Preprints of the 20th World Congress
The International Federation of Automatic Control
Toulouse, France, July 9-14, 2017

LPV Model Order Selection from Noise-corrupted Output and Scheduling Signal Measurements ⋆

Manas Mejari * **Dario Piga** ** **Alberto Bemporad** *

* *IMT School for Advanced Studies Lucca, 55100 Lucca, Italy (e-mail: {manas.mejari, alberto.bemporad} @imtlucca.it)*
** *IDSIA Dalle Molle Institute for Artificial Intelligence SUPSI-USI, 6928 Manno, Switzerland (e-mail: dario@idsia.ch)*

Abstract: In parametric identification of *Linear Parameter-Varying* (LPV) systems, it is important to achieve a low variance of the model estimate by limiting the number of parameters to be identified. This is the well known "model order selection" problem, which consists of selecting the number of input and output delays and the basis functions characterizing the dependence of the LPV model parameters on the scheduling signal. Ignoring the effect of noise on the observations of the scheduling signals may lead to a bias in the final estimate and, as a consequence, also to an incorrect selection of the model order. In this paper, we introduce a "bias-corrected cost function" for the identification of LPV systems from noise-corrupted observations of the output and scheduling variable. The introduced cost function provides a bias-free parameter estimation along with model order selection. The proposed identification approach has two main advantages: (i) the problem of model order selection can be handled by adding a LASSO-like penalty term to the bias-corrected cost function; (ii) it provides a bias-free cost as a criterion to tune some hyper-parameters influencing the final parameter estimate.

Keywords: Bias-correction methods, Linear parameter-varying systems, Model order selection.

1. INTRODUCTION

Identification of *Linear Parameter-Varying* (LPV) systems has gained significant attention in the past years thanks to the capabilities offered by LPV models to describe the behaviour of many nonlinear and time-varying systems. Motivated by the need of accurate and low-complexity LPV models, significant efforts have been devoted for developing efficient methods for the identification of LPV models. One of the main challenge in parametric LPV identification is the choice of the model structure, which requires to specify the model order (in terms of number of input and output delays) as well as the type of the so-called *basis functions*. The basis functions characterize the dependence of the model coefficients on the scheduling signal and thus, in order to avoid the use of under-parametrized models, a large set of basis functions is typically chosen to adequately describe the underlying system. However, there is a risk of over-parametrization of the model which may cause a large variance in the estimate of the model parameters. This is the well known "bias-variance trade-off" problem and it can be partially overcome by using sparse parametric estimation methods based on *LASSO* and on the *Non-Negative Garrote* (NNG). Application of the NNG and a sparse LASSO-type estimator for sparse identification of LPV-ARX models is presented in (Tóth et al., 2009) and (Tóth et al., 2012), respectively. These methods can be used to select a subset of p-dependent nonlinear basis functions by penalizing, together with the fitting error, the ℓ_1-norm of the model parameters, thus enforcing sparsity in the final estimate of the parameter vector. Extensions of NNG and LASSO for order selection in a nonparametric LPV identification are discussed in (Mejari et al., 2016) and (Piga and Tóth, 2013), respectively.

Most of the LPV identification and sparse estimation methods available in the literature assume that only the output measurements are corrupted by noise, while the observations of the scheduling signal are noise free. However, in practice, this is an unrealistic assumption in most cases, as the scheduling variable is often measured by a sensor, and thus inherently affected by measurement noise. To the best of the authors' knowledge, the only contributions on LPV identification based on noisy measurements of the scheduling signals are given in (Cerone et al., 2013; Butcher et al., 2008; Piga et al., 2015). The contribution in Cerone et al. (2013) addresses the identification of LPV systems in the set-membership framework, where an outer bounding box on the feasible parameter set is computed under the assumption that the perturbing noise is bounded. The work in Butcher et al. (2008) proposes an *instrumental variable* (IV) approach but it is limited to the case in which the dependence on the scheduling signal is linear. In Piga et al. (2015), these limitations are overcome by using a bias-corrected IV approach. The method provides consistent estimates of LPV models with

⋆ This work was partially supported by the European Commission under project H2020-SPIRE-636834 "DISIRE - Distributed In-Situ Sensors Integrated into Raw Material and Energy Feedstock" (`http://spire2030.eu/disire/`).

polynomial dependence on the scheduling variable and the instruments are only required to be uncorrelated with the noise corrupting the output observations. However, no penalty function is introduced in Piga et al. (2015), and thus ℓ_1-regularization methods (like the LASSO) cannot be used to select the model structure.

In this paper, we extend the approach presented in Piga et al. (2015), to address the issue of model-order selection by introducing "a bias-corrected cost" function. An ℓ_1-regularization term is then added to this cost to achieve accurate model-order selection. Furthermore, by using the introduced bias-corrected cost as a criterion for cross-validation, an unbiased tuning of the hyper-parameters influencing the parameter estimates is achieved.

The paper is organized as follows. In Section 2, the notation used throughtout the paper is introduced. The considered identification problem is formulated in Section 3. In Section 4, the instrumental-variable identification method is reviewed, and the asymptotic properties of the estimated parameters are discussed. The bias-corrected version of the instrumental-variable method is presented in Section 5. A simulation example illustrating the capabilities of the proposed method is presented in Section 6.

