Frequentist Techniques for Model Calibration

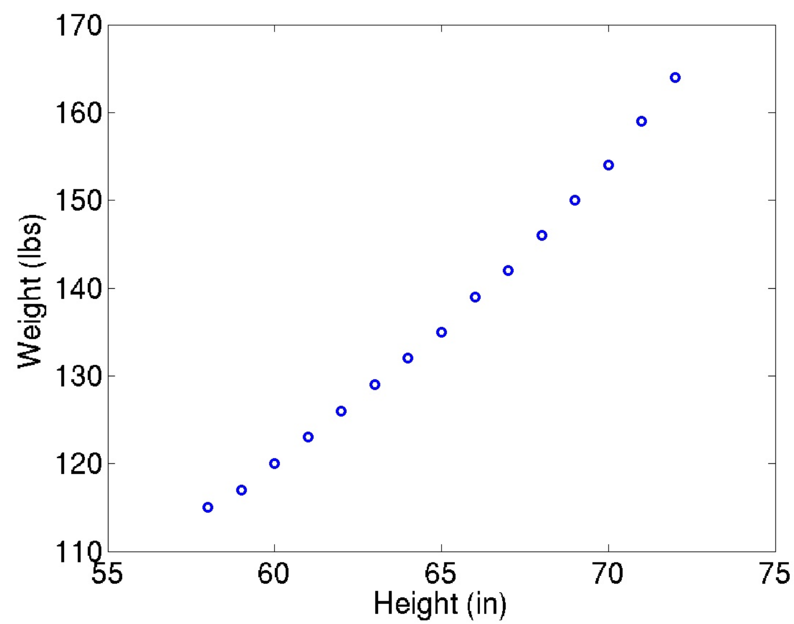
Reading: Chapter 7

Example: Consider the height-weight data from the *1975 World Almanac and Book of Facts*

Height (in)	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
Weight (lbs)	115	117	120	123	126	129	132	135	139	142	146	150	154	159	164

Consider the model

$$\Upsilon_i = q_1 + q_2(x_i/12) + q_3(x_i/12)^2 + \varepsilon_i$$



Linear Regression

Consider

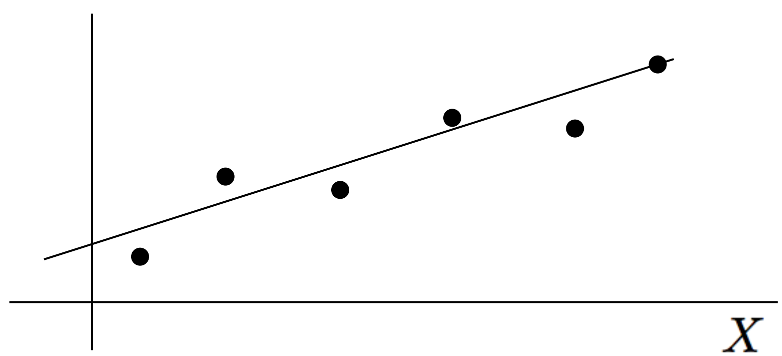
$$\Upsilon = Xq_0 + \varepsilon$$

where

$$\Upsilon = \begin{bmatrix} \Upsilon_1 \\ \vdots \\ \Upsilon_n \end{bmatrix}, \; X = \begin{bmatrix} X_{11} & \cdots & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & \cdots & X_{np} \end{bmatrix}, \; q_0 = \begin{bmatrix} q_1 \\ \vdots \\ q_p \end{bmatrix}, \; \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Observations — Design Matrix — Unknown Parameters — Errors

Example: $\Upsilon_i = (q_0 + q_1 X_i) + \varepsilon_i \; , \; i = 1, \cdots, n$



Linear Regression

Statistical Model:

$$\Upsilon = Xq_0 + \varepsilon$$

Assumptions:

(i) $\mathbb{E}(\varepsilon_i) = 0$

(ii) ε_i iid (independent and identically distributed)

$$\Rightarrow \quad \text{var}(\varepsilon_i) = \sigma_0^2$$

$$\mathbb{E}[(\varepsilon_i - \mathbb{E}(\varepsilon_i))(\varepsilon_j - \mathbb{E}(\varepsilon_j))] = \text{cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ for } i \neq j$$

(iii*) If $\varepsilon_i \sim N(0, \sigma_0^2)$ and hence $\Upsilon_i \sim N(Xq_0, \sigma_0^2)$, MLE provides expression for q_0 and σ_0^2.

(iv*) Generalized least squares employed if nonconstant variance.

Goals:

(1) Construct a 'good' estimator $\hat{q}$ for q.

(2) Construct an estimator $\hat{\sigma}^2$ for σ_0^2.

