

SELECTION OF ARX MODELS ESTIMATED BY THE PENALIZED WEIGHTED LEAST SQUARES METHOD

Qin, Pan
Faculty of Mathematics, Kyushu University

Nishii, Ryuei
Faculty of Mathematics, Kyushu University

<https://doi.org/10.5109/25904>

出版情報 : Bulletin of informatics and cybernetics. 42, pp.35-43, 2010-12. Research Association of Statistical Sciences

バージョン :

権利関係 :



SELECTION OF ARX MODELS ESTIMATED BY THE PENALIZED
WEIGHTED LEAST SQUARES METHOD

by

Pan QIN and Ryuei NISHII

*Reprinted from the Bulletin of Informatics and Cybernetics
Research Association of Statistical Sciences, Vol.42*

FUKUOKA, JAPAN
2010

SELECTION OF ARX MODELS ESTIMATED BY THE PENALIZED WEIGHTED LEAST SQUARES METHOD

By

Pan QIN* and Ryuei NISHII†

Abstract

We consider the selection problem of auto-regressive time-series models with exogenous variables (ARX) estimated by Penalized Weighted Least Squares (PWLS) method. AIC and BIC are developed based on the maximum likelihood estimation. Therefore, in this research, we evaluate GIC for the ARX models estimated by PWLS. In a numerical experiment, the model selected by GIC shows an excellent performance, especially, in a target region.

Key Words and Phrases: AIC, auto-regressive time-series models with exogenous variables, BIC, GIC, MLE, Model selection, System identification.

1. Introduction

Auto-Regressive time-series models with exogenous variables (ARX) is useful to describe dynamic behavior of a system or process in the discrete time domain. Selecting an appropriate model structure is most crucial for a successful modeling application. Akaike's Information Criterion (AIC) (Akaike (1974)) and Bayesian Information Criterion (BIC) (Schwarz (1978)) can be used for the ARX model evaluation and selection when the models are obtained by the Maximum Likelihood Estimation (MLE).

However, various estimation procedures other than MLE may also be utilized for ARX models in practical applications. In this paper, the ARX models estimated by Penalized Weighted Least Squares (PWLS) method are taken into consideration. The reason for using PWLS is two-fold: In many modeling applications, the different measurements should be assigned different weights (Ljung (1996)). For example, if we want to construct an ARX model which has good prediction performance on some specified regions, then the measurements in such regions should be put on more weight. On the other hand, regularization techniques are often used in the estimation procedures to control the trade-off between the smoothness of the model and the goodness of fitting to the data (Konishi and Kitagawa (2008)). Especially for the estimation of ARX models, Ljung (1996, Sect. 7.4) pointed out the importance of regularization techniques in the cases that the Hessian is an ill-conditioned matrix and/or the model is specified by too many parameters in comparison with the sample size. Therefore, the selection of ARX

* Faculty of Mathematics, Kyushu University, 744 Motooka, Nishi-Ku, Fukuoka 819-0395, Japan.
Email: pan@math.kyushu-u.ac.jp

† Faculty of Mathematics, Kyushu University, 744 Motooka, Nishi-Ku, Fukuoka 819-0395, Japan.
Email: nishii@math.kyushu-u.ac.jp

models estimated by PWLS is not only to find the proper auto-regressive order and variable subsets but to find an appropriate regularization parameter.

Because the ARX models obtained by PWLS are different from the ones obtained by MLE, AIC and BIC are not applicable. Nevertheless, this situation may be remedied by the Generalized Information Criterion (GIC) (Konishi and Kitagawa (1996)). GIC is a general framework for constructing information criteria in the context of functional statistics and can be applied to evaluate statistical models constructed by various types of estimation procedures other than MLE. In this research, we investigate what GIC can offer for evaluating and selecting the ARX model estimated by PWLS.

The rest of the paper is organized as follows. Section 2 describes the ARX models estimated by PWLS. GIC for selecting the ARX models estimated PWLS will be specified in Section 3. In Section 4, numerical results are given. Finally, the discussion is stated in Section 5.

