ARX models for time-varying systems estimated by recursive penalized weighted least squares method

Qin, Pan
Faculty of Mathematics, Kyushu University

Nishii, Ryuei
Faculty of Mathematics, Kyushu University

Nakagawa, Tadashi
MAZDA Motor Corporation

Nakamoto, Takayoshi
MAZDA Motor Corporation

出版情報：Journal of Math-for-Industry (JMI). 2 (A), pp.109-114, 2010-04-08. 九州大学大学院数理学研究院
バージョン：
権利関係：

Journal of Math-for-Industry, Vol.2(2010A-11) , pp.109-114

ARX models for time-varying systems estimated by recursive penalized weighted least squares method

Pan Qin, Ryuei Nishii, Tadashi Nakagawa and Takayoshi Nakamoto

Received on April 2, 2010 / Revised on April 8, 2010

Abstract. We consider the modeling problem for time-varying systems by Auto-Regressive models with eXogenous variables (ARX) models. To track the variations of time-varying systems, we propose a new Recursive Penalized Weighted Least Squares (RPWLS) method to estimate the ARX models. Furthermore, by virtue of Generalized Information Criterion, the proper ARX models by RPWLS are selected. Numerical examples are provided to verify the performance of the proposed RPWLS method.

Keywords. ARX model, Time-varying systems, GIC, Model selection

1. Introduction

We consider systems in which variables of different kinds interact mutually. While time-invariant systems are most common way of describing a dynamical phenomenon, it is also quite often useful or necessary to employ time-varying systems. A time-varying system has parameters which are dependent to time, in contrast to a time-invariant one. A typical example for time-varying systems is the flight of the aircraft. Its time variant characteristics are caused by different configuration of control surfaces during take off, cruise and landing as well as constantly decreasing weight due to fuel consumption.

It is not only theoretical but also practical interests to investigate modeling methods for time-varying systems. For example, in adaptive control designs, it is necessary to construct a time-varying model, based on which the adaptive control can be modified. In this paper, we discuss the estimation method of Auto-Regressive models with eXogenous variables (ARX) models with time-varying parameters.

To estimate the ARX models for time-varying systems, recursive modeling methods, such as recursive least squares method and recursive instrumental variable method, have been proposed [5]. These two recursive methods do not take the regularizations of the ARX coefficients into consideration. However, it is known that regularization techniques are important when the Hessian is an ill-conditioned matrix and/or the model is specified by too many parameters in comparison with the sample size ([5], Sect. 7.4). To deal with such a problem, we devised the recursive penalized weighted least squares (RPWLS) method. There are two reasons for employing RPWLS. First, to make the estimated ARX models capture the variations of the time-varying system, we should assign less weight to the older measurements that are no longer representative of the system. Therefore, samples need to have different weights.

Second, regularization techniques for parameter estimation are often used for balancing between the fitting and the smoothness as well as for stabilization of the estimated parameters [4].

Selection of an appropriate model structure is crucial for a successful modeling application. Akaike's Information Criterion (AIC) [1] and Schwarz's Bayesian Information Criterion (BIC) [7] can be used for the ARX model evaluation and selection when the models are obtained by maximum likelihood estimation (MLE). However, AIC and BIC cannot be used in the scenario for estimating the ARX models by RPWLS. This problem may be solved by using the Generalized Information Criterion (GIC) [3]. GIC can be used to evaluate statistical models estimated by various procedures other than MLE. In our previous study [6], we derived the GIC for time-invariant ARX models estimated by the penalized weighted least squares (PWLS) method. As a natural continuation of that study, in this paper, we investigate what GIC can offer for evaluation and selection of ARX models estimated by RPWLS.

The rest of the paper is organized as follows. Section 2 describes the ARX models estimated by RPWLS. GIC for selecting the time-varying ARX models by RPWLS will be specified in Section 3. In Section 4, the numerical examples are provided. Finally, the discussion is stated in Section 5.

