A Multi-Step Least-Squares Method for Nonlinear Rational Models*

Mingliang Wang[1], Elling W. Jacobsen[1], Véronique Chotteau[2] and Håkan Hjalmarsson[1]

***Abstract*— Models rational in the parameters arise frequently in biosystems and other applications. As with all models that are non-linear in the parameters, direct parameter estimation, using e.g. nonlinear least-squares, can become challenging due to the issues of local minima and finding good initial estimates. Here we propose a multi-step least-squares method for a class of nonlinear rational models. The proposed method is applied to an extended Monod-type model. Numerical simulations indicate that the proposed method is consistent.**

I. INTRODUCTION

Nonlinear rational models (NRM) are widely used to describe various real-world phenomena, e.g. the dynamics of biological reactions [1], [2], the kinetics of chemical reactions [3], and physical applications [4]. The NRMs are a special class of nonlinear models in which the numerator and denominator consist of linear combinations of input variables and unknown parameters. Mathematically, they can be expressed as

$$y(t) = \frac{\sum_i \alpha_i f_i(t)}{\sum_j \beta_j g_j(t)} + e(t), \tag{1}$$

where $y(t)$ is the output variable, $\{f_i(t), g_j(t)\}$ are the noise-free input variables, $e(t)$ is the measurement noise, $\{\alpha_i, \beta_j\}$ are the unknown parameters. Estimating the unknown parameters in the model is essential for the purpose of understanding, predicting and control of the real-world systems. However, it is not a computationally easy task.

In the literature, there are a number of contributions that concern parameter estimation of NRMs. They can be split into linear and nonlinear approaches. In the nonlinear approaches, the model parameters are directly estimated using the nonlinear rational structure, e.g. the nonlinear least-squares method [5] is commonly used for parameter estimation of NRMs in biosystems [6]. Other nonlinear methods include the prediction error method [7], Newton-type methods [8], [9], and back propagation [10]. Unfortunately, these nonlinear approaches are based on local optimizations and hence are highly dependent on the initial values.

To avoid the issue of local optima, linear approaches are proposed as alternative solutions. They transform the rational models into linear ones by taking advantage of the rational structures, e.g. multiplying the denominator in (1) to both sides of the equation makes the model linear in parameters. However, they pay a two-fold penalty in this step:

(i) the parameter estimation problem becomes an errors-in-variables (EIV) problem [11];
(ii) the noise term $(\sum_j \beta_j g_j(t) e(t))$ becomes non-stationary.

For (i), the least-squares estimate can be adjusted by bias-compensation operations, leading to bias-compensated least squares (BCLS) [12]. Based on BCLS, various methods have been developed for NRMs, including the implicit least-squares [4], the extended least-squares [13], and the corrected least-squares [14]. However, little attention has been paid to the implications of (ii) and how to address these.

In addition, a further complication can be observed in many applications, namely that the numerator and denominator terms are not linear in the unknown parameters. This happens, for example, in the catalytic dehydration of n-hexyl alcohol process [3],

$$y(t) = \frac{\gamma_3 \gamma_1 u_1(t)}{1 + \gamma_1 u_1(t) + \gamma_2 u_2(t)}, \tag{2}$$

where the numerator is the product of unknown parameters γ_1, γ_3 and the input variable $u_1(t)$. Thus, in this case, despite the rational structure can be transformed by multiplying the denominator to both sides of the equation, the parameters remain nonlinear in the transformed equation.

Considering the above facts, we propose a multi-step Least-Squares method for nonlinear RATional models (LS-RAT). We illustrate our method using an extended Monod-type model. This is a widely used first-principle model for the kinetics in bio-processes [1]. In the proposed method, following the linear approaches, we transform the nonlinear rational model into a reparameterized linear model that is amenable to least-squares estimation. In order to handle issues (i)–(ii), we extend the BCLS method to the case with non-stationary noise by adding a statistically motivated weighting. By exploiting the structural relation between the reparametrized model and the original model, the original parameters can be estimated from the parameters of the reparameterized model by the least-squares method.

The advantages of the proposed LS-RAT are thoroughly validated in a simulation study.

*This work was supported by the strategic innovation program Processindustriell IT och Automation (PIIA), a common effort by Vinnova, Formas and the Swedish Department of Energy, under the grant SmartFD, contract 2016-02398, by the VINNOVA Competence Centre AdBIOPRO, contract 2016-05181, and by the Swedish Research Council through the research environment NewLEADS - New Directions in Learning Dynamical Systems, contract 2016-06079.

