The Annals of Statistics
1987, Vol. 15, No. 1, 99–111

ROBUST SPECTRAL REGRESSION[1]

BY ALEXANDER M. SAMAROV

Massachusetts Institute of Technology and University of Lowell

This paper addresses the problem of linear regression estimation when the disturbances follow a stationary process with its spectral density known only to be in a neighborhood of some specified spectral density, for instance, that of white noise. Rather than trying to adapt to a small unspecified autocorrelation, we follow here the robustness approach, and establish the extent of the regressors and disturbance spectra interaction which require serial correlation correction. We consider a class of generalized least-squares estimates, and find the estimator in this class which optimally robustifies the least-squares estimator against serial correlation. The estimator, when considered in the frequency domain, is of a form of weighted least squares with the most prominent frequencies of the regression spectrum being downweighted in a way similar to Huber's robust regression estimator.

1. Introduction. Consider the linear regression model

$$y_t = x_t'\beta + u_t, \tag{1}$$

where the disturbances u_t follow a second-order stationary process with $Eu_t = 0$. Let $\rho(s)$ be the correlation function, $Eu_t u_{t+s} = \sigma^2\rho(s)$, and denote the corresponding normalized spectral density $f(\lambda)$. Using methods developed in robust statistics, we address here the problem of the efficiency of linear (in the y's) estimators of β when the spectral density $f(\lambda)$ is unknown but belongs to a neighborhood of some specified spectral density $f_0(\lambda)$.

The asymptotic efficiency of the ordinary least square (OLS) estimator relative to the best linear unbiased estimator (BLUE) is known to depend upon the degree of variation of the disturbance spectral density $f(\lambda)$ at the frequencies where the spectral mass of the regressors x_t is concentrated [Grenander and Rosenblatt (1957)]. In order to write the asymptotic variances in terms of spectral distributions we will assume, following Grenander and Rosenblatt (1957), that the regressors satisfy the following asymptotic conditions:

G.1. $d_{jn}^2 = \sum_{t=1}^n x_{jt}^2 \to \infty$ as $n \to \infty$, $j = 1, \dots, p$. Denote D_n the $p \times p$ diagonal matrix with elements d_{jn}, $j = 1, \dots, p$;
G.2. $\lim_{n\to\infty} x_{jn}^2/d_{jn}^2 = 0$, $j = 1, \dots, p$;
G.3. $\lim_{n\to\infty} \sum_{t=1}^{n-s} x_{jt}x_{k,t+s}/d_{jn}d_{kn} = r_{jk}(s)$, $j, k = 1, \dots, p$; $s = 0, 1, 2, \dots$;
G.4. The matrix $R(0) = \{r_{jk}(0),\ j, k = 1, \dots, p\}$ is nonsingular.

Received February 1985; revised May 1986.
[1]This research was supported in part by National Science Foundation Grants MCS-8202688 and DMS-8408971.
AMS 1980 *subject classifications*. Primary 62J02, 62F35; secondary 62M10, 62M15.
Key words and phrases. Serial correlation, regression spectrum, efficiency robustness, minimax robustness, generalized least squares.

Conditions G.1 and G.2 mean that the amount of information contained in each regressor grows with n, but not too fast. Condition G.3 requires that asymptotically the x_t's have the second-order properties of a vector-valued stationary time series. It turns out that the asymptotic behavior of the x_t's, relevant to the problem in question, can be described in terms of the correlation sequence $R(s) = \{r_{jk}(s),\ j, k = 1, \dots, p\}$ or its Fourier transform, the regression spectral distribution $M(\lambda)$:

$$R(s) = \int_{-\pi}^{\pi} e^{i\lambda s}\, dM(\lambda), \tag{2}$$

where $M(\lambda)$ is a Hermitian $p \times p$ matrix whose increments, $M(\lambda_2) - M(\lambda_1)$, $\lambda_2 > \lambda_1$, are nonnegative definite. The asymptotic covariance matrix of the standardized OLS estimator $D_n(b_{\mathrm{OLS}} - \beta)$ can then be written as

$$V(f, b_{\mathrm{OLS}}, M) = 2\pi R^{-1}(0) \int f(\lambda)\, dM(\lambda)\, R^{-1}(0), \tag{3}$$

provided $f(\lambda)$ is continuous on $[-\pi, \pi]$. (Integration over frequencies λ will always be performed over the interval $[-\pi, \pi]$ and the integration limits here and in the rest of the paper will be dropped.)

We will also assume that $M(\lambda)$ is absolutely continuous with the spectral density $m(\lambda)$, which is a natural assumption in many areas of application, for example, in economics. [The spectral density $m(\lambda)$ is a $p \times p$ Hermitian nonnegative definite matrix. Since the components of the x's are real, the real part $\mathrm{Re}(m(\lambda))$ of $m(\lambda)$ is a symmetric nonnegative definite matrix the components of which are even functions of λ. The imaginary part $\mathrm{Im}(m(\lambda))$ is a skew-symmetric matrix whose components are odd in λ; see, e.g., Hannan (1970, pages 34–38).] In this case, the OLS estimates will be fully asymptotically efficient only if $f(\lambda)$ takes not more than p different values [Grenander and Rosenblatt (1957)]. This rather unrealistic behavior of the noise is unlikely to be satisfied in practice, and as a result, the efficiency of the OLS estimator relative to the BLUE can be very low, cf. Watson (1955), Hannan (1963, 1970), Bloomfield and Watson (1975) and Rao and Griliches (1969). In fact, the example given later in this section shows that even small departures from a flat spectrum for the noise may inflate the variance of the OLS estimator by arbitrarily large amounts.