2. ARX Models estimated by PWLS

Let t denote the present time, then an ARX model for Δ -step-ahead prediction can be formulated as the following:

$$y(t + \Delta) = \sum_{i=1}^m \alpha_i y(t - i + 1) + \sum_{j=1}^n \beta'_j \mathbf{u}(t - j + 1) + e(t), \quad e(t) \sim N(0, \sigma^2), \quad (1)$$

for $t = 1, \dots, N$, where $y(t) \in R$ is the output (target variable), $\mathbf{u}(t) \in R^p$ denotes p external inputs (explanatory variables), $\alpha_i \in R$ and $\beta_j \in R^p$ are coefficients, Δ is a natural number, m and n are the ARX model orders satisfying $m \geq n$.

In this paper, we estimate the ARX models by minimizing the following function:

$$J_{PWLS} = \sum_{t=1}^N w_t [y(t + \Delta) - \mathbf{h}'_t \boldsymbol{\xi}]^2 + \lambda \boldsymbol{\xi}' K \boldsymbol{\xi} \quad (2)$$

where

$$\begin{aligned} \mathbf{h}_t &= [y(t) \ y(t-1) \ \cdots \ y(t-m+1) \ \mathbf{u}(t)' \ \mathbf{u}(t-1)' \\ &\quad \cdots \ \mathbf{u}(t-n+1)']' : (m+np) \times 1, \\ \boldsymbol{\xi} &= [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_m \ \beta'_1 \ \beta'_2 \ \cdots \ \beta'_n]' : (m+np) \times 1 \end{aligned}$$

and $w_t > 0$ are weights, K is an $(m+pn) \times (m+pn)$ positive-semi definite matrix, and $\lambda > 0$ is a regularization parameter. The function (2) can be also reformulated as follows:

$$J_{PWLS} = (\mathbf{y} - H\boldsymbol{\xi})' W (\mathbf{y} - H\boldsymbol{\xi}) + \lambda \boldsymbol{\xi}' K \boldsymbol{\xi} \quad (3)$$

where

$$\begin{aligned} \mathbf{y} &= (y(1+\Delta) \ y(2+\Delta) \ \cdots \ y(N+\Delta))' : N \times 1, \\ H &= (\mathbf{h}_1 \ \mathbf{h}_2 \ \cdots \ \mathbf{h}_N)' : N \times (m+pn), \\ W &= \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_N \end{pmatrix} : N \times N. \end{aligned}$$

By solving the equation $\partial J_{PWLS}/\partial \xi = \mathbf{0}$, the PWLS estimator $\hat{\xi}_{PWLS}$ which minimizes J_{PWLS} is given as the following:

$$\hat{\xi}_{PWLS} = (H'WH + \lambda K)^{-1} H'W\mathbf{y}. \quad (4)$$

The selection problem of the ARX models achieved by $\hat{\xi}_{PWLS}$ is to find the proper regularization parameter λ and the orders m and n . Note that MLE for ARX models is obtained by $\hat{\xi}_{MLE} = (H'H)^{-1}H\mathbf{y}$ (Ljung (1996)). It is obvious that $\hat{\xi}_{PWLS}$ is different from $\hat{\xi}_{MLE}$. Thus AIC and BIC cannot be applied to the selection of ARX models estimated by $\hat{\xi}_{PWLS}$. In the following section, GIC will be derived for the ARX models estimated by $\hat{\xi}_{PWLS}$.

3. GIC for the ARX models estimated by PWLS

In this section, we will calculate GIC for the ARX models estimated by PWLS. Define $\theta = (\xi', \sigma')'$ and let $f(y(t + \Delta) | \theta_{PWLS})$ denote an estimated ARX model, in which θ_{PWLS} is an M-estimator obtained by the solution of the following implicit equation:

$$\sum_{t=1}^N \psi(y(t + \Delta), \hat{\theta}_{PWLS}) = \mathbf{0} \quad (5)$$

with ψ being referred to as ψ -function (Huber (2004)). Then GIC for the models obtained by $\hat{\theta}_{PWLS}$ is given by the following theorem in the context of functional statistics.

