MASSACHUSETTS INSTITUTE of TECHNOLOGY
LIBRARIES

HD28
.M414
no.1448-83
c.2

LIBRARIES

WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

TESTING AN ASPECT OF MODEL PERFORMANCE USING SPECTRAL ANALYSIS

by

Robert L. Eberlein

WP 1448-83 March 1983

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

TESTING AN ASPECT OF MODEL PERFORMANCE USING SPECTRAL ANALYSIS

by

Robert L. Eberlein

WP 1448-83 March 1983

Testing an Aspect of Model Performance Using Spectral Analysis

Robert L. Eberlein
March, 1983

Abstract

A test for a dynamic model's ability to reproduce behavior modes of interest is developed and evaluated. The test consists of using the model to make one observation ahead predictions, and performing spectral analysis on the errors made in this prediction. This procedure is considered as a means of evaluating the ability of a linear model to reproduce a oscillatory behavior mode. Through the use of model generated data the test is evaluated in terms of its performance under the circumstances of model order failure, incorrect noise specification and incorrect model specification. The spectral analysis of residuals seems to be a useful test for determining a model's ability to reproduce an observed oscillatory behavior mode. Other tests for determining whether a model is useful in describing an aspect of observed behavior are called for.

This paper has benefited from the comments and criticisms of Alan Graham, Jim Hines, Ed Kuh, Jim Powell, George Richardson and Qifan Wang.

Introduction

Models are designed to give insights into reality. Reality is complex but models, if they are to be useful, cannot be. For this reason, the comparison of models and reality is very difficult. The purpose of this paper is to motivate and evaluate one simple technique for this comparison. This technique is the spectral analysis of one step model prediction residuals and is applicable to models designed to examine cyclic aspects of observed time series.

The basic idea underlying the technique is very simple. Given a model of a process, the model can be used to predict measured time series. The use of information available until time t for prediction at time $t+1$ allows for the best prediction. The errors generated by this prediction process can be recorded and their spectral characteristics evaluated through the use of a Fourier transform. An oscillatory behavior mode describes a frequency range of interest. Should the residuals exhibit inordinately great power in this frequency range the ability of the model to produce the behavior mode is called into question. This technique and some of its properties will be considered in this paper.

There are many statistical tests available for the evaluation of models. These tests are normally aimed at determining whether the model is explaining a time series up to an error which is white and contemporaneously uncorrelated with the model's explanatory variables. If a model is intended only to deal with a limited number of aspects of observed behavior such tests are not appropriate because to pass existing statistical tests the model must be able to generate the observed time

series up to a white noise process. Statistical tests which can be used to evaluate model performance with respect to a small subset of the observed behavior are called for. The technique considered in this paper is in this class.

The models which will be considered in this paper are dynamic time series models; that is, models which have dynamics determined by the model structure as well as exogenous inputs (Hannan, 1976). Specific consideration will be given only to linear models in this paper, though extension to the nonlinear case will be discussed. The general state space form (Harvey, 1981) of a dynamic time series model is

$$\underline{x}_{t+1} = \underline{A}\underline{x}_t + \underline{B}\underline{u}_t + \underline{D}\underline{w}_t$$
$$\underline{y}_t = \underline{C}\underline{x}_t + \underline{E}\underline{u}_t + \underline{F}\underline{v}_t \quad . \tag{1}$$

In these equations $\underline{x}$[1] represents a state vector of length k_x which summarizes the current position of the process. The vector $\underline{y}$ is of length $k_y \leq k_x$ and represents the observations actually made on the process. $\underline{u}$ is a vector of length k_u of exogenous inputs into the process and is often referred to as the control vector. It is assumed that $\underline{u}$ is observed exactly. $\underline{w}$ and $\underline{v}$, of length k_w and k_v respectively, are vectors of noise entering into the model's dynamic and measurement processes respectively. $\underline{A}$ (k_x by k_x) is the state transition matrix, $\underline{B}$ (k_x by k_u) the input matrix, $\underline{D}$ (k_x by k_w) the model noise matrix, $\underline{C}$ (k_y by k_x) the measurement matrix, $\underline{E}$ (k_y by k_u) the feed through matrix and $\underline{F}$ (k_y by k_v) the measurement noise matrix.

---

[1] Throughout the paper lower case underlined boldfaced letters will refer to vectors and uppercase underlined boldfaced letters will refer to matrices.

Model Purpose and Performance

The usual assumptions made in dealing with the above model are that **w** and **v** are independent white noise processes and that the matrices **A**, **B** and **C** satisfy certain rank conditions which makes the problem well specified (Mehra, 1974). Under these circumstances there are statistical techniques for estimating and evaluating the above models. The properties of system estimators are not, however, well documented (Schweppe, 1972).

If, in addition to the above assumptions, all states are observed (**C** = **I**) the statistical theory is quite well developed. This is the case considered by Hannan (1976). Under these circumstances the test for model validity tests the null hypothesis that the errors are contemporaneously uncorrelated with the state variables (**x**) and themselves uncorrelated over time.

The best known test of the errors is the Durbin-Watson test for first order autocorrelation. This test is performed and the results reported for most models based on time series data. The problem with the Durbin-Watson test is that it only tests for a very limited type of dynamic error structure. Testing for higher order autocorrelation and correlation between errors and explanatory variables was recommended by Box and Jenkins (1976), and this technique has more recently been extended to the vector time series problem (for example Poskitt and Tremayne 1982, Deistler, Dunsmuir and Hannan, 1978, Dunsmuir and Hannan 1976).

The statistical tests which have been developed for time series models are designed to detect non-white noise or correlations between the explanatory variables and the errors. For many models of interest it is highly unlikely that such tests would fail to detect problems. Many models are meant to be simple representations of complex processes. As a

consequence it is unlikely that the model can generate the multitude of behavior contained in a data series. If one or more modes present in the data are not reproduced by the model, the modes will appear in the time series of error terms. In this case the existing tests are not the appropriate ones. Rather, what is needed is a test of the model's explanation of certain aspects of the processes generating the data.

In contrast to using statistical techniques to evaluate model performance, it is possible to evaluate models on the basis of their dynamic characteristics. Recently improved computational techniques (CCREMS 1983, Perez 1981) have made the rigorous evaluation of model dynamics feasible for quite large models. This analysis can be applied to both theoretically oriented models (Forrester 1982) and econometric models (Kuh 1983). The analysis of models from this perspective can yield insights into the elements of the system determining different behavior modes.

One question which remains when performing such an analysis is whether the model behavior modes correspond to something which is a part of the system, and if so whether the model is accurately capturing the dynamics associated with that behavior mode. This is the question addressed by Laffargue (1979) and by Howrey and Klein (1972). These authors use two methods to answer this question. The first is the comparison of the dynamics implied by the linearizaton of the homogeneous model structure with the dynamics apparent in time series. The second is the simulation of the model with exogenous shocks to determine whether the output has the same dynamic characteristics as the observed data. This second comparison is normally done in the frequency domain.

The second of these techniques seems to get very close to the issues

with which this paper is concerned. There are, however, problems with the comparison of model output and observed time series. Translation to the frequency domain overcomes some, but not all, of these problems. The power spectrum of the output will depend on the product of the power spectrum of the noise and the transfer function from the noise to the output. Distinct noise inputs driving the same model can yield distinct outputs. Figure 1[2] shows the power spectrum for a second order linear model driven by white noise and by autocorrelated noise. The two spectra are quite different, and there is no indication that the output was generated by two structurally identical models.

