Refined instrumental variable methods for identification of Hammerstein continuous-time Box–Jenkins models

Vincent Laurain, Marion Gilson, Hugues Garnier, Peter C. Young

Refined Instrumental Variable Methods for Identification of Hammerstein Continuous-time Box–Jenkins Models

V. Laurain*, M. Gilson*, H. Garnier*, P.C. Young**

***Abstract*— This article presents instrumental variable methods for direct continuous-time estimation of a Hammerstein model. The non-linear function is a sum of known basis functions and the linear part is a Box–Jenkins model. Although the presented algorithm is not statistically optimal, this paper further shows the performance of the presented algorithms and the advantages of continuous-time estimation on relevant simulations.**

I. INTRODUCTION

The need for non-linear identification grows as the studied system complexity increases and non-linear models apply in many fields [13]. Many different approaches were developed to deal with black-box model identification, whether they are non parametric, using Volterra series approach [10], semi-parametric using neural network methods and support vector machine classification [15], [8], or parametric such as state dependent parameters [23] or extended Kalman filter [4]. Other references can be found in *e.g.* [7]. Semi-parametric approaches, even if performing efficiently, lack the possibility of giving an *a posteriori* physical representation of the studied system. On the other hand, transfer function models provide a generic approach to data-based modelling of linear systems, encompass both discrete-time and continuous-time applications and are in an ideal form to interpret serial, parallel connections of sub-systems which often have a physical significance.

Hammerstein block diagram model is widely represented for modelling non-linear systems [5], [3], [18]. The non-linear block can be represented as a piecewise linear function [2] or as a sum of basis functions [14], [6]. The available methods are often designed for discrete-time (DT) model estimation and usually, extended least squares are used to minimize a prediction error [5].

Even if acquired data are sampled, the underlying dynamic of a real system is continuous. Therefore, direct continuous-time model identification methods regained interest in the recent years [9]. A survey by Rao and Unbehauen [17] shows that CT model identification methods applied to Hammerstein models are poorly represented in literature, and to the best of our knowledge, no method uses instrumental variable (IV) techniques to handle Hammerstein CT model identification. Instrumental variables have the advantage of producing an unbiased estimation independently on the noise model assumed with an acceptable variance in parameter estimation in many practical cases. Moreover, an optimal choice of these instruments leads to a minimal variance estimator [19], [22].

*Centre de Recherche en Automatique de Nancy (CRAN), Nancy-université, CNRS, BP 239, 54506 Vandoeuvre-les-Nancy Cedex, France vincent.laurain@cran.uhp-nancy.fr

**Centre for Research on Environmental Systems and Statistics,Lancaster University, UK; and Integrated Catchment Assessment and Management Centre, School of Resources, Environment and Society, Australian National University, Canberra. p.young@lancaster.ac.uk

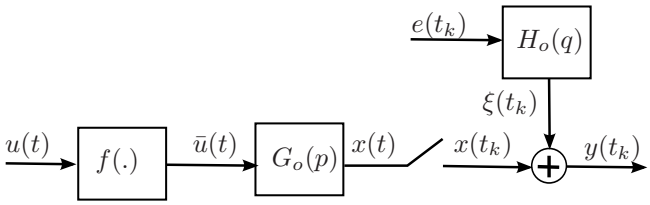


Fig. 1. Hammerstein block representation

Therefore, the main contribution of this paper is to present IV estimation methods for Hammerstein CT models where the non-linear function is a sum of known basis functions $\gamma_1, \gamma_2, \ldots, \gamma_l$ given as:

$$\bar{u}(t) = \sum_{i=1}^{l} \alpha_i \gamma_i(u(t)). \tag{1}$$

The proposed algorithms are based on the multi-input single-output refined instrumental variable for CT systems (MISO RIVC) first introduced by Young and Jakeman [26] and recently extended to handle the case of Box–Jenkins models [24], [25], [12].

II. PROBLEM DESCRIPTION

Consider the Hammerstein system represented in Figure 1 and assume that both input and output signals, $u(t)$ and $y(t)$ are uniformly sampled at a constant sampling time T_s over N samples. Notice first that this Hammerstein system produces the same input-output data for any pair $(\beta f(u), \frac{G_o(p)}{\beta})$. Therefore, to get a unique parametrization, one of the gains of $f(u)$ or $G_o(p)$ has to be fixed [5], [1]. Hence, the first coefficient of the function $f(.)$ equals 1, *i.e.* $\alpha_1 = 1$ in (1). Moreover, $\bar{u}(t)$ in (1) is an internal variable and is actually not directly accessible. The Hammerstein system $\mathcal{S}$, is described by the following input-output relationship:

$$\mathcal{S} \begin{cases} x(t) = \sum_{i=1}^{l} G_{o,i}(p)\gamma_i(u(t)) \\ \xi(t_k) = H_o(q)e(t_k), \\ y(t_k) = x(t_k) + \xi(t_k), \end{cases} \tag{2}$$

where

$$G_{o,i}(p) = \frac{B_{o,i}(p)}{A_o(p)} = \frac{\alpha_i B_o(p)}{A_o(p)}. \tag{3}$$

