Employ the Principle Components in the Detection of Feedback

Dr. Heyam A.A Hayawi
Asst. Prof.
Dept. of Statistics and Informatics, College of Computer Sciences and Mathematic, University of Mosul/Iraq
heyamhayawi@gmail.com

Abstract. In linear dynamic feedback models, the relationship between two input series input and output shows that the approach is a propose to discover a feedback in linear dynamic system through examining autocorrelation function and partial autocorrelation function for principle components by using few tests of time series identification and exact using Ljung-Box test depending on Simulation approach to show the efficiency of the suggested approach and application on linear and nonlinear models within and without feedback. The obtained result is good and encourage.

Keyword: principle Component, State Space, Feedback

1. Introduction. A model is defined as a hypothetical description, an easy representation, or a description designed for a particular process entity. Scientifically, the model is a mathematical or logical representation of a system of entities, phenomena, or processes. The system is influenced by external influences processed by the user and called inputs or triggers, including non-controlled signals known as disturbances or disturbances and the relationship between them is determined by the conversion function that reflects the change in the Acetate to turn into outputs [3].

The control system is described as open-loop. If the control work is independent of the output, otherwise it is closed-loop. In the case of closed-circuit control systems, the inputs to the system are based on a synchronous basis with output observations such as autopilot and aircraft control in the air.

Modeling is the process of building models that summarize the conceptual or graphical phenomena of a particular entity or system. Useful, reliable and insightful and able to distinguish whether the model reflects the truth and deals with deviations between theory and data[4].

Examining the data to ensure the presence of feedback between more than one time series, i.e. the input chain and the output chain, is necessary to know the appropriate kinetic model that should be used to obtain the best estimates of the parameters of the model. This measure is then used to identify the inputs that subsequently reduce the error and then improve the performance. Undoubtedly, feedback is a way of describing and understanding the indirect interference found in specific physical systems. The idea of reverse feeding can be illustrated in Figures (1) and (2) as follows:
Disturbances

In this research, the main components method and the partial and subjective self-correlation analysis were used for the output data series as well as the data of the first main component to detect the presence of the feedback or not by adopting the simulation method with linear and non-linear models.

a. **AutoCorrelation Function (ACF)**
   The self-correlation coefficient is the statistical key in the analysis of the time series because it represents a measure of the correlation strength between the observations of the same random variable at k time periods and is calculated as shown in equation (1) as follows [11],[9]:
   \[
   r_k = \frac{\sum_{t=1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}, \quad k = 0,1,2,\ldots,(n-1)
   \]  
   \(\text{(1)}\)

b. **Partial AutoCorrelation Function (PACF)**
   It is used to measure the correlation between Yt and Yt-k string values with the Yt value constant for the rest of the periods, ie to measure the degree of correlation between Yt and Yt-k by determining the effect of other values at other displacements and is calculated as given in equation (2) as follows: [11]

![Diagram of a feedback system in a motor system](image1)

Figure (1): A feedback system in a motor system

![Diagram of a system](image2)

Figure (2): Lack of a feedback system in a motor system
2. Principal Components Analysis. The method of analyzing the main components is an exploratory method and is preferred to simplify the description of a group of related variables that are treated equally. Be the number of original variables So that each major component is a linear structure of the original variables, and these components are eligible to explain most of the total variance, so they are arranged in descending order according to their variability, that is, what the first major component explains more than the second main component and the information that the second main component explains is greater than the main component. Third and so on to the rest of the other major components. [1], [5] Key Component Analysis is a tool for coordinate Axis Transformation and Dimensionality Reduction, in which a new set of coordinates is calculated by maximizing the contrast of the sample data points with those coordinates. [12] In order to clarify and simplify the concept of the main components, two variables are taken and let X1 and X2 be present with N of observations, as the data of the two variables are converted to standard data by subtracting the sample mean from each view to convert to data with zero mean and variance respectively: - [1]

\[ x_1 = X_1 - \bar{X}_1 \]
\[ x_2 = X_2 - \bar{X}_2 \]  

The basic idea of creating the main components is to obtain two new variables, C1 and C2, and each of them are linear functions of the two variables respectively, that is:-

\[ C_1 = a_{11}x_1 + a_{12}x_2 \]
\[ C_2 = a_{21}x_1 + a_{22}x_2 \]  

As: The coefficients for the main components, i.e. the eigenvalues, represent Eigenv Values, and the mean of the two main components C1 and C2 is: -

\[ mean \ C_1 = mean \ C_2 = 0 \]  

While contrasting

\[ Var \ C_1 = a_{11}^2S_1^2 + a_{12}^2S_2^2 + 2a_{11}a_{12}rS_1S_2 \]
\[ Var \ C_2 = a_{21}^2S_1^2 + a_{22}^2S_2^2 + 2a_{21}a_{22}rS_1S_2 \]  

As:

\[ S_1^2 = Var \ X_1 \]
\[ S_2^2 = Var \ X_2 \]

The main component parameters are chosen to fulfill three requirements: -
1- Variation of the first component, C1, is as large as possible.
2- The values of observations in the major components C1 and C2 are not related.
3. \( a_{11}^2 + a_{22}^2 = a_{21}^2 + a_{22}^2 = 1 \).