[1]Mingliang Wang, Håkan Hjalmarsson and Elling W. Jacobsen are with the Centre for Advanced Bioproduction and the Department of Automatic Control, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Brinellvägen 8, 114 28 Stockholm, Sweden miwan@kth.se; hjalmars@kth.se; elling.w.jacobsen@ee.kth.se

[2]Véronique Chotteau is with the Centre for Advanced Bioproduction and the Cell Technology Group, Department of Industrial Biotechnology/Bioprocess Design, School of Biotechnology, KTH Royal Institute of Technology, Brinellvägen 8, 114 28 Stockholm, Sweden veronique.chotteau@biotech.kth.se

The rest of this paper is organized as follows. Section 2 introduces extended Monod-type models, which is the type of model we will use to illustrate our ideas. We emphasize, however, that the method is applicable to a wide class of nonlinear rational models, e.g. NRMs in the form of (1). Section 3 shows how to develop the proposed method on the extended Monod-type model: 1) re-write the NRM such that it becomes a linear regression, 2) discuss the impact of measurement noise and handle errors-in-variables problems, 3) estimate the original parameters reliably by a least-squares method. The performance of the proposed method is validated through simulations in Section 4, and Section 5 concludes the paper.

II. THE EXTENDED MONOD MODEL

In this paper, we use an extended Monod-type (EMT) model [15] as a fairly general example for nonlinear rational models. The EMT model is typically used to model the macroscopic reactions in a metabolic process [2]. In the EMT model,

$$v_0(t) = v_{max} \underbrace{\prod_{i\in S} \frac{C_i(t)}{k_i + C_i(t)}}_{\text{Activation}} \underbrace{\prod_{j\in Q} \frac{r_j}{r_j + C_j(t)}}_{\text{Inhibition}}, \tag{3}$$

where S, Q are the sets of rational components that account for activation and inhibition effects in a macroscopic reaction, respectively; $\{C_i(t), C_j(t)\}$ are the corresponding concentrations in S, Q; $\{v_{max}, k_i, r_j\}$ are the parameters. $v_0(t)$ is the reaction rate.

For ease-of-notation we will often limit ourselves to a simple model with two rational components,

$$v_0(t) = v_{max} \frac{C_1(t)}{k_1 + C_1(t)} \frac{r_2}{r_2 + C_2(t)}. \tag{4}$$

Introduce

$$\rho = [r_2, k_1, v_{max}]^T \tag{5}$$

as the unknown parameter vector.

III. A MULTI-STEP LEAST-SQUARES APPROACH

In this section, we introduce the proposed method using the EMT model (4) as an example.

A. Linear Reparameterization

Following the idea of reparameterization in [16], we can transform the rational structure in (4) to a reparameterized linear model. Multiplying both sides by the denominator, (4) can be expressed as

$$(k_1 + C_1(t))\,(r_2 + C_2(t))\,v_0(t) = v_{max} r_2 C_1(t), \tag{6}$$

which, after some re-arrangements, can be written on matrix form,

$$\underbrace{C_1(t)C_2(t)v_0(t)}_{\triangleq y_0(t)} = \underbrace{\begin{bmatrix} -C_1(t)v_0(t) \\ -C_2(t)v_0(t) \\ -v_0(t) \\ C_1(t) \end{bmatrix}^T \begin{bmatrix} r_2 \\ k_1 \\ k_1 r_2 \\ r_2 v_{max} \end{bmatrix}}_{\triangleq \varphi_0^T(t)\theta}, \tag{7}$$

where θ is a set of new parameters that are functions of original parameters ρ in (5),

$$\theta(\rho) = \left[r_2, k_1, k_1 r_2, r_2 v_{max}\right]^T. \tag{8}$$

Our reparameterized linear model is given by

$$y_0(t) = \varphi_0^T(t)\theta. \tag{9}$$

The model can be over-parameterized as the number of parameters increases geometrically with the number of rational components in (3). However, in real practice, the number of rational components are often limited.

B. Taking the measurement noise into account

While (9) is a linear regression, we have neglected any noise. Here, we consider the case that the reaction rate is subject to measurement errors, i.e. we measure

$$v(t) = v_0(t) + e(t), \tag{10}$$

where we assume that $e(t)$ is Gaussian white noise $N(0, \sigma^2)$.