$B_o(p)$ and $A_o(p)$ are polynomials in differential operator p ($p^i x(t) = \frac{\mathrm{d}^i x(t)}{\mathrm{d}t^i}$) of respective degree n_b and n_a ($n_a \geq n_b$). The method presented is based on the identification of a Box–Jenkins model, where the linear and the noise models are not constrained to have common polynomials. Given the discrete-time, sampled nature of the data, an obvious assumption is that the model of the basic dynamic system is in CT, differential equation form while the coloured noise associated with the sampled output measurement $y(t_k)$ has rational spectral density and can be represented by a discrete-time autoregressive moving average ARMA model ([27], [28]):

$$\xi(t_k) = H_o(q)e(t_k) = \frac{C_o(q^{-1})}{D_o(q^{-1})}e(t_k) \tag{4}$$

where $C_o(q^{-1})$ and $D_o(q^{-1})$ are polynomials in shift operator q^{-1} ($q^{-r}x(t_k) = x(t_{k-r})$) with respective degree n_c and n_d. $e(t_k)$ is a zero-mean, normally distributed, discrete-time white noise sequence: $e(t_k) \sim \mathcal{N}(0, \sigma_e^2)$. Consequently, the Hammerstein model estimation problem can be treated under the previous assumptions using a MISO RIVC algorithm where all inputs have a common denominator $A_o(p)$ and $B_{o,i}(p) = \alpha_i B_o(p)$. This method will be called Hammerstein RIVC (HRIVC).

A. Refined IV for Hammerstein CT Models

The model set to be estimated, denoted as $\mathcal{M}$ with system ($\mathcal{G}$) and noise ($\mathcal{H}$) models parameterized independently, then takes the form,

$$\mathcal{M} \quad : \{G_i(p, \boldsymbol{\rho}), H(q, \boldsymbol{\eta})\}, i = 1 \dots l \tag{5}$$

where $\boldsymbol{\rho}$ and $\boldsymbol{\eta}$ are parameter vectors that characterise the system and noise model, respectively. In particular, the system model is formulated in CT terms:

$$\begin{aligned} \mathcal{G} : G_i(p, \boldsymbol{\rho}) &= \frac{B_i(p, \boldsymbol{\rho})}{A(p, \boldsymbol{\rho})}, \\ &= \frac{\alpha_i(b_0 p^{n_b} + b_1 p^{n_b - 1} + \dots + b_{n_b})}{p^{n_a} + a_1 p^{n_a - 1} \dots + a_{n_a}}, \end{aligned} \tag{6}$$

with $i = 1 \dots l$. The associated model parameters are stacked columnwise in the parameter vector,

$$\boldsymbol{\rho} = \begin{bmatrix} \mathbf{a} \\ \alpha_1 \mathbf{b} \\ \vdots \\ \alpha_l \mathbf{b} \end{bmatrix} \in \mathbb{R}^{n_\rho}, \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n_a} \end{bmatrix} \in \mathbb{R}^{n_a}, \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{n_b} \end{bmatrix} \in \mathbb{R}^{n_b+1}, \tag{7}$$

with $n_\rho = n_a + l(n_b+1)$ while the noise model is in discrete-time form

$$\mathcal{H} : H(q, \boldsymbol{\eta}) = \frac{C(q^{-1}, \boldsymbol{\eta})}{D(q^{-1}, \boldsymbol{\eta})} = \frac{1 + c_1 q^{-1} + \dots + c_{n_c} q^{-n_c}}{1 + d_1 q^{-1} + \dots + d_{n_d} q^{-n_d}} \tag{8}$$

where the associated model parameters are stacked columnwise in the parameter vector,

$$\boldsymbol{\eta} = \begin{bmatrix} c_1 & \cdots & c_{n_c} & d_1 & \cdots & d_{n_d} \end{bmatrix}^T \in \mathbb{R}^{n_c + n_d} \tag{9}$$

Consequently, the noise transfer function takes the usual ARMA model form:

$$\xi(t_k) = \frac{C(q^{-1}, \boldsymbol{\eta})}{D(q^{-1}, \boldsymbol{\eta})}e(t_k). \tag{10}$$

More formally, there exists the following decomposition of the parameter vector $\boldsymbol{\theta}$ for the whole hybrid model,

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\rho} \\ \boldsymbol{\eta} \end{pmatrix} \tag{11}$$

such that the model can be written in the form

$$y(t_k) = \frac{1}{A(p, \boldsymbol{\rho})} \sum_{i=1}^{l} B_i(p, \boldsymbol{\rho}) \gamma_i(u(t_k)) + \frac{C(q^{-1}, \boldsymbol{\eta})}{D(q^{-1}, \boldsymbol{\eta})} e(t_k),$$

with $B_i(p, \boldsymbol{\rho}) = \alpha_i B(p, \boldsymbol{\rho})$. The HRIVC method derives from the RIV algorithm for DT systems. This was evolved by converting the maximum likelihood estimation equations to a pseudo-linear form involving optimal prefilters [21], [26]. A similar analysis can be utilised in the present situation since the problem is very similar, in both algebraic and statistical terms.

