Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2017) 25: 4311 – 4325

doi:10.3906/elk-1701-326

# Regularized estimation of Hammerstein systems using a decomposition-based iterative instrumental variable method

**Vikram SAINI*, Lillie DEWAN**
Department of Electrical Engineering, National Institute of Technology, Kurukshetra, India

**Received:** 31.01.2017 • **Accepted/Published Online:** 30.05.2017 • **Final Version:** 05.10.2017

**Abstract:** This paper presents a two-step instrumental variable (IV) method to obtain the regularized and consistent parameter estimates of the Hammerstein ARMAX model based on the bilinear parameterized form. The two-step identification method consists of estimating the bilinear parameters in the first step, followed by parameter reduction in the second step. An iterative identification method is proposed, based on the idea of separating the bilinear form in the two separable forms with partial parameters and solving the decomposed model forms iteratively. The IV-based estimation is integrated into the formulated decomposed structure by introducing the instruments constructed from the estimated auxiliary model outputs. It is shown that in a stochastic environment the proposed IV method produces consistent estimates in the presence of correlated noise disturbances. The validity of the proposed algorithm is verified with the help of extensive simulations using a Monte Carlo study.

**Key words:** Parameter estimation, least square method, instrumental variable method, identification, nonlinear models, Hammerstein models, stochastic models

## 1. Introduction

A Hammerstein system consists of an interconnection of the linear dynamic system and static non-linearity. Several identification methods have been developed to identify the Hammerstein systems (see [1–3] and references therein). These methods can be broadly classified as iterative and noniterative methods (see [4] for further classifications). The estimation problem using the noniterative method was first proposed by Chang and Luus [5] and an iterative identification method was proposed by Narendra and Gallman [1].

Most of the iterative identification methods transform the original single-input, single-output (SISO) model form into a multi-input, single-output (MISO) form by using bilinear parameterization. The identification scheme estimates the bilinear parameters in the first step followed by parameter reduction, resulting in a two-step procedure [1,6–8]. Therefore, the efficiency of the estimation procedure depends on both steps. On the other hand, some iterative methods estimate the two sets of parameters corresponding to the static nonlinearity and the linear system separately. It estimates the two sets of parameters within the framework of prediction error identification using nonlinear optimization. In general, the iterative procedure provides a simple and effective mechanism and reduces the optimization to a least square (LS) problem under the assumption that the noise affecting the output is white noise [5–9]. However, this assumption is too restrictive in practical applications and the choice of noise model structure has great importance in identifying the consistent estimates [9,10].

*Correspondence: vikram_1430@nitkkr.ac.in

Even though there has been extensive research on Hammerstein system identification, there are limited results on consistent and regularized estimation in the presence of correlated noise. An extended LS (ELS) identification method and the nonlinear adaptive algorithm have been reported for Hammerstein ARMAX models [11,12]. An iterative Hierarchical least squares method has been reported for controlled Hammerstein nonlinear controlled autoregressive systems [13]. Ding et al. presented an auxiliary model based on a LS method for Hammerstein OE models [14]. A data filtering-based iterative LS algorithm was proposed for MIMO linear and Hammerstein models [15,16]. An instrumental variable (IV) method using the nonlinear transformed instruments has been studied to improve the estimates' consistency [17]. For Hammerstein Box–Jenkin models, a refined IV method has been reported based on the extended framework [18]. Although the IV methods are consistent, the parameter reduction step and the regularization of nonlinear basis functions were not discussed. Recently, the regularization of basis function expansion was considered for the identification of a NL state space model [19].

The objective of this paper is to present a regularized iterative identification method, based on the IV estimation scheme to cope with correlated noise disturbances. The idea is to separate the Hammerstein model form into two separable forms with partial parameters and transform the problem into estimation of two decomposed forms. The consistent estimates are obtained by integrating the IV estimation in the proposed decomposed model form, by introducing the instruments constructed from the estimated auxiliary model outputs. The advantage of this procedure is the simultaneous regularization and an unbiased estimation of model parameters. The second contribution is to provide the efficient transformation of the bilinear parameters to the Hammerstein model parameters. The proposed method has been compared with the LS and the ELS method for different noise levels through extensive Monte Carlo simulations.

The paper is organized as follows: Section 2 describes the problem formulation. In Section 3, the regularized estimation based on the IV method along with certain conditions is proposed, followed by the investigation of an efficient parameter reduction step. In Section 4, the proposed method is compared with other methods through simulation examples. Finally, conclusions are drawn in Section 5.

In this paper $R$, $R^n$, $R^{n\times m}$, and $I_m$ denote the sets of real numbers, $n-$dimensional real vectors, $n\times m$ real matrices, and identity matrices of order $m$.The superscript $T$ denotes the matrix transpose. For $M=M^T\in R^{n\times n}$, $M\succ 0$ means that $M$ is positive definite.

## 2. Problem formulation and preliminaries

Consider the following discrete time Hammerstein ARMAX system:

$$y\left(k\right)=\frac{B\left(q^{-1}\right)}{A\left(q^{-1}\right)}f\left(u\left(k\right)\right)+\vartheta\left(k\right)=\frac{B\left(q^{-1}\right)}{A\left(q^{-1}\right)}f\left(u\left(k\right)\right)+\frac{C\left(q^{-1}\right)}{A\left(q^{-1}\right)}f\left(u\left(k\right)\right) \tag{1}$$

where $u\left(k\right)$ is the input signal, $y\left(k\right)$ is the output signal, and $\varepsilon\left(k\right)$ is a white noise with zero mean. Further, $A\left(\cdot\right)$, $B\left(\cdot\right)$ and $C\left(\cdot\right)$ are polynomials in the unit backward shift operator $q^{-1}$ $\left(q^{-1}y\left(k\right)=y(k-1)\right)$ of known orders $n_a$, $n_b$, and $n_c$, and are defined by

$$A\left(q^{-1}\right)=1+a_1q^{-1}+a_2q^{-2}+\ldots+a_{n_a}q^{-n_a} \tag{2a}$$

$$B\left(q^{-1}\right)=b_1q^{-1}+b_2q^{-2}+\ldots+b_{n_b}q^{-n_b} \tag{2b}$$

$$C\left(q^{-1}\right) = 1 + c_1 q^{-1} + c_2 q^{-2} + \ldots + c_{n_c} q^{-n_c} \tag{2c}$$

Let the nonlinearity $f(\cdot)$ be modelled using the linear combination of known basis functions such that $f\left(u\left(k\right)\right) = \sum_{i=1}^{n_d} d_i f_i\left(u\left(k\right)\right)$. Introducing the notations $\mathbf{a} = \left[a_1 \cdots a_{n_a}\right]^T \in R^{n_a}$, $\mathbf{b} = \left[b_1 \cdots b_{n_b}\right]^T \in R^{n_b}$, $\mathbf{d} = \left[d_1 \cdots d_{n_d}\right]^T \in R^{n_d}$ and substituting $f\left(u\left(k\right)\right)$ in (1), we have

$$y\left(k\right) = \boldsymbol{\varphi}_y^T\left(k\right)\mathbf{a} + \boldsymbol{\varphi}_u^T\left(k\right)\boldsymbol{\beta} + e\left(k\right) \;=\; \mathbf{z}^T\left(k\right)\boldsymbol{\theta} + e(k), \tag{3a}$$

where

$$e\left(k\right) = C\left(q^{-1}\right)\varepsilon\left(k\right) \tag{3b}$$

$\boldsymbol{\beta} = \left[b_1\mathbf{d}^T \cdots b_{n_b}\mathbf{d}^T\right]^T \in R^{n_b n_d}$, $\quad\boldsymbol{\varphi}_y\left(k\right) = \left[-y(k-1) \cdots -y(k-n_a)\right]^T \in R^{n_a}$,

$\boldsymbol{\varphi}_u\left(k\right) = \left[f_1(u(k-1) \cdots f_{n_d}(u(k-1) \cdots f_1(u(k-n_b)) \cdots f_{n_d}(u(k-n_b))\right]^T$ $\boldsymbol{\theta} = \left[\mathbf{a}^T \;\; \boldsymbol{\beta}^T\right]^T \in R^{n_a+n_b n_d}$ and $\mathbf{z}\left(k\right) = \left[\boldsymbol{\varphi}_y^T(k) \;\; \boldsymbol{\varphi}_u^T(k)\right]^T \in R^{n_a+n_b n_d}$.

