}$$

where Z_t is a stationary time series with properties described in Theorem 9.8.2.

There are several methods of instrumental variable estimation. In econometrics, the two common instrumental variable procedures are called two-stage least squares and limited information maximum likelihood. The two procedures have the same asymptotic efficiency. For the model (9.8.6), the first step of the two-stage procedure is the regression of Y_{t-1} and Y_{t-2} on φ_{t1}, $\varphi_{t-1,1}$, $\varphi_{t-2,1}$, φ_{t2}, $\varphi_{t-1,2}$, and $\varphi_{t-2,2}$. The predicted values $\hat{Y}_{t-1}$ and $\hat{Y}_{t-2}$ are obtained from these regressions, and the instrumental variable estimates are obtained by regressing Y_t on $(\varphi_{t1}, \varphi_{t2}, \hat{Y}_{t-1}, \hat{Y}_{t-2})$.

To apply an instrumental variable procedure to the model (9.8.6), we assume the matrix of sums of squares and products of $(\varphi_{t1}, \varphi_{t2}, \varphi_{t-1,1}, \varphi_{t-2,1}, \varphi_{t-1,2}, \varphi_{t-2,2})$ to be nonsingular and to satisfy assumption 1 preceding Theorem 5.6.1. We require only two instrumental variables in order to estimate the parameters λ_1 and λ_2, and some of the variables may be omitted if singularities appear in practice. For example, if $\mathbf{\Phi} = (\varphi_{.0}, \varphi_{.1}, \varphi_{.2})$, where $\varphi_{t0} \equiv 1$ and $\varphi_{t1} = t$, then the matrix with rows $(1, t, \varphi_{t2}, t-1, \varphi_{t-1,2}, t-2, \varphi_{t-2,2})$ will be singular. One can delete variables from the regression to create a nonsingular regression problem. In the example case one can omit $\varphi_{t-1,1} = t-1$ and $\varphi_{t-2,1} = t-2$.

In Theorem 9.8.2, we show that the instrumental variable estimator is consistent for the model (9.8.6). The conclusions generalize immediately to any number of φ_{ti} and any number of lagged Y's.

Theorem 9.8.2. Let $\{Y_t: t \in (1, 2, \ldots)\}$ satisfy the model (9.8.6), where

$$Z_t = \sum_{j=0}^{\infty} b_j e_{t-j},$$

the b_j are absolutely summable, the roots of $m^2 - \lambda_1 m - \lambda_2 = 0$ are less than one in absolute value, (Y_{-1}, Y_0) are bounded in probability, and the e_t are independent $(0, \sigma^2)$ random variables with $E\{|e_t|^{2+\delta}\} < L^{2+\delta}$ for some finite L and some $\delta > 0$.

Let φ_{t1} and φ_{t2} satisfy the assumptions (9.1.8) and (9.1.9). Let

$$\mathbf{Q} = \lim_{n\to\infty} \mathbf{Q}_n$$

be nonsingular, where

$$\mathbf{Q}_n = \mathbf{D}_{1n}^{-1}(\Phi; \psi)'(\Phi; \psi)\mathbf{D}_{1n}^{-1},$$

$(\Phi; \psi)$ is the matrix whose tth row is $(\varphi_{t1}, \varphi_{t2}, \varphi_{t-1,1}, \varphi_{t-2,1}, \varphi_{t-1,2}, \varphi_{t-2,2})$ for $t = 3, 4, \ldots, n$, and

$$\mathbf{D}_{1n}^2 = \operatorname{diag}\left[\sum_{t=3}^{n} \varphi_{t1}^2, \sum_{t=3}^{n} \varphi_{t2}^2, \sum_{t=2}^{n-1} \varphi_{t1}^2, \sum_{t=1}^{n-2} \varphi_{t1}^2, \sum_{t=2}^{n-1} \varphi_{t2}^2, \sum_{t=1}^{n-2} \varphi_{t2}^2\right].$$

Let $\hat{\boldsymbol{\theta}} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\lambda}_1, \hat{\lambda}_2)$ be the instrumental variable estimator, and let

$$\mathbf{D}_{2n} = \operatorname{diag}\left\{\left(\sum_{t=1}^{n} \varphi_{t1}^2\right)^{1/2}, \left(\sum_{t=1}^{n} \varphi_{t2}^2\right)^{1/2}, S_n, S_n\right\},$$

where $S_n^2 = \sum_{t=1}^{n}(\beta_1\varphi_{t1} + \beta_2\varphi_{t2})^2$. Then

