

American Journal of Economics and Business Management

Vol. 9 Issue 10 | **pp.** 222-235 | **ISSN:** 2576-5973

Available online @ https://www.globalresearchnetwork.us/index.php/ajebm

Article

Identifying Functional Relationships and Reconstructing Parameters from Experimental Data

Sherali Baratovich Ochilov[1], Nargiza Rozikovna Khalliyeva[2]

1. Associate Professor of Department of Economics and Management, Bukhara State Technical University
2. PhD, Associate Professor of Department of Economics and Management, Bukhara State Technical University

* Correspondence: -

Abstract: Turning measurements into a quantitative law involves two linked problems: choosing the structure of the functional relationship and estimating (reconstructing) its parameters together with their uncertainty. In practice the two are often treated separately, which leads to over-fitted empirical formulas, parameters that cannot be determined from the available data, and unstable solutions of ill-posed inverse problems. This paper proposes and tests an integrated workflow that combines information-theoretic structure selection (AICc, BIC, Akaike weights), cross-validation and residual diagnostics with nonlinear least-squares estimation by the Levenberg–Marquardt method, local and bootstrap uncertainty quantification, identifiability analysis based on the conditioning of the Fisher information matrix, and Tikhonov regularisation for linear ill-posed problems. The workflow is evaluated in four controlled numerical experiments with known ground truth: selection among six candidate laws, a Monte Carlo study of estimator accuracy versus noise level and experimental design, reconstruction of the parameters of a damped oscillator from a noisy trajectory, and deconvolution of a Gaussian-blurred signal. The exponential law was recovered with an Akaike weight of 0.947 although a cubic polynomial fitted almost equally well (R2 = 0.9925 vs 0.9931). At 5 % noise, the relative RMSE of the parameters stayed below 4.2 %, while shortening the observation window from 12 to 3 units increased the condition number of JTJ from 127 to 1.27·104 and the error of the asymptote from 3.2 % to 37.8 %. Oscillator parameters were recovered to within 0.3 % (ω) and 1.9 % (ζ) from all six starting points; Tikhonov regularisation reduced the reconstruction error of an operator with condition number 3.8·1011 from 6·108 to 3.4 %. The results show that the experimental design, not the fitting algorithm, is usually the limiting factor.

Keywords: functional dependence; system identification; inverse problem; nonlinear least squares; Levenberg–Marquardt method; model selection; Akaike information criterion; identifiability; bootstrap; Tikhonov regularisation

Citation: Ochilov, S. B & Khalliyeva, N. R. Identifying Functional Relationships and Reconstructing Parameters from Experimental Data. American Journal of Economics and Business Management 2026, 9(10), 222-235

Received: 13th Jul 2026
Revised: 28th Aug 2026
Accepted: 15th Sep 2026
Published: 08th Oct 2026

Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

Almost every empirical science reaches a point at which a table of measurements must be converted into a formula. A chemist estimates a rate constant from a concentration curve, an engineer recovers stiffness and damping from a vibration record, an economist fits a growth or saturation law to a time series, and a physician estimates the elimination half-life of a drug from blood samples. In each case two questions have to be answered: *which* functional relationship links the variables, and *what values* of its parameters are consistent with the observations. The first is a problem of structural identification; the second is a problem of parameter reconstruction, which in its general form is an inverse problem.[1][2]

The two questions are tightly coupled. A wrongly chosen structure produces biased parameters even when the fit looks excellent, while a correct structure can still yield meaningless parameters if the data do not contain enough information to determine them. Hadamard's classical definition of a well-posed problem — existence, uniqueness and continuous dependence of the solution on the data — is regularly violated in practice: many parameter-estimation problems are ill-conditioned, and small measurement errors are amplified into large errors in the reconstructed quantities.[3][4]

The methodological literature offers mature tools for each stage taken separately. Least-squares estimation goes back to Gauss and Legendre, and its nonlinear version is routinely solved by the damped Gauss–Newton (Levenberg–Marquardt) method. Model selection is commonly based on the Akaike and Schwarz criteria and their small-sample correction, with a systematic treatment of multi-model inference given by Burnham and Anderson. Cross-validation offers a criterion that does not rely on likelihood assumptions. Uncertainty of nonlinear estimates is assessed through the asymptotic covariance matrix or by resampling. The concept of identifiability, introduced by Bellman and Åström, has been developed into practical tools such as the profile likelihood, and the observation that many multi-parameter models are «sloppy», with eigenvalues of the Fisher information spread over many orders of magnitude, has changed how parameter uncertainty is interpreted. For linear ill-posed problems, Tikhonov regularisation with the regularisation parameter chosen by the discrepancy principle or the L-curve remains the standard. More recently, sparse regression, symbolic regression and physics-informed neural networks have extended the search for structure to large libraries of candidate terms.[5][6]

Despite this richness, applied studies — including many dissertations and articles in economics, engineering and the life sciences — frequently stop after maximising R^2 for a single chosen formula. Three typical shortcomings follow: (i) the structure is selected by goodness of fit alone, so that flexible polynomials win over mechanistically meaningful laws; (ii) parameters are reported without uncertainty or without checking whether they are identifiable from the data; and (iii) the role of the experimental design (the range and density of the sampled variable) is ignored.

The aim of this study is to formulate a compact, reproducible workflow that treats structure selection, parameter reconstruction, identifiability and regularisation as one procedure, and to quantify its performance in controlled experiments with known ground truth. The specific objectives are: (1) to compare information criteria, cross-validation and goodness-of-fit measures as tools for recovering the true functional form; (2) to measure the accuracy of nonlinear least-squares reconstruction as a function of noise level and observation window, and relate it to the conditioning of the information matrix; (3) to test the procedure on a dynamic model given by an ordinary differential equation; and (4) to demonstrate the necessity and the effect of regularisation for an ill-posed linear inverse problem.[7]

2. Materials and Methods

2.1 Problem statement

Let the experiment produce n pairs (x_i, y_i), $i = 1, \ldots, n$, where x is the controlled variable and y the measured response. The observation model is

$$y_i = f(x_i; \theta) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \qquad (1)$$

where f belongs to a finite set of candidate structures $\mathcal{M} = \{f_1, \ldots, f_M\}$ and $\theta \in \mathbb{R}^p$ is the parameter vector. The task is to select f from $\mathcal{M}$, to estimate θ and to quantify the uncertainty of the estimate. The overall workflow is shown in Figure 1.

