Intermediate Econometrics

Multiple Regression Analysis

OLS Asymptotics (updated)

Text: Wooldridge, Chapter 5 and Appendix C

July 10, 2011

Introduction and Motivation for Asymptotic Analysis

- We have seen that, under the Gauß–Markov Assumptions, the OLS estimator is unbiased and efficient (BLUE), with covariance matrix

$$\text{Cov}(\widehat{\beta}) = \sigma^2 (X'X)^{-1}.$$

- Moreover, hypothesis tests (t and F tests) have been developed under the assumption of normally distributed errors.

- However, in many situations, these assumptions will be unrealistic:
  - For many types of data (e.g., time series data), the random sampling assumption and thus perhaps strong exogeneity ($\text{E}(u_i|\boldsymbol{X}) = 0$) will be violated, and OLS may be biased.
  - The assumption of homoskedasticity is unrealistic even in many cross–sectional situations, so our simple formulas for the variances are no longer valid.

  - There is also nothing that guarantees normally distributed errors. In fact, economic data often show pronounced deviations from the normal distribution.
  - If errors are not normally distributed, then the OLS estimator will likewise not be normally distributed, which in turn implies that the t and F statistics will not have t and F distributions, respectively.

- Obtaining the exact distributions of $\widehat{\beta}$ and the test statistics is in general not feasible if the classical assumptions do not hold.

- Asymptotic or large sample theory allows to derive *approximate* properties of estimators by assuming that the sample size n "is large".

- Use of estimators and test statistics may then be justified by referring to favorable large sample properties.

Convergence in Probability and Consistency

•

- We consider a sequence of random variables θ_n indexed by n, $n \in \mathbb{N}$ (in our applications, n is the sample size).

- Sequence θ_n is said to *converge to a constant* θ *in probability* if

$$\lim_{n\to\infty} \Pr(|\theta_n - \theta| \geq \epsilon) = 0 \quad \text{for any } \epsilon > 0, \tag{1}$$

  or equivalently

$$\lim_{n\to\infty} \Pr(|\theta_n - \theta| < \epsilon) = 1 \quad \text{for any } \epsilon > 0. \tag{2}$$

- (1) and (2) essentially state that, for large n, the probability is high that θ_n will be close to θ.

- Convergence in probability is also denoted as

$$\theta_n \xrightarrow{p} \theta,$$

or

$$\text{plim}_{n\to\infty}\theta_n = \theta,$$

where "plim" is short for "probability limit".

- Now suppose that θ is a population parameter and $\widehat{\theta}_n$ is a sequence of estimators of θ.

- We say that $\widehat{\theta}_n$ is a **consistent** estimator of θ if

$$\widehat{\theta}_n \xrightarrow{p} \theta. \tag{3}$$

- That is, via increasing the sample size, $\widehat{\theta}_n$ can be made arbitrarily close to θ with arbitrarily high probability.

- As n increases, the sampling distribution of $\widehat{\theta}_n$ becomes more concentrated about θ, i.e., $\widehat{\theta}_n$ is less likely to be very far from θ.

- It turns out that a sufficient condition for consistency of an unbiased estimator is that its variance converges to zero as $n \to \infty$.

- To see this, recall Tschebyschev's inequality: For random variable x with density $f(x)$, a function $g(x) \geq 0$, and a constant $a > 0$,

$$
\begin{aligned}
\mathsf{E}(g(x)) &= \int_{-\infty}^{\infty} g(x) f(x) dx \\
&= \int_{g(x)<a} g(x) f(x) dx + \int_{g(x)\geq a} g(x) f(x) dx \\
&\geq \int_{g(x)\geq a} g(x) f(x) dx \\
&\geq a \int_{g(x)\geq a} f(x) dx = a\mathsf{Pr}(g(x) \geq a).
\end{aligned}
$$

- That is

$$
\mathsf{Pr}(g(x) \geq a) \leq \frac{\mathsf{E}(g(x))}{a}. \tag{4}
$$

- Now put $g(x) = (x - \mu_x)^2$, where $\mu_x = \mathsf{E}(x)$, then

$$\mathsf{E}(g(x)) = \mathsf{E}((x - \mu_x)^2) = \mathsf{Var}(x) =: \sigma_x^2,$$

and, for all $\epsilon > 0$,

$$\mathsf{Pr}((x - \mu_x)^2 \geq \epsilon^2) = \mathsf{Pr}(|x - \mu_x| \geq \epsilon) \leq \frac{\sigma_x^2}{\epsilon^2}. \tag{5}$$

- Relation (5) can be used to derive a (weak) *law of large numbers* (LLN):

- Let $x_1, \ldots, x_n$ be independently and identically distributed with mean $\mathsf{E}(x_i) = \mu_x$ and variance $\sigma^2 < \infty$. Then

$$\frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}_n \xrightarrow{p} \mu_x. \tag{6}$$

- This follows from (5),

$$\mathsf{E}(\overline{x}_n) = \mathsf{E}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_i \mathsf{E}(x_i) = \mu_x,$$

and

$$\begin{aligned}
\mathsf{Var}(\overline{x}_n) &= \mathsf{Var}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \\
&= \frac{1}{n^2}\sum_i \mathsf{Var}(x_i) + \frac{1}{n^2}\sum_i\sum_{j\neq i} \underbrace{\mathsf{Cov}(x_i, x_j)}_{=0} \\
&= \frac{1}{n^2}\sum_i \underbrace{\mathsf{Var}(x_i)}_{=\sigma^2} = \frac{\sigma^2}{n}.
\end{aligned}$$

- That is, the variance of the sample average shrinks to zero as the sample size increases.

Properties of plim and consistency of OLS

- To investigate the properties of the OLS estimator, the following properties of the probability limit are useful:

- If x_n and y_n are two sequences of random variables satisfying

$$x_n \xrightarrow{p} x, \quad \text{and} \quad y_n \xrightarrow{p} y, \tag{7}$$

then

$$x_n + y_n \xrightarrow{p} x + y, \tag{8}$$

$$x_n y_n \xrightarrow{p} xy \tag{9}$$

$$\frac{x_n}{y_n} \xrightarrow{p} \frac{x}{y} \quad \text{provided } y \neq 0. \tag{10}$$

- We want to show that the OLS estimator is consistent, i.e.,

$$\widehat{\beta}_j \xrightarrow{p} \beta_j, \tag{11}$$

where for simplicity we abstain from explicitly indexing $\widehat{\beta}_j$ by n.

- Let us consider the simple linear model

$$y = \beta_0 + \beta_1 x + u.$$

- Instead of the random sampling assumption, which implies strong exogeneity:

$$\mathsf{E}(u_i|\boldsymbol{X}) = 0, \tag{12}$$

  we only need the weaker assumption that x_i and u_i are uncorrelated in the population,

$$\mathsf{Cov}(x_i, u_i) = \mathsf{E}(x_i u_i) = 0. \tag{13}$$

- Note that strong exogeneity of the regressors, i.e.,

$$\mathsf{E}(u_i|\boldsymbol{X}) = 0, \tag{14}$$

  implies uncorrelatedness, i.e., (13).

- The converse is not true, however, i.e,

$$\mathsf{Cov}(x_i, u_i) = 0 \nRightarrow \mathsf{E}(u_i|\boldsymbol{X}) = 0. \tag{15}$$

- Thus, if only (13) is assumed, OLS is biased in general but consistent.

- To see consistency of the slope coefficient, we write

$$\widehat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_i (x_i - \overline{x}) u_i}{\frac{1}{n}\sum_i (x_i - \overline{x})^2}. \tag{16}$$

- The numerator and denominator of (16) are the sample covariance between x and u and the sample variance of x, respectively.

- By invoking the law of large numbers, it turns out that

$$\begin{aligned} \frac{1}{n}\sum_i (x_i - \overline{x}) u_i &\overset{p}{\to} \text{Cov}(x, u) = 0 \\ \frac{1}{n}\sum_i (x_i - \overline{x})^2 &\overset{p}{\to} \text{Var}(x). \end{aligned} \tag{17}$$

- Thus, by (8) and (10),

$$\text{plim}_{n\to\infty}\widehat{\beta}_1 = \beta_1 + \frac{\text{Cov}(x,u)}{\text{Var}(x)} = \beta_1 + \frac{0}{\text{Var}(x)} = \beta_1. \tag{18}$$

- Consistency can also be established for the OLS parameter vector in the multivariate model.

