1. Basic Concepts of Nonlinear Dynamics
Attractors and bifurcations

**Preview**
Attractors and stability
Bifurcations and instability
Catastrophes
Data analysis

**Why Nonlinear Dynamics?**
Explains changes that occur over time.
Allows for structural comparison of models across situations, even theories that are very different by appearances.
Better explanations of data, $R^2$

**How Dynamics Affect Our Thinking**

**OLD WAY**
Linear relationships
Static situations
Normal pdf
Determined by simple functions

**NEW WAY**
Nonlinear relationships
Changes over time
Different kinds of changes display different dynamics
Seemingly random process
Exponential pdf. Catastrophe models have more complex exponential distributions
Implications for Interpreting Systems

**OLD WAY**
Outcomes are proportional in inputs
Simple cause and effect
Ignore, stifle random blips
Maintain equilibrium and control

**NEW WAY**
Little things can have big consequences, vice-versa
Control variables that behave differently
Emergent phenomena
It was no blip
Navigate a repertoire of nonlinear change processes
Self-Organization occurs from the bottom-up, while hierarchies operate from the top down.
Gives rise to new concepts of networks, and the onset of emergent business or other social activity.

Psychology is not the first science to break out of the linear rut. According to Stewart (1989), the physical sciences made the transition more than a half century ago:

So ingrained became the linear habit, that by the 1940s and 1950s many scientist[s] and engineers knew little else . . . [W]e live in a world which for centuries acted as if the only animal in existence was the elephant, which assumed that holes in the skirting-board must be made by tiny elephants, which saw the soaring eagle as a wing-eared Dumbo, [and] the tiger as an elephant with a rather short trunk and stripes (1989, p. 83-84).

**ATTRACTION**
A chunk of space.
An object enters and does not exit
Unless a very strong force is applied.

A fixed point
all points entering the space are drawn to the epicenter directly.

A fixed point attractor that is also known as a sink. Points entering the space reach the epicenter in a spiral motion.

**Attractor Basins**
Area around an attractor where the attracting force can operate.
Some attractors are stronger than others

**Research Strategy**
Sometimes a nonlinear behavior can be represented by more than one nonlinear function.
One hypothesized model could be better than another. Test both and compare.
Default: Is the nonlinear model better than a linear interpretation?

**Periodic Attractors or Limit Cycles**
Arrange a set of fixed points in a circle and see what happens...

Periodic attractor
There are many in real life –
Economics
Biology
Signal processing

Basic form, sine wave:
General form: \( y^2 = \theta_1 \sin(y_1 - \theta_2) \)
For compound oscillators:
\( y_2 = \theta_1 \sin(y_1 - \theta_2) - \theta_3 \cos(y_1 - \theta_4) + ... \)

If \( \theta_2 \) becomes a variable, \( f(y) \) becomes aperiodic.
Peaks flattened, cycles irregular in catastrophes.

Dampened oscillators too.

**Repellors**
Incoming points (objects) are deflected outward in any direction.

**Saddle point**
Has properties of an attractor and a repellor. You can visit, but you can't stay there too long.

The perturbed pendulum produces a system with 2 attractors and a saddle in the middle.

**Structural Stability**
All points are behaving according to the same mathematical rule.

**Attractor + Saddle**
Unfolding of a romantic relationship over time.

**Chaotic or “Strange” Attractor**
More on these later as time permits. Not usually germane to catastrophes, however.

**Bifurcations**
A split in a dynamical field where different dynamics exist in each local region.
Represent a pattern of instability
Could be as simple as a point, or a more complex pattern
**Types of Bifurcations: Subtle**
Simple splitting of a dynamical field as a control parameter changes value.
Most common type in the social sciences thus far.

![Change In Dining Speed](Image)

**Types of Bifurcations: Catastrophic**
Entire dynamical regions come and go suddenly as a control parameter changes value.
Related to, but different from, catastrophe models in catastrophe theory.
Bifurcations found in catastrophe models are actually subtle.

**Types of Bifurcation: Annihilation**
Collision between 2 or more dynamical structures changes dynamics irreversibly over time.

**Separatrix**
An area separating two dynamics.
Can take the form of a repellor.
Can be a passive “ambivalent” space.

**ENIGMA:**
It is sometimes difficult to determine whether a dynamic is really the result of a repellor that is located between two attractors or a limit cycle.
Empirical study will help, as usual!