Following the usual *prediction error minimisation* approach in the present hybrid situation, a suitable error function $\varepsilon(t_k)$, at kth sampling instant, is given as:

$$\varepsilon(t_k) = \frac{D(q^{-1}, \boldsymbol{\eta})}{C(q^{-1}, \boldsymbol{\eta})} \left\{ y(t_k) - \sum_{i=1}^{l} \frac{B_i(p, \boldsymbol{\rho})}{A(p, \boldsymbol{\rho})} \gamma_i(u(t_k)) \right\} \tag{12}$$

which can be written as

$$\varepsilon(t_k) = \frac{D(q^{-1}, \boldsymbol{\eta})}{C(q^{-1}, \boldsymbol{\eta})} \left\{ \frac{1}{A(p, \boldsymbol{\rho})} \left[A(p, \boldsymbol{\rho}) y(t_k) - \sum_{i=1}^{l} B_i(p, \boldsymbol{\rho}) \gamma_i(u(t_k)) \right] \right\}, \tag{13}$$

where the DT prefilter $D(q^{-1}, \boldsymbol{\eta})/C(q^{-1}, \boldsymbol{\eta})$ will be recognised as the inverse of the ARMA(n_c,n_d) noise model. However, since the polynomial operators commute in this linear case, (13) can be considered in the alternative form, by using for sake of clarity $u_i(t) = \gamma_i(u(t))$:

$$\varepsilon(t_k) = A(p, \boldsymbol{\rho}) y_{\mathrm{f}}(t_k) - \sum_{i=1}^{l} B_i(p, \boldsymbol{\rho}) u_{i\mathrm{f}}(t_k) \tag{14}$$

where $y_{\mathrm{f}}(t_k)$ and $u_{i\mathrm{f}}(t_k)$ represent the *sampled* outputs of the complete CT and DT prefiltering operation, involving the CT filtering operations using the filter (see [25], [24]):

$$f_{\mathrm{c}}(p, \boldsymbol{\rho}) = \frac{1}{A(p, \boldsymbol{\rho})}, \tag{15}$$

as well as DT filtering operations, using the inverse noise model filter:

$$f_{\mathrm{d}}(q^{-1}, \boldsymbol{\eta}) = \frac{D(q^{-1}, \boldsymbol{\eta})}{C(q^{-1}, \boldsymbol{\eta})}. \tag{16}$$

Therefore, from (14), the associated linear-in-the-parameters model then takes the form [25]:

$$y_{\mathrm{f}}^{(n_a)}(t_k) = \boldsymbol{\varphi}_{\mathrm{f}}^T(t_k) \boldsymbol{\rho} + \varsigma(t_k) \tag{17}$$

where

$$\boldsymbol{\varphi}_{\mathrm{f}}(t_k) = \begin{bmatrix} -\mathbf{y}_{\mathrm{f}} \\ \mathbf{u_{1f}} \\ \vdots \\ \mathbf{u_{lf}} \end{bmatrix}, \mathbf{y}_{\mathrm{f}} = \begin{bmatrix} y_{\mathrm{f}}^{(n_a-1)}(t_k) \\ y_{\mathrm{f}}^{(n_a-2)}(t_k) \\ \vdots \\ y_{\mathrm{f}}(t) \end{bmatrix},$$

$$\mathbf{u_{if}} = \begin{bmatrix} u_{i\mathrm{f}}^{(n_b)}(t_k) \\ u_{i\mathrm{f}}^{(n_b-1)}(t_k) \\ \vdots \\ u_{i\mathrm{f}}(t_k) \end{bmatrix},$$

$v^{(n)}(t_k)$ is the nth time derivative of $v(t)$ sampled at the kth sampling instant and $\varsigma(t_k) = A(p, \boldsymbol{\rho})\xi(t_k)$.

Of course none of $A(p, \boldsymbol{\rho})$, $B_i(p, \boldsymbol{\rho})$, $C(q^{-1}, \boldsymbol{\eta})$ or $D(q^{-1}, \boldsymbol{\eta})$ is known and only their estimates are available. Therefore, IV estimation normally involves an iterative (or relaxation) algorithm in which, at each iteration, the 'auxiliary model' used to generate the instrumental variables, as well as the associated prefilters, are updated, based on the parameter estimates obtained at the previous iteration [24], [25].