For $N$ measurements of the input–output data pair $\left(u(k), y(k)\right)$ for $k \in T$ where $Card\left(T\right) = N \, and \, T \subset Z^+$, (3a) can be represented in a vector-matrix equation as given below:

$$\mathbf{Y} = \boldsymbol{\phi}_y\mathbf{a} + \boldsymbol{\phi}_u\boldsymbol{\beta} + \boldsymbol{E} \;=\; \boldsymbol{Z\theta} + \mathbf{E}, \tag{4}$$

where $\boldsymbol{\phi}_y = \left[\boldsymbol{\varphi}_y(1) \cdots \boldsymbol{\varphi}_y(N)\right]^T \in R^{N \times n_a}$, $\boldsymbol{\phi}_u = \left[\boldsymbol{\varphi}(1) \cdots \boldsymbol{\varphi}_u(N)\right]^T \in R^{N \times n_b n_d}$, $\mathbf{Y} = \left[y(1) \cdots Y(N)\right]^T \in R^N$, $\mathbf{E} = \left[e(1) \cdots e(N)\right]^T \in R^N$ and $\mathbf{Z} = \left[\mathbf{z}(1) \cdots \mathbf{z}(N)\right]^T \in R^{N \times (n_a+n_b n_d)}$.

Let us introduce an assumption to make the parameterization of (4) unique. For this the coefficient of the first basis function is assumed to be 1, i.e. $d_1 = 1$ [1,7,11–15,17–18]. Then the estimates of the coefficient vectors $\mathbf{a}$ and $\boldsymbol{\beta}$ can be obtained by considering the following minimization problem:

$$\left\|\mathbf{Y} - \boldsymbol{\phi}_y\mathbf{a} - \boldsymbol{\phi}_u\boldsymbol{\beta}\right\|_{\mathbf{2}}^{\mathbf{2}}, \tag{5}$$

which provides LS estimates. However, the LS estimation of the coefficients of basis functions to be used to model the nonlinearity often yields unregularized estimates and can suffer from high variance. On the other hand, the regularized estimates can be obtained by imposing the constraint on $\boldsymbol{\beta}$, i.e. by minimizing the following cost function:

$$J\left(\mathbf{a}, \boldsymbol{\beta}\right) := \frac{1}{2}\left(\mathbf{Y} - \boldsymbol{\phi}_y\mathbf{a} - \boldsymbol{\phi}_u\boldsymbol{\beta}\right)^T\left(\mathbf{Y} - \boldsymbol{\phi}_y\mathbf{a} - \boldsymbol{\phi}_u\boldsymbol{\beta}\right) + \frac{\lambda}{2}\boldsymbol{\beta}^T\boldsymbol{\Omega}\boldsymbol{\beta} \tag{6}$$

where $\boldsymbol{\Omega}$ is some positive semidefinite weight matrix and $\lambda \geq 0$ is a smoothing parameter that controls the trade-off between fidelity to the data and roughness of the basis functions. In particular, when $\boldsymbol{\Omega} = \mathbf{I}$, the penalty function in (6) takes the form of a ridge penalty [20]. Moreover, the large values of $\lambda$ will provide low variance in the estimates because the estimates concentrate around zero.

The following standard assumptions are made:

$A_1$ The input $u(k)$ is a persistently exciting ergodic random signal with zero mean so that ensemble average may be replaced by time averages over one sample function.

$A_2$ The signals $u(k)$ and $\varepsilon(k)$ are assumed to be independent with each other.

### 3. The extended IV identification method for the Hammerstein model

#### 3.1. The decomposition based method

In this paper, we propose to employ the decomposition form to estimate $\mathbf{a}$ and using the resultant estimate to yield a regularized estimate of $\boldsymbol{\beta}$. Now differentiating (6) with respect to $\mathbf{a}$ and $\boldsymbol{\beta}$ respectively, we obtain

$$\boldsymbol{\phi}_y^T\boldsymbol{\phi}_{\mathbf{y}}\mathbf{a}+\boldsymbol{\phi}_y^T\boldsymbol{\phi}_u\,\boldsymbol{\beta}=\boldsymbol{\phi}_y^T\mathbf{Y} \tag{7a}$$

$$\boldsymbol{\phi}_u^T\boldsymbol{\phi}_y\mathbf{a}+(\boldsymbol{\phi}_u^T\boldsymbol{\phi}_u+\lambda\boldsymbol{\Omega})\,\boldsymbol{\beta}=\boldsymbol{\phi}_u^T\mathbf{Y} \tag{7b}$$

Now two schemes can be adopted to obtain the estimates of $\mathbf{a}$ and $\boldsymbol{\beta}$. In the first scheme, we rearrange (7a) in the following block matrix equation:

$$\begin{bmatrix} \boldsymbol{\phi}_y^T\boldsymbol{\phi}_{\mathbf{y}} & \boldsymbol{\phi}_y^T\boldsymbol{\phi}_u \\ \boldsymbol{\phi}_u^T\boldsymbol{\phi}_y & \left(\boldsymbol{\phi}_u^T\boldsymbol{\phi}_u+\lambda\boldsymbol{\Omega}\right) \end{bmatrix}\begin{bmatrix} \mathbf{a} \\ \boldsymbol{\beta} \end{bmatrix}=\begin{bmatrix} \boldsymbol{\phi}_y^T\mathbf{Y} \\ \boldsymbol{\phi}_u^T\mathbf{Y} \end{bmatrix} \tag{8}$$

If the matrix on the left-hand side is nonsingular then the block matrix equation can be solved iteratively to find the estimates of $\mathbf{a}$ and $\boldsymbol{\beta}$. The second scheme employs the decomposed form. In order to proceed let $\boldsymbol{\phi}_u^T\boldsymbol{\phi}_u+\lambda\boldsymbol{\Omega}\succ 0$; then using (7b), the partial term $\boldsymbol{\phi}_u\boldsymbol{\beta}$ can be obtained as

$$\boldsymbol{\phi}_u\,\boldsymbol{\beta}=\boldsymbol{\phi}_u(\boldsymbol{\phi}_u^T\boldsymbol{\phi}_u+\lambda\boldsymbol{\Omega})^{-1}\boldsymbol{\phi}_u^T(\mathbf{Y}-\boldsymbol{\phi}_y\mathbf{a}) \tag{9}$$

Let $\mathbf{K}=\boldsymbol{\phi}_u(\boldsymbol{\phi}_u^T\boldsymbol{\phi}_u+\lambda\boldsymbol{\Omega})^{-1}\boldsymbol{\phi}_u^T\in R^{N\times N}$. Substituting (9) into (4), we obtain a decomposed form as follows:

$$\tilde{Y}=\tilde{\boldsymbol{\phi}}_y\mathbf{a}+\mathbf{E}, \tag{10}$$

where $\tilde{\mathbf{Y}}=[\tilde{\mathbf{y}}(1)\cdots\tilde{\mathbf{y}}(N)]^{\mathbf{T}}=(\mathbf{I}_N-\mathbf{K})\,\mathbf{Y}\in R^N$ and $\tilde{\boldsymbol{\phi}}_y=\left[\boldsymbol{\varphi}_y(1)\cdots\boldsymbol{\varphi}_y(N)\right]^T=(\mathbf{I}_N-\mathbf{K})\,\boldsymbol{\phi}_y\in R^{N\times n_a}$. The model form (10) is independent of the parameter vector $\boldsymbol{\beta}$ and is linear in the parameter vector $\mathbf{a}$ and therefore can be solved using the LS method, i.e.

$$\hat{\mathbf{a}}_{LS}=\left(\tilde{\boldsymbol{\phi}}_y^T\tilde{\boldsymbol{\phi}}_y\right)^{-1}\tilde{\boldsymbol{\phi}}_y^T\tilde{\mathbf{Y}} \tag{11}$$

Substituting the estimate $\hat{\mathbf{a}}_{LS}$ into 7(b) and solving yields the following:

$$\hat{\boldsymbol{\beta}}_{LS}=(\boldsymbol{\phi}_u^T\boldsymbol{\phi}_u+\lambda\boldsymbol{\Omega})^{-\mathbf{1}}\boldsymbol{\phi}_u^T(\mathbf{Y}-\boldsymbol{\phi}_y\hat{\mathbf{a}}_{LS}) \tag{12}$$

Note that the least square estimates $\hat{\mathbf{a}}_{LS}$ are biased and inconsistent due to the presence of correlated noise disturbance $e\,(k)$.

