VALIDATION OF OSCILLATORY BEHAVIOR MODES USING SPECTRAL ANALYSIS

Robert L. Eberlein
Qifan Wang

System Dynamics Group
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Abstract

In this paper we outline and evaluate a simple technique for analyzing the ability of a model to reproduce an oscillatory behavior mode. The technique consists of using a model as a predictor, and then performing spectral analysis on the prediction errors. The technique is referred to as the spectral analysis of residuals or SAR test. The paper motivates the use of prediction residuals and illustrates the technique with a simple model of inventory oscillation. The SAR test appears to yield a substantial amount of information about the performance of a model. However, the technique breaks down if the observed behavior is a result of the system being subjected to shocks with similar dynamic characteristics to the system output or if the system has more than one set of mechanisms generating the behavior of interest. The SAR test is not capable of distinguishing between models which can explain the behavior equally well using different state space representations.

Introduction

The process of model building in System Dynamics involves the consideration of the behavior modes of the system being modeled. The reproduction of these behavior modes is considered to be an important part of the model validation process [1]. The evaluation of the model's ability to reproduce behavior modes is difficult and requires a great deal of time consuming analysis of model output. The purpose of this paper is to present one statistical aid for behavior-mode verification. The use of statistical tools can increase both the efficacy of behavior-mode validation and the ability of the modeler to communicate validity.

Validation in System Dynamics modeling has been the focus of a good deal of criticism of the field [2]. The responses to this criticism have been varied. The strongest defense of the existing validation techniques in System Dynamics has been the inability of proposed alternatives to adequately deal with the problem [3]. However, there has been some work done to develop statistical techniques for the model validation process.

Peterson [4,5] advocates the use of system identification techniques developed for engineering models. Sterman [6] considers the use of statistical tools not for the determination of model structure, but for the comparison of the model with observed data. This paper is more in line with the latter approach.

In this paper we concentrate on the validation of a model's ability to reproduce oscillatory behavior modes. Much of what will be said regarding the validation process in this special case does have more general applicability. We concentrate on the single behavior mode validation because it helps in the motivation of the approach chosen. The aspects of the approach which generalize will be discussed in the conclusions. These generalizations are areas in which research is currently being done.

This paper is organized as follows. We first discuss the value of considering the spectral density of a time series when evaluating the dynamic characteristics of that time series. We then motivate the use of model prediction residuals as the appropriate data for testing. The proposed testing technique is then sketched out. The situations in which the proposed test is and is not useful as a diagnostic tool are then considered. Finally, areas that warrant further research are indicated.

Oscillations and the Power Spectrum

To facilitate discussion of oscillatory behavior modes we will consider a very simple model which has an oscillatory mode. This is the workforce/inventory oscillator [7,8]. The model was simplified to a second-order model and written in linear form. The model is of interest because of the importance of inventory/workforce interactions in the business cycle [8]. The model equations written in DYNAMO are given in Figure 1. A noise term is introduced into the workforce level and is meant to represent the randomness of the results of advertising that jobs are available, and unpredictable variations which occur in the number of quits.

When subjected to a 10% step increase in orders the model yields a production rate that displays damped oscillatory behavior as can be seen in Figure 2. However, when the model is subject to a noise input, as in Figure 3, the oscillations are not as easy to analyze. The reason for this is the ability of the noise to generate cycles of different frequencies when viewed in the time domain.

An alternative way to view an oscillatory pattern is in the frequency domain. This is done by taking the Fourier transform of the model output [9]. What the Fourier transform contains is information on the frequencies that dominate the model behavior. Using the fourier transform the power spectrum for a time series can be obtained. The power spectrum for the workforce inventory

Equations

```
A P.K=WF.K*PROD             - Production (output units per month)
L WF.K=WF.J+DT*CWF.JK       - Workforce (workers)
N WF=OR.K/PROD
R CWF.KL=(DWF.K-WF.K)/      - Change in workforce (workers per
                 month)
X   TAWF+NOIS.K
A DWF.K=DP.K/PROD           - Desired workforce (workers)
A DP.K=OR.K+IC.K            - Desired production (output units per
                 month)
A IC.K=(DI.K-I.K)/TAI       - Inventory correction (output units per
                                month)
A DI.K=DIC*OR.K             - Desired inventory (output units)
L I.K=I.J+DT*(P.J-O.J)      - Inventory  (output units)
N I=DIC*O
```

