

Lecture Notes on

Info-Gap Estimation and Forecasting

Yakov Ben-Haim
 Yitzhak Moda'i Chair in
 Technology and Economics
 Faculty of Mechanical Engineering
 Technion — Israel Institute of Technology
 Haifa 32000 Israel
<http://www.technion.ac.il/yakov>
yakov@technion.ac.il

¶ Source material:

- Yakov Ben-Haim, 2005, Info-gap Decision Theory For Engineering Design. Or: Why ‘Good’ is Preferable to ‘Best’, appearing as chapter 11 in *Engineering Design Reliability Handbook*, Edited by Efstratios Nikolaidis, Dan M.Ghiocel and Surendra Singhal, CRC Press, Boca Raton.
- Yakov Ben-Haim, 2006, *Info-Gap Decision Theory: Decisions Under Severe Uncertainty*, 2nd edition, section 3.2.13, Academic Press, London.
- Yakov Ben-Haim, 2008, Info-gap forecasting and the advantage of sub-optimal models, *European Journal of Operational Research*, to appear.
- Yakov Ben-Haim, 2008, Info-Gap Economics: An Overview, working paper. (`\papers\BoE2008\ige03.tex`)

A Note to the Student: These lecture notes are not a substitute for the thorough study of books. These notes are no more than an aid in following the lectures.

Contents

| | | |
|----------|---|----------|
| 1 | Linear Regression | 3 |
| 2 | System Identification | 8 |
| 2.1 | Model Uncertainty: Preliminary Example | 8 |
| 2.2 | Optimal System Identification | 9 |
| 2.3 | Uncertainty | 10 |
| 2.4 | Robustness | 11 |
| 2.5 | Performance and Robustness | 11 |
| 2.6 | Example | 12 |
| 2.7 | An Interpretation: Focus of Uncertainty | 16 |

| | | |
|----------|---|-----------|
| 2.8 | Robustness and Opportuneness | 17 |
| 2.9 | Forecasting and looseness of model prediction | 20 |
| 3 | Tychonov Up-Dating of a Linear System with Model Uncertainty | 22 |
| 3.1 | Formulation of the Up-Dating Problem | 22 |
| 3.2 | Robustness to Uncertainty | 23 |
| 3.3 | Robustness of the Tychonov Regularized Model | 24 |
| 3.4 | Proofs | 26 |
| 4 | Estimating an Uncertain Probability Density | 28 |
| 5 | Forecasting | 32 |
| 5.1 | Preliminary Example: European Central Bank Interest Rates | 32 |
| 5.2 | Info-Gap Forecasting: Formulation | 33 |
| 5.2.1 | The Estimated System and its Uncertainty | 33 |
| 5.2.2 | Forecasting with Slope Adjustment | 33 |
| 5.2.3 | Definition of the Robustness Function | 34 |
| 5.2.4 | Evaluating the Robustness Function | 34 |
| 5.2.5 | Crossing of Robustness Curves and the Advantage of Sub-Optimal Models | 35 |
| 5.3 | Example: 1-Dimensional System | 36 |
| 5.4 | Robustness and Probability of Forecast Success | 38 |

1 Linear Regression

§ Modelling is a decision problem. We will consider modelling the Phillips curve.¹

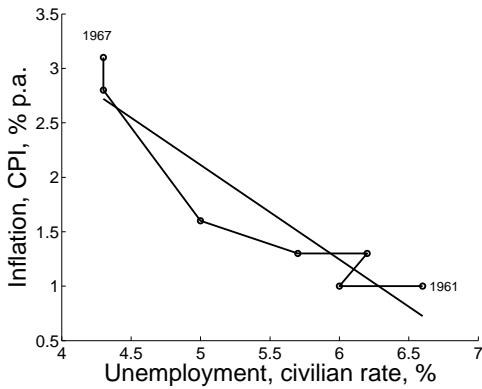


Figure 1: Inflation vs. unemployment in the US, 1961–1967.



Figure 2: 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:

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

§ 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?

¹Source: Yakov Ben-Haim, 2008, Info-Gap Economics: An Overview, working paper. (\papers\BoE2008\ige03.tex)

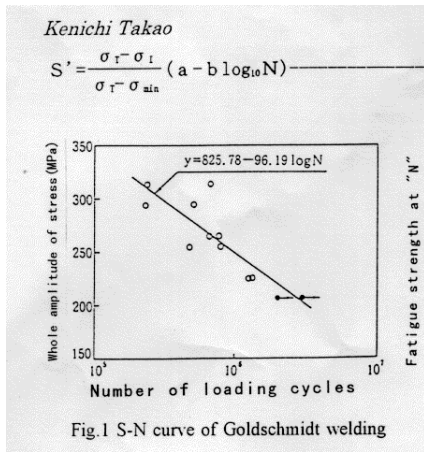


Figure 3: S-N curves.

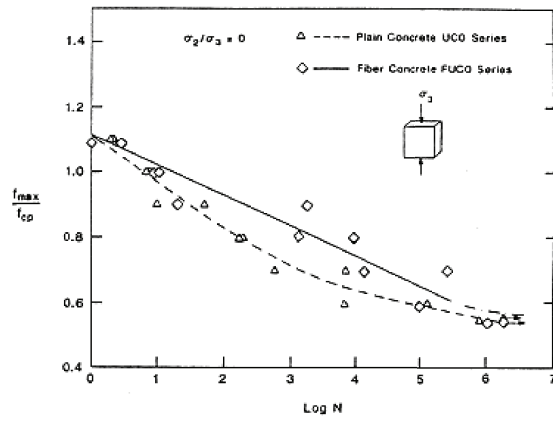


Figure 4: S-N curves.

§ Challenge: Two foci of uncertainty:

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

§ 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?

§ Data:

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

§ Least-squares estimate of slope:

- Linear regression:

$$c = su + b \quad (1)$$

Mean squared error:

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

MSE estimate of the slope:

$$\tilde{s} = \arg \min_s \text{MSE} \quad (3)$$

One finds:

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

where:

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

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

§ Robustness question:

How much can the data err, and the slope's error will be acceptable?

§ Moments:

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

§ Unknown fractional errors:

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

§ Fractional-error info-gap model:

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

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

§ Robustness of LS estimate \tilde{s} :

Max horizon of uncertainty in moments

at which \tilde{s} errs no more than r_c :

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

§ Robustness of LS estimate \tilde{s} :

$$\hat{h}(\tilde{s}, \rho) = \frac{\rho}{2 + \rho}, \quad \rho = -r_c/\tilde{s} \quad (8)$$

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

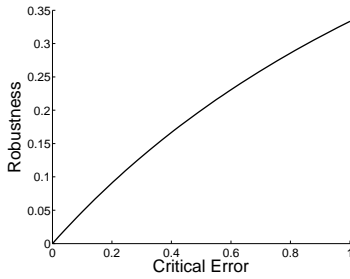


Figure 5: Robustness of estimated slope, $\hat{h}(\tilde{s}, \rho)$, vs. critical error, ρ . Eq.(8).

- 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?

§ Estimates of Phillips slope:

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

$$\hat{h}(s_e, r_c) = \max \left\{ h : \left(\max_{\gamma, \sigma^2 \in \mathcal{U}(h)} |s(\gamma, \sigma^2) - s_e| \right) \leq r_c \right\} \quad (9)$$

One finds:

$$\hat{h}(s_e, \rho) = \begin{cases} \frac{\rho + \zeta - 1}{\rho + \zeta + 1} & \text{if } \rho^2 \geq \zeta^2 - 1 \text{ and } \rho \geq 1 - \zeta \\ \frac{\rho - \zeta + 1}{-\rho + \zeta + 1} & \text{if } \rho^2 \leq \zeta^2 - 1 \text{ and } \rho \geq \zeta - 1 \end{cases} \quad (10)$$

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

- Eq.(10) includes eq.(8) as a special case, when $\zeta = 1$.
- When $\zeta > 1$, the robustness follows the lower line of eq.(10) (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. 7 and 8. (The two lines in eq.(10) are equal when $\rho^2 = \zeta^2 - 1$.)