**Put some elements together**
Two fixed point attractors
A pitchfork bifurcation
A repellor

And you get a cusp catastrophe!

All discontinuous changes of events can be explained by one of 7 elementary topological forms.

Forms differ in their levels of complexity. Cusp is the most commonly used.

**Singularity Theorem**
Catastrophe models are a set of universal unfoldings from Taylor Series expansions

Given a maximum of 4 control parameters there is only one behavior response surface

Ignore mirror images
Ignore minor permutations
Singularity actually holds up to 5 control parameters, after which multiple surfaces are possible.

**Classification Theorem**
Given a fixed number of control parameters there is only one surface associated with them.

Similarly, if we know the number of stable states of the behavioral surface, we know how many control parameters are operating.

The number of control variables and the behavioral spectrum are unbreakable packages.
Allows us to infer a global structure from a local structure.

Viz: If we know how many behavioral states there are, we know the number of control parameters in the process and what they do.

**Structural stability**
**Morphogenesis**

**Elementary models with 1 order parameter**
The fold
The cusp
The swallowtail
The butterfly
This group the CUSPOID series.

**The Fold**
df(y)/dy = y^2 – a

Its bifurcation set consists of a single point.

---

**2. Catastrophe Theory**

Rene Thom 1972, 1975
Each model has 3 characteristic equations:
The response surface is perhaps the most important, e.g. \( \frac{df(y)}{dy} = y^2 - a \) for the fold.
Set response surface equation equal to 0 to solve for the equilibrium points of \( y \).

The **bifurcation set**, is the set of critical points where behavior changes.
It is the derivative of the response surface: e.g. 
\[ \frac{df(y)^2}{d^2(y)} = 2y \] for the fold.
Set this function = 0 to solve for the bifurcation set.

The **potential or energy function** captures the dynamics when the system is standing still.
It is the integral of the response surface, e.g. \( f(y) = y^3 - ay \) for the fold.
Useful for defining statistical distributions associated with each catastrophe model.

**The Cusp**
BIFURCATION explains large v small differences.
ASYMMETRY explains proximity to the threshold of change.
\( \frac{df(y)}{dy} = y^3 - by - a \)

Often shown with its bifurcations set, which was used in many of the early behavioral applications.
Illustrates gradients instead of control parameters.

**Cusps in optics**

**Hysteresis** – Behavior changes back and forth, or up and down the manifold, with a different pathway going up compared to going down. Presence strongly suggests a cusp dynamic.

**SOME APPLICATIONS**
Hysteresis example: Group polarization

Zeeman’s Dog

Hardy’s model for sports performance

Zeeman’s Stock Market Model
Zeeman-Weintraub-Guastello stock market

Zeeman’s Prison Riots
Buckling of an Elastic Beam
Mechanics & materials problem, Zeeman 1976

**Delay Rule**
Once the control parameters are in place to promote a behavior change, there could be small time delay before the change occurs. "Calm before the storm".

**Maxwell Convention**
One gradient may be stronger than the other. The system is likely to remain in the stable state that is most probable. "we're [not] used to rioting here"?

**Phase Shifts**
Catastrophes, phase shifts, s/o
Discontinuous changes of events are implicated in self-organizing process

**The Swallowtail**
\[ \frac{df(y)}{dy} = y^6 - cy^3 - by - a \]
2 3-d sections

Leadership Emergence in Production Groups

**The Butterfly Catastrophe**
Response surface
\[ \frac{df(y)}{dy} = y^5 - dy^3 - cy^2 - by - a \]

Slice of bifurcation set

Motivation in Organizations

**Elementary models with 2 order parameters**
The wave crest (also known as the hyperbolic umbilic)
The hair (also known as the elliptic umbilic)
The mushroom (also known as the parabolic umbilic)

**Wave Crest**
Response surface requires 2 equations:
\[ \frac{df(x)}{dx} = 3x^2 + a + cy \]
\[ \frac{df(y)}{dy} = 3y^2 + b + cx \]
The crest of every wave breaks at a 120° angle. Applications in the behavioral sciences have not been reported.
Response surface has 2 equations
\[ \frac{df(y)}{dy} = 3y^2 - x^2 + a + 2cy \]
\[ \frac{df(x)}{dx} = -2xy + b + 2cx \]

Note that the two behaviors interact.