Constants

```
C PROD=1     - Productivity (output units per month per
                             worker)
C TAWF=12    - Time to adjust workforce (months)
C TAI=6.5    - Time to adjust inventory (months)
C DIC=1      - Desire iventory coverage (months)
```

Exogenous inputs

```
A NOIS.K=NORMRN(0,1)          - Noise (dimensionless)
A OR.K=100*(1+STEP(SS,ST))   - Orders (units per month)
```

Figure 1. Simple Workforce/Inventory Equations

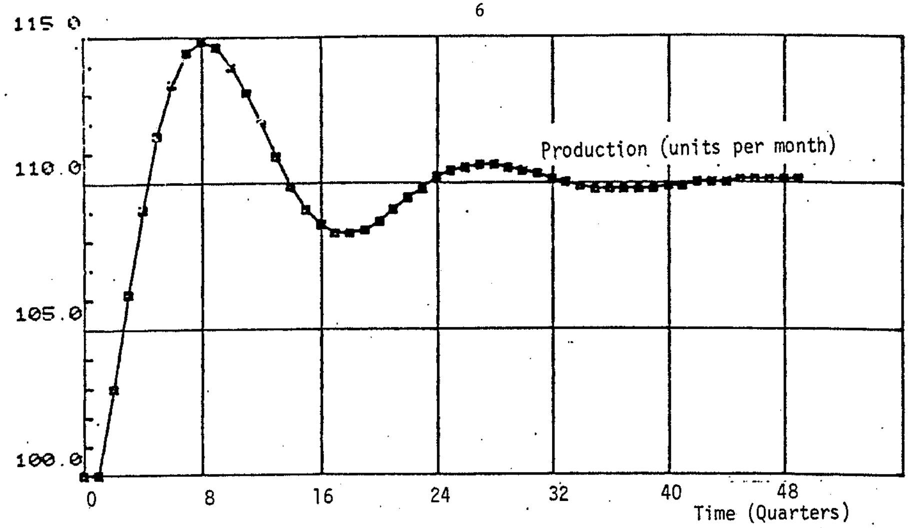

Figure 2 The response of the workforce/inventory model to a 10% increase in orders.

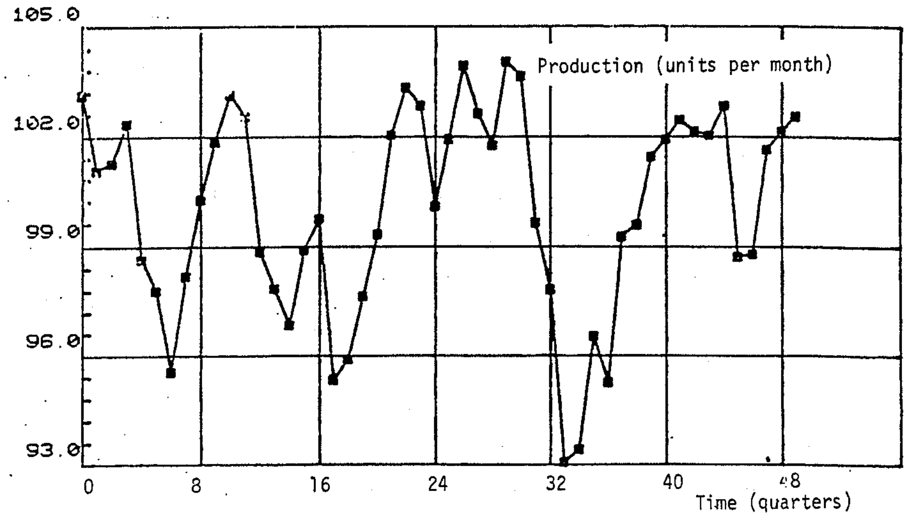

Figure 3 The response of the inventory/ workforce model to noise in the hire fire rate

model output of Figure 3 is shown in Figure 4. The horizontal axis in Figure 4 represents the frequency (per month) at which the tendency of the model to oscillate is being evaluated. The vertical axis represents the power, or the tendency of the model to show oscillations at that period. The peak in the power spectrum corresponds to the period for which the system is most oscillatory.