Considering the model (7), $y_0(t)$ and $\varphi_0(t)$ are not available. Instead we have to contend with

$$y(t) := C_1(t)C_2(t)v(t) = y_0(t) + \tilde{y}(t) \text{ and} \tag{11}$$

$$\varphi(t) := \begin{bmatrix} -C_1(t)v(t) \\ -C_2(t)v(t) \\ -v(t) \\ C_1(t) \end{bmatrix} = \varphi_0(t) + \tilde{\varphi}(t), \tag{12}$$

where

$$\tilde{y}(t) := C_1(t)C_2(t)e(t),\ \tilde{\varphi}(t) := \begin{bmatrix} -C_1(t)e(t) \\ -C_2(t)e(t) \\ -e(t) \\ 0 \end{bmatrix}. \tag{13}$$

This means that we are in an errors-in-variables setting and using $\varphi(t)$ and $y(t)$ instead of $\varphi_0(t)$ and $y_0(t)$, respectively, in a least-squares estimate will result in a biased estimate. In the appendix we review bias-compensated least-squares (BCLS) method [12] which we use as a base for the developments below.

It is important to notice that the error terms $\tilde{y}(t)$ and $\tilde{\varphi}(t)$ are non-stationary. For example, the variance of $\tilde{y}(t)$, given by

$$\mathrm{E}[\tilde{y}^2(t)] = C_1^2(t)C_2^2(t)\sigma^2, \tag{14}$$

is time-dependent since $C_1(t)$ and $C_2(t)$ vary over time. To make best use of the information in the data, this should be accounted for when estimating the model parameters. Using (11)–(13), we can write

$$\begin{aligned} y(t) &:= y_0(t) + \tilde{y}(t) = \varphi_0^T(t)\theta + \tilde{y}(t) \\ &= \varphi^T(t)\theta + \varepsilon(t), \end{aligned} \tag{15}$$

where $\varepsilon(t) = \tilde{y}(t) - \tilde{\varphi}^T(t)\theta$. It is well known [17] that measurements should be weighted with the inverse of the variance of the error term, in this case $\varepsilon(t)$. This variance is conveniently computed by observing that

$$\varepsilon(t) = D(t)e(t), \tag{16}$$

where $D(t)$ is the denominator of the ration model, e.g. for our model (4),

$$D(t) = (k_1 + C_1(t))(r_2 + C_2(t)). \quad (17)$$

This can be seen from the example in (4) where we substitute v_0 with the noisy measurement $v(t)$. Thus, since the noise $\{e(t)\}$ is assumed to be Gaussian i.i.d., $\varepsilon(t) \sim N(0, D(t)^2\sigma^2)$. Based on the above facts, we propose the following weighted BCLS estimate

$$\hat{\theta} = (\hat{R} - \tilde{R})^{-1} (\hat{r} - \tilde{r}), \quad (18)$$

where

$$\hat{R} = \frac{1}{N}\sum_{t=1}^{N} W_t \varphi(t)\varphi(t)^T,\ \tilde{R} = \frac{1}{N}\sum_{t=1}^{N} W_t \mathrm{E}\left[\tilde{\varphi}(t)\tilde{\varphi}(t)^T\right],$$
$$\hat{r} = \frac{1}{N}\sum_{t=1}^{N} W_t \varphi(t) y(t),\ \tilde{r} = \frac{1}{N}\sum_{t=1}^{N} W_t \mathrm{E}\left[\tilde{\varphi}(t)\tilde{y}(t)\right]. \quad (19)$$

To take the statistical properties into account, we set the weight

$$W_t = 1/(D(t)^2\sigma^2). \quad (20)$$

Computing the bias compensating terms $\tilde{R}$ and $\tilde{r}$ requires the second order statistics $E\left[\tilde{\varphi}(t)\tilde{\varphi}(t)^T\right]$ and $E\left[\tilde{\varphi}(t)\tilde{y}(t)\right]$. These can be readily computed from the expressions for how these quantities depend on $e(t)$ and the noise variance σ^2. From (13) follows

$$\begin{aligned} \mathrm{E}\left[\tilde{\varphi}(t)\tilde{\varphi}(t)^T\right] &= \begin{bmatrix} C_1^2 & C_1C_2 & C_1 & 0 \\ C_1C_2 & C_2^2 & C_2 & 0 \\ C_1 & C_2 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \sigma^2, \\ \mathrm{E}\left[\tilde{\varphi}(t)\tilde{y}(t)\right] &= -\begin{bmatrix} C_1^2C_2 & C_1C_2^2 & C_1C_2 & 0 \end{bmatrix}^T \sigma^2, \end{aligned} \quad (21)$$

where C_1, C_2 are short for $C_1(t), C_2(t)$. Thus, given the noise variance σ^2, the required statistics can be computed.

So far we have proposed the estimate $\hat{\theta}$ in (18). However, in order to compute this estimate, we need to know ρ since the weighting W_t depends on ρ (see (20) and (17)). We address this issue in the sequel.

