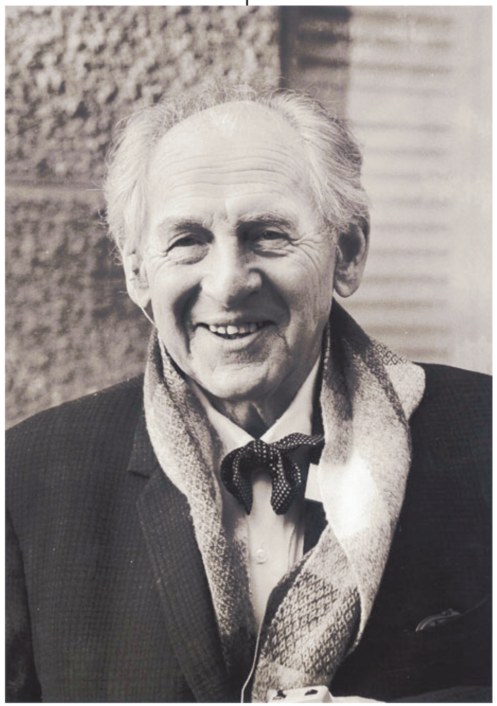
4. Economies of Scale

„Wo immer etwas falsch ist, ist es zu groß“

Leopold Kohr
(1909 -1994)



Definition:

By how much increases a firms product, when simultaneously all inputs are increased by the same factor k ?

E.g.

$$f(kx_1, kx_2) > k\, f(x_1, x_2) \qquad \text{for } k>1.$$

The cost function

shows the minimum costs for which a production level of Y can be reached. Usually the long term average cost curve

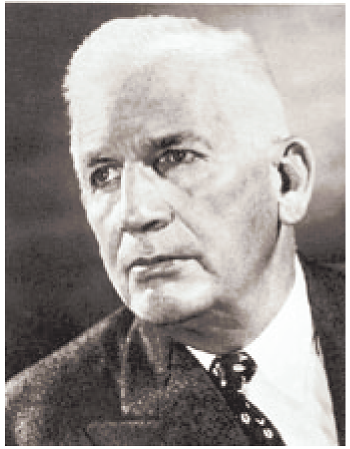
$$AC(Y) = \frac{c(p_1, p_2, Y)}{Y}$$

exhibits a minimum, which indicates the optimal amount of production.

The Cobb-Douglas production function: (p65)

usually in the form $Y = c \cdot (K^a \cdot L^b)$.

Johan Gustaf Knut Wicksell
(1851 - 1926)

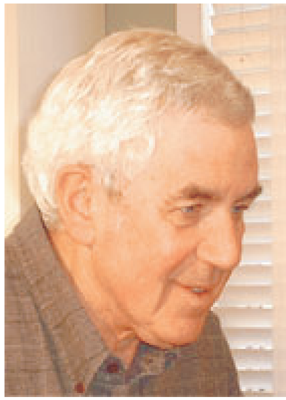


Paul Howard Douglas
(1892 - 1976)

C-D production function for Electricity Supply Enterprises

$$Y = a_0 x_l^{a_l} x_k^{a_k} x_f^{a_f} u$$

Nerlove, Marc. “Returns to Scale in Electricity Supply.” In C.Christ, ed., Measurement in Economics:Studies in Mathemetical Economics and Econometrics in Memory of Yehuda Grunfeld. Stanford, CA: Stanford University Press, 1963



Marc Nerlove (1933-)

Implies the following cost function:

$$\log c = \log a_0 + \frac{1}{r}\log Y + \frac{a_l}{r}\log p_l + \frac{a_k}{r}\log p_k + \frac{a_f}{r}\log p_f + \frac{1}{r}\log u$$

where

$$r = a_l + a_k + a_f$$

measures the amount of returns to scale.

first least squares estimation

Dependent Variable: LOG(TC)

Method: Least Squares

Sample: 1 145

Included observations: 145

$$\hat{r} = 1/0.720394 = 1.388$$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.526503	1.774367	-1.987471	0.0488
LOG(Y)	0.720394	0.017466	41.24448	0.0000
LOG(PL)	0.436341	0.291048	1.499209	0.1361
LOG(PK)	-0.219888	0.339429	-0.647819	0.5182
LOG(PF)	0.426517	0.100369	4.249483	0.0000

R-squared	0.925955	Mean dependent var	1.724663
Adjusted R-squared	0.923840	S.D. dependent var	1.421723
S.E. of regression	0.392356	Akaike info criterion	1.000578
Sum squared resid	21.55201	Schwarz criterion	1.103224
Log likelihood	-67.54189	F-statistic	437.6863
Durbin-Watson stat	1.013062	Prob(F-statistic)	0.000000

