8-2017

Efficient Parameter-Estimating Algorithms for Symmetry-Motivated Models: Econometrics and Beyond

Vladik Kreinovich
University of Texas at El Paso, vladik@utep.edu

Anh H. Ly
Banking University of Ho Chi Minh City

Olga Kosheleva
University of Texas at El Paso, olgak@utep.edu

Songsak Sriboonchitta
Chiang Mai University, songsakecon@gmail.com

Comments:
Technical Report: UTEP-CS-17-81

Efficient Parameter-Estimating Algorithms for Symmetry-Motivated Models: Econometrics and Beyond

Vladik Kreinovich, Anh H. Ly, Olga Kosheleva, and Songsak Sriboonchitta

Abstract It is known that symmetry ideas can explain the empirical success of many non-linear models. This explanation makes these models theoretically justified and thus, more reliable. However, the models remain non-linear and thus, identification or the model's parameters based on the observations remains a computationally expensive nonlinear optimization problem. In this paper, we show that symmetry ideas can not only help to select and justify a nonlinear model, they can also help us design computationally efficient almost-linear algorithms for identifying the model's parameters.

1 Formulation of the Problem

Need for prediction. In many real-life situations, we have a quantity x that changes with time t, and we want to use the previous values of this quantity to predict its future values. For example, we know how the stock price has changed with time, and we want to use this information to predict future stock prices.

In many cases, such a prediction is possible. For example, when weather records show clear yearly cycles, it is reasonable to predict that a similar yearly cycle will be observed in the future as well.

Vladik Kreinovich and Olga Kosheleva
University of Texas at El Paso, 500 W. University,
El Paso, Texas 79968, USA, e-mail: olgak@utep.edu, vladik@utep.edu

Anh H. Ly
Banking University of Ho Chi Minh City, 56 Hoang Dieu 2, Quan Thu Duc, Thu Duc
Ho Ch Minh City, Vietnam

Songsak Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai 50200 Thailand
e-mail: songsakecon@gmail.com

How can we predict: main idea. A usual approach to prediction is that we select some *model*, i.e., some parametric family of functions $f(t, c_1, \ldots, c_\ell)$. Based on the available observations, we find the parameters $\widetilde{c}_i$ which provide the best fit, and then we use these values $\widetilde{c}_j$ to predict the future values of the quantity x as

$$x(t) \approx f(t, \widetilde{c}_1, \ldots, \widetilde{c}_\ell).$$

Examples of models. In some cases, the dependence of the quantity x on time t is polynomial, in which case

$$f(t, c_1, \ldots, c_\ell) = c_1 + c_2 \cdot t + c_3 \cdot t^2 + \ldots + c_\ell \cdot t^{\ell-1}.$$

For a simple periodic process, the dependence of the quantity x on time is described by a sinusoid, in which case

$$f(t, c_1, c_2, c_3) = c_1 \cdot \sin(c_2 \cdot t + c_3).$$

To get a more realistic description of a periodic process, we need to take into account higher harmonics, i.e., assume that

$$f(t, c_1, c_2, \ldots) = c_1 \cdot \sin(c_2 \cdot t + c_3) + c_4 \cdot \sin(2c_2 \cdot t + c_5) + \ldots$$

For a simple radioactive decay, the amount of radioactive material decreases exponentially:

$$f(t, c_1, c_2) = c_1 \cdot \exp(-c_2 \cdot t).$$

A more realistic model takes into account that often, a radioactive material is a mixture of several different isotopes, with different half-lives. In this case,

$$f(t, c_1, c_2, \ldots) = c_1 \cdot \exp(-c_2 \cdot t) + c_3 \cdot \exp(-c_4 \cdot t) + \ldots$$

Other models include *log-periodic model*

$$f(t, c_1, c_2, \ldots, c_7) = c_1 + c_2 \cdot (c_3 - t)^{c_4} + c_5 \cdot (c_3 - t)^{c_4} \cdot \cos(c_6 \cdot \ln(c_3 - t) + c_7)$$

which is used to predict economic crashes [2, 4, 5, 7, 8, 9, 10, 11, 12, 14, 21, 22, 23, 24, 25, 26], or a model

$$f(t, c_1, c_2, c_3) = c_1 \cdot \ln(t - c_2) + c_3$$

that describes, for some software packages, the dependence of the number of uncovered faults on time t; see, e.g., [15, 16].

