Why would anyone ever use a linear model?

Saturday, January 12, 2013
1:14 PM

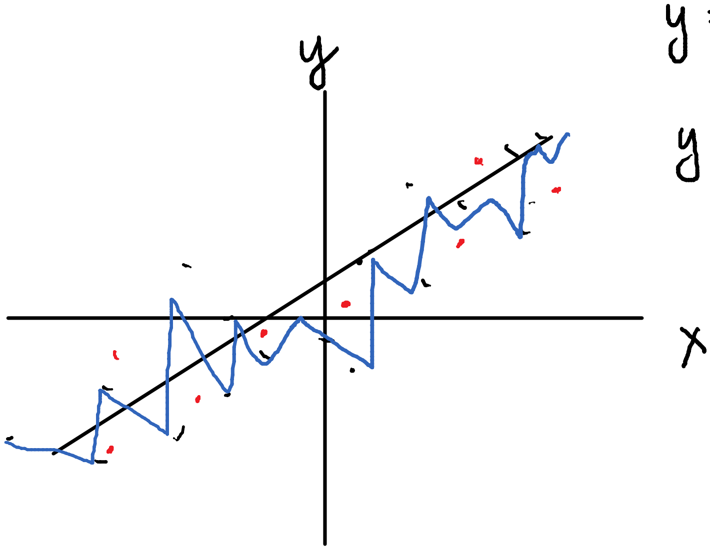
(a) test hypothesis (test theory)

(b) predict future observations

- policy initiative
- forecasting
- validation (out-of-sample)

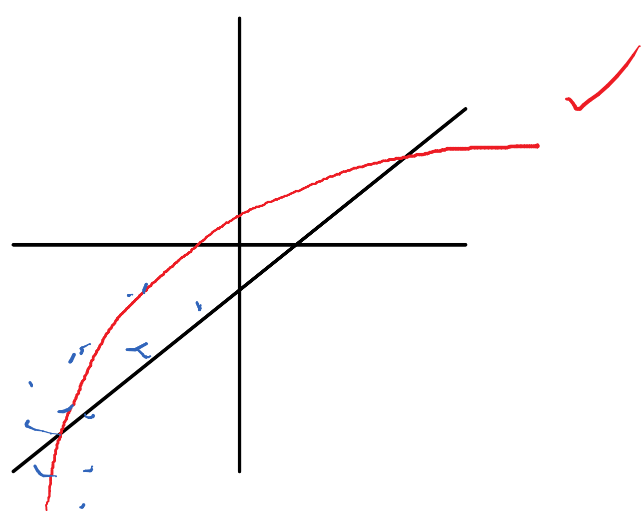
statistical model

inferences + predictions = data + model assumptions



$y = mx + b + \varepsilon$

$y = \beta_0 + \beta_1 \cdot x + \varepsilon$

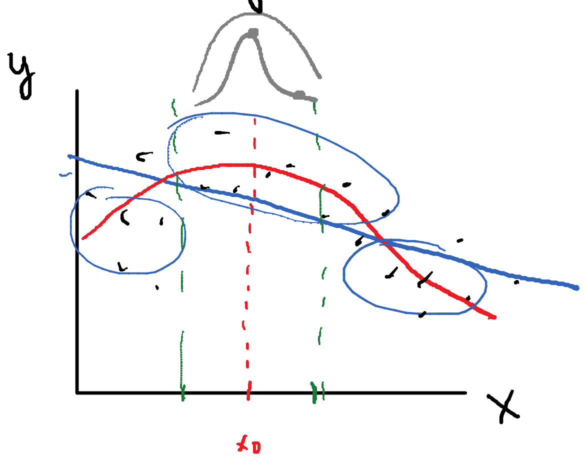


Low-assumption approaches

Saturday, January 12, 2013
1:20 PM

non-parametric models (no formal mathematical structures imposed on the data that are flexible only through a discrete set of tunable parametric)

Nadaraya-Watson



$$\hat{y}(x_0) = \frac{\sum_i y_i \cdot K(x_i - x_0)}{\sum_i K(x_i - x_0)}$$

when assumptions are good, they add value to an analysis.