Following the usual robustness approach we will assume that the spectral density of the noise f, although unknown, belongs to some neighborhood of the specified spectral density f_0. The most important case is when $f_0(\lambda) \equiv 1/2\pi$, when the disturbances at the "central" model are uncorrelated. More specifically, we consider the gross errors type neighborhoods [Huber (1981)], $\mathscr{U}_\varepsilon(f_0)$: we say that $f \in \mathscr{U}_\varepsilon(f_0)$ if

$$f(\lambda) = (1 - \varepsilon) f_0(\lambda) + \varepsilon p(\lambda), \tag{4}$$

where $0 < \varepsilon < 1$ is a small number and $p(\lambda)$ is an arbitrary symmetric probability density on $[-\pi, \pi]$. We will consider only normalized spectral densities f, $\int f(\lambda)\, d\lambda = 1$, so that the variance σ^2 of all time series with $f \in \mathscr{U}_\varepsilon(f_0)$ will be the same, and we will assume without loss of generality that $\sigma^2 = 1$. Neighborhoods of this type have also been considered by Hosoya (1978) and Franke (1981) for the problem of robust prediction of time series.

The neighborhood $\mathcal{U}_\varepsilon(f_0)$ corresponds to the additive contamination of the disturbances of the central model:

$$u_t = e_t + v_t, \tag{5}$$

where $Ee_t = Ev_t = 0$, $Ee_t^2 = 1 - \varepsilon$, $Ev_t^2 = \varepsilon$, $\{e_t\}$ and $\{v_t\}$ are stationary time series, independent of each other, with normalized spectral densities $f_0(\lambda)$ and $p(\lambda)$, respectively. For $f_0(\lambda) \equiv 1/2\pi$ this means that an arbitrarily correlated but weak (small ε) noise is allowed to be added to the uncorrelated errors e_t.

Other types of neighborhoods: L_2-neighborhoods or those with certain degree of smoothness of the perturbation density $p(\lambda)$ in (4) could be used instead of $\mathcal{U}_\varepsilon(f_0)$ in the context of the minimax problem considered in the next section.

To illustrate the lack of robustness of the OLS variance consider the case of one centered regressor ($p = 1$) and the neighborhood of the white-noise spectral density $\mathcal{U}_\varepsilon(1/2\pi)$. It is easy to see from (3) that

$$\sup_{f \in \mathcal{U}_\varepsilon(1/2\pi)} V(f, b_{\text{OLS}}, m) = 1 - \varepsilon + 2\pi\varepsilon \sup_{\lambda \in [-\pi, \pi]} m(\lambda), \tag{6}$$

while $V(1/2\pi, b_{\text{OLS}}, m) = 1$. So, if the regressor spectral density has sufficiently high peaks, the asymptotic variance of the OLS estimator can be made arbitrarily large by even very small departure from uncorrelated errors. If, on the other hand, $m(\lambda)$ is relatively flat, $V(f, b_{\text{OLS}}, m)$ will not be much inflated by small perturbations of f. In the extreme case, when $m(\lambda) \equiv 1/2\pi$, $\sup_f V(f, b_{\text{OLS}}, 1/2\pi) = 1$, and the variance does not change with f at all.

This simple example shows that the OLS variance may be very sensitive to departures from the uncorrelated errors. In the next section we introduce a class of generalized least-squares (GLS) estimators which can robustify OLS against these departures. Before proceeding to the next section we will briefly comment on the more common, adaptive methods of serial correlation correction.

Most of the known methods of dealing with serial correlation are adaptive in the sense that they first use the OLS residuals to estimate, parametrically or nonparametrically, the unknown autocorrelation, or spectrum, of the disturbances, and then use these estimates to obtain the improved (and, hopefully, asymptotically efficient) regression estimates. The parametric adaptation methods, such as the most often used fixed order autoregression models, may sometimes do more harm than good when compared with OLS, if the specification of the disturbance process is not quite correct [see, for example, Engle (1974) and Newbold and Davies (1978)]. The nonparametric spectrum estimation approach, proposed by Hannan (1963, 1970), leads to the asymptotically efficient estimates but, as most nonparametric methods do, requires a large sample size.

Instead of trying to adapt to a small unspecified autocorrelation, we follow here the robustness approach, and establish the extent of the regressors and disturbance spectra interaction which require serial correlation correction. We also propose the frequency domain weighted least-squares estimators which optimally robustify the OLS estimators against that autocorrelation.

Admittedly, the approach we propose may require nonparametric estimation of $m(\lambda)$. However, unlike the nonparametric adaptive estimation, we can estimate the regression spectrum directly from observations and not from regression residuals, and $m(\lambda)$ has to be estimated only at the "high energy" frequencies which is usually easier. Also, the spectrum $m(\lambda)$ may be known either completely or up to a finite-dimensional parameter.

2. Results. We consider a class of GLS estimators defined with a continuous, positive function $g(\lambda)$, even on $[-\pi, \pi]$, as follows:

$$b_g = (X'G^{-1}X)^{-1}X'G^{-1}Y, \tag{7}$$

where X is the $n \times p$ matrix of regressors, Y is the n vector of response y_t and the $n \times n$ Toeplitz matrix $G = \{g_{k-j}, k, j = 1, \dots, n\}$ corresponds to $g(\lambda)$:

$$g_t = \int g(\lambda) \cos t\lambda \, d\lambda, \qquad t = 0, \pm 1, \pm 2, \dots . \tag{8}$$

Notice that the estimator b_g does not change when $g(\lambda)$ is multiplied by a nonzero constant.