$$\mathbf{D}_{2n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = O_p(1).$$

Proof. By the difference equation properties of our model, we may write

$$Y_t = \beta_1 \sum_{j=0}^{t-1} w_j \varphi_{t-j,1} + \beta_2 \sum_{j=0}^{t-1} w_j \varphi_{t-j,2} + U_t + c_{1t} Y_0 + c_{2t} Y_{-1}, \qquad (9.8.7)$$

where $U_t = \sum_{j=0}^{t-1} w_j Z_{t-j}$, the w_j are defined in terms of λ_1 and λ_2 by Theorem 2.6.1, and c_{1t} and c_{2t} go to zero as t goes to infinity. Equation (9.8.7) gives Y_{t-1} and Y_{t-2} as functions of lagged values of φ_{t1} and φ_{t2} and of a random term U_t. By our assumptions on the φ_{ti} and since β_1 and β_2 are not both zero, the current and lagged values of φ_{ti} satisfy the conditions placed on Φ and ψ in Theorem 5.6.1. Likewise, Z_t satisfies the conditions on Z_t of Theorem 5.6.1.

The $\mathbf{B}_n$ of assumption 4 preceding Theorem 5.6.1 is

$$\mathbf{B}_n = \mathbf{D}_{1n}^{-1}(\Phi; \psi)'\Gamma(\Phi; \psi)\mathbf{D}_{1n}^{-1},$$

where the ijth element of Γ is $\gamma_Z(|i-j|)$, and the matrix $\mathbf{B} = \lim_{n\to\infty} \mathbf{B}_n$ is of the same form as the matrix $\mathbf{B}$ of Theorem 9.1.1. The matrices $\mathbf{G}_{nij}$ of assumption 4 preceding Theorem 5.6.1 are of a similar form, with the covariance matrix of U_{t-1}, of U_{t-2}, or of U_{t-1} with U_{t-2} replacing that of Z_t.

Therefore, the conditions of Theorem 5.6.1 are satisfied, and the order in probability of the error in the estimators is given by that theorem. ▲

By Theorem 9.8.2 the instrumental variable estimator is consistent when the error in the equation is a stationary moving average time series with absolutely summable coefficients. Although the instrumental variable estimators are not the most efficient estimators, they are not dependent on a specification for the error process. Hence, they can be used as preliminary estimators. In some applications, one will use the instrumental variable estimators to obtain information about the error process. Given the instrumental variable estimates $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\lambda}_1$, $\hat{\lambda}_2$ of the model (9.8.6), we can calculate the estimated residuals

$$\hat{Z}_t = Y_t - \hat{\beta}_1 \varphi_{t1} - \hat{\beta}_2 \varphi_{t2} - \hat{\lambda}_1 Y_{t-1} - \hat{\lambda}_2 Y_{t-2} . \tag{9.8.8}$$

The presence of the lagged Y-values in the equation means that the results of Theorem 9.3.1 do not hold for autocorrelations computed from these residuals. However, we are able to obtain a somewhat weaker result.

Theorem 9.8.3. Given the assumptions of Theorem 9.8.2, the sample covariances calculated from the instrumental variable residuals satisfy

$$\hat{\gamma}_{\hat{Z}}(h) - \hat{\gamma}_Z(h) = O_p(r_n) ,$$

where

$$\hat{\gamma}_{\hat{Z}}(h) = \frac{1}{n} \sum_{t=3}^{n-h} \hat{Z}_t \hat{Z}_{t+h} , \quad \hat{\gamma}_Z(h) = \frac{1}{n} \sum_{t=3}^{n-h} Z_t Z_{t+h} , \qquad h = 0, 1, 2, \ldots, n-1 ,$$

$\hat{Z}_t$ is defined in (9.8.8), $r_n = \max\{n^{-1/2}, S_n^{-1}\}$, and S_n is defined in Theorem 9.8.2.

Proof. Using the definition of $\hat{Z}_t$, we have, for $h = 0, 1, \ldots, n-1$,

$$\begin{aligned}
\hat{\gamma}_{\hat{Z}}(h) &= \frac{1}{n} \sum_{t=3}^{n-h} \hat{Z}_t \hat{Z}_{t+h} \\
&= \frac{1}{n} \left[\sum_{t=3}^{n-h} Z_t Z_{t+h} - (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \sum_{t=3}^{n-h} \mathbf{A}_t' Z_{t+h} \right. \\
&\quad - \sum_{t=3}^{n-h} Z_t \mathbf{A}_{t+h,\cdot} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\
&\quad \left. + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \sum_{t=3}^{n-h} \mathbf{A}_t' \mathbf{A}_{t+h,\cdot} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right] ,
\end{aligned}$$

where $\mathbf{A}_t = (\varphi_{t1}, \varphi_{t2}, Y_{t-1}, Y_{t-2})$. Now, for example,

$$\frac{1}{n}\sum_{t=3}^{n-h} Y_{t-1}Z_{t+h} \leq n^{-1/2}\left(\sum_{t=3}^{n-h} Y_{t-1}^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=3}^{n-h} Z_{t+h}^2\right)^{1/2}$$
$$= O_p(\max\{1, n^{-1/2}S_n\}).$$