- The condition for this to be the case is a generalization of (13), namely

$$\text{Cov}(x_{i1}, u_i) = \text{Cov}(x_{i2}, u_i) = \cdots = \text{Cov}(x_{ik}, u_i) = 0.$$

OLS is inconsistent when Cov$(x_j, u) \neq 0$

- We have seen that OLS is biased in case there is correlation between the independent variables and the error term (e.g., omitted variable bias).

- It is important to note that this bias is **not** a small sample–phenomenon, i.e., it does not disappear as the sample size grows.

- For example, in the simple linear regression model, $y = \beta_0 + \beta_1 x + u$, we have

$$\widehat{\beta}_1 = \beta_1 + \frac{n^{-1}\sum_i (x_i - \bar{x}) u_i}{n^{-1}\sum_i (x_i - \bar{x})^2} = \beta_1 + \frac{s_{xu}}{s_x^2}.$$

- As the sample size grows, by the law of large numbers,

$$\text{plim}_{n\to\infty} s_{xu} = \text{Cov}(x, u), \quad \text{plim}_{n\to\infty} s_x^2 = \text{Var}(x),$$

so the asymptotic bias is

$$\text{plim}_{n\to\infty} \widehat{\beta}_1 - \beta_1 = \frac{\text{Cov}(x, u)}{\text{Var}(x)}.$$

- To illustrate further, consider model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,$$

where u satisfies the Gauß–Markov assumptions.

- If we estimate model

$$y = \beta_0 + \beta_1 x_1 + \tilde{u}$$

instead, we have

$$\tilde{u} = u + \beta_2 x_2,$$

and

$$\begin{aligned} \text{Cov}(x_1, \tilde{u}) &= \text{Cov}\{x_1, (u + \beta_2 x_2)\} \\ &= \underbrace{\text{Cov}(x_1, u)}_{=0} + \beta_2 \text{Cov}(x_1, x_2) = \beta_2 \text{Cov}(x_1, x_2). \end{aligned}$$

- Hence, the estimator

$$\widehat{\beta}_1 = \beta_1 + \frac{n^{-1}\sum_i (x_{i1} - \overline{x}_1)\tilde{u}_i}{n^{-1}\sum_i (x_{i1} - \overline{x}_1)^2}$$

has probability limit

$$\widehat{\beta}_1 \xrightarrow{p} \beta_1 + \frac{\mathsf{Cov}(x_1, \tilde{u})}{\mathsf{Var}(x_1)} = \beta_1 + \beta_2 \frac{\mathsf{Cov}(x_1, x_2)}{\mathsf{Var}(x_1)},$$

and is consistent only if either $\beta_2 = 0$ or $\mathsf{Cov}(x_1, x_2) = 0$.

- The asymptotic bias is

$$\mathsf{plim}_{n\to\infty} \widehat{\beta}_1 - \beta_1 = \beta_2 \delta_1,$$

where

$$\delta_1 = \frac{\mathsf{Cov}(x_1, x_2)}{\mathsf{Var}(x_1)} = \mathsf{plim}_{n\to\infty} \frac{s_{x_1,x_2}}{s^2_{x_1}},$$

cf. Equation (39) of Lecture Slides 2.

- With regard to the direction of the asymptotic bias, the same reasoning applies as in the previous discussion of the omitted variable bias.

Asymptotic Normality

- The testing procedures we have considered so far have been developed under the assumption of normally distributed errors.

- However, we cannot expect this to be true in general.

- Indeed, economic data often show pronounced deviations from the normal distribution.

- This is an important issue, since if the errors are not normal, then the OLS estimator will not be normal, which in turn implies that the t and F statistics will not follow t and F distributions, respectively.

- We have data on arrests during 1986 (and further variables) for $n = 2725$ men born in either 1960 or 1961 in California.[1]

- Each man in the sample was arrested at least once prior to 1986.

- A linear model explaining arrests is

$$\begin{aligned} narr86 &= \beta_0 + \beta_1 \cdot pcnv + \beta_2 \cdot avgsen + \beta_3 \cdot tottime \\ &\quad + \beta_4 \cdot ptime86 + \beta_5 \cdot qemp86 + u, \end{aligned}$$

  where

  - *narr86* is the number of times a man was arrested
  - *pcnv* is the proportion of prior arrests leading to conviction
  - *avgsen* is average sentence served from past convictions
  - *tottime* is total time the man has spent in prison prior to 1986
  - *ptime86* is months spent in prison in 1986
  - *qemp86* is number of quarters in 1986 during which the man was (legally) employed

[1] Cf. Example 3.5 of Wooldridge.

- As a rough check of normality, we can look at a plot of the regression residuals,

$$\begin{aligned}\widehat{\boldsymbol{u}} &= \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}) \\ &= (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{u} \\ &= \boldsymbol{M}\boldsymbol{u},\end{aligned}$$

  where $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, $\boldsymbol{M}^2 = \boldsymbol{M}$ ($\boldsymbol{M}$ is *idempotent*).

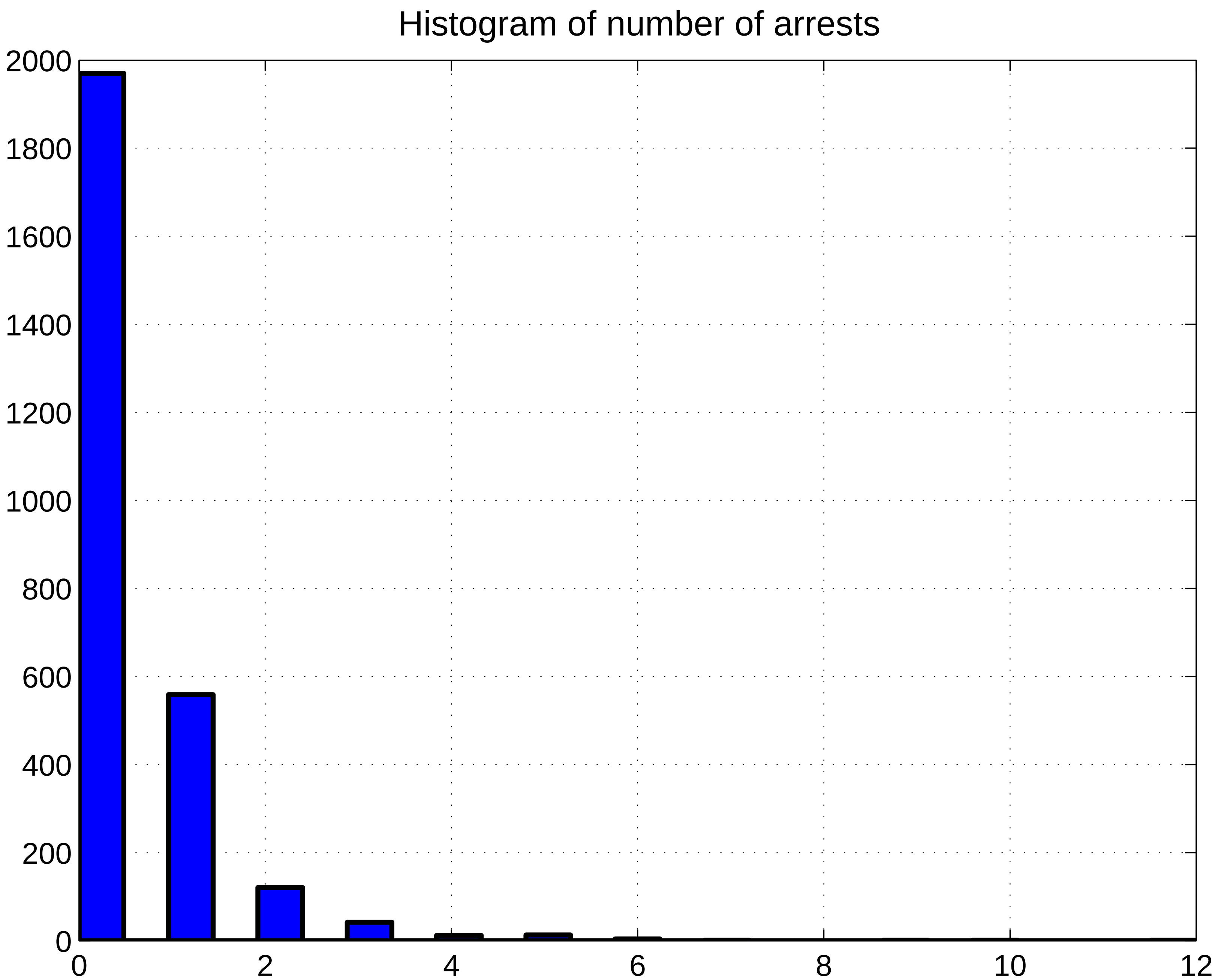
- Hence normality of u implies that $\widehat{\boldsymbol{u}}$ is also normal with mean zero and covariance matrix $\sigma^2\boldsymbol{M}$.