The class (7) is quite rich: b_g is the OLS estimator when $g(\lambda) \equiv 1/2\pi$, and the BLUE when $g(\lambda) = f(\lambda)$. Most of the linear estimates with one form or another of serial correlation correction can be represented in the form (7) with either G or g estimated from the data. Rozanov and Kozlov (1969) [see also Ibragimov and Rozanov (1978)] studied the estimates (7), writing them in the spectral form, and obtained the asymptotic covariance matrix of b_g under the assumptions that the regressors satisfy Grenander's conditions G.1–G.4, and that f and g be positive, continuous and even on $[-\pi, \pi]$. Assuming also, as we do, that the regression spectrum is absolutely continuous their result can be written as follows:

$$\begin{aligned} &\lim_{n\to\infty} D_n \mathrm{Cov}(b_g) D_n \\ &\quad = 2\pi \left(\int \frac{m(\lambda)}{g(\lambda)} \, d\lambda \right)^{-1} \int \frac{f(\lambda) m(\lambda) \, d\lambda}{g^2(\lambda)} \left(\int \frac{m(\lambda)}{g(\lambda)} \, d\lambda \right)^{-1} \\ &\quad = V(f, g, m), \quad \text{say.} \end{aligned} \tag{9}$$

[Notice that since f and g are even we can replace $m(\lambda)$ in (9) by its real part, which is nonnegative and even in λ. We therefore assume from now on that the matrix $m(\lambda)$ is replaced by its real part without changing the notation.]

Here D_n is the $p \times p$ diagonal matrix defined in the condition G.1, and regularity of the matrix $\int [m(\lambda)/g(\lambda)] \, d\lambda$ follows from the condition G.4 and the positiveness of $g(\lambda)$. Of course, for $g(\lambda) \equiv 1/2\pi$ and $g(\lambda) = f(\lambda)$, (9) gives the well-known expressions for the asymptotic covariance matrices of OLS [see (3)] and BLUE, respectively, cf. Grenander (1981). It is not difficult to show, as in Ibragimov and Rozanov (1978), that for any g

$$V(f, g, m) \geq V(f, f, m) = 2\pi \left(\int \frac{m(\lambda)}{f(\lambda)} \, d\lambda \right)^{-1}, \tag{10}$$

in the sense that $V(f, g, m) - V(f, f, m)$ is a nonnegative definite matrix.

When the regression spectral density $m(\lambda)$ is assumed to be continuous, as is the case in this work, the continuity of f and g assumed in (9) could be somewhat relaxed. This will not, however, be pursued here, rather we will consider neighborhoods of the form

$$\mathscr{P}_\varepsilon(f_0) = \mathscr{U}_\varepsilon(f_0) \cap C, \tag{11}$$

with $f_0 \in C^+$, where C is the class of continuous functions on $[-\pi, \pi]$ and C^+ the class of positive functions in C.

As a measure of efficiency we will use a natural scalar summary of the asymptotic covariance matrix $V(f, g, m)$, the asymptotic mean square error (MSE), possibly weighted with a nonnegative definite matrix $\mathscr{N}$:

$$\operatorname{tr}(\mathscr{N}V(f, g, m)),$$

where $\operatorname{tr}(\cdot)$ denotes the trace of a matrix. The matrix $\mathscr{N}$ provides a certain degree of flexibility in designing the minimax estimates, allowing us, for example, to minimize only the variance of certain coefficients or particular linear combinations.

Our goal is to find the estimator in the class defined by (7), or equivalently, the function $g \in C^+$, which has the highest efficiency uniformly over $f \in \mathscr{P}_\varepsilon(f_0)$, i.e., to find $g^* \in C^+$ such that

$$\sup_{f \in \mathscr{P}_\varepsilon(f_0)} \operatorname{tr}(\mathscr{N}V(f, g^*, m)) = \inf_{g \in C^+} \sup_{f \in \mathscr{P}_\varepsilon(f_0)} \operatorname{tr}(\mathscr{N}V(f, g, m)). \tag{12}$$

REMARK. Note that for $f \in \mathscr{P}_\varepsilon(f_0)$ we have from (4) and (9):

$$V(f, g, m) = V(f_0, g, m) + \varepsilon(V(p, g, m) - V(f_0, g, m)), \tag{13}$$

where the matrix difference in the last term is, in fact, the change-of-variance of the estimator b_g with respect to spectral or autocorrelation perturbations [cf. Ronchetti and Rousseeuw (1985)]. We have therefore

$$\begin{aligned} &\sup_{f \in \mathscr{P}_\varepsilon(f_0)} \operatorname{tr}(\mathscr{N}V(f, g, m)) \\ &\quad = \operatorname{tr}(\mathscr{N}V(f_0, g, m)) + \varepsilon \sup_p \operatorname{tr}(\mathscr{N}(V(p, g, m) - V(f_0, g, m))), \end{aligned} \tag{14}$$

where $p \in C^+$, $\int p(\lambda)\, d\lambda = 1$, and we may call the upper bound in the last term in (14) the spectral (or serial correlation) change-of-variance sensitivity (SCVS). The SCVS represents the largest possible variance increase resulted from the misspecification of the disturbance correlation structure. As we saw in Section 1 the SCVS of OLS may be arbitrarily large if $m(\lambda)$ has high enough peaks.

