Lecture 5

Estimation

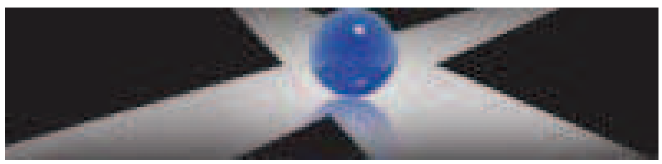
with

Info-Gap Uncertainties

Yakov Ben-Haim

Technion

Israel Institute of Technology



Contents

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:**

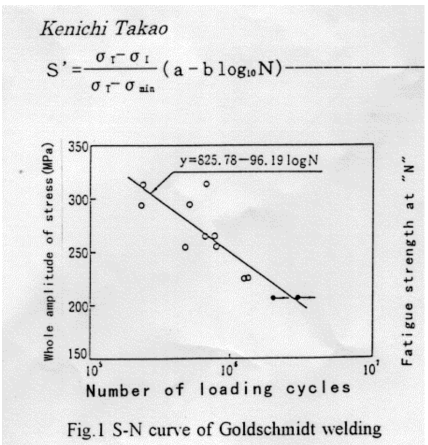


Fig.1 S-N curve of Goldschmidt welding

Figure 1: S-N curves.

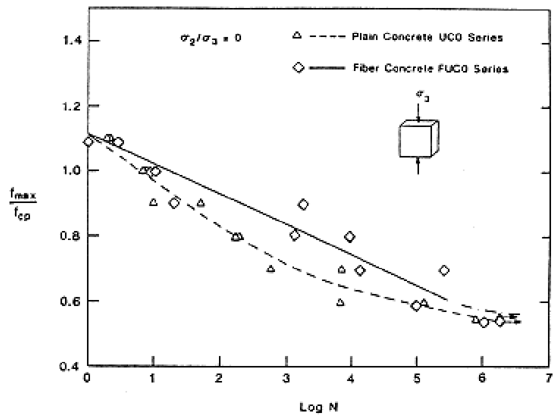


Figure 2: S-N curves.

§ 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?

Source: Yakov Ben-Haim, 2010, *Info-Gap Economics: An Operational Introduction,* Palgrave-Macmillan.

§ **Economic Phillips curve:**

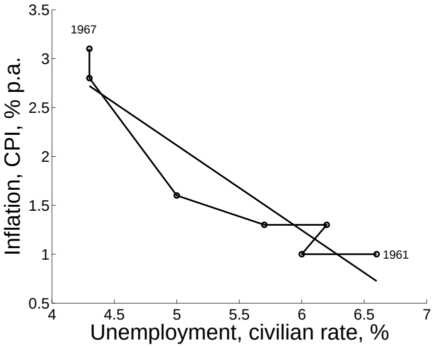


Figure 3: Inflation vs. unemployment in the US, 1961–1967.

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: $c_1, \ldots, c_n$.
- Unemployment, mechanical stress, etc: $u_1, \ldots, u_n$.

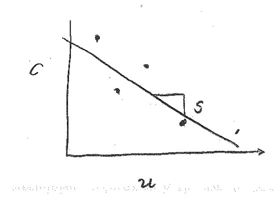


Figure 5: Paired data.

§ **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:

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

One finds:

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

where:

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

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

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

§ **Robustness question:**

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

§ **Moments:**

$\gamma =$ covariance, $\mathrm{cov}(u, c)$. $\widetilde{\gamma} =$ estimate.
$\sigma^2 =$ variance, $\mathrm{var}(u)$. $\widetilde{\sigma}^2 =$ estimate.

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

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

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

$$\mathcal{U}(h) = \left\{ (\gamma, \sigma^2) : \ \left|\frac{\gamma - \widetilde{\gamma}}{\widetilde{\gamma}}\right| \le h, \ \ \left|\frac{\sigma^2 - \widetilde{\sigma}^2}{\widetilde{\sigma}^2}\right| \le h, \ \sigma^2 \ge 0 \right\}, \quad h \ge 0$$

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

§ **Performance requirement:** $|s(\gamma, \sigma^2) - \widetilde{s}| \le r_{\mathrm{c}}$.