2. ARX models estimated by RPWLS

Let t denote the present time. The ARX model can be formulated as follows.

$$y_t = \sum_{i=1}^{m} \alpha_i y_{t-i} + \sum_{j=0}^{n} \boldsymbol{\beta}_j' \boldsymbol{u}_{t-j} + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma^2) \tag{1}$$

for $t = 1, \cdots, N$, where $y_t \in \mathbb{R}$ is the output (target variable), $\boldsymbol{u}_t \in \mathbb{R}^p$ denotes p external inputs (explanatory variables), $\alpha_i \in \mathbb{R}$

and $\boldsymbol{\beta}_j \in \mathbb{R}^p$ are coefficients, and m and n are the ARX model orders satisfying $m \geq n$.

2.1. Parameter estimation by PWLS

The ARX models are estimated by minimizing the following function based on the measurements $\{y(t), \boldsymbol{u}(t)\}$ with $t = 1 - m, 2 - m, ..., 0, 1, \ldots, N$:

$$J_{PWLS} = \sum_{t=1}^{N} w_t \left\{ [y_t - \boldsymbol{h}_t' \boldsymbol{\xi}]^2 + \sigma^2 \boldsymbol{\xi}' K \boldsymbol{\xi} \right\} \tag{2}$$

where $w_t \geq 0$ is the weight of the t-th sample, and

$$\begin{aligned}
\boldsymbol{h}_t &= \left(y_{t-1}\ y_{t-2}\ \cdots\ y_{t-m}\ \boldsymbol{u}_{t-1}'\ \boldsymbol{u}_{t-2}'\ \cdots\ \boldsymbol{u}_{t-n}' \right)' : (m + np) \times 1 \\
\boldsymbol{\xi} &= \left(\alpha_1\ \alpha_2\ \cdots\ \alpha_m\ \boldsymbol{\beta}_1'\ \boldsymbol{\beta}_2'\ \cdots\ \boldsymbol{\beta}_n' \right)' : (m + np) \times 1 \\
K &= \begin{pmatrix} \lambda_1 K_1 & O & \cdots & O \\ O & \lambda_2 K_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & \lambda_2 K_2 \end{pmatrix} : (m + pn) \times (m + pn)
\end{aligned}$$

Here, $K_1 : m \times m$ is a positive-semi definite matrix corresponding to the auto-regressive term of the formula (1), and $K_2 : p \times p$ is also positive-semi definite corresponding to the exogenous term. Now, $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ are the regularization parameters to be tuned.

The target function (2) can be also reformulated as follows:

$$J_{PWLS} = (\boldsymbol{y} - H\boldsymbol{\xi})' W (\boldsymbol{y} - H\boldsymbol{\xi}) + \mathrm{tr}(W)\sigma^2 \boldsymbol{\xi}' K \boldsymbol{\xi} \tag{3}$$

where

$$\begin{aligned}
\boldsymbol{y} &= \begin{pmatrix} y_1 & y_2 & \cdots & y_N \end{pmatrix}' : N \times 1 \\
H &= \begin{pmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_2 & \cdots & \boldsymbol{h}_N \end{pmatrix}' : N \times (m + pn) \\
W &= \mathrm{diag}\,(w_1, w_2, \ldots, w_N) : N \times N.
\end{aligned}$$

Here, diag$(\cdots)$ means a diagonal matrix. Let $\hat{\boldsymbol{\xi}}_N$ denote the PWLS estimator obtained by the measurements up to time N. Then by solving the equation $\partial J_{PWLS} / \partial \boldsymbol{\xi} = \boldsymbol{0}$, the PWLS estimator $\hat{\boldsymbol{\xi}}_N$ that minimizes J_{PWLS} can be obtained as

$$\hat{\boldsymbol{\xi}}_N = \left(H'WH + \mathrm{tr}(W)\sigma^2 K \right)^{-1} H'W\boldsymbol{y}. \tag{4}$$

The numerical results in our previous study [6] indicated that the ARX models estimated by PWLS performed very well for time-invariant systems. However, as a kind of off-line estimate procedures, PWLS cannot adaptively handle the variations of time-varying systems adaptively. Recursive methods are a popular way of solving such a problem.

2.2. Proposal of parameter estimation by RPWLS

A natural way to track the variations of time-varying systems, the natural way is to assign less weights to older measurements that are no longer representative of the systems. Accordingly, we can specify the weights in (3) up to the current time N as follows:

$$w_t = \begin{cases} 0 & \text{if } t \leq N - \tau \\ \rho^{N-t} & \text{if } N - \tau < t \leq N \end{cases} \tag{5}$$

where $0 < \rho < 1$, and τ is a pre-assigned natural number less than N.