#### 3.2. The extended IV identification method

In the identification literature, the IV method has been widely used to obtain unbiased estimates [14,21–24]. Let $\boldsymbol{\psi}_y(k)\in R^{n_{\psi_y}}$ be a vector of dimensions $n_{\psi_y}\geq n_a$, called the vector of instruments. Define the following matrix and vector in asymptotic terms as

$$\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\psi}_y\,(k)\,\tilde{\boldsymbol{\varphi}}_y^T(k)=\hat{\mathbf{R}}_{\psi\tilde{\varphi}}\,(N)=\mathbf{R}_{\psi\tilde{\varphi}}=E\left[\boldsymbol{\psi}_y(k)\tilde{\boldsymbol{\varphi}}_y^T(k)\right]\;w.p.1 \tag{13a}$$

$$\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\psi}_y\left(k\right)\tilde{y}\left(k\right)\;=\hat{\mathbf{r}}_{\psi\tilde{y}}\left(N\right)=\mathbf{r}_{\psi\tilde{y}}=E\left[\boldsymbol{\psi}_y\left(k\right)\tilde{y}\left(k\right)\right]\;w.p.1, \tag{13b}$$

where $E$ represents the expectation operator and $w.p.1$ means convergence in probability. An extended IV (EIV) estimator of model form (10) is given by

$$\hat{\mathbf{a}}_{IV}(N)=\left(\hat{\mathbf{R}}_{\psi\tilde{\varphi}}^{T}\mathbf{W}\hat{\mathbf{R}}_{\psi\tilde{\varphi}}\right)^{-1}\hat{\mathbf{R}}_{\tilde{\varphi}}\mathbf{W}\hat{\mathbf{r}}_{\psi\tilde{y}} \tag{14}$$

The EIV estimates are consistent if the IV vector satisfies the following properties [21]:

$$\text{Rank }\hat{\mathbf{R}}_{\boldsymbol{\psi}\tilde{\varphi}}^{T}\hat{\mathbf{R}}_{\boldsymbol{\psi}\tilde{\varphi}}{=}n_a\;wp.1,\;\text{and} \tag{15a}$$

$$\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\psi}_y\left(k\right)e\left(k\right)=\hat{\mathbf{r}}_{\psi e}\left(N\right)=\mathbf{r}_{\psi e}=E\left[\boldsymbol{\psi}_y\left(k\right)e\left(k\right)\right]=0\;w.p.1 \tag{15b}$$

The IV vector can be defined by using noise-free output corresponding to the model form (10). Therefore, the instruments are constructed based on the definition of an auxiliary model that has a structure similar to that of (10). In order to proceed, the auxiliary model is defined by using the known input signals as

$$y_{aux}\left(k\right)=\frac{B_{aux}\left(q^{-1}\right)}{A_{aux}\left(q^{-1}\right)}f\left(u\left(k\right)\right), \tag{16}$$

where $A_{aux}\left(q^{-1}\right)$ and $B_{aux}\left(q^{-1}\right)$ are constant coefficient polynomials of the same order as polynomials $A\left(q^{-1}\right)$ and $B\left(q^{-1}\right)$ respectively. In addition, the function $f_{aux}(\cdot)$ is given in terms of basis functions of the order $n_d$ (see (3)).

Note that the computation of $y_{aux}\left(k\right)$ in (16) requires knowledge of the unknown coefficients of polynomials and the basis functions. In order to overcome this problem, first the auxiliary polynomials $A_{aux}\left(q^{-1}\right)$ and $B_{aux}\left(q^{-1}\right)$, and the coefficients of the basis functions are identified from signals $u(k)$ and $y(k)$ using the LS method (see Eqs. (11) and (12)). Then the noise-free signal $y_{aux}\left(k\right)$ is generated with the identified model. The assumptions $A_1$ and $A_2$ together with (3) imply that $u(k)$ and $e(s)$ are independent for $\forall k$ and $s$; therefore, the functions $f_i\left(u(k-j)\right) i=1,\ldots,n_d$ are also independent of the noise $e(k)$ (page 49 [25]), provided the basis functions satisfy

(i) $f_i$ is continuous over $R$ and is bounded in all bounded intervals in $R$.

(ii) $f_i(x)=O\left(e^{|x|^{\alpha}}\right)$ as $|x|\to\infty$ where $\alpha\in R$ and $\alpha<2$.

Then the estimated $y_{aux}\left(k\right)$ is independent of the noise $e(s)$ for all $k$ and $s$. Now referring to (10), the auxiliary noise-free output is obtained by the following:

$$\tilde{\mathbf{Y}}_{aux}=\left(\mathbf{I}_N{-}\mathbf{K}\right)\mathbf{Y}_{aux} \tag{17}$$

where $\tilde{\mathbf{Y}}_{aux}:=[\tilde{y}_{aux}(1)\cdots\tilde{y}_{aux}(N)]\in R^N$ and $\mathbf{Y}_{aux}:=[y_{aux}(1)\cdots y_{aux}(N)]\in R^N$. To formulate the IV vector, collect the obtained auxiliary outputs in (17) and rearrange in the vector form with delay to give

$$\boldsymbol{\psi}_y\left(k\right)=\left[-\tilde{y}_{aux}(k-1)\cdots-\tilde{y}_{aux}(k-n_{\psi_y})\right]^T \tag{18}$$

Since the asymptotic efficiency of the estimates depends on the instruments, the choice of instruments in (18) can be refined to include some optimality properties. This allows an additional filtering of the instruments with a filter such that the IV estimate obeys the minimum variance optimality criterion [9,24]. Therefore, the efficient IV estimate can be obtained by using the following filtered IV vector:

$$\boldsymbol{\psi}_y(k) = \left[-\tilde{y}_{aux}(k-1)\cdots-\tilde{y}_{aux}(k-n_{\psi_y})\right]^T \tag{19}$$

In the above equation, $\hat{y}_{aux}(k-i)\,\forall\; i=1,\ldots,n_{\psi_y}$ are obtained by filtering the estimated signal $\tilde{y}_{aux}(k-i)\,\forall i = 1,\ldots,n_{\psi_y}$ respectively with a linear time invariant filter given as

$$\hat{y}_{aux}(k-i) = C\left(q^{-1}\right)^{-1}\tilde{y}_{aux}(k-i)\;\;\forall\; i=1,2,\ldots,n_{\psi_y}, \tag{20}$$

where $C\left(q^{-1}\right)^{-1}$ is the inverse of $C\left(q^{-1}\right)$. It is assumed that $C(z)$ ($z$ being an arbitrary complex variable replacing $q^{-1}$) is stable and it has all zeros outside the unit circle. However, it is not feasible to use $C\left(q^{-1}\right)$ to form efficient instruments. This requires a prior knowledge of $C\left(q^{-1}\right)$, which is not available initially. In this case, the parameter vector $\mathbf{a}$ is replaced with its estimate $\hat{\mathbf{a}}$ in (10) to compute the residuals $\hat{\mathbf{E}} = \tilde{\mathbf{Y}} - \tilde{\boldsymbol{\phi}}_{\mathbf{y}}\hat{\mathbf{a}}$ and therefore allows an iterative LS procedure to update the estimate $\hat{C}\left(q^{-1}\right)$. Then the iterative estimate $\hat{\text{a}}_{IV}$ is computed using (14) and the IV estimate $\hat{\boldsymbol{\beta}}_{IV}$ can be obtained as

$$\hat{\boldsymbol{\beta}}_{IV}=\left((\hat{\mathbf{Q}}_{\psi\varphi}^T\hat{\mathbf{Q}}_{\psi\varphi}+\lambda\boldsymbol{\Omega}\right)^{-1}\hat{\mathbf{Q}}_{\psi\varphi}^T\hat{\mathbf{q}}_{\psi y} \tag{21}$$