A more complex example is a neural network, when c_j are the corresponding weights; see, e.g., [1, 6].

How do we estimate the parameters? Usually, the Least Squares method is used to estimate the values of the parameters $c_1, \ldots, c_\ell$.

In other words, based on the values $x(t_i)$ observed at different moments of time t_i, $1 \le i \le n$, we find the values c_j for which the mean square approximation error is the smallest possible, i.e., for which the following expression is minimized:

$$\sum_{i=1}^{n} (x_i - f(t_i, c_1, \ldots, c_\ell))^2. \tag{1}$$

Identifying the model's parameters is often computationally intensive. In some cases – e.g., for the polynomial dependence – the model $f(x, c_1, \ldots, c_\ell)$ linearly depends on the values of the parameters c_j. In this case, the minimized expression (1) is quadratic in c_j.

We can find the minimum of a function of several variables by equating all its partial derivatives to 0. For a quadratic objective function (1), all the partial derivatives are linear functions of c_j. Thus, by equating them all to 0, we get a system of linear equations for the unknowns c_j. For solving systems of linear equations, there are many efficient algorithms, so in this case, the problem of identifying the model's parameters is computationally easy.

On the other hand, in general, the dependence of the model on the parameters c_j is non-linear. Thus, the objective function (1) is more complex than quadratic. It is known that, in general, optimization is computationally intensive – for example, it has been proven that optimization is an NP-hard problem, meaning that it as complex as a computational problem can be; see, e.g., [13, 17, 18].

It is therefore desirable to select models for which identification is easier. This bring us to a question of how we select models in the first place.

How are models selected in the first place? Sometimes, we have an good understanding of the processes that cause the quantity x to change. In such situations, we have a theoretically justified model.

In most cases, however, the model is selected empirically. We try different models, and we select the one for which, for the same number of parameters, the approximation error is the smallest.

In many cases, the empirical efficiency of selected models can be explained by symmetry ideas. In an empirical choice, we only compare a few possible models. As a result, the fact that the selected model turned out to be better than others does not necessarily mean that this model is indeed the best for a given phenomenon: there are, in principle, many other models that we did not consider in our empirical comparison.

Good news is that in many cases, the empirical selection can be confirmed by a theoretical analysis. For example, often, it turns out that the empirically successful model can be derived from the natural symmetry requirements; see, e.g., [16]. This theoretical justification compares the selected model not just with a few others, but with *all* possible models – thus, it makes us more confident that the selected model is indeed the best.

But the model remains computationally intensive. The fact that the empirically selected model is theoretically justified does not change its formulas. So, if the dependence of this model on the corresponding parameters c_j is non-linear, the problem of identifying parameters of this model remains computationally intensive.

What we do in this paper: we show that symmetries can help in parameter identification too. In this paper, we show that symmetries are not only helpful in selecting a model, they can also help design computationally efficient algorithms for identifying parameters of the selected model.

Structure of this paper. In Section 2, we briefly recall what symmetries are used to derive the corresponding models, how exactly these models are derived, and what are the resulting models. In Section 3, we analyze the problem of determining parameters of these models, and we show how to make this identification computationally easier.

2 How Symmetries Justify Models: A Brief Reminder

Preliminaries In some practical cases, the changes in the quantity x come from a single and simple process – this is the situation, e.g., with most oscillations. In most practical cases, however, many different factors lead to changes in x. Some of these changes are independent, and may have different intensity. Thus, the resulting value of the quantity x can be represented as a linear combination of the dependencies corresponding to different factors.

In precise terms, this means that we consider models of the type

$$C_1 \cdot e_1(t) + \ldots + C_m \cdot e_m(t) \tag{2}$$

for some functions $e_j(t)$ (which may depend on other parameters as well).

- This is the case for polynomials, when $e_1(t) = 1$, $e_2(t) = t$, $e_3(t) = t^2$, etc.
- This is the case for periodic processes, when $e_1(t)$ is the main sinusoid, $e_2(t)$ is the sinusoid corresponding to double frequency, $e_3(t)$ is the sinusoid corresponding to triple frequency, etc.
- This is the case for radioactive decay, where $e_j(t)$ are exponential functions with different hall-life.

