[redacted]

IDENTIFICATION OF NONLINEAR OUTPUT-AFFINE SYSTEMS USING AN ORTHOGONAL LEAST SQUARES ALGORITHM

S.A. Billings[+], M.J. Korenberg[*] and S. Chen[+]

[+] Department of Control Engineering

University of Sheffield

Mappin Street

Sheffield S1 3JD

[*] Department of Electrical Engineering

Queen's University

Kingston

Canada

Research Report No.313

March 1987

IDENTIFICATION OF NONLINEAR OUTPUT-AFFINE SYSTEMS USING AN ORTHOGONAL LEAST SQUARES ALGORITHM

Abstract

The identification of a nonlinear output-affine difference equation model is considered. An orthogonal least squares algorithm is derived which can determine which term to regress upon, detect significant terms in the model expansion and provide the final parameter estimates. It is shown that the orthogonal property of the algorithm results in a particularly simple estimation routine and simulated examples are included to illustrate the techniques.

1. Introduction

The output-affine difference equation model which relates sampled output signals to sampled inputs was introduced by Sontag (1979) following a study of polynomial response maps for nonlinear systems. The advantage of the output-affine model is that it is globally valid. However, the lack of practical identification algorithms for this model has to date limited its usefulness in control.

The internal behaviours of a wide range of nonlinear discrete-time systems are governed by state-affine models (Sontag, 1979). Cyrot-Normand and Dang Van Mien (1980) demonstrated that state-affine models can be constructed by combining several linear state-space models identified from different plant operating points and input variations. Several processes were modelled using this approach and the technique was shown to work well for slowly varying levels of plant operation.

In the present study, the least squares method is used as a basic identification tool for obtaining estimates of the unknown parameters in an output-affine difference equation model based on input-output data measurements. To overcome the excessive computational burden, an orthogonal estimator (Korenberg, 1985; Korenberg, Billings and Liu, 1986) derived originally for the NARMAX model (Leontaritis and Billings, 1985) is extended to the output-affine case which is considerably more complex because the term to regress upon has to be determined prior to final estimation. The technique provides both the important structure information and the final parameter estimates in a very simple and efficient manner.

After a brief description of the output-affine difference equation model, the problems associated with using the least squares method are discussed, and an orthogonal regression procedure is introduced. Finally, it is shown that bias occurs using the least squares method whenever outputs are corrupted by noise. Further research is proposed in order to overcome this difficulty.

2. Output-affine difference equation model

Let U be the input set and Y be the output set. Denote as U^+ the union of all U^t, $t \geq 1$, where U^t is the set of all sequences (u(1),...,u(t)) of length t

$U^t = ((u(1),...,u(t)): u(i) \in U, i=1,...,t)$

A polynomial input-output map or response function is a function $f: U^+ \rightarrow Y$ such that, for each t, $f_t: U^t \rightarrow Y$ is a polynomial of finite degree on all variables, although this degree may tend to ∞ as $t \rightarrow \infty$. A strictly causal response function is such that, for each t, $f_t(u(1),...,u(t))$ is independent of u(t). f is said to be bounded if there exists an integer d such that, for all t, the maximum power which any individual input is raised to in f_t is less than d. f is finitely realizable iff f has a finite dimensional realization.

Sontag (1979) proved that f is finitely realizable and bounded iff it satisfies the affine difference equation

$$\sum_{i=0}^{r} a_i(u(t-1),...,u(t-r))y(t-i) = a_{r+1}(u(t-1),...,u(t-r)) \tag{1}$$

where r is the order of the system, and a_i, $i=0,1,...,r+1$ are polynomials. The results can be generalised to the multi-input, multi-output case but for simplicity, only the single-input, single-output case is considered in this paper.

The output-affine difference equation (1) is an input-output model valid everywhere (Sontag, 1979). Its potential application, therefore, is very broad. The obvious disadvantage of using eqn. (1) in the identification of nonlinear systems, however, is that there is no simple explicit recursion in y(t). This may explain why so few results have been achieved in this area.