Therefore,

$$\frac{1}{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\mathbf{D}_{2n}\mathbf{D}_{2n}^{-1}\sum_{t=3}^{n-h}\mathbf{A}_{t\cdot}'Z_{t+h} = O_p(\max\{n^{-1/2}, S_n^{-1}\}),$$

$$\frac{1}{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\mathbf{D}_{2n}\mathbf{D}_{2n}^{-1}\sum_{t=3}^{n-h}\mathbf{A}_{t\cdot}'\mathbf{A}_{t+h,\cdot}\mathbf{D}_{2n}^{-1}\mathbf{D}_{2n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) = O_p(\max\{n^{-1/2}, S_n^{-1}\}),$$

where $\mathbf{D}_{2n}$ is defined in Theorem 9.8.2. ▲

Because the estimated autocorrelations are consistent estimators, the estimated autocorrelations of the residuals can be used to develop a model for the error process. Given an original model such as (9.8.6) and a specified form for the error process, it is natural to estimate the entire model by a nonlinear fitting procedure.

To introduce such models, let Z_t be the first order autoregressive process

$$Z_t = \rho Z_{t-1} + e_t, \tag{9.8.9}$$

where $|\rho| < 1$ and the e_t are independent $(0, \sigma^2)$ random variables. We may then write (9.8.6) as a nonlinear model with lagged values of Y_t among the explanatory variables, and whose error term e_t is a sequence of independent random variables. That is, substituting (9.8.6) into (9.8.9), we obtain

$$\begin{aligned} Y_t &= \beta_1\varphi_{t1} + \beta_2\varphi_{t2} + (\lambda_1+\rho)Y_{t-1} + (\lambda_2 - \rho\lambda_1)Y_{t-2} \\ &\quad - \rho\beta_1\varphi_{t-1,1} - \rho\beta_2\varphi_{t-1,2} - \rho\lambda_2 Y_{t-3} + e_t \\ &\overset{\text{(say)}}{=} f(\mathbf{W}_t;\boldsymbol{\theta}) + e_t, \qquad t = 4, 5, \ldots, n, \end{aligned} \tag{9.8.10}$$

where $\mathbf{W}_t = (\varphi_{t1}, \varphi_{t2}, \varphi_{t-1,1}, \varphi_{t-1,2}, Y_{t-1}, Y_{t-2}, Y_{t-3})$ and $\boldsymbol{\theta}' = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (\beta_1, \beta_2, \lambda_1, \lambda_2, \rho)$. Note that we have expanded the vector of parameters to include ρ.

By Theorem 5.5.3 and its corollaries, the least squares estimator of $\boldsymbol{\theta}$ will be consistent and, when properly standardized, asymptotically normally distributed under mild assumptions. Note that the model (9.8.10) is of the same form as (9.8.1) with coefficients that satisfy nonlinear restrictions.

Example 9.8.1. To illustrate the estimation of a model containing lagged values of the dependent variable, we use a model studied by Ball and St. Cyr (1966). The data, displayed in Table 9.8.1, are for production and employment in the bricks, pottery, glass, and cement industry of the United Kingdom. The

Table 9.8.1. Production Index and Employment of the Bricks, Pottery, Glass, and Cement Industry of the United Kingdom

Year-Quarter	Logarithm of Production Index	Logarithm of Employment (000)
1955–1	4.615	5.844
2	4.682	5.846
3	4.644	5.852
4	4.727	5.864
1956–1	4.625	5.855
2	4.700	5.844
3	4.635	5.841
4	4.663	5.838
1957–1	4.625	5.823
2	4.644	5.817
3	4.575	5.823
4	4.635	5.817
1958–1	4.595	5.802
2	4.625	5.790
3	4.564	5.790
4	4.654	5.784
1959–1	4.595	5.778
2	4.700	5.781
3	4.654	5.802
4	4.727	5.814
1960–1	4.719	5.820
2	4.787	5.826
3	4.727	5.838
4	4.787	5.841
1961–1	4.771	5.838
2	4.812	5.844
3	4.796	5.846
4	4.828	5.855
1962–1	4.762	5.852
2	4.868	5.864
3	4.804	5.864
4	4.844	5.861
1963–1	4.754	5.844
2	4.875	5.835
3	4.860	5.841
4	4.963	5.852
1964–1	4.956	5.849
2	5.004	5.858