C. Estimating the original parameters

Next we consider how to to estimate the original parameter vector ρ from an estimate $\hat{\theta}$ of (19). One possibility is to use indirect PEM [18]. Here one forms the cost function as

$$J(\rho) = \left(\hat{\theta} - \theta(\rho)\right)^T \Delta \left(\hat{\theta} - \theta(\rho)\right), \quad (22)$$

where the weighting matrix Δ is taken as $\Delta = \mathrm{cov}(\hat{\theta})^{-1}$. However, this leads to a non-convex optimization problem

$$\hat{\rho} = \arg\min_{\rho} J(\rho), \quad (23)$$

which means that not much is gained with our procedure compared to directly minimizing the prediction error criterion

$$J_{PE}(\rho) := \sum_{t=1}^{N} \left[v(t) - \frac{v_{max}C_1(t)}{k_1 + C_1(t)} \frac{r_2}{r_2 + C_2(t)} \right]^2. \quad (24)$$

Instead, we will explore the structure of our problem. We begin by noticing that the original parameter vector is a part of θ. Thus, by picking out certain elements of $\hat{\theta}$, denoted by $\hat{\theta}_A$, we can get an estimate of ρ. However, this means we discard the information about ρ in the unpicked elements of $\hat{\theta}$, denoted by $\hat{\theta}_B$, since all elements of $\hat{\theta}$ actually contain information about ρ. For example, in (7) we have $\theta\{1\} = \rho\{1\}$, $\theta\{2\} = \rho\{2\}$ and $\theta\{3\} = \rho\{1\}\rho\{2\}$, where $\theta\{i\}$ is the i_{th} element of θ. To fully extract the information, we take advantage of the structure relation between θ and ρ by substituting the picked elements $\hat{\theta}_A$ to the unpicked $\hat{\theta}_B$ in a way amenable to least-squares. For example, we have

$$\hat{\theta}\{1\} \approx \rho\{1\}, \quad (25)$$
$$\hat{\theta}\{3\} \approx \hat{\theta}\{2\}\rho\{1\}, \quad (26)$$

which is indeed two observation equations that can be used to estimate $p\{1\}$ in a least-squares sense. Based on this idea, the observation equations can be constructed as

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hat{\theta}\{2\} & 0 & 0 \\ 0 & 0 & \hat{\theta}\{1\} \end{bmatrix} \rho = H\rho = \hat{\theta}. \quad (27)$$

Based on (27), a least-square estimate of the original parameter ρ can be obtained,

$$\hat{\rho} = H^{\dagger}\hat{\theta}, \quad (28)$$

where $H^{\dagger}$ is the pseudo-inverse of H.

The proposed method is to first estimate θ using the ordinary least-squares method,

$$\hat{\theta} = \hat{R}^{-1}\hat{r}, \quad (29)$$

where we set $W_t = 1, t = 1, \cdots, N$. Then we use (28) to estimate ρ. This estimate is subsequently used in the weighting W_t, i.e. the estimate of ρ is used in (17) to form an estimate of $D(t)$ and then use this estimate subsequently to form an estimate of the weighting in (20). We then re-estimate θ using (18) with this weighting. This procedure can then be repeated.

In summary, the proposed LS-RAT is given by:

(i) Set $k = 0$ and calculate $\hat{\theta}_k$ by (29).
(ii) Use $\hat{\theta}_k$ to calculate the estimate $\hat{\rho}_{k+1}$ by (28).
(iii) Use $\hat{\rho}_{k+1}$ to replace ρ in (20) (by way of (17)) and then calculate the estimate $\hat{\theta}_{k+1}$ using (18). Set $k \to k+1$.

Steps (ii) and (iii) can be repeated. One possible stopping criterion is to check when the prediction error criterion starts to increase, that is $J_{PE}(\hat{\rho}_{k+1}) > J_{PE}(\hat{\rho}_k)$. The convergence of the proposed method will be established in future work.

IV. SIMULATIONS

To validate the proposed method, Monte Carlo (MC) experiments are conducted in this section. The experiments are carried out using MATLAB® on a 2.80 GHz laptop PC with Windows 10.

