Optimal Refined Instrumental Variable Methods for Identification of LPV Output-Error and Box-Jenkins Models

V. Laurain*, M. Gilson*, R. Tóth**, H. Garnier*

***Abstract*— Identification of Linear Parameter-Varying (LPV) systems in an Input-Output (IO) setting is investigated, focusing on the case when the noise part of the data generating system is an additive colored noise. In the Box-Jenkins (BJ) and Output-Error (OE) cases, it is shown that the currently available linear regression and Instrumental Variable (IV) methods from the literature are not optimal in terms of bias and variance of the estimates. To overcome the underlying problems, a statistically optimal Refined Instrumental Variable (RIV) method is introduced. The proposed approach is compared to the existing methods via a representative simulation example.**

I. Introduction

The common need for accurate and efficient control of today's industrial applications is driving the system identification field to face the constant challenge of providing better models of physical phenomena. Systems encountered in practice are often nonlinear or have time-varying nature. Dealing with models of such kind without any structure is often found infeasible in practice. This rises the need for system descriptions that form an intermediate step between *Linear Time-Invariant* (LTI) systems and nonlinear/time-varying plants. To cope with these expectations, the model class of *Linear Parameter-Varying* (LPV) systems provides an attractive candidate. In LPV systems the signal relations are considered to be linear just as in the LTI case, but the parameters are assumed to be functions of a measurable time-varying signal, the so-called *scheduling variable* $p : \mathbb{Z} \rightarrow \mathbb{P}$. The compact set $\mathbb{P} \subseteq \mathbb{R}^{n_\mathbb{P}}$ denotes the *scheduling space*. The LPV system class has a wide representation capability of physical processes and this framework is also supported by a well worked out and industrially reputed control theory. Despite the advances of the LPV control field, identification of such systems is not well developed.

The existing LPV identification approaches are almost exclusively formulated in discrete-time, commonly assuming static dependence on the scheduling parameter (dependence only on the instantaneous value of the scheduling variable), and they are mainly characterized by the type of LPV model structure used: *Input-Output* (IO) [2], [3], [21] *State Space* (SS) [14], [20], [5] or *Orthogonal Basis Functions* (OBFs) based models [19] (see [18] for an overview of existing methods). In the field of system identification, IO models are widely used as the stochastic meaning of estimation is much better understood for such models, for example via the *Prediction-Error* (PE) setting, than for other model structures. Often an important advantage of IO models is that they can be directly derived from physic/chemical laws in their continuous form. Therefore, it is more natural to express a given physical system through an IO operator form or transfer function modeling. A comparison between IO and SS models based approaches can be found in [17] for linear systems.

Among the available identification approaches of IO models, the interest for *Instrumental Variable* (IV) methods has been growing in the last years. The main reason of this increasing interest is that IV methods offer similar performance as extended *Least Square* (LS) methods or other *Prediction Error Minimization* (PEM) methods (see [15], [13]) and provide consistent results even for an imperfect noise structure which is the case in most practical applications. These approaches have been used in many different frameworks such as direct continuous-time [15], [10], nonlinear [11] or closed-loop identification [9], [8] and lead to optimal estimates if the system belongs to the model set defined.

In the LPV case, most of the methods developed for IO models based identification are derived under a linear regression form [21], [2], [1]. By using the concepts of the LTI PE framework, recursive LS and IV methods have been also introduced [7], [4]. However, it has been only recently understood how a PE framework can be established for the estimation of general LPV models [18]. Due to the linear regressor based estimation, the usual model structure in existing methods is assumed to be *auto regressive with exogenous input* (ARX). Even if a non-statistically optimal IV method has been recently introduced in [4] for LPV *Output Error* (OE) models, no method has been proposed so far to deal with colored noise in a statistically optimal way. Moreover, it can be shown that it is generally impossible to reach statistically optimal estimates by using linear regression as presented so far in the literature. These imply, that there is lack of an LPV identification approach, which is capable of statistically optimal estimation of LPV-IO models under colored noise conditions, e.g. as in a *Box-Jenkins* (BJ) setting, which is the case in many practical applications.

By aiming at fulfilling this gap, an optimal estimation method is developed in this paper for LPV-IO BJ discrete-time models in the SISO case. The properties of the method are compared to the existing theory showing the increased statistical performance of the estimation.

The paper is organized as follows: In Section II, the general class of LPV systems in an IO representation form

*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

**Delft Center for Systems and Control (DCSC), Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands, r.toth@tudelft.nl

is introduced pointing out the main difficulties presented. In Section III, linear regression in the LPV prediction error framework is analyzed and it is shown that such an estimation scheme even in a IV setting is statistically not optimal if the noise is not white. Moreover, a reformulation of the dynamical description of LPV data generating plants in the considered setting is introduced which makes possible the extension of LTI-IV methods to the LPV framework. In Section IV, statistically optimal LPV-IV methods are introduced and analyzed, while their performance increase compared to other methods is shown in Section V. Finally in Section VI, the main conclusions of the paper are drawn and directions of future research are indicated.

II. Problem description

A. System description

Consider the data generating LPV system described by the following equations:

$$\mathcal{S}_\mathrm{o}\begin{cases} A_\mathrm{o}(p_k,q^{-1})\chi_\mathrm{o}(t_k)=B_\mathrm{o}(p_k,q^{-1})u(t_{k-d}) \\ y(t_k)=\chi_\mathrm{o}(t_k)+v_\mathrm{o}(t_k) \end{cases} \tag{1}$$

where d is the delay, χ_o is the noise-free output, u is the input, v_o is the additive noise with bounded spectral density, y is the noisy output of the system and q is the time-shift operator, i.e. $q^{-i}u(t_k)=u(t_{k-i})$. $A_\mathrm{o}(p_k,q^{-1})$ and $B_\mathrm{o}(p_k,q^{-1})$ are polynomials in q^{-1} of degree n_a and n_b respectively:

$$A_\mathrm{o}(p_k,q^{-1})=1+\sum_{i=1}^{n_\mathrm{a}} a_i^\mathrm{o}(p_k)q^{-i}, \tag{2a}$$

$$B_\mathrm{o}(p_k,q^{-1})=\sum_{j=0}^{n_\mathrm{b}} b_j^\mathrm{o}(p_k)q^{-i}, \tag{2b}$$

where the coefficients a_i and b_j are real meromorphic functions, e.g. rational functions, with static dependence on p. It is assumed that these coefficients are non-singular on $\mathbb{P}$, thus the solutions of $\mathcal{S}_\mathrm{o}$ are well-defined and the process part is completely characterized by the coefficient functions $\{a_i^\mathrm{o}\}_{i=1}^{n_\mathrm{a}}$ and $\{b_j^\mathrm{o}\}_{j=0}^{n_\mathrm{b}}$.

Most of existing methods in the literature assume an ARX type of data generating system, which means that the noise process v_o can be written as

$$e_\mathrm{o}(t_k)=A_\mathrm{o}(p_k,q^{-1})v_\mathrm{o}(t_k), \tag{3}$$

where e_o is a zero-mean, discrete-time white noise process with a normal distribution $\mathcal{N}(0,\sigma_\mathrm{o}^2)$, where σ_o^2 is the variance. This assumption is unrealistic in most practical applications as it assumes that both the noise and the process part of $\mathcal{S}_\mathrm{o}$ contain the same dynamics. Often the colored noise associated with the sampled output measurement $y(t_k)$ has a rational spectral density which has no dependence on p. Therefore, it is a more realistic assumption that v_o is represented by a discrete-time *autoregressive moving average* (ARMA) model:

$$v_\mathrm{o}(t_k)=H_\mathrm{o}(q)e_\mathrm{o}(t_k)=\frac{C_\mathrm{o}(q^{-1})}{D_\mathrm{o}(q^{-1})}e_\mathrm{o}(t_k), \tag{4}$$

where $C_\mathrm{o}(q^{-1})$ and $D_\mathrm{o}(q^{-1})$ are monic polynomials with constant coefficients and with respective degree n_c and n_d. Furthermore, all roots of $z^{n_\mathrm{d}}D_\mathrm{o}(z^{-1})$ are inside the unit disc. It can be noticed that in case $C_\mathrm{o}(q^{-1})=D_\mathrm{o}(q^{-1})=1$, (4) defines an OE noise model.

B. Model considered

Next we introduce a discrete-time LPV Box-Jenkins (BJ) type of model structure that we propose for the identification of the data-generating system (1) with noise model (4). In the proposed model structure the noise model and the process model is parameterized separately.

1) Process model: The process model is denoted by $\mathcal{G}_\rho$ and defined in a form of a LPV-IO representation:

$$\mathcal{G}_\rho:\big(A(p_k,q^{-1},\rho),B(p_k,q^{-1},\rho)\big)=(\mathcal{A}_\rho,\mathcal{B}_\rho) \tag{5}$$

where the p-dependent polynomials A and B are parameterized as

$$\mathcal{A}_\rho\begin{cases} A(p_k,q^{-1},\rho)=1+\sum_{i=1}^{n_\mathrm{a}} a_i(p_k)q^{-i}, \\ a_i(p_k)=a_{i,0}+\sum_{l=1}^{n_\alpha} a_{i,l}f_l(p_k) \quad i=1,\dots,n_\mathrm{a} \end{cases}$$

$$\mathcal{B}_\rho\begin{cases} B(p_k,q^{-1},\rho)=\sum_{j=0}^{n_\mathrm{b}} b_j(p_k)q^{-i}, \\ b_j(p_k)=b_{j,0}+\sum_{l=1}^{n_\beta} b_{j,l}g_l(p_k) \quad j=0,\dots,n_\mathrm{b} \end{cases}$$

In this parametrization, $\{f_l\}_{l=1}^{n_\alpha}$ and $\{g_l\}_{l=1}^{n_\beta}$ are meromorphic functions of p, with static dependence, allowing the identifiability of the model (pairwise orthogonal functions on $\mathbb{P}$ for example). The associated model parameters ρ are stacked columnwise in the parameter vector,

$$\rho=\begin{bmatrix} \mathtt{a}_1 & \dots & \mathtt{a}_{n_\mathrm{a}} & \mathtt{b}_0 & \dots & \mathtt{b}_{n_\mathrm{b}} \end{bmatrix}^\top\in\mathbb{R}^{n_\rho}, \tag{7}$$

where

$$\mathtt{a}_i=\begin{bmatrix} a_{i,0} & a_{i,1} & \dots & a_{i,n_\alpha} \end{bmatrix}\in\mathbb{R}^{n_\alpha+1}$$
$$\mathtt{b}_j=\begin{bmatrix} b_{j,0} & b_{j,1} & \dots & b_{j,n_\beta} \end{bmatrix}\in\mathbb{R}^{n_\beta+1}$$

and $n_\rho=n_\mathrm{a}(n_\alpha+1)+(n_\mathrm{b}+1)(n_\beta+1)$. Introduce also $\mathcal{G}=\{\mathcal{G}_\rho\mid\rho\in\mathbb{R}^{n_\rho}\}$, as the collection of all process models in the form of (5).