SOURCE: Ball and St. Cyr (1966). Reproduced from *Review of Economic Studies* Vol. 33, No. 3 with the permission of the authors and editor.

quarterly data cover the period 1955–1 to 1964–2. The model studied by Ball and St. Cyr can be written as

$$Y_t = \beta_0 + \beta_1 \varphi_t + \beta_2 t + \beta_3 D_{t1} + \beta_4 D_{t2} + \beta_5 D_{t3} + \lambda Y_{t-1} + Z_t, \quad (9.8.11)$$

where

Y_t = the logarithm of employment in thousands of the bricks, pottery, glass, and cement industry of the United Kingdom,

φ_t = the logarithm of the production index of the industry,

t = time with 1954–4 as the origin.

$$D_{tj} = \begin{cases} 1 & \text{if observation } t \text{ occurs in quarter } j, \\ -1 & \text{if observation } t \text{ occurs in quarter 4}, \\ 0 & \text{otherwise}. \end{cases}$$

We assume Z_t is a zero mean stationary time series, $\beta_1 \neq 0$, and $|\lambda| < 1$.

To construct the instrumental variable estimator, we use the limited information maximum likelihood procedure SYSLIN of SAS/ETS®. In the terminology of that program Y_t and Y_{t-1} are endogenous variables and φ_t, φ_{t-1}, t, D_{t1}, D_{t2}, and D_{t3} are instruments. The instrumental variable estimated equation is

$$\hat{Y}_t = 0.27 + 0.077\varphi_t - 0.00042t - 0.00492D_{t1} - 0.00336D_{t2} + 0.00728D_{t3} + 0.894Y_{t-1}.$$

The program calculates estimated standard errors under the assumption that the errors are uncorrelated. The estimated standard errors are biased if the errors are autocorrelated. Because we are considering the possibility that the errors are autocorrelated, we do not present the estimated standard errors.

An analysis of the estimated residuals $\hat{Z}_t$, such as that in Table 9.7.2, indicated that the data are consistent with the hypothesis that Z_t is a first order autoregressive process with a parameter of about 0.42.

To estimate the model with autoregressive errors, we use the maximum likelihood option of procedure ARIMA of SAS/ETS®. We specify the model (9.8.11) with a first order autoregressive error. While the error in the first observation is correlated with the lagged dependent variable for that observation, we chose to include that observation in the fit. Thus, the program is used to maximize a pseudolikelihood that treats Y_{t-1} as an ordinary explanatory variable. The estimated model is

$$\begin{aligned} \hat{Y}_t = {} & \underset{(0.41)}{0.88} + \underset{(0.032)}{0.097}\,\varphi_t - \underset{(0.00030)}{0.00055}\,t - \underset{(0.00186)}{0.00368}\,D_{t1} \\ & - \underset{(0.00181)}{0.00432}\,D_{t2} + \underset{(0.00162)}{0.00734}\,D_{t3} + \underset{(0.081)}{0.773}\,Y_{t-1}, \end{aligned}$$

and $\hat{\rho} = 0.386$ (0.194), where the numbers in parentheses are the estimated

standard errors. Although this is a fairly small sample with a fairly large number of explanatory variables, the maximum likelihood estimates are similar to the instrumental variable estimates. Because of the small sample and the presence of time among the explanatory variables, it is a reasonable conjecture that the estimator $\hat{\rho}$ possesses a negative bias.

There are two important reasons for recognizing the potential for autocorrelated errors in an analysis such as this one. First, the simple least squares estimators are no longer consistent when the errors are autocorrelated. Second, the estimators of the variances of the estimators may be seriously biased. A hypothesis of some interest in the study of employment was the hypothesis that $\beta_1 + \lambda = 1$. This corresponds to the hypothesis of constant returns to scale for labor input. In the simple least squares analysis the sum of these coefficients is 0.89 with an estimated standard error of 0.045. The resulting "t-statistic" of -2.5 suggests that there are increasing returns to scale to labor in this industry. On the other hand, the analysis recognizing the possibility of autocorrelated errors estimates the sum as 0.88, but with a standard error of 0.072. The associated "t-statistic" of -1.7 and the potential for small sample bias in the estimators could lead one to accept the hypothesis that the sum is one. ▲▲

REFERENCES

Section 9.1. Anderson (1971), Grenander (1954), Grenander and Rosenblatt (1957), Hannan (1970).

Section 9.2. Ahlberg, Nilson, and Walsh (1967), Fuller (1969), Gallant and Fuller (1973), Greville (1969), Poirier (1973).