B. Iterative HRIVC Algorithm

Let us consider the jth iteration where we have access to the estimate:

$$\hat{\boldsymbol{\theta}}^{j-1} = \begin{pmatrix} \hat{\boldsymbol{\rho}}^{j-1} \\ \hat{\boldsymbol{\eta}}^{j-1} \end{pmatrix} \tag{18}$$

obtained at iteration $j-1$. The most important aspect of optimal IV estimation is the definition of the instrumental variable. It has been shown that this instrument requires the knowledge of the noise free regressor [19], [22]. Therefore, in this context, the associated optimal IV vector $\hat{\varphi}_{\mathrm{f}}(t_k)$, is then an estimate of the noise-free version of the vector $\varphi_{\mathrm{f}}(t_k)$ in (17) and is defined as follows:

$$\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t_k) = \begin{bmatrix} -\hat{\mathbf{x}}_{\mathrm{f}} \\ \mathbf{u_{1f}} \\ \vdots \\ \mathbf{u_{lf}} \end{bmatrix}, \hat{\mathbf{x}}_{\mathrm{f}} = \begin{bmatrix} \hat{x}_{\mathrm{f}}^{(n_a-1)}(t_k) \\ \hat{x}_{\mathrm{f}}^{(n_a-2)}(t_k) \\ \vdots \\ \hat{x}_{\mathrm{f}}(t_k) \end{bmatrix} \tag{19}$$

where the filtered noise-free output $\hat{x}_{\mathrm{f}}(t_k)$ is obtained from:

$$\hat{x}(t, \hat{\boldsymbol{\rho}}^{j-1}) = \sum_{i=1}^{l} G_i(p, \hat{\boldsymbol{\rho}}^{j-1}) u_i(t). \tag{20}$$

The IV optimisation problem can now be stated in the form

$$\hat{\boldsymbol{\rho}}^j(N) = \arg\min_{\boldsymbol{\rho}} \left\| \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t_k) \boldsymbol{\varphi}_{\mathrm{f}}(t_k)^T \right] - \left[\frac{1}{N} \sum_{k=1}^{N} \hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t_k) y_{\mathrm{f}}^{(n_a)}(t_k) \right] \right\|_{\mathbf{Q}}^2 \tag{21}$$

where $\|\mathbf{x}\|^{\mathbf{2}} = \mathbf{x}^{\mathbf{T}}\mathbf{Q}\mathbf{x}$ and $\mathbf{Q} = \mathbf{I}$. This results in the solution of the IV estimation equations:

$$\hat{\boldsymbol{\rho}}^j(N) = \left[\sum_{k=1}^{N} \hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t_k) \boldsymbol{\varphi}_{\mathrm{f}}^T(t_k) \right]^{-1} \sum_{k=1}^{N} \hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t_k) y_{\mathrm{f}}^{(n_a)}(t_k) \tag{22}$$

where $\hat{\boldsymbol{\rho}}^j(N)$ is the IV estimate of the system model parameter vector at the jth iteration based on the appropriately prefiltered input/output data $Z^N = \{u(t_k); y(t_k)\}_{k=1}^N$. if $G_{0,i} \in \mathcal{G}_i$, HRIVC provides a consistent estimate under the conditions: $\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mathbb{E}\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t_k) \boldsymbol{\varphi}_{\mathrm{f}}(t_k)^T$ is full column rank and $\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mathbb{E}\hat{\boldsymbol{\varphi}}_{\mathrm{f}}(t_k) \varsigma_{\mathrm{f}}(t_k) = 0$.

An estimate of the sampled noise signal $\xi(t_k)$, at the jth iteration, is obtained by subtracting the sampled output of the auxiliary model equation (20) from the measured output $y(t_k)$, *i.e.*:

$$\hat{\xi}(t_k) = y(t_k) - \hat{x}(t_k, \hat{\boldsymbol{\rho}}^{j-1}). \tag{23}$$

This estimate provides the basis for the estimation of the noise model parameter vector $\boldsymbol{\eta}^j$, using in this case the MATLAB identification toolbox ARMA estimation algorithm. The process is iterated until a stopping criterion or a certain number of iterations is reached. At the end of the iterative process, coefficients $\hat{\alpha}_i$ are not directly accessible. They are however deduced from polynomial $\hat{B}_i(p)$ as $B_i(p, \boldsymbol{\rho}) = \alpha_i B(p, \boldsymbol{\rho})$. The hypothesis $\alpha_1 = 1$ guarantees that $\hat{B}_1(p, \boldsymbol{\rho}) = \hat{B}(p, \boldsymbol{\rho})$ and $\hat{\alpha}_i$ can be computed from:

$$\hat{\alpha}_i = \frac{1}{n_b + 1} \sum_{k=0}^{n_b} \frac{\hat{b}_{i,k}}{\hat{b}_{1,k}}, \tag{24}$$

where $\hat{b}_{i,k}$ is the kth coefficient of polynomial term $\hat{B}_i(p, \boldsymbol{\rho})$ for $i = 2 \dots l$.

