Least Squares Prediction

Least Squares Prediction, Part I

- The ability to predict is important to:
  - Business economists and financial analysts who attempt to forecast the sales and revenues of specific firms
  - Government policymakers who attempt to predict the rates of growth in national income, inflation, investment, saving, social insurance program expenditures, and tax revenues
  - Local businesses who need to have predictions of growth in neighborhood populations and income so that they may expand or contract their provision of service
- Accurate predictions provide a basis for better decision making in every type of planning context

Least Squares Prediction, Part II

- In order to use regression analysis as a basis for prediction, we must assume that y_0 and x_0 are related to one another by the same regression model that describes our sample of data, so that, in particular, SR1 holds for these observations.

  (4.1) $y_0 = \beta_1 + \beta_2 x_0 + e_0$

  where e_0 is a random error

Least Squares Prediction, Part III

- The task of predicting y_0 is related to the problem of estimating $E(y_0|x_0) = \beta_1 + \beta_2 x_0$
- Although $E(y_0|x_0) = \beta_1 + \beta_2 x_0$ is not random, the outcome y_0 is random
- Consequently, as we will see, there is a difference between the **interval estimate** of $E(y_0|x_0) = \beta_1 + \beta_2 x_0$ and the **prediction interval** for y_0
- The **least squares predictor** of y_0 comes from the fitted regression line

(4.2) $\hat{y}_0 = b_1 + b_2 x_0$

Least Squares Prediction, Part IV

- To evaluate how well this predictor performs, we define the forecast error, which is analogous to the least squares residual

(4.3) $$f = y_0 - \hat{y}_0 = \left(\beta_1 + \beta_2 x_0 + e_0\right) - \left(\mathrm{b}_1 + \mathrm{b}_2 x_0\right)$$

- We would like the forecast error to be small, implying that our forecast is close to the value we are predicting

Least Squares Prediction, Part V

- Taking the expected value of *f,* we find that:

$$E(f|x) = \beta_1 + \beta_2 x_0 + E(e_0) - [E(b_1) + E(b_2)x_0] = \beta_1 + \beta_2 x_0 + 0 - [\beta_1 + \beta_2 x_0] = 0$$

Which means, on average, the forecast error is zero and $\hat{y}_0$ is an **unbiased predictor** of y_0

- However, unbiasedness does not necessarily imply that a particular forecast will be close to the actual value.
- y_0 is the **best linear unbiased predictor** (*BLUP*) of y_0 if assumptions SR1–SR5 hold

Least Squares Prediction, Part VI

- The variance of the forecast is equation 4.4: $var(f|x) = \sigma^2\left[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right]$]
- The variance of the forecast is smaller when:
  - The overall uncertainty in the model is smaller, as measured by the variance of the random errors σ^2
  - The sample size *N* is larger
  - The variation in the explanatory variable is larger
  - The value of $(x_0 - x)^2$ is small

Least Squares Prediction, Part VII

- In practice we use $\widehat{var}(f|x) = \hat{\sigma}^2\left[1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right]$ for the variance
- The **standard error of the forecast** is equation 4.5: $\mathrm{se}(f) = \sqrt{\widehat{var}(f|x)}$
- The 100(1 − α)% **prediction interval** is:
  - (4.6) $\hat{y}_0 \pm t_c \mathrm{se}(f)$

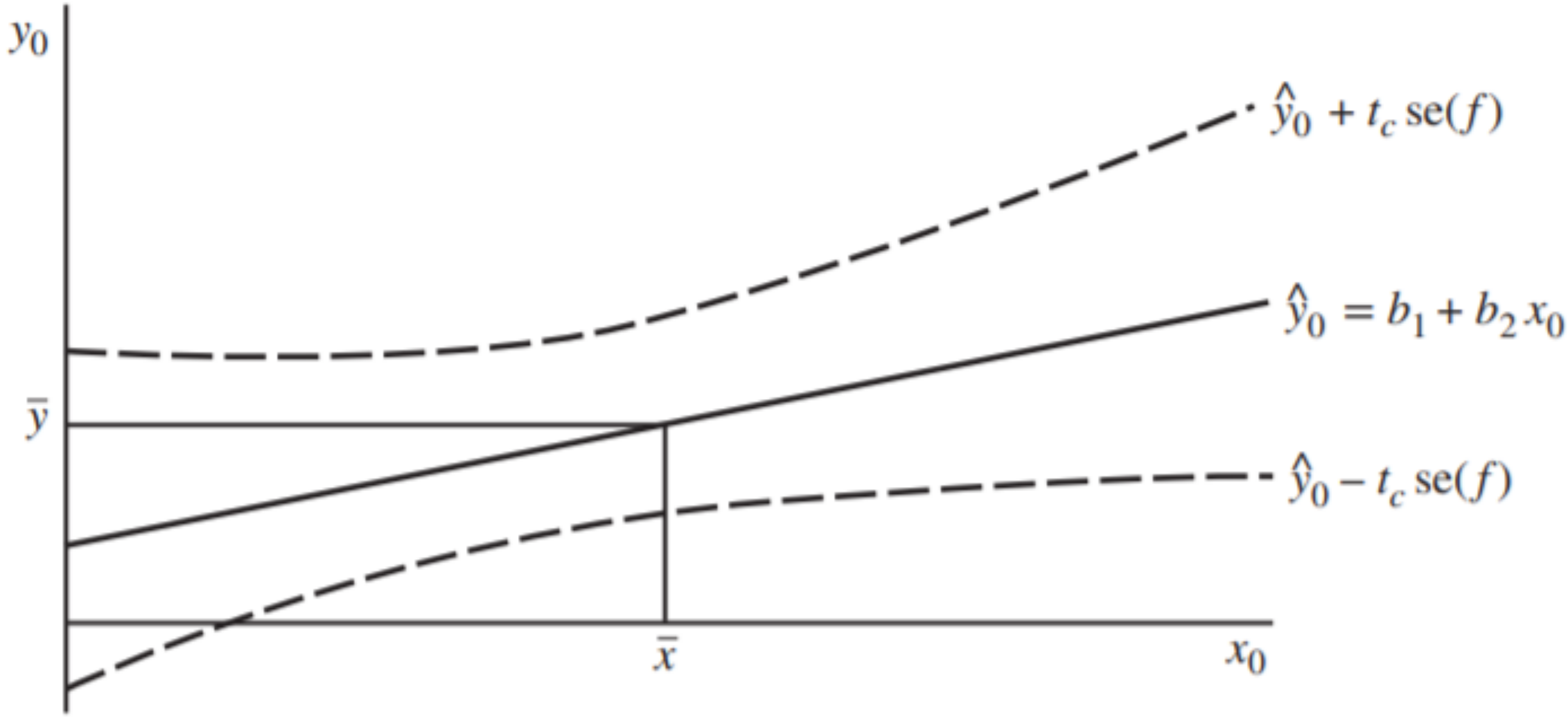


FIGURE 4.2 Point and interval prediction.