THEOREM 3.1. *GIC for $f(y(t + \Delta) | \hat{\theta}_{PWLS})$ can be calculated as follows:*

$$\text{GIC} = -2 \sum_{t=1}^N \log f(y(t + \Delta) | \hat{\theta}_{PWLS}) + 2\text{tr}(R^{-1}Q), \quad (6)$$

where matrices R and Q given by

$$R = -\frac{1}{N} \sum_{t=1}^N \left. \frac{\partial \psi(y(t + \Delta), \theta)'}{\partial \theta} \right|_{\theta = \hat{\theta}_{PWLS}}, \quad (7)$$

$$Q = \frac{1}{N} \sum_{t=1}^N \psi(y(t + \Delta), \hat{\theta}_{PWLS}) \left. \frac{\partial \log f(y(t + \Delta) | \theta)}{\partial \theta'} \right|_{\theta = \hat{\theta}_{PWLS}}. \quad (8)$$

□

For the proof, the readers are referred to Konishi and Kitagawa (1996).

An ARX model means that a conditional distribution of $y(t + \Delta)$ can be given by the normal distribution having mean $\mathbf{h}_t' \xi$ and variance σ^2 . Based on this fact, we construct a ψ -function as follows:

$$\sum_{t=1}^N \psi(y(t + \Delta), \theta) = \sum_{t=1}^N w_t \frac{\partial}{\partial \theta} \left\{ \log f(y(t + \Delta) | \theta) - \frac{1}{2} \lambda^* \xi' K \xi \right\} \quad (9)$$

$$= \begin{pmatrix} \frac{1}{\sigma^2} H' W (\mathbf{y} - H \boldsymbol{\xi}) - \text{tr}(W) \lambda^* K \boldsymbol{\xi} \\ -\frac{\text{tr}(W)}{\sigma} + \frac{1}{\sigma^3} (\mathbf{y} - H \boldsymbol{\xi})' W (\mathbf{y} - H \boldsymbol{\xi}) \end{pmatrix} = \mathbf{0} \quad (10)$$

where $\lambda^* = \lambda / (\text{tr}(W) \sigma^2)$. By solving the equation (10), the estimator $\hat{\boldsymbol{\theta}}_{PWLS}$ for ARX models are obtained as

$$\hat{\boldsymbol{\theta}}_{PWLS} = \begin{pmatrix} \hat{\boldsymbol{\xi}}_{PWLS} \\ \hat{\sigma}_{PWLS} \end{pmatrix} = \begin{pmatrix} \Xi H' W \mathbf{y} \\ \sqrt{\frac{(\mathbf{y} - H \hat{\boldsymbol{\xi}}_{PWLS})' W (\mathbf{y} - H \hat{\boldsymbol{\xi}}_{PWLS})}{\text{tr}(W)}} \end{pmatrix}$$

where

$$\Xi = (H' W H + \text{tr}(W) \hat{\sigma}_{PWLS}^2 \lambda^* K)^{-1}.$$

Then, the GIC for the ARX models obtained by $\hat{\boldsymbol{\theta}}_{PWLS}$ is given as the following proposition.

PROPOSITION 3.2. GIC for the ARX models achieved by $\hat{\boldsymbol{\theta}}_{PWLS}$ are calculated as follows:

$$\begin{aligned} \text{GIC} = & -2 \sum_{t=1}^N \left\{ \log f(y(t + \Delta) \mid \hat{\boldsymbol{\theta}}_{PWLS}) \right. \\ & \left. - \hat{\sigma}_{PWLS}^2 w_t \text{tr} \left[\begin{pmatrix} \Xi^{-1} & \mathbf{a} \\ \mathbf{a}' & 2\text{tr}(W) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{b}_t - \lambda^* K \hat{\boldsymbol{\xi}}_{PWLS} \\ c_t \end{pmatrix} \begin{pmatrix} \mathbf{b}_t \\ c_t \end{pmatrix}' \right] \right\}, \quad (11) \end{aligned}$$

where

$$\mathbf{a} = 2\text{tr}(W) \hat{\sigma}_{PWLS} \lambda^* K \hat{\boldsymbol{\xi}}_{PWLS}, \quad (12)$$

$$\mathbf{b}_t = \frac{1}{\hat{\sigma}_{PWLS}^2} (y(t + \Delta) - \mathbf{h}_t' \hat{\boldsymbol{\xi}}_{PWLS}) \mathbf{h}_t, \quad (13)$$

$$c_t = -\frac{1}{\hat{\sigma}_{PWLS}} + \frac{1}{\hat{\sigma}_{PWLS}^3} (y(t + \Delta) - \mathbf{h}_t' \hat{\boldsymbol{\xi}}_{PWLS})^2. \quad (14)$$