C. Comments

- A simplified version of HRIVC algorithm named HSRIVC follows the exact same theory for estimation of Hammerstein CT output-error models. It is mathematically described by, $C(q^{-1}, \boldsymbol{\eta}^j) = C_o(q^{-1}) = 1$ and $D(q^{-1}, \boldsymbol{\eta}^j) = D_o(q^{-1}) = 1$. All previous given equations remain true, and it suffices to estimate $\boldsymbol{\rho}^j$ as $\boldsymbol{\theta}^j = \boldsymbol{\rho}^j$. The implementation of HSRIVC is much simpler than HRIVC as there is no model noise estimation in the algorithm.
- The present paper considers CT model identification. However, the DT versions of both IV-based methods can be easily developed and will be also evaluated in the next section,
- Even if the proposed algorithm performs well, it is not statistically optimal as discussed in section III-C.

III. Numerical Examples

This section presents numerical evaluation of both suggested HRIVC and HSRIVC methods. For all presented examples, the non-linear block has a polynomial form, *i.e.* $\gamma_i(u(t)) = u^i(t), \forall i$ and $\bar{u}(t) = u(t) + 0.5u^2(t) + 0.25u^3(t)$, where $u(t)$ follows a uniform distribution with values between -2 and 2. To highlight the performance of CT model IV-based methods, two simulated systems are considered. All systems are simulated with a zero order hold on the input.

A. Second-order System

The linear dynamic block is first a second-order system described by:

$$G_o(p) = \frac{10p + 30}{p^2 + p + 5}. \tag{25}$$

The sampling time equals $T_s = 0.48$s. Based on this process, two different systems $\mathcal{S}_1$ and $\mathcal{S}_2$ are defined. $\mathcal{S}_1$ is a Hammerstein output error model and therefore

$$H_o(q) = 1.$$

while $\mathcal{S}_2$ is a Hammerstein Box–Jenkins model with:

$$H_o(q) = \frac{1}{1 - q^{-1} + 0.2q^{-2}}.$$

The models considered for estimation are:

$$\mathcal{M}_{HRIVC} \begin{cases} G(p, \boldsymbol{\rho}) = \dfrac{b_0 p + b_1}{p^2 + a_1 p + a_2}, \\ H(q, \boldsymbol{\eta}) = \dfrac{1}{1 + d_1 q^{-1} + d_2 q^{-2}}, \\ f(u(t)) = u(t) + \alpha_1 u^2(t) + \alpha_2 u^3(t) \end{cases} \tag{26}$$

for the HRIVC method and

$$\mathcal{M}_{HSRIVC} \begin{cases} G(p, \boldsymbol{\rho}) = \dfrac{b_0 p + b_1}{p^2 + a_1 p + a_2}, \\ H(q, \boldsymbol{\eta}) = 1, \\ f(u(t)) = u(t) + \alpha_1 u^2(t) + \alpha_2 u^3(t) \end{cases} \tag{27}$$

for HSRIVC.

500 Monte Carlo simulation runs with a new noise realization for each run were realized under a signal to noise ratio (SNR) of 30dB and 10dB with:

$$\text{SNR} = 20\log\left(\frac{P_e}{P_x}\right), \tag{28}$$

P_g being the average power of signal g. The number of samples is $N = 2000$. Table I exhibits the mean value of the estimated parameters, their standard deviation and their normalised root mean square error (RMSE) defined as:

$$\text{RMSE}(\hat{\theta}_j) = \sqrt{\frac{1}{N_{exp}} \sum_{i=1}^{N_{exp}} \left(\frac{\theta_j^o - \hat{\theta}_j(i)}{\theta_j^o}\right)^2},$$

with $\hat{\theta}_j$ the jth estimated parameter of $\boldsymbol{\theta}$.

Table I shows that the HRIVC and HSRIVC methods provide similar, unbiased estimates of the model parameters with reasonable standard deviations. Results obtained using the HRIVC algorithm, have standard deviations which are always smaller than the ones produced by HSRIVC. Even though, the HSRIVC algorithm based on an output-error model is a reasonable alternative to the full HRIVC algorithm based on a Box–Jenkins model.

B. Fourth-order System

The aim of this paper is not to compare direct continuous-time and indirect discrete-time model estimation methods. However, authors show through a chosen example the interest of using the direct CT methods with respect to the traditional DT methods. The linear part of the second system is based on a benchmark proposed by Rao and Garnier in [16] (see also [11]). It is a fourth-order, non-minimum phase system with complex poles. Its transfer function is given by:

$$G_o(p) = \frac{-6400p + 1600}{p^4 + 5p^3 + 408p^2 + 416p + 1600}. \tag{29}$$

The sampling frequency is chosen to be about ten times the bandwidth of the system under study which leads to $T_s = 0.0314$s. White noise is added to the output samples. 500 Monte Carlo simulation runs were realized with a SNR of 10dB using the proposed HSRIVC method and its discrete-time version HSRIV. The models take the forms:

$$\mathcal{M}_{HSRIVC} \begin{cases} G(p, \boldsymbol{\rho}) = \dfrac{b_o p + b_1}{p^4 + a_1 p^3 + a_2 p^2 + a_3 p + a_4}, \\ H(q, \boldsymbol{\eta}) = 1, \\ f(u(t)) = u(t) + \alpha_1 u^2(t) + \alpha_2 u^3(t) \end{cases} \tag{30}$$

for HSRIVC and

$$\mathcal{M}_{HSRIV} \begin{cases} G(q, \boldsymbol{\rho}) = \frac{\tilde{b_0} q^{-1} + \tilde{b_1} q^{-2} + \tilde{b_2} q^{-3} + \tilde{b_3} q^{-4}}{1 + \tilde{a_1} q^{-1} + \tilde{a_2} q^{-2} + \tilde{a_3} q^{-3} + \tilde{a_4} q^{-4}}, \\ H(q, \boldsymbol{\eta}) = 1, \\ f(u(t)) = u(t) + \alpha_1 u^2(t) + \alpha_2 u^3(t) \end{cases} \tag{31}$$

for HSRIV. Figures 2(a) and 3(a) display the magnitude Bode plots of the DT and CT estimated linear models. It can be firstly noticed that both models present similar results for low frequencies whereas for high frequencies, the CT method exhibits a superiority in model estimation. Both methods correctly estimate both resonance peaks. On the other side, the DT method appears to be less reliable, as for some realizations, the algorithm did not converge to acceptable values even though the initialization step is the same for both methods. By only looking at Bode diagram and considering only realizations which converged, both methods give satisfactory results. However, when looking at non-linear function estimations (Figures 2(b) and 3(b)), the DT method hands out results with a very large variance while the CT approach delivers a set of estimated functions centered nearly exactly on the true non-linear function. This can be explained by two facts: the DT version of the Hammerstein model (assuming the appropriate zero order hold) rises the number of parameters to be estimated for the numerator polynomial and therefore results in worse estimation. Furthermore, in the DT case, the numerator coefficients are so close to null that a small absolute error produces a large relative error. Estimated $\hat{\alpha}_i$ coefficients, which are directly deduced from $\hat{B}$ (see (24)), dramatically suffer from this particular situation.

TABLE I
ESTIMATION RESULTS FOR DIFFERENT NOISE MODELS

system	SNR	method		b_0	b_1	a_1	a_2	α_1	α_2	d_1	d_2
$\mathcal{S}_1$			true value	10	30	1	5	0.5	0.25		
	30	HSRIVC	$mean(\hat{\theta})$	9.9869	30.0251	1.0003	4.9996	0.5005	0.2507		
			$std(\hat{\theta})$	0.3053	0.6984	0.0084	0.0236	0.0113	0.0116		
			RMSE	0.0305	0.0233	0.0084	0.0047	0.0227	0.0464		
	10	HSRIVC	$mean(\hat{\theta})$	9.9834	29.8845	0.9987	4.9960	0.5061	0.2553		
			$std(\hat{\theta})$	0.9508	2.3145	0.0267	0.0789	0.0359	0.0383		
			RMSE	0.0950	0.0772	0.0267	0.0158	0.0728	0.1544		
$\mathcal{S}_2$			true value	10	30	1	5	0.5	0.25	-1	0.2
	30	HSRIVC	$mean(\hat{\theta})$	9.9957	29.8760	1.0001	4.9991	0.5026	0.2523		
			$std(\hat{\theta})$	0.3670	1.5660	0.0170	0.0436	0.0201	0.0180		
			RMSE	0.0367	0.0523	0.0169	0.0087	0.0405	0.0723		
		HRIVC	$mean(\hat{\theta})$	9.9906	30.0172	1.0006	5.0020	0.5008	0.2506	-1.0002	0.2005
			$std(\hat{\theta})$	0.2497	0.8954	0.0119	0.0265	0.0118	0.0115	0.0219	0.0223
			RMSE	0.0250	0.0298	0.0119	0.0053	0.0236	0.0460	0.0218	0.1112
	10	HSRIVC	$mean(\hat{\theta})$	10.0882	29.6146	1.0010	4.9814	0.5080	0.2604		
			$std(\hat{\theta})$	1.0764	4.4585	0.0517	0.1291	0.0610	0.0542		
			RMSE	0.1079	0.1490	0.0517	0.0261	0.1230	0.2208		
		HRIVC	$mean(\hat{\theta})$	10.049	30.0277	0.9998	4.9980	0.5015	0.2522	-0.9997	0.1994
			$std(\hat{\theta})$	0.7861	2.8278	0.0379	0.0871	0.0369	0.0366	0.0227	0.0219
			RMSE	0.0787	0.0942	0.0378	0.0174	0.0738	0.1466	0.0227	0.1096