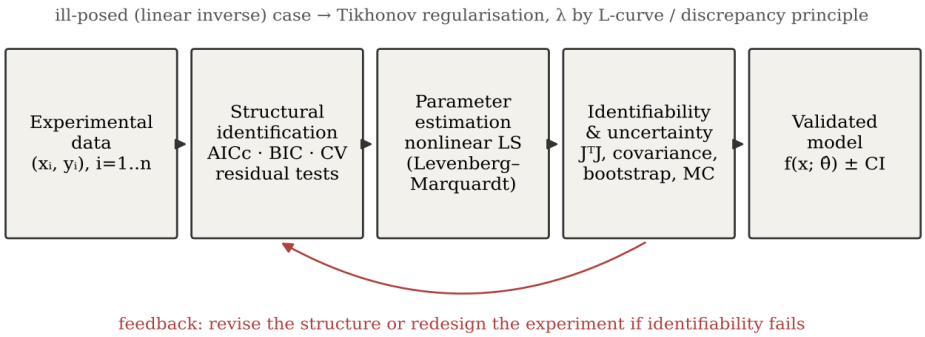


Figure 1. Integrated workflow for identifying a functional relationship and reconstructing its parameters

2.2 Parameter estimation

For a fixed structure, the parameters are found by minimising the residual sum of squares

$$S(\boldsymbol{\theta}) = \Sigma_i\,[y_i - f(x_i;\,\boldsymbol{\theta})]^2 \to \min, \qquad (2)$$

which coincides with the maximum-likelihood estimate under Gaussian errors. The minimisation uses the Levenberg–Marquardt iteration

$$(\mathbf{J}^{\mathrm{T}}\mathbf{J} + \mu_k \operatorname{diag} \mathbf{J}^{\mathrm{T}}\mathbf{J})\,\Delta\boldsymbol{\theta} = \mathbf{J}^{\mathrm{T}}\mathbf{r}, \quad \boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \Delta\boldsymbol{\theta}, \qquad (3)$$

where $\mathbf{J}$ is the $n\times p$ Jacobian $\partial f(x_i;\,\boldsymbol{\theta})/\partial\theta_j$, $\mathbf{r}$ is the residual vector and μ_k is a damping parameter adapted at each step: large μ gives a short gradient-descent step, small μ gives a Gauss–Newton step. For models defined by differential equations, f is evaluated by numerical integration (an explicit Runge–Kutta 4(5) scheme with relative tolerance 10^{-9}), and the Jacobian is obtained by finite differences. Because nonlinear problems may possess local minima, the optimisation is started from several dispersed initial points (multi-start), and the solution with the smallest S is retained.[8][9]

2.3 Selecting the structure

Candidate structures are compared by the corrected Akaike criterion and the Bayesian (Schwarz) criterion

$$\mathrm{AICc} = n\ln(S/n) + 2k + 2k(k+1)/(n-k-1), \quad \mathrm{BIC} = n\ln(S/n) + k\ln n, \qquad (4)$$

where $k = p + 1$ counts the parameters including σ^2. Differences $\Delta_m = \mathrm{AICc}_m - \min \mathrm{AICc}$ are converted into Akaike weights $w_m = \exp(-\Delta_m/2) / \Sigma \exp(-\Delta_j/2)$, interpretable as the relative support for each model within $\mathcal{M}$. As a likelihood-free check, the root-mean-square prediction error is estimated by 5-fold cross-validation (CV-RMSE). Finally, the residuals of each fitted model are examined for systematic structure using the Durbin–Watson statistic $\mathrm{DW} = \Sigma(r_i - r_{i-1})^2 / \Sigma r_i^2$: values well below 2 indicate serial correlation of residuals, i.e. a mis-specified functional form.[10][11]

2.4 Uncertainty and identifiability

The asymptotic covariance of the estimate is

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \approx \hat{\sigma}^2\,(\mathbf{J}^{\mathrm{T}}\mathbf{J})^{-1}, \quad \hat{\sigma}^2 = S(\hat{\boldsymbol{\theta}})/(n-p), \qquad (5)$$

from which standard errors and the parameter correlation matrix follow. Because (5) relies on local linearisation, it is complemented by a residual bootstrap with $B = 2000$ replicates: residuals are resampled with replacement, added to the fitted curve and the model is re-estimated; percentile intervals give 95 % confidence bounds. Practical identifiability is diagnosed by the condition number $\kappa(\mathbf{J}^{\mathrm{T}}\mathbf{J}) = \lambda_{\max}/\lambda_{\min}$: when κ grows by orders of magnitude, some combination of parameters is weakly determined by the data, and small noise produces large parameter errors. To separate the effect of noise from that of the experimental design, Monte Carlo experiments with 1000 (noise study) and 600 (design study) independent noise realisations were performed, and accuracy was summarised by the relative bias and the relative root-mean-square error (rRMSE) of each parameter.

2.5 Regularisation of ill-posed linear problems

When the unknown is a function rather than a few parameters, discretisation leads to a linear system $\mathbf{K}\,\mathbf{f} = \mathbf{g}$ with a severely ill-conditioned matrix $\mathbf{K}$. The Tikhonov-regularised solution is

$$\mathbf{f}_\lambda = \operatorname{argmin} \{ \|\mathbf{K}\,\mathbf{f} - \mathbf{g}\|^2 + \lambda \|\mathbf{D}\,\mathbf{f}\|^2 \} = (\mathbf{K}^T\mathbf{K} + \lambda \mathbf{D}^T\mathbf{D})^{-1}\,\mathbf{K}^T\mathbf{g}, \qquad (6)$$

where $\mathbf{D}$ is the first-difference operator, which penalises roughness. The regularisation parameter λ is chosen by two independent rules: the corner (point of maximum curvature) of the L-curve, i.e. the plot of $\|\mathbf{D}\,\mathbf{f}_\lambda\|$ against $\|\mathbf{K}\,\mathbf{f}_\lambda - \mathbf{g}\|$ in log–log scale, and Morozov's discrepancy principle, which selects λ such that the residual norm equals the expected noise norm $\delta = \sigma\sqrt{n}$.

