PREDICTION FOR NON-STATIONARY STOCHASTIC PROCESSES

Michael D. Godfrey

Econometric Research Program
Research Paper No. 12
December 14, 1965

The research described in this paper was supported by National Science Foundation Grant NSF-GS 551. The computer facilities used are supported by National Science Foundation Grant NSF-GP 579.

Princeton University
Econometric Research Program
92-A Nassau Street
Princeton, New Jersey

ABSTRACT

This paper presents a procedure for the prediction of time series for which the generating model is non-stationary. The procedure requires that the non-stationary change in the process be fairly smooth over time in order that this change may be estimated and used in the prediction. Since we characterize the process by its - time changing - spectrum, we are only using information which is approximately linear.

The procedure is applied to a time series of economic data and found to give results which encourage us to consider further developments of the method.

PREDICTION FOR NON-STATIONARY STOCHASTIC PROCESSES

I. Introduction

While prediction theory and practice for stationary processes is will developed, the practical applications of this development have not been uniformly successful. Frequently the lack of success of predictors based on Weiner theory is ascribed to non-stationarity of the generating process.

In this paper we will outline a practical procedure for forming predictions which requires only "local stability" of the generating process. In addition, this procedure will permit us to conveniently treat processes with almost periodic variations, such as certain seasonal patterns.

II. Prediction Model

For the purposes of prediction we will suppose that the observed time series for which future values are to be predicted is generated by a process with a representation:

$$x_t = \int_{-\pi}^{\pi} e^{i\omega t} K(\omega, t) d\omega \qquad 2.1$$

where $K(\omega, t)$ = a random function of ω and t with independent increments on ω.

In order to describe the process, we wish to form an estimate of

$$F(\omega, t) = E(K(\omega, t)\, K^*(\omega, t)) \qquad 2.2$$

In the case of a stationary process where $F(\omega, t) = F(\omega)$, we form an estimate of the spectral density:

$$f(\omega) = \frac{d\, F(\omega)}{d\omega} \qquad 2.3$$

and derive a prediction scheme from the estimated spectrum. Since $\hat{f}(\omega)$ is not a function of time, there is no problem in using it in order to make inferences about future values of the realization. However, in the non-stationary case considered here, the spectral density $f(\omega, t)$ is changing over any realization which we may have. Therefore, it seems unreasonable to estimate the spectral density by taking a simple average down the realization since this results in an average which treats each observation equally. In particular, we may consider the case of a linear trend in the frequency components of the spectral density function. Then the spectral density may be represented by:

$$f(\omega, t) = g(\omega) + A(\omega)t \qquad 2.4$$

By averaging down a realization of n observations from $t = 1, \ldots, n$, we will be estimating the density function

$$f(\omega) = g(\omega) + A(\omega)\, \frac{n + 1}{2} \qquad 2.5$$

However, if we are interested in prediction of x_t at some point in time $n + \tau$, then we wojld like to have an estimate of the spectral density of the process up to the time $n + \tau$. In the simple case of a linear trend in the spectrum, an estimate based on equation 2.5 will be closest to the true changing spectrum at $t = \frac{n+1}{2}$ and farthest at $t = 1$ and $t = n$. This is clearly an undesirable property for prediction.

The traditional procedure for estimation of spectra, and subsequent computation of optimal predicting filters has involved estimation of serial correlation coefficients. Some assumption of stationarity over the entire length of the realization is required in order to make inferences from these estimates. However, an alternative procedure for estimation of spectra is to apply digital filtering.

In simplest form we may think of estimating the spectrum of a process by first band pass filtering the realization at a set of frequencies (presumably equispaced) on the range 0 to π, and then computing the variance of each filtered series. We have been able to perform the frequency decomposition without requiring stationarity. Thus, by following this kind of procedure we may consider possible schemes for handling non-stationarity at each frequency.

III. The Prediction Algorithm

If we think in terms of estimating the spectrum of a process by digital filtering it is natural in this context to consider the possibility that the variance of each digital filtered series is not a constant over the realization. In particular, we may consider the variance of each series to be a stochastic variable with a representation:

$$\sigma_x^2(\omega, t) = K_x(\omega) + a_x(\omega)t + \sum_{s=1}^{\ell} a_{x,s}(\omega)\, \sigma_x^2(\omega, t-s) \qquad 3.1$$

where $K_x(\omega)$, $a_x(\omega)$, $a_{x,s}(\omega)$ = constants.