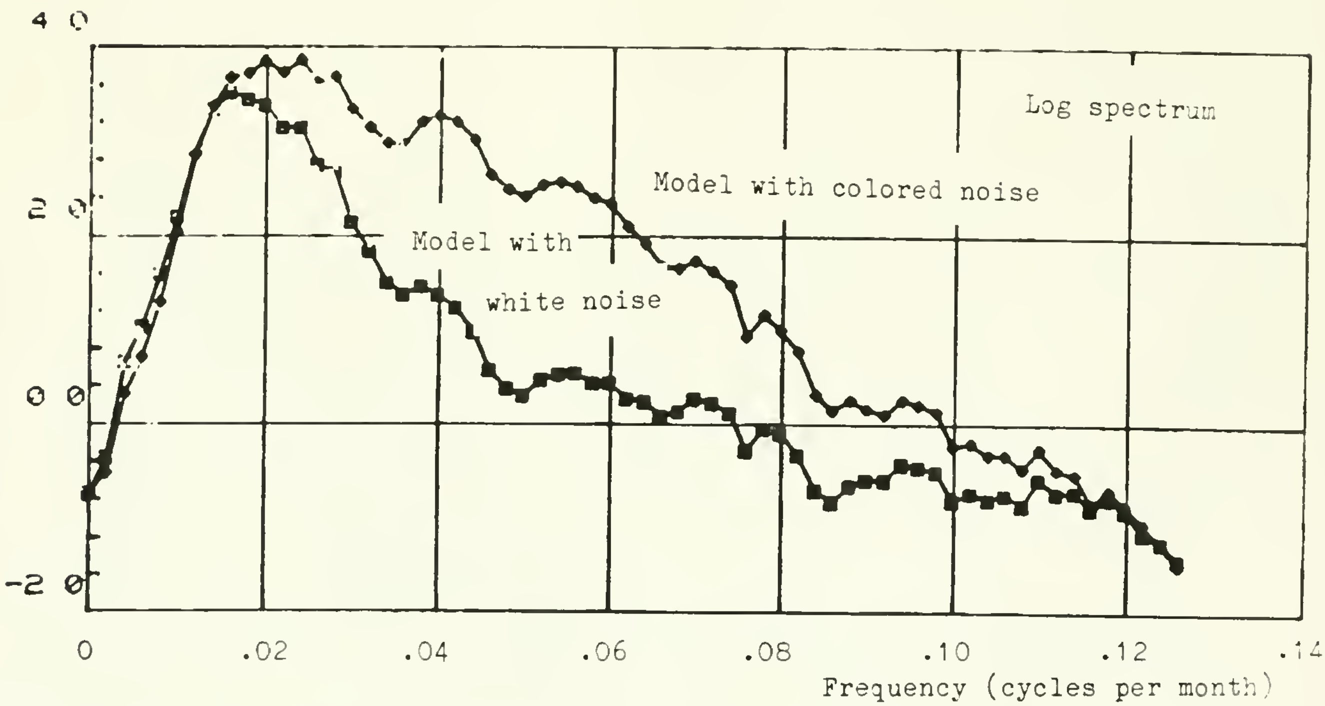

Figure 1 The power spectra for a second order model excited by two different types of noise.

---

[2] The complete description of how all the Figures were generated is contained in Appendix B.

Aspects of Behavior

The design of a model will depend on the issues which the model is meant to address. Very simple models can be the source of great insights into complex processes. The model, precisely because of its simplicity, cannot be expected to generate all components of a measured time series. Model purpose determines model boundary and the selection of a model boundary may violate the conditions necessary for standard statistical tests to be valid. The purpose of this section is to consider which statistical assumptions might be violated without detracting from the model's purpose.

The most obvious, and least troubling, way in which a model can fail to capture a process is through the assumptions on noise implicit in any statistical representation of a model. The noise entering a system may display a wide variety of dynamic characteristics attributable to a number of causes. For example the noise may display seasonal characteristics outside the model bounds. A case of this would be the attempt to explain housing starts based on demographic variables, disposable income per household, the stock of housing and the interest rate. Certainly such a model would be incapable of generating the seasonality observed in housing starts, and yet the model may be quite useful in other respects. The usual procedure in this case is to use seasonally adjusted data in order to get at the more fundamental characteristics of the problem.

The solution to the problem of separating dynamic modes is often not this simple and in the context of a dynamic system an approach such as seasonal adjustment may destroy a great deal of valuable information. This is because when noise is passed through a system the output dynamics will depend both on the dynamic characteristics of the noise going in and on the

system. Thus, seasonal adjustment of output variable is not the same as the correction of the system output for the seasonal characteristics of the noise. The actual effect of seasonal adjustment is not clear. This issue has, of course, been raised before (see for example Wallis 1974).

The second reason that a model may fail to match the processes of interest is that the model is often of lower dimension. In building a model it is desirable to keep the model small, for a dynamic model this corresponds to keeping the dimension of the model low. Keeping model dimension low while maintaining the model's applicability is difficult. For deterministic models Perez (1982) has given some indication of how to generate lower dimension counterparts to large models which already exist and have been verified. Unfortunately the task which most often faces the researcher is not the reduction of an existing model, but rather the verification of the already reduced model. In the case of economic models the only numerical data available for such verification are normally historical time series.

There are two types of model failure which are very serious and which a good series of performance tests should make clear. The first is simpler and just involves diagnosing the inability of a model to reproduce the behavior modes of interest in the data. The second is somewhat more subtle and involves the correctness of the explanation that the model is giving. It is the first type of failure at which the test considered in this paper will be aimed. The second type of failure is more difficult to discern and the reasons for this are worth considering.

A model can be chosen to match as closely as possible an observed time series without giving any understanding into the processes driving the observed series. This is the case with many of the time series models used

in forecasting. The purpose of a forecasting model is prediction, therefore such a model cannot be criticized on the basis of its inability to inform us about reality. The predictive time series model does, however point out a potential problem. There are an infinite number of state space representations of a given input output process (Chen 1970). There may exist more than one physically meaningful model which will generate a given output series. As an example of this consider two explanations of business cycles, the workforce-inventory model (Metzler 1956, Mass 1976) and the multiplier-accelerator model (Samuelson 1936, Low 1980). The two models can, by appropriate choice of parameters, be made dynamically equivalent in terms of the output paths generated (aggregate production in this case). However, conceptually the two models are distinct. The results of an analysis of policies intended to ameliorate the business cycle will be sensitive to the choice of model.

There is a need to be able to test aspects of model performance. Model failure attributable to undesirable noise characteristics, or incorrect model order may not be a significant criticism of a model designed to represent only one of several behavior modes. Model failure due to the inabilityy of the model to generate the desired behavior modes should be diagnosed. The rest of the paper will consider a simple test which is applicable when oscillatory behavior modes are of interest. The ability of the test to discern important model failures, while remaining insensitive to unimportant model failures, will be considered.

The Frequency Domain

In order to analyze historical time series it is useful to transform them from the time domain to the frequency domain. This approach allows the easy comparison of the phase and gain relationships of different time series at different frequencies (see for example Sargent 1979). Such an approach is useful because the noise entering the time series can obscure these relationships. In addition, the fourier decomposition of a series separates behavior modes associated with different complex roots of the homogeneous dynamic system matrix **A** . It is very difficult to do this in the time domain.

The easiest way to evaluate model output in the frequency domain is to generate output by simulating the model with noise entering. The output can then be compared to the observed time series in terms of the phase gain and power relationships at each frequency. This is the approach Laffargue (1979) uses. However this approach can be misleading for a number of reasons. The two most obvious problems are the inability of the power spectrum to distinguish the nature of the noise entering the system or the degree of damping of the mode.

Figure 1 showed how one system with two different noise input series could generate distinct spectra. The output spectra contain a combination of the characteristics of the noise exciting the system and the inherent system dynamics. The dynamics of the system are normally the things of interest. Consideration of the power spectra alone could quite easily lead one to reject an essentially correct model.

Consideration of the power spectrum can also lead one to accept a flawed model. Two systems with different dynamics can generate very similar spectra. Again this is most clearly shown by an example. It is

easy to design two second order linear systems to display oscillations with the same frequency but different degrees of damping. The power spectra for output from the models with the two sets of parameters is shown in Figure 2. The noise entering the models was white. The two spectra are very similar and certainly would not lead one to reject one model in favor of the other.