3. Least squares estimation

In order to apply the well-known least squares estimation method, a linear-in-the-parameters expression must be derived. Expanding each a_i as an L degree polynomial, keeping only one term of $a_0(u(t-1),...,u(t-r))y(t)$ on the left hand side (LHS) of the equation and moving the rest of the terms to the right hand side (RHS) yields

$$Y(t) = \sum_{i=1}^{M} p_i(t)\theta_i + \xi(t) \tag{2}$$

where θ_i, $i=1,...,M$ are unknown parameters to be estimated and $\xi(t)$ is some modelling error. For example, consider a second order system with a linear expansion for a_i

$$(c_{00}+c_{01}u(t-1)+c_{02}u(t-2))y(t) + (c_{10}+c_{11}u(t-1)+c_{12}u(t-2))y(t-1) + (c_{20}+c_{21}u(t-1)+c_{22}u(t-2))y(t-2)$$
$$= c_{30}+c_{31}u(t-1)+c_{32}u(t-2)$$

Inspection of this model clearly illustrates that before any estimation routine can be applied the term to regress upon must be determined since in general c_{00} may not exist and there is no strict recursion in y(t). This creates a significant problem for parameter estimation.

Assume that $u(t-1)y(t)$ is chosen as $Y(t)$, the term on which to regress upon, then

$$\theta_1 = \frac{c_{30}}{c_{01}},\ p_1(t) = 1;\ \theta_2 = \frac{c_{31}}{c_{01}},\ p_2(t) = u(t-1);\ \theta_3 = \frac{c_{32}}{c_{01}},\ p_3(t) = u(t-2);$$

$$\theta_4 = -\frac{c_{00}}{c_{01}},\ p_4(t) = y(t);\ \theta_5 = -\frac{c_{02}}{c_{01}},\ p_5(t) = u(t-2)y(t);\ \theta_6 = -\frac{c_{10}}{c_{01}},\ p_6(t) = y(t-1);\ etc.$$

The total terms in the expansion of each a_i are

$$n = \sum_{i=0}^{L} n_i \tag{3}$$

where $n_0 = 1$, $n_i = n_{i-1}(r+i-1)/i$, $i=1,...,L$. Therefore, there are n possible candidates for $Y(t)$. For each expression eqn. (2), the total number of unknown parameters is

$$M = (r + 2)n - 1 \tag{4}$$

TABLEs 1 and 2 give values of n and M for some typical values of r and L. It is seen that n and M increase rapidly as the system order r and the degree L of the polynomial expansion increase.

The immediate difficulty is however which term of $a_0(u(t-1),...,u(t-r))y(t)$ should be chosen as $Y(t)$. There appears to be no a priori information regarding this matter. If a term is used as $Y(t)$ which does not exist in the real system, the final model fit will be useless. Such a situation can be detected only after analysing the estimation results and comparing them with the results obtained using different choices of $Y(t)$. This means that each possible choice of $Y(t)$ has to be tried first before the conclusion can be reached. It is apparent that the identification task is complicated. Many statistical tests can be used to analyse the estimation results in order to determine the correct choice of $Y(t)$. Fortunately, a simple calculation of the sum of the squared errors often provides adequate information

$$sqe = \sum_{t=1}^{N} (y(t) - \hat{y}(t))^2 \tag{5}$$

where N is the data length

$$\hat{y}(t) = \frac{1}{\hat{a}_0(u(t-1),...,u(t-r))}\{\hat{a}_{r+1}(u(t-1),...,u(t-r)) - \sum_{i=1}^{r}\hat{a}_i(u(t-1),...,u(t-r))y(t-i)\}$$

and $\hat{a}_i[.]$, $i=0,1,...,r+1$ are the estimates of $a_i[.]$.

Even when a correct term which exists in the data generating mechanism is used as $Y(t)$, the determination of the structure or which terms to include in the model is essential if a parsimonious model is to be determined from the large number of candidates TABLEs 1 and 2. Direct least squares estimation based on eqn. (2) may involve an excessive number of terms. Simply increasing the system order r and the degree L of the polynomial expansion to achieve the desired accuracy will in general result in an excessively complex model and possibly numerical ill-conditioning. An orthogonal least squares algorithm (Korenberg, Billings and Liu,

1986) can however be derived to overcome these difficulties.

4. Orthogonal regression

The orthogonal least squares algorithm (Korenberg, Billings and Liu, 1986) is much more efficient than the ordinary least squares algorithm. The strength of the algorithm lies mainly in the fact that it provides information regarding which terms in the model are significant. This is often vital in the identification of nonlinear systems. Since the detailed derivation can be found in the paper mentioned above, only the main results are quoted. An orthogonal regression procedure is then proposed for the output-affine problem outlined above.