2.6 Design of the numerical experiments

To evaluate the procedure objectively, the data were generated in controlled numerical experiments in which the true structure and parameters are known; real measurements were imitated by adding independent Gaussian noise. This «synthetic experiment» design is the standard way to validate identification methods, because only with a known ground truth can bias and error be measured. Four experiments were carried out:

E1 (structure). $n = 40$ equally spaced points on generated by $y = a\,e^{-bx} + c$ with $a = 4.0$, $b = 0.35$, $c = 1.2$; noise $\sigma = 5$ % of the mean signal ($\sigma = 0.108$). Six candidates were fitted: linear, quadratic, cubic, power $a(x+1)^{-b} + c$, exponential and rational $(a + cbx)/(1 + bx)$.

E2 (accuracy and design). The exponential model of E1 at noise levels of 1, 3, 5, 10 and 15 %, and with observation windows $[0, x_{max}]$, $x_{max} \in \{3, 6, 9, 12, 18\}$, with $n = 40$ and 5 % noise.

E3 (dynamic model). The damped oscillator $\ddot{x} + 2\zeta\omega\dot{x} + \omega^2 x = 0$ with $\omega = 2.0$, $\zeta = 0.15$, $x(0) = 1$, $\dot{x}(0) = 0$, observed at 120 points on s with noise $\sigma = 0.04$; four parameters (ω, ζ, x_0, v_0) were reconstructed from six starting points.

E4 (ill-posed problem). Deconvolution on [0, 1] with $N = 80$ nodes: $\mathbf{K}$ is a Gaussian blurring kernel of width 0.03, the true signal is the sum of two Gaussian peaks, and noise equals 1 % of max g.

All computations were performed in Python 3 with NumPy and SciPy (functions *curve_fit*, *least_squares* and *solve_ivp*) using a fixed random seed, so all results are exactly reproducible.[12]

3. Results

In natural sciences, engineering, economics, and many other fields, the researcher often receives not a ready mathematical law but a set of experimental or observational points:

$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

The main question is: what functional law connects these points?

The relationship may be linear, quadratic, exponential, power-type, harmonic, or more complicated. In many applications the type of model is chosen first, and only afterwards are the coefficients fitted. Yet this strategy has a natural limitation. Over a short interval, different functions may look almost identical. A short arc of a sine function can appear nearly linear, and a segment of an exponential curve can be approximated well by a line or a quadratic polynomial. Therefore, the statement "this function fits the points well" is not equivalent to the statement "this is the true functional structure of the data." In the present approach, two tasks are separated: first, identification of the structure; second, estimation of the parameters. The central principle is: identify the structure first, then determine the parameters.[13]

Mathematical statement of the problem

Assume that the experimental observations are:

$(x_i, y_i), \quad i = 1, 2, \ldots, n$

and that the analytical form of the function is unknown.

The aim is to identify the functional class, estimate the principal parameters, separate different local regimes when necessary, construct a differential equation from the detected local relation, and reconstruct the function by solving that equation.

The first step is to evaluate the local slope between neighboring points:

$d_i = (y_{i+1} - y_i)/(x_{i+1} - x_i)$

This quantity is a finite-difference estimate of the derivative.

In this work, the term local differential invariant means a differential combination that, for a given function class, remains constant in the ideal case and expresses a characteristic parameter of that class.

Telescoping property of finite differences

Define the first finite difference by:

$\Delta y_i = y_{i+1} - y_i$

Summing all first differences gives:

$\Sigma_{i=1}^{n-1} \Delta y_i = y_n - y_1$

because the intermediate terms cancel pairwise. This identity is independent of the function class; it is an algebraic property of the finite-difference operator.

If the step in x is constant, $x_{i+1} - x_i = h$, then $d_i = \Delta y_i/h$ and the mean local slope is:

$\bar{d} = (y_n - y_1)/(x_n - x_1)$

This is important because averaging can hide local structure. The sequences 1, 3, 1, 3 and 2, 2, 2, 2 have the same mean value, 2, but the first oscillates while the second is locally constant. Hence local values must be examined rather than only their average.[14][15]

Identification of a linear function

Let the model be:

$y = ax + b$

Its first derivative is:

$y' = a$

Thus, a constant local slope is the characteristic signature of a linear function.

Example. Let x = 0, 1, 2, 3, 4 and y = 3, 5, 7, 9, 11. The local slopes are 2, 2, 2, 2. Hence a = 2. Using the point x = 0, y = 3 gives b = 3, and therefore:

$y = 2x + 3$

Two-stage parameter estimation

In real data the local slopes are not necessarily identical. Suppose they are 2.1, 1.8, 2.3, and 1.8. In the first stage, the slope parameter is found by minimizing:

$S_1(a) = \Sigma(d_i - a)^2 \rightarrow \min$

The minimizer is the mean local slope.

In the second stage, after a has been estimated, the intercept b is obtained from the original observations by minimizing:

$S_2(b) = \Sigma[y_i - (ax_i + b)]^2 \rightarrow \min$

This leads to:

$b = (1/n)\Sigma(y_i - ax_i)$

The first stage therefore uses local differential information, whereas the second uses the original data.

Quadratic function

Consider:

$y = ax^2 + bx + c$

Its first derivative is $y' = 2ax + b$, which is not constant. The second derivative is:

$y'' = 2a$

Thus, constancy of the second derivative is the characteristic signature of a quadratic function.

