Digital Signal Processing 114 (2021) 103071

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Degree and noise power estimation from noisy polynomial data via AR modelling

Asoke K. Nandi [a,b,*]

[a] Electronic and Electrical Engineering, Brunel University London, Uxbridge, UB8 3PH, UK
[b] School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

ARTICLE INFO

Article history:
Available online 19 April 2021

Keywords:
Data modelling
Noisy polynomial data
Polynomial degree estimation
Noise variance
Time-series representations of polynomials
Regression

ABSTRACT

An accurate estimation of the noise power from noisy data leads to better estimation of signal-to-noise ratio (SNR) and is useful in detection, estimation, and prediction. The major contributions of this paper are to estimate the polynomial degree and the noise power from data coming from an underlying polynomial with additive Gaussian noise, using an AR model. The two proposed methods have been inspired by the recent results that all finite degree polynomials have equivalent representation in finite order autoregressive (AR) models, with known AR coefficients and different constant terms. Preliminary experiments in a variety of scenarios provide estimations of the constant term and the standard deviation of these estimations, which are then used as a guide to developing theoretically the probability density functions. In the first stage, the degree of a polynomial is selected by minimizing the variance of the estimations of the constant term in the equivalent AR model. In the second stage, the noise variance is estimated using the estimated degree of a polynomial, a combination of the variance of the estimations of the constant term, and another known parameter. Further computer experiments have been carried out for evaluating the proposed methods for degree and noise power estimations. Four well-known and well-regarded maximum likelihood-based approaches have been used for comparisons.

1. Introduction

Estimation of noise power from noisy data is useful in detection, estimation, and prediction [1]. This paper addresses the class of data coming from an underlying polynomial with additive, zero-mean, independent and identically distributed Gaussian noise. Noise variance can be estimated from the differences between the available noisy data and reconstructed noise-free data when polynomial coefficients are available. Polynomial coefficients can be estimated from data by fitting polynomial regression models with the Least-Squares method. In 1805 Legendre published the Least-Squares method [2] and Gauss published it in 1809 [3] and later in 1823 [4]. In 1815 Gergonne wrote a paper on "The application of the method of least squares to the interpolation of sequences" [5] and its English translation by St. John and Stigler [6]. In the last 120 years, polynomial regressions contributed much to the development of regression analysis [7–9], which have many diverse applications, including an interesting one in polymerase chain reaction bias correction in quantitative DNA methylation studies [10].

The estimation of the degree of the polynomial can be achieved using many model order selection techniques [11,12]. In this paper comparisons of the proposed methods are carried out with four chosen model order selection techniques that have been developed around the maximum likelihood method, namely Akaike Information Criterion (AIC), corrected Akaike Information Criterion (AICc), Generalized Information Criterion (GIC), and Bayesian Information Criterion (BIC) [13].

The proposed methods to estimate the polynomial degree and the noise power from data coming from an underlying polynomial with additive Gaussian noise are fundamentally different to what have been practised in the past. The proposed methods have been inspired by the very recent results that all polynomials of degree q give rise to the same set of q known time-series coefficients of autoregressive models and an additional constant term μ. This paper extends the novel and interesting approaches initiated in [26] and [29]. The study in [26] is very different from this paper, in that it neither explored noisy data nor considered polynomial degree estimation. The study in [29], amongst other things, proposed one polynomial degree estimator. But it is very different from this paper, as it did not develop theoretically the probability density functions of estimations of the constant term μ, did not propose the degree estimator in this paper, and did not consider any noise

* Correspondence to: Electronic and Electrical Engineering, Brunel University London, Uxbridge, UB8 3PH, UK.
E-mail address: asoke.nandi@brunel.ac.uk

https://doi.org/10.1016/j.dsp.2021.103071

variance estimator. This is primarily a novel technique paper with a solid theoretical foundation and much success with simulated data, using no polynomial coefficients.

Section 2 offers some brief theoretical background, including introduction to four chosen model order selection criteria based on the maximum likelihood. Section 3 offers some experimental results in a variety of scenarios and provides estimations of the constant term and the standard deviation of these estimations. These help to evince patterns that were a guide to developing the probability density functions of the estimations in section 4. Based on these, section 5 proposes a new polynomial degree estimator, while section 6 offers a novel way to estimate the noise variance. Section 7 contains discussions and Section 8 records conclusions.

2. Theoretical background

To make the paper self-contained we introduce briefly some relevant existing knowledge. Section 2.1 offers brief outlines of the maximum likelihood method and the four chosen information criteria which are used in this paper to compare the results from the proposed estimators.

A polynomial of degree q in continuous time can be written as $y(t) = \sum_{i=0}^{q} c(i)\ t^i$. For uniformly sampled discrete time, the continuous time, t, is represented as $t = nT$, where n is an integer and T is the sampling period. In this scenario, the above equation can be rewritten as $y(nT) = \sum_{i=0}^{q} c(i)\ (nT)^i$. For the sake of simplicity in notations and without the loss of any generalisations, this can be written as

$$y(n) = \sum_{i=0}^{q} c(i)\, n^i \tag{1}$$

where the new value of $c(i)$ is the old value of $c(i)$ multiplied by T^i. Thus, a set of real-valued noisy data from polynomials in uniformly sampled discrete time, can be represented by

$$x(n) = \sum_{i=0}^{q} c(i)\, n^i + e(n), \tag{2}$$

where $e(n)$ represents noise.

2.1. Maximum log-likelihood

Maximum likelihood-based techniques use the degree of the polynomial and values of its coefficients to estimate noise variance. The log-likelihood function for N independent and identically distributed samples from a Gaussian distribution is given by

$$lnL(\sigma, y(.)) = -\left(\frac{N}{2}\right)\ln(\pi) - \left(\frac{N}{2}\right)\ln\left(\sigma^2\right) - \left(\frac{1}{2\sigma^2}\right)\sum_{i=1}^{N}(x(i) - y(i))^2 \tag{3}$$

For a polynomial of degree 1, i.e., $q = 1$, $y(i) = c(0) + c(1)\,i$. Thus, $\partial lnL/\partial c(0) = \sum_{i=1}^{N}(x(i) - c(1)\,i - c(0))/\sigma^2$, $\partial lnL/\partial c(1) = \sum_{i=1}^{N}(x(i) - c(1)\,i - c(0))i/\sigma^2$, and $\partial lnL/\partial(\sigma^2) = -(N/2)\,1/\sigma^2 + (1/2)\,1/(\sigma^4)\sum_{i=1}^{N}(x(i) - y(i))^2$. Maximising this log-likelihood function one obtains the following three equations:

$$\sum_{i=1}^{N}(x(i) - c(1)\,i - c(0)) = 0 \tag{4}$$

$$\sum_{i=1}^{N}(x(i) - c(1)\,i - c(0))\,i = 0 \tag{5}$$

$$\sum_{i=1}^{N}(x(i) - y(i))^2 = N\sigma^2 \tag{6}$$

Using equations (4) and (5), one obtains the ordinary least squares estimates of $c(0)$ and $c(1)$. If one sets up two matrices, $X^T = [x(1)\ \ x(2)\dots x(N)]$ and $A^T = [1\ 2\ \dots\ N;\ 1\ 1\ \dots\ 1]$, then $[c(1)\ \ c(0)]^{\mathrm{T}} = \left(A^T A\right)^{-1} A^T X$. It should be remarked that one cannot calculate σ^2 from equation (6) in the absence of the knowledge of the noise-free data, $y(i)$. Instead one can estimate σ^2 using the data values and fitted values, i.e., $\hat{y}(n) = c(1)\,n + c(0)$,

$$\hat{\sigma^2} = \frac{\left(\sum_{i=1}^{N}\left(x(i) - \hat{y}(i)\right)^2\right)}{N} \tag{7}$$

Hence, for each chosen value of the degree of a polynomial (q), one can estimate the corresponding coefficients of the polynomial, $[c(0)\ \ c(1)\dots c(q)]$, and estimate the corresponding value of σ^2. Thus, for a given dataset, M different choices for the value of q will produce M different estimates of σ^2. The challenge is to decide the appropriate value of q. To make this decision four commonly used and well-regarded model order selection techniques, i.e., AIC, AICc, GIC, and BIC, are chosen and briefly described below.

2.1.1. Akaike Information Criterion (AIC)

Given a set of models, AIC [12–18] aims to select the best model from this set. Thus, the selected model is not guaranteed to be the best model as it represents a relative choice within the set of given models. AIC tries to balance between the risk of overfitting and the risk of underfitting, so it is a compromise between the best fitted model and the simplicity of the model.

AIC uses the log-likelihood to provide a measure of the goodness of fit. Suppose that there is a statistical model of the data, that q is the number of estimated parameters, and that $\mathcal{L}$ is the maximum value of the likelihood function for the model. AIC is defined as

$$AIC(q) = 2q - 2\ln(\mathcal{L}) \tag{8}$$

The selected model will correspond to the one for which AIC is the minimum. The first term in equation (8) attempts to keep the polynomial degree small while the second term attempts to obtain the maximum value of the likelihood.

2.1.2. Corrected Akaike Information Criterion (AICc)

As the number of data tends to ∞, AIC has certain desirable properties. However, whenever the number of data (N) is small, there is a significant chance that AIC will choose models with too many parameters, i.e., AIC will overfit. In fact, equation (4) does not have any dependence on the number of data values. To address this, AICc was introduced by Sugiura [19]. Since then, many researchers [12,16,17,20] have extended the applicability of AICc, which can be defined as

$$AICc(q) = \frac{2qN}{N - q - 1} - 2\ln(\mathcal{L}) \tag{9}$$

AICc clearly depends, amongst other factors, on the number of data (N). The procedure for selecting the best model from a given set of models, i.e., the degree of a polynomial (q) and the corresponding polynomial coefficients, $[c(0)\ \ c(1)\dots c(q)]$, requires the minimisation of AICc.

2.1.3. Generalised Information Criterion (GIC)

In AIC, the factor of $2q$ has been designed to address the issue of overfitting. Intuitively, the probability of overfitting will be reduced as the number of data increases. In finite sample situations,

Table 1
Estimates of μ using data from $x(n) = n + 1 + \mathbb{N}(0, \sigma)$ for six choices of a polynomial degree.