Section 9.3. Bollerslev, Chou, and Kroner (1992), Durbin and Watson (1950, 1951), Engle (1982), Granger and Teräsvirta (1993), Hannan (1969), Harvey, Ruiz, and Shephard (1994), Melino and Turnbull (1990), von Neumann (1941, 1942).

Section 9.4. Davis (1941), Grether and Nerlove (1970), Jorgenson (1964), Kendall and Stuart (1966), Lovell (1963), Malinvaud (1970a), Nerlove (1964), Stephenson and Farr (1972), Tintner (1940, 1952).

Section 9.5. Harvey (1989).

Section 9.6. Davis (1941), Durbin (1962, 1963), Hannan (1958), Tintner (1940, 1952), Rao and Tintner (1963).

Section 9.7. Amemiya (1973), Cochran and Orcutt (1949), Durbin (1960), Hildreth (1969), Malinvaud (1970a), Prest (1949).

Section 9.8. Aigner (1971), Amemiya and Fuller (1967), Ball and St. Cyr (1966), Dhyrmes (1971), Fuller and Martin (1961), Griliches (1967), Hatanaka (1974), Malinvaud (1970a), Nagaraj and Fuller (1991, 1992), Wallis (1967).

EXERCISES

1. Let the time series X_t be defined by

$$X_t = \beta_0 + \beta_1 t + Z_t ,$$

where

$$Z_t = \rho Z_{t-1} + e_t, \qquad |\rho| < 1, \quad t = 0, \pm 1, \pm 2, \ldots,$$

and $\{e_t\}$ is a sequence of independent $(0, \sigma^2)$ random variables. Compute the covariance matrices of $\hat{\boldsymbol{\beta}}_S$ of (9.1.2) and $\hat{\boldsymbol{\beta}}_G$ of (9.1.3) for $\rho = 0.9$ and $n = 5$ and 10.

2. The data in the accompanying table are the forecasts of the yield of corn in Illinois released by the United States Department of Agriculture in August of the production year.
   (a) For this set of data estimate a mean function meeting the following restrictions:
      (i) The mean function is linear from 1944 through 1956, quadratic from 1956 through 1960, linear from 1960 through 1965, quadratic from 1965 through 1969, and linear from 1969 through 1974.
      (ii) The mean function, considered as a continuous function of time, has a continuous first derivative.
   (b) Reestimate the mean function subject to the restriction that the derivative is zero for the last five years. Test the hypothesis that the slope for the last five years is zero.
   (c) Plot the data and the estimated mean function.

August Forecast of Corn Yield for Illinois

Year	Yield	Year	Yield	Year	Yield
1944	45.5	1955	58.0	1965	88.0
1945	43.0	1956	59.0	1966	82.0
1946	55.0	1957	52.0	1967	96.0
1947	45.0	1958	64.0	1968	103.0
1948	58.0	1959	63.0	1969	95.0
1949	61.0	1960	63.0	1970	93.0
1950	58.0	1961	73.0	1971	92.0
1951	56.0	1962	79.0	1972	97.0
1952	56.0	1963	81.0	1973	105.0
1953	58.0	1964	85.0	1974	86.0
1954	45.0				

SOURCE: U.S. Department of Agriculture, *Crop Production*, various issues.

3. Evaluate (9.3.5) and (9.3.6) for $n = 10$ and the model of Exercise 9.1. The covariance between $\sum_{t=2}^{n} Z_t Z_{t-1}$ and $\sum_{t=1}^{n} Z_t^2$ for the first order autoregressive process can be evaluated using Theorem 6.2.1. Using these approximations and the approximation

$$E\left\{\frac{A}{B}\right\} \doteq \frac{E\{A\}}{E\{B\}} + [E\{B\}]^{-2}\left[\operatorname{Cov}\{A, B\} - \left(\frac{E\{A\}}{E\{B\}}\right)\operatorname{Var}\{B\}\right],$$

find the approximate expected value of $\hat{r}_2(1)$ for the model and sample sizes of Exercise 9.1.

4. Using a period of three observations, construct a moving average to remove a linear trend from the third observation and hence show that the second difference operator is a multiple of the moving average constructed to remove a linear trend from the third observation in a group of three observations.

5. Define X_t by

$$X_t = \rho X_{t-1} + e_t, \qquad |\rho| < 1,$$

where the e_t are independent $(0, \sigma^2)$ random variables. Find the autocorrelation functions and spectral densities of

$$Y_t = X_t - X_{t-1} \quad \text{and} \quad W_t = X_t - 2X_{t-1} + X_{t-2}.$$

Plot the spectral densities for $\rho = 0.5$.