For a suitable choice of $a_{x,s}(\omega)$ this process will generate variances which are fairly slowly changing functions of time. The procedure which we will present will, in effect, use this representation of the variance at each frequency to derive a prediction scheme at each point in time. We are, in a sense, predicting the spectrum and deriving the prediction of the time series from the prediction of the spectrum. A computationally simple way of doing this is as follows:

We first complex-demodulate the series x_t by forming

$$x_{j,t} = L(e^{i\omega_j t} x_t) \qquad j = 0, \ldots, m \qquad 3.2$$

where $x_{j,t}$ = the complex-demodulate of x_t around frequency ω_j

$L(\,)$ = a low pass filter.

The m + 1 frequencies at which we demodulate may, if we have no prior knowledge about the series, simply be chosen to lie uniformly between 0 and π. The appropriate value of m will depend on estimation criteria to be discussed later. However, if we have some knowledge of the structure of the process of which we have a realization, we may choose the demodulation frequencies in order to obtain estimates centered on frequencies of particular importance. In particular, if there seems to exist some kind of changing seasonal pattern in the record, then we should demodulate at the fundamental seasonal frequency and each of the harmonics if we intend to predict the seasonal pattern as a part of the process. If, on the other hand, we want to get rid of the seasonal and predict only the "non-seasonal" variation, then we should pick the demodulation frequencies so as to avoid the seasonals.

We should now observe that

$$f(\omega_j) = E(x_{j,t}\, x_{j,t}^*) \qquad 3.3$$

Thus we may consider the formation of estimates of the changing spectrum by taking moving averages of $x_{j,t}\, x_{j,t}^*$.

However, we may form our prediction of x_t directly by extrapolating each complex-demodulate, then remodulating and summing up the resulting series. We extrapolate the $x_{j,t}$ by a scheme based on equation 3.1

$$\hat{x}_{j,t} = K(\omega_j) + a(\omega_j)t + \sum_{s=1}^{\ell_1} \alpha_s(\omega_j)\, \hat{x}_{j,t-s} + \sum_{s=0}^{\ell_2} \beta_s(\omega_j) \qquad 3.4$$

We carry out the extrapolation to $t + \tau$, where τ is the number of time periods into the future which we wish to forecast.

We then remodulate the $\hat{x}_{j,t}$ by

$$\overset{\vee}{x}_{j,t} = \hat{x}_{j,t} e^{-i\omega_j t} \qquad 3.5$$

and form our final prediction by

$$\hat{x}_t = \text{Re} \sum_{j=0}^{m} \overset{\vee}{x}_{j,t} \qquad 3.6$$

Since we estimate the variance at each frequency by equation 3.4 nothing in this procedure requires stationarity of either first or second moments. All that we require for success in prediction is predictable change in the complex-demodulates. If the original series is from a stationary white noise process then this procedure will not prove more successful than the simple procedure of predicting the mean of the series. However, the more time varying structure there is in the process the better will be the forecasts by this procedure. In particular, if there is a deterministic component given by

$$x_t' = (\alpha + \beta t) e^{-i\omega_j t} \qquad 3.7$$

then its complex-demodulate at ω_j will be

$$x_{j,t} = \alpha + \beta t \qquad 3.8$$

Given a long enough realization to permit the formation of reasonable estimates of α and β we can predict this component with a very small error.

Incidentally, it is interesting to point out in this context that we can easily introduce into the computation a weighting function which represents our relative valuation of errors at each frequency. Thus we would replace equation 3.6 by

$$\hat{x}_t' = \mathrm{Re} \sum_{j=0}^{m} c_j \overset{\vee}{x}_{j,t} \qquad 3.9$$

where c_j = arbitrary weights constrained by

$$\sum_{j=0}^{m} c_j = m + 1 \qquad 3.10$$

However, we may choose that set of m weights which best reflect our determination of the costs of prediction errors at each frequency.