The Likelihood function (pp172)

Assumption: Gaussian errors, $u \sim N(0, \sigma^2 I)$

Density for variables $(X_1,Y_1), \ldots,(X_n,Y_n)$

$$p(y;X,\beta,\sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right)$$

Likelihood-function

$$L(\beta,\sigma^2;y,X) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right)$$

Log-likelihood-function

$$\log L = \ell(\beta,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{S(\beta)}{2\sigma^2}$$

Sir Ronald Aylmer Fisher 1890-1962

Maximum Likelihood Estimator

Derivatives:

$$\frac{\partial \ell}{\partial \beta} = -\frac{\partial S(\beta)}{\partial \beta}\frac{1}{2\sigma^2}$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{S(\beta)}{2\sigma^4}$$

Daniel Bernoulli

(1700 - 1782)

Likelihood-equations: set derivatives zero

Maximum-Likelihood (ML)-estimator:

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

$$\tilde{\sigma}^2 = \frac{1}{n}(y - X\tilde{\beta})'(y - X\tilde{\beta})$$

Note: ML-estimates for β are identical to OLS-estimates (if noises are Gaussian)

Diagnostics based on the likelihood (p61)

Usually functions of *SSE*

- Logarithmic Likelihood

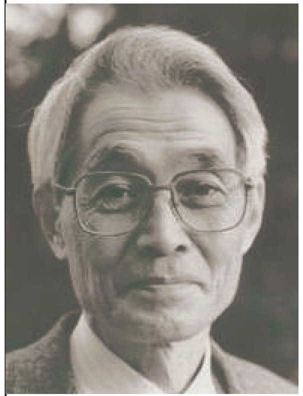
$$\ell(\hat{\beta}) = -\frac{n}{2}\left(1 + \log(2\pi) + \log\hat{\sigma}_e^2\right)$$

- Akaike's Informationcriterion

$$AIC = -\frac{2\ell(\hat{\beta})}{n} + \frac{2k}{n}$$

- Schwarz' Bayesian Informationcriterion

$$BIC = -\frac{2\ell(\hat{\beta})}{n} + \frac{k\log n}{n}$$



Hirotugu Akaike (1927 -)

Gideon E. Schwarz (1933 -2007)

Test against the value 1, confidence intervals (p25)

New t-statistics from $(\hat{\beta}_i - 1) / \hat{\sigma}_\varepsilon \sqrt{(X'X)^{-1}_{ii}}$.

Generally from a confidence interval:

$$[\hat{\beta}_i \pm \hat{\sigma}_\varepsilon \sqrt{(X'X)^{-1}_{ii}}\, t_{1-a/2}] = [0.686, 0.755].$$

The Delta-method:

Standard error for $\hat{r} = f(\hat{\beta}_1) = 1/\hat{\beta}_1$:

Taylorexpansion around the estimated expected value of the desired parameter, i.e. $f(\beta) \approx f(\hat{\beta}) + (\beta - \hat{\beta})\,\partial f/\partial\beta$ and thus also for a variance-covariance matrix $\mathrm{Kov}(f,f') \approx \Delta\mathrm{Kov}(\beta,\beta')\Delta'$, where $\Delta_{ij} = \partial f_i/\partial\beta_j$.

From the diagonal of $\mathrm{K\hat{o}v}(f,f')$ we yield $\mathrm{V\hat{a}r}(\hat{r}) = (\partial f/\partial\hat{\beta}_1)^2\,\mathrm{V\hat{a}r}(\hat{\beta}_1) = \mathrm{V\hat{a}r}(\hat{\beta}_1)/\hat{\beta}_1^4$ and a confidence interval of $[1.322, 1.454]$ for r.

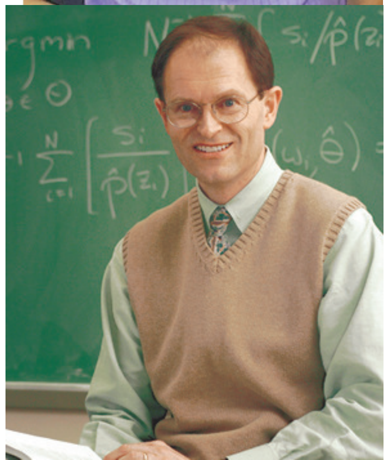
The trick of Papke and Wooldridge:

One estimates the parameters untransformed and evaluates the gradient at this value. Then the regressors are transformed such that

$\tilde{x}_i = [x_i - (\hat{\beta}_i / \hat{\beta}_k)x_k]$ for $i \neq k$ and $\tilde{x}_k = x_k / \hat{\beta}_k$

and performs a regression on the regressand. The desired standard error can then be simply read of the result table in the row for $\tilde{x}_k$. Since in the present example we have $\hat{\beta} = (0, 1/0.72, 0, 0, 0)$, one just needs to multiply $\log Y$ accordingly:

Leslie E. Papke **Jeffrey M. Wooldridge**



Dependent Variable: LOG(TC)

Method: Least Squares

Sample: 1 145

Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.526503	1.774367	-1.987471	0.0488
LOG(Y)*0.720394	1.000000	0.024246	41.24448	0.0000
LOG(PK)	-0.219888	0.339429	-0.647819	0.5182
LOG(PL)	0.436341	0.291048	1.499209	0.1361
LOG(PF)	0.426517	0.100369	4.249483	0.0000

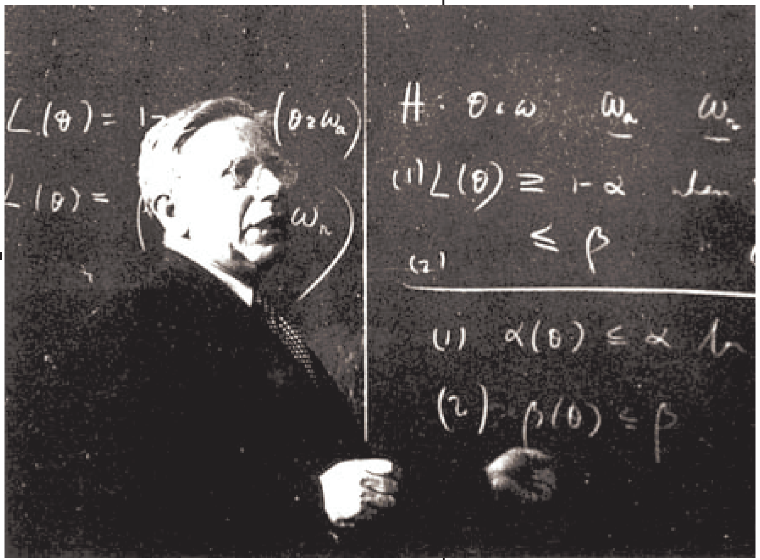
R-squared	0.925955	Mean dependent var	1.724663
Adjusted R-squared	0.923840	S.D. dependent var	1.421723
S.E. of regression	0.392356	Akaike info criterion	1.000578
Sum squared resid	21.55201	Schwarz criterion	1.103224
Log likelihood	-67.54189	F-statistic	437.6863
Durbin-Watson stat	1.013062	Prob(F-statistic)	0.000000

4.2 Constraints, the Wald-test (p31)

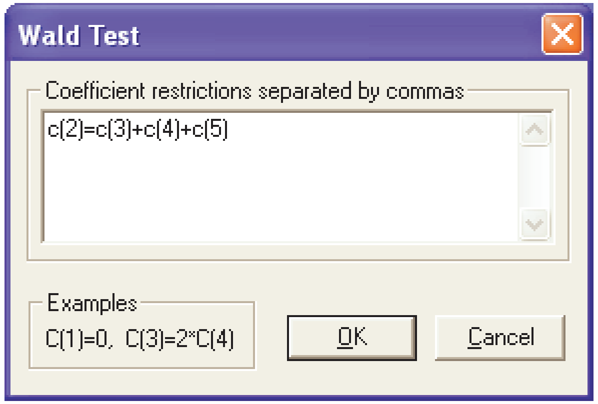
Linear restrictions on the parameters, i.e. they can be presented in the form $H\beta=h$, where H denotes a $g\times m$ matrix with the restriction coefficients for g constraints.

Test statistics:

$T_W = d'(H(X'X)^{-1}H')^{-1}/\hat{\sigma}^2 \sim \chi^2_g$



Abraham Wald (1902-1950)



Wald Test:

Equation: EQ01

Test Statistic	Value	df	Probability
F-statistic	0.026985	(1, 140)	0.8698
Chi-square	0.026985	1	0.8695

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2) - C(3) - C(4) - C(5)	0.077424	0.471323

Restrictions are linear in coefficients.

$$W \geq LR \geq LM$$

Nerloves Modell A:

$$\log(c/p_f) = \log a_0 + \frac{1}{r}\log Y + \frac{a_l}{r}\log\left(p_l/p_f\right) + \frac{a_k}{r}\log\left(p_k/p_f\right) + \frac{1}{r}\log u.$$

Dependent Variable: LOG(TC/PF)

Method: Least Squares

Sample: 1 145

Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.690789	0.884871	-5.301098	0.0000
LOG(Y)	0.720688	0.017436	41.33398	0.0000
LOG(PK/PF)	-0.007381	0.190736	-0.038698	0.9692
LOG(PL/PF)	0.592910	0.204572	2.898291	0.0044