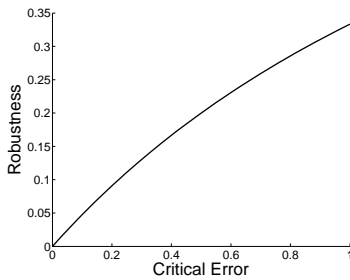


Figure 6: $\hat{h}(\tilde{s}, \rho)$ vs. ρ .

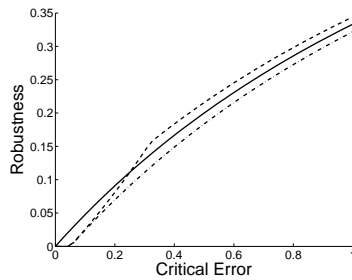


Figure 7: $\hat{h}(s_e, \rho)$ vs. ρ . $\zeta = 1$ (solid), 1.05 (dash), 0.95 (dot dash).

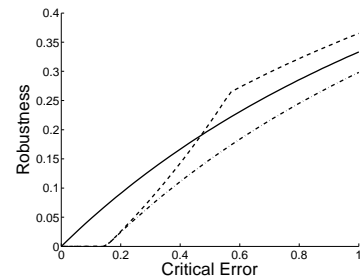


Figure 8: $\hat{h}(s_e, \rho)$ vs. ρ . $\zeta = 1$ (solid), 1.15 (dash), 0.85 (dot-dash).

- 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 System Identification

¶ Optimal system identification: Adjusting a model to conform to data.

¶ Main thesis:

Optimal identification has no robustness to residual errors in the model.

¶ Corollaries:

- Sub-optimal models can be robust.
- Sub-optimal models can
 - be more robust than, and
 - reproduce data as well as, the optimal model.

2.1 Model Uncertainty: Preliminary Example

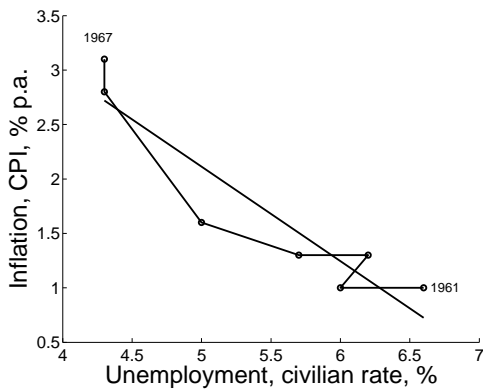


Figure 9: Inflation vs. unemployment in the US, 1961–1967.



Figure 10: Inflation vs. unemployment in the US, 1961–1993.

§ From fig. 9, US unemployment vs. inflation for 1961–1967 looks linear:

$$\pi = aU + b \tag{11}$$

§ From fig. 10 shows more complicated dynamics.

§ Slopes in other periods:

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

§ Info-gaps:

- Uncertain data and process.
- Unknown functional relation.

§ We studied this example in section 1. Now consider more general case.

2.2 Optimal System Identification

¶ Notation:

Y_I = I th data set, $I = 1, \dots, N$,

$F_I(Q)$ = Model prediction of Y_I .

Q = Parameters and properties of model.

\mathcal{Y} = $\{Y_1, \dots, Y_N\}$.

$\mathcal{F}(Q)$ = $\{F_1(Q), \dots, F_N(Q)\}$.

$R[\mathcal{Y}, \mathcal{F}(Q)]$ = Performance of predictor, e.g. mean-square error:

$$R[\mathcal{Y}, F(Q)] = \frac{1}{N} \sum_{I=1}^N \|F_I(Q) - Y_I\|^2 \quad (12)$$

¶ Optimal model, Q^\bullet , minimizes performance-measure:

$$Q^\bullet = \arg \min_Q R[\mathcal{Y}, F(Q)] \quad (13)$$

¶ We will show: fidelity of model to data as good as $R[\mathcal{Y}, F(Q^\bullet)]$ is

- obtainable but not feasible.
- not robust to info-gaps in model.

2.3 Uncertainty

¶ Model structure $F_I(Q)$ is wrong. Relevant factors are missing:

- Non-linearities.
- Time dependence.
- Dimensionality.
- Etc.

¶ Complete model:

$$\phi_I = F_I(Q) + U_I \quad (14)$$

$F_I(Q)$ = Best known model structure.

ϕ_I = Correct model structure.

U_I = Unknown info-gap.

¶ Info-gap model of uncertainty: Unbounded family of nested sets (of models):

$$F_I(Q) \in \mathcal{U}(h, F_I(Q)), \quad h \geq 0 \quad (15)$$

$$h < h^\bullet \implies \mathcal{U}(h, F_I(Q)) \subset \mathcal{U}(h^\bullet, F_I(Q)) \quad (16)$$

2.4 Robustness

¶ Fidelity of model to data:

$R[\mathcal{Y}, \mathcal{F}(Q)]$ = Fidelity of model $F_I(Q)$ to data.

$R[\mathcal{Y}, \mathcal{F}_U(Q)]$ = Fidelity of model $F_I(Q) + U_I$ to data.

r_c = Acceptable fidelity of model to data.

¶ Robustness of model $F_I(Q)$:

- How wrong can $F_I(Q)$ be without exceeding acceptable fidelity?
- Epistemic, not ontological question.
- Max horizon of uncertainty, h , which does not jeopardize fidelity:

$$\hat{h}(Q, r_c) = \max \left\{ h : \max_{\substack{\phi_I \in \mathcal{U}(h, F_I(Q)) \\ I=1, \dots, N}} R[\mathcal{Y}, \mathcal{F}_U(Q)] \leq r_c \right\} \quad (17)$$

2.5 Performance and Robustness

¶ $R[\mathcal{Y}, \mathcal{F}(Q)]$ = Fidelity of model, $F_I(Q)$, to data.

¶ $\hat{h}(Q, r_c)$ = Robustness of model, $F_I(Q)$, with fidelity-aspiration r_c .

¶ Theorem:

$$r_c = R[\mathcal{Y}, \mathcal{F}(Q)] \quad \text{implies} \quad \hat{h}(Q, r_c) = 0 \quad (18)$$

Meaning:

No model can be relied upon to perform “as advertised”.

¶ This holds also for optimal model, Q^\bullet :

$$R[\mathcal{Y}, F(Q^\bullet)] = \min_Q R[\mathcal{Y}, F(Q)] \quad (19)$$

$$R_C^\bullet = R[\mathcal{Y}, \mathcal{F}(Q^\bullet)] \quad \text{implies} \quad \hat{h}(Q^\bullet, R_C^\bullet) = 0 \quad (20)$$

¶ Implication:

Sub-optimal models can be more robust than optimal model at same fidelity.

