$\{$Designs$\} \times \{$Robust$\}$ =

Robust Designs ?

As many uses of the word as there are users

Stanislav Kolenikov
skolenik@unc.edu
117 New West, Cameron Ave,
University of North Carolina,
Chapel Hill, NC 27599-3260, US

December 10, 2001

Advanced Search Preferences Language Tools Search Tips

Google

robust | Google Search

I'm Feeling Lucky

Web Images Groups Directory

Searched **English and Russian** pages for **robust**. Results **1 - 50** of about **1,340,000**. Search took **0.39** seconds.

Fast **Robust** Predicates for Computational Geometry
Adaptive Precision Floating-Point Arithmetic and Fast
Robust Predicates for Computational Geometry. **...**
Description: C code for orientation and incircle tests using adaptive-precision floating-point arithmetic
Category: Science > Math > Geometry > Computational Geometry > Software
www.cs.cmu.edu/~quake/robust.html - 7k - Cached - Similar pages

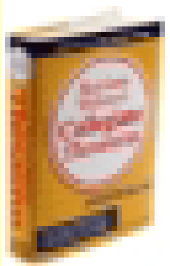
CMU **Robust** Speech Recognition Home Page
CMU **Robust** Speech Recognition Group. Rita Singh, Mike Seltzer, Jon Nedel, Bhiksha
Raj, Juan Huerta, and Richard Stern at the Great Wall of China, an excellent **...**
www.cs.cmu.edu/afs/cs.cmu.edu/user/robust/www/home.html - 6k - Cached - Similar pages
[More results from www.cs.cmu.edu]

www3.interscience.wiley.com/cgi-bin/jtoc?ID=5510
Similar pages

The MathWorks - **Robust** Control Toolbox
... Design **robust** multivariable feedback control systems. This toolbox contains specialized
functions for advanced "**robust**" multivariable feedback control system **...**
www.mathworks.com/products/robust/ - 11k - Cached - Similar pages

Robust Hyperlinks and **Robust** Locations Home Page
... Fri, 05 Oct 2001 00:01:41 GMT 404 - Page Not Found.
Just kidding! This is not the first time **...**
www.cs.berkeley.edu/~phelps/Robust/ - 1k - Cached - Similar pages

Merriam-Webster's Collegiate Thesaurus



Search

Dictionary

Entry Word: **robust**
Function: *adjective*
Text: **1**
Synonyms FLOURISHING, booming, prospering, prosperous, roaring, thrifty, thriving
2
Synonyms STRONG 3, concentrated, full-bodied, lusty, potent

What is robust in robust designs?

- Distributional robustness
- Model misspecification robustness
  1. nonlinearity
  2. heteroskedasticity
  3. time trend
- Noise (Taguchi designs)
- Missing data
- Errors in the design variables

Distributional robustness - I

What if the errors in $y = X\theta + \epsilon$ are not normal?

$$\widehat{\theta}_{OLS} = (X^T X)^{-1} X^T y \Longrightarrow$$

$$y_k \to \infty \text{ for some } k \Rightarrow \widehat{\theta}_{OLS} \to \infty$$

Theory (Huber 1981): functional derivatives

Tools: influence curve (Hampel 1974), breakdown point

Distributional robustness - II

Estimation: M-estimates

$$\sum_i \rho(y_i - \sum_j x_{ij}\theta_j) \to \min \theta,$$

where $\rho(z) = o(z^2)$ as $z \to \infty$, or equivalently ($\psi(\cdot) = \rho'(\cdot)$)

$$\sum_i \psi(y_i - \sum_j x_{ij}\theta_j)x_{ik} \to \min, \quad k = 1, \ldots, p$$

Examples:

Tukey's biweight: $\psi(x) = x(1 - x^2)^2 \, I(|x| < 1)$

Huber's minimal information: $\psi(x) = \max[-k, \min(k, x)]$

Distributional robustness - III

Also important: geometry of factors. Huber (1975):

- Avoid outliers among the independent variables

- Always calculate the diagonal h_{ii} of the projection matrix $H = X(X^T X)^{-1} X^T$. If a particular h_{ii} is close to 1, then decrease it by (approximate) replication of that observation.