Following the bounded-influence regression approach, see, e.g., Krasker and Welsch (1982), we may set an upper bound to the SCVS, and, subject to that bound, try to minimize the MSE at the central model f_0, i.e., the first term in (14). This approach, together with the optimal choice of the bound, is taken in Theorem 1 below.

In order to formulate the solution of our minimax problem we have to consider the following matrix equation:

$$D = \int m(\lambda) \min\left(\frac{1}{f_0(\lambda)}, \frac{c}{\operatorname{tr}^{1/2}(\mathscr{N}D^{-1}m(\lambda)D^{-1})} \right) d\lambda, \tag{15}$$

where D is a $p \times p$ matrix and $c > 0$. Equation (15) always has the trivial solution $D = 0$ for all $c > 0$, and the positive definite solution

$$D_0 = \int \frac{m(\lambda)}{f_0(\lambda)} \, d\lambda, \tag{16}$$

for

$$c \geq c_0 = \max_{-\pi < \lambda < \pi} \left(\operatorname{tr}^{1/2}(\mathscr{N} D_0^{-1} m(\lambda) D_0^{-1}) / f_0(\lambda) \right), \tag{17}$$

where D_0 is invertible because of condition G.4 and $f_0 \in C^+$. Denote $c_{\min} > 0$ the smallest value of c for which (15) has a positive definite solution. Using a fixed point theorem argument, as in Proposition 1 of Krasker (1980), we can show that $c_{\min} < c_0$, i.e., a positive definite solution of (15) exists for some $c \nless c_0$. Clearly, any such solution $D(c)$ for $c < c_0$ satisfies the inequality

$$D(c) \leq D_0, \tag{18}$$

i.e., $D_0 - D(c)$ is nonnegative definite. We show in the Appendix that

$$c_{\min} \geq \frac{\operatorname{tr} \mathscr{N}}{\int \operatorname{tr}^{1/2}(\mathscr{N} m(\lambda)) \, d\lambda}, \tag{19}$$

and for $p = 1$ the equality in (19) is attained.

THEOREM 1. *Let conditions* G.1–G.4 *be satisfied and the regression spectral density* $m(\lambda)$ *be a continuous* (*elementwise*) *matrix. Then*

$$\begin{aligned} &\inf_{g \in C^+} \sup_{f \in \mathscr{P}_\varepsilon(f_0)} \operatorname{tr}(\mathscr{N} V(f, g, m)) \\ &\quad = \min_{c \geq c_{\min}} \sup_{f \in \mathscr{P}_\varepsilon(f_0)} \operatorname{tr}(\mathscr{N} V(f, g_c^*, m)) \\ &\quad = 2\pi \min_{c \geq c_{\min}} \left[\varepsilon c^2 + (1 - \varepsilon) \int \min\left\{ c^2, \frac{1}{f_0^2(\lambda)} \right.\right. \\ &\qquad\qquad \left.\left. \times \operatorname{tr}(\mathscr{N} D^{-1} m(\lambda) D^{-1}) \right\} f_0(\lambda) \, d\lambda \right] \\ &\quad = \min_{c \geq c_{\min}} J(c), \quad \textit{say}, \end{aligned} \tag{20}$$

where

$$g_c^*(\lambda) = a \max\left(f_0(\lambda), \frac{\operatorname{tr}^{1/2}(\mathscr{N} D^{-1} m(\lambda) D^{-1})}{c} \right), \qquad a > 0, \tag{21}$$

and $D = D(c)$ *is the solution of* (15).

The proof of Theorem 1 is given in the Appendix.

Several comments on this result are in order. We note first that the optimal function $g_c^*(\lambda)$ in (21) coincides with $f_0(\lambda)$ everywhere except at the frequencies where the regression spectral density $m(\lambda)$ has high enough "peaks." At those frequencies it follows $m(\lambda)$. Notice also that if $c \geq c_0$, then $g_c^*(\lambda) = af_0(\lambda)$ and

$$\min_{c \geq c_0} J(c) = J(c_0) = 2\pi\left(\varepsilon c_0^2 + (1-\varepsilon)\mathrm{tr}\left(\mathscr{N} D_0^{-1}\right)\right), \tag{22}$$

which means that we may restrict minimization over c in (20) to the interval $[c_{\min}, c_0]$.

In the most important case of uncorrelated errors at the central model, i.e., $f_0(\lambda) \equiv 1/2\pi$, the right-hand side of (22),

$$\varepsilon \max_{-\pi \leq \lambda \leq \pi} \mathrm{tr}\left(\mathscr{N} R^{-1}(0) m(\lambda) R^{-1}(0)\right) + (1-\varepsilon)\mathrm{tr}\left(\mathscr{N} R^{-1}(0)\right), \tag{23}$$

is the maximum MSE of the OLS estimator when $f \in \mathscr{P}_\varepsilon(1/2\pi)$. Expression (23) shows that the maximum increase in MSE due to small departures from uncorrelated errors depends on the size of "peaks" of $m(\lambda)$, i.e., of the maximum of the trace in (23). If $m(\lambda)$ is relatively "flat," i.e., the regressors do not have very prominent frequencies, the increase in MSE will not be significant, and the OLS estimator may be safely used. If, on the other hand, the power of regressors is highly concentrated near certain frequencies, the increase in MSE may be very large.