FORDHAM

THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Measuring Goodness of Fit

Measuring Goodness of Fit, Part I

- There are two major reasons for analyzing the model:
  - (4.7) $y_i = \beta_1 + \beta_2 x_i + e_i$

1. To explain how the dependent variable (y_i) changes as the independent variable (x_i) changes
2. To predict y_0 given an x_0

Measuring Goodness of Fit, Part II

- To develop a measure of the variation in y_i that is explained by the model, we begin by separating y_i into its explainable and unexplainable components
  - (4.8) $y_i = E(y_i|x) + e_i$
  - $E(y_i|x)$ is the explainable or systematic part
  - e_i is the random, unsystematic, and unexplainable component

Measuring Goodness of Fit, Part III

- Recall that the sample variance of y_i is $s_y^2 = \frac{\sum(\hat{y}_i - \bar{y})}{N-1}$
- Squaring and summing both sides of (4.10), and using the fact that:

$$\sum(\hat{y}_i - \bar{y})\hat{e}_i = 0 \quad \text{we get: (4.11)}$$

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum \hat{e}_i^2$$

- Equation 4.11 decomposition of the "total sample variation" in *y* into explained and unexplained components
  - These are called "sums of squares"

Measuring Goodness of Fit, Part IV

- Specifically:

$$\sum \left(y_i - \bar{y} \right)^2 = \text{total sum of squares} \;=\; \text{SST}$$

$$\sum \left(\hat{y}_i - \bar{y} \right)^2 = \text{sum of squares due to regression} \;=\; \text{SSR}$$

$$\sum \hat{e}_i^2 = \text{sum of squares due to error} \;=\; \text{SSE}$$

- Using these abbreviations, equation 4.11 becomes ***SST = SSR + SSE***

4.2 Measuring Goodness of Fit, Part V

- Let's define the **coefficient of determination**, or R^2, as the proportion of variation in y explained by x within the regression model:
  - (4.12) $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$
- The closer R2 is to 1, the closer the sample values y_i are to the fitted regression equation

Measuring Goodness of Fit, Part VI

- If R^2 = 1, then all the sample data fall exactly on the fitted least squares line, so SSE = 0, and the model fits the data "perfectly"
- If the sample data for y and x are uncorrelated and show no linear association, then the least squares fitted line is "horizontal" and identical to $\bar{y}$, so that SSR = 0 and R^2 = 0