In the above equation, the matrix and vector are defined as follows:

$$\hat{\mathbf{Q}}_{\psi\varphi}=\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\psi}_u(k)\,\boldsymbol{\varphi}_u^T(k) \tag{22a}$$

$$\hat{\mathbf{q}}_{\psi y}=\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{\psi}_u(k)\left(y(k)-\boldsymbol{\varphi}_y^T(k)\,\hat{\mathbf{a}}_{IV}\right) \tag{22b}$$

and the filtered IV vector $\boldsymbol{\psi}_u(k)$ is given as

$$\boldsymbol{\psi}_u(k) = \left[\hat{f}(u(k-1))\cdots\hat{f}_{n_d}(u(k-1)\cdots\hat{f}_1(u(k-_b))\cdots\hat{f}_{n_d}(u(k-n_b))\right]^T, \tag{23}$$

where

$$\hat{f}_i(u(k-j)) = C\left(q^{-1}\right)^{-1}f_i(u(k-j))\;\;\forall\; i=1,\ldots,n_d,\quad j=1,\ldots,n_b. \tag{24}$$

The resulting two-step regularized EIV algorithm based on the decomposed form for the parameter estimation of the Hammerstein ARMAX model is as follows:

***Step-1***

(i) Collect the data set $(u(k),y(k))_{k=1}^N$. Fit the data set $(u(k),y(k))$ using the LS method (c.f. (11) and (12)). Let the estimates obtained are $\hat{\mathbf{a}}_{LS}$ and $\hat{\boldsymbol{\beta}}_{LS}$. Set $\hat{\mathbf{a}}_1=\hat{\mathbf{a}}_{LS}$, $\hat{\boldsymbol{\beta}}_1=\hat{\boldsymbol{\beta}}_{LS}$ and $l=2$.

(ii) Compute the residuals $\hat{\mathbf{E}}_l = \tilde{\mathbf{Y}} - \tilde{\boldsymbol{\phi}}_y \hat{\mathbf{a}}_{l-1}$. Compute the estimate $\hat{C}_l\left(q^{-1}\right)$ using the estimated residual vector.

(iii) Form the data vectors $\mathbf{Y}$ and $\tilde{\mathbf{Y}}$, and the information matrices $\boldsymbol{\varphi}_u(k)$, $\boldsymbol{\varphi}_y(k)$, $\boldsymbol{\phi}_u, \boldsymbol{\phi}_y$ and $\tilde{\boldsymbol{\phi}}_y$ (c.f. (**??**), (4), and (10)).

(iv) Compute $y_{aux}(k)$ using $\hat{\mathbf{a}}_{l-1}$ and $\hat{\boldsymbol{\beta}}_{l-1}$ (c.f. (16)). Filter $\hat{y}_{aux}(k-r)$, $r = 1, \ldots, n_{\psi_y}$ and $f_i\left(u\left(k-j\right)\right)$ $i = 1, \ldots, n_d$, $j = 1, \ldots, n_b$ using $\hat{C}_l\left(q^{-1}\right)^{-1}$ (c.f. (20) and (24)).

(v) Form the instrumental vectors $\boldsymbol{\psi}_y(k)$ and $\boldsymbol{\psi}_u(k)$ using (19) and (23), respectively. Compute the IV estimates $\hat{\mathbf{a}}_{IV}$ and $\hat{\boldsymbol{\beta}}_{IV}$ according to (14) and (21).

(vi) If convergence occurs stop, else set $\mathbf{a}_l = \hat{\mathbf{a}}_{IV}$, $\hat{\boldsymbol{\beta}}_l = \hat{\boldsymbol{\beta}}_{IV}$, $l = l + 1$ and go to (ii)

***Step-2:*** Perform the parameter reduction step to get the estimate of the Hammerstein model.

### 3.3. Efficient parameter reduction

The parameter reduction step transforms the estimated bilinear parameters $\hat{\boldsymbol{\beta}}_{IV}$ to the model parameters $\hat{\mathbf{b}}$ and $\hat{\mathbf{d}}$ owing to the assumption $d_1 = 1$. In the identification literature, the transformation is based on the averaging principal [1,7,11–15,17,18]. In particular, let $\hat{\boldsymbol{\beta}}_{IV}(i)$ represents the $i$th element of $\hat{\boldsymbol{\beta}}_{IV}$; then referring to (3) and assumption $d_1 = 1$, the estimates of $d_j$, $j = 2, 3, \ldots, n_d$, can be computed by

$$\hat{b}_i = \hat{\boldsymbol{\beta}}_{IV}\left(\left(i-1\right)n_d + 1\right), \qquad i = 1, 2, 3, \ldots, n_b \tag{25a}$$

$$\hat{d}_j = \frac{1}{n_b}\sum_{i=1}^{n_b} \frac{\hat{\boldsymbol{\beta}}_{IV}\left(\left(i-1\right)n_d + j\right)}{\hat{b}_i} \qquad j = 2, 3, \ldots, n_d \tag{25b}$$

The estimates $\hat{d}_j$ are the arithmetic average of $n_b$ estimates. However, in this paper, the estimates $\hat{d}_j$ are calculated using the weighted average as follows:

$$\hat{d}_j = \frac{\sum\limits_{i=1}^{n_b} \hat{\boldsymbol{\beta}}_{IV}\left(\left(n-1\right)n_d + j\right)}{\sum\limits_{n=1}^{n_b} \hat{b}_n} \qquad j = 2, 3, \ldots\ldots, n_d \tag{26}$$

with the assumption that $\sum\limits_{i=1}^{n_b} \hat{b}_i \neq 0$. However, this assumption can be relaxed by redefining (26) in absolute terms. The transformation of bilinear estimates is given by

$$\hat{d}_j = \left(\frac{sign(\hat{\boldsymbol{\beta}}_{IV}\left(j\right))}{sign(\hat{b}_1)}\right)\left(\frac{\sum\limits_{i=1}^{n_b}\left|\hat{\boldsymbol{\beta}}_{IV}\left(\left(i-1\right)n_d + j\right)\right|}{\sum\limits_{i=1}^{n_b}\left|\hat{b}_i\right|}\right) \qquad j = 2, 3, \ldots\ldots, n_d, \tag{27}$$

where $|\cdot|$ and $sign$ represents the absolute value and the signum function, respectively.

## 4. Results and discussion

In this section, two numerical examples are presented to illustrate the performance of the proposed EIV method. For both examples, a data set $(N = 3000)$ is generated from the system driven by input $u(k)$ having zero mean unit variance with normal distribution, covering the sufficient range of the input nonlinearity. The proposed EIV method is implemented using two choices of weight matrix. First the EIV estimates are obtained using identity matrix $(\mathbf{W} = \mathbf{I})$. The second choice of the weight matrix is $\mathbf{W} = \left(\hat{\mathbf{R}}_{\boldsymbol{\psi\psi}}\right)^{-1}$ [22]. Further, the penalty function in (6) is implemented in the form of a ridge penalty such that matrix $\mathbf{\Omega} = \mathbf{I}$ and the parameter $\lambda$ is calculated using cross validation. The stochastic properties are investigated based on the performance criterion, i.e. statistical estimation error (SEE) and the average coefficient of variation (ACV). The SEE is defined as

$$SEE = \sqrt{\frac{\left\| m(\hat{\boldsymbol{\theta}}^{red}) - \boldsymbol{\theta}^{red} \right\|_2}{\left\| \boldsymbol{\theta}^{red} \right\|_2}} \tag{28}$$

where $m(\hat{\theta}^{red})$ represents the sample mean of estimator for 300 Monte Carlo simulation runs and $\hat{\theta}^{red} = \begin{bmatrix} \hat{a}^T & \hat{b}^T & \hat{d}^T \end{bmatrix}^T$, $d_1 = 1$. The ACV is defined as

$$ACV = \frac{1}{n_a + n_b + n_d} \sum_{i=1}^{n_a+n_b+n_d} \frac{\sigma\left(\hat{\theta}_i^{red}\right)}{\left| m(\hat{\theta}_i^{red}) \right|}, \tag{29}$$

where $\sigma\left(\hat{\theta}_i^{red}\right)$ represents the standard deviation of estimates.