2. NOTATIONS

Let $\mathbb{R}^n$ be the set of real vectors of dimension n. The 2−norm of the vector $x \in \mathbb{R}^n$ is denoted by $\|x\|_2$. For matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$, the Kronecker product between A and B is denoted by $A\otimes B \in \mathbb{R}^{mp\times nq}$. Let $\mathbb{I}_a^b$ be the sequence of successive integers $\{a, a+1, \cdots, b\}$, with $b > a$. The floor function is denoted by $\lfloor\cdot\rfloor$, with $\lfloor m\rfloor$ being the largest integer less than or equal to m. The expected value of a random vector $\mathbf{x}$ is denoted by $\mathbb{E}[\mathbf{x}]$.

3. PROBLEM FORMULATION

3.1 Data-generating system

As a data-generating system, let us consider a discrete-time *single-input single-output* (SISO) LPV system described by the *Output Error* (OE) structure:

$$y_{\text{o}}(k) = -\sum_{i=1}^{n_a^{\text{o}}} a_i^{\text{o}}(p_{\text{o}}(k))y_{\text{o}}(k-i) + \sum_{j=0}^{n_b^{\text{o}}} b_j^{\text{o}}(p_{\text{o}}(k))u(k-j) \quad \text{(1a)}$$

$$y(k) = y_{\text{o}}(k) + \nu_{\text{o}}(k), \quad \text{(1b)}$$

where $u(k) \in \mathbb{R}$ and $y(k) \in \mathbb{R}$ are the measured input and output signals of the system at time k, respectively, $\nu_{\text{o}}(k)$ is an additive zero-mean white noise corrupting the output, $p_{\text{o}}(k)$ is the noise-free scheduling signal, which is assumed to take value in the compact set $\mathbb{P}$. The functions $a_i^{\text{o}}(\cdot)$ and $b_j^{\text{o}}(\cdot)$ are assumed to be polynomials of maximum degree n_g^{o} in the scheduling variable $p_{\text{o}}(k)$, and they have to be estimated along with the parameters $n_a^{\text{o}}, n_b^{\text{o}}, n_g^{\text{o}}$ defining the structure of the system (1). For the clarity of exposition, in the rest of the paper we consider the case of scalar scheduling signal $p_{\text{o}}(k)$. Extension to the multidimensional case is straightforward.

In order to describe the data-generating system (1) in a compact form, let us introduce the matrix notation:

$$\begin{aligned}
\mathbf{p}_{\text{o}}(k) &= \left[1 \;\; p_{\text{o}}(k) \;\; p_{\text{o}}^2(k) \cdots p_{\text{o}}^{n_g^{\text{o}}}(k)\right]^\top, \\
\bar{a}_i^{\text{o}} &= \left[\bar{a}_{i,0}^{\text{o}} \;\; \bar{a}_{i,1}^{\text{o}} \cdots \;\; \bar{a}_{i,n_g^{\text{o}}}^{\text{o}}\right]^\top, \; a_i^{\text{o}}(p_{\text{o}}(k)) = (\bar{a}_i^{\text{o}})^\top \mathbf{p}_{\text{o}}(k), \\
\bar{b}_j^{\text{o}} &= \left[\bar{b}_{j,0}^{\text{o}} \;\; \bar{b}_{j,1}^{\text{o}} \cdots \;\; \bar{b}_{j,n_g^{\text{o}}}^{\text{o}}\right]^\top, \; b_j^{\text{o}}(p_{\text{o}}(k)) = \left(\bar{b}_j^{\text{o}}\right)^\top \mathbf{p}_{\text{o}}(k), \\
\theta_{\text{o}} &= \left[(\bar{a}_1^{\text{o}})^\top \cdots \left(\bar{a}_{n_a^{\text{o}}}^{\text{o}}\right)^\top (\bar{b}_0^{\text{o}})^\top \cdots \left(\bar{b}_{n_b^{\text{o}}}^{\text{o}}\right)^\top\right]^\top, \\
\chi_{\text{o}}(k) &= [-y_{\text{o}}(k-1) \cdots - y_{\text{o}}(k-n_a^{\text{o}}) \;\; u(k) \cdots u(k-n_b^{\text{o}})]^\top, \\
\phi_{\text{o}}(k) &= \chi_{\text{o}}(k) \otimes \mathbf{p}_{\text{o}}(k).
\end{aligned}$$

The data-generating system (1) can be then rewritten as:

$$y(k) = \phi_{\text{o}}^\top(k)\theta_{\text{o}} + \nu_{\text{o}}(k). \quad (2)$$

3.2 Scheduling signal observations

The measurements $p(k)$ of the scheduling signal are assumed to be corrupted by an additive noise $\eta_{\text{o}}(k)$, i.e.,

$$p(k) = p_{\text{o}}(k) + \eta_{\text{o}}(k), \quad (3)$$

where $\eta_{\text{o}}(k) \sim \mathcal{N}(0, \sigma_\eta^2)$ is a zero mean white Gaussian noise uncorrelated with the noise corrupting the output signal, i.e., $\mathbb{E}[\eta_{\text{o}}(k)\nu_{\text{o}}(t)] = 0$ for all time indexes k and t.