In all these cases, the functions $e_j(t)$ are differentiable (smooth). So, without losing generality, we can assume that these functions are smooth.

In these terms, selecting a model means selecting the corresponding functions

$$e_1(t), \ldots, e_m(t).$$

What natural symmetries should we consider? Many physical processes – such as radioactive decay – do not have a starting point, their general properties do not

change whether we consider the piece of a radioactive material now or in a hundred years. The exact amount of the material will decrease, but its properties – and its rate of decay – will remain the same. In such situations, the observed value $x(t)$ changes with time, but the whole family of functions (2) should not change if we simply start counting time from a different starting point.

If we start to count time from a starting point which is t_0 moments in the future, then moment t in the new scale corresponds to moment $t+t_0$ in the original scale. Thus, if in the new scale, the set of functions has the form (2), then these same functions in the original time scale have the form

$$C_1 \cdot e_1(t+t_0) + \ldots + C_m \cdot e_m(t+t_0). \tag{3}$$

The above natural requirement then says that the families (2) and (3) must coincide – i.e., that:

- every function of type (2) can be expressed in the form (3) (with, of course, different constants C_j), and
- vice versa, every function of type (3) can be expressed in the form (2).

In other cases, there *is* a natural starting (or ending) point t_0, but there is no preferred time unit. In such cases, it is reasonable to require that if we use a different unit for measuring time, nothing will change – in particular, the class (2) of possible dependencies should not change.

If we keep t_0 as the starting point, and choose a measuring unit which is λ times smaller, then we get a new numerical value $t' = t_0 + \lambda \cdot (t - t_0)$. It is therefore reasonable to require that if we make this change, the family of approximating functions remains the same, i.e., that the family

$$C_1 \cdot e_1(t_0 + \lambda \cdot (t - t_0)) + \ldots + C_m \cdot e_m(t_0 + \lambda \cdot (t - t_0)) \tag{4}$$

coincides with the original family (2).

What can we conclude from these symmetry requirements. Let us consider the two cases separately:

- first, the case (3) of shift-invariance, and
- then, the case (4) of scale-invariance.

Case of shift-invariance. In the shift-invariant case, every function from the family (3) also belongs to the family (2).

In particular, for every j and t_0, the function $e_j(t+t_0)$ belongs to the family (3): it corresponds to the case when $C_j = 1$ and $C_{j'} = 0$ for all $j' \neq j$. Thus, we conclude that the function $e_j(t+t_0)$ belongs to the family (2), i.e., that

$$e_j(t+t_0) = C_{1j}(t_0) \cdot e_1(t) + \ldots + C_{mj}(t_0) \cdot e_m(t) \tag{5}$$

for some coefficients $C_{j'j}(t_0)$ depending on the shift t_0.

For each t, if we consider the equation (5) at m different moments of time $t = t_1, \ldots, t_m$, then we get the following system of m linear equations with m linear unknowns $C_{1j}(t_0), \ldots, C_{mj}(t_0)$:

$$\begin{aligned} e_j(t_1+t_0) &= C_{1j}(t_0)\cdot e_1(t_1)+\ldots+C_{mj}(t_0)\cdot e_m(t_1),\\ e_j(t_2+t_0) &= C_{1j}(t_0)\cdot e_1(t_2)+\ldots+C_{mj}(t_0)\cdot e_m(t_2),\\ &\ldots\\ e_j(t_m+t_0) &= C_{1j}(t_0)\cdot e_1(t_m)+\ldots+C_{mj}(t_0)\cdot e_m(t_m). \end{aligned} \tag{6}$$

The solution to a linear system can be explicitly described by the Cramer's rule (see. e.g., [19]), according to which this solution is a ratio of two determinants – i.e., a differentiable function of the right-hand sides and of the coefficients at the unknowns. Since the functions $e_j(t)$ are smooth, the right-hand sides and the coefficients are also smooth, and thus, thus the solution $C_{j'j}(t_0)$ is a differentiable function of differentiable functions – thus, a smooth function itself.

