What Could We Do Better?

Alternative Statistical Methods

Curtis Storlie
Review of Current Statistical Model

- We take measurements of the displacement, $y$, at times $t_1, \ldots, t_n$

- What we actually observe is $\tilde{y}$ which is a noisy version of $y$,

$$\tilde{y}(t_j) = y(t_j) + \varepsilon(t_j)$$

OR

$$\tilde{y}_j = y_j + \varepsilon_j$$

- $\varepsilon_j$ is the error resulting from imperfect measurement at time $t_j$
A Statistical Model For Displacement

- Under the spring model it is assumed that the displacement over time is governed by the spring model

\[ y(t; C, K) \]

So we could write

\[ \tilde{y}_j = y(t_j; C, K) + \varepsilon_j \]
A Statistical Model For Displacement

- Assumptions about the error terms, $\varepsilon_j$
  - We assume that $\varepsilon_j \sim N(0, \sigma^2)$
  - It was also assumed that $\varepsilon$ is independent over time so that $\varepsilon_i$ is independent of $\varepsilon_j$ for $i \neq j$.
  - These assumptions of independent and identically distributed errors are often abbreviated by iid
  - We will take a closer look at these assumptions later

- With the assumptions above, the model is equivalent to

$$\tilde{y}_j \sim N(y(t_j; C, K), \sigma^2)$$
A Statistical Model For Displacement

- We also have several replicates (repeated experiments) of the beam vibration

- Let $t_{ij}$ be the $j^{th}$ time at which displacement is measured for the $i^{th}$ replicate, $i = 1, \ldots, r$ and $j = 1, \ldots, n_i$

- Denote the measured displacement over time for the $i^{th}$ replicate by $\tilde{y}_{i}(t_j) = \tilde{y}_{ij}$.

- If it is reasonable to assume $C$ and $K$ are the same for each replicate then we could assume

$$\tilde{y}_{ij} \sim N \left(y(t_{ij}; C, K), \sigma^2 \right)$$

independent over time and replicate
A Statistical Model For Displacement

- The likelihood of this model would be the product of all of the individual density functions since \( \tilde{y}_{ij} \) is independent of \( \tilde{y}_{i'j'} \) unless \( i = i' \) and \( j = j' \)

\[
L(C, K, \sigma; \tilde{y}) = \prod_{i=1}^{r} \prod_{j=1}^{n_i} \mathcal{N} (\tilde{y}_{ij}; f_s(C, K, t_{ij}), \sigma^2)
\]

where \( \mathcal{N} (x; \mu, \sigma^2) \) is the normal density function evaluated at \( x \) with parameters \( \mu \) and \( \sigma^2 \)
The so-called maximum likelihood estimates (MLE’s) for $C$ and $K$, denoted $\hat{C}$ and $\hat{K}$, are the values of $C$ and $K$ that maximize the likelihood function.

We could of course use the likelihood function to form a region of plausible values for $C$ and $K$ as well (recall Floyd talked about this).

Because of the form of the normal density, the MLE is the same as the least squares estimate you’re familiar with.
How Good is this Model?

- We can assess the goodness of fit by using the spring model to predict the observed measurements.

- The predicted (or fitted) values, \( \hat{y}(t) \) are obtained by evaluating the spring model with \( C = \hat{C} \) and \( K = \hat{K} \):

\[
\hat{y}(t) = y(t; \hat{C}, \hat{K})
\]

- We can compare the fitted values at the observed times, \( \hat{y}(t_{ij}) = \hat{y}_{ij} \) to the observed values \( \tilde{y}_{ij} \). 

Spring Model Fitted Values

![Graph showing the fitted values for different repetitions of a spring model.](image-url)
Spring Model Fitted Values

The figure shows a graph with a timeline on the x-axis and a value scale on the y-axis. The graph includes multiple replications (rep 1, rep 2, rep 3, rep 4) and a fitted line. The values are represented in scientific notation, $\times 10^{-5}$. The graph visualizes the oscillatory behavior over time.
Spring Model Fitted Values

![Graph showing the fitted values of a spring model over time. The graph plots w (vertical axis) against time (horizontal axis), with multiple replications and a fitted curve.]
Model Residuals

- The residuals for a model are obtained by taking the difference between the observed values and the fitted values

\[ e_{ij} = \tilde{y}_{ij} - \hat{y}_{ij} \]

- The residuals show us how far our model predictions are from the observed data.