3.3 LPV model structure

The following model structure ($\mathcal{M}$) is considered to estimate the data-generating system (1):

$$y(k) = -\sum_{i=1}^{n_a} a_i(p(k))y(k-i) + \sum_{j=0}^{n_b} b_j(p(k))u(k-j) + \epsilon(k), \quad (4)$$

with $\epsilon(k)$ denoting the residual term, modelling the mismatch between the true system and the model output. For the true system to belong to the model class $\mathcal{M}$, the parameters n_a and n_b defining the dynamical order of the model in (4) are chosen large enough so that $n_a \geq n_a^{\text{o}}$ and $n_b \geq n_b^{\text{o}}$. In other words, an overparametrized model structure is used. Moreover, the functions $a_i : \mathbb{R} \to \mathbb{R}$ and $b_j : \mathbb{R} \to \mathbb{R}$ are parametrized with polynomial basis functions as follows:

$$a_i(p(k)) = \bar{a}_{i,0} + \sum_{s=1}^{n_g} \bar{a}_{i,s}p^s(k) = (\bar{a}_i)^\top \mathbf{p}(k), \quad \text{(5a)}$$

$$b_j(p(k)) = \bar{b}_{j,0} + \sum_{s=1}^{n_g} \bar{b}_{j,s}p^s(k) = \left(\bar{b}_j\right)^\top \mathbf{p}(k), \quad \text{(5b)}$$

where the degree n_g of the polynomials in (5) is also chosen large enough so that $n_g \geq n_g^{\text{o}}$ and $\mathbf{p}(k) = \left[1 \;\; p(k) \;\; p^2(k) \cdots p^{n_g}(k)\right]^\top$.

Using the matrix notations introduced in Section 3.1, model (4) can be compactly written as:

$$y(k) = \phi^\top(k)\theta + \epsilon(k), \quad (6)$$

where, $\theta = [\bar{a}_1^\top \cdots \bar{a}_{n_a}^\top \;\; \bar{b}_0^\top \cdots \bar{b}_{n_b}^\top]^\top \in \mathbb{R}^{n_\theta}$ is the vector of model parameters to be identified, $\bar{a}_i$ and $\bar{b}_i$ are defined similarly to $\bar{a}_i^{\text{o}}$ and $\bar{b}_i^{\text{o}}$, and $\phi(k)$ is the regressor with measured (thus, noise-corrupted) outputs and scheduling signals at time k, defined as $\phi(k) = \chi(k) \otimes \mathbf{p}(k)$, with

$$\chi(k) = [-y(k-1) \cdots - y(k-n_a) \;\; u(k) \cdots u(k-n_b)]^\top.$$

The identification problem addressed in this paper aims at obtaining an asymptotically unbiased estimate of the "true" parameter vector θ_o along with the unknown parameters n_a^o, n_b^o and n_g^o from an N-length observed data sequence $\mathcal{D}_N = \{u(k), y(k), p(k)\}_{k=1}^N$ generated by (1).

4. INSTRUMENTAL-VARIABLE ESTIMATE

As proposed in Laurain et al. (2010) and Piga et al. (2015), an *instrumental variable* approach can be used to handle the bias due to the noise $\nu_o(k)$ affecting the output signal measurements. This bias can be removed by choosing the instruments $z(k) \in \mathbb{R}^{n_\theta}$ in such a way that they are uncorrelated with the output noise $\nu_o(k)$, i.e., $\mathbb{E}[z(k)\nu_o(k)] = 0$ for all k. In the following, we show that, because of the effect of the noise on the scheduling signal, using only instrumental variables is not enough to achieve a consistent parameter estimate and thus an accurate selection of the model structure.

Consider the following IV-LASSO optimization problem for estimating a sparse model parameter vector θ:

$$\hat{\theta}_{\mathrm{IV}} = \underset{\theta}{\operatorname{argmin}} \mathcal{J}_{\mathrm{IV}}(\theta, \lambda, N), \tag{7}$$

with

$$\mathcal{J}_{\mathrm{IV}}(\theta, \lambda, N) = \left\| \frac{1}{N} \left(Z^\top Y - Z^\top \Phi \theta \right) \right\|_2^2 + \lambda \|\theta\|_1, \tag{8}$$

where $Z = [z(1) \cdots z(N)]^\top$ is the matrix of instruments; $\Phi = [\phi(1) \cdots \phi(N)]^\top$ is the regressor matrix, $Y = [y(1) \cdots y(N)]^\top$ is the noise-corrupted output observation vector. Note that the quadratic term in the definition of $\mathcal{J}_{\mathrm{IV}}$ is the loss function minimized in standard IV identification schemes (Söderström, 2007). The second term is used to enforce sparsity in the estimate of θ, and the hyper-parameter $\lambda \geq 0$ is tuned to balance the trade-off between model complexity and data fitting. The chosen instruments $z(k)$ must be independent of the output noise realization $\nu_o(k)$. Thus, a possible choice of $z(k)$ is:

$$z(k) = [-\hat{y}(k-1) \cdots -\hat{y}(k-n_a) \;\; u(k) \cdots u(k-n_b)]^\top \otimes \mathbf{p}(k),$$

where $\hat{y}$ is an approximation of the noise-free output, independent of the noise ν_o, which can be obtained from an estimated (not necessarily unbiased) model of the system. An iterative algorithm can be implemented to 'refine' the instruments by computing, at each iteration, an estimate θ of the model parameters, based on which the simulated output $\hat{y}$ is generated and used as an instrument at the next run.

Due to noise in the measurements of p, the quadratic cost in equation (8) is asymptotically biased, in the sense that, asymptotically, its minimum is not achieved at the true system parameter vector θ_o (see the Appendix for a proof). In order to overcome this drawback, instead of the cost in equation (8), a bias-corrected cost function achieving a consistent estimate of θ_o, along with an accurate model order selection, is introduced in the following section.