2.6 Example

¶ 1-dimensional system:

$$\begin{aligned} Y_I &= \text{Scalar measurements.} \\ F_I(Q) &= QI. \quad \text{Nominal linear model} \\ R[\mathcal{Y}, \mathcal{F}(Q)] &= \text{Mean-squared error:} \end{aligned}$$

$$R[\mathcal{Y}, F(Q)] = \frac{1}{N} \sum_{I=1}^N (QI - Y_I)^2 \quad (21)$$

¶ Q^\bullet = Least-squares optimal model:

$$Q^\bullet = \arg \min_Q R[\mathcal{Y}, F(Q)] = \frac{\eta_1}{\eta_0} \quad (22)$$

$$\eta_1 = \frac{1}{N} \sum_{i=1}^N I Y_I, \quad \eta_0 = \frac{1}{N} \sum_{i=1}^N I^2 \quad (23)$$

¶ Model error: Uncertain quadratic term.

$$\phi_I = QI + UI^2 \quad (24)$$

¶ Info-gap model for quadratic uncertainty:

$$\mathcal{U}(h, QI) = \{ \phi_I = QI + UI^2 : |U| \leq h \}, \quad h \geq 0 \quad (25)$$

¶ **Robustness:**

Max horizon of uncertainty, h , with acceptable fidelity to data.

$$\hat{h}(Q, r_c) = \max \left\{ h : \max_{|U| \leq h} R[\mathcal{Y}, \mathcal{F}_U(Q)] \leq r_c \right\} \quad (26)$$

$$\hat{h}(Q, r_c) = \begin{cases} 0, & r_c \leq \xi_2 \\ \frac{|\xi_1|}{\xi_0} \left(-1 + \sqrt{1 + \frac{r_c - \xi_2}{\xi_1^2}} \right), & \xi_2 < r_c \end{cases} \quad (27)$$

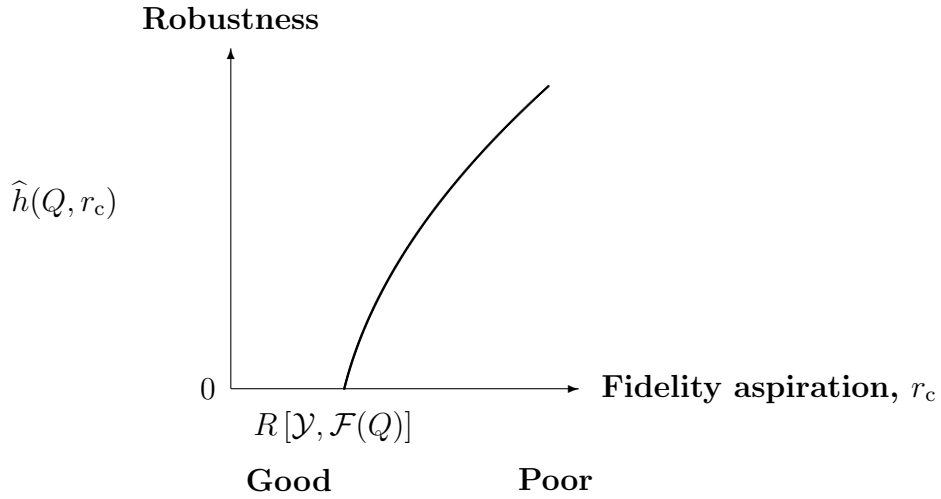
$$\xi_2 = \frac{1}{N} \sum_{i=1}^N (QI - YI)^2 \quad (28)$$

$$= R[\mathcal{Y}, \mathcal{F}(Q)] \quad (29)$$

$$\xi_1 = \frac{1}{N} \sum_{i=1}^N I^2 (QI - YI) \quad (30)$$

$$\xi_0 = \frac{1}{N} \sum_{i=1}^N I^4 \quad (31)$$

¶ Trade-off: robustness vs. fidelity.



¶ No robustness for aspiration at nominal performance:

$$\hat{h}(Q, r_c) = 0 \quad \text{if} \quad r_c = R[\mathcal{Y}, \mathcal{F}(Q)] \quad (32)$$

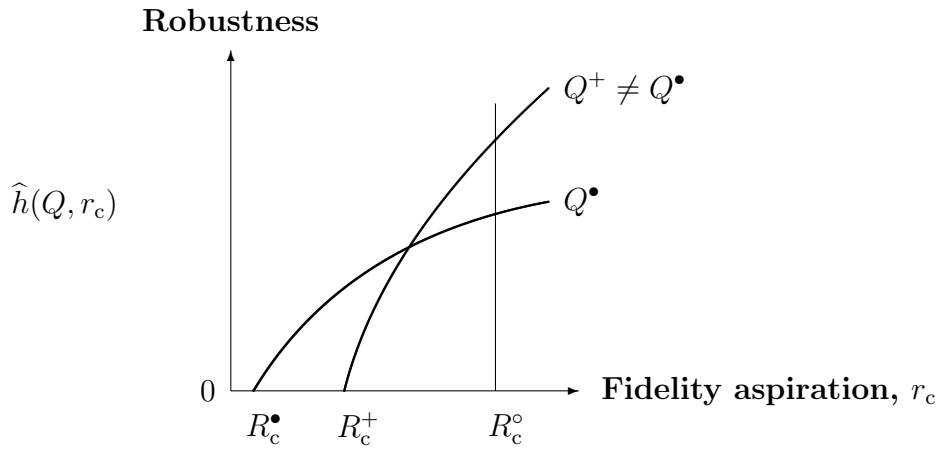
¶ Preference for sub-optimal model:

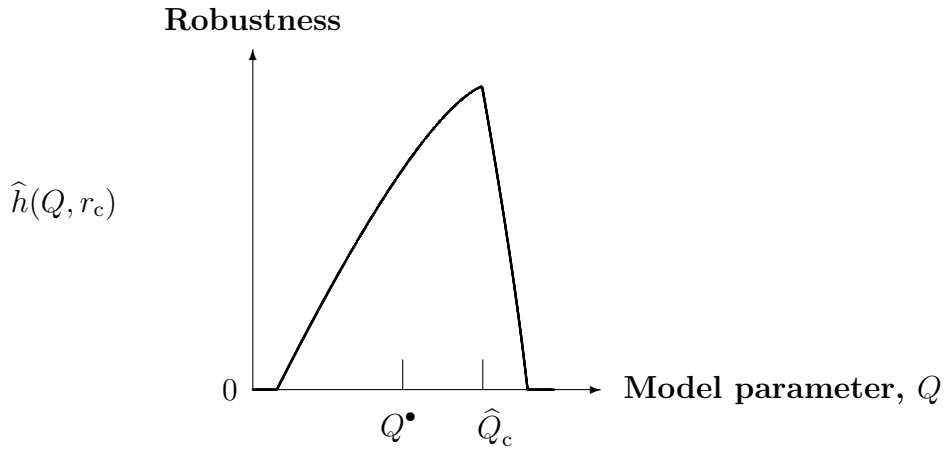
Q^\bullet = L.S.-optimal model.

Q^+ = L.S.-sub-optimal model.

R_c° = Acceptable fidelity.

Q^+ preferred to Q^\bullet at R_c° .





¶ Robust-satisficing model:

Q^* = L.S.-optimal model. R^* = L.S. optimal error.

\hat{Q}_C = Robust-satisficing model. Maximizes $\hat{h}(Q, r_c)$.

R_C only slightly $> R^*$. $\hat{h}(\hat{Q}_C, r_c) \gg \hat{h}(Q^*, r_c)$.

\hat{Q}_C preferred to Q^* .

¶ Conclusions:

- Any model, $F_I(Q)$,
 - has no immunity to unknown quadratic term:

$$\hat{h}(Q, r_c) = 0 \quad \text{if} \quad r_c = R[\mathcal{Y}, \mathcal{F}(Q)].$$

- is reliable only at less-than-nominal fidelity.
- Also holds for least-square optimal model, Q^* .

- Robustness curves can cross:
 - Sub-optimal model Q^+
 - more robust than optimal model Q^*
 - at same fidelity to data.

- Info-gap strategy:
 - Satisfice fidelity to data.
 - Optimize robustness to model-deficiency.