□

PROOF. Based on (10), R in (7) can be derived as follows:

$$\begin{aligned} R &= -\frac{1}{N} \sum_{t=1}^N \frac{\partial \psi(y(t + \Delta), \boldsymbol{\theta})'}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{PWLS}} \\ &= -\frac{1}{N} \begin{pmatrix} -\frac{1}{\hat{\sigma}_{PWLS}^2} H' W H - 2\text{tr}(W) \lambda^* K \\ -\frac{2}{\hat{\sigma}_{PWLS}^3} (\mathbf{y} - H \hat{\boldsymbol{\xi}}_{PWLS})' W H \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
& \left(\begin{array}{c} -\frac{2}{\hat{\sigma}_{PWLS}^3} H' W (\mathbf{y} - H \hat{\boldsymbol{\xi}}_{PWLS}) \\ \frac{\text{tr}(W)}{\hat{\sigma}_{PWLS}^2} - \frac{3}{\hat{\sigma}_{PWLS}^4} (\mathbf{y} - H \hat{\boldsymbol{\xi}}_{PWLS})' W (\mathbf{y} - H \hat{\boldsymbol{\xi}}_{PWLS}) \end{array} \right) \\
&= \frac{1}{N \hat{\sigma}_{PWLS}^2} \left(\begin{array}{cc} \Xi^{-1} & 2\text{tr}(W) \hat{\sigma}_{PWLS} \lambda^* K \hat{\boldsymbol{\xi}}_{PWLS} \\ 2\text{tr}(W) \hat{\sigma}_{PWLS} \lambda^* \hat{\boldsymbol{\xi}}_{PWLS}' K & 2\text{tr}(W) \end{array} \right) \\
&= \frac{1}{N \hat{\sigma}_{PWLS}^2} \left(\begin{array}{cc} \Xi^{-1} & \mathbf{a} \\ \mathbf{a}' & 2\text{tr}(W) \end{array} \right). \tag{15}
\end{aligned}$$

with \mathbf{a} defined by the formula (12). Q in (6) can be derived as follows

$$\begin{aligned}
Q &= \frac{1}{N} \sum_{t=1}^N \psi(y(t+\Delta), \hat{\boldsymbol{\theta}}_{PWLS}) \frac{\partial \log f(y(t+\Delta) | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{PWLS}} \\
&= \frac{1}{N} \sum_{t=1}^N w_t \left(\frac{\partial \left\{ \log f(y(t+\Delta) | \boldsymbol{\theta}) - \frac{1}{2} \lambda^* \boldsymbol{\xi}' K \boldsymbol{\xi} \right\}}{\partial \boldsymbol{\theta}} \right) \\
&\quad \left(\frac{\partial \log f(y(t+\Delta) | \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)' \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{PWLS}} \\
&= \frac{1}{N} \sum_{t=1}^N w_t \left(\begin{array}{c} \frac{1}{\hat{\sigma}_{PWLS}^2} \left(y(t+\Delta) - \mathbf{h}_t' \hat{\boldsymbol{\xi}}_{PWLS} \right) \mathbf{h}_t - \lambda^* K \hat{\boldsymbol{\xi}}_{PWLS} \\ -\frac{1}{\hat{\sigma}_{PWLS}} + \frac{1}{\hat{\sigma}_{PWLS}^3} \left(y(t+\Delta) - \mathbf{h}_t' \hat{\boldsymbol{\xi}}_{PWLS} \right)^2 \end{array} \right) \\
&\quad \left(\begin{array}{c} \frac{1}{\hat{\sigma}_{PWLS}^2} \left(y(t+\Delta) - \mathbf{h}_t' \hat{\boldsymbol{\xi}}_{PWLS} \right) \mathbf{h}_t \\ -\frac{1}{\hat{\sigma}_{PWLS}} + \frac{1}{\hat{\sigma}_{PWLS}^3} \left(y(t+\Delta) - \mathbf{h}_t' \hat{\boldsymbol{\xi}}_{PWLS} \right)^2 \end{array} \right)' \\
&= \frac{1}{N} \sum_{t=1}^N w_t \left(\begin{array}{c} \mathbf{b}_t - \lambda^* K \hat{\boldsymbol{\xi}}_{PWLS} \\ c_t \end{array} \right) \left(\begin{array}{c} \mathbf{b}_t \\ c_t \end{array} \right)' \tag{16}
\end{aligned}$$