Applications in optics
Little published in the behavioral sciences

**The Mushroom**
Response surface

The response surface has 2 equations
\[ \frac{df(y)}{dy} = 4y^3 + 2dy + x^2 + b \]
\[ \frac{df(x)}{dx} = 2xy + 2cx + a \]

Note that the two behavior variables interact.

Creative Problem Solving
General participation, especially creative contributions.
Old School “Equilibrium”
Used to refer to stable, unstable states, attractors and saddles.

Catastrophe stable states are attractors. Catastrophe attractors may be fixed points of limit cycles.

**Sussmann & Zahler (1978) Critiques**
TRIVIALITY – Some applications were based on only superficial rationale and merely restated that discontinuities existed. Did not use the deep math.
ABSORB PREDICTIONS – Some proposed models made predictions that were not wholly believable or were counterintuitive.

**Oliva & Capdevielle (1980)**
Verifiability problem with response surfaces isn’t any different from the problems we solve with linear regression all the time.

**Guastello (1981) replies:**
The entire point is to use the products of the math in the form of response surfaces and their functions. Absurd? Test the models empirically. The results could be exciting.
Parsimony is all relative: Its parsimonious to use 1 model to encompass many different-looking phenomena

**Catastrophe Theory’s Statistical Era**
Cobb – nonlinear regression for cusp pdfs. Tests a pre-specific model.
Guastello – polynomial regression for cusp etc response surfaces. Tests a pre-specific model.
Oliva – combinational approach for pdfs with factoring many possible variables in a data set. Exploratory procedure.
Guastello – nonlinear regression, more direct than earlier NLR version.

**Second Statistical Wave (1995)**
Chaos, Catastrophe, and Human Affairs
Metaphors and Easter Bunnies
Compendium of catastrophe models
Chaos studies with nonlinear regression, exponential structures.
Ave R² nonlinear: linear = 2:1.

Prison riots revisited:
Statistically it worked just fine.
Rationalized as a special case of the butterfly theory of motivation in organizations.

**3. Statistical Testing of Hypotheses**
Here we will assume that you have an idea that you would like to test -- a model that seems to explain something of importance in your research.

**Which nonlinear model to use when?**

Define a model in the form of an equation
Test it statistically with real data.
Employ methods that separate the deterministic portion of the data from noise.
"Noise" is that portion of the data variance that is not explained by the deterministic equation.

**Three Traditions in Data Analysis**

- **Math**
  - No limit to data quantity
  - It's always clean
  - You always know what it contains because you put it there.

- **Physics and biology**
  - Filtering
  - Surrogate data
  - Theory helps you guess what it should contain

- **Social sciences**
  - Could always use more data
  - It’s noisy
  - Transient dynamics make it VERY noisy
  - You don’t know what it contains, but nonlinear theory helps develop a cogent hypothesis.

**Hierarchical Models**

- Each model in the hierarchy subsumes properties of the simpler models.
- Each progressively complex model adds a new dynamical feature.

**Structural vs. functional equations**

- **Linear vs. nonlinear** – classes of functions, error structures and what it means to be I I D.
- **Linearization**: Testing for a power law
- **Autocorrelation**, dependent error, and the structure of behavioral measurements
- **Type of data and amounts that are required**
- **Probability functions**, location, and scale
- The catastrophe models, which can be tested through power polynomial regression
- The exponential series of models, which are tested through nonlinear regression.

**Catastrophe models** are also testable as static probability functions through nonlinear regression.

**Functional Equations**

- Anytime $Y = f(X)$, linear or nonlinear
- We’ll see numerous examples where
- $Y_2 = f(Y_1)$, where $y$ is measured at 2 or more points in time.

**Structural Equations**

- Bear a close analogy to the relationship between the final factors of a factor analysis and original variables that were involved.
- Variation of nonlinear regression techniques:
  - Correlate research variables with estimated model parameters.

$$\Delta z = \beta_0 + \beta_1 z_1^3 + \beta_2 z_1^2 + \beta_3 bz_1 + \beta_4 a$$

- Expand this equation so that multiple research variables contribute to $b$ or $a$.
- OR two very different problems would involve the same structure but involve very different $b$ or $a$ variables.