2.7 An Interpretation: Focus of Uncertainty

¶ **Least-squares estimation** focusses on managing error in data, Y_I :

$$\text{Minimize: } \sum_{I=1}^N (F_I(Q) - Y_I)^2 \quad (33)$$

¶ **Info-gap estimation** focusses on managing

- error in data, Y_I :

$$\text{Satisfice: } \sum_{I=1}^N (F_I(Q) - Y_I)^2$$

- error in model, $F_I(Q)$:

$$\text{Maximize: } \hat{h}(Q, r_c).$$

2.8 Robustness and Opportuneness

¶ Robustness of model $F_I(Q)$:

how wrong can $F_I(Q)$ be without exceeding acceptable fidelity?

$$\hat{h}(Q, r_c) = \max \left\{ h : \max_{\substack{\phi_I \in \mathcal{U}(h, F_I(Q)) \\ I=1, \dots, N}} R[\mathcal{Y}, \mathcal{F}_U(Q)] \leq r_c \right\} \quad (34)$$

¶ Opportuneness of model $F_I(Q)$:

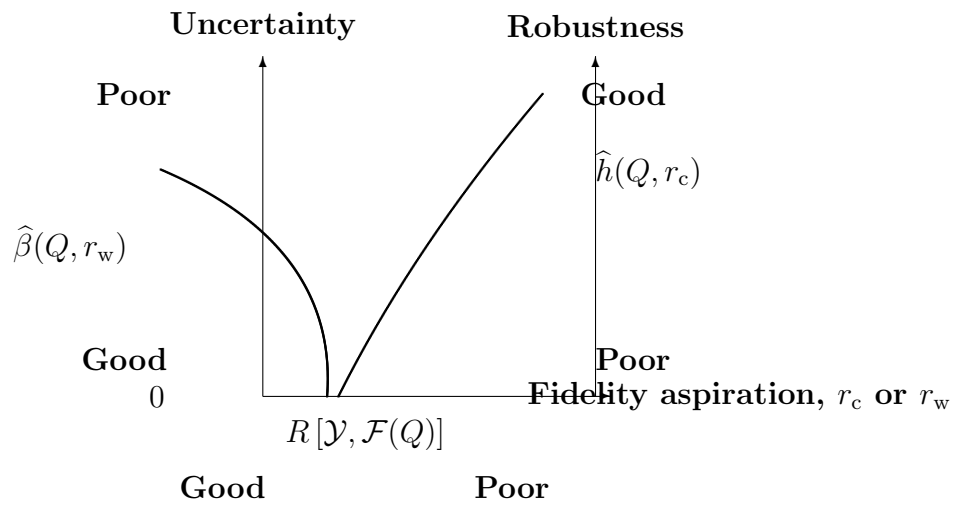
how wrong must $F_I(Q)$ be to enable windfall fidelity?

$$r_w \ll r_c$$

$$\hat{\beta}(Q, r_w) = \min \left\{ h : \min_{\substack{\phi_I \in \mathcal{U}(h, F_I(Q)) \\ I=1, \dots, N}} R[\mathcal{Y}, \mathcal{F}_U(Q)] \leq r_w \right\} \quad (35)$$

¶ Preferences:

- Robustness:
 - Immunity to failure.
 - Satisficing at critical fidelity.
 - Bigger is better
- Opportuneness:
 - Immunity to windfall.
 - Windfalling at wildest-dream fidelity.
 - Big is bad.



¶ Trade-offs:

- Robustness vs. critical fidelity.
- Opportuneness vs. windfall fidelity.

¶ **Sympathetic immunities:**

change in model, Q , which improves \hat{h}
 also improves $\hat{\beta}$.

$$\frac{\partial \hat{h}}{\partial Q} \frac{\partial \hat{\beta}}{\partial Q} < 0 \tag{36}$$

¶ **Antagonistic immunities:**

change in model, Q , which improves \hat{h}
 also degrades $\hat{\beta}$.

$$\frac{\partial \hat{h}}{\partial Q} \frac{\partial \hat{\beta}}{\partial Q} > 0 \tag{37}$$

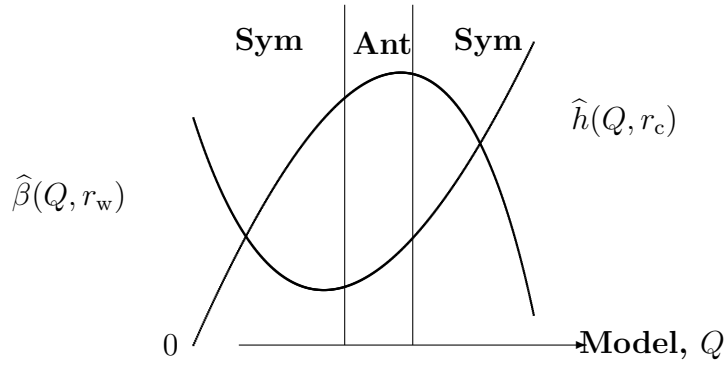


Figure 11: Schematic immunity curves

$$\left(\frac{\hat{h}}{|U_M|} + 1 \right)^2 - \frac{\xi_0 r_c}{\xi_1} = \left(\frac{\hat{\beta}}{|U_M|} - 1 \right)^2 - \frac{\xi_0 r_w}{\xi_1} \tag{38}$$

2.9 Forecasting and looseness of model prediction

¶ The issue of prediction looseness:

- At high robustness,
Many models have same fidelity.
- Do their predictions agree?

¶ Unknown complete model:

$$\phi_I = F_I(Q) + U_I \quad (39)$$

¶ The info-gap uncertainty model is:

$$\mathcal{U}[h, F_I(Q)], \quad h \geq 0.$$

¶ For design Q define:

$$\begin{aligned} h^* &= \hat{h}(Q, r_c) \\ &= \text{Robustness of } Q \text{ at } r_c. \end{aligned}$$

$$\begin{aligned} \Lambda(Q) &= \mathcal{U}[h^*, F_I(Q)] \\ &= \text{set of all models, } \phi_I, \text{ which satisfy the prediction error at } r_c. \\ &= \text{Predictions of fidelity-equivalent models.} \\ &= \text{Prediction-looseness of model } Q. \end{aligned}$$

¶ Fidelity–robustness trade-off:

$$r_c < R_C^\bullet \implies \hat{h}(Q, r_c) \leq \hat{h}(Q, R_C^\bullet) \quad (40)$$

Robustness decreases as fidelity improves.

¶ Robustness–prediction-looseness trade-off:

$$\hat{h}(Q, r_c) < \hat{h}(Q^\bullet, r_c) \implies \Lambda(Q) \subseteq \Lambda(Q^\bullet) + \mu \quad (41)$$

Robustness decreases as looseness improves.

¶ **The dilemma:**

- Fidelity to data necessary for trueness of model.
- Robustness to model uncertainty verifies fidelity.
- Looseness of model prediction results from
fidelity-robustness to model-uncertainty.

¶ Dilemma due to **conflict of two uncertainties:**

- Measurement error (spread of data).
Causes need for fidelity.
- Model error (epistemic limitation).
Causes need for robustness.

¶ **Hume and the problem of induction:**

- The past does not bind the future.
- Experience cannot validate scientific induction.

¶ **Robustness-fidelity-looseness trade-offs:**

Measurement error and limited understanding
impose prediction looseness.

¶ **Epistemological warrant:**

- Basis for theory (model) selection.
- Obtained by:
 - High fidelity to data.
 - High robustness to model error.

¶ Question: **Is warrant warranted?**

Warrant = Hi fidelity and high robustness
= High prediction looseness.

Answer: Doesn't look like it.

3 Tychonov Up-Dating of a Linear System with Model Uncertainty

This section based on:

Yakov Ben-Haim and Scott Cogan, Up-Dating a Linear System with Model Uncertainty: An Info-Gap Approach, Intl. Conf. on Uncertainty in Structural Dynamics, University of Sheffield, UK. 15–17.6.2009.

3.1 Formulation of the Up-Dating Problem

§ **Measurements:**

$f \in \mathfrak{R}^J$ is the force vector.

$y^{(m)} \in \mathfrak{R}^N$ is the response vector, for $m = 1, \dots, M$.

§ **Model we will up-date:** choose V in:

$$y = Vf \quad (42)$$

§ **Uncertainty:** inhomogeneous input/output relation is:

$$y = Vf + u \quad (43)$$

§ **Tychonov-regularized least squared error** is:

$$S = \lambda \|\tilde{y} - y\|^2 + \frac{1}{M} \sum_{m=1}^M \|y^{(m)} - y\|^2 \quad (44)$$

where:

- \tilde{y} is a prior estimate of the response.
- We are using the Euclidean norm: $\|x\|^2 = x^T x$.