The assigning method of the weights in (5) implies that the data with τ steps before and more are totally discarded. Furthermore, the data are put less weights exponentially as time steps apart from the current time N. Hence, τ should be chosen to be small if the system seems to be varying in a short period.

Suppose that the data at times $t = 1 - m, 2 - m, ..., 0, 1, ..., N$ are available. Let W_N denote the weight matrix at time N, The weight matrices can be formulated as follows:

$$W_N = \mathrm{diag}\left(0, \ldots, 0, 0, \rho^{\tau-1}, \rho^{\tau-2}, \ldots, \rho, 1\right) : N \times N \tag{6}$$

Then, the estimated coefficient vector $\boldsymbol{\xi}$ based on the weight matrix (6) is given by the following formula:

$$\begin{aligned}
\hat{\boldsymbol{\xi}}_N &= \left(H'W_N H + \mathrm{tr}(W_N)\sigma^2 K \right)^{-1} H'W_N \boldsymbol{y} \\
&= \Xi_N \left(\sum_{i=N-\tau+1}^{N} \rho^{N-i} \boldsymbol{h}_i y_i \right) \qquad (7)
\end{aligned}$$

where

$$\Xi_N = \left(\sum_{i=N-\tau-1}^{N} \rho^{N-i} \boldsymbol{h}_i \boldsymbol{h}_i' + \frac{1 - \rho^\tau}{1 - \rho} \sigma^2 K \right)^{-1}.$$

2.3. Non-regularized case

In the special case of $K = O$, i.e., without regularization case, the parameter estimation formula (7) has a recursive formula, which is found in ([5], Sect. 11.2).

$$\hat{\boldsymbol{\xi}}_N = \hat{\boldsymbol{\xi}}_{N-1} + c_N \Xi_{N-1} \boldsymbol{h}_N \tag{8}$$

$$c_N = \frac{y_N - \boldsymbol{h}_N' \hat{\boldsymbol{\xi}}_{N-1}}{\rho + \boldsymbol{h}_N' \Xi_{N-1} \boldsymbol{h}_N} \tag{9}$$

$$\Xi_N = \rho^{-1} \left(\Xi_{N-1} - \frac{\Xi_{N-1} \boldsymbol{h}_N \boldsymbol{h}_N' \Xi_{N-1}}{\rho + \boldsymbol{h}_N' \Xi_{N-1} \boldsymbol{h}_N} \right) \tag{10}$$

Here, Ξ_{N-1} and ξ_{N-1} are derived by using the weight matrix:

$$W_{N-1} = \mathrm{diag}\left(0, \ldots, 0, 0, \rho^{\tau-1}, \rho^{\tau-2}, \ldots, \rho, 1, 0\right) : N \times N.$$

3. GIC for the ARX models estimated by RPWLS

Here, we will calculate GIC for the ARX models estimated by RPWLS at time N. Define $\boldsymbol{\theta} = (\boldsymbol{\xi}',\ \sigma)'$ and let $f(y_t \mid \hat{\boldsymbol{\theta}}_N)$ denote an estimated ARX model, in which $\hat{\boldsymbol{\theta}}_N$ is an M-estimator

obtained by solving the following implicit equation:

$$\sum_{t=1}^{N} \boldsymbol{\psi}(y_t, \hat{\boldsymbol{\theta}}_N) = \mathbf{0} \tag{11}$$

with $\boldsymbol{\psi}$ being referred to as $\boldsymbol{\psi}-$function [2]. According to [3], the following theorem gives GIC for the models.