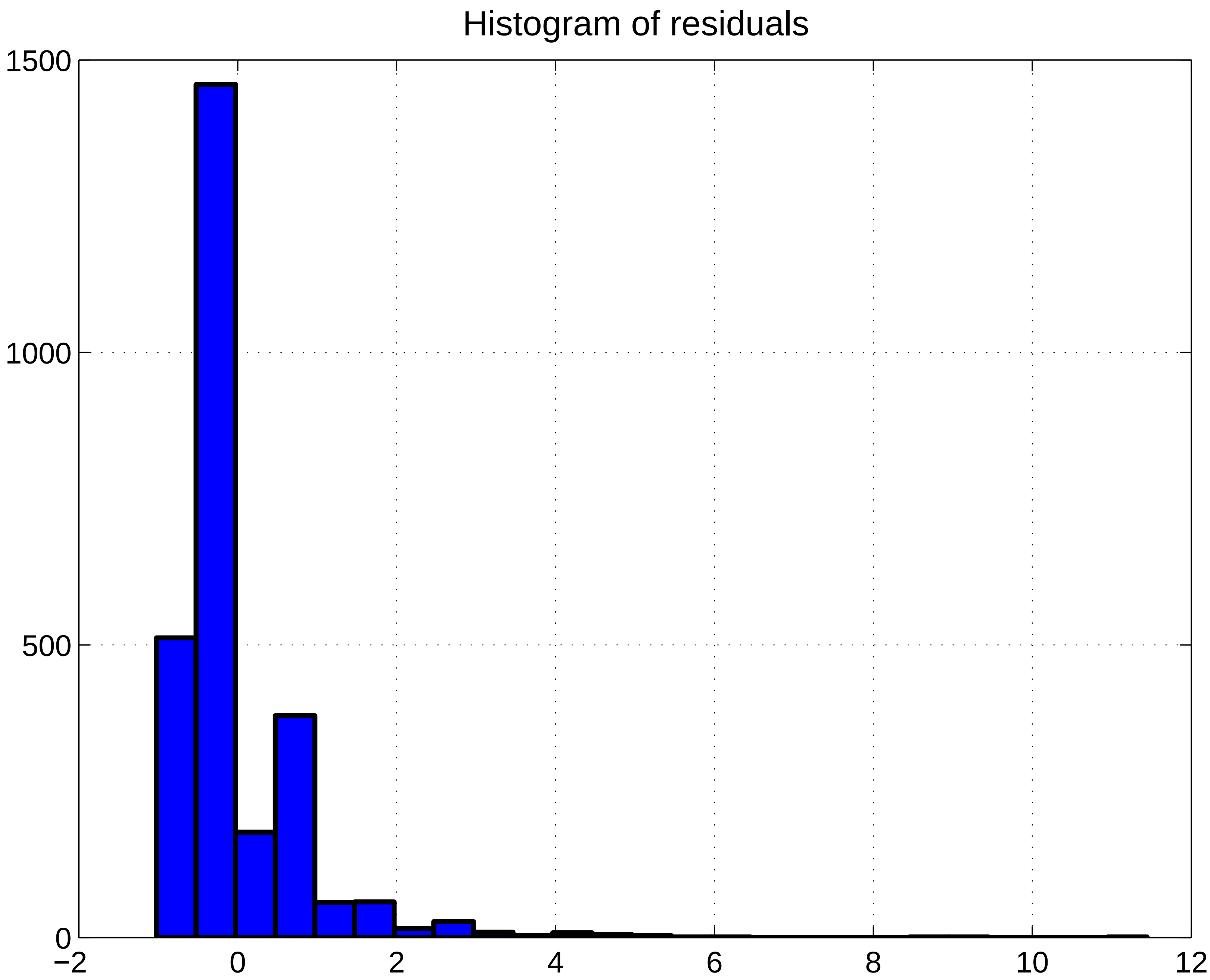
- Hence, quantities

$$\check{u}_i = \frac{\widehat{u}_i}{\sqrt{m_{ii}}}, \quad i = 1, \ldots, n, \tag{19}$$

  where m_{ii} is the ith diagonal element of $\boldsymbol{M}$, should have an identical normal distribution.[2]

[2] Although they are not iid even if the original errors are iid. However, since $(\boldsymbol{X}'\boldsymbol{X})^{-1} \stackrel{n\to\infty}{\longrightarrow} \boldsymbol{0}$, we have $\boldsymbol{M} \stackrel{n\to\infty}{\longrightarrow} \boldsymbol{I}$, so that, asymptotically, $\boldsymbol{u}$ and $\widehat{\boldsymbol{u}}$ have the same distribution.


Histogram of number of arrests
2000
1800
1600
1400
1200
1000
800
600
400
200
0
0
2
4
6
8
10
12


Histogram of residuals
1500
1000
500
0
−2
0
2
4
6
8
10
12

Convergence in distribution and asymptotic normality

- We consider a sequence of random variables x_n with an associated sequence of cumulative distribution functions (cdf) F_n,

$$F_n(y) = \Pr(x_n \leq y). \tag{20}$$

- If x is a random variable with cdf F, then we say that x_n *converges in distribution to* x, denoted as

$$x_n \xrightarrow{d} x, \quad \text{or} \quad x_n \xrightarrow{d} F, \tag{21}$$

  if

$$\lim_{n\to\infty} F_n(y) = F(y) \tag{22}$$

  at all points y where F is continuous.

- The practical usefulness of the concept of convergence in distribution lies in establishing an approximation to the true distribution (for n "large enough") when the true distribution is unknown or intractable.

- One of the most important examples of convergence in distribution are central limit theorems (CLT).

Central Limit Theorem

- We have seen that the distribution of the sample mean $\overline{x}_n$ depends on n.
- In particular, its variance is σ^2/n and shrinks to zero as $n \to \infty$, i.e., the limiting distribution has all its mass concentrated in one point.
- This is clearly not useful as an approximative distribution of $\overline{x}_n$ for large n.
- Obviously, we need to standardize $\overline{x}_n$ in order to "fix" its variance.
- Recall the general rule for random variable x and constant c,

$$\text{Var}(cx) = c^2\text{Var}(x). \tag{23}$$

- Then since

$$\mathsf{Var}\left(\sqrt{n}\overline{x}_n\right) = n\mathsf{Var}\left(\overline{x}_n\right) = n(\sigma^2/n) = \sigma^2, \tag{24}$$

we study the behavior of the standardized variable

$$\sqrt{n}\frac{\overline{x}_n - \mu}{\sigma} = \frac{\overline{x}_n - \mu}{\sigma/\sqrt{n}}. \tag{25}$$

- The classical central limit theorem (CLT) is as follows:

- Let $x_1, \ldots, x_n$ be independently and identically distributed with $\mathsf{E}(x_i) = \mu$ and $\mathsf{var}(x_i) = \sigma^2 < \infty$. Then

$$\sqrt{n}\frac{\overline{x}_n - \mu}{\sigma} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{x_i - \mu}{\sigma} \xrightarrow{d} \mathsf{N}(0, 1), \tag{26}$$

or equivalently

$$\sqrt{n}(\overline{x}_n - \mu) \xrightarrow{d} \mathsf{N}(0, \sigma^2). \tag{27}$$

- This is a remarkable result as it states that the mean of any long sequence of iid variables is approximately normally distributed, *no matter what their distribution* (provided the variance is finite).