- The residuals are also our best guess at the values of the \( \varepsilon_{ij} \). Hence for our current model, we would expect the \( e_{ij} \) to look iid. Let’s see.
Spring Model Residuals
Spring Model Residuals
Spring Model Residuals

![Diagram of Spring Model Residuals](image)
**Coefficient of Determination $R^2$**

- For a linear model

\[
R^2 = 1 - \frac{\text{SSE}}{\text{SSTot}} = 1 - \frac{\sum(\tilde{y}_j - \hat{y}_j)^2}{\sum(\tilde{y}_j - \bar{y})^2}
\]

- Conceptually, SSE represents the amount of error (variability) left unaccounted for in the data after fitting our model.
- SSTot represents the amount of error there that would be left over after simply fitting a constant (mean) function to the data.
- So $R^2$ represents the amount of variability in $\tilde{y}$ that is explained by our model.
Coefficient of Determination $R^2$

- We can define $R^2$ for nonlinear models (like the spring model) in the same way.
- In this particular case,

$$R^2 = 1 - \frac{3.60 \times 10^{-6}}{6.74 \times 10^{-6}} = .4662$$

- This means the spring model accounts for about 47% of the variability in our displacement measurements.
How Good is this Model?

- Visually our model doesn’t appear to fit the data well.
- \( R^2 = .47 \) which is fairly low, but perhaps the rest of the variability is purely due to measurement error.
- How can we find out?
How Good is this Model?

- We can investigate this further by fitting a more flexible model and compare the results to the spring model.

- “Nonparametric” Models
  - Smoothing Splines
  - Local Regression
  - Wavelets
  - Many others
How Good is this Model?

- We can investigate this further by fitting a more flexible model and compare the results to the spring model.

- “Nonparametric” Models
  - Smoothing Splines
  - Local Regression
  - Wavelets
  - Many others
Smoothing Splines

- A cubic spline $f(x)$ is a function that is a piecewise cubic polynomial. Between each set of knot points along the x-axis it is a cubic polynomial.

- At the knot points the cubic polynomials are required to meet continuously and with continuous first and second derivative.

- One possible form for this function is

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{k=1}^{K} \beta_{3+k} \left( |x - x_k|_+ \right)^3$$

where $x_k$ are the knot locations
A Cubic Spline

Cubic Spline with knots at $x=1$ and $x=3$
Smoothing Splines

- A smoothing spline, $\hat{f}(x)$, is a cubic spline with knots at every observed x-axis location (in our case time points).
- For given data values $(x_j, y_j)$, $\hat{f}$ is determined to be the cubic spline $f$ that minimizes

$$\sum_{j=1}^{n} (y_j - f(x_j))^2 + \lambda \int_{a}^{b} (f''(x))^2 \, dx$$

- $\lambda$ is referred to as the smoothing parameter
Smoothing Splines

- Larger values of $\lambda$ result in a bigger penalty for large values of the second derivative and hence results in a smoother function $\hat{f}$.
  - as $\lambda \to \infty$, $\hat{f}$ becomes a line

- Smaller values of $\lambda$ result in a more wiggly function $\hat{f}$
  - as $\lambda \to 0$, $\hat{f}$ becomes an interpolating spline (if all of the $x$ values are distinct).
Smoothing Splines

- Choosing an appropriate value for $\lambda$ is not always easy.

- Often “leave one out” Cross Validation is used
  
  - Fit $\hat{f}$ to all of the data points except one.
  
  - Then predict the one we left out.

