I.4.3 The Log-linear Regressionmodel

The log-linear regression model is a nonlinear relation between Y and X:

$$Y = \tilde{\beta}_0 \cdot X^{\beta_1} \cdot e^u. \tag{19}$$

By taking the natural logarithm on both sides we obtain a linear (in the parameters) regression model for the transformed variables $\log Y$ and $\log X$, where $\beta_0 = \log \tilde{\beta}_0$:

$$\log Y = \beta_0 + \beta_1 \log X + u, \tag{20}$$

Visualizing the log-linear regression model

- Simulate data from a simple log-linear regression model with $\tilde{\beta}_0 = 0.2$ and $\beta_1 = -1.8$:

$$Y = 0.2 \cdot X^{-1.8} e^{u},$$

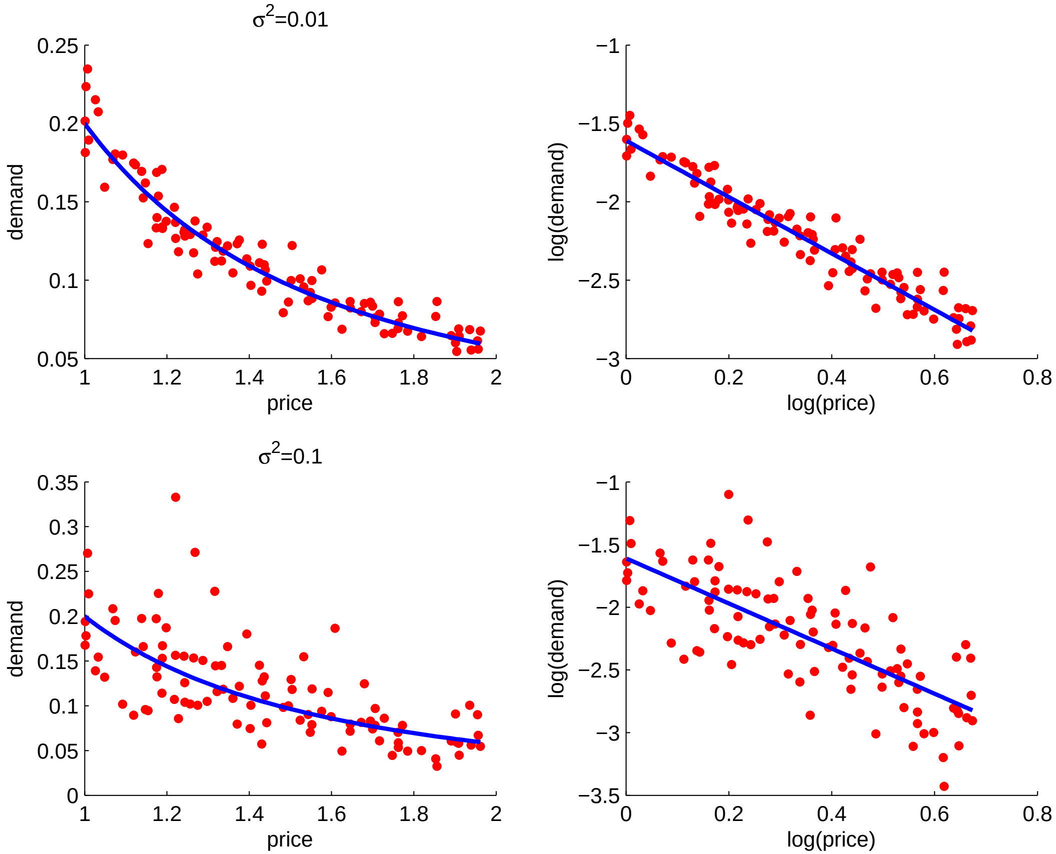
- Specification of the error term:

$$u \sim \mathrm{Normal}\left(0, \sigma^2\right)$$

- Demonstration $\Rightarrow$

MATLAB Code: `regsimlog.m`

Visualizing the log-linear regression model



Understanding the parameters

In economics, elasticity measures of how changing one variable affects other variables. If $y = f(x)$, then the elasticity is the ratio of the percentage change $\%\Delta y$ in y to the percentage change $\%\Delta x$ in the variable x:

$$\frac{\partial \log y}{\partial \log x} = \frac{\partial y}{y} \Big/ \frac{\partial x}{x} \approx \frac{\%\Delta y}{\%\Delta x}.$$