- When applying this result, we treat $\overline{x}_n$ as approximately normal with mean μ and variance σ^2/n, written as

$$\overline{x}_n \overset{a}{\sim} \mathsf{N}(\mu, \sigma^2/n), \tag{28}$$

where "$\overset{a}{\sim}$" stands for "approximately in large samples".

CLT illustration

- As an example for the CLT, we consider a uniform distribution over the interval $[0, 1]$, i.e., a random variable x with density function

$$f(x) = \begin{cases} 1 & x \in [0, 1] \\ 0 & \text{otherwise.} \end{cases} \tag{29}$$

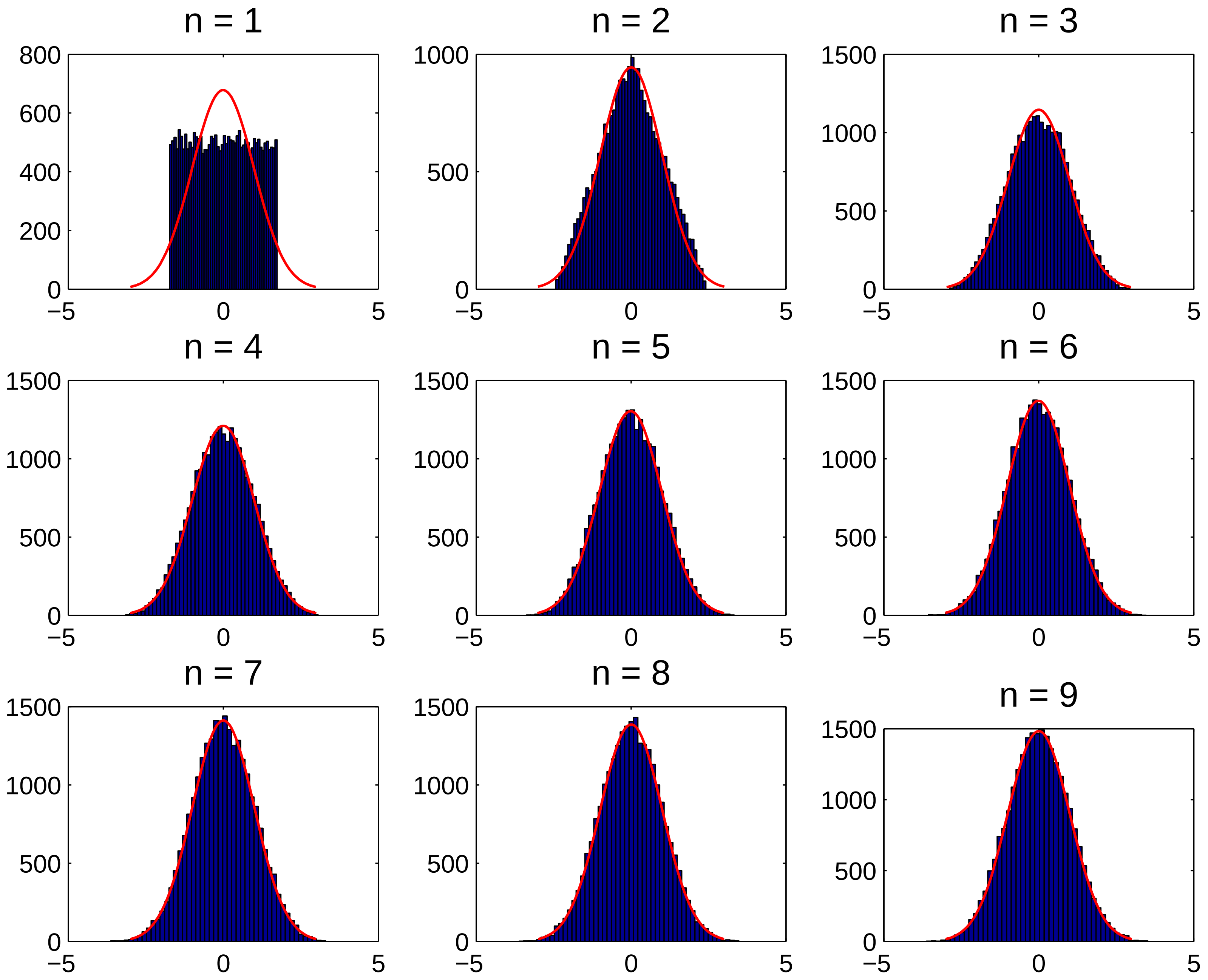
- We generate random samples of different sample sizes n.

- For each n, 25000 samples are generated, and for each sample, quantity

$$\sqrt{n}\frac{\overline{x}_n - \mu}{\sigma} \tag{30}$$

  is calculated (with $\mu = 0.5$ and $\sigma = 1/\sqrt{12}$).

- Then we can plot histograms of (30) for each n and compare with a normal density curve, see next slide.


n = 1
n = 2
n = 3
n = 4
n = 5
n = 6
n = 7
n = 8
n = 9
800
600
400
200
0
1000
500
1500
−5
0
5

- The limiting distribution of $\sqrt{n}(\overline{x}_n - \mu)$ given by the CLT does not depend on the distribution of the x_is.

- The sample size required for the normal approximation to be adequate, however, does.

- In particular, the more this distribution deviates from the normal distribution, the larger the sample has to be.

- This can be illustrated by means of the binomial distribution.

- In particular, define iid Bernoulli random variables x_i, $i = 1, \ldots, n$, such that

$$x_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p. \end{cases} \tag{31}$$

- Then $S_n = \sum_{i=1}^{n} x_i$ has a binomial distribution with

$$\mathsf{E}(S_n) = np, \quad \mathsf{Var}(S_n) = npq, \quad q = 1 - p.$$

- The first CLT was provided by De Moivre in 1738 for the binomial distribution, i.e.,
$$S_n \overset{a}{\sim} \text{Normal}(np, npq). \tag{32}$$

- For $p = 0.5$ (i.e., equal probability of 0 and 1), the probability mass function is symmetric.

