

Lecture 5
Estimation
with
Info-Gap Uncertainties

Yakov Ben-Haim
Technion
Israel Institute of Technology



Contents

1	Linear Regression	3
2	Estimating an Uncertain Probability Density	9

1 Linear Regression

§ **Modelling is a decision problem.** We will consider 2 examples:

- Modelling a mechanical S-N curve.
- Modelling the economic Phillips curve.

§ **Mechanical S-N curve:**

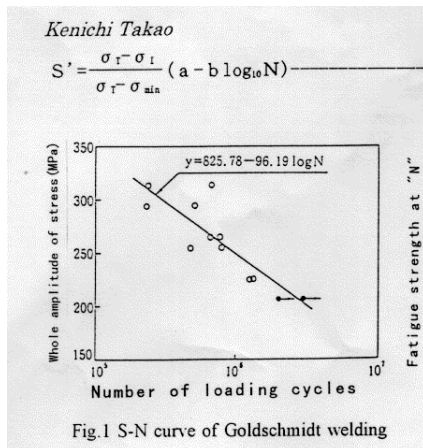


Figure 1: S-N curves.

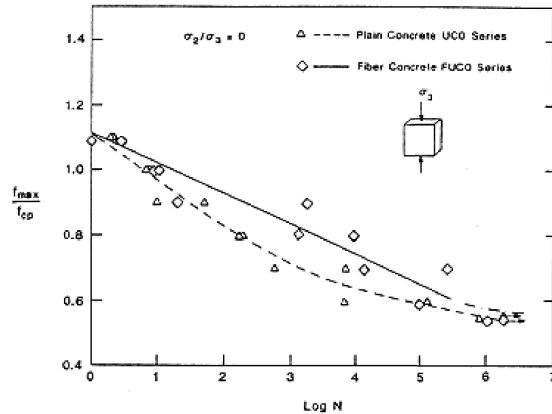


Figure 2: S-N curves.

§ **Challenge: Two foci of uncertainty:**

- **Randomness:**
 - Noisy data (statistics).
- **Info-gaps:**
 - Changing fundamentals.
 - Material variability.
 - Environmental variability.

§ **Questions:**

- How to use empirical data to model uncertain material?
- Is optimal estimation (e.g. least-squares) a good strategy?
- Can we do better?
- How to manage both statistical and info-gap uncertainty?
- How to evaluate estimate vis a vis info-gaps?

§ Economic Phillips curve:

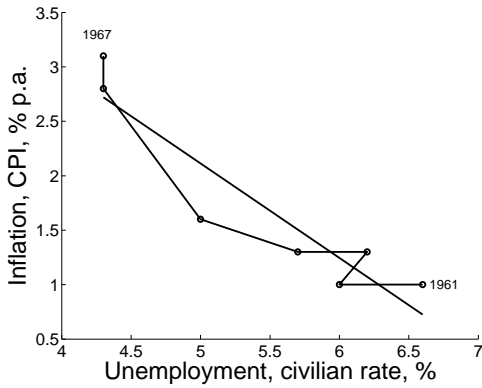


Figure 3: Inflation vs. unemployment in the US, 1961–1967.



Figure 4: Inflation vs. unemployment in the US, 1961–1993.

§ Inflation vs. unemployment, US, '61–'67:

- Approximately linear.
- Slope ≈ -0.87 %CPI/%unemployment.

§ Slopes in other periods:

- '61–'67: -0.87
- '80–'83: -3.34
- '85–'93: -1.08
- '70–'78: ???

§ Challenge: Two foci of uncertainty:

- Randomness:
 - Noisy data (statistics).
- Info-gaps:
 - Changing fundamentals.
 - Data revision.

§ Questions:

- How to use historical data to model the future?
- Is optimal estimation (e.g. least-squares) a good strategy?
- Can we do better?
- How to manage both statistical and info-gap uncertainty?
- How to evaluate estimate vis a vis info-gaps?