R-squared	0.931585	Mean dependent var	-1.484195
Adjusted R-squared	0.930129	S.D. dependent var	1.482087
S.E. of regression	0.391762	Akaike info criterion	0.990874
Sum squared resid	21.64032	Schwarz criterion	1.072991
Log likelihood	-67.83836	F-statistic	639.9802
Durbin-Watson stat	1.015369	Prob(F-statistic)	0.000000

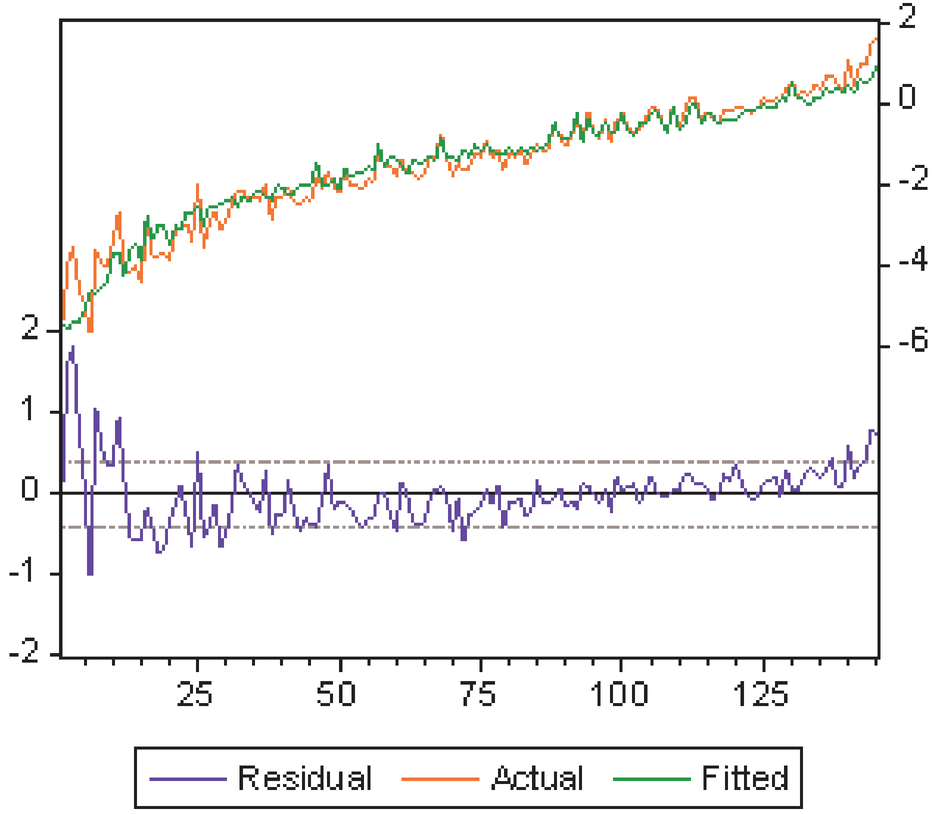
$$\hat{\beta}_S = H_S^{-1}(h - H_{\acute{S}}\hat{\beta}_{\acute{S}}),$$

which in the case of only one constraint can be fulfilled by entering it into the equation, and here yields $1+0.007-0.593 = 0.414$.

$$\mathrm{Var}(\beta_S) = H_S^{-1} H_{\acute{S}}\, \mathrm{Var}(\beta_{\acute{S}})\, H_{\acute{S}}' H_S^{-1\,\prime}$$

$$=0.099.$$

4.3 Nonlinearities, Feasible Generalized Least Squares



Dependent Variable: LOG(TC/Y/PF)