The function g_c^*, suggested in Theorem 1, defines the GLS estimator which will keep the maximum increase in MSE, i.e., the SCVS, under the bound c while minimizing the MSE at the central model, and the theorem also offers the method for selecting the best c. In the next section we give an example illustrating the choice of c as a function of ε, and showing that the gain in efficiency of our minimax estimator over OLS may be significant.

The next theorem gives the complete solution of our minimax problem and specifies the least favorable spectral density in $\mathscr{P}_\varepsilon(f_0)$.

THEOREM 2. *Let the conditions of Theorem* 1 *be satisfied. Then for sufficiently small* $\varepsilon > 0$ *the game with the loss function,*

$$Q(f, g) = \mathrm{tr}(\mathscr{N} V(f, g, m)),$$

has a saddle point (f^*, af^*) *with any* $a > 0$,

$$\inf_{g \in C^+} \sup_{f \in \mathscr{P}_\varepsilon(f_0)} Q(f, g) = \sup_{f \in \mathscr{P}_\varepsilon(f_0)} \inf_{g \in C^+} Q(f, g) = Q(f^*, af^*),$$

where

$$f^*(\lambda) = (1-\varepsilon)\max\left(f_0(\lambda), \frac{\mathrm{tr}^{1/2}\left(\mathscr{N} D^{-1} m(\lambda) D^{-1}\right)}{c^*}\right), \tag{24}$$

$D = D(c^*)$ *is the solution of* (15), *and* c^* *is such that*

$$\int f^*(\lambda)\, d\lambda = 1. \tag{25}$$

The value of the game is

$$Q(f^*, af^*) = 2\pi(1-\varepsilon)\mathrm{tr}(\mathscr{N} D^{-1}). \tag{26}$$

The proof is given in the Appendix.

It turns out that it is easier to interpret and compute the minimax estimator b_{g^*} suggested by Theorem 1 if we Fourier-transform the data and consider the frequency domain approximation of b_{g^*}.

Applying the discrete Fourier transform (DFT) to the matrix of regressors X and the vector of responses Y we have

$$\dot{X} = \Omega X, \qquad \dot{Y} = \Omega Y,$$

where Ω is the $n \times n$ unitary matrix with (k, j)th element given by $1/\sqrt{n}\exp\{i2\pi kj/n\}$, $k, j = 0, 1, \dots, n-1$. The estimator b_g given in (7) can then be written as

$$b_g = (\dot{X}^*\Omega G^{-1}\Omega^*\dot{X})^{-1}\dot{X}^*\Omega G^{-1}\Omega^*\dot{Y}, \tag{27}$$

where A^* is the transpose of the complex conjugate of A. Now, it is well known [see Grenander and Szegö (1958)] that for large n the matrix $\Omega G^{-1}\Omega^*$ is nearly diagonal with the elements $1/g(2\pi k/n)$, $k = 0, 1, \dots, n-1$ [$g(\lambda)$ is assumed to be periodically extended outside $[-\pi, \pi]$]. More precisely, it can be shown using, for example, results from Grenander (1981) and Davies (1973) that

$$D_n\left(b_g - \tilde{b}_g\right) \to 0 \quad \text{in probability as } n \to \infty,$$

where

$$\tilde{b}_g = \left(\dot{X}^*\operatorname{diag}\left(\frac{1}{g\left(\frac{2\pi k}{n}\right)}\right)\dot{X}\right)^{-1}\dot{X}^*\operatorname{diag}\left(\frac{1}{g\left(\frac{2\pi k}{n}\right)}\right)\dot{Y} \tag{28}$$

and D_n is defined in G.1. The estimator $\tilde{b}_g$ has, therefore, the same asymptotic covariance matrix as b_g, but is easier to compute from a given function $g(\lambda)$. The estimator $\tilde{b}_g$ is of the weighted least-squares form with the weight $w_k = 1/[g(2\pi k/n)]$ being applied at the frequency $\lambda_k = 2\pi k/n$. In particular, for the optimal function $g^*(\lambda)$ and $f_0(\lambda) \equiv 1/2\pi$, we have the optimal weights

$$w_k^* = \min\left(1, \frac{c}{2\pi \operatorname{tr}^{1/2}\left(\mathcal{N}D^{-1}m\left(\frac{2\pi k}{n}\right)D^{-1}\right)}\right), \tag{29}$$

where we have chosen the factor a in (21) so that the maximum weight is 1. We can clearly see now that the optimal estimator downweights the frequencies where the regression spectral density $m(\lambda)$ has "high peaks," i.e., where the trace in (29) is large.

The computation of the estimator $\tilde{b}_g$ involves the estimation of the regression spectral density $m(\lambda)$ from the x-data and the iterative solution of the matrix equation (15).

3. The case of $p = 1$ and an example. In the case of one centered regressor,

$$y_t = \beta x_t + u_t, \tag{30}$$

the interpretation and computation of the minimax estimator b_{g^*} considerably simplifies. Set $p = 1$, $f_0(\lambda) \equiv 1/2\pi$, $\mathscr{N} = 1$ and denote $\mu = cD/2\pi$. Then (15) can be written as

$$\mu = c\int m(\lambda)\min\left(1, \frac{\mu}{\sqrt{m(\lambda)}}\right) d\lambda. \tag{31}$$