§ Paired data, fig. 5:

- CPI, system lifetime, etc: c_1, \dots, c_n .
- Unemployment, mechanical stress, etc: u_1, \dots, u_n .

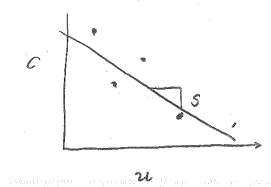


Figure 5: Paired data.

§ **Least-squares estimate of slope:**

- Linear regression:

$$c = su + b \tag{1}$$

- Mean squared error:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N [c_i - (su_i + b)]^2 \tag{2}$$

- MSE estimate of the slope:

$$\tilde{s} = \arg \min_s \text{MSE} \tag{3}$$

One finds:

$$\tilde{s} = \frac{\text{cov}(u, c)}{\text{var}(u)} \tag{4}$$

where:

$$\text{cov}(u, c) = \frac{1}{n} \sum_{i=1}^n c_i u_i - \left(\frac{1}{n} \sum_{i=1}^n c_i \right) \left(\frac{1}{n} \sum_{i=1}^n u_i \right) \tag{5}$$

and $\text{var}(u) = \text{cov}(u, u)$.

- In our case, fig. 5, $\tilde{s} < 0$.

§ **Robustness question:**

How much can the data err due to info-gaps, and the slope's error will be acceptable?

§ **Moments:**

$\gamma = \text{covariance, cov}(u, c)$. $\tilde{\gamma} = \text{estimate}$.

$\sigma^2 = \text{variance, var}(u)$. $\tilde{\sigma}^2 = \text{estimate}$.

§ **Consider info-gap in data.** Specifically, unknown fractional errors of moments:

$$\left| \frac{\gamma - \tilde{\gamma}}{\tilde{\gamma}} \right|, \quad \left| \frac{\sigma^2 - \tilde{\sigma}^2}{\tilde{\sigma}^2} \right| \tag{6}$$

§ **Fractional-error info-gap model:**

$$\mathcal{U}(h) = \left\{ (\gamma, \sigma^2) : \left| \frac{\gamma - \tilde{\gamma}}{\tilde{\gamma}} \right| \leq h, \quad \left| \frac{\sigma^2 - \tilde{\sigma}^2}{\tilde{\sigma}^2} \right| \leq h, \quad \sigma^2 \geq 0 \right\}, \quad h \geq 0$$

§ **Least-squares estimate:** $\tilde{s} = \tilde{\gamma}/\tilde{\sigma}^2$.

Actual value: $s = \gamma/\sigma^2$.

§ **Performance requirement:** $|s(\gamma, \sigma^2) - \tilde{s}| \leq r_c$.

§ **Robustness of LS estimate \tilde{s} :**

Max horizon of uncertainty in moments
at which \tilde{s} errs no more than r_c :

$$\hat{h}(\tilde{s}, r_c) = \max \left\{ h : \left(\max_{\gamma, \sigma^2 \in \mathcal{U}(h)} |s(\gamma, \sigma^2) - \tilde{s}| \right) \leq r_c \right\} \quad (7)$$

§ **Derivation of the robustness:**

- $m(h)$ = inner maximum in eq.(7).
- $m(h)$ occurs at $\gamma = (1 + h)\tilde{\gamma}$, $\sigma^2 = (1 - h)^2\tilde{\sigma}^2$.
- Thus, for $h \leq 1$:

$$m(h) = \left| \frac{(1 + h)\tilde{\gamma}}{(1 - h)\tilde{\sigma}^2} - \frac{\tilde{\gamma}}{\tilde{\sigma}^2} \right| \quad (8)$$

$$= \left(\frac{1 + h}{1 - h} - 1 \right) \left| \frac{\tilde{\gamma}}{\tilde{\sigma}^2} \right| \quad (9)$$

$$= \frac{2h}{1 - h} |\tilde{s}| \quad (10)$$

- Equate $m(h) = r_c$ and solve for h (recall $\tilde{s} < 0$):

$$\frac{2h}{1 - h} = -\frac{r_c}{\tilde{s}} = \rho \text{ (definition)} \implies \hat{h} = \frac{\rho}{2 + \rho} \ (\leq 1) \quad (11)$$

§ **Robustness of LS estimate \tilde{s} :**

$$\hat{h}(\tilde{s}, \rho) = \frac{\rho}{2 + \rho}, \quad \rho = -r_c/\tilde{s} \quad (12)$$

Recall: $\tilde{s} < 0$ so $\rho > 0$.