3. **Methods of detection of feedback.** There are many methods for detecting feedback, including:

1. **Cross Correlation Function:** The cross correlation function is an important analytical tool that shows the strength of the link between two observations, each of which belongs to a separate time series from the other, the first is called the input chain and the second is called the output chain, and the cross-link measures the relationship between the current values. For the output chain and between the past and current values of the input chain, when modeling kinematic systems the input and output chains are refined and then the cross-link between the two purified series is observed and the values of the cross-link function are observed, so if the order function appears the cross-over between the two purified series is at least one significant value at a given displacement, indicating a reverse between the two series [16].

2. **ARMA Vector Operations:** ARMA model vector operations are an extension of single variable ARIMA models because they contain feedback between two or more series. To illustrate these models, we assume we have two series and that each series is a linear equation for its previous values at displacement 1 and previous values for another series at displacement 1 also as well as errors. Randomness is written as shown in equation (8) as follows [13],[8]:

\[
Y_{1,t} = \phi_{11}Y_{1,t-1} + \phi_{12}Y_{2,nl} + a_{1,t} \\
Y_{2,t} = \phi_{21}Y_{1,t-1} + \phi_{22}Y_{2,nl} + a_{2,t} \quad \ldots \ (8)
\]

Since: \( \phi \) The coefficients of the current values of \( Y_{1,t} \) and with the previous values \( Y_{2,t} \) of the two series \( (a_{1,t}) \) and \( (a_{2,t}) \) are represented, and independent random errors with a mean of zero and the variance \( (\sigma_1, t) \) respectively, and equation (8) represents the \( (\sigma_2, t) \) operations of the self-regression vector (AR). Any multiple AR model is used to detect feedback.

4. **Diagnostic tests.** One of the most important diagnostic tests is the portmanteau test, where it is known that the test statistic is used to determine whether the residual series of the model is white noise (WN) or not, and is called the portmanteau statistic (portmanteau statistic) and it tests the following hypothesis: [14]

\[
H_0 : \rho_1 = \ldots = \rho_m = 0 \\
H_a : \rho_k \neq 0 \quad ; k \in \{1, \ldots, m\} \quad \ldots \ (9)
\]

In general, model adequacy tests based mainly on this statistic are called portmanteau tests[7]. Initial tests of this nature, such as the B-P and the L-B, have been shown to be ineffective[6]. It was mentioned[11] that these two tests sometimes fail to reject poorly-matched models of data, and care must be taken to accept a model based on portmanteau tests only. Because the portmanteau test statistic is affected by the number of self-correlation coefficients (m) used in its calculation, an increase in the number of self-correlation coefficients can lead to the acceptance of the null hypothesis [15],[2].

**Ljung-Box Test (L-B)**

In 1978, researchers Ljung & Box made a simple modification to the B-P test, based on what was observed by Box & Pierce that rk is distributed according to the normal distribution with a mean of zero and the variance of its amount \( (n-k)/[n(n+2)] \), i.e.:[10].

\[
r_k \sim N \left( 0, \frac{(n-k)}{n(n+2)} \right) \quad \ldots \ (10)
\]
As:
\[
\sqrt{\frac{n(n+2)}{n-k}} r_k \sim N(0,1)
\]  \hspace{1cm} (11)

By squaring equation (11), we get:
\[
\frac{n(n+2)}{n-k} r_k^2 \sim \chi_i^2
\]  \hspace{1cm} (12)

Taking the sum of m from the self-correlations after estimating the parameters q, p of the model, then:
\[
\left[ n(n+2) \sum_{k=1}^{m} \frac{r_k^2}{(n-k)} \right] \sim \chi^2_{m-p-q}
\]  \hspace{1cm} (13)

This is known as the L-B or QL-B statistic:
\[
Q_{L-B} = n(n+2) \sum_{k=1}^{m} \frac{r_k^2}{(n-k)} \sim \chi^2_{m-p-q}
\]  \hspace{1cm} (14)

Researchers Ljung & Box found that this modification of BP test gave them a statistic closer to the $\chi^2$ distribution of BP stats because their mean levels were close to theoretical values [7] and it was observed to give better results at small sample sizes [17]. Where the calculated values $Q_{L-B}$ are compared with the values of the tabular $\chi^2$ with a degree of freedom $(m-p-q)$ in the case of studying the rest of the model. If the calculated value is less than the tabular value, this leads to the conclusion that the remaining series are white noise, and vice versa.