5. BIAS-CORRECTED LASSO FOR SPARSE LPV IDENTIFICATION

In this section, we formulate a bias-corrected version of the IV-LASSO cost $\mathcal{J}_{\mathrm{IV}}(\theta, \lambda, N)$ to obtain a consistent estimate of the model parameters as well as an accurate model order selection. It will be proved that the proposed biased-corrected cost function converges asymptotically (as $N \to \infty$) to the "true cost" function (i.e., a non-negative loss function which achieves its minimum at the true parameter vector θ_o). Such a bias-corrected cost function is also used as an optimal criterion to tune, via cross-validation, the regularization parameter λ.

To perform model order selection, let us solve the optimization problem:

$$\hat{\theta}_{\mathrm{CIV}} = \underset{\theta}{\operatorname{argmin}} \mathcal{J}_{\mathrm{CIV}}(\theta, \lambda, N), \tag{9}$$

with

$$\mathcal{J}_{\mathrm{CIV}}(\theta, \lambda, N) = \left\| \frac{1}{N} \left(Z^\top Y - \Psi \theta \right) \right\|_2^2 + \lambda \|\theta\|_1. \tag{10}$$

The function $\mathcal{J}_{\mathrm{CIV}}(\theta, \lambda, N)$ will be referred to as "bias-corrected cost". In case $\lambda = 0$, $\mathcal{J}_{\mathrm{CIV}}(\theta, \lambda, N)$ will be referred to as "non-regularized bias-corrected cost".

The matrix Ψ appearing in the definition of $\mathcal{J}_{\mathrm{CIV}}(\theta, \lambda, N)$ was originally introduced in Piga et al. (2015) and it is given by $\Psi = \sum_{k=1}^N \Psi_k$, where each matrix Ψ_k is constructed in a way to satisfy the following conditions:

- C1. the matrix Ψ_k only depends on the noise-corrupted observations of the data $\{u(k), p(k), y(k)\}_{k=1}^N$ and on the variance σ_η^2 of the noise $\eta_o(k)$ corrupting the scheduling variable measurement $p(k)$.
- C2. let $\Omega_k = z(k)[\chi(k) \otimes \mathbf{p}_o(k)]^\top$. Then,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N \Omega_k = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^N \Psi_k \;\text{ w.p.1.}$$

5.1 Construction of Ψ_k

As explained in Piga et al. (2015), the matrices Ψ_k satisfying conditions C1 and C2 can be constructed through the procedure outlined below (inspired by Piga and Tóth (2014)), under the assumption that the variance σ_η^2 of the noise corrupting the scheduling observations $p(k)$ is known.

1. Compute the analytic expression of the conditional expectation $\mathbb{E}[\Omega_k | Y]$. By construction, since each element of the vector $\mathbf{p}_o(k)$ is a polynomial in $p_o(k)$, the entries of $\mathbb{E}[\Omega_k | Y]$ are described by an affine combination of the monomials $p_o(k), p_o^2(k), p_o^3(k), \ldots$.
2. Express the n^{th}-order monomial of the noise-free scheduling signal $p_o^n(k)$ in terms of the expected value of the noise-corrupted monomial $p^n(k)$ and the noise variance σ_η^2 in terms of the Hermite polynomial:

$$p_o^n(k) = \mathbb{E}\left[(n!) \sum_{m=0}^{\lfloor n/2 \rfloor} \frac{(-1)^m \sigma_\eta^{2m}}{m!(n-2m)!} \frac{p^{n-2m}(k)}{2^m} \right] \tag{11}$$

3. Compute the matrix Ψ_k by replacing each of the monomials $p_o(k)$, $p_o^2(k)$, $p_o^3(k), \ldots$ appearing in the analytic expression of $\mathbb{E}[\Omega_k | Y]$, with the term inside the expectation operator in (11). In this way, the matrix Ψ_k satisfies the following condition:

$$\mathbb{E}\left[\frac{1}{N} \Omega_k | Y \right] = \mathbb{E}\left[\frac{1}{N} \Psi_k | Y \right] \quad \forall k \in \mathbb{I}_1^N. \tag{12}$$

Property 1. The computed matrices Ψ_k satisfy conditions C1 and C2 under the assumption that the amplitude of the measured output and of the scheduling signals is bounded. See (Piga et al., 2015, Appendix A2) for a detailed proof.

5.2 Consistency of the bias-corrected cost function

In this section, we prove that minimizing the non-regularized bias-corrected cost $\mathcal{J}_{\mathrm{CIV}}(\theta, 0, N)$ leads to a consistent estimate of the system parameters θ_{o}. This basically means that the bias on the estimated parameters $\hat{\theta}_{\mathrm{CIV}}$ due to the noise on the data vanishes as the length N of the training dataset increases.

In the following, we prove that the non-regularized bias-corrected cost $\mathcal{J}_{\mathrm{CIV}}(\theta, 0, N)$ converges pointwise (as $N \to \infty$) to the "true" cost $\mathcal{V}_{\mathrm{o}}(\theta, N)$ defined as:

$$\mathcal{V}_{\mathrm{o}}(\theta, N) = \left\| \frac{1}{N} \sum_{k=1}^{N} z(k) \left(y_{\mathrm{o}}(k) - [\chi_{\mathrm{o}}(k) \otimes \mathbf{p}_{\mathrm{o}}(k)]^\top \theta \right) \right\|_2^2, \quad (13)$$

which achieves its minimum $\mathcal{V}_{\mathrm{o}}(\theta, N) = 0$ at $\theta = \theta_{\mathrm{o}}$. In the following, we will assume that θ_{o} is the only minimizer of $\mathcal{V}_{\mathrm{o}}(\theta, N)$.