2) Noise model: The noise model is denoted by $\mathcal{H}$ and defined as a LTI transfer function:

$$\mathcal{H}_\eta:(H(q,\eta)) \tag{8}$$

where H is a monic rational function given in the form of

$$H(q,\eta)=\frac{C(q^{-1},\eta)}{D(q^{-1},\eta)}=\frac{1+c_1q^{-1}+\ldots+c_{n_\mathrm{c}}q^{-n_\mathrm{c}}}{1+d_1q^{-1}+\ldots+d_{n_\mathrm{d}}q^{-n_\mathrm{d}}}. \tag{9}$$

The associated model parameters η are stacked columnwise in the parameter vector,

$$\eta=\begin{bmatrix} c_1 & \dots & c_{n_\mathrm{c}} & d_1 & \dots & d_{n_\mathrm{d}} \end{bmatrix}^\top\in\mathbb{R}^{n_\eta}, \tag{10}$$

where $n_\eta=n_\mathrm{c}+n_\mathrm{d}$. Additionally, denote $\mathcal{H}=\{\mathcal{H}_\eta\mid\eta\in\mathbb{R}^{n_\eta}\}$, the collection of all noise models in the form of (8).

3) Whole model: With respect to a given process and noise part $(\mathcal{G}_\rho, \mathcal{H}_\eta)$, the parameters can be collected as

$$\theta = \begin{bmatrix} \rho \\ \eta \end{bmatrix}, \tag{11}$$

and the signal relations of the LPV-BJ model, denoted in the sequel as $\mathcal{M}_\theta$, are defined as:

$$\mathcal{M}_\theta \begin{cases} A(p_k, q^{-1}, \rho)\chi(t_k) = B(p_k, q^{-1}, \rho)u(t_{k-d}) \\ v(t_k) = \dfrac{C(q^{-1}, \eta)}{D(q^{-1}, \eta)}e(t_k) \\ y(t_k) = \chi(t_k) + v(t_k) \end{cases} \tag{12}$$

Based on the previously defined model structure, the model set, denoted as $\mathcal{M}$, with process $(\mathcal{G}_\rho)$ and noise $(\mathcal{H}_\eta)$ models parameterized independently, takes the form

$$\mathcal{M}\left\{(\mathcal{G}_\rho, \mathcal{H}_\eta) \ \mid \ \mathrm{col}(\rho, \eta) = \theta \in \mathbb{R}^{n_\rho + n_\eta}\right\}. \tag{13}$$

This set corresponds to the set of candidate models in which we seek the model that explains data gathered from $\mathcal{S}_\mathrm{o}$ the best, under a given identification criterion (cost function).

C. Predictors and prediction error

Similar to the LTI case, in the LPV prediction error framework, one is concerned about finding a model in a given LPV model structure $\mathcal{M}$, which minimizes the statistical mean of the squared prediction error based on past samples of (y, u, p). However in the LPV case, no transfer function representation of systems is available. Furthermore, multiplication with q is not commutative over the p-dependent coefficients, meaning that $q^{-1}B(p_k, q^{-1})u(t_k) = B(p_{k-1}, q^{-1})u(t_{k-1})$ which is not equal to $B(p_k, q^{-1})u(t_{k-1})$. Therefore to define predictors with respect to models $\mathcal{M}_\theta \in \mathcal{M}$, a convolution type representation of the system dynamics, i.e. a LPV *Impulse Response Representation* (IRR), is used where the coefficients has dynamic dependence on p (dependence on future and past samples of p) [18]. This means that $\mathcal{S}_0$ with an asymptotically stable process and noise part is written as

$$y(t_k) = \underbrace{(G_\mathrm{o}(q) \diamond p)(t_k)u(t_k)}_{\chi_\mathrm{o}(t_k)} + \underbrace{(H_0(q) \diamond p)(t_k)}_{v_\mathrm{o}(t_k)} e_\mathrm{o}(t_k) \tag{14}$$

where

$$(G_\mathrm{o}(q) \diamond p)(t_k) = \sum_{i=0}^{\infty}(\alpha_i^\mathrm{o} \diamond p)(t_k)q^{-i}, \tag{15a}$$

$$(H_\mathrm{o}(q) \diamond p)(t_k) = 1 + \sum_{i=1}^{\infty}(\beta_i^\mathrm{o} \diamond p)(t_k)q^{-i}, \tag{15b}$$

with $\alpha_i^\mathrm{o} \diamond p$ expressing dynamic dependence of α_i on p, i.e. $\alpha_i^\mathrm{o} \diamond p = \alpha_i(p, qp, q^{-1}p, q^2p, \ldots)$. Now if p is deterministic and there exits a convergent adjoint $H_\mathrm{o}^\dagger$ of H_o such that

$$e_\mathrm{o}(t_k) = (H_\mathrm{o}^\dagger(q) \diamond p)(t_k)v_\mathrm{o}(t_k), \tag{16}$$

then it is possible to show (see [18]) that the *one-step ahead predictor* of y is

$$\begin{aligned} y(t_k \mid t_{k-1}) = \big((H_\mathrm{o}^\dagger(q)G_\mathrm{o}(q)) \diamond p\big)(t_k)\ u(t_k) \\ + \big((1 - H_\mathrm{o}^\dagger(q)) \diamond p\big)(t_k)\ y(t_k). \end{aligned} \tag{17}$$

In case the noise model is not dependent on p, like in (4), $(H_\mathrm{o}(q) \diamond p)(t_k) = \frac{C_\mathrm{o}(q^{-1})}{D_\mathrm{o}(q^{-1})}$ and $(H_\mathrm{o}^\dagger(q) \diamond p)(t_k) = \frac{D_\mathrm{o}(q^{-1})}{C_\mathrm{o}(q^{-1})}$. With respect to a parameterized model structure, we can define the *one-step ahead prediction error* as

$$\varepsilon_\theta(t_k) = y(t_k) - \hat{y}(t_k \mid t_{k-1}), \tag{18}$$

where

$$\begin{aligned} \hat{y}(t_k \mid t_{k-1}) = \big((H^\dagger(q, \theta)G(q, \theta)) \diamond p\big)(t_k)\ u(t_k) \\ + \big((1 - H^\dagger(q, \theta)) \diamond p\big)(t_k)\ y(t_k) \end{aligned} \tag{19}$$

with $G(q, \theta)$ and $H(q, \theta)$ the IRR's of the process and noise part respectively. Denote $\mathcal{D}_N = \{y(t_k), u(t_k), p(t_k)\}_{k=1}^N$ a data sequence of $\mathcal{S}_\mathrm{o}$. Then to provide an estimate of θ based on the minimization of ε_θ, an identification criterion $W(\mathcal{D}_N, \theta)$ can be introduced, like the *least squares* criterion

$$W(\mathcal{D}_N, \theta) = \frac{1}{N}\sum_{k=1}^{N}\varepsilon_\theta^2(t_k), \tag{20}$$

such that the parameter estimate is

$$\hat{\theta}_N = \arg \min_{\theta \in \mathbb{R}^{n_\rho + n_\eta}} W(\mathcal{D}_N, \theta). \tag{21}$$

D. Persistency of excitation

In order to estimate an adequate model in a given model set, most PEM algorithms like least squares or instrumental variables methods require that a *persistency of excitation* condition with respect to $\mathcal{D}_N$ collected from the system is satisfied. Such condition is required to guarantee consistency and convergence of the algorithm provided estimates. In order to analyze the estimation of the defined LPV-BJ model structure, the characterization of the data sets satisfying this condition with respect to $\mathcal{M}$ is needed.

In the LTI case, persistency of excitation is associated with the notion of an *informative data set*. Let $\mathcal{D}_N = \{y(t_k), u(t_k)\}_{k=1}^N$ be a data set of quasi-stationary u and y collected from the data generating system and let $W(\mathcal{D}_N, \theta)$ be an identification criterion. $\mathcal{D}_N$ is called informative with respect to a parametric model set $\mathcal{M}$ with parameters θ and a given $W(\mathcal{D}_N, \theta)$ if any two models in $\mathcal{M}$ can be distinguished under $W(\mathcal{D}_N, \theta)$ [6]. Basically this means that if the model set $\mathcal{M}$ is identifiable (no two parameters θ_1 and θ_2 given models correspond to the same predictor) and the data set $\mathcal{D}_N$ is informative, then $W(\mathcal{D}_N, \theta)$ has a global optimum in the statistical sense. The latter is the essential requirement for consistency of any minimization method. The notion of persistency of excitation with order n means in the LTI case an informative $\mathcal{D}_N$ with respect to a model structure with n parameters. The latter is equivalent to the possibility of statistically uniquely estimating a n^th order FIR filter based on $\mathcal{D}_N$.