§ **Robustness of LS estimate $\widetilde{s}$:**

Max horizon of uncertainty in moments
at which $\widetilde{s}$ errs no more than $r_{\rm c}$:

$$\widehat{h}(\widetilde{s}, r_{\rm c}) = \max\left\{h:\ \left(\max_{\gamma,\sigma^2\in\mathcal{U}(h)} |s(\gamma,\sigma^2)-\widetilde{s}|\right)\le r_{\rm c}\right\} \tag{7}$$

§ **Derivation of the robustness:**

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

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

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

$$= \frac{2h}{1-h}|\widetilde{s}| \tag{10}$$

- Equate $m(h)=r_{\rm c}$ and solve for h (recall $\widetilde{s}<0$):

$$\frac{2h}{1-h} = -\frac{r_{\rm c}}{\widetilde{s}} = \rho \text{ (definition)} \quad\Longrightarrow\quad \widehat{h}=\frac{\rho}{2+\rho}\ \ (\le 1) \tag{11}$$

§ **Robustness of LS estimate $\widetilde{s}$:**

$$\widehat{h}(\widetilde{s},\rho)=\frac{\rho}{2+\rho},\quad \rho=-r_{\rm c}/\widetilde{s} \tag{12}$$

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

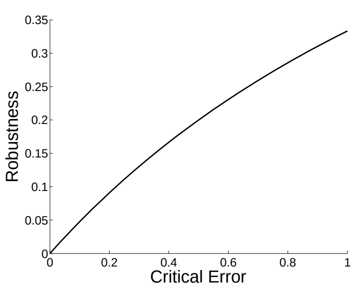


Figure 6: Robustness of estimated slope, $\widehat{h}(\widetilde{s},\rho)$, vs. critical error, ρ. Eq.(12).

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

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

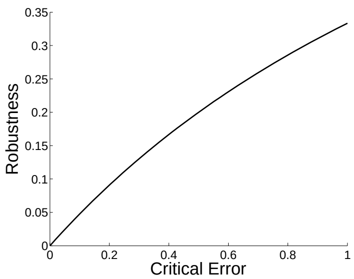


Figure 7: $\widehat{h}(\widetilde{s}, \rho)$ vs. ρ.

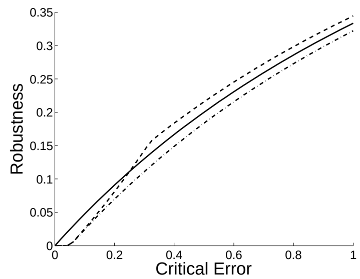


Figure 8: $\widehat{h}(s_{\rm e}, \rho)$ vs. ρ. $\zeta = 1$ (solid), 1.05 (dash), 0.95 (dot dash).

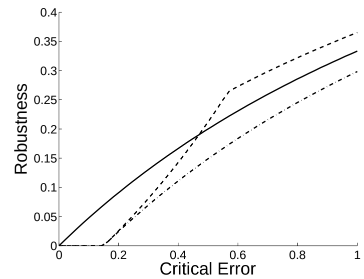


Figure 9: $\widehat{h}(s_{\rm e}, \rho)$ vs. ρ. $\zeta = 1$ (solid), 1.15 (dash), 0.85 (dot-dash).

§ **Estimates of Phillips slope:**

- $\widetilde{s}$ = LS estimate, with robustness $\widehat{h}(\widetilde{s}, r_{\rm c})$.
- $s_{\rm e}$= any estimate, with robustness $\widehat{h}(s_{\rm e}, r_{\rm c})$.
- Definitions: $\zeta = s_{\rm e}/\widetilde{s}$, $\quad \rho = -r_{\rm c}/\widetilde{s}$. (Recall: $\widetilde{s} < 0$.)
- Robustness of $s_{\rm e}$, in analogy to eq.(7):

$$\widehat{h}(s_{\rm e}, r_{\rm c}) = \max\left\{ h :\ \left(\max_{\gamma, \sigma^2 \in \mathcal{U}(h)} |s(\gamma, \sigma^2) - s_{\rm e}| \right) \le r_{\rm c} \right\} \tag{13}$$