From equation (37) we obtain the following expected value of $\log Y$, if the predictor is equal to X:

$$\mathrm{E}(\log Y) = \beta_0 + \beta_1 \log X.$$

Therefore:

Understanding the parameters

$$E(\frac{\%\Delta Y}{\%\Delta X}) \approx \mathrm{E}(\frac{\partial \log Y}{\partial \log X}) = \frac{\partial \mathrm{E}(\log Y)}{\partial \log X} = \beta_1.$$

- The parameter β_1 is the expected change (in percent) of the response variable Y, if the predictor X is increased by 1% (**elasticity**).

- The sign of β_1 shows the direction of the expected change. If $\beta_1 = 0$, then a change in X has no influence on Y.

- If X is increased by $p\%$, then the expected change of Y is equal to $\beta_1 p\%$.

I.4.4 Statistical Properties of OLS Estimation

Econometric inference: learning from the data about the unknown parameter $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ in the regression model.

- Use the OLS estimator $\hat{\boldsymbol{\beta}}$ to learn about the regression parameter.

- Is this estimator equal to the true value?

- How large is the difference between the OLS estimator and the true parameter?

- Is there a better estimator than the OLS estimator?

Understanding the estimation problem

- Simulate data from a simple regression model with $\beta_0 = 0.2$ and $\beta_1 = -1.8$:

$$y_i = 0.2 - 1.8x_i + u_i, \quad u_i \sim \mathrm{Normal}\left(0, \sigma^2\right) \tag{21}$$

- Run OLS estimation to obtain $(\hat{\beta}_0, \hat{\beta}_1)$ and compare the estimated values with the true values $\beta_0 = 0.2$ and $\beta_1 = -1.8$.

- Repeat this experiment several times

- Demonstration $\Rightarrow$

MATLAB Code: `regestdemo.m`

Understanding the estimation problem

- Although we are estimating the true model, the OLS estimator differs from the true value.

- Many different data sets of size N may be generated by the same regression model due to the stochastic error term. The estimated parameters differ, as the sample mean, sample variance and correlation coefficient are different for each data set:

$$\hat{\beta}_1 = \frac{s_y}{s_x} r_{xy}, \qquad \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}.$$

What is the expected error of the OLS estimator?

Obviously, the estimators are random variables. Hence it makes sense to study the statistical properties of OLS estimation:

- Are the OLS estimates unbiased, i.e. is the expected difference between the OLS estimator and the true parameter equal to 0?
- How precise are these parameter estimates, i.e., how large is the variance of the two estimators?
- Are the OLS coefficients correlated?
- How are the OLS coefficients distributed?

What is the expected error of the OLS estimator?

For fixed values of $x_1, \ldots, x_N$ the sampling properties of $\overline{y}$, s_y^2 and r_{xy} determine the estimation error. In general, the estimation error