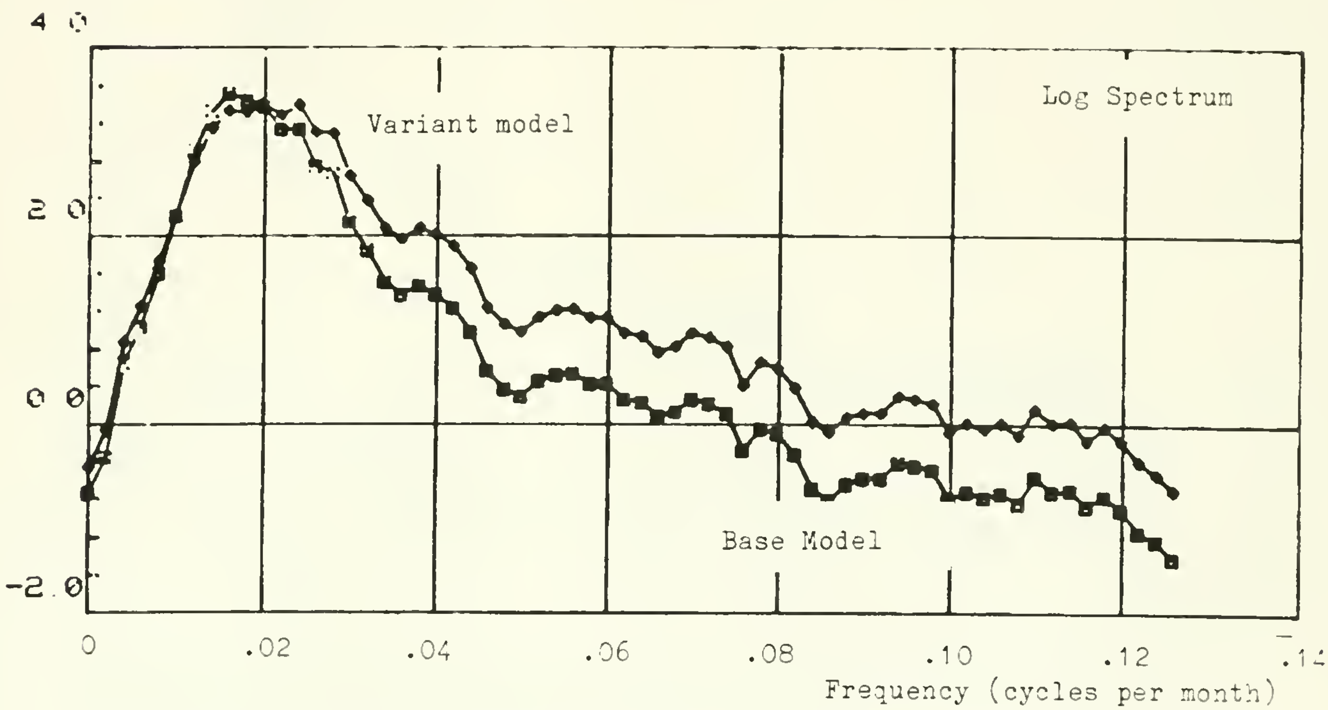

Figure 2 Power spectra for a second order model with two distinct sets of parameters.

There is an approach to the evaluation of model system correspondence which gets around this. This approach involves the evaluation of model prediction residuals. In Figure 2 the power spectra for two different models were plotted. If the second model was designed to generate the output of the first model, then it would make sense to use the second model to predict the output of the first model. The results of doing this are

shown in Figure 3. The power spectra for both the residuals from this prediction and for the output of the first model are plotted. The residuals show a peak at the frequency at which the output spectrum peaks, indicating that the model is not accurately representing the behavior mode of interest. The residuals distinguish between the two models.

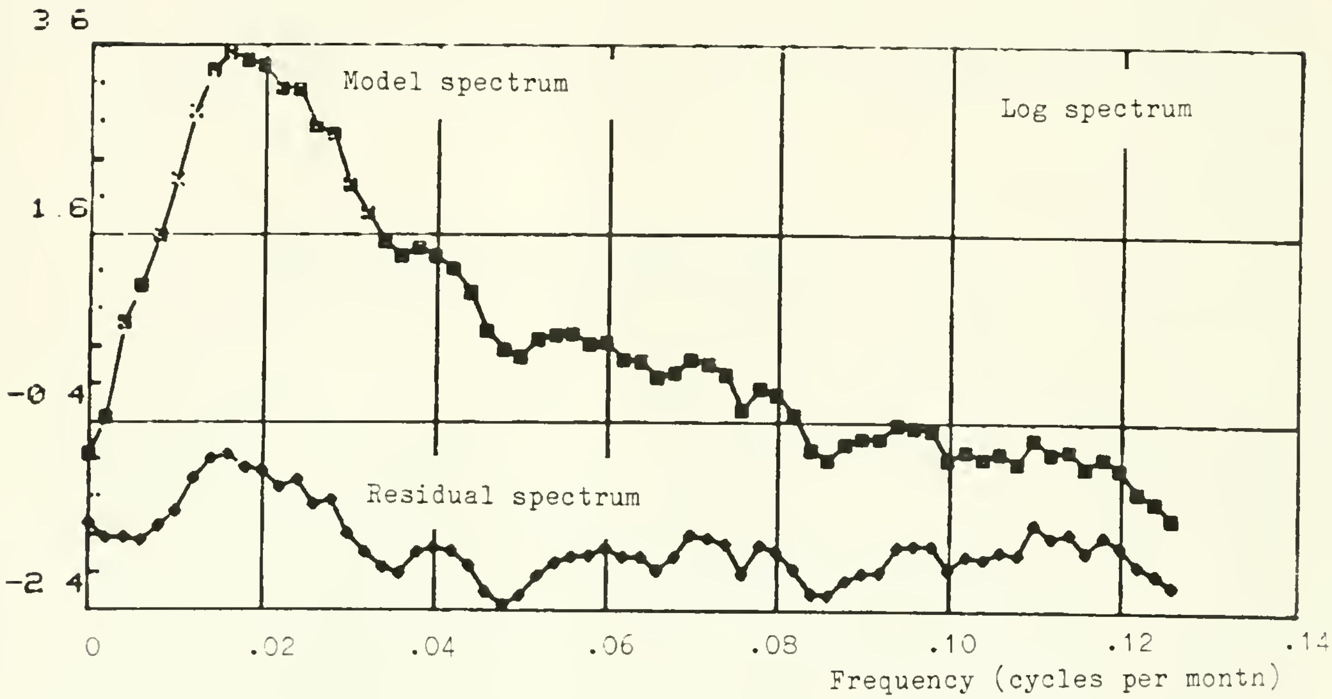

Figure 3 Plotted are the power spectra of the Base Model and of the residuals which result when the Variant Model of Figure 2 is used to explain the Base Mode output.

The use of residuals is clearly called for but the generation of residuals can be difficult and expensive. The residuals which are of interest are those generated with all components of the model active. For a large econometric model this would mean the full model simulation errors. That is, the errors which would be made if, given the model and the information at time t, the "best" forecasts of series at time $t+1$ were

made. If there is measurement noise, or some of the endogenous variables (for example expectations) are not observed the problem is somewhat more difficult. Kalman (1961) determined the best way to make this prediction given measurement errors and unobserved states. Kalman's solution will be described briefly below.

For the rest of this paper the properties of single time series will be considered. Phase and gain relationships between variables can be dealt with in a manner similar to that in which single output time series will be treated. However, this area is one in which more research needs to be done.

Spectral Analysis of Residuals (SAR)

The proposed "test" of model performance with respect to oscillatory behavior modes is quite simple. Given the model prediction errors, does their spectrum indicate that the behavior mode of interest is still present after using the model to "explain" the data? An indication of this would be a peak in the spectrum of the residuals near the point where the original time series's spectrum peaked. This test is not presented as a statistic with associated significance levels, though such an approach is possible in special cases. This method for evaluating model performance considered in this paper is somewhat more qualitative. Quantitative tests with general applicability are called for but have not yet been formulated.

It is possible to consider at a theoretical level the dynamic properties of the residuals when all the dynamic elements of the system are observed. This corresponds to perfect knowledge of $\underline{x}_t$ in equation 1 or equivalently in the notation of equation 1 $\underline{C} = \underline{I}$ and $\underline{F} = \underline{0}$. For notational convenience the matrix $\underline{C}$ will be taken to represent the

transformation to the scaler time series of interest, rather than the transformation to all the observed time series. Under these conditions the prediction error $\hat{y}_t - y_t$ is given by

$$e_{t+1} = (\underline{\tilde{C}}\underline{\tilde{A}} - \underline{CA})\underline{x}_t + (\underline{\tilde{C}}\underline{\tilde{B}} - \underline{CB})\underline{u}_t + (\underline{\tilde{E}} - \underline{E})\underline{u}_{t+1} - \underline{CDw}_t - \underline{Fv}_{t+1} \quad . \qquad (2)$$

In equation 2 the matrices $\underline{\tilde{C}}$, $\underline{\tilde{B}}$, $\underline{\tilde{B}}$ and $\underline{\tilde{E}}$ are from the model while $\underline{A}$, $\underline{B}$, $\underline{C}$, $\underline{D}$, $\underline{E}$ and $\underline{F}$ are from the system.

For colored noise entering the system with other elements of the model specification correct equation 2 takes the form

$$e_t = - \underline{CDw}_t - \underline{Ev}_t \quad . \qquad (3)$$

The dynamic characteristics of the prediction errors will precisely match those of a linear combination of the noise entering the system. If the noise entering displays dynamics distinct from the time series of interest then the spectral analysis of the residuals will support the model's ability to reproduce the behavior mode. If, on the other hand, the entering noise possess precisely the dynamics of interest there will be something of a problem. Such a situation might be considered a rather degenerate one[3], but if it occurs it becomes very difficult to disentangle the process and the noise. When such an identification problem exists,

---

[3]There are reasons why such a situation might arise. If, for example, there are two firms with one taking orders from the other then the orders from the first firm may have a random component similar to the output of the second firm. If only orders given to the first firm and the output of the second are observed, then the modeling of the two firms as one will leave residuals with dynamic characteristics similar to the orders made by the first firm. But it is not clear that the model is wrong.

there is no method which can successfully determine the model performance at the specified frequency.

