2.6 Influential Observations and Leverage

One important feature of OLS estimation, which we have not stressed up to this point, is that each element of the vector of parameter estimates $\hat{\boldsymbol{\beta}}$ is simply a weighted average of the elements of the vector $\boldsymbol{y}$. To see this, define $\boldsymbol{c}_i$ as the i^{th} row of the matrix $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ and observe from (2.02) that $\hat{\beta}_i = \boldsymbol{c}_i\boldsymbol{y}$. This fact will prove to be of great importance when we discuss the statistical properties of least-squares estimation in the next chapter.

Because each element of $\hat{\boldsymbol{\beta}}$ is a weighted average, some observations may affect the value of $\hat{\boldsymbol{\beta}}$ much more than others do. Consider Figure 2.14. This figure is an example of a **scatter diagram**, a long-established way of graphing the relation between two variables. Each point in the figure has Cartesian coordinates (x_t, y_t), where x_t is a typical element of a vector $\boldsymbol{x}$, and y_t of a vector $\boldsymbol{y}$. One point, drawn with a larger dot than the rest, is indicated, for reasons to be explained, as a high leverage point. Suppose that we run the regression

$$\boldsymbol{y} = \beta_1\boldsymbol{\iota} + \beta_2\boldsymbol{x} + \boldsymbol{u}$$

twice, once with, and once without, the high leverage observation. For each regression, the fitted values all lie on the so-called **regression line**, which is the straight line with equation

$$y = \hat{\beta}_1 + \hat{\beta}_2 x.$$

The slope of this line is just $\hat{\beta}_2$, which is why β_2 is sometimes called the **slope coefficient**; see Section 1.1. Similarly, because $\hat{\beta}_1$ is the intercept that the

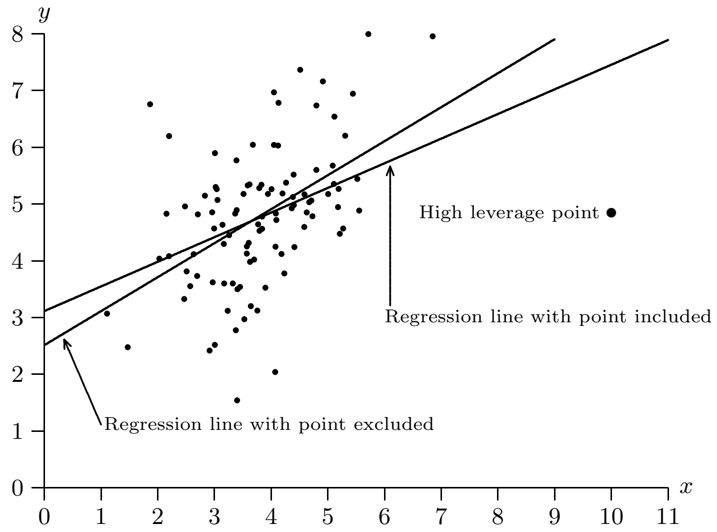

Figure 2.14 An influential observation