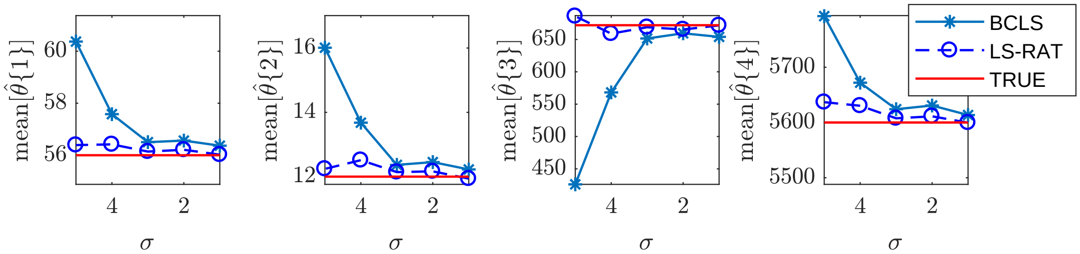

Fig. 1. The mean values of $\hat{\theta}$ (over 100 MC simulations) estimated by LS-RAT and BCLS under different noise levels.

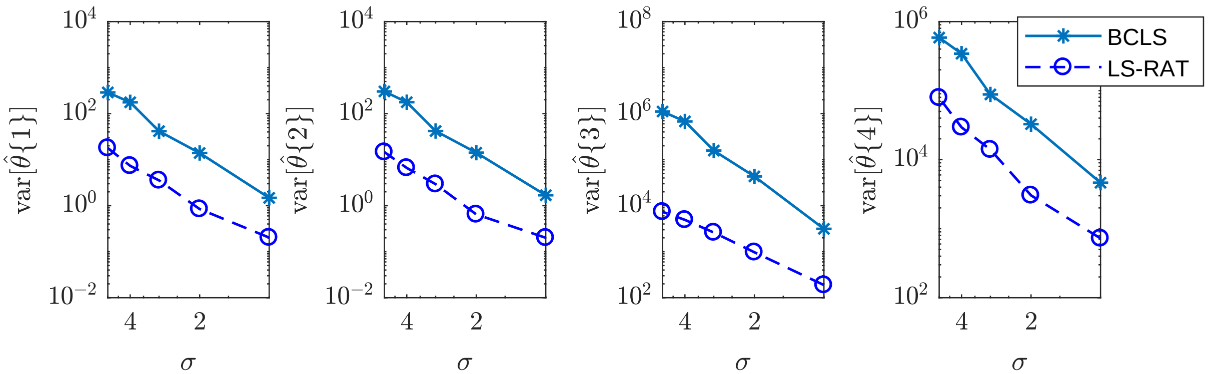

Fig. 2. The variances of $\hat{\theta}$ (over 100 MC simulations) estimated by LS-RAT and BCLS under different noise levels.

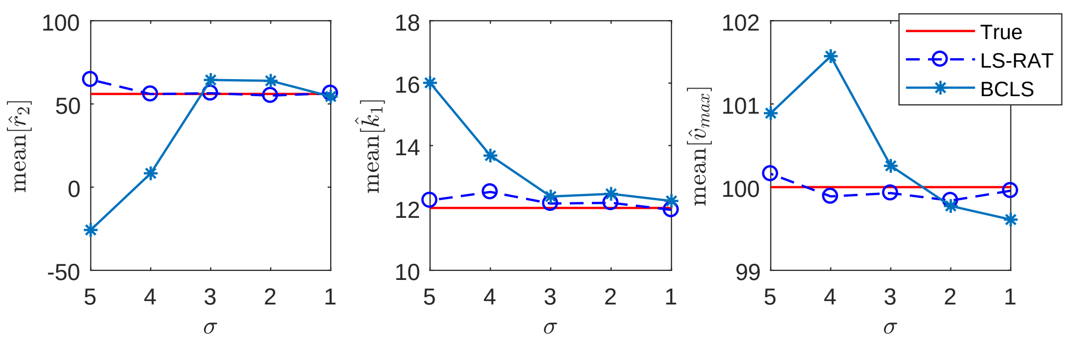

Fig. 3. The mean values of original parameters $\hat{\rho}$ (over 100 MC simulations) estimated by LS-RAT and BCLS under different noise levels.

A. Experimental settings

This simulation is conducted with the rational model in (4). The measurement noise $e(t) \sim N(0, \sigma^2)$ and the system inputs $C_1(t)$ and $C_2(t)$ are uniformly sampled from $[1, 200]$ and $[1, 300]$, respectively. The parameter vector ρ is taken as follows, $\rho = [56, 12, 100]$ which corresponds to $\theta = [56, 12, 672, 5600]$. The proposed method is applied to 100 Monte Carlo realization of noise with different noise variances. The sample size N is fixed at 500. Steps (ii) and (iii) in LS-RAT were iterated until $\|\hat{\theta}_{k+1} - \hat{\theta}_k\| < 10^{-3}$.

B. Results

We compare the accuracy of $\hat{\theta}$ estimated by LS-RAT and the following algorithm:

(a) Estimate θ using BCLS
(b) Use the estimate from (a) to obtain an estimate of ρ via (28).