The other and more serious way in which a model can diverge from reality is in lacking the actual fine structure of the system, or misrepresenting the structure. In the framework of equation 1 both of these are equivalent to errors in the $\underline{\tilde{A}}$ matrix. A model of lower order then the system corresponds to having zero entries in various places in the $\underline{\tilde{A}}$ matrix as well as the $\underline{\tilde{C}}$ matrix. In this case equation 2 gives an error which can take on almost any dynamic characteristics.

In particular, as long as there is error in the $\underline{\tilde{A}}$ matrix, the errors will be correlated with the values of $\underline{x}$. It follows that the errors may take on some of the dynamic characteristics of the observed time series. Whether these dynamic characteristics will be relevant to the behavior mode of interest is the important question. If they are the test proposed for model evaluation will give evidence against the model.

The above analysis was based on the assumption that the observation matrix $\underline{C}$ was the identity. There are two reasons why this is quite unlikely to be the case. The first is that when a model is put into a state space form, the observation of all states requires that the measurement noise be zero. The second is that the model may have in it certain states which cannot, by their nature, be measured.

If there is measurement noise in the observation of dynamic time series then such constructions as distributed lags will prevent us from exact state measurement. This is so because in the state space representation a formulation such as

$$y_{t+1} = .5x_t + .3x_{t-1} + .2x_{t-2} \tag{4}$$

becomes

$$z1_{t+1} = x_t \tag{5a}$$

$$z2_{t+1} = z1_t \tag{5b}$$

$$y_{t+1} = .5x_t + .3z1_t + .2z2_t \tag{5c}$$

If there is error in the measurement of x, then both $z1$ and $z2$ will diverge from the values which would be obtained by lagging x once and twice respectively. In such circumstances we would want to use information about the behavior of y as well as x in getting the best estimates of the z's .

The second obvious reason that the observation matrix might fail to be the identity is the explicit inclusion of unobservable variables into the model. If, for example, the model were to include an expectations variable z determined by the observed variables y and x according to the equation

$$z_{t+1} = .5x_t + .5y_t + w_t \tag{6}$$

with w representing noise then the value of z will not be known. yet z may determine the dynamic progression of the system.

When there are unobserved states, the use of a model to predict one step ahead is somewhat more difficult. However, there is a general solution to the problem for linear models known as Kalman filtering (Kalman 1962, Kalman and Bucy 1962) which can be extended to non-linear models

(Sage and Melsa, 1976). The Kalman filter is a technique for answering the following question. Given the best estimate we can get of the states at time t and the available observations at time t+1 what is our best estimate of the states at time t+1?

There are two ways to answer this question. The first is to simulate the model from time t to time t+1 . The resulting $\underline{x}$ would then represent our estimate of the states. The second is to compare the relationships between the states and the observed variable. Observed states could be set equal to their observed values. Unobserved states are still related to observations in some manner and therefore some guess for these is possible. The Kalman filter gives the optimal weights for combining these estimates. (Optimal under the correct model specification and white Guassian noise.)

Applying the SAR Test to Synthetic Data

The theoretical discussion has left it somewhat uncertain what the merits of the spectral analysis of residuals (SAR) test are. The purpose of this section is to simulate various types of model failures and to consider the ability of the SAR test to yield information on model performance. The simulations were done using second and fourth order linear models. The models and the specifics of the tests done are described in appendices A and B respectively.

For the purposes of the simulation it was assumed that there was only one observed time series. A Kalman filter was used to evaluate the states and generate the prediction error. The resulting prediction error was then used as the basis for analyzing model performance. The second order model was chosen to display oscillatory behavior. That is, the model was chosen

to have complex eigenvalues. The test of performance is therefore the amount of power which the residuals retain at the frequencies where model output power is greatest.

The first simulations were done using a model which matched the process up to the dynamic characteristics of the noise. Different types of noise were used to excite the second order System.[4] The model used to generate residuals was correct except that its specification did not include assumptions on the dynamic characteristics of the noise. The types of noise considered were: white noise for which the model is exactly correct, first order autocorrelated noise, noise with relatively high frequency oscillatory characteristics and noise with oscillatory characteristics near the frequency of the System output.

When white noise is used to excite the System the model is as correct as it is possible for a model to be. The model can be used to match the observed time series up to white noise. This can be clearly seen in Figure 4, which shows the power spectrum of the model output, the noise input and the calculated residuals. Other applicable model tests will not reject this model either. The Durbin Watson statistic is near 2 and averaged 1.9? over ten different noise seeds. The Kolmogorov-Smirnoff statistics (Jenkins and Watts 1968) testing for the whiteness of the spectra give a statistic of 18.2 distributed X^2 with 20 degrees of freedom which does not allow the hypothesis of whiteness to be rejected.

The introduction of first order autocorrelated noise into the System changes the power spectrum of the output. The power spectrum in this case peaks at a slightly lower frequency, consistent with the introduction of

---

[4]System with upper case S will refer to the model used in the simulations to generate the data. Thus the model being correct means that the model being used to generate the residuals is the same as the model used to generate the data.

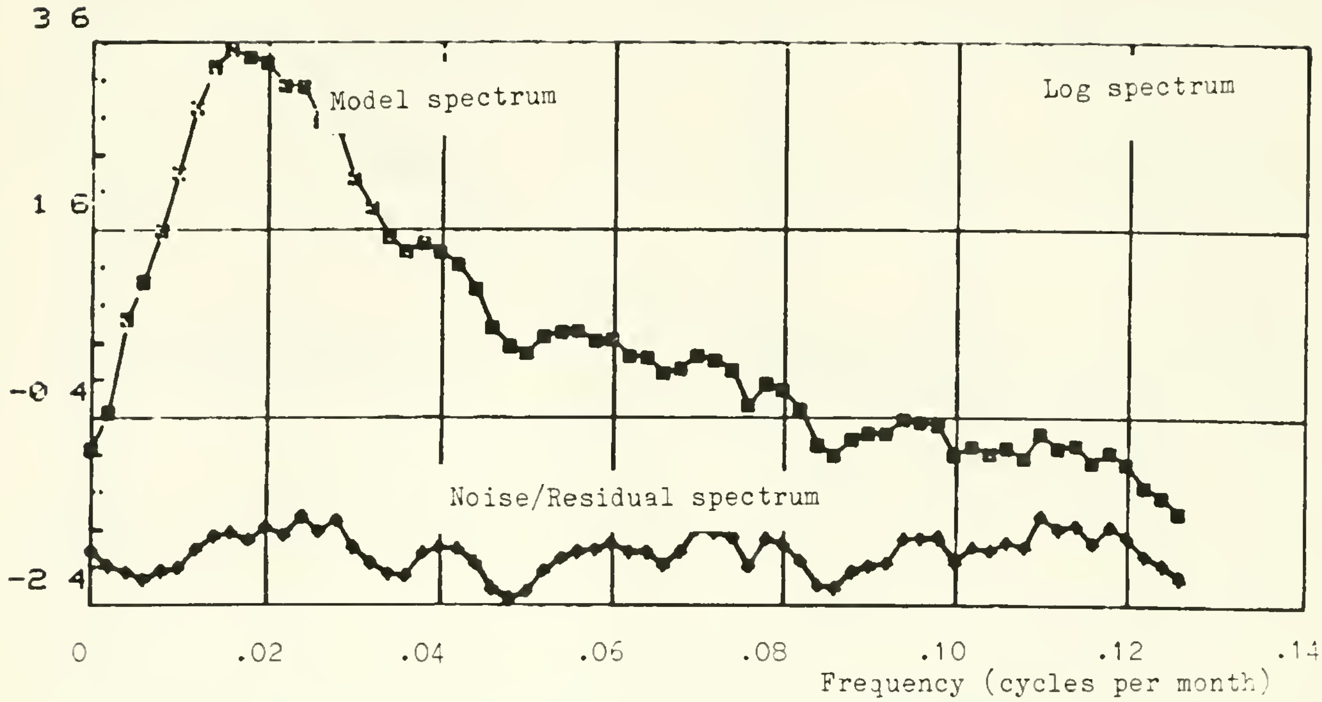

Figure 4 Power spectra for the Base Model excited by white noise, the white input noise and the residuals obtained when the Base Model is used to explain the output.

the autocorrelated noise which has a monotonically decreasing power spectrum. This can be seen in Figure 5. The power spectrum of the residuals matches fairly closely that of the autocorrelated noise and does not show the peak associated with the model output.