Least Squares Problem

Minimize

$$\mathcal{J}(q) = (\Upsilon - Xq)^T(\Upsilon - Xq)$$

Note: General result for quadratic forms

$$T = A^T \Phi A$$

$$\Rightarrow \nabla_q T = 2(\nabla_q A^T)\Phi A$$

Thus

$$\nabla_q \mathcal{J} = 2[\nabla_q(\Upsilon - Xq)^T][\Upsilon - Xq] = 0$$

where

$$\nabla_q(\Upsilon - Xq)^T = -\nabla_q q^T X^T = -X^T$$

Least Squares Estimate: $q_{OLS} = (X^T X)^{-1} X^T \upsilon$

Least Squares Estimator: $\hat{q}_{OLS} = (X^T X)^{-1} X^T \Upsilon$

Parameter Estimator Properties

Estimator Mean:

$$\begin{aligned}\mathbb{E}(\hat{q}) &= \mathbb{E}\left[(X^TX)^{-1}X^T\Upsilon\right] \\ &= (X^TX)^{-1}X^T\mathbb{E}(\Upsilon) \\ &= q_0\end{aligned}$$

Estimator Covariance Let $A = (X^TX)^{-1}X^T$

$$\begin{aligned}V(\hat{q}) = &\ \mathbb{E}[(\hat{q}-q_0)(\hat{q}-q_0)^T] \\ = &\ \mathbb{E}[(q_0 + A\varepsilon - q_0)(q_0 + A\varepsilon - q_0)^T]\ , \text{ since } \hat{q} = A\Upsilon = A(Xq_0 + \varepsilon) \\ = &\ A\mathbb{E}(\varepsilon\varepsilon^T)A^T \\ = &\ \sigma_0^2(X^TX)^{-1}\end{aligned}$$

Variance Estimator Properties

Goal: Construct an estimator $\hat{\sigma}^2$ for σ_0^2

Residual: $\widehat{R} = \Upsilon - X\hat{q}$

Variance Estimator: $\hat{\sigma}^2 = \frac{1}{n-p}\widehat{R}^T\widehat{R}$

Note:

$$\begin{aligned} \widehat{\Upsilon} &= \Upsilon - X(X^TX)^{-1}X^T\Upsilon \\ &= (I_n - H)\Upsilon \end{aligned}$$

where

$$H = X(X^TX)^{-1}X^T$$

Properties of H:

$$H^T = H \quad \text{(symmetric)}$$

$$H^2 = H \quad \text{(idempotent)}$$

$$(I_n - H)^2 = (I_n - H)$$

$$(I_n - H)X = X - X(X^TX)^{-1}X^TX = 0$$

Variance Estimator Properties

Note:

$$\begin{aligned}\widehat{R} &= (I_n - H)\Upsilon \\ &= (I_n - H)(Xq_0 + \varepsilon) \\ &= (I_n - H)\varepsilon\end{aligned}$$

so

$$\begin{aligned}\widehat{R}^T\widehat{R} &= \left[(I_n - H)\varepsilon\right]^T\left[(I_n - H)\varepsilon\right] \\ &= \varepsilon^T(I_n - H)\varepsilon\end{aligned}$$

Thus

$$\widehat{R}^T\widehat{R} = \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}\varepsilon_i\varepsilon_j$$

so

$$\begin{aligned}\mathbb{E}(\widehat{R}^T\widehat{R}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}\mathbb{E}(\varepsilon_i\varepsilon_j) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}\text{cov}(\varepsilon_i,\varepsilon_j) \;\; \text{since } \mathbb{E}(\varepsilon_i) = \mathbb{E}(\varepsilon_j) = 0 \\ &= \sum_{i=1}^{n} h_{ii}\text{var}(\varepsilon_i) = \sigma_0^2\text{tr}(I_n - H)\end{aligned}$$

Variance Estimator Properties

Note:

$$\mathrm{tr}(A+B) = \mathrm{tr}(B+A)$$

$$\mathrm{tr}(AB) = \mathrm{tr}(BA)$$

Thus

$$\begin{aligned} \mathrm{tr}(I_n - H) &= n - \mathrm{tr}\left[(X^T X)^{-1} X^T X\right] \\ &= n - p \end{aligned}$$