Standard deviation of noise, σ	$< \mu >$ value if degree = 1	$< \mu >$ value if degree = 2	$< \mu >$ value if degree = 3	$< \mu >$ value if degree = 4	$< \mu >$ value if degree = 5	$< \mu >$ value if degree = 6
1	1.00	0.00	0.00	-0.01	0.00	-0.03
2	1.00	0.00	0.00	0.00	0.01	0.02
3	1.00	0.00	0.01	-0.01	0.04	0.00
4	1.00	0.00	0.01	0.02	0.02	0.05
5	1.01	0.00	0.00	-0.02	-0.01	-0.12

Table 2
Estimates of $\sigma_{\mu(q)}$ using data from $x(n) = n + 1 + \mathbb{N}(0, \sigma)$ for six choices of a polynomial degree.

Standard deviation of noise, σ	$< \sigma_{\mu(q)} >$ if degree = 1	$< \sigma_{\mu(q)} >$ if degree = 2	$< \sigma_{\mu(q)} >$ if degree = 3	$< \sigma_{\mu(q)} >$ if degree = 4	$< \sigma_{\mu(q)} >$ if degree = 5	$< \sigma_{\mu(q)} >$ if degree = 6
1	1.41	2.45	4.46	8.34	15.8	30.3
2	2.84	4.91	8.95	16.7	31.7	60.5
3	4.26	7.37	13.4	25.1	47.5	90.9
4	5.70	9.87	18.0	33.6	63.7	121.7
5	7.08	12.2	22.3	41.7	79.1	151.3

extensive simulation studies have demonstrated that the following generalised information criterion (GIC) [21]

$$GIC(q) = \alpha q - 2\ln(\mathcal{L}) \tag{10}$$

can outperform AIC if $\alpha > 2$. Values of α in the range of 2 to 6 appear to offer the best performance. The optimal value for α depends on many factors but there is no clear hint on how to choose its value in a specific scenario. Note that $\alpha = 2$ corresponds to AIC. In the following investigations the value for α has been set to 4. Note that GIC does not explicitly depend on the number of data values, unlike AICc.

2.1.4. Bayesian Information Criterion (BIC)

The form of BIC [22–25,12,16] is very similar to AIC, in that they both have two terms – a negative log-likelihood one and a penalty term for the number of parameters, although their origins are different. The log-likelihood term is identical in both cases. The penalty term is $2q$ in AIC, while it is $\ln(N)(q)$ in BIC. Note that AIC does not depend on the number of data (N), but BIC does include a dependence on N. In that sense, BIC captures something of AIC and AICc, and it can be written as

$$BIC(q) = [\ln(N)]q - 2\ln(\mathcal{L}) \tag{11}$$

3. Experiments for $< \mu >$ and $< \sigma_{\mu(q)} >$

All noise-free data from uniformly sampled polynomials of finite degree q can be perfectly represented by an autoregressive time-series model of order q and a constant [26], such that

$$y(n) = \sum_{i=1}^{q} a(i)\, y(n-i) + \mu \tag{12}$$

where

$$a(i) = (-1)^{i+1}\binom{q}{i} \tag{13}$$

for $i = 1, 2, \ldots, q$, and

$$\mu = c(q)(q!) \tag{14}$$

Here is a hint to why equation (12) and (14) are true. Given a polynomial of degree q, all differentials of order higher than $(q + 1)$ are zero. The differential of order $(q + 1)$ is a constant and is equal to $c(q)\ (q!)$, which is represented in equation (14). In uniformly discrete domain (time or space or anything else), this polynomial can be represented by a q-th order difference equation with this constant term, which is essentially the equation (12). Equation (12) can be rewritten as

$$-\sum_{i=1}^{q} a(i)\, y(n-i) + y(n) = \mu \tag{15}$$

3.1. Experiments with noisy data

As noise-free $y(n)$ values are not available, this equation can be recast with known noisy data values $x(n)$ as follows

$$-\sum_{i=1}^{q} a(i)\, x(n-i) + x(n) = \mu(n, q) \tag{16}$$

where $\mu(n, q)$ may depend on both n and q. Equation (16) can be written in matrix form as XA = M, where X is a $(f-q)$x$(q+1)$ matrix and X = $[(x(1)\ldots x(f-q))^T;\ (x(2)\ldots x(f-q+1))^T;\ \ldots(x(q+1)\ldots x(f))^T]$, A is a $(q+1)$x1 matrix and A = $[(-a(1)\ldots - a(q)1)^T]$, M is a $(f-q)$x1 matrix and M = $[\mu(q+1, q)\ldots \mu(f, q)]^T]$, as well as f is the number of data values being used for estimation.

All the entries in matrix X are known as they represent the noisy data values. Also, all the entries in matrix A are known from equation (13). Therefore, M can be obtained from XA, containing $(f-q)$ values. For a chosen value of q, these $(f-q)$ values of $\mu(n, q)$ are estimates of the same constant term, $\mu(q)$, for a polynomial of degree q. From these $(f-q)$ values, one can estimate the mean value, $\mu(q)$, and the root-mean square value, $\sigma_{\mu(q)}$, where

$$\mu(q) = \frac{\sum_{n=1}^{q} \mu(n, q)}{(f-q)} \tag{17}$$

and

$$\sigma_{\mu(q)} = \sqrt{\left[\frac{\left(\sum_{n=q+1}^{f} (\mu(n, q) - \mu(q))^2\right)}{(f-q)}\right]} \tag{18}$$

Thus, for every value of q, there are two parameters - the mean value, $\mu(q)$, and the root-mean square value, $\sigma_{\mu(q)}$.

Table 3
Estimates of μ using data from $x(n) = n^2 + n + 1 + \mathbb{N}(0, \sigma)$ for six choices of a polynomial degree.

Standard deviation of noise, σ	$<\mu>$ value if degree = 1	$<\mu>$ value if degree = 2	$<\mu>$ value if degree = 3	$<\mu>$ value if degree = 4	$<\mu>$ value if degree = 5	$<\mu>$ value if degree = 6
1	-40.0	2.00	0.00	-0.01	0.00	-0.02
2	-40.0	2.00	0.00	0.00	0.00	0.03
3	-40.0	2.00	0.00	0.02	0.03	0.01
4	-40.0	2.00	0.00	-0.02	0.01	-0.06
5	-40.0	2.00	0.01	0.01	0.04	-0.02

Table 4
Estimates of $\sigma_{\mu(q)}$ using data from $x(n) = n^2 + n + 1 + \mathbb{N}(0, \sigma)$ for six choices of a polynomial degree.

Standard deviation of noise, σ	$<\sigma_{\mu(q)}>$ if degree = 1	$<\sigma_{\mu(q)}>$ if degree = 2	$<\sigma_{\mu(q)}>$ if degree = 3	$<\sigma_{\mu(q)}>$ if degree = 4	$<\sigma_{\mu(q)}>$ if degree = 5	$<\sigma_{\mu(q)}>$ if degree = 6
1	34.4	2.44	4.44	8.30	15.7	30.1
2	34.5	4.90	8.93	16.7	31.6	60.5
3	34.6	7.35	13.4	25.0	47.3	90.4
4	34.8	9.85	18.0	33.6	63.6	121.5
5	35.1	12.3	22.4	41.8	79.2	151.3

In this novel conceptual framework some experiments are carried out with generated data from polynomials of different degrees and additive zero-mean Gaussian noise to help evince patterns from particular cases through a pedagogical development.

First a polynomial of degree 1 has been considered for data generation: $y(n) = n + 1$. For this experiment, 1,000 sets of 101 data values have been generated for each value of the standard deviation (σ) of noise using the zero-mean Gaussian distribution, $\mathbb{N}(0, \sigma)$. Thus, the generated noisy data can be described by

$$\begin{aligned} &x(n, \sigma) = n + 1 + \mathbb{N}(0, \sigma), \\ &\text{for } n = -50 : 1 : 50 \quad \text{and} \quad \sigma = 1 : 1 : 5 \end{aligned} \tag{19}$$

This notation indicates that all integer values of n from -50 to 50 and of σ from 1 to 5 are used. In general, the correct degree of the polynomial will be unknown. So, for each set of f data values, different degrees of polynomial values are chosen, i.e., degree of 1, 2, 3, 4, 5, and 6. In each set of 101 data values, the first 60 data values, i.e., $f = 60$, have been used for estimating $\mu(q)$ and $\sigma_{\mu(q)}$. Thus, in this experiment there are 1000 pairs of these values. Three parameters are estimated from these 1000 pairs – the expected value of μ, denoted as $<\mu>$, the expected value of the standard deviation, $\sigma_{\mu(q)}$, denoted as $<\sigma_{\mu(q)}>$, and the standard deviation of $\sigma_{\mu(q)}$, denoted by $\sigma(\sigma_{\mu(q)})$. These are estimated in the following ways:

$$<\mu> = \frac{\sum_{i=1}^{1000} (\mu(q))_i}{1000} \tag{20}$$

$$<\sigma_{\mu(q)}> = \frac{\sum_{i=1}^{1000} (\sigma_{\mu(q)})_i}{1000} \tag{21}$$

$$\sigma\left(\sigma_{\mu(q)}\right) = \sqrt{\left[\frac{\left(\sum_{i=1}^{1000} \left((\sigma_{\mu(q)})_i - <\sigma_{\mu(q)}>\right)^2\right)}{1000}\right]} \tag{22}$$

Table 1 and Table 2 record $<\mu>$ and $<\sigma_{\mu(q)}>$, respectively, for the linear polynomial. Similar experiments have been carried out with quadratic, cubic and quartic polynomials. The corresponding results are presented in Tables 3 to 8.