The introduction of noise which has a power spectrum with a distinct peak, in this case at a higher frequency than for the System yields a output spectrum as shown in Figure 6. The output spectrum in this case is double peaked which clearly suggests that there is more than one oscillatory behavior mode active. The relevant question in this case is whether the model can explain one of the behavior modes. Using the model to generate the residuals yields quite good results

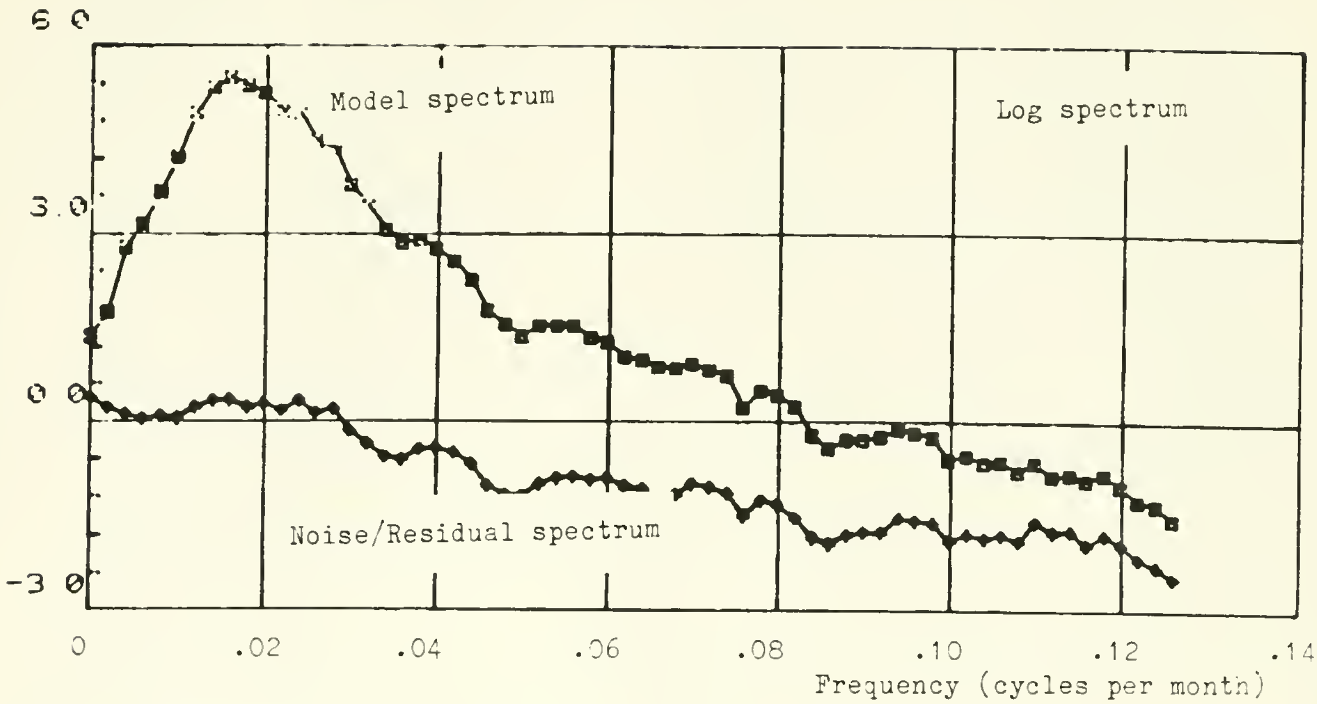

Figure 5 Power spectra for the Base Model excited by first order autocorrlated noise, the autocorrelated input noise and the residuals obtained when the Base Model is used to explain the output.

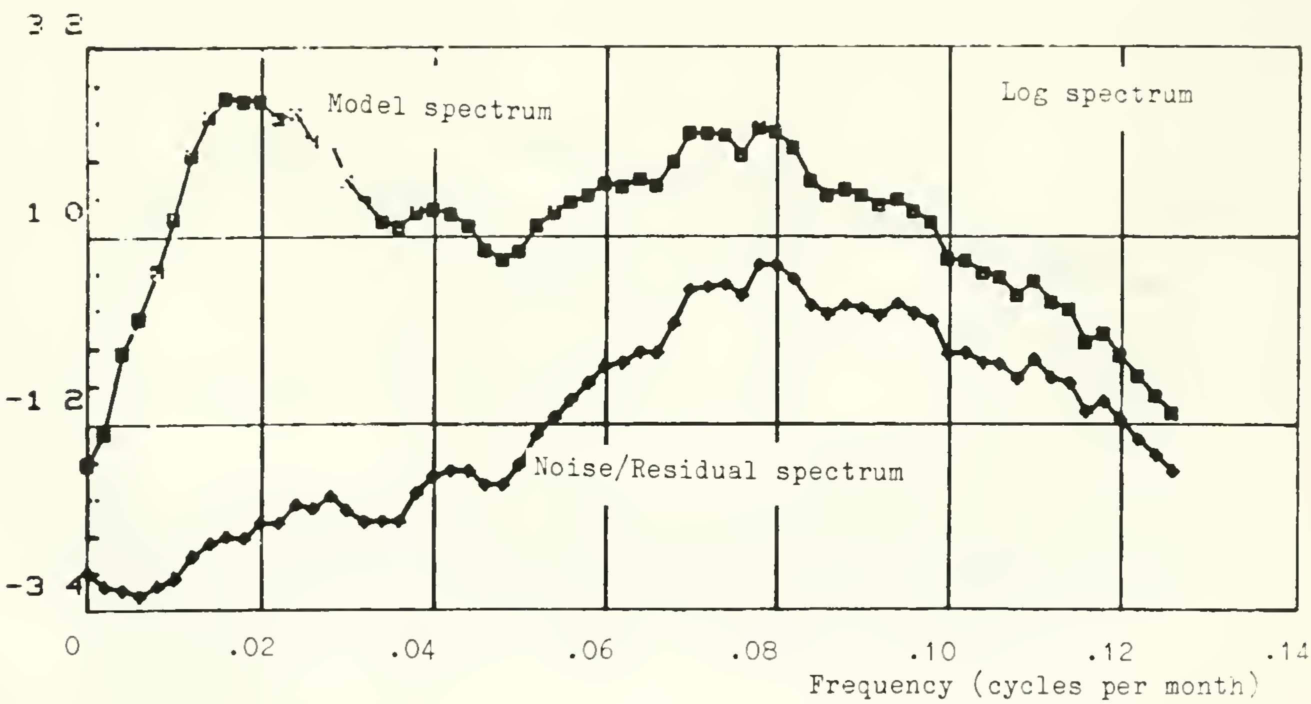

Figure 6 Power spectra for the Base Model excited by second order colored noise, the colored input noise and the residuals obtained when the Base Model is used to explain the output.

The introduction of noise with the same dynamic characteristics as the model output makes model performance difficult to evaluate. The power spectrum of the noise is very similar to that of the System driven by white noise as can be seen in Figure 7. However, when the System is excited by such noise the Power spectrum of the output indicates a much greater degree of variability. The reason for this is that the System is being excited at its natural frequency, and therefore tends to amplify the noise. The power spectrum of the residuals shows a peak at the same frequency as the System. The SAR test suggests that performance is poor. The SAR test, like any other test, yields misleading information in this situation.