In the LPV case, there are numerous differences. First of all, the requirements for identifiability imply that the linear combinations of the used f_l and g_l functions in the coefficient parametrization provide inequivalent dynamical behaviors of the model structure for each θ. With respect to the considered LPV-BJ structure and a LS criterion, a sufficient condition of identifiability is that $\{f_l\}_{l=1}^{n_\alpha}$ and $\{g_l\}_{l=1}^{n_\beta}$ are orthogonal on $\mathbb{P}$ and the polynomials A, B, C, D are co-prime for all $p_k \in \mathbb{P}$. Moreover, the notion of an informative data set in the LPV case is not equivalent to the condition of persistency of excitation with a given order. First of all, the model parameters θ are related to signals $f_l(p_k)q^{-i}y(t_k)$, $g_l(p_k)q^{-j}u(t_k)$, and not only the time-shifted versions of u and y, thus the functions f_l and g_l and the scheduling trajectory p together also influence the estimation of θ. Moreover, the estimation of the parameters of a LPV-FIR filter, irrelevant to the coefficient parametrization, is not equivalent with the estimation of a LPV-BJ model due to the non-commutativity of multiplication by q. This means that the terminology of persistency of excitation with order n is ill-defined in the LPV case. Instead, the informativity of the data sets with respect to the assumed coefficient parametrization and model order is needed to be satisfied in order to ensure consistency and convergence of the estimation. However, conditions of informative data sets have not been investigated directly in the LPV literature.

In [3] and [21], for the case of LPV-IO ARX models with polynomial dependence of the coefficients on the parameters, i.e. $f_l(p_k) = g_l(p_k) = p_k^l$, conditions for persistency of excitation has been investigated, unknowingly addressing the question of informativity of the data set. In these works it is assumed that a LPV system is a family of LTI systems associated with each points in $\mathbb{P}$. If the dependence on p is a polynomial of order $n_{\max} = \max(n_\beta, n_\alpha)$, then the LPV system identification can be realized by identifying at least $n_{\max}$ LTI models operating at distinct values of $\mathbb{P}$, and use them to determine the coefficients of the polynomial dependence based on the interpolation principle. This means that a data set $\mathcal{D}_N$ is informative if

- p visits at least $n_{\max}$ different points $\bar{\mathrm{p}}_1, \ldots, \bar{\mathrm{p}}_n \in \mathbb{P}$.
- Each sequence of u associated with a $\bar{\mathrm{p}}_l$ must be persistently exciting with respect to the LTI model corresponding to $\bar{\mathrm{p}}_l$.
- The number of revisits of each $\bar{\mathrm{p}}_l$ must be large enough to approximate the ergodicity condition in a given neighborhood around the considered points of $\mathbb{P}$.

However, this condition is rather conservative from a number of viewpoints. A LPV system can be considered as a set of LTI systems associated with each point of $\mathbb{P}$, but these systems share a common memory so they can describe the continuation of the signal trajectories when p changes. This means that in terms of the above given condition the variation of p must be infinitely slow in order to consider these systems to be independent LTI systems. However, ergodicity requires basically that the number of revisits of the chosen points must be infinite in the general case, which means that the p should vary as fast as possible to revisit these points more often. This concludes that the above given condition is too conservative for practical use. In [21], an improved version of this approach has been developed which tries to overcome the problem of conservativeness, however as it is based on the same principle of independent LTI system estimation, the question whether a data set is informative in the LPV case remains open.

E. Identification problem statement

Based on the previous considerations, the identification problem addressed in the sequel can now be defined.

Problem 1: Given a discrete time LPV data generating system $\mathcal{S}_\mathrm{o}$ defined as (1) and a data set $\mathcal{D}_N$ collected from $\mathcal{S}_\mathrm{o}$. Based on the LPV-BJ model structure $\mathcal{M}_\theta$ defined by (12), estimate the parameter vector θ using $\mathcal{D}_N$ under the following assumptions:

- A1 $\quad \mathcal{S}_\mathrm{o} \in \mathcal{M}$, i.e. there exits a $\mathcal{G}_\mathrm{o} \in \mathcal{G}$ and a $\mathcal{H}_\mathrm{o} \in \mathcal{H}$ such that $(\mathcal{G}_\mathrm{o}, \mathcal{H}_\mathrm{o})$ is equal to $\mathcal{S}_\mathrm{o}$.
- A2 $\quad$ In the parametrization $\mathcal{A}_\rho$ and $\mathcal{B}_\rho$, $\{f_l\}_{l=1}^{n_\alpha}$ and $\{g_l\}_{l=1}^{n_\beta}$ are chosen such that $(\mathcal{G}_\mathrm{o}, \mathcal{H}_\mathrm{o})$ is identifiable for any trajectory of p.
- A3 $\quad u(t_k)$ is not correlated to $e_\mathrm{o}(t_k)$.
- A4 $\quad \mathcal{D}_N$ is informative with respect to $\mathcal{M}$.
- A5 $\quad \mathcal{S}_\mathrm{o}$ is uniformly frozen stable, i.e. for any $\bar{\mathrm{p}} \in \mathbb{P}$, the roots of $z^{n_\mathrm{a}}(A(\bar{\mathrm{p}}, z^{-1}))$ are in the unit disc [18].

III. On the use of linear regression framework and statistical optimality

All methods for LPV-IO parametric identification proposed in the literature so far are based on linear regression methods such as least squares or instrumental variables [3] [4]. The currently accepted view in the literature is that if the system belongs to the model set defined in (13), then $y(t_k)$ can be written in the linear regression form:

$$y(t_k) = \varphi^\top(t_k)\rho + \tilde{v}(t_k) \tag{22}$$

with ρ as defined in (7),

$$\varphi(t_k) = \begin{bmatrix} -\mathrm{y}(t_k) f_0(p_k) \\ \vdots \\ -\mathrm{y}(t_k) f_{n_\alpha}(p_k) \\ \mathrm{u}(t_k) g_0(p_k) \\ \vdots \\ \mathrm{u}(t_k) g_{n_\beta}(p_k) \end{bmatrix} \in \mathbb{R}^{n_\rho}, \tag{23a}$$

$$\mathrm{y}(t_k) = \begin{bmatrix} y(t_{k-1}) \\ \vdots \\ y(t_{k-n_\mathrm{a}}) \end{bmatrix}, \quad \mathrm{u}(t_k) = \begin{bmatrix} u(t_{k-d}) \\ \vdots \\ u(t_{k-n_\mathrm{b}-d}) \end{bmatrix}, \tag{23b}$$

and

$$\tilde{v}(t_k) = A(p_k, q^{-1}, \rho) v(t_k). \tag{23c}$$

In this section it is shown why such a linear regression framework cannot lead to statistically optimal (unbiased and minimal variance) estimates when the model structure is a LPV Box–Jenkins. Let us first introduce the adjoint $A^\dagger$ of A, such that $\chi = A^\dagger(p_k, q^{-1}, \rho)u \Leftrightarrow A(p_k, q^{-1}, \rho)\chi = u$. Note that the adjoint always exits in a IRR sense with respect to an asymptotically stable A. In the LTI case, $A^\dagger = \frac{1}{A}$, however, in the LPV case, $A^\dagger \neq \frac{1}{A}$ due to the non-commutativity of the multiplication by q.

A. The conclusion brought in [4]

By considering (22) and the associated extended regressor in (23a), it is well known that the LS method leads to optimal estimate only if the noise model is ARX ($\tilde{v}(t_k)$ is a white noise). This condition implies that $v(t_k) = A^{\dagger}(p_k, q^{-1}, \rho)e(t_k)$ and is not fulfilled in many practical situations as $v_{\rm o}$ is often not related to the process itself and does not depend on p_k. Therefore it is proposed in [4] to use an IV method where the instrument is built using the simulated data generated from an estimated auxiliary ARX model. Instrumental variables have the particularity to produce unbiased estimates if the instrument is not correlated to the measurement noise. The algorithm proposed in [4] can be summed up as follows:

Algorithm 1 (One-step IV method):

Step 1 Estimate an ARX model by the LS method (minimizing (20)) using the extended regressor (23a).

Step 2 Generate an estimate $\hat{\chi}(t_k)$ of $\chi(t_k)$ based on the resulting ARX model of the previous step. Build an instrument based on $\hat{\chi}(t_k)$ and then estimate ρ using the IV method.

Based on the numerical simulation given in [4], the following conclusions have been proposed:

- In case $\mathcal{S}_{\rm o}$ corresponds to a LPV-OE model ($v_{\rm o} = e_{\rm o}$), Algorithm 1 leads to an unbiased estimate.
- The variance of the estimated parameters is much larger than in a LS estimation process as it is well-known.
- The estimation result can be improved if one uses a multi-step algorithm such as in [12].

B. Existing methods and optimal estimates

In the present paper, the authors only partially agree with the conclusions stated in [4]. It is true that the results can be improved and that the IV estimates are unbiased but this paper claims that:

- Even by using multi-step algorithm of [12], the optimal estimate cannot be reached with the linear regression form (22).
- The optimal estimates can be reached for LPV-BJ models by using IV methods and the variance in the estimated parameters might be lower than the LS estimator in given situations.

In the following part it is shown why these statements hold true. In order to show why statistically optimal estimation of the model (12) cannot be reached under the viewpoint (22), it is necessary to revisit the result of optimal prediction error in the LTI case.