3.2. Summary

Computer experiments have recorded $<\mu>$, $<\sigma_{\mu(q)}>$, and $\sigma\left(\sigma_{\mu(q)}\right)$ for different degrees of polynomials and noise powers. Experimental results from section 3.1 can be summarized and generalized as follows:

a) The $<\mu>$ values within each column appear to be the same, i.e., they appear to be independent of the noise standard deviations.
b) When the chosen degree is the same as the correct one, the $<\mu>$ values appear to be $q!$, where q is the correct degree of the polynomial.
c) When the chosen degree is larger than the correct one, the $<\mu>$ values appear to be 0.
d) When the chosen degree is smaller than the correct one, the $<\mu>$ values appear to be different from $q!$ and 0.
e) Whenever the chosen degree is the same as or larger than the correct degree, the $<\sigma_{\mu(q)}>$ values within each column appear to increase linearly, i.e., they are linearly dependent on the noise standard deviations.
f) Whenever the chosen degree is smaller than the correct one, the $<\sigma_{\mu(q)}>$ values within each column appear to be the same, i.e., they appear to be independent of the noise standard deviations.
g) For each value of σ, as the chosen degree is increased from the correct one, the $<\sigma_{\mu(q)}>$ values increase non-linearly.

4. Analysis of $<\mu>$ and $<\sigma_{\mu(q)}>$

In this section analysis are presented to explain all the aforementioned observations in three different scenarios – 1) when the chosen degree is the same as the correct degree, 2) when the chosen degree is larger than the correct degree, and 3) when the chosen degree is smaller than the correct degree. In the final subsection the analysis is summarized and verified.

4.1. Chosen degree is the same as the correct degree

In [26], it has been proven that, for the correct degree, $\mu = c(q)\ (q!)$, for noise-free data. Taking the expectation of equation (16), one can write

$$\begin{aligned} <\mu> \;=\; <\mu(q)> &= \ll -\sum_{i=1}^{q} a(i)\, x(n-i) + x(n) \gg \\ &= -\sum_{i=1}^{q} a(i)\, y(n-i) + y(n) = c(q)\,(q!) \end{aligned} \tag{23}$$

Table 5
Estimates of μ using data from $x(n) = n^3 + n^2 + n + 1 + \mathbb{N}(0, \sigma)$ for six choices of a polynomial degree.

Standard deviation of noise, σ	$<\mu>$ value if degree = 1	$<\mu>$ value if degree = 2	$<\mu>$ value if degree = 3	$<\mu>$ value if degree = 4	$<\mu>$ value if degree = 5	$<\mu>$ value if degree = 6
1	2091	-121	6.00	0.00	0.00	-0.01
2	2091	-121	5.99	0.00	-0.01	0.03
3	2091	-121	6.01	-0.02	0.03	-0.04
4	2091	-121	6.01	0.02	0.02	0.03
5	2091	-121	6.01	0.00	0.05	-0.03

Table 6
Estimates of $\sigma_{\mu(q)}$ using data from $x(n) = n^3 + n^2 + n + 1 + \mathbb{N}(0, \sigma)$ for six choices of a polynomial degree.

Standard deviation of noise, σ	$<\sigma_{\mu(q)}>$ if degree = 1	$<\sigma_{\mu(q)}>$ if degree = 2	$<\sigma_{\mu(q)}>$ if degree = 3	$<\sigma_{\mu(q)}>$ if degree = 4	$<\sigma_{\mu(q)}>$ if degree = 5	$<\sigma_{\mu(q)}>$ if degree = 6
1	2221	101	4.45	8.32	15.8	30.2
2	2221	101	8.95	16.7	31.7	60.5
3	2221	102	13.3	24.9	47.1	90.1
4	2221	102	17.7	33.1	63.7	120.0
5	2221	102	22.4	41.8	79.3	151.7

Table 7
Estimates of μ using data from $x(n) = n^4 + n^3 + n^2 + n + 1 + \mathbb{N}(0, \sigma)$ for six choices of a polynomial degree.

Standard deviation of noise, σ	$<\mu>$ value if degree = 1	$<\mu>$ value if degree = 2	$<\mu>$ value if degree = 3	$<\mu>$ value if degree = 4	$<\mu>$ value if degree = 5	$<\mu>$ value if degree = 6
1	-103730	8287	-486	24.0	-0.01	0.00
2	-103730	8287	-486	24.0	0.02	0.02
3	-103730	8287	-486	24.0	-0.01	0.01
4	-103730	8287	-486	24.0	0.01	0.01
5	-103730	8287	-486	24.0	0.04	-0.13

Table 8
Estimates of $\sigma_{\mu(q)}$ using data from $x(n) = n^4 + n^3 + n^2 + n + 1 + \mathbb{N}(0, \sigma)$ for six choices of a polynomial degree.

Standard deviation of noise, σ	$<\sigma_{\mu(q)}>$ if degree = 1	$<\sigma_{\mu(q)}>$ if degree = 2	$<\sigma_{\mu(q)}>$ if degree = 3	$<\sigma_{\mu(q)}>$ if degree = 4	$<\sigma_{\mu(q)}>$ if degree = 5	$<\sigma_{\mu(q)}>$ if degree = 6
1	137037	8750	398	8.31	15.7	30.1
2	137037	8750	398	16.6	31.5	60.4
3	137037	8750	398	24.9	47.3	90.4
4	137037	8750	399	33.5	63.4	121.4
5	137037	8750	399	41.7	79.0	151.1

Moreover, the fact that $c(q) = 1$ in these experiments explains the above observation (b) in section 3.2, i.e., $<\mu> = q!$. Thus, the column 2 of Table 1, the column 3 of Table 3, the column 4 of Table 5, and the column 5 of Table 7 are explained.

Given that $\mu(n, q) = x(n) - \sum_{i=1}^{q} a(i)\ x(n-i)$, variations in $\mu(n, q)$ values will come from a combination of incorrect structure, i.e., incorrect choice of the polynomial degree, and variations in $x(n)$. In the current scenario of having chosen the correct degree, variations will come from variations in $x(n)$ only. Therefore, the variance is given by

$$<\sigma_{\mu(q)}>^2 = \left((1)^2 + \sum_{i=1}^{q} a(i)^2\right)\sigma^2 \tag{24}$$

Combining equations (13) and (16), one obtains $<\sigma_{\mu(q)}>^2$ values of 2, 6, 20, 70, 252, and 924 for polynomial degree (q) values of 1, 2, 3, 4, 5, and 6 respectively. Thus, the column 2 of Table 2, the column 3 of Table 4, the column 4 of Table 6, and the column 5 of Table 8 are all explained.

4.2. Chosen degree is larger than the correct degree

First, consider the case of a linear polynomial, i.e. the true degree is 1, such that $y(n) = c(0) + c(1)n$. When the chosen degree is 2, $\mu(n, 2) = x(n) - 2x(n-1) + x(n-2)$. Therefore, $<\mu(n, 2)> = <x(n)> - 2<x(n-1)> + <x(n-2)> = y(n) - 2y(n-1) + y(n-2) = [c(0) + c(1)n] - 2[c(0) + c(1)(n-1)] + [c(0) + c(1)(n-2)] = 0$. There is another way to appreciate this result. Essentially, $[x(n) - 2x(n-1) + x(n-2)]$ represents d^2x/dt^2. Given three consecutive data values, one can write $dx/dt = (x(n) - x(n-1))/T$ and $dx/dt = (x(n-1) - x(n-2))/T$. These lead to $d^2x/dt^2 = [x(n) - 2x(n-1) + x(n-2)]/T^2$. The second differential of every linear polynomial is zero.

In fact, for each value of q, the corresponding set of $a(i)$ coefficients, $[1, a(1), \ldots, a(q)]$, can represent the q-th differential of the polynomial. This is precisely why $\mu = c(q)\ (q!)$, which is the q-th differential of a polynomial of true degree q. Thus, for a linear polynomial, $<\mu(n, q)> = 0$, whenever $q > 1$. More generally, of course, $<\mu(n, q)> = 0$, when q is larger than the correct degree. Therefore, the columns 3, 4, 5, 6, and 7 of Table 1, the columns 4, 5, 6, and 7 of Table 3, the columns 5, 6, and 7 of Table 5, and the columns 6 and 7 of Table 7 are all explained. Whenever the chosen degree is the same as or larger than the correct degree, $<\mu(n,q)>$ is equal to the q-th differential of a polynomial.

Given that $\mu(n,q)=x(n)-\sum_{i=1}^{q}a(i)\ x(n-i)$, for the choice of a degree larger than the correct degree of the polynomial, variations will come from statistical variations in $x(n)$. Therefore, the variance is given by $<\sigma_{\mu(q)}>^2=\left((1)^2+\sum_{i=1}^{q}a(i)^2\right)\sigma^2$ as in equation (24). Combining equations (13) and (24), one obtains $<\sigma_{\mu(q)}>^2$ values of 2, 6, 20, 70, 252, and 924 for polynomial degree (q) values of 1, 2, 3, 4, 5, and 6 respectively. Thus, the columns 3, 4, 5, 6, and 7 of Table 2, the columns 4, 5, 6, and 7 of Table 4, the columns 5, 6, and 7 of Table 6, and the columns 6 and 7 of Table 8 are explained. Whenever the chosen degree is the same as or larger than the correct degree, $<\sigma_{\mu(q)}>^2$ comes from equation (24).

4.3. Chosen degree is smaller than the correct degree

Choosing a degree q is equivalent to assuming the underlying noise-free time-series model to be of order q and

$$\begin{aligned}\mu(n,q)&=x(n)-\sum_{i=1}^{q}a(i)\,x(n-i)\\&=y(n)-\sum_{i=1}^{q}a(i)\,y(n-i)+e(n)-\sum_{i=1}^{q}a(i)\,e(n-i)\end{aligned}\tag{25}$$

where $e(n)$ are noise, while the true polynomial, assuming the correct degree is D, can be written as $y(n)=\sum_{j=0}^{D}c(j)\ n^j$. Here $q<D$, and

$$\begin{aligned}\sum_{i=1}^{q}a(i)\,y(n-i)&=\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}y(n-i)\\&=\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}\sum_{j=0}^{D}c(j)\,(n-i)^j\end{aligned}\tag{26}$$

Below each of the j values is considered separately.