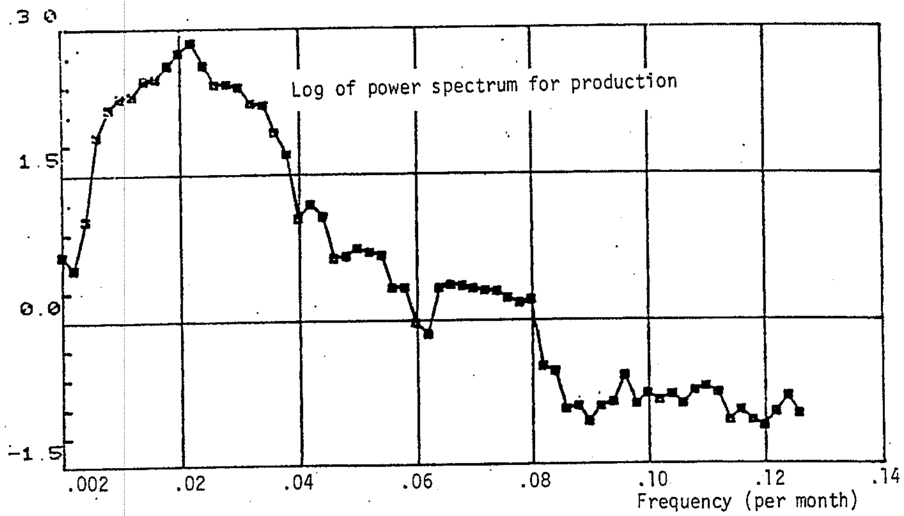

Figure 4 Power spectrum for inventory/workforce output shown in Figure 3.

The peak in the power spectrum of the model output occurs at a period of approximately 50 months. The spectrum gives information about oscillatory tendencies in a clear form not available from the observation of the noise run shown in Figure

3. For a real system the only time-series available will be those corresponding to noisy runs. The analysis and comparison of the oscillatory tendencies of actual data and model output is much easier in the frequency domain.

Senge [10] used this technique of translating time-series into the frequency domain in the comparison of model generated output with available data on investment. The comparison of two time-series in the frequency domain is much easier than is the comparison in the time domain. The reason for this is simple. Different noise inputs generate different outputs, but when translated to the frequency domain the similarities of the dynamics are preserved. This can be clearly seen in Figure 5 and 6. Figure 5 represents the simple inventory model run under two different noise seeds. The noise for the two runs has the same statistical characteristics, but different actual values after time t_1. The two series are clearly different in the time domain after the noise seed changes. Figure 6 represents the power spectrum of the model output for each noise input. Unlike the model output, the spectra are almost the same.

There are, however, shortcomings to this approach. Two models with different parameters can produce similar spectra. Consider for example the choice of model parameters which will generate oscillations of the same period, but with different degrees of damping. This can be accomplished with the inventory oscillator by changing the time to adjust inventory from 12.0 to

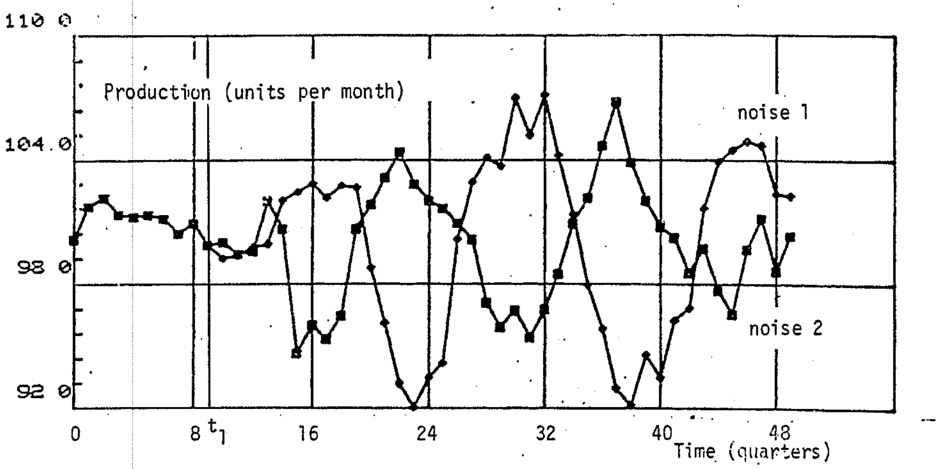

Figure 5 Inventory/workforce model output with two different noise seeds.