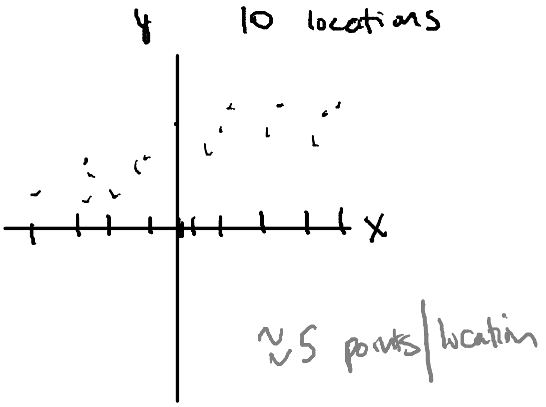
① extract better information ...

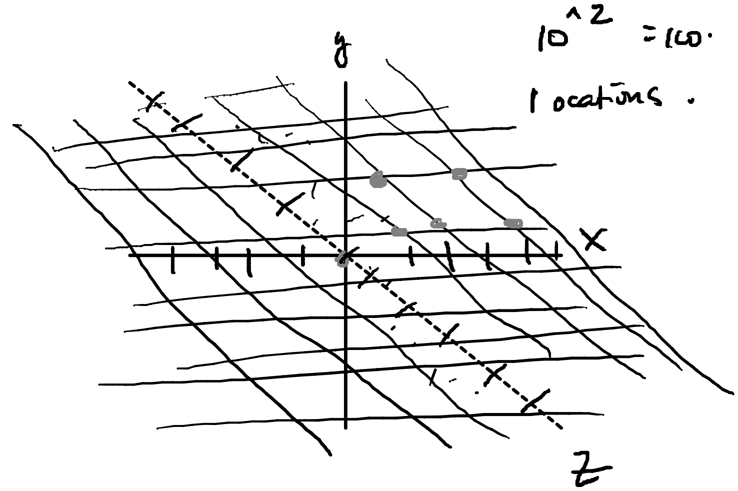
② ... with greater confidence.

these advantages grow as the # of predictor variables grows.

N = 50 $\frac{1}{2}$ point/location.

4 10 locations 4. $10^2 = 100$.





"Curse of dimensionality."

Getting to the canonical OLS model

Saturday, January 12, 2013
1:21 PM

things we don't know

$$\underset{n \times 1}{\underline{\underline{y}}} = \underset{n \times k}{\underline{\underline{X}}}\ \underset{k \times 1}{\underline{\underline{\beta}}} + \underset{n \times 1}{\underline{\underline{u}}}$$

things we know

n = # of observations

k = # of independent variables (includes the constant)

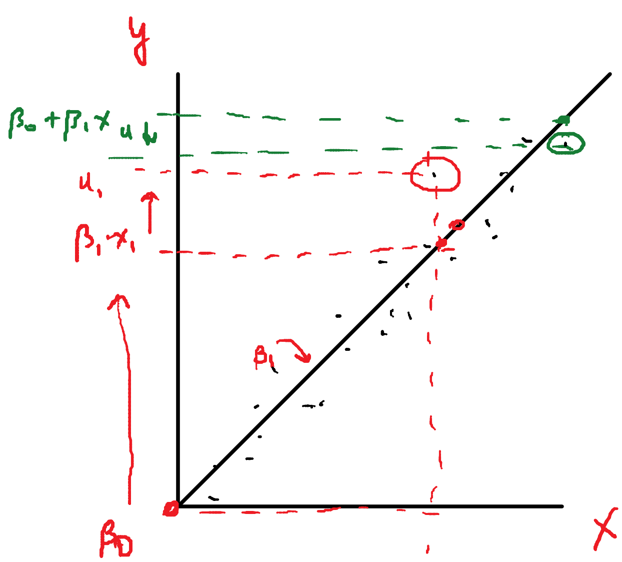
$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + u_1$$