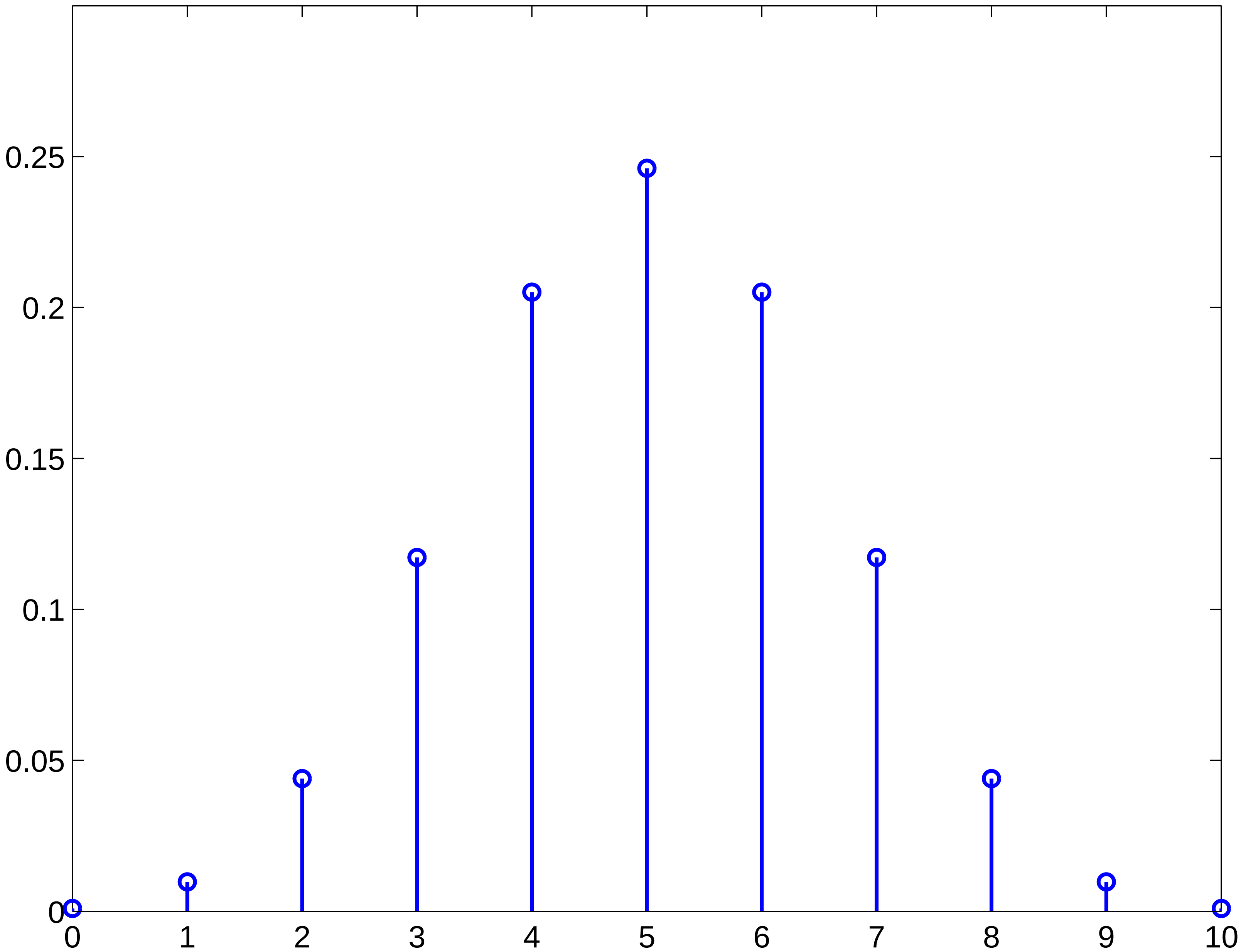
- Thus is is close to the normal distribution along this dimension.

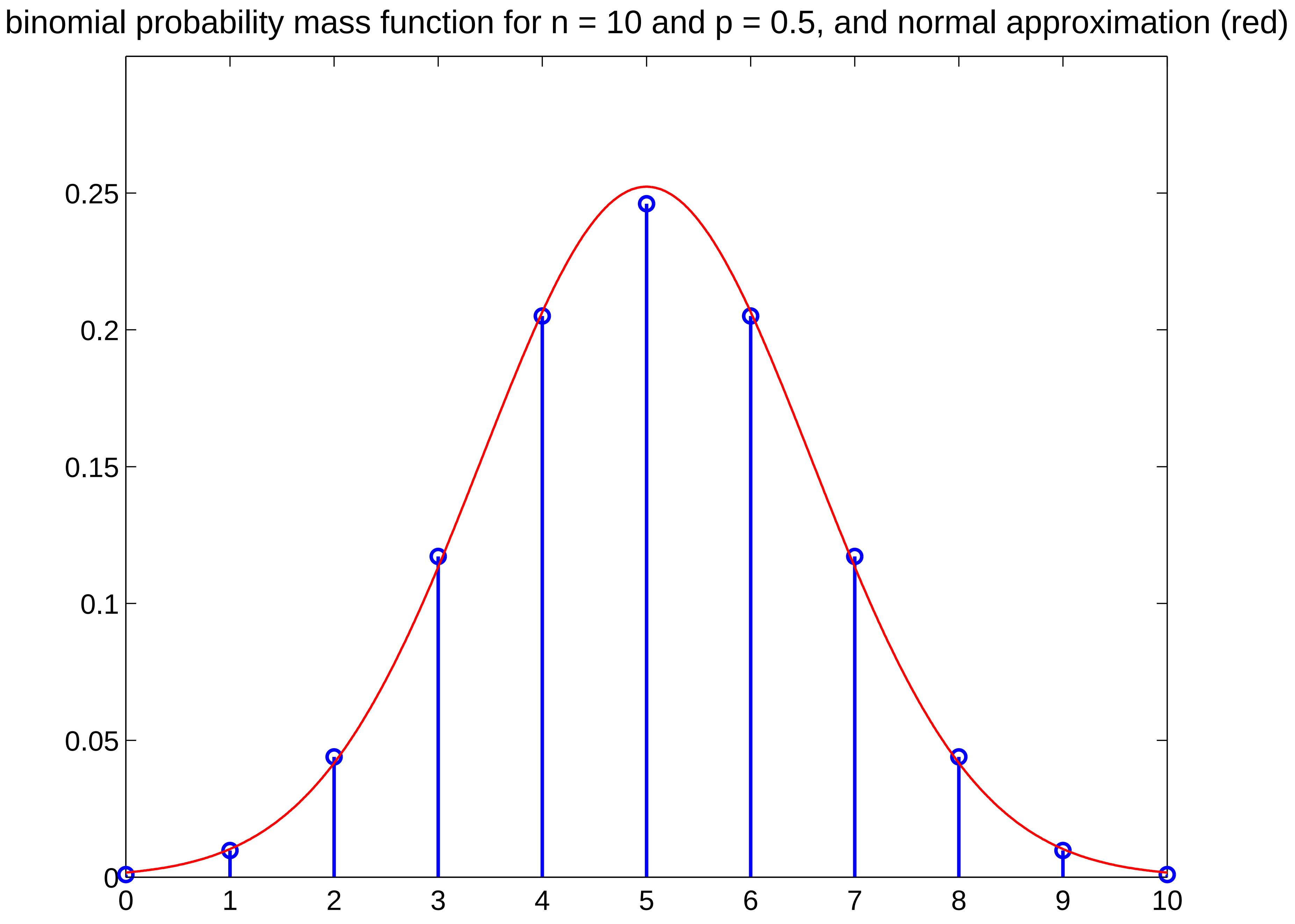
- Then the quality of approximation via the normal density with mean np and variance npq, i.e.,

$$\text{Pr}(S_n = s) = \binom{n}{s} p^s (1-p)^{n-s} \approx \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{(s-np)^2}{2npq}\right\} \tag{33}$$

  appears to be reasonable even for relatively small n.