Proposition 1. Let us define the cost function:

$$\mathcal{J}_{\mathrm{o}}(\theta, N) = \left\| \frac{1}{N} \sum_{k=1}^{N} z(k) \left(y(k) - [\chi(k) \otimes \mathbf{p}_{\mathrm{o}}(k)]^\top \theta \right) \right\|_2^2. \quad (14)$$

Then, for any compact set $\Theta \subset \mathbb{R}^{n_\theta}$, the following property holds:

$$\lim_{N\to\infty} \mathcal{J}_{\mathrm{o}}(\theta, N) = \lim_{N\to\infty} \mathcal{V}_{\mathrm{o}}(\theta, N) \quad \forall \theta \in \Theta. \quad (15)$$

Proof: Consider the limit of the argument of the 2-norm in the cost (14):

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} z(k) \left(y(k) - [\chi(k) \otimes \mathbf{p}_{\mathrm{o}}(k)]^\top \theta \right) \quad (16a)$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} z(k) y_{\mathrm{o}}(k) \quad (16b)$$

$$+ \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} z(k) \left(\nu_{\mathrm{o}}(k) - [(\chi(k) - \chi_{\mathrm{o}}(k)) \otimes \mathbf{p}_{\mathrm{o}}(k)]^\top \theta \right) \quad (16c)$$

$$- \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \left(z(k) [\chi_{\mathrm{o}}(k) \otimes \mathbf{p}_{\mathrm{o}}(k)]^\top \theta \right), \quad (16d)$$

where the decomposition above is obtained by splitting $y(k) = y_{\mathrm{o}}(k) + \nu_{\mathrm{o}}(k)$ and $\chi(k) = \chi(k) - \chi_{\mathrm{o}}(k) + \chi_{\mathrm{o}}(k)$ in (16a). We now analyse the asymptotic behaviour of (16). The term (16c) converges to zero as N tends to infinity. This follows from the fact that $\nu_{\mathrm{o}}(k)$ and $(\chi(k) - \chi_{\mathrm{o}}(k))$ are zero mean noises and they are uncorrelated with the instrument $z(k)$. The remaining term is the sum of (16b) and (16d), which is equal to the argument of the 2-norm of the the true cost $\mathcal{V}_{\mathrm{o}}(\theta, N)$ (13). Thus, since the argument of the 2-norm in (13) converges to the argument of the 2-norm in (14), because of continuity of the 2-norm, it follows that, $\lim_{N\to\infty} \mathcal{J}_{\mathrm{o}}(\theta, N) = \lim_{N\to\infty} \mathcal{V}_{\mathrm{o}}(\theta, N)$ for any $\theta \in \Theta$. ■

Next step is to prove that the non-regularized cost $\mathcal{J}_{\mathrm{CIV}}(\theta, 0, N)$ (10) converges asymptotically to $\mathcal{J}_{\mathrm{o}}(\theta, N)$, or equivalently, because of Proposition 1, to the true cost function $\mathcal{V}_{\mathrm{o}}(\theta, N)$. This also implies that minimizing the non-regularized bias-corrected cost $\mathcal{J}_{\mathrm{CIV}}(\theta, 0, N)$ provides a consistent parameter estimate $\hat{\theta}_{\mathrm{CIV}}$ which asymptotically converges to true parameter vector θ_{o}.

Proposition 2. For any compact set $\Theta \subset \mathbb{R}^{n_\theta}$, the following property holds:

$$\lim_{N\to\infty} \mathcal{J}_{\mathrm{CIV}}(\theta, 0, N) = \lim_{N\to\infty} \mathcal{J}_{\mathrm{o}}(\theta, N) \quad \forall \theta \in \Theta. \quad (17)$$

Proof: Let us rewrite the argument of the 2-norm of the bias-corrected cost $\mathcal{J}_{\mathrm{CIV}}(\theta, 0, N)$ as:

$$\frac{1}{N} \sum_{k=1}^{N} z(k) y(k) - \frac{1}{N} \sum_{k=1}^{N} \Psi_k \theta. \quad (18)$$

By construction of matrix Ψ_k, we have (see Condition C2):

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \Psi_k = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \Omega_k. \quad (19)$$

Taking the limit of (18) and substituting (19) into (18) we obtain:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} z(k) y(k) - \frac{1}{N} \sum_{k=1}^{N} \Psi_k \theta \quad (20a)$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} z(k) y(k) - \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \Omega_k \theta = \quad (20b)$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} z(k) y(k) - z(k) [\chi(k) \otimes \mathbf{p}_{\mathrm{o}}(k)]^\top \theta. \quad (20c)$$

Note that the argument of the limit in (20c) is the argument of the 2-norm of the cost $\mathcal{J}_{\mathrm{o}}(\theta, N)$. Thus, because of continuity of the 2-norm function, eq. (17) follows. ■

Combining Proposition 1 and 2, we obtain:

$$\lim_{N\to\infty} \mathcal{J}_{\mathrm{CIV}}(\theta, 0, N) = \lim_{N\to\infty} \mathcal{V}_{\mathrm{o}}(\theta, N) \quad \forall \theta \in \Theta. \quad (21)$$

Condition (21) implies that the minimum $\hat{\theta}_{\mathrm{CIV}}$ of the the non-regularized bias-corrected cost function $\mathcal{J}_{\mathrm{CIV}}(\theta, 0, N)$ converges to the minimum of the true cost $\mathcal{V}_{\mathrm{o}}(\theta, N)$, i.e.,

$$\lim_{N\to\infty} \hat{\theta}_{\mathrm{CIV}} = \theta_{\mathrm{o}}. \quad (22)$$

The advantages of introducing the bias-corrected cost $\mathcal{J}_{\mathrm{CIV}}(\theta, 0, N)$ are twofold:

- it allows us to perform model order selection by adding an ℓ_1-regularization term and it gives a 'correct' quadratic error-fitting term in order to remove the bias (asymptotically) due to noise affecting the scheduling variable observations.
- the bias-corrected cost $\mathcal{J}_{\mathrm{CIV}}(\theta, 0, N)$ is an unbiased criterion which can be used to assess the performance of the estimated model, and thus to tune the hyper-parameter λ via cross validation.