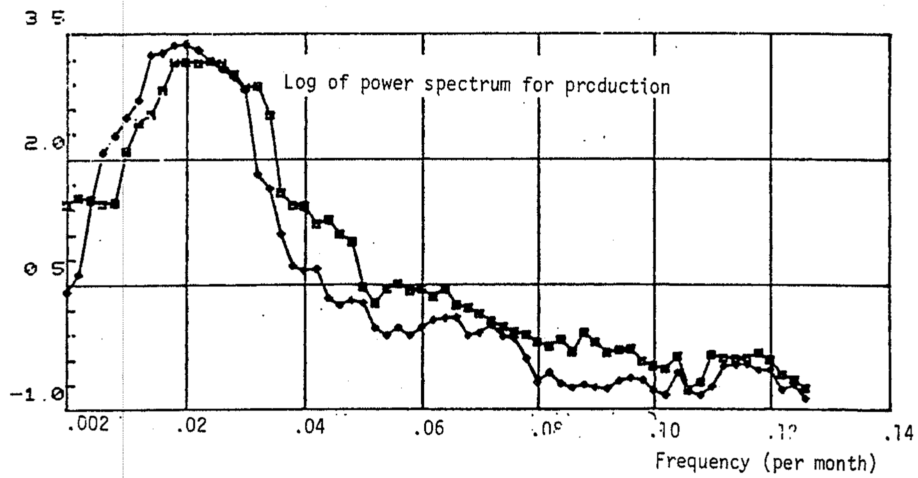

Figure 6 Power spectra for inventory/workforce model output shown in figure 5

8.0 months and the time to adjust workforce from 6.5 to 8.4 months. This changes the damping ratio from .73 to .97. The power spectra for the two models are, however, quite similar as can be seen in Figure 7.

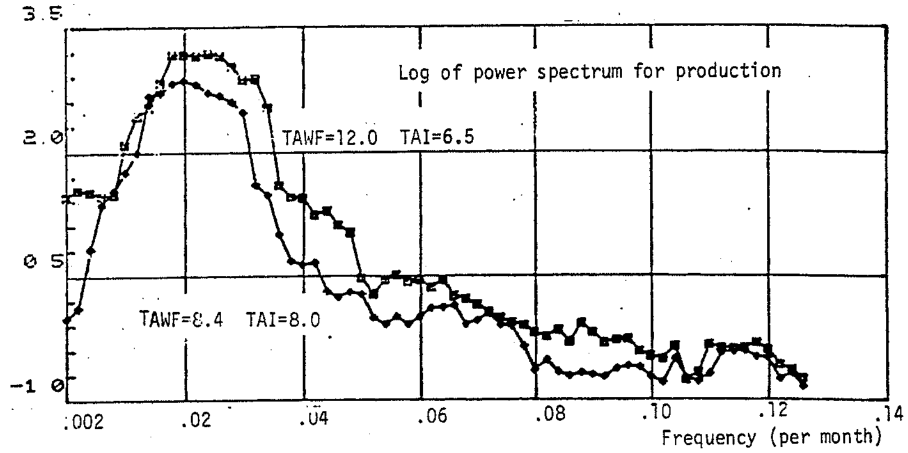

Figure 7 Inventory/workforce power spectra for two values of adjustment times

There are two ways to get around the problem of different models generating very similar output. The first is to perform a more quantitative analysis of the spectrum. It is possible to construct statistical tests for the equivilance of two power spectra over limited frequency ranges [11]. The alternative approach is to evaluate the model in terms of its ability to explain the actual data. The major disadvantage of the first approach is that it requires the specification of a selection criterion at an early stage in model evaluation. The second is somewhat more difficult to implement, but has the advantage of being useful in a number of different contexts. It is the second which we will pursue.

Model Prediction Residuals

In Industrial Dynamic [7] Forrester points out that two models with identical structure can, when subjected to different noise inputs, display substantially different time paths. We have seen above that in the frequency domain these differences are lessened. But we have also seen that in the frequency domain the differences that actually exist in a model can also be hidden. In order to get around this problem it is useful to consider the output in the time domain and consider the question: Is there a time interval over which the point by point comparison of model output is informative? The answer to this question is yes and the reason for this is simple. If two models are doing the same thing up to time t, then the models will probably be very close right after time t. This is true because any noise entering in must be integrated before it will have any effect on the system. In the noise runs of Figure 5 the noise entering was identical to the point t_1. Close inspection will show that shortly after t_1 the two model output paths are quite close. It is only later that the divergence occurs.