Since the functions $e_{j'}(t)$ and $C_{j'j}(t_0)$ are all differentiable, we can differentiate both sides of equation (5) by t_0 and take $t_0 = 0$. As a result, for each j, we get the following differential equation:

$$e_j'(t) = c_{1j}\cdot e_1+\ldots+c_{mj}\cdot e_m, \tag{7}$$

where e_j', as usual, denotes the derivatives, and $c_{j'j} \stackrel{\text{def}}{=} C_{j'j}'(0)$.

Thus, m functions $e_1(t)$, ..., $e_m(t)$ satisfy a system of m linear differential equations (7) with constant coefficients. A general solution to this system of equations is well known: it is a linear combination of functions of the type $t^k\cdot\exp(\lambda\cdot t)$, where λ are eigenvalues of the matrix $c_{j'j}$ and factors $t, t^2, \ldots, t^q$ appear if the corresponding eigenvalue is multiple, with multiplicity q; see, e.g., [20]. Please note that the eigenvalues are, in general, complex numbers $\lambda = a+b\cdot\mathrm{i}$, in which case

$$\exp(\lambda\cdot t) = \exp(a\cdot t)\cdot(\cos(b\cdot t)+\mathrm{i}\cdot\sin(b\cdot t)).$$

In real-valued terms, each function $e_j(t)$ is thus a linear combination of functions of the type

$$t^k\cdot\exp(a\cdot t)\cdot(\cos(b\cdot t)+\mathrm{i}\cdot\sin(b\cdot t)).$$

Case of scale-invariance. Let us now consider the case of scale-invariance with respect to the special point t_0. To simplify our analysis, let us consider, instead of time, an auxiliary variable $\tau \stackrel{\text{def}}{=} \ln(t-t_0)$. In terms of this auxiliary variable, we have $t = t_0+\exp(\tau)$, and the original functions $e_i(t)$ take the form $E_i(\tau) = e_i(t_0+\exp(\tau))$.

In terms of the new variable τ, the scaling transformation takes the form $\tau\to\tau+\tau_0$, where $\tau_0 \stackrel{\text{def}}{=} \ln(\lambda)$. Thus, for the new functions $E_j(\tau)$, scale-invariance means that the original class of functions

$$C_1 \cdot E_1(\tau) + \ldots + C_m \cdot E_m(\tau)$$

coincides with the transformed family

$$C_1 \cdot E_1(\tau + \tau_0) + \ldots + C_m \cdot E_m(\tau + \tau_0).$$

We already know what this condition implies: that each function $E_j(\tau)$ is a linear combination of functions

$$\tau^k \cdot \exp(\lambda \cdot \tau) = \tau^k \cdot \exp(a \cdot \tau) \cdot (\cos(b \cdot \tau) + \mathrm{i} \cdot \sin(b \cdot \tau)).$$

Substituting τ's definition $\tau = \ln(t - t_0)$ into this formula, and taking into account that $\exp(\tau) = \exp(\ln(t - t_0)) = t - t_0$ and thus, $\exp(a \cdot \tau) = (\exp(\tau))^a = (t - t_0)^a$, we conclude that each function $e_j(t) = E_j(\tau) = E_j(\ln(t - t_0))$ is a linear combination of functions of the type

$$(\ln(t - t_0))^k \cdot (t - t_0)^\lambda =$$

$$(\ln(t - t_0))^k \cdot (t - t_0)^a \cdot (\cos(b \cdot \ln(t - t_0) + \mathrm{i} \cdot \sin(b \cdot \ln(t - t_0))).$$

Comments.

- While it is good that we get expressions similar to what we have empirically observed, be it in case of predicting economic crashes or the case of predicting the number of discovered software faults, the dependence of these expressions on the corresponding parameters t_0, a, and b is highly nonlinear. So, it is computationally difficult to identify the parameters of these models from observations.
- What if we have both shift- and scale-invariance? In this cases, the expression should be both a linear combination of the terms $t^k \cdot \exp(\lambda \cdot t)$ and a combination of the terms of the type $(\ln(t - t_0))^k \cdot (t - t_0)^\lambda$. The need for the second interpretation excludes exponential terms, so such functions should be linear combinations of terms x^k, i.e., polynomials, with C_j as the only parameters. This is the only case when the dependence on the parameters is linear and so, identification of these parameters is computationally easy.