IV. Details of Estimation Procedure

In attempting to form predictions using the scheme outlined above, choices need to be made concerning the bandwidth of the low-pass filter used in demodulation and the coefficients for the extrapolation of the demodulates. These choices are clearly interdependent as the bandwidth of the filter determines the number of observations which are lost at the end of the series and the apparent smoothness of the series.

In the translation of the basic procedure indicated above into an operational numerical algorithm one important addition was made. The raw and partially filtered observations near the end of the realization were included with the demodulates and used in the extrapolation. This appears to produce somewhat improved estimates as it permits using a long moving average, from which better trend and low frequency estimates can be obtained, without completely losing the terms near the end of the realization.

The steps involved in the algorithm are as follows:

1. In order to form the complex demodulates we use convolutions of unweighted moving averages. This generates the kind of low pass filtering operation which we want in many fewer numerical operations than would be required if we used the weights given by the Fourier transform of the filter function. In addition, we preserve the end terms at each step of the convolution for use in the prediction equation. Thus we carry out the following operations

$$x_{j,t}^{(0)} = x_t e^{i\omega_j t} \qquad t = 1, \ldots, n$$

$$x_{j,t}^{(1)} = \sum_{s=-k_r}^{k} x_{j,t+s}^{(0)} \qquad 4.1$$

$$x_{j,t}^{(r)} = \sum_{s=-k_r}^{k} x_{j,t+s}^{(r-1)}$$

If we apply r convolutions we have remaining a series of n terms which has at each end k_1 terms which are $x_{j,t}^{(0)}$ $(t = n-k_1, \ldots, n)$, k_2 terms which are $x_{j,t}^{(1)}$ $(t = n-k_1+k_2, \ldots, n-k_1-1)$, etc. In practice we have used the special case of this scheme where

$$r = 4 \quad \text{and}$$

$$k = k_1 = k_2 = k_3 = k_4 \qquad 4.2$$

Thus the one variable which we have at our disposal at this point is k.

2. After forming a set of m + 1 complex valued series $x_{j,t}^{(4)}$ for $j = 0, 1, \ldots, m$, we must then form extrapolations of these series. Since these series are necessarily fairly smooth functions of time it is natural to apply a low order linear regressive scheme. However, before applying the linear regressive we have estimated trend coefficients according to

$$x_{j,t}^{(4)} = a_j + b_j t + y_{j,t} \qquad 4.3$$

We then apply the linear regression to $y_{j,t}$ and add back in the mean and trend terms to form the predicted complex-demodulates by forming

$$\tilde{x}_{j,t} = \sum_{s=1}^{\ell_1} \alpha_s \tilde{x}_{j,t-s} + \sum_{s=0}^{\ell_2} \beta_s y_{j,t-s} \qquad 4.4$$

where $y_{j,t} = 0$ for $t > n$ and then

$$\tilde{x}_{j,t} = \tilde{x}_{j,t} + a_j + b_j t$$

Since it is the magnitude and phase of $x_{j,t}^{(4)}$ which we hope will be slowly changing functions of time, equation 4.4 should be interpreted in terms of polar coordinates. Thus we extrapolate $|y_{j,t}|$ and $\theta_{j,t}$ rather than the real and imaginary parts of $y_{j,t}$. $\theta_{j,t}$ is, of course, given by

$$\theta_{j,t} = \tan^{-1} \frac{\mathrm{Im}[y_{j,t}]}{\mathrm{Re}[y_{j,t}]} \qquad 4.5$$

At this point we have to choose values for the coefficients α_s and β_s. If we consider equation 4.4 as a filtering scheme then we will want to choose the α_s and β_s on the basis of the properties of the transfer function which they determine. This transfer function is given by

$$T(z) = \frac{\sum_{0}^{\ell_2} \beta_s z^s \lambda(n + s - t)}{1 - \sum_{1}^{\ell_1} \alpha_s z^s} \qquad 4.6$$

where $\lambda(\ell) = 1 \qquad \ell \geq 0$

$= 0 \qquad \ell < 0$

Thus we want to choose T(z) according to criteria of phase and gain properties which depend on the number to time periods which we want to predict (t - n). We could consider estimating the α_s and β_s from the complex-demodulates using Weiner prediction procedures. However, in the present study we have simply chosen the coefficients on the basis of general gain and phase criteria.