For $j=0$, the right-hand side of equation (26) can be written as RHS $(j=0)=c(0)\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}=c(0)$, adapting the identity 0.151.4 on page 3 in [27]. For $j=1$, the right-hand side of equation (26) can be written as RHS $(j=1)=nc(1)\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}-c(1)\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}i=c(0)$. It can be shown that the first term is $nc(1)$, adapting the identity 0.151.4 on page 3 in [27], and that the second term is 0, adapting the identity 0.154.6 on page 4 in [27]. Hence, RHS $(j=1)=nc(1)$. Using the same two identities, it can be shown, for up to and including $j=(q-1)$, that RHS $(j)=n^jc(j)$.

For $j=q$, the right-hand side of equation (26) can be written as

$$\begin{aligned}RHS(j=q)&=\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}c(q)\,(n-i)^q\\&=c(q)\,n^q\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}\\&\quad+c(q)\,n^{q-1}\binom{q}{1}\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}(-i)+\ldots\\&\quad+c(q)\,n^{q-q}\binom{q}{q}\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}(-i)^q\end{aligned}$$

The first term is $n^qc(q)$, adapting the identity 0.151.4 on page 3 in [27], and all other terms up to and including the penultimate one containing $(-i)^{q-1}$ are 0, adapting the identity 0.154.6 on page 4 in [27]. The last term is

$$\begin{aligned}c(q)\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}(-i)^q&=c(q)\,(-1)\sum_{i=1}^{q}(-1)^{i}\binom{q}{i}(-i)^q\\&=c(q)\,(-1)\,(-1)^q\,(-1)^q\,(q!)\\&=-(q!)\,c(q)\end{aligned}$$

adapting the identity 0.154.4 on page 4 in [27]. Hence, RHS $(q)=n^qc(q)-(q!)\,c(q)$.

Now the case of $j=(q+1)$ is considered. In this scenario, the right-hand side of equation (26) can be written as

$$\begin{aligned}&RHS\,(j=q+1)\\&=\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}c\,(q+1)\,(n-i)^{q+1}\\&=c\,(q+1)\,n^{q+1}\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}\\&\quad+c\,(q+1)\,n^{q}\binom{q+1}{1}\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}(-i)+\ldots\\&\quad+c\,(q+1)\,n^{q+1-q}\binom{q+1}{q}\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}(-i)^q\\&\quad+c\,(q+1)\,n^{0}\binom{q+1}{q+1}\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}(-i)^{q+1}\end{aligned}$$

The first term is $n^{q+1}c(q+1)$, adapting the identity 0.151.4 on page 3 in [27], and all other terms up to and including the one containing $(-i)^{q-1}$ are 0, adapting the identity 0.154.6 on page 4 in [27]. The penultimate term is $c\,(q+1)\,n\,(q+1)\,(-1)\,(-1)^q\,(-1)^q\,(q!)=-((q+1)!)\,nc(q+1)$. The last term is $c\,(q+1)\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}(-i)^{q+1}$. Hence, $RHS(q+1)=n^{q+1}c\,(q+1)-((q+1)!)\,nc\,(q+1)+c\,(q+1)\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}(-i)^{q+1}$

Putting all the components of the RHS of equation (26) together, one can write, $RHS=\sum_{i=1}^{q+1}RHS\,(i)=\sum_{j=1}^{q+1}c\,(j)\,n^j-(q!)\,c\,(q)-((q+1)!)\,nc\,(q+1)+c\,(q+1)\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}(-i)^{q+1}$. Hence, equation (25) can be written as

$$\begin{aligned}\mu(n,q)&=y(n)-\sum_{i=1}^{q}a(i)\,y(n-i)+e(n)-\sum_{i=1}^{q}a(i)\,e(n-i)\\&=y(n)-\sum_{j=0}^{q+1}c(j)\,n^j+(q!)\,c(q)+((q+1)!)\,nc\,(q+1)\\&\quad-c\,(q+1)\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}(-i)^{q+1}+e(n)\\&\quad-\sum_{i=1}^{q}a(i)\,e(n-i)\\\mu(n,q)&=((q+1)!)\,nc\,(q+1)+(q!)\,c\,(q)\\&\quad-c\,(q+1)\sum_{i=1}^{q}(-1)^{i+1}\binom{q}{i}(-i)^{q+1}+e(n)\\&\quad-\sum_{i=1}^{q}a(i)\,e(n-i)\end{aligned}\tag{27}$$

According to equation (27), for $D = 2$ (an actual quadratic polynomial) and $q = 1$ (choosing a linear polynomial), $\mu(n, 1) = 2nc(2) + c(1) - c(2) + e(n) - e(n-1)$. Thus, $< \mu(n, 1) > = < 2n > c(2) + c(1) - c(2)$. Furthermore, $< \sigma_{\mu(1)} >^2 =$ variance of $2nc(2) + 2\sigma^2$.

In a similar manner, according to equation (27), for $D = 3$ (an actual cubic polynomial) and $q = 2$ (choosing a quadratic polynomial), one finds that $\mu(n, 2) = 6nc(3) + 2c(2) - 6c(3) + e(n) - 2e(n-1) + e(n-2)$. Thereby, $< \mu(n, 2) > = < 6n > c(3) + 2c(2) - 6c(3)$, and, of course, $< \sigma_{\mu(2)} >^2 =$ variance of $6nc(3) + 6\sigma^2$.

Similarly, according to equation (27), for $D = 4$ (an actual quartic polynomial) and $q = 2$ (choosing a cubic polynomial), one finds $\mu(n, 3) = 24nc(4) + 6c(3) - 36c(4) + e(n) - 3e(n-1) + 3e(n-2) - e(n-3)$. Therefore, $< \mu(n, 3) > = < 24n > c(4) + 6c(3) - 36c(4)$. Also, $< \sigma_{\mu(3)} >^2 =$ variance of $24c(3) + 20\sigma^2$.

From the experimental setup outlined above for $D = 2$,

$$< \mu(n, 1) > = \left(\sum_{n=-49}^{9} 2n \right) / 59 = -40 \tag{28}$$

This explains the column 2 of Table 3. Similar calculations can explain the columns 2 and 3 of Table 5, as well as the columns 2, 3, and 4, of Table 7. Basically, whenever the chosen degree is smaller than the correct degree, the $< \mu >$ comes from noise-free polynomial data values and independent of zero-mean noise standard deviations. This is different from the other two scenarios, in which $< \mu(n, q) >$ is equal to the q-th differential of a polynomial. This then completes the understanding of Tables 1, 3, 5, and 7.

Unlike in the other scenarios (i.e., choosing the correct degree or larger than the correct degree), here $\mu(n, 1)$ contains two types of terms – one is statistical, i.e., noise, and is the same as before but the other is noise-free data dependent, i.e., $2n$. As these two types of terms are independent of each other, one can write variance of $\mu(n, 1) =$ variance of $(2n)$ + $2\sigma^2$.

$$\begin{aligned} < \sigma_{\mu(q)} >^2 &= \frac{\left(\sum_{n=-49}^{9} (2n)^2 \right)}{59} - (-40)^2 + 2\sigma^2 \\ &= 2760 - (-40)^2 + 2\sigma^2 = 1160 + 2\sigma^2 \end{aligned} \tag{29}$$

This agrees well with the results in column 2 of Table 4. Similar calculations can explain the columns 2 and 3 of Table 6, as well as the columns 2, 3, and 4, of Table 8. Basically, whenever the chosen degree is smaller than the correct degree, the $< \sigma_{\mu(q)} >^2$ values come from a combination of noise-free polynomial data values and zero-mean noise standard deviations. This is different from the other two scenarios, in which $< \sigma_{\mu(q)} >^2$ values come only from noise standard deviations as given by equation (24). This then completes the understanding of Tables 2, 4, 6, and 8.

4.4. Standard deviation of standard deviation

In section 3 it was stated that values of the standard deviations of $\sigma_{\mu(q)}$, denoted by $\sigma(\sigma_{\mu(q)})$, were obtained in the above experiments. First some theoretical backgrounds are offered here before discussing the results in section 4.5. Consider n samples $\{r(1), r(2), \ldots, r(n)\}$ from a population that follows a Gaussian distribution. The sample standard deviation can be written as

$$s = \sqrt{\left[\frac{\left(\sum_{i=1}^{n} (r(i) - < r(i) >)^2 \right)}{n} \right]}$$

The distribution of s is given by [28]

$$p(n, s) = \frac{2 \left(\frac{n}{2\sigma^2} \right)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \exp\left(-\frac{ns^2}{2\sigma^2} \right) s^{n-2} \tag{30}$$

where $\Gamma(.)$ is the Gamma function and $\sigma^2 = ns^2/(n-1)$. Hence, the variance of s can be derived as

$$var(s) = \frac{1}{n} \left[n - 1 - \frac{2\Gamma^2\left(\frac{n}{2}\right)}{\Gamma^2\left(\frac{n-1}{2}\right)} \right] \sigma^2 \tag{31}$$

Therefore, the standard deviation of s can be written as

$$SD(s) = \sigma \sqrt{\frac{1}{n} \left[n - 1 - \frac{2\Gamma^2\left(\frac{n}{2}\right)}{\Gamma^2\left(\frac{n-1}{2}\right)} \right]} \tag{32}$$

which shows that, for a fixed value of n, $SD(s)$ increases linearly with σ.

4.5. Summary and verification

From the explorations in subsections 4.1, 4.2, and 4.3, one can summarise the following for $\mu(n, q)$:

1) if q is the correct degree of the polynomial, $\mu(n,q) = c(q)(q!) + \alpha(q)\ \mathbb{N}(0, \sigma)$, where $\alpha^2(q)$ = 2, 6, 20, 70, 252, and 924 for q = 1, 2, 3, 4, 5, and 6 respectively; these values of $\alpha^2(q)$ can be derived from equation (13). Therefore, it is clear that $\mu(n,q) = \mathbb{N}(c(q)(q!), \alpha(q)\sigma)$, which is a Gaussian distribution.
2) if q is larger than the correct degree of the polynomial, $\mu(n,q) = \alpha(q)\ \mathbb{N}(0, \sigma)$, where $\alpha^2(q)$ = 2, 6, 20, 70, 252, and 924 for q = 1, 2, 3, 4, 5, and 6 respectively. Therefore, $\mu(n,q) = \mathbb{N}(0, \alpha(q)\sigma)$, which is another Gaussian distribution.
3) if q is smaller than the correct degree of the polynomial, $\mu(n,q) = f(n) + \beta(q)\ \mathbb{N}(0, \sigma)$. Therefore, $\mu(n,q) = \mathbb{N}(f(n), \beta(q)\sigma)$, which is yet another Gaussian distribution.