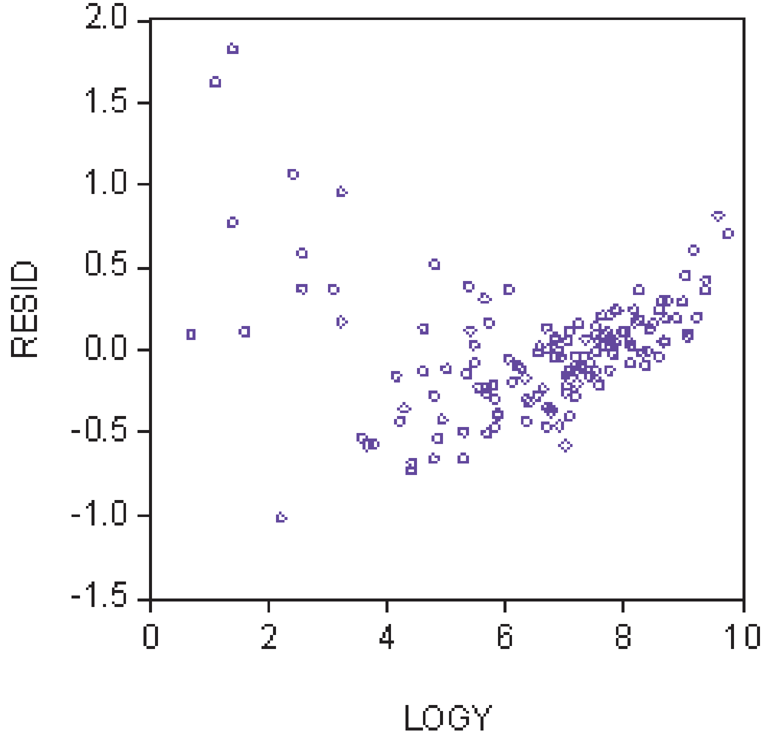
Method: Least Squares

Sample: 1 145

Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.690789	0.884871	-5.301098	0.0000
LOG(Y)	-0.279312	0.017436	-16.01956	0.0000
LOG(PK/PF)	-0.007381	0.190736	-0.038698	0.9692
LOG(PL/PF)	0.592910	0.204572	2.898291	0.0044

R-squared	0.653553	Mean dependent var	-8.040846
Adjusted R-squared	0.646182	S.D. dependent var	0.658616
S.E. of regression	0.391762	Akaike info criterion	0.990874
Sum squared resid	21.64032	Schwarz criterion	1.072991
Log likelihood	-67.83836	F-statistic	88.66281
Durbin-Watson stat	1.015369	Prob(F-statistic)	0.000000



$$\log(c/p_f) = \log a_0 + \alpha \log Y + \tfrac{1}{2}\gamma \log^2 Y + \frac{a_l}{r}\log\left(p_l / p_f\right) + \frac{a_k}{r}\log\left(p_k / p_f\right) + \varepsilon$$

Dependent Variable: LOG(TC/PF)

Method: Least Squares

Sample: 1 145

Included observations: 145

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.764649	0.701727	-5.364837	0.0000
LOG(Y)	0.152547	0.061860	2.465992	0.0149
LOG(PK/PF)	0.074166	0.150016	0.494390	0.6218
LOG(PL/PF)	0.480586	0.161072	2.983663	0.0034
LOG(Y)^2/2	0.101028	0.010727	9.417823	0.0000

R-squared	0.958118	Mean dependent var	-1.484195
Adjusted R-squared	0.956922	S.D. dependent var	1.482087
S.E. of regression	0.307612	Akaike info criterion	0.513919
Sum squared resid	13.24751	Schwarz criterion	0.616565
Log likelihood	-32.25909	F-statistic	800.6873
Durbin-Watson stat	1.665259	Prob(F-statistic)	0.000000

A feasible GLS estimator (p88)

Hayashi, 2000 suggests the specification $\mathrm{Var}(\varepsilon) \propto \lambda_0 + \lambda_1 Y^{-1}$.

Simple, two-stage procedure:

- $\hat{V}$ from the residuals as above,
- then $\hat{\beta}_{FGLS} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y$.

Separating the data

in 5 groups of 29 observations each.

Employ dummy-variables

q1 – q5

Programming with Eviews.

Dependent Variable: LOG(TC/PL)
Method: Least Squares
Date: 04/17/08 Time: 11:42
Sample: 1 145
Included observations: 145

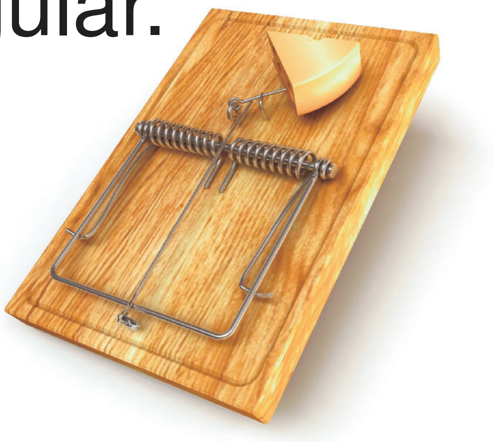
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Q1	-3.343348	1.652646	-2.023027	0.0452
Q2	-6.488974	1.825601	-3.554432	0.0005
Q3	-7.332942	2.671684	-2.744689	0.0069
Q4	-6.546049	3.025269	-2.163791	0.0324
Q5	-6.714258	2.181091	-3.078394	0.0026
Q1*LOG(Y)	0.400290	0.044370	9.021601	0.0000
Q1*LOG(PF/PL)	0.466182	0.169372	2.752424	0.0068
Q1*LOG(PK/PL)	-0.081356	0.371103	-0.219227	0.8268
Q2*LOG(Y)	0.658151	0.150263	4.380006	0.0000
Q2*LOG(PF/PL)	0.528265	0.189693	2.784841	0.0062
Q2*LOG(PK/PL)	0.377936	0.357290	1.057785	0.2922
Q3*LOG(Y)	0.938279	0.313146	2.996304	0.0033
Q3*LOG(PF/PL)	0.347734	0.246322	1.411703	0.1605
Q3*LOG(PK/PL)	0.250008	0.295837	0.845087	0.3997
Q4*LOG(Y)	0.912044	0.279177	3.266899	0.0014
Q4*LOG(PF/PL)	0.399690	0.158800	2.516935	0.0131
Q4*LOG(PK/PL)	0.093352	0.426200	0.219033	0.8270
Q5*LOG(Y)	1.044390	0.135455	7.710217	0.0000
Q5*LOG(PF/PL)	0.686848	0.277439	2.475674	0.0146
Q5*LOG(PK/PL)	-0.289436	0.364519	-0.794021	0.4287