§ **Actual mean-squared error.** Substituting eq.(43) into eq.(44):

$$\begin{aligned} S(V, u) &= \underbrace{(1 + \lambda)f^T V^T V f - 2(\lambda\tilde{y} + \bar{y})^T V f + \lambda\|\tilde{y}\|^2 + \overline{\|y\|^2}}_{S_o} \\ &\quad + \underbrace{(1 + \lambda)u^T u - 2(\lambda\tilde{y} + \bar{y} - (1 + \lambda)Vf)^T u}_{2z^T u} \end{aligned} \quad (45)$$

$$= S_o + (1 + \lambda)u^T u - 2z^T u \quad (46)$$

where:

$$\bar{y} = \frac{1}{M} \sum_{m=1}^M y^{(m)} \quad (47)$$

$$\overline{\|y\|^2} = \frac{1}{M} \sum_{m=1}^M \|y^{(m)}\|^2 \quad (48)$$

- S_o is the ordinary Tychonov-regularized least-squares error function for the linear model, $y = Vf$.

- S_u contains the uncertain inhomogeneous terms in the model in eq.(43). S_u also contains the measurements, f and $y^{(1)}, \dots, y^{(M)}$, in the vector z and in S_o .

§ **Goal.** We wish to choose V but we cannot actually minimize $S(V, u)$ since u is unknown. The approach we take is to choose V to make $S(V, u)$ adequately small for a maximal range of possible realizations of u .

3.2 Robustness to Uncertainty

§ **System model:** $S(V, u)$ in eq.(46).

§ **Uncertainty model:** spherical info-gap model for uncertain vector u in eq.(43):

$$\mathcal{U}(h) = \{u : u^T u \leq h^2\}, \quad h \geq 0 \quad (49)$$

§ **Performance requirement:** regularized squared error must not exceed S_c :

$$S(V, u) \leq S_c \quad (50)$$

§ **Robustness function.**

$$\hat{h}(V, S_c) = \max \left\{ h : \left(\max_{u \in \mathcal{U}(h)} S(V, u) \right) \leq S_c \right\} \quad (51)$$

§ Note that robustifying w.r.t. data which does not include the non-homogeneous term is a bit like the Tychonov concept of biasing the estimate towards a prior value.

§ One can readily show that:

$$\hat{h}(V, S_c) = \frac{1}{1 + \lambda} \left(-\sqrt{z^T z} + \sqrt{z^T z + (1 + \lambda)(S_c - S_o)} \right) \quad (52)$$

or zero if $S_c \leq S_o$. The dependence of the robustness on the model matrix, V , and on the observations f and $y^{(m)}$, arises through S_o and z , defined in eq.(45).

3.3 Robustness of the Tychonov Regularized Model

§ **Preview.** In this section we:

- Derive an explicit expression for the robustness of an up-dated model which minimizes the Tychonov-regularized mean squared error, S_o in eq.(45).
- Theorem 1 asserts that Tychonov-optimal matrices are more robust to uncertainty than all other matrices, at fixed Tychonov weight.
- Theorem 2 asserts that the robustness of Tychonov optimal matrices increases as the Tychonov weight decreases.
- Proofs appear in appendix 3.4.

§ **Tychonov-regularized mean squared error**, S_o in eq.(45), can be written:

$$S_o(V) = (1 + \lambda) \left[(Vf - a)^T (Vf - a) \right] + b \quad (53)$$

where:

$$a = \frac{1}{1 + \lambda} (\lambda \tilde{y} + \bar{y}) \quad (54)$$

$$b = \lambda \|\tilde{y}\|^2 + \|\bar{y}\|^2 - (1 + \lambda) a^T a \quad (55)$$

§ **Minimization of $S_o(V)$:**

- If $f \neq 0$ then the matrix V can always be chosen to precisely satisfy $Vf = a$, which minimizes $S_o(V)$ in eq.(53).
- Let V_T denote any such choice of V , which we will refer to as a *Tychonov optimal matrix*.
- It then results that z , defined in eq.(45), is identically zero.
- Furthermore one finds that $S_o(V_T) = b$.
- One now finds the robustness in eq.(52), for any Tychonov optimal matrix V_T , to be:

$$\hat{h}(V_T, S_c) = \sqrt{\frac{S_c - b}{1 + \lambda}} \quad (56)$$

or zero if $S_c \leq b$.

Theorem 1 *A Tychonov optimal matrix, V_T , is strictly more robust than any other matrix V , at fixed Tychonov weight λ :*

$$\hat{h}(V_T, S_c) > \hat{h}(V, S_c) \quad (57)$$

for all values of $S_c > b$.

§ Note relation to result by Zacksenhouse *et al*:

Zacksenhouse *et al.*² [proposition 2] derive a similar result though they consider info-gap un-

²Zacksenhouse, M., S.Nemets, M.A.Lebedev and M.A.L.Nicolelis, 2008, Robust-satisficing linear regression: Performance/robustness trade-off and consistency criterion, *Mechanical Systems and Signal Processing*, to appear.

certainty in the data, rather than uncertainty in the model structure as we have done.

Theorem 2 *Robustness of a Tychonov optimal matrix decreases as the Tychonov weight increases.*

Given two Tychonov weights, $\lambda < \lambda'$, with corresponding Tychonov optimal matrices V_T and V'_T , respectively. Then:

$$\hat{h}(V_T, S_c) > \hat{h}(V'_T, S_c) \quad (58)$$

for all values of $S_c > b(\lambda)$.

The proof of this theorem depend on the following lemma. First define the variance of the measured responses as:

$$\overline{\|y - \bar{y}\|^2} = \overline{\|y\|^2} - \|\bar{y}\|^2 \quad (59)$$

where the two terms on the right are defined in eqs.(47) and (48).

Lemma 1 *The coefficient b in eq.(48) can be expressed:*

$$b = \frac{\lambda}{1 + \lambda} \|\tilde{y} - \bar{y}\|^2 + \overline{\|y - \bar{y}\|^2} \quad (60)$$

§ Note from lemma 1 that:

$$\tilde{y} = \bar{y} \quad \text{implies} \quad b = \overline{\|y - \bar{y}\|^2} \quad (61)$$

Thus, if the Tychonov estimate, \tilde{y} , equals the measured average, \bar{y} , then:

- b is independent of λ .
- $\hat{h}(V_T, S_c)$ decreases with increasing λ , but does not shift to the right.

Theorems 1 and 2 are illustrated in figs. 12 and 13. The data are in the footnotes³ and⁴. Combining theorems 1 and 2 we observe that Tychonov optimal matrices are more robust than other matrices (evaluated at the same Tychonov weight) but that increasing the Tychonov weight causes a reduction in the robustness of the Tychonov optimal matrix.

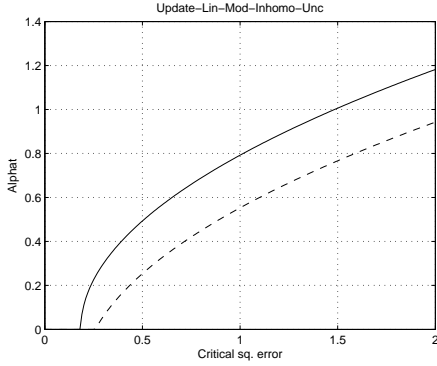


Figure 12: Robustness curves illustrating theorem 1. Tychonov-optimal (solid) and a different V matrix (dash). Tychonov weight: $\lambda = 0.3$

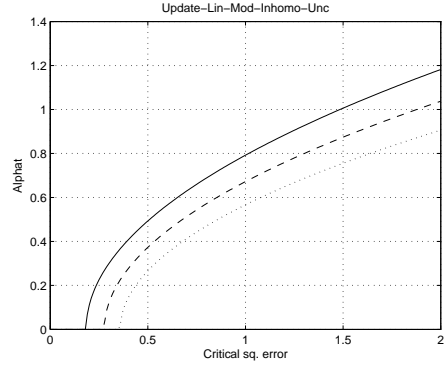


Figure 13: Robustness curves illustrating theorem 2. Tychonov optimal matrices with different weights: $\lambda = 0.3$ (solid), 0.6 (dash) and 1.0 (dot).

