Ch06. Multiple Regression Analysis: Further Issues

Ping Yu

HKU Business School
The University of Hong Kong

Effects of Data Scaling on OLS Statistics

Section 2.4a: SLR

- The fitted regression line in the example of CEO salary and return on equity is

$$\begin{aligned} \widehat{salary} &= 963.191 + 18.501 roe, \\ n &= 209, R^2 = .0132 \end{aligned}$$

- How will the intercept and slope estimates change when the units of measurement of the dependent and independent variables changes?
- Suppose the salary is measured in dollars rather than thousands of dollars, the intercept should be $963,191$ and the slope should be $18,501$. (why?)
- Solution: convert a new problem to an old problem whose solution is known.

A Brute-Force Solution

- Suppose $y_i^* = w_1 y_i$ and $x_i^* = w_2 x_i$.
- Then
$$\widehat{\beta}_1^* = \frac{\sum_{i=1}^n \left(x_i^* - \overline{x}^*\right) y_i^*}{\sum_{i=1}^n \left(x_i^* - \overline{x}^*\right)^2} = \frac{w_1 w_2 \sum_{i=1}^n \left(x_i - \overline{x}\right) y_i}{w_2^2 \sum_{i=1}^n \left(x_i - \overline{x}\right)^2} = \frac{w_1}{w_2}\widehat{\beta}_1,$$
and
$$\widehat{\beta}_0^* = \overline{y}^* - \overline{x}^* \widehat{\beta}_1^* = w_1 \overline{y} - w_2 \overline{x} \frac{w_1}{w_2}\widehat{\beta}_1 = w_1 \left(\overline{y} - \overline{x}\widehat{\beta}_1\right) = w_1 \widehat{\beta}_0,$$
where $\overline{x}^* = w_2 \overline{x}$ and $\overline{y}^* = w_1 \overline{y}$.
- But when the number of regressors is large, the formula of $\widehat{\beta}$ is complicated.
- The textbook derives the relationship between $\widehat{\beta}^*$ and $\widehat{\beta}$ by FOCs, while we employ the objective functions.

Relationship Between Objective Functions

- Recall that
$$\left(\widehat{\beta}_0, \widehat{\beta}_1\right) = \arg\min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2, \tag{1}$$
where arg means arguments.
- Now, for the rescaled data,
$$\begin{aligned} & \min_{\beta_0^*,\beta_1^*} \sum_{i=1}^{n} (y_i^* - \beta_0^* - \beta_1^* x_i^*)^2 \\ = & \min_{\beta_0^*,\beta_1^*} \sum_{i=1}^{n} (w_1 y_i - \beta_0^* - \beta_1^* w_2 x_i)^2 \\ = & \min_{\beta_0^*,\beta_1^*} w_1^2 \sum_{i=1}^{n} \left(y_i - \frac{\beta_0^*}{w_1} - \beta_1^* \frac{w_2}{w_1} x_i\right)^2, \end{aligned}$$
where note that since (β_0^*, β_1^*) can be freely chosen, $\left(\frac{\beta_0^*}{w_1}, \beta_1^* \frac{w_2}{w_1}\right)$ can also be freely chosen although (w_1, w_2) are fixed, just as (β_0, β_1) in (1).

Relationship Between $\widehat{\beta}$'s and $\widehat{\sigma}^2$'s

- So (why?)

$$
\begin{aligned}
\widehat{\beta}_0 &= \frac{\widehat{\beta}_0^*}{w_1} \text{ and } \widehat{\beta}_1 = \widehat{\beta}_1^* \frac{w_2}{w_1} \\
&\Longrightarrow \widehat{\beta}_0^* = w_1 \widehat{\beta}_0 \text{ and } \widehat{\beta}_1^* = \frac{w_1}{w_2} \widehat{\beta}_1
\end{aligned}
$$

- Also,

$$
\begin{aligned}
\widehat{\sigma}^{*2} &= \frac{1}{n-2} \sum_{i=1}^{n} \widehat{u}_i^{*2} = \frac{1}{n-2} \sum_{i=1}^{n} \left(y_i^* - \widehat{\beta}_0^* - \widehat{\beta}_1^* x_i^* \right)^2 \\
&= \frac{w_1^2}{n-2} \sum_{i=1}^{n} \left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i \right)^2 \\
&= \frac{w_1^2}{n-2} \sum_{i=1}^{n} \widehat{u}_i^2 \\
&= w_1^2 \widehat{\sigma}^2.
\end{aligned}
$$

- Or, the standard error of the regression (SER) $\widehat{\sigma}^* = w_1 \widehat{\sigma}$.