Unbiased Estimator:

$$S^2 = \frac{R^T R}{n-p}$$

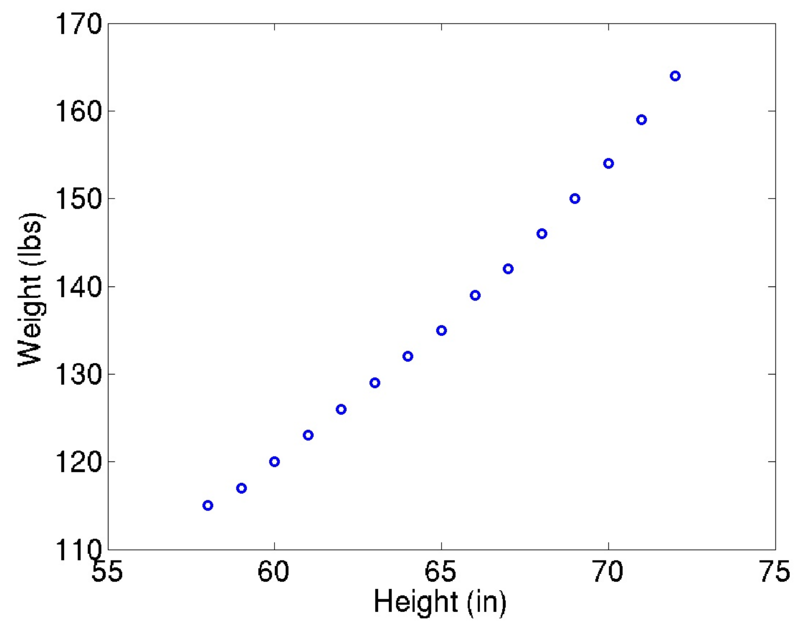
Example

Example: Consider the height-weight data from the *1975 World Almanac and Book of Facts*

Height (in)	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
Weight (lbs)	115	117	120	123	126	129	132	135	139	142	146	150	154	159	164

Consider the model

$$\Upsilon_i = q_1 + q_2(x_i/12) + q_3(x_i/12)^2 + \varepsilon_i$$



Example

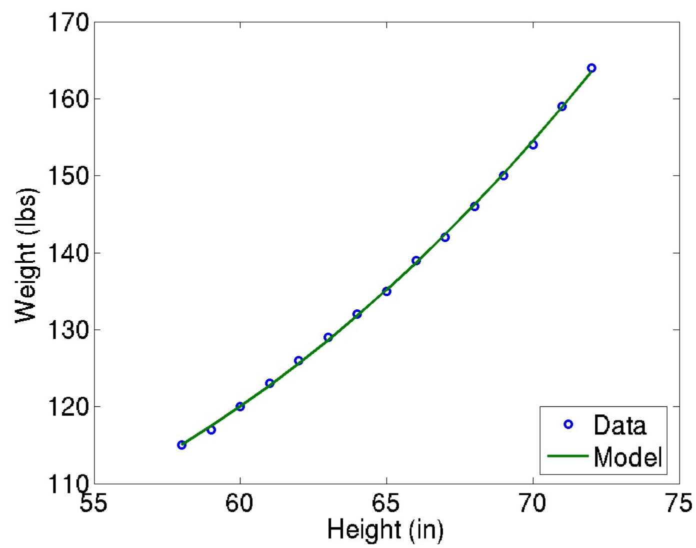
Here

$$X = \begin{bmatrix} 1 & 4.83 & 112.91 \\ 1 & 4.92 & 118.85 \\ 1 & 5.00 & 125.00 \\ 1 & 5.08 & 131.35 \\ 1 & 5.17 & 137.92 \\ 1 & 5.25 & 144.70 \\ 1 & 5.33 & 151.70 \\ 1 & 5.42 & 158.93 \\ 1 & 5.50 & 166.38 \\ 1 & 5.58 & 174.05 \\ 1 & 5.67 & 181.96 \\ 1 & 5.75 & 190.11 \\ 1 & 5.83 & 198.50 \\ 1 & 5.92 & 207.12 \\ 1 & 6.00 & 216.00 \end{bmatrix}$$

Least Square Estimate:

$$(X^T X)q = X^T v$$

$$\Rightarrow \quad \begin{aligned} q_1 &= 261.88 \\ q_2 &= -88.18 \\ q_3 &= 11.96 \end{aligned}$$



Note: $\mathrm{cond}(X^T X) = 6.7 \times 10^7$

Note: $\mathrm{eig}(H) = \mathrm{eig}(X(X^T X)^{-1} X^T) = 0, 1$

Example

Variance Estimate:

$$\sigma^2 = 0.15$$

Parameter Covariance Estimate:

$$V = \begin{bmatrix} 634.88 & -235.04 & 21.66 \\ -235.04 & 87.09 & -8.03 \\ 21.66 & -8.03 & 0.74 \end{bmatrix}$$