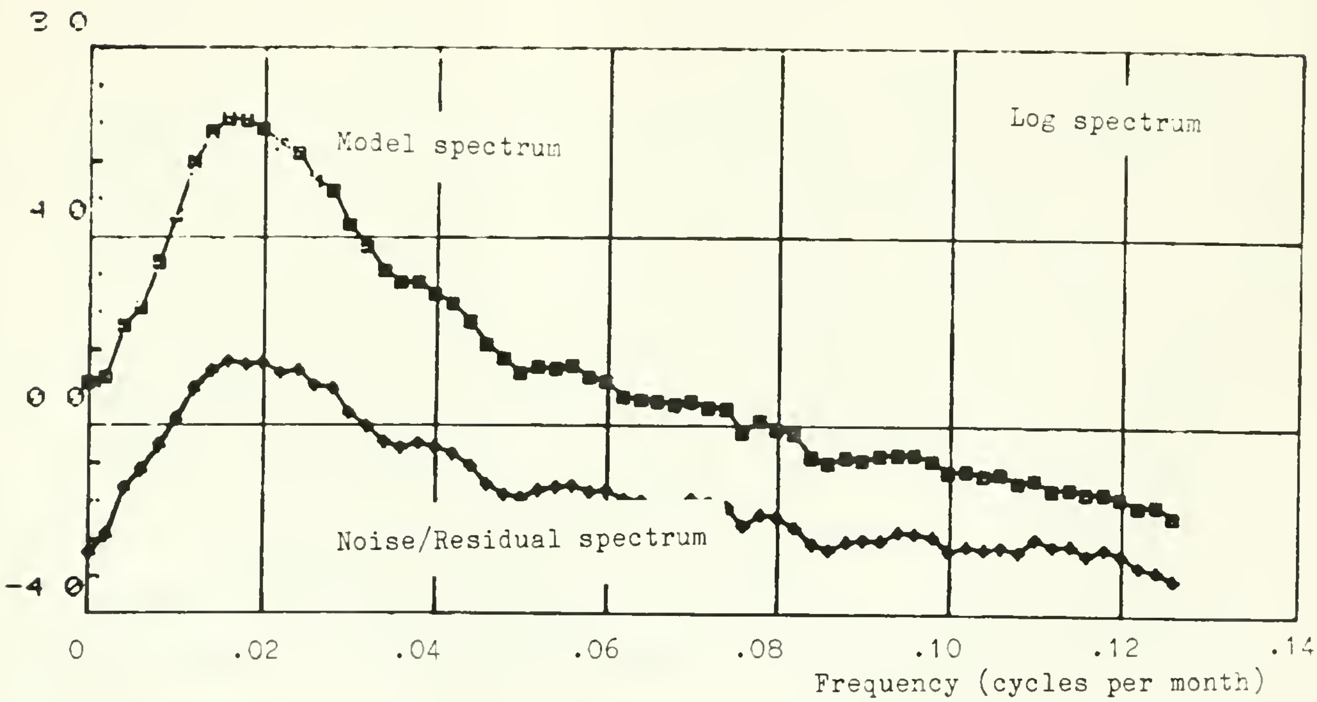

Figure 7 Power spectra for the Base Model excited by colored noise, the colored input noise and the residuals obtained when the Base Model is used to explain the output. The input noise was chosen to have dynamic characteristics similar to those of the base model.

The second situation simulated was the case of incorrect parameters. In this case parameters were given other than their actual value, though the model was the same structurally as the System. (The case of a structurally different system will be considered below.) There are two cases of interest in considering a model with wrong parameters. The first is that in which the model generates a spectrum quite similar to that of the System. The second is that in which the model spectrum is distinct from the system.

The model can be chosen to have a similar spectrum with distinctive damping characteristics. Consideration of the spectra in these circumstances does not yield any decisive information about the ability of the model to reproduce the behavior mode. The use of the residuals, as we saw above, does offer somewhat more information.

If the parameters of the model are chosen so that the power spectra of the model and that of the System diverge then it would be expected that the residuals would show a great deal of power at the frequencies of interest. Indeed this is the case as can be seen in Figure 8. The power spectrum of the residuals shows a marked peak at the System frequency, and a marked trough at the model frequency. In essence, the model is able to account for an excessive amount of the variability at the higher frequencies.

It is quite common to have small models represent complicated processes. This is something that a modeler often aims at in getting an understanding of a system. The ability of a small model to capture a behavior mode of interest will depend of the characteristics of the overall system. If there is a high degree of decoupling (distinct behavior modes associated with distinct state variables) a small model will do quite well in capturing the behavior. The greater the degree of coupling the less

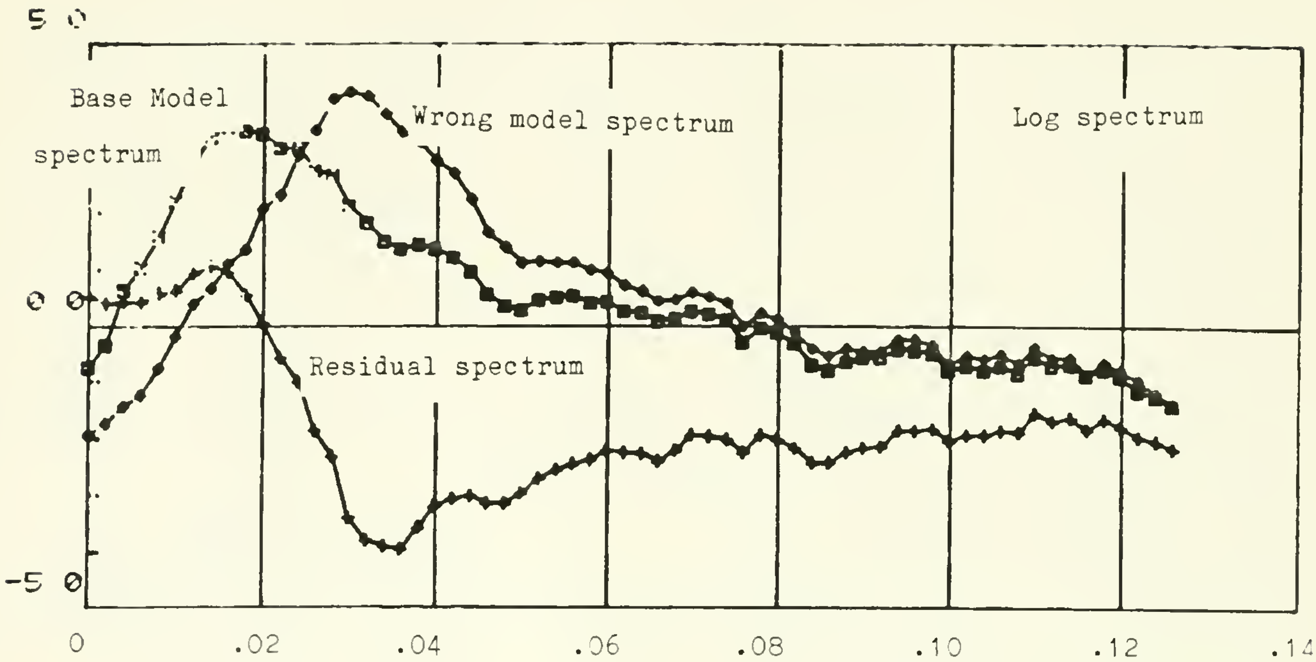

Figure 8 Power spectra for the Base Model, a model with different parameters from those of the base model and residuals generated when the model with different parameters is used to explain the Base Model output. The noise exciting the models is white.

this will be true. In order to get at the usefulness of the SAR test in identifying embedded behavior modes a fourth order system was considered. The fourth order system was chosen to have the basic second order system embedded in it, but to display two oscillatory behavior modes. The second order model was then used to generate residuals.

The SAR test was run on the residuals generated using a second order model to explain a fourth order System. The resulting residuals have very low power at the frequencies of interest as can be seen in Figure 9. The residuals do show a pronounced peak at the frequency of the second oscillatory mode of the System.

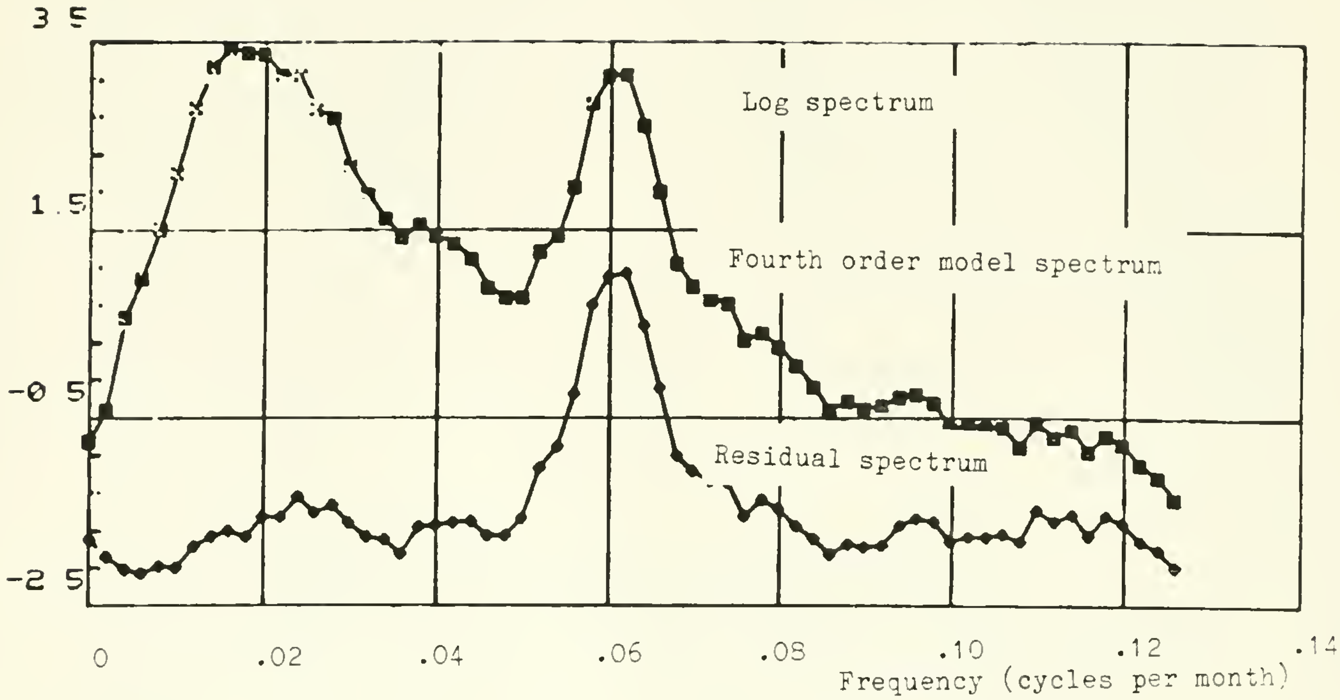

Figure 9 The power spectra for a more complicated model of fourth order and the residuals whic result when the Base Model is used to explain the more complicated model output.