Thus, $\mu(n, q)$ for every chosen degree (correct or not) follows a Gaussian distribution. Therefore, all the above narratives in subsection 4.4 are applicable to $\sigma(\sigma_{\mu(q)})$.

Relevant results from the aforementioned computer experiments in section 3 are presented in three Figures. Fig. 1 contains four sets of values corresponding to data from a linear polynomial with the chosen degree of 1 (star signs in cyan), from a quadratic polynomial with the chosen degree of 2 (square signs in blue), from a cubic polynomial with the chosen degree of 3 (open circle signs in green), and from a quartic polynomial with the chosen degree of 4 (diamond signs in magenta). The horizontal axis represents the values of noise standard deviations and the vertical axis represents the normalised $\sigma(\sigma_{\mu(q)})$ values. The dashed black line, with a slope of 1 and intercept of 0, has been added for reference only. Experimental results are in good agreement with the theoretical expectation that the normalised $\sigma(\sigma_{\mu(q)})$ values vary linearly with the noise standard deviation (σ) with a slope of 1 and intercept of 0, in the cases of chosen orders being the same as the correct orders.

Fig. 2 contains fourteen sets of values corresponding to data from a linear polynomial with the chosen degree of 2 (star signs in blue), from a linear polynomial with the chosen degree of 3 (star signs in green), from a linear polynomial with the chosen degree of 4 (star signs in magenta), from a linear polynomial with the chosen degree of 5 (star signs in black), from a linear polynomial with the chosen degree of 6 (star signs in red), from a quadratic polynomial with the chosen degree of 3 (square signs in green), from a quadratic polynomial with the chosen degree of 4 (square signs

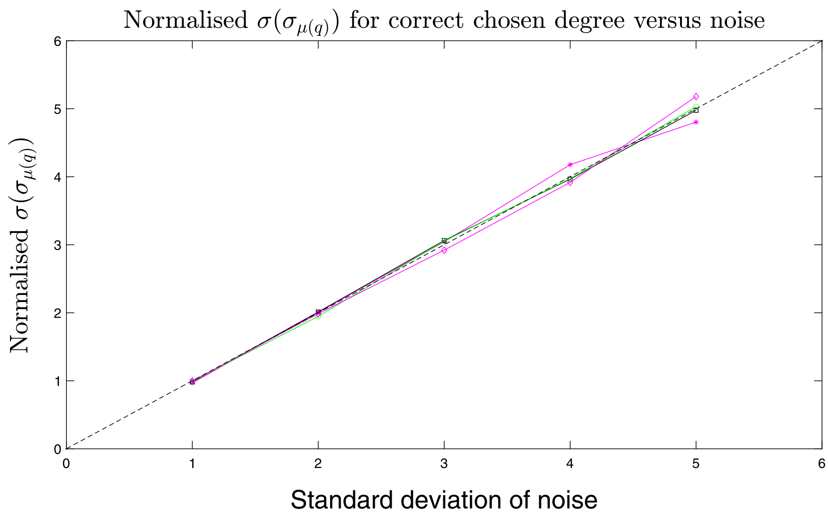

Fig. 1. Normalised $\sigma(\sigma_{\mu(q)})$ for correct chosen degree versus noise. The horizontal axis represents the values of noise standard deviations and the vertical axis represents the normalised $\sigma(\sigma_{\mu(q)})$ values. It contains four sets of values corresponding to data from four different polynomials. The dashed black line, with a slope of 1 and intercept of 0, has been added for reference only. (For the interpretation of the colours in the seven figures, the reader is referred to the web version of this article.)

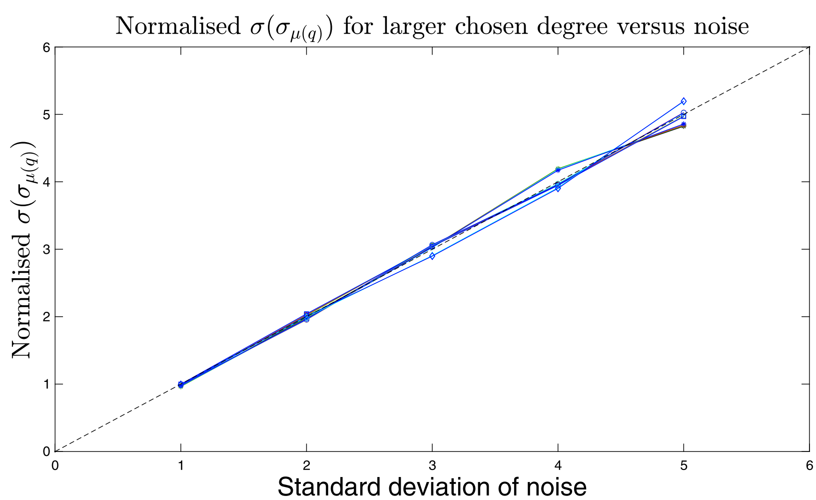

Fig. 2. Normalised $\sigma(\sigma_{\mu(q)})$ for larger chosen degree versus noise. The horizontal axis represents the values of noise standard deviations and the vertical axis represents the normalised $\sigma(\sigma_{\mu(q)})$ values. It contains fourteen sets of values corresponding to data from different polynomials and different chosen degrees. The dashed black line, with a slope of 1 and intercept of 0, has been added for reference only.

in magenta), from a quadratic polynomial with the chosen degree of 5 (square signs in black), from a quadratic polynomial with the chosen degree of 6 (square signs in red), from a cubic polynomial with the chosen degree of 4 (open circle signs in magenta), from a cubic polynomial with the chosen degree of 5 (open circle signs in black), from a cubic polynomial with the chosen degree of 6 (open circle signs in red), from a quartic polynomial with the chosen degree of 5 (diamond signs in black), and from a quartic polynomial with the chosen degree of 6 (diamond signs in red). The horizontal axis represents the values of noise standard deviations and the vertical axis represents the normalised $\sigma(\sigma_{\mu(q)})$ values. The dashed black line, with a slope of 1 and intercept of 0, has been added for reference only. Experimental results are in good agreement the theoretical expectation that the normalised $\sigma(\sigma_{\mu(q)})$ values vary linearly with the noise standard deviation (σ) with a slope of 1 and intercept of 0, in the cases of chosen orders being larger than the correct orders.

Fig. 3 contains six sets of values corresponding to data from a quadratic polynomial with the chosen degree of 1 (square signs in cyan), from a cubic polynomial with the chosen degree of 1 (open circle signs in cyan), from a cubic polynomial with the chosen degree of 2 (open circle signs in blue), from a quartic polynomial with the chosen degree of 1 (diamond signs in cyan), from a quartic polynomial with the chosen degree of 2 (diamond signs in blue), and from a quartic polynomial with the chosen degree of 3 (diamond signs in green). The horizontal axis represents the values of noise standard deviations and the vertical axis represents the normalised $\sigma(\sigma_{\mu(q)})$ values. The dashed black line, with a slope of 1 and intercept of 0, has been added for reference only. Again, experimental results are in good agreement with the theoretical expectation that the normalised $\sigma(\sigma_{\mu(q)})$ values vary linearly with the noise standard deviation (σ) with a slope of 1 and intercept of 0, in the cases of chosen orders being smaller than the correct orders.

It is clear that the experimental results are in good agreement with the theoretical expectation that the normalised $\sigma(\sigma_{\mu(q)})$ values vary linearly with the noise standard deviation (σ) with a slope of 1 and intercept of 0, in the cases of chosen orders being the same as the correct orders (Fig. 1) as well as in the cases of chosen orders being larger than the correct orders (Fig. 2) and for chosen orders being smaller than the correct orders (Fig. 3).

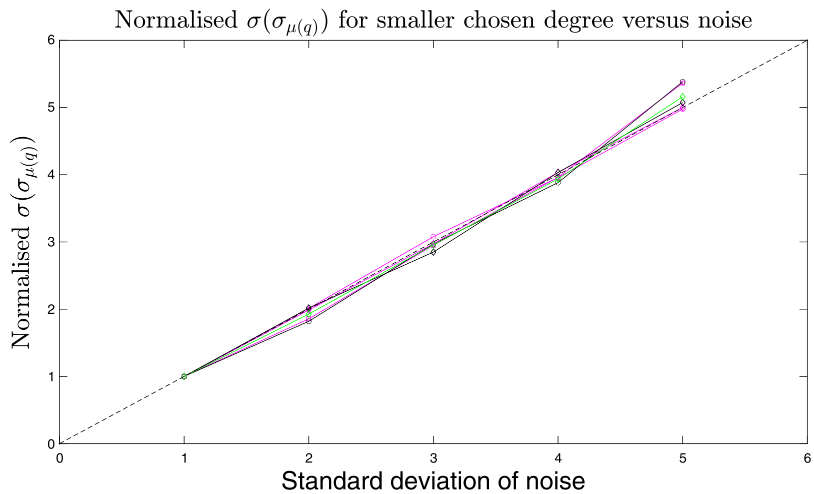

Fig. 3. Normalised $\sigma(\sigma_{\mu(q)})$ for smaller chosen degree versus noise. The horizontal axis represents the values of noise standard deviations and the vertical axis represents the normalised $\sigma(\sigma_{\mu(q)})$ values. It contains six sets of values corresponding to data from different polynomials and different chosen degrees. The dashed black line, with a slope of 1 and intercept of 0, has been added for reference only.

5. Polynomial degree estimators

Here, two polynomial degree estimators are introduced – the recently published PTS1 [29] and the proposed PTS2. These two are compared with the chosen four of the existing estimators.