Thus while it is true that System Dynamics models are not

good point predictors, they can do quite well for a short interval of time. This is generally true, and forms the basis for generating many useful statistics on model performance. We need to consider the ability of a model to predict conditions in the very near future. The closer we are to the last point of model/system coincidence the greater the model's ability to match the actual system state.

The problem of determining the short term prediction error can be stated as follows: If we have a set of predicted system states which are the best possible at time t , and if we have an observation at t+d of a number of the states, then what is our best prediction at t+d of all of the states? The obvious solution is to integrate the model of the system from t to t+d and look at the output. If the predicted outputs are the same as the observed then the model states at time t+d are probably correct. If, however, the two series diverge, then it is likely that the model at time t+d has incorrect estimates of some of the states.

We can use the difference between the predicted and observed output to tell us which way the states are off. If, for example we observe that inventory is 10 and the model predicted 9 then the inventory level should probably be adjusted upward. In addition the disparity between predicted and actual inventory levels might indicate that the workforce level was predicted incorrectly. The question of how much to adjust which level was answered by Kalman in the now famous Kalman filter [12].

The above outlines the technique used for generating the prediction errors. The errors thus generated can be used for other purposes and form the basis for the estimation techniques as discussed by Peterson [5]. This approach of generating prediction errors is necessary for any rigorous statistical treatment of model behavior. Unless the state is updated in such a manner comparison of model and system output is of limited use. As Richardson [1] and Forrester [7] have pointed out, a constant will often, if not always, do better than a simulated model in predicting exact states.

The SAR Test

We have seen that different models can generate similar spectra. The purpose of the filtering of the data to get prediction residuals is to bring out more fully the differences between models or between a model and a system. If the wrong model parameters are used to explain the data then how do the residuals reveal these errors in the model parameters? As in the case of the model output the best way to analyze error characteristics is in terms of their spectral density. This technique will hereafter be referred to as the spectral analysis of residuals or SAR test.

The SAR test can be performed to judge the ability of a model to explain a data series. In Figure 7 we saw the power

spectra for two workforce/inventory models under different adjustment time choices. Calling the model with TAWF=12 and TAI=6.5 the "true model" we can consider the ability of the incorrect model to explain the true model. Using the model with the wrong parameters to explain the data series generated by the correct parameters we have used the Kalman filter to generate prediction residuals. The power spectrum of the residuals is shown in Figure 8. The residuals using the correct model to explain model output also have their spectrum plotted in figure 8. The two spectra show the difference between the two models. Using the incorrect model to explain the data yields residuals which show strength near the frequency of interest. This is an indication that the model has failed to explain some component of this behavior mode.

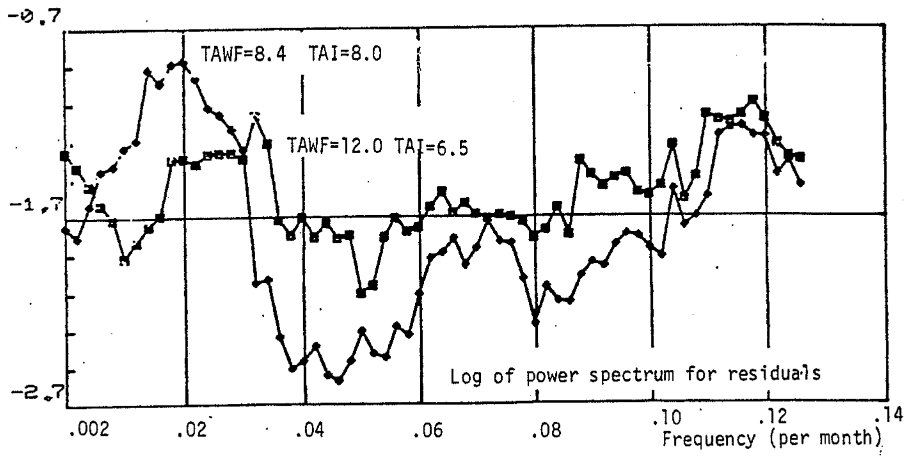

Figure 8 Power spectra for residuals when the inventory/workforce model is used to explain output

To summarize, implementating the SAR test requires

1) The existence of a model.