Theorem 1 (GIC for M-estimators)**.** GIC *for* $f(y_t \mid \hat{\boldsymbol{\theta}}_N)$ *can be calculated as follows :*

$$\mathrm{GIC} = -2\sum_{t=1}^{N} \log f(y_t \mid \hat{\boldsymbol{\theta}}_N) + 2\mathrm{tr}\left(R^{-1}Q\right) \tag{12}$$

where matrices R and Q given by

$$R = -\frac{1}{N}\sum_{t=1}^{N} \left.\frac{\partial \boldsymbol{\psi}(y_t, \boldsymbol{\theta})'}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_N} \tag{13}$$

$$Q = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{\psi}(y_t, \hat{\boldsymbol{\theta}}_N) \left.\frac{\partial \log f(y_t \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_N}. \tag{14}$$

□

An ARX model means that a conditional distribution of y_t is a normal distribution having mean $\boldsymbol{h}_t'\boldsymbol{\xi}$ and variance σ^2. Based on this fact, we can construct a $\boldsymbol{\psi}-$function as follows:

$$\begin{aligned}\sum_{t=1}^{N} \boldsymbol{\psi}(y_t, \boldsymbol{\theta}) &= \sum_{t=1}^{N} w_t \frac{\partial}{\partial \boldsymbol{\theta}}\left\{\log f(y_t \mid \boldsymbol{\theta}) - \frac{1}{2}\boldsymbol{\xi}' K \boldsymbol{\xi}\right\} \\ &= \begin{pmatrix} \frac{1}{\sigma^2} H' W_N(\boldsymbol{y} - H\boldsymbol{\xi}) - \mathrm{tr}(W_N)K\boldsymbol{\xi} \\ \frac{-\mathrm{tr}(W_N)}{\sigma} + \frac{1}{\sigma^3}(\boldsymbol{y} - H\boldsymbol{\xi})' W_N (\boldsymbol{y} - H\boldsymbol{\xi}) \end{pmatrix} = \mathbf{0}. \end{aligned} \tag{15}$$

By solving the equation (15), we arrive at the estimator $\hat{\theta}_N$ for the ARX models at time N:

$$\hat{\boldsymbol{\theta}}_N = \begin{pmatrix} \hat{\boldsymbol{\xi}}_N \\ \hat{\sigma}_N \end{pmatrix} = \begin{pmatrix} \Xi_N H' W_N \boldsymbol{y} \\ \sqrt{\dfrac{\left(\boldsymbol{y} - H\hat{\boldsymbol{\xi}}_N\right)' W_N \left(\boldsymbol{y} - H\hat{\boldsymbol{\xi}}_N\right)}{\mathrm{tr}(W_N)}} \end{pmatrix}.$$

The GIC for the ARX models obtained by $\hat{\boldsymbol{\theta}}_N$ can be gotten from the following proposition.

Proposition 1. *The* GICs *for the ARX models obtained by* $\hat{\boldsymbol{\theta}}_N$ *are calculated as follows:*

$$\begin{aligned}\mathrm{GIC} &\equiv \mathrm{GIC}(m, n, p, \rho, \lambda_1, \lambda_2) \\ &= -2\sum_{t=1}^{N}\Bigg[\log f(y_t \mid \hat{\boldsymbol{\theta}}_N) \\ &\quad - w_t \hat{\sigma}_N^2 \mathrm{tr}\left\{\begin{pmatrix} \Xi_N^{-1} & \boldsymbol{a} \\ \boldsymbol{a}' & 2\mathrm{tr}(W_N) \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{b}_t - K\hat{\boldsymbol{\xi}}_N \\ c_t \end{pmatrix}\begin{pmatrix} \boldsymbol{b}_t \\ c_t \end{pmatrix}'\right\}\Bigg] \end{aligned} \tag{16}$$

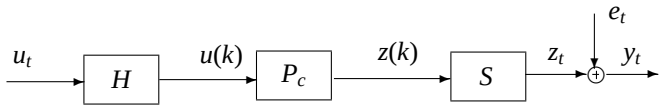

Figure 1: Input-output data sampled from a continuous-time process

where

$$\boldsymbol{a} = 2\mathrm{tr}(W_N)\hat{\sigma}_N K \hat{\boldsymbol{\xi}}_N \tag{17}$$

$$\boldsymbol{b}_t = \frac{1}{\hat{\sigma}_N^2}\left(y_t - \boldsymbol{h}_t'\hat{\boldsymbol{\xi}}_N\right)\boldsymbol{h}_t \tag{18}$$

$$c_t = -\frac{1}{\hat{\sigma}_N} + \frac{1}{\hat{\sigma}_N^3}\left(y_t - \boldsymbol{h}_t'\hat{\boldsymbol{\xi}}_N\right)^2. \tag{19}$$

□

The readers are referred to [6] for the proof.