Note that, in the bias-corrected LASSO cost (10), the quadratic fitting-error term is asymptotically unbiased and the ℓ_1 regularization enforces sparsity in the final estimate, shrinking the component of the vector θ towards zero. Then, the finale estimate $\hat{\theta}_{\mathrm{CIV}}$ is actually biased. Nevertheless, once the model order is selected based on the regularized cost $\mathcal{J}_{\mathrm{CIV}}(\theta, \lambda, N)$, the zero components of the estimated parameter vector $\hat{\theta}_{\mathrm{CIV}}$ are discarded and a lower complex model is re-identified by minimizing the non-regularized biased-corrected cost $\mathcal{J}_{\mathrm{CIV}}(\theta, 0, N)$, thus obtaining a consistent estimate.

5.3 Estimation with unknown noise variance

In computing the bias correcting matrix Ψ_k (Section 5.1), the variance of the noise σ_η^2 corrupting the scheduling signal measurements is assumed to be known. This is quite a restrictive assumption. Nevertheless, an exhaustive grid search over the scalar σ_η^2 can be performed, and the non-regularized bias-corrected cost $\mathcal{J}_{\text{CIV}}(\hat{\theta}_{\text{CIV}}(\sigma_\eta^2), 0, N)$ can be used as a performance metric, on calibration data, to tune the "optimal" variance σ_η^2 via cross-validation. This tuning of σ_η^2 is simpler than the one used in Piga et al. (2015), which requires to solve a set of nonlinear equations.

6. SIMULATION EXAMPLES

This section illustrates the effectiveness of the proposed method on a simulation example.

6.1 Data-generating system and model structure

The considered data-generating system is of the form:

$$y_{\text{o}}(k) = a_{1,2}^{\text{o}} p_{\text{o}}^2(k) y_{\text{o}}(k-1) + b_{0,2}^{\text{o}} p_{\text{o}}^2(k) u(k),$$
$$y(k) = y_{\text{o}}(k) + \nu_{\text{o}}(k), \;\; p(k) = p_{\text{o}}(k) + \eta_{\text{o}}(k),$$

with true parameters $a_{1,2}^{\text{o}} = -0.8$ and $b_{0,2}^{\text{o}} = 0.4$. The input $u(k)$ is taken as a white-noise with uniform distribution in the interval $[0 \;\; 1]$. The noise-free scheduling variable is given by $p_{\text{o}}(k) = \sin(0.1k) + 0.2\delta(k)$, with $\delta(k)$ being a random variable with Gaussian distribution $\mathcal{N}(0,1)$. The noise signals η_{o} and ν_{o} are white Gaussian noise processes with standard deviation $\sigma_\eta = 0.25$ and $\sigma_\nu = 0.06$. The influence of the noise on the signal measurements is quantified in terms of the signal-to-noise ratios:

$$\text{SNR}_{\text{y}} = 10 \log \frac{\sum_{k=1}^{N} \left(y_{\text{o}}(k) - \bar{y}_{\text{o}}(k)\right)^2}{\sum_{k=1}^{N} \left(\nu_{\text{o}}(k)\right)^2} = 9 \;\; \text{dB},$$
$$\text{SNR}_{\text{p}} = 10 \log \frac{\sum_{k=1}^{N} \left(p_{\text{o}}(k) - \bar{p}_{\text{o}}(k)\right)^2}{\sum_{k=1}^{N} \left(\eta_{\text{o}}(k)\right)^2} = 10 \;\; \text{dB},$$

with $\bar{y}_{\text{o}}$ and $\bar{p}_{\text{o}}$ being the sample mean of y_{o} and p_{o}.

The following over-parametrized LPV model structure $\mathcal{M}$ is used to describe the behaviour of the system:

$$y(k) = \sum_{i=1}^{n_{\text{a}}} a_i(p(k), \theta) y(k-i) + \sum_{j=0}^{n_{\text{b}}} b_j(p(k), \theta) u(k-j) + \epsilon(k),$$

with $n_a = 4$ and $n_b = 2$. Each coefficient $a_i(.)$ and $b_j(.)$ is parametrized as a second order polynomial in p, i.e.,

$$a_i(p(k), \theta) = a_{i,0} + a_{i,1} p(k) + a_{i,2} p^2(k), \quad (23a)$$
$$b_j(p(k), \theta) = b_{j,0} + b_{j,1} p(k) + b_{j,2} p^2(k). \quad (23b)$$

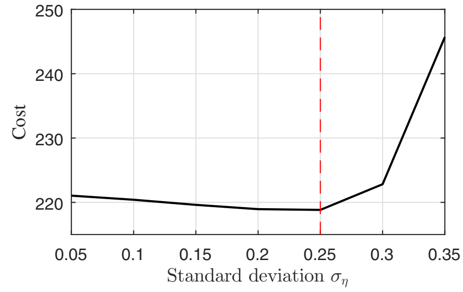

Fig. 1. Non-regularized bias-corrected cost $\mathcal{J}_{\text{CIV}}(\hat{\theta}_{\text{CIV}}(\sigma_\eta), 0, N_c)$ vs noise standard deviation σ_η.