What we plan to do now. Now that we have described the symmetry-motivated models, let us described how to make identification of the parameters of these models easy.

3 Analysis of the Problem and Resulting Computationally Efficient Parameter Identification

Main idea. What we would like to do is come up with a linear differential equation with linear coefficients that describes all linear combinations of symmetry-

motivated models. To describe such an equation, let us denote the differentiation operation by D, so that $(Df)(t) \stackrel{\text{def}}{=} f'(t)$.

Shift-invariant case: analysis of the problem. Let us start with describing shift-invariant models in these terms. In these models, every function $e_j(t)$ is a linear combination of functions of the type $x^k \cdot \exp(\lambda \cdot t)$.

To find an appropriate differential equation for these functions, let us start with the case $k = 1$, when this function takes the form $\exp(\lambda \cdot t)$. For the function

$$\exp(\lambda \cdot t),$$

we have $D\exp(\lambda \cdot t) = \lambda \cdot \exp(\lambda \cdot t)$, thus $(D - \lambda)\exp(\lambda \cdot t) = 0$.

For the next ($k = 1$) function $e(t) = t \cdot \exp(\lambda t)$, we have

$$(De)(t) = \exp(\lambda \cdot t) + \lambda \cdot \exp(\lambda \cdot t),$$

thus $((D - \lambda)e)(t) = \exp(\lambda \cdot t)$. We already know that

$$(D - \lambda)\exp(\lambda \cdot t) = 0,$$

thus we have $((D - \lambda)^2 e)(t) = 0$.

Similarly, for the function $e(t) = t^k \cdot \exp(\lambda \cdot t)$, we have

$$(De)(t) = k \cdot t^{k-1} \cdot \exp(\lambda \cdot t) + \lambda \cdot t^k \cdot \exp(\lambda \cdot t),$$

thus

$$((D - \lambda)e)(t) = k \cdot t^{k-1} \cdot \exp(\lambda \cdot t).$$

So, by induction, we can prove that for this function $e(t)$, we have $(D - \lambda)^k e = 0$.

Different expressions forming $e_j(t)$ correspond to different eigenvalues λ_ℓ, so each of them annihilated by a corresponding differential operation $D - \lambda_\ell$, or, if this eigenvalue if multiple with multiplicity q_ℓ, by an operator $(D - \lambda_\ell)^{q_\ell}$. Thus, if we apply all these operators one after another, all the terms in $e_j(t)$ will be annihilated and thus, we will have a differential operator

$$\widetilde{D} \stackrel{\text{def}}{=} (D - \lambda_1)^{q_1} (D - \lambda_2)^{q_2} \ldots (D - \lambda_m)^{q_m}$$

for which $\widetilde{D} e_j = 0$ for all j. Since each model $x(t)$ is a linear combination of the functions $e_j(t)$, the function $x(t)$ also satisfies the equation $\widetilde{D}x = 0$.

If we open the parentheses, we conclude that $\widetilde{D}$ is a polynomial of m-th order in terms of D, i.e., that it has the form

$$\widetilde{D} = D^m + a_1 \cdot D^{m-1} + a_2 \cdot D^{m-2} + \ldots + a_m.$$

Thus, the equation $(\widetilde{D}x)(t) = 0$ takes the form

$$\frac{d^m x}{dt^m} + a_1 \cdot \frac{d^{m-1} x}{dt^{m-1}} + a_2 \cdot \frac{d^{m-2} x}{dt^{m-2}} + \ldots + a_m \cdot x = 0. \tag{8}$$

This is the desired differential equation with constant coefficients.

Examples. For a polynomial of order $\le m-1$, all eigenvalues are zeros, so $\widetilde{D} = D^m$, and the corresponding differential equation has the form

$$\frac{d^m x}{dt^m} = 0.$$

One can see that solutions to this differential equation are indeed exactly polynomials of order $\le m-1$.

For a simple sinusoidal signal $x(t) = A \cdot \cos(\omega \cdot t + \varphi)$, we get a second order differential equation with constant coefficients

$$\frac{d^2 x}{dt^2} + a_1 \cdot \frac{dx}{dt} + a_2 \cdot x = 0.$$

To be more precise, the sinusoid correspond to the case when $a_1 = 0$ and $a_2 > 0$; other cases correspond to exponential functions or functions of the type

$$A \cdot \exp(-a \cdot t) \cdot \cos(\omega \cdot t + \varphi).$$

How can we easily identify a model: towards an algorithm. Instead of the original parameters of the model – parameters on which depends highly non-linearly – we can instead identify the parameters $a_1, \ldots, a_m$ of the corresponding differential equation (8).