- decreases within increasing number N of observations $\Rightarrow$

MATLAB Code: `regestall.m`

- increases within increasing variance σ^2 $\Rightarrow$

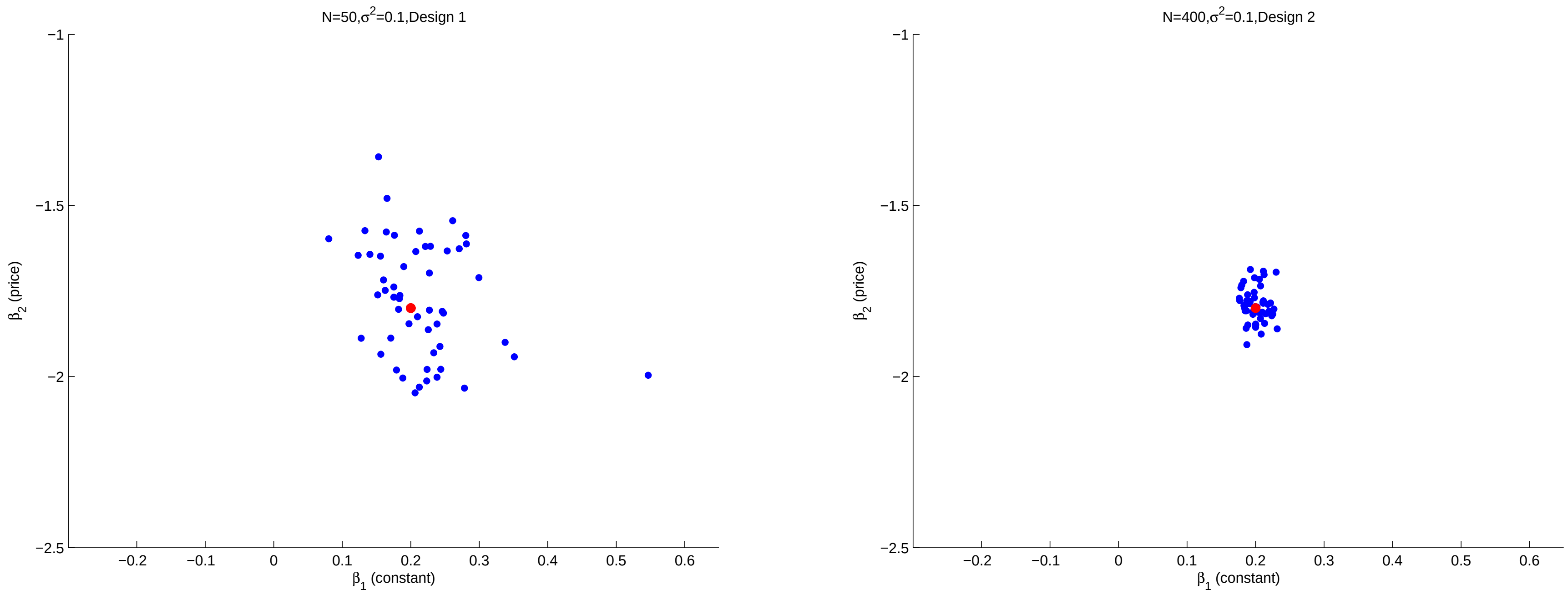
MATLAB Code: `regestall.m`

- depends on the predictor variables through s_x^2 $\Rightarrow$

MATLAB Code: `regestall.m`

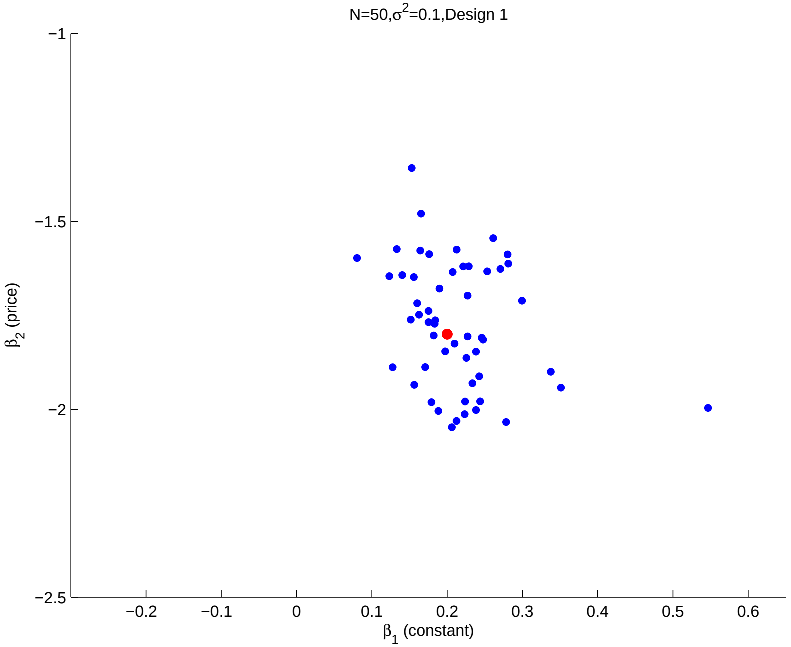
What is the expected error of the OLS estimator?