Finally, we can select the ARX models estimated by RPWLS up to time N by choosing the tuning parameters m, n, p, ρ, λ_1, and λ_2 that minimize GIC in **Proposition 1**.

Note here that the tuning parameters λ_1 and λ_2 are used for specifying the matrix K. Note also that the time-range parameter τ can not be determined by minimizing GIC because the likehood monotonically increases as τ becomes small.

4. Numerical examples

4.1. Setup of ARX models

Consider the Single-Input Single-Output (SISO) continuous-time process P_c observed by a sampling period $T = 0.2$ (second) in Figure 1.

In Figure 1, H is a zero-order holder which transforms the discrete-time input signal u_t into a continuous-time signal $u(k)$:

$$u(k) = u_t \quad \text{if} \quad (t-1)T < k \le tT.$$

S is a sampler which transforms the continuous-time output $z(k)$ into a discrete-time signal z_t:

$$z_t = z(k)\delta(k - tT) \quad \text{for} \quad t = 0, 1, \ldots, N$$

where

$$\delta(k) = \begin{cases} 1 & \text{if} \quad k = 0 \\ 0 & \text{else.} \end{cases}$$

The discrete-time output z_t is corrupted by measurement noise e_t and y_t is the measurement output. The continuous-time process P_c is described by the following time-varying transfer function:

$$P_c : G(s, k) = \frac{Y(s)}{U(s)} = \frac{a_1(k)s + a_0(k)}{b_2(k)s^2 + b_1(k)s + b_0(k)}$$

where $Y(s)$ and $U(s)$ are the Laplace transforms of $y(k)$ and $u(k)$ respectively, and are defined as follows:

$$Y(s) = \int_0^\infty y(k)e^{-sk}dk$$

$$U(s) = \int_0^\infty u(k)e^{-sk}dk.$$

Here, $a_1(k)$, $a_0(k)$, $b_2(k)$, $b_1(k)$ and $b_0(k)$ are continuous time-varying coefficients. In system and control theory, such a $G(s,k)$ is often used to describe the relation between the input and the output of a time-varying continuous-time system [5].

To obtain the input-output data, the system in Figure 1 is excited by using the input u_t which is a zero mean white signal with variance 1. The simulation is conducted with the measurement noise of NSR (noise to signal ratio)= 10%. NSR is defined as the ratio of σ_e/σ_z, where σ_e and σ_z are the standard deviations of the measurement noise and of the noise-free output respectively. The time-varying coefficients are operated in the following subsections: 4.2. time-invariant pattern, 4.3. switching pattern and 4.4. constantly varying pattern.

Based on GIC, we select the proper tuning parameters m, n, λ and ρ for the ARX models. The m and n are selected among the set $\{1, 2, \ldots, 5\}$, λ_i is selected from $\lambda_i = 10^{-1}, 10^{-2}, \ldots, 10^{-10}$ $(i = 1, 2)$ and ρ is selected from $\rho = 0.80, 0.81, \ldots, 0.99$. Because the system in Figure 1 is SISO, $p = 1$ is fixed. K_1 and K_2 are two diagonal matrices in which the diagonal elements are the inverses of sample variances of y_t and u_t respectively. Then with the selected tuning parameters, the ARX models are estimated by the proposed RPWLS methods. All the simulations in the following subsections were run for 80 seconds. The numerical results are shown in the following subsections.

Note that if the proper tuning parameters are selected by GIC m and n are fixed in the recursive estimations of ARX models. However, in some cases m and n are also time-varying. In such cases, the recursive calculation procedure of GIC is necessary and this will be studied in the further research.

4.2. Case 1: Time-invariant pattern

In the time-invariant pattern, the transfer function $G(s)$ is supposed to be as the following:

$$G(s) = \frac{3s+1}{2s^2+3s+1}.$$

We used time-invariant pattern to simulate a steady industrial process. The input and output data are shown in Figure 2. The numerical results selected by GIC for $\tau = 20, 40, 60$ are listed in Table 1. Numerals in boldface denote the maximum coefficients of determination.

The numerical results indicate that GIC selected $m = 2$ and $n = 1$ for $\tau = 40, 60$. This is identical to the orders of $G(s)$. Therefore, the minimal realizations of ARX models are selected for $\tau = 40, 60$. The coefficients of determination indicated that τ produces trivial effects on the selection of ARX models. This is in accordance with the fact that $G(s)$ is time-invariant.