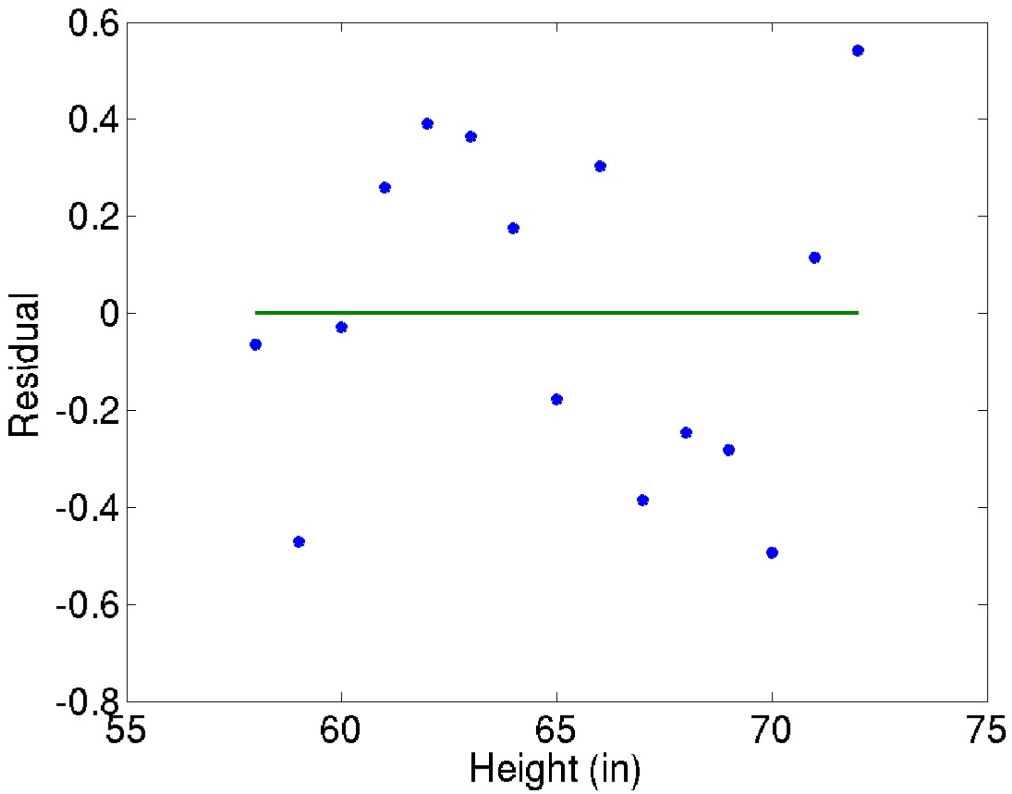
Note: This yields variances and standard deviations for parameter estimates

$$\begin{array}{lcl} q_1 = 261.88 \pm 50.39 & & q_1 \in [211.48, 312.27] \\ q_2 = -88.18 \pm 18.66 & \Rightarrow & q_2 \in [-106.84, -69.51] \\ q_3 = 11.96 \pm 1.72 & & q_3 \in [10.24, 13.68]. \end{array}$$

Goal: Can we additionally compute confidence intervals? Yes, but we need a little more statistics.

Example

Hypothesis: One way to check the hypothesis of iid is to plot the residuals



Random Variables Related to the Normal

Chi-Square Random Variables:

If $X \sim N(0,1)$, then the random variable $Y = X^2$ is said to have a χ^2-distribution with 1 dof. Furthermore, if $Y_i, i = 1, \cdots, n$ are independent χ^2 random variables with 1 dof, then their sum

$$\chi^2(n) = \sum_{i=1}^{n} Y_i$$

is a χ^2 random variable with n dof.

T Random Variables:

If $Z \sim N(0,1)$ and $\chi^2(n)$ is a chi-square random variable with n dof that is independent from Z, then

$$T = \frac{Z}{\sqrt{\chi^2(n)/n}}$$

has a (Student's) t-distribution with n dof.

Variance Estimator Properties

Assumption: Assume that $\varepsilon_j \sim N(0, \sigma_0^2)$ are iid (use asymptotic results if not normal).

Property 1: $(n-p)\hat{\sigma^2}/\sigma_0^2$ has a χ^2-distribution with $n-p$ dof.

Property 2: If δ_k denotes the k^{th} diagonal element of $(X^T X)^{-1}$, then

$$T_k = \frac{\hat{q}_k - q_{0_k}}{\hat{\sigma}\sqrt{\delta_k}}$$

has a t-distribution with $n-p$ dof.