6. Consider the model

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \sum_{i=1}^{4} \alpha_i \delta_{ti} + Z_t,$$

where

$$\delta_{ti} = \begin{cases} 1 & \text{for quarter } i \ (i = 1, 2, 3, 4), \\ 0 & \text{otherwise}, \end{cases}$$

$$\sum_{i=1}^{4} \alpha_i = 0,$$

and Z_t is a stationary time series. Assuming that the model holds for a period of 20 observations (5 years), construct the weights needed to decompose the most recent observation in a set of 20 observations into "trend," "seasonal," and "remainder." Define trend to be $\beta_0 + \beta_1 t + \beta_2 t^2$ for the tth observation, and seasonal for quarter i to be the α_i associated with that quarter.

7. Obtain the conclusion of Proposition 9.6.3 directly by differencing the autoregressive moving average time series.

8. Prove Lemma 9.4.1.

9. Prove Lemma 9.4.2.

10. Using the data of Exercise 2 and the model of part b of that exercise, construct the statistics (9.3.8) and (9.3.9).

11. Assume that the data of Exercise 2 of Chapter 7 satisfy the model

$$Y_t = \sum_{j=1}^{4} \alpha_j \delta_{tj} + \beta t + Z_t,$$

where

$$\delta_{tj} = \begin{cases} 1 & \text{for quarter } j, \\ 0 & \text{otherwise,} \end{cases}$$

Z_t is the stationary autoregressive process

$$Z_t = \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + e_t,$$

and $\{e_t\}$ is a sequence of normal independent $(0, \sigma^2)$ random variables.

(a) Estimate the parameters of the model. Estimate the standard errors of your estimators. Do the data support the model for the error time series?

(b) Using the model of part a, estimate gasoline consumption for the four quarters of 1974. Construct 95% confidence intervals for your predictions, using normal theory. Actual consumption during 1974 was [2223, 2540, 2596, 2529]. What happened? *Computational hint:* If a computer program is available that fits regression models with autoregressive errors, the parameters and predictions can be obtained in one fitting operation. Add four "observations" and four independent variables to the data set. The four added independent variables are defined by

$$\varphi_{t,5+i} = \begin{cases} -1 & \text{for 1974-}i, \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, 3, 4$. The four observations on the original data for 1973 and the four added "observations" for 1974 are displayed in the accompanying table. The values of the original independent variables for the prediction period are the true values of these independent variables, while the values of Y_t for the prediction period are zero. The augmented regression model with autoregressive errors is then estimated. The coefficients for the four added independent variables are the predictions, and the estimated standard errors are appropriate for the predictions.

Year	Quarter	Dependent Variable	δ_{t1}	δ_{t2}	δ_{t3}	δ_{t4}	t	Independent Variables for Predictions			
1973	1	2454	1	0	0	0	53	0	0	0	0
	2	2647	0	1	0	0	54	0	0	0	0
	3	2689	0	0	1	0	55	0	0	0	0
	4	2549	0	0	0	1	56	0	0	0	0
1974	1	0	1	0	0	0	57	−1	0	0	0
	2	0	0	1	0	0	58	0	−1	0	0
	3	0	0	0	1	0	59	0	0	−1	0
	4	0	0	0	0	1	60	0	0	0	−1

12. Using the Prest data of Table 9.7.1, estimate the parameters of the model

$$Y_t = \beta_0 + \beta_1 t + \beta_2 (t-35)^2 + \beta_3 \varphi_{t1} + \beta_4 \varphi_{t2} + \lambda Y_{t-1} + Z_t,$$
$$Z_t = \rho Z_{t-1} + e_t,$$

where $\{e_t\}$ is a sequence of independent $(0, \sigma^2)$ random variables with $E\{e_t^4\} = \eta\sigma^4$ and $|\rho| < 1$, $|\lambda| < 1$.

13. Let Y_t satisfy

$$Y_t = \alpha\varphi_t + Z_t,$$
$$Z_t = e_t + \beta e_{t-1}, \qquad t = 1, 2, \ldots,$$

where $|\beta| < 1$, $\{e_t\}$ is a sequence of independent $(0, \sigma^2)$ random variables with $E\{e_t^4\} = \eta\sigma^4$, and $\{\varphi_t\}$ satisfies the assumptions (9.1.8) and (9.1.9). Let