This algorithm is denoted as BCLS. The first step is identical to the first step used in [14] (which proceeds by using non-linear least-squares). The results under various noise levels (σ) are presented in Fig. 1 and Fig. 2. It is seen that LS-RAT performs much better than BCLS in term of both the mean and variance of the estimates.

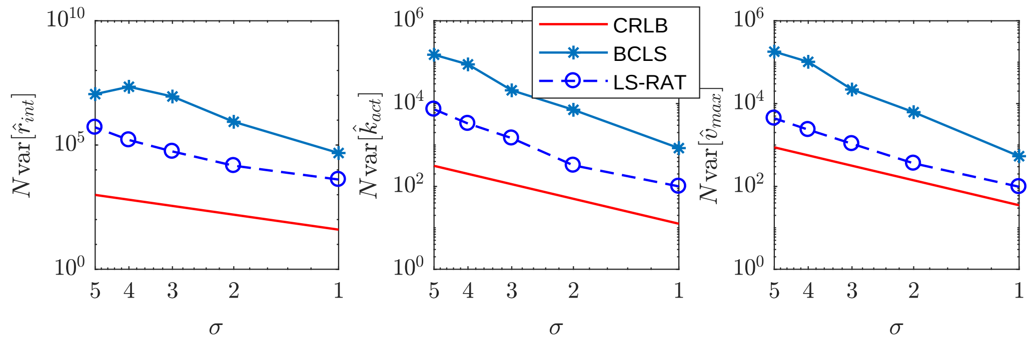

Fig. 4. Normalized variance of $\hat{r}_2$, $\hat{k}_1$ and $\hat{v}_{max}$ (y-axis) under different noise levels (x-axis) for the BCLS, LS-RAT, Cámer-Rao lower bound (the best variance we can achieve).

Similarly, we study the performance of the two methods on estimating the original parameters $\hat{\rho}$. The mean of the estimates for each original parameter is presented in Fig. 3 which indicates the excellence of LS-RAT under all noise levels. The asymptotic normalized variances of the estimates of $\hat{\rho}$ are presented in Fig. 4, from which we can see the higher statistical accuracy of our method (LS-RAT) when compared with $\hat{\rho}$ obtained by BCLS.

Then the fit of proposed method is examined with the normalized mean-square-error metric

$$Err_j = \|\hat{\rho}_j - \rho\|^2 \,/\, \|\hat{\rho}_j\|^2 \,, \tag{30}$$

where $\hat{\rho}_j$ is the estimate from the j_{th} Monte Carlo run. To compare the mean performance, the average error over 100 Monte Carlo run is defined as

$$AvErr = \frac{1}{100}\sum_{j=1}^{100} Err_j. \tag{31}$$

The data size is set to $N = \{50, 100, 300, 500, 700, 900\}$ and the noise level $\sigma = 2$. We present the averaged error of the parameter vectors in Fig. 5, where the proposed method (LS-RAT) is compared with BCLS. As shown in Fig. 5, the errors decrease with increasing data size N and the proposed method shows much lower error than the other methods at all N.

In addition, we also compare the proposed method (LS-RAT) with the nonlinear least-squares (NLLS) method. The experimental settings are the same except $v_{max} = 500$, $\sigma = 2$ and the sample size fixed at a large number $N = 5000$. We implement the NLLS using the function "lsqnonlin" in Matlab. The NLLS starts with initial value $[1, 1, 1]$. The simulation results are presented in Table I which clearly shows the NLLS trapped in a local optimal while the proposed method is computationally reliable.

V. CONCLUSIONS

In this paper, we proposed a multi-step least-squares method for parameter estimation of nonlinear rational models. For convenience, we use an EMT model to illustrate our method. The nonlinear rational model is re-parameterized into a linear regressions model to avoid direct nonlinear optimization. However, when noise is present, this transformation leads to an EIV regression model with a new parameter vector θ and a non-stationary error $\varepsilon(t)$. We addressed this by incorporating bias compensation and proper weighting procedures in our method. Moreover, the estimates of the new parameter vector θ are fully used to re-estimate the original vector by solving another least-squares problems. In the simulation experiments, we demonstrated that our method can avoid the local minima encountered by the NLLS method. We also validated that the proposed weighting improved the accuracy in comparison with the non-weighted BCLS method. Finally, a study based on increasing sample size indicates that the novel estimator is consistent as the error decays with increasing sample size.