Verification of 1: Recall that

$$\begin{aligned}
(n-p)\hat{\sigma}^2/\sigma_0^2 &= \frac{1}{\sigma_0^2}\widehat{R}^T\widehat{R} \\
&= \frac{1}{\sigma_0^2}\varepsilon^T(I_n - H)\varepsilon \\
&= \frac{1}{\sigma_0^2}\left\langle \varepsilon, (I_n - H)\varepsilon \right\rangle \\
&= \frac{1}{\sigma_0^2}\left\langle \varepsilon, U\Lambda U^T\varepsilon \right\rangle \quad , \; U \text{ orthogonal} \\
&= \frac{1}{\sigma_0^2}\left\langle U^T\varepsilon, \Lambda U^T\varepsilon \right\rangle
\end{aligned}$$

Variance Estimator Properties

Note: The eigenvalues of H are either 0 or 1 and $\mathrm{tr}(I_n - H) = \mathrm{rank}(I_n - H) = n - p$.
Thus

$$\Lambda = \begin{bmatrix} I_{n-p} & 0 \\ 0 & 0 \end{bmatrix}$$

Note: U^T orthogonal implies that $q = U^T\varepsilon \sim N(0, \sigma_0)$ since $\varepsilon \sim N(0, \sigma_0)$. Thus

$$\nu = \frac{(n-p)\hat{\sigma}^2}{\sigma_0^2} = \frac{\langle u, \Lambda u \rangle}{\sigma_0^2} = \sum_{i=1}^{n-p} \frac{u_i^2}{\sigma_0^2}$$

is the sum of squares of $n - p$ independent $N(0,1)$ random variables.

Result: $(n-p)\hat{\sigma}^2/\sigma_0^2$ is $\chi^2(n-p)$

Variance Estimator Properties

Verification of 2: Because $Z = \frac{\hat{q}_k - q_{0_k}}{\sigma_0 \sqrt{\delta_k}} \sim N(0,1)$, it follows that

$$
\begin{aligned}
T_k &= \frac{\hat{q}_k - q_{0_k}}{\hat{\sigma}\sqrt{\delta_k}} \\
&= \frac{\hat{q}_k - q_{0_k}}{\sigma_0 \sqrt{\delta_k}} \frac{\sigma_0}{\hat{\sigma}\sqrt{n-p}} \cdot \sqrt{n-p} \\
&= \frac{Z}{\sqrt{\nu/(n-p)}} \quad , \; Z \sim N(0,1) \quad , \; \nu \sim \chi^2(n-p),
\end{aligned}
$$

Result: T_k has a t-distribution with $n-p$ dof.

Confidence Interval: Because

$$P\left(\hat{q}_k - t_{n-p,1-\alpha/2} \cdot \hat{\sigma}\sqrt{\delta_k} < q_{0_k} < \hat{q}_k + t_{n-p,1-\alpha/2} \cdot \hat{\sigma}\sqrt{\delta_k}\right) = 1 - \alpha$$

the $(1-\alpha) \times 100\%$ confidence interval is

$$\left[q_k - t_{n-p,1-\alpha/2} \cdot \sigma\sqrt{\delta_k}, q_k + t_{n-p,1-\alpha/2} \cdot \sigma\sqrt{\delta_k}\right]$$

Example

Previous Example:

For $\alpha = .05$, $t_{n-p,1-\alpha/2} = 2.2$. The 95% confidence intervals are thus

$$q_1 \in [206.45, 317.31]$$

$$q_2 \in [-108.71, -67.65]$$

$$q_3 \in [10.07, 13.86].$$

Note: This is consistent with the results on Slide 11 with $2\sigma = 94.45\%$.

Summary of Linear Theory

Statistical Model:

$$\Upsilon = Xq_0 + \varepsilon \ , \ q \in \mathbb{R}^p$$

$$\upsilon = Xq_0 + \epsilon \ , \ \text{(realization)}$$

Assumptions: $\mathbb{E}(\varepsilon_i) = 0$, ε_i iid with $\text{var}(\varepsilon_i) = \sigma_0^2$

Least Squares Estimator and Estimate:

$$\hat{q} = (X^T X)^{-1} X^T \Upsilon \quad , \ \mathbb{E}(\hat{q}) = q_0 \ , \ V(\hat{q}) = \sigma_0^2 (X^T X)^{-1}$$

$$q = (X^T X)^{-1} X^T \upsilon$$

Error Variance Estimator and Estimate: $\widehat{R} = \Upsilon - X\hat{q}$, $R = \upsilon - Xq$

$$\hat{\sigma}^2 = \frac{1}{n-p} \widehat{R}^T \widehat{R} \quad , \quad \sigma^2 = \frac{1}{n-p} R^T R$$

Covariance Matrix Estimator and Estimate:

$$V(\hat{q}) = \hat{\sigma}^2 (X^T X)^{-1} \quad , \quad V = \sigma^2 (X^T X)^{-1}$$