$$\hat{\alpha}_S = \left(\sum_{t=1}^{n} \varphi_t^2\right)^{-1} \sum_{t=1}^{n} \varphi_t Y_t,$$
$$\hat{Z}_t = Y_t - \hat{\alpha}_S \varphi_t,$$
$$e_t(\hat{Z}; \tilde{\tilde{\beta}}) = \begin{cases} 0, & t = 1, \\ \hat{Z}_t - \tilde{\tilde{\beta}} e_{t-1}(\hat{Z}; \tilde{\tilde{\beta}}), & t = 2, 3, \ldots, n, \end{cases}$$
$$W_t(Y; \tilde{\tilde{\beta}}) = \begin{cases} 0, & t = 1, \\ e_{t-1}(\hat{Z}; \tilde{\tilde{\beta}}) - \tilde{\tilde{\beta}} W_{t-1}(Y; \tilde{\tilde{\beta}}), & t = 2, 3, \ldots, n, \end{cases}$$

and $\tilde{\tilde{\beta}}$ is constructed from the first order autocorrelation of the $\hat{Z}_t$ by the rule (8.3.1) and restricted to $(-1, 1)$.

(a) Show that

$$\hat{\hat{\beta}} - \hat{\beta} = O_p(n^{-1}),$$

where

$$\hat{\hat{\beta}} = \tilde{\tilde{\beta}} + \Delta\hat{\hat{\beta}},$$

$$\Delta\hat{\hat{\beta}} = \frac{\sum_{t=1}^{n} e_t(\hat{Z}; \tilde{\tilde{\beta}}) W_t(\hat{Z}; \tilde{\tilde{\beta}})}{\sum_{t=1}^{n} [W_t(\hat{Z}; \tilde{\tilde{\beta}})]^2},$$

and $\hat{\beta}$ is the analogous estimator constructed by replacing $\hat{Z}$ with Z.

(b) Let $\hat{\boldsymbol{\epsilon}} = \hat{\mathbf{T}}\mathbf{z}$, where

$$\hat{\epsilon}_t = \hat{\mathbf{T}}_{t\cdot}\mathbf{z} = \begin{cases} (1+\hat{\beta}^2)^{-1}Z_1, & t=1, \\ Z_t - \hat{\beta}\hat{\epsilon}_{t-1}, & t=2,3,\ldots,n, \end{cases}$$

and $\hat{\mathbf{T}}_{t\cdot}$ is the tth row of $\hat{\mathbf{T}}$. Show that

$$\left[\sum_{t=1}^{n} \varphi_t^2\right]^{1/2} (\hat{\alpha}_G - \tilde{\alpha}) = O_p(n^{-1/2}),$$

where

$$\tilde{\alpha} = [\boldsymbol{\Phi}'\hat{\mathbf{T}}'\hat{\mathbf{T}}\boldsymbol{\Phi}]^{-1}\boldsymbol{\Phi}'\hat{\mathbf{T}}'\hat{\mathbf{T}}\mathbf{y},$$

$$\hat{\alpha}_G = [\boldsymbol{\Phi}'\mathbf{V}^{-1}\boldsymbol{\Phi}]^{-1}\boldsymbol{\Phi}'\mathbf{V}^{-1}\mathbf{y},$$

and

$$\mathbf{V} = E\{\mathbf{z}\mathbf{z}'\}.$$

14. Let $\{Y_t\}$ satisfy

$$Y_t = \sum_{i=0}^{r} \beta_i \varphi_{ti} + \sum_{j=1}^{s} \lambda_j Y_{t-j} + e_t, \qquad t = 1, 2, \ldots,$$

where $\{e_t\}$ is a sequence of independent $(0, \sigma^2)$ random variables with $E\{e_t^4\} = \eta\sigma^4$ and the roots of $m^s - \sum_{j=1}^{s} \lambda_j m^{s-j} = 0$ are less than one in absolute value. State and prove Theorem 9.8.1 for this model.

15. Add a linear term to the wheat model of Section 9.2, and fit the model with a first order autoregressive error.

16. Let $\varphi_{t\cdot} = (1, \varphi_{t1}, \varphi_{t2}) = (1, t, t + (-1)^t)$. Does this vector satisfy the assumptions (9.1.8) and (9.1.9) with $\mathbf{A}_0$ nonsingular? Is there a linear transformation of $\varphi_{t\cdot}$ that satisfies the two assumptions with $\mathbf{A}_0$ nonsingular?

17. Let the model (9.7.1) hold with φ_{ti} that satisfy the assumptions (9.1.8) and (9.1.9). Show that the conditions of Theorem 5.7.4 are satisfied for $\{Z_t\}$ that is a stationary autoregressive moving average with iid$(0, \sigma^2)$ errors.

18. Using the estimates of Example 9.5.1, construct the vector $\mathbf{H}_m$ to estimate μ_m using $(Y_{m-3}, Y_{m-2}, \ldots, Y_{m+2}, Y_{m+3})$. What is the estimated variance of $\hat{\mu}_m$ constructed with the vector?