Relationship Between R^2's and Standard Errors

- It turns out that $R^{*2} = R^2$:

$$\begin{aligned} R^{*2} &= 1 - \frac{SSR^*}{SST^*} \\ &= 1 - \frac{\sum_{i=1}^n \left(y_i^* - \widehat{\beta}_0^* - \widehat{\beta}_1^* x_i^*\right)^2}{\sum_{i=1}^n \left(y_i^* - \overline{y}^*\right)^2} \\ &= 1 - \frac{w_1^2 \sum_{i=1}^n \widehat{u}_i^2}{w_1^2 \sum_{i=1}^n \left(y_i - \overline{y}\right)^2} \\ &= 1 - \frac{SSR}{SST} = R^2 \end{aligned}$$

- Standard Errors:

$$\begin{aligned} \widehat{\beta}_0^* &= w_1 \widehat{\beta}_0 \text{ and } \widehat{\beta}_1^* = \frac{w_1}{w_2} \widehat{\beta}_1 \\ \Longrightarrow & \quad se\left(\widehat{\beta}_0^*\right) = w_1 \cdot se\left(\widehat{\beta}_0\right) \text{ and } se\left(\widehat{\beta}_1^*\right) = \frac{w_1}{w_2} \cdot se\left(\widehat{\beta}_1\right) \end{aligned}$$

Relationship Between t Statistics and CIs (Section 6.1: MLR)

- It is natural to predict that t statistic is the same as before:

$$t_{\widehat{\beta}_1^*} = \frac{\widehat{\beta}_1^*}{se\left(\widehat{\beta}_1^*\right)} = \frac{\frac{w_1}{w_2}\widehat{\beta}_1}{\frac{w_1}{w_2}\cdot se\left(\widehat{\beta}_1\right)} = \frac{\widehat{\beta}_1}{se\left(\widehat{\beta}_1\right)} = t_{\widehat{\beta}_1}.$$

- The CI for β_1 is the original CI multiplied by $\frac{w_1}{w_2}$:

$$\begin{aligned} & \left[\widehat{\beta}_1^* - 1.96\cdot se\left(\widehat{\beta}_1^*\right), \widehat{\beta}_1^* + 1.96\cdot se\left(\widehat{\beta}_1^*\right)\right] \\ = \quad & \frac{w_1}{w_2}\left[\widehat{\beta}_1 - 1.96\cdot se\left(\widehat{\beta}_1\right), \widehat{\beta}_1 + 1.96\cdot se\left(\widehat{\beta}_1\right)\right]. \end{aligned}$$

- The results for β_0 are similar; the only difference is to replace $\frac{w_1}{w_2}$ by w_1.

Unit Change in Logarithmic Form

- Changing the unit of measurement of y and x, when they appear in logarithmic form, does not affect any of the slope estimates, but may affect the intercept estimate.
- why?

$$
\begin{aligned}
& \min_{\beta_0^*,\beta_1^*} \sum_{i=1}^n \left[\log(y_i^*) - \beta_0^* - \beta_1^* \log(x_i^*)\right]^2 \\
= & \min_{\beta_0^*,\beta_1^*} \sum_{i=1}^n \left[\log(y_i) + \log(w_1) - \beta_0^* - \beta_1^* \log(x_i) - \beta_1^* \log(w_2)\right]^2 \\
= & \min_{\beta_0^*,\beta_1^*} \sum_{i=1}^n \left[\log(y_i) - (\beta_0^* + \beta_1^* \log(w_2) - \log(w_1)) - \beta_1^* \log(x_i)\right]^2,
\end{aligned}
$$

so

$$
\begin{aligned}
\widehat{\beta}_0 &= \widehat{\beta}_0^* + \widehat{\beta}_1^* \log(w_2) - \log(w_1) \text{ and } \widehat{\beta}_1 = \widehat{\beta}_1^* \\
&\Longrightarrow \widehat{\beta}_0^* = \widehat{\beta}_0 - \widehat{\beta}_1 \log(w_2) + \log(w_1) \text{ and } \widehat{\beta}_1^* = \widehat{\beta}_1.
\end{aligned}
$$

- I.e., the elasticity is invariant to the units of measurement of either y or x, and the intercept is related to both the original intercept and slope.