where \mathbf{b}_t and c_t are defined by the formulas (13) and (14) respectively. Then, substitute R and Q in (6) by (15) and (16), the conclusion in PROPOSITION 3.2. can be proved. \square

Finally, we can select the ARX models estimated by PWLS through choosing the regularization parameter λ and ARX model orders m and n which minimize GIC in PROPOSITION 3.2..

4. Numerical example

The dataset acquired from a winding process (Moor (1997)) are used in the numerical study. Figure 1 shows a scheme of the winding process. The main part of the plant is composed of a plastic web that is unwinded from first reel r_1 (unwinding reel), goes over the traction reel r_2 and is finally rewinded on the rewinding reel r_3 . The reel r_1 is driven by the motor M_1 and the reel r_3 is driven by the motor M_3 . The motors are controlled by the input currents I_1 and I_3 (A) respectively. S_1 , S_2 and S_3 (rad/s) denote the angular speeds of the three reels and are measured by dynamo tachometers respectively. T_1 (kg) is the tension between r_1 and r_2 and is measured by a tension meter. The inputs S_1 , S_2 , S_3 , I_1 and I_3 and the output T_1 are shown in Figure 2.

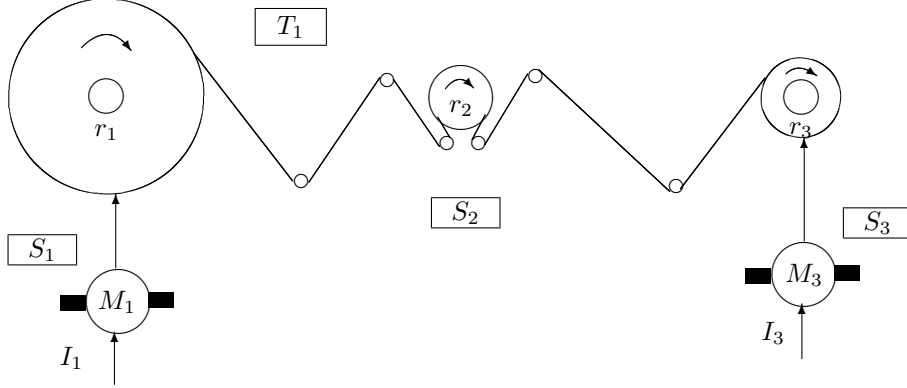


Figure 1: A winding system.

Large tension will lead to the damage on the plastic web. Therefore, we want to construct ARX models which have good performance on prediction of large tension ($|T_1(t)| \geq 1.5kg$ in this study). Figure 3 depicts the samples of tension satisfy $|T_1(t)| \geq 1.5kg$. In the numerical study, we select two ARX models obtained by MLE based on AIC and BIC respectively. Let $w_t = 1$, an ARX model estimated by Penalized Least Squares (PLS) is selected by GIC. Two ARX models estimated by PWLS with $w_t = \sqrt{|T_1(t)|}$ and $|T_1(t)|$ are selected by GIC. Let $\lambda = (0.1)^\delta \text{tr}(W)$ for $\delta = 1, \dots, 10$, then λ can be selected by selecting δ based on GIC. The order m of ARX models is selected from $m = 1, \dots, 5\Delta$ and n is selected from $n = 1, \dots, m$. We used the identity matrix as matrix K in PLS and PWLS. There are 2500 samples which are sampled by $10Hz$ in the dataset. The samples are divided into two parts. The first 1250 samples are used for estimating the ARX models and the rest samples are used for testing the prediction performance of the estimated models. The numerical results are listed in Table 1. Numerals in bold face denote the maximum coefficients of determination in each setting.