See also: Stat 174 and Richard Smith's notes to it (Smith and Young 2001) for discussion of regression diagnostics.

Misspecification robustness - I

What if the true relation is not $y = X\theta + \epsilon$ but some other $y = g(x, \beta)+$errors? Then the estimates $\widehat{\theta}_{OLS}$ will lose all nice properties such as being the MLE estimates, being unbiased, and being of smallest variance.

Box and Draper (1959) analyzed misspecified polynomial case with the requirement on the design that it should represent the true function as well as possible within the region of interest, and lend itself to a verification of the specification. They concluded that:

> . . . *at least, in the cases considered, the optimal design in typical situations in which both variance and bias occur is very nearly the same as would be obtained if* variance were ignored completely *and the experiment designed so as to* minimize the bias alone.

Misspecification robustness - II

In particular, if a polynomial of insufficient order $d_1 < d_2$ is fitted to the data, then MSE

> ... *arising from bias alone is minimized when all moments of the design up to that of order* $(d_2 + 1)$ *are equal to the moments of a uniform distribution* ...

and the optimal design does not differ much from the one outlined above when variance and bias contributions to the MSE are “equal”, or even in the case when variance is somewhat “greater” than the bias.

Modernized approach: smooth functions with bounded d*-th derivative* $\Rightarrow$ *splines?*

Demonstration - I

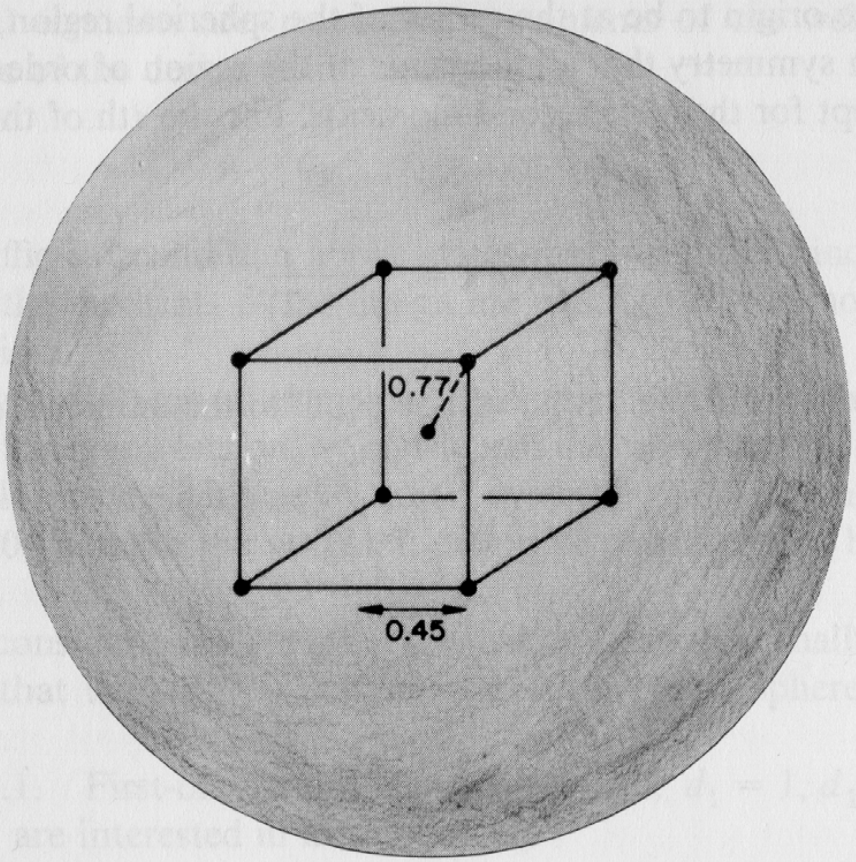


FIGURE 13.6. A first-order (two-level factorial) design in three factors which minimizes squared bias from second-order terms when the region of interest is a sphere of unit radius.