Example: Japanese Learning

- Suppose we want to study the SLR,

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

where

$y_i =$ exam mark in Japanese language course

$x_i =$ hours of study per day during the semester (average hours)

- The fitted regression line is

$$\begin{array}{rcl} \widehat{y}_i & = & \underset{(5.6)}{20} + \underset{(3.5)}{14}x_i \end{array}$$

- If you do not study at all, the predicted mark is 20. One additional hour of study per day increases exam mark by 14 marks. At the mean value $\overline{x} = 3$ hours of study per day is expected to result in a mark of $\overline{y} = 62$.

continue

- Now if we report hours of study per week (x_i^*) rather than per day, the variable has been scaled:
$$x_i^* = w_2 x_i = 7x_i \text{ and } y_i^* = w_1 y_i = y_i.$$
- If we run the regression based on x_i^* and y_i, we get
$$\begin{array}{rcl}\widehat{y}_i & = & \underset{(5.6)}{20} + \underset{(0.5)}{2}x_i^*\end{array}$$
- Each additional hour of study per week increases the exam mark by 2 marks [intuition here].
- t statistic remains the same: $\frac{2}{0.5} = \frac{14}{3.5}$.
- The CI for β_1^*, $[2-1.96\times 0.5, 2+1.96\times 0.5] = [1.02, 2.98]$, is $1/7$ of the CI for β_1, which is $[14-1.96\times 3.5, 14+1.96\times 3.5] = [7.14, 20.98]$.
- Note that the means $\overline{x}^*$ and $\overline{y}^*$ will still be on the newly estimated regression line:
$$\begin{aligned}\overline{y}^* &= 20 + 2\overline{x}^* \\ 62 &= 20 + 2\times 21\end{aligned}$$

More on Functional Form

a: More on Using Logarithmic Functional Forms

- Logarithmic transformations have the convenient percentage/elasticity interpretation.
- Slope coefficients of logged variables are invariant to rescalings.
- Taking logs often eliminates/mitigates problems with outliers. (why? figure here)
- Taking logs often helps to secure normality (e.g., log(wage) vs. wage) and homoskedasticity (see Chapter 8 for an example).
- Variables measured in units such as years (e.g., education, experience, tenure, age, etc) should not be logged.
- Variables measured in percentage points (e.g., unemployment rate, participation rate of a pension plan, etc.) should also not be logged.
- Logs must not be used if variables take on zero or negative values (e.g., hours of work during a month).
- It is hard to reverse the log-operation when constructing predictions. (we will discuss more on this point later in this chapter)

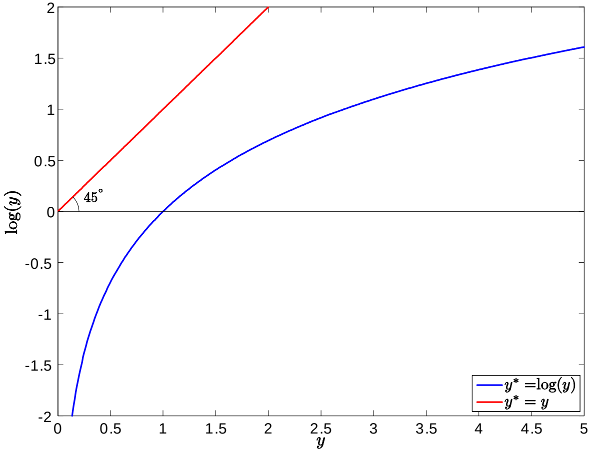


Figure: $\log y < y$ and $\lim_{y\to\infty} \frac{\log y}{y} = 0$

b: Models with Quadratics

- Example: Suppose the fitted regression line for the wage equation is

$$
\begin{aligned}
\widehat{wage} &= 3.73 + .298exper - .0061exper^2 \\
& \quad (.35) \quad (.041) \qquad\quad (.0009) \\
n &= 526, R^2 = .093 \text{ (quite small)}
\end{aligned}
$$

- The predicted *wage* is a concave function of *exper*. [figure here]
- The marginal effect of *exper* on wage is

$$\frac{\partial wage}{\partial exper} = \widehat{\beta}_1 + 2\widehat{\beta}_2 exper = .298 - 2 \times .0061 exper.$$

- The first year of experience increases the wage by some $\$.30$, the second year by $.298 - 2(.0061)(1) = \$.29 < \$.30$ etc.