Although the proposed method was presented for an EMT

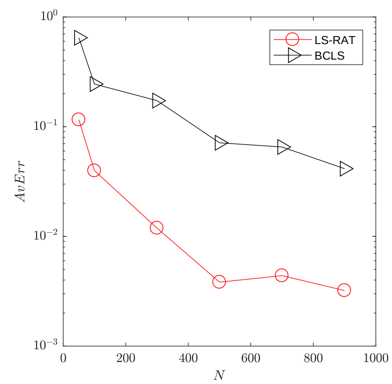

Fig. 5. The average error, as defined in (31), of $\hat{\rho}$ over 100 realization of noise for different sample size N.

TABLE I
COMPARISON OF NLLS AND LS-RAT

Performance \ Methods	NLLS	LS-RAT
$\hat{r}_{inh}$	$2{\times}10^{-9}$	55.95
$\hat{k}_{act}$	16.4	12.01
$\hat{v}_{max}$	307.105	500.21

model, it is also applicable to the nonlinear rational models in the form of (1) which is linear in parameters after multiplying the denominator to both sides of the model. Therefore, our method are suitable for many real world applications [19], e.g. the catalytic dehydration of n-hexyl alcohol model [3] and rational function neural network structures [20].

APPENDIX

BIAS-COMPENSATED LEAST-SQUARES

In the appendix we review bias-compensated least-squares [12]. Consider a linear model,

$$y_0(t) = \varphi_0^T(t)\theta_0, \tag{32}$$

where $y_0(t)$ is the regressed variable, $\varphi_0(t)$ is an m-vector of regression variables, θ_0 is an m-vector of unknown parameters. In EIV problems, noises exist in the measurements $y(t)$ and $\varphi(t)$,

$$y(t) = y_0(t) + \tilde{y}(t), \tag{33}$$

$$\varphi(t) = \varphi_0(t) + \tilde{\varphi}(t), \tag{34}$$

where $\tilde{\varphi}(t)$ and $\tilde{y}(t)$ are the noise terms. The least-squares estimate (LS) of θ_0 is

$$\begin{aligned} \hat{\theta}_{LS} &= \left(\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi(t)^T\right)^{-1}\left(\frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t)\right) \\ &\to \left(\mathrm{E}\left[\varphi(t)\varphi(t)^T\right]\right)^{-1}\mathrm{E}\left[\varphi(t)y(t)\right], N \to \infty, \end{aligned} \tag{35}$$

where $\mathrm{E}[\cdot]$ denotes the statistical expectation, N is the number of measurements. The LS estimate can be viewed as a special case of weighted LS estimates [17],

$$\hat{\theta}_{WLS} = (\Phi^T W \Phi)^{-1}\Phi^T W y \tag{36}$$

where $\Phi^T = [\varphi(1), \cdots, \varphi(N)]$, $y = [y(1), \cdots, y(N)]^T$, the weighting matrix W is an identity matrix in the LS estimate. Assume $\tilde{\varphi}(t)$ and $\tilde{y}(t)$ are stochastic variables with zero mean,

$$\mathrm{E}\left[\varphi(t)\varphi(t)^T\right] = \varphi_0(t)\varphi_0(t)^T + \mathrm{E}\left[\tilde{\varphi}(t)\tilde{\varphi}(t)^T\right], \tag{37}$$

$$\mathrm{E}\left[\varphi(t)y(t)^T\right] = \varphi_0(t)y_0(t) + \mathrm{E}\left[\tilde{\varphi}(t)\tilde{y}(t)^T\right], \tag{38}$$

where the covariance $E\left[\tilde{\varphi}(t)\tilde{\varphi}(t)^T\right] \neq 0$ and $E\left[\tilde{\varphi}(t)\tilde{y}(t)\right] \neq 0$ when $\tilde{\varphi}(t)$ and $\tilde{y}(t)$ are correlated. Thus, the LS estimate is biased when applied to EIV problems. Bias-compensated least-squares (BCLS) adjust this estimate for the bias in the following way,

$$\begin{aligned} \hat{\theta}_{BCLS} &= \left(\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi(t)^T - \frac{1}{N}\sum_{t=1}^{N}\tilde{\varphi}(t)\tilde{\varphi}(t)^T\right)^{-1} \\ &\times \left(\frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t) - \frac{1}{N}\sum_{t=1}^{N}\tilde{\varphi}(t)\tilde{y}(t)\right). \end{aligned} \tag{39}$$

Alternatives to the bias estimates $\frac{1}{N}\sum_{t=1}^{N}\tilde{\varphi}(t)\tilde{y}(t)$ and $\frac{1}{N}\sum_{t=1}^{N}\tilde{\varphi}(t)\tilde{\varphi}(t)^T$ have been proposed, e.g. Koopmans-Levin [21], bias-eliminating least-squares (BELS) [22], [23]. A thorough comparison of these methods can be found in [11]. Though the estimation of the bias terms is important, it is not essential for this paper and we will stick to the straightforward sample means above.