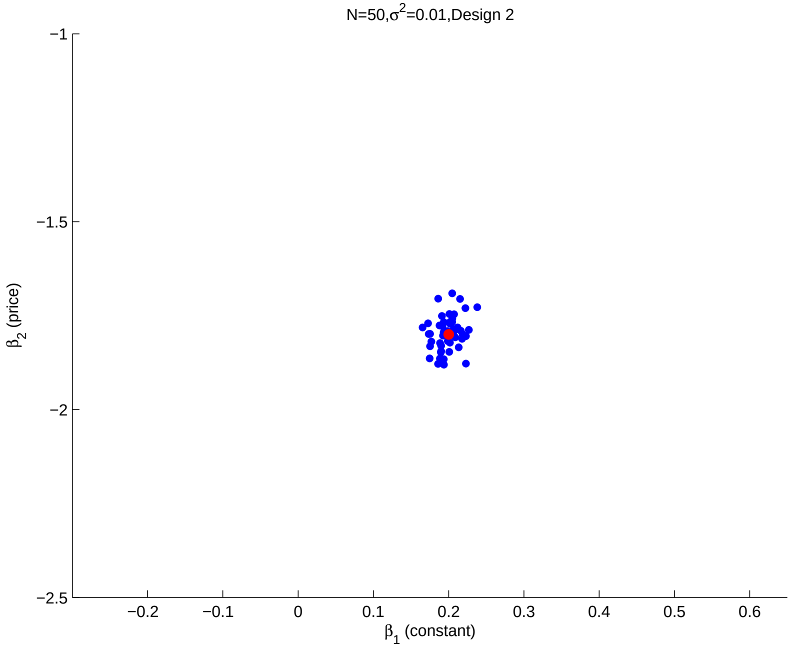
$N = 50$ versus $N = 400$ ($\sigma^2 = 0.1$, Design 1)



What is the expected error of the OLS estimator?

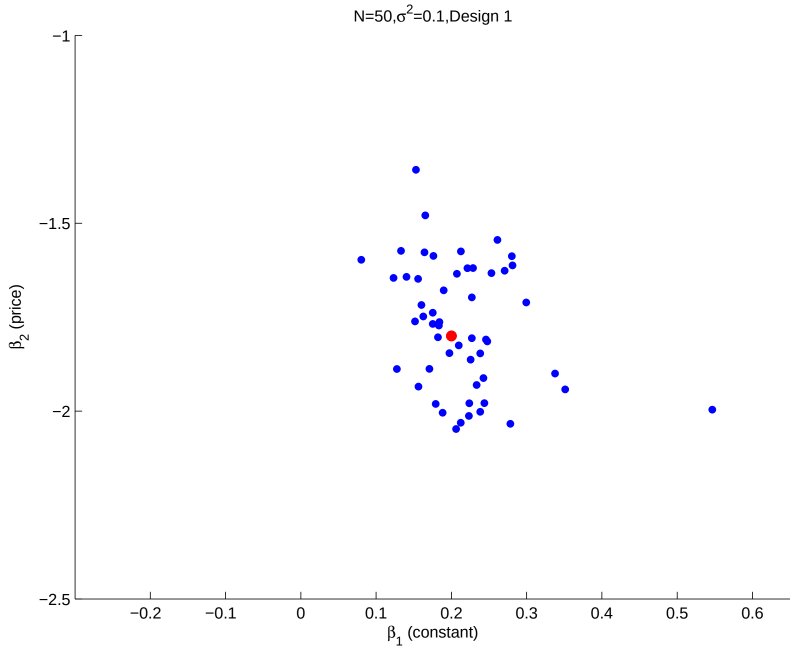
$\sigma^2 = 0.1$ versus $\sigma^2 = 0.01$ ($N = 50$, Design 1)