binomial probability mass function for n = 10 and p = 0.5

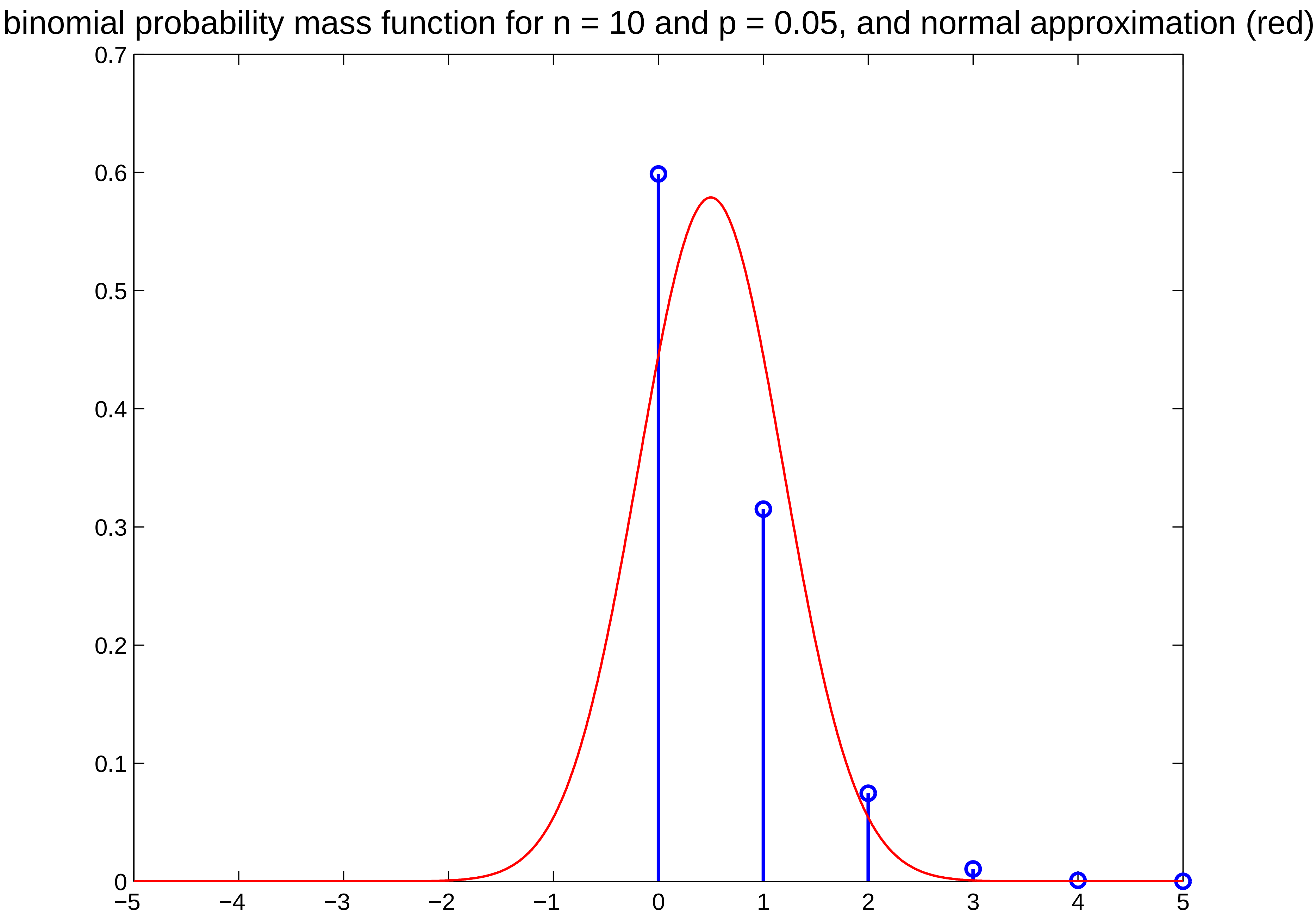



binomial probability mass function for n = 10 and p = 0.5, and normal approximation (red)
0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5
6
7
8
9
10

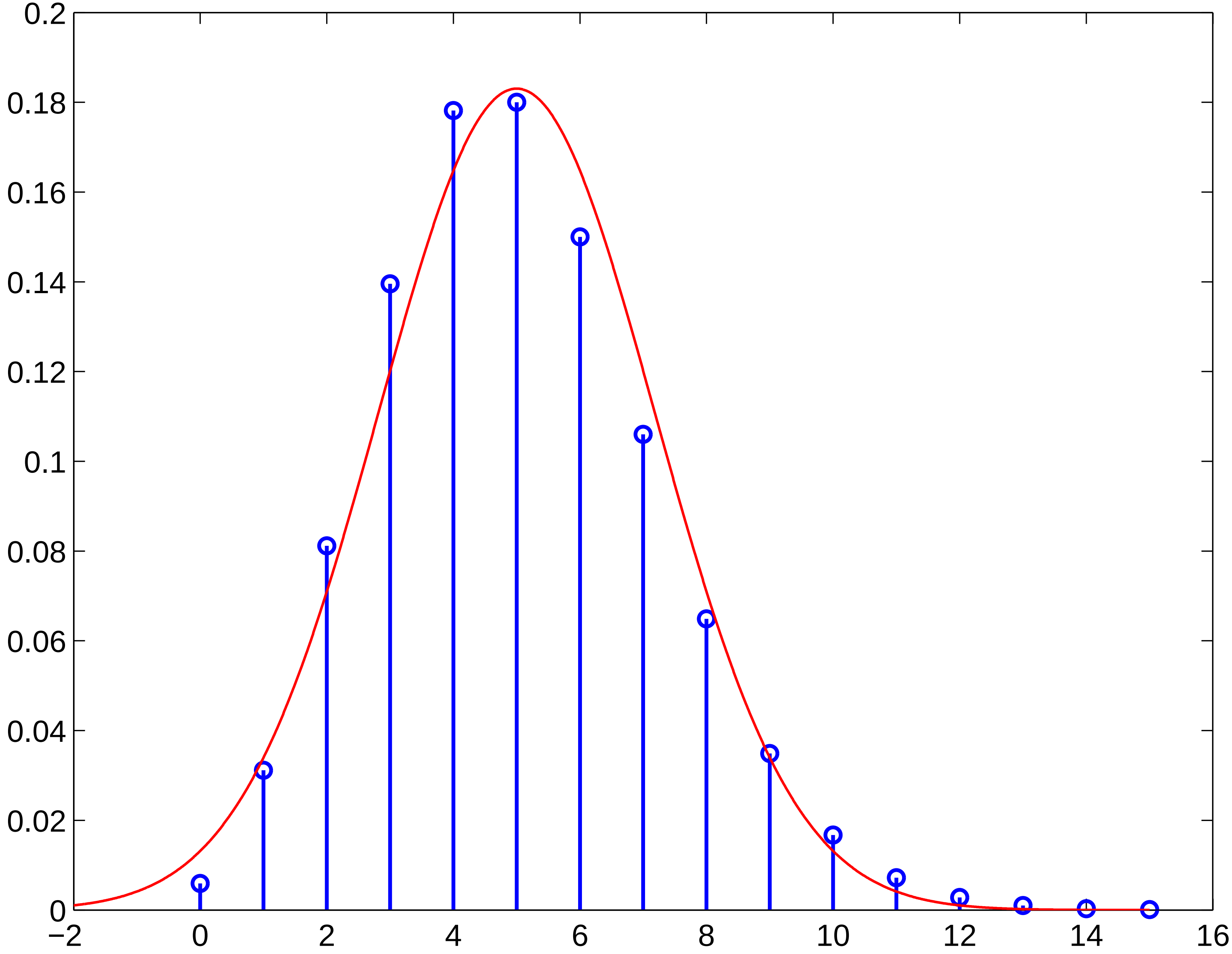
- It's a bit different if p is far from 0.5.

- Then the probability mass function is highly asymmetric for small sample sizes (i.e., very "nonnormal")