◦ Let $m(h)$ denote the inner maximum:

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

◦ For $h \le 1$ this occurs at one of the following:

$$\text{Either:} \qquad \gamma = (1+h)\widetilde{\gamma},\ \ \sigma^2 = (1-h)\widetilde{\sigma}^2 \tag{15}$$
$$\text{Or:} \qquad \gamma = (1-h)\widetilde{\gamma},\ \ \sigma^2 = (1+h)\widetilde{\sigma}^2 \tag{16}$$

◦ Denote the corresponding $m(h)$'s:

$$m_1(h) \quad = \quad \left| \frac{(1+h)\widetilde{\gamma}}{(1-h)\widetilde{\sigma}^2} - s_{\rm e} \right| \tag{17}$$
$$m_2(h) \quad = \quad \left| \frac{(1-h)\widetilde{\gamma}}{(1+h)\widetilde{\sigma}^2} - s_{\rm e} \right| \tag{18}$$

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

$$m(h) = \max[m_1(h),\ m_2(h)] \tag{19}$$

The maximum depends on the value of h.

◦ After some algebra, and equating $m(h) = r_{\rm c}$, one finds:

$$\widehat{h}(s_{\rm e}, \rho) = \begin{cases} \dfrac{\rho + \zeta - 1}{\rho + \zeta + 1} & \text{if } \rho^2 \ge \zeta^2 - 1 \text{ and } \rho \ge 1 - \zeta \\ \\ \dfrac{\rho - \zeta + 1}{-\rho + \zeta + 1} & \text{if } \rho^2 \le \zeta^2 - 1 \text{ and } \rho \ge \zeta - 1 \end{cases} \tag{20}$$

$\widehat{h}(s_{\rm e}, \rho)$ is zero otherwise. Note $\widehat{h} \le 1$.

• 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 $\widetilde{s}$) 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.

• $X = (x_1, \ldots, x_N) =$ random sample.

• $\widetilde{p}(x|\lambda) =$ be a pdf for x with parameters λ.

¶ Likelihood function:

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

¶ Maximum likelihood estimate (MLE):

$$\lambda^{\star} = \arg\max_{\lambda} L(X, \widetilde{p}) \tag{22}$$

¶ Examples of MLE.

• **Exponential distribution:** The pdf is:

$$\widetilde{p}(x|\lambda) = \lambda \mathrm{e}^{-\lambda x}, \;\; x \ge 0 \tag{23}$$

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

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

Thus:

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

Equating to zero and solving for λ yields the MLE:

$$0 = \frac{\partial L}{\partial \lambda} \quad \Longrightarrow \quad 0 = N\lambda^{N-1} - \lambda^N \sum_{i=1}^{N} x_i \quad \Longrightarrow \quad \boxed{\frac{1}{\lambda_{\mathrm{MLE}}} = \frac{1}{N} \sum_{i=1}^{N} x_i} \tag{26}$$

Note that:

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

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

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

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

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

Note that:

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

Thus MLE and LSE agree. Define the squared error:

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

Thus:

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

¶ **Robust-satisficing:**

- Form of the pdf is not certain.
- $\widetilde{p}(x|\lambda)$ 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:**

$$\mathcal{U}(h, \widetilde{p}) = \left\{ p(x) : \; p(x) \in \mathcal{P}, \; |p(x) - \widetilde{p}(x|\lambda)| \le h\psi(x) \right\}, \quad h \ge 0 \tag{33}$$

- $\mathcal{P}$ 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 $\mathcal{U}(h, \widetilde{p})$, how should we choose the parameters of the nominal pdf $\widetilde{p}(x|\lambda)$?