### 4.1. Example 1

**Case 1:** Consider the following discrete Hammerstein model given as

$$A\left(q^{-1}\right) = 1 - 1.60q^{-1} + 0.80q^{-2}, \quad B\left(q^{-1}\right) = 0.85q^{-1} + 0.65q^{-2}$$

$$C\left(q^{-1}\right) = 1 - 1.0q^{-1} + 0.20q^{-2} \text{ and } f\left(u\left(k\right)\right) = 1.0u\left(k\right) + 0.50u^2\left(k\right) + 0.25u^3(k)$$

In this case, two different noise levels corresponding signal to noise Ratio[1] ($SNR = 14.29$ dB, and 06.94 dB) have been considered. For the two noise levels, the estimates computed using the parameter reduction steps discussed in section 3.3 are presented in Table 1. Note that the estimates obtained using (25a) and (25b) are briefly written as arithmetic parameter reduction (APR) and using (25a) and (27) as weighted parameter reduction (WPR). In addition, the proposed IV method is compared with the LS method (implemented without decomposition and regularization) [1] and the ELS method [11]. For the EIV estimates, the dimension of IV vector $n_{\psi_y}$ is found to be 8, i.e. $n_{\psi_y} = 8 > n_a = 2$. This choice is selected among different values of $n_{\psi_y}$ ranges between 3 and 10 based on the best performance criterion, i.e. the one with minimum SEE. The statistical properties, i.e. mean, standard deviation of estimated parameters, SEE, and ACV for $SNR = 14.29$ dB and 06.94 dB using the different methods are presented in Table 2.

[1] *The SNR is calculated as* $SNR = 10log_{10}\left(\frac{\sum_{k=1}^{N}(y(k)-\vartheta(k))^2}{\sum_{k=1}^{N}(\vartheta(k))^2}\right)$ *dB*

**Table 1.** Comparative performance of parameters reduction for Example 1 Case 1 ( $N = 3000$, 300 $runs$).

| | Parameters | | $b_1$ | $b_2$ | $d_2$ | $d_3$ |
|---|---|---|---|---|---|---|
| $SNR = 14.29$ | APR | mean | 0.8550 | 0.6484 | 0.5023 | 0.2521 |
| | | Std | 0.0648 | 0.0467 | 0.0164 | 0.0142 |
| | WPR | mean | 0.8550 | 0.6484 | 0.4996 | 0.2497 |
| | | Std | 0.0648 | 0.0467 | 0.0162 | 0.0143 |
| $SNR = 06.94$ | APR | mean | 0.8635 | 0.6452 | 0.5136 | 0.2624 |
| | | Std | 0.1467 | 0.1072 | 0.0449 | 0.0376 |
| | WPR | mean | 0.8635 | 0.6452 | 0.5002 | 0.2502 |
| | | Std | 0.1467 | 0.1072 | 0.0371 | 0.0330 |
| True value | | | 0.85 | 0.65 | 0.50 | 0.25 |

**Table 2.** Comparative performance of estimates for Example 1 Case 1 ( $N = 3000$, 300 runs).

| | Parameters | | $a_1$ | $a_2$ | $b_1$ | $b_2$ | $d_2$ | $d_3$ | SEE | ACV |
|---|---|---|---|---|---|---|---|---|---|---|
| $SNR = 14.29$ | LS | mean | –1.4864 | 0.6952 | 0.8542 | 0.7480 | 0.4947 | 0.2494 | 0.2775 | 0.0510 |
| | | Std | 0.0108 | 0.0101 | 0.0639 | 0.0640 | 0.0224 | 0.0196 | | |
| | ELS | mean | –1.5959 | 0.7943 | 0.8512 | 0.6554 | 0.4957 | 0.2498 | 0.0645 | 0.0432 |
| | | Std | 0.0041 | 0.0039 | 0.0480 | 0.0668 | 0.0163 | 0.0151 | | |
| | EIV $\mathbf{W} = \mathbf{I}$ | mean | –1.6002 | 0.8002 | 0.8550 | 0.6484 | 0.4996 | 0.2497 | 0.0472 | 0.0404 |
| | | Std | 0.0028 | 0.0025 | 0.0648 | 0.0467 | 0.0162 | 0.0143 | | |
| | EIV $\mathbf{W} = \left(\hat{\mathbf{R}}_{\psi\psi}\right)^{-1}$ | mean | –1.6000 | 0.8000 | 0.8542 | 0.6499 | 0.4996 | 0.2497 | 0.0420 | 0.0404 |
| | | Std | 0.0037 | 0.0033 | 0.0645 | 0.0463 | 0.0162 | 0.0143 | | |
| $SNR = 06.94$ | LS | mean | –1.2199 | 0.4578 | 0.8596 | 0.9760 | 0.4957 | 0.2505 | 0.5050 | 0.1079 |
| | | Std | 0.0253 | 0.0230 | 0.1373 | 0.1350 | 0.0502 | 0.0443 | | |
| | ELS | mean | –1.5824 | 0.7758 | 0.8529 | 0.6681 | 0.4844 | 0.2512 | 0.1272 | 0.0993 |
| | | Std | 0.0100 | 0.0096 | 0.1104 | 0.1527 | 0.0371 | 0.0358 | | |
| | EIV $\mathbf{W} = \mathbf{I}$ | mean | –1.6003 | 0.8004 | 0.8635 | 0.6452 | 0.5002 | 0.2502 | 0.0777 | 0.0923 |
| | | Std | 0.0066 | 0.0058 | 0.1467 | 0.1072 | 0.0371 | 0.0330 | | |
| | EIV $\mathbf{W} = \left(\hat{\mathbf{R}}_{\psi\psi}\right)^{-1}$ | mean | –1.5998 | 0.8000 | 0.8615 | 0.6489 | 0.5000 | 0.2502 | 0.0697 | 0.0920 |
| | | Std | 0.0086 | 0.0077 | 0.1455 | 0.1057 | 0.0370 | 0.0329 | | |
| | True value | | –1.60 | 0.80 | 0.85 | 0.65 | 0.50 | 0.25 | | |

Table 1 shows the qualitative form of the variances and the biases for the estimated coefficients of basis functions using APR and WPR steps. The variances of estimates $(d_1\, and\, d_2)$ using the WPR step are lower than those obtained from the APR step. From Table 2, it is observed that the estimates obtained using the LS method are heavily biased. This is expected since the LS method estimates the parameters based on the assumption of noise to be white. Therefore, the biases in the estimates are due to the misfit of the noise model. The ELS method improves the consistency and efficiency of estimates by suitably incorporating the noise model parameters in the LS estimation procedure. However, in both SNR cases the estimates obtained using the ELS method are slightly biased. The proposed method provides consistent estimates than the LS and the ELS methods, which is evident from the lower SEE and ACV values. Further, the weighted solution of the proposed method when implemented using $\mathbf{W} = \left(\hat{\mathbf{R}}_{\psi\psi}\right)^{-1}$ slightly improves the consistency and efficiency of the estimates than the unweighted ( $\mathbf{W} = \mathbf{I}$) estimates.

**Case 2:** The entire configuration is the same as case 1 except that the nonlinearity is described using the wavelet basis functions. In particular, the Mexican hat wavelet is chosen due to its attractive properties (see [26,27] and references therein for details about the wavelet functions). Then the function $f\left(u\left(k\right)\right)$ can be expanded as

$$f\left(u\left(k\right)\right)=\sum_{i=i_0}^{i_{\max}}\sum_{j\in J_i}d_{i,j}f_{i,j}\left(u\left(k\right)\right) \tag{30}$$

where $J_i(i=i_0,\, i_0+1,\ldots,i_{max})$ is a set, $f_{i,j}\left(u\left(k\right)\right)=2^{i/2}g\left(2^iu\left(k\right)-j\right)$ and $g(\cdot)$ represents the Mexican hat wavelet. Here $i_0\, and\, i_{max}$ are taken as 1 and 7, respectively. Note that the choice of $i_{max}$ is arbitrary to include the multiresolution property. Given resolution level $i$, the translation parameter $j$ can be determined as $-4\le j\le 2^i+4$ [26]. Since the presented method includes the regularization for the wavelet basis functions, the range of translation parameter is taken as $-4\le j\le t2^i$ for some $t\in Z^+$ such that $t2^i>2^i+4$. This means that the coefficients have negligible contribution for $j>2^i+4$ (for each resolution level $i$) and hence redundant to estimate. The value of $t$ is taken as 4 to satisfy $t2^i>2^i+4$. The dimension of IV vector $\boldsymbol{\psi}_y\left(k\right)$ is chosen to be 80 (i.e. $n_{\psi_y}=80$) using the cross-validation technique, i.e. this choice provides minimum SEE. In this case, the Monte Carlo study has been done with 100 simulation runs. For the case of SNR = 14.29 dB and SNR = 06.94 dB, the estimates are shown in Table 3, and the plots of estimated wavelet basis coefficients are shown in Figure.