REFERENCES

[1] A. Grosfils, A. V. Wouwer, and P. Bogaerts, “On a general model structure for macroscopic biological reaction rates,” *Journal of biotechnology*, vol. 130, no. 3, pp. 253–264, 2007.

[2] E. Hagrot, H. Æ. Oddsdóttir, J. G. Hosta, E. W. Jacobsen, and V. Chotteau, “Poly-pathway model, a novel approach to simulate multiple metabolic states by reaction network-based model–application to amino acid depletion in cho cell culture,” *Journal of biotechnology*, vol. 259, pp. 235–247, 2017.

[3] G. E. Box and W. G. Hunter, “The experimental study of physical mechanisms,” *Technometrics*, vol. 7, no. 1, pp. 23–42, 1965.

[4] Q. Zhu, “An implicit least squares algorithm for nonlinear rational model parameter estimation,” *Applied mathematical modelling*, vol. 29, no. 7, pp. 673–689, 2005.

[5] D. W. Marquardt, “An algorithm for least-squares estimation of nonlinear parameters,” *Journal of the society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[6] T. L. Toulias and C. P. Kitsos, “Fitting the michaelis–menten model,” *Journal of Computational and Applied Mathematics*, vol. 296, pp. 303–319, 2016.

[7] S. Billings and S. Chen, “Identification of non-linear rational systems using a prediction-error estimation algorithm,” *International Journal of Systems Science*, vol. 20, no. 3, pp. 467–494, 1989.

[8] R. F. Heiser and W. R. Parrish, “Representing physical data with rational functions,” *Industrial & engineering chemistry research*, vol. 28, no. 4, pp. 484–489, 1989.

[9] S. Dimitrov and D. Kamenski, “A parameter estimation method for rational functions,” *Computers & chemical engineering*, vol. 15, no. 9, pp. 657–662, 1991.

[10] Q. Zhu, “A back propagation algorithm to estimate the parameters of non-linear dynamic rational models,” *Applied Mathematical Modelling*, vol. 27, no. 3, pp. 169–187, 2003.

[11] T. Söderström, “Errors-in-variables methods in system identification,” *Automatica*, vol. 43, no. 6, pp. 939–958, 2007.

[12] P. Stoica and T. Söderström, “Bias correction in least-squares identification,” *International Journal of Control*, vol. 35, no. 3, pp. 449–457, 1982.

[13] S. Billings and Q. Zhu, “Rational model identification using an extended least-squares algorithm,” *International Journal of Control*, vol. 54, no. 3, pp. 529–546, 1991.

[14] B. Mu, E.-W. Bai, W. X. Zheng, and Q. Zhu, “A globally consistent nonlinear least squares estimator for identification of nonlinear rational systems,” *Automatica*, vol. 77, pp. 322–335, 2017.

[15] J. F. Andrews, “A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates,” *Biotechnology and Bioengineering*, vol. 10, no. 6, pp. 707–723, 1968.

[16] T. Doeswijk and K. Keesman, “Linear parameter estimation of rational biokinetic functions,” *water research*, vol. 43, no. 1, pp. 107–116, 2009.

[17] T. Söderström and P. Stoica, *System identification*. Prentice Hall, 1989.

[18] T. Söderström, P. Stoica, and B. Friedlander, “An indirect prediction error method for system identification,” *Automatica*, vol. 27, no. 1, pp. 183–188, 1991.

[19] D. M. Bates and D. G. Watts, “Nonlinear regression analysis and its applications,” *John Wiles & Sons, Inc*, 2007.

[20] H. Leung and S. Haykin, “Rational function neural network,” *Neural Computation*, vol. 5, no. 6, pp. 928–938, 1993.

[21] K. V. Fernando and H. Nicholson, “Identification of linear systems with input and output noise: the koopmans-levin method,” in *IEE Proceedings D (Control Theory and Applications)*, vol. 132, no. 1. IET, 1985, pp. 30–36.

[22] W. X. Zheng, “A bias correction method for identification of linear dynamic errors-in-variables models,” *IEEE transactions on automatic control*, vol. 47, no. 7, pp. 1142–1147, 2002.

[23] M. Hong, T. Soderstrom, and W. X. Zheng, “A simplified form of the bias-eliminating least squares method for errors-in-variables identification,” *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1754–1756, 2007.