**Models considered here have both functional and structural properties.**

**Linear Models**

- $Y = B_0 + B_1 X$
- Or expand:
- $Y = B_0 + B_1 X_1 + B_2 X_2 + ... + B_N X_N$

**Advantages:**

- Only one structure (function) worry about
- Can assume normally distributed error
- Can assume homoscedasticity
- Error dispersion is minimized compared to nonlinear functions

**Standard error of $B_i$ calculated the same way for all $B_i$.**

**Nonlinear Models**

**Myriad possible functions, e.g.,**

- $Y = e^x$
- $Y = cx^a$

$$pr(x) = \frac{1}{1 + \exp(-c - ax)}$$

- Statistically these 3 functions would be:
- $Y = e^{ix}$
- $Y = cx^a$

$$pr(x) = \frac{1}{1 + \exp(\theta_1 + \theta_2 x)}$$

- Where $\theta_i$ are nonlinear regression weights.

**Errors in nonlinear models** could be:
Additive as they are in linear models
\[ Y = f(x) + \varepsilon \]
Where \( \varepsilon \) is IID. 
IID = “independently and identically distributed.” Common assumption in nonlinear computational studies.

Proportional:
\[ Y = w_1f(x) + w_2(x)\varepsilon \]

Multiplicative:
\[ Y = f(x)(1 + \varepsilon) \]
The std error of \( \theta \) will depend on the argument associated with it. Will revisit this situation in the context of error functions for nonlinear regression.

Linearization
Polynomial regression type where a nonlinear component is substituted where a simple variable would go in the general linear model.
\[ Y = B_0 + B_1X_1 + B_2X_2^2 \]
OR start with
\[ Y = e^{\theta x} \]
And take a log of both sides to convert it to a linear model:
\[ \log(y) = B_0 + B_1X \]
Errors are now more likely to be normally distributed rather than exponential. Usual benefits of the linear model.

Power Law Distributions
Freq \( Y = aY^{-b} \)
Log(Freq\[y\]) = log(a) −b*log(y)
\[ a = \text{regression constant} \]
\[ b = \text{fractal dimension and calculated as regression weight.} \]
The two models for testing power laws are mathematically the same, but statistically equivalent only if \( X \) is error-free.

Structure of Behavioral Measurements
In the classic definition, a measurement consists of true score (T) plus error (e)
The classic assumption is that all errors are independent of true scores and other errors.

In nonlinear dynamics, our true score is the result of a linear and determinant process, a nonlinear component, dependent error, and independent error; the latter is now called IID (independently and identically distributed).

Autocorrelation
Typical problem: US-Soviet arms expenditures over time:
\[ Y_2 = X_1 \]
Regression coefficient \( r \) is greatly inflated by correlation of \( X \) and \( Y \) with themselves. Better model:
\[ Y_2 = Y_1 + X_1 \]
OK to use OLS regression so long as errors are IID. Alt: Residuals are correlated from one point to the next.
ARIMA (Box-Jenkins) GLS developed for this purpose with linear models. ARIMA figures optimal lag length as part of the calculation process.

Durbin-Watson statistic
Null Hyp: \( d = 2 \).
Two examples from CCHA:
Prison riots, cusp catastrophe
Workforce stability, May-Oster logistic map (population dynamics) model.
Linear model displayed significant autocorrelation of residuals while Nonlinear model did not.

Conjecture: The large amount of variance accounted for by nonlinear models results in part from the transformation of non-IID into the function that is specified by the model.

Proof (Brock et al., and others): Presence of non-IID error indicates the presence of nonlinearity in the data.

BDS statistic – doesn’t specify which type of nonlinearity, however.

Where might non-IID error come from?
An external shock at \( T_x \) becomes iterated through the function at \( T_{(x+1)} \).

AKA sensitivity to initial conditions!

Types of Data Format and Structures
These procedures require dependent measures (order parameters) that are measured at two points in time at least.
Many entities that are measured at two points in time, OR
One long time series of observations from one entity, OR
An ensemble of shorter time series taken from several entities.

How Many Observations?
Need enough data to cover all the actual underlying dynamics.

Problem of stationarity in real data: Are the dynamics for one time epoch the same as the dynamics for another time epoch?

It is better to have a smaller number of observations that cover the full dynamical character of your phenomenon than to have a larger number of observations that cover the underlying topology poorly.

Because these are statistical procedures, all the usual rules, and caveats, pertaining to statistical power apply. (Not always available for nonlinear regression)

Generalizability out of sample: N, number of regression parameters.