Consider a linear-in-the-parameters expression of eqn. (2). Assume that no two $p_i(t)s$ are identical and $\xi(t)$ is a zero mean white sequence uncorrelated with $p_i(t)$, i=1,..,M. The orthogonal algorithm involves a procedure of transferring eqn. (2) into an equivalent orthogonal equation

$$Y(t) = \sum_{i=1}^{M} w_i(t) g_i + \xi(t) \tag{6}$$

where

$$\begin{aligned} w_1(t) &= p_1(t) \\ w_k(t) &= p_k(t) - \sum_{i=1}^{k-1} \alpha_{ik} w_i(t), \quad k=2,\ldots,M \end{aligned} \tag{7}$$

and

$$\alpha_{ij} = \frac{\overline{w_i(t) p_j(t)}}{\overline{w_i^2(t)}}, \quad i<j,\ j=2,\ldots,M \tag{8}$$

Here, the bar $\overline{}$ denotes time averaging. The estimates of the coefficients g_i are given by

$$\hat{g}_i = \frac{\overline{w_i(t) Y(t)}}{\overline{w_i^2(t)}}, \quad i=1,\ldots,M \tag{9}$$

The coefficients of the original equation can easily be obtained according to the formula

$$\begin{aligned} \hat{\theta}_M &= \hat{g}_M \\ \hat{\theta}_i &= \hat{g}_i - \sum_{j=i+1}^{M} \alpha_{ij} \hat{\theta}_j, \quad i=M-1, M-2, \ldots, 1 \end{aligned} \tag{10}$$

The estimates eqn. (10) are unbiased and the standard deviations of the estimates are given by

$$std(\hat{\theta}_i) = \sigma \sqrt{\frac{1}{N} \sum_{j=i}^{M} \frac{t_{ij}^2}{\overline{w_j^2(t)}}}, \quad i=1,\ldots,M \tag{11}$$

where

$$t_{ij} = \begin{cases} 1 & i=j \\ -\sum_{k=i+1}^{j} \alpha_{ik} t_{kj} & 0<i<j \end{cases} \quad j=1,\ldots,M \tag{12}$$

and $\sigma^2 = E\{\xi^2(t)\}$. An estimate of σ can be obtained from

$$\hat{\sigma} = \sqrt{\frac{1}{N-M}\sum_{t=1}^{N}(Y(t) - \sum_{i=1}^{M} w_i(t)\hat{g}_i)^2} \tag{13}$$

Because of the orthogonality

$$\overline{Y^2(t)} = \sum_{i=1}^{M} g_i^2\overline{w_i^2(t)} + \sigma^2 \tag{14}$$

The ratio of the reduction in the sum of squared errors due to the i'th term is therefore given by

$$[rer]_i = \frac{g_i^2\overline{w_i^2(t)}}{\overline{Y^2(t)}}, \quad i=1,...,M \tag{15}$$

Eqn. (15) can be used as a criterion for determining the significance of the terms in the model expansion and insignificant terms can be removed from the orthogonal equation. This is equivalent to removing corresponding $p_i(t)s$ from the original model leading to a less complex model.

However, care is required in interpreting eqn. (15). It can be shown that if the position of each $p_i(t)$ in eqn. (2) is changed a different value of $[rer]_i$ will be obtained. In general, a term in the first position of eqn. (2) will have a larger $[rer]_i$ value than that obtained by changing its position to appear later in the equation. As a result, simply orthogonalizing the $p_i(t)s$ into the orthogonal equation in the order in which they happened to be written down in eqn. (2) may produce the wrong information regarding their significance using eqn. (15). To avoid this difficuty, an orthogonal regression procedure motivated by the idea of stepwise regression (Smillie, 1966) is proposed

In the initial stage, all $p_i(t)$, $i=1,...,M$ are considered as possible candidates for $w_1(t)$. For $i=1,...,M$, calculate

$$w_1^{(i)}(t) = p_i(t)$$

$$\hat{g}_1^{(i)} = \frac{\overline{w_1^{(i)}(t)Y(t)}}{\overline{(w_1^{(i)}(t))^2}}$$

$$[rer]_1^{(i)} = \frac{(g_1^{(i)})^2\overline{(w_1^{(i)}(t))^2}}{\overline{Y^2(t)}}$$

Find the maximum of $[rer]_1^{(i)}$, say, $[rer]_1^{(j)} = \max\{[rer]_1^{(i)}, 1 \le i \le M\}$. Then the one-term orthogonal equation is selected

$$Y(t) = w_1(t)g_1 + \xi(t)$$

with $w_1(t) = w_1^{(j)}(t)$, $\hat{g}_1 = \hat{g}_1^{(j)}$ and $[rer]_1 = [rer]_1^{(j)}$.