Solving (31) for c and plugging it in (20) we have

$$\inf_{g \in c^+} \sup_{f \in \mathscr{P}_\varepsilon(1/2\pi)} V(f, g, m)$$

$$= \min_{\mu_{\min} \leq \mu \leq \mu_{\max}} \frac{\mu^2\left[2\pi\varepsilon + (1-\varepsilon)\int m(\lambda)\min\left(1, \frac{m(\lambda)}{\mu^2}\right) d\lambda\right]}{\left(\int m(\lambda)\min\left(1, \frac{\mu}{\sqrt{m(\lambda)}}\right) d\lambda\right)^2}, \tag{32}$$

where $\mu_{\min} = \min_\lambda \sqrt{m(\lambda)}$ and $\mu_{\max} = \max_\lambda \sqrt{m(\lambda)}$, as follows from (17), (19) and the comment after Theorem 1. It follows now from Theorem 2 that the minimum in (32) occurs at $\mu = \mu^*$ such that [cf. (25)]

$$\operatorname{mes}\{A^c\} + \frac{1}{\mu^*}\int_A \sqrt{m(\lambda)}\, d\lambda = \frac{2\pi}{1-\varepsilon}$$

(here $A = \{\lambda \colon \sqrt{m(\lambda)} \leq \mu^*\}$ and $\operatorname{mes}\{\cdot\}$ is the Lebesgue measure on $[-\pi, \pi]$) and is equal to

$$Q(f^*, af^*) = V(f^*, af^*, m) = \frac{1-\varepsilon}{\int m(\lambda)\min\left(1, \frac{\mu^*}{\sqrt{m(\lambda)}}\right) d\lambda},$$

with

$$f^*(\lambda) = \frac{(1-\varepsilon)}{2\pi}\max\left(1, \frac{\sqrt{m(\lambda)}}{\mu^*}\right).$$

To give a numerical example we consider (30) with an autoregressive carrier $x_t = \rho x_{t-1} + e_t$, where $\{e_t\}$ are i.i.d., independent of $\{u_t\}$, with the first two moments 0 and 1, respectively, and $|\rho| < 1$. The normalized spectral density of x_t is

$$m(\lambda) = \frac{1-\rho^2}{2\pi(1+\rho^2 - 2\rho\cos\lambda)}.$$

We have now, using (6), for the OLS estimator

$$\sup_{f \in \mathscr{P}_\varepsilon(1/2\pi)} V(f, b_{\mathrm{OLS}}, m) = 1 - \varepsilon + \frac{1+\rho}{1-\rho}\varepsilon.$$

The following table gives the maximum variances of the OLS and minimax estimators for several values of ε when $\rho = 0.9$, together with the values of μ^*:

ε	0	0.01	0.05	0.1	0.15	0.2
OLS	1	1.18	1.9	2.8	3.7	4.6
minimax	1	1.125	1.425	1.67	1.86	2
μ^*	1.73	1.2	0.7	0.5	0.4	0.3

Notice that for $\varepsilon = 0$ the minimax estimator coincides with OLS while for $\varepsilon = 0.15$ its maximum variance is half that of OLS.

APPENDIX

PROOF OF (19). Left multiply (15) by $\mathscr{N}D^{-1}$ and take the trace of both sides:

$$\text{(A.1)}\quad \operatorname{tr}\mathscr{N} = \int \operatorname{tr}(\mathscr{N}D^{-1}m(\lambda))\min\left(\frac{1}{f_0(\lambda)}, \frac{c}{\operatorname{tr}^{1/2}(\mathscr{N}D^{-1}m(\lambda)D^{-1})}\right)d\lambda.$$

Denoting by $\mathscr{N}^{1/2}$ and $m^{1/2}(\lambda)$ the symmetric, nonnegative definite square roots of the matrices $\mathscr{N}$ and $m(\lambda)$, we have

$$\text{(A.2)}\quad \begin{aligned} |\operatorname{tr}(\mathscr{N}D^{-1}m(\lambda))| &= |\operatorname{tr}(\mathscr{N}^{1/2}D^{-1}m^{1/2}(\lambda)m^{1/2}(\lambda)\mathscr{N}^{1/2})| \\ &\le \operatorname{tr}^{1/2}(m^{1/2}(\lambda)D^{-1}\mathscr{N}^{1/2}\mathscr{N}^{1/2}D^{-1}m^{1/2}(\lambda)) \\ &\quad\times \operatorname{tr}^{1/2}(\mathscr{N}^{1/2}m^{1/2}(\lambda)m^{1/2}(\lambda)\mathscr{N}^{1/2}) \\ &= \operatorname{tr}^{1/2}(\mathscr{N}D^{-1}m(\lambda)D^{-1})\operatorname{tr}^{1/2}(\mathscr{N}m(\lambda)), \end{aligned}$$

where we used the inequality $|\operatorname{tr}(AB)| \le \operatorname{tr}^{1/2}(A'A)\operatorname{tr}^{1/2}(B'B)$ which is a form of the Cauchy–Schwarz inequality. Now (19) follows immediately from (A.1) and (A.2). □

When $p = 1$ (15) becomes

$$\text{(A.3)}\quad D = \int m(\lambda)\min\left(\frac{1}{f_0(\lambda)}, \frac{cD}{\sqrt{\mathscr{N}m(\lambda)}}\right)d\lambda,$$

where all quantities are now scalar. Consider the function

$$F(D, c) = D - \int m(\lambda)\min\left(\frac{1}{f_0(\lambda)}, \frac{cD}{\sqrt{m(\lambda)}}\right)d\lambda,$$

for $c \ge c_{\min} = \sqrt{\mathscr{N}}/\int\sqrt{m(\lambda)}\,d\lambda$. For $c \ge c_{\min} F(0, c) = 0$, $dF(D, c)/dD = 1 - c\int m(\lambda)\,d\lambda/\mathscr{N} \le 0$, for small enough $D > 0$, and $F(D_0, c) \ge 0$, where D_0 is defined in (16). Therefore, by the mean value theorem there exists $D = D(c)$ in the interval $(0, D_0]$ such that $F(D(c), c) = 0$.