2) Using the model to get the best one observation ahead predictions for observed time series.

3) Creating residuals by comparing the predicted time series with the observed time series.

4) Considering the power spectrum of the residuals.

Evaluation of the SAR Test

We have developed a method for looking at a model's ability to reproduce an oscillatory behavior mode. The method is simple, and is informative in situations which might naturally be expected to be encountered. In addition, the technique has avoided the obvious problems that simpler techniques might encounter. When is this model evaluation technique capable of yielding information about model performance? The characteristics of the SAR test were evaluated primarily through synthetic data experiments using variants of the workforce inventory model. For details on the types of tests performed the interested reader is referred to [13]. The results of this evaluation are summarized below.

There are four situations in which a model is likely to fail to reproduce the observed dynamics, or to mislead the modeler with regard to what the system is actually doing. These situations are certainly related and the distinctions are drawn

essentially for convenience of discussion.

1) Though the model is accurately reflecting the processes of concern, the noise process effecting the system is different from the noise process modeled.

2) The model has the wrong set of parameters.

3) The model and the processes of concern are of two different orders.

4) The model may be representing processes which are not in fact active in the real world.

These four cases will be taken up in order.

Ideally models should explain behavior with noise acting only to excite or add energy to models. Unfortunately this is not always the case. The noise entering a system often has dynamic characteristics of its own and these dynamic characteristics can be transferred to the system output yielding substantially more complicated dynamics than the system of interest would generate. This situation arises because it is necessary to limit model boundaries in order to say something interesting. Treating certain processes as noise for the purpose of a model is a valid modeling technique. It does not, however, guarantee that the system and model dynamics will match in every respect.

The SAR test is insensitive to the characteristics of the noise entering the system in the following sense. If the dynamic characteristics of the noise are different from those of the system, then the SAR test will not indicate the presence of system dynamics in the residuals. To make this clear suppose that the noise entering a system has an annual cycle. Then the residuals will contain this annual cycle. As long as the model does not have an annual cycle as a reference mode, the SAR test will favor the model.

The case in which the system and noise do have similar characteristics (an annual cycle in the above example), is more dificult. There is an essential sense in which the noise and system cannot be separated or identified. The SAR test cannot tell that this is the problem, but combining the SAR test with other tests might yield more information.

Models may have wrong parameters, but be correct in other respects. This is the situation in the plots of Figures 6 and 7. The two models are the same except for the choice of parameters. The spectra of the residuals are distinct. The technique is able to distinguish models with wrong parameters from those with correct parameters. This is true of both the special case considered above in which the model has the same natural frequency as the data and the cases in which the two frequencies differ.

Models of processes are normally of a lower order than the processes themselves. Lower order models fill the need to have relatively simple and understandable models. On the other hand,

the effect of this approach to modeling is to have models with substantially simpler dynamics than the processes they represent. The problem in model validation becomes one of showing that the model reproduces the dynamics of interest. This problem is intensified because it is desirable to show that the dynamics can be produced not only by a simple model, but by a simple model with a counterpart in reality. The reproduction of dynamics by simple models can always be accomplished by the modal decomposition of a system [14]. However this decomposition will not necessarily yield a model which has an interpretable state space representation.

The application of the SAR test to models attempting to explain higher order behavior is quite informative. The residuals tend to be lacking power in the frequencies which the model is designed to explain. There can, of course, be some problems if there are two mechanisms generating similar dynamics. Under these circumstances unless both mechanisms are incorporated into the model the SAR test will always indicate that something is wrong. Such a situation is very similar to the case in which the system is being driven by noise which is dynamically similar to the system output. The separation of the two processes may not be possible.

The final case which needs to be considered is a situation in which a model is attempting to explain the behavior of a system on the basis of the wrong relationship between variables.

An example of this would be the use of the multiplier/accelerator model to explain inventory/workforce dynamics. Under the appropriate parameter choice the two models can be constructed to have the same inherent dynamics. The two models are using different transmission mechanisms and different state variables to explain the same process.

It is not possible to distinguish between two such models using a SAR test or any other simple analysis of model fit. It is necessary to look much more closely at the interactions between the states in this situation. The reason for this is that there are an infinite number of state space representations of the same process. All are capable of generating the dynamics, but normally only a few have any meaningful interpretation. The discrimination between these few will require either more data or a closer look at the available data. Such things as the implied phase relationships between state variables can potentially yield information about the validity of a model.