3. Once the series $\hat{x}_{j,t}$ have been formed for values of t to $n + \tau$, where τ is the number of periods which we wish to predict, then we simply form

$$\hat{x}_t = \mathrm{Re} \sum_{j=0}^{m} \hat{x}_{j,t} e^{-i\omega_j t} \qquad 4.7$$

as the final forecast.

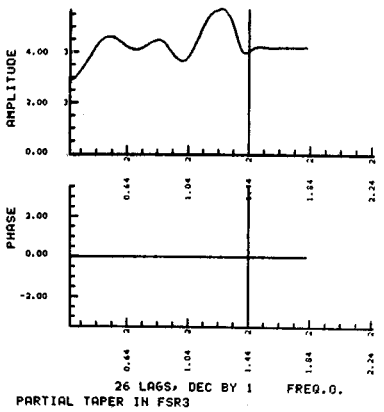
V. Empirical Results

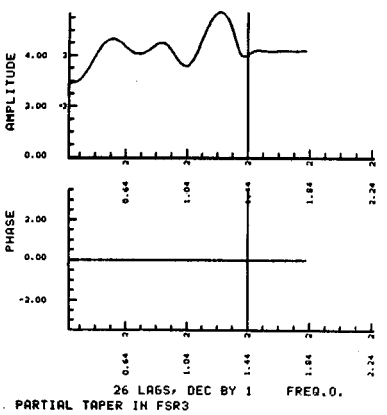
The following figures show the results of the application of this procedure to an actual time series. The application of the procedure is, at present, at an early exploratory stage. However, the results shown below seem to suggest interesting possibilities.

The series that was used was chosen partly due to its apparent time-varying seasonal pattern. The series is composed of observations taken at 10 day intervals. The data were recorded in such a way that there are exactly 36 observations per year. Thus the fundamental seasonal and its harmonics occur at:

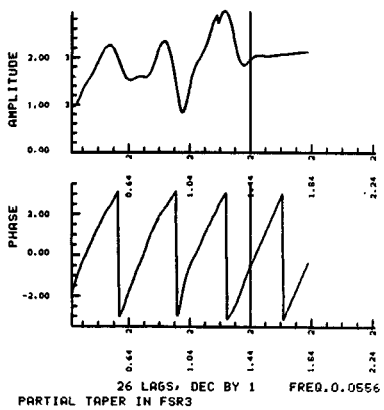
$$\omega = \frac{k\pi}{18} \qquad k = 1, \ldots, 18$$

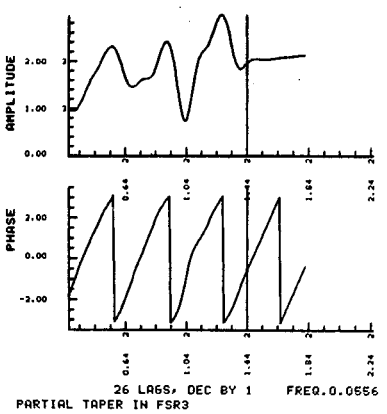
We used these 18 frequencies and $\omega = 0$ for the demodulation frequencies. The value of k for the filtering of the demodulates was 6. Figure 4.1 shows the demodulates and extrapolations of the demodulates for three selected frequencies. The graphs on the left of the figure are the demodulates, while



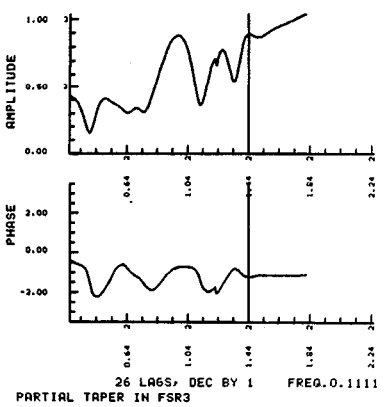


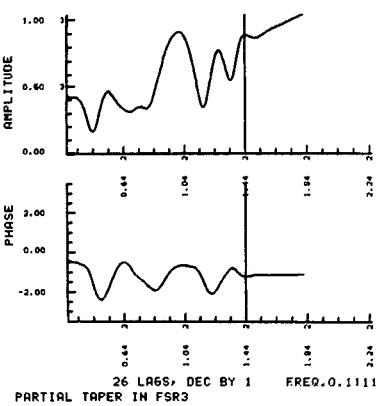
$\omega = 0$





$\omega = \pi/18$





$\omega = \pi/9$

Figure 4.1

Complex-demodulates ($\chi_{i,t}$) and predicted
Complex-demodulates ($\hat{\chi}_{i,t}$)