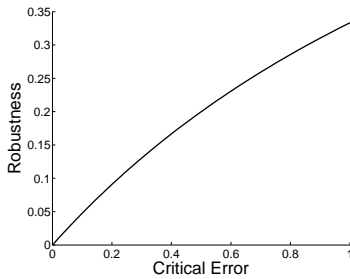


Figure 6: Robustness of estimated slope, $\hat{h}(\tilde{s}, \rho)$, vs. critical error, ρ . Eq.(12).

- Best-estimate: zero robustness.
- Trade-off: robustness vs. estim. error.
- Example: $\rho = 0.2$, $\hat{h} = 0.09$.

§ **Can we do better than LS estimate?**

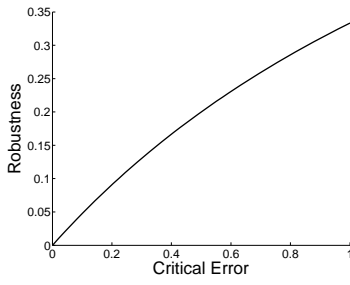


Figure 7: $\hat{h}(\tilde{s}, \rho)$ vs. ρ .

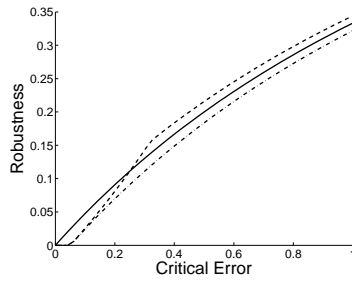


Figure 8: $\hat{h}(s_e, \rho)$ vs. ρ . $\zeta = 1$ (solid), 1.05 (dash), 0.95 (dot dash).

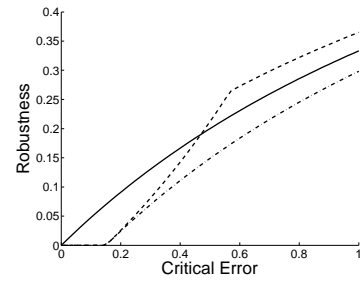


Figure 9: $\hat{h}(s_e, \rho)$ vs. ρ . $\zeta = 1$ (solid), 1.15 (dash), 0.85 (dot-dash).

§ Estimates of Phillips slope:

- \tilde{s} = LS estimate, with robustness $\hat{h}(\tilde{s}, \mathcal{R}_C)$.
- s_e = any estimate, with robustness $\hat{h}(s_e, \mathcal{R}_C)$.
- Definitions: $\zeta = s_e/\tilde{s}$, $\rho = -\mathcal{R}_C/\tilde{s}$. (Recall: $\tilde{s} < 0$.)
- Robustness of s_e , in analogy to eq.(7):

$$\hat{h}(s_e, \mathcal{R}_C) = \max \left\{ h : \left(\max_{\gamma, \sigma^2 \in \mathcal{U}(h)} |s(\gamma, \sigma^2) - s_e| \right) \leq \mathcal{R}_C \right\} \quad (13)$$

- Let $m(h)$ denote the inner minimum:

$$m(h) = \max_{\gamma, \sigma^2 \in \mathcal{U}(h)} \left| \frac{\gamma}{\sigma^2} - s_e \right| \quad (14)$$