5. Simulation experiences. The simulation method has wide uses in all fields, and its importance has emerged after the rapid development in providing applications (ready software). Simulation can be defined as a process of simulating the actual reality of real models, whether this simulation was done manually or by computer and it is a mirror of some aspects of the real world and appears dependent on random processes (Stochastic) [3]. In this research, simulation experiments were conducted on two models, one Linear and the other nonlinear, as observations were generated with four different volumes of observations for both models (n=50,100,250,50) view and the Ut model inputs were random signals generated from the standard random Gaussian signals distribution "rgs". As for jamming, random signals were generated that follow the normal distribution. The following linear model was used without back-up:
\[
y_u(t) = 0.7*y(t-1) + e(t)
\]  \hspace{1cm} ... (15)

And the following linear model is feedback:
\[
y_f(t) = 0.7*y(t-1) + 0.3*y(t-1) + e(t)
\]  \hspace{1cm} ... (16)

While using the non-linear model and without feedback as follows:
\[
y_u(t) = (1 + 0.5*\exp(-0.6*u(t-2)^2)) + e_t
\]  \hspace{1cm} ... (17)

The following non-linear model is feedback by:
\[
y_f(t) = (1 + 0.5*\exp(-0.6*u(t-2)^2)*0.8*y(t-1) + e_t
\]  \hspace{1cm} ... (18)

The experiment was repeated 1000 times and using the ready program MINITAB and using the macros (MACROS). The self-link and partial self-function of the original series of outputs were examined and then found the main components of the two series as well as examining the two functions above for the first main component using the LB test at a level of significance of 0.05 and the number of ACF and PACF equal For a quarter of the sample size and since it is expected that as long as the first major component explains most of the
variance and reflects the strength of the link between the original variables, then in the case of a reverse feed between the original variables (the input and output chains), it is expected that the self-linkages and sequences S partial self In this case, whether for the output chain data or the first main component data, these associations are significant, leading to rejection $H_0$ in equation (9) and acceptance $H_1$ which reveals the existence of the feedback and in return is accepted $H_0$ in the absence of a feedback.

6. Simulation results. Applied to the output chain of the linear model and with the presence of the feedback in the model and when ($n = 50$) it was found that the Ljung-Box test statistic revealed the significance of the self-correlations by 99.8%, i.e. rejecting the null hypothesis which states that the values of the self-correlations are not significant as well as revealed the LB statistic. The significance of the self-correlations of the first major component is 95%, which indicates that it is possible to detect the feedback of the original data series, which is better than it is when finding the main components of the data. However, when the sample size increases, i.e. ($n = 100,250,500$) views), it turns out that the LB statistic revealed the significance of the self-correlations of the first component by 100%, while the ratio with the original output chain at the above sizes was (93.7%, 95%, 95). Respectively, these results provide a probability for the first major component when employed to detect the presence of feedback. This indicates that revealing the feedback using the first main component of the main components is better than the original data series.

Tables (1) and (2) show the percentages of the significance of the correlation functions of the linear model and the non-linear model and with the presence of the opposite or not as follows:

| Sample Size | The nature of data | The first major component | The first major component | The first major component | The first major component |
|-------------|-------------------|--------------------------|--------------------------|--------------------------|--------------------------|
|             | Output series     |                          |                          |                          |                          |
| N=50        |                   |                          |                          |                          |                          |
| Linear model without feedback | 6.9% | 7.1% | 13.9% | 14.1% | 22.9% | 22.2% | 23.0% | 22.2% |
| Linear model with feedback | 99.8% | 95.0% | 93.7% | 100% | 95.0% | 100% | 95.0% | 100% |
| Non-Linear model without feedback | 7.6% | 8.1% | 15.1% | 14.5% | 22.5% | 22.0% | 22.0% | 22.5% |
| Non-Linear model with feedback | 98.5% | 72.0% | 100% | 97.7% | 98.5% | 72.5% | 98.5% | 72.5% |

By examining the results in table (1) above, it is clear that:

1. Relying on the L-B test in examining the significance of the self-correlation function, whether for the original output chain data or for the data of the first major component, gave very encouraging results in its ability to distinguish between the models used in terms of the presence of the feedback in its movement or not.

2. On the other hand, it is also noted that there is a strong indication that the use of the first major component with the linear model in the presence of the feedback gave very strong results in the detection of the presence of the feedback as the size of the sample
increased using the proposed method than with the non-linear model with the feedback.