  - $\lambda$ is chosen to minimize the sum of the squared “leave one out” errors
Smoothing Spline Fitted Values

![Graph showing Smoothing Spline Fitted Values with time on the x-axis and w on the y-axis, with multiple repetitions and a smooth line.](image)
Smoothing Spline Fitted Values

![Graph showing smoothing spline fitted values with multiple repetitions and a fitted line.](image-url)
Smoothing Spline Fitted Values

![Graph showing smoothing spline fitted values](image)

- The graph displays fitted values over time with different replicates labeled as rep 1, rep 2, rep 3, and rep 4.
- Each replicate is represented by a distinct color, with a smooth line also plotted for comparison.
- The x-axis represents time ranging from 0 to 0.5, while the y-axis represents the fitted values ranging from $-6 \times 10^{-5}$ to 6.6.
Smoothing Splines

- Do the residuals from the smoothing spline appear to be iid?
- Do they appear to be normally distributed?
Smoothing Spline Residuals
Smoothing Spline Residuals
Smoothing Spline Residuals

![Graph showing Smoothing Spline Residuals]

- x \times 10^{-5}

- Y-axis range from -1 to 1

- X-axis range from 0 to 0.5
Histogram of the Standardized Residuals
QQ-plot of the Standardized Residuals

QQ Plot of Sample Data versus Standard Normal

Quantiles of Input Sample vs. Standard Normal Quantiles
Test for Lack of Fit

- We wish to decide between the following two hypotheses
  - $H_0$: $y(t) = y(t; C, K)$ for some values of $C$ and $K$
  - $H_a$: $y(t) \neq y(t; C, K)$ for any values of $C$ and $K$
- We can accomplish this by comparing the residuals from the spring model to those of the smoothing spline.
- If $y(t; C, K)$ is a good model, then the residuals should be similar in magnitude to those for the smoothing spline.
Test for Lack of Fit

- Specifically, we will calculate the Sum of Squares Error (SSE) for both the spring model and smoothing spline

\[ SSE = \sum_{i=1}^{r} \sum_{j=1}^{n_i} e_{ij}^2 \]

- An estimate for \( \sigma^2 \) is given by

\[ \hat{\sigma}^2 = \frac{SSE}{n - df} \]

where \( df \) is the number of parameters in the model (or the degrees of freedom)
Test for Lack of Fit

- Let the $SSE$ and $df$ for the spring model and smoothing spline be denoted $SSE_1$, $df_1$ and $SSE_2$, $df_2$ respectively and

\[
\hat{\sigma}_1^2 = \frac{SSE_1}{n - df_1}
\]

\[
\hat{\sigma}_2^2 = \frac{SSE_2}{n - df_2}
\]

- If $H_0$ is true (the spring model is correct), $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ should both be estimates of the same quantity $\sigma^2$
Test for Lack of Fit

- We will then calculate

\[ F^* = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{\hat{\sigma}^2 \text{ from spring model}}{\hat{\sigma}^2 \text{ from smoothing spline}} \]

- If \( H_0 \) is true we would expect \( F^* \) to be close to ???
- If \( H_0 \) is false we would expect \( F^* \) to be ???
Test for Lack of Fit

- We will then calculate

\[ F^* = \frac{\hat{\sigma}^2}{\tilde{\sigma}^2} \]

- If \( H_0 \) is true we would expect \( F^* \) to be close to 1
- If \( H_0 \) is false we would expect \( F^* \) to be LARGE
Test for Lack of Fit

- When $H_0$ is true, $F^*$ has approximately an $F$ distribution with parameters $df_1$ and $df_2$.
- So we can calculate the probability of observing an $F^*$ as large as the one we observed or larger if $H_0$ really was true.

$$p\text{-val} = P(F \geq F^*)$$
P-value from F Distribution

F Distribution with df1 = 10 and df2 = 20
Test for Lack of Fit

- If the p-val is small then we can reject $H_0$, because there is a small chance this data could have been produced by the spring model.

- Usually we use the rule that if p-val $< .05$ we reject $H_0$ and conclude that the model $y(t; C, K)$ has a lack of fit.
Test for Lack of Fit

- For our data $SSE_1 = 3.598 \times 10^{-6}$, $df_1 = 2$, $SSE_2 = 2.450 \times 10^{-7}$, $df_2 = 1723$, $n = 30673$,
- This gives us $\hat{\sigma}^2 = 1.173 \times 10^{-10}$, $\tilde{\sigma}^2 = 8.464 \times 10^{-12}$, and
  
  \[ F^* = \frac{1.173 \times 10^{-10}}{8.464 \times 10^{-12}} = 13.861 \]
- the p-value is then given by
  
  \[ p-val = P(F \geq 13.861) < 0.001 \]
- This is a very small p-value so we reject $H_0$ and conclude there is a lack of fit with the spring model.
How Good is this Model?