Box and Draper (1987)

Demonstration - II

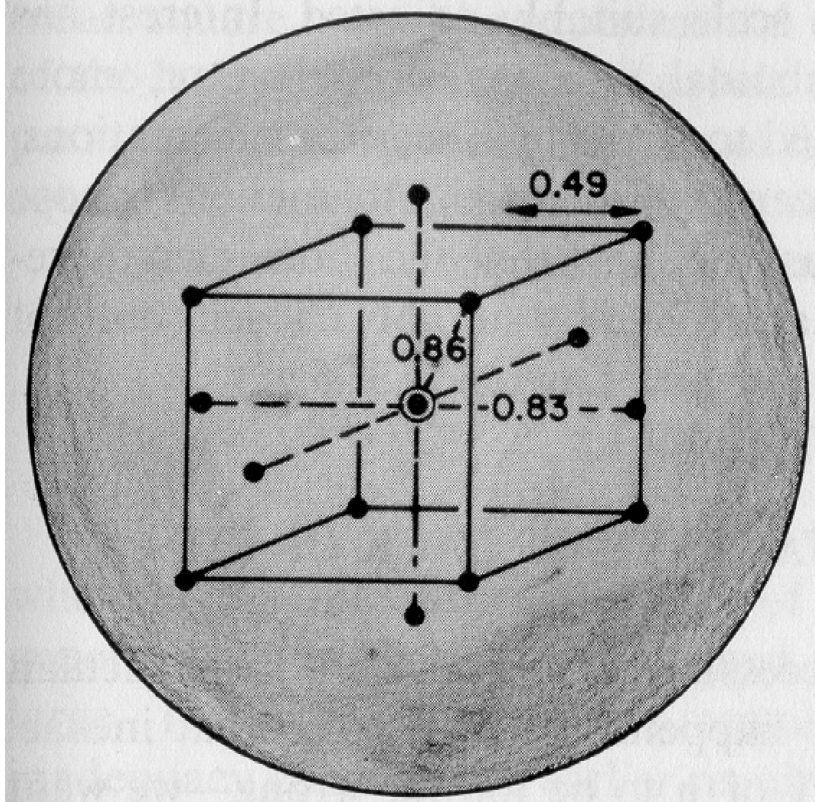


FIGURE 13.7. A second-order composite rotatable design which minimizes squared bias from third-order terms when a weight function uniform over a spherical region of interest of radius 1 is assumed.

Box and Draper (1987)

Misspecification robustness - III

Huber (1975): need to consider broader class of deviations and minimax estimators.

Global fit: least favorable (symmetric) true function for a design on [-1,1] is approximately quadratic (to be precise, $(ax^2+b)^+$). Hence the minimax design is a multiple of $\{-1, 0, 0, 1\}$.

Slope estimation: least favorable true function is approximately cubic. The “uniform” design performs best.

Extrapolation: $g^{(h)}$ is bounded $\Rightarrow$ the optimal design points correspond to the roots of Chebyshev polynomial of order h, and the weights are found from the design moments of the order up to $h-1$.

Misspecification robustness - IV

Wiens (1991): other problems to consider.

$$\text{Fitted: } Y_{ij} = \theta^T z(x_i) + \epsilon_{ij}$$

$$\text{True: } Y_{ij} = \mu_i + \epsilon_{ij},\ \mu_i = \theta^T z(x_i) + f(x_i)$$

Problem 1: to maximize the minimum power in determining misspecification.

Problem 2: to minimize the bias in estimation of σ^2.