$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + u_2$$

$$y_3 = \_\_\_\_\_\_$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots u_n$$



A first take on OLS: assume correct specification

Saturday, January 12, 2013
1:37 PM

$$\underset{k\times n}{X'}\underset{n\times 1}{y} = \underset{n\times k}{X'X}\beta + X'u$$

$$(X'X)^{-1}X'y = \cancel{(X'X)^{-1}X'X}\,\beta + (X'X)^{-1}X'u$$

$$(X'X)^{-1}X'y = \beta + \underbrace{(X'X)^{-1}X'u}_{=0}$$

$$\boxed{(X'X)^{-1}X'y = \hat{\beta}}$$

↳ coefficients of X on the noise term, u.

assume that X and u are uncorrelated, this coef $= 0$.

Assumptions

1. $\boxed{\text{correct specification}}$ → questionable?
2. X and u are uncorrelated.

The invertibility of X'X

Saturday, January 12, 2013
1:35 PM

$$\boxed{\underset{k \times k}{(\underset{k\times n}{X'}\ \underset{n\times k}{X})^{-1}}}$$

$\underset{n\times k}{X}$ be of full rank (column rank).

$$\boxed{\text{Theorem: } \operatorname{rank}(X'X) = \operatorname{rank}(X).}$$

Pf. : omitted.

Invertible Matrix Theorem:

① $(X'X)^{-1}$ exists.

② $(X'X)^{-1}$ has full rank.

X has rank k when all its k variables (columns) are linearly independent (not perfectly collinear).

Regression only works when independent variables are not linear functions of one another.

$$\begin{array}{cccc} \text{const} & = c1 & + c2 & + c3 \\ \end{array}$$
$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ & \vdots & & \end{bmatrix}$$

$X'X$ not invertible.

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$X'X$ not invertible.

$2\,inc + \frac{1}{2} tax = \boxed{SI}$ inc tax X

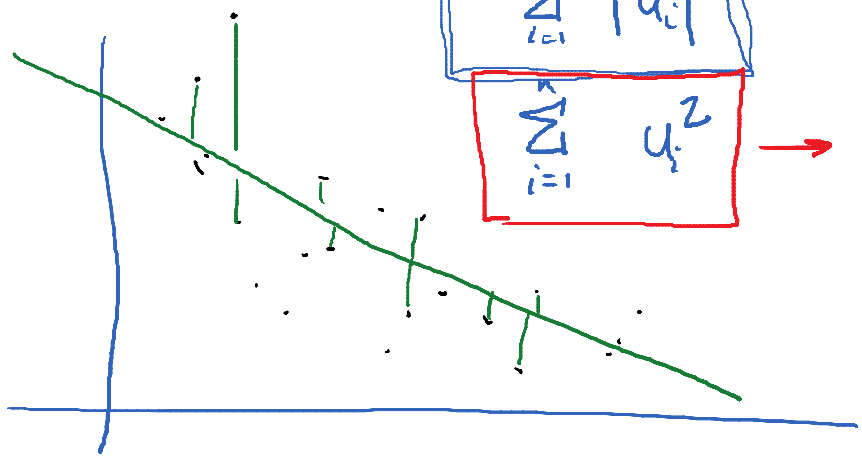
Optimization theory (from calculus)