- OK, our current spring model is not adequate.
- There are two possibilities.
  1. We have incorrectly specified the mean structure.
  2. We have incorrectly specified the error structure.
     - It is possible that the spring model is still a good model for the mean structure, but observations are not independent over time and/or replicate.
Improving our Model

- If we are not willing to give up on the spring model just yet, perhaps each replicate could get its own values for $C$ and $K$.

- We could assume that

$$C_i \sim N(\mu_c, \sigma_c^2)$$

$$K_i \sim N(\mu_k, \sigma_k^2)$$

and

$$\tilde{y}_{ij} = y(t_{ij}; C_i, K_i) + \varepsilon_{ij}$$

- This is typically referred to as a random effects model. That is, the parameters $C_i$ and $K_i$ are random variables that take on different values for each replicate.
Spring Model Fitted Values

The graph represents the fitted values for a spring model over time. The x-axis represents time, ranging from 0 to 0.5, and the y-axis represents the fitted values, scaled by $10^{-5}$. The graph includes multiple replications (rep 1 to rep 4) and a fitted line for comparison.
Plot of the Residuals for Random Effects model

![Plot of the Residuals for Random Effects model](image)
Improving our Model

- There still seems to be a systematic lack of fit for this random effects model as well.

- It seems fairly clear that we need a better underlying mathematical model.
  - Fortunately we have one, the “beam model”!

- Let us now try the model

\[ \tilde{y}_j = y_b(\theta, t_j) + \varepsilon_j \]

where \( y_b(\theta, t) \) is the beam model Ralph presented earlier today.
Beam Model Fitted Values

![Graph showing data and fitted values with time on the x-axis and y on the y-axis. The graph includes a legend for data and fitted with line styles.]
Beam Model Fitted Values
Smoothing Spline Fitted Values

![Graph showing Smoothing Spline Fitted Values](image)
Beam Model Residuals

Graph showing Beam Model Residuals with x-axis from 0 to 4 and y-axis from -3 to 3. The residuals are plotted with a blue line.
Beam Model Residuals
Beam Model Residuals

![Graph showing beam model residuals with x-axis ranging from 1 to 2 and y-axis ranging from $-1 \times 10^{-5}$ to $1 \times 10^{-5}$, with scattered data points.]
How Good is the Beam Model?

- For this data set $R^2 = .94$ for the beam model, which is MUCH better than .47 for the spring model.

- As far as for testing lack of fit

\[ F^* = \frac{3.348 \times 10^{-11}}{1.495 \times 10^{-12}} = 22.39 \]

- so the p-value is

\[ p\text{-val} = P(F \geq 13.861) < 0.001 \]

- So there is still some lack of fit with the beam model, but it is clearly a much better fit than the spring model
Can we improve our current use of the Beam Model?

- Of course we can.
- Since the residual plot shows that the errors, $\varepsilon_j$’s, are clearly not independent across time, perhaps we could fit the model

$$\tilde{y}(t) = y_b(\theta, t) + \varepsilon(t)$$

where $\varepsilon(t)$ is a random process that is NOT iid over time.
- This could be reasonable if the measuring sensor is affected by the beam vibration for instance.
Can we improve our current use of the Beam Model?

- Perhaps $\varepsilon(t)$ is a Brownian Motion
- We need something that oscillates though
- Maybe we could use something like the spring model with random parameters + $iid$ errors to model the measurement error.
- The systematic structure of the residuals however, suggests that we still do not have the mean structure of $\tilde{y}$ quite right.
- We may need to go back to Ralph and ask how we might make the beam model even more “kick ass”.

Conclusions

- We have determined that the Spring Model does not adequately model the beam displacement data.
- Nonparametric Statistical Models helped us to test and improve upon our mathematical model.
- It appears that after fitting a smoothing spline to the data, the measurement error is close to iid normal.
- Why don’t we just use a Smoothing Spline and call it good?
- The Beam Model still has some lack of fit, but visually fits the data quite well.
- There is never going to be a magic model that predicts perfectly. Need to decide how good is good enough.