19. The U.S. wheat yields for 1992 and 1993 are 39.4 and 38.3. Are these yields consistent with the model of Example 9.7.2? Fit the model of Example 9.7.2 to the data for 1908 through 1993, and predict yields for 1994, 1995, and 1996. Fit the model in which (9.7.8) is replaced with

$$Y_t = \beta_0 + \beta_1 \varphi_{t2} + \beta_2 \varphi_{t3} + Z_t ,$$

where

$$\varphi_{t3} = \begin{cases} 0, & 1 \leq t \leq 70, \\ (t-70)^2, & 70 \leq t \leq 80, \\ 100 + 20(t-80), & t \geq 80. \end{cases}$$

Use the fitted model to predict yields for 1994, 1995, and 1996.

20. The U.S. wheat yields for 1992 and 1993 are given in Exercise 19 as 39.4 and 38.3, respectively. Fit the structural model of Example 9.5.1 to the data for 1908 through 1993. Predict yields for 1994, 1995, and 1996.

21. Let Z_t be a stationary time series with absolutely summable covariance function. Show that the assumptions (9.1.8) and (9.1.9) are sufficient for the limit (9.1.7) to exist. Give an example of a Z_t and φ_t such that **B** of (9.1.7) is singular.

22. Let Z_t satisfy the ARCH model of (9.3.21). Show that Z_t satisfies the assumptions of Proposition 9.3.3.

23. Consider the model

$$Z_t = \theta Z_{t-1} + a_t , \qquad |\theta| < 1 ,$$
$$a_t = (k_1 + k_2 a_{t-1}^2)^{1/2} e_t ,$$

where the e_t are NI(0, 1) random variables and $0 \leq k_2 \leq 0.3$. Let $\hat{\theta}_{\text{OLS}}$ and $(\hat{k}_{1,\text{OLS}}, \hat{k}_{2,\text{OLS}})$ denote the least squares estimators obtained by regressing Z_t on Z_{t-1} and $\hat{a}_t^2$ on $(1, \hat{a}_{t-1}^2)$, respectively, where $\hat{a}_t = Z_t - \hat{\theta}_{\text{OLS}} Z_{t-1}$. Define $\hat{h}_t = (\hat{k}_{1,\text{OLS}} + \hat{k}_{2,\text{OLS}} \hat{a}_{t-1}^2)$. Let $(\tilde{k}_{1,\text{GLS}}, \tilde{k}_{2,\text{GLS}})$ be the estimator of (k_1, k_2) obtained by regressing $\hat{h}_t^{-1} \hat{a}_t^2$ on $(\hat{h}_t^{-1}, \hat{h}_t^{-1} \hat{a}_{t-1}^2)$. Define $\tilde{h}_t = (\tilde{k}_{1,\text{GLS}} + \tilde{k}_{2,\text{GLS}} \hat{a}_{t-1}^2)$.

Let $\tilde{\theta}_{\text{GLS}}$ denote the estimated generalized least squares estimator obtained by regressing $\tilde{h}_t^{-1/2} Z_t$ on $\tilde{h}_t^{-1/2} Z_{t-1}$.

(a) Find the asymptotic distribution of

$$n^{1/2}(\hat{\theta}_{\text{OLS}} - \theta, \hat{k}_{1,\text{OLS}} - k_1, \hat{k}_{2,\text{OLS}} - k_2)' .$$

(b) Find the asymptotic distribution of

$$n^{1/2}[\tilde{\theta}_{\text{GLS}} - \theta, \tilde{k}_{1,\text{GLS}} - k_1, \tilde{k}_{2,\text{GLS}} - k_2)' .$$

(c) Consider the conditional maximum likelihood estimators $\hat{\theta}_{\text{ML}}$, $\hat{k}_{1,\text{ML}}$, and $\hat{k}_{2,\text{ML}}$ obtained by minimizing

$$L_n(\theta, k_1, k_2) = \sum_{t=2}^{n} \log h_t + \sum_{t=2}^{n} h_t^{-1}(Z_t - \theta Z_{t-1})^2 ,$$

where $h_t = k_1 + k_2(Z_{t-1} - \theta Z_{t-2})^2$. Find the asymptotic distribution of $n^{1/2}(\hat{\theta}_{ML} - \theta, \hat{k}_{1,ML} - k_1, \hat{k}_{2,ML} - k_2)'$. Compare the limiting distributions of $\hat{\theta}_{OLS}$, $\hat{\theta}_{GLS}$, and $\hat{\theta}_{ML}$.