In the numerical results, the coefficient of determination R_f^2 is calculated from all the 1250 test samples of tension and their predictions. R_t^2 is calculated from the 145 test samples of tension which satisfy $|T_1(t)| \geq 1.5kg$ and their predictions. R_f^2 shows

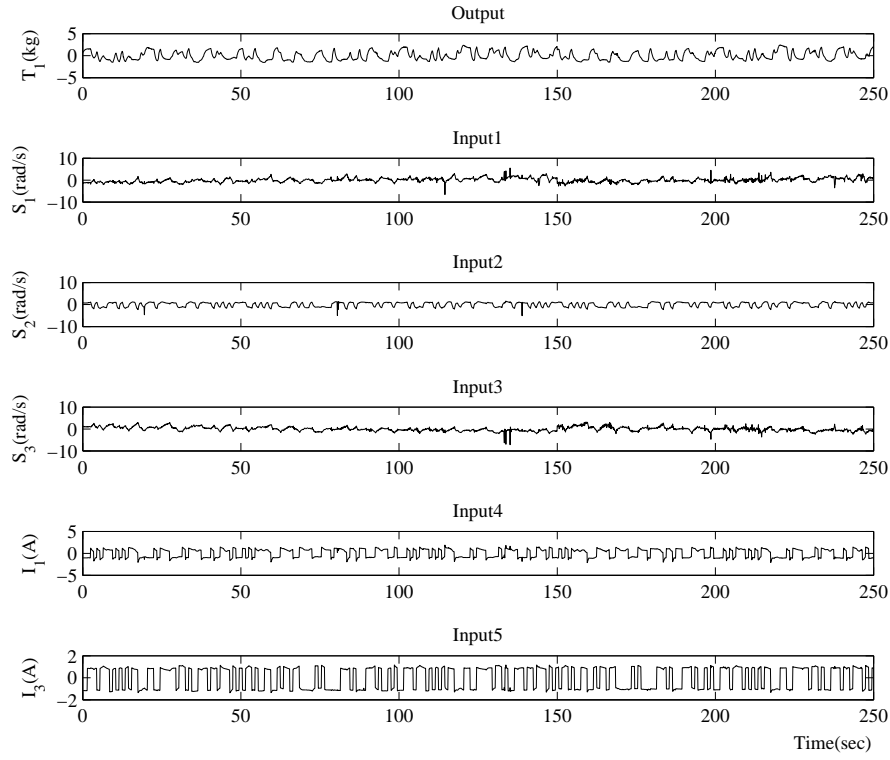


Figure 2: Inputs and output data

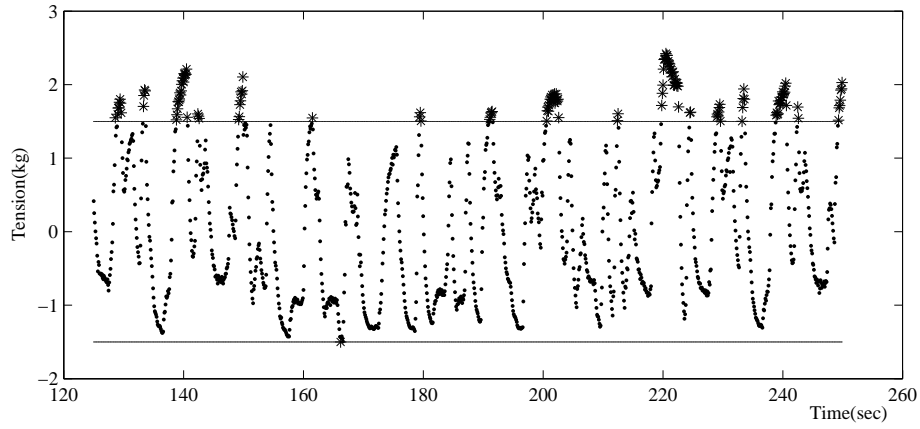
Figure 3: The samples of tension in the test data (starts indicate the samples satisfying $|T_1(t)| \geq 1.5 \text{ kg}$)