1) The LTI case: In analogy with (12), consider the LTI-BJ model as

$$\mathcal{M}_{\theta}^{\rm LTI}\begin{cases} A(q^{-1},\rho)\chi(t_k) = B(q^{-1},\rho)u(t_{k-d}) \\ v(t_k) = \dfrac{C(q^{-1},\eta)}{D(q^{-1},\eta)}e(t_k), \\ y(t_k) = \chi(t_k) + v(t_k). \end{cases} \tag{24}$$

where $A(q^{-1}, \rho)$ and $B(q^{-1}, \rho)$ are polynomials in q^{-1} with constant real coefficients and have degree $n_{\rm a}$ and $n_{\rm b}$ respectively and e is a white noise with $e(t_k) \in \mathcal{N}(0, \sigma^2)$. $y(t_k)$ can be written in the linear regression form:

$$y(t_k) = \varphi^{\top}(t_k)\rho + \tilde{v}(t_k), \tag{25}$$

with

$$\rho = \left[\begin{array}{cccccc} a_1 & \ldots & a_{n_{\rm a}} & b_0 & \ldots & b_{n_{\rm b}} \end{array}\right]^{\top} \in \mathbb{R}^{n_{\rm a}+n_{\rm b}+1}$$

$$\varphi = \left[\, y(t_{k-1}) \ \ldots \ y(t_{k-n_{\rm a}}) \ \ u(t_{k-d}) \ \ldots \ u(t_{k-n_{\rm b}-d}) \,\right]^{\top}$$

and

$$\tilde{v}(t_k) = A(q^{-1},\rho)v(t_k). \tag{26}$$

Following the conventional PEM approach of the LTI framework (which is maximum likelihood estimation because of the normal distribution assumption on $e(t_k)$), the prediction error $\varepsilon_{\theta}(t_k)$ of (25) with respect to (24) is

$$\varepsilon_{\theta}(t_k) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)A(q^{-1},\rho)}\Big(A(q^{-1},\rho)y(t_k) - B(q^{-1},\rho)u(t_k)\Big), \tag{27}$$

where the filter $D(q^{-1}, \eta)/C(q^{-1}, \eta)$ can be recognized as the inverse of the ARMA($n_{\rm c}$,$n_{\rm d}$) noise model in (24). The polynomial operators commute and therefore $\varepsilon_{\theta}(t_k)$ is equivalent to the error function $\varepsilon_{*}(t_k)$ defined as:

$$\varepsilon_{*}(t_k) = A(p_k, q^{-1},\rho)y_{\rm f}(t_k) - B(p_k, q^{-1},\rho)u_{\rm f}(t_k), \tag{28}$$

where $y_{\rm f} = Q(q^{-1}, \theta)y$ and $u_{\rm f} = Q(q^{-1}, \theta)u$ represent the outputs of the prefiltering operation with

$$Q(q^{-1},\theta) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)A(q^{-1},\rho)}. \tag{29}$$

Therefore (25) is equivalent to:

$$y_{\rm f}(t_k) = \varphi_{\rm f}^{\top}(t_k)\rho + \tilde{v}_{\rm f}(t_k) \tag{30}$$

with

$$\tilde{v}_{\rm f}(t_k) = A(q^{-1},\rho)v_{\rm f}(t_k) = e(t_k). \tag{31}$$

In other words, if the optimal filter (29) is known *a priori*, it is possible to filter the data such that the estimation problem is reduced to the maximum likelihood estimation. This implies that a simple LS algorithm applied to the data prefiltered with (29) leads to the statistically optimal estimate under minor conditions.

2) The LPV case: Following the above introduced PEM approach in the LPV case (which is again maximum likelihood estimation because of the normal distribution assumption on $e(t_k)$), the prediction error $\varepsilon_{\theta}(t_k)$ of (22) with respect to (12) is

$$\varepsilon_{\theta}(t_k) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)}A^{\dagger}(p_k, q^{-1},\rho)\Big(A(p_k, q^{-1},\rho)y(t_k) - B(p_k, q^{-1},\rho)u(t_k)\Big) \tag{32}$$

where $D(q^{-1}, \eta)/C(q^{-1}, \eta)$ can be again recognized as the inverse of the ARMA($n_{\rm c}$,$n_{\rm d}$) noise model of (12).

In opposition to the LTI case, the polynomial operators do not commute in the LPV case as it has been shown in

Section II-C. Hence, no filter can be chosen such that both conditions

$$A(p_k, q^{-1}, \rho) y_{\mathrm{f}}(t_k) = \tfrac{D(q^{-1}, \eta)}{C(q^{-1}, \eta)} A^{\dagger}(p_k, q^{-1}, \rho) A(p_k, q^{-1}, \rho) y(t_k)$$
$$B(p_k, q^{-1}, \rho) u_{\mathrm{f}}(t_k) = \tfrac{D(q^{-1}, \eta)}{C(q^{-1}, \eta)} A^{\dagger}(p_k, q^{-1}, \rho) B(p_k, q^{-1}, \rho) u(t_k)$$

are fulfilled simultaneously. Consequently, no filtering of the data can lead to a regression equation

$$y_{\mathrm{f}}(t_k) = \varphi_{\mathrm{f}}^{\top}(t_k)\rho + \tilde{v}_{\mathrm{f}}(t_k) \tag{33}$$

which is equivalent to (22) and where $\tilde{v}_{\mathrm{f}}$ is white. In other words, by choosing φ such as in (23a) and therefore by assuming (22) (as in [3] and [4]) it is not possible to transform the estimation problem of (12) into a maximum likelihood estimation problem. The latter implies that no method proposed so far in the literature for solving the estimation of LPV-IO models or LTI-IO models can lead to optimal estimate in the LPV-BJ case by assuming the regression form (22). As a consequence, the existing theory needs to be modified in order to solve the identification problem stated in Section II-E.

C. Reformulation of the model equations

In order to introduce a method which provides a solution to the identification problem of LPV-BJ models, rewrite the signal relations of (12) as

$$\mathcal{M}_{\theta} \begin{cases} \underbrace{\chi(t_k) + \sum_{i=1}^{n_{\mathrm{a}}} a_{i,0}\chi(t_{k-i})}_{F(q^{-1})\chi(t_k)} + \sum_{i=1}^{n_{\mathrm{a}}}\sum_{l=1}^{n_{\alpha}} a_{i,l} \underbrace{f_l(p_k)\chi(t_{k-i})}_{\chi_{i,l}(t_k)} = \\ \qquad \sum_{j=0}^{n_{\mathrm{b}}}\sum_{l=0}^{n_{\beta}} b_{j,l} \underbrace{g_l(p_k)u(t_{k-d-j})}_{u_{j,l}(t_k)} \\ v(t_k) = \dfrac{C(q^{-1},\eta)}{D(q^{-1},\eta)} e(t_k) \\ y(t_k) = \chi(t_k) + v(t_k) \end{cases} \tag{34}$$

where $F(q^{-1}) = 1 + \sum_{i=1}^{n_{\mathrm{a}}} a_{i,0} q^{-i}$. Note that in this way the LPV-BJ model is rewritten as a *Multiple-Input Single-Output* (MISO) system with $(n_{\mathrm{b}}+1)(n_{\beta}+1) + n_{\mathrm{a}} n_{\alpha}$ inputs $\{\chi_{i,l}\}_{i=1,l=1}^{n_{\mathrm{a}},n_{\alpha}}$ and $\{u_{j,l}\}_{j=0,l=0}^{n_{\mathrm{b}},n_{\beta}}$ as represented in Fig. 1. Given the fact that the polynomial operator commutes in this representation ($F(q^{-1})$ does not depend on p_k), (34) can be rewritten as

$$\begin{aligned} y(t_k) = &-\sum_{i=1}^{n_{\mathrm{a}}}\sum_{l=1}^{n_{\alpha}} \frac{a_{i,l}}{F(q^{-1})}\chi_{i,l}(t_k) \\ &+ \sum_{j=0}^{n_{\mathrm{b}}}\sum_{l=0}^{n_{\beta}} \frac{b_{j,l}}{F(q^{-1})} u_{k,j}(t_k) + H(q)e(t_k), \end{aligned} \tag{35}$$

which is a LTI representation. As (35) is an equivalent form of the model (12), thus under the Assumption A1, it holds that the data generating system $\mathcal{S}_{\mathrm{o}}$ has also a MISO-LTI interpretation.

IV. Refined Instrumental variable for LPV systems

Based on the MISO-LTI formulation (35), it becomes possible to achieve optimal PEM using linear regression and to extend the *Refined Instrumental Variable* (RIV) approach of the LTI identification framework to provide an efficient way of identifying LPV-BJ models.

A. Optimal PEM for LPV-BJ models

Using (35), $y(t_k)$ can be written in the regression form:

$$y(t_k) = \varphi^{\top}(t_k)\rho + \tilde{v}(t_k) \tag{36}$$

where,

$$\begin{aligned} \varphi(t_k) = \big[&-y(t_{k-1}) \ \dots \ -y(t_{k-n_{\mathrm{a}}}) \ -\chi_{1,1}(t_k) \ \dots \\ &-\chi_{n_{\mathrm{a}},n_{\alpha}}(t_k) \ u_{0,0}(t_k) \ \dots \ u_{n_{\mathrm{b}},n_{\beta}}(t_k) \big]^{\top} \\ \rho = \big[& a_{1,0} \ \dots \ a_{n_{\mathrm{a}},0} \ a_{1,1} \ \dots \ a_{n_{\mathrm{a}},n_{\alpha}} \ b_{0,0} \ \dots \ b_{n_{\mathrm{b}},n_{\beta}} \big]^{\top} \end{aligned}$$

and

$$\tilde{v}(t_k) = F(q^{-1},\rho)v(t_k). \tag{37}$$

It is important to notice that (36) and (22) are not equivalent. The extended regressor in (36) contains the noise-free output terms $\{\chi_{i,k}\}$. By following the conventional PEM approach on (36), the prediction error $\varepsilon_{\theta}(t_k)$ is given as:

$$\begin{aligned} \varepsilon_{\theta}(t_k) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)F(q^{-1},\rho)} \Bigg(& F(q^{-1},\rho)y(t_k) - \\ & - \sum_{i=1}^{n_{\mathrm{a}}}\sum_{l=1}^{n_{\alpha}} a_{i,l}\chi_{i,l}(t_k) + \sum_{j=0}^{n_{\mathrm{b}}}\sum_{l=0}^{n_{\mathrm{b}}} b_{j,l}u_{k,j}(t_k) \Bigg) \end{aligned} \tag{38}$$