PROOF OF THEOREM 1. Denoting

$$k(\lambda) = \frac{1}{g^2(\lambda)}\operatorname{tr}\left(\mathcal{N}\left(\int \frac{m(\lambda)}{g(\lambda)}\,d\lambda\right)^{-1} m(\lambda)\left(\int \frac{m(\lambda)}{g(\lambda)}\right)^{-1}\right),$$

we can write

$$\operatorname{tr}(\mathcal{N}V(p,g,m)) = 2\pi\int p(\lambda)k(\lambda)\,d\lambda.$$

We now use the fact that

$$\text{(A.4)} \qquad \sup_{\substack{p\in C^+\\ \int p(\lambda)\,d\lambda=1}} \int p(\lambda)k(\lambda)\,d\lambda = \sup_{-\pi\le\lambda\le\pi} k(\lambda),$$

for $k \in C$, to rewrite (14) as follows:

$$\text{(A.5)} \qquad \begin{aligned}&\sup_{f\in\mathcal{P}_\varepsilon(f_0)} \operatorname{tr}(\mathcal{N}V(f,g,m))\\ &\quad= (1-\varepsilon)\operatorname{tr}(\mathcal{N}V(f_0,g,m)) + 2\pi\varepsilon \sup_{-\pi\le\lambda\le\pi} k(\lambda).\end{aligned}$$

Denoting now $\mathcal{B}_c = \{g \in C^+,\ \sup_{-\pi\le\lambda\le\pi} k(\lambda) \le c^2\}$ for some $c > 0$ we can, using (A.5), reduce the original minimax problem to a constrained optimization problem:

$$\text{(A.6)} \qquad \begin{aligned}&\inf_{g\in C^+}\sup_{f\in\mathcal{P}_\varepsilon(f_0)} \operatorname{tr}(\mathcal{N}V(f,g,m))\\ &\quad= \inf_c \inf_{g\in\mathcal{B}_c} \sup_{f\in\mathcal{P}_\varepsilon(f_0)} \operatorname{tr}(\mathcal{N}V(f,g,m))\\ &\quad= \inf_c\left[(1-\varepsilon)\inf_{g\in\mathcal{B}_c} \operatorname{tr}(\mathcal{N}V(f_0,g,m)) + 2\pi\varepsilon c^2\right].\end{aligned}$$

Now to solve the minimization problem in (A.6) we consider the following problem:

Problem A. Minimize

$$\text{(A.7)} \qquad \int \operatorname{tr}(\mathcal{N}h'(\lambda)h(\lambda))f_0(\lambda)\,d\lambda$$

over $p \times p$ matrix-valued functions $h(\lambda)$, subject to

$$\text{(A.8)} \qquad \sup_{-\pi\le\lambda\le\pi} \operatorname{tr}(\mathcal{N}h'(\lambda)h(\lambda)) \le c^2,$$

and

$$\text{(A.9)} \qquad \int B(\lambda)h(\lambda)\,d\lambda = I,$$

where I is the $p \times p$ identity matrix and the $p \times p$ matrix $B(\lambda) = m^{1/2}(\lambda)$ is the symmetric, nonnegative definite square root of $m(\lambda)$.

It is easy to check that minimizing (A.7) subject to (A.9) is equivalent to minimizing

$$\text{(A.10)}\qquad \int \operatorname{tr}\left[\left(h(\lambda) - \frac{B(\lambda)P}{f_0(\lambda)}\right)\mathscr{N}\left(h(\lambda) - \frac{B(\lambda)P}{f_0(\lambda)}\right)'\right] f_0(\lambda)\, d\lambda,$$

with arbitrary $p \times p$ matrix P. Now the minimum of (A.10) subject to (A.8) is achieved [cf. Lemma 2 in Samarov (1985)] at

$$\text{(A.11)}\qquad h^*(\lambda) = B(\lambda)P\min\left(\frac{1}{f_0(\lambda)}, \frac{c}{\operatorname{tr}^{1/2}(\mathscr{N}P'm(\lambda)P)}\right).$$

For $h^*(\lambda)$ to be a solution of Problem A it should satisfy (A.9), which means that the matrix P should satisfy the equation

$$\text{(A.12)}\qquad P^{-1} = \int m(\lambda)\min\left(\frac{1}{f_0(\lambda)}, \frac{c}{\operatorname{tr}^{1/2}(\mathscr{N}P'm(\lambda)P)}\right) d\lambda,$$

which coincides with equation (15) with $P = D^{-1}$.