You can often get the job done with 50 data points if you have:
a good model,
decent measurements,
only one or two parameters to estimate,
carefully collected data

Optimal time lag
ARIMA approach
Theiler windows
Put simply, the time lag between observations is optimal if it reflects the real time frame in which data points are generated. For instance:
Catastrophe models are usually lagged “before” and “after” a discrete event.
Macroeconomic variables might be studied best at lags equal to an economic quarter of the year.

Probability Density Functions
It is convenient that any differential function can be transformed into a probability density function, using the Wright-Ito transformation.

\[ p_d(z) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(z-\lambda)^2}{2\sigma^2}} \]

The variable \( y \) is a “dependent measure” that exhibits the dynamical character that you are studying.
\( y \) is then transformed into \( z \) with respect to location and scale:
\[ z = \frac{(y - \lambda)}{\sigma} \]

Location Parameter
In most discussions of probability functions, “location” refers to the mean.
In this case it refers to the lower limit of the function, or the lowest observed value.
This is our way of fixing a zero point and transforming interval scales (common in the social sciences) into ratio scales.
By fixing the location point, we define where the nonlinear function is going to start.
We would not be able to draw a conclusion about the presence of a particular function if only part of the function was allowed into the data.

Scale Parameter
The scale parameter in common discussions of probability density functions is the standard deviation of the distribution.
We usually use the standard deviation here also.

The use of the scale parameter later in nonlinear statistical equations eliminates bias between two or more variables that are multiplied together.
The results of linear regression are not affected by values of location and scale so long as the variables in the linear model are strictly additive.

Statistical interactions: \( y = \ldots + \beta_n X_1 X_2 \) are affected by the std dev of \( X_1 \) and \( X_2 \).
Nonlinear models are so affected.

CATASTROPHE MODELS
For discontinuous changes over time.
Analysis that follows requires the polynomial form of multiple linear regression.
You can do the analysis with most any standard statistical software package.
The instructions that follow pertain to the cuspoid group.
Although they have been generalized to the umbilic group, we must save that exercise for another installment.
Choose a Model
One begins the task of modeling by choosing a model that appears closest to the phenomenon as it is known to the researcher. The cusp is the most often used model, so for demonstration purposes we will start there.

**Cusp Response Surface**
Depicts 2 stable states of behavior and requires 2 control variables. The ASYMMETRY parameter governs how close the system is to a sudden discontinuous change of events. The BIFURCATION parameter governs how large a change will take place. Two or more experimental variables may be hypothesized for each parameter without changing the basic model or analytic procedure.

**Deterministic Cusp Equation**
The deterministic equation for the cusp is shown below, and followed by its probability function using the Wright-Ito transformation.

**Equation for the cusp response surface:**
\[
\delta f(y)/\delta y = y^3 - by - a.
\]

Its pdf would be, therefore:
\[
pdf(z) = \xi \exp \left[ -z^4/4 + bz^2/2 + \alpha \right].
\]
This is an example of what a cusp pdf could look like. Note the two small, separated density regions in the front, and the single density region toward the rear.

Next we take the deterministic equation, insert regression weights, and a quadratic term. The quadratic term is an additional correction for location. It corrects for one side of the cusp being better developed than the other side.

Several hypotheses are being tested in the power polynomial equation:
\[
F \text{ test for the model overall; get the } R^2 \text{ and save it for later use.}
\]

t tests on the beta weights; these tell you what parts of the model account for unique portions of variance.

In regard to the model elements, some parts are more important than others. The cubic term expresses whether the model is consistent with cusp structure; the correct level of complexity for a catastrophe model is captured by the leading power term.

If there is a cusp structure, then you have to have a bifurcation variable, as represented by the \( b_{z_1} \) term. Always hypothesize one, or shabby results for the model as a whole will often result. The asymmetry term, \( b_{a} \) is important in the model, but failing to find one does not negate the cusp structure; it only means the model is not complete.

The quadratic term is the most expendable. It is not part of the formal deterministic cusp structure. In the event that unique weights are not obtained for all model components, delete the quadratic term and try again.

Another procedural contrast: In ordinary linear regression, when a variable does not attain significant weight, we simply delete that variable. In nonlinear dynamics, we delete variables based on their relative importance to the theoretical function.

**Linear Contrasts and Drawing Conclusions**
Construct these two models, compute their \( R^2 \) coefficients, and compare the results against the \( R^2 \) obtained for the cusp.