**Table 3.** Comparative performance of estimates for Example 1 Case 2 ($N=3000,\ 1$ runs).

| | Parameters | | $a_1$ | $a_2$ | $d_1$ | $d_2$ | SEE | ACV |
|---|---|---|---|---|---|---|---|---|
| $SNR=14.29\,dB$ | LS | mean | –1.4874 | 0.6957 | 0.9059 | 0.9475 | 0.4035 | 0.1308 |
| | | Std | 0.0118 | 0.0114 | 0.3386 | 0.3664 | | |
| | EIV $\mathbf{W}=\mathbf{I}$ | mean | –1.6004 | 0.8009 | 0.8625 | 0.6654 | 0.0976 | 0.0593 |
| | | Std | 0.0125 | 0.0125 | 0.1207 | 0.1282 | | |
| $SNR=06.94\,dB$ | LS | mean | –1.2224 | 0.4593 | 0.7818 | 1.3890 | 0.6570 | 0.2806 |
| | | Std | 0.0255 | 0.0237 | 0.8073 | 0.8035 | | |
| | EIV $\mathbf{W}=\mathbf{I}$ | mean | –1.5969 | 0.7998 | 0.8453 | 0.6321 | 0.0948 | 0.1299 |
| | | Std | 0.0509 | 0.0525 | 0.2385 | 0.2526 | | |
| True value | | | –1.60 | 0.80 | 0.85 | 0.65 | | |

It is observed that in low as well as high noise the LS estimates provide inconsistent estimates with high SEE and ACV as compared to the EIV method. Note that in this particular case the SEE and ACV are calculated using the estimates $\hat{a}$ and $\hat{b}$. Further, Figure shows that the estimates of wavelet basis coefficients are regularized and unbiased using the proposed EIV method. The presented method provides regularization and shrinks the wavelet basis coefficients estimates towards zero relative to the LS estimates. The shrinkage of coefficients has the effect of reducing variance of estimates and the selection of appropriate wavelets basis coefficients.

### 4.2. Example 2

The considered discrete Hammerstein model is given as follows:

$$A\left(q^{-1}\right)=1-1.50q^{-1}+1.10q^{-2}-0.35q^{-3},\ B\left(q^{-1}\right)=0.80q^{-1}-0.55q^{-2}+0.90q^{-3}$$

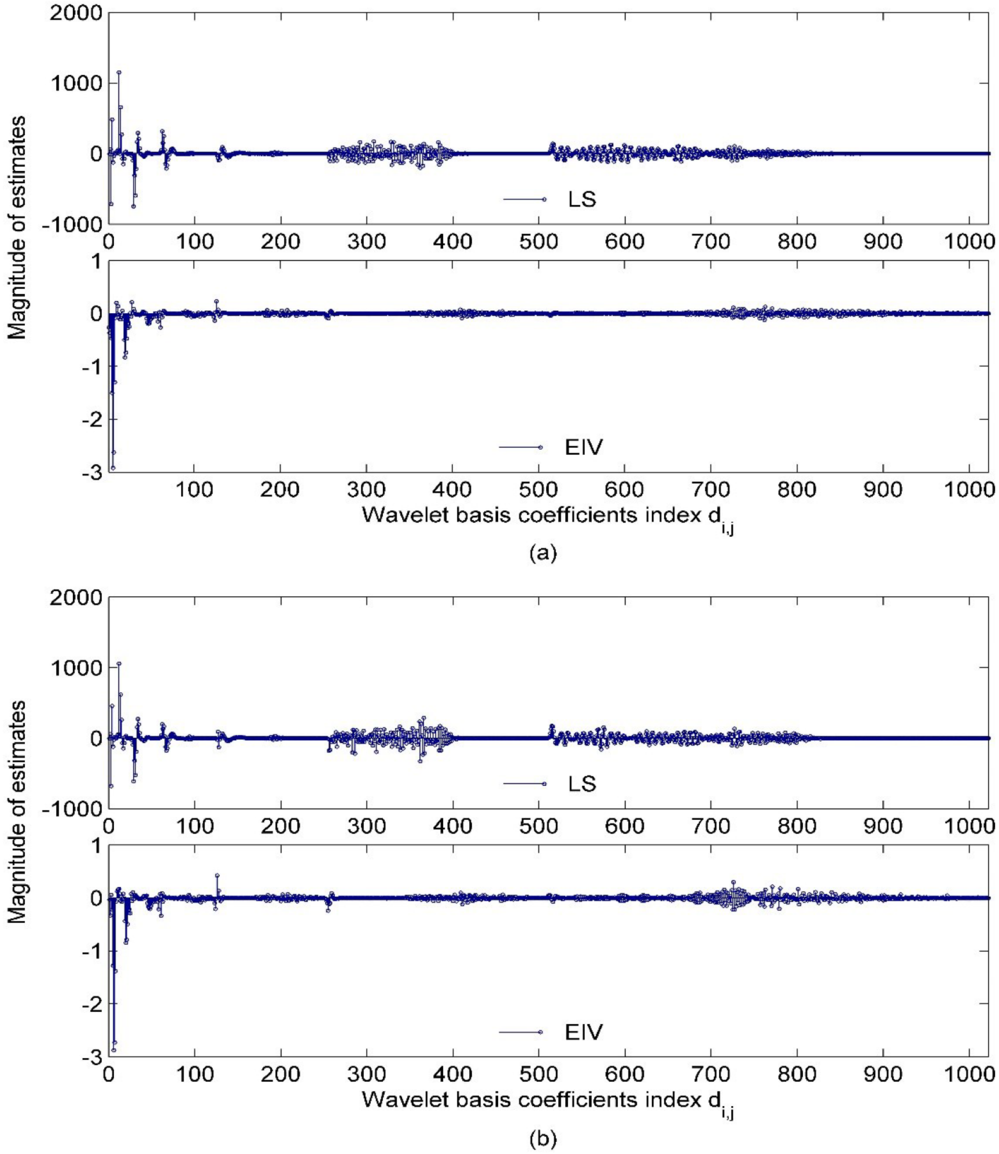


**Figure.** Comparative performance of wavelet basis coefficients estimates using the LS and the EIV $(W = I)$ method for the case (a) SNR = 14.29 dB (b) SNR = 06.94 dB.

$$f\left(u\left(k\right)\right) = 1.0u\left(k\right) + 0.80u^{2}\left(k\right) + 0.45u^{3}(k)$$

Again, the statistical properties are investigated using the 300 Monte Carlo simulations. Two different stochastic disturbance structures have been considered.

**Case 1:** The noise disturbances are generated by the constrained ARMA process, where the denominator polynomial is forced to be $A\left(q^{-1}\right)$. In particular, the noise is described as

$$v\left(k\right) = \frac{C\left(q^{-1}\right)}{A\left(q^{-1}\right)}\varepsilon\left(k\right) = \frac{1 - 0.70q^{-1} + 0.85q^{-2} - 0.50q^{-3}}{1 - 1.50q^{-1} + 1.10q^{-2} - 0.35q^{-3}}\varepsilon\left(k\right) \tag{31}$$

Two different noise levels corresponding to SNR = 13.61 dB and 05.70 dB are considered. Table 4 shows in the qualitative form the variances and the biases for the estimated coefficients of basis functions using APR and WPR steps. The comparative performances of the LS, the ELS, and the EIV method are listed in Tables 5 and 6.