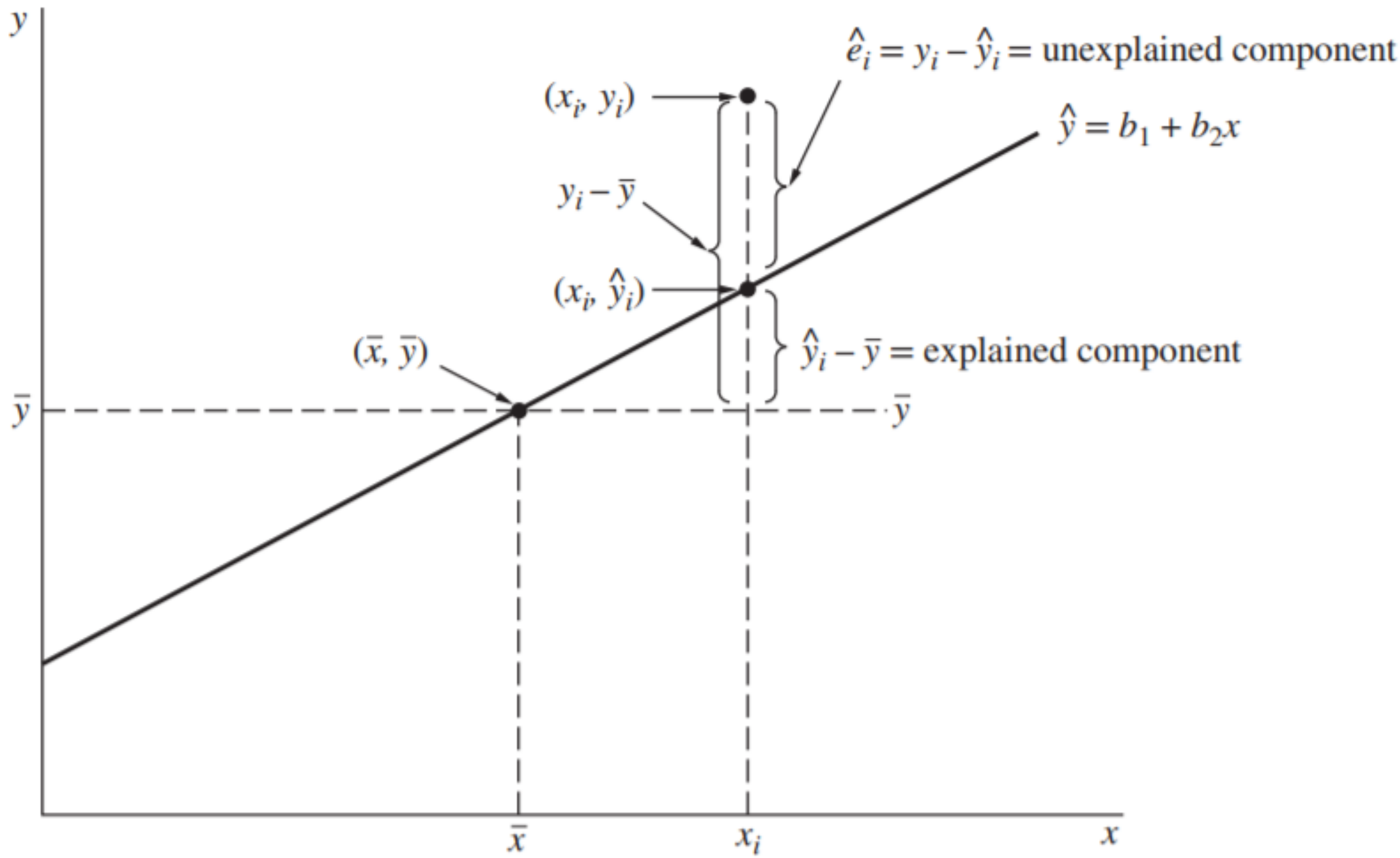


FIGURE 4.3 **Explained and unexplained components of y_i.**

Correlation Analysis

- The correlation coefficient ρ_{xy} between *x* and *y* is defined as:
  - (4.13) $\rho_{xy} = \dfrac{\operatorname{cov}(x, y)}{\sqrt{\operatorname{var}(x)\sqrt{\operatorname{var}(y)}}} = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$
- Substituting sample values, we get the sample correlation coefficient:

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

Correlation Analysis (cont.)

- Where:

$$s_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) / (N - 1)$$

$$s_x = \sqrt{\sum (x_i - \bar{x})^2 / (N - 1)}$$

$$s_y = \sqrt{\sum (y_i - \bar{y})^2 / (N - 1)}$$

- The sample correlation coefficient r_{xy} has a value between −1 and 1, and it measures the strength of the linear association between observed values of x and y

Correlation Analysis and R^2

Two relationships between R^2 and r_{xy}

1. $r^2_{xy} = R^2$
2. R^2 can also be computed as the square of the sample correlation coefficient between y_i and $b_1 + b_2 x_i$

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School
of Business

The Effects of Scaling the Data

The Effects of Scaling the Data, Part I

- What are the effects of scaling the variables in a regression model?
- Consider the food expenditure example
- We report weekly expenditures in dollars, but we report income in \$100 units, so a weekly income of \$2,000 is reported as x = 20
- If we had estimated the regression using income in dollars, the results would have been:

  FOOD_EXP = 83.42 + 0.1021 INCOME(\$) R^2 = 0.385 (se) (43.41) $*$(0.0209) $***$

The Effects of Scaling the Data, Part II

- Possible effects of scaling the data

1. Changing the scale of *x*: the coefficient of *x* must be multiplied by *c*, the scaling factor
   - When the scale of *x* is altered, the only other change occurs in the standard error of the regression coefficient, but it changes by the same multiplicative factor as the coefficient, so that their ratio, the *t*-statistic, is unaffected
   - All other regression statistics are unchanged

The Effects of Scaling the Data, Part III

- Possible effects of scaling the data

2. Changing the scale of y: If we change the units of measurement of y, but not x, then all the coefficients must change in order for the equation to remain valid
   - Because the error term is scaled in this process, the least squares residuals will also be scaled.
   - This will affect the standard errors of the regression coefficients, but it will not affect *t*-statistics or R^2

The Effects of Scaling the Data, Part IV

- Possible effects of scaling the data

3. Changing the scale of *y* and *x* by the same factor: there will be no change in the reported regression results for b_2, but the estimated intercept and residuals will change
    - t-statistics and R^2 are unaffected
    - The interpretation of the parameters is made relative to the new units of measurement

Choosing a Functional Form, Part I

- The starting point in all econometric analyses is economic theory
- What does economics really say about the relation between food expenditure and income, holding all else constant?
- We expect there to be a positive relationship between these variables because food is a normal good
- But nothing says the relationship must be a straight line

Choosing a Functional Form, Part II

- By transforming the variables *y* and *x*, we can represent many curved, nonlinear relationships and still use the linear regression model
  - Choosing an algebraic form for the relationship means choosing transformations of the original variables
  - The most common are:
    - **Power:** If x is a variable, then x^p means raising the variable to the power p
      - Quadratic (x^2)
      - Cubic (x^3)
    - **Natural logarithm:** If x is a variable, then its natural logarithm is $\ln(x)$