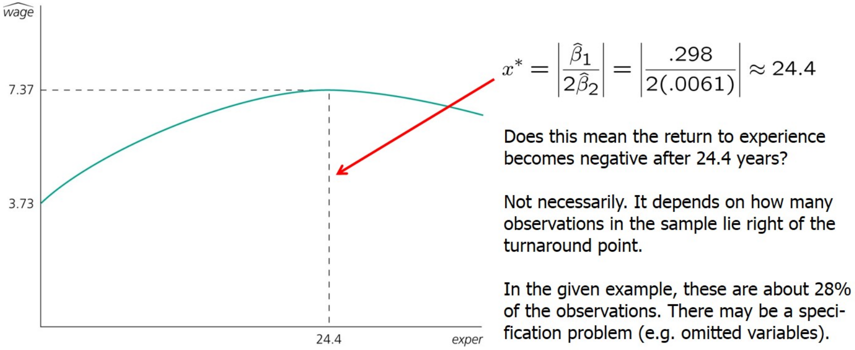


Figure: Wage Maximum with Respect to Work Experience

- Mincer (1974, *Schooling, Experience and Earnings*) assumed $\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + u$. [photo here]

History of the Wage Equation

Jacob Mincer (1922-2006), Columbia,
father of modern labor economics

Example: Effects of Pollution on Housing Prices

- The fitted regression line is

$$\begin{aligned}
\widehat{\log(price)} &= \underset{(.57)}{13.39} - \underset{(.115)}{.902}\log(nox) - \underset{(.043)}{.087}\log(dist) \\
&\quad - \underset{(.165)}{.545}rooms + \underset{(.013)}{.062}rooms^2 - \underset{(.006)}{.048}stratio \\
n &= 506, R^2 = .603
\end{aligned}$$

  where
  nox = nitrogen oxide in air
  $dist$ = distance from employment centers, in miles
  $stratio$ = student/teacher ratio
- The predicted $\log(price)$ is a convex function of *rooms*. [figure here]
- The coefficient of *rooms* is negative. Does this mean that, at a low number of rooms, more rooms are associated with lower prices?
- The marginal effect of *rooms* on $\log(price)$ is

$$\frac{\partial \log(price)}{\partial rooms} = \frac{\partial price/price}{\partial rooms} = -.545 + 2 \times .062 rooms.$$

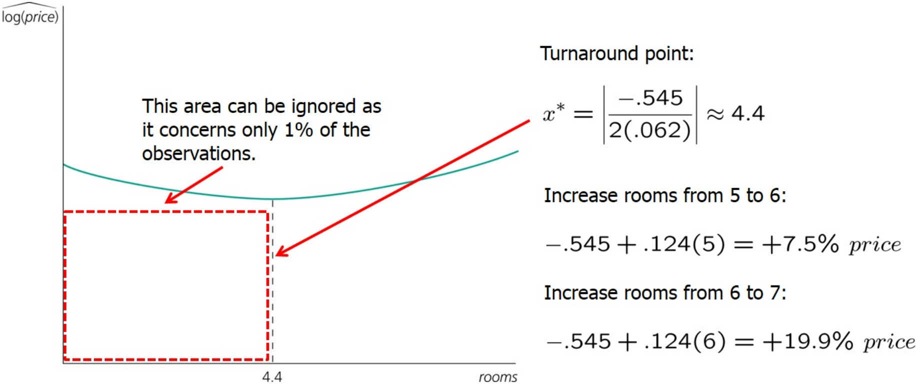


Figure: Calculation of the Turnaround Point

Other Possibilities

- Using Quadratics Along with Logarithms:

$$\begin{aligned}\log(price) &= \beta_0 + \beta_1 \log(nox) + \beta_2 \log(nox)^2 \\ &\quad + \beta_3 crime + \beta_4 rooms + \beta_5 rooms^2 + \beta_6 stratio + u,\end{aligned}$$

  which implies

$$\frac{\partial \log(price)}{\partial \log(nox)} = \frac{\%\partial price}{\%\partial nox} = \beta_1 + 2\beta_2 \log(nox).$$