where $D(q^{-1},\eta)/C(q^{-1},\eta)$ can be recognized again as the inverse of the ARMA(n_{c},n_{d}) noise model in (12). However, since the system written as in (35) is equivalent to a LTI system, the polynomial operators commute and (38) can be considered in the alternative form

$$\begin{aligned} \varepsilon_{\theta}(t_k) = F(q^{-1},\rho)y_{\mathrm{f}}(t_k) &- \sum_{i=1}^{n_{\mathrm{a}}}\sum_{l=1}^{n_{\alpha}} a_{i,l}\chi_{i,l}^{\mathrm{f}}(t_k) \\ &+ \sum_{j=0}^{n_{\mathrm{b}}}\sum_{l=0}^{n_{\beta}} b_{j,l}u_{k,j}^{\mathrm{f}}(t_k) \end{aligned} \tag{39}$$

where $y_{\mathrm{f}}(t_k)$, $u_{k,j}^{\mathrm{f}}(t_k)$ and $\chi_{i,l}^{\mathrm{f}}(t_k)$ represent the outputs of the prefiltering operation, using the filter (see [24]):

$$Q(q^{-1},\theta) = \frac{D(q^{-1},\eta)}{C(q^{-1},\eta)F(q^{-1},\rho)}. \tag{40}$$

Based on (39), the associated linear-in-the-parameters model takes the form [24]:

$$y_{\mathrm{f}}(t_k) = \varphi_{\mathrm{f}}^{\top}(t_k)\rho + \tilde{v}_{\mathrm{f}}(t_k), \tag{41}$$

where

$$\begin{aligned} \varphi_{\mathrm{f}}(t_k) = \big[&-y_{\mathrm{f}}(t_{k-1}) \ \dots \ -y_{\mathrm{f}}(t_{k-n_{\mathrm{a}}}) \ -\chi_{1,1}^{\mathrm{f}}(t_k) \ \dots \\ &-\chi_{n_{\mathrm{a}},n_{\alpha}}^{\mathrm{f}}(t_k) \ u_{0,0}^{\mathrm{f}}(t_k) \ \dots \ u_{n_{\mathrm{b}},n_{\beta}}^{\mathrm{f}}(t_k) \big]^{\top} \end{aligned}$$

and

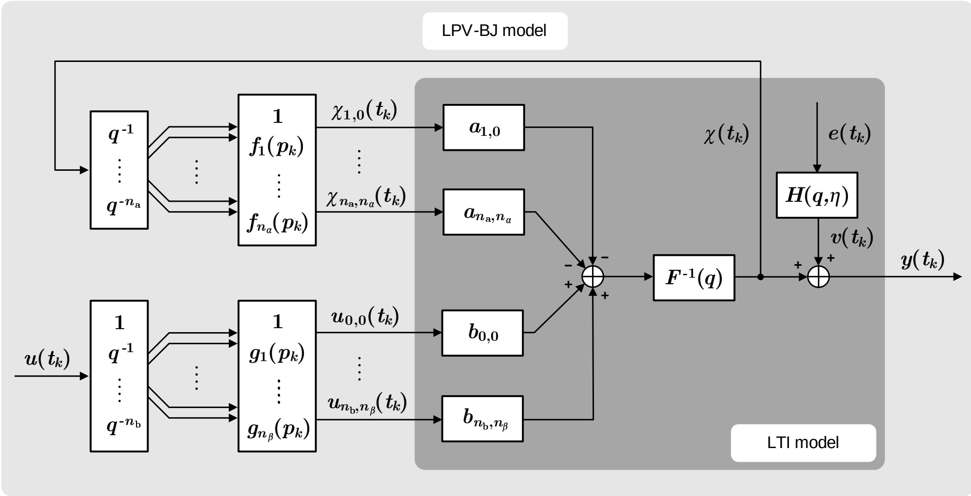

Fig. 1. MISO LTI interpretation of the LPV-BJ model

$$\tilde{v}_{\mathrm{f}}(t_k) = F(q^{-1},\rho)v_{\mathrm{f}}(t_k) = F(q^{-1},\rho)\frac{D(q^{-1},\eta)}{C(q^{-1},\eta)F(q^{-1},\rho)}v(t_k) = e(t_k). \quad (42)$$

B. The refined instrumental variable estimate

Many methods of the LTI identification framework can be used to provide an efficient estimate of ρ given (41) where $\tilde{v}_{\mathrm{f}}(t_k)$ is a white noise. Here, the refined instrumental variable method is chosen for the following reasons:

- RIV methods lead to optimal estimates if $\mathcal{S}_{\mathrm{o}} \in \mathcal{M}$, see [16]. This statement is true as well for usual prediction error methods such as the extended LS approach.
- In practical situation of identification, $\mathcal{G}_{\mathrm{o}} \in \mathcal{G}$ might be fulfilled due to first principle or expert's knowledge. However, it is commonly fair to assume that $H_{\mathrm{o}} \notin \mathcal{H}$. In such case, RIV methods has the advantage that they still provide consistent estimates whereas methods such as extended LS are biased and more advanced PEM methods needs robust initialization [13].

Aiming at the application of the RIV approach for the estimation of LPV-BJ models, consider the relationship between the process input and output signals as in (36). Based on this form, the extended-IV estimate can be given as [16]:

$$\hat{\rho}_{\mathtt{XIV}}(N) = \arg\min_{\rho\in\mathbb{R}^{n_\rho}} \left\| \left[\frac{1}{N}\sum_{k=1}^{N} L(q)\zeta(t_k)L(q)\varphi^{\top}(t_k)\right] - \left[\frac{1}{N}\sum_{t=1}^{N} L(q)\zeta(t_k)L(q)y(t_k)\right]\right\|_W^2, \quad (43)$$

where $\zeta(t_k)$ is the instrumental vector, $\|x\|_W^2 = x^T W x$, with W a positive definite weighting matrix and $L(q)$ is a stable prefilter. if $G_{\mathrm{o}} \in \mathcal{G}$, the extended-IV estimate is consistent under the following conditions[1]:

C1 $\bar{\mathbb{E}}\{L(q)\zeta(t_k)L(q)\varphi^{\top}(t_k)\}$ is full column rank.

C2 $\bar{\mathbb{E}}\{L(q)\zeta(t_k)L(q)\tilde{v}(t_k)\} = 0.$

Moreover it has been shown in [16] and [22] that the minimum variance estimator can be achieved if:

C3 $W = I$.

C4 ζ is chosen as the noise-free version of the extended regressor in (36) and is therefore defined in the present LPV case as:

$$\zeta(t_k) = \big[-\chi(t_{k-1}) \ \ldots \ -\chi(t_{k-n_a}) \ -\chi_{1,1}(t_k) \ \ldots \ -\chi_{n_a,n_\alpha}(t_k) \ u_{0,0}(t_k) \ \ldots \ u_{n_b,n_\beta}(t_k)\big]^{\top}$$

C5 $\mathcal{G}_{\mathrm{o}} \in \mathcal{G}$ and n_ρ is equal to the minimal number of parameters required to represent $\mathcal{G}_{\mathrm{o}}$ with the considered model structure.

C6 $L(q)$ is chosen as in (40).

C. Remarks on the use of the RIV approach

- Full column rank of $\bar{\mathbb{E}}\{L(q)\varphi(t_k)L(q)\varphi^{\top}(t_k)\}$ follows under Assumption A4 [3]. To fulfill C1 under A4, the discussion can be found in [16].
- C5 cannot be satisfied if the data generating LPV system $\mathcal{S}_{\mathrm{o}}$ is quasi, i.e. p_k is correlated to $u(t_k)$ or $y(t_k)$, and $\mathcal{S}_{\mathrm{o}}$ can be described as a Hammerstein or Wiener model [11].
- In a practical situation none of $F(q^{-1},\rho)$, $\{a_{i,l}(\rho)\}_{i=1,l=0}^{n_a,n_\alpha}$, $\{b_{j,l}(\rho)\}_{j=0,l=0}^{n_b,n_\beta}$, $C(q^{-1},\eta)$, $D(q^{-1},\eta)$ is known *a priori*. Therefore, the RIV estimation normally involves an iterative (or relaxation) algorithm in which, at each iteration, an 'auxiliary model' is used to generate the instrumental variables (which guarantees C2), as well as the associated prefilters. This auxiliary model is based on the parameter estimates obtained at the previous iteration. Consequently, if convergence occurs, C4 and C6 are fulfilled. It is important to note that convergence of the iterative RIV has not been proved so far and is only empirically assumed [23].

[1]The notation $\bar{\mathbb{E}}\{.\} = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\mathbb{E}\{.\}$ is adopted from the prediction error framework of [12].

D. Iterative LPV-RIV Algorithm

Based on the previous considerations, the iterative scheme of the RIV algorithm can be extended to the LPV case as follows.

Algorithm 2 (Optimal LPV-RIV):

Step 1 Assume that as an initialization, an ARX estimate of $\mathcal{M}_\theta$ is available by the LS approach, i.e. $\hat{\theta}^{(0)} = [\ (\hat{\rho}^{(0)})^\top \quad (\hat{\eta}^{(0)})^\top\ ]^\top$ is given. Set $\tau = 0$.

Step 2 Compute an estimate of $\chi(t_k)$ by simulating the auxiliary model:

$$A(p_k, q^{-1}, \hat{\rho}^{(\tau)})\hat{\chi}(t_k) = B(p_k, q^{-1}, \hat{\rho}^{(\tau)})u(t_{k-d})$$

based on the estimated parameters $\hat{\rho}^{(\tau)}$ of the previous iteration. Deduce the output terms $\{\hat{\chi}_{i,l}(t_k)\}_{i=1,l=0}^{n_\mathrm{a},n_\alpha}$ as given in (34) and using the model $\mathcal{M}_{\hat{\theta}^{(\tau)}}$.