The final area of failure for a model is that of a model which is an alternative state space representation of a process. Because there are an infinite number of state space representations of the same process (see for example Chen, 1972) it will be possible to fit many dynamic models equally well to a process. This is, of course, the point of the debates on observational equaivalence (see for example Sargent 1976). This normally does not represent a problem because it is rare for two state space representations with plausible physical or behavioral interpretations to coincide. Should this be the case, however, the SAR test will not be capable of distinguishing the two.

The example mentioned earlier of the workforce-inventory versus the

multiplier-accelerator explanation of business cycles brings this point out. The output generated by the two models under appropriate parameter specifications is indistinguishable. However, the models posit different state variables as being active in the process. The models can therefore be used to make statements about the relationships between different variables. These statements can then be checked empirically. One technique for doing this would be the phase and gain analysis of variable, or residuals, in the frequency domain. The SAR test cannot distinguish between some types of models. However, natural extensions to the SAR test could be used on many cases that come to mind.

Conclusions and Directions for Future Research

This paper has identified an area of model building, analysis and diagnosis which warrants further research. Specifically, the question of how to assess the ability of a model to deal with aspects of behavior has been addressed. One simple and fairly obvious technique, the spectral analysis of model prediction residuals, has been considered as a means of assessing aspects of model performances. When the aspect of model performance being considered is the ability to reproduce an oscillatory behavior mode the SAR is a useful diagnostic tool. The SAR technique breaks down if the noise dynamics are similar to the system dynamics. The SAR technique cannot distinguish between two models capable of reproducing the dynamics. In both these cases any univariate test is bound to fail.

The SAR test has the obvious extension to the spectral analysis of a vector of residuals. In this case both power spectra and the phase and gain relationships of the residuals could be considered. The model, if it is performing well, will remove the phase and gain relationships of

interest from the data. If the phase and gain relationships are still present in the residuals this would suggest that the model is failing, much as a peak in the power spectrum at the frequency of interest suggests a model failure.

The test has been considered only for linear models but there is nothing which prevents its extension to the nonlinear case. Prediction residuals and the fourier transform can be applied to nonlinear models. The interpretation of the power spectrum is somewhat less clear for nonlinear models. In order to deal with nonlinear models correctly it is necessary to define what is meant by a behavior mode in each particular case. Once this is done scrutinizing the residuals for evidence of this behavior mode is the natural nonlinear extension of the SAR test.

The problem of evaluating a model's ability to match the gross qualitative characteristics of observed behavior is very important, and very difficult. Techniques which are practical and informative in this task are doubtless called for. Tests such as the SAR test are one vehicle toward this end. But what is probably more important than any technique is a unified framework for approaching this problem.

Appendix A: Data Generating Models

The basic data generating model used was the workforce inventory oscillator (Forrester 1960, Mass 1976). The model was simplified to a second order model and written in linear form. The interaction between inventories and production are often considered to be the primary mechanisms producing business cycles and the model therefore seemed of some interest.

The model equations are given by

$$
\begin{aligned}
&P = WF*PROD \\
&WF = WF_{-1} + CWF_{-1} \\
&CWF = (DWF - WF)/TAWF + NOIS \\
&DWF = DP/PROD \\
&DP = O + IC \\
&IC = (DI - I)/TCI \\
&DI = DIC*O \\
&I = I + P_{-1} - O_{-1}
\end{aligned}
\qquad (7)
$$

P - Production (output units per month)
PROD - Productivity (output units per month per worker)
WF - Workforce (workers)
CWF - Change in workforce (workers per month)
DWF - Desired workforce (workers)
TAWF - Time to adjust workforce (months)
NOIS - A random component in the hire fire rate (workers per month)
DP - Desired production (output units per month)
O - Orders (output units per months)
IC - Inventory correction (output units per month)
DI - Desired inventory (output units)
I - Inventory (output units)
TCI - Time to correct inventory (months)
DIC - Desired inventory coverage (months)

Put in the form of equations 1 the equations become

$$\begin{vmatrix} WF \\ I \end{vmatrix} = \begin{vmatrix} 1-1/TAWF & -1/(PROD*TAWF*TCI) \\ PROD & 1 \end{vmatrix} \begin{vmatrix} WF-1 \\ I-1 \end{vmatrix} +$$

$$\begin{vmatrix} (T1+DIC)/(PROD*T1*T2) \\ -1 \end{vmatrix} O_{-1} + \begin{vmatrix} 1 \\ 0 \end{vmatrix} NOIS$$

$$P = WF*PROD$$

The introduction of the noise term corresponds to uncertainty in the results of advertising that jobs are available, and uncertainty about how many people will quit.

The base MODEL has an orders level of 100 units per month, a workforce of 100 people, a productivity of 1 unit per month per man and a desired inventory coverage of 1 unit per unit of output per month. The noise was chosen to have a variance of one.

The above MODEL with the choice of parameters TAWF=12 and TCI=6.5 is oscillatory with a period of approximately 55 months.

Higher Order Model

The higher order model is generated by adding some additional structure to the above model. It is quite simple and somewhat unrealistic. The desired speculative inventory change is a linear function of the difference between price and normal price. When the price is above normal price people want to buy and hold on to inventories. However, when inventories are above their desired level the price is driven down. The inventories are separated into two categories, distributors and producers. The producers are modeled as in the paper. The distributors act only as middle men and do not produce. The four levels in the model are the price, the distributor's inventory, the producer's inventory and the workforce. The equations are given by

$$P = PROD*WF$$

$$WF = WF_{-1} + (DWF_{-1} - WF_{-1})/TAWF$$

$$DWF = DP/PROD$$

$$DP = DRO + PIC$$

$$PIC = (PDI-PI)/TPCI$$

$$PDI = PDIC*DRO$$

$$PI = PI_{-1} + P_{-1} - DRO_{-1}$$

$$DRO = O + DIC$$

$$DIC = RDIC + SDIC$$

$$RDIC = (DDI - DI)/TDCI$$

$$DDI = DDIC*O$$

$$SDIC = (PRC - NPRC)*MPI$$

$$DI = DI_{-1} + DRO_{-1} - O_{-1}$$

$$PRC = PRC_{-1} + CPRC_{-1} + (NPRC - PRC_{-1})/TNP$$

$$CPRC = (DDI-DI)*MIP1 + (PDI-PI)*MIP2$$

P - production (units per month)
PROD - productivity (units per month per worker)
WF - workforce (workers)
DWF - desired workforce (workers)
TAWF - time to adjust workforce (months)
DP - desired production (units per month)
DRO - distributors orders (output units per month)
PIC - producer's inventory correction (units per month)
PDI - producer's desired inventory (units)
PI - producer's inventory (units)
TPCI - time for producers to correct inventory (months)
PDIC - producer's desired inventory coverage (months)
O - orders (units per month)
DIC - distributor's inventory correction (units per month)
RDIC - real distributor's inventory correction (units per month)
SDIC - speculative inventory correction (units per month)
DDI - distributor's desired inventory (units)
DI - distributors inventory (units)
TDCI - time for distributors to correct inventory (months)
DDIC - distributors desired inventory correction (units per month)

PRC - price (dolloars per unit)

NPRC - normal price (dollars)

MPI - multiplier from price for inventory (units per month per dollar)

CPRC - change in price (dollar per month)

MIP1 - multiplier from inventory for price (dollars per month per unit)

MIP1 - multiplier from inventory for price (dollars per month per unit)

The model run used MPI=5, NPRC=1, PROD=1, PDIC=.6, DDIC=.4, O=100, TDCI=15, TNP=15, TAWF=12, and TPCI=6.5, MIP1 = .03 and MIP2=.001

The noise used in the model was generated by passing white noise through a second order linear filter. The general equations for the filter are given as

$$X = aX_{-1} + bY_{-1} + eWN$$
$$Y = cX_{-1} + dY_{-1} + fWN$$
$$NOIS = gX + hY$$

Where the term WN is a white noise term, the variance of WN is adjusted so that the variance of NOIS is one. The output NOIS has color characteristics. The specifics of the parameter values chosen for each plot are described in appendix B.