Of course, we have to approximate each derivative by a finite difference, so that if we start with a sequence of values $x_1, \ldots, x_i, \ldots$ corresponding to moments of time

$$t_1, \;\; t_2 = t_1 + \Delta t, \;\; t_3 = t_1 + 2\Delta t, \ldots, t_i = t_1 + (i-1) \cdot \Delta t,$$

then we form finite difference $(\Delta x)_i \stackrel{\text{def}}{=} \dfrac{x_i - x_{i-1}}{\Delta t}$. Then, instead of the second derivatives, we will use the values

$$(\Delta^2 x)_i \stackrel{\text{def}}{=} (\Delta(\Delta x))_i = \frac{(\Delta x)_i - (\Delta x)_{i-1}}{\Delta t} = \frac{x_i - 2x_{i-1} + x_{i-2}}{(\Delta t)^2}.$$

Similarly, in the general case, we have

$$(\Delta^k x)_i = (\Delta(\Delta^{k-1} x))_i = \frac{x_i - k \cdot x_{t-1} + C_2^k \cdot t_{i-1} - C_3^k \cdot t_{i-2} + \ldots + (-1)^k \cdot t_{i-k}}{(\Delta t)^k}.$$

So, instead of equation (8), we have an approximate equation

$$(\Delta^m x)_i + a_1 \cdot (\Delta^{m-1} x)_i + a_2 \cdot (\Delta^{m-2} x)_i + \ldots + x_i = 0. \tag{9}$$

The values $(\Delta^k x)_i$ are computed based on the observations x_i, so we get an (overdetermined) system of linear equations from which we can easily find the unknowns $a_1, \ldots, a_m$ by using the Least Squares method.

Shift-invariant case: resulting algorithm. Based on the sequence of observations $x_i = x(t_i)$, we compute the sequence of values $(\Delta x)_i = \dfrac{x_i - x_{i-1}}{\Delta t}$, then the sequence $(\Delta^2 x)_i = (\Delta(\Delta x))_i$, etc., until we have computed $(\Delta^m x)_i$. Based on thus computed sequences, we find the parameters a_j by applying the Least Squares Method to the equations (9).

Important comments.

- No problem if observations are not equally spaced in time: just take $(\Delta x)_i = \dfrac{x_i - x_{i-1}}{\Delta t_i}$, where we denoted $\Delta t_i \stackrel{\text{def}}{=} t_i - t_{i-1}$.
- It should be mentioned that even when the measurements of $x_i = x(t_i)$ at different moments of time are uncorrelated, their linear combinations (as in the left-hand side of formula (9)) are correlated, since the expressions for i and for $i-1$ now depend on the same value x_i. Thus, we need to use the Least Squares in the presence of this easy-to-compute correlation. This does not affect the computational easiness – the expression is still quadratic and equating its derivatives to 0 still leads to a system of linear equations.
- If needed, we can convert the new parameters $a_1, \ldots, a_m$ into the more traditional ones. All we need for this is to compute the derivatives of the original expressions $f(t, c_1, \ldots, c_\ell)$ and find the values a_j for which the linear combinations of these derivatives are 0s. Then, we get expressions describing a_j in terms of c_j: $a_j = f_j(c_1, \ldots, c_\ell)$. Once we know a_j, we can solve the corresponding system of equations $f_j(c_1, \ldots, c_\ell) = a_j$. This system is non-linear, but when the number of parameters is small, it is not that difficult to solve.