R-squared	0.957322	Mean dependent var	1.052996
Adjusted R-squared	0.950835	S.D. dependent var	1.412561
S.E. of regression	0.313208	Akaike info criterion	0.643546
Sum squared resid	12.26243	Schwarz criterion	1.054130
Log likelihood	-26.65706	Durbin-Watson stat	1.774708

4.4 Multicollinearity (pp43)

Perfekt collinearity: $X'X$ is singular.

(Dummy variable trap)



Almost collinearity: some regressors are strongly linearly dependent: leads to imprecise estimation.

(but OLS remains the BLUE!)

Effects

on the variances of the coefficients:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{(1-R_j^2)\sum_{i=1}^{T}\left(x_{ij}-\overline{x}_j\right)}$$

thus depends upon

- variance of noise
- variance of regressors
- number of observations
- collinearity of regressors

Micronumerosity

= not enough data!

Statistically everything fine.

No problem for predictions!

Arthur S. Goldberger

Indicators for multicollinearity

- high R^2 and little signifikances
- high correlations among regressors
- not robust against data changes
- Condition number of the design matrix:

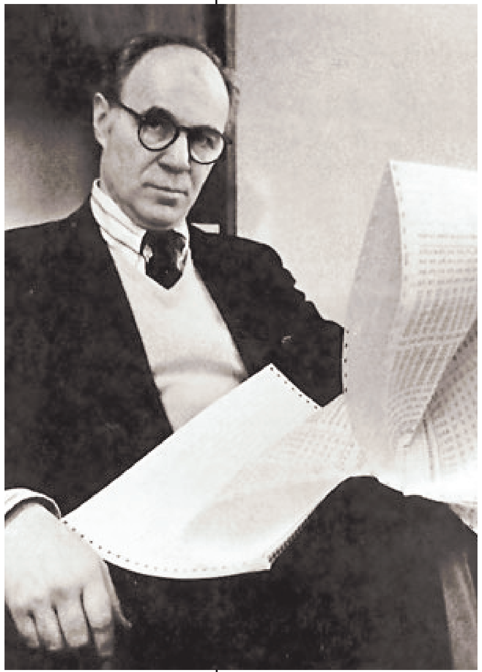
$$CI = \sqrt{[\max \mathrm{eig}(X'X) / \min \mathrm{eig}(X'X)]}$$

Supporting regressions (p43)

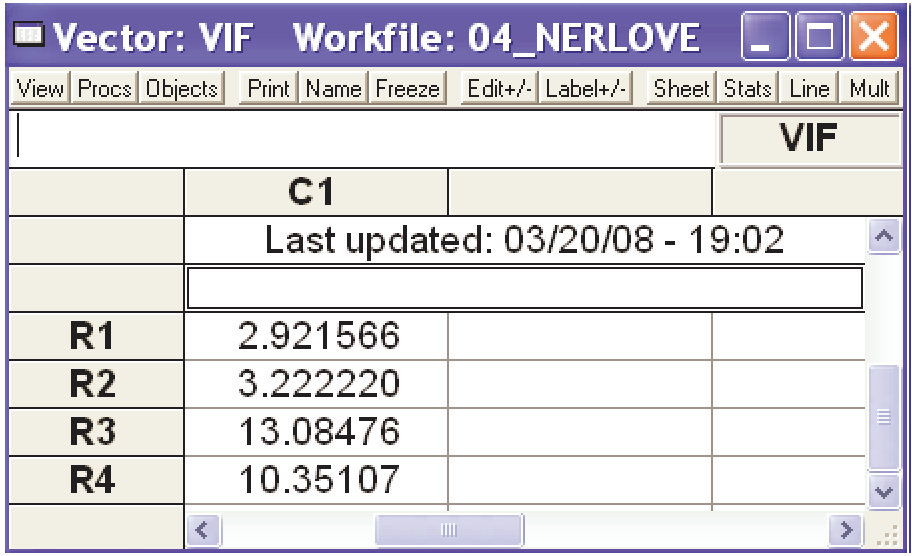
Regressions of single regressors on all the rest, leads to R_j^2.