Table 2. BFR computed on validation data.

Method	BFR
IV method	0.6061
Bias-Corrected IV	0.9146

The IV-LASSO method (based on the minimization of $\mathcal{J}_{\text{IV}}(\theta, \lambda, N)$ in (8)) and the Bias-corrected IV-LASSO method (based on the minimization of $\mathcal{J}_{\text{CIV}}(\theta, \lambda, N)$ in (9)) are compared. The model parameters are estimated based on a training set with $N = 8000$ samples. A recursive scheme is used to refine the instruments $z(k)$ as described in (Piga et al., 2015), in order to maximize the accuracy of the estimated parameters. To study the statistical properties of the algorithms, a Monte-Carlo study with 100 runs is performed. At each run, new inputs, scheduling variables and noise signals are generated.

A calibration set with $N_c = 1000$ samples is used to calibrate the hyper-parameter λ and to estimate the noise-variance σ_η^2 (see Section 5.3) through cross-validation. A grid search over the λ and σ_η is then performed, and the combined values of λ and σ_η which provide the best data fit over the calibration dataset are selected. Specifically, the non-regularized IV cost $\mathcal{J}_{\text{IV}}(\hat{\theta}_{\text{IV}}(\sigma_\eta), 0, N_c)$ and the bias-corrected cost $\mathcal{J}_{\text{CIV}}(\hat{\theta}_{\text{CIV}}(\sigma_\eta), 0, N_c)$, computed on the calibration dataset and for different values of λ and σ_η, are used to assess the performance of the estimated models. The computed value of λ is 9.2 for the IV scheme and 22 for the bias-corrected IV. As far as the estimate of the noise variance σ_η^2 is concerned, Fig. 1 shows the non-regularized bias-corrected cost $\mathcal{J}_{\text{CIV}}(\hat{\theta}_{\text{CIV}}(\sigma_\eta), 0, N_c)$ (computed w.r.t. the calibration dataset) as a function of σ_η and for $\lambda = 22$. Note that the minimum of the bias-corrected cost is achieved for $\sigma_\eta = 0.25$, which is exactly the true value of the noise standard deviation. This shows that the bias-corrected cost function provides an efficient criterion to tune the parameter σ_η.

The estimates of the model parameters computed through the IV-LASSO and the bias-corrected IV-LASSO approach are reported in Table 1, which shows the mean and standard deviations of the estimated parameters $a_{i,j}$ over the Monte Carlo runs. The obtained results show that the bias-corrected IV-LASSO provides a parameter estimate close to the true parameters and it detects the structure of the underlying system quite accurately. On the other hand, the noise on the scheduling variable deteriorates the performance of the IV-LASSO scheme, which provides biased parameter estimates and, as a consequence, also a systematic error in the selection of the model structure. Similar results are obtained for the parameters $b_{i,j}$.

The performance of the IV-LASSO and the biased corrected IV-LASSO methods is tested on a validation dataset of length $N_{\text{val}} = 1000$. The Best Fit Rates (BFR) are reported in Table 2, and the true output and the estimated model output $\hat{y}$ are plotted in Fig. 2. The obtained results show the better performance of the bias-corrected IV-LASSO w.r.t. the classical IV-LASSO approach.

7. CONCLUSIONS

This paper has presented an extension of the the bias-corrected IV method (Piga et al., 2015) for sparse identifi-

Table 1. Average and standard deviation (over 100 Monte-Carlo runs) of the LPV model coefficients $a_i(p(t))$ estimated through IV-LASSO and bias-corrected (BC) IV-LASSO.

Coefficients	$a_{1,0}$	$a_{1,1}$	$a_{1,2}$	$a_{2,0}$	$a_{2,1}$	$a_{2,2}$	$a_{3,0}$	$a_{3,1}$	$a_{3,2}$	$a_{4,0}$	$a_{4,1}$	$a_{4,2}$
True Value	0	0	-0.8	0	0	0	0	0	0	0	0	0
Mean (IV)	-0.1715	0	-0.3882	0.1619	0	-0.0040	0.0044	0	0	0.0146	0	-0.0011
Mean (BC-IV)	-0.0302	0	-0.7120	0.0026	0	0.0107	0	0	0	0	0	0.0015
std (IV)	0.0279	0.0050	0.0241	0.0313	0.0040	0.0129	0.0264	0.0027	0.0041	0.0117	0.0022	0.0057
std (BC-IV)	0.0358	0.0019	0.0457	0.0066	0.0031	0.0152	0	0.0029	0.0016	0.0012	0.0014	0.0051