**Table 4.** Comparative performance of parameter reduction for Example 2 Case 1 ($N = 3000$, $300\,runs$).

| | Parameters | | $b_1$ | $b_2$ | $b_3$ | $d_2$ | $d_3$ |
|---|---|---|---|---|---|---|---|
| $SNR = 13.61$ | APR | mean | 0.8030 | –0.5509 | 0.9016 | 0.7992 | 0.4496 |
| | | Std | 0.0353 | 0.0326 | 0.0246 | 0.0331 | 0.0253 |
| | WPR | mean | 0.8030 | –0.5509 | 0.9016 | 0.7981 | 0.4489 |
| | | Std | 0.0353 | 0.0326 | 0.0246 | 0.0262 | 0.0220 |
| $SNR = 05.70$ | APR | mean | 0.7994 | –0.5485 | 0.9074 | 0.8103 | 0.4576 |
| | | Std | 0.0860 | 0.0768 | 0.0564 | 0.0860 | 0.0669 |
| | WPR | mean | 0.7994 | –0.5485 | 0.9074 | 0.7987 | 0.4500 |
| | | Std | 0.0860 | 0.0768 | 0.0564 | 0.0613 | 0.0484 |
| True value | | | 0.80 | 0.55 | 0.90 | 0.80 | 0.45 |

**Table 5.** Comparative performance of estimates for Example 2 Case 1 ($N = 3000,\ SNR = 13.61\,dB$, 300 runs).

| Parameters | | $a_1$ | $a_2$ | $a_3$ | $b_1$ | $b_2$ | $b_3$ | $d_2$ | $d_3$ | SEE | ACV |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LS | mean | –1.3037 | 0.7471 | –0.1728 | 0.8023 | –0.3949 | 0.7462 | 0.8387 | 0.4504 | 0.4292 | 0.0526 |
| | std | 0.0165 | 0.0285 | 0.0144 | 0.0343 | 0.0376 | 0.0362 | 0.0341 | 0.0270 | | |
| ELS | mean | –1.4614 | 1.0227 | –0.3036 | 0.8026 | –0.5205 | 0.8646 | 0.8159 | 0.4497 | 0.2022 | 0.0361 |
| | std | 0.0108 | 0.0189 | 0.0103 | 0.0286 | 0.0341 | 0.0359 | 0.0285 | 0.0231 | | |
| EIV $\mathbf{W} = \mathbf{I}$ | mean | –1.5002 | 1.1002 | –0.3499 | 0.8030 | –0.5509 | 0.9016 | 0.7981 | 0.4489 | 0.0393 | 0.0306 |
| | std | 0.0079 | 0.0112 | 0.0059 | 0.0353 | 0.0326 | 0.0246 | 0.0262 | 0.0220 | | |
| EIV $\mathbf{W} = \left(\hat{\mathbf{R}}_{\psi\psi}\right)^{-1}$ | mean | –1.4997 | 1.0994 | –0.3496 | 0.8018 | –0.5497 | 0.9006 | 0.7988 | 0.4495 | 0.0303 | 0.0291 |
| | std | 0.0068 | 0.0102 | 0.0058 | 0.0340 | 0.0292 | 0.0241 | 0.0258 | 0.0216 | | |
| True value | | -1.50 | 1.1 | -0.35 | 0.80 | -0.55 | 0.90 | 0.80 | 0.45 | | |

**Table 6.** Comparative performance of estimates for Example 2 Case 1 ($N = 3000,\ SNR = 05.70\,dB$, 300 runs).

| Parameters | | $a_1$ | $a_2$ | $a_3$ | $b_1$ | $b_2$ | $b_3$ | $d_2$ | $d_3$ | SEE | ACV |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LS | mean | -1.0199 | 0.2586 | 0.0923 | 0.7949 | -0.1658 | 0.5459 | 0.9486 | 0.4552 | 0.6680 | 0.1625 |
| | std | 0.0223 | 0.0363 | 0.0195 | 0.0698 | 0.0777 | 0.0716 | 0.0940 | 0.0638 | | |
| ELS | mean | -1.3439 | 0.8202 | -0.1909 | 0.7975 | -0.4240 | 0.7830 | 0.8685 | 0.4524 | 0.3877 | 0.0938 |
| | std | 0.0276 | 0.0449 | 0.0226 | 0.0656 | 0.0762 | 0.0787 | 0.0706 | 0.0511 | | |
| EIV $\mathbf{W} = \mathbf{I}$ | mean | -1.5016 | 1.1042 | -0.3524 | 0.7994 | -0.5485 | 0.9074 | 0.7987 | 0.4500 | 0.0588 | 0.0726 |
| | std | 0.0218 | 0.0309 | 0.0156 | 0.0860 | 0.0768 | 0.0564 | 0.0613 | 0.0484 | | |
| EIV$\mathbf{W} = \left(\hat{\mathbf{R}}_{\psi\psi}\right)^{-1}$ | mean | -1.4991 | 1.1000 | -0.3505 | 0.7966 | -0.5452 | 0.9031 | 0.8013 | 0.4523 | 0.0519 | 0.0689 |
| | std | 0.0189 | 0.0284 | 0.0154 | 0.0825 | 0.0684 | 0.0558 | 0.0598 | 0.0470 | | |
| True value | | –1.50 | 1.1 | –0.35 | 0.80 | –0.55 | 0.90 | 0.80 | 0.45 | | |

It can be seen from Table 4 that the parameter reduction step using the WPR steps provides estimates of coefficients of basis functions with less bias and variance than using the APR steps, which is clearly evident in the low SNR case. From Tables 5 and 6, it is observed that the estimates obtained from the LS method are inconsistent. In both SNR cases, in spite of improving the statistical properties of estimates, the ELS estimates provide biased estimates. For the case of $SNR = 14.29\,dB$, the proposed method provides more consistent estimates than the ELS method and reduces both SEE and ACV even in the case when the SNR reduces to $06.94\,dB$. The weighted EIV estimates are identical to the estimates obtained from the unweighted EIV estimates with a slight improvement in the statistical properties.

**Case 2:** This case illustrates the robustness when the noise disturbances are generated using the low pass finite impulse response (FIR) filter. The Hammerstein ARMAX structure can effectively represent the stochastic FIR noise together with the deterministic part (Hammerstein system without noise). Further, the deterministic part has no structural errors. The transfer function describing the noise is given as

$$v(k) = P(z^{-1})\varepsilon\,(k)\,, \tag{32}$$

where $P(z^{-1})$ is low pass FIR digital filter of 12th order with cut-off frequency equal to 0.1 using a Hamming window technique in MATLAB. The ELS method is implemented using the ARMAX model structure with noise model order taken as the 2nd and 3rd order polynomial, i.e. $n_c = 2, 3$. The selection of the choice of noise model order is based on the performance criterion, i.e. the one with minimum SEE. The noise model order and the dimension of IV vector ($n_{\psi_y}$) for the proposed EIV method are found as $n_c = 3$ and $n_{\psi_y} = 7$, respectively. The statistical properties for the cases of SNR = 13.47 dB and SNR = 05.94 dB are listed in Tables 7 and 8, respectively.

**Table 7.** Comparative performance of estimates for Example 2 Case 2 ($N = 3000,\ SNR = 13.47\,dB$, 300 runs).

| Parameters | | $a_1$ | $a_2$ | $a_3$ | $b_1$ | $b_2$ | $b_3$ | $d_2$ | $d_3$ | SEE | ACV |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LS | mean | –1.5314 | 1.1066 | –0.3409 | 0.7994 | –0.5756 | 0.8849 | 0.7877 | 0.4510 | 0.1315 | 0.0160 |
| | Std | 0.0036 | 0.0026 | 0.0033 | 0.0099 | 0.0105 | 0.0111 | 0.0240 | 0.0179 | | |
| ELS $n_c = 2$ | mean | –1.5070 | 1.0796 | –0.3341 | 0.8002 | –0.5558 | 0.8791 | 0.7906 | 0.4504 | 0.1154 | 0.0144 |
| | Std | 0.0023 | 0.0036 | 0.0036 | 0.0074 | 0.0093 | 0.0112 | 0.0211 | 0.0155 | | |
| ELS $n_c = 3$ | mean | –1.5066 | 1.0784 | –0.3334 | 0.8001 | –0.5557 | 0.8783 | 0.7903 | 0.4505 | 0.1178 | 0.0146 |
| | Std | 0.0022 | 0.0037 | 0.0037 | 0.0075 | 0.0092 | 0.0115 | 0.0213 | 0.0156 | | |
| EIV $\mathbf{W} = \mathbf{I}$ | mean | –1.5001 | 1.1000 | –0.3500 | 0.7994 | –0.5502 | 0.9000 | 0.8007 | 0.4505 | 0.0201 | 0.0100 |
| | Std | 0.0023 | 0.0023 | 0.0016 | 0.0091 | 0.0070 | 0.0045 | 0.0147 | 0.0112 | | |
| EIV $\mathbf{W} = \left(\hat{\mathbf{R}}_{\psi\psi}\right)^{-1}$ | mean | –1.5001 | 1.1000 | –0.3500 | 0.7994 | –0.5502 | 0.8999 | 0.8007 | 0.4505 | 0.0198 | 0.0083 |
| | Std | 0.0022 | 0.0021 | 0.0014 | 0.0076 | 0.0053 | 0.0037 | 0.0130 | 0.0089 | | |
| True value | | –1.50 | 1.1 | –0.35 | 0.80 | –0.55 | 0.90 | 0.80 | 0.45 | | |

It is observed from Tables 5 and 6 that the LS estimates are highly biased and provide higher SEE and ACV. The ELS method reduces the SEE and ACV compared to the LS method. However, it still provides biased estimation for the case of SNR = 13.61, which is more prominent when the SNR further reduces to 05.70 dB. On the other hand, the EIV method provides consistent estimates and reduces both SEE and ACV significantly when compared with the values obtained using the LS and the ELS method.