3.4 Proofs

Proof of theorem 1. Since λ is non-negative, we see from eq. (53) that:

$$b \leq S_o(V) \tag{62}$$

with strict inequality unless V is itself a Tychonov optimal matrix.

Hence, since V is *not* a Tychonov optimal matrix:

$$S_c - S_o < S_c - b \tag{63}$$

for all values of S_c . Hence:

$$(1 + \lambda)(S_c - S_o) < (1 + \lambda)(S_c - b) \tag{64}$$

Thus:

$$z^T z + (1 + \lambda)(S_c - S_o) < z^T z + (1 + \lambda)(S_c - b) \tag{65}$$

³The data for these figures are:

$$\begin{aligned} \tilde{y}^T &= (2.3 \ 1.2), \quad f^T = (1 \ 0.7 \ 0.3) \\ Y &= \begin{pmatrix} 3.0 & 3.2 & 2.8 & 3.1 \\ 1.5 & 1.4 & 1.6 & 1.7 \end{pmatrix} \end{aligned}$$

⁴The non-Tychonov matrix is $V = \begin{pmatrix} 1.2 & 1.6 & 2.5 \\ 0.5 & 0.9 & 1.5 \end{pmatrix}$.

Hence:

$$z^T z + (1 + \lambda)(S_c - S_o) < z^T z + \sqrt{z^T z} \sqrt{(1 + \lambda)(S_c - b)} + (1 + \lambda)(S_c - b) \quad (66)$$

$$= \left(\sqrt{z^T z} + \sqrt{(1 + \lambda)(S_c - b)} \right)^2 \quad (67)$$

Thus:

$$\sqrt{z^T z + (1 + \lambda)(S_c - S_o)} < \sqrt{z^T z} + \sqrt{(1 + \lambda)(S_c - b)} \quad (68)$$

Hence

$$\frac{1}{1 + \lambda} \left(-\sqrt{z^T z} + \sqrt{z^T z + (1 + \lambda)(S_c - S_o)} \right) < \sqrt{\frac{S_c - b}{1 + \lambda}} \quad (69)$$

which, by referring to eqs.(52) and (56) and recalling that $S_c > b$, proves the result. ■

Proof of lemma 1. Combining eqs.(54) and (55) we can write:

$$b = \lambda \|\tilde{y}\|^2 + \|\bar{y}\|^2 - \frac{1}{1 + \lambda} \left(\lambda^2 \|\tilde{y}\|^2 + 2\lambda \tilde{y}^T \bar{y} + \|\bar{y}\|^2 \right) \quad (70)$$

$$= \frac{\lambda}{1 + \lambda} \|\tilde{y}\|^2 - \frac{2\lambda}{1 + \lambda} \tilde{y}^T \bar{y} + \|\bar{y}\|^2 - \frac{1}{1 + \lambda} \|\bar{y}\|^2 \quad (71)$$

Completing the square in the first two terms in eq.(71):

$$b = \frac{\lambda}{1 + \lambda} \left(\|\tilde{y}\|^2 - 2\tilde{y}^T \bar{y} + \|\bar{y}\|^2 \right) - \frac{\lambda}{1 + \lambda} \|\bar{y}\|^2 + \|\bar{y}\|^2 - \frac{1}{1 + \lambda} \|\bar{y}\|^2 \quad (72)$$

$$= \frac{\lambda}{1 + \lambda} \|\tilde{y} - \bar{y}\|^2 + \|\bar{y}\|^2 - \|\bar{y}\|^2 \quad (73)$$

which, with the definition in eq.(59), completes the proof. ■

Proof of theorem 2. Eqs.(56) and (60) enable explicit derivation of the partial derivative of $\hat{h}(V_T, S_c)$ with respect to λ , which is found to be strictly negative for all values of S_c for which the robustness is positive ($S_c > b$):

$$\frac{\partial \hat{h}(V_T, S_c)}{\partial \lambda} = - \frac{\|\tilde{y} - \bar{y}\|^2 + (1 + \lambda)(S_c - b)}{2(1 + \lambda)^3} \sqrt{\frac{1 + \lambda}{S_c - b}} \quad (74)$$

■

4 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 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:

$$L(X, \tilde{p}) = \prod_{i=1}^N \tilde{p}(x_i|\lambda) \quad (75)$$

¶ Maximum likelihood estimate:

$$\lambda^* = \arg \max_{\lambda} L(X, \tilde{p}) \quad (76)$$

¶ 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:

$$\mathcal{U}(h, \tilde{p}) = \{p(x) : p(x) \in \mathcal{P}, |p(x) - \tilde{p}(x|\lambda)| \leq h\psi(x)\}, \quad h \geq 0 \quad (77)$$

- \mathcal{P} is the set of all normalized and non-negative pdfs on the domain of x .
- $\psi(x)$ is the known envelope function.
- 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:**

$$\hat{h}(\lambda, L_c) = \max \left\{ h : \left(\min_{p \in \mathcal{U}(h, \tilde{p})} L(X, p) \right) \geq L_c \right\} \quad (78)$$

¶ $\mu(h)$ = **inner minimum** in eq.(78).

For the info-gap model in eq.(77) $\mu(h)$ is obtained for the following choices of the pdf at the data points X :

$$p(x_i) = \begin{cases} \tilde{p}(x_i) - h\psi(x_i) & \text{if } h \leq \tilde{p}(x_i)/\psi(x_i) \\ 0 & \text{else} \end{cases} \quad (79)$$

Define:

$$h_{\max} = \min_i \frac{\tilde{p}(x_i)}{\psi(x_i)} \quad (80)$$

Since $\mu(h)$ is the product of the densities in eq.(79) we find:

$$\mu(h) = \begin{cases} \prod_{i=1}^N [\tilde{p}(x_i) - h\psi(x_i)] & \text{if } h \leq h_{\max} \\ 0 & \text{else} \end{cases} \quad (81)$$

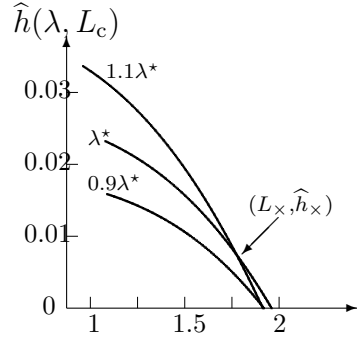
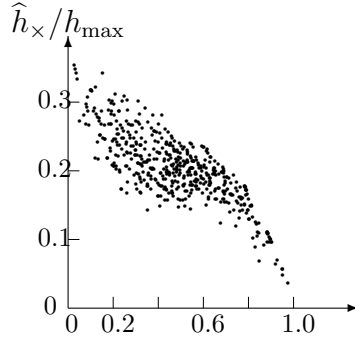
¶ $\mu(h)$ and $\hat{h}(\lambda, L_c)$:

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

$$\mu(h) = L_c \quad \text{implies} \quad \hat{h}(\lambda, L_c) = h \quad (82)$$

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

Robustness

Critical likelihood, $\log_{10} L_c$ Figure 14: Robustness curves. $\lambda^* = 3.4065$. $L_\times / L[X, \tilde{p}(x|\lambda^*)]$ Figure 15: Loci of intersection of robustness curves $\hat{h}(\lambda^*, L_c)$ and $\hat{h}(1.1\lambda^*, L_c)$.¶ **Robustness curves** in fig. 14 based on:

- Eqs.(81) and (82).
- Nominal pdf is exponential, $\tilde{p}(x|\lambda) = \lambda \exp(-\lambda x)$ with $\lambda = 3$.
- Envelope function is constant, $\psi(x) = 1$.
- Random sample, X , with $N = 20$.
- MLE of λ , eq.(76): $\lambda^* = 1/\bar{x}$ 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.14, $\hat{h}(\lambda, L_c)$ vs. L_c , hits horizontal axis when $L_c =$ likelihood:

$$\hat{h}(\lambda, L_c) = 0 \quad \text{if} \quad L_c = L[X, \tilde{p}(x|\lambda)] \quad (83)$$

- 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 $(L_\times, \hat{h}_\times)$:
 - $\lambda^* \succ 1.1\lambda^*$ for $L_c > L_\times$ and $h < \hat{h}_\times$.
 - $1.1\lambda^* \succ \lambda^*$ else.