- For $h \leq 1$ this occurs at one of the following:

$$\text{Either: } \gamma = (1 + h)\tilde{\gamma}, \quad \sigma^2 = (1 - h)\tilde{\sigma}^2 \quad (15)$$

$$\text{Or: } \gamma = (1 - h)\tilde{\gamma}, \quad \sigma^2 = (1 + h)\tilde{\sigma}^2 \quad (16)$$

- Denote the corresponding $m(h)$'s:

$$m_1(h) = \left| \frac{(1 + h)\tilde{\gamma}}{(1 - h)\tilde{\sigma}^2} - s_e \right| \quad (17)$$

$$m_2(h) = \left| \frac{(1 - h)\tilde{\gamma}}{(1 + h)\tilde{\sigma}^2} - s_e \right| \quad (18)$$

- $m(h)$ is the greater of these two alternatives:

$$m(h) = \max[m_1(h), m_2(h)] \quad (19)$$

The maximum depends on the value of h .

- After some algebra, and equating $m(h) = \mathcal{R}_C$, one finds:

$$\hat{h}(s_e, \rho) = \begin{cases} \frac{\rho + \zeta - 1}{\rho + \zeta + 1} & \text{if } \rho^2 \geq \zeta^2 - 1 \text{ and } \rho \geq 1 - \zeta \\ \frac{\rho - \zeta + 1}{-\rho + \zeta + 1} & \text{if } \rho^2 \leq \zeta^2 - 1 \text{ and } \rho \geq \zeta - 1 \end{cases} \quad (20)$$

$\hat{h}(s_e, \rho)$ is zero otherwise. Note $\hat{h} \leq 1$.

- Eq.(20) includes eq.(12) as a special case, when $\zeta = 1$.
- When $\zeta > 1$, the robustness follows the lower line of eq.(20) (which has greater slope than the robustness curve for \tilde{s}) for small ρ , and then follows the upper line of the equation for larger ρ . This causes crossing of robustness curves as illustrated by the solid and dashed lines in figs. 8 and 9. (The two lines in eq.(20) are equal when $\rho^2 = \zeta^2 - 1$.)
- LS estimate: 0 error, 0 robustness.
- Trade-off: robustness vs. estim. error.
- Curve crossing: preference reversal.

§ **Can we do better than least-squares?** Yes, but at a price:

Robust-satisficing estimate is more robust to uncertainty at positive estimation error.

2 Estimating an Uncertain Probability Density

¶ The problem:

- Estimate parameters of a probability density function (pdf) based on observations.
- Common approach: select parameter values to maximize the likelihood function for the class of pdfs.
- In this section: simple example of a situation where the **form** of the pdf is uncertain, not only **parameters**.

¶ Notation:

- x = random variable.
- $X = (x_1, \dots, x_N)$ = random sample.
- $\tilde{p}(x|\lambda)$ = be a pdf for x with parameters λ .

¶ Likelihood function:

$$L(X, \tilde{p}) = \prod_{i=1}^N \tilde{p}(x_i|\lambda) \quad (21)$$

¶ Maximum likelihood estimate (MLE):

$$\lambda^* = \arg \max_{\lambda} L(X, \tilde{p}) \quad (22)$$

¶ Examples of MLE.

- **Exponential distribution:** The pdf is:

$$\tilde{p}(x|\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0 \quad (23)$$

The likelihood function, from eq.(21), is:

$$L = \prod_{i=1}^N \tilde{p}(x_i|\lambda) = \lambda^N \exp\left(-\lambda \sum_{i=1}^N x_i\right) \quad (24)$$

Thus:

$$\frac{\partial L}{\partial \lambda} = \left(N\lambda^{N-1} - \lambda^N \sum_{i=1}^N x_i \right) \exp\left(-\lambda \sum_{i=1}^N x_i\right) \quad (25)$$

Equating to zero and solving for λ yields the MLE:

$$0 = \frac{\partial L}{\partial \lambda} \implies 0 = N\lambda^{N-1} - \lambda^N \sum_{i=1}^N x_i \implies \boxed{\frac{1}{\lambda_{\text{MLE}}} = \frac{1}{N} \sum_{i=1}^N x_i} \quad (26)$$