For equally spaced data:

$\Delta^2 y = 2ah^2$

and therefore:

$a = \Delta^2 y/(2h^2)$

Example: $y = x^2 + 2x + 1$ at x = 0, 1, 2, 3, 4 gives y = 1, 4, 9, 16, 25. The first differences are 3, 5, 7, 9; the second differences are 2, 2, 2. Hence a = 1. The remaining coefficients are b = 2 and c = 1.

Polynomial of degree n

A general polynomial has the form:

$y = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$

After n differentiations:

$y^{(n)} = n! \cdot a_n$

For equally spaced points:

$\Delta^n y = n! \cdot a_n \cdot h^n$

so that:

$a_n = \Delta^n y/(n! \cdot h^n)$

In ideal, noise-free data, the first order at which a constant finite difference appears indicates the polynomial degree.

Cubic example and sequential coefficient recovery

Take x = 0, 1, 2, 3, 4 and y = 5, 14, 41, 98, 197. The first differences are 9, 27, 57, 99; the second differences are 18, 30, 42; the third differences are 12, 12.

Hence the leading coefficient is $a_3 = 12/6 = 2$.

Subtracting $2x^3$ from the original values leaves 5, 12, 25, 44, 69. The second differences of this remainder are 6, 6, 6, so $a_2 = 3$. Subtracting $3x^2$ leaves 5, 9, 13, 17, 21, whose first differences are 4, 4, 4, 4. Thus $a_1 = 4$, and from x = 0 we obtain $a_0 = 5$.

The reconstructed polynomial is:

$y = 2x^3 + 3x^2 + 4x + 5$

The derivative need not be constant

The method should not be restricted to y′ = constant or y″ = constant. More generally, the experimental data may reveal relations of the form:

$y' = f(x)$

$y'' = f(x)$

$y^{(k)} = F(x, y, y', \ldots)$

For example, if the local calculations suggest $y' \approx 2x + 3$, one adopts the differential relation $y' = 2x + 3$ and integrates to obtain:

$y = x^2 + 3x + C$

Similarly, if $y'' = 6x + 4$, then:

$y' = 3x^2 + 4x + C_1$

$y = x^3 + 2x^2 + C_1x + C_2$

Thus, the general idea is not merely to find a constant derivative, but to identify the functional law obeyed by the derivative.

Exponential function

For:

$y = A \cdot e^{(\beta x)}$

we have:

$y' = \beta y$

and therefore:

$y'/y = \beta$

If this ratio is locally constant, the differential equation $y' = \beta y$ is obtained. Separation of variables gives $dy/y = \beta dx$, and integration yields:

$y = A \cdot e^{(\beta x)}$

Exact discrete invariant for the exponential function

For two neighboring points of the exponential law:

$y_i = A \cdot e^{(\beta x_i)}$

$y_{i+1} = A \cdot e^{(\beta x_{i+1})}$

taking the ratio and logarithm gives:

$\beta_i = \ln(y_{i+1}/y_i)/(x_{i+1} - x_i)$

For ideal exponential data, all β_i values are identical. This is an exact discrete relation and does not require numerical differentiation.

Power function

For:

$y = A \cdot x^{\alpha}$

we have:

$y' = A\alpha x^{(\alpha-1)}$

and hence:

$x \cdot y'/y = \alpha$

If this relation is constant, then $x \cdot y' = \alpha y$. Separation of variables yields $dy/y = \alpha dx/x$ and finally:

$y = A \cdot x^{\alpha}$

Exact discrete invariant for the power function

For two neighboring points:

$y_{i+1}/y_i = (x_{i+1}/x_i)^{\alpha}$

Hence:

$\alpha_i = \ln(y_{i+1}/y_i)/\ln(x_{i+1}/x_i)$

For ideal power-law data, all α_i are identical. Direct application in the real domain usually requires positive x_i and y_i.

Harmonic function

Consider:

$y = A \cdot \sin(\omega x + \varphi)$

It satisfies:

$y'' + \omega^2 y = 0$

Using the ratio y''/y is inconvenient near zero crossings. Therefore let $q = \omega^2$ and determine q from:

$S(q) = \Sigma(d^2_i + q y_i)^2 \rightarrow \min$

The least-squares minimizer is:

$q = -\Sigma(d^2_i y_i)/\Sigma(y_i^2)$

If $q > 0$, then $\omega = \sqrt{q}$. The remaining coefficients are obtained from:

$y = C_1 \cos(\omega x) + C_2 \sin(\omega x)$

Multivariable linear function

Consider the model:

$y = a_1 x_1 + a_2 x_2 + \ldots + a_p x_p + b$

The partial derivatives are:

$\partial y/\partial x_j = a_j$

Thus each coefficient can be estimated along its own local direction, provided that the effects of the other variables are controlled or mathematically separated.

For each j:

$S_j(a_j) = \Sigma(D_{ij} - a_j)^2 \rightarrow \min$

and for equal weighting:

$a_j = (1/n)\Sigma D_{ij}$

After all a_j are estimated, the intercept can be obtained from:

$b = (1/n)\Sigma[y_i - (a_1 x_{i1} + \ldots + a_p x_{ip})]$

Multivariable power function

Consider:

$y = A \cdot x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot \ldots \cdot x_p^{\alpha_p}$

For each variable:

$\partial y/\partial x_j = \alpha_j \cdot y/x_j$

Hence:

$\alpha_j = (x_j/y) \cdot \partial y/\partial x_j$

From data, if D_{ij} approximates the corresponding partial derivative, α_j may be estimated from:

$S_j(\alpha_j) = \Sigma[D_{ij} - \alpha_j(y_i/x_{ij})]^2 \rightarrow \min$

$\alpha_j = \Sigma[D_{ij}(y_i/x_{ij})] / \Sigma[(y_i/x_{ij})^2]$

Each α_j remains attached to its own variable. Replacing several different exponents by a single geometric mean is generally incorrect.

After all exponents are found, define:

$g_i = x_{i1}^{\alpha_1} \cdot x_{i2}^{\alpha_2} \cdot \ldots \cdot x_{ip}^{\alpha_p}$

Then the scale parameter is:

$A = \Sigma(y_i g_i)/\Sigma(g_i^2)$

Decomposition and composition principle

For multivariable models, the general logic is to separate the local contribution of each variable, estimate the corresponding parameter, and only then combine the parameters

into the full model. This may be summarized as: decomposition → parameter estimation → composition.