¶ **Robustness:**

$$\widehat{h}(\lambda, L_{\rm c}) = \max \left\{ h : \; \left(\min_{p \in \mathcal{U}(h, \widetilde{p})} L(X, p) \right) \ge L_{\rm c} \right\} \tag{34}$$

¶ $m(h) =$ **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:

$$p(x_i) = \begin{cases} \widetilde{p}(x_i) - h\psi(x_i) & \text{if } h \le \widetilde{p}(x_i)/\psi(x_i) \\ 0 & \text{else} \end{cases} \tag{35}$$

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

$$h_{\max} = \min_i \frac{\widetilde{p}(x_i)}{\psi(x_i)} \tag{36}$$

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

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

¶ $m(h)$ **and** $\widehat{h}(\lambda, L_{\rm c})$**:**

- Robustness is the max h at which $m(h) \ge L_{\rm c}$.
- $m(h)$ strictly decreases as h increases.
- Hence robustness is the solution of $m(h) = L_{\rm c}$.
- Hence $m(h)$ is the inverse of $\widehat{h}(\lambda, L_{\rm c})$:

$$m(h) = L_{\rm c} \quad \text{implies} \quad \widehat{h}(\lambda, L_{\rm c}) = h \tag{38}$$

- Plot of $m(h)$ vs. h is plot of $L_{\rm c}$ vs. $\widehat{h}(\lambda, L_{\rm c})$.

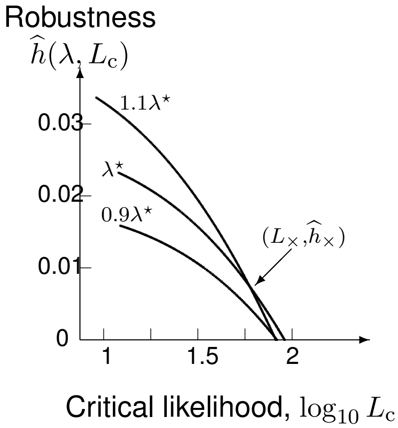


Figure 10: Robustness curves. $\lambda^\star = 3.4065$.

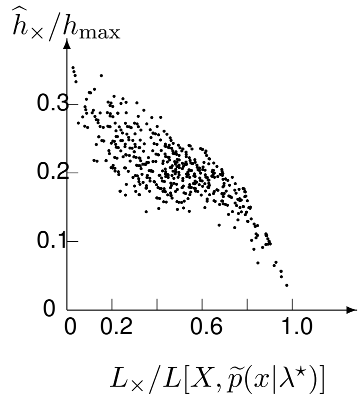


Figure 11: Loci of intersection of robustness curves $\widehat{h}(\lambda^\star, L_c)$ and $\widehat{h}(1.1\lambda^\star, L_c)$.

¶ **Robustness curves** in fig. 10 based on:

- Eqs.(37) and (38).
- Nominal pdf is exponential, $\widetilde{p}(x|\lambda) = \lambda \exp(-\lambda x)$ 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): $\lambda^\star = 1/\overline{x}$ where $\overline{x} = (1/N)\sum_{i=1}^{N} x_i$ is the sample mean.
- Robustness curves for 3 λ's: $0.9\lambda^\star$, $\lambda^\star$, and $1.1\lambda^\star$.

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

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

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

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

¶ **Preferences between estimates of λ:**

- $\widehat{h}(\lambda^\star, L_c) > \widehat{h}(0.9\lambda^\star, L_c) \implies \lambda^\star \succ 0.9\lambda^\star$.
- $\widehat{h}(\lambda^\star, L_c)$ and $\widehat{h}(1.1\lambda^\star, L_c)$ cross at $(L_\times, \widehat{h}_\times)$:
  - ∘ $\lambda^\star \succ 1.1\lambda^\star$ for $L_c > L_\times$ and $h < h_\times$.
  - ∘ $1.1\lambda^\star \succ \lambda^\star$ else.

¶ 500 repetitions:

- $\lambda^\star$ dominates $0.9\lambda^\star$.
- Preferences reverse between $\lambda^\star$ and $1.1\lambda^\star$.
- Normalized $(h_\times,\ L_\times)$ in fig. 11.
- Center of cloud: (0.5, 0.2). Typical cross of robustness curves at:
  - L_c about half of best-estimated value.
  - $\widehat{h}$ 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.