In the second stage, all $p_i(t)$, $i=1,...,M$, $i \ne j$ are considered as possible candidates for $w_2(t)$. For $i=1,...,M$, $i \ne j$, calculate

$$w_2^{(i)}(t) = p_i(t) - \alpha_{12}^{(i)}w_1(t)$$

$$\hat{g}_2^{(i)} = \frac{\overline{w_2^{(i)}(t)Y(t)}}{\overline{(w_2^{(i)}(t))^2}}$$

$$[rer]_2^{(i)} = \frac{(\hat{g}_2^{(i)})^2\overline{(w_2^{(i)}(t))^2}}{\overline{Y^2(t)}}$$

where

$$\alpha_{12}^{(i)} = \frac{\overline{w_1(t)p_i(t)}}{\overline{w_1^2(t)}}$$

Find the maximum of $[rer]_2^{(i)}$, say, $[rer]_2^{(k)} = \max\{[rer]_2^{(i)}, 1 \le i \le M, i \ne j\}$. Then the two-term orthogonal equation is formed

$$Y(t) = w_1(t)g_1 + w_2(t)g_2 + \xi(t)$$

with $w_2(t) = w_2^{(k)}(t) = p_k(t) - \alpha_{12}w_1(t)$, $\hat{g}_2 = \hat{g}_2^{(k)}$ and $[rer]_2 = [rer]_2^{(k)}$, where $\alpha_{12} = \alpha_{12}^{(k)}$.

The procedure is terminated at the m'th stage when

$$1 - \sum_{i=1}^{m} [rer]_i < \underline{a\ desired\ tolerance}$$

where $m < M$. The final orthogonal equation is selected as

$$Y(t) = \sum_{i=1}^{m} w_i(t)g_i + \xi(t)$$

Using eqn. (10), it is straightforward to calculate the corresponding parameters θ_i in the model containing only m significant terms.

The above procedure has advantages over stepwise regression. Because of the orthogonality, before adding the i+1'th term to the orthogonal equation, there is no need to remove the effect of the i'th terms. In stepwise regression, a term that was significant at an earlier stage may become insignificant after several other terms are included in the model. To overcome this difficuty, before adding a new term, those terms that are already in the model are tested for their significance using some statistical test (Billings and Voon, 1986). If a term is found to become insignificant it is deleted from the model. Such a precaution is not necessary in orthogonal regression.

In orthogonal regression, it is not required to remove the mean levels. This is desired because removing mean levels from signals when estimating nonlinear models can change the structure of the model, and will almost always induce input sensitivity (Billings and Voon, 1984).

5. Simulation study

Two simulated examples are used to illustrate the application of the orthogonal regression algorithm to output-affine systems.

Example 1

$$[u(t-1)+u(t-2)]y(t)=0.2u(t-1)y(t-1)+0.7u(t-1)u(t-2)$$

200 pairs of data were generated using the input $u(t)=sin\{(\pi/6)t\}$. An output-affine model with $r=2$ and $L=2$ was used to fit the data. Depending on the choice of $Y(t)$, the total possible linear-in-the-parameters models of eqn. (2) is 6. Each such model has 23 parameters in the right hand side of the expression. Each expression is fitted using the orthogonal regression algorithm. The results are given in TABLE 3 where the order in which the selected terms appear is according to the order in which they were actually selected. Comparing *sqe* values in each case there is little doubt that the correct choice of $Y(t)$ is either $u(t-1)y(t)$ or $u(t-2)y(t)$ which correspond to the existing terms $a_0(u(t-1),u(t-2))y(t)$ in the real system. Both choices of $Y(t)$ produce the correct estimates of the real system parameters.