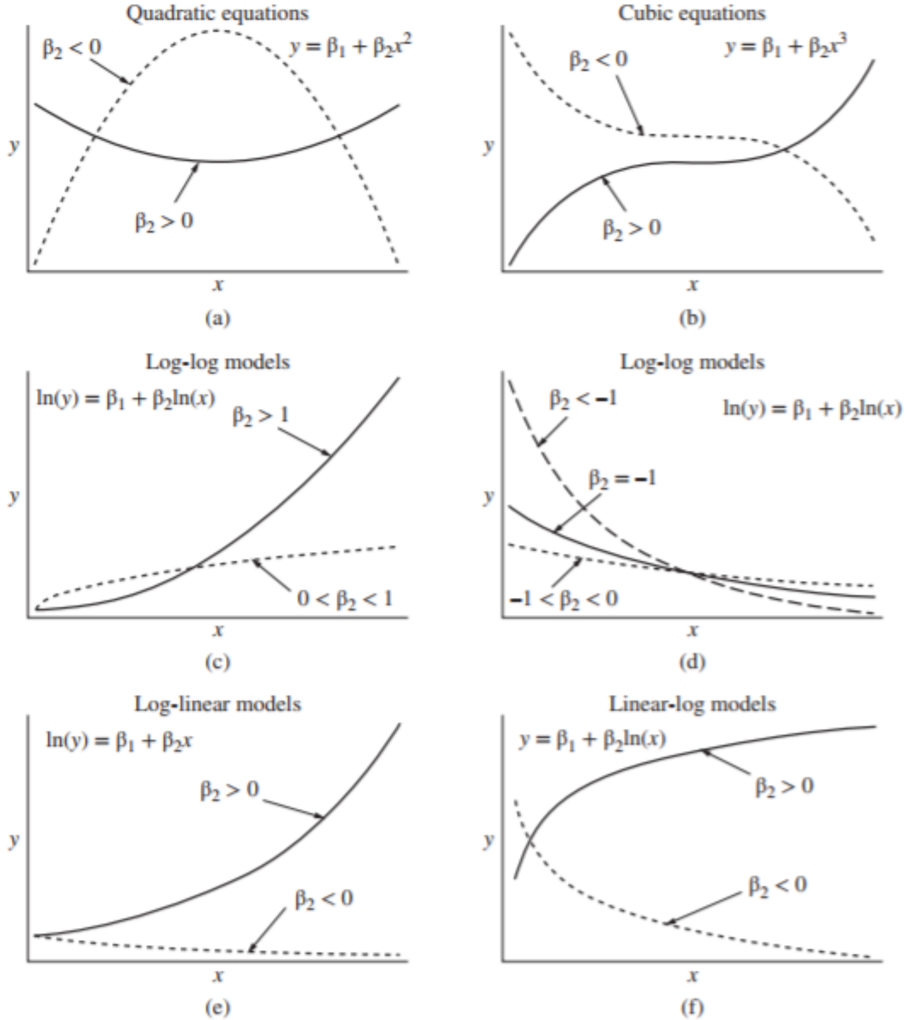


FIGURE 4.5 **Alternative functional forms.**

TABLE 4.1 **Some Useful Functions, Their Derivatives, Elasticities, and Other Interpretation**

Name	Function	Slope = dy/dx	Elasticity
Linear	$y = \beta_1 + \beta_2 x$	β_2	$\beta_2 \frac{x}{y}$
Quadratic	$y = \beta_1 + \beta_2 x^2$	$2\beta_2 x$	$(2\beta_2 x)\frac{x}{y}$
Cubic	$y = \beta_1 + \beta_2 x^3$	$3\beta_2 x^2$	$(3\beta_2 x^2)\frac{x}{y}$
Log-log	$\ln(y) = \beta_1 + \beta_2 \ln(x)$	$\beta_2 \frac{y}{x}$	β_2
Log-linear	$\ln(y) = \beta_1 + \beta_2 x$	$\beta_2 y$	$\beta_2 x$
	or, a 1 unit change in x leads to (approximately) a $100\beta_2\%$ change in y		
Linear-log	$y = \beta_1 + \beta_2 \ln(x)$	$\beta_2 \frac{1}{x}$	$\beta_2 \frac{1}{y}$
	or, a 1% change in x leads to (approximately) a $\beta_2/100$ unit change in y		

Choosing a Functional Form, Part II

- Summary of three configurations:
  1. In the log-log model both the dependent and independent variables are transformed by the "natural" logarithm
     - The parameter β_2 is the elasticity of y with respect to *x*
  2. In the log-linear model only the dependent variable is transformed by the logarithm
  3. In the linear-log model the variable *x* is transformed by the natural logarithm

A Linear-Log Food Expenditure Model

- A linear-log equation has a linear, untransformed term on the left-hand side and a logarithmic term on the right-hand side:
  $y = \beta_1 + \beta_2 \ln(x)$
  - The elasticity of y with respect to x is $\varepsilon = \text{slope} \times x/y = \beta_2/y$
  - A convenient interpretation is:
    - The change in y, represented in its units of measure, is approximately β_2 =100 times the percentage change in x

$$\Delta y = y_1 - y_0 = \beta_2 \left[\ln(x_1) - \ln(x_0) \right]$$

$$= \frac{\beta_2}{100} \times 100 \left[\ln(x_1) - \ln(x_0) \right]$$

$$\approx \frac{\beta_2}{100} (\%\Delta\, x)$$

A Linear-Log Food Expenditure Model (cont.)

- Given alternative models that involve different transformations of the dependent and independent variables, and some of which have similar shapes, what are some guidelines for choosing a functional form?

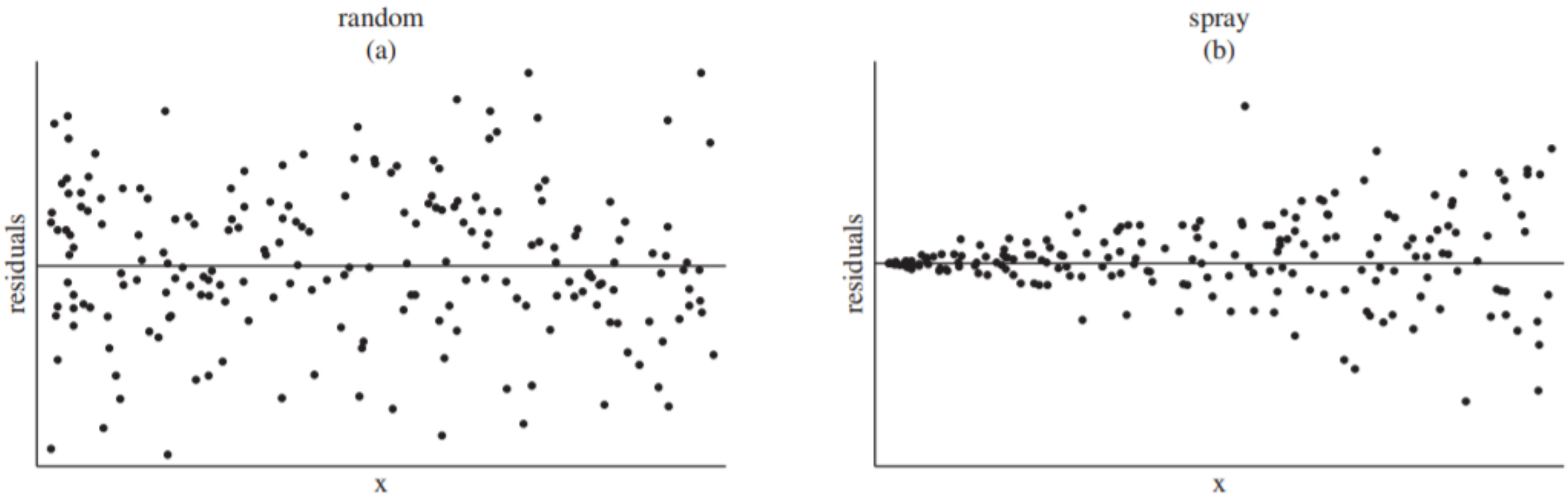
1. Choose a shape that is consistent with what economic theory tells us about the relationship
2. Choose a shape that is sufficiently flexible to “fit” the data
3. Choose a shape so that assumptions SR1–SR6 are satisfied, ensuring that the least squares estimators have the desirable properties described in Chapters 2 and 3