5.1. PTS1

The equation (16) can be rearranged and approximated as follows

$$\hat{x}(n) = \sum_{i=1}^{q} a(i)\, x(n-i) + x(n) + <\mu(q)>$$

As everything on the right hand of the above equation is known, these $(f-q)$ values of $\hat{x}(n)$ are calculated and can be regarded as time-series “fitted” values. In these experiments, f is equal to 60. The relevant root-mean square time-series estimation error, $fe(q)$, is defined as

$$fe(q) = \sqrt{\left[\frac{\left(\sum_{i=q}^{f}\left(\hat{x}(i) - x(i)\right)^2\right)}{(f-q)}\right]} \tag{33}$$

where $(f-q)$ is the number of data values being estimated and $\hat{x}(i)$ are the estimated values. It should be noted that $fe(q)$ generally decreases as q increases.

Recall that each of AIC, AICc, GIC, and BIC attempts to balance between overfitting and underfitting scenarios. In one scenario errors reduce, while errors increase in the other scenario with increasing values of q. A similar scenario arises here, in that $fe(q)$ generally decreases as q increases, while $\sigma^2_{\mu(q)}$ increases with q. In [29], the following parameter is defined

$$PTS1(q) = \sigma^2_{\mu(q)} + (fe(q))^2 \tag{34}$$

The estimated value of q is the one for which PTS1(q) is the minimum.

5.2. PTS2

From the observations in Tables 2, 4, 6, and 8 as well as the analysis in sections 4.1, 4.2, and 4.3, the following can be concluded:

1) When the chosen degree (q) is the same as the correct degree (D), $<\sigma_{\mu(q)}>^2 = \alpha^2(q)\ \sigma^2$, where $\alpha^2(q) = 2, 6, 20, 70, 252,$ and 924 for $q = 1, 2, 3, 4, 5,$ and 6.
2) When the chosen degree (q) is larger than the correct degree (D), $<\sigma_{\mu(q)}>^2 = \alpha^2(q)\ \sigma^2$.
3) When the chosen degree (q) is smaller than the correct degree (D), $<\sigma_{\mu(q)}>^2 =$ (variance of some known function of n) + $\alpha^2(q)\ \sigma^2$.

When the chosen degree (q) is larger than the correct degree (D), $<\sigma_{\mu(q)}>^2 = \alpha^2(q>D)\ \sigma^2 > \alpha^2(D)\ \sigma^2$, which is the variance for the correct degree. Also, when the chosen degree (q) is smaller than the correct degree (D), for sufficient number of data values, $<\sigma_{\mu(q)}>^2 =$ (variance of some known function of n) + $\alpha^2(q<D)\ \sigma^2 > \alpha^2(D)\sigma^2$. Hence, the following parameter is defined

$$PTS2(q) = \sigma^2_{\mu(q)} \tag{35}$$

The estimated value of q is the one for which PTS2(q) is the minimum; PTS2(q) is the proposed degree estimator.

For each degree of polynomial from 1 to 4 (see section 3) and each of the five values of noise standard deviations from 1 to 5, 1000 sets of noisy data were generated. In each set of 101 data values, the first 60 data values have been used for estimating the degree of the polynomial. Fig. 4, Fig. 5, Fig. 6, and Fig. 7 display the accuracy (%) of degree estimation using AIC (green lower triangles), AICc (black upper triangles), GIC (magenta + signs), BIC (blue circles), PTS1 (red stars), and PTS2 (black diamonds) versus noise standard deviations for the linear polynomial, quadratic polynomial, cubic polynomial, and quartic polynomial respectively. Each of AIC, AICc, GIC, and BIC calculates log-likelihoods, which require the value of σ. When estimated values of σ were used the results were poorer. So, the exact values of σ were used for the above results, even though these are not available in reality. In each of these Figures AICc is always better than AIC, while GIC and BIC give very similar results, and they are always much better than AIC and AICc. PTS1 and the proposed PTS2 are always the best by far.

6. Noise variance estimator

All the preparatory work has been completed above to propose an estimator for noise standard deviation and to evaluate the same. In subsections 4.1 and 4.2 it has been shown that, when the chosen order is either the correct order or larger than the correct order, the variance is given by

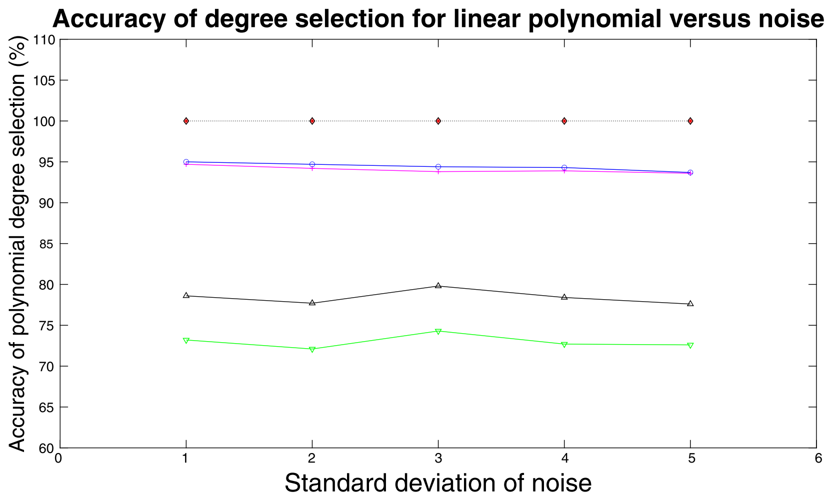

Fig. 4. Accuracy of degree selection for linear polynomial versus noise. The lines display the accuracy (%) of degree estimation using AIC (green lower triangles), AICc (black upper triangles), GIC (magenta + signs), BIC (blue circles), PTS1 (red stars), and PTS2 (black diamonds) versus noise standard deviations for the linear polynomial.

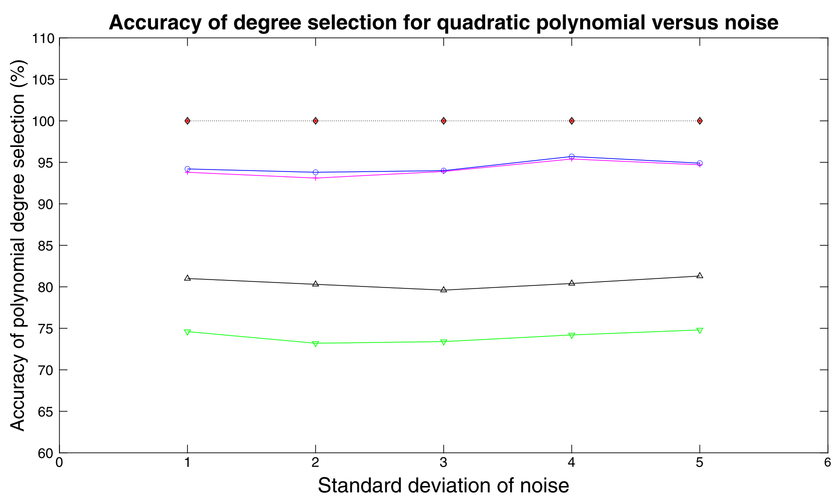

Fig. 5. Accuracy of degree selection for quadratic polynomial versus noise. The lines depict the accuracy (%) of degree estimation using AIC (green lower triangles), AICc (black upper triangles), GIC (magenta + signs), BIC (blue circles), PTS1 (red stars), and PTS2 (black diamonds) versus noise standard deviations for the quadratic polynomial.

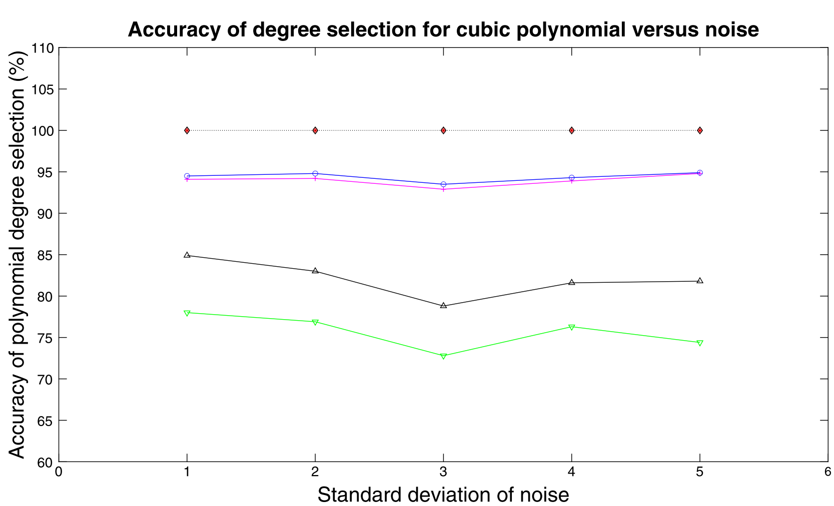

Fig. 6. Accuracy of degree selection for cubic polynomial versus noise. The lines show the accuracy (%) of degree estimation using AIC (green lower triangles), AICc (black upper triangles), GIC (magenta + signs), BIC (blue circles), PTS1 (red stars), and PTS2 (black diamonds) versus noise standard deviations for the cubic polynomial.