Step 3 Compute the estimated filter:

$$\hat{Q}(q^{-1}, \hat{\theta}^{(\tau)}) = \frac{D(q^{-1}, \hat{\eta}^{(\tau)})}{C(q^{-1}, \hat{\eta}^{(\tau)})F(q^{-1}, \hat{\rho}^{(\tau)})}$$

and the associated filtered signals $\{u^\mathrm{f}_{j,l}(t_k)\}_{j=0,l=0}^{n_\mathrm{b},n_\beta}$, $y_\mathrm{f}(t_k)$ and $\{\chi^\mathrm{f}_{i,l}(t_k)\}_{i=1,l=0}^{n_\mathrm{a},n_\alpha}$.

Step 4 Build the filtered estimated regressor $\hat{\varphi}_\mathrm{f}(t_k)$ and in terms of C4 the filtered instrument $\hat{\zeta}_\mathrm{f}(t_k)$ as:

$$\begin{aligned}\hat{\varphi}_\mathrm{f}(t_k) = \big[&-y_\mathrm{f}(t_{k-1}) \ \ldots \ -y_\mathrm{f}(t_{k-n_\mathrm{a}}) \ -\hat{\chi}^\mathrm{f}_{1,1}(t_k) \\ &\ldots \ -\hat{\chi}^\mathrm{f}_{n_\mathrm{a},n_\alpha}(t_k) \ u^\mathrm{f}_{0,0}(t_k) \ \ldots \ u^\mathrm{f}_{n_\mathrm{b},n_\beta}(t_k)\big]^\top \\ \hat{\zeta}_\mathrm{f}(t_k) = \big[&-\hat{\chi}_\mathrm{f}(t_{k-1}) \ \ldots \ -\hat{\chi}_\mathrm{f}(t_{k-n_\mathrm{a}}) \ -\hat{\chi}^\mathrm{f}_{1,1}(t_k) \\ &\ldots \ -\hat{\chi}^\mathrm{f}_{n_\mathrm{a},n_\alpha}(t_k) \ u^\mathrm{f}_{0,0}(t_k) \ \ldots \ u^\mathrm{f}_{n_\mathrm{b},n_\beta}(t_k)\big]^\top\end{aligned}$$

Step 5 The IV optimization problem can now be stated in the form

$$\hat{\rho}^{(\tau+1)}(N) = \arg\min_{\rho\in\mathbb{R}^{n_\rho}} \left\| \left[\frac{1}{N}\sum_{k=1}^{N}\hat{\zeta}_\mathrm{f}(t_k)\hat{\varphi}_\mathrm{f}^\top(t_k)\right]\rho - \left[\frac{1}{N}\sum_{k=1}^{N}\hat{\zeta}_\mathrm{f}(t_k)y_\mathrm{f}(t_k)\right]\right\|^2 \quad (44)$$

This results in the solution of the IV estimation equations:

$$\hat{\rho}^{(\tau+1)}(N) = \left[\sum_{k=1}^{N}\hat{\zeta}_\mathrm{f}(t_k)\hat{\varphi}_\mathrm{f}^\top(t_k)\right]^{-1}\sum_{k=1}^{N}\hat{\zeta}_\mathrm{f}(t_k)y_\mathrm{f}(t_k)$$

where $\hat{\rho}^{(\tau+1)}(N)$ is the IV estimate of the process model associated parameter vector at iteration $\tau+1$ based on the prefiltered input/output data.

Step 6 An estimate of the noise signal v is obtained as

$$\hat{v}(t_k) = y(t_k) - \hat{\chi}(t_k, \hat{\rho}^{(\tau)}). \quad (45)$$

Based on $\hat{v}$, the estimation of the noise model parameter vector $\hat{\eta}^{(\tau+1)}$ follows, using in this case the `ARMA` estimation algorithm of the `MATLAB` identification toolbox (an IV approach can also be used for this purpose, see [23]).

Step 7 If $\theta^{(\tau+1)}$ has converged or the maximum number of iterations is reached, then stop, else increase τ by 1 and go to Step 2.

Based on a similar concept, a statistically optimal method, the so-called *simplified* LPV-RIV (LPV-SRIV), can also be developed for the estimation of LPV-OE models. This method is based on a model structure (12) with $C(q^{-1}, \eta) = D(q^{-1}, \eta) = 1$ and consequently, Step 6 of Algorithm 2 can be skipped. Naturally, the LPV-SRIV is not statistically optimal for LPV-BJ models, however it still has a certain degree of robustness as it is shown in Section V.

V. Simulation Example

As a next step, the performance of the proposed and of the existing methods in the literature are compared based on a representative simulation example. The robustness with respect to noise and modeling error is investigated as well.

A. Data generating system

The system taken into consideration is inspired by the example in [4] and is mathematically described as follows:

$$\mathcal{S}_\mathrm{o}\begin{cases} A_\mathrm{o}(q, p_k) = 1 + a_1^\mathrm{o}(p_k)q^{-1} + a_2^\mathrm{o}(p_k)q^{-2} \\ B_\mathrm{o}(q, p_k) = b_0^\mathrm{o}(p_k)q^{-1} + b_1^\mathrm{o}(p_k)q^{-2} \\ H_\mathrm{o}(q) = \dfrac{1}{1 - q^{-1} + 0.2q^{-2}} \end{cases} \quad (46)$$

where $v(t_k) = H_\mathrm{o}(q)e(t_k)$ and

$$a_1^\mathrm{o}(p_k) = 1 - 0.5p_k - 0.1p_k^2 \quad (47a)$$
$$a_2^\mathrm{o}(p_k) = 0.5 - 0.7p_k - 0.1p_k^2 \quad (47b)$$
$$b_0^\mathrm{o}(p_k) = 0.5 - 0.4p_k + 0.01p_k^2 \quad (47c)$$
$$b_1^\mathrm{o}(p_k) = 0.2 - 0.3p_k - 0.02p_k^2 \quad (47d)$$

In the upcoming examples, the scheduling signal p is considered as a periodic function of time:

$$p_k = 0.5\sin(0.35\pi k) + 0.5, \quad (48)$$

and $u(t_k)$ is taken as a white noise with a uniform distribution $\mathcal{U}(-1, 1)$ and with length $N = 4000$ to generate data sets $\mathcal{D}_N$ of $\mathcal{S}_\mathrm{o}$.

B. Model structures

In the sequel, the instrumental variable method presented in [4] named here *One Step Instrumental Variable* (`OSIV`) and the conventional *Least Square* (`LS`) method such as the one used in [3] are compared to the proposed methods. Both methods assume the following model structure:

$$\mathcal{M}_\theta^{\mathtt{LS,OSIV}}\begin{cases} A(p_k, q^{-1}, \rho) = 1 + a_1(p_k)q^{-1} + a_2(p_k)q^{-2} \\ B(p_k, q^{-1}, \rho) = b_0(p_k)q^{-1} + b_1(p_k)q^{-2} \\ H(p_k, q, \rho) = A^\dagger(p_k, q^{-1}, \rho) \end{cases}$$

where

$$a_1(p_k) = a_{1,0} + a_{1,1}p_k + a_{1,2}p_k^2 \quad (49a)$$
$$a_2(p_k) = a_{2,0} + a_{2,1}p_k + a_{2,2}p_k^2 \quad (49b)$$
$$b_0(p_k) = b_{0,0} + b_{0,1}p_k + b_{0,2}p_k^2 \quad (49c)$$
$$b_1(p_k) = b_{1,0} + b_{1,1}p_k + b_{1,2}p_k^2 \quad (49d)$$

In contrast with these model structures, the proposed LPV *Refined Instrumental Variable* method (LPV-RIV) represents the situation $\mathcal{S}_{\rm o} \in \mathcal{M}$ and assumes the following LPV-BJ model:

$$\mathcal{M}_\theta^{\texttt{LPV-RIV}} \begin{cases} A(p_k, q^{-1}, \rho) = 1 + a_1(p_k)q^{-1} + a_2(p_k)q^{-2} \\ B(p_k, q^{-1}, \rho) = b_0(p_k)q^{-1} + b_1(p_k)q^{-2} \\ H(p_k, q, \eta) = \dfrac{1}{1 + d_1 q^{-1} + d_2 q^{-2}} \end{cases}$$

with $a_1(p_k)$, $a_2(p_k)$, $b_0(p_k)$, $b_1(p_k)$ as given in (49a-d), while the LPV *Simplified Refined Instrumental Variable* method (LPV-SRIV) represents the case when $\mathcal{G}_{\rm o} \in \mathcal{G}$, $\mathcal{H}_{\rm o} \notin \mathcal{H}$ and assumes the following LPV-OE model:

$$\mathcal{M}_\theta^{\texttt{LPV-SRIV}} \begin{cases} A(p_k, q^{-1}, \rho) = 1 + a_1(p_k)q^{-1} + a_2(p_k)q^{-2} \\ B(p_k, q^{-1}, \rho) = b_0(p_k)q^{-1} + b_1(p_k)q^{-2} \\ H(p_k, q, \eta) = 1 \end{cases}$$

C. Example 1: Robustness to noise

In this part, the robustness of the proposed and existing algorithms are investigated with respect to different *signal-to-noise ratios* (SNR)

$$\mathrm{SNR} = 10 \log \frac{P_{\chi_{\rm o}}}{P_{e_{\rm o}}}, \tag{50}$$

where $P_{\chi_{\rm o}}$ and $P_{e_{\rm o}}$ are the average power of signals $\chi_{\rm o}$ and $e_{\rm o}$ respectively. To provide representative results, a Monte-Carlo simulation of $N_{\rm MC} = 100$ runs with new noise realization is accomplished at different noise levels: 15dB, 10dB, 5dB and 0dB. An example of input/output signals at $\mathrm{SNR} = 0$dB is shown in Figure 2.