Misspecification robustness - V

Formalization:

$$\xi : \left(2^S, \mathcal{B}(S)\right) \to \mathbf{R} \text{ — design measure;}$$

$$\mathcal{F}_\eta^+ = \left\{ f \mid \int f^2(x)\, dx \geq \eta^2,\ \int z(x) f(x)\, dx = 0 \right\}$$

$$\mathcal{F}_\eta^- = \left\{ f \mid \int f^2(x)\, dx \leq \eta^2,\ \int z(x) f(x)\, dx = 0 \right\}$$

$$B_\xi = \int z(x) z^T(x)\, d\xi(x), \quad b_{f,\xi} = \int z(x) f(x)\, d\xi(x)$$

$$\mathcal{B}(f, \xi) = \int f^2(x)\, d\xi(x) - b_{f,\xi}^T B_\xi^{-1} b_{f,xi}$$

(L_2 distance from f to the space of $\theta^T z(x)$)

Misspecification robustness - VI

Then the solutions to the above two problems are

$$\xi^{(1)} = \arg\max_{\mathcal{P}} \min_{\mathcal{F}_\eta^+} \mathcal{B}(f, \xi)$$
$$\xi^{(2)} = \arg\min_{\mathcal{P}} \max_{\mathcal{F}_\eta^-} \mathcal{B}(f, \xi)$$

In fact, the solutions coincide, and it is the uniform design:

$$\xi^{(1)} = \xi^{(2)} = U(S)$$

In practice:

- discrete uniform design: points equally spaced across S;
- randomized design: point drawn at random from $U(S)$.

Heteroskedasticity - I

$$Y(x) = z(x)^T\theta + f(x) + \epsilon(x),$$
$$\text{Var}\,\epsilon(x) = \sigma^2 g(x).$$

Typical approach: use weights inversely proportional to $\widehat{g}(x)$.

What does minimax do for the L_2 loss, i.e. integrated mean squared error (Wiens 1998)?

$$IMSE = \int_S E\Big[\hat{Y}(x) - E[Y|x]\Big]^2 dx$$

Heteroskedasticity - II

$$\nu = \sigma^2/n\eta^2 \text{ — variance to squared bias ratio;}$$
$$S \text{ is an ellipsoid} \in \mathbf{R}^q, \quad \Omega = \int_S dx$$
$$l(u;\gamma) = 1 + \frac{u^2}{(q+2)\gamma^2}, \quad h_0(u;\gamma) = \frac{a\nu(b+u^2)}{1+c\nu l^2(u;\gamma)}$$
$$\gamma_0 = \arg\min_{\gamma\geq 0}\Big(\Omega^{-1}a(\gamma)(b(\gamma)+q\gamma)+(4\Omega^2c(\gamma))^{-1}\Big),$$

where $a = a(\gamma)$, $b = b(\gamma)$, $c = c(\gamma)$ are certain normalizing constants. Then the minimax design ξ_0 has the density $h_0(\|x\|;\gamma_0)$ (Wiens 1998, Theorem 3).

Heteroskedasticity - III

If the model is correctly specified ($f \equiv 0$), then the minimax design density is

$$k_0(x) = \frac{\left(z(x)^T A^{-1} z(x)\right)^{2/3}}{\int_S \left(z(x)^T A^{-1} z(x)\right)^{2/3} dx},$$

where $A = \int_S z(x) z(x)^T \, dx$ (Wiens 1998, Theorem 5). The regression weights are then $w_0(x) = \Omega / k_0(x)$; the least favorable skedasticity function: $g(x) = w_0(x)^{-1/2}$.

Robustness to time trend - I

Experiments are performed sequentially — what if there is trend and/or serial correlation?

- Confound the trend with the factor interactions assuming linear or quadratic trend (Atkinson 1996)

- Assume AR in errors and construct an optimal design.
  Constantine (1989) builds a modification of the no-correlation-optimal design based on the available information on the (signs of) error correlation by (counter-)matching signs of the factors in consequent experiments.