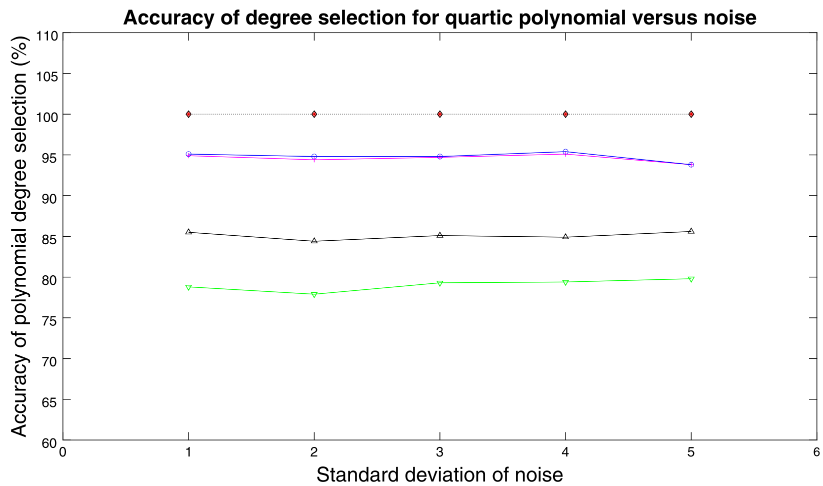

Fig. 7. Accuracy of degree selection for quartic polynomial versus noise. The lines display the accuracy (%) of degree estimation using AIC (green lower triangles), AICc (black upper triangles), GIC (magenta + signs), BIC (blue circles), PTS1 (red stars), and PTS2 (black diamonds) versus noise standard deviations for the quartic polynomial.

$$< \sigma_{\mu(q)} >^2 = \left((1)^2 + \sum_{i=1}^{q} a\,(i)^2 \right) \sigma^2 = \alpha^2(q)\sigma^2 \tag{36}$$

with known values of $\alpha^2(q)$. Therefore, the noise variance can be written as

$$\sigma\,(q)^2 = \frac{< \sigma_{\mu(q)} >^2}{\alpha^2(q)} \tag{37}$$

Of course, this is for a single value of q. There are infinitely many possible q values are available starting from the correct degree to any value larger than this. Defining D as the correct degree and M is the largest chosen degree, an alternative and more general estimator can be written as

$$\hat{\sigma}^2 = \left(\prod_{q=D}^{M} \left(\frac{< \sigma_{\mu(q)} >^2}{\alpha^2(q)} \right) \right)^{\frac{1}{M-D+1}} \tag{38}$$

Choosing a larger number of values of q allows one to combine many more estimations of σ^2 than choosing just a single value for q.

Some computer experiments have been carried out for each of the four polynomials (linear, quadratic, cubic, and quartic) and five different values of noise standard deviations σ of 1, 2, 3, 4, and 5. The experimental setup is the same as in sections 3. For each degree of a polynomial, 1,000 sets of 101 data values have been generated for each value of the standard deviation (σ) of noise using the zero-mean Gaussian distribution, $\mathbb{N}\,(0, \sigma)$. As before, in each set of 101 data values, the first 60 data values, i.e., $f = 60$, have been used for estimating the noise standard deviation, σ, using each of the five methods – AIC, AICc, GIC, BIC, and the proposed one. The true value of the noise standard deviation has been used for AIC, AICc, GIC, and BIC to obtain a more accurate estimate of the polynomial degree. For the proposed method, the value of M was set at 6 for illustration, though this can be made as large as one would like. The average standard deviation is calculated from the 1000 estimates of σ. These estimates are recorded in Table 9.

For each combination of the polynomial degree and the noise standard deviation, both the mean and the RMS values from AIC, AICc, GIC, and BIC, are extremely similar to each other; they appear to underestimate the true values. Expected values from the proposed method are pretty much the true values, even though RMS values from the proposed method are larger than those from AIC, AICc, GIC, and BIC.

7. Discussion

In the following three items are discussed.

Zero-mean Gaussian noise: all experimental results in previous sections have come from additive, zero-mean, independent and identically distributed Gaussian noise. The statistical significance of polynomial degree estimations from the six estimators for each combination of a polynomial degree (4 in total) and a noise standard deviation (5 in total) have been investigated. For each of these 20 combinations and each of the six estimators, there are 1000 estimated values of the polynomial degree. The MATLAB "poissfit" function has been used to obtain the maximum likelihood estimate of the degree and its 95% confidence interval (i.e., the significance level of 0.05). The maximum likelihood estimates of the degree in all 20 combinations for each of the chosen four of the existing estimators AIC, AICc, GIC, and BIC are different from the true value of the degree. For AIC and AICc, 95% confidence intervals do not contain the true degree 100% of these cases. For GIC and BIC, confidence intervals do not contain the true degree 25% of these cases. In contrast, PTS1 and PTS2 have selected the correct polynomial degree in all of these 20 combinations; these results are consistent with the true degree for 100% of these cases of zero-mean Gaussian noise.

Zero-mean non-Gaussian noise: As an example of non-Gaussian noise, the effects for adding Uniform noise have been investigated, using the identical procedure to the above. At the 95% confidence interval (i.e., the significance level of 0.05), the maximum likelihood estimate of the degree in all 20 combinations for each of the chosen four of the existing estimators is different from the true value of the degree. For AIC and AICc, confidence intervals do not contain the true polynomial degree 100% of these cases. For GIC and BIC, confidence intervals do not contain the true degree 25% of these cases. Yet, PTS1 and PTS2 have selected the correct polynomial degree in all of these 20 combinations and these results are consistent with the true degree in 100% of these cases of zero-mean Uniform noise.

For each estimator, 20 (4 degrees * 5 noise standard deviations) values of the percentage accuracy with Gaussian noise and the corresponding percentage accuracy with Uniform noise are very similar. For each estimator, both the average of these difference percentage accuracies and the standard deviation of these difference percentage accuracies are 0.65% ± 2.10% for AIC, -0.24% ± 1.78% for AICc, -0.37% ± 0.97% for GIC, -0.37% ± 0.97% for BIC, 0.0% ± 0.0% for PTS1 and 0.0% ± 0.0% for PTS2. Also, the PTS1 and PTS2 are the only two to select the correct degree of a polynomial

Table 9
Estimates of the noise standard deviation ($\hat{\sigma}$) using AIC, AICc, GIC, BIC, and the proposed method.

True Noise Standard Deviation (σ)	Method	Degree = 1 Linear Polynomial	Degree = 2 Quadratic Polynomial	Degree = 3 Cubic Polynomial	Degree = 4 Quartic Polynomial
1	AIC	0.96 ± 0.09	0.95 ± 0.09	0.95 ± 0.09	0.94 ± 0.09
	AICc	0.97 ± 0.09	0.96 ± 0.09	0.96 ± 0.09	0.94 ± 0.09
	GIC	0.97 ± 0.09	0.96 ± 0.09	0.96 ± 0.09	0.95 ± 0.09
	BIC	0.97 ± 0.09	0.96 ± 0.09	0.96 ± 0.09	0.95 ± 0.09
	Proposed	**1.00 ± 0.14**	**0.99 ± 0.15**	**0.99 ± 0.15**	**0.99 ± 0.16**
2	AIC	1.93 ± 0.18	1.91 ± 0.18	1.90 ± 0.18	1.88 ± 0.18
	AICc	1.94 ± 0.18	1.92 ± 0.18	1.90 ± 0.18	1.89 ± 0.18
	GIC	1.96 ± 0.18	1.93 ± 0.18	1.91 ± 0.18	1.90 ± 0.18
	BIC	1.96 ± 0.18	1.93 ± 0.18	1.91 ± 0.18	1.90 ± 0.18
	Proposed	**2.00 ± 0.28**	**1.99 ± 0.30**	**2.00 ± 0.30**	**1.99 ± 0.32**
3	AIC	2.91 ± 0.27	2.88 ± 0.27	2.84 ± 0.27	2.82 ± 0.28
	AICc	2.92 ± 0.28	2.89 ± 0.27	2.85 ± 0.26	2.83 ± 0.28
	GIC	2.94 ± 0.28	2.91 ± 0.27	2.87 ± 0.26	2.84 ± 0.28
	BIC	2.94 ± 0.28	2.91 ± 0.27	2.87 ± 0.26	2.84 ± 0.28
	Proposed	**3.00 ± 0.42**	**2.99 ± 0.45**	**2.99 ± 0.47**	**2.98 ± 0.47**
4	AIC	3.87 ± 0.37	3.84 ± 0.35	3.77 ± 0.35	3.78 ± 0.36
	AICc	3.88 ± 0.37	3.85 ± 0.35	3.78 ± 0.35	3.79 ± 0.36
	GIC	3.92 ± 0.37	3.88 ± 0.36	3.80 ± 0.35	3.80 ± 0.36
	BIC	3.92 ± 0.37	3.88 ± 0.36	3.80 ± 0.35	3.80 ± 0.36
	Proposed	**4.02 ± 0.59**	**4.01 ± 0.59**	**3.95 ± 0.60**	**4.00 ± 0.63**
5	AIC	4.83 ± 0.44	4.82 ± 0.44	4.76 ± 0.44	4.72 ± 0.46
	AICc	4.85 ± 0.44	4.83 ± 0.45	4.77 ± 0.44	4.73 ± 0.46
	GIC	4.89 ± 0.45	4.87 ± 0.45	4.80 ± 0.45	4.74 ± 0.46
	BIC	4.80 ± 0.45	4.87 ± 0.45	4.80 ± 0.45	4.74 ± 0.46
	Proposed	**4.99 ± 0.67**	**5.00 ± 0.74**	**4.99 ± 0.77**	**4.97 ± 0.84**

every time; their performances were always the best. BIC and GIC performances are very similar to each other, and these are better than AICc, which performed better than AIC.

In the cases of additive, zero-mean, independent and identically distributed, and symmetric non-Gaussian noise the results are expected to be similar, since the theoretical results in sections 4.1 and 4.2 do not assume more than the noise being additive, zero-mean, and independent and identically distributed. Indeed, these results also hold for additive, zero-mean, independent and identically distributed, and non-symmetric non-Gaussian noise. However, in this more general case the numerical values could be very different.

Note on applications: This is primarily a novel technique paper with a solid theoretical foundation and much success with simulated data. Nonetheless, there are many applications of this study, including satellite navigation, marine navigation, and digital mammography [30,31]. Yet another interesting new one can be in polymerase chain reaction (PCR) bias correction in quantitative DNA methylation studies. PCR plays a fundamental role in genetics as it facilitates the quantification of small amounts of genetic materials. In the last two or three decades of PCR's existence, the ultimate goal of reliable estimation of a basic parameter has proved to be elusive [10,32–35]. This has led to the development of different methods to analyze amplification curves. In [33], authors stated, "In published comparisons of these methods, available algorithms were typically applied in a restricted or outdated way" and went on to develop "a framework for robust and unbiased assessment of curve analysis performance". Subsequently, authors in [34] reexamined the study in [33] and cast doubt on it. In the calibration fitting they tried linear, quadratic, and cubic response functions [34]. One major challenge lies in choosing how to fit the data – hyperbolic, linear, quadratic, cubic, quartic. The study in [10] using 10 genes demonstrates that conventional hyperbolic models can fit the data but not so well. Cubic models can fit the data from 9 of the 10 genes better than hyperbolic models, except for one gene (SFRP1). The framework of this paper can fit data from all 10 genes well with no exceptions; the details of this study will be presented later.