Table 1: The selected ARX models and the coefficients of determination for $\Delta = 1, 2, 3$

Δ	IC	w_i	Estimator	m	n	δ	λ	R_f^2	R_l^2
								All test data ($N=1250$)	$ T_1(t) \geq 1.5kg$ ($N=145$)
1	AIC	1	MLE	5	5	-	0	0.9801	0.7650
	BIC	1	MLE	3	3	-	0	0.9786	0.7392
	GIC	1	PLS	5	3	7	1.25×10^{-4}	0.9791	0.7404
	GIC	$\sqrt{ T_1(t) }$	PWLS	3	3	7	1.11×10^{-4}	0.9813	0.7907
	GIC	$ T_1(t) $	PWLS	3	3	7	1.08×10^{-4}	0.9825	0.8229
2	AIC	1	MLE	10	10	-	0	0.9572	0.6318
	BIC	1	MLE	7	2	-	0	0.9549	0.5871
	GIC	1	PLS	10	10	7	1.25×10^{-4}	0.9575	0.6321
	GIC	$\sqrt{ T_1(t) }$	PWLS	10	10	8	1.11×10^{-5}	0.9588	0.6835
	GIC	$ T_1(t) $	PWLS	7	3	7	1.08×10^{-4}	0.9600	0.6900
3	AIC	1	MLE	11	11	-	0	0.9315	0.4605
	BIC	1	MLE	7	2	-	0	0.9310	0.4478
	GIC	1	PLS	11	11	7	1.25×10^{-4}	0.9318	0.4592
	GIC	$\sqrt{ T_1(t) }$	PWLS	11	11	8	1.11×10^{-5}	0.9328	0.5258
	GIC	$ T_1(t) $	PWLS	11	11	8	1.08×10^{-5}	0.9322	0.5845

the prediction performance for all testing samples and R_l^2 shows the performance of predicting the large tension in the test samples.

Table 1 indicates that the ARX models estimated PLS shows the similar performance of models estimated by MLE. However, the models estimated by PWLS with $w_t = |T_1(t)|$ work well on predicting large tension than other selected models.

5. Discussion

In this paper, we addressed the selection of the ARX models estimated by PWLS. Because Akaike's AIC and Schwarz's BIC cover only models obtained by MLE, they cannot be applied to the ARX models estimated by PWLS. However, GIC is applicable to models achieved by various estimation methods in the context of functional statistics. Therefore we specified GIC for the ARX models estimated by PWLS. Based on the specified GIC, the proper regularization parameters λ and the orders m and n can be selected for the ARX models estimated by PWLS. It is worth noting that the specified GIC can be extended to the ARX models estimated by PLS naturally. The numerical experiments were conducted to examine the performance of the specified GIC, and GIC selected excellent models.

Acknowledgement

The authors are very grateful for the reviewer and the editor of BIC for those careful reading and valuable comments. This research is motivated by the joint research between Mazda Motor Corporation and Kyushu University. The authors would like to thank Mazda Motor Corporation for supporting the research.

References

- Akaike, H. (1974). A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control* **AC-19**, 716-723.
- De Moor, B.L.R. (ed.) (1997). *DaISy: Database for the Identification of Systems*, URL: <http://homes.esat.kuleuven.be/smc/daisy/>. [Used dataset: Data from a test setup of an industrial winding process, *Process Industry Systems*, 97-003.], Department of Electrical Engineering, ESAT/SISTA, K.U. Leuven, Belgium
- Huber, P. (2004). *Robust Statistics*, Wiley, New York.
- Konishi, S and Kitagawa, G. (1996). Generalized Information Criteria in Model Selection, *Biometrika* **83**, 875-890.
- Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*, Springer.
- Ljung, L. (1996). *System Identification, Theory for the User*, Prince Hall PTR.
- Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics* **6**, 461-464.

Received October 13, 2009

Revised March 25, 2010