- Econometrics: covariance matrix estimates robust to autocorrelation (Newey and West 1987)

Infinitesimal Approach - I

Wiens and Zhou (1997), following Hampel et. al. (1986).

$$Y_i = \theta^T z(x_i) + f(x_i) + \epsilon_{ij}, \quad \mathsf{Cov}\,\epsilon = \sigma^2 P$$

Objective:

$$\mathcal{D}(f, \xi, P) = \sigma^{2p} |B_\xi|^{-2} \left| \frac{Z^T P Z}{n} \right| \left(1 + \frac{n}{\sigma^2} b_{f,\xi}^T \left(\frac{Z^T P Z}{n} \right)^{-1} b_{f,\xi} \right)$$

Apparatus: Gateuax derivatives.

Infinitesimal Approach - II

Change of variance function ($P_0 = I$):

$$CVF(\xi, P) = \frac{\frac{d}{dt}\mathcal{D}(f, \xi, (1-t)P_0 + tP)|_{t=0}}{\mathcal{D}(f_0, \xi, P_0)}$$

Change of bias function ($f_0 = 0$):

$$CBF(\xi, P) = \frac{\frac{1}{2}\frac{d^2}{dt^2}\mathcal{D}(1-t)f_0 + tf, \xi, P_0)|_{t=0}}{\mathcal{D}(f_0, \xi, P_0)}$$

Change of variance sensitivity:

$$CVS(\xi, \mathcal{P}) = \sup_{P \in \mathcal{P}} \mathrm{trace}\Big(\frac{Z^T(P-I)Z}{n} B_\xi^{-1}\Big)$$

Change of bias sensitivity:

$$CBS(\xi, \mathcal{F}) = \sup_{f \in \mathcal{F}} \Big(n b_{f,\xi}^T B_\xi^{-1} b_{f,\xi}\Big)$$

Infinitesimal Approach - III

V-robust design for given α minimizes $\mathcal{D}(f_0, \xi, P_0)$ s.t. $CVS(\xi, \mathcal{P}) \leq \alpha$; it is *most V-robust* design if $\alpha = \inf\left\{CVS(\xi, \mathcal{P}) : \xi \in \Xi\right\}$ for a certain class Ξ of designs.

B-robust design for given β minimizes $\mathcal{D}(f_0, \xi, P_0)$ s.t. $CBS(\xi, \mathcal{F}) \leq \beta$; it is *most B-robust* design if $\alpha = \inf\left\{CBS(\xi, \mathcal{F}) : \xi \in \Xi\right\}$.

M-robust design: the one that is both V- and B-robust; *most M-robust* if it is both most V-robust and B-robust (do such designs exist?).

Infinitesimal Approach - IV

Mild autocorrelation:

$$\mathcal{P}_1 = \{P|\rho(s) = 0, |s| > 1; 0 < c_0 \leq \rho(1) < 1\},$$
$$\mathcal{P}_2 = \{P|\rho(s) = 0, |s| > 1; -1 < \rho(1) \leq -c_1 < 0\}$$

The V-robust designs are similar to Constantine (1989) and Jenkins and Chanmugam (1962): $\langle\, 1, -1, \ldots, 1, -1, (0)\,\rangle$ for $\mathcal{P}_1$ and $\langle\, 1, 1, \ldots, 1, (0), \ldots, -1, \ldots, -1\,\rangle$ for $\mathcal{P}_2$, where (0) indicates that zero is included for an odd size design.

The most V-robust designs are given by some trigonometric functions of the index. Their empirical distribution functions converge weakly to $1/2 + \pi^{-1} \arcsin x$. (Also the optimal polynomial regression design.)

Infinitesimal Approach - V

B-robustness: bounded bias designs robust to biases in $\mathcal{F}_\eta^-$. Only $\xi \ll$ Lebesgue measure are admissible for this problem under L_2 loss $\Rightarrow$ no "good" finite design.