Local stability criterion

In real data a local derivative or invariant need not be exactly constant. A relative stability criterion may therefore be introduced:

$|I_i - \bar{I}|/|\bar{I}| \le \varepsilon$

Here I_i is the local invariant, $\bar{I}$ is its characteristic or mean value, and ε is a tolerance selected by the researcher. For example, $\varepsilon = 0.05$ represents a five-percent tolerance and $\varepsilon = 0.10$ a ten-percent tolerance. These values are not universal mathematical constants; they are admissible local-variation thresholds.

Relation between derivative fluctuation and function deviation

Assume:

$y'(x) = \bar{d} + e(x)$

where $\bar{d}$ is the characteristic derivative and $e(x)$ is the deviation. If:

$|e(x)| \le \varepsilon|\bar{d}|$

then:

$y(x) = y(a) + \bar{d}(x-a) + \int_a^x e(t)dt$

Define the main trend:

$y_0(x) = y(a) + \bar{d}(x-a)$

Then the deviation $r(x) = y(x) - y_0(x)$ satisfies:

$|r(x)| \le \varepsilon|\bar{d}|(x-a)$

Hence a relative bound on derivative fluctuation yields a corresponding bound on the deviation of the function from the main trend.

Stability in the exponential case

If:

$y'/y = \beta + e(x)$

and $|e(x)| \le \varepsilon|\beta|$, then:

$y(x) = y(a)e^{[\beta(x-a)]}\cdot e^{[\int_a^x e(t)dt]}$

For the ideal model $y_0(x) = y(a)e^{[\beta(x-a)]}$, we obtain:

$y(x)/y_0(x) = e^{[\int_a^x e(t)dt]}$

If $L = x - a$, then:

$e^{(-\varepsilon|\beta|L)} \le y(x)/y_0(x) \le e^{(\varepsilon|\beta|L)}$

This shows that the same local tolerance ε can have different global effects for different function classes and interval lengths.

Curie law as a methodological example

The ideal Curie law is:

$\chi = C/T$

It is a special case of a power law with exponent $\alpha = -1$. Hence:

$T\cdot\chi'/\chi = -1$

which gives $T\chi' = -\chi$. Separation of variables produces $d\chi/\chi = -dT/T$ and integration restores:

$\chi = C/T$

Thus the Curie law can be reconstructed from the local differential relation.

Kapitza-type example and local regimes

Real experimental data need not obey a single law over the whole interval. In the previously examined Kapitza-type example, local analysis separated several regimes: 0.40–0.48 approximately linear; 0.49–1.25 exponential; 1.26–1.96 exponential with a different parameter; 1.97–2.29 approximately linear.

In the first exponential interval, $\beta \approx 2.7$, giving $y \approx A\cdot e^{(2.7x)}$. In the second, $\beta \approx 5.4$, giving $y \approx A\cdot e^{(5.4x)}$. The function class remains exponential, but the local parameter changes. This illustrates how the method can reveal regime changes.

General algorithm. The differential-invariant method can be summarized as follows:

1) acquire experimental data; 2) compute local finite differences; 3) test simple signatures; 4) test higher-order polynomial signatures; 5) use an exact discrete invariant when available; 6) otherwise test differential invariants; 7) construct the differential equation; 8) reconstruct the function class; 9) estimate structural parameters; 10) estimate

remaining parameters from original data; 11) decompose multivariable effects; 12) compose the final model; 13) test local stability with ε; 14) separate local regimes when necessary.[16]

Relation to least squares. The differential-invariant method is not opposed to least squares. Their roles are different. Differential-invariant analysis is primarily used to identify the functional structure, while least squares may be used to refine the parameters of the identified model. In this sense, the two approaches are complementary rather than competing. The principal feature of the method is that the functional structure is inferred from the local law of change rather than imposed in advance. Two information sources are used: local differential information and the original function values. Local information supports structural identification; the original data support parameter estimation and model verification.

The approach is also broader than the search for constant derivatives. The derivative itself may follow a nonconstant law, such as $y' = f(x)$, $y'' = f(x)$, or, more generally, $y^{(k)} = F(x, y, y', \ldots)$. Identifying that law from data and solving the resulting differential equation is the broader form of the method.

Identification of the functional form (E1). Table 1 summarises the six candidate models. Goodness of fit alone does not discriminate clearly: the cubic polynomial ($R^2 = 0.9925$) is almost indistinguishable from the true exponential law ($R^2 = 0.9931$), and both are well above 0.97 for the quadratic and power laws.

Table 1. Comparison of candidate structures for the data of experiment E1 ($n = 40$)

Model	p	R^2	RMSE	AICc	ΔAICc	w_{AIC}	BIC	CV-RMSE	DW
Linear	2	0.7760	0.498	-49.0	136.7	$<10^{-4}$	-44.6	0.539	0.12
Quadratic	3	0.9716	0.177	-129.2	56.5	$<10^{-4}$	-123.5	0.196	0.67
Power	3	0.9810	0.145	-145.2	40.5	$<10^{-4}$	-139.6	0.196	1.06
Rational	3	0.9871	0.120	-160.7	25.0	$<10^{-4}$	-155.0	0.140	1.42
Cubic	4	0.9925	0.091	-179.9	5.8	0.053	-173.3	0.105	2.33
Exponential	3	0.9931	0.088	-185.7	0.0	0.947	-180.1	0.092	2.50

Note: p — number of regression parameters; w_{AIC} — Akaike weight; CV-RMSE — 5-fold cross-validation error; DW — Durbin–Watson statistic. Models are ordered by increasing AICc support.