Using Diagnostic Residual Plots, Part I

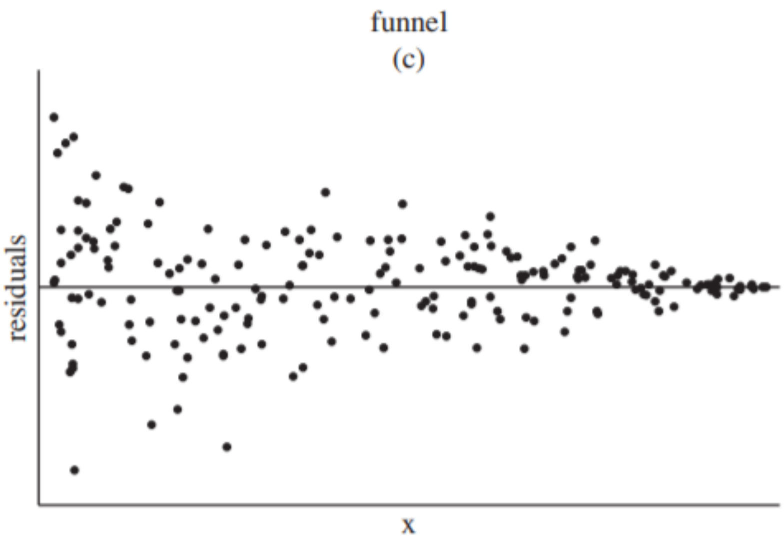
- When specifying a regression model, we may inadvertently choose an inadequate or incorrect functional form

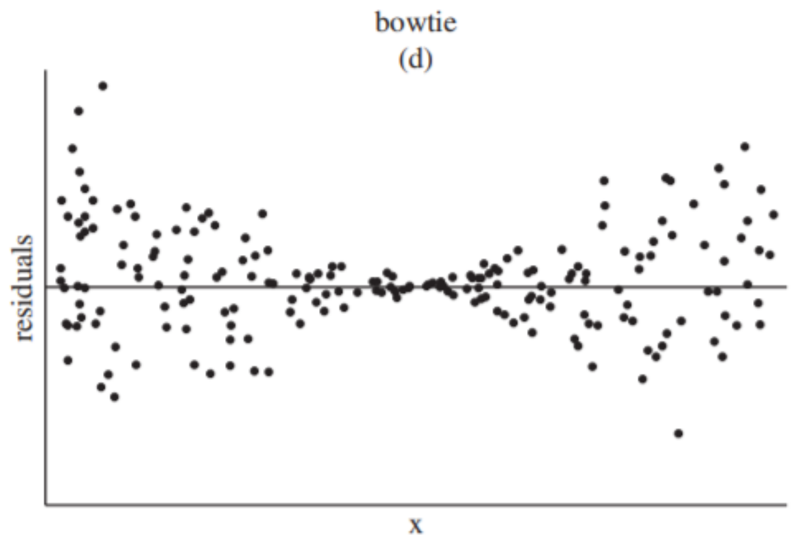
1. Examine the regression results
    - There are formal statistical tests to check for:
        - Homoskedasticity
        - Serial correlation
2. Use residual plots

Using Diagnostic Residual Plots, Part II

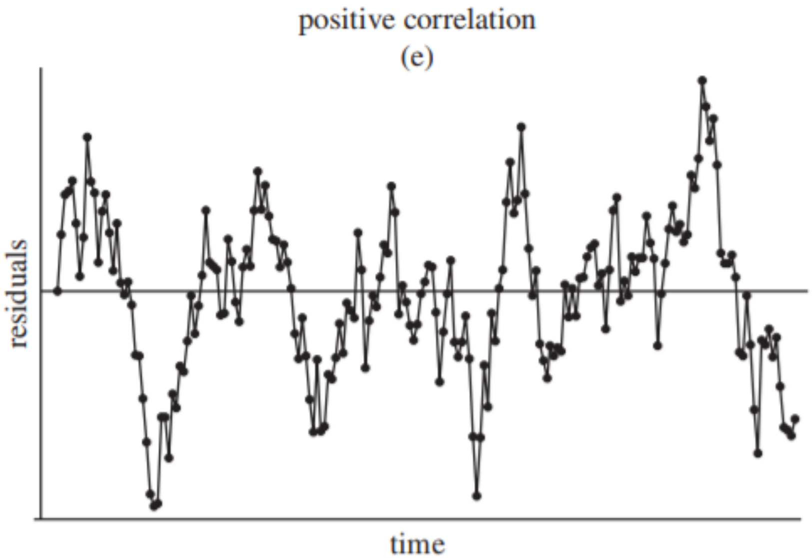


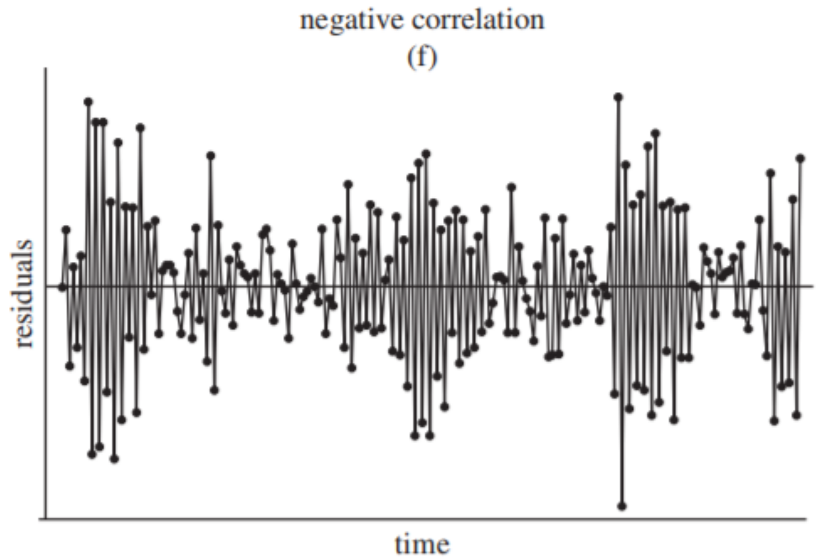
Using Diagnostic Residual Plots, Part III