- Higher Order Polynomials: It is often assumed that the total cost takes the following form,

$$cost = \beta_0 + \beta_1 quantity + \beta_2 quantity^2 + \beta_3 quantity^3 + u,$$

  which implies a U-shaped marginal cost (MC), where β_0 is the total fixed cost. [figure here]

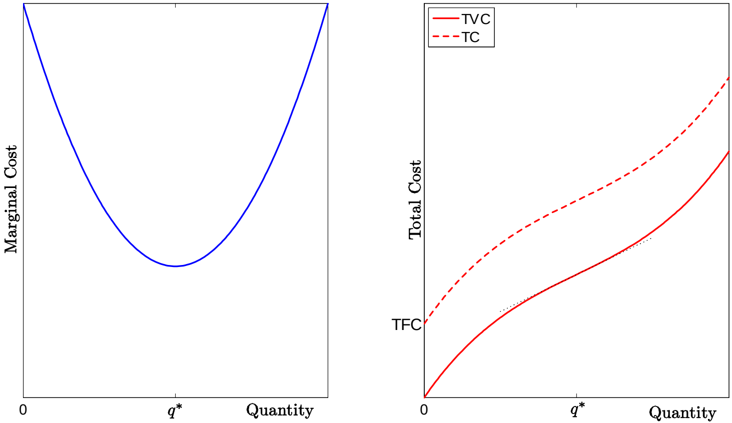


Figure: Quadratic MC Implies Cubic TC: q^* is the inflection point

c: Models with Interaction Terms

- In the model

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + \beta_3 sqrft \cdot bdrms + \beta_4 bthrms + u,$$

$sqrft \cdot bdrms$ is the interaction term.

- The marginal effect of *bdrms* on *price* is

$$\frac{\partial price}{\partial bdrms} = \beta_2 + \beta_3 sqrft.$$

- The effect of the number of bedrooms depends on the level of square footage.
- Interaction effects complicate interpretation of parameters: β_2 is the effect of number of bedrooms, but for a square footage of zero.
- How to avoid this interpretation difficulty?

Reparametrization of Interaction Effects

- The model
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$$
can be reparametrized as
$$y = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 - \mu_1)(x_2 - \mu_2) + u,$$
where $\mu_1 = E[x_1]$ and $\mu_2 = E[x_2]$ are population means of x_1 and x_2, and can be replaced by their sample means.
- What is the relationship between $\left(\widehat{\alpha}_0, \widehat{\delta}_1, \widehat{\delta}_2\right)$ and $\left(\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2\right)$? (Exercise)
- Now,
$$\frac{\partial y}{\partial x_2} = \delta_2 + \beta_3 (x_1 - \mu_1),$$
i.e., δ_2 is the effect of x_2 if all other variables take on their mean values.
- Advantages of reparametrization:
  - It is easy to interpret all parameters.
  - Standard errors for partial effects at the mean values are available.
  - If necessary, interaction may be centered at other interesting values.

More on Goodness-of-Fit and Selection of Regressors

a: Adjusted R-Squared

- General remarks on R-squared:
  - A high R-squared does not imply that there is a causal interpretation.
  - A low R-squared does not preclude precise estimation of partial effects.
- Recall that

$$R^2 = 1 - \frac{SSR/n}{SST/n} = 1 - \frac{\tilde{\sigma}_u^2}{\tilde{\sigma}_y^2},$$

so R^2 is estimating the population R-squared

$$\rho^2 = 1 - \frac{\sigma_u^2}{\sigma_y^2},$$

the proportion of the variation in y in the population explained by the independent variables.

- Adjusted R-Squared:

$$\overline{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{\hat{\sigma}_u^2}{\hat{\sigma}_y^2},$$

is sometimes also called R-bar squared [photo here], where $\hat{\sigma}_u^2$ and $\hat{\sigma}_y^2$ are unbiased estimators of σ_u^2 and σ_y^2 due to the correction of dfs.