The information criteria, by contrast, are decisive. The exponential model has the minimum AICc (-185.7) and BIC, and receives an Akaike weight of 0.947; the cubic polynomial, despite its similar R^2, is penalised for its extra parameter ($\Delta AICc = 5.8$, $w = 0.053$). The cross-validation error confirms the ranking (0.092 vs 0.105), and its value is close to the true noise level $\sigma = 0.108$, as expected for a correctly specified model. The residual diagnostics expose mis-specification directly: the linear (DW = 0.12), quadratic (0.67) and power (1.06) laws leave strongly autocorrelated residuals, whereas the exponential and cubic models give DW ≈ 2.3–2.5, consistent with white noise (Figure 2b). Thus residual tests rule out the clearly wrong structures, and the information criterion then chooses between the remaining adequate ones on the basis of parsimony.

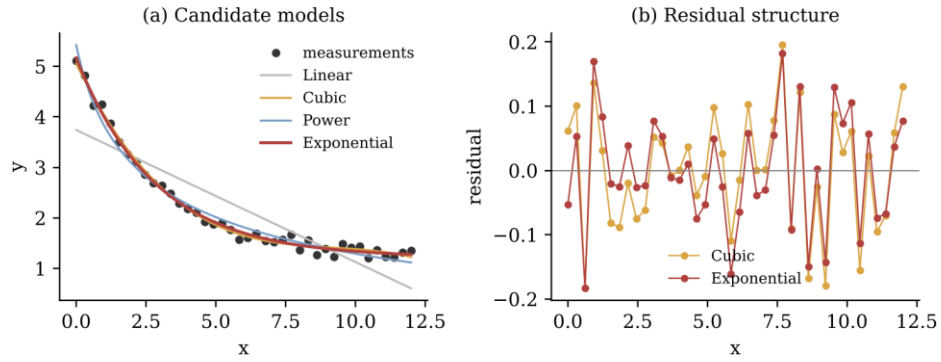


Figure 2. Experiment E1: (a) data and fitted candidate models; (b) residuals of the cubic and exponential models

Accuracy of parameter reconstruction (E2)

For the selected exponential model, the Levenberg–Marquardt estimate from the E1 data is $a = 3.949 \pm 0.055$, $b = 0.3500 \pm 0.0122$, $c = 1.219 \pm 0.033$ (estimate ± standard error from (5)); the true values 4.0, 0.35 and 1.2 lie inside all intervals. The bootstrap 95 % intervals — [3.847; 4.055] for a, [0.328; 0.374] for b and [1.153; 1.279] for c — agree closely with ±1.96 standard errors, which shows that the linear approximation (5) is adequate here. The condition number of $\mathbf{J}^{\mathrm{T}}\mathbf{J}$ is 124, and the strongest parameter correlation is between the decay rate b and the asymptote c (r = 0.81), visible as the elongated cloud of bootstrap estimates in Figure 3b: a faster decay can be partly compensated by a higher asymptote.

Table 2. Monte Carlo accuracy of the reconstructed parameters versus noise level (1000 replicates, x ∈ [0, 12])

Noise, %	Bias a, %	Bias b, %	Bias c, %	rRMSE a, %	rRMSE b, %	rRMSE c, %
1	-0.02	-0.05	-0.03	0.32	0.81	0.61
3	0.01	-0.02	-0.04	1.02	2.38	1.91
5	0.08	0.19	0.00	1.65	4.20	3.35
10	0.17	0.42	-0.18	3.37	7.73	6.52
15	0.26	0.60	-0.61	5.05	12.81	10.21

Table 2 shows that the estimator is practically unbiased (|bias| < 0.7 % even at 15 % noise), and its error grows almost linearly with noise. The decay rate b is consistently the least accurate parameter: its rRMSE is 4.2 % at 5 % noise and 12.8 % at 15 %, about 2.5 times the error of the amplitude a (Figure 3a). A practical rule follows: the relative accuracy of a rate constant is roughly 0.8 of the relative noise level for this design, so a target accuracy of 5 % for b requires noise below about 6 %.

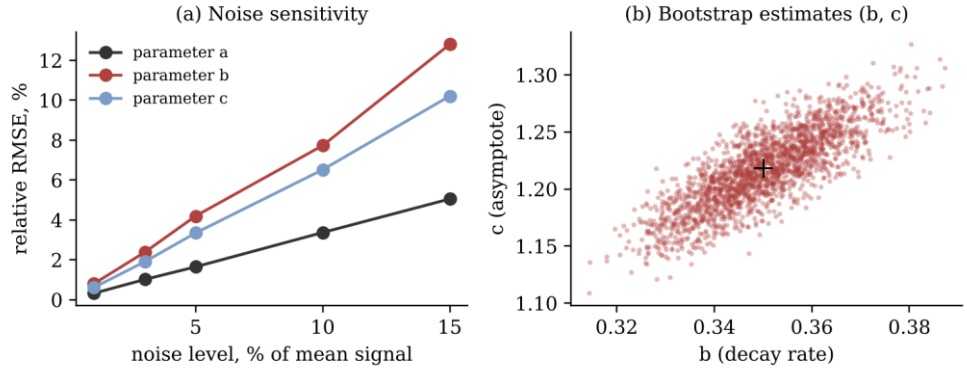


Figure 3. Experiment E2: (a) relative RMSE of parameters versus noise level; (b) 2000 bootstrap estimates in the (b, c) plane, the cross marks the point estimate

Table 3. Influence of the observation window on identifiability and accuracy (5 % noise, n = 40, 600 replicates)

x_{max}	$x_{max} \cdot b$	$\kappa(J^TJ)$	rRMSE a, %	rRMSE b, %	rRMSE c, %
3	1.05	$1.27 \cdot 10^4$	10.6	16.6	37.8
6	2.10	902	2.3	6.6	9.2
9	3.15	200	1.6	4.9	5.0
12	4.20	127	1.7	4.0	3.2
18	6.30	111	1.8	3.8	2.1

Note: $x_{max} \cdot b$ is the number of characteristic decay times covered by the experiment.