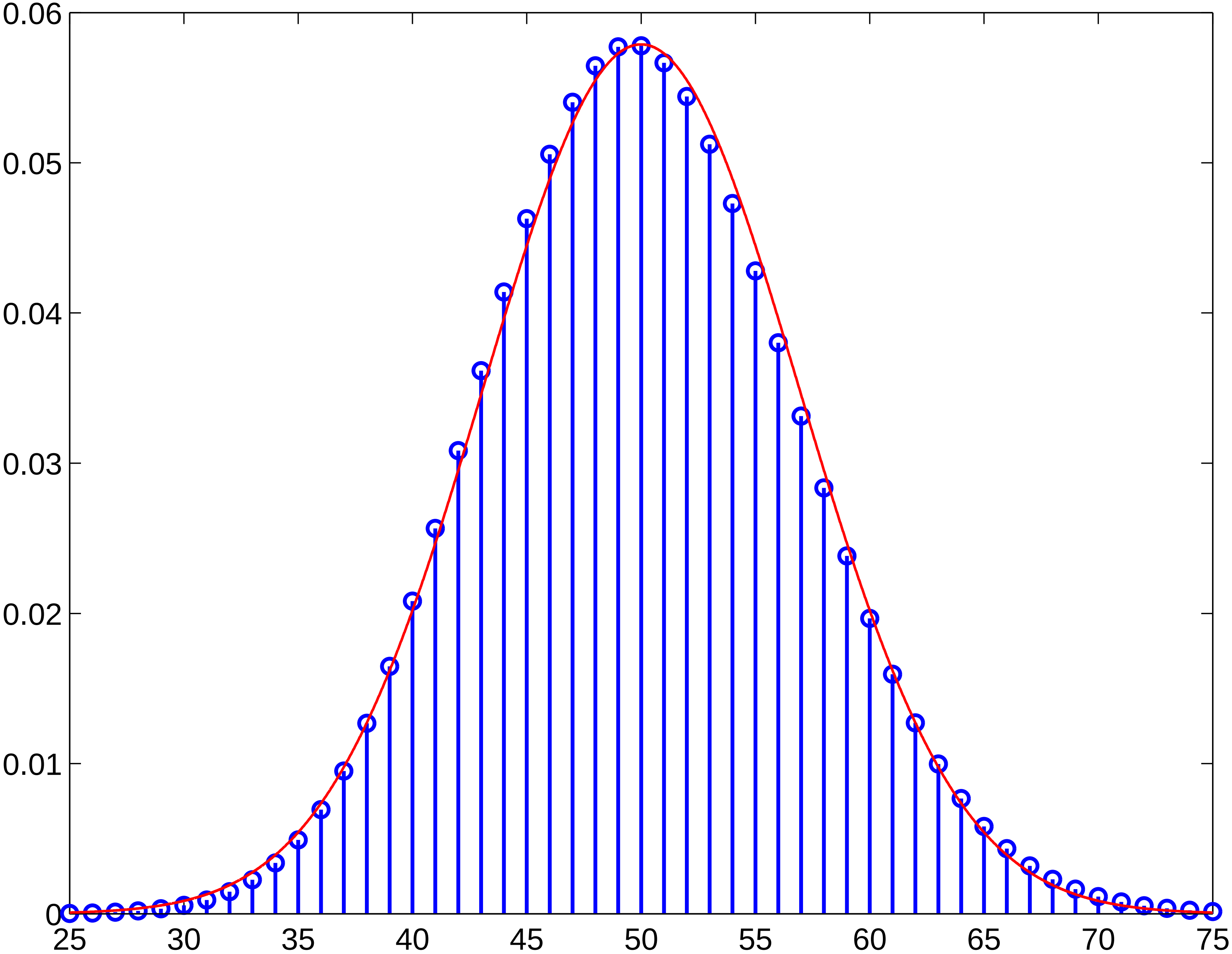
- We need a larger sample size for the approximation to be adequate.


binomial probability mass function for n = 10 and p = 0.05, and normal approximation (red)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−5
−4
−3
−2
−1
0
1
2
3
4
5

binomial probability mass function for n = 100 and p = 0.05, and normal approximation (red)



binomial probability mass function for n = 1000 and p = 0.05, and normal approximation (red



Asymptotic normality of OLS

- Asymptotic normality can also be established for the OLS estimator under different sets of assumptions.

- We first consider the simple linear model.

- Slutzky's Theorem (next slide) is useful in outlining the asymptotic normality of the ordinary least squares estimator.

Slutzky's Theorem

- If $x_n \xrightarrow{d} x$, $a_n \xrightarrow{p} a$, $b_n \xrightarrow{p} b$, then

$$a_n + b_n x_n \xrightarrow{d} a + bx. \tag{34}$$

- If $x_n \xrightarrow{d} x$ and $b_n \xrightarrow{p} 0$, then

$$b_n x_n \xrightarrow{p} 0. \tag{35}$$

- For example, from (34) it follows that if

$$x_n \xrightarrow{d} \mathsf{N}(0, \sigma^2), \quad b_n \xrightarrow{p} b, \quad a_n \xrightarrow{p} a, \tag{36}$$

then

$$a_n + b_n x_n \xrightarrow{d} a + b\mathsf{N}(0, \sigma^2) \stackrel{d}{=} \mathsf{N}(a, b^2\sigma^2), \tag{37}$$

since in general, for constants a and b, $\mathsf{E}(a + bx) = a + b\mathsf{E}(x)$, and $\mathsf{Var}(bx) = b^2\mathsf{Var}(x)$.

Asymptotic normality of OLS

- Write $\widehat{\beta}_1$ as

$$\widehat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_i (x_i - \overline{x})u_i}{\frac{1}{n}\sum_i (x_i - \overline{x})^2}. \tag{38}$$

Hence,

$$\sqrt{n}(\widehat{\beta}_1 - \beta_1) = \frac{\frac{1}{\sqrt{n}}\sum_i (x_i - \overline{x})u_i}{\frac{1}{n}\sum_i (x_i - \overline{x})^2}. \tag{39}$$

- Let $\mu_x = \mathsf{E}(x)$, and write the numerator in (39) as

$$\frac{1}{\sqrt{n}}\sum_i (x_i - \overline{x})u_i = \frac{1}{\sqrt{n}}\sum_i (x_i - \mu_x + \mu_x - \overline{x})u_i \tag{40}$$

$$= \frac{1}{\sqrt{n}}\sum_i (x_i - \mu_x)u_i + (\mu_x - \overline{x})\frac{1}{\sqrt{n}}\sum_i u_i. \tag{41}$$

- In the second term in (41), we have, by the LLN and the CLT, respectively,

$$\mu_x - \overline{x} \xrightarrow{p} 0, \quad \frac{1}{\sqrt{n}} \sum_i u_i \xrightarrow{d} \mathsf{N}(0, \sigma^2), \tag{42}$$

where $\sigma^2 = \mathsf{Var}(u_i)$.

- Thus, by (35),

$$(\mu_x - \overline{x}) \frac{1}{\sqrt{n}} \sum_i u_i \xrightarrow{p} 0. \tag{43}$$

- In the first term in (41), under random sampling and zero correlation between x_i and u_i, variables $(x_i - \mu)u_i$ are independently and identically distributed with

$$\mathsf{E}\{(x_i - \mu_x)u_i\} = \mathsf{Cov}(x_i, u_i) = 0, \quad \mathsf{Var}\{(x_i - \mu_x)u_i\} = \mathsf{E}\{(x_i - \mu_x)^2 u_i^2\}.$$

- Thus,

$$\frac{1}{\sqrt{n}} \sum_i (x_i - \mu_x)u_i \xrightarrow{d} \mathsf{N}(0, \mathsf{E}\{(x_i - \mu_x)^2 u_i^2\}). \tag{44}$$

As before (cf. (17)), for the denominator in (39),

$$\frac{1}{n}\sum_i (x_i - \overline{x})^2 \xrightarrow{p} \mathsf{Var}(x). \tag{45}$$

- Summarizing, (34) then implies

$$\sqrt{n}(\widehat{\beta}_1 - \beta_1) \xrightarrow{d} \mathsf{N}\left(0, \frac{\mathsf{E}\{(x_i - \mu_x)^2 u_i^2\}}{\mathsf{Var}(x)^2}\right). \tag{46}$$

- Result (46) will be useful in our discussion of heteroskedasticity.

- In case of homoskedasticity, i.e.,

$$\mathsf{E}(u_i^2 | x_i) = \sigma^2, \tag{47}$$

it can be shown[3] that

$$\mathsf{E}\{(x_i - \mu_x)^2 u_i^2\} = \sigma^2 \mathsf{E}(x_i - \mu_x)^2 = \sigma^2 \mathsf{Var}(x), \tag{48}$$

[3] Via the law of iterated expectations.

so that the variance in (46) then has the more familiar form[4]

$$\frac{\sigma^2}{\mathsf{Var}(x)}. \tag{49}$$

- However, the more general formula in (39) will be useful later when we discuss heteroskedasticity.

- Under homoskedasticity, we will then use results (46) along with (49) by treating $\widehat{\beta}_1$ as

$$\widehat{\beta}_1 \overset{a}{\sim} \mathsf{N}\left(\beta_1, \frac{\sigma^2}{n\mathsf{Var}(x)}\right), \tag{50}$$

where we estimate $\mathsf{Var}(x)$ via

$$\widehat{\mathsf{Var}}(x) = n^{-1}\sum_i (x_i - \overline{x})^2. \tag{51}$$

[4]Note that this is the asymptotic variance of $\sqrt{n}(\widehat{\beta}_1 - \beta_1)$, so that, due to the factor $\sqrt{n}$, the n in the denominator is missing here.

Asymptotic normality of OLS

- Asymptotic normality can also be established for the OLS estimator in the multiple regression model under different sets of assumptions.