Table (2): The percentages of the significance of the partial self-correlation function

| Sample size | N=50  | N=100 | N=250 | N=500 |
|-------------|-------|-------|-------|-------|
| The nature of data | Output series | The first major component | Output series | The first major component | Output series | The first major component |
| Linear model without feedback | 7.5% | 9.2% | 12.5% | 13.2% | 22.5% | 22.7% | 22.0% | 24.2% |
| Linear model with feedback | 95.8% | 93.0% | 89.7% | 95% | 93.0% | 95% | 92.0% | 93% |
| Non-Linear model without feedback | 8.6% | 10.1% | 17.1% | 18.5% | 25.5% | 28.0% | 21.0% | 27.5% |
| Non-Linear model with feedback | 95.5% | 70.0% | 95% | 97.7% | 95.5% | 75.5% | 93.5% | 70.5% |

From the results of the above table, it is noted:

1. The employment of the L-B test in examining the significance of the partial self-correlation function gave encouraging results in its ability to distinguish between the models used in terms of whether or not the feedback was in its movement and whether the application was with the data of the original output chain or the main component.
2. The proposed method gave very encouraging and close results despite the difference in the sample sizes in detecting the feedback in the linear model, whether by using it for the output chain data or data for the first major component.
3. Although the proposed method is presented for good results in revealing the feedback using the data of the first major component of the non-linear model of different sample sizes, the use of the output chain data for the non-linear model with the reverse feeding gave strong results in revealing the presence of the feedback.

The behavior of the linear model and the non-linear model can be observed through the self-linking and partially self-correlating functions of one of the iterations through Figures (3) and (4) as follows:
Conclusions

The research reached some conclusions, including:

The proposed method based on the L-B test gave very encouraging results in distinguishing between the models used in terms of the presence of the feedback in its mobility or not, whether it was adopted in applying the proposed method to the self-correlation coefficients or partial self-correlation coefficients.

The process of applying the proposed method based on the self-correlation coefficients of the first major component led to very strong results, especially with the linear model with the opposite feedback in addition to obtaining encouraging results when using the non-linear model.

When applying the proposed method relying on the self-correlation coefficients for the first major component, the results were very encouraging and converging with those obtained when using the original output chain data for the linear model while the results obtained with the non-linear model were more robust with string data. The original outputs with the main main component data.

The increase in the sample size in general played a positive role in raising the percentage of the ability to detect the presence of adverse reactions in the kinetic models used.

Reference

1. Afifi, A.A., (1984). "Computer Aided Multivariate Analysis", Lifetime Learning Publications Belmont, California U.S.A.
2. Alhamoshy, Wasaan Raad Thanoon, (2011):” Simulation study of the Schwartz standard of information behavior and some time series diagnostic tests”, Master Thesis, College of Computer and Mathematic Sciences, University of Mosul, Iraq.
3. AlKayat, Bassel Thanoon, (2010), "Introduction to Computer Static Simulation and Modeling Using MATLAB, Dar Iben Alhaytham For printing and publishing, University of Mosul.

4. AlKayat, Bassel Thanoon, (2011), "Markovian modeling with practical applications, Dar Iben Alhaytham For printing and publishing, University of Mosul.

5. Al-Rawi, Kasha Mahmood, (1987), "Introduction to regression analysis", Dar Al-Kutub for Printing and Publishing, University of Mosul, Iraq.

6. Arranz, M.A., (2005), "Portmanteau Test Statistics in Time Series", http://www.tol-project.org/.

7. Davies, N., & Newbold, P., (1979), "Some power studies of a Portmanteau Test of Time Series Model Specification", Biometrika, 66, pp. 153-155.

8. Fleming, W.H., (1989), "Future Directios in Control Theory A Mathematical Perspective", Report, Society for Industrial and Applied Mathematics, Philadelphia, U.S.A.

9. Kanjilal, P.P., (1995), "Adaptive Prediction and Predictive Control", Peter Peregrines Ltd, London.

10. Ljung, G.M., Box, G.E.P., (1978), "On A Measure of Lack of Fit in Time series Models", Biometrika, 65, pp. 297-303.

11. Makridakis, S., Wheelwright, S., & Hyndman, R., (1998), "Forecasting Methods and Applications", 3rd ed., John Wiley & Sons, New York.

12. Nelles, O., (2001), "Nonlinear System Identification from Classical Approach to Neural Network and Fuzzy Models", Springer Verlag Berlin Heidelberg Germany.

13. Pankratz, Alan, (1991), "Forecasting with Dynamic Regression Models", John Wiley and Sons, New York, U.S.A.

14. Tsay, R.S., (2002), "Analysis of Financial Time Series: Financial Econometrics", John Wiley & Sons, New York.

15. Vandel, Walter, (1992), "Time series from an applied point of view and Box Jenkins models", Abd al-Mardi Azzam, Arabization, Mars Publishing House, Saudi Arabia.

16. Wei, W. W. S., (1990), "Time series Analysis Univariate Multivariate Methods", Addison-Wesley Publishing Company, Inc., The Advanced Book Program, California, U.S.A.

17. Yaffee, R. A., & McGee, M., (2000), "Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS", Academic Press, San Diego.