C. Discussions

It can be noticed that results present a higher parameter variance than for a linear model estimation problem. This comes mainly from the redundancy of the $B(p)$ parameters contained in $\boldsymbol{\theta}$ and by the higher number of estimated parameters: when the Hammerstein model relies on only $n_a + l - 1 + n_b$ parameters, the proposed algorithm needs to estimate $n_a + l(n_b + 1)$ parameters. Hence, even if not optimal, this algorithm can produce a very good starting value for statistically optimal prediction error methods. However, the low variance in estimated parameters makes it an interesting method for practical data. An alternative RIV approach that can handle other types of nonlinearity, including nonlinear terms in variables other than the input, is 'state-dependent parameter' (SDP) estimation (e.g. [20]). Here, the parameters in the nonlinear function are estimated by a nonlinear, iterative optimization procedure in which the RIV estimation algorithm is incorporated to estimate the linear TF parameters, based on the nonlinearly transformed input. Although computationally less efficient, this is statistically more efficient than the method proposed in the present paper. Some further research about introducing constraint to avoid the parameters redundancy might be therefore relevant.

IV. CONCLUSION

The theory of multi-input single-output refined instrumental variable for CT systems has been applied to a non linear Hammerstein model composed of a linear dynamic CT Box-Jenkins transfer function and a non-linear function defined as the sum of known basis functions. The performance and consistency for both HSRIVC and HRIVC methods have been highlighted. Finally, some advantages of using the suggested CT method with respect to its DT version have been illustrated.

REFERENCES

[1] E-W. Bai. A blind approach to the Hammerstein-Wiener model identification. *Automatica*, 38, Issue 6:967–979, 2002.

[2] E-W. Bai. Identification of linear systems with hard input nonlinearities of known structure. *Automatica*, 38, Issue 5:853–860, May 2002.

[3] E-W. Bai and K-S. Chan. Identification of an additive nonlinear system and its applications in generalized Hammerstein models. *Automatica*, 44, Issue 2:430–436, February 2008.

[4] C. Bohn and H. Unbehauen. The application of matrix differential calculus for the derivation of simplified expressions in approximate non-linear filtering algorithms. *Automatica*, 36, Issue 10:1553–1560, October 2000.

[5] F. Ding and T. Chen. Identification of Hammerstein nonlinear ARMAX systems. *Automatica*, 41, Issue 9:1479–1489, September 2005.

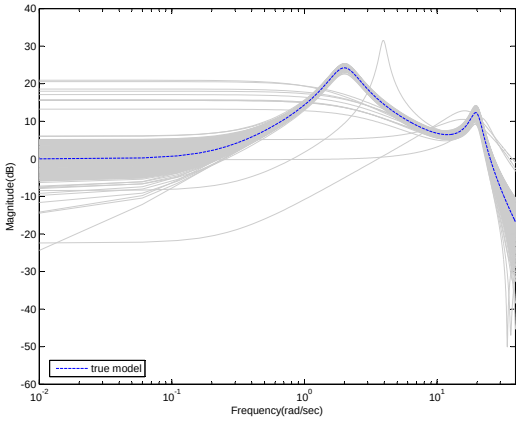
[6] F. Ding, Y. Shi, and T. Chen. Auxiliary model-based least-squares identification methods for Hammerstein output-error systems. *Systems & Control Letters*, 56, Issue 5:373–380, 2007.

[7] G. B. Giannakis and E. Serpedin. A bibliography on nonlinear system identification. *Signal Processing*, 81, Issue 3:533–580, March 2001.

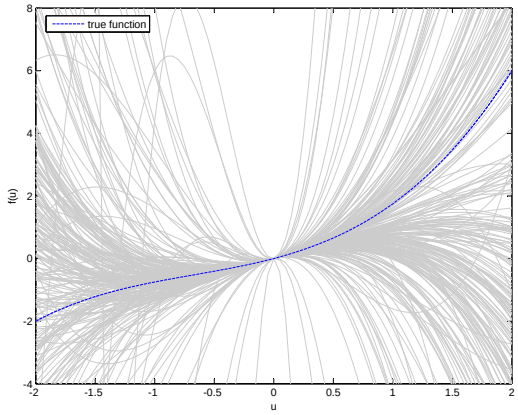
[8] I. Goethals, K. Pelckmans, J. A. K. Suykens, and B. De Moor. Identification of MIMO Hammerstein models using least squares support vector machines. *Automatica*, 41, Issue 7:1263–1272, July 2005.

[9] H. Garnier and L. Wang (Eds). *Identification of continuous-time models from sampled data*. Springer-Verlag, London, March 2008.

[10] Z. Q. Lang, S. A. Billings, R. Yue, and J. Li. Output frequency response function of nonlinear volterra systems. *Automatica*, 43, Issue 5:805–816, May 2007.