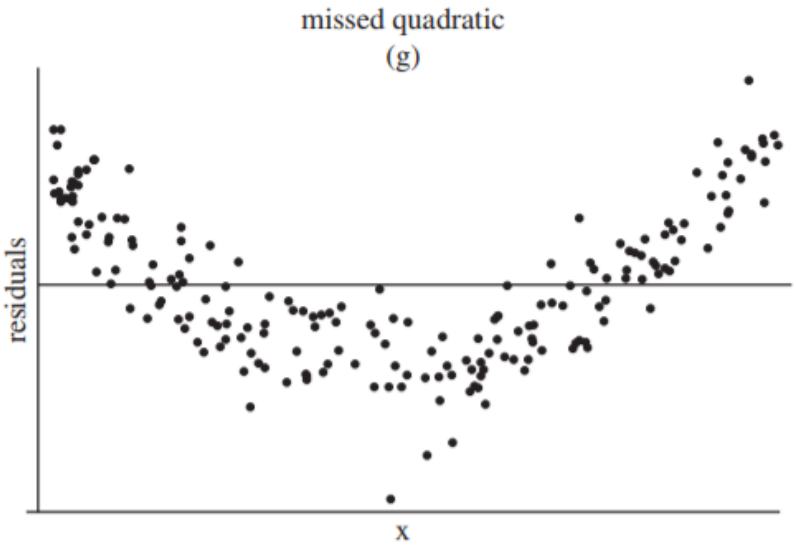


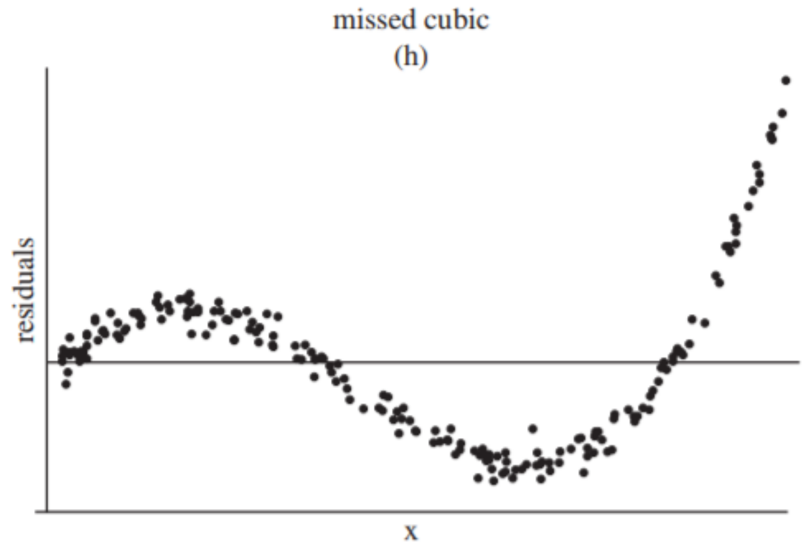
Using Diagnostic Residual Plots, Part IV





Using Diagnostic Residual Plots, Part V





Are the Regression Errors Normally Distributed?

- Hypothesis tests and interval estimates for the coefficients rely on the assumption that the errors, and hence the dependent variable *y*, are normally distributed.
- A histogram of the least squares residuals gives us a graphical representation of the empirical distribution.
- There are many tests for normality.
  - The Jarque–Bera test for normality is valid in large samples.
  - It is based on two measures: **skewness** and **kurtosis**.

Identifying Influential Observations

- One worry in data analysis is that we may have some unusual and/or **influential observations**. Sometimes, these are termed "outliers."
  - If an unusual observation is the result of a data error, then we should correct it.
  - Understanding how it came about, the story behind it, can be informative.
- One way to detect whether an observation is influential is to delete it and re-estimate the model.

Identifying Influential Observations (cont.)

- The **studentized residual** is the standardized residual based on the delete-one sample.
- If the studentized residual falls outside the 95% interval estimate interval, then the observation is worth examining because it is “unusually” large.
- Another measure of the influence of a single observation on the least squares estimates is called DFBETAS.

FORDHAM

THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Polynomial Models

Polynomial Models

- In addition to estimating linear equations, we can also estimate quadratic and cubic equations.
- Economics students will have seen many average and marginal cost curves (U-shaped) and average and marginal product curves (inverted-U shaped) in their studies.

Quadratic and Cubic Equations

- The general form of a quadratic equation is:

  $$y = a_0 + a_1 x + a_2 x^2$$

- The general form of a cubic equation is:

  $$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

- A problem with the linear equation is that it implies an increase at the same constant rate, when one might expect a rate to be increasing
- Polynomial models may provide a better fit

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Log-Linear Models

Log-Linear Models

- Econometric models that employ natural logarithms are very common
- Logarithmic transformations are often used for variables that are monetary values
  - Wages, salaries, income, prices, sales, and expenditures
  - In general, for variables that measure the "size" of something
  - These variables have the characteristic that they are positive and often have distributions that are positively skewed, with a long tail to the right

Log-Linear Models (cont.)