The experimental design matters more than the noise level (Table 3). When the observation window covers only about one decay time ($x_{max} = 3$), the curve does not reach its plateau, the asymptote c is barely constrained, κ rises to $1.27 \cdot 10^4$ and the error of c reaches 37.8 %, i.e. more than ten times the error obtained with the same number of points and the same noise on [0, 12] (3.2 %). Extending the window beyond about four decay times brings diminishing returns for a and b but continues to improve c (2.1 % at $x_{max} = 18$). The condition number therefore serves as an inexpensive, data-free predictor of reconstruction quality that can be computed before an experiment is run.

Reconstruction of a dynamic model (E3)

For the damped oscillator, all six Levenberg–Marquardt runs, started from points as far apart as $(\omega, \zeta) = (0.8, 0.8)$ and $(3.0, 0.3)$, converged to the same minimum ($S = 0.1516$) in 6–14 function evaluations. The reconstructed values are $\omega = 2.0053 \pm 0.0086$ (true 2.0; error 0.26 %), $\zeta = 0.1472 \pm 0.0038$ (true 0.15; error 1.9 %), $x_0 = 1.0012 \pm 0.0184$ and $v_0 = 0.038 \pm 0.047$ (true 0, inside one standard error). The residual RMSE (0.0355) is consistent with the noise level 0.04, so the model explains all systematic variation (Figure 4a).

The sum-of-squares landscape (Figure 4b) explains the robust convergence: within the explored region it has a single, well-defined basin centred on the true parameters. The basin is elongated along ζ, which is why the damping ratio is determined about five times less precisely in relative terms than the frequency: the frequency is fixed by the positions of the zero crossings over the whole record, whereas the damping is fixed only by the envelope, which decays into the noise after a few periods.

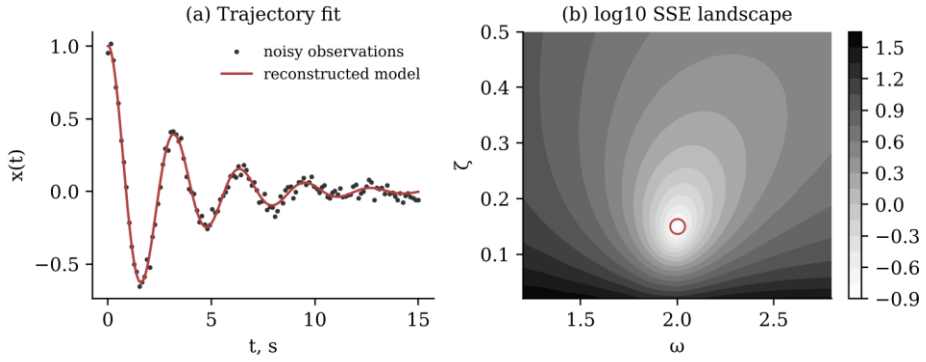


Figure 4. Experiment E3: (a) noisy trajectory and reconstructed model; (b) $\log_{10}$ of the sum of squared residuals over (ω, ζ), the circle marks the true parameters

Ill-posed reconstruction and regularisation (E4)

The blurring operator in E4 has the condition number $\kappa(\mathbf{K}) = 3.84 \cdot 10^{11}$. The unregularised least-squares solution is therefore useless: its relative error is $6.16 \cdot 10^8$, i.e. the 1 % noise is amplified by more than ten orders of magnitude. Tikhonov regularisation removes this instability (Table 4, Figure 5).

Table 4. Reconstruction error of the deconvolution problem for different choices of the regularisation parameter

Method	λ	Relative error $\|f_\lambda - f\| / \|f\|$
Unregularised least squares	0	$6.16 \cdot 10^{8}$
Tikhonov, L-curve corner	$9.62 \cdot 10^{-3}$	3.41 %
Tikhonov, discrepancy principle	$1.48 \cdot 10^{-1}$	4.12 %
Tikhonov, oracle (best possible λ)	$3.77 \cdot 10^{-2}$	2.96 %

Note: the oracle λ uses the unknown true signal and is shown only as a benchmark.

Both data-driven rules give errors close to the theoretical optimum of 2.96 %: 3.41 % for the L-curve and 4.12 % for the discrepancy principle. The L-curve slightly under-regularises and the discrepancy principle slightly over-regularises (it smooths the narrower peak more), which is consistent with the known behaviour of these rules. In both cases the positions and relative heights of the two peaks are recovered, while an under-regularised solution ($\lambda \approx 10^{-7}$) shows strong spurious oscillations (Figure 5b).

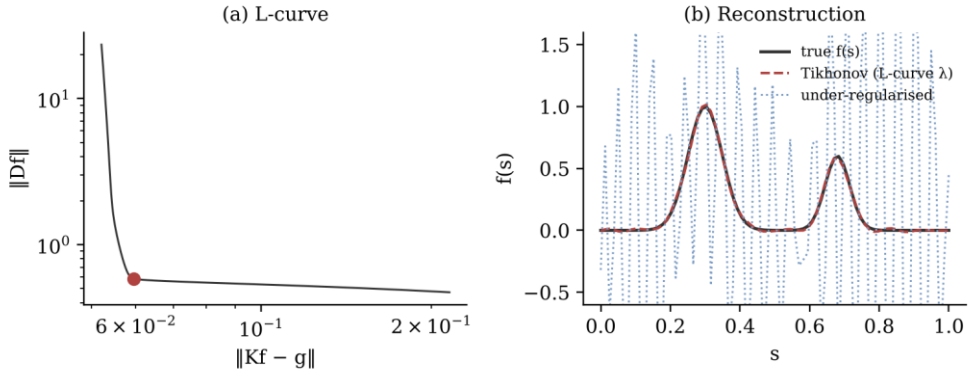


Figure 5. Experiment E4: (a) L-curve with the selected corner; (b) true signal, regularised and under-regularised reconstructions

4. Discussion

The four experiments lead to several conclusions that are relevant for anyone who derives quantitative laws from measurements.