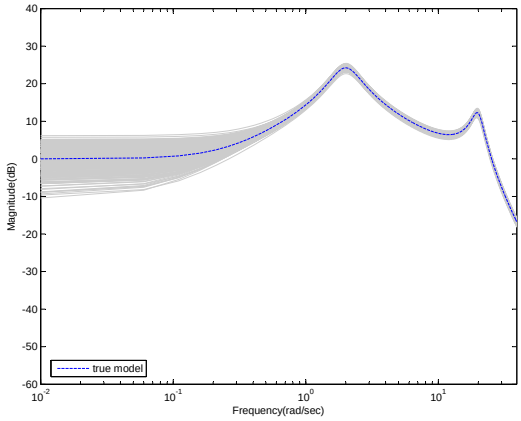


(a) Magnitude Bode plots of the identified DT HSRIV models together with the true system.

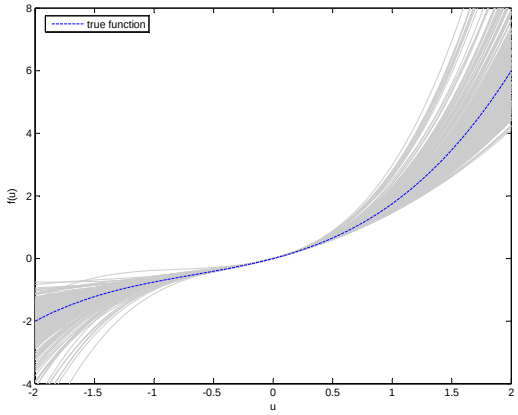


(b) Non-linear function estimated with DT HSRIV models together with the true non-linear function.

Fig. 2. DT model identification



(a) Magnitude Bode plots of the identified CT HSRIVC models together with the true system.



(b) Non-linear function estimated with CT HSRIVC models together with the true non-linear function.

Fig. 3. CT model identification

[11] L. Ljung. Initialisation aspects for subspace and output-error identification methods. In *European Control Conference (ECC'2003)*, Cambridge (U.K.), September 2003.

[12] K. Mahata and H. Garnier. Identification of continuous-time Box-Jenkins models with arbitrary time-delay. In *46th Conference on Decision and Control (CDC'2007)*, New Orleans, LA, USA, 12-14 December 2007.

[13] O. Nelles. *Nonlinear system identification*. Springer-Verlag, Berlin, 2001.

[14] H. J. Palanthandalam-Madapusi, B. Edamana, D. S. Bernstein, W. Manchester, and A. J. Ridley. Narmax identification for space weather prediction using polynomial radial basis functions. *46th IEEE Conference on Decision and Control*, New Orleans, LA, USA, 2007.

[15] A. S. Poznyak and L. Ljung. On-line identification and adaptive trajectory tracking for nonlinear stochastic continuous time systems using differential neural networks. *Automatica*, 37, Issue 8:1257–1268, August 2001.

[16] G. P. Rao and H. Garnier. Numerical illustrations of the relevance of direct continuous-time model identification. In *15th Triennial IFAC World Congress on Automatic Control*, Barcelona (Spain), 2002.

[17] G. P . Rao and H. Unbehauen. Identification of continuous-time systems. *IEE Proceedings on Control Theory Appl.*, 153(2), March 2006.

[18] J. Schoukens, W. D. Widanage, K. R. Godfrey, and R. Pintelon. Initial estimates for the dynamics of a Hammerstein system. *Automatica*, 43, Issue 7:1296–1301, July 2007.

[19] T. Söderström and P. Stoica. *Instrumental Variable Methods for System Identification*. Springer-Verlag, New York, 1983.

[20] P. C. Young. The identification and estimation of nonlinear stochastic systems, in A. I. Mees (Ed), *Nonlinear Dynamics and Statistics*, pages 127–166. Birkhauser: Boston, 2001.

[21] P. C. Young. Some observations on instrumental variable methods of time-series analysis. *International Journal of Control*, 23:593–612, 1976.

[22] P. C. Young. *Recursive Estimation and Time-series Analysis*. Springer-Verlag, Berlin, 1984.

[23] P. C. Young and H. Garnier. Identification and estimation of continuous-time, data-based mechanistic (DBM) models for environmental systems. *Environmental Modelling & Software*, 21, Issue 8:1055–1072, August 2006.

[24] P. C. Young, H. Garnier, and M. Gilson. An optimal instrumental variable approach for identifying hybrid continuous–time Box–Jenkins models. *14th IFAC Symposium on System Identification*, Newcastle, Australia:225–230, March 2006.

[25] P. C. Young, H. Garnier, and M. Gilson. Refined instrumental variable identification of continuous-time hybrid Box-Jenkins models, in H. Garnier and L. Wang (Eds), *Identification of continuous-time models from sampled data*, pages 91–132. Springer-Verlag, London, March 2008.

[26] P. C. Young and A. Jakeman. Refined instrumental variable methods of recursive time-series analysis - part III. extensions. *International Journal of Control*, 31, Issue 4:741–764, 1980.

[27] R. Johansson. Identification of Continuous–Time Models. *IEEE Transactions on Signal Processing*, 42, Issue 4:887–897, 1994.

[28] R. Pintelon, J. Schoukens and Y. Rolain Box–Jenkins continuous–time modeling. *Automatica*, 36, Issue 7:983–991, 2000.