In a clear result, the cusp model will account for at least as much variance as the nearest linear alternative, and often more so.

Next, evaluate the elements of the cusp model. If you have all the necessary parts, and the \( R^2 \)'s compare favorably, you have a clear case of the cusp.

**Example: Occupational Health**
*Diathesis-stress model*
First published in Guastello (1995)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>r</th>
<th>R2 (step)</th>
<th>t(weight) F(model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cusp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z13</td>
<td>-.61</td>
<td>.37</td>
<td>4.86****</td>
</tr>
<tr>
<td>z12</td>
<td>-.71</td>
<td>.60</td>
<td>-6.69****</td>
</tr>
<tr>
<td>z1*Job stress</td>
<td>.20</td>
<td>.60</td>
<td>-1.77*</td>
</tr>
<tr>
<td>Age-experience</td>
<td>.16</td>
<td>.63</td>
<td>5.23****</td>
</tr>
<tr>
<td>Anxiety</td>
<td>.24</td>
<td>.69</td>
<td>5.34****</td>
</tr>
<tr>
<td>Job-stress</td>
<td>.20</td>
<td>.70</td>
<td>2.46** 88.74****</td>
</tr>
</tbody>
</table>

*p<.10, **p<.05, ***p<.01, ****p<.001

Difference control model

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>.35</td>
<td>.12</td>
<td>4.74****</td>
</tr>
<tr>
<td>Age-experience</td>
<td>.25</td>
<td>.20</td>
<td>5.00****</td>
</tr>
<tr>
<td>Job-stress</td>
<td>.28</td>
<td>.23</td>
<td>2.90*** 23.85****</td>
</tr>
</tbody>
</table>

Pre-post control model

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>.36</td>
<td>.13</td>
<td>4.92****</td>
</tr>
<tr>
<td>Age-experience</td>
<td>.26</td>
<td>.21</td>
<td>5.26****</td>
</tr>
<tr>
<td>Time-1 illnesses</td>
<td>.19</td>
<td>.25</td>
<td>3.43***</td>
</tr>
<tr>
<td>Job-stress</td>
<td>.29</td>
<td>.28</td>
<td>3.17*** 22.29****</td>
</tr>
</tbody>
</table>

Resilience in Organizations
Adequate preparation for a possible disaster
Adaptive responses to handle it well while it's happening
Minimize damage or casualties afterwards

Resilience in Individuals
How well can they keep their lives together in the event of a serious illness?
How well can they bounce back, stay optimistic?

Buckling of an Elastic Beam
Human performance application, Guastello (1985)

What other human applications can you think of for the beam buckling dynamic?
Resilience variables contribute to the bifurcation parameter.

Severity of the illness, stress, etc. contribute to the asymmetry parameter

Expand, but Avoid Fishing Expeditions.
You can expand: It is possible to hypothesize more than one research variable as a bifurcation or asymmetry variable.
Have a clear notion of which variables should behave in which capacity. A well-conceived theory will help this task enormously.

NONLINEAR REGRESSION
Start with the model (function) that you wish to test.
Identify where the regression weights are supposed to appear.
In NLR, you must specify the placement of constants in a model. Unlike the general linear model, constants could appear anywhere at all in nonlinear functions.

Transform variables by location and scale.
See handout for commands.
1. Specifying the model – SPSS accepts roman characters instead of my greek θ.
2. Specify starting values of the regression parameters.
3. Other options we don’t need right away.

Newton-Raphson search: iterative procedure to determine the final values of the regression parameters.

Based on principle: calculate \( dF(y) = 0 \), solve for roots. (Hartley 1961, Marquardt, 1963.)
Suppose we have \( F(z) = cx^θ \)

Suppose we have an estimate \( θ = t_n \)
Plug \( t_n \) into \( F(z) \).
Plug \( t_n \) into \( dF(z) \).
Calculated \( t_n \) at the next step:
\( t_{n+1} = t_n - \frac{F(z)}{dF(z)} \)
Does the new \( t \) reduce the variance of the predicted values of \( z \) significantly? Continue until the answer is NO then stop.

Output:
ANOVA table showing SS, df, MS
SPSS doesn’t give an F-test; controversial which of several to apply here.
\( R^2 \) – fmri machines use this!
Tests on each regression weight.
Correlation among parameter estimates.
Tests on the regression weights are important, however!
If you do not obtain statistical significance for all three weights, delete the multiplicative constant and try again.
If that doesn’t look good, drop the additive constant ($\theta_3$) and you’ll be back to the simplest model again.