Goodness of fit is not a criterion of structure. In E1 a cubic polynomial reproduced the data almost as well as the true law. If the analyst had reported the polynomial, its four coefficients would have had no physical meaning and the formula would have failed outside the sampled range, since a cubic eventually diverges while the true response tends to a plateau. Information criteria and cross-validation, which reward predictive ability and penalise unnecessary parameters, consistently recovered the correct structure, in agreement with the multi-model inference framework. The Akaike weight is particularly useful in reporting because it expresses model uncertainty on a probability-like scale rather than as a binary decision. We recommend that residual autocorrelation (DW, runs test) be checked first, since it rejects inadequate structures for a reason that is easy to interpret, and that information criteria be used only to choose among structures whose residuals are compatible with noise.

Parameters must be reported with uncertainty and correlations. The strong correlation between b and c in E2 means that their individual standard errors understate how loosely the pair is determined; reporting the correlation matrix or a joint confidence region avoids this. The close agreement between the analytical covariance (5) and the bootstrap indicates that, for well-designed experiments with moderate noise, the cheap local estimate is sufficient; the bootstrap becomes necessary when the confidence regions are visibly non-elliptical, i.e. for poorly conditioned designs.

The design limits accuracy more than the algorithm. All estimators in this study converged reliably, but the error of the same parameter differed by a factor of more than ten depending on the observation window (Table 3). This is the practical meaning of the «sloppiness» phenomenon: the information matrix, not the optimiser, determines what

can be learned. Since $\kappa(J^TJ)$ can be evaluated at the planning stage using prior guesses of the parameters, it provides a simple tool for experimental design — for instance, choosing the duration of an observation so that it covers at least three to four characteristic times of the process. The same reasoning underlies optimal experimental design based on the determinant or the eigenvalues of the Fisher information.

Ill-posedness requires prior information. E4 shows that for linear inverse problems the data alone cannot determine the solution; some form of prior knowledge — here smoothness through the operator D — must be introduced, and the regularisation parameter acts as the balance between fidelity to the data and fidelity to that prior. The fact that two independent selection rules gave nearly optimal and mutually consistent results suggests using both as a cross-check in practice.

Limitations. The study is based on controlled numerical experiments with Gaussian, homoscedastic and independent noise. Real data often contain outliers, heteroscedastic or correlated errors and errors in the independent variable; these call for robust or weighted loss functions, errors-in-variables methods and profile-likelihood intervals. The candidate set $\mathcal{M}$ was specified in advance; when the structure is unknown, sparse or symbolic regression can generate candidates, which can then be evaluated by the same workflow. Finally, the dynamic example contained only four parameters; for high-dimensional mechanistic models structural identifiability should be analysed before estimation.

5. Conclusion

This paper presented an integrated workflow for identifying functional relationships and reconstructing parameters from experimental data and evaluated it in four controlled experiments with known ground truth. The main results are as follows.

Residual diagnostics combined with AICc/BIC and cross-validation recovered the true exponential law with an Akaike weight of 0.947, whereas R^2 could not distinguish it from a cubic polynomial (0.9931 vs 0.9925).

Levenberg–Marquardt estimation was practically unbiased; the error of the parameters grew linearly with noise, the rate constant being the least accurate (rRMSE 4.2 % at 5 % noise). Analytical and bootstrap confidence intervals agreed for well-conditioned designs.

The observation window was the dominant factor of accuracy: shortening it from 12 to 3 units raised $\kappa(\mathbf{J}^T\mathbf{J})$ by two orders of magnitude and the error of the asymptote from 3.2 % to 37.8 %. The condition number can therefore be used to plan experiments before data are collected.

Parameters of an ODE model were reconstructed robustly from dispersed starting points (errors 0.26 % for ω and 1.9 % for ζ), and Tikhonov regularisation with L-curve or discrepancy-based λ reduced the error of an ill-posed deconvolution from $6.16 \cdot 10^8$ to about 3–4 %.

The workflow is general and can be applied without modification to kinetic, mechanical, biological and economic models. Further work will extend it to non-Gaussian and correlated errors, to automatic generation of candidate structures, and to sequential optimal design of experiments.

REFERENCES

[1] A. Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation. Philadelphia, PA, USA: SIAM, 2005, 342 p.

[2] K. P. Burnham and D. R. Anderson, Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed. New York, NY, USA: Springer, 2002, 488 p.

[3] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, "Universally sloppy parameter sensitivities in systems biology models," PLoS Computational Biology, vol. 3, no. 10, Art. no. e189, 2007.

[4] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," Proceedings of the National Academy of Sciences, vol. 113, no. 15, pp. 3932–3937, 2016.

[5] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Data-driven discovery of partial differential equations," Science Advances, vol. 3, no. 4, Art. no. e1602614, 2017.
[6] M. Schmidt and H. Lipson, "Distilling free-form natural laws from experimental data," Science, vol. 324, no. 5923, pp. 81–85, 2009.
[7] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," Journal of Computational Physics, vol. 378, pp. 686–707, 2019.
[8] P. Virtanen et al., "SciPy 1.0: Fundamental algorithms for scientific computing in Python," Nature Methods, vol. 17, pp. 261–272, 2020.
[9] A. C. Atkinson, A. N. Donev, and R. D. Tobias, Optimum Experimental Designs, with SAS. Oxford, U.K.: Oxford University Press, 2007, 511 p.
[10] L. Ljung, System Identification: Theory for the User, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1999.
[11] R. Isermann and M. Münchhof, Identification of Dynamic Systems: An Introduction with Applications. Berlin, Germany: Springer, 2011.
[12] B. Fornberg, "Generation of finite difference formulas on arbitrarily spaced grids," Mathematics of Computation, vol. 51, pp. 699–706, 1988.
[13] B. Fornberg, "Calculation of weights in finite difference formulas," SIAM Review, vol. 40, no. 3, pp. 685–691, 1998.
[14] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," Analytical Chemistry, vol. 36, no. 8, pp. 1627–1639, 1964.
[15] P. A. Gorry, "General least-squares smoothing and differentiation by the convolution (Savitzky–Golay) method," Analytical Chemistry, vol. 62, no. 6, pp. 570–573, 1990.
[16] J. Cullum, "Numerical differentiation and regularization," SIAM Journal on Numerical Analysis, vol. 8, no. 2, pp. 254–265, 1971