Example 2

$$u(t-2)y(t)=[u(t-1)+u(t-2)]y(t-1)-u(t-1)y(t-2)+u(t-1)u(t-2)$$

A uniformly distributed random variable with amplitude range ±0.5 was used as input to generate 200 pairs of data. An output-affine model with $r=2$ and $L=2$ was used as a candidate model. The results obtained using the orthogonal regression algorithm are shown in TABLE 4. It is clear that the correct choice of $Y(t)$ is $u(t-2)y(t)$ which is the only term of $a_0(u(t-1),u(t-2))y(t)$ in the real system. The model obtained in this case is a good approximation to the real system. Because the estimated coefficient of $u(t-1)y(t)$ is so near zero, it is reasonable to test whether this term actually exists. Restarting the regression with this term deleted gives the final estimated model

$$u(t-2)y(t)=1.0u(t-2)y(t-1)+1.0u(t-1)u(t-2)+1.0u(t-1)y(t-1)-1.0u(t-1)y(t-2)$$

with a smaller $sqe = 3.526\times10^{-6}$. This confirms that the coefficient of $u(t-1)y(t)$ is indeed zero.

6. Noise corrupted outputs

If the system output is corrupted by noise bias occurs using the least squares method. This is because some of the $p_i(t)$s become correlated with $\xi(t)$. The extended least squares method, which has been shown to give unbiased estimates in the identification of NARMAX models (Billings and Voon, 1984), in this case will also produce biased estimates. Let

$$z(t) = y(t) + e(t)$$

where z(t) is the output corrupted by the white noise e(t). Substituting for y(t) using $z(t)-e(t)$ in eqn. (1) gives

$$\sum_{i=0}^{r} a_i(u(t-1),...,u(t-r))z(t-i) = a_{r+1}(u(t-1),...,u(t-r)) + \sum_{i=0}^{r} a_i(u(t-1),...,u(t-r))e(t-i) \tag{16}$$

Defining

$$\xi(t) = a_0(u(t-1),...,u(t-r))e(t) \tag{17}$$

and expanding the rest of polynomials yields the model

$$Y(t) = \sum_{i=1}^{M} p_i(t)\theta_i + \xi(t) \tag{18}$$

Now

$$M = (2\,r + 2)n - 1 \tag{19}$$

The number of unknown parameters is almost doubled and because $\xi(t)$ is correlated with those $p_i(t)s$ which have $z(t)$ as their component, bias occurs in the estimates.

Other alternative estimation methods, therefore, must be employed in order to obtain unbiased estimates. Rearrange eqn. (16) into

$$z(t) = \frac{1}{a_0(u(t-1),...,u(t-r))}\{a_{r+1}(u(t-1),...,u(t-r)) - \sum_{i=1}^{r} a_i(u(t-1),...,u(t-r))(z(t-i) - e(t-i))\} + e(t) \tag{20}$$

which is no longer linear-in-the-parameters. The prediction-error estimation algorithm (Goodwin and Payne, 1977; Billings and Voon, 1986) is applicable provided that some structure determination technique is incorporated with the algorithm. The extended orthogonal algorithm (Korenberg, Billings and Liu, 1986) can be modified using the idea of orthogonal regression. The resulting extended orthogonal regression procedure may then be used to detect significant terms in the model while the prediction-error algorithm provides optimal estimates of the final model parameters. Further research is required in order to develop a suitable estimation algorithm for stochastic output-affine systems.

7. Conclusions

The least squares method is employed in the identification of the output-affine difference equation model. The problem is not as simple as it first appeared. An orthogonal regression algorithm was introduced which combines structure determination with parameter estimation efficiently.

In order to obtain unbiased estimates of the stochastic output-affine system, further research is needed. A possible unbiased estimation algorithm is the prediction-error algorithm coupled with some suitable structure determination technique. This is currently under investigation. The main challenge is to develop an efficient structure determination routine to detect significant terms in the output-affine model prior to the final estimation.

It is obvious that parsimonious models of linear systems and other tpyes of nonlinear systems can be identified using the orthogonal regression algorithm.

Acknowledgments

The authors gratefully acknowledge SERC Grant Ref. GR/D/30587 for financial support for the work reported above.

References

[1] Billings, S.A., and W.S.F. Voon (1984). Least squares parameter estimation algorithms for non-linear systems. *Int. J. Systems Sci.*, Vol.15, No.6, pp601-615.

[2] Billings, S.A., and W.S.F. Voon (1986). A prediction-error and stepwise-regression estimation algorithm for non-linear systems. *Int. J. Control,* Vol.44, No.3, pp803-822.

[3] Cyrot-Normand, D., and H. Dang Van Mien (1980). Nonlinear state-affine identification methods: applications to electrical power plants. *Preprints of IFAC Symp. on Auto. Control in Power Generation, Distribution and Protection.* Pretoria, South Africa, 1980, pp449-462.