Appendix B: Description of the Simulations Done

The simulations described were all repeated with 10 different noise seeds to try to get at general characteristics of the processes. The plots shown in the paper are all the average spectra over all 10 runs. The spectra were smoothed with a window width of 3. The runs were all done for a length of 420 months. 512 was chosen as te basis for the Fourier transform. The simulations were all carried out using FORTRAN programs which are available on request.

Figure 1

Figure one was produced by using white noise in the inventory-workforce model and by using colored noise in the inventory-workforce model. The colored noise was generated by setting a=.8, b=-.07, c=1.0, d=.9, e=1.0, f=0.0, g=1.0 and h=1.0 . The model was run with TAWF=12.0 and TCI=6.5.

Figure 2

Figure 2 was created by using two different sets of time constants in the workforce-inventory model. The first were TAWF=12.0 and TCI=6.5 . The second were TAWF=8.0 and TCI=8.4 . The noise seeds to the model with the second constants were chosen to have twice the variance of the noise seeds to the model with the first set of constants.

Figure 3

Figure three plots the spectrum from the workforce-inventory model with TAWF=12.0 and TCI=6.5 being excited by white noise. Under the assumption that TAWF=8.0 and TCI=8.4 the workforce-inventory model was used to generate residuals.

Figure 4

Figure 4 plots the model spectrum, the input noise spectrum and the residual spectrum for white noise. The w-i model was run with TAWF=12.0 and TCI=6.5 .

Figure 5

Figure 5 plots the model spectrum, the input noise spectrum and the residual spectrum for autocorrelated noise. The noise was generated by setting a=.8, b=0.0, c=0.0, d=0.0, e=1.0, f=0.0, g=1.0 and h=0.0 . The w-i model was run with TAWF=12.0 and TCI=6.5 .

Figure 6

Figure 6 plots the model spectrum, the input noise spectrum and the residual spectrum for second order colored noise. The noise was generated by setting a=.85, b=-.2, c=1.0, d=.75, e=1.0, f=0.0, g=1.0 and h=0.0 . The w-i model was run with TAWF=12.0 and TCI=6.5 .

Figure 7

Figure 7 plots the model spectrum, the input noise spectrum and the residual spectrum for second order colored noise with the same characteristics as the model output. The noise was generated by setting a=.9, b=-.0135, c=1.0, d=1.0, e=1.0, f=0.0, g=1.0 and h=0.0 . The w-i model was run with TAWF=12.0 and TCI=6.5 .

Figure 8

Figure 8 plots the spectrum from the workforce-inventory model with TAWF=12.0 and TCI=6.5 being excited by white noise and the spectrum of the w-i model with TAWF=12 and TCI=2. Under the assumption that TAWF=12.0 and TCI=2.0 the w-i model was used to generate residuals from the TAWF=12, TCI=6.5 data.

Figure 9

Figure 9 plots the spectrum fot the more complicated model with hoarding and the residuals from explaining this time series using the w-i model. The hoarding model values ae as in appendix A. The w-i values assumed for generating the residuals were TAWF=12.0 and TCI=6.5

Bibliography

Anderson, B.O. and J.B. Moore, Optimal Filtering Techniques, Prentice Hall Englewood Cliffs, New Jersey: 1979

Blaszcynski, G.M. Principles of Model Determination, PhD Theses, MIT Electrical Engineering, 1982

Box, G.M. and G.M. Jenkins, Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco: 1976

CCREMS, "TROLL PROGRAM LIMO", Technical Report #34, Center for Computational Research in Economics and Management Science, MIT, Cambridge, MA. 1983

Chen, C.T., Introduction to Linear System Theory, Holt, Rinehart and Winston, Inc. New York: 1970

Deistler, M., W. Dunsmuir and E.J. Hannan, "Vector Linear Time Series Models: Corrections and Extensions," Advances in Applied Probablilty, 10 (1978) 332-72

Dunsmuir, W. and E.J. Hannan, "Vector Linear Time Series Models," Advances in Applied Probability, 8 (1976) 339-64

Forrester, J.W. Industrial Dynamics, Wright Allen Press, Cambridge, Massachusetts; 1961

Forrester, N.B., A Dynamic Synthesis of Basic Macroeonomic Theory: Implications for Stabilizations Policy Analysis, Ph.D. Theses, Sloan School of Management, MIT, Cambridge, 1982

Godfrey, L.G. "Testing the Adequacy of a Time Series Model," Biometrica, 66 (1979) 67-72

Granger, C.W. and P. Newbold, Forecasting Economic Time Series, Academic Press, New York: 1977

Hannan, E.J. Multiple Time Series, John Wiley, 1970; New York

Harvey, A.C., Time Series Models, John Wiley and Sons, New York: 1981

Howrey, E.P. and L.R. Klein, "Dynamic Properties of Nonlinear Econometric Models," International Economic Review, 13 (1972) 599-617

Jenkins, G.M. and D.G. Watts, Spectral Analysis Holden Day, San Francisco: 1968

Kalman, R. "A New Approach to Linear Prediction and Filtering Problems" Journal of Basic Engineering, Series D, 82 (1960) 35-45

Kalman, R. and R. Bucy "New Results in Prediction and Filter Theory," Journal of Basic Engineering, Series D, 83 (1961) 95-108

Kuh, E. Applying Dynamic Model Analysis Techniques: MQEM A Case Study in Inflation, CCREMS, MIT, 1983

Laffargue, J.P. "Properties Cyclique des Modeles Economitriques," Unpublished, CEPREMAP, Paris, 1979

Low, G.W. "The Multiplier Accelarator Model of Business Cycles Interpreted From a System Dynamics Perspective", TIMS Series in the Management Sciences 14 (1980) 107-24

Harvey, A.C. Time Series Models, John Wiley and Sons, New York; 1981

Mass, N.J., Economic Cycles: An Analysis of Underlying Causes, Wright Allen Press, Cambridge, Massachusetts; 1975

Meadows, D.L., Dynamics of Commodity Cycles, The MIT Press, Cambridge Mass.: 1970

Mehra, R.K. "Identification in Control and Econometrics: Similarities and Differences," Annals of Social and Economic Measurement 3 (1974)

Metzler, L.A. "The Nature and Stability of Inventory Cycles," The Review of Economics and Statistics, 23 (1941) 113-29

Perez-Arriaga, J.I. Selective Model Analysis With Applications to Electrical Power Systems, Ph.D. Thesis, Electrical Engineering and Computer Science, MIT, Cambridge, 1981

Pierce, D.A., "Residuals Correlations and Diagnostic Checking in Dynamic Dusturbance Time Series Models," Journal of the American Statistical Association, 67 (1972) 636-40

Poskitt, D.S. and A.R. Tremayne, "Diagnostic Tests for Multiple Time Series Models," The Annals of Statistics, 10 (1982) 114-20

Richardson, G.P. and A.L. Pugh III, Introduction to System Dynamics Modeling with DYNAMO, MIT Press, Cambridge: 1981

Sage, A.P. and J.L. Melsa, Estimation Theory With Applications to Communications and Control, McGraw Hill, New York, 1971

Samuelson, P.A. "Interactions Between the Multiplier Analysis and the Pinciple of the Accelerator," Review of Economics and Statistics, 21 (1939) 75-79

Sargent, T.J., "Observational Equivilance of natural and Unnatural Rate Theories of Macroeconomics," Journal of Political Economy, 84 (1976),

-----------, Macroeconomic Theory, Academic Press, New York:1979

Schweppe, F.C., Uncertain Dynamic Systems, Prentice Hall Inc., Englewood Cliffs, N.J.; 1973

Sims, C.A. "Money Income and Causality," The American Economic Review, 62 (1972), 540-52

Wallis, K.F., "Seasonal Adjustment and Relations Between Variables," Journal of the American Statistical Association, 69 (1974) 18-31

4353 C44

MIT LIBRARIES
3 9080 004 492 929

Date Due

FEB 12 1991

Lib-26-67

BASEMENT

Bar Code
On Last
Page