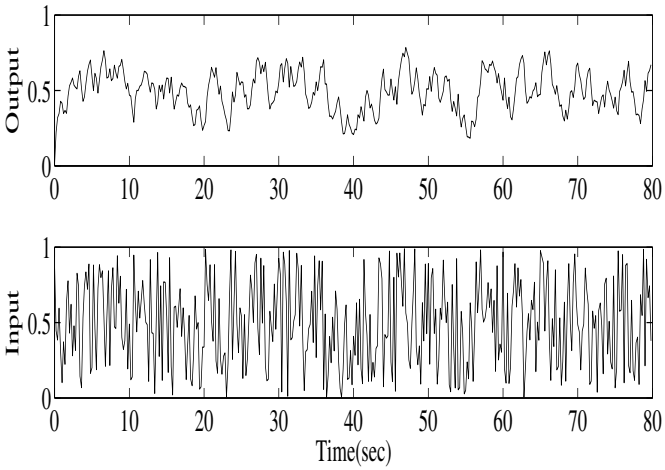

Figure 2: Input and output of time-invariant pattern

Table 1: Selected ARX models and the coefficient of determination for time-invariant pattern

τ	m	n	ρ	λ_1	λ_2	R^2
20	2	2	0.99	10^{-7}	10^{-3}	0.9845
40	2	1	0.99	10^{-7}	10^{-3}	0.9875
60	2	1	0.99	10^{-7}	10^{-3}	**0.9877**

4.3. Case 2: Switching pattern

In this pattern, the transfer function $G(s)$ is switched from one model to another at 40 second as the following:

$$G(s,k) = \begin{cases} \dfrac{3s+1}{2s^2+3s+1} & \text{for} \quad 0 \le k < 40 \\[2ex] \dfrac{10s+5}{2s^2+5s+5} & \text{for} \quad 40 \le k \le 80. \end{cases}$$

The switching pattern simulates the variations of loading and/or operation conditions in industry. The input and output of the switching pattern are shown in Figure 3.

The numerical results selected by GIC for $\tau = 20, 40, 60$ are listed in Table 2. Numerals in boldface denote the maximum coefficients of determination.

The numerical results indicate that GIC selected $m = 2$ and $n = 1$ for $\tau = 20, 40$. This is identical to the orders of $G(s,k)$. Therefore, the minimal realizations of models are selected for $\tau = 20, 40$. The coefficients of determination shows that τ should not be too large.

4.4. Case 3: Constantly varying pattern

In the constantly varying pattern, the transfer function $G(s,k)$ constantly changes over time k:

$$G(s,k) = \frac{(3+0.2k)\,s+1}{(2-0.02k)\,s^2+(3+0.06k)\,s+1}.$$

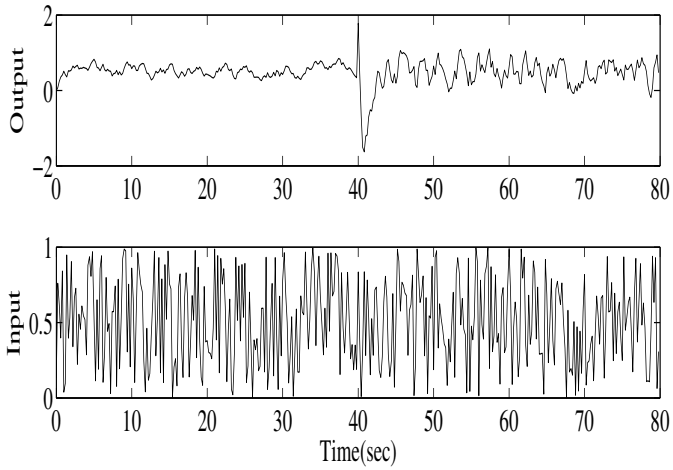

Figure 3: Input and output of switching pattern

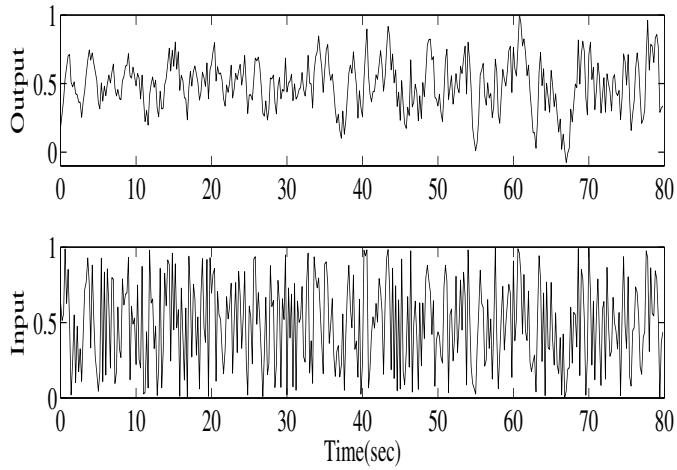

Figure 4: The input and output of constantly varying pattern

Table 2: Selected ARX models and the coefficient of determination for switching pattern

τ	m	n	ρ	λ_1	λ_2	R^2
20	2	1	0.95	10^{-7}	10^{-3}	**0.9715**
40	2	1	0.98	10^{-7}	10^{-4}	0.9455
60	2	2	0.99	10^{-7}	10^{-3}	0.9251

Table 3: The selected ARX models and the coefficient of determination for constantly varying pattern for constantly varying pattern

τ	m	n	ρ	λ_1	λ_2	R^2
20	2	1	0.94	10^{-5}	10^{-4}	**0.9469**
40	2	1	0.96	10^{-5}	10^{-4}	0.9448
60	2	1	0.99	10^{-5}	10^{-4}	0.9270

We used such a pattern to simulate the constantly variations such as the consumptions of fuel, energy or material. The input and output data are shown in Figure 3. Numerals in boldface denote the maximum coefficients of determination.

The ARX models selected by GIC for $\tau = 20, 40, 60$ are listed in Table 3. Numerals in boldface denote the maximum coefficients of determination.

The numerical results indicates that $m = 2$ and $n = 1$ are selected by GIC for $\tau = 20, 40, 60$. These values are identical to the orders of $G(s, k)$. Therefore, the minimal realizations of ARX models are selected for $\tau = 20, 40, 60$. The coefficients of determination shows that τ should not be too large.

5. Discussion

In this paper, we addressed the modeling problems for time-varying systems. To track the variations of time-varying systems, we proposed the RPWLS method that use the discounted weights and a regularization parameter. Because AIC and Schwarz's BIC cannot be applied to RPWLS, we calculate GIC to select the proper weights, regularization parameters and orders for the ARX models. The numerical examples indicated that the the proposed method selected the proper ARX models and the selected ARX models performed very well.

Acknowledgments

This research is motivated by the joint research between Mazda Motor Corporation and Kyushu University. The first two authors would like to thank Mazda Motor Corporation for supporting the research. The authors are very grateful to the reviewer for carefully reading and pointing out an error in the draft.

References

[1] Akaike, H. A new look at the statistical model identification, *IEEE Tran. Automat. Control.* **AC-19**, 716-723, 1974.

[2] Huber, P. *Robutst Statistics*, Wilely, New York, 2004.

[3] Konishi, S. and Kitagawa, G. Generalized information criteria in model selection, *Biometrika* **83**, 875-890, 1996.

[4] Konishi, S. and Kitagawa, G. *Information Criteria and Statistical Modeling*, Springer, 2008.

[5] Ljung, L. *System Identification, Theory for the User*, Prince Hall PTR, 1996.

[6] Qin, P. and Nishii, R. Selection of ARX models estimated by the penalized weighted leastsquares method, *Bull. Inform. Cybernet.* To appear.

[7] Scharwz, G. Estimating the dimension of a model, *Ann. Statist* **6**, 461-464, 1978.

Pan Qin
Faculty of Mathematics, Kyushu University
E-mail: pan(at)math.kyushu-u.ac.jp

Ryuei Nishii
Faculty of Mathematics, Kyushu University
E-mail: nishii(at)math.kyushu-u.ac.jp

Tadashi Nakagawa
Mazda Motor Corporation
E-mail: nakagawa.t(at)mazda.co.jp

Takayoshi Nakamoto
Mazda Motor Corporation
E-mail: nakamoto.tak(at)mazda.co.jp