History of $\overline{R}^2$

Henri Theil (1924-2000)[1], Chicago and Florida

[1]He is a Dutch econometrician. Two other Dutch econometricians, Jan Tinbergen (1903-1994) and Tjalling Koopmans (1910-1985) won the Nobel Prize in economics in 1969 and 1975, respectively.

continue

- $\overline{R}^2$ takes into account degrees of freedom of the numerator and denominator, so is generally a better measure of goodness-of-fit.
- $\overline{R}^2$ imposes a penalty for adding new regressors: $k \uparrow \Longrightarrow \overline{R}^2 \downarrow$
- $\overline{R}^2$ increases if and only if the t statistic of a newly added regressor is greater than one in absolute value. [proof not required]
  - Compared with $y = \beta_0 + \beta_1 x_1 + u$, the regression $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ has a larger $\overline{R}^2$ if and only if

$$\left| t_{\widehat{\beta}_2} \right| > 1.$$

- Relationship between R^2 and $\overline{R}^2$: by

$$1 - R^2 = \frac{SSR}{SST} = \frac{n-k-1}{n-1} \frac{SSR/(n-k-1)}{SST/(n-1)} = \frac{n-k-1}{n-1}\left(1 - \overline{R}^2\right),$$

  we have

$$\overline{R}^2 = 1 - \left(1 - R^2\right) \frac{n-1}{n-k-1} < R^2$$

  unless $k = 0$ or $R^2 = 1$. [figure here]
- Note that $\overline{R}^2$ even gets negative if $R^2 < \frac{k}{n-1}$.

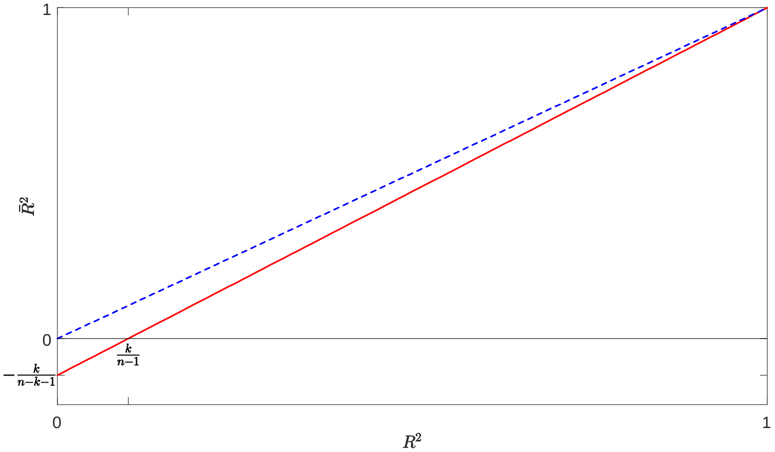


Figure: Relationship Between $\overline{R}^2$ and R^2

b: Using Adjusted *R*-squared to Choose between Nonnested Models

- Models are nonnested if neither model is a special case of the other.
- For example, to incorporate diminishing return of *sales* to R&D, we consider two models:

$$\begin{aligned} rdintens &= \beta_0 + \beta_1 \log(sales) + u, \\ rdintens &= \beta_0 + \beta_1 sales + \beta_2 sales^2 + u, \end{aligned}$$

  where
  $rdintens$ = R&D intensity.
- $R^2 = .061$ and $\overline{R}^2 = .030$ in model 1 and $R^2 = .148$ and $\overline{R}^2 = .090$ in model 2.
- A comparison between the R-squared of both models would be unfair to the first model because the first model contains fewer parameters.
- In the given example, even after adjusting for the difference in degrees of freedom, the quadratic model is preferred.