- The log-linear model, $\ln(y) = \beta_1 + \beta_2 x$, has a logarithmic term on the left-hand side of the equation and an untransformed (linear) variable on the right-hand side
  - Both its slope and elasticity change at each point and are the same sign as β_2
  - In the log-linear model, a one-unit increase in x leads, approximately, to a 100 β_2 % change in y

$$100\left[\ln\left(y_1\right)-\ln\left(y_0\right)\right] \approx \%\Delta\, y = 100\beta_2\left(x_1 - x_0\right) = \left(100\beta_2\right) \times \Delta x$$

Prediction in the Log-Linear Model, Part I

- In a log-linear regression, the R^2 value automatically reported by statistical software is the percent of the variation in ln(*y*) explained by the model
- However, our objective is to explain the variations in *y*, not ln(*y*)
- Furthermore, the fitted regression line predicts:
  - $\widehat{\ln(y)} = b_1 + b_2 x$
  - Whereas we want to predict *y*

Prediction in the Log-Linear Model, Part II

- A natural choice for prediction is:
  - $\hat{y}_n = \exp\left(\widehat{\ln(y)}\right) = \exp(b_1 + b_2 x)$
  - The subscript "*n*" is for "natural"
  - But a better alternative is:
    - $\hat{y}_c = \widehat{E(y)} = \exp(b_1 + b_2 x + \hat{\sigma}^2/2) = \hat{y}_n{}^{e\wedge\left(\hat{\sigma}^2/2\right)}$
    - The subscript "*c*" is for "corrected"
    - This uses the properties of the **log-normal distribution**

Prediction in the Log-Linear Model, Part III

- Recall that σ^2 must be greater than zero and $e^0 = 1$
  - Thus, the effect of the correction is always to increase the value of the prediction because $e^{(\hat{\sigma}^2/2)}$ is always greater than one
- The natural predictor tends to systematically underpredict the value of y in a log-linear model, and the correction offsets the downward bias in large samples

Example: Prediction in the Log-Linear Model

- The wage equation is:

  $\ln(\widehat{WAGE})$= 1.5968 + 0.0988 × EDUC = 1.5968 + 0.0988 × 12 = 2.7819

- The natural predictor is $\hat{y}_n = \exp\left(\widehat{\ln(y)}\right) = \exp(2.7819) = 16.1493$
- The corrected predictor is:

$$\hat{y}_c = \widehat{E(y)} = \exp(b_1 + b_2 x + \hat{\sigma}^2/2) = \hat{y}_n{}^{e^\wedge\left(\hat{\sigma}^2/2\right)}$$ = 16.1493 × 1.1246 = 18.1622

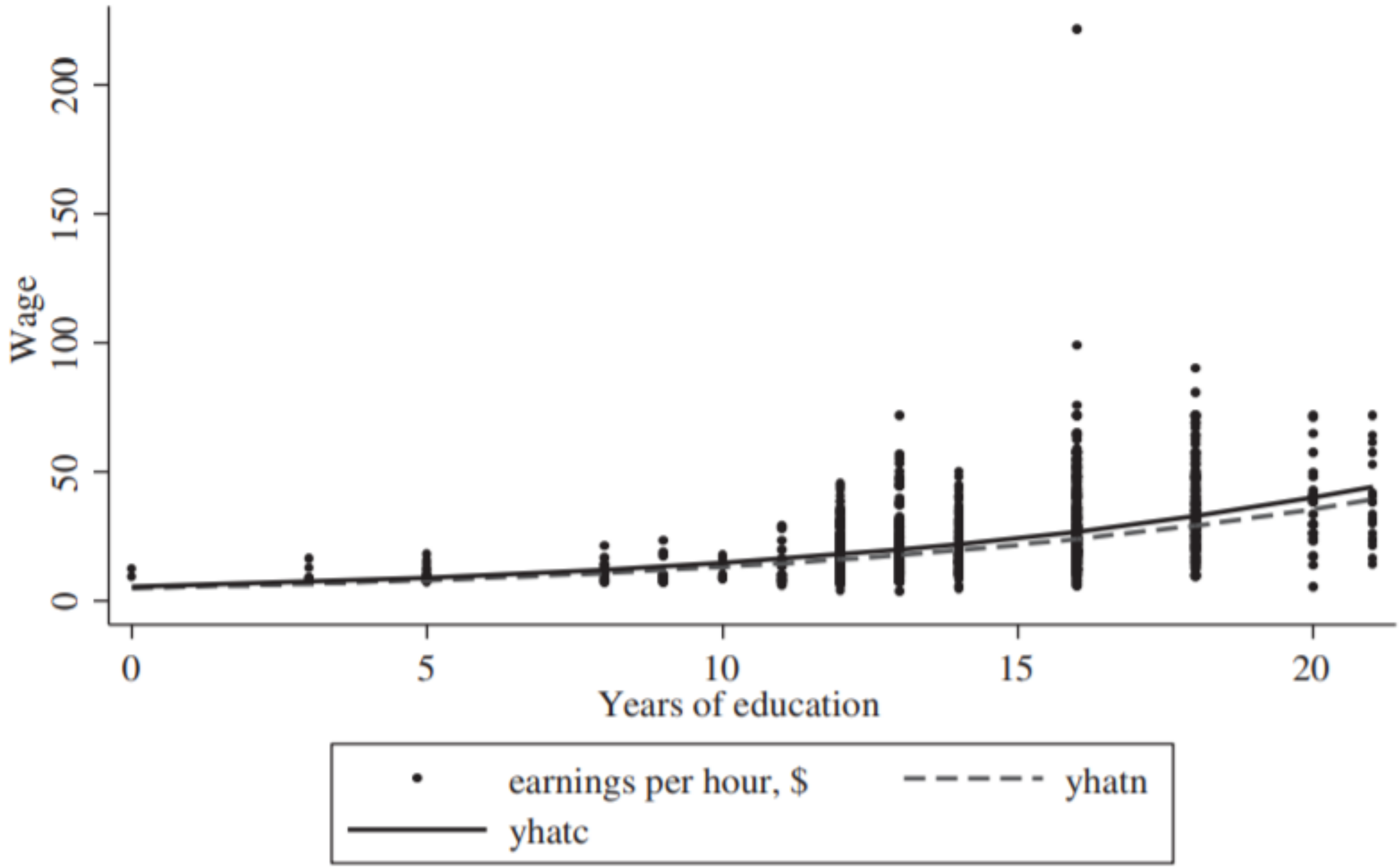


FIGURE 4.13 **The natural and corrected predictors of wage.**

FORDHAM THE JESUIT UNIVERSITY OF NEW YORK | Gabelli School of Business

A Generalized Measure

- A general goodness-of-fit measure, or general R^2, is:

$$R_g^2 = \left[\operatorname{corr}\left(y, \hat{y}\right)\right]^2 = r_{y\hat{y}}^2$$