[4] Goodwin, G.C., and R.L. Payne (1977). *Dynamic System Identification: Experiment Design and Data Analysis.* Academic Press, New York.

[5] Korenberg, M.J. (1985). Orthogonal identification of nonlinear difference equation models. *Mid. West Symp. on Cts and Systems,* Louisville.

[6] Korenberg, M.J., S.A. Billings, and Y.P. Liu (1986). An orthogonal parameter estimation algorithm for nonlinear stochastic systems. *Research Report No.307,* Department of Control Eng., University of Sheffield, Sheffield, U.K. (Submitted for publication).

[7] Leontaritis, I.J., and S.A Billings (1985). Input-output parametric models for non-linear systems, Part I: deterministic non-linear systems; Part II: stochastic non-linear systems. *Int. J. Control,* Vol.41, No.2, pp303-344.

[8] Smillie, K.W. (1966). *An Introduction to Regression and Correlation.* Academic Press, New York.

[9] Sontag, E.D. (1979). *Polynomial Response Maps,* Lecture Notes in Control and Information Sciences 13. Springer Verlag, Berlin.

TABLE 1. Values of n

L	r = 1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	6	10	15	21	28
3	4	10	20	35	56	84

TABLE 2. Values of M

L	r = 1	2	3	4	5	6
1	5	11	19	29	41	55
2	8	23	49	89	146	223
3	11	39	99	209	391	671

TABLE 3. Results of example 1

$Y(t)$	RHS terms	*sqe*
$y(t)$	$160.3u(t-1)u(t-2)y(t)-32.61u(t-2)+65.72y(t-1)$ $+20.24u^2(t-2)y(t)-178.9u^2(t-1)y(t-1)$	3.075×10^{-7}
$u(t-1)y(t)$	$0.7000u(t-1)u(t-2)-1.000u(t-2)y(t)+0.2000u(t-1)y(t-1)$	3.044×10^{-11}
$u(t-2)y(t)$	$0.7000u(t-1)u(t-2)-1.000u(t-1)y(t)+0.2000u(t-1)y(t-1)$	9.325×10^{-11}
$u^2(t-1)y(t)$	$1.594u(t-1)u(t-2)y(t)-0.8180u(t-1)u(t-2)y(t-1)+0.06525y(t)$	1.644
$u(t-1)u(t-2)y(t)$	$0.6270u^2(t-1)y(t)+0.5130u(t-1)u(t-2)y(t-1)-0.04074y(t)$	1.644
$u^2(t-2)y(t)$	$0.8180u(t-1)u(t-2)y(t-1)+0.1848y(t)+0.1377u(t-1)u(t-2)y(t)$	1.644

TABLE 4. Results of example 2

$Y(t)$	RHS terms	*sqe*
$y(t)$	fifteen terms were selected	16.74
$u(t-1)y(t)$	$-3.556u(t-1)u(t-2)y(t)+1.000u(t-2)y(t)$ $-1.490\times10^{-7}u(t-1)y(t-1)-1.803\times10^{-5}u^2(t-1)-0.3039y(t)$ $+1.778u^2(t-2)y(t)+1.778u^2(t-1)y(t)$	4.334×10^5
$u(t-2)y(t)$	$1.000u(t-2)y(t-1)+1.000u(t-1)u(t-2)+3.576\times10^{-7}u(t-1)y(t)$ $-1.000u(t-1)y(t-2)+1.000u(t-1)y(t-1)$	4.366×10^{-6}
$u^2(t-1)y(t)$	$0.1709y(t)-0.5624u(t-2)y(t)+2.000u(t-1)u(t-2)y(t)$ $+0.5624u(t-1)y(t)-1.000u^2(t-2)y(t)$	4.443×10^5
$u(t-1)u(t-2)y(t)$	$-0.2812u(t-1)y(t)+0.2812u(t-2)y(t)+0.500u^2(t-1)y(t)$ $-0.08547y(t)+0.5000u^2(t-2)y(t)$	4.443×10^5
$u^2(t-2)y(t)$	$1.001u^2(t-2)y(t-1)+0.04245u(t-1)+0.1400u^2(t-2)$ $-3.089\times10^{-4}u(t-1)u(t-2)y(t)-1.003u(t-1)u(t-2)y(t-2)$ $+1.003u(t-1)u(t-2)y(t-1)$	3.709×10^4