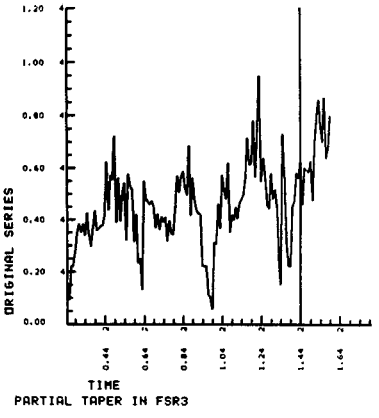


Figure 4.2
Original Series

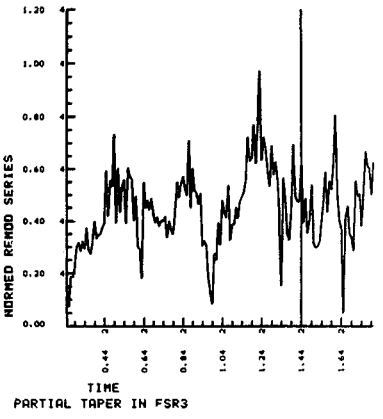


Figure 4.3
Remodulated predicted Series

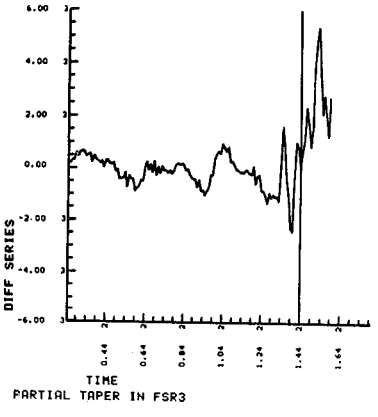


Figure 4.4
Difference between Original and Remodulated Series

those on the right are the corresponding smoothed and extrapolated series. The extrapolation equation used was:

$$\tilde{x}_{j,t} = 1.57\,\tilde{x}_{j,t-1} - 0.7\,\tilde{x}_{j,t-2} + 0.07\,y_{t-1} + 0.06\,y_{t-2} \qquad t = 4k+2,\ n$$

$$\tilde{x}_{j,t} = 1.57\,\tilde{x}_{j,t-1} - 0.7\,\tilde{x}_{j,t-2} \qquad t = n+1,\ n+\tau$$

In this case we used

$$n = 144$$

$$\tau = 20$$

Figure 4.2 shows the original series from which the demodulates were computed. Figure 4.3 shows the remodulation of the smoothed and extrapolated demodulates ($\hat{x}_{i,t}$). Finally, Figure 4.4 shows the difference between the original series and $\hat{x}_t$. We had 159 observations of the original series x_t. However, we used only the first 144 observations in computing $\hat{x}_t$. Therefore, there are 15 points for which we have known values of x_t for comparison with the predicted values.

The point at which we stopped using the original series for the computation is indicated by a vertical line in each figure. In Figure 4.2 the points to the right of the line are observations which were ignored. In Figure 4.3 these points are the predicted values. From inspection of Figures 4.3 and 4.4 the predicted values seem to reproduce the periodic

structure of the series with some indication that the prediction has properly adapted to the changing structure of the series.

It has not been possible, so far, to examine more formally the practical performance of this procedure. However, we should mention that many of the classical tests of prediction error cannot be directly applied as they rely on stationarity for their meaning.

VI. Conclusion

We have stated a prediction procedure for realizations from non-stationary stochastic processes which requires only that the spectral density function vary relatively slowly with time. We consider the series as if it had been generated by an approximately linear, but time varying, process.

We have written a program, in preliminary form, for the computation of this procedure. The application of this procedure to actual data yields interesting and encouraging results.