¶ **500 repetitions:**

- λ^* dominates $0.9\lambda^*$.
- Preferences reverse between λ^* and $1.1\lambda^*$.
- Normalized (h_x, L_x) in fig. 15.
- Center of cloud: $(0.5, 0.2)$. Typical cross of robustness curves at:
 - L_c about half 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.(77), 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.

5 Forecasting

¶ Source material: Yakov Ben-Haim, 2008, Info-gap forecasting and the advantage of sub-optimal models, *European Journal of Operational Research*, to appear.

5.1 Preliminary Example: European Central Bank Interest Rates

| Date | Interest rate | Implied λ |
|-------------|---------------|-------------------|
| 1 Jan 1999 | 4.50 | |
| 9 Apr 1999 | 3.50 | 0.778 |
| 5 Nov 1999 | 4.00 | 1.143 |
| 4 Feb 2000 | 4.25 | 1.063 |
| 17 Mar 2000 | 4.50 | 1.059 |
| 28 Apr 2000 | 4.75 | 1.056 |
| 9 Jun 2000 | 5.25 | 1.105 |
| 28 Jun 2000 | 5.25 | 1.000 |
| 1 Sep 2000 | 5.50 | 1.048 |
| 6 Oct 2000 | 5.75 | 1.045 |
| 11 May 2001 | 5.50 | 0.957 |
| 31 Aug 2001 | 5.25 | 0.955 |

Table 1: Interest rates for overnight loans at the European Central Bank (marginal lending facility). Source: <http://www.ecb.int/stats/monetary/rates/html/index.en.html>

¶ ECB overnight interest rates: table 1.

- First loans: 1999.
- Data through August 2001.

¶ El-Qaeda attacks in US: 11 Sept 2001.

- Predict next ECB interest rate?
- Asymmetric uncertainty: rate will go down.

¶ Questions:

- How to forecast the rate?
- How to assess confidence in the forecast?

5.2 Info-Gap Forecasting: Formulation

5.2.1 The Estimated System and its Uncertainty

¶ **N -dimensional system** whose average behavior is:

$$y_t = A_t y_{t-1} \quad (84)$$

Zero-mean, additive, random disturbances are ignored and all other inputs are incorporated in the multi-dimensional state vector y_t .

¶ **Solution of eq.(84):**

$$y_{T+k} = \left(\prod_{i=1}^k A_{T+i} \right) y_T \quad (85)$$

where the product operator is lefthand matrix multiplication: $\prod_{i=1}^k A_{T+i} = A_{T+k} \prod_{i=1}^{k-1} A_{T+i}$.

¶ **1-Step and k -Step Forecast:**

- From eq.(85): a k -step process is a 1-step process with coefficient matrix $A^{(k)} = \prod_{i=1}^k A_{T+i}$.
- If the matrices A_{T+i} belong to an info-gap model, $\mathcal{U}(h, \tilde{A})$, then the product matrix $A^{(k)}$ also belongs to an info-gap model, $\mathcal{U}_k(h, \tilde{A}^k)$:

$$\mathcal{U}_k(h, \tilde{A}^k) = \left\{ A = \prod_{i=1}^k A_i : A_i \in \mathcal{U}(h, \tilde{A}) \right\}, \quad h \geq 0 \quad (86)$$

- Conclusions about 1-step forecasts hold for k -step forecasts also.

¶ **Info-gap uncertainty** in transition matrices A_t . E.g.:

$$\mathcal{U}(h, \tilde{A}) = \left\{ A_t, t > T : \tilde{A}_{ij} - h v_{ij} \leq [A_t]_{ij} \leq \tilde{A}_{ij} + h w_{ij}, i, j = 1, \dots, N \right\}, \quad h \geq 0 \quad (87)$$

5.2.2 Forecasting with Slope Adjustment

¶ **“Slope-adjusted” predictor:**

$$y_t^s = B y_{t-1}^s \quad (88)$$

where B is a constant matrix which we are free to choose. The question is: how to choose B ?

¶ **Vector of average forecast errors** for time $t = T + k$ (ignoring zero-mean, additive, random disturbances), based on knowledge of y_T , is:

$$\eta_k(B, A_t) = y_{T+k}^s - y_{T+k} = \left(B^k - \prod_{i=1}^k A_{T+i} \right) y_T \quad (89)$$

Judicious choice of B can reliably compensate for deviation of A_{T+i} from \tilde{A} .

5.2.3 Definition of the Robustness Function

¶ **Requirement:** satisfice the forecast error:

$$|\eta_{k,m}(B, A_t)| \leq \varepsilon_c \quad (90)$$

¶ **Robustness:**

$$\hat{h}(B, \varepsilon_c) = \max \left\{ h : \left(\max_{\substack{A_{T+i} \in \mathcal{U}(h, \tilde{A}) \\ i=1, \dots, k}} |\eta_{k,m}(B, A_t)| \right) \leq \varepsilon_c \right\} \quad (91)$$

5.2.4 Evaluating the Robustness Function

¶ We evaluate the robustness for 1-step forecast.

¶ The robustness in eq.(91) can be written:

$$\hat{h}(B, \varepsilon_c) = \max \left\{ h : \left(\max_{A_{T+1} \in \mathcal{U}(h, \tilde{A})} \eta_{1,m}(B, A_{T+1}) \right) \leq \varepsilon_c \right. \\ \left. \text{and} \left(\min_{A_{T+1} \in \mathcal{U}(h, \tilde{A})} \eta_{1,m}(B, A_{T+1}) \right) \geq -\varepsilon_c \right\} \quad (92)$$

¶ The 1-step forecast error for the m th state variable, from eq.(89), is:

$$\eta_{1,m}(B, A_{T+1}) = \underbrace{\sum_{n=1}^N [B - \tilde{A}]_{mn} y_{T,n}}_{\delta} - \sum_{n=1}^N [A_{T+1} - \tilde{A}]_{mn} y_{T,n} \quad (93)$$

δ can be positive or negative and is controlled through the choice of the forecast matrix B .

¶ **Define:**

$$\theta_c(h) = \max_{A_{T+1} \in \mathcal{U}(h, \tilde{A})} \sum_{n=1}^N [A_{T+1} - \tilde{A}]_{mn} y_{T,n} \quad (94)$$

$$\theta_a(h) = - \min_{A_{T+1} \in \mathcal{U}(h, \tilde{A})} \sum_{n=1}^N [A_{T+1} - \tilde{A}]_{mn} y_{T,n} \quad (95)$$

- Contraction axiom implies that $\theta_a(0) = \theta_c(0) = 0$.
- Nesting axiom then implies that $\theta_a(h) \geq 0$ and $\theta_c(h) \geq 0$ and monotonic for all $h \geq 0$.

¶ From eqs.(93)–(95), the robustness is:

$$\hat{h}(B, \varepsilon_c) = \max \{h : \delta + \theta_a(h) \leq \varepsilon_c \quad \text{and} \quad -\delta + \theta_c(h) \leq \varepsilon_c\} \quad (96)$$

¶ Plotting the robustness.