Klein's rule of thumb: R_j^2 should be all less than R^2.

This leads to variance inflation factor: $\mathrm{VIF} = (1-R_j^2)^{-1}$



Lawrence Robert Klein (1920 -)



	C1
	Last updated: 03/20/08 - 19:02
R1	2.921566
R2	3.222220
R3	13.08476
R4	10.35107

Remedies for multicollinearity

- **Do nothing!**

- **More data!**
- Formalising of relationships between regressors (equation systems)
- Spezification of relationships between parameters (restrictions and particular lag structures)
- Incorporate estimates from other studies or prior information (Bayesian estimation)
- Form principal components
- (remove variabls – leads to bias)
- Shrinkage-estimators

Ridge Regression

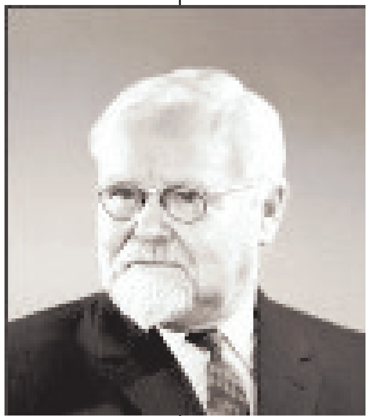
Estimate a regularised system:

$$\hat{\beta}_R = \left(X^T X + kI\right)^{-1} X^T Y$$

(Hoerl & Kennard, 1970)

Is equivalent to the constraint:

$$\beta'\beta = c^2$$



Andrei N. Tikhonov
(1906 – 1993)

For c^2=1.5

Dependent Variable: LOG(TC/PF)
Method: Least Squares
Date: 03/20/08 Time: 19:02
Sample: 1 145
Included observations: 145
Convergence achieved after 94 iterations
LOG(TC/PF) =-@SQRT(1.5-C(1)^2-C(2)^2-C(3)^2-C(4)^2)+C(4)* LOG(Y)
+C(1)*LOG(PK/PF)+C(2)*LOG(PL/PF)+C(3)*LOG(Y)^2/2

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.484111	0.137611	-3.517969	0.0006
C(2)	1.032177	0.153576	6.720935	0.0000
C(3)	0.143087	0.009481	15.09227	0.0000
C(4)	-0.126783	0.048722	-2.602153	0.0103

R-squared	0.946138	Mean dependent var	-1.484195
Adjusted R-squared	0.944992	S.D. dependent var	1.482087
S.E. of regression	0.347604	Akaike info criterion	0.751693
Sum squared resid	17.03682	Schwarz criterion	0.833810
Log likelihood	-50.49773	Durbin-Watson stat	1.501916

The Lasso

is equivalent to the constraint:

$$\sum |\beta_i| \leq c$$

special case of least angle regression (LARS)

removes unnecessary regressors

Robert Tibshirani

Lasso c=3.6

Dependent Variable: LOG(TC/PF)
Method: Least Squares
Date: 03/20/08 Time: 15:18
Sample: 1 145
Included observations: 145
Convergence achieved after 4 iterations
LOG(TC/PF) =-(3.6-@ABS(C(1))-@ABS(C(2))-@ABS(C(3))-@ABS(C(4)))
+C(1)* LOG(Y)+C(2)*LOG(PK/PF)+C(3)*LOG(PL/PF)+C(4)*LOG(Y)^2/2

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.100061	0.055717	1.795878	0.0747
C(2)	-0.154582	0.088768	-1.741419	0.0838
C(3)	0.739026	0.085006	8.693801	0.0000
C(4)	0.108797	0.009991	10.88994	0.0000

R-squared	0.957058	Mean dependent var	-1.484195
Adjusted R-squared	0.956144	S.D. dependent var	1.482087
S.E. of regression	0.310375	Akaike info criterion	0.525124
Sum squared resid	13.58286	Schwarz criterion	0.607241
Log likelihood	-34.07153	Durbin-Watson stat	1.682427

EViews Standard-Output

Dependent Variable: LOG_M/V0
Method: Least Squares
Date: 01/24/08 Time: 13:54
Sample: 1 36
Included observations: 36

Variable	Coefficient	Std. Error	t-Statistic	Prob.
1/V0	0.021083	0.193736	0.108825	0.9141
LOG_Y/V0	0.882881	0.048982	18.02473	0.0000
RN/V0	-11.21374	3.285429	-3.413172	0.0019
GY/V0	-0.212149	0.291022	-0.728979	0.4717
S/V0	0.205901	0.027388	7.517951	0.0000
W/V0	0.013782	0.005821	2.367800	0.0245