The randomized approximation (weakly convergent to a B-robust design) is of the form $x_{i,j} = t_{i,j}u_i$ where $t_{i,1}, \ldots, t_{1,2^q}$ are distributed uniformly on a unit sphere in $\mathbf{R}^q$ and u_i are distributed according to a certain density dependent on ν. Uniform distribution on S is B-robust.

Counterintuitively, most B-robust design is δ_0.

Infinitesimal Approach - VI

M-robust designs are constructed from B-robust designs by appropriate permutation of points.

If ξ^* is B-robust in a class of designs, and there exists a permutation of the design points $\xi^* \mapsto \xi^{**}$ s. t. $CVS(\xi^{**}, \mathcal{P}) \leq \alpha$, then ξ^{**} maximizes B_ξ among the designs with small CBS, and hence among a smaller class of bounded CBS and CVS.

It is unclear whether most M-robust designs exist.

Infinitesimal Approach - VII

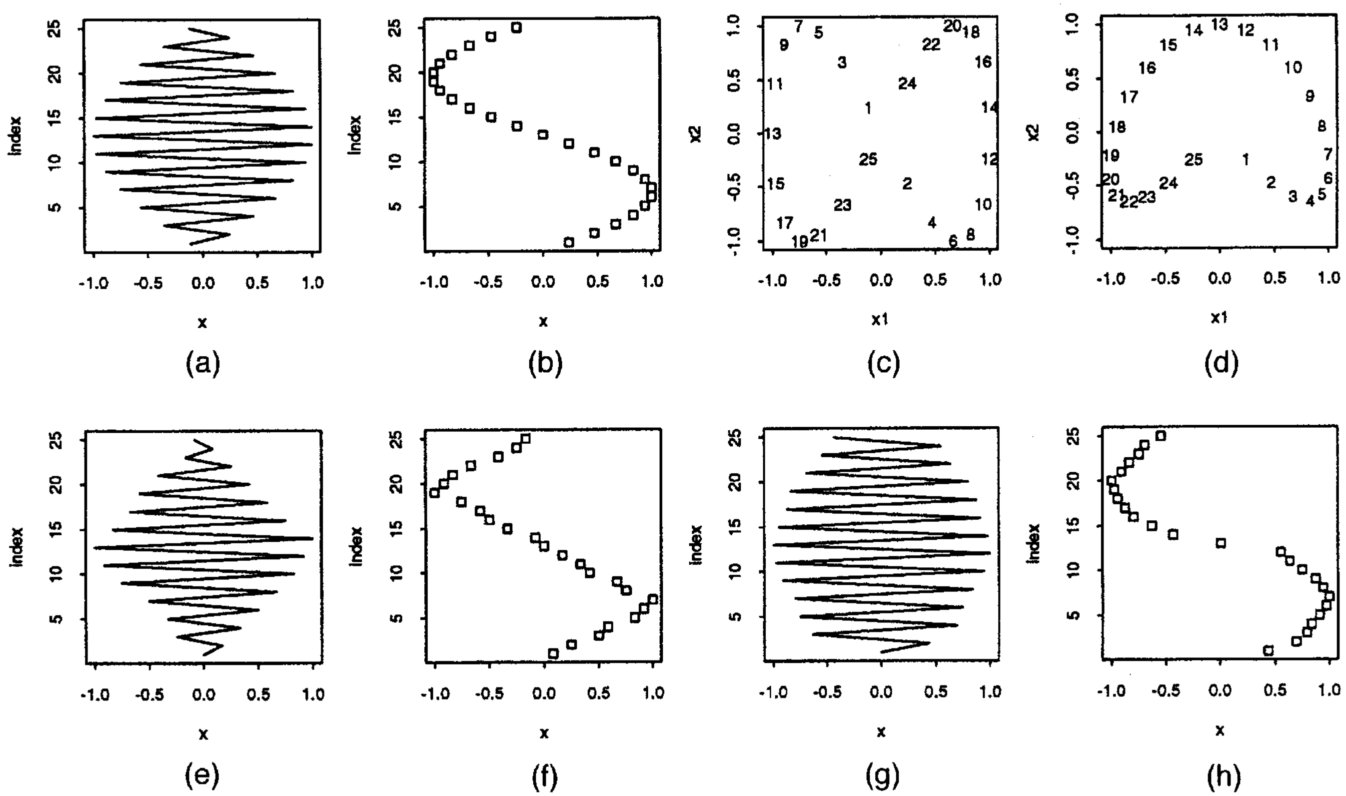


Figure 1. Most V-, Most B-, and M-Robust Designs, $n = 25$. (a), (b): Most V-robust designs in $[-1, 1]$ for $\mathcal{P}_1$ and $\mathcal{P}_2$; (c), (d): most V-robust designs in $[-1, 1] \times [-1, 1]$, with the indices of design points plotted for $\mathcal{P}_1$ and $\mathcal{P}_2$; (e), (f): most B-robust designs for $\mathcal{F}_1$, ordered for $\mathcal{P}_1$ and $\mathcal{P}_2$; (g), (h): M-robust designs for $\mathcal{F}_1$, ordered for $\mathcal{P}_1$ and $\mathcal{P}_2$.

Fig. 1 of Wiens and Zhou (1997).

Robustness to missing data

The property we want to preserve: estimability. Then *robustness*, in this context, is the property of the design to remain connected after t observations are lost. Usually the analysis is performed for small t=1, 2, 3 or so.

An observation in a design is missing $\Leftrightarrow$ there is an extra dummy variable for this observation. Lal, Gupta and Bhar (2001): sufficient conditions of the missingness-robustness of parameters related to the subset of variables (say X_1):