Comparing Models with Different Dependent Variables

- R-squared or adjusted R-squared must not be used to compare models which differ in their definition of the dependent variable.
- Example (CEO Compensation and Firm Performance):

$$\begin{aligned} \widehat{salary} &= \underset{(223.90)}{830.63} + \underset{(.0089)}{.0163}sales + \underset{(11.08)}{19.63}roe \\ n &= 209, R^2 = .029, \overline{R}^2 = .020, SST = 391,732,982 \end{aligned}$$

and

$$\begin{aligned} \widehat{lsalary} &= \underset{(0.29)}{4.36} + \underset{(.033)}{.275}lsales + \underset{(.0040)}{.0179}roe \\ n &= 209, R^2 = .282, \overline{R}^2 = .275, SST = 66.72 \end{aligned}$$

- There is much less variation in $\log(salary)$ that needs to be explained than in $salary$, so it is not fair to compare R^2 and $\overline{R}^2$ of the two models. (we will discuss how to compare the fitting of these two models later in this chapter)

c: Controlling for Too Many Factors in Regression Analysis

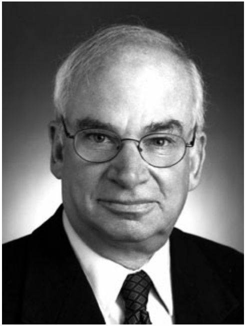
- In some cases, certain variables should not be held fixed:
  - In a regression of traffic fatalities on state beer taxes (and other factors) one should not directly control for beer consumption.
  - why? Beer taxes influence traffic fatalities only through beer consumption; if beer consumption is controlled, then the coefficient of beer taxes measures the indirect effect of beer taxes, which is hardly interesting.
  - In a regression of family health expenditures on pesticide usage among farmers one should not control for doctor visits.
  - why? Health expenditures include doctor visits, and we would like to pick up all effects of pesticide use on health expenditure.
- Different regressions may serve different purposes:
  - In a regression of house prices on house characteristics, one would include price assessments and also housing attributes if the purpose of the regression is to study the validity of assessments; one should not include price assessments if the purpose of the regression is to estimate a hedonic price model,[2] which measures the marginal values of various housing attributes.

[2] What consumers are seeking to acquire is not goods themselves (e.g. cars or train journeys) but the characteristics they contain (e.g., display of fashion sense, transport from A to B).

History of Hedonic Price Model

- Hedonic Utility: Lancaster, Kelvin J., 1966, A New Approach to Consumer Theory, *Journal of Political Economy*, 74, 132-157.
- Hedonic Pricing: Rosen, S., 1974, Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *Journal of Political Economy*, 82, 34-55.

Kelvin J. Lancaster (1924-1999), Columbia



Sherwin Rosen (1938-2001)[3], Chicago

[3] His student Robert H. Thaler (1945-) at the University of Chicago won the Nobel Prize in economics in 2017.

d: Adding Regressors to Reduce the Error Variance

- Recall that

$$Var\left(\widehat{\beta}_j\right) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)}.$$

  - Adding regressors may exacerbate multicollinearity problems ($R_j^2 \uparrow$).
  - On the other hand, adding regressors reduces the error variance ($\sigma^2 \downarrow$).
- Variables that are uncorrelated with other regressors should be added because they reduce error variance ($\sigma^2 \downarrow$) without increasing multicollinearity (R_j^2 remains the same).
- However, such uncorrelated variables may be hard to find.
- Example (Individual Beer Consumption and Beer Prices): Including individual characteristics in a regression of beer consumption on beer prices leads to more precise estimates of the price elasticity if individual characteristics are uncorrelated with beer prices.

$$\log(cons) = \beta_0 + \beta_1 \log(price) + \underbrace{indchar}_{\text{uncorrelated with } \log(price)} + u.$$

(**) Prediction and Residual Analysis

c: Predicting y When $\log(y)$ is the Dependent Variable

- We study only this prediction problem as promised.
- Note that

$$\log(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

implies

$$y = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) \exp(u) = m(\mathbf{x}) \exp(u).$$

- Under the additional assumption that u is independent of $(x_1, \cdots, x_k)$, we have

$$\begin{aligned} E[y|\mathbf{x}] &= \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) E[\exp(u)|\mathbf{x}] \\ &= \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) E[\exp(u)] \\ &\equiv m(\mathbf{x})\alpha_0, \end{aligned}$$

where the second equality is due to the independence between u and $\mathbf{x}$, so the predicted y is

$$\widehat{y} = \widehat{m}(\mathbf{x})\widehat{\alpha}_0$$

where

$$\widehat{m}(\mathbf{x}) = \exp\left(\widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \cdots + \widehat{\beta}_k x_k\right) \text{ and } \widehat{\alpha}_0 = \frac{1}{n}\sum_{i=1}^n \exp(\widehat{u}_i).$$