8. Conclusion

The major contribution of this paper is the novel way to estimate the noise variance, with the relevant theoretical foundations, from a set of polynomial data values from an underlying polynomial with additive zero-mean, independent and identically distributed Gaussian noise without using any polynomial coefficients. Very recent results have demonstrated that all polynomials of degree q can be represented by the same set of known time-series coefficients of autoregressive models and a constant term μ [26]. Computer experiments have recorded $<\mu>$, $<\sigma_{\mu(q)}>$, and $\sigma\left(\sigma_{\mu(q)}\right)$ for different degrees of polynomials and noise powers. Models have been developed to explain all these results. These models provided inspiration to develop methods for the degree estimation and the noise power estimation. Existing techniques estimate the degree of the underlying polynomial and its corresponding coefficients. With this knowledge they can calculate the fitted values of the data and comparing them with the noisy data values can estimate the noise in the data. Four well-known and well-regarded maximum likelihood-based techniques have been used to estimate the degree of the polynomial and the noise variance.

The following of the proven items in this paper are highlighted:

1) The $<\mu>$ values, for a given polynomial and a chosen value for the degree of this polynomial, are independent of the noise standard deviations.
2) If the chosen degree is the same as the correct one, $<\mu>$ values are $c(q)q!$, where q is the correct degree of the polynomial and $c(q)$ is its leading degree coefficient.
3) Whenever the chosen degree is larger than the correct one, the $<\mu>$ values are 0.

4) Whenever the chosen degree is smaller than the correct one, the $<\mu>$ values are very different from $c(q)q!$ and 0.

Below are some important experimental observations about the degree and the noise power estimators:

1) The polynomial degree estimates obtained from the proposed PTS1 and PTS2 are significantly better than those from the existing methods – AIC, AICc, GIC, and BIC.
2) Estimated values of noise power from the proposed method are pretty much the true values and appear to be unbiased.

The proposed methods for the degree estimation and the noise power estimation from noisy polynomial data are uncommonly different in that they do not use polynomial coefficient values and yet they are remarkably successful.

CRediT authorship contribution statement

All aspects of conceptualization, methodology, software creation, data curation, investigation, supervision, validation, writing, and typing have been carried out by the sole author of this paper, Professor A.K. Nandi.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The author acknowledges Dr Chao Liu for formatting the manuscript.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.dsp.2021.103071.

References

[1] A.K. Nandi, D.J. Roberts, A.K. Nandi, Improved long-term time-series predictions of total blood use data from England, Transfusion 60 (10) (2020) 2307–2318, https://doi.org/10.1111/trf.15966.
[2] A.M. Legendre, Nouvelles méthodes pour la détermination des orbites des comètes, Chez Firmin Didot, Paris, France, 1805, “Sur la Méthode des moindres quarrés” appears as an appendix.
[3] C.F. Gauss, Theoria Motus Corporum Coelestium, Perthes, Hamburg, Germany, 1809, Translation reprinted as “Theory of motions of heavenly bodies moving about the sun in conic sections”, Dover, New York, 1963.
[4] C.F. Gauss, Theoria Combinationis Observationum Erroribus Minimis Obnoxiae, Henricus Dieterich, Gottingen, Germany, 1823.
[5] J.D. Gergonne, The application of the method of least squares to the interpolation of sequences, Hist. Math. 1 (4) (November 1974) 439–447, https://doi.org/10.1016/0315-0860(74)90034-2 (Translated by Ralph St. John and S.M. Stigler from the 1815 French ed.).
[6] S.M. Stigler, Gergonne’s 1815 paper on the design and analysis of polynomial regression experiments”, Hist. Math. 1 (4) (1974) 431–439, https://doi.org/10.1016/0315-0860(74)90033-0.
[7] G.U. Yule, On the theory of correlation, J. R. Stat. Soc. 60 (4) (1897) 812–854, https://doi.org/10.2307/2979746.
[8] K. Pearson, G.U. Yule, N. Blanchard, A. Lee, The law of ancestral heredity, Biometrika 2 (2) (1903) 211–236, https://doi.org/10.1093/biomet/2.2.211.
[9] R.A. Fisher, The goodness of fit of regression formulae, and the distribution of regression coefficients, J. R. Stat. Soc. 85 (4) (1922) 597–612, https://doi.org/10.2307/2341124.
[10] E.A. Moskalev, et al., Correction of PCR-bias in quantitative DNA methylation studies by means of cubic polynomial regressions, Nucleic Acids Res. 39 (11) (2011) e77, https://doi.org/10.1093/nar/gkr213.
[11] H. Linhart, W. Zucchini, Model Selection, Wiley, New York, 1986.
[12] K.P. Burnham, D.R. Anderson, Model Selection and Multi-Model Inference, Springer-Verlag, New York, 2002.
[13] P. Stoica, Y. Selen, Model-order selection, IEEE Signal Process. Mag. 21 (4) (2004) 36–47, https://doi.org/10.1109/MSP.2004.1311138.
[14] H. Akaike, A new look at statistical model identification, IEEE Trans. Autom. Control 19 (6) (1974) 716–723, https://doi.org/10.1109/TAC.1974.1100705.
[15] H. Akaike, On the likelihood of a time series model, Statistician 27 (3) (1978) 217–235.
[16] K.P. Burnham, D.R. Anderson, Multi-model inference: understanding AIC and BIC in model selection, Sociol. Methods Res. 33 (2) (2004) 261–304, https://doi.org/10.1177/0049124104268644.
[17] J.E. Cavanaugh, Unifying the derivations of the Akaike and corrected Akaike information criteria, Stat. Probab. Lett. 31 (2) (1997) 201–208, https://doi.org/10.1016/s0167-7152(96)00128-9.
[18] G. Claeskens, N.L. Hjort, Model Selection and Model Averaging, Cambridge University Press, 2008.
[19] N. Sugiura, Further analysis of the data by Akaike’s information criterion and the finite corrections, Commun. Stat., Theory Methods 7 (1978) 13–26, https://doi.org/10.1080/03610927808827599.
[20] S. Konishi, S. Kitagwa, Information Criteria and Statistical Modeling, Springer, New York, 2008.
[21] R.J. Bhansali, D.Y. Downham, Some properties of the order of an autoregressive model selected by a generalization of Akaike’s FPE criterion, Biometrika 64 (3) (1977) 547–551, https://doi.org/10.1093/biomet/64.3.547.
[22] G.E. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (2) (1978) 461–464.
[23] R.I. Kashyap, Optimal choice of AR and MA parts in autoregressive moving average models, IEEE Trans. Pattern Anal. Mach. Intell. 4 (2) (1982) 99–104.
[24] Y. Yang, Can the strengths of AIC and BIC be shared?, Biometrika 92 (4) (2005) 937–950, https://doi.org/10.1093/biomet/92.4.937.
[25] S.I. Vrieze, Model selection and psychological theory: a discussion of the differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), Psychol. Methods 17 (2) (2012) 228–243, https://doi.org/10.1037/a0027127.
[26] A.K. Nandi, Data modelling with polynomial representations and autoregressive representations, and their connections, IEEE Access 8 (2020) 110412–110424, https://doi.org/10.1109/ACCESS.2020.3000860.
[27] I.S. Gradshteyn, I.M. Ryzhik, Tables of Integrals, Series, and Products, Academic Press, New York, 1980, ISBN 10: 0122947606 and ISBN 13: 978-0122947605. (Corrected and Enlarged Edition Prepared by Alan Jeffrey).
[28] J.F. Kenney, E.S. Keeping, The distribution of the standard deviation, in: Mathematics of Statistics, Pt. 2, 2nd ed., Van Nostrand, Princeton, NJ, 1951, pp. 160–173.
[29] A.K. Nandi, Model order selection from noisy polynomial data without using any polynomial coefficients, IEEE Access 8 (2020) 130417–130430, https://doi.org/10.1109/ACCESS.2020.3008527.
[30] N. Owsley, Fitting polynomials to data in the presence of noise, in: Proceedings of the IEEE ICASSP, Paris, France, 1982, pp. 1505–1508.
[31] P. Monnin, et al., Comparison of the polynomial model against explicit measurements of noise components for different mammography systems, Phys. Med. Biol. 59 (19) (2014) 5741–5761, https://doi.org/10.1088/0031-9155/59/19/5741.
[32] N. Fernandez-Jimenez, et al., Cubic regression-based degree of correction predicts the performance of whole bisulfitome amplified DNA methylation analysis, Epigenetics 7 (12) (2012) 1349–1354, https://doi.org/10.4161/epi.22846.
[33] J.M. Ruijter, et al., Evaluation of qPCR curve analysis methods for reliable biomarker discovery: bias, resolution, precision, and implications, Methods 59 (2013) 32–46, https://doi.org/10.1016/j.ymeth.2012.08.011.
[34] J. Tellinghuisen, A-N. Spiess, Comparing real-time quantitative polymerase chain reaction analysis methods for precision, linearity, and accuracy of estimating amplification efficiency, Anal. Biochem. 449 (2014) 76–82, https://doi.org/10.1016/j.ab.2013.12.020.
[35] J. Tellinghuisen, A-N. Spiess, qPCR data analysis: better results through iconoclasm, Biomolecular Detection and Quantification 17 (2019) 100084, https://doi.org/10.1016/j.bdq.2019.100084.

Professor Nandi received Ph.D. from the University of Cambridge. He worked in several universities, including Oxford and Imperial College London, and was Finland Distinguished Professor. He co-discovered three fundamental particles (W^+, W^- and Z^0), for which the Nobel Prize for Physics was awarded to his two team leaders in 1984. His current research interests lie in signal processing and machine learning. He authored over 600 technical publications, including 250 journal papers and five books. The H-index of his publications is 80 (Google Scholar). He is a Fellow of the Royal Academy of Engineering and seven other institutions, including the IEEE.