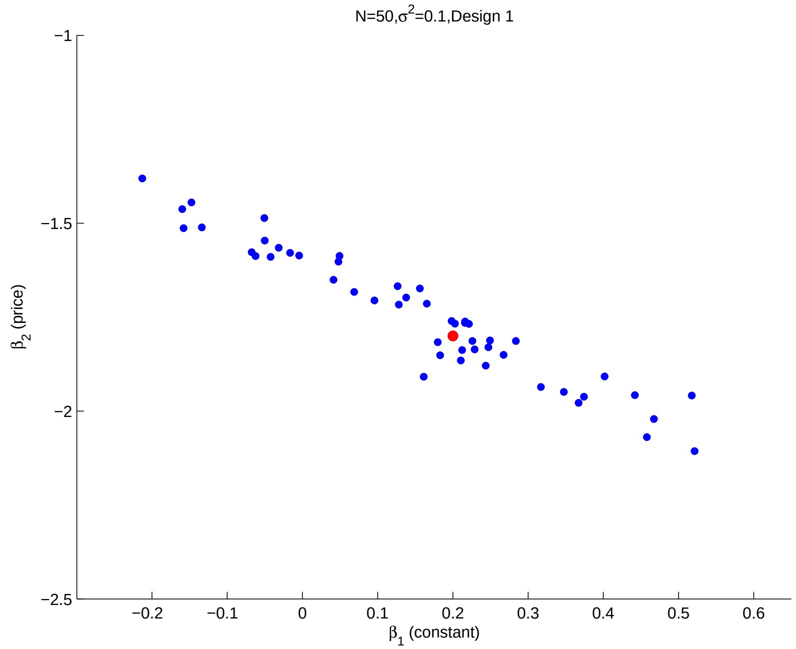




What is the expected error of the OLS estimator?

Design 1: $x_i \sim -.5 + \text{Uniform}\,[0,1]$ (left hand side) versus Design 2: $x_i \sim 1 + \text{Uniform}\,[0,1]$ ($N = 50$, $\sigma^2 = 0.1$) (right hand side)





Unbiasedness

Under assumption (5) and (6), the OLS estimator is unbiased, i.e. on average the estimated value is equal to the true one:

$$\mathrm{E}(\hat{\beta}_1) = \beta_1, \qquad \mathrm{E}(\hat{\beta}_0) = \beta_0.$$

Relation between $\hat{\beta}_1$ and β_1:

$$\hat{\beta}_1 = \beta_1 + \frac{1}{N s_x^2} \sum_{i=1}^{N} (x_i - \overline{x}) u_i. \tag{22}$$

Unbiasedness

Proof of (22). Note that $\sum_{i=1}^{N}(x_i - \overline{x}) = 0$. Therefore:

$$\hat{\beta}_1 = \frac{1}{Ns_x^2}\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{x}) = \frac{1}{Ns_x^2}\sum_{i=1}^{N}(x_i - \overline{x})y_i.$$

Use that $y_i = \beta_0 + \beta_1 x_i + u_i$ is generated from model (4) and plug in:

$$\hat{\beta}_1 = \frac{\beta_0}{Ns_x^2}\sum_{i=1}^{N}(x_i - \overline{x}) + \frac{\beta_1}{Ns_x^2}\sum_{i=1}^{N}(x_i - \overline{x})x_i + \frac{1}{Ns_x^2}\sum_{i=1}^{N}(x_i - \overline{x})u_i$$

$$= \beta_1 + \frac{1}{Ns_x^2}\sum_{i=1}^{N}(x_i - \overline{x})u_i.$$

Unbiasedness

Since $\mathrm{E}(u_i) = 0$, we obtain from (22) that $\mathrm{E}(\hat{\beta}_1) = \beta_1$:

$$\mathrm{E}(\hat{\beta}_1) = \beta_1 + \frac{1}{N s_x^2} \sum_{i=1}^{N} (x_i - \overline{x}) \mathrm{E}(u_i) = \beta_1.$$

Hence $\mathrm{E}(\beta_1 - \hat{\beta}_1) = 0$, we obtain $\mathrm{E}(\hat{\beta}_0) = \beta_0$:

$$\hat{\beta}_0 = \beta_0 + (\beta_1 - \hat{\beta}_1)\overline{x} + \frac{1}{N} \sum_{i=1}^{N} u_i,$$

$$\Rightarrow \mathrm{E}(\hat{\beta}_0) = \beta_0 + \mathrm{E}(\beta_1 - \hat{\beta}_1)\overline{x} + \frac{1}{N} \sum_{i=1}^{N} \mathrm{E}(u_i) = \beta_0.$$