Note that:

$$E(x) = \frac{1}{\lambda} \quad (27)$$

- **Normal distribution: MLE of the mean.** The pdf is:

$$\tilde{p}(x|\lambda) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2} \quad (28)$$

The likelihood function, from eq.(21), is:

$$L = \prod_{i=1}^N \tilde{p}(x_i|\lambda) = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2\right) \quad (29)$$

Note that:

$$\mu_{MLE} = \arg \max_{\mu} L = \arg \min_{\mu} \sum_{i=1}^N (x_i - \mu)^2 = \text{Least Squares Estimate} \quad (30)$$

Thus MLE and LSE agree. Define the squared error:

$$S = \sum_{i=1}^N (x_i - \mu)^2 \quad (31)$$

Thus:

$$\frac{\partial S}{\partial \mu} = 0 = -2 \sum_{i=1}^N (x_i - \mu) \implies \boxed{\mu_{MLE} = \frac{1}{N} \sum_{i=1}^N x_i} \quad (32)$$

¶ **Robust-satisficing:**

- Form of the pdf is not certain.
- $\tilde{p}(x|\lambda)$ is most reasonable choice of the form of the pdf. We will estimate λ .
- Actual form of the pdf is unknown.
- We wish to choose those parameters to:
 - Satisfice the likelihood.
 - To be *robust* to the info-gaps in the shape of the actual pdf which generated the data, or which might generate data in the future.

¶ **Info-gap model:**

$$\mathcal{U}(h, \tilde{p}) = \{p(x) : p(x) \in \mathcal{P}, |p(x) - \tilde{p}(x|\lambda)| \leq h\psi(x)\}, \quad h \geq 0 \quad (33)$$

- \mathcal{P} is the set of all normalized and non-negative pdfs on the domain of x .
- $\psi(x)$ is the known envelope function. E.g. $\psi(x) = 1$, implying severe uncertainty on tail.
- h is the unknown horizon of uncertainty.

¶ **Question:**

Given the random sample X , and the info-gap model $\mathcal{U}(h, \tilde{p})$, how should we choose the parameters of the nominal pdf $\tilde{p}(x|\lambda)$?

¶ **Robustness:**

$$\hat{h}(\lambda, L_c) = \max \left\{ h : \left(\min_{p \in \mathcal{U}(h, \tilde{p})} L(X, p) \right) \geq L_c \right\} \quad (34)$$

¶ $m(h) =$ **inner minimum** in eq.(34).

For the info-gap model in eq.(33) $m(h)$ is obtained for the following choices of the pdf at the data points X :

$$p(x_i) = \begin{cases} \tilde{p}(x_i) - h\psi(x_i) & \text{if } h \leq \tilde{p}(x_i)/\psi(x_i) \\ 0 & \text{else} \end{cases} \quad (35)$$

Choose $p(x) = \tilde{p}(x)$ for all other x 's.

Define:

$$h_{\max} = \min_i \frac{\tilde{p}(x_i)}{\psi(x_i)} \quad (36)$$

Since $m(h)$ is the product of the densities in eq.(35) we find:

$$m(h) = \begin{cases} \prod_{i=1}^N [\tilde{p}(x_i) - h\psi(x_i)] & \text{if } h \leq h_{\max} \\ 0 & \text{else} \end{cases} \quad (37)$$

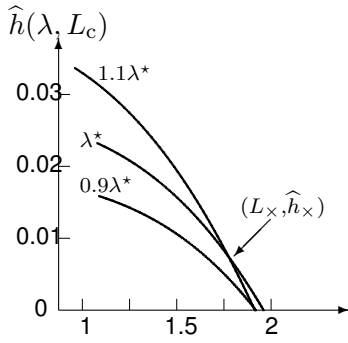
¶ $m(h)$ and $\hat{h}(\lambda, L_c)$:

- Robustness is the max h at which $m(h) \geq L_c$.
- $m(h)$ strictly decreases as h increases.
- Hence robustness is the solution of $m(h) = L_c$.
- Hence $m(h)$ is the inverse of $\hat{h}(\lambda, L_c)$:

$$m(h) = L_c \quad \text{implies} \quad \hat{h}(\lambda, L_c) = h \quad (38)$$

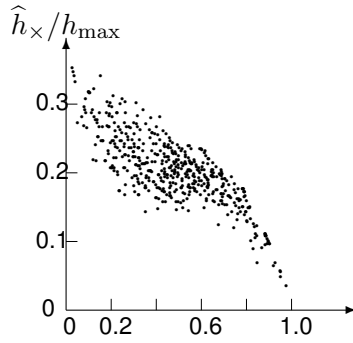
- Plot of $m(h)$ vs. h is plot of L_c vs. $\hat{h}(\lambda, L_c)$.

Robustness



Critical likelihood, $\log_{10} L_c$

Figure 10: Robustness curves. $\lambda^* = 3.4065$.



$L_x/L[X, \tilde{p}(x|\lambda^*)]$

Figure 11: Loci of intersection of robustness curves $\hat{h}(\lambda^*, L_c)$ and $\hat{h}(1.1\lambda^*, L_c)$.

¶ Robustness curves in fig. 10 based on:

- Eqs.(37) and (38).
- Nominal pdf is exponential, $\tilde{p}(x|\lambda) = \lambda \exp(-\lambda x)$ with $\lambda = 3$.
- Envelope function is constant, $\psi(x) = 1$. Note severe uncertainty on the tail.
- Random sample, X , with $N = 20$.
- MLE of λ , eq.(22): $\lambda^* = 1/\bar{x}$ where $\bar{x} = (1/N) \sum_{i=1}^N x_i$ is the sample mean.
- Robustness curves for 3 λ 's: $0.9\lambda^*$, λ^* , and $1.1\lambda^*$.

¶ Robustness of the estimated likelihood is zero for any λ :

- Likelihood function for λ is $L[X, \tilde{p}(x|\lambda)]$.
- Each curve in fig.10, $\hat{h}(\lambda, L_c)$ vs. L_c , hits horizontal axis when $L_c =$ likelihood:

$$\hat{h}(\lambda, L_c) = 0 \quad \text{if} \quad L_c = L[X, \tilde{p}(x|\lambda)] \tag{39}$$

- λ^* is the MLE of λ . Thus $\hat{h}(\lambda^*, L_c)$ hits horizontal axis to the right of $\hat{h}(\lambda, L_c)$.

¶ Preferences between estimates of λ :

- $\hat{h}(\lambda^*, L_c) > \hat{h}(0.9\lambda^*, L_c) \implies \lambda^* \succ 0.9\lambda^*$.
- $\hat{h}(\lambda^*, L_c)$ and $\hat{h}(1.1\lambda^*, L_c)$ cross at (L_x, \hat{h}_x) :
 - $\lambda^* \succ 1.1\lambda^*$ for $L_c > L_x$ and $h < h_x$.
 - $1.1\lambda^* \succ \lambda^*$ else.

¶ 500 repetitions:

- λ^* dominates $0.9\lambda^*$.
- Preferences reverse between λ^* and $1.1\lambda^*$.
- Normalized (h_x, L_x) in fig. 11.
- Center of cloud: (0.5, 0.2). Typical cross of robustness curves at:
 - L_c about half of best-estimated value.
 - \hat{h} about 20% of maximum robustness.

¶ Past and future data-generating processes:

- Data in this example generated from exponential distribution.
- Nothing in data to suggest that exponential distribution is wrong.
- Motivation for info-gap model, eq.(33), is that,
 - while the *past* has been exponential,
 - the *future* may not be.
- The robust-satisficing estimate of λ accounts not only for the historical evidence (the sample X) but also for the future uncertainty about relevant family of distributions.