R-squared	0.996072	Mean dependent var	7.333740
Adjusted R-squared	0.995417	S.D. dependent var	1.530226
S.E. of regression	0.103593	Akaike info criterion	-1.545673
Sum squared resid	0.321948	Schwarz criterion	-1.281753
Log likelihood	33.82211	Durbin-Watson stat	1.503403

Adjusted coefficient of determination (p23)

- Adding a regressor to a model yields:
  - R^2 is increased
  - Increase of R^2 not necessarily means that the new regressor is relevant!

- Adjusted coefficient of determination:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k}\frac{RSS}{TSS}$$

- Suitable for the comparison of models

$$\bar{R}^2 < R^2$$

- For large n we have $(n-1)/(n-k) \approx 1$

The Durbin-Watson-Test (1950) (pp110)

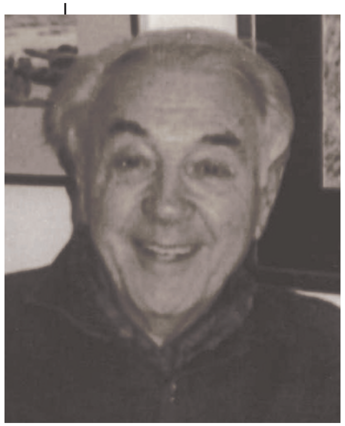
Null hypothesis: no autocorrelation of first order

Test statistic:

$$d = \frac{\sum_{t=2}^{n} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{n} \hat{\varepsilon}_t^2} \cong 2(1 - \hat{\rho}_1)$$

- For positive autocorrelation d lies in the interval (0,2)
- for negative in the interval (2,4);
- if d is close to 2, there is no significant autocorrelation in the noise;
- for values of d close to 0 or 4 the errors are highly correlated.

James Durbin (1923 –)



Geoffrey Stuart Watson (1921 –1998)

Durbin-Watson-Test, critical values

- the citical values for d depend upon matrix X.
- Thus D & W have lower (d_L) and upper bounds (d_U) for the critical values
  - $d < d_L$: H_0 is rejected
  - $d > d_U$: H_0 is not rejected
  - $d_L < d < d_U$: no decision

Critical bounds for $\alpha = 0.05$:

n	k=2		k=3		k=10	
	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.17	3.22
20	1.20	1.41	1.10	1.54	0.42	2.70
100	1.65	1.69	1.63	1.71	1.48	1.87

Ramsey‘s (1969) RESET-Test (p66)

RESET: *Regression Equation Specification Error*

For checking of the functional form of the model.

Tests H_0: $\gamma = 0$ für $Y = X\beta + Z\gamma + v$ with $z_t' = (\hat{Y}_t^2, \hat{Y}_t^3, \ldots)$

If H_0 is true, the functional form (linearity) of the response is correct.

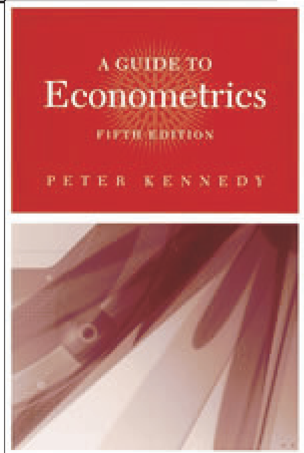
Tests:

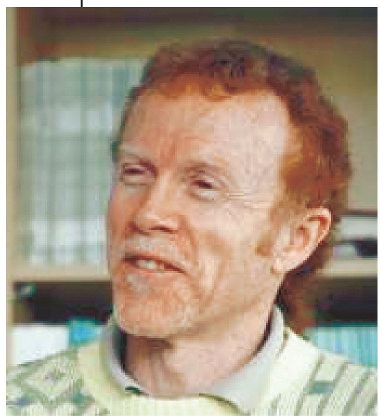
- *t*-Test, if $g = 1$
- *F*-Test
- asymptotical Chi-square-test (test statistic gF)

James Bernard Ramsey

The Ten Commandments of Applied Econometrics

- ***Thou shalt use common sense and economic theory.***
- ***Thou shalt ask the right questions.***
- ***Thou shalt know the context.***
- ***Thou shalt inspect the data.***
- ***Thou shalt not worship complexity.***
- ***Thou shalt look long and hard at thy results.***
- ***Thou shalt beware the costs of data mining.***
- ***Thou shalt be willing to compromise.***
- ***Thou shalt not confuse significance with substance.***
- ***Thou shalt confess in the presence of sensitivity.***





Peter Kennedy