$$I_n - C_*^{-1/2} V^T C_{\theta_1}^{-} V C_*^{-1/2} > 0, \quad C_* = U^T P U,\ V = X_1^T P U,$$

$$P = I - X_2 (X_2' X_2)^{-} X_2^T,\ U = \mathrm{diag}(I\{x_i \text{ is missing}\}),\ X = \left(X_1\, X_2\right)$$

Efficiency of the residual design can also be established.

Errors in factor levels - I

What if factor levels are not achieved precisely? Easy to see: β_{OLS} are biased, as well as the predictions $\widehat{y}$.

"Good" design (Draper and Beggs 1971): minimize the discrepancy measured e.g. by

$$G = E[(\widehat{y}_X - \widehat{y}_Z)^T(\widehat{y}_X - \widehat{y}_Z)],$$

where predictions $\widehat{y}_X$ are predictions based on the selected (but not achieved) factor levels X, and $\widehat{y}_Z$ are predictions based on the true (but unobserved) factor levels Z.

Errors in factor levels - II

G is comprised of a bias term and a variance term:

$$Z = X + \epsilon, \quad C_1 = \epsilon^T X(X^T X)^{-1} X^T \epsilon,$$

$$C_2 = X(X^T X)^{-1} X^T - X(Z^T Z)^{-1} Z^T -$$

$$Z(Z^T Z)^{-1} X^T + Z(Z^T Z)^{-1} X^T X(Z^T Z)^{-1} Z^T$$

$$G = \beta^T E[C_1]\beta + \sigma_y^2 \mathrm{tr}\, E[C_2],$$

$$\approx \beta^T E[C_1]\beta + \sigma_y^2 \mathrm{tr}\,(X^T X)^{-1} E[\epsilon^T \epsilon]$$

assuming that the errors in factors ϵ are small, so that their third powers are negligible.

Errors in factor levels - III

The results of Draper and Beggs (1971):

- dim$X = 1 \Rightarrow$ the design with $\sum_u x_u = 0$ minimizes G;
- dim$X \geq 2 \Rightarrow$ the design with $\sum_u x_{iu} = \sum_u x_{iu} x_{ju} = 0 \; \forall \, i, j$ minimizes G;
- dim$X \geq 2$, errors in factors are correlated $\Rightarrow$ need $\sum_u x_{iu} = 0 \; \forall \, i$ + conditions on the design second moments depending on the covariance of the factor errors. Small correlations $\Rightarrow$ orthogonal design is still OK.