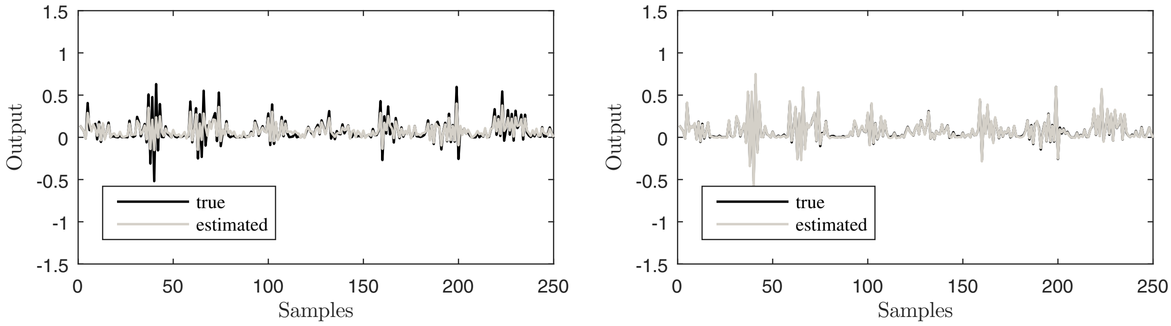

Fig. 2. Validation data. Results achieved by IV-LASSO (left panels) and by the bias-corrected IV-LASSO (right panels).

cation of LPV models from noisy scheduling variable measurements. The LPV model-structure selection problem is solved along with the asymptotically bias-free parameter estimation, through the formulation of a bias-corrected cost function. An ℓ_1-regularization term is used to enforce sparsity in the parameter vector estimate, and the trade-off between model complexity and data-fitting is balanced through cross-validation, using the introduced bias-corrected cost as a performance criterion. Future research includes the extension to the identification of LPV systems under different noise conditions (e.g., Box-Jenkins model structures), and its generalization to non-polynomial dependencies on the scheduling variable.

APPENDIX

We prove that the quadratic term in (8) is an asymptotically biased cost, in the sense that the true parameter $\theta_{\rm o}$ is not guaranteed to be the minimizer of $\mathcal{J}_{\rm IV}(\theta, 0, N)$ as $N \to \infty$. Let us rewrite the quadratic term in (8) as:

$$\mathcal{J}_{\rm IV}(\theta,0,N) = \left\| \frac{1}{N}\sum_{k=1}^{N} z(k)\Big(y(k) - [\chi(k)\otimes \mathbf{p}(k)]^\top \theta\Big) \right\|_2^2 =$$

$$= \left\| \frac{1}{N}\sum_{k=1}^{N} z(k)\left(y_{\rm o}(k) - [\chi_{\rm o}(k)\otimes \mathbf{p}_{\rm o}(k)]^\top \theta\right) \right. \tag{.1a}$$

$$+\frac{1}{N}\sum_{k=1}^{N} z(k)\left(\nu_{\rm o}(k) - [(\chi(k)-\chi_{\rm o}(k))\otimes \mathbf{p}_{\rm o}(k)]^\top \theta\right) \tag{.1b}$$

$$\left. -\frac{1}{N}\sum_{k=1}^{N} z(k)\left([\chi(k)\otimes(\mathbf{p}(k)-\mathbf{p}_{\rm o}(k))]^\top \theta\right) \right\|_2 . \tag{.1c}$$

As $N \to \infty$, (.1b) converges to zero with probability 1 since $z(k)$ is uncorrelated with the output noise $\nu_{\rm o}(k)$ and $\chi(k) - \chi_{\rm o}(k)$ depends linearly on the past samples $\nu_{\rm o}(k)$. The term (.1a) is the ideal cost, since it quantifies the error between the noise-free and the estimated output. For $\theta = \theta_{\rm o}$, (.1a) is equal to 0. The asymptotically non-zero term (.1c) causes the optimum of the asymptotic cost $\mathcal{J}_{\rm IV}(\theta, 0, N)$ to deviate from the optimum of the ideal cost.

REFERENCES

Butcher, M., Karimi, A., and Longchamp, R. (2008). On the consistancy of certain identification methods for linear parameter varying systems. In *Proc. of the 17th IFAC world congress*, 4018–4023. Seoul, South Korea.

Cerone, V., Piga, D., and Regruto, D. (2013). A convex relaxation approach to set-membership identification of LPV systems. *Automatica*, 49(9), 2853–2859.

Laurain, V., Gilson, M., Tóth, R., and Garnier, H. (2010). Refined instrumental variable methods for identification of LPV Box–Jenkins models. *Automatica*, 46(6), 959–967.

Mejari, M., Piga, D., and Bemporad, A. (2016). Regularized least square support vector machines for order and structure selection of LPV-ARX models. In *Proceedings of 15th European Control Conference*, 1649–1654. Aalborg, Denmark.

Piga, D., Cox, P., Tóth, R., and Laurain, V. (2015). LPV system identification under noise corrupted scheduling and output signal observations. *Automatica*, 53, 329–338.

Piga, D. and Tóth, R. (2013). LPV model order selection in an LS-SVM setting. In *52nd IEEE Conference on Decision and Control*, 4128–4133. Florence, Italy.

Piga, D. and Tóth, R. (2014). A bias-corrected estimator for nonlinear systems with output-error type model structures. *Automatica*, 50(9), 2373–2380.

Söderström, T. (2007). Errors-in-variables methods in system identification. *Automatica*, 43(6), 939–958.

Tóth, R., Hjalmarsson, H., and Rojas, C.R. (2012). Order and structural dependence selection of LPV-ARX models revisited. In *Proc. of the 51st IEEE Conference on Decision and Control*, 6271–6276. Maui, Hawaii (USA).

Tóth, R., Lyzell, C., Enqvist, M., Heuberger, P., and Van den Hof, P. (2009). Order and structural dependence selection of LPV-ARX models using a nonnegative garrote approach. In *Proc. of the 48th IEEE Conf. on Decision and Control*, 7406–7411. Shanghai, China.