Extensions of and Alternatives to the SAR Test

This last problem helps to point the way for future research in the area of statistical validation of dynamic models. It is clear that the use of a single time series can yield only a limited amount of information. Normally models have a number of observable variables associated with a given behavior mode, and the simultaneous analysis of all of these seems appropriate. The procedures developed in this paper do have some applicability to

the higher dimension cases.

The most obvious extension of the SAR test is to conduct the same test in higher dimensions. The consideration of phase relationships can be equally well accomplished in the frequency domain by considering cross spectra. Cross spectra yield information about the phase shift and the strengh of coupling of two time series at different frequencies. The cross spectra for different model outputs can be compared to the observed cross spectra of the data. As with the single spectrum it is probably misleading to look at only model generated output. The use of the prediction residuals can again be helpful in the determination of model validity. The spectral analysis of prediction residuals gets at the notions of period, damping and phase relationships. These are all important elements in the model validation process.

A related but distinct way [15] of analyzing the dynamic relationships between different variable is in terms of their Granger causality [16]. The techniques developed by Sims [17] for determining causality are easy to implement. A test of Granger causality of variable A on variable B essentially tests whether variable A contains information not in variable B which will help to predict future values of variable B. If variable A does contain such information then variable A is said to Granger cause variable B. This type of test could be run on model output when the model is excited by noise inputs.

The notion of prediction residuals is probably not appropriate for this test.

There are a wealth of techniques for dealing with time series [18]. However, when used in isolation many of the techniques in the literature on time series seem to lack the ability to inform us about reality . But these techniques, when used in the context of understanding how a model relates to the system of interest they can be quite useful. ARIMA and vector autoregressive models constitute the natural reduced form models against which a dynamic structural model can be tested [19]. In addition such "black box" models can potentially form the basis for judging structural models in terms of their dynamic characteristics. One possible technique would be the consideration of the ability of the structural model to account for the modes of interest in the reduced form model.

Conclusion

In this paper we have outlined and evaluated a simple technique for analyzing the ability of a model to reproduce an oscillatory behavior mode. This technique appears to yield a substantial amount of information about the performance of a model. However the technique breaks down if the observed behavior is a result of the system being subjected to shocks with similar dynamic characteristics to the system output or if the system has more than one set of mechanisms generating the behavior. The SAR test is not capable of distinguishing between

models that can explain the behavior equally well using different state space representations.

The SAR test should be considered as one in a series for evaluating mdel performance. Failure of the model when the SAR test is applied is a strong indication of problems. Good performance under the SAR test is yet another forward step in the long validation process.

References

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

[2] Nordhaus, W.D. "World Dynamics: Modeling Without Data," The Economic Journal, 83(1973):1156-83

[3] Senge, P.M. "Statistical Estimation of Feedback Models," Simulation, 28(1977):177-84

[4] Peterson, D.W. Hypothesis, Estimation and Validation of Dynamic Social Models: Energy Demand Modeling, PhD thesis, MIT, 1975

[5] Peterson, D.W. "Statistical Tools for System Dynamics," in J. Randers, ed Elements of the System Dynamics Method, MIT Press Cambridge, Mass. 1980, pp. 224-45

[6] Sterman, J.D., "Appropriate Summary Statistics for Evaluating the Historical Fit of System Dynamics Models," 1983, presented at this conference.

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

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

[9] Bloomfield, P. Fourier Analysis of Time Series, John Wiley and Sons, New York: 1976

[10] Senge, P.M. The System Dynamics National Model Investment Function: A comparison to the Neoclassical Investment Function, PhD Theses, MIT Sloan School of Management, 1978

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

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

[13] Eberlein, R.L. "Testing an Aspect of Model Performance Using Spectral Analysis," Unpublished, 1983

[14] Perez Arriaga, J.I. Selective Modal Analysis With Applications to Electric Power Systems, PhD Theses, MIT Department of Electrical Engineering, 1981

[15] Sargent, T.J., Macroeconomic Theory, Academic Press, New York:1979

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

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

[18] Box, G.E.P. and G.M. Jenkins, Time Series Analysis Forecasting and and Control, Holden-Day, San Francisco: 1976

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