- For the wage equation, the general R^2 is:

$$\mathrm{R}_{\mathrm{g}}^2 = \left[\operatorname{corr}\left(\mathrm{y}, \hat{\mathrm{y}}\right)\right]^2 = 0.4647^2 = 0.2159$$

- Compare this to the reported R^2 = 0.2577

Prediction Intervals in the Log-Linear Model

- If we prefer a prediction or forecast interval over a "point" predictor for y, then we must rely on the natural predictor y^n
- A 100(1 – α)% prediction interval for *y* is:

$$\left[\exp\left(\widehat{\ln(y)} - t_c se(f)\right), \exp\left(\widehat{\ln(y)} + t_c se(f)\right)\right]$$

Example: Prediction Intervals for a Log-Linear Model

- For the wage equation, a 95% prediction interval for the wage of a worker with 12 years of education is:
  - $[\exp(2.7819 - 1.96 \times 0.4850), \exp(2.7819 + 1.96 \times 0.4850)] = [6.2358, 41.8233]$
- The interval prediction is \$6.24–\$41.82, which is so wide that it is basically useless
- Our model is not an accurate predictor of individual behavior in this case

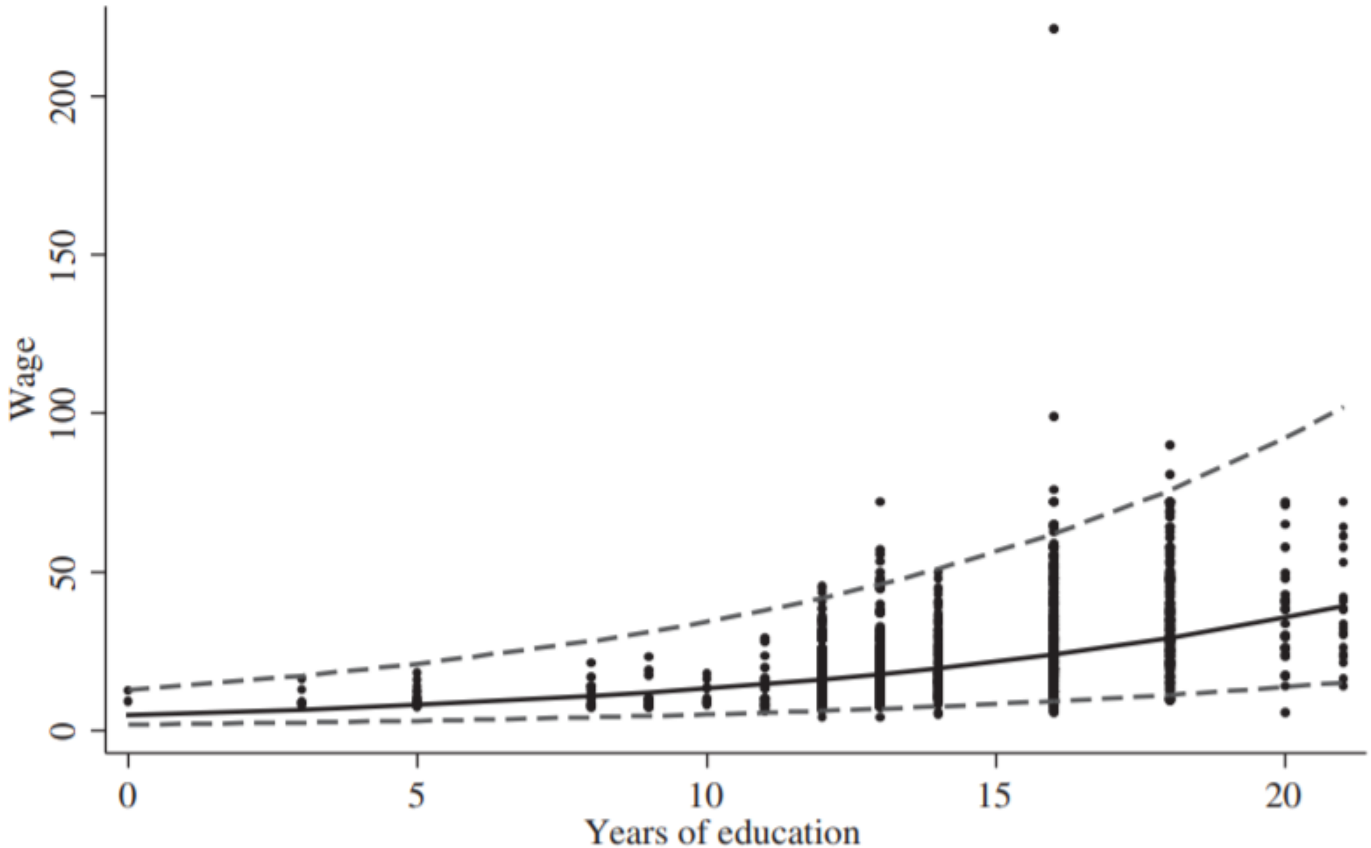


FIGURE 4.14 **The 95% prediction interval for wage.**

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Key Words

- Coefficient of determination
- Correlation
- Forecast error
- Functional form
- Goodness of fit
- Growth model
- Influential observations
- Jarque–Bera test
- Kurtosis
- Least squares predictor
- Linear model
- Linear relationship
- Linear-log model
- Log-linear model
- Log-log model
- Log-normal distribution
- Prediction
- Prediction interval
- R^2
- Residual diagnostics
- Scaling data
- Skewness
- Standard error of the forecast

Copyright

FORDHAM

THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business