Scale-invariant case: analysis of the problem. As we have shown earlier, the scale-invariant case reduces to the shift-invariant case if we introduce an auxiliary variable $\tau = \ln(t - t_0)$. Thus, similarly to the above-described shift-invariant case, with respect to this new variable τ, we get a differential equation

$$\frac{d^m x}{d\tau^m} + a_1 \cdot \frac{d^{m-1} x}{d\tau^{m-1}} + \ldots + a_m \cdot x = 0. \tag{10}$$

Differentiating the relation between τ and t, we conclude that $d\tau = \dfrac{dt}{t - t_0}$. Thus, $\dfrac{d}{d\tau} = (t - t_0) \cdot \dfrac{d}{dt}$, and the equation (1) takes the following form:

$$(t - t_0)^m \cdot \frac{d^m x}{dt^m} + a_1 \cdot (t - t_0)^{m-1} \cdot \frac{d^{m-1} x}{dt^{m-1}} + \ldots + a_m \cdot x = 0. \tag{11}$$

There are two possibilities:

- it may be that we know t_0, or
- it may be that we need to determine t_0 from observations.

In the first subcase, all we need is to find the values a_j.

In the second subcase, to make the problem linear, we expand all the polynomials

$$(t-t_0)^j = x^j + (-j\cdot t_0)\cdot t^{j-1} + \ldots,$$

then each term $a_j\cdot(t-t_0)^{m-j}\cdot\dfrac{d^{m-j}x}{dt^{m-j}}$ becomes a linear combination of the following terms:

$$t^{m-j}\cdot\frac{d^{m-j}x}{dt^{m-j}},\ \ t^{m-j-1}\cdot\frac{d^{m-j}x}{dt^{m-j}},\ \ \ldots,\ \ \frac{d^{m-j}x}{dt^{m-j}}.$$

Let us denote the coefficients at $t^{m-j-k}\cdot\dfrac{dx^{m-j}}{dt^{m-j}}$ by a_{jk}. Then, the formula (11) takes the following form:

$$\begin{gathered}
t^m\cdot\frac{dx^m}{dt^m} + a_{01}\cdot t^{m-1}\cdot\frac{dx^m}{dt^m} + \ldots + a_{0m}\cdot\frac{dx^m}{dt^m} + \\
a_{10}\cdot t^{m-1}\cdot\frac{dx^{m-1}}{dt^{m-1}} + a_{11}\cdot t^{m-2}\cdot\frac{dx^{m-1}}{dt^{m-1}} + \ldots + a_{1,m-1}\cdot\frac{dx^{m-1}}{dt^{m-1}} + \\
\ldots + \\
a_{m0}\cdot x = 0.
\end{gathered} \tag{12}$$

Thus, depending on whether we know t_0 or we don't, we arrive at the following linear algorithms.

Scale-invariant case: resulting algorithms. Based on the original sequence of observations $x_i = x(t_i)$, we compute the finite differences $(\Delta^k x)_i$ for all possible values $k \le m$.

Then, if we know the value t_0, we compute the parameters $a_1,\ldots,a_m$ of the corresponding model by applying the Least Squares method to the following system of linear equations:

$$(t_i-t_0)^m\cdot(\Delta^m x)_i + a_1\cdot(t_i-t_0)^{m-1}\cdot(\Delta^{m-1}x)_i + \ldots + a_m\cdot x_i = 0. \tag{13}$$

When we do not know the value t_0, then we need to find the parameters a_{jk} of the model by applying the Least Squares method to the following system of equations:

$$\begin{gathered}
t_i^m\cdot(\Delta^m x)_i + a_{01}\cdot t_i^{m-1}\cdot(\Delta^m x)_i + \ldots + a_{0m}\cdot(\Delta^m x)_i + \\
a_{10}\cdot t_i^{m-1}\cdot(\Delta^{m-1}x)_i + a_{11}\cdot t^{m-2}\cdot(\Delta^{m-1}x)_i + \ldots + a_{1,m-1}\cdot(\Delta^{m-1}x)_i + \\
\ldots + \\
a_{m0}\cdot x = 0.
\end{gathered} \tag{14}$$

Acknowledgments

We acknowledge the partial support of the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Thailand. This work was also supported in part by the National Science Foundation grant HRD-1242122 (Cyber-ShARE Center of Excellence).