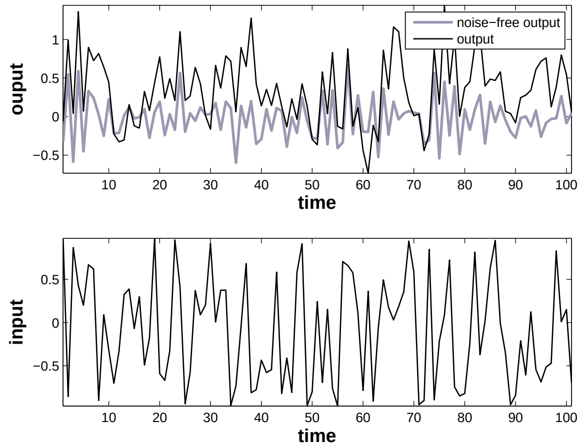

Fig. 2. I/O signals at $\mathrm{SNR} = 0$dB

With respect to the considered methods, Table I shows the norm of the bias and variance of the estimated parameter vector:

$$\text{Bias norm} = ||\rho_o - \bar{\mathbb{E}}(\hat{\rho})||_2 \tag{51a}$$

$$\text{Variance norm} = ||\bar{\mathbb{E}}(\hat{\rho} - \bar{\mathbb{E}}(\hat{\rho}))||_2 \tag{51b}$$

where $\bar{\mathbb{E}}$ is the mean operator over the Monte-Carlo simulation and $||.||_2$ is the $\mathcal{L}_2$ norm. The table also displays the mean number of iterations the algorithms needed to converge to the estimated parameter vector.

For the Monte-Carlo simulation at $\mathrm{SNR} = 15$dB, Table II and III show the detailed results about mean and standard deviation of the estimated parameters. In some practical application, only one realization is accessible and therefore it is not possible to compute the uncertainty through *Monte-Carlo simulation* (MCS). In this latter case it is important to be able to determine the *standard error* (SE) on the estimated parameters with a *single realization* (SR). Therefore the results of SR are also given in these tables. Note that it is possible to compute the SR standard error from the covariance matrix defined as [23]

$$\hat{P}_\rho = \hat{\sigma}_e^2 \left[\sum_{k=1}^{N} \hat{\zeta}_{\rm f}(t_k) \hat{\zeta}_{\rm f}^\top(t_k) \right]^{-1} \tag{52}$$

by using the relation

$$\mathrm{SE} = \sqrt{\mathrm{diag}(\hat{P}_\rho)}. \tag{53}$$

It can be seen from Table I that the IV methods are unbiased according to the theoretical results. It might not appear clearly for the OSIV method when using SNR under 10dB but considering the variances induced, the bias is only due to the relatively low number of simulation runs. Under 10dB, the results of the OSIV cannot be considered as relevant as they induce such large values. In the present BJ system, the OSIV method does not lead to satisfying results and cannot be used in practical applications. It can be seen that for SNR down to 5dB, the optimal LPV-RIV produces variance in the estimated parameters which are very close to the one obtained with the LS method, not mentioning that the bias has been completely suppressed. The suboptimal LPV-SRIV methods offers satisfying results, considering that the noise model is not correctly assumed. The variance in the estimated parameters is twice as much as in the optimal LPV-RIV case and it is in close range to the variance of the LS method. Finally, it can be pointed out that the number of iterations is high in comparison to the linear case for RIV methods (typically, 4 iterations are needed in a second order linear case). Table II and III shows that detailed results lead to the same conclusion as when looking at Table I. It can be finally seen from III that the optimal LPV-RIV method estimates accurately the noise model and that the standard error obtained from a single realization is well correlated to the standard deviation obtained through Monte-Carlo simulation.

D. Example 2: Robustness to modeling error

All the tests realized in the previous section has been achieved in the case when $\mathcal{G}_{\rm o} \in \mathcal{G}$. This assumption is fair for linear systems as most process can be derived from first principle laws. In nonlinear systems however, the non-linearity and therefore the scheduling parameter dependence is often an approximation of the real function. Therefore it is relevant to investigate the case when $\mathcal{G}_{\rm o} \notin \mathcal{G}$. However, it is not possible in this case to compare the estimated parameters to the true parameters. Consequently, the authors choose to

TABLE I

ESTIMATOR BIAS AND VARIANCE NORM AT DIFFERENT SNR

Method		15dB	10dB	5dB	0dB
`LS`	Bias norm	2.9107	3.2897	3.0007	2.8050
	Variance norm	0.0074	0.0151	0.0215	0.0326
`OSIV`	Bias norm	0.1961	1.8265	6.9337	10.8586
	Variance norm	1.3353	179.4287	590.7869	11782
`LPV-SRIV`	Bias norm	0.0072	0.0426	0.1775	0.2988
	Variance norm	0.0149	0.0537	0.4425	0.4781
	mean iteration number	22	22	25	30
`LPV-RIV`	Bias norm	0.0068	0.0184	0.0408	0.1649
	Variance norm	0.0063	0.0219	0.0696	0.2214
	mean iteration number	31	30	30	32

TABLE II

MEAN AND STANDARD DEVIATION OF THE ESTIMATED A POLYNOMIAL PARAMETERS AT SNR $= 15$DB

		$a_{1,0}$	$a_{1,1}$	$a_{1,2}$	$a_{2,0}$	$a_{2,1}$	$a_{2,2}$
method	true value	1	-0.5	-0.1	0.5	-0.7	-0.1
`LS`	mean	-0.3794	2.2373	-2.0584	-0.1085	-0.0755	-0.4786
	std	0.0219	0.0663	0.0591	0.0125	0.0600	0.0558
`OSIV`	mean	1.0259	-0.6161	0.0205	0.5092	-0.7510	-0.0377
	std	0.3023	1.0330	0.8605	0.1227	0.4348	0.3986
`LPV-SRIV` MCS	mean	1.0003	-0.5013	-0.0971	0.5007	-0.7047	-0.0943
	std	0.0313	0.1022	0.0893	0.0106	0.0650	0.074
`LPV-SRIV` SR	$\hat{\rho}$	0.9801	-0.3743	-0.2120	0.4978	-0.7154	-0.0736
	SE	0.0377	0.1567	0.1486	0.0171	0.1010	0.1099
`LPV-RIV` MCS	mean	0.9999	-0.5020	-0.0989	0.5005	-0.7050	-0.0962
	std	0.0170	0.0589	0.0610	0.0084	0.0488	0.0523
`LPV-RIV` SR	$\hat{\rho}$	0.9947	-0.5053	-0.0506	0.4981	-0.7303	-0.0350
	SE	0.0120	0.0479	0.0435	0.0050	0.0330	0.0368

TABLE III

MEAN AND STANDARD DEVIATION OF THE ESTIMATED B AND D POLYNOMIAL PARAMETERS AT SNR $= 15$DB

		$b_{0,0}$	$b_{0,1}$	$b_{0,2}$	$b_{1,0}$	$b_{1,2}$	$b_{2,2}$	d_1	d_2
method	true value	0.5	-0.4	0.01	0.2	-0.3	-0.02	-1	0.2
`LS`	mean	0.5043	-0.4045	0.0085	-0.3201	0.7890	-0.7335	X	X
	std	0.0039	0.0233	0.0219	0.0097	0.0284	0.0238	X	X
`OSIV`	mean	0.4986	-0.3991	0.0110	0.2096	-0.3409	0.0181	X	X
	std	0.0115	0.0564	0.0503	0.1151	0.3731	0.2922	X	X
`LPV-SRIV` MCS	mean	0.4996	-0.3998	0.0101	0.1997	-0.3004	-0.0190	X	X
	std	0.0038	0.0183	0.0171	0.0104	0.0367	0.0300	X	X
`LPV-SRIV` SR	$\hat{\rho}$	0.4998	-0.3783	-0.0108	0.1996	-0.2885	-0.0273	X	X
	SE	0.0044	0.0221	0.0216	0.0138	0.0561	0.0493	X	X
`LPV-RIV` MCS	mean	0.4998	-0.3993	0.0092	0.1998	-0.3008	-0.0194	-1.003	0.2042
	std	0.0020	0.0106	0.0106	0.0055	0.0228	0.0214	0.0171	0.0172
`LPV-RIV` SR	$\hat{\rho}$	0.5020	-0.4142	0.0245	0.1971	-0.2889	-0.0219	X	X
	SE	0.0016	0.0075	0.0072	0.0042	0.0165	0.0141	X	X

expose the fitness score [13]

$$\text{fit} = 100\% \cdot \left(1 - \frac{||\hat{\chi}(t_k) - \chi_\text{o}(t_k)||_2}{||\chi_\text{o}(t_k) - \bar{\mathbb{E}}(\chi_\text{o}(t_k))||_2}\right) \tag{54}$$

between the noise-free output $\chi_\text{o}(t)$ and the output $\hat{\chi}(t)$ simulated using the identified model.