Sampling Distribution: Requires $\varepsilon_i \sim N(0, \sigma_0^2)$ or sufficiently large n

- $\hat{q} \sim N(q_0, \sigma_0^2 (X^T X)^{-1})$
- $(1-\alpha) \times 100\%$ Confidence Intervals: $\delta_k = [(X^T X)^{-1}]_{kk}$

$$\left[q_k - t_{n-p, 1-\alpha/2} \sigma \sqrt{\delta_k} \ , \ q_k + t_{n-p, 1-\alpha/2} \sigma \sqrt{\delta_k} \right]$$

Summary of Nonlinear Theory

Statistical Model:

$$\Upsilon = f(q_0) + \varepsilon \ , \ q \in \mathbb{R}^p, \ \upsilon \in \mathbb{R}^1$$

$$\upsilon = f(q_0) + \epsilon \ , \ \text{(realization)}$$

Assumptions: $\mathbb{E}(\varepsilon_i) = 0$, ε_i iid with $\text{var}(\varepsilon_i) = \sigma_0^2$

Least Squares Estimator and Estimate:

$$\hat{q} = \underset{q \in \mathcal{Q}}{\text{argmin}} \sum_{i=1}^{n} [\Upsilon_i - f_i(q)]^2 \quad , \quad q = \underset{q \in \mathbb{Q}}{\text{argmin}} \sum_{i=1}^{n} [\upsilon_i - f_i(q)]^2$$

Error Variance Estimator and Estimate: $\widehat{R} = \Upsilon - f(\hat{q})$, $R = \upsilon - f(q)$

$$\hat{\sigma}^2 = \frac{1}{n-p} \widehat{R}^T \widehat{R} \quad , \quad \sigma^2 = \frac{1}{n-p} R^T R$$

Covariance Matrix Estimator and Estimate: $\mathcal{X}_{ik}(q) = \frac{\partial f_i(q)}{\partial q_k}$

$$V(\hat{q}) = \hat{\sigma}^2 [\mathcal{X}^T(\hat{q}) \mathcal{X}(\hat{q})]^{-1} \quad , \quad V = \sigma^2 [\mathcal{X}^T(q) \mathcal{X}(q)]^{-1}$$

Statistical Properties: Requires $\varepsilon_i \sim N(0, \sigma_0^2)$ or sufficiently large n

- $\hat{q} \sim N\left(q_0, \sigma_0^2 \left[\mathcal{X}^T(q_0) \mathcal{X}(q_0)\right]^{-1}\right)$
- $(1-\alpha) \times 100\%$ Confidence Intervals: $\delta_k = [(\mathcal{X}^T(q)\mathcal{X}(q))^{-1}]_{kk}$

$$\left[q_k - t_{n-p, 1-\alpha/2} \sigma \sqrt{\delta_k} \ , \ q_k + t_{n-p, 1-\alpha/2} \sigma \sqrt{\delta_k}\right]$$