## 5. Conclusions

In this paper, an IV-based formulation of the Hammerstein model based on the decomposed form has been analyzed in order to cope with the correlated noise disturbance. The basic advantage of the presented method

**Table 8.** Comparative performance of estimates for Example 2 Case 2 ( $N = 3000,\ SNR = 05.94\,dB$, 300 runs).

| Parameters | | $a_1$ | $a_2$ | $a_3$ | $b_1$ | $b_2$ | $b_3$ | $d_2$ | $d_3$ | SEE | ACV |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LS | mean | –1.6382 | 1.1391 | –0.3155 | 0.7969 | –0.6619 | 0.8410 | 0.7421 | 0.4566 | 0.2753 | 0.0360 |
| | Std | 0.0115 | 0.0075 | 0.0066 | 0.0209 | 0.0245 | 0.0231 | 0.0502 | 0.0432 | | |
| ELS $n_c = 2$ | mean | –1.5368 | 1.0124 | –0.2758 | 0.7985 | –0.5812 | 0.8049 | 0.7580 | 0.4551 | 0.2461 | 0.0344 |
| | Std | 0.0068 | 0.0106 | 0.0092 | 0.0164 | 0.0214 | 0.0248 | 0.0441 | 0.0369 | | |
| ELS $n_c = 3$ | mean | –1.5355 | 1.0064 | –0.2721 | 0.7989 | –0.5802 | 0.8009 | 0.7564 | 0.4548 | 0.2516 | 0.0346 |
| | Std | 0.0065 | 0.0109 | 0.0097 | 0.0157 | 0.0210 | 0.0257 | 0.0439 | 0.0363 | | |
| EIV $\mathbf{W} = \mathbf{I}$ | mean | –1.5000 | 1.0997 | –0.3501 | 0.7990 | –0.5503 | 0.8992 | 0.8005 | 0.4517 | 0.0290 | 0.0197 |
| | std | 0.0052 | 0.0056 | 0.0028 | 0.0184 | 0.0124 | 0.0090 | 0.0295 | 0.0221 | | |
| EIV $\mathbf{W} = \left(\hat{\mathbf{R}}_{\psi\psi}\right)^{-1}$ | mean | –1.4999 | 1.0997 | –0.3502 | 0.7990 | –0.5505 | 0.8991 | 0.8006 | 0.4519 | 0.0304 | 0.0191 |
| | Std | 0.0052 | 0.0050 | 0.0028 | 0.0179 | 0.0116 | 0.0089 | 0.0292 | 0.0214 | | |
| True value | | –1.50 | 1.1 | –0.35 | 0.80 | –0.55 | 0.90 | 0.80 | 0.45 | | |

is that it preserves the simpler structure and computational attractive properties of the IV method and provides regularized estimates of bilinear parameters. Specifically, it allows the transformation of the model into two decomposed forms by suitably incorporating the regularization of bilinear parameters. For the decomposed model form, a suitable choice of instruments based on the auxiliary model outputs has been discussed. It has been shown that the proposed algorithm allows consistent estimation of the model in low SNR environment without losing too much estimation accuracy. The performance of the proposed method has been verified and compared with the least square variants through relevant Monte Carlo study.

## References


[1] Narendra KS, Gallman PG. An iterative method for the identification of nonlinear systems using Hammerstein model. IEEE T Automat Contr 1966; 11: 546-550.

[2] Thathachar MAL, Ramaswamy S. Identification of a class of nonlinear systems. Int J Control 1973; 18: 741-752.

[3] Fruzzetti KP, Palazoglu A, McDonald KA. Nonlinear model predictive control using Hammerstein models. J Process Contr 1977; 7: 31-41.

[4] Janczak A. Identification of Nonlinear Systems Using Neural Networks and Polynomial Models: A Block-Oriented Approach. Berlin, Germany: Springer-Verlag, 2005.

[5] Chang FHI, Luus R. A noniterative method for identification using Hammerstein models. IEEE T Automat Contr 1971; 16: 464-468.

[6] Bai EW. An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems Automatica 1998; 34: 333-338.

[7] Vörös J. Iterative algorithm for parameter identification of Hammerstein systems with two-segment nonlinearities. IEEE T Automat Contr 1999; 44: 2145-2149.

[8] Bai EW, Chan K.S. Identification of an additive nonlinear system and its applications in generalized Hammerstein models. Automatica 2008; 44: 430-436.

[9] Ljung L. System Identification: Theory for the User, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[10] Hsia TC. System Identification, Least-squares Methods. Lexington, MA, USA: Lexington Books, 1977.

[11] Ding F, Chen T. Identification of Hammerstein nonlinear ARMAX systems. Automatica 2005; 41: 1479-1489.

[12] Chaudhary NI, Raja MAZ. Identification of Hammerstein nonlinear ARMAX systems using nonlinear adaptive algorithms. Nonlinear Dynam 2015; 79: 1385-1397.

[13] Chen H, Ding F. Hierarchical least squares identification for Hammerstein nonlinear controlled autoregressive systems. Circ Syst Signal Pr 2015; 34: 61-75.

[14] Ding F, Shi Y, Chen T. Auxiliary model-based least-squares identification methods for Hammerstein output-error systems. Syst Control Lett 2007; 56: 373-380.

[15] Wang DQ, Liu HB, Ding F. Highly efficient identification methods for dual-rate Hammerstein systems. IEEE T Contr Syst T 2015; 23: 1952-1960.

[16] Wang Y, Ding F. Filtering-based iterative identification for multivariable systems. IET Control Theory A 2016; 10: 894-902.

[17] Saini V, Dewan L. Instrument variable method based on nonlinear transformed instruments for Hammerstein system identification. J Vib Control published online ahead of print; January 2017; DOI: 10.1177/1077546317694770.

[18] Laurain V, Garnier H, Gilson M, Young P. Refined instrumental variable method for identification of Hammerstein continuous-time Box-Jenkins models. In: 47th IEEE Conference on Decision and Control; 9–11 December 2008; Cancun, Mexico. pp. 1386-1391.

[19] Svensson A, Schön T.B, Solin A, Särkkä S. Nonlinear state space model identification using a regularized basis function expansion. In: 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing; 13–16 December 2015; Cancun, Mexico. pp. 481-484.

[20] Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 1970; 12: 55-67.

[21] Soderstromt T, Stoica P. Comparison of some instrumental variable methods— consistency and accuracy aspects. Automatica 1981; 17: 101-115.

[22] Stoica P, Nehorai A. Non-iterative optimal min-max instrumental variable method for system identification. Int J Control 1988; 47: 1759-1769.

[23] Young PC. An instrumental variable method for real time identification of a noisy process. Automatica 1970; 6: 271-287.

[24] Stoica P, Soderstrom T. Optimal instrumental variable estimation and approximate implementations. IEEE T Automat Contr 1983; 28: 757-771.

[25] Grimmett G, Stirzaker D. Probability and Random Processes. Oxford, UK: Oxford University Press, 2001.

[26] Billings SA, Wei HL. A new class of wavelet networks for nonlinear system identification. IEEE T Neural Networ 2005; 16: 862-874.

[27] Truong NV, Wang L, Young PC. Nonlinear system modeling based on non- parametric identification and linear wavelet estimation of SDP models. Int J Control 2007; 80: 774-788.