References

1. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.
2. J. A. Feigenbaum, "A statistical analyses of log-periodic precursors to financial crashes", *Quantitative Finance*, 2001, Vol. 1, No. 5, pp. 527–532.
3. J. A. Feigenbaum and P. Freund, "Discrete scaling in stock markets before crashes", *International Journal of Modern Physics*, 1996, Vol. 12, pp. 57–60.
4. L. Gazola, C. Fenandez, A. Pizzinga, and R. Riera, "The log-periodic-AR(1)-GARCH(1,1) model for financial crashes", *The European Physical Journal B*, 2008, Vol. 61, No. 3, pp. 355–362.
5. P. Geraskin and D. Fantazzinin, "Everything you always wanted to know about log periodic power laws for bubble modelling but were afraid to ask", *European Journal of Finance*, 2013, Vol. 19, No. 5, pp. 366–391.
6. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Leaning*, MIT Press, Cambridge, Massachusetts, 2016.
7. Z. Q. Jiang, W. H. Zhou, D. Sornette, R. Woodard, K. Bastiaensen, and P. Cauwels, "Bubble diagnosis and prediction of the 2005–2007 and 2008–2009 Chinese stock market bubbles", *Journal of Economic Behavior and Organization*, 2010, Vol. 74, pp. 149–162.
8. A. Johansen, "Characterization of large price variations in financial markets", *Physica A*, 2003, Vol. 324, pp. 157–166.
9. A. Johansen, O. Ledoit, and D. Sornette, "Crashes as critical points", *International Journal of Theoretical and Applied Finance*, 2000, Vol. 3, No. 2, pp. 219–255.
10. A. Johansen and D. Sornette, "Financial anti-bubbles: Log-periodicity in Gold and Nikkei collapses", *International Journal of Modern Physics C*, 1999, Vol. 10, No. 4, pp. 563–575.
11. A. Johansen and D. Sornette, "Large Stock market price drawdowns are outliers", *The Journal of Risk*, 2002, Vol. 4, No. 2, pp. 69–110.
12. A. Johansen and D. Sornette, "Endogenous versus exogenous crashes in financial markets", in: *Contemporary Issues in International Finance*, Nova Science Publishers, 2004; reprinted as a special issue of *Brussels Economic Review*, 2006, Vol. 49, No. 3/4.
13. V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational Complexity and Feasibility of Data Processing and Interval Computations*, Kluwer, Dordrecht, 1998.
14. V. Kreinovich, H. T. Nguyen, and S. Sriboonchitta, "Log-periodic power law as a predictor of catastrophic events: a new mathematical justification", *Proceedings of the International Conference on Risk Analysis in Meteorological Disasters RAMD'2014*, Nanjing, China, October 12–13, 2014.
15. V. Kreinovich, T. Swenson, and A. Elentukh, "Interval approach to testing software", *Interval Computations*, 1994, No. 2, pp. 90–109.
16. H. T. Nguyen and V. Kreinovich, *Applications of Continuous Mathematics to Computer Science*, Kluwer, Dordrecht, 1997.
17. C. H. Papadimitriou, *Computational Complexity*, Pearson, Boston, Massachusetts, 1993.
18. P. Pardalos, *Complexity in Numerical Optimization*, World Scientific, Singapore, 1993.
19. D. Poole, *Linear Algebra: A Modern Introduction*, Cengage Learning, Independence Kentucky, 2014.

20. J. C. Robinson, *An Introduction to Ordinary Differential Equations*, Cambridge University Press, Cambridge, UK, 2004.
21. D. Sornette, "Critical market crashes", *Physics Reports*, 2003, Vol. 378, No. 1, pp. 1–98.
22. D. Sornette, *Why Stock Markets Crash: Critical Events in Compelx Financial Systems*, Princeton University Press, Princeton, New Jersey, 2003.
23. D. Sornette and A. Johansen, "Significance of log-periodic precursors to financial crashes", *Quantitative Finance*, 2001, Vol. 1, No. 4, pp. 452–471.
24. D. Sornette and W. Z. Zhou, "The US 2000–2002 market descent: how much longer and deeper?" *Quantitative Finance 2*, Vol. 2, No. 6, pp. 468–481.
25. J. M. Weatherall, *The Physics of Wall Street: the History of Predicting the Unpredictable*, Houghton Mifflin Harcourt, New York, 2013.
26. W. Z. Zhou and D. Sornette, "Evidence of a worldwide stock market log-periodic anti-bubble since mid-2000", *Physics A*, 2003, Vol. 330, No. 3, pp. 543–583.