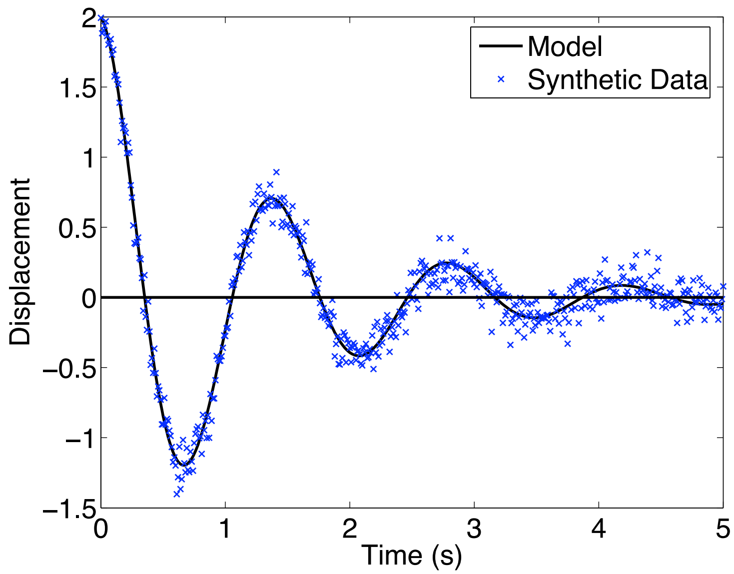
Nonlinear Parameter Estimation

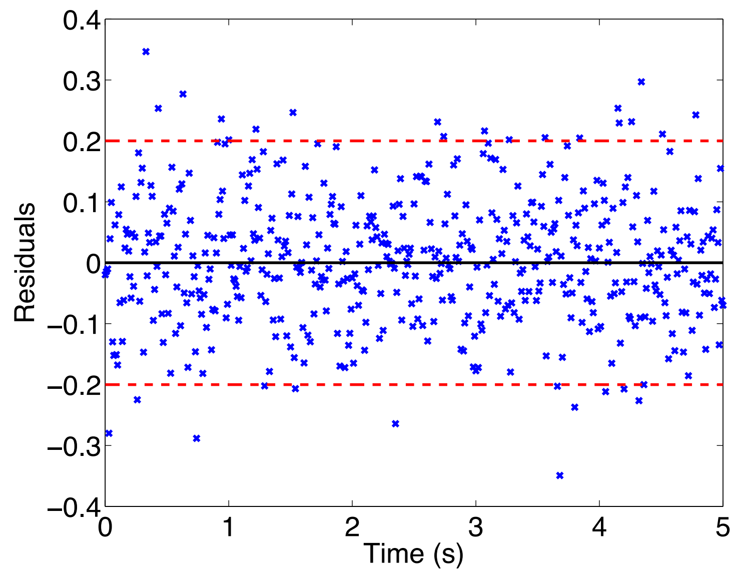
Example: Consider the spring model

$$\ddot{z} + C\dot{z} + Kz = 0$$
$$z(0) = 2 \; , \; \dot{z}(0) = -C$$

with displacement observations y(t,q) = z(t,q), parameter q = C, and stiffness K=20.5. The solution is

$$z(t) = 2e^{-Ct/2}\cos\left(\sqrt{K - C^2/4}\cdot t\right)$$





Synthetic Data: Generated with $\sigma_0 = 0.1$

Nonlinear Parameter Estimation

Example: The sensitivity matrix is

$$\mathcal{X}(q) = \left[\frac{\partial y}{\partial C}(t_1, q), \cdots, \frac{\partial y}{\partial C}(t_n, q)\right]^T$$

where

$$\frac{\partial y}{\partial C} = e^{-Ct/2}\left[\frac{Ct}{\sqrt{4K - C^2}}\sin\left(\sqrt{K - C^2/4}\cdot t\right) - t\cos\left(\sqrt{K - C^2/4}\cdot t\right)\right]$$

Here

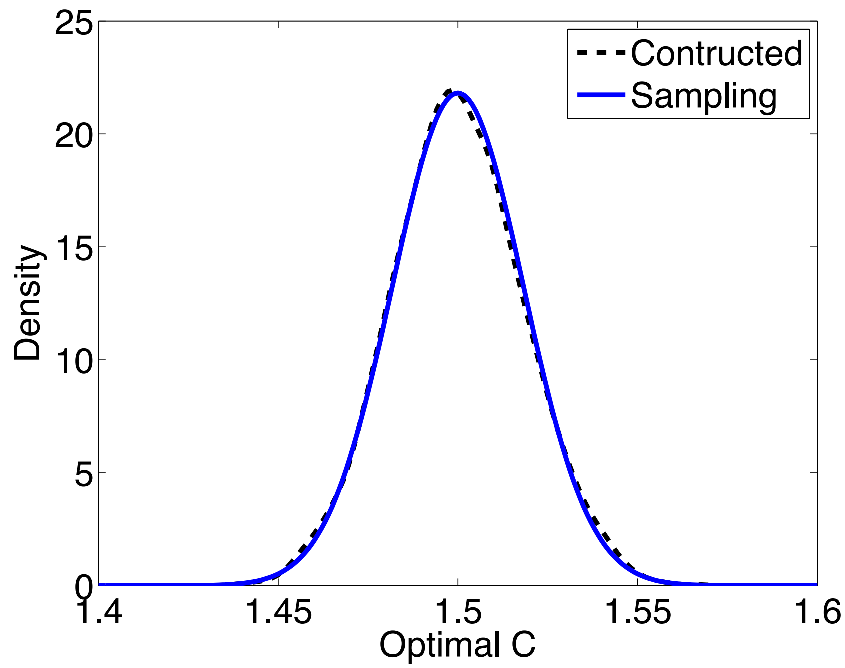
$$V = \sigma_c^2 = \sigma_0^2\left[\mathcal{X}^T(q)\mathcal{X}(q)\right]^{-1} = 3.35 \times 10^{-4}$$

so that

$$\widehat{C} \sim N\left(C_0, \sigma_c^2\right) \;,\; \sigma_c = 0.0183$$

Note: In 10,000 simulations, 9455 of confidence intervals contained true parameter value.

Figure: Sampling distribution compared with that constructed using 10,000 estimated values of C.