The fitness score is 100% if the simulated output coincides exactly to the noise-free output. If the score equals 0% the estimated output fits the noise-free output as good as its mean value. A score under -100% is set to -100. The minimum, maximum and mean value of the fitness score for different approaches are displayed in Table IV ($\text{SNR} = 10\text{dB}$) and Table V ($\text{SNR} = 0\text{dB}$) for a Monte-Carlo simulation of 20 runs. The right model for the presented system is $n_\text{a} = 2$, $n_\text{b} = 2$, $n_\text{c} = 0$, $n_\text{d} = 2$, $n_\alpha = 2$ and $n_\beta = 2$ denoted as $[n_\text{a}\, n_\text{b}\, n_\text{c}\, n_\text{d}\, n_\alpha\, n_\beta] = [220222]$. For example the model $[210101]$ corresponds to

$$\mathcal{M}_{210101} \begin{cases} A(p_k, q^{-1}, \rho) = 1 + a_1(p_k)q^{-1} + a_2(p_k)q^{-2}, \\ B(p_k, q^{-1}, \rho) = b_0(p_k)q^{-1}, \\ H(p_k, q, \eta) = \dfrac{1}{1 + d_1 q^{-1}}, \end{cases}$$

with

$$a_1(p_k) = a_{1,0}, \quad a_2(p_k) = a_{2,0}, \quad b_0(p_k) = b_{0,0} + b_{0,1}p_k.$$

Table IV and Table V show that the `LS` method fails to correctly fit the output. At $\text{SNR} = 0\text{dB}$, the LS method gives very poor fitting score. Table IV shows that at 10dB, even if the `OSIV` estimator is not efficient, it can compete with the `LPV-RIV` estimator in some runs but the values of the fitting score are more sparse. Table V shows that the `OSIV` estimator cannot be used reliably. It can be noticed that in all simulations, the best fitting score of the `OSIV` estimated model is less than the average fitting score of the `LPV-RIV` estimated model. The `LPV-RIV` estimator seems to be the less affected by strong noise and is reliable even for strong modeling error (model $[110011]$) and at $\text{SNR} = 0\text{dB}$ (see Figure 3).

VI. Conclusion

This paper highlighted the lack of efficient methods in the literature to handle the estimation of LPV Box-Jenkins models. It has been shown that the conventional formulation of least squares estimation cannot lead to statistically optimal parameter estimates. As a solution, the LPV identification problem is reformulated and a method to estimate efficiently LPV-BJ models was proposed. The introduced method has been compared to the existing methods of the literature both in terms of theoretical analysis and in terms of a representative numerical example. The presented example has shown that the proposed procedure is robust to noise and modeling error and outperforms the existing methods. As continuation of the presented work, extensions of the method to closed-loop and continuous-time LPV system identification are intended.

References

[1] B. Bamieh and L. Giarré. Identification of linear parameter varying models. In *Proceedings of the 38th IEEE Conference on Decision and Control*, pages 1505–1510, Phoenix, USA, Dec 1999.

[2] B. Bamieh and L. Giarré. Identification for a general class of LPV models. In *Proceedings of the 11th IFAC Symposium on System Identification*, Santa Barbara, California, USA, June 2000.

[3] B. Bamieh and L. Giarré. Identification of linear parameter-varying models. *International Journal of Robust and Nonlinear Control*, 12:841–853, 2002.

[4] M. Butcher, A. Karimi, and R. Longchamp. On the consistency of certain identification methods for linear parameter varying systems. In *Proceedings of the 17th IFAC World Congress*, pages 4018–4023, Seoul, Korea, July 2008.

[5] P. L. dos Santos, J. A. Ramos, and J. L. M. de Carvalho. Identification of linear parameter varying systems using an iterative deterministic-stochastic subspace approach. In *Proc. of the European Control Conference*, pages 4867–4873, Kos, Greece, July 2007.

[6] M. Gevers, A. S. Bazanella, and L. Miskovic. Informative data: how to get just sufficiently rich? In *proceedings of the 47th IEEE Conference on Decision and Control*, pages 1962–1967, Cancun, Mexico, December 2008.

[7] L. Giarré, D. Bauso, P. Falugi, and B. Bamieh. LPV model identification for gain scheduling control: An application to rotating stall and surge control problem. *Control Engineering Practice*, 14:351–361, 2006.

[8] M. Gilson and P. Van den Hof. Instrumental variable methods for closed-loop system identification. *Automatica*, 41, Issue 2:241–249, 2005.

[9] M. Gilson, H. Garnier, P. C. Young, and P. Van den Hof. Refined instrumental variable methods for closed-loop system identification methos. In *proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, June 2009.

[10] H. Garnier and L. Wang (Editors). *Identification of Continuous-time Models from Sampled Data*. Springer-Verlag, London, March 2008.

[11] V. Laurain, M. Gilson, H. Garnier, and P. C. Young. Refined instrumental variable methods for identification of Hammerstein continuous time Box-Jenkins models. In *proceedings of the 47th IEEE Conference on Decision and Control*, Cancun, Mexico, Dec 2008.

[12] L. Ljung. *System identification : theory for the user - Second Edition*. Prentice-Hall, 1999.

[13] L. Ljung. Experiments with identification of continuous time models. In *proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, July 2009.

[14] M. Lovera and G. Mercère. Identification for gain-scheduling: a balanced subspace approach. In *Proc. of the American Control Conference*, pages 858–863, New York City, USA, July 2007.

[15] G. P. Rao and H. Garnier. Identification of continuous-time systems: direct or indirect ? *Systems Science*, 30, Issue 3:25–50, 2004.

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

[17] P. Stoica and M. Jansson. MIMO system identification: state-space and subspace approximations versus transfer function and instrumental variables. *IEEE Transaction on Signal Processing*, 48(11):3087 – 3099, 2000.

[18] R. Tóth. *Modeling and Identification of Linear Parameter-Varying Systems: An Orthonormal Basis Function Approach*. PhD thesis, Delft University of Technology, 2008.

[19] R. Tóth, P. S. C. Heuberger, and P. M. J. Van den Hof. An LPV identification framework based on orthonormal basis functions. In *proceedings of the 15th IFAC Symposium on System Identification*, Saint-Malo, France, July 2009.

[20] J. W. van Wingerden and M. Verhaegen. Subspace identification of bilinear and LPV systems for open- and closed-loop data. *Automatica*, 45:372–381, 2009.

[21] X. Wei and L. Del Re. On persistent excitation for parameter estimation of quasi-LPV systems and its application in modeling of diesel engine torque. In *Proceedings of the 14th IFAC Symposium on System Identification*, pages 517–522, Newcastle, Australia, March 2006.

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

[23] P. C. Young. The refined instrumental variable method: Unified estimation of discrete and continuous-time transfer function models. *Journal Européen des Systèmes Automatisés*, 42:149–179, 2008.

TABLE IV

MODELING ERROR MONTE-CARLO SIMULATION WITH $\mathrm{SNR} = 10$DB

Model parameters						LS			OSIV			LPV-RIV		
n_a	n_b	n_c	n_d	n_α	n_β	min(fit)	max(fit)	mean(fit)	min(fit)	max(fit)	mean(fit)	min(fit)	max(fit)	mean(fit)
2	2	0	2	2	2	60.2268	63.631	61.8351	19.0787	95.5119	80.9420	98.2614	99.2459	98.7147
2	2	0	0	1	1	61.2401	63.8981	62.7319	-1.8039	97.1807	86.3854	72.4847	99.1952	97.2106
1	1	0	2	2	2	47.7804	52.058	50.4318	74.9416	78.1466	76.4985	79.2946	79.6643	79.549
1	1	0	0	1	1	47.5383	52.6303	50.0925	71.5279	76.8744	74.8501	78.3167	78.6013	78.5231

TABLE V

MODELING ERROR MONTE-CARLO SIMULATION WITH $\mathrm{SNR} = 0$DB

Model parameters						LS			OSIV			LPV-RIV		
n_a	n_b	n_c	n_d	n_α	n_β	min(fit)	max(fit)	mean(fit)	min(fit)	max(fit)	mean(fit)	min(fit)	max(fit)	mean(fit)
2	2	0	2	2	2	28.2972	35.866	32.9784	-28.3212	84.5957	62.4347	93.4777	97.5705	95.9271
2	2	0	0	1	1	33.0264	41.1713	36.1755	-100	83.4528	34.4322	48.6399	97.7966	87.4178
1	1	0	2	2	2	1.6346	6.818	4.5074	-100	73.4932	14.8833	22.5242	79.4046	75.8615
1	1	0	0	1	1	1.1722	6.0193	4.2104	-100	75.2366	28.2461	42.8853	79.244	71.9069

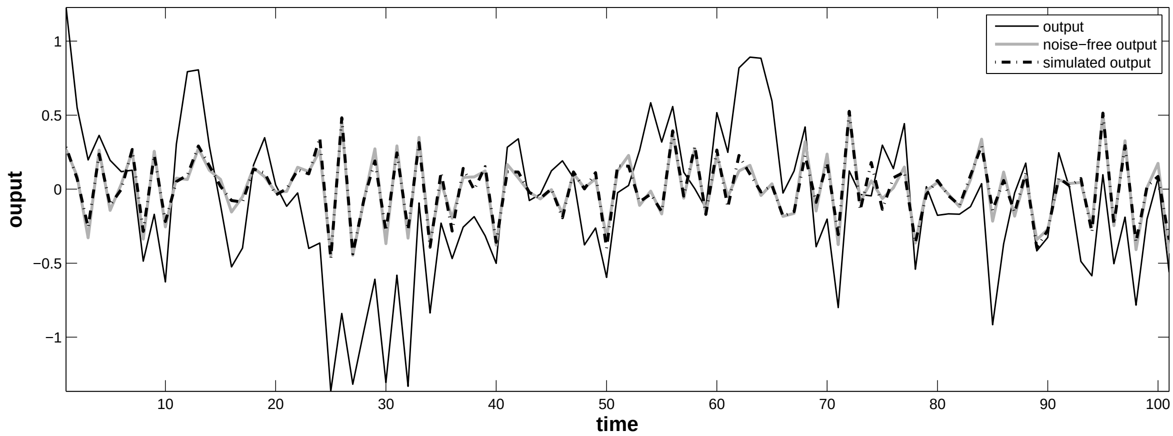

Fig. 3. Noise-free output, noisy output and simulated output for model $[110011]$ identified with LPV-RIV ($\mathrm{SNR} = 0$dB and fit$= 79\%$)

[24] P. C. Young, H. Garnier, and M. Gilson. *Identification of continuous-time models from sampled data*, chapter Refined instrumental variable identification of continous-time hybrid Box-Jenkins models. Springer-Verlag, London, 2008.