- Define:

$$\varepsilon(h) = \min \{\delta + \theta_a(h), -\delta + \theta_c(h)\} \quad (97)$$

- Plot of h vs. $\varepsilon(h)$ is the same as a plot of $\hat{h}(B, \varepsilon_c)$ vs. ε_c .
- Fig. 16: The vertical axis is h or $\hat{h}(B, \varepsilon_c)$, while the horizontal axis is $\varepsilon(h)$ or ε_c .

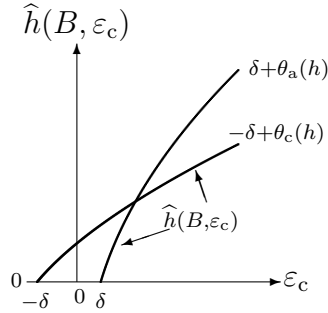


Figure 16: Robustness function based on eqs.(96) and (97).

5.2.5 Crossing of Robustness Curves and the Advantage of Sub-Optimal Models

¶ 1-step forecast error, eq.(93):

$$\eta_{1,m}(B, A_{T+1}) = \underbrace{\sum_{n=1}^N [B - \tilde{A}]_{mn} y_{T,n}}_{\delta} - \sum_{n=1}^N [A_{T+1} - \tilde{A}]_{mn} y_{T,n} \quad (98)$$

¶ Applies also to k -step error, with notational change.

¶ If A_t will be constant at \tilde{A} in the future, then the k -step prediction error for the m th state variable is:

$$\eta_{k,m}(B, \tilde{A}) = \underbrace{\sum_{n=1}^N \left(B^k - \tilde{A}^k \right)_{mn}}_{\varepsilon^*} y_{T,n} \quad (99)$$

- One is tempted to choose $B = \tilde{A}$ in order to minimize the anticipated prediction error ε^* .
- Is this a good choice?

¶ **Theorem:** There exist sub-optimal models for 1-step forecasting which are more robust than optimal models.

5.3 Example: 1-Dimensional System

¶ **The system.** Consider a scalar system whose average behavior evolves as:

$$y_t = \lambda_t y_{t-1} \quad (100)$$

¶ **Asymmetric uncertainty:** λ_t tends to drift up.

$$\mathcal{U}(h, \tilde{\lambda}) = \left\{ \lambda_t, t > T : 0 \leq \frac{\lambda_t - \tilde{\lambda}}{\tilde{\lambda}} \leq h \right\}, \quad h \geq 0 \quad (101)$$

¶ **Slope-adjusted forecaster.**

$$y_t^s = \ell y_{t-1}^s \quad (102)$$

¶ **Robustness of k -step forecast** with growth coefficient ℓ , defined in eq.(91):

$$\hat{h}(\ell, \varepsilon_c) = \begin{cases} 0 & \text{if } \varepsilon_c \leq (\ell^k - \tilde{\lambda}^k) y_T \\ \left(\frac{\varepsilon_c + \ell^k y_T}{\tilde{\lambda}^k y_T} \right)^{1/k} - 1 & \text{else} \end{cases} \quad (103)$$

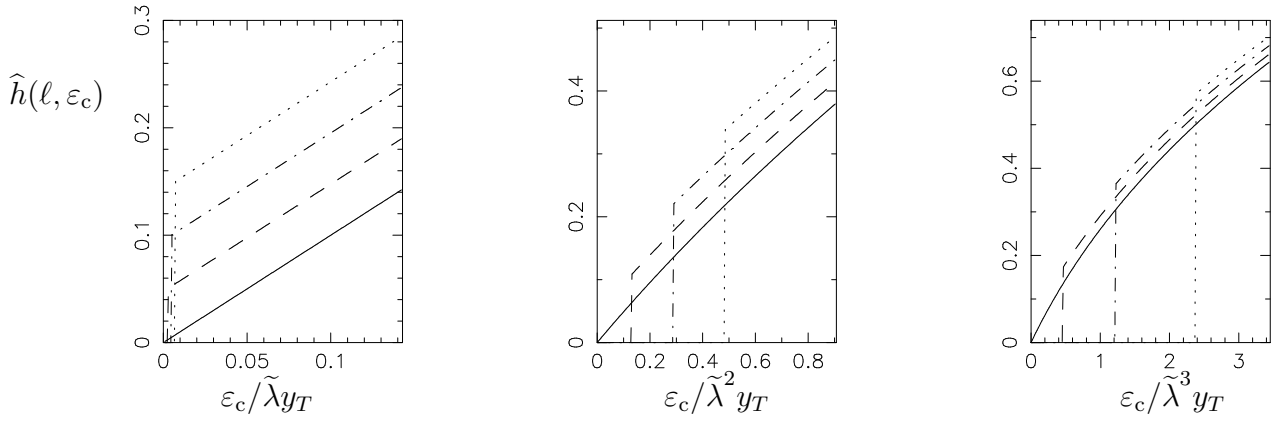


Figure 17: Robustness vs. normalized forecast error, eq.(103), for $\ell = 1.05, 1.1, 1.15, 1.2$ from bottom to top curve. $\tilde{\lambda} = 1.05, y_T = 1. k = 1$ (left), 2 (mid), 3 (right).

¶ **Numerical example**, fig. 17.

- Lowest curve in each frame is nominal forecaster: $\ell = \tilde{\lambda} = 1.05$.
- ℓ increases by 0.05 with each higher curve.
- Horizontal axis: satisfied forecast error, ε_c , normalized to nominal forecast value, $\tilde{\lambda}^k y_T$.
- **1-step forecast** (left frame):

◦ Slope-adjusted predictors are far more robust than the nominal predictor for essentially all levels of forecast error ε_c .

◦ For instance, consider 5% fractional forecast error, $\varepsilon_c/\tilde{\lambda}^k y_T = 0.05$.

$$\hat{h}(1.05, \varepsilon_c) = 0.050 \text{ (bottom curve), and } \hat{h}(1.2, \varepsilon_c) = 0.19 \text{ (top curve).}$$

The slope-adjusted predictor is about 4 times more robust than the nominal predictor.

- **2- and 3-step forecast** (middle and right frames):

◦ robustness premium of slope-adjusted forecaster, $\ell > \tilde{\lambda}$, compared to the nominal predictor, $\ell = \tilde{\lambda}$, becomes smaller as the horizon of the prediction increases.

5.4 Robustness and Probability of Forecast Success

¶ **1-step forecast error** of m variable, from eq.(89), is:

$$\eta_{1,m}(B, A_{T+1}) = \sum_{n=1}^N [B - A_{T+1}]_{mn} y_{T,n} \quad (104)$$

¶ **Forecast is successful if:**

$$|\eta_{1,m}(B, A_{T+1})| \leq \varepsilon_c \quad (105)$$

- This can be written explicitly as:

$$-\varepsilon_c + \sum_{n=1}^N [B - \tilde{A}]_{mn} y_{T,n} \leq \underbrace{\sum_{n=1}^N [A_{T+1} - \tilde{A}]_{mn} y_{T,n}}_u \leq \varepsilon_c + \sum_{n=1}^N [B - \tilde{A}]_{mn} y_{T,n} \quad (106)$$

which defines the variable u .

- Recalling the definition of δ in eq.(93), the condition for forecast success in eq.(106) becomes:

$$\delta - \varepsilon_c \leq u \leq \delta + \varepsilon_c \quad (107)$$

¶ **Probability of forecast success:**

- $F(u)$ is unknown cumulative probability distribution of u .
- Probability of forecast success with model B :

$$P_s(B) = F(\delta + \varepsilon_c) - F(\delta - \varepsilon_c) \quad (108)$$

¶ **Is robustness, $\hat{h}(B, \varepsilon_c)$, a proxy for probability of success, $P_s(B)$?**