**Parameter estimates**
In multiple linear regression we might inspect correlations among the IVs.
In NLR we might inspect correlations among the parameter estimates.
Parameter estimates don’t have linear counterparts.
Imagine what we would have to do to the parameters to bend the curve so that it could accommodate each data point.
Each point has its own (set of) parameter(s), which in turn have std errors associated with them.
If the 95% CI doesn’t include 0.0, the test is significant. No concern about $p < .01$ or smaller.
If parameters are highly correlated – equivalent to multicollinearity. Consider dropping a parameter from the model.
Option: Save set of parameters and correlate (ordinary linear r) research variables with these estimates to find variables that explain them.

**Less is more:**
Multiple linear reg: $R^2$ increases (trivially) just by adding another variable to the model.
Not true in NLR. New (trivial) variable can add substantial error, lowering $R^2$.
Parameter values and $R^2$ are not connected in NLR as they are in MLR.

**Tips for Nonlinear Regression**
On the “model parameter” command line, the names of the regression weights are specified with initial values. Use either the initial values given in the narrative, or pick your own.
When in doubt, give the weights equal value.
It often doesn’t matter whether they start off equal or not.
If you have the option to choose constrained or unconstrained NLR, use unconstrained, which is the default.
If you have the option to choose least squares or maximum likelihood error term specification, I’d recommend least squares, which is what I regularly use.
If you try NLR with your own data set and find that the fit is so bad that you obtain a negative $R^2$, do not be alarmed. This happens. Just treat the negative $R^2$ as if $R^2 = .00$, which is poor enough, and move on.
When testing for significance, the tests on the weights are VERY important. Some people value them more than the overall $R^2$.
Tests for weights are made using the principle of confidence intervals. An alpha of $p < .05$ is regarded as universally sufficient.
A nonsignificant weight with a high overall $R^2$, could be the result of high correlation among the parameter estimates; this is akin to multicollinearity in ordinary linear models.

**TEST CATASTROPHES THROUGH NONLINEAR REGRESSION?**
a PDF that looks like that of an elementary catastrophe, and we want to see if that’s true.
a catastrophe process, but, all our time-1 measurements are the same value -- 0.
Static data

**Cusp model:**
$$pdf(z) = \xi \exp \left[ \beta_1 z^4 + \beta_2 z^3 + \beta_3 b z^2 + \beta_4 a z \right]$$

**Basic Linear Model:**
$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

**Linear Interaction Model**
$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

**Entire Sample**

<table>
<thead>
<tr>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Linear</td>
</tr>
<tr>
<td>Linear Interaction</td>
</tr>
<tr>
<td>Cusp Model</td>
</tr>
</tbody>
</table>

Significant weights obtained for: $\beta_1 z^4$, $\beta_2 z^3$, $\beta_3 b z^2$, $\beta_4 a z$

**Swallowtail catastrophe model**
The swallowtail response surface is:
$$df(y)/dy = y^4 - cy^3 - by^2 - a$$

Note placement of regression weights
Signs of weights determined empirically
The PDF is tested through nonlinear regression. If we don’t know what the control variables are yet we just ignore the variables $a$, $b$, and $c$ in the equation. (In principle, we’re treating them as constants, or part of the regression weight). (We need to have a hypothesis about them in the polynomial regression method).

**Easiest means of finding the modes & antimodes;**
The equation below is a polynomial regression model where the FREQUENCY of $y$ is a function of the value of $y$.

These were the final parameter estimates for the degree to which the data fit the swallowtail distribution. They were determined through NLR.

**Test Control Variables Too**
Hypotheses concerning control variables can be tested by specifying the variables in the places marked by arrows.

**NLR: Effect of adding variables**
Attenuation of $R^2$ when control variables are added

Adding variables can make $R^2$ go down instead of up because it just adds more opportunities for error. Local linearity in time series
Any curve can be approximated by a line if it’s short enough.
Need to see full unfolding to see global nonlinearity. Importance of an adequate time lag.
Otherwise the linear model “cheats.”

**Least Squares vs. Maximum Likelihood**
Definition of the error term
Either can be used with linear or nonlinear models. Linear ML models are rare.
ML permits capitalization on chance → less generalization out of sample.

---