Saturday, January 12, 2013
1:21 PM

- approximating a DGP w/ linear model
- pick β coefficients that minimize the magnitude of estimated errors, $\hat{u}$

$$\boxed{\sum_{i=1}^{n} |u_i|}$$

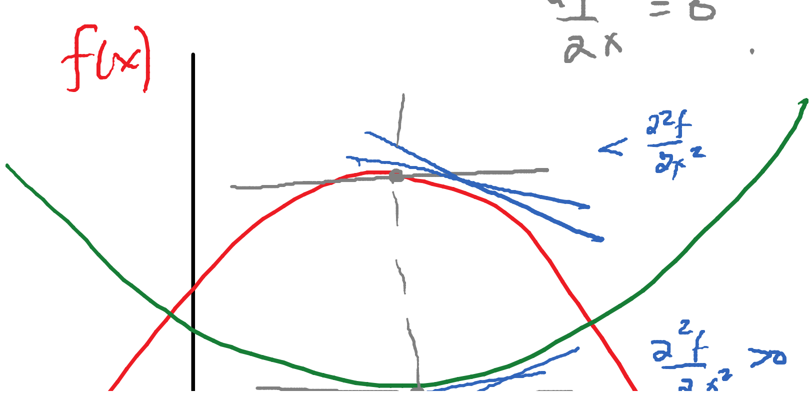
$$\boxed{\sum_{i=1}^{n} u_i^2} \rightarrow \text{SSE}$$



optimization

$f(x)$, pick x to minimize $f(x)$.

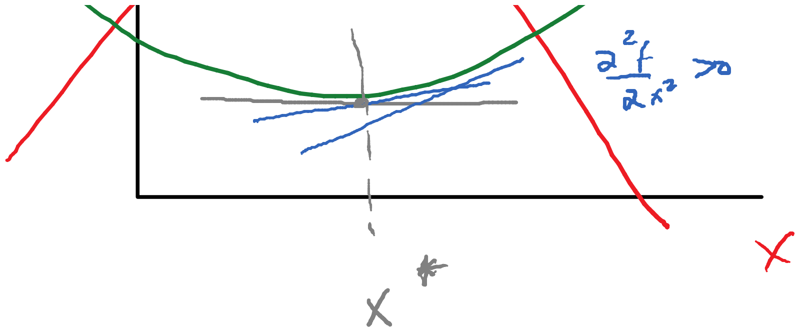
$$\frac{\partial f}{\partial x} = 0$$



$< \frac{\partial^2 f}{\partial x^2}$

$\frac{\partial^2 f}{\partial x^2} > 0$

? maximum or minimum?



minimum!

min $f(\beta_0, \beta_1, \dots, \beta_k)$ wrt $\beta_0, \beta_1, \dots, \beta_k$

$$\left.\begin{aligned} \frac{\partial f}{\partial \beta_0} &= 0 \\ \frac{\partial f}{\partial \beta_1} &= 0 \\ &\vdots \\ \frac{\partial f}{\partial \beta_k} &= 0 \end{aligned}\right\} \text{ Solve simultaneously.}$$

k

$$\underset{k}{H} = \overset{k}{\begin{bmatrix} \frac{\partial^2 f}{\partial \beta_0^2} & \boxed{\frac{\partial^2 f}{\partial \beta_0 \partial \beta_1}} & \frac{\partial^2 f}{\partial \beta_0 \partial \beta_2} & \cdots & \frac{\partial^2 f}{\partial \beta_0 \beta_k} \\ \boxed{\frac{\partial^2 f}{\partial \beta_1 \partial \beta_0}} & \frac{\partial^2 f}{\partial \beta_1^2} & \frac{\partial^2 f}{\partial \beta_1 \partial \beta_2} & \cdots & \frac{\partial^2 f}{\partial \beta_1 \partial \beta_k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial \beta_k \partial \beta_0} & \frac{\partial^2 f}{\partial \beta_k \partial \beta_1} & \frac{\partial^2 f}{\partial \beta_k \partial \beta_2} & \cdots & \frac{\partial^2 f}{\partial \beta_k^2} \end{bmatrix}}$$

Young's theorem: implies H is symmetric.