Nonlinear Parameter Estimation

Example: Consider the heat model

$$\frac{d^2 T_s}{dx^2} = \frac{2(a+b)}{ab}\frac{h}{k}\left[T_s(x) - T_{amb}\right]$$

$$\frac{dT_s}{dx}(0) = \frac{\Phi}{k} \quad , \quad \frac{dT_s}{dx}(L) = \frac{h}{k}\left[T_{amb} - T_s(L)\right]$$

with parameters $q = [\Phi, h]$. The observed solution is

$$y_i(q) = T_s(x_i, q) = c_1(q)e^{-\gamma x_i} + c_2(q)e^{\gamma x_i} + T_{amb}$$

where $\gamma = \sqrt{\frac{2(a+b)h}{abk}}$ and

$$c_1(q) = -\frac{\Phi}{k\gamma}\left[\frac{e^{\gamma L}(h+k\gamma)}{e^{-\gamma L}(h-k\gamma) + e^{\gamma L}(h+k\gamma)}\right] \quad , \quad c_2(q) = \frac{\Phi}{k\gamma} + c_1(q)$$

Data:

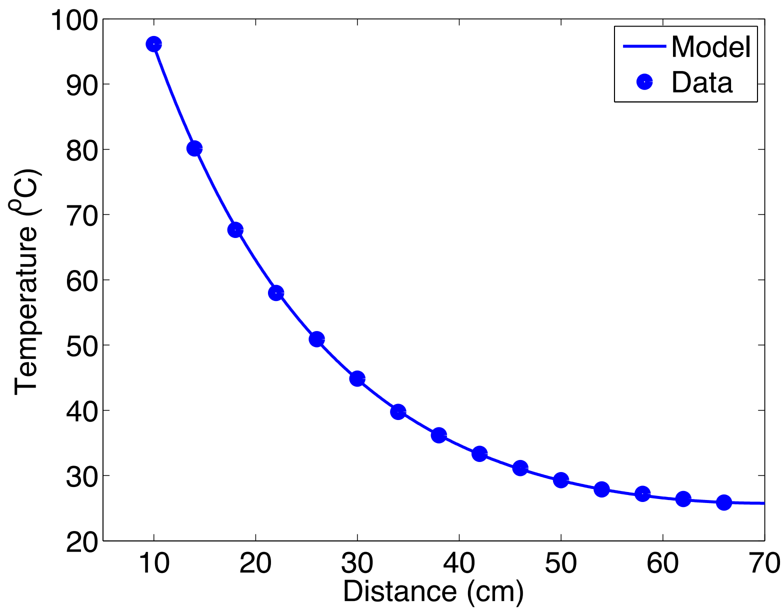
x (cm)	10	14	18	22	26	30	34	38
Temp (^{o}C)	96.14	80.12	67.66	57.96	50.90	44.84	39.75	36.16

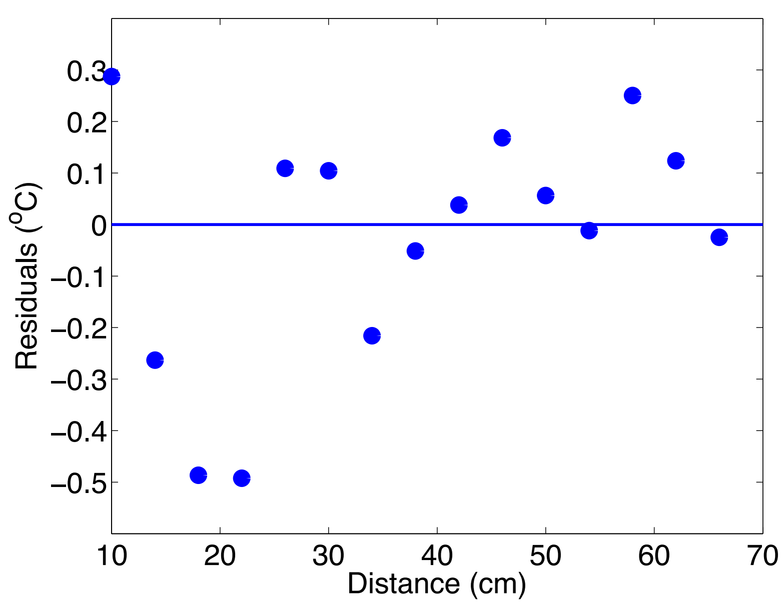
x (cm)	42	46	50	54	58	62	66	
Temp (^{o}C)	33.31	31.15	29.28	27.88	27.18	26.40	25.86	

Nonlinear Parameter Estimation

Heat Example: Standard deviations

$$\sigma = 0.2504 \,, \sigma_\Phi = 0.1450 \,, \; \sigma_h = 1.4482 \times 10^{-5}$$





Note: We will revisit this example in the context of Bayesian inference.