Denoting now

$$\text{(A.13)}\qquad h_g(\lambda) = \frac{B(\lambda)}{g(\lambda)}\left(\int \frac{m(\lambda)}{g(\lambda)}\, d\lambda\right)^{-1},$$

we have

$$\operatorname{tr}(\mathscr{N}V(f_0, g, m)) = 2\pi \int \operatorname{tr}\left(\mathscr{N}h_g^T(\lambda)h_g(\lambda)\right) f_0(\lambda)\, d\lambda,$$

and $h_g(\lambda)$ satisfies the constraints (A.9) and (A.8) if $g \in \mathscr{B}_c$ where $\mathscr{B}_c$ is defined in (A.6). It is now easy to check that if

$$\text{(A.14)}\quad g(\lambda) = g_c^*(\lambda) = a\max\left(f_0(\lambda), \frac{\operatorname{tr}^{1/2}(\mathscr{N}D^{-1}m(\lambda)D^{-1})}{c}\right), \qquad a > 0,$$

and $D^{-1} = P$, the $h_g(\lambda)$ from (A.13) coincides with the optimal $h^*(\lambda)$ given by (A.11) and (A.12). We have, therefore, that $g_c^*(\lambda)$ solves the minimization problem in (A.6). The claims of Theorem 1 are obtained by inserting $g_c^*(\lambda)$ in (A.6). □

Proof of Theorem 2. It is sufficient to show that

$$\text{(A.15)}\qquad \begin{aligned} \inf_{g \in C^+} \sup_{f \in \mathscr{P}_\varepsilon(f_0)} Q(f, g) &\le Q(f^*, af^*) = 2\pi \operatorname{tr}\left(\mathscr{N}\left(\int \frac{m(\lambda)}{f^*(\lambda)}\, d\lambda\right)^{-1}\right) \\ &\le \sup_{f \in \mathscr{P}_\varepsilon(f_0)} \inf_{g \in C^+} Q(f, g). \end{aligned}$$

Observe first that from (10)

$$\sup_{f \in \mathscr{P}_\varepsilon(f_0)} \inf_{g \in C^+} Q(f, g) = \sup_{f \in \mathscr{P}_\varepsilon(f_0)} Q(f, af) = 2\pi \sup_{f \in \mathscr{U}_\varepsilon(f_0)} \operatorname{tr}\left(\mathscr{N}\left(\int \frac{m(\lambda)}{f(\lambda)}\, d\lambda\right)^{-1}\right),$$

and the second inequality in (A.15) follows since $f^* \in \mathscr{P}_\varepsilon(f_0)$.

Choose now in Theorem 1 $a = 1 - \varepsilon$ and $c = c^*$ such that

$$\int g^*_{c^*}(\lambda)\, d\lambda = 1. \tag{A.16}$$

Since g^*_c is continuous in c, such c^* can be chosen close enough to c_0 for sufficiently small $\varepsilon > 0$, so that $c^* \geq c_{\min}$. For this choice of a and c we have $g^* = f^* \in \mathscr{P}_\varepsilon(f_0)$ where f^* is defined by (24). It is now straightforward to check using (A.16) and (15) that $\mathscr{J}(c^*) = Q(f^*, af^*)$, where $\mathscr{J}(c)$ is defined in (20), and the first inequality in (A.15) follows. □

Acknowledgments. The author is grateful to E. Kuh and W. Oldford for helpful discussions.

REFERENCES

BLOOMFIELD, P. and WATSON, G. (1975). The inefficiency of least squares. *Biometrika* **62** 121–128.

DAVIES, R. (1973). Asymptotic inference in stationary Gaussian time series. *Adv. in Appl. Probab.* **5** 469–497.

ENGLE, R. (1974). Specification of the disturbance for efficient estimation. *Econometrica* **42** 135–147.

FRANKE, J. (1981). Minimax-robust prediction of discrete time series. Preprint.

GRENANDER, U. (1981). *Abstract Inference*. Wiley, New York.

GRENANDER, U. and ROSENBLATT, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.

GRENANDER, U. and SZEGÖ, G. (1958). *Toeplitz Forms and Their Applications*. Univ. California Press, Berkeley.

HANNAN, E. J. (1963). Regression for time series. In *Proc. of the Symposium On Time Series Analysis* (M. Rosenblatt, ed.) 17–37. Wiley, New York.

HANNAN, E. J. (1970). *Multiple Time Series*. Wiley, New York.

HOSOYA, Y. (1978). Robust linear extrapolations of second-order stationary processes. *Ann. Probab.* **6** 574–584.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

IBRAGIMOV, I. and ROZANOV, Y. (1978). *Gaussian Random Processes*. Springer, New York (originally published in Russian in 1970 by Nauka, Moscow).

KRASKER, W. S. (1980). Estimation in linear regression models with disparate data points. *Econometrica* **48** 1333–1346.

KRASKER, W. and WELSCH, R. (1982). Efficient bounded-influence regression estimation. *J. Amer. Statist. Assoc.* **77** 595–604.

NEWBOLD, P. and DAVIES, N. (1978). Error misspecification and spurious regressions. *Internat. Econom. Rev.* **19** 513–519.

RAO, P. and GRILICHES, Z. (1969). Small-sample properties of several two-stage regression methods in the context of auto-correlated errors. *J. Amer. Statist. Assoc.* **64** 253–272.

RONCHETTI, E. and ROUSSEEUW, P. (1985). Change-of-variance sensitivities in regression analysis. *Z. Wahrsch. verw. Gebir.* **68** 503–519.

ROZANOV, Y. and KOZLOV, M. (1969). On asymptotically efficient estimation of regression coefficients. *Soviet Math. Dokl.* **10** 1075–1078.

SAMAROV, A. M. (1985). Bounded-influence regression via local minimax mean squared error. *J. Amer. Statist. Assoc.* **80** 1032–1040.

WATSON, G. (1955). Serial correlation in regression analysis I. *Biometrika* **42** 327–341.

CENTER FOR COMPUTATIONAL RESEARCH
IN ECONOMICS AND MANAGEMENT SCIENCE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
1 AMHERST STREET
CAMBRIDGE, MASSACHUSETTS 02139