H must be positive definite

p.d. = a $k \times k$ matrix A is p.d. iff

$$x' A x > 0 \quad \forall x \in \mathbb{R}^k$$

Solving for $\hat{\beta}$ via optimization

Saturday, January 12, 2013
1:22 PM

min

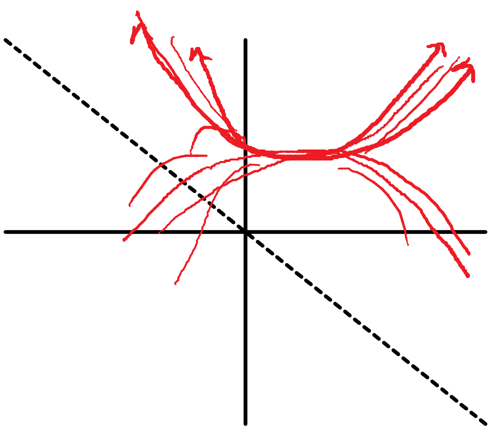
$$\frac{\partial f}{\partial \beta} = 0 \quad \forall \beta$$

$\underline{\underline{H}}$ positive definite.

max

$$\frac{\partial f}{\partial \beta} = 0 \quad \forall \beta$$

H negative definite.



$(A+B)' = A' + B'$

$$\min \sum_i \hat{u}_i^2 = \hat{u}'\hat{u}$$

$$= (y - x\hat{\beta})'(y - x\hat{\beta})$$

$[y'x\hat{\beta}]'$

$$= y'y - \underbrace{y'x\hat{\beta} - (x\hat{\beta})'y} - (x\hat{\beta})'x\hat{\beta}$$

$$= y'y - 2(X\beta)'y - (X\beta)'X\beta$$

Quadratic form: $x'Ax$

$$\frac{\partial x'Ax}{\partial x} = \begin{bmatrix} \frac{\partial x'Ax}{\partial x_1} \\ \frac{\partial x'Ax}{\partial x_2} \end{bmatrix} =$$

$$= \begin{bmatrix} 2ax_1 + bx_2 + cx_2 \\ 2dx_2 + bx_1 + cx_2 \end{bmatrix} = x'(A + A')$$

$$x = [x_1 \quad x_2] \qquad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$x'Ax \longrightarrow ax_1^2 + bx_1x_2 + cx_1x_2 + dx_2^2$$

$$\frac{\partial}{\partial x_1} = 2ax_1 + bx_2 + cx_2$$

$$\frac{\partial}{\partial x_2} = bx_1 + cx_1 + 2dx_2$$

if $\underline{\underline{A}}$ is symmetric, then

$$\frac{\partial x'Ax}{\partial x} = x'(A + A') = \underline{\underline{2x'A'}} = \underline{\underline{2Ax}}$$

back to problem...

$$\hat{u}'\hat{u} = y'y - 2(x\hat{\beta})'y + (x\hat{\beta})'x\hat{\beta}$$

$$\hat{u}'\hat{u} = y'y - 2(x\hat{\beta})'y + \boxed{\hat{\beta}'x'x\hat{\beta}}$$

$$- 2\hat{\beta}'x'y$$

$$\boxed{\frac{\partial \hat{u}'\hat{u}}{\partial \hat{\beta}} = -2x'y + 2\hat{\beta}'x'x}$$

$$0 = -2x'y + 2x'x\hat{\beta}$$

$$2x'y = 2x'x\hat{\beta}$$

$$(x'x)^{-1} x'y = (x'x)^{-1}x'x\hat{\beta}$$

$$(x'x)^{-1}x'y = \hat{\beta} \quad //$$

H? Theorem: if any $n \times k$ matrix X has full rank (k), it can be shown that any matrix $X'X$ is positive definite.