Robustness to noise factors - I

- *Control factors*: fully controlled in the experiment and in production;

- *Noise factors*: difficult/expensive/etc. to control in production, but controllable in the lab;

  or: the factors that affect variance rather than the level of the response.

Aim: minimize variance of the response while keeping its level fixed at a pre-specified level.

Atkinson (1996): the dependence of the variance on the mean can evidence the lack of transformation to normality.

Robustness to noise factors - II

Taguchi setup

1. Perform (complete) factorial experiment on the noise factors for each combination of the control factors levels in the design. Control factors may form fractional factorial design.

2. Find estimates of both the level and the variance of the response.

3. Determine which of the control factors affect the level of the response, and which affect the variance.

4. Choose the combination of the control factors that minimizes the variance of the response.

5. Given the above combination of the control factors, choose the combination of the control factors affecting the level only to bring it to the desired value.

Robustness to noise factors - III

Standard setup

Single (fractional) factorial experiment is performed on both control and noise factors. The focus is on interactions of the noise and control factors, so all such interactions should be estimable.

$$y = X\beta + Z\gamma + X * Z\delta + \epsilon,$$

where X represent the control factors, Z, the noise factors, and $X * Z$, the interaction between the two. In the experiment Z can be controlled, but in production, they are difficult to control, so it will be assumed that Z are random with variance σ_Z^2. Then $\operatorname{Var} y = \sigma_\epsilon^2 + (\gamma + X\delta)^2\sigma_Z^2 \to \min_X$.

References -I

Atkinson, A. C. (1996). The Usefulness of Optimum Experimental Designs. *JRSS (B)*, **58**, 59–76.

Box, G. E. P., and N. R. Draper (1959). A Basis for the Selection of a Response Surface Design. *The American Statistical Association Journal*, **54**, 622–654.

Box, G. E. P., and N. R. Draper (1987). *Empirical Model-Building and Response Surfaces*. Wiley & Sons, NY.

Constantine, G. M. (1989). Robust designs for serially correlated observations. *Biometrika*, **76**, 245–251.

Draper, N. R., and W. J. Beggs (1971). Errors in the Factor Levels and Experimental Design. *Ann. Math. Stat*, **42**, 46–58.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *JASA*, **69**, 383–393.

References -II

Hampel, F. R., E. Ronchetti, R. J. Rousseauw, and W. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, NY.

Huber, P. J. (1975). Robustness and Designs. In: *A Survey of Statistical Design and Linear Models*, ed. J. N. Srivastava. North-Holland.

Huber, P. J. (1981). *Robust Statistics*. Wiley, NY.

Jenkins, G. M., and J. Chanmugam (1962). The Estimation of Slope When the Errors are Autocorrelated. *JRSS*, B, **24**, 199–214.

Lal, K., V. K. Gupta, and L. Bhar (2001). Robustness of designed experiments against missing data. *J. of Applied Statistics*, Routledge, **28**, 63–79.

References -III

Newey, W. K., and K. D. West. A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, **55**, 703–708.

Smith, R. L., and K. D. S. Young (2001). *Linear Regression*. Manuscript, UNC.

Wiens, Douglas P. Homepage: http://www.stat.ualberta.edu/˜wiens/.

Wiens, D. P. (1991). Designs for approximately linear regression: Two optimality properties of uniform designs. *Stat. and Prob. Letters*, **12**, 217–221.

Wiens, D. P., and J. Zhou. Robust Designs Based on the Infinitesimal Approach. *JASA*, **92**, 1503–1511.

Wiens, D. P. (1998). Minimax Robust Designs and Weights for Approximately Specified Regression Models With Heteroskedastic Errors. *JASA*, **93**, 1440–1450.