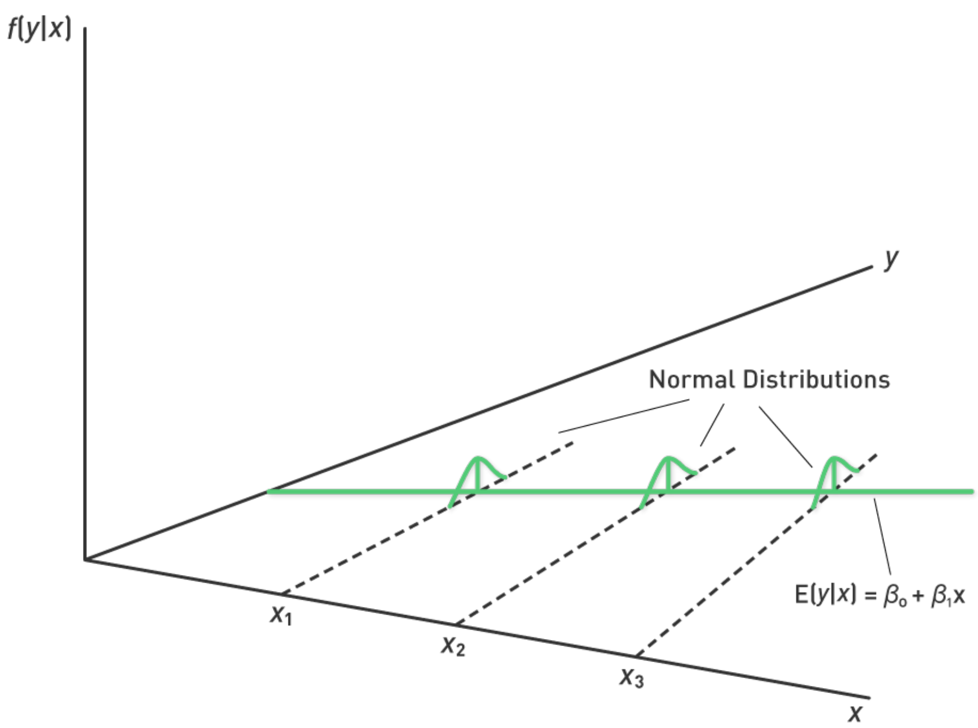
## Normal Error Term Assumption

To infer the sampling distribution of our OLS coefficients, we can add an assumption about the shape of the error distribution.

- **Assumption MLR.6 (Normality of error terms)**
    - $u_i \sim N(0, \sigma^2)$
      independently of $x_{i1}, x_{i2}, ..., x_{ik}$
- Assume the errors are drawn from a normal distribution with mean zero.
- Also assume the errors are independent of our *x*s, so the distribution looks the same conditional on any values of the *x*s.



## Theory and Practice

- How realistic is the normality assumption?
- There is a purely theoretical argument that the error term is the sum of many different unobserved factors.
    - Sums of many independent factors are normally distributed by a version of the central limit theorem.
- Some problems:
    - How many different factors? Is the number great enough?
    - What if a few factors are more influential than others?
    - How independent are the different factors?

## Theory and Practice (cont.)

- In practice, the residuals are often not normal at all.
    - E.g., if there is a highly skewed $y$ variable, the errors are often skewed as well.
- We will soon discuss how to test for normality.
- For now, bear in mind that this is a rather strong assumption.

## Classical Linear Model Assumptions

- When normality is added to the five Gauss-Markov assumptions, the resulting collection is known as the classical linear model (CLM).
  - MLR.1–MLR.5: Gauss-Markov assumptions
  - MLR.1–MLR.6: classical linear model (CLM) assumptions
- **Theorem 4.1 (Normal sampling distributions) states:**
  - Under assumptions MLR.1–MLR.6, the OLS coefficients are normally distributed.
  - $\hat{\beta}_j \sim N(\beta_j, \mathrm{Var}\left(\hat{\beta}_j\right))$
- Each $\hat{\beta}_j$ is normally distributed around the true parameter.
  - We already calculated the variance of $\hat{\beta}_j$ earlier, so we know the exact distribution.

## Standardizing the Distribution

- How can normality be used to test hypotheses?
- First, normalize the estimator by subtracting its mean and dividing by its standard deviation.
  - $\dfrac{\hat{\beta}_j - \beta_j}{sd\left(\hat{\beta}_j\right)} \sim N(0, 1)$
- This provides the standard normal distribution.
- No matter what the true parameters are, we always get the exact same distribution.

## Standardizing the Distribution (cont.)

- In practice, we don't know the standard deviation to put in the denominator.
  - It has to be estimated using the standard error of the sample.
- This changes the normal distribution to a *t*-distribution.

$$\frac{\hat{\beta}_j - \hat{\beta}_j}{se\left(\hat{\beta}_j\right)} \sim t_{\mathrm{n-k-1}}$$

- The *t*-distribution is similar to a normal distribution, but the process of estimation introduces more variance and makes the tails of the distribution heavier.

## Formulating a Null Hypothesis

- Most of the time, our null will be that the population parameter is equal to zero.
  - Controlling for all other independent variables, there is no effect of $x_j$ on *y*.
- $H_0 : \hat{\beta}_j = 0$
- We are working in the frequentist framework.
  - Our null hypothesis is specific enough (given our population model) to identify the distribution of our standardized coefficient.
  - $\frac{\hat{\beta}_j}{se\left(\hat{\beta}_j\right)} \sim t_{\mathrm{n-k-1}}$
- It is distributed according to a *t*-distribution and so it's called a *t*-statistic.

## The *t*-Statistic

- We collect our data and compute our OLS estimate, $\hat{\beta}_j$
- The farther our estimate is from zero, the more evidence we have against our null hypothesis, but we have to normalize by a measure of variability.
- We therefore divide by an estimate of standard deviation to get our *t*-statistic,
  - $t_{\hat{\beta}_j} \frac{\hat{\beta}_j}{se\left(\hat{\beta}_j\right)} \sim t_{\mathrm{n-k-1}}$
- The *t*-statistic measures how many estimated standard deviations the estimated coefficient is away from zero.
- It is distributed according to a *t*-distribution, so we can talk about how significant the difference is.