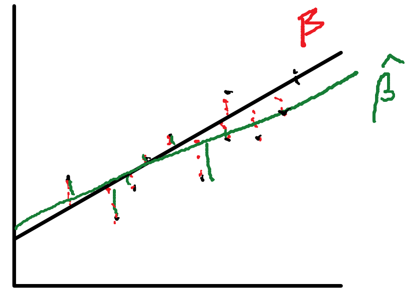
$$\frac{\partial^2 \hat{u}'\hat{u}}{\partial \hat{\beta}^2} = 2\underset{\sim}{(X'X)}$$

Expected value of the error term

Saturday, January 12, 2013
1:37 PM

$\hat{u}$ u

estimated actual error



$$\frac{\partial \hat{u}'\hat{u}}{\partial \hat{\beta}} = 0$$

$$\frac{\partial}{\partial \beta} \sum_{i=1}^{n} (y_i - x_i\hat{\beta})^2 = 0$$

$$2 \sum_i (y - x\hat{\beta}) \cdot (-x) = 0$$

$$-2x \sum_i (y - x\hat{\beta}) = 0$$

$$-2x \sum_i \hat{u}_i = 0$$

$$\boxed{\sum_i \hat{u}_i = 0}$$

$$\boxed{\mu_{\hat{u}} = \frac{1}{n} \sum_{i=1}^{n} \hat{u}_i = 0}$$

Another take on OLS: the CEF

Saturday, January 12, 2013
1:36 PM

Angrist and Pischke, pp. 28-40

$$E[y|x] = \int y \cdot f(y|x)\, dy$$

We just proved that $\hat{\beta}_{OLS}$ is the best linear predictor of x, in a least-squares sense.

$$\hat{\beta}_{OLS} = \arg\min_{\beta} E\Big[\underbrace{E[y|x]}_{CEF} - \underbrace{x\hat{\beta}}_{prediction}\Big]$$

Pf: $\hat{\beta}$ minimizes $(y - x\hat{\beta})^2$.

$$(y - x\hat{\beta})^2 = \left(y - E[y|x] + E[y|x] - x\hat{\beta}\right)^2$$

$$= \left(y - E[y|x]\right)^2 + \left(E[y|x] - x\hat{\beta}\right)^2 - \underline{2\left(y - E[y|x]\right)}\left(E[y|x] - x\hat{\beta}\right)$$

$$= \left(y - E[y|x]\right)^2 + \left(E[y|x] - x\hat{\beta}\right)^2 - \underline{2\varepsilon\left(E[y|x] - x\hat{\beta}\right)}$$

$$=\left(y-E[y|x]\right)+\underbrace{\left(E[y|x]-x\beta\right)}_{(\text{gap between CEF and } x\beta)^2}-2\varepsilon\left(E[y|x]-x\beta\right)$$

CEF decomposition property:

$$y = E[y|x] + \varepsilon$$

(i) $\underline{E[\varepsilon|x]=0}$ and (ii) ε is uncorrelated with any function of x.

(i): $$E[\varepsilon|x] = E\big[y - E[y|x]\,\big|\,x\big] =$$

$$E[y|x] - E\big[E[y|x]\,\big|\,x\big]$$

$$E[y|x] - E[y|x] = 0$$

(ii): for any function of x, $h(x)$:

$$E[h(x)\varepsilon] = E\big[E[h(x)\varepsilon|x]\big] =$$

$$= E\big[h(x)\,E[\varepsilon|x]\big]$$

$$E\left[\underline{h(x)} \cdot \underline{E[\varepsilon|x]} \right]$$

$= 0$ by (i).

Law of Iterated Expectations:

$E[E[a|b]] = E[a]$.

$$-2\underbrace{\varepsilon}_{\varepsilon}\overbrace{\left(E[y|x] - x\underline{\hat{\beta}}\right)}^{h(x)}$$

$$E\left[\left(y - E[y|x]\right)^2 + \underbrace{\left(E[y|x] - x\hat{\beta}\right)^2}_{\text{(gap between CEF and } x\beta)^2} - \cancel{2\varepsilon\left(E[y|x] - x\hat{\beta}\right)} \right]$$

$$E\left[\left(E[y|x] - x\hat{\beta}\right)^2 \right]$$

OLS minimizes this term.

$=$

$=$

$h(x)$

$\sum_{i} \left(E[y|x] - X\hat{\beta} \right)$