$E\left[\exp(u)\right] \geq 1$

- Recall that $E[u] = 0$, so

$$E\left[\exp(u)\right] \geq \exp(E[u]) = \exp(0) = 1.$$

- In the following figure, suppose u takes only two values u_1 and u_2 with probability $\frac{1}{2}$ and $\frac{1}{2}$, respectively. Since $E[u] = \frac{1}{2}(u_1 + u_2) = 0$, $u_1 = -u_2$.
- Now,

$$\begin{aligned} E\left[\exp(u)\right] &= \frac{1}{2}\left(\exp(u_1) + \exp(u_2)\right) \\ &\geq \exp\left(\frac{1}{2}(u_1 + u_2)\right) \\ &= \exp(E[u]) = 1, \end{aligned}$$

where the equality is achieved only if $u_1 = u_2 = 0$, i.e., $u = 0$.
- As a result, $\widetilde{y} = \widehat{m}(\mathbf{x}) = \exp\left(\widehat{\log(y)}\right)$ under-estimates $E[y|\mathbf{x}]$!

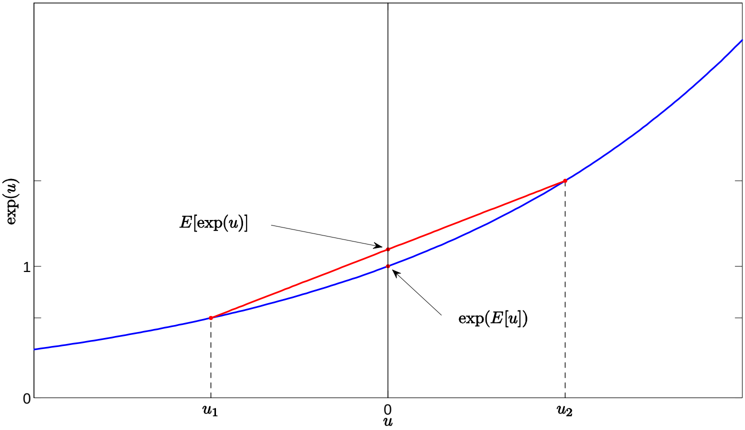


Figure: Illustration of Jensen's Inequality

Comparing R-Squared of a Logged and an Unlogged Specification

- Reconstider the CEO salary problem:

$$\begin{aligned} \widehat{salary} &= 613.43 + .0190 sales + .0234 mktval + 12.70 ceoten \\ & \quad (65.23)\ (.0100) \qquad (.0095) \qquad (5.61) \\ n &= 177, R^2 = .201 \end{aligned}$$

  and

$$\begin{aligned} \widehat{lsalary} &= 4.504 + .163 lsales + .0109 mktval + .0117 ceoten \\ & \quad (0.257)(.039) \qquad (.050) \qquad (.0053) \\ n &= 177, \widetilde{R}^2 = .318 \end{aligned}$$

- R^2 and $\widetilde{R}^2$ are the R-squareds for the predictions of the unlogged salary variable (although the second regression is originally for logged salaries). Both R-squareds can now be directly compared

About $\widetilde{R}^2$

- Recall that

$$R^2 = \widehat{Corr}\left(y, \widehat{y}\right)^2,$$

where $\widehat{y}$ is the predicted value of y.

- When *lsalary* is the dependent variable, the predicted value of y is $\widehat{m}(\mathbf{x})\widehat{\alpha}_0 = \widehat{\alpha}_0\widetilde{y}$.
- Since $\widehat{\alpha}_0 > 0$,

$$\widehat{Corr}\left(y, \widehat{y}\right) = \widehat{Corr}\left(y, \widehat{\alpha}_0\widetilde{y}\right) = \widehat{Corr}\left(y, \widetilde{y}\right)$$

invariant to $\widehat{\alpha}_0$, where recall that for any $a > 0$,

$$Corr\left(X, aY\right) = Corr\left(X, Y\right) = \frac{Cov\left(X, Y\right)}{\sqrt{Var\left(X\right)Var\left(Y\right)}}.$$

- As a result,

$$\widetilde{R}^2 = \widehat{Corr}\left(y, \widetilde{y}\right)^2.$$