- E.g., it will hold under the Gauss–Markov Assumptions.

- Namely, in this case (now assuming homoskedasticity)

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \text{MVN}(\mathbf{0}, \sigma^2 \boldsymbol{Q}^{-1}), \tag{52}$$

  where

$$\boldsymbol{Q} = \text{plim}_{n\to\infty} \frac{\boldsymbol{X}'\boldsymbol{X}}{n},$$

  so that we act according to

$$\widehat{\boldsymbol{\beta}} \overset{a}{\sim} \text{MVN}\left(\boldsymbol{\beta}, \frac{\sigma^2}{n}\boldsymbol{Q}^{-1}\right) = \text{MVN}\left(\boldsymbol{\beta}, \sigma^2 (n\boldsymbol{Q})^{-1}\right), \tag{53}$$

where we estimate the asymptotic covariance matrix as

$$\widehat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

- We also have

$$t_{\widehat{\beta}_j} = \frac{\widehat{\beta}_j - \beta_j}{\widehat{\sigma}_{\widehat{\beta}_j}} \xrightarrow{d} \text{Normal}(0, 1). \tag{54}$$

- In the present context, (54) and $\widehat{\sigma}_{\widehat{\beta}_j}$ are also referred to as the **asymptotic t statistic** and **asymptotic standard error**, respectively.

- Asymptotic confidence intervals can also be constructed using the quantiles of the normal distribution

- In summary, for large enough sample sizes, nothing changes from what we have done before when it comes to hypothesis testing.

- Note that **homoskedasticity is still required** for the usual t and F statistics to be valid in large samples.

- We will later discuss what to do in case of heteroskedasticity.

- Regarding the asymptotic F test, we note that this can also be carried out by means of an χ^2 distribution with q degrees of freedom (q = number of linear restrictions).

- In particular, as $n \to \infty$,

$$\nu_1 F_{\nu_1,\nu_2} \overset{\nu_2 \to \infty}{\longrightarrow} \chi^2(\nu_1). \tag{55}$$

- For example, take the 95% quantiles of the $F_{\nu_1,\infty}$ distribution with $\nu_1 = 2, 3, 4$, given by 2.9957, 2.6049, and 2.3719, respectively.

- We compute

$$2 \times 2.9957 = 5.9915 \tag{56}$$
$$3 \times 2.6049 = 7.8147 \tag{57}$$
$$4 \times 2.3719 = 9.4877, \tag{58}$$

which are the respective quantiles of the $\chi^2(\nu_1)$ distribution.

- In this form, the test is known as *Wald Test* (named after Abraham Wald).

Lagrange Multiplier Test

- Based on asymptotic arguments, further methods exist that can be used for hypothesis testing.

- A popular example with an intuitively appealing form of the test statistic is the **Lagrange multiplier test**, or **LM test**, for exclusion restrictions.

- As it is based on asymptotic arguments, this requires the Gauß–Markov assumptions, but not Gaussianity of the errors.

- Consider the multiple regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u, \tag{59}$$

- We want to test, in (59), whether

$$H_0 : \beta_{k-q+1} = \beta_{k-q+2} = \cdots = \beta_k = 0, \tag{60}$$

  i.e., we have q exclusion restrictions.

- To do the test, we run the restricted regression, i.e., we get estimates

$$y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \cdots + \tilde{\beta}_{k-q} x_{k-q} + \tilde{u}. \tag{61}$$

- If null hypothesis (60) is correct, variables $x_{k-q+1}, x_{k-q+2}, \ldots, x_k$ should approximately be uncorrelated with the residuals of the restricted model, $\tilde{u}$.

- Thus, we run the regression of $\tilde{u}$ on *all* the independent variables in (59), that is

$$\tilde{u} = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k + v, \tag{62}$$

- Finally, we calculate the usual R^2 from regression (62), since

$$nR^2 \xrightarrow{d} \chi^2(q), \tag{63}$$

  where q is the number of restrictions imposed by (60).

- $\chi^2(q)$ denotes a χ^2 distribution with q degrees of freedom.

LM Test Example

- As an example, we consider the data on the number arrests, where $n = 2725$.

- Suppose we want to test

$$H_0 : \beta_2 = \beta_3 = 0. \tag{64}$$

- We thus estimate the equation

$$narr86 = \beta_0 + \beta_1 \cdot pcnv + \beta_4 \cdot ptime86 + \beta_5 \cdot qemp86 + u, \tag{65}$$

and obtain its residuals, $\tilde{u}$.

- Subsequently, we calculate the regression

$$\begin{aligned}\tilde{u} &= \delta_0 + \delta_1 \cdot pcnv + \delta_2 \cdot avgsen + \delta_3 \cdot tottime \\ &\quad + \delta_4 \cdot ptime86 + \delta_5 \cdot qemp86 + v,\end{aligned}$$

and obtain its R^2.

- In this example, the R^2 turns out to be 0.0015.
- The LM test statistic is thus

$$\text{LM} = nR^2 = 2725 \times 0.0015 = 4.0707.$$

- From Table 1, we can infer that the critical value at the 10% level is 4.6, so we cannot reject H_0 at the 10% level.

- The p–value is

$$P(\chi(2)^2 > \mathsf{LM}) = 1 - F_{\chi^2(2)}(\mathsf{LM}) = 0.1306, \tag{66}$$

where $F_{\chi^2(2)}$ is the cumulative distribution function (cdf) of the χ^2 distribution with two degrees of freedom.

- For purpose of comparison, we shall also consider the F test for (60).

- The F statistic is

$$\begin{aligned} F = \frac{\mathsf{SSR}_r - \mathsf{SSR}_{ur}}{\mathsf{SSR}_{ur}} \times \frac{n-k-1}{q} &= \frac{R^2_{ur} - R^2_r}{1 - R^2_{ur}} \times \frac{n-k-1}{q} \\ &= \frac{0.0428 - 0.0413}{1 - 0.0428} \times \frac{2725}{2} \\ &= 2.0339, \end{aligned}$$

with p–value 0.1310.

Table 1: Quantiles of the χ^2 distribution (ν denotes degrees of freedom)

ν	0.9	0.95	0.975	0.99
1	2.7055	3.8415	5.0239	6.6349
2	4.6052	5.9915	7.3778	9.2103
3	6.2514	7.8147	9.3484	11.3449
4	7.7794	9.4877	11.1433	13.2767
5	9.2364	11.0705	12.8325	15.0863
6	10.6446	12.5916	14.4494	16.8119
7	12.0170	14.0671	16.0128	18.4753
8	13.3616	15.5073	17.5345	20.0902
9	14.6837	16.9190	19.0228	21.6660
10	15.9872	18.3070	20.4832	23.2093
11	17.2750	19.6751	21.9200	24.7250
12	18.5493	21.0261	23.3367	26.2170
13	19.8119	22.3620	24.7356	27.6882
14	21.0641	23.6848	26.1189	29.1412
15	22.3071	24.9958	27.4884	30.5779

Table 2: 95% Quantiles of the F distribution (= 5% critical values) (ν_1 numerator degrees of freedom; ν_2 denominator degrees of freedom)

ν_2/ν_1	2	3	4	5	6	7	8	9	10
10	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782
15	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437
20	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479
25	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365
30	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646
35	3.2674	2.8742	2.6415	2.4851	2.3718	2.2852	2.2167	2.1608	2.1143
40	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772
45	3.2043	2.8115	2.5787	2.4221	2.3083	2.2212	2.1521	2.0958	2.0487
50	3.1826	2.7900	2.5572	2.4004	2.2864	2.1992	2.1299	2.0734	2.0261
60	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926
70	3.1277	2.7355	2.5027	2.3456	2.2312	2.1435	2.0737	2.0166	1.9689
80	3.1108	2.7188	2.4859	2.3287	2.2142	2.1263	2.0564	1.9991	1.9512
90	3.0977	2.7058	2.4729	2.3157	2.2011	2.1131	2.0430	1.9856	1.9376
100	3.0873	2.6955	2.4626	2.3053	2